1 Convex Sets, and Convex Functions

In this section, we introduce one of the most important ideas in the theory of optimization, that of a convex set. We discuss other ideas which stem from the basic definition, and in particular, the notion of a convex function which will be important, for example, in describing appropriate constraint sets.

1.1 Convex Sets

Intuitively, if we think of $\mathbb{R}^2$ or $\mathbb{R}^3$, a convex set of vectors is a set that contains all the points of any line segment joining two points of the set (see the next figure).

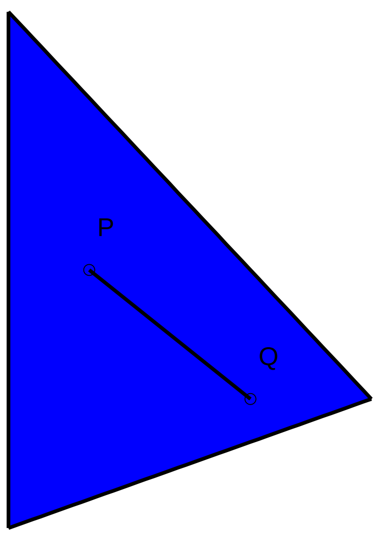

Figure 1: A Convex Set

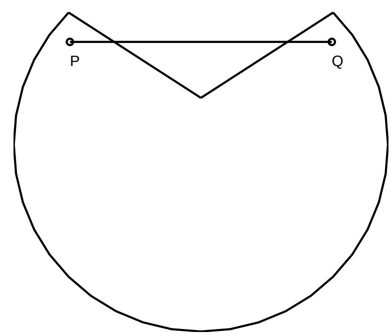

Figure 2: A Non-convex Set

To be more precise, we introduce some definitions. Here, and in the following, V will always stand for a real vector space.

Definition 1.1 *Let $u, v \in V$. Then the set of all* **convex combinations** *of u and v is the set of points*

$$\{w_\lambda \in V \ : \ w_\lambda = (1-\lambda)u + \lambda v, 0 \leq \lambda \leq 1\}. \tag{1.1}$$

In, say, $\mathbb{R}^2$, this set is exactly the line segment joining the two points u and v. (See the examples below.)
Next, is the notion of a **convex set**.

Definition 1.2 *Let $K \subset V$. Then the set K is said to be* **convex** *provided that given two points $u, v \in K$ the set (1.1) is a subset of K.*

We give some simple examples:

Examples 1.3

(a) An interval of $[a, b] \subset \mathbb{R}$ is a convex set. To see this, let $c, d \in [a, b]$ and assume, without loss of generality, that $c < d$. Let $\lambda \in (0, 1)$. Then,

$$\begin{aligned} a &\leq c = (1-\lambda)c + \lambda c < (1-\lambda)c + \lambda d \\ &< (1-\lambda)d + \lambda d = d \\ &\leq b. \end{aligned}$$

(b) A disk with center $(0, 0)$ and radius c is a convex subset of $\mathbb{R}^2$. This may be easily checked by using the usual distance formula in $\mathbb{R}^2$ namely $\|x-y\| := \sqrt{(x_1-y_1)^2 + (x_2-y_2)^2}$ and the triangle inequality $\|u + v\| \leq \|u\| + \|v\|$. (Exercise!)

(c) In $\mathbb{R}^n$ the set $H := \{x \in \mathbb{R}^n \ : \ a_1 x_1 + \ldots + a_n x_n = c\}$ is a convex set. For any particular choice of constants a_i it is a hyperplane in $\mathbb{R}^n$. Its defining equation is a generalization of the usual equation of a plane in $\mathbb{R}^3$, namely the equation $ax + by + cz + d = 0$.

To see that H is a convex set, let $x^{(1)}, x^{(2)} \in H$ and define $z \in \mathbb{R}^3$ by $z := (1 - \lambda)x^{(1)} + \lambda x^{(2)}$. Then

$$\begin{aligned} z &= \sum_{i=1}^{n} a_i[(1-\lambda)x_i^{(1)} + \lambda x_i^{(2)}] = \sum_{i=1}^{n} (1-\lambda)a_i x_i^{(1)} + \lambda a_i x_i^{(2)} \\ &= (1-\lambda)\sum_{i=1}^{n} a_i x_i^{(1)} + \lambda \sum_{i=1}^{n} a_i x_i^{(2)} = (1-\lambda)c + \lambda c \\ &= c. \end{aligned}$$

Hence $z \in H$.

(d) As a generalization of the preceeding example, let A be an $m \times n$ matrix, $b \in \mathbb{R}^m$, and let $S := \{(x \in \mathbb{R}^n \,:\, Ax = b\}$. (The set S is just the set of all solutions of the linear equation $Ax = b$.) Then the set S is a convex subset of $\mathbb{R}^n$. Indeed, let $x^{(1)}, x^{(2)} \in S$. Then

$$A\left((1-\lambda)x^{(1)} + \lambda x^{(2)}\right) \;=\; (1-\lambda)A\left(x^{(1)}\right) + \lambda A\left(x^{(2)}\right) \;=\; (1-\lambda)b + \lambda b = b.$$

Note: In (c) above, we can take $A = (a_1, a_2, \ldots, a_n)$ so that with this choice, the present example is certainly a generalization.

We start by checking some simple properties of convex sets.
A first remark is that $\mathbb{R}^n$ is, itself, obviously convex. Moreover, the unit ball in $\mathbb{R}^n$, namely $B_1 := \{x \in \mathbb{R}^n \,|\, \|x\| \leq 1\}$ is likewise convex. This fact follows immediately from the triangle inequality for the norm. Specifically, we have, for arbitrary $x, y \in B_1$ and $\lambda \in [0,1]$

$$\|(1-\lambda)\, x + \lambda\, y\| \leq (1-\lambda)\, \|x\| + \lambda\, \|y\| \leq (1-\lambda) + \lambda \;=\; 1\,.$$

The ball B_1 is a closed set. It is easy to see that, if we take its interior

$$\overset{\circ}{B} := \{x \in \mathbb{R}^n \,|\, \|x\| < 1\}\,,$$

then this set is also convex. This gives us a hint regarding our next result.

Proposition 1.4 *If $C \subset \mathbb{R}^n$ is convex, the $c\ell(C)$, the closure of C, is also convex.*

Proof: Suppose $x, y \in c\ell(C)$. Then there exist sequences $\{x_n\}_{n=1}^\infty$ and $\{y_n\}_{n=1}^\infty$ in C such that $x_n \to x$ and $y_n \to y$ as $n \to \infty$. For some $\lambda, 0 \leq \lambda \leq 1$, define $z_n := (1-\lambda)\, x_n + \lambda\, y_n$. Then, by convexity of C, $z_n \in C$. Moreover $z_n \to (1-\lambda)\, x + \lambda\, y$ as $n \to \infty$. Hence this latter point lies in $c\ell(C)$. $\square$

The simple example of the two intervals $[0,1]$ and $[2,3]$ on the real line shows that the union of two sets is not necessarily convex. On the other hand, we have the result:

Proposition 1.5 *The intersection of any number of convex sets is convex.*

Proof: Let $\{K_\alpha\}_{\alpha \in A}$ be a family of convex sets, and let $\mathcal{K} := \cap_{\alpha \in A} K_\alpha$. Then, for any $x, y \in \mathcal{K}$ by definition of the intersection of a family of sets, $x, y \in K_\alpha$ for all $\alpha \in A$ and each of these sets is convex. Hence for any $\alpha \in A$, and $\lambda \in [0,1]$, $(1-\lambda)x + \lambda y \in K_\alpha$. Hence $(1-\lambda)x + \lambda y \in \mathcal{K}$. $\square$

Relative to the vector space operations, we have the following result:

Proposition 1.6 *Let C, C_1, and C_2 be convex sets in $\mathbb{R}^n$ and let $\beta \in \mathbb{R}$ then*

(a) $\beta\, C := \{z \in \mathbb{R}^n \,|\, z = \beta x, x \in C\}$ *is convex.*

(b) $C_1 + C_2 := \{z \in \mathbb{R}^n \,|\, z = x_1 + x_2, x_1 \in C_1, x_2 \in C_2\}$ *is convex.*

Proof: We leave part (a) to the reader. To check that part (b) is true, let $z_1, z_2 \in C_1 + C_2$ and take $0 \le \lambda \le 1$. We take $z_1 = x_1 + x_2$ with $x_1 \in C_1, x_2 \in C_2$ and likewise decompose $z_2 = y_1 + y_2$. Then

$$\begin{aligned} (1-\lambda)\, z_1 + \lambda\, z_2 &= (1-\lambda)\,[x_1 + x_2] + \lambda\,[y_1 + y_2] \\ &= [(1-\lambda)\,x_1 + \lambda\, y_1] + [(1-\lambda)\,x_2 + \lambda\, y_2] \in C_1 + C_2\,, \end{aligned}$$

since the sets C_1 and C_2 are convex. $\square$

We recall that, if A and B are two non-empty sets, then the Cartesian product of these two sets $A \times B$ is defined as the set of ordered pairs $\{(a,b) \,:\, a \in A, b \in B\}$. Notice that order *does* matter here and that $A \times B \neq B \times A$! Simple examples are

1. Let $A = [-1,1], B = [-1,1]$ so that $A \times B = \{(x,y) \,:\, -1 \le x \le 1, -1 \le y \le 1\}$ which is just the square centered at the origin, of side two.

2. $\mathbb{R}^2$ itself can be identified (and we usually do!) with the Cartesian product $\mathbb{R} \times \mathbb{R}$.

3. let $C \subset \mathbb{R}^2$ be convex and let $S := \mathbb{R}^+ \times C$. Then S is called a right cylinder and is just $\{(z,x) \in \mathbb{R}^3 \,:\, z > 0, x \in C\}$. If, in particular $C = \{(u,v) \in \mathbb{R}^2 \,:\, u^2 + v^2 \le 1\}$, then S is the usual right circulinder lying above the x, y-plane (without the bottom!).

This last example shows us a situation where $A \times B$ is convex. In fact it it a general result that if A and B are two non-empty convex sets in a vector space V, then $A \times B$ is likewise a convex set in $V \times V$.

Exercise 1.7 *Prove this last statement.*

While, by definition, a set is convex provided all convex combinations of two points in the set is again in the set, it is a simple matter to check that we can extend this statement to include convex combinations of more than two points. Notice the way in which the proof is constructed; it is often very useful in computations!

Proposition 1.8 *Let K be a convex set and let $\lambda_1, \lambda_2, \ldots, \lambda_p \geq 0$ and $\sum_{i=1}^{p} \lambda_i = 1$. If $x_1, x_2, \ldots x_p \in K$ then*

$$\sum_{i=1}^{p} \lambda_i \, x_i \in K. \tag{1.2}$$

Proof: We prove the result by induction. Since K is convex, the result is true, trivially, for $p = 1$ and by definition for $p = 2$. Suppose that the proposition is true for $p = r$ (induction hypothesis!) and consider the convex combination $\lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_{r+1} x_{r+1}$. Define $\Lambda := \sum_{i=1}^{r} \lambda_i$. Then since $1 - \Lambda = \sum_{i=1}^{r+1} \lambda_i - \sum_{i=1}^{r} \lambda_i = \lambda_{r+1}$, we have

$$\left(\sum_{i=1}^{r} \lambda_i \, x_i \right) + \lambda_{r+1} \, x_{r+1} \; = \; \Lambda \left(\sum_{i=1}^{r} \frac{\lambda_i}{\Lambda} x_i \right) + (1 - \Lambda) \, x_{r+1}.$$

Note that $\sum_{i=1}^{r} \left(\dfrac{\lambda_i}{\Lambda} \right) = 1$ and so, by the induction hypothesis, $\sum_{i=1}^{r} \left(\dfrac{\lambda_i}{\Lambda} \right) \, x_i \in K$. Since $x_{r+1} \in K$ it follows that the right hand side is a convex combination of two points of K and hence lies in K $\square$

Remark: We will also refer to combinations of the form (1.2) as **convex combinations** of the p points $x_1, x_2, \ldots, x_p$.

