Math 4025

Primer on Convex Sets and Functions Part I

1 Closed convex sets

Definition 1.1. *Suppose $C \subseteq \mathbb{R}^n$ is some set. Then we say*

(a) *C is closed if whenever $\{\mathbf{x}_k\} \subseteq C$ satisfies $\mathbf{x}_k \to \bar{\mathbf{x}}$, we have $\bar{\mathbf{x}} \in C$;*

(b) *C is convex provided for all $\mathbf{x}_0 \in C$, $\mathbf{x}_1 \in C$, and $0 \le \lambda \le 1$, one has*

$$\mathbf{x}_\lambda := (1-\lambda)\mathbf{x}_0 + \lambda \mathbf{x}_1 \in C; \tag{1}$$

(c) *C is bounded if there exists $r > 0$ so that $\|\mathbf{x}\| \le r \; \forall \mathbf{x} \in C$.*

(d) *C is a cone if for all $\mathbf{x} \in C$ and $r > 0$, we have $r\mathbf{x} \in C$.*

These words are used in various combinations as a general description of a set, as the Figure 1 demonstrates.

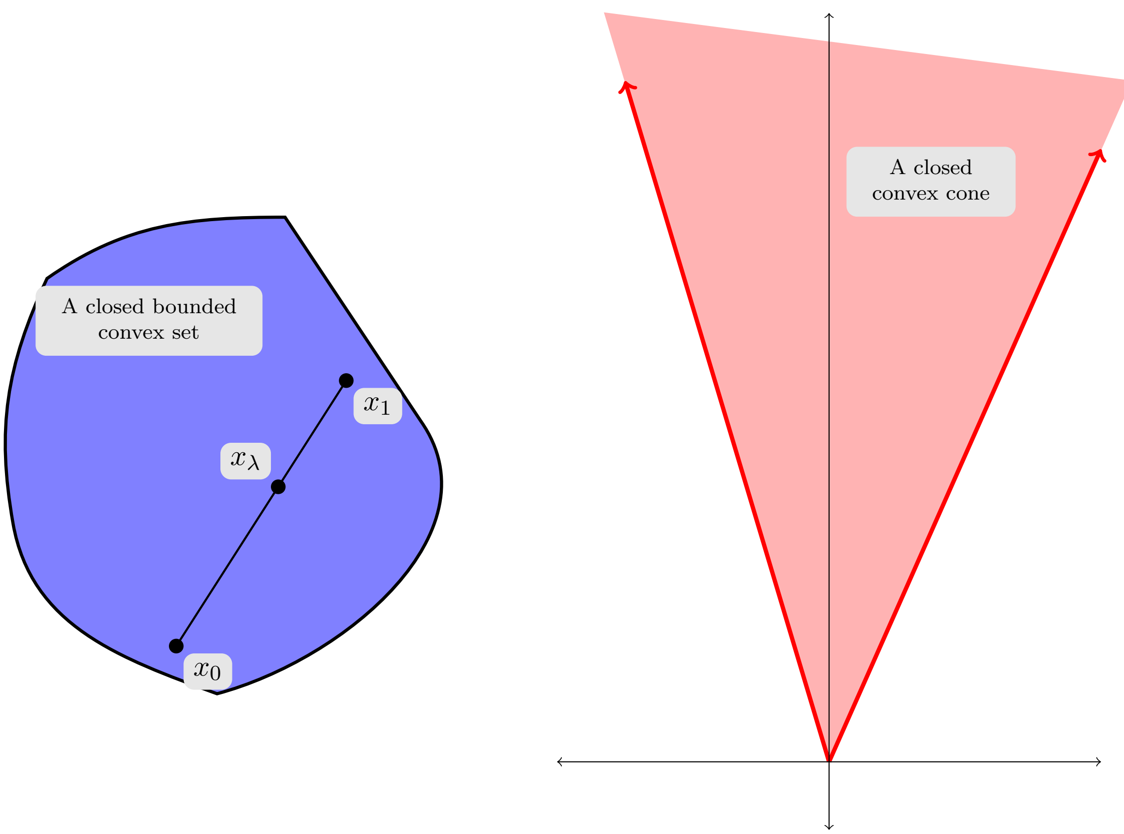

Figure 1: A closed convex set and cone

Exercise 1.1. *Suppose $\{C_\alpha\}_{\alpha \in A}$ is any collection of subsets of $\mathbb{R}^n$. Show that if each C_α is closed (resp. convex, a cone), then $\bigcap_{\alpha \in A} C_\alpha$ is closed (resp. convex, a cone).*

Convex sets are relatively simple to picture in 2 dimensions.

Exercise 1.2. *Sketch the following subsets of* $\mathbb{R}^2 = \left\{\begin{pmatrix} x \\ y \end{pmatrix} : x \in \mathbb{R},\ y \in \mathbb{R}\right\}$*, and determine if they are convex and/or closed and/or bounded. (Here we are writing the components of a vector in* $\mathbb{R}^2$ *with the more familiar variables* x *and* y *rather than* x_1 *and* x_2*).*

(a) $\left\{\begin{pmatrix} x \\ y \end{pmatrix} : |x| + |y| \le 1\right\}$ **(b)** $\left\{\begin{pmatrix} x \\ y \end{pmatrix} : |x|^2 + |y|^2 \le 4\right\}$ **(c)** $\left\{\begin{pmatrix} x \\ y \end{pmatrix} : \max\{|x|, |y|\} \le 2\right\}$

(d) $\left\{\begin{pmatrix} x \\ y \end{pmatrix} : y < -3x + 4\right\}$ **(e)** $\left\{\begin{pmatrix} x \\ y \end{pmatrix} : y \ge x^2 - 5x - 6\right\}$ **(f)** $\left\{\begin{pmatrix} x \\ y \end{pmatrix} : x^2 + y^2 = 1\right\}$

(g) $\left\{\begin{pmatrix} x \\ y \end{pmatrix} : y \ge e^x\right\}$ **(h)** $\left\{\begin{pmatrix} x \\ y \end{pmatrix} : y \le \ln(x),\ x > 0\right\}$ **(i)** $\left\{\begin{pmatrix} x \\ y \end{pmatrix} : 1 \le x \le 4,\ -1 \le y < 5\right\}$

(j) $\left\{\begin{pmatrix} x \\ y \end{pmatrix} : y < x^2 - 5x - 6\right\}$ **(k)** $\left\{\begin{pmatrix} x \\ y \end{pmatrix} : y \ge -2x - 2\right\}$ **(l)** $\left\{\begin{pmatrix} x \\ y \end{pmatrix} : -2 \le y \le x + 5\right\}$

(m) $\left\{\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \end{pmatrix}\right\}$ **(n)** $\left\{\begin{pmatrix} x \\ 0 \end{pmatrix} : x \ge 0\right\} \cup \left\{\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right\}$ **(o)** $\left\{\begin{pmatrix} x \\ 1 \end{pmatrix} : -1 \le x \le 1\right\} \cup \left\{\begin{pmatrix} 0 \\ -1 \end{pmatrix}\right\}$

