A Appendix

In section 1 we have used a number of elementary results from linear algebra. In particular, this includes the following facts:

- The bipolar set of a closed, convex set in $\mathbb{R}^d$ containing the origin is the set itself.
- A set containing the origin is polyhedral iff its polar is polyhedral.
- The projection of a polyhedral cone is again a polyhedral cone.

For the convenience of the reader we provide proofs and present the underlying theory in a rather self-contained way in this appendix.

Let E be a vector space over the real numbers with finite dimension d and E' its dual. The space E then is isomorphic to $\mathbb{R}^d$ and we will use this fact in some of the discussion below, in which case we will denote the origin by $0 \in \mathbb{R}^d$ and the canonical basis by $\{e_1, \ldots, e_d\}$.

A.1 Polar sets

We start with some basic definitions following [87] and [33]; shorter introductions to the geometry of convex sets can be found in [32] and [37]. For any set $A \subseteq E$, the smallest closed convex set containing A is called the *closed convex hull* of A, i.e. $\overline{\text{conv}}(A)$ is the intersection of all closed convex sets containing A. A closed convex set $C \subseteq E$ is called a *closed convex cone* if $\lambda a \in C$ for every a in C and $\lambda \geqslant 0$. The *closed convex cone generated by a set* $W \subseteq E$ is the closure of the convex cone

$$\text{cone}(W) := \left\{ \sum_{i \in I} \mu_i w_i : \; w_i \in W, \; \mu_i \geqslant 0 \right\},$$

where I is finite. It is the smallest closed convex cone containing W. We define $\text{cone}(\varnothing) := \{0\}$. The following properties of cones can be checked easily:

- Every closed convex cone contains the origin.
- The intersection of two closed convex cones is again a closed convex cone.

For a set $A \subseteq E$ we define the *polar* A° of E as

$$A^\circ := \{y \in E' : \langle x, y \rangle \leqslant 1, \text{ for all } x \in A\}.$$

If A is a cone, we may equivalently define A° as

$$A^\circ = \{y \in E' : \langle x, y\rangle \leqslant 0, \text{ for all } x \in A\}.$$

If A is a linear space, we even may equivalently define A° as the annihilator

$$A^\circ = \{y \in E' : \langle x, y\rangle = 0, \text{ for all } x \in A\}.$$

The *Minkowski sum* of two sets $A, B \subseteq E$ is defined as the set

$$A + B := \{a + b, \ a \in A, \ b \in B\}.$$

It is easy to verify that, for any two sets $A \subseteq B \subseteq E$, we have $A^\circ \supseteq B^\circ$. If $C_1, C_2 \subseteq E$ are cones, then $(C_1 + C_2)^\circ = C_1^\circ \cap C_2^\circ$. Note that the polar of a cone is a closed convex cone.

The following theorem is a version of the celebrated Hahn-Banach theorem. The proof presented here can be found in [82]; for a more general discussion see for example [76].

Proposition A.1 (Bipolar Theorem)**.** *For a set $A \subseteq E$ the* bipolar $A^{\circ\circ} = (A^\circ)^\circ$ *equals the closed convex hull of $A \cup \{0\}$.*

Proof: Let $B = \text{conv}(A \cup \{0\})$. Since $B \supseteq A$ we have $B^\circ \subseteq A^\circ$.

On the other hand, let $y \in A^\circ$ and $M \in \mathfrak{N}$ and pick $\lambda_i \in [0, 1]$, for $1 \leqslant i \leqslant M$, such that $\sum_{i=1}^M \lambda_i = 1$. Then we have, for any $a_i \in A \cup \{0\}$:

$$1 \geqslant \sum_{i=1}^M \lambda_i \langle y, a_i\rangle = \sum_{i=1}^M \langle y, \lambda_i a_i\rangle = \langle \sum_{i=1}^M \lambda_i a_i, y\rangle.$$

Every $x \in B$ can be written as $x = \sum_{i=1}^M \lambda_i a_i$. It follows that $B^\circ \supseteq A^\circ$ and hence $A^\circ = B^\circ$.