For any given set which is *not* convex, we often want to find a set which is convex and which contains the set. Since the entire vector space V is obviously a convex set, there is always at least one such convex set containing the given one. In fact, there are infinitely many such sets. We can make a more economical choice if we recall that the intersection of any number of convex sets is convex.
Intuitively, given a set $C \subset V$, the intersection of all convex sets containing C is the "smallest" subset containing C. We make this into a definition.

Definition 1.9 *The* **convex hull** *of a set C is the intersection of all convex sets which contain the set C. We denote the convex hull by* $\mathrm{co}\,(C)$.

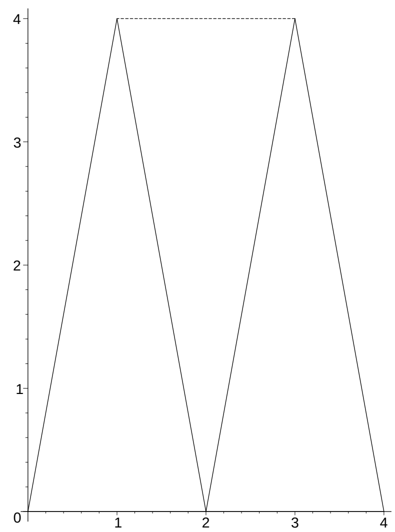

Figure 3: The Formation of a Convex Hull

We illustrate this definition in the next figure where the dotted line together with the original boundaries of the set for the boundary of the convex hull.

Examples 1.10

(a) Suppose that $[a, b]$ and $[c, d]$ are two intervals of the real line with $b < c$ so that the intervals are disjoint. Then the convex hull of the set $[a, b] \cup [c, d]$ is just the interval $[a, d]$.

(b) In $\mathbb{R}^2$ consider the elliptical annular region $\mathcal{E}$ consisting of the disk $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq R$ for a given positive number R without the interior points of an elliptical region as indicated in the next figure.

Clearly the set $\mathcal{E}$ is not convex for the line segment joining the indicated points P and Q has points lying in the "hole" of region and hence not in $\mathcal{E}$. Indeed, this is the case for any line segment joining two points of the region which are, say, symmetric with respect to the origin. The entire disk of radius R, however, is convex and indeed is the convex hull, $\text{co}\,(\mathcal{E})$.

These examples are typical. In each case, we see that the convex hull is obtained by adjoining all convex combinations of points in the original set. This is indeed a general result.

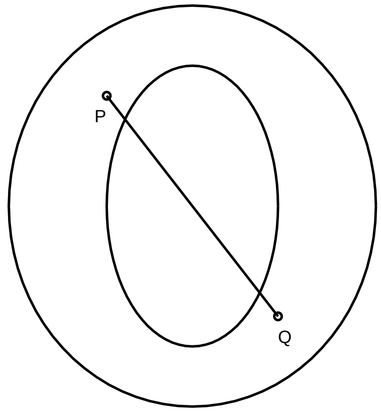

Figure 4: The Elliptically Annular Region

Theorem 1.11 *Let $S \subset V$. Then the set of all convex combinations of points of the set S is exactly* $\mathrm{co}\,(S)$.

Proof: Let us denote the set of all convex combinations of p points of S by $C_p(S)$. Then the set of *all possible* convex combinations of points of S is $C(S) := \cup_{p=1}^{\infty} C_p(S)$. If $x \in C(S)$ then it is a convex combination of points of S. Since $S \subset \mathrm{co}(S)$ which is convex, it is clear from Proposition 1.8 that $x \in \mathrm{co}(S)$ and, hence $C(S) \subset \mathrm{co}\,(S)$. To see that the opposite inclusion holds, it is sufficient to show that $C(S)$ is a convex set. Then, since $\mathrm{co}(S)$ is, by definition, the smallest convex set containing the points of S it follows that $\mathrm{co}(S) \subset C(S)$.

Now to see that $C(S)$ is indeed convex, let $x, y \in C(S)$. Then for some positive integers p and q, and $p-$ and q-tuples $\{\mu_i\}_{i=1}^p, \{\nu_j\}_{j=1}^q$ with $\sum_1^p \mu_i = 1$ and $\sum_1^q \nu_j = 1$, and points $\{x_1, x_2, \ldots, x_p\} \subset S$ and $\{y_1, y_2, \ldots, y_q\} \subset S$, we have

$$x = \sum_{i=1}^{p} \mu_i \, x_i \,, \quad \text{and} \quad y = \sum_{j=1}^{q} \nu_j \, y_j \,.$$

Now, let λ be such that $0 \leq \lambda \leq 1$ and look at the convex combination $(1-\lambda)x + \lambda y$. From the representations we have

$$
\begin{aligned}
(1-\lambda)\,x+\lambda\,y &= (1-\lambda)\,\left(\sum_{i=1}^{p}\mu_i\,x_i\right)+\lambda\,\left(\sum_{j=1}^{q}\nu_j\,y_j\right)\\
&= \sum_{i=1}^{p}(1-\lambda)\,\mu_i\,x_i+\sum_{j=1}^{q}\lambda\,\nu_j\,y_j
\end{aligned}
$$

But we have now, a combination of $p+q$ points of S whose coefficients are all non-negative and, moreover, for which

$$
\sum_{i=1}^{p}(1-\lambda)\,\mu_i+\sum_{j=1}^{q}\lambda\,\nu_j \;=\; (1-\lambda)\,\sum_{i=1}^{p}\mu_i+\lambda\,\sum_{j=1}^{q}\nu_j \;=\; (1-\lambda)\cdot 1+\lambda\cdot 1 \;=\; 1\,,
$$

which shows that the convex combination $(1-\lambda)x+\lambda y\in C(S)$. Hence this latter set is convex. $\square$

Convex sets in $\mathbb{R}^n$ have a very nice characterization discovered by Carathèodory.

Theorem 1.12 *Let S be a subset of $\mathbb{R}^n$. Then every element of $co\,(S)$ can be represented as a convex combination of no more than $(n+1)$ elements of S.*

Proof: Let $x\in\mathrm{co}\,(S)$. Then we can represent x as $\displaystyle\sum_{i=1}^{m}\alpha_i x^{(i)}$ for some vectors $x^{(i)}\in S$ and scalars $\alpha_i\geq 0$ with $\displaystyle\sum_{i=1}^{m}\alpha_i=1$. Let us suppose that m is the *minimal* number of vectors for which such a representation of x is possible. In particular, this means that for all $i=1,\ldots,m$, $\alpha_i>0$. If we were to have $m>n+1$, then the vectors $x^{(i)}-x$, $i=1,2,\ldots,m$ must be linearly dependent since there are more vectors in this set than the dimension of the space. It follows that there exist scalars $\lambda_2,\ldots,\lambda_m$, at least one of which is positive (why?) such that

$$
\sum_{i=2}^{m}\lambda_i\,(x^{(i)}-x) \;=\; 0\,.
$$

Let $\mu_1:=-\sum_{i=2}^{m}\lambda_i\,,$ and $\mu_i:=\lambda_i\,,i=2,3,\ldots,m$. Then

$$
\sum_{i=1}^{m}\mu_i\,x_i \;=\; 0\,,\quad\text{and}\quad \sum_{i=1}^{m}\mu_i \;=\; 0\,,
$$

while at least one of the scalars $\mu_2, \mu_3, \dots, \mu_m$ is positive.
The strategy for completing the proof is to produce a convex combination of vectors that represents x and which has *fewer* than m summands which would then contradict our choice of m as the *minimal* number of non-zero elements in the representation. To this end, $\hat{\alpha}_i := \alpha_i - \hat{\gamma}\mu_i\,, i = 1, \dots, m$, where $\hat{\gamma} > 0$ is the largest γ such that $\alpha_i - \gamma\,\mu_i \geq 0$ for all i. Then, since $\sum_{i=1}^m \mu_i x_i = 0$ we have

$$\begin{aligned}\sum_{i=1}^m \hat{\alpha}_i\, x_i &= \sum_{i=1}^m (\alpha_i - \hat{\gamma}\,\mu_i)\, x_i \\ &= \sum_{i=1}^m \alpha_i\, x_i - \hat{\gamma}\sum_{i=1}^m \mu_i\, x_i \;=\; \sum_{i=1}^m \alpha_i\, x_i \;=\; x\,,\end{aligned}$$

the last equality coming from our original representation of x. Now, the $\hat{\alpha}_i \geq 0$, at least one is zero, and

$$\sum_{i=1}^m \hat{\alpha}_i \;=\; \sum_{i=1}^m \alpha_i - \hat{\gamma}\sum_{i=1}^m \mu_i \;=\; \sum_{i=1}^m \alpha_i \;=\; 1\,,$$

the last equality coming from the choice of α_i. Since at least one of the $\hat{\alpha}_i$ is zero, this gives a convex representation of x in terms of *fewer* than m points in S which contradicts the assumption of minimality of m. $\square$

Carathéodory's Theorem has the nature of a **representation** theorem somewhat analogous to the theorem which says that any vector in a vector space can be represented as a linear combination of the elements of a basis. One thing both theorems do, is to give a *finite* and *minimal* representation of all elements of an infinite set. The drawback of Carathéodory's Theorem, unlike the latter representation, is that the choice of elements used to represent the point is neither uniquely determined for that point, *nor* does the theorem guarantee that the *same* set of vectors in C can be used to represent all vectors in C; the representing vectors will usually change with the point being represented. Nevertheless, the theorem is useful in a number of ways a we will see presently. First, a couple of examples.

Examples 1.13

(a) In $\mathbb{R}$, consider the interval $[0, 1]$ and the subinterval $\left(\frac{1}{4}, \frac{3}{4}\right)$. Then $\mathrm{co}\left(\frac{1}{4}, \frac{3}{4}\right) = \left(\frac{1}{4}, \frac{3}{4}\right)$. If we take the point $x = \frac{1}{2}$, then we have both

$$x = \frac{1}{2}\left(\frac{3}{8}\right) + \frac{1}{2}\left(\frac{5}{8}\right) \text{ and } x = \frac{3}{4}\left(\frac{7}{16}\right) + \frac{1}{4}\left(\frac{11}{16}\right).$$

So that certainly there is no uniqueness in the representation of $x = \dfrac{1}{2}$.

(b) In $\mathbb{R}^2$ we consider the two triangular regions, T_1, T_2, joining the points $(0,0)$, $(1,4)$, $(2,0)$, $(3,4)$ and $(4,0)$ as pictured in the next figures. The second of the figures indicates that joining the apexes of the triangles forms a trapezoid which is a convex set. It is the convex hull of the set $T_1 \cup T_2$.