Definition 1.2. *For any set* $E \subseteq \mathbb{R}^n$*, the closed convex hull of* E *is denoted by* $\overline{\text{co}}\,(E)$*, and is defined as the smallest closed convex set that contains* E*. In other words,* $\overline{\text{co}}\,(E)$ *satisfies the two properties:*

(i) $\overline{\text{co}}\,(E)$ *is closed and convex with* $E \subseteq \overline{\text{co}}\,(E)$*, and*

(ii) If S *is any closed convex set containing* E*, then* $\overline{\text{co}}\,(E) \subseteq S$*.*

Exercise 1.3. *(a) Find* $\overline{\text{co}}\,E$ *for each case where* E *is the set in Exercise 1.2*(f),(m),(n), *and* (o)*.*

(b) Show that for any set C*, we have*

$$\overline{\text{co}}\,E \;=\; \bigcap\{C :\ C \text{ is closed and convex and } E \subseteq C\} \tag{2}$$

If $\mathbf{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$, the hyperplane (resp. halfspaces) with *normal vector* $\mathbf{a}$ and level b is defined by

$$\begin{aligned} \mathbf{H}_{\mathbf{a},b} &:= \{\mathbf{x} : \langle \mathbf{x}, \mathbf{a}\rangle = b\} \\ \mathbf{H}^+_{\mathbf{a},b} &:= \{\mathbf{x} : \langle \mathbf{x}, \mathbf{a}\rangle \le b\} \\ \mathbf{H}^-_{\mathbf{a},b} &:= \{\mathbf{x} : \langle \mathbf{x}, \mathbf{a}\rangle \ge b\}. \end{aligned}$$

Exercise 1.4. *(a) Which set(s) in Exercise 1.2 equal* $\mathbf{H}^+_{\mathbf{a},b}$ *for some* $\mathbf{a} \in \mathbb{R}^2$ *and* $b \in \mathbb{R}$*? Identify* $\mathbf{a}$ *and* b*.*

(b) Show every hyperspace and halfspace is closed, convex, and not bounded.

(c) Show that $\mathbf{H}^-_{\mathbf{a},b} = \mathbf{H}^+_{-\mathbf{a},-b}$ *for any* $\mathbf{a} \in \mathbb{R}^n$*,* $b \in \mathbb{R}$*.*

(d) Show for any $\mathbf{a} \in \mathbb{R}^n$*,* $b \in \mathbb{R}$ *that* $\text{bdry}(\mathbf{H}^-_{\mathbf{a},b}) = \text{bdry}(\mathbf{H}^+_{\mathbf{a},b}) = \mathbf{H}_{\mathbf{a},b}$*.*

(e) If a set $E \subseteq \mathbb{R}^n$ *is such that* $E \subseteq H$*, where* $H = \mathbf{H}_{\mathbf{a},b}$ *or* $H = \mathbf{H}^\pm_{\mathbf{a},b}$ *for some* $\mathbf{a} \in \mathbb{R}^n$ *and* $b \in \mathbb{R}$*, show that* $\overline{\text{co}}\,E \subseteq H$*.*

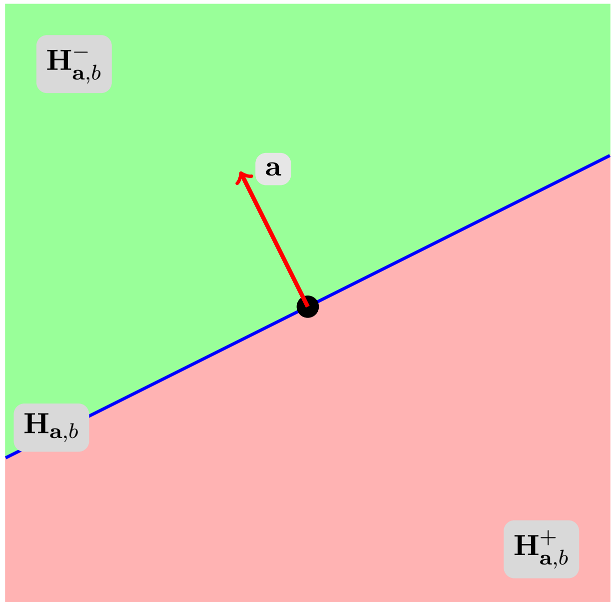

Figure 2: Two half-spaces with the hyperplane between them

(f) Find normal vectors $\mathbf{a}_1$, $\mathbf{a}_2 \in \mathbb{R}^2$ and levels b_1, $b_2 \in \mathbb{R}$ so that $\mathbf{H}^-_{\mathbf{a}_1,b_1} \cap \mathbf{H}^-_{\mathbf{a}_2,b_2} = \emptyset$.

(g) Under what general conditions on normal vectors $\mathbf{a}_1$, $\mathbf{a}_2 \in \mathbb{R}^n$ and levels b_1, $b_2 \in \mathbb{R}$ can it be that $\mathbf{H}^-_{\mathbf{a}_1,b_1} \cap \mathbf{H}^-_{\mathbf{a}_2,b_2} = \emptyset$? What about $\mathbf{H}^-_{\mathbf{a}_1,b_1} \subseteq \mathbf{H}^-_{\mathbf{a}_2,b_2}$?

We will show $\overline{\text{co}}\, E$ is the intersection of all half-spaces that contain E.

Suppose $\mathbf{a}_1, \dots, \mathbf{a}_m$ are m vectors in $\mathbb{R}^n$ and $b_1, \dots, b_m$ are real numbers. Let $\mathbf{A} \in \mathcal{M}_{m\times n}$ (:= the set of all $m \times n$ matrices) and $\mathbf{b} \in \mathbb{R}^m$ be given by

$$\mathbf{a}_i = \begin{pmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{in} \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} \mathbf{a}_1^\top \\ \vdots \\ \mathbf{a}_m^\top \end{pmatrix} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ a_{21} & \dots & a_{2n} \\ \vdots & \dots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}, \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

A *Feasible* set $F_{\mathbf{A},\mathbf{b}}$ is defined as

$$F_{\mathbf{A},\mathbf{b}} := \left\{ \mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} \le \mathbf{b} \right\} = \bigcap_{i=1}^m \mathbf{H}^-_{\mathbf{a}_i,b_i}$$

Exercise 1.5. *Show that if $E \subseteq \mathbf{H}^-_{\mathbf{a}_i,b_i}$ for each $i = 1, \dots, m$, then $\overline{\text{co}}\, E \subseteq F_{\mathbf{A},\mathbf{b}}$. In fact, show that more generally the inclusion "$\subseteq$" in (2) always holds.*

Definition 1.3. *Suppose $C \subseteq \mathbb{R}^n$ is closed and convex, and let $\mathbf{x} \in C$. A vector $\mathbf{n} \in \mathbb{R}^n$ is called a* normal vector *at $\mathbf{x}$ provided*

$$\langle \mathbf{n}, \mathbf{y} - \mathbf{x} \rangle \leq 0 \quad \forall \mathbf{y} \in C.$$

The set of all normal vectors at $\mathbf{x}$ is denoted by $N_C(\mathbf{x})$.