We will now prove that $B^{\circ\circ} = \overline{B}$ which finishes the proof. Let $x \in \overline{B}$. For any $y \in B^\circ$ we have $\langle x, y\rangle \leqslant 1$ by definition and continuity, from which it follows that $x \in B^{\circ\circ}$ and therefore $\overline{B} \subseteq B^{\circ\circ}$. Conversely, assume $x_1 \notin \overline{B}$. Then there exists an $y \in E'$ and a constant c such that $\langle x, y\rangle \leqslant c$, for $x \in B$, and $\langle x_1, y\rangle > c$ (this follows from the Hahn-Banach theorem in its version as separating hyperplanes theorem, see for example [82]).

Because $0 \in B$ we have $c \geqslant 0$. We can even assume $c > 0$. It follows that $\langle x, y/c\rangle \leqslant 1$, for $x \in B$, and thus $y/c \in B^\circ$. But from $\langle x_1, y/c\rangle > 1$ we see that $x_1 \notin B^{\circ\circ}$. ∎

Corollary A.2. *If $C \subseteq E$ is a closed convex cone then $C^{\circ\circ} = C$.*

A.2 Polyhedral sets

We will now introduce the concept of *polyhedral sets*, which can be defined in two distinct ways. The first definition builds a polyhedron "from inside": Let V and W be two finite sets in E. The Minkowski sum of conv(V) and the cone generated by W

$$P = \operatorname{conv}(V) + \operatorname{cone}(W)$$

is called a *V-polyhedron*, where the name comes from the fact that such a polyhedron is defined using its vertices. Note that P is *closed*.

Polyhedral sets can also be built "from outside". A set $P \subseteq E$ is called an *H-polyhedron*, if it can be expressed as the finite intersection of closed halfspaces, that is

$$P = \bigcap_{i=1}^{N} \{x \in E : \langle x, y_i \rangle \leqslant c_i\},$$

for some elements $y_i \in E'$, and some constants c_i, $i \in \{1, \dots N\}$. As a subset of $\mathbb{R}^d$ such a polyhedron can be written as

$$P = P(A, z) := \left\{x \in \mathbb{R}^d : Ax \leqslant z\right\} \qquad \text{for some } A \in \mathbb{R}^{N \times d}, z \in \mathbb{R}^N.$$

Note that an H-polyhedron with all $c_i = 0$, i.e. of the form $P(A, 0)$, is in fact a closed convex cone: we shall encounter such *polyhedral cones* quite often.

These two distinct characterizations for polyhedral sets are useful for calculations and will play an important part in the following discussion. As we will verify below, the notions of V- and H-polyhedral sets are equivalent.

Our first Lemma deals with the projection of H-cones. The proof and a more thorough discussion can be found in [87].

Proposition A.3. *A projection of an H-cone along any coordinate directions e_k, $1 \leqslant k \leqslant d$, is again an H-cone. More specifically, if C is an H-cone in $\mathbb{R}^d$, then so is its elimination cone $elim_k(C) := \{x + \mu e_k : x \in C, \mu \in \mathbb{R}\}$ and its projection cone $proj_k(C) := elim_k(C) \cap \{x \in \mathbb{R}^d : \langle x, e_k \rangle = 0\}$.*

Proof: Note that it suffices to show that the set $\operatorname{elim}_k(C)$ is an H-cone, for any k, because the projection cone is the intersection of the elimination cone with the two halfspaces $\{x \in \mathbb{R}^d : \langle x, e_k \rangle \leqslant 0\}$ and $\{x \in \mathbb{R}^d : \langle x, -e_k \rangle \leqslant 0\}$.

Suppose that $C = P(A, 0)$ and $a_1, a_2, \dots, a_N$ are the row vectors of A. We will construct a new matrix A^k such that $\operatorname{elim}_k(C) = P(A^k, 0)$.
Claim: $A^k = \{a_i : a_{ik} = 0\} \cup \{a_{ik}a_j - a_{jk}a_i : a_{ik} > 0, a_{jk} < 0\}$
If $x \in C$ then $Ax \leqslant 0$. But then we also have $A^k x \leqslant 0$, because A^k consists of nonnegative linear combinations of rows of A. Therefore $C \subseteq P(A^k, 0)$. As

the k^{th} component of A^k is zero by construction, we even have $\text{elim}_k(C) \subseteq P(A^k, 0)$.

On the other hand, let $x \in P(A^k, 0)$. We want to show that there is a $\mu \in \mathbb{R}$ such that $x - \mu e_k \in C$, i.e. $A(x - \mu e_k) \leqslant 0$. Writing these equations out, we obtain the inequalities $a_j x - a_{jk}\mu \leqslant 0$, or

$$\begin{aligned} \mu &\geqslant \tfrac{a_i x}{a_{ik}}, \quad \text{if } a_{ik} > 0, \\ \mu &\leqslant \tfrac{a_j x}{a_{jk}}, \quad \text{if } a_{jk} < 0. \end{aligned}$$

Such a μ exists, because if $a_{ik} > 0$ and $a_{jk} < 0$, then $(a_{ik}a_j - a_{jk}a_i)x \leqslant 0$, since $x \in P(A^k, 0)$, which can be written as

$$\frac{a_i x}{a_{ik}} \leqslant \frac{a_j x}{a_{jk}}.$$

It follows that $P(A^k, 0) \subseteq \text{elim}_k(C)$, finishing the proof. ■

Proposition A.4. *Every V-polyhedron is an H-polyhedron and vice versa.*

We split the proposition into two claims for the two directions, which we prove independently.

Claim: Every V-polyhedron is an H-polyhedron.

Remark: Proving the claim directly turns out to be rather tedious, due to the difficulty of manipulating the necessary sets. There is, however, an elegant proof using *homogenization*: Every polyhedron in d-dimensional space can be regarded as a polyhedral cone in dimension $d + 1$. The equivalence between V-cones and H-cones is easier to show. The direct proof uses Fourier-Motzkin elimination to calculate the sets explicitly. It can be found, together with the indirect proof given here, in [87].

Proof: By mapping a point $x \in \mathbb{R}^d$ to $\left(\begin{smallmatrix} 1 \\ x \end{smallmatrix}\right) \in \mathbb{R}^{d+1}$ we associate with every polyhedral set P in $\mathbb{R}^d$ a cone in $\mathbb{R}^{d+1}$ in the following way: If $P = P(A, z)$ is a H-polyhedral set, define

$$C(P) := P\left(\left(\begin{smallmatrix} -1 & 0 \\ -z & A \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right)\right).$$

Conversely, if $C \in \mathbb{R}^{d+1}$ is an arbitrary H-cone, then $\{x \in \mathbb{R}^d : \left(\begin{smallmatrix} 1 \\ x \end{smallmatrix}\right) \in C\}$ is a (possibly empty) H-polyhedral set.

On the other hand, if $P = \text{conv}(V) + \text{cone}(W)$ is a V-polyhedral set for some finite sets V and W, we define

$$C(P) := \text{cone}\left(\begin{smallmatrix} 1 & 0 \\ V & W \end{smallmatrix}\right),$$

that is, we add a zeroth coordinate to the vectors in V and W before generating the cone, namely 1 and 0, respectively. As before, a straightforward calculation shows that if C is a V-cone in $\mathbb{R}^{d+1}$, then $\{x \in \mathbb{R}^d : \binom{1}{x} \in C\}$ is a V-polyhedral set in $\mathbb{R}^d$.

If we can now show that every V-cone is an H-cone we are done, since every V-polyhedral set in $\mathbb{R}^d$ can be identified with a V-cone in $\mathbb{R}^{d+1}$ and the H-cone then translated back to the H-polyhedral set. Consider thus a V-cone, which can be written as

$$C = \left\{ x \in \mathbb{R}^d : \exists \lambda_i \geqslant 0 : x = \sum_i \lambda_i w_i,\ w_i \in W \right\},$$

or equivalently as

$$C = \left\{ (x, \lambda) \in \mathbb{R}^{d+n} : \lambda_i \geqslant 0, x = \sum_i \lambda_i w_i,\ w_i \in W \right\},$$

the latter set being an H-cone in $\mathbb{R}^{d+n}$. By successsively projecting the cone onto the hyperplanes for which the k^{th} coordinate equals zero, for $d < k \leqslant d + n$, we obtain a cone in $\mathbb{R}^d$ since we already showed that such a projection of an H-cone is again an H-cone. This finishes the proof of the claim.