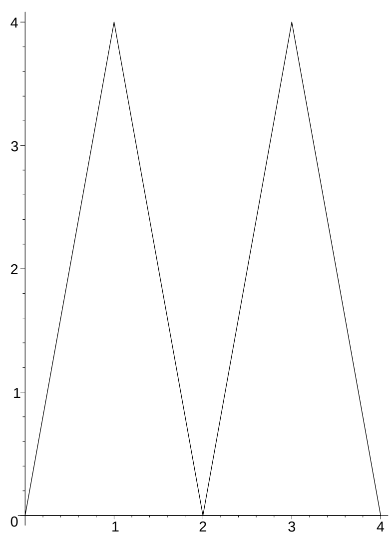

Figure 5: The Triangles T_1 and T_2

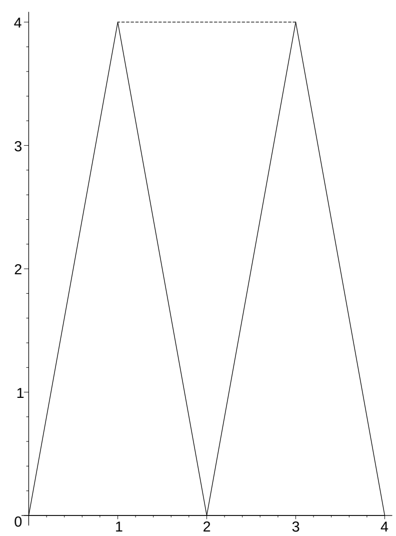

Figure 6: The Convex Hull of $T_1 \cup T_2$

Again, it is clear that two points which both lie in one of the original triangles have more than one representation. Similarly, if we choose two points, one from T_1 and one from T_2, say the points $(1,2)$ and $(3,2)$, the point

$$\frac{1}{2}\left[\begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 3 \\ 2 \end{pmatrix}\right] = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

does not lie in the original set $T_1 \cup T_2$, but does lie in the convex hull. Moreover, this point can also be represented by

$$\frac{1}{2}\begin{pmatrix} \frac{3}{2} \\ \\ 2 \end{pmatrix} + \frac{2}{2}\begin{pmatrix} \frac{5}{2} \\ \\ 2 \end{pmatrix}$$

as can easily be checked.

The next results depend on the notion of norm in $\mathbb{R}^n$ and on the convergence of a sequence of points in $\mathbb{R}^n$. In particular, it relies on the fact that, in $\mathbb{R}^n$, or for that matter in any complete metric space, Cauchy sequences converge.

Recall that a set of points in $\mathbb{R}^n$ is called **compact** provided it is closed and bounded. One way to characterize such a set in $\mathbb{R}^n$ is that if $C \subset \mathbb{R}^n$ is compact, then, given any sequence $\{x^{(1)}, x^{(2)}, \ldots, x^{(n)}, \ldots\} \subset C$ there is a subsequence which converges, $\lim_{k\to\infty} x^{(n_k)} = x^o \in C$.

As a corollary to Carathèodory's Theorem, we have the next result about compact sets:

Corollary 1.14 *The convex hull of a compact set in $\mathbb{R}^n$ is compact.*

Proof: Let $C \subset \mathbb{R}^n$ be compact. Notice that the simplex

$$\sigma := \left\{(\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{R}^n \, : \, \sum_{i=1}^{n} \lambda_i \, = \, 1\right\}$$

is also closed and bounded and is therefore compact. (Check!) Now suppose that $\{v^{(j)}\}_{j=1}^{\infty} \subset \text{co}\,(C)$. By Carathèodory's Theorem, each $v^{(j)}$ can be written in the form

$$v^{(k)} \, = \sum_{i=1}^{n+1} \lambda_{k,i} x^{(k,i)}, \quad \text{where} \quad \lambda_{k,i} \geq 0, \sum_{i=1}^{n+1} \lambda_{k,i} \, = \, 1, \quad \text{and} \quad x^{(k,i)} \in C.$$

Then, since C and σ are compact, there exists a sequence $k_1, k_2, \ldots$ such that the limits $\lim_{j\to\infty} \lambda_{k_j,i} = \lambda_i$ and $\lim_{j\to\infty} x^{(k_j,i)} = x^{(i)}$ exist for $i = 1, 2, \ldots, n+1$. Clearly $\lambda_i \geq 0$, $\sum_{i=1}^{n+1} \lambda_i = 1$ and $x_i \in C$.

Thus, the sequence $\{v^{(k)}\}_{k=1}^{\infty}$ has a subsequence, $\{v^{(k_j)}\}_{j=1}^{\infty}$ which converges to a point of co (C) which shows that this latter set is compact. □

The next result shows that if C is closed and convex (but perhaps not bounded) is has a smallest element in a certain sense. It is a simple result from analysis that involves the fact that the function $x \to \|x\|$ is a continuous map from $\mathbb{R}^n \to \mathbb{R}$ and the fact that Cauchy sequences in $\mathbb{R}^n$ converge.

Before proving the result, however, we recall that the norm in $\mathbb{R}^n$ satisfies what is known as the **parallelogram law**, namely that for any $x, y \in \mathbb{R}^n$ we have

$$\|x+y\|^2 + \|x-y\|^2 \;=\; 2\,\|x\|^2 + 2\,\|y\|^2\,.$$

This identity is easily proven by expanding the terms on the left in terms of the inner product rules. It is left as an exercise.

Theorem 1.15 *Every closed convex subset of $\mathbb{R}^n$ has a unique element of minimum norm.*

Proof:
Let K be such a set and note that $\iota := \inf\limits_{x \in K} \|x\| \geq 0$ so that the function $x \to \|x\|$ is bounded below on K. Let $x^{(1)}, x^{(2)}, \ldots$ be a sequence of points of K such that $\lim\limits_{i \to \infty} \|x^{(i)}\| = \iota$. [1] Then, by the parallelogram law, $\|x^{(i)} - x^{(j)}\|^2 = 2\,\|x^{(i)}\|^2 + 2\,\|x^{(j)}\|^2 - 4\,\|\frac{1}{2}\left(x^{(i)} + x^{(j)}\right)\|^2$. Since K is convex, $\frac{1}{2}\left(x^{(i)} + x^{(j)}\right) \in K$ so that $\|\frac{1}{2}\left(x^{(i)} + x^{(j)}\right)\| \geq \iota$. Hence

$$\|x^{(i)} - x^{(j)}\|^2 \;\leq\; 2\,\|x^{(i)}\|^2 + 2\,\|x^{(j)}\|^2 - 4\,\iota^2.$$

As $i, j \to \infty$, we have $2\,\|x^{(i)}\|^2 + 2\,\|x^{(j)}\|^2 - 4\,\iota \to 0$. Thus, $\{x^{(j)}\}_{j=1}^{\infty}$ is a Cauchy sequence and has a limit point x. Since K is closed, $x \in K$. Moreover, since the function $x \to \|x\|$ is a continuous function from $\mathbb{R}^n \to \mathbb{R}$,

$$\iota \;=\; \lim_{j \to \infty} \|x^{(j)}\| \;=\; \|x\|.$$

In order to show uniqueness of the point with minimal norm, suppose that there were two points, $x, y \in K, x \neq y$ such that $\|x\| = \|y\| = \iota$. Then by the parallelogram law,

$$\begin{aligned} 0 \quad < \quad & \|x-y\|^2 \;=\; 2\,\|x\|^2 + 2\,\|y\|^2 - 4\,\|\frac{1}{2}\,(x+y)\,\|^2 \\ = \quad & 2\,\iota^2 + 2\,\iota^2 - 4\,\|\frac{1}{2}\,(x+y)\,\|^2 \end{aligned}$$

so that $4\,\iota^2 \;>\; 4\,\|\frac{1}{2}\,(x+y)\,\|^2$ or $\|\frac{1}{2}\,(x+y)\,\| < \iota$ which would give a vector in K of norm less than the infimum ι. □

[1] Here, and throughout this course, we shall call such a sequence a **minimizing sequence**.

Example 1.16 *It is easy to illustrate the statement of this last theorem in a concrete case. Suppose that we define three sets in* $\mathbb{R}^2$ *by* $H_1^+ := \{(x,y) \in \mathbb{R}^2 \,:\, 5x - y \geq 1\}$, $H_2^+ := \{(x,y) \in \mathbb{R}^2 \,:\, 2x + 4y \geq 7\}$ *and* $H_3^- := \{(x,y) \in \mathbb{R}^2 \,:\, 2x + 2y \leq 6\}$ *whose intersection (the intersection of half-spaces) forms a convex set illustrated below. The point of minimal norm is the closest point in this set to the origin. From the projection theorem in* $\mathbb{R}^2$, *that point is determined by the intersection of the boundary line* $2x + 4y = 7$ *with a line perpendicular to it and which passes through the origin as illustrated here.*

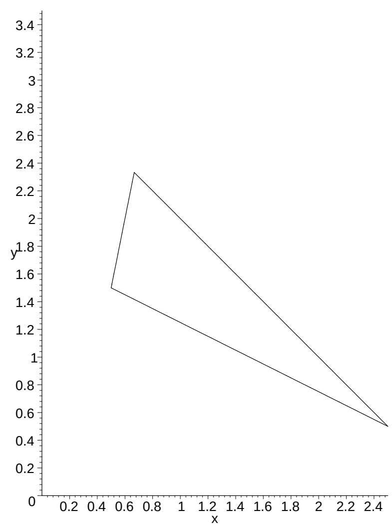

Figure 7: The Convex Set

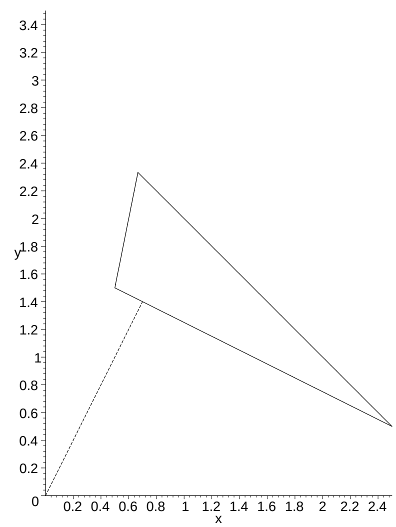

Figure 8: The Point of Minimal Norm

1.2 Separation Properties

There are a number of results which are of crucial importance in the theory of convex sets, and in the theory of mathematical optimization, particularly with regard to the development of *necessary conditions*, as for example in the theory of Lagrange multipliers. These results are usually lumped together under the rubric of *separation theorems*. We discuss two of these results which will be useful to us. We will confine our attention to the finite dimensional case.[2]

[2]In a general inner product space or in Hilbert space the basic theorem is called the Hahn-Banach theorem and is one of the central results of functional analysis.

The idea goes back to the idea of describing a circle in the plane by the set of all tangent lines to points on the boundary of the circle.

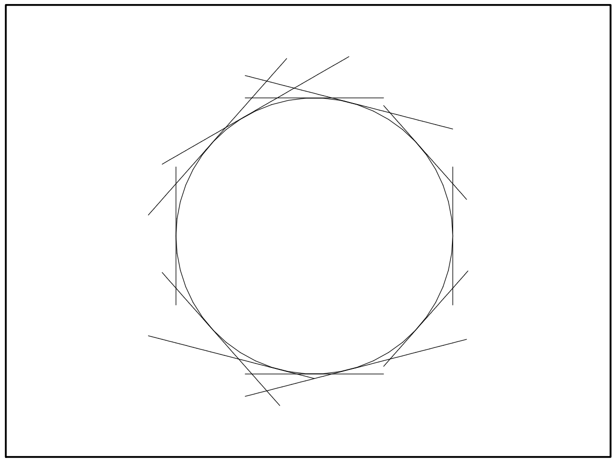

Figure 9: The Circle with Supporting Hyperplanes

Our proof of the Separation Theorem (and its corollaries) depends on the idea of the projection of a point onto a convex set. We begin by proving that result.
The statement of the Projection Theorem is as follows:

Theorem 1.17 *Let $C \subset \mathbb{R}^n$ be a closed, convex set. Then*

(a) For every $x \in \mathbb{R}^n$ there exists a unique vector $z^\star \in C$ that minimizes $\|z - x\|$ over all $z \in C$. We call $z^\star$ the projection of x onto C.

(b) $z^\star$ is the projection of x onto C if and only if

$$\langle y - z^\star, x - z^\star \rangle \leq 0, \text{ for all } y \in C\,.$$

Proof: Fix $x \in \mathbb{R}^n$ and let $w \in C$. Then minimizing $\|x - z\|$ over all $z \in C$ is equivalent to minimizing the same function over the set $\{z \in C \,:\, \|x - z\| \leq \|x - w\|\}$. This latter set is both closed and bounded and therefore the continuous function $g(z) := \|z - x\|$, according to the theorem of Weierstrass, takes on its minimum at some point of the set.
We use the paralellogram identity to prove uniqueness as follows. Suppose that there are two distinct points, z_1 and $z2$, which both minimize $\|z - x\|$ and denote this minimum by ι. Then we have

$$
\begin{aligned}
0 \;&<\; \|(z_1 - x) - (z_2 - x)\|^2 \;=\; 2\,\|z_1 - x\|^2 + 2\,\|z_2 - x\|^2 - 4\left\|\frac{1}{2}[(z_1 - x) + (z_2 - x)]\right\|^2 \\
&=\; 2\,\|z_1 - x\|^2 + 2\,\|z_2 - x\|^2 - 4\left\|\frac{z_1 + z_2}{2} - x\right\|^2 \;=\; 2\,\iota^2 + 2\,\iota^2 - 4\,\|\hat{z} - x\|^2\,,
\end{aligned}
$$

where $\hat{z} = (z_1 + z_2)/2 \in C$ since C is convex. Rearranging, and taking square roots, we have

$$
\|\hat{z} - x\| \;<\; \iota
$$

which is a contradiction of the fact that z_1 and z_2 give minimal values to the distance. Thus uniqueness is established.

To prove the inequality in part (b), and using $\langle \cdot, \cdot \rangle$ for the inner product, we have, for all $y, z \in C$, the inequality

$$
\begin{aligned}
\|y - x\|^2 \;&=\; \|y - z\|^2 + \|z - x\|^2 - 2\,\langle (y - z), (x - z)\rangle \\
&\geq\; \|z - x\|^2 - 2\,\langle (y - z), (x - z)\rangle\;.
\end{aligned}
$$

Hence, if z is such that $\langle (y - z), (x - z)\rangle \leq 0$ for all $y \in C$, then $\|y - x\|^2 \geq \|z - x\|^2$ for all $y \in C$ and so, by definition $z = z^\star$.

To prove the necessity of the condition, let $z^\star$ be the projection of x onto C and let $y \in C$ be arbitrary. For $\alpha > 0$ define $y_\alpha = (1 - \alpha) z^\star + \alpha y$ then

$$
\begin{aligned}
\|x - y_\alpha\|^2 \;&=\; \|(1 - \alpha)(x - z^\star) + \alpha(x - y)\|^2 \\
&=\; (1 - \alpha)^2 \|x - z^\star\|^2 + \alpha^2 \|x - y\|^2 + 2\,(1 - \alpha)\,\alpha\,\langle (x - z^\star), (x - y)\rangle\,.
\end{aligned}
$$

Now consider the function $\varphi(\alpha) := \|x - y_\alpha\|^2$. Then we have from the preceeding result

$$
\left.\frac{\partial \varphi}{\partial \alpha}\right|_{\alpha=0} \;=\; -2\,\|x - z^\star\|^2 + 2\,\langle (x - z^\star), (x - y)\rangle \;=\; -2\,\langle (y - z^\star), (x - z^\star)\rangle\,.
$$

Therefore, if $\langle (y - z^\star), (x - z^\star)\rangle > 0$ for some $y \in C$, then

$$
\left.\frac{\partial}{\partial \alpha}\left\{\|x - y_\alpha\|^2\right\}\right|_{\alpha=0} \;<\; 0
$$

and, for positive but small enough α, we have $\|x - y_\alpha\| < \|x - z^\star\|$. This contradicts the fact that $z^\star$ is the projection of x onto C and shows that $\langle (y - z^\star), (x - z^\star)\rangle \leq 0$ for all $y \in C$. $\square$

In order to study the separation properties of convex sets we need the notion of a **hyperplane**.

Definition 1.18 *Let $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$ and assume $a \neq 0$. Then the set*

$$H := \{x \in \mathbb{R}^n \,|\, \langle a, x\rangle \,=\, b\}\,,$$

is called a hyperplane with normal vector a.

We note, in passing, that hyperplanes are convex sets, a fact that follows from the bilinearity of the inner product. Indeed, if $x, y \in H$ and $0 \leq \lambda \leq 1$ we have

$$\langle a, (1-\lambda)\,x + \lambda\,y\rangle \;=\; (1-\lambda)\,\langle a, x\rangle + \lambda\,\langle a, y\rangle \;=\; (1-\lambda)\,b + \lambda\,b \;=\; b\,.$$

Each such hyperplane defines the half-spaces

$$H^+ := \{x \in \mathbb{R}^n \,|\, \langle a, x\rangle \,\geq\, b\}\,, \;\text{ and }\; H^- := \{x \in \mathbb{R}^n \,|\, \langle a, x\rangle \,\leq\, b\}\,.$$

Note that $H^+ \cap H^- = H$. These two half-spaces are *closed sets*, that is, they contain all their limit points. Their interiors [3] are given by

$$\overset{\circ}{H}{}^{+} \;=\; \{x \in \mathbb{R}^n \,|\, \langle a, x\rangle \,>\, b\}\,, \;\text{ and }\; \overset{\circ}{H}{}^{-} := \{x \in \mathbb{R}^n \,|\, \langle a, x\rangle \,<\, b\}\,.$$

Whether closed or open, these half-spaces are said to be generated by the hyperplane H. Each is a convex set.

Of critical importance to optimization problems is a group of results called **separation theorems**.

Definition 1.19 *Let $S, T \subset \mathbb{R}^n$ and let H be a hyperplane. Then H is said to* **separate** *S from T if S lies in one closed half-space determined by H while T lies in the other closed half-space. In this case H is called a* **separating hyperplane**. *If S and T lie in the* open *half-spaces, then H is said to* **strictly separate** *S and T.*

[3] Recall that a point x_o is an interior point of a set $S \subset \mathbb{R}^n$ provided there is a ball with center x_o of sufficiently small radius, which contains only points of S. The interior of a set consists of all the interior points (if any) of that set.

We can prove the *Basic Separation Theorem* in $\mathbb{R}^n$ using the result concerning the projection of a point on a convex set. The second of our two theorems, the *Basic Support Theorem* is a corollary of the first.

Theorem 1.20 *Let $C \subset \mathbb{R}^n$ be convex and suppose that $y \notin c\ell(C)$. Then there exist an $a \in \mathbb{R}^n$ and a number $\gamma \in \mathbb{R}$ such that $\langle a, x\rangle > \gamma$ for all $x \in C$ and $\langle a, y\rangle \leq \gamma$.*

Proof: Let $\hat{c}$ be the projection of y on $c\ell(C)$, and let $\gamma = \inf_{x\in C} \|x - y\|$, i.e., γ is the distance from y to its projection $\hat{c}$. Note that since $y \notin c\ell(C), \gamma > 0$.

Now, choose an arbitrary $x \in C$ and $\lambda \in (0, 1)$ and form $x_\lambda := (1-\lambda)\hat{c} + \lambda x$. Since $x \in C$ and $\hat{c} \in c\ell(C)$, $x_\lambda \in c\ell(C)$. So we have

$$\begin{aligned}
\|x_\lambda - y\|^2 &= \|(1-\lambda)\hat{c} + \lambda x - y\|^2 = \|\hat{c} + \lambda(x - \hat{c}) - y\|^2 \\
&\geq \|\hat{c} - y\|^2 > 0.
\end{aligned}$$

But we can write $\|\hat{c} + \lambda(x - \hat{c}) - y\|^2 = \|(\hat{c} - y) + \lambda(x - \hat{c})\|^2$ and we can expand this latter expression using the rules of inner products.

$$\begin{aligned}
0 &< \|\hat{c} - y\|^2 \leq \|(\hat{c} - y) + \lambda(x - \hat{c})\|^2 = \langle(\hat{c} - y) + \lambda(x - \hat{c}), (\hat{c} - y) + \lambda(x - \hat{c})\rangle \\
&= \langle\hat{c} - y, \hat{c} - y\rangle + \langle\hat{c} - x, \lambda(x - \hat{c})\rangle + \langle\lambda(x - \hat{c}), \hat{c} - y\rangle + \langle\lambda(x - \hat{c}), \lambda(x - \hat{c})\rangle,
\end{aligned}$$

and from this inequality, we deduce that

$$2\,\lambda\langle\hat{c} - y, x - \hat{c}\rangle + \lambda^2\langle x - \hat{c}, x - \hat{c}\rangle \;\geq\; 0.$$

From this last inequality, dividing both sides by $2\,\lambda$ and taking a limit as $\lambda \to 0^+$, we obtain

$$\langle\hat{c} - y, x - \hat{c}\rangle \;\geq\; 0.$$

Again, we can expand the last expression $\langle\hat{c} - y, x - \hat{c}\rangle = \langle\hat{c} - y, x\rangle + \langle\hat{c} - y, -\hat{c}\rangle \geq 0$. By adding and subracting y and recalling that $\|\hat{c} - y\| > 0$, we can make the following estimate,

$$\begin{aligned}
\langle\hat{c} - y, x\rangle &\geq \langle\hat{c} - y, \hat{c}\rangle \;=\; \langle\hat{c} - y, y - y + \hat{c}\rangle \\
&= \langle\hat{c} - y, y\rangle + \langle\hat{c} - y, \hat{c} - y\rangle \;=\; \langle\hat{c} - y, y\rangle + \|\hat{c} - y\|^2 > \langle\hat{c} - y, y\rangle.
\end{aligned}$$

In summary, we have $\langle \hat{c} - y, x \rangle > \langle \hat{c} - y, y \rangle$ for all $x \in C$. Finally, define $a := \hat{c} - y$. Then this last inequality reads

$$\langle a, x \rangle \;>\; \langle a, y \rangle \;\text{ for all }\; x \in C.$$

$\square$

Before proving the Basic Support Theorem, let us recall some terminology. Suppose that $S \subset \mathbb{R}^n$. Then a point $s \in S$ is called a boundary point of S provided that every ball with s as center intersects both S and its complement $\mathbb{R}^n \setminus S := \{x \in \mathbb{R}^n \;:\; x \notin S\}$. Note that every boundary point is a limit point of the set S, but that the converse is not true. Moreover, we introduce the following definition:

Definition 1.21 *A hyperplane containing a convex set C in one of its closed half spaces and containing a boundary point of C is said to be a* **supporting hyperplane** *of C.*

Theorem 1.22 *Let C be a convex set and let y be a boundary point of C. Then there is a hyperplane containing y and containing C in one of its half spaces.*

Proof: Let $\{y^{(k)}\}_{k=1}^{\infty}$ be a sequence in $\mathbb{R}^n \setminus c\ell(C)$ with $y^{(k)} \to y$ as $k \to \infty$. Let $\{a_{(k)}\}_{k=1}^{\infty}$ be the sequence of vectors constructed in the previous theorem and define $\hat{a}_{(k)} := a_{(k)}/\|a_{(k)}\|$. Then, for each k, $\langle \hat{a}_{(k)}, y^{(k)} \rangle < \inf_{x \in C} \langle \hat{a}_{(k)}, x \rangle$.
Since the sequence $\{\hat{a}_{(k)}\}_{k=1}^{\infty}$ is bounded, it contains a convergent subsequence $\{\hat{a}_{(k_j)}\}_{j=1}^{\infty}$ which converges to a limit $\hat{a}_{(o)}$. Then, for any $x \in C$,

$$\langle a_{(o)}, y \rangle \;=\; \lim_{j \to \infty} \langle \hat{a}_{(k_j)}, y^{(k_j)} \rangle \;\leq\; \lim_{j \to \infty} \langle \hat{a}_{(k_j)}, x \rangle \;=\; \langle \hat{a}_{(o)}, x \rangle.$$

$\square$

1.3 Extreme Points

For our discussion of extreme points, it will be helpful to recall the definition of polytope and polygon.

Definition 1.23 *A set which can be expresssed as the intersection of a finite number of half-spaces is said to be a convex* **polytope**. *A non-empty, bounded polytope is called a* **polygon**. *The following figures give examples of polytopes and polygons.*

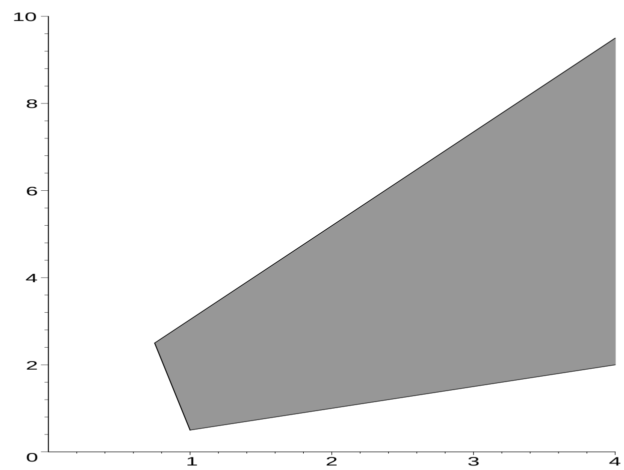

Figure 10: A Convex Polytope

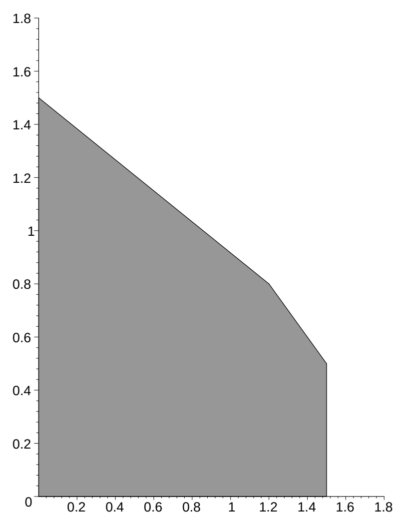

Figure 11: A Convex Polygon in $\mathbb{R}^2$

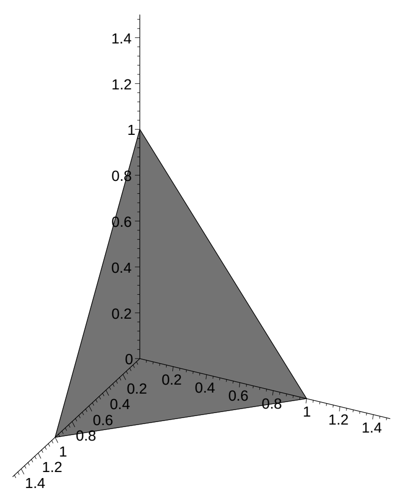

Figure 12: A Convex Polygon in $\mathbb{R}^3$

Exercise 1.24 *Given vectors $a_{(1)}, a_{(2)}, \ldots, a_{(p)}$ and real numbers $b_1, b_2, \ldots, b_p$, each inequality $\langle x, a_{(i)} \rangle \le b_i$, $i = 1, 2, \ldots, p$ defines a half-space.*
Show that if one of the $a_{(i)}$ is zero, the resulting set can still be expressed as the intersection of a finite number of half-planes.

If we look at a typical convex polygon in the plane, we see that there are a finite number of corner points, pairs of which define the sides of the polygon.

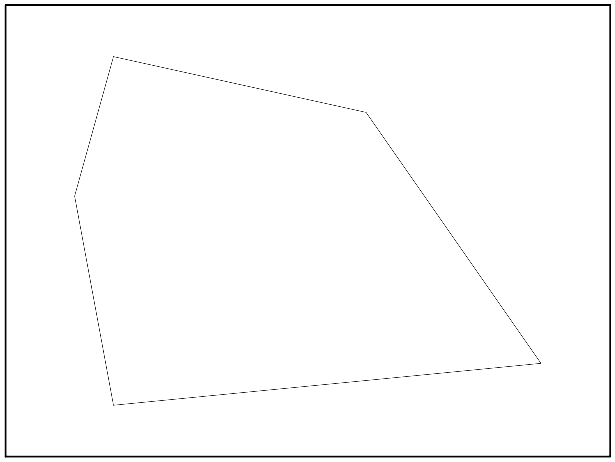

Figure 13: A Convex Polygon

These five corner points are called **extreme points** of the polygon. They have the property that there are no two distinct points of the convex set in terms of which such a corner point can be written as a convex combination. What is interesting is that every other point of the convex set *can* be written as a linear combination of these extreme points, so that, in this sense, the finitely many corner points can be used to describe the entire set. Put another way, the convex hull of the extreme points is the entire polygon. Let us go back to an earlier example, Example 2.4.13.

Example 1.25 Let H_1, H_2, and H_3 be hyperplanes described, respecitvely, by the equations $5x - y = 1$, $2x + 4y = 7$, and $x + y = 3$. The region is the intersection of the half-spaces defined by these hyperplanes and is depicted in the next figure.

The extreme points of this bounded triangular region are the three vertices of the region, and are determined by the intersection of two of the hyperplanes. A simple calculation shows that they are $P_1 : \left(\frac{1}{2}, \frac{5}{2}\right)$, $P_2 := \left(\frac{2}{3}, \frac{7}{3}\right)$, $P_3 := \left(\frac{5}{2}, \frac{1}{2}\right)$.

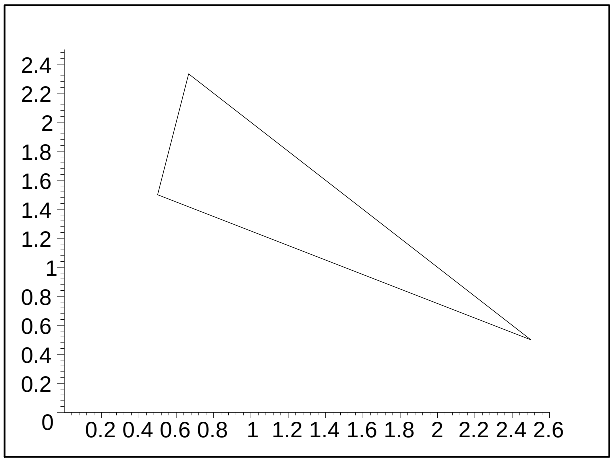

Figure 14: The Trianglular Region

We can illustrate the statement by first observing that any vertex is a trivial convex combination of the three vertices, e.g., $P_1 = 1\,P1 + 0\,P_2 + 0\,P_3$. Moreover, each *side* is a convex combination of the two vertices which determine that side. Thus, say, all points of H_3 are of the form

$$(1-\lambda)\begin{pmatrix} \frac{1}{2} \\ \frac{5}{2} \end{pmatrix} + \lambda \begin{pmatrix} \frac{5}{2} \\ \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} + 2\,\lambda \\ \frac{5}{2} - 2\,\lambda \end{pmatrix}, \quad \text{for} \;\; \lambda \in [0,1].$$

We can in fact check that all such points lie on H_3 by just adding the components. Thus

$$\left(\frac{1}{2} + 2\,\lambda\right) + \left(\frac{5}{2} - 2\,\lambda\right) = \frac{1}{2} + \frac{5}{2} = 3.$$

As for the points *interior* to the triangle, we note that, as illustrated in the next figure, that all such points lie on a line segment from P_1 to a point $\hat{P}$ lying on the side determined by the other two vertices. Thus, we know that for some $\lambda \in (0,1)$, $\hat{P} = (1-\lambda)\,P_2 + \lambda\,P_3$.

Suppose that the point Q is a point interior to the triangle, and that it lies on a line segment $\overline{P_1\hat{P}}$. The point Q has the form $Q:\,(1-\mu)\,P_2 + \mu\,\hat{P}$. So we can write

$$\begin{aligned} Q:\,(1-\mu)\,P_2 + \mu\,\hat{P} &= (1-\mu)\,P_2 + \mu\,[(1-\lambda)\,P_2 + \lambda\,P_3] \\ &= (1-\mu)\,P_2 + \mu\,(1-\lambda)\,P_2 + \mu\,\lambda P_3. \end{aligned}$$

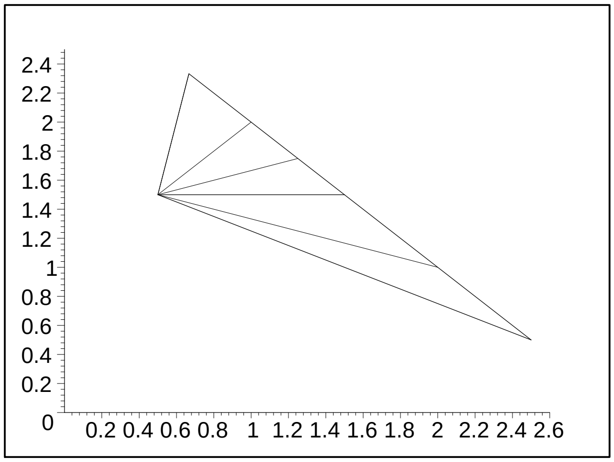

Figure 15: Rays from P_1

Note that all the coefficients in this last expression are non-negative and moreover

$$(1-\mu) + \mu\,(1-\lambda) + \mu\,\lambda \;=\; 1 - \mu + \mu - \mu\,\lambda + \mu\,\lambda \;=\; 1 - \mu + \mu \;=\; 1.$$

so that the point Q is a convex combination of three corner points of the polygon.

Having some idea of what we mean by an extreme point in the case of a polygon, we are prepared to introduce a definition.

Definition 1.26 *A point x in a convex set C is said to be an* **extreme point** *of C provided there are not two distinct points $x_1, x_2 \in C, x_1 \neq x_2$ such that $x = (1-\lambda)\,x_1 + \lambda\,x_2$, for some $\lambda \in (0,1)$.*

Notice that the corner points of a polygon satisfy this condition and are therefore extreme points. In the case of a polygon there are only finitely many. However, this may not be the case. We need only to think of the convex set in $\mathbb{R}^2$ consisting of the unit disk centered at the origin. In that case, the extreme points are *all* the points of the unit circle, i.e., the *boundary points* of the disk.

There are several questions that we wish to answer: (a) Do all convex sets have an extreme point? (ANS.: NO!); (b) Are there useful conditions which *guarantee* that an extreme point exists? (ANS. YES!); (c) What is the relationship, *under the conditions of part* (b) between the convex set,C, and the set of its extreme points, ext (C)? (ANS.:

$c\ell$ [co ext (C)] $= C$.), and, (d) What implications do these facts have with respect to optimization problems?

It is very easy to substantiate our answer to (a) above. Indeed, if we look in $\mathbb{R}^2$ at the set $\{(x,y) \in \mathbb{R}^2 \,:\, 0 \le x \le 1, y \in \mathbb{R}\}$ (draw a picture!) we have an example of a convex set with no extreme points. In fact this example is in a way typical of sets without extreme points, but to understand this statment we need a preliminary result.

Lemma 1.27 *Let C be a convex set, and H a supporting hyperplane of C. Define the (convex) set $T := C \cap H$. Then every extreme point of T is also an extreme point of C.*

Remark: The set T is illustrated in the next figure. Note that the intersection of H with C is not just a single point. It is, nevertheless, closed and convex since both H and C enjoy those properties.

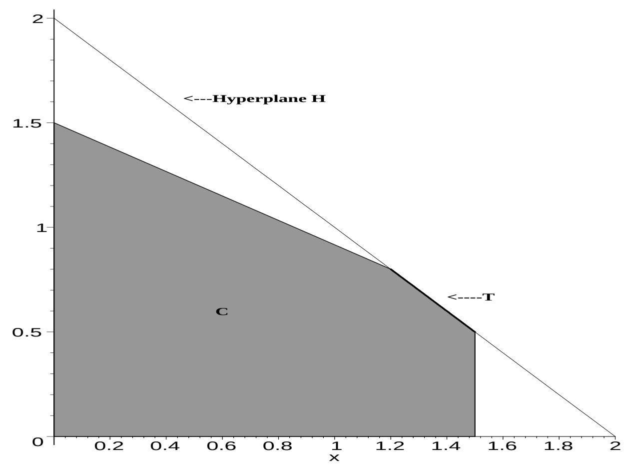

Figure 16: The set $T = H \cap C$

Proof: : Supose $\tilde{x} \in T$ is *not* an extreme point of C. Then we can find a $\lambda \in (0,1)$ such that $\tilde{x} = (1-\lambda)x_1 + \lambda\, x_2$ for some $x_1, x_2 \in C$, $x_1 \ne x_2$. Assume, without loss of generality, that H is described by $\langle x, a\rangle = c$ and that the convex set C lies in the positive half-space determined by H. Then $\langle x_1, a\rangle \ge c$ and $\langle x_2, a\rangle \ge c$. But since $\tilde{x} \in H$,

$$c \;=\; \langle \tilde{x}, a\rangle = (1-\lambda)\,\langle x_1, a\rangle + \lambda\,\langle x_2, a\rangle,$$

and thus x_1 and x_2 lie in H. Hence $x_1, x_2 \in T$ and hence $\tilde{x}$ is not an extreme point of T. $\square$

We can now prove a theorem that guarantees the existence of extreme points for certain convex subsets of $\mathbb{R}^n$.

Theorem 1.28 *A non-empty, closed, convex set $C \subset \mathbb{R}^n$ has at least one extreme point if and only if it does not contain a line, that is, a set L of the form $L = \{x + \alpha\, d \,:\, \alpha \in \mathbb{R}, d \neq 0\}$.*

Proof: Suppose first that C has an extreme point $x^{(e)}$ and contains a line $L = \{\overline{x} + \alpha\, d \,:\, \alpha \in \mathbb{R}, d \neq 0\}$. We will see that these two assumptions lead to a contradiction. For each $n \in \mathbb{N}$, the vector

$$x_{\pm}^{(n)} \;:=\; \left(1 - \frac{1}{n}\right) x^{(e)} + \frac{1}{n}\,(\overline{x} \pm n\,d) \;=\; x^{(e)} \pm d + \frac{1}{n}(\overline{x} - x^{(e)})$$

lies in the line segment connecting $x^{(e)}$ and $\overline{x} \pm n\,d$, and so it belongs to C. Since C is closed, $x^{(e)} \pm d = \lim\limits_{n\to\infty} x_{\pm}^{(n)}$ must also belong to C. It follows that the three vectors $x^{(e)} - d$, $x^{(e)}$ and $x^{(e)} + d$, all belong to C contradicting the hypothesis that $x^{(e)}$ is an extreme point of C.

To prove the converse statement, we will use induction on the dimension of the space. Suppose then that C does *not* contain a line. Take, first, $n = 1$. The statement is obviously true in this case for the only closed convex sets which are not all of $\mathbb{R}$ are just closed intervals.

We assume, as induction hypothesis, that the statement is true for $\mathbb{R}^{n-1}$. Now, if a nonempty, closed, convex subset of $\mathbb{R}^n$ contains no line, then it must have boundary points. Let $\overline{x}$ be such a point and let H be a supporting hyperplane of C at $\overline{x}$. Since H is an $(n-1)$-dimensional manifold, the set $C \cap H$ lies in an $(n-1)$-dimensional subspace and contains no line, so by the induction hypothesis, $C \cap H$ must have an extreme point. By the preceeding lemma, this extreme point of C. $\square$

From this result we have the important representation theorem:

Theorem 1.29 *A closed, bounded, convex set $C \subset \mathbb{R}^n$ is equal to $c\ell\,[\,\text{co ext}\,(C)]$.*

Proof: We begin by observing that since the set C is bounded, it can contain no line. Moreover, the smallest convex set that contains the non-empty set $\text{ext}\,(C)$ is just the convex hull of this latter set. So we certainly have that $C \supset c\ell[\,\text{co ext}\,(C)] \neq \emptyset$. Denote the closed convex hull of the extreme points by K. We remark that, since C is bounded, K is necessarily bounded.

We need to show that $C \subset K$. Assume the contrary, namely that there is a $y \in C$ with $y \notin K$. Then by the first separation theorem, there is a hyperplane $\hat{H}$ separating y and K. Thus, for some $a \neq 0$, $\langle y, a\rangle < \inf\limits_{x \in K} \langle x, a\rangle$. Let $c_o = \inf\limits_{x \in C} \langle x, a\rangle$. The number c_o is finite and there is and $\hat{x} \in C$ such that $\langle \hat{x}, a\rangle = c_o$ because, by the theorem of Weierstraß, a continuous function (in this case $x \mapsto \langle x, a\rangle$) takes on its minimum value over any closed bounded set.[4]

It follows that the hyperplane $H = \{x \in \mathbb{R}^n : \langle x, a\rangle = c_o\}$ is a supporting hyperplane to C. It is disjoint from K since $c_o < \inf\limits_{x \in K} \langle x, a\rangle$. The preceeding two results then show that, the set $H \cap C$ has an extreme point which is also *a fortiori* an extreme point of C and which cannot then be in K. This is a contradiction. $\square$.

Up to this point, we have answered the first three of the questions raised above. Later, we will explore the relationship of these results to the simplex algorithm of linear programming.

1.4 Convex Functions

Our final topic in this first part of the course is that of convex functions. Again, we will concentrate on the situation of a map $f : \mathbb{R}^n \to \mathbb{R}$ although the situation can be generalized immediately by replacing $\mathbb{R}^n$ with any real vector space V.

The idea of a convex functions is usually first met in elementary calculus when the necessary conditions for a maximum or minimum of a differentiable function are discussed. There, one distinguishes between local maxima and minima by looking at the "concavity" of a function; functions have a local minimum at a point if they are "concave up" and a local maximum if they are concave down. The second derivative test is one which checks the concavity of the function.

It is unfortunate that the terminology persists. As we shall see, the mathematical properties of these functions can be fully described in terms of convex sets and in such a way that the analysis carries over easily from one to multiple dimensions. For these reasons, among others, we take, as our basic notion, the idea of a convex set in a vector space.

We will find it useful, and in fact modern algorithms reflect this usefulness, to consider functions $f\mathbb{R}^n \to \mathbb{R}^*$ where $\mathbb{R}^*$ is the set of **extended real numbers**. This set is obtained from the real numbers by adjoining two symbols, $+\infty$ and $-\infty$ to $\mathbb{R}$ together with the

[4]The continuity of this map follows immediately from the Cauchy-Schwarz Inequality!

following relations: $-\infty < s < +\infty$ for all $x \in \mathbb{R}$. Thus the set $\mathbb{R}^* = \mathbb{R} \cup \{-\infty, +\infty\}$, and any element of the set is called an extended real number. The arithmetic operations are extended, but with some exceptions. Thus $x \pm \infty = \pm\infty, \pm\infty + \pm\infty = \pm\infty, x \cdot \pm\infty = \pm\infty$ if $x > 0, x \cdot \pm\infty = \mp\infty$ if $x < 0$, and so forth. The relations $+\infty + (-\infty)$ and $0 \cdot (\pm\infty)$ are not defined.
In fact, we will have little need for doing arithemetic with these symbols; for us, they are merely a convenience, so that, for example, instead of saying that a given function f is not bounded below, we can simply write inf $f = -\infty$.
We want to give an interesting example here, but before we do that, we need to recall the definition of the graph of a function.

Definition 1.30 *Let $f : A \to B$, where A and B are non-empty sets. Then the* **graph of** f *is defined by*

$$\mathrm{Gr}\ (f) := \{(a, b) : a = f(b), a \in A, b \in B\}.$$

A probably new definition is the following one for the so-called **indicator function** of a set.

Definition 1.31 *Let $S \subset \mathbb{R}^n$ be a non-empty set. Then the* **indicator function of** S *is defined as the function ψ_S given by*

$$\psi_S(x) := \begin{cases} 0 & \textit{if } x \in S \\ +\infty & \textit{if } x \notin S. \end{cases}$$

Example 1.32 Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ and let $S \subset \mathbb{R}$ be non-empty. then the **restriction** of f to S, written $f\Big|_S$, can be defined in terms of its graph as

$$\mathrm{Gr}\ \left(f\Big|_S\right) = \{(y, x) : x \in S, y = f(x)\}.$$

On the other hand, consider the function $f + \psi_S$. This function maps $\mathbb{R}^n$ to $\mathbb{R}^*$ and, by definition is

$$(f + \psi_S)(x) = \begin{cases} f(x), & \text{if } x \in S \\ +\infty, & \text{if } x \notin S. \end{cases}$$

Note that since we assume here that $f(\mathbb{R}^n) = \mathbb{R}$, the function does not take on the value $+\infty$ so that the arithmetic makes sense!

The point is that the function $f + \psi_S$ can be identified with the restriction of the function f to the set S.

Before beginning with the main part of the discussion, we want to keep a couple of examples in mind.
The primal example of a convex function is $x \mapsto x^2$, $x \in \mathbb{R}$. As we learn in elementary calculus, this function is infinitely often differentiable and has a single critical point at which the function in fact takes on, not just a relative minimum, but an *absolute minimum*.

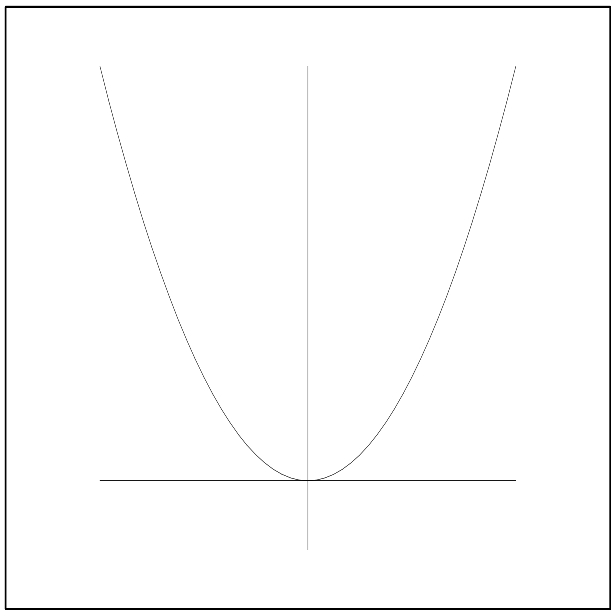

Figure 17: The Generic Parabola

The critical points are, by definition, the solution of the equation $\dfrac{d}{dx}\, x^2 = 2\,x$ or $2\,x = 0$. We can apply the second derivative test at the point $x = 0$ to determine the nature of the critical point and we find that, since $\dfrac{d^2}{dx^2}\left(x^2\right) = 2 > 0$, the function is "concave up" and the critical point is indeed a point of relative minimum. That this point gives an *absolute minimum* to the function, we need only see that the function values are bounded below by zero since $x^2 > 0$ aofr all $x \neq 0$.
We can give a similar example in $\mathbb{R}^2$.

Example 1.33 We consider the function

$$(x, y) \mapsto \frac{1}{2}\, x^2 + \frac{1}{3}\, y^2 := z,$$

whose graph appears in the next figure.

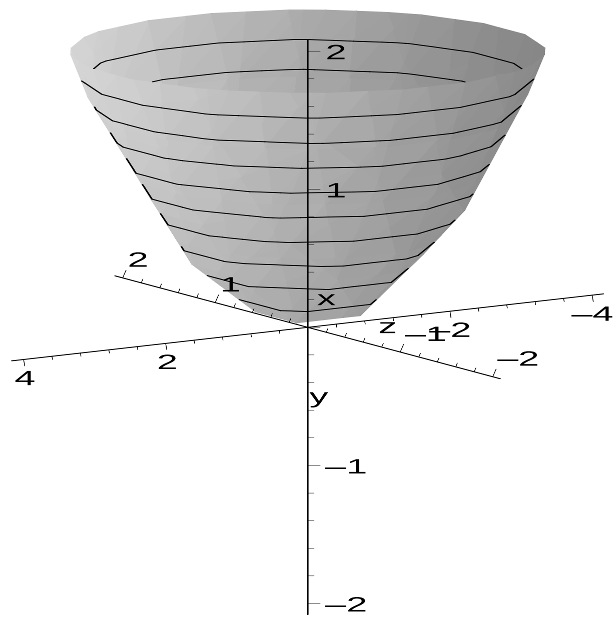

Figure 18: An Elliptic Parabola

This is an elliptic paraboloid. In this case we expect that, once again, the minimum will occur at the origin of coordinates and, setting $f(x, y) = z$, we can compute

$$\text{grad }\ (f)\,(x, y)\ =\ \begin{pmatrix} x \\ \\ \frac{2}{3}\,y \end{pmatrix},\ \ \text{and}\ \ \mathrm{H(f(x,y))}\ =\ \begin{pmatrix} 1 & 0 \\ \\ 0 & \frac{2}{3} \end{pmatrix}.$$

Notice that, in our terminology, the Hessian matrix $H(f)$ is positive definite at *all* points $(x, y) \in \mathbb{R}^2$.

The critical points are exactly those for which $\mathrm{grad\ [f(x,y)]} = 0$ whose only solution is $x = 0, y = 0$. The second derivative test is just that

$$\det\ \mathrm{H(f(x,y))}\ =\ \left(\frac{\partial^2 \mathrm{f}}{\partial \mathrm{x}^2}\right)\,\left(\frac{\partial^2 \mathrm{f}}{\partial \mathrm{y}^2}\right) - \frac{\partial^2 \mathrm{f}}{\partial \mathrm{x}\,\partial \mathrm{y}}\ >\ 0$$

which is clearly satisfied.
Again, since for all $(x, y) \neq (0, 0), z > 0$, the origin is a point where f has an absolute minimum.

As the idea of *convex set* lies at the foundation of our analysis, we want to describe the set of convex functions in terms of convex sets. To do this, we introduce a crucial object associated with a function, namely its **epigraph**.

Definition 1.34 *Let V be a real vector space and $S \subset V$ be a non-empty set. If $f : S \to \mathbb{R}$ then epi(f) is defined by*

$$\operatorname{epi}(f) := \{(z, s) \in \mathbb{R}^* \times S \,:\, z \geq f(s)\}.$$

Convex functions are defined in terms of their epigraphs as indicated next.

Definition 1.35 *Let $C \subset \mathbb{R}^n$ be convex and $f : C \longrightarrow \mathbb{R}^*$. Then the function f is called a* **convex function** *provided epi$(f) \subset \mathbb{R} \times \mathbb{R}^n$ is a convex set.*

We emphasize that this definition has the advantage of directly relating the theory of convex sets to the theory of convex functions. However, a more traditional definition is that a function is convex provided that, for any $x, y \in C$ and any $\lambda \in [0, 1]$

$$f\,(\,(1-\lambda)\,x + \lambda\,y) \,\leq\, (1-\lambda)\,f(x) + \lambda\,f(y)\,.$$

In fact, these definitions turn out to be equivalent. Indeed, we have the following result.

Theorem 1.36 *Let $C \subset \mathbb{R}^n$ be convex and $f : C \longrightarrow \mathbb{R}^*$. Then the following are equivalent:*

(a) epi(f) is convex.

(b) For all $\lambda_1, \lambda_2, \ldots, \lambda_n$ with $\lambda_i \geq 0$ and $\sum_{i=1}^n \lambda_i = 1$, and points $x_i \in C, i = 1, 2, \ldots, n$, we have

$$f\left(\sum_{i=1}^n \lambda_i\,x_i\right) \,\leq\, \sum_{i=1}^n \lambda_i\,f(x_i)\,.$$

(c) For any $x, y \in C$ and $\lambda \in [0,1]$,

$$f\left((1-\lambda)\, x + \lambda\, y \right) \;\le\; (1-\lambda)\, f(x) + \lambda\, f(y)\,.$$

Proof: To see that (a) implies (b) we note that, for all $i = 1, 2, \ldots, n$, $(f(x_i), x_i) \in$ epi (f). Since this latter set is convex, we have that

$$\sum_{i=1}^{n} \lambda_i\, (f(x_i), x_i) \;=\; \left(\sum_{i=1}^{n} \lambda_i\, f(x_i), \sum_{i=1}^{n} \lambda_i\, x_i \right) \in\; \text{epi}\,(f)\,,$$

which, in turn, implies that

$$f\left(\sum_{i=1}^{n} \lambda_i\, x_i \right) \;\le\; \sum_{i=1}^{n} \lambda_i\, f(x_i)\,.$$

This establishes (b). It is obvious that (b) implies (c). So it remains only to show that (c) implies (a) in order to establish the equivalence.
To this end, suppose that $(z_1, x_1), (z_2, x_2) \in$ epi (f) and take $0 \le \lambda \le 1$. Then

$$(1-\lambda)\,(z_1, x_1) + \lambda\,(z_2, x_2) \;=\; ((1-\lambda)\, z_1 + \lambda\, z_2, (1-\lambda)\, x_1 + \lambda\, x_2)\;,$$

and since $f(x_1) \le z_1$ and $f(x_2) \le z_2$ we have, since $(1-\lambda) > 0$, and $\lambda > 0$, that

$$(1-\lambda)\, f(x_1) + \lambda\, f(x_2) \;\le\; (1-\lambda)\, z_1 + \lambda\, z_2\,.$$

Hence, by the assumption (c), $f\left((1-\lambda)\, x_1 + \lambda\, x_2 \right) \le (1-\lambda)\, z_1 + \lambda\, z_2$, which shows this latter point is in epi(f). $\square$

In terms of the indicator function of a set, which we defined earlier, we have a simple result

Proposition 1.37 *A non-empty subset $D \subset \mathbb{R}^n$ is convex if and only if its indicator function is convex.*

Proof: The result follows immediately from the fact that epi$\,(\psi_D) = D \times \mathbb{R}_{\ge 0}$.

Certain simple properties follow immediately from the analytic form of the definition (part (c) of the equivalence theorem above). Indeed, it is easy to see, and we leave it as an exercise for the reader, that if f and g are convex functions defined on a convex set C,

then $f+g$ is likewise convex on C. The same is true if $\beta \in real\,, \beta > 0$ and we consider βf.
We also point out that if $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is a *linear* or *affine*, then f is convex. Indeed, suppose that for a vector $a \in \mathbb{R}^n$ and a real number b, the function f is given by $f(x) =< a, x > +b$. Then we have, for any $\lambda \in [0, 1]$,

$$\begin{aligned} f(\,(1-\lambda)\,x+\lambda\,y) &= \;< a, (1-\lambda)\,x+\lambda\,y > +b \;=\; (1-\lambda) \;< a, x > +\lambda \;< a, y > +(1-\lambda)\,b+\lambda\,b \\ &= \;(1-\lambda)\,(< a, x > +b)+\lambda\,(< a, y > +b) \;=\; (1-\lambda)\,f(x)+\lambda\,f(y)\,, \end{aligned}$$

and so f is convex, the weak inequality being an equality in this case.

In the case that f is *linear*, that is $f(x) =< a, x >$ for some $a \in \mathbb{R}^n$ then it is easy to see that the map $\varphi : x \rightarrow [f(x)]^2$ is also convex. Indeed, if $x, y \in \mathbb{R}^n$ then, setting $\alpha = f(x)$ anad $\beta = f(y)$, and taking $0 < \lambda < 1$ we have

$$\begin{aligned} (1-\lambda)\,\varphi(x)+\lambda\,\varphi(y)-\varphi(\,(1-\lambda)\,x &+ \lambda\,y) \\ &= (1-\lambda)\,\alpha^2+\lambda\,\beta^2-(\,(1-\lambda)\,\alpha+\lambda\,\beta)^2 \\ &= (1-\lambda)\,\lambda\,(\alpha-\beta)^2 \;\geq 0\,. \end{aligned}$$

Note, that in particular for the function $f : \mathbb{R} \longrightarrow \mathbb{R}$ given by $f(x) = x$ is linear and that $[f(x)]^2 = x^2$ so that we have a proof that the function that we usually write $y = x^2$ is a convex function.

A simple sketch of the parabola $y = x^2$ and any horizontal cord (which necessarily lies above the graph) will convince the reader that all points in the domain corresponding to the values of the function which lie below that horizontal line, form a convex set in the domain. Indeed, this is a property of convex functions which is often useful.

Proposition 1.38 *If C is a convex set and $f : C \longrightarrow \mathbb{R}$ is a convex function, then the level sets $\{x \in C \,|\, f(x) \leq \alpha\}$ and $\{x \in C \,|\, f(x) < \alpha\}$ are convex for all scalars α.*

Proof: We leave this proof as an exercise.

Notice that, since the intersection of convex sets is convex, the set of points simultaneously satisfying m inequalities $f_1(x) \leq c_1, f_2(x) \leq c_2, \ldots, f_m(x) \leq c_m$ where each f_i is a convex function, defines a convex set. In particular, the polygonal region defined by a set of such inequalities when the f_i are affine is convex.

From this result, we can obtain an important fact about points at which a convex function attain a minimum.

Proposition 1.39 *Let C be a convex set and $f : C \longrightarrow \mathbb{R}$ a convex function. Then the set of points $M \subset C$ at which f attains its minumum is convex. Moreover, any relative minimum is an absolute minimum.*

Proof: If the function does not attain its minimum at any point of C, then the set of such points in empty, which is a convex set. So, suppose that the set of points at which the function attains its minimum is non-empty and let m be the minimal value attained by f. If $x, y \in M$ and $\lambda \in [0,1]$ then certainly $(1-\lambda)x + \lambda y \in C$ and so

$$m \leq f(\,(1-\lambda)\,x + \lambda\,y)\,) \;\leq\; (1-\lambda)\,f(x) + \lambda\,f(y) \;=\; m\,,$$

and so the point $(1-\lambda)x + \lambda y \in M$. Hence M, the set of minimal points, is convex.

Now, suppose that $x^\star \in C$ is a relative minimum point of f, but that there is another point $\hat{x} \in C$ such that $f(\hat{x}) < f(x^\star)$. On the line $(1-\lambda)\hat{x} + \lambda x^\star\,, 0 < \lambda < 1$, we have

$$f((1-\lambda)\,\hat{x} + \lambda\,x^\star) \;\leq\; (1-\lambda)\,f(\hat{x}) + \lambda\,f(x^\star) \;<\; f(x^\star)\,,$$

contradicting the fact that $x^\star$ is a relative minimum point. □

Again, the example of the simple parabola, shows that the set M may well contain only a single point, i.e., it may well be that the minimum point is unique. We can guarantee that this is the case for an important class of convex functions.

Definition 1.40 *A real-valued function f, defined on a convex set C is said to be* **strictly convex** *provided, for all $x, y \in C, x \neq y$ and $\lambda \in (0,1)$, we have*

$$f(\,(1-\lambda)\,x + \lambda\,y)\,) \;<\; (1-\lambda)\,f(x) + \lambda\,f(y)\,.$$

Proposition 1.41 *If C is a convex set and $f : C \longrightarrow \mathbb{R}$ is a strictly convex function then f attains its minimum at, at most, one pont.*

Proof: Suppose that the set of minimal points M is not empty and contains two distinct points x and y. Then, for any $0 < \lambda < 1$, since M is convex, we have $(1-\lambda)x + \lambda y \in M$. But f is *strictly* convex. Hence

$$m \;=\; f(\,(1-\lambda)\,x + \lambda\,y\,) \;<\; (1-\lambda)\,f(x) + \lambda\,f(y) \;=\; m\,,$$

which is a contradiction. $\square$

If a function is *differentiable* then, as in the case in elementary calculus, we can give characterizations of convex functions using derivatives. If f is a continuously differentiable function defined on an open convex set $C \subset \mathbb{R}^n$ then we denote its gradient at $x \in C$, as usual, by $\nabla f(x)$. The **excess function**

$$E(x,y) \;:=\; f(y) - f(x) - \langle \nabla f(x), y - x \rangle$$

is a measure of the discrepancy between the value of f at the point y and the value of the tangent approximation at x to f at the point y. This is illustrated in the next figure.

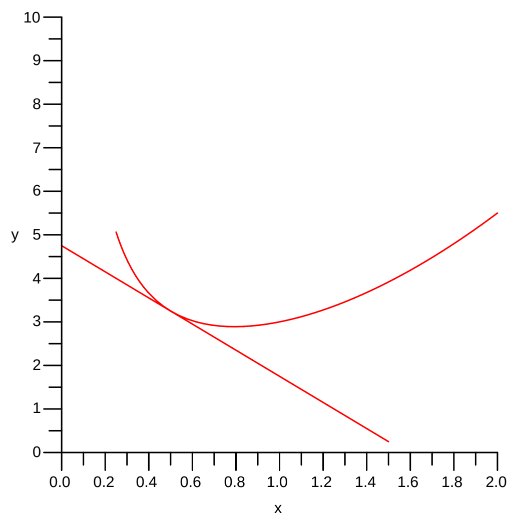

Figure 19: The Tangent Approximation

Now we introduce the notion of a monotone derivative

Definition 1.42 *The map $x \mapsto \nabla f(x)$ is said to be* **monotone** *on C provided*

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0 \,,$$

for all $x, y \in C$*.*

We can now characterize convexity in terms of the function E and the monotonicity concept just introduced. However, before stating and proving the next theorem, we need a lemma.

Lemma 1.43 *Let* f *be a real-valued, differentiable function defined on an open interval* $I \subset \mathbb{R}$*. Then if the first derivative* f' *is a non-decreasing function on* I*, the function* f *is convex on* I*.*

Proof: Choose $x, y \in I$ with $x < y$, and for any $\lambda \in [0, 1]$, define $z_\lambda := (1-\lambda)x + \lambda y$. By the Mean Value Theorem, there exist $u, v \in \mathbb{R}\,, x \leq v \leq z_\lambda \leq u \leq y$ such that

$$f(y) \,=\, f(z_\lambda) + (y - z_\lambda)\, f'(u)\,, \quad \text{and} \quad f(z_\lambda) \,=\, f(x) + (z_\lambda - x)\, f'(v)\,.$$

But, $y - z_\lambda = y - (1-\lambda)x - \lambda y = (1-\lambda)(y-x)$ and $z_\lambda - x = (1-\lambda)x + \lambda y - x = \lambda(y-x)$ and so the two expressions above may be rewritten as

$$f(y) \,=\, f(z_\lambda) + \lambda\, (y - x)\, f'(u)\,, \quad \text{and} \quad f(z_\lambda) \,=\, f(x) + \lambda\, (y - x)\, f'(v)\,.$$

Since, by choice, $v < u$, and since f' is non-decreasing, this latter equation yields

$$f(z_\lambda) \,\leq\, f(x) + \lambda\, (y - x)\, f'(u)\,.$$

Hence, multiplying this last inequality by $(1-\lambda)$ and the expression for $f(y)$ by $-\lambda$ and adding we get

$$(1-\lambda)\, f(z_\lambda) - \lambda\, f(y) \,\leq\, (1-\lambda)\, f(x) + \lambda(1-\lambda)(y-x)f'(u) - \lambda f(z_\lambda) - \lambda(1-\lambda)(y-x)f'(u)\,,$$

which can be rearranged to yield

$$(1-\lambda)\, f(z_\lambda) + \lambda\, f(z_\lambda) \,=\, f(z_\lambda) \,\leq\, (1-\lambda)\, f(x) + \lambda\, f(y)\,,$$

and this is just the condition for the convexity of f □

We can now prove a theorem which gives three different characterizations of convexity for continuously differentiable functions.

Theorem 1.44 *Let f be a continuously differentiable function defined on an open convex set $C \subset \mathbb{R}^n$. Then the following are equivalent:*

(a) $E(x,y) \geq 0$ for all $x, y \in C$;

(b) the map $x \mapsto \nabla f(x)$ is monotone in C;

(c) the function f is convex on C.

Proof: Suppose that (a) holds, i.e. $E(x,y) \geq 0$ on $C \times C$. Then we have both

$$f(y) - f(x) \;\geq\; \langle \nabla f(x), y - x \rangle \,,$$

and

$$f(x) - f(y) \;\geq\; \langle \nabla f(y), x - y \rangle \;=\; -\langle \nabla f(y), y - x \rangle \,.$$

Then, from the second inequality, $f(y) - f(x) = \langle \nabla f(y), x - y \rangle$, and so

$$\begin{aligned} \langle \nabla f(y) - \nabla f(x), y - x \rangle &= \langle \nabla f(y), y - x \rangle - \langle \nabla f(x), y - x \rangle \\ &\geq (f(y) - f(x)) \;-\; (f(y) - f(x)) \;\geq\; 0 \,. \end{aligned}$$

Hence, the map $x \mapsto \nabla f(x)$ is monotone in C.

Now suppose the map $x \mapsto \nabla f(x)$ is monotone in C, and choose $x, y \in C$. Define a function $\varphi : [0,1] \to \mathbb{R}$ by $\varphi(t) := f(x + t(y - x))$. We observe, first, that if φ is convex on $[0,1]$ then f is convex on C. To see this, let $u, v \in [0,1]$ be arbitrary. On the one hand,

$$\varphi((1-\lambda)u + \lambda v) \;=\; f(x + [(1-\lambda)u + \lambda v](y - x)) \;=\; f(\,(1 - [(1-\lambda)u + \lambda v])\,x + ((1-\lambda)u + \lambda v)\,y)\,,$$

while, on the other hand,

$$\varphi((1-\lambda)u + \lambda v) \;\leq\; (1-\lambda)\, f(x + u(y - x)) + f(x + v(y - x))\,.$$

Setting $u = 0$ and $v = 1$ in the above expressions yields

$$f((1-\lambda)\,x) + \lambda\, y) \;\leq\; (1-\lambda)\, f(x) + \lambda\, f(y)\,.$$

so the convexity of φ on $[0,1]$ implies the convexity of f on C.

Now, choose any α, β, $0 \leq \alpha < \beta \leq 1$. Then

$$\varphi'(\beta) - \varphi'(\alpha) \;=\; \langle (\nabla f(x + \beta(y - x)) - \nabla f(x + \alpha(y - x))\,,\, y - x\rangle\,.$$

Setting $u := x + \alpha(y - x)$ and $v := x + \beta(y - x)$[5] we have $v - u = (\beta - \alpha)(y - x)$ and so

$$\varphi'(\beta) - \varphi'(\alpha) \;=\; \langle (\nabla f(v) - \nabla f(u), v - u\rangle \;\geq\; 0\,.$$

Hence φ' is non-decreasing, so that the function φ is convex.

Finally, if f is convex on C, then, for fixed $x, y \in C$ define

$$h(\lambda) \;:=\; (1 - \lambda)\, f(x) + \lambda\, f(y) - f((1 - \lambda)\, x + \lambda\, y)\,.$$

Then $\lambda \mapsto h(\lambda)$ is a non-negative, differentiable function on $[0, 1]$ and attains its minimum at $\lambda = 0$. Therefore $0 \leq h'(0) = E(x, y)$, and the proof is complete. □

As an immediate corollary, we have

Corollary 1.45 *Let f be a continuously differentiable convex function defined on a convex set C. If there is a point $x^\star \in C$ such that, for all $y \in C$, $\langle \nabla\, f(x^\star), y - x^\star\rangle \geq 0$, then $x^\star$ is an absolute minimum point of f over C.*

Proof: By the preceeding theorem, the convexity of f implies that

$$f(y) - f(x^\star) \;\geq\; \langle \nabla\, f(x^\star), y - x^\star\rangle\,,$$

and so, by hypothesis,

$$f(y) \;\geq\; f(x^\star) + \langle \nabla\, f(x^\star), y - x^\star\rangle \;\geq\; f(x^\star)\,.$$

□

Finally, let us recall some facts from multi-variable calculus. Suppose that $D \subset \mathbb{R}^n$ is open and that the function $f : D \longrightarrow \mathbb{R}$ has continuous second partial derivatives in D. Then we define the **Hessian** of f to be the $n \times n$ matrix of second partial derivatives, denoted $\nabla^2 f(x)$ or $F(x)$:

$$\nabla^2 f(x) \;:=\; \left[\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right]\,.$$

[5] Note that u and v are convex combinations of points in the convex set C and so $u, v \in C$.

Since

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_i \partial x_i},$$

it is clear that the Hessian is a symmetric matrix.

A result of great usefulness in analysis is the result known as **Taylor's Theorem** or sometimes **generalized mean value theorems**. For a function f which is continuously differentiable on its domain $D \subset \mathbb{R}^n$, which contains the line segment joining two points $x_1, x_2 in D$ the theorem asserts the existence of a number $\theta \in [0,1]$ such that

$$f(x_2) = f(x_1) + \langle \nabla f(\theta x_1 + (1-\theta) x_2), x_2 - x_1 \rangle .$$

Moreover, if f is twice continuously differentiable then there is such a number $\theta \in [0,1]$ such that

$$f(x_2) = f(x_1) + \langle \nabla f(x_1), x_2 - x_1 \rangle + \frac{1}{2} \left\langle x_2 - x_1, \nabla^2 f(\theta x_1 + (1-\theta) x_2)(x_2 - x_1) \right\rangle .$$

Now, the Hessian that appears in the above formula is a symmetric matrix, and for such matrices we have the following definition.

Definition 1.46 *An $n \times n$ symmetric matrix, Q, is said to be* **positive semi-definite** *provided, for all $x \in \mathbb{R}^n$,*

$$\langle x, Qx \rangle \geq 0 .$$

The matrix Q is said to be **positive definite** *provided for all $x \in \mathbb{R}^n, x \neq 0$,*

$$\langle x, Qx \rangle > 0 .$$

We emphasize that the notions of positive definite and positive semi-definite are defined *only* for symmetric matrices.

Now, for twice continuously differentiable functions there is another ("second order") characterization of convexity.

Proposition 1.47 *Let $D \subset \mathbb{R}^n$ be an open convex set and let $f : D \longrightarrow \mathbb{R}$ be twice continuously differentiable in D. Then f is convex if and only if the Hessian matrix of f is positive semi-definite throughout D.*

Proof: By Taylor's Theorem we have

$$f(y) \;=\; f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2}\left\langle y - x, \nabla^2 f(x + \lambda\,(y - x))(y - x) \right\rangle \,,$$

for some $\lambda \in [0, 1]$. Clearly, if the Hessian is positive semi-definite, we have

$$f(y) \;\geq\; f(x) + \langle \nabla f(x), y - x \rangle \,,$$

which in view of the definition of the excess function, means that $E(x, y) \geq 0$ which implies that f is convex on D.

Conversely, suppose that the Hessian is *not* positive semi-definite at some point $x \in D$. Then, by the continuity of the Hessian, there is a $y \in D$ so that, for all $\lambda \in [0, 1]$,

$$\langle y - x, \nabla^2 f(x + \lambda(y - x))\,(y - x) \rangle \;<\; 0 \,,$$

which, in light of the second order Taylor expansion implies that $E(x, y) < 0$ and so f cannot be convex. $\square$