It is clear that we always have $\mathbf{0} \in N_C(\mathbf{x})$ whenever $\mathbf{x} \in C$, since $\langle \mathbf{0}, z \rangle = 0 \, \forall \mathbf{z} \in \mathbb{R}^n$ and hence in particular when $\mathbf{z} = \mathbf{y} - \mathbf{x}$ with $\mathbf{y} \in C$.

Exercise 1.6. *Show that $N_C(\mathbf{x})$ is a closed convex cone.*

The next proposition says that nontrivial normal cones can occur only at the boundary points of C.

Proposition 1.1. *If $\mathbf{x} \in \operatorname{int}(C)$, then $N_C(\mathbf{x}) = \{\mathbf{0}\}$.*

Proof. Suppose $\mathbf{x} \in C$. Since $\mathbf{x}$ belongs the open set $\operatorname{int}(C)$, there exists $r_0 > 0$ so that $\mathbb{B}(\mathbf{x}; r_0) \subseteq C$. Suppose there exists $\mathbf{n} \in N_C(\mathbf{x})$ with $\mathbf{n} \neq \mathbf{0}$. Let $r := \|\mathbf{n}\| > 0$ and set $\mathbf{y} := \mathbf{x} + \frac{r_0}{2\,r}\mathbf{n}$. Then $\|\mathbf{y} - \mathbf{x}\| \frac{r_0}{2\,r}\|\mathbf{n}\| = \frac{r_0}{2}$. Hence $\mathbf{y} \in \mathbb{B}(\mathbf{x}; r_0) \subseteq C$, and

$$\langle \mathbf{n}, \mathbf{y} - \mathbf{x} \rangle = \langle \mathbf{n}, \frac{r_0}{2\,r}\mathbf{n} \rangle = \frac{r_0}{2\,r}\langle \mathbf{n}, \mathbf{n} \rangle = \frac{r_0\, r}{2} > 0.$$

□

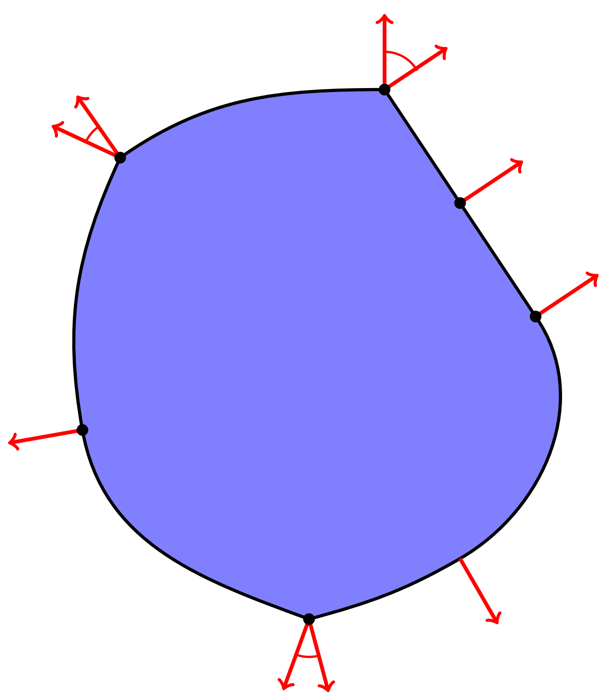

Figure 3: Some normal vectors

Exercise 1.7. *Let $\mathbf{a} \in \mathbb{R}^n$, $b \in \mathbb{R}$, and set $C = \mathbf{H}_{\mathbf{a},b}$ and $C^{\pm} := \mathbf{H}^{\pm}_{\mathbf{a},b}$.*

(a) Show that $N_C(\mathbf{x}) = \{r\mathbf{a} : r \in \mathbb{R}\}$ for all $\mathbf{x} \in C$.

(b) Show that $N_{C^+}(\mathbf{x}) = \{r\mathbf{a} : r \geq 0\}$ *for all* $\mathbf{x} \in \mathrm{bdry} C$.

(c) Show that $N_{C^-}(\mathbf{x}) = \{r\mathbf{a} : r \leq 0\}$ *for all* $\mathbf{x} \in \mathrm{bdry} C$.

(d) For $\mathbf{x} \in \mathbf{H}_{\mathbf{a},b}$*, show that* $N_{\mathbf{H}^{\pm}_{\mathbf{a},b}}(\mathbf{x})$ *does not depend on the level* b.

(e) Suppose $\mathbf{a}_1 \neq \mathbf{a}_2$ *are linearly independent vectors, and* $\mathbf{x} \in C := \mathbf{H}^+_{\mathbf{a}_1,b_1} \cap \mathbf{H}^+_{\mathbf{a}_2,b_2}$ *for some levels* b_1, $b_2 \in \mathbb{R}$. *Calculate* $N_C(\mathbf{x})$. *(Hint: there are four different cases depending on which of the hyperplanes* $\mathbf{H}_{\mathbf{a}_1,b_1}$, $\mathbf{H}_{\mathbf{a}_2,b_2}$ $\mathbf{x}$ *belongs.*

Definition 1.4. *For any closed set* $C \subseteq \mathbb{R}^n$*, the* distance function $d_C(\cdot) : \mathbb{R}^n \to [0, \infty)$ *is defined by*

$$d_C(\mathbf{x}) = \inf\Big\{ \|\mathbf{y} - \mathbf{x}\| : \mathbf{y} \in C \Big\}.$$

The set of closest *elements* $\mathrm{proj}_C(\mathbf{x})$ *in* C *to* $\mathbf{x} \in \mathbb{R}^n$ *(also called the* projection *of* $\mathbf{x}$ *onto* C*) is defined as*

$$\mathrm{proj}_C(\mathbf{x}) = \Big\{ \mathbf{y} \in C : d_C(\mathbf{x}) = \|\mathbf{y} - \mathbf{x}\| \Big\}.$$

Proposition 1.2. *Suppose* $C \subseteq \mathbb{R}^n$ *is a nonempty closed set.*

(a) The distance function is Lipschitz continuous of rank 1 *on* $\mathbb{R}^n$*. This means*

$$\big|d_C(\mathbf{x}_1) - d_C(\mathbf{x}_2)\big| \leq \|\mathbf{x}_1 - \mathbf{x}_2\| \quad \forall \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n. \tag{3}$$

(b) For all $\mathbf{x} \in \mathbb{R}^n$*, the projection set* $\mathrm{proj}_C(\mathbf{x})$ *is nonempty.*

(c) If $\mathbf{x} \in C$ *or if* C *is convex, then* $\mathrm{proj}_C(\mathbf{x})$ *consists of a single element.*

Proof. (a) Suppose $\mathbf{x}_1$, $\mathbf{x}_2 \in C$, and let $\varepsilon > 0$. There exists $\mathbf{y} \in C$ so that $\|\mathbf{x}_2 - \mathbf{y}\| < d_C(\mathbf{x}_2) + \varepsilon$, or that $-d_C(\mathbf{x}_2) < -\|\mathbf{x}_2 - \mathbf{y}\| + \varepsilon$. Since $\mathbf{y} \in C$, we also have $d_C(x_1) \leq \|\mathbf{x}_1 - \mathbf{y}\|$. Thus adding these two inequalities gives

$$d_C(\mathbf{x}_1) - d_C(\mathbf{x}_2) \leq \|\mathbf{x}_1 - \mathbf{y}\| - \|\mathbf{x}_2 - \mathbf{y}\| + \varepsilon \leq \|\mathbf{x}_1 - \mathbf{x}_2\| + \varepsilon.$$

Letting $\varepsilon \to 0$, and then switching the roles of $\mathbf{x}_1$ and $\mathbf{x}_2$ yields (3).
(b) Let $\mathbf{x} \in \mathbb{R}^n$. For each $k \in \mathbb{N}$, there exists $\mathbf{x}_k \in C$ so the $\|\mathbf{x}_k - \mathbf{x}\| \leq d_C(\mathbf{x}) + \frac{1}{k}$. Since $\{\mathbf{x}_k\}$ is bounded, there exists a convergent subsequence of $\{\mathbf{x}_k\}$ which without loss of generality we take to be the entire sequence $\{\mathbf{x}_k\}$. Say $\mathbf{x}_k \to \mathbf{y}$, and since C is closed, we have $\mathbf{y} \in C$. Since the norm is a continuous function, we have $\|\mathbf{y} - \mathbf{x}\| = \lim_{k\to\infty} \|\mathbf{x}_k - \mathbf{x}\| = d_C(\mathbf{x})$, or that $\mathbf{y} \in \mathrm{proj}_C(\mathbf{x})$.
(c) If $\mathbf{x} \in C$, then $d_C(\mathbf{x}) = 0$ and clearly $\mathrm{proj}_C(\mathbf{x}) = \{\mathbf{x}\}$.

Let us note a general fact: if $\mathbf{z}_1$ and $\mathbf{z}_2$ are two distinct vectors with $\|z_1\| = \|z_2\| = r$, then $\big\|\frac{1}{2}(\mathbf{z}_1 + \mathbf{z}_2)\big\| < r$. Indeed,

$$\left\|\frac{1}{2}(\mathbf{z}_1 + \mathbf{z}_2)\right\|^2 = \frac{1}{4}\big\langle \mathbf{z}_1 + \mathbf{z}_2, \mathbf{z}_1 + \mathbf{z}_2 \big\rangle = \frac{1}{4}\Big[\|\mathbf{z}_1\|^2 + \|\mathbf{z}_2\|^2 + 2\langle \mathbf{z}_1, \mathbf{z}_2\rangle\Big] = \frac{r^2}{2}\big[1 + \cos(\theta)\big] < r^2,$$

where θ is the angle between $\mathbf{z}_1$ and $\mathbf{z}_2$. The final strict inequality above is justified since $\cos(\theta) = 1$ only when the there exists a positive constant ρ with $\mathbf{z}_1 = \rho \mathbf{z}_2$. But since $\mathbf{z}_1 \neq \mathbf{z}_2$ and have the same norm, this can't happen.

Now suppose C is convex and $\mathbf{y}_1$, $\mathbf{y}_2 \in \text{proj}_C(\mathbf{x})$, and so $\|\mathbf{y}_1 - \mathbf{x}\| = d_C(\mathbf{x}) = \|\mathbf{y}_1 - \mathbf{x}\|$. Since C is convex and $\mathbf{y}_1$, $\mathbf{y}_2 \in C$, we have $\frac{1}{2}\big(\mathbf{y}_1 + \mathbf{y}_2\big) \in C$. Therefore

$$\left\| \frac{1}{2}\big(\mathbf{y}_1 + \mathbf{y}_2\big) - \mathbf{x} \right\| \geq d_C(\mathbf{x}). \tag{4}$$

However, if $\mathbf{y}_1 \neq \mathbf{y}_2$, by the fact noted above (with $\mathbf{z}_1 := \mathbf{y}_1 - \mathbf{x}$, $\mathbf{z}_2 := y_2 - \mathbf{x}$ and $r := d_C(\mathbf{x})$), we have

$$\left\| \frac{1}{2}\big(\mathbf{y}_1 + \mathbf{y}_2\big) - \mathbf{x} \right\| = \left\| \frac{1}{2}\big[(\mathbf{y}_1 - \mathbf{x}) + (\mathbf{y}_2 - \mathbf{x})\big] \right\| < d_C(\mathbf{x}). \tag{5}$$

The inequalities in (4) and (5) are incompatible, and therefore we must have $\mathbf{y}_1 = \mathbf{y}_2$, or that $\text{proj}_(\mathbf{x})$ consists of a single element.

□

The next important result is called the Separation Theorem (between a point and a convex sets), since it says every point $\mathbf{x} \notin C$ can be separated from the convex set C by a hyperplane.

Theorem 1.1. *Suppose $C \subseteq \mathbb{R}^n$ nonempty, closed and convex, and $\mathbf{x} \notin C$. Then there exists $\mathbf{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$ so that $\mathbf{x} \in \text{int}(\mathbf{H}^-_{\mathbf{a},b})$ and $C \subseteq \text{int}(\mathbf{H}^+_{\mathbf{a},b})$. One can in fact take $\mathbf{a} = (\mathbf{x} - \mathbf{y})$ where $\text{proj}_C(\mathbf{x}) = \{\mathbf{y}\}$ and $b = \frac{1}{2}\big[\|\mathbf{x}\|^2 - \|\mathbf{y}\|^2\big]$.*

Proof. Suppose C is nonempty, closed and convex, and $\mathbf{x} \notin C$, and let $\mathbf{y}$ be the unique projection of $\mathbf{x}$ into S. Set $\mathbf{a} = \mathbf{x} - \mathbf{y} \neq \mathbf{0}_n$ and $b = \big\langle \mathbf{a}, \frac{\mathbf{x}+\mathbf{y}}{2} \big\rangle = \frac{1}{2}\big[\|\mathbf{x}\|^2 - \|\mathbf{y}\|^2\big]$. We use the simple fact that $\left\langle \mathbf{a}, \dfrac{\mathbf{x}-\mathbf{y}}{2} \right\rangle = \dfrac{1}{2}\langle \mathbf{a}, \mathbf{a} \rangle > 0$, and note that

$$\begin{aligned} \langle \mathbf{a}, \mathbf{y} \rangle &< \langle \mathbf{a}, \mathbf{y} \rangle + \left\langle \mathbf{a}, \frac{\mathbf{x}-\mathbf{y}}{2} \right\rangle = \left\langle \mathbf{a}, \frac{\mathbf{x}+\mathbf{y}}{2} \right\rangle = b \\ &< \left\langle \mathbf{a}, \frac{\mathbf{x}+\mathbf{y}}{2} \right\rangle + \left\langle \mathbf{a}, \frac{\mathbf{x}-\mathbf{y}}{2} \right\rangle = \langle \mathbf{a}, \mathbf{x} \rangle. \end{aligned}$$

Thus $\mathbf{y} \in \text{int}\mathbf{H}^+_{\mathbf{a},b}$ and $\mathbf{x} \in \text{int}\mathbf{H}^-_{\mathbf{a},b}$. Now let $\mathbf{z} \in C$, and since C is convex, we have $\mathbf{z}_\lambda := \lambda\mathbf{z} + (1-\lambda)\mathbf{y} = \mathbf{y} + \lambda(\mathbf{z} - \mathbf{y}) \in C$ for all $0 \leq \lambda \leq 1$. Recall $\mathbf{y} = \text{proj}_C(\mathbf{x})$, and thus for all $0 \leq \lambda \leq 1$

$$\begin{aligned} \big\|\mathbf{y} - \mathbf{x}\big\|^2 &= d_C(\mathbf{x})^2 \leq \big\|\mathbf{z}_\lambda - \mathbf{x}\big\|^2 \\ &= \big\|\mathbf{y} - \mathbf{x} + \lambda(\mathbf{z} - \mathbf{y})\big\|^2 \\ &= \big\|\mathbf{y} - \mathbf{x}\big\|^2 + 2\lambda\big\langle \mathbf{y} - \mathbf{x}, \mathbf{z} - \mathbf{y} \big\rangle + \lambda^2\big\|\mathbf{z} - \mathbf{y}\big\|^2 \end{aligned}$$

Rearranging terms and dividing by $\lambda > 0$ gives

$$\big\langle \mathbf{a}, \mathbf{z} - \mathbf{y} \big\rangle = \big\langle \mathbf{x} - \mathbf{y}, \mathbf{z} - \mathbf{y} \big\rangle \leq \frac{1}{2}\lambda\big\|\mathbf{z} - \mathbf{y}\big\|^2$$

Letting $\lambda \downarrow 0$ implies $\langle \mathbf{a}, \mathbf{z} \rangle \leq \langle \mathbf{a}, \mathbf{y} \rangle < b$, or that $\mathbf{z} \in \text{int}\mathbf{H}^+_{\mathbf{a},b}$. □

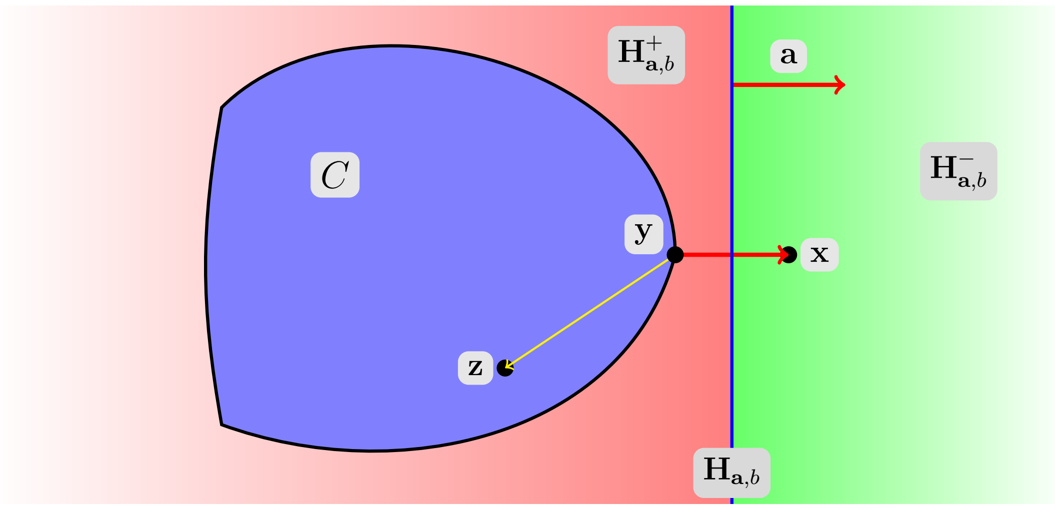

Figure 4: Separation

Corollary 1. *For any set $E \subseteq \mathbb{R}^n$, we have*

$$\overline{\mathrm{co}}\,(E) = \bigcap \Big\{ H : \; H \textit{ is a halfspace with } E \subseteq H \Big\}. \tag{6}$$

Proof. The inclusion "$\subseteq$" is clear, as noted in Exercise 1.5. As for the "$\supseteq$" inclusion, suppose $\mathbf{x} \notin \overline{\mathrm{co}}\,(E)$. Since $\overline{\mathrm{co}}\,(E)$ is closed and convex, by Theorem 1.1, there exists $\mathbf{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$ so that $E \subseteq \overline{\mathrm{co}}\,(E) \subseteq \mathrm{int}\big(H^+_{\mathbf{a},b}\big)$ and $\mathbf{x} \in \mathrm{int}\big(H^-_{\mathbf{a},b}\big)$. This shows $\mathbf{x}$ cannot belong to the right hand side of (6). $\square$

Exercise 1.8. *Here are some further variations of Theorem 1.1. Suppose C_1 and C_2 are both closed and convex and $C_1 \cap C_2 = \emptyset$.*

(a) Assume at least one of C_1, C_2 is also bounded. Show there exists $\mathbf{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$ so that $C_1 \in \mathrm{int}\big(\mathbf{H}^-_{\mathbf{a},b}\big)$ and $C_2 \subseteq \mathrm{int}\big(\mathbf{H}^+_{\mathbf{a},b}\big)$.

(b) Provide an example where part (a) is false if both C_1 and C_2 are unbounded.

(c) Show there exists $\mathbf{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$ so that $C_1 \in \mathbf{H}^-_{\mathbf{a},b}$ and $C_2 \subseteq \mathbf{H}^+_{\mathbf{a},b}$.

Remark 1.1. The hallmark of convexity is the dual way nearly every concept can be treated. Even the definition of a (closed) convex set has two possible approaches. The so-called *primal* way is the direct statement given in Definition 1.1 . The second *dual* way is contained in Corollary 1, where we can characterize the property that $\mathbf{x}$ does not belong to a closed convex set C by asserting the existence of a half space containing C and not $\mathbf{x}$.

2 Lower semicontinuous convex functions

The concepts that a set is closed and convex has an analogue for functions. We first define these concepts for functions that are finite-valued, and then later extend them to allow a function that "takes on" the value $+\infty$.

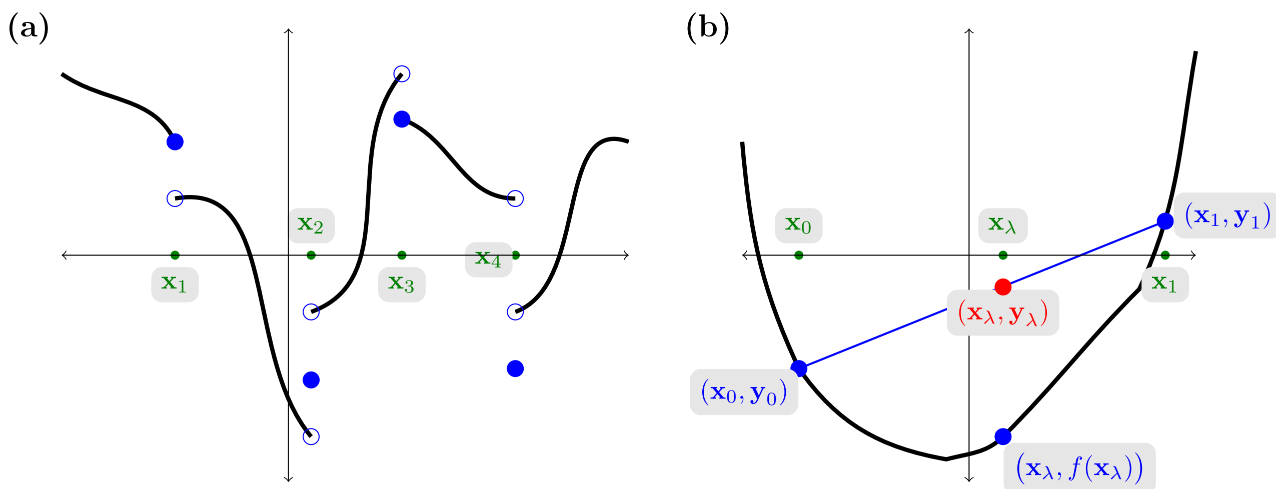

Figure 5: Illustration of lsc and convex functions

Definition 2.1. *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a function.*

(a) We say $f(\cdot)$ is lower semicontinuous *(lsc) at $\mathbf{x} \in \mathbb{R}^n$ provided for all $\varepsilon > 0$, there exists $\delta > 0$ so that*

$$\big\|\mathbf{y} - \mathbf{x}\big\| < \delta \quad \Rightarrow \quad f(\mathbf{y}) \geq f(\mathbf{x}) - \varepsilon. \tag{7}$$

If $f(\cdot)$ is lsc at each point of some set C, then $f(\cdot)$ is lsc on C.

(b) We say $f(\cdot)$ is convex *provided for all $\mathbf{x}_0$, $\mathbf{x}_1 \in \mathbb{R}^n$ and $0 \leq \lambda \leq 1$, one has*

$$f\big((1-\lambda)\mathbf{x}_0 + \lambda\mathbf{x}_1\big) \leq (1-\lambda) f\big(\mathbf{x}_0\big) + \lambda f\big(\mathbf{x}_1\big). \tag{8}$$

An illustration of the lower semicontinuous concept is given in Figure 5(a), where the function is not lsc at $\mathbf{x}_1$ and $\mathbf{x}_2$, but is lsc at $\mathbf{x}_3$ and $\mathbf{x}_4$. Figure 5(b) illustrates the condition in (8): If we set $x_\lambda := \big((1-\lambda)\mathbf{x}_0 + \lambda\mathbf{x}_1\big)$, $\mathbf{y}_0 := f(\mathbf{x}_0)$, $y_1 := f(\mathbf{x}_1)$, and $\mathbf{y}_\lambda := \big((1-\lambda)\mathbf{y}_0 + \lambda\mathbf{y}_1\big)$, then (8) can be more succinctly written as $f(\mathbf{x}_\lambda) \leq \mathbf{y}_\lambda$.

Exercise 2.1. *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a function, and define the* epigraph $\mathrm{epi}(f)$ *of $f(\cdot)$ by*

$$\mathrm{epi}(f) := \left\{ \begin{pmatrix} \mathbf{x} \\ r \end{pmatrix} \in \mathbb{R}^n \times \mathbb{R} : \; r \geq f(\mathbf{x}) \right\}. \tag{9}$$

See Figure 6.

(a) Show that $f(\cdot)$ is lsc on $\mathbb{R}^n$ if and only if $\mathrm{epi}(f)$ is a closed subset of $\mathbb{R}^{n+1}$.

(b) Show that $f(\cdot)$ is convex on $\mathbb{R}^n$ if and only if $\mathrm{epi}(f)$ is a convex subset of $\mathbb{R}^{n+1}$.

It turns out that in optimization it is very convenient to extend the notion of a function to allow it to take $\pm\infty$ values, but one must be careful since $\pm\infty$ are not numbers and do not obey the usual rules of arithmetic. Nonetheless, it is easy to extend the notions in the usual way and have an effective arithmetic.

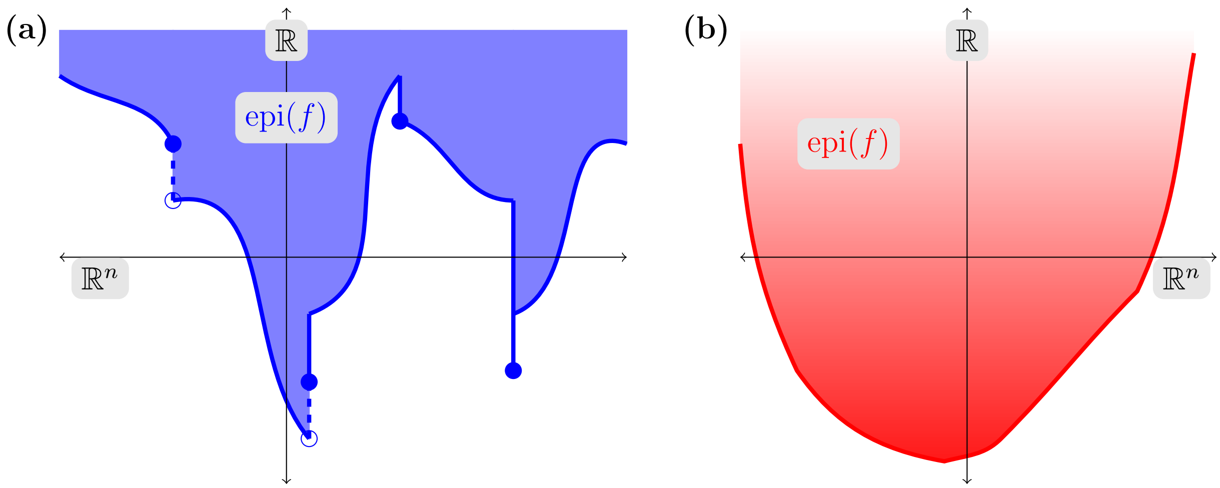

Figure 6: Epigraphs

Our point of view is tilted toward minimization rather than maximization, and that affects the conventions. The conventions are:

$$\infty - \infty = \infty; \quad b \pm \infty = \pm\infty \ \forall b \in \mathbb{R}; \quad r \cdot (\pm\infty) = \pm\infty \ \forall r > 0; \text{ and } -\infty < b < +\infty \ \forall b \in \mathbb{R}. \tag{10}$$

The multiplication $0 \cdot \pm\infty$ and remains undefined. So, using these conventions, we can assume the function in Definition 2.1 is allowed to take on $\pm\infty$ and use the conditions in (7) and (8). Another way, which is employed here, is to use the results in Exercise 2.1 as the definitions.

Definition 2.2. *Suppose $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a function.*

(a) We say $f(\cdot)$ is lower semicontinuous *(lsc) on a closed set C if* $\mathrm{epi}(f) \cap \{\mathbb{R}^n \times \mathbb{R}\}$ *is closed.*

(b) We say $f(\cdot)$ is convex *provided* $\mathrm{epi}(f) \cap \{\mathbb{R}^n \times \mathbb{R}\}$ *is a convex set in $\mathbb{R}^{n+1}$.*

Since a function may no longer be finite-valued, it is convenient to introduce notation for the (effective) domain $\mathrm{dom}(f) := \{\mathbf{x} \in R^n : \ f(\mathbf{x}) < +\infty\}$. In this way, we need only consider functions defined on $\mathbb{R}^n$ but keep track of its most important values through $\mathrm{dom}(f)$. An important class of functions consists of the those functions that are lsc with $\mathrm{dom}(f) \neq \emptyset$, and we denote this class by $\mathcal{F}$; the class that in addition requires its members to be convex is denoted by $\mathcal{F}_c$.

A philosophical viewpoint of Convex Analysis is to treat sets and functions in an equal manner. Given a function $f \in \mathcal{F}$, we often invoke $\mathrm{epi}(f)$ to discuss properties of $f(\cdot)$. Conversely, if $C \subseteq \mathbb{R}^n$, the indicator function $I_C : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ defined by

$$I_C(\mathbf{x}) = \begin{cases} 0 & \text{if } \mathbf{x} \in C \\ +\infty & \text{if } \mathbf{x} \notin C \end{cases}$$

is associated to C. If C is nonempty and closed, then $I_C(\cdot) \in \mathcal{F}$, and if in addition C is convex, then $I_C(\cdot) \in \mathcal{F}_c$.

Exercise 2.2. *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is continuous and $C \subseteq \mathbb{R}^n$ is nonempty. Consider the optimization problem*

$$\inf f(\mathbf{x}) \quad \textit{subject to } \mathbf{x} \in C. \tag{$\mathcal{P}$}$$

Let $\tilde{f}(\mathbf{x}) := f(\mathbf{x}) + I_C(\mathbf{x})$. Consider the optimization problem

$$\inf \tilde{f}(\mathbf{x}) \quad \textit{subject to } \mathbf{x} \in \mathbb{R}^n. \tag{$\tilde{\mathcal{P}}$}$$

Show that problems ($\mathcal{P}$) and ($\tilde{\mathcal{P}}$) are equivalent in the sense that a point $\mathbf{x} \in \mathbb{R}^n$ is an optimal solution of one of the problems if and only if it a solution of the other.

The cases of most import is where $f(\cdot)$ is lsc and C is closed, in which case $\tilde{f}(\cdot) \in \mathcal{F}$, and where C is convex and $f(\cdot)$ are convex with $\mathrm{dom}(f) \cap C \neq \emptyset$, in which case $\tilde{f}(\cdot) \in \mathcal{F}_c$).

3 Calculus

Differentiable properties of functions $f : \mathbb{R}^n \to \mathbb{R}$ were reviewed earlier. We now see how many of these concepts have systematic counterparts for convex functions.

Theorem 3.1. *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is lsc. Then $f(\cdot)$ is convex if and only if for every $\mathbf{x}_0$, $\mathbf{x}_1 \in \mathbb{R}^n$, and $0 < \lambda < 1$, we have*

$$\frac{f(\mathbf{x}_\lambda) - f(\mathbf{x}_0)}{\lambda} \leq f(\mathbf{x}_1) - f(\mathbf{x}_0) \leq \frac{f(\mathbf{x}_1) - f(\mathbf{x}_\lambda)}{1 - \lambda} \tag{11}$$

where $\mathbf{x}_\lambda := (1 - \lambda)\mathbf{x}_0 + \lambda \mathbf{x}_1$.

Proof. Assume first $f(\cdot)$ is convex, and let $\mathbf{x}_0$, $\mathbf{x}_1 \in \mathbb{R}^n$ and $0 < \lambda < 1$. Then (8) implies

$$f(\mathbf{x}_\lambda) = f\big((1 - \lambda)\mathbf{x}_0 + \lambda \mathbf{x}_1\big) \leq (1 - \lambda) f\big(\mathbf{x}_0\big) + \lambda f\big(\mathbf{x}_1\big). \tag{12}$$

Assume first that $\mathbf{x}_0$, $\mathbf{x}_1 \in \mathrm{dom}(f)$. The right-hand side of (12) equals $f\big(\mathbf{x}_0\big) + \lambda\big(f\big(\mathbf{x}_1\big) - f\big(\mathbf{x}_0\big)\big)$, and then rearranging terms in (12) and dividing by λ gives the first inequality in (11). The right-hand side of (12) also equals $f\big(\mathbf{x}_1\big) + (1 - \lambda)\big(f\big(\mathbf{x}_0\big) - f\big(\mathbf{x}_1\big)\big)$, and again rearranging terms in (12) and dividing by $(1 - \lambda)$ gives the second inequality in (11).

Conversely, suppose $f(\cdot)$ is not convex, and so there exists $\mathbf{x}_0$, $\mathbf{x}_1 \in \mathbb{R}^n$ and $0 \leq \lambda \leq 1$ so that $f(\mathbf{x}_\lambda) > (1-\lambda) f\big(\mathbf{x}_0\big) + \lambda f\big(\mathbf{x}_1\big)$. Then arguing exactly as above, we can deduce the reverse strict inequalities in (11) hold.

Finally, if one, two, or all three of $\mathbf{x}_0$, $\mathbf{x}_\lambda$, or $\mathbf{x}_1$ do not belong to $\mathrm{dom}(f)$, then (12) holds according to the conventions introduced in (10). $\square$

Remark 3.1. The geometrical insight into understanding where (11) comes from is to consider the three secant lines with slopes (here, $n = 1$)

$$m_1 = \frac{f(\mathbf{x}_\lambda) - f(\mathbf{x}_0)}{\lambda(\mathbf{x}_1 - \mathbf{x}_0)}; \quad m_2 = \frac{f(\mathbf{x}_1) - f(\mathbf{x}_0)}{\mathbf{x}_1 - \mathbf{x}_0}; \quad \text{and} \quad m_3 = \frac{f(\mathbf{x}_1) - f(\mathbf{x}_\lambda)}{(1 - \lambda)(\mathbf{x}_1 - \mathbf{x}_0)}. \tag{13}$$

Note that $\mathbf{x}_\lambda - \mathbf{x}_0 = \lambda(\mathbf{x}_1 - \mathbf{x}_0)$ and $\mathbf{x}_1 - \mathbf{x}_\lambda = (1 - \lambda)(\mathbf{x}_1 - \mathbf{x}_0)$, so the three quotients really are the slopes of secant lines as depicted in Figure 7 (where for definiteness we took $\mathbf{x}_0 < \mathbf{x}_1$). This figure suggests if $f(\cdot)$ is convex, then $m_1 \leq m_2 \leq m_3$. The only difference between (11) and (13) is division by $\mathbf{x}_1 - \mathbf{x}_0 > 0$.

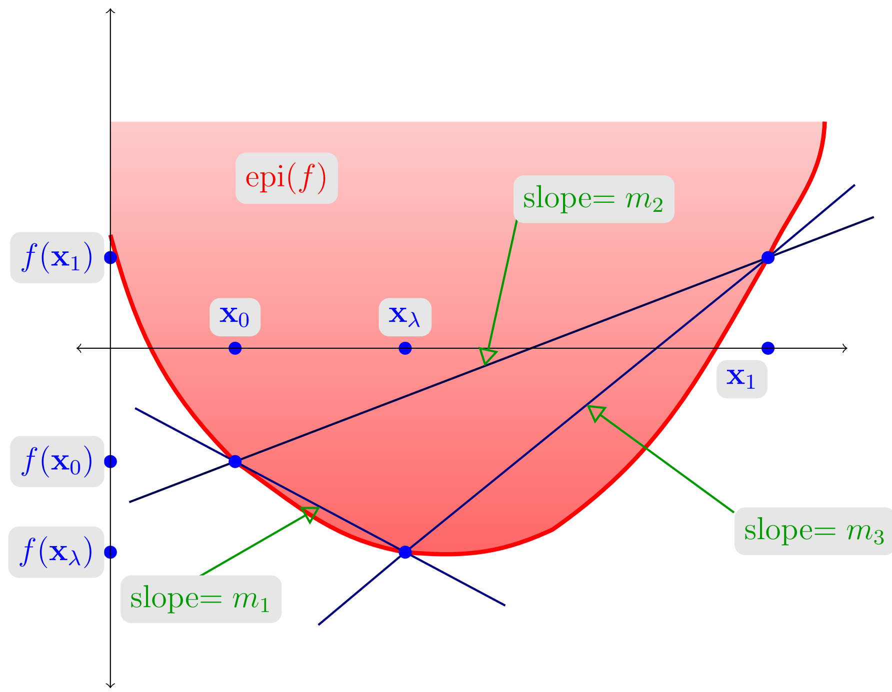

Figure 7: Three point property

We'll pursue several consequences that accrue from Theorem 3.1.

Suppose $f \in \mathcal{F}$, $\mathbf{x} \in \mathrm{dom}(f)$, and $v \in \mathbb{R}^n$. The directional derivative $f'(\mathbf{x}; \mathbf{v})$ of $f(\cdot)$ at $\mathbf{x}$ in direction $\mathbf{v}$ is defined by

$$f'(\mathbf{x}; \mathbf{v}) := \lim_{h \to 0^+} \frac{f(\mathbf{x} + h\,\mathbf{v}) - f(\mathbf{x})}{h} \tag{14}$$

provided the limit exists (it could be $\pm\infty$). If $f(\cdot)$ is differentiable at $\mathbf{x}$ with gradient equal to $\nabla f(\mathbf{x})$, then the directional derivative (14) exists and equals $\nabla f(\mathbf{x})\,\mathbf{v}$, however (14) may exist and be finite in all directions $\mathbf{v}$ with $f(\cdot)$ not being differentiable at $\mathbf{x}$. A simple example is to take $\mathbf{x} = 0$, $f(\mathbf{x}) = \|\mathbf{x}\|$, and note then that $f'(\mathbf{x}; \mathbf{v}) = \|\mathbf{v}\|$.

Proposition 3.1. *Suppose $f \in \mathcal{F}_c$, $\mathbf{x} \in \mathrm{dom}(f)$, and $\mathbf{v} \in \mathbb{R}^n$ are such that $\mathbf{x} + h\mathbf{v} \in \mathrm{dom}(f)$ for all $h \in [0, h_0]$. Then $f'(\mathbf{x}; \mathbf{v})$ exists (or equals $-\infty$). It is finite if h_0 can be chosen so that $\mathbf{x} \pm h\mathbf{v} \in \mathrm{dom}(f)$ for all $h \in [0, h_0]$.*

Proof. Let $g : (0, h_0] \to \mathbb{R}$ be given by $g(h) = \dfrac{f(\mathbf{x} + h\,\mathbf{v}) - f(\mathbf{x})}{h}$, and we claim $g(\cdot)$ is nondecreasing on $(0, h_0]$. Suppose $0 < h < h' \le h_0$, and consider (11) with $\mathbf{x}_0 := \mathbf{x}$, $\mathbf{x}_1 := \mathbf{x} + h'\mathbf{v}$ and $\lambda := h/h'$, in which case $\mathbf{x}_\lambda := \mathbf{x} + h\mathbf{v}$. Then the first inequality in (11) says

$$h'\,g(h) = h'\frac{f(\mathbf{x}_\lambda) - f(\mathbf{x}_0)}{\lambda} \le f(\mathbf{x}_1) - f(\mathbf{x}_0) = h'\,g(h'),$$

and this implies $g(h) \le g(h')$, or that $g(\cdot) \nearrow$. So if $h \downarrow 0$, then $g(h) \to -\infty$ unless it is bounded below, in which case it converges.

As for the finiteness assertion, suppose $\mathbf{x} \pm h\mathbf{v} \in \mathrm{dom}(f)$ for all $h \in [0, h_0]$. Let $h \in (0, h_0]$, and consider the three point property (11) with $\mathbf{x}_0 := \mathbf{x} - h_0\mathbf{v}$, $\mathbf{x}_1 = \mathbf{x} + h\mathbf{v}$, and $\lambda := \frac{h_0}{h+h_0}$. The choice of λ was to ensure $\mathbf{x}_\lambda = \mathbf{x}$, and so

$$(h + h_0)\frac{f(\mathbf{x}) - f(\mathbf{x} - h_0\mathbf{v})}{h_0} = \frac{f(\mathbf{x}_\lambda) - f(\mathbf{x}_0)}{\lambda} \underbrace{\leq}_{\text{by (11)}} \frac{f(\mathbf{x}_1) - f(\mathbf{x}_\lambda)}{(1-\lambda)} = (h + h_0)\frac{f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x})}{h}$$

This shows $g(h)$ is bounded below by $\dfrac{f(\mathbf{x}) - f(\mathbf{x} - h_0\mathbf{v})}{h_0}$, and therefore $\lim\limits_{h\to 0^+} g(h) = f'(\mathbf{x}; \mathbf{v})$ exists and is finite. $\square$

Exercise 3.1. *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is lsc. Show that $f(\cdot)$ is convex if and only if for all $\mathbf{x}_0, \mathbf{x}_1 \in \mathbb{R}^N$, the function $g : [0, 1] \to \mathbb{R}$ defined by $g(t) = f\big((1-t)\mathbf{x}_0 + t\mathbf{x}_1\big)$ is convex on $[0, 1]$.*