The second part of the equivalence can also be shown directly or via homogenization, but we will give a third proof, which makes use of an elegant induction argument. For a thorough discussion of these concepts (and the proof of the following claim) see also [33].

Claim: Every H-polyhedron is a V-polyhedron.

Proof: Let P be an intersection of finitely many closed halfspaces in $\mathbb{R}^d$. We may assume w.l.g. that the dimension of P is d and will prove the claim by induction on d. If $d = 1$, then P is a halfline or a closed interval and the claim is clear. For $d \geqslant 2$ we will show that every point in P can be represented as the convex combination $a = (1 - \lambda)b + \lambda c,\ 0 \leqslant \lambda \leqslant 1$, where b and c belong to two distinct facets F and G of P respectively, i.e.

$$F = \text{conv}(V_F) + \text{cone}(W_F) \quad \text{and} \quad G = \text{conv}(V_G) + \text{cone}(W_G),$$

for some finite sets V_F, W_F, V_G, W_G. This suffices to prove the claim since the Minkowski sum of two V-polyhedral sets is again a V-polyhedral set.

Since every facet has dimension $d - 1$, we know that the boundaries of P are polyhedral sets. Let a be any point in the interior of P. Then there is some line l through a that intersects two facets of P, which is not parallel to any of the generating hyperplanes and intersects them in distinct points. Since a must lie between two such intersection points it is the linear combination of finitely many elements of V-polyhedral sets and because a was an arbitrary point in the interior of P, it follows that P itself is V-polyhedral. ∎

The next proof can also be found in [33], along with other constructive results regarding polyhedra.

Proposition A.5. *Let $A \subseteq E$ be a polyhedral set. Then its polar A° also is a polyhedral set.*

Proof: We show that the polar of a V-polyhedron is an H-polyhedron, which we calculate explicitly. Let therefore A be of the form

$$A = \mathrm{conv}(V) + \mathrm{cone}(W) = \mathrm{conv}(\{v_1, \ldots, v_N\}) + \mathrm{cone}(\{w_1, \ldots, w_K\}),$$

for some finite sets V and W. By definition, we have

$$\begin{aligned} A^\circ &= \left\{ y \in E' : \left\langle \sum_{i=1}^N \lambda_i v_i + \sum_{j=1}^K \mu_j w_j, y \right\rangle \leqslant 1, \lambda_i \geqslant 0, \mu_j \geqslant 0, \sum \lambda_i = 1 \right\} \\ &= \left\{ y \in E' : \sum_i \lambda_i \langle v_i, y\rangle + \sum_j \mu_j \langle w_j, y\rangle \leqslant 1, \lambda_i \geqslant 0, \mu_j \geqslant 0, \sum \lambda_i = 1 \right\}. \end{aligned}$$

We therefore find that

$$A^\circ = \bigcap_{i=1}^N \{y \in E' : \langle v_i, y\rangle \leqslant 1\} \cap \bigcap_{j=1}^K \{y \in E' : \langle w_j, y\rangle \leqslant 0\},$$

which is an H-polyhedron. ∎

Corollary A.6. *A convex, closed set containing the origin is polyhedral iff its polar is so too.*

Proof: This follows immediately from the previous proposition and the bipolar theorem, since then A° is polyhedral and $A = A^{\circ\circ}$. ∎

B The Legendre Transformation

Definition B.1. *Let $u : \mathbb{R} \to \mathbb{R} \cup \{-\infty\}$ be a concave upper semi-continuous function and $D = \mathrm{int}\{u > -\infty\} \neq \varnothing$ its domain, which we assume to be non-empty. The* conjugate *v of u is the function*

$$v(y) := \sup\{u(x) - xy,\ x \in \mathbb{R}\}.$$

The function v is the Legendre transform of $-u(-x)$ and is therefore convex rather than concave.[4] Given the conjugate function v, the original function u can be recovered via the transformation

$$u(x) := \inf\{v(y) + xy,\ y \in \mathbb{R}\}.$$

From these definitions it is immediately clear that for every $(x, y) \in \mathbb{R}^2$ we have *Fenchel's inequality*:

$$u(x) - v(y) \leqslant xy. \tag{251}$$

Note that equality holds when the supremum (respectively the infimum) in the above definitions is attained for the corresponding values of x and y.

Definition B.2. *The* subdifferential $\partial v(y_0)$ *of a convex function* v *at* y_0 *is the set of* $x \in \mathbb{R}$ *such that*

$$v(y) \geqslant v(y_0) + x \cdot (y - y_0), \quad \textit{for all } y \in \mathbb{R}.$$

For a concave function u *we define the* superdifferential $\partial u(x_0)$ *of at* x_0 *equivalently as the set of* $y \in \mathbb{R}$ *satisfying*

$$u(x) \leqslant u(x_0) + y \cdot (x - x_0), \quad \textit{for all } x \in \mathbb{R}.$$

If $\partial u(x_0)$ *consists of one single element* y*, then* u *is* differentiable at x_0 *and* $\nabla u(x_0) = y$*. Equivalently if* $\partial v(y_0)$ *consists of one single element* x*, then* v *is* differentiable at y_0 *and* $\nabla v(y_0) = x$*.*

Our first duality result links the super- and subdifferential of the conjugate functions u and v:

Proposition B.3. *The superdifferential* $\partial u(x_0)$ *contains* y_0 *iff* $-x_0 \in \partial v(y_0)$*.*

Proof: Let y_0 be in $\partial u(x_0)$. Then we have, for every x,

$$\begin{aligned} u(x) &\leqslant u(x_0) + y_0(x - x_0) \\ u(x) - y_0 x &\leqslant u(x_0) - y_0 x_0. \end{aligned}$$

Since this also holds for the supremum and using Fenchel's inequality on the right hand side, we obtain for every y in $\mathbb{R}$

$$\begin{aligned} v(y_0) &\leqslant u(x_0) - x_0 y_0 \leqslant v(y) + x_0 y - x_0 y_0 \\ v(y_0) &\leqslant v(y) + x_0(y - y_0), \end{aligned}$$

[4] In fact, the classical duality theory considers the (convex) function $-u(-x)$ to obtain a perfectly symmetric setting.

which is exactly the requirement for $-x_0$ to be in the subdifferential $\partial v(y_0)$. The other direction can be proved analogously. ■

There is an important duality regarding the smoothness and the strict concavity of the dual functions u and v. The following proof can be found in [45].

Proposition B.4. *The following are equivalent:*

1. $u : D \to \mathbb{R}$ *is strictly concave.*

2. v *is differentiable on the interior of its domain.*

Proof: $(i) \Rightarrow (ii)$. Suppose that (ii) fails, i.e. there is some y such that $\partial v(y)$ contains two distinct points, and call them $-x_1$ and $-x_2$. We may suppose that $x_1 < x_2$. This is equivalent to the requirement that $y \in \partial u(x_1) \cap \partial u(x_2)$. For $i = 1, 2$ we have

$$u(x_i) - v(y) = x_i y$$

and using Fenchel's inequality, we get (for $0 \leqslant \lambda \leqslant 1$):

$$\lambda u(x_1) + (1 - \lambda)u(x_2) - v(y) = y \cdot (\lambda x_1 + (1 - \lambda)x_2) \tag{252}$$

$$\geqslant u(\lambda x_1 + (1 - \lambda)x_2) - v(y). \tag{253}$$

But this implies that u is affine on $[x_1, x_2]$, a contradiction since u is strictly concave. Therefore $\partial v(y)$ must be single-valued for all $y \in \text{int dom } (v)$, i.e. v is continuously differentiable.

$(ii) \Rightarrow (i)$. Suppose that there are two distinct points x_1 and x_2 such that u is affine on the line segment $[x_1, x_2]$. If we set $x := \frac{1}{2}(x_1 + x_2)$, there is an y such that $\nabla v(y) = x$, i.e. $y \in \partial u(x)$. We can write

$$0 = u(x) - v(y) - xy = \frac{1}{2} \sum_{i=1}^{2} \left[u(x_i) - v(y) - y x_i \right],$$

which implies (using Fenchel's inequality), that each of the terms in the bracket on the right hand side must vanish. We therefore have $y \in \partial u(x_1) \cap \partial u(x_2)$, i.e. $\partial v(y)$ contains more than one point, which contradicts the assumption that v is differentiable. ■