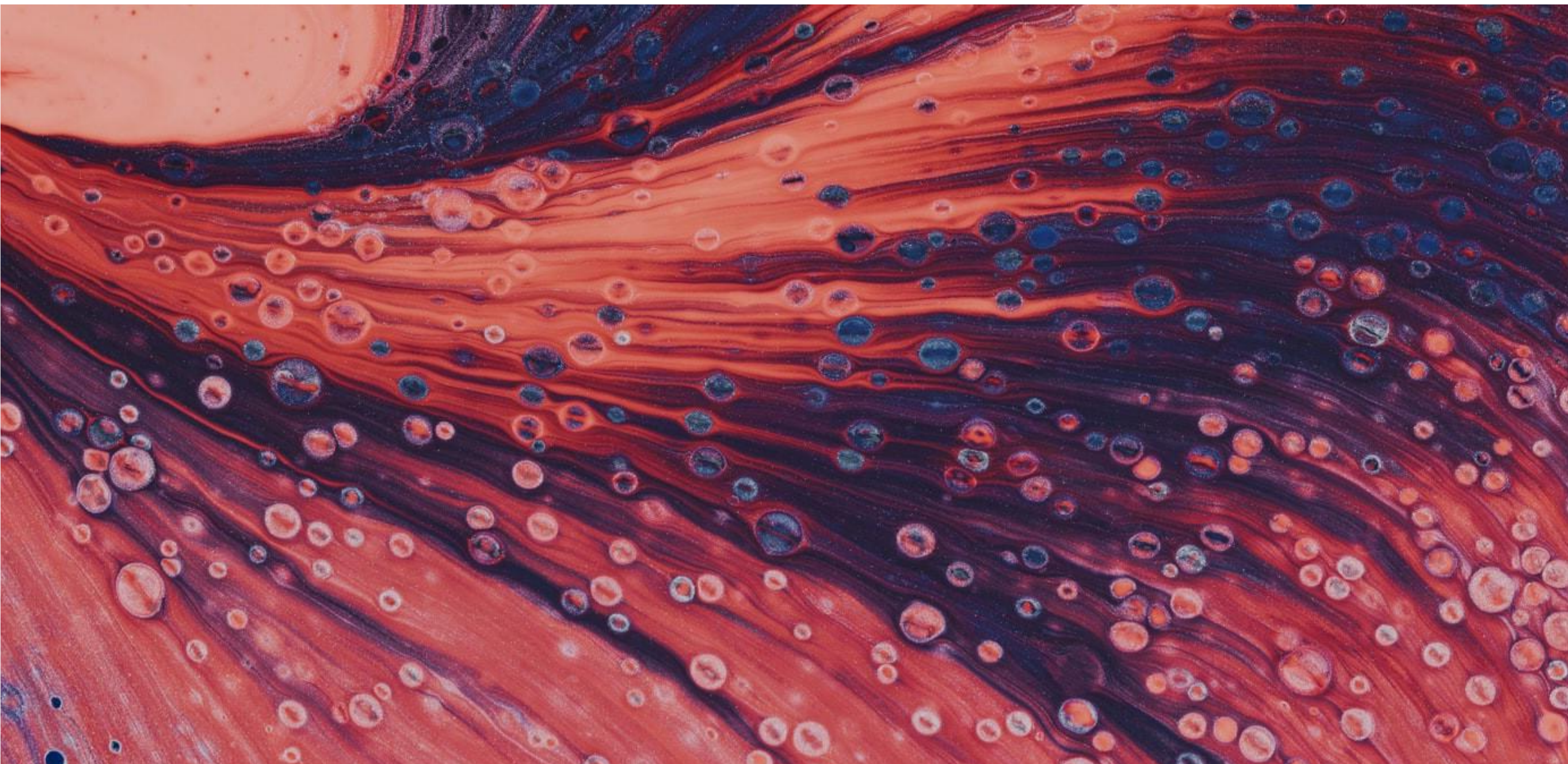
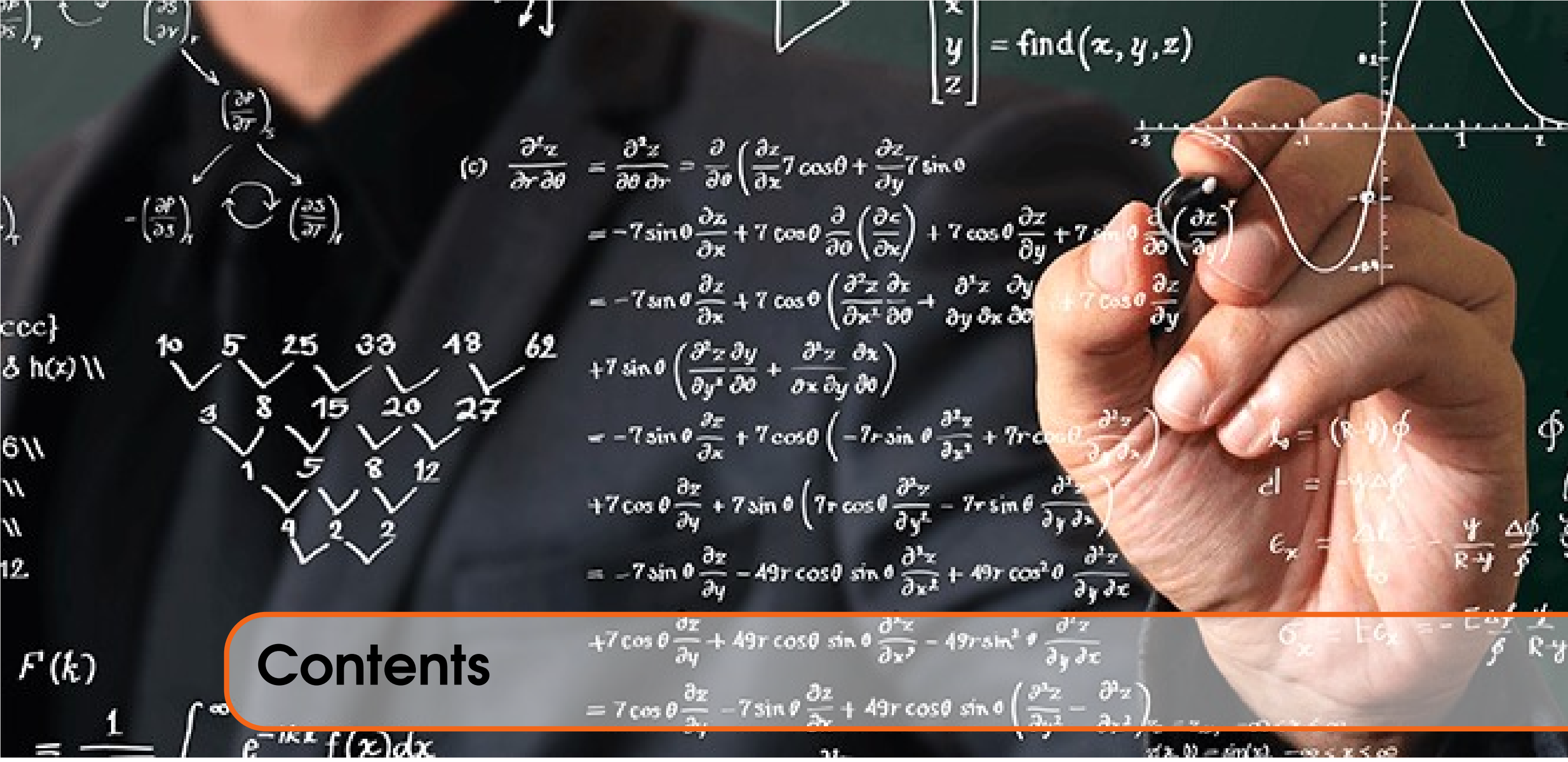
CO 463 Course Notes

University of Waterloo

The One And Only
Waterloo 76er
Bill Zhuo

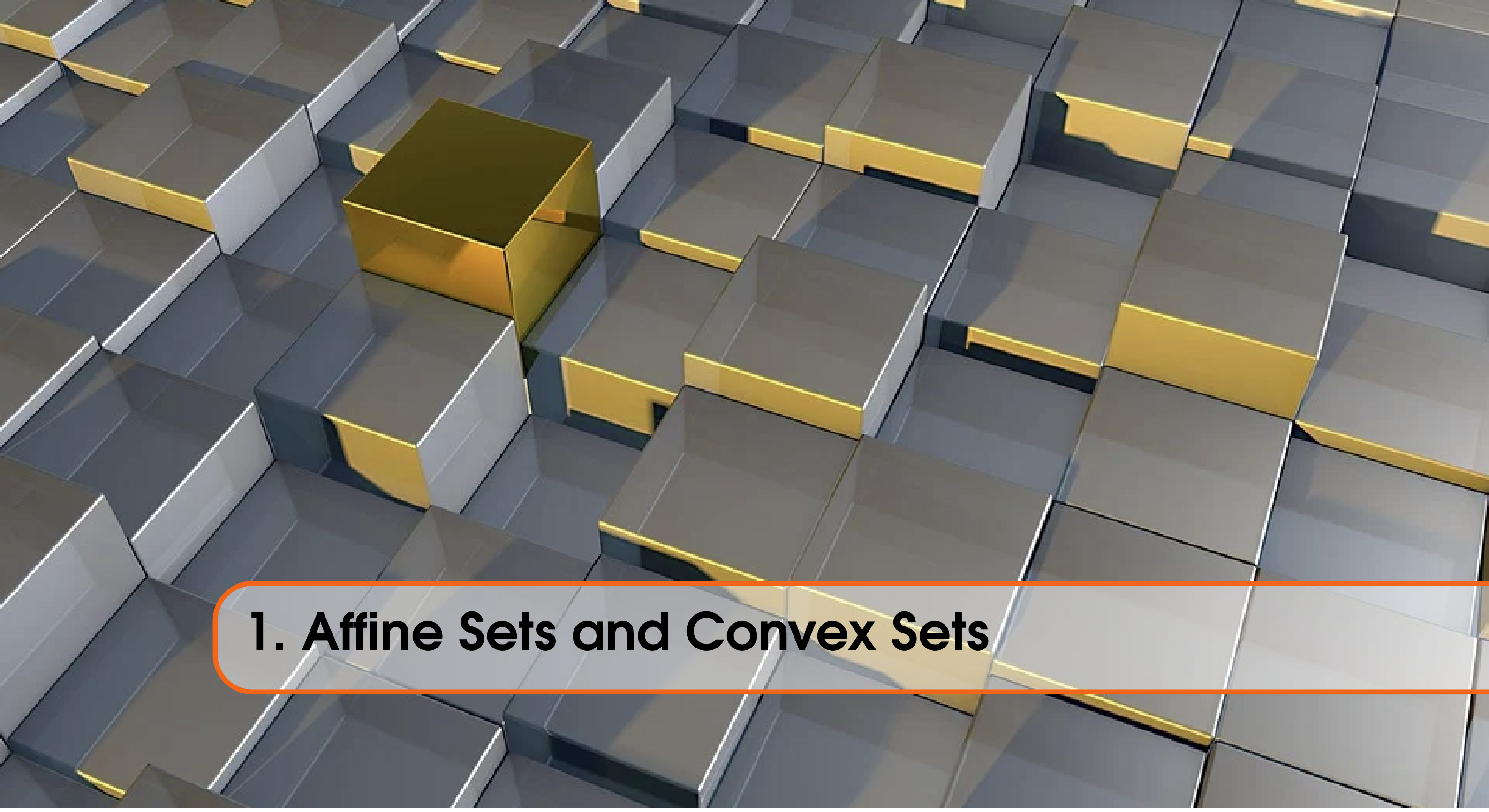




Contents

I CO463 Main Content

CO463 Main Content



1. Affine Sets and Convex Sets

1.1 introduction

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable. Consider the problem

$$(P) \qquad \begin{array}{ll} \min & f(x) \\ s.t. & x \in C \subseteq \mathbb{R}^n \end{array}$$

In the special case, when $C = \mathbb{R}^n$, the minimizers of f (if any) will occur at the critical points of f, namely, $x \in \mathbb{R}^n$ such that $\nabla f(x) = 0$. This is known as **Fermat's Rule**. We will discuss and learn convexity of sets and functions and how we can approach problem (P) in the more general settings of:

1. Absence of differentiability of the objective function f, f is convex
2. $\emptyset \neq C \subsetneq \mathbb{R}^n$, convex C is the constraint set.

1.1.1 Affine Sets and Affine Subspaces in $\mathbb{R}^n$

Definition 1.1.1 — Affine set, affine subspace, and affine hull. Let $S \subseteq \mathbb{R}^n$. Then:

1. S is an affine set if for all $x, y \in S$ and for all $\lambda \in \mathbb{R}$,

$$\lambda x + (1-\lambda)y \in S$$

   Trivially, $\emptyset, \mathbb{R}^n$ are affine sets.
2. S is an affine subspace if it is a non-empty affine set.
3. The affine hull of S, denoted by $\mathbf{aff}(S)$, is the intersection of all affine sets containing S

■ **Example 1.1 — Affine Sets of $\mathbb{R}^n$.** 1. Any linear subspace of $\mathbb{R}^n$
2. $a + L$ where $a \in \mathbb{R}^n$ and L is any linear subspace
3. $\emptyset, \mathbb{R}^n$

■

R Geometrically speaking, a non-empty subset $S \subseteq \mathbb{R}^n$ is affine if the line connecting any two points in the set lies entirely in the set. For example, $S = \{(x_1, x_2) | x_2 \leq 0\}$ is not affine.

1.1.2 Convex Sets in $\mathbb{R}^n$

Definition 1.1.2 — Convex set. A subset C of $\mathbb{R}^n$ is convex if for all $\lambda \in (0,1)$ and $x, y \in C$ we have $\lambda x + (1-\lambda)y \in C$.

Example 1.2 — Convex sets. In $\mathbb{R}^n$,

1. $\emptyset, \mathbb{R}^n$
2. Balls
3. Affine sets
4. Any half-space,

$$C = \{x \in \mathbb{R}^n | \langle x, u \rangle \leq \eta\}, u \in \mathbb{R}^n, \eta \in \mathbb{R}$$

■

R Geometrically speaking, a subset $C \subseteq \mathbb{R}^n$ is convex if given any two points $x, y \in C$, the line segment joining x, y, denoted by $[x, y]$ lies entirely in C.

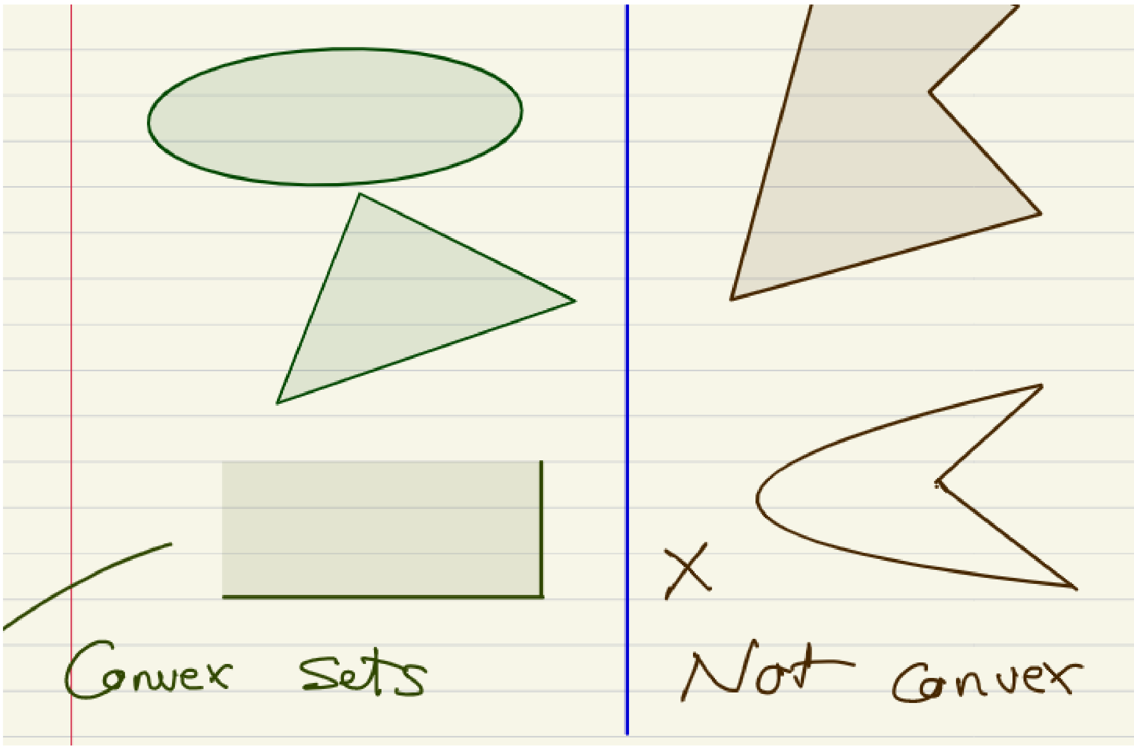


Figure 1.1.1: Convex sets and non-convex sets

Theorem 1 The intersection of an arbitrary collection of convex sets is convex.

Proof. Let I be an index set (not necessarily finite). Let $(C_i)_{i \in I}$ be a collection of convex subsets of $\mathbb{R}^n$. Consider

$$C := \bigcap_{i \in I} C_i$$

Let $\lambda \in (0,1)$ and let $(x, y) \in C \times C$. Since C_i is convex for all $i \in I$. We have

$$\lambda x + (1-\lambda)y \in C_i$$

Thus,

$$\lambda x + (1-\lambda)y \in \bigcap_{i\in I} C_i = C$$

Hence, C is convex. ■

Corollary 1.1.1 Let $b_i \in \mathbb{R}^n$ and $\beta_i \in \mathbb{R}$ for $i \in I$ where I is an arbitrary index set. Then the set

$$C = \{x \in \mathbb{R}^n | \langle x, b_i \rangle \leq \beta_i, \forall i \in I\}$$

is convex.

1.2 Convex Combinations of Vectors

Definition 1.2.1 — Convex Combination. A vector sum

$$\lambda_1 x_1 + \cdots + \lambda_m x_m$$

is called a convex combination of vectors $x_1, \cdots, x_m$ if for $i = 1, \cdots, m$, $\lambda_i \geq 0$ and $\sum_{i=1}^m \lambda_i = 1$.

Theorem 2 A subset C of $\mathbb{R}^n$ is convex if and only if it contains all the convex combinations of its elements.

Proof.
1. **Easy:** suppose C contains all the convex combinations of its elements. Let $\lambda \in (0,1)$ and let $x, y \in C$. By assumption, the convex combination $\lambda x + (1-\lambda)y \in C$. Thus, C is convex.
2. **Hard:** suppose C is convex. Induction on m, the number of elements in the convex combination.
   (a) **Base case:** when $m = 2$, the conclusion is clear by the convexity of C.
   (b) **Induction step:** suppose that for some $m > 2$ it holds that any convex combination of m vectors lies in C. Let $\{x_1, \cdots, x_m, x_{m+1}\} \subseteq C$, let $\lambda_1, \cdots, \lambda_m, \lambda_{m+1} \geq 0$, such that $\sum_{i=1}^{m+1} \lambda_i = 1$. We want to show that

$$z := \sum_{i=1}^{m+1} \lambda_i x_i \in C$$

   Note that there must exist at least one $\lambda_i \in [0,1)$ or else if all $\lambda_i = 1$ then the sum will be greater than 3, which is non-sense. Without loss of generality, we can and do assume that $\lambda_{m+1} \in [0,1)$. Now:

$$\begin{aligned}
z &= \sum_{i=1}^{m+1} \lambda_i x_i \\
&= \sum_{i=1}^{m} \lambda_i x_i + \lambda_{m+1} x_{m+1} \\
&= (1-\lambda_{m+1}) \sum_{i=1}^{m} \frac{\lambda_i}{1-\lambda_{m+1}} x_i + \lambda_{m+1} x_{m+1} \\
&= (1-\lambda_{m+1}) \sum_{i=1}^{m} \lambda_i' x_i + \lambda_{m+1} x_{m+1}
\end{aligned}$$

observe that $\lambda_i' = \frac{\lambda_i}{1-\lambda_{m+1}} \geq 0$ and

$$\sum_{i=1}^{m} \lambda_i' = \frac{1-\lambda_{m+1}}{1-\lambda_{m+1}} = 1$$

Then by inductive hypothesis, we know that

$$z = (1-\lambda_{m+1}) \underbrace{\sum_{i=1}^{m} \frac{\lambda_i}{1-\lambda_{m+1}} x_i}_{\in C} + \lambda_{m+1} \underbrace{x_{m+1}}_{\in C} \in C$$

since C is convex.

We are done. ∎

Definition 1.2.2 — Convex hull. Let $S \subseteq \mathbb{R}^n$. The intersection of all convex sets containing S is called the convex hull of S and is denoted by $\mathbf{conv}(S)$.

By Theorem 1, $\mathbf{conv}(S)$ is convex. In fact, it is the smallest convex set containing S.

Theorem 3 Let $S \subseteq \mathbb{R}^n$. Then $\mathbf{conv}(S)$ consists of all the convex combinations of the elements of S, i.e.,

$$\mathbf{conv}(S) := \left\{ \sum_{i \in I} \lambda_i x_i : I \text{ is a finite index set}, x_i \in S, \lambda_i \geq 0, \sum_{i=1} \lambda_i = 1 \right\}$$

Proof. Let

$$D := \left\{ \sum_{i \in I} \lambda_i x_i : I \text{ is a finite index set}, x_i \in S, \lambda_i \geq 0, \sum_{i=1} \lambda_i = 1 \right\}$$

1. $\mathbf{conv}(S) \subseteq D$: note that $S \subseteq D$. It remains to show D is convex. Let $d_1, d_2 \in D$ and let $\lambda \in (0,1)$. Then, there exists

$$\lambda_1, \cdots, \lambda_k \geq 0, \sum_{i=1}^{k} \lambda_i = 1$$

$$\mu_1, \cdots, \mu_r \geq 0, \sum_{j=1}^{r} \mu_j = 1,$$

$$d_1 = \sum_{i=1}^{k} \lambda_i x_i, \{x_1, \cdots, x_k\} \subseteq S$$

$$d_2 = \sum_{j=1}^{r} \mu_j y_j, \{y_1, \cdots, y_r\} \subseteq S$$

Therefore,

$$\begin{aligned} &\lambda d_1 + (1-\lambda) d_2 \\ =&\lambda \sum_{i=1}^{k} \lambda_i x_i + (1-\lambda) \sum_{j=1}^{r} \mu_j y_j \end{aligned}$$

note that $\lambda\lambda_i, (1-\lambda)\mu_j \geq 0$ for all $i \in \{1, \cdots, k\}$ and $j \in \{1, \cdots, r\}$, and that

$$\lambda \sum_{i=1}^{k} \lambda_i + (1-\lambda) \sum_{j=1}^{r} \mu_j = 1$$

Thus, D is convex and $\mathbf{conv}(S) \subseteq D$.

2. $D \subseteq \mathbf{conv}(S)$: observed that $S \subseteq \mathbf{conv}(S)$. By Theorem 2, all the convex combinations of elements in S are in $\mathbf{conv}(S)$

Thus, we are done. ■

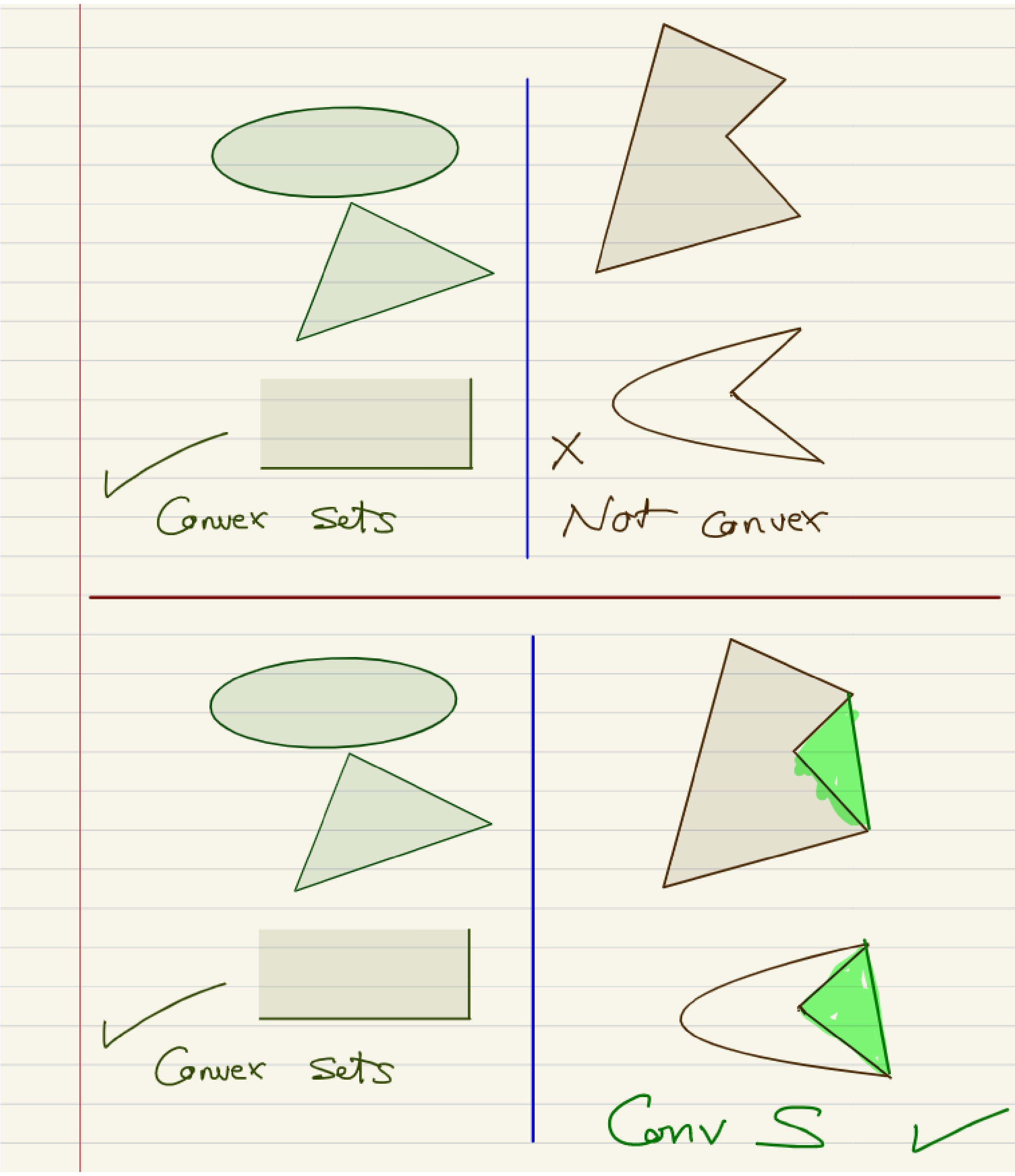


Figure 1.2.1: Convex hulls

1.3 Convex Sets: Best Approximations

Definition 1.3.1 — Distance function. Let $S \subseteq \mathbb{R}^n$. The distance to S is the function $d_S : \mathbb{R}^n \to [0,\infty]$ defined by

$$d_S(x) = \inf_{s\in S} \|x-s\|$$

Definition 1.3.2 — Projection onto a set. Let $\emptyset \neq C \subseteq \mathbb{R}^n$, $x \in \mathbb{R}^n$, and let $p \in C$. Then, p is a projection of x onto C, denoted by $P_C(x)$ if

$$d_C(x) = \|x-p\|$$

R Note that this projection is not necessarily unique. By the projection theorem introduced later, we need C to be convex to have the unique projection.

R Recall that in $\mathbb{R}^n$, every Cauchy sequence converges since $\mathbb{R}^n$ is complete. We also recall sequential continuity in $\mathbb{R}^n$. Consider $\|\cdot\|$, the Euclidean norm on $\mathbb{R}^n$. It is continuous on $\mathbb{R}^n$.

Lemma 1.4 — Auxiliary I. Let $x,y,z \in \mathbb{R}^n$. Then,

$$\|x-y\|^2 = 2\|z-x\|^2 + 2\|z-y\|^2 - 4\left\|z-\frac{x+y}{2}\right\|^2$$

Proof. For the RHS, we handle it term by term

$$\begin{aligned}
2\|z-x\|^2 &= 2\|z\|^2 - 4\langle z,x\rangle + 2\|x\|^2 \\
2\|z-y\|^2 &= 2\|z\|^2 - 4\langle z,y\rangle + 2\|y\|^2 \\
4\left\|z-\frac{x+y}{2}\right\|^2 &= 4\left[\|z\|^2 + \frac{1}{4}\|x+y\|^2 - \langle z,x+y\rangle\right] \\
&= 4\|z\|^2 + \|x+y\|^2 - 4\langle z,x\rangle - 4\langle z,y\rangle
\end{aligned}$$

Now, add them together,

$$\begin{aligned}
RHS &= 2\|x\|^2 + 2\|y\|^2 - \|x+y\|^2 \\
&= 2\|x\|^2 + 2\|y\|^2 - \|x\|^2 - \|y\|^2 - 2\langle x,y\rangle \\
&= \|x-y\|^2 = LHS
\end{aligned}$$

■

Lemma 1.5 — Auxiliary II. Let $x,y \in \mathbb{R}^n$. Then,

$$\langle x,y\rangle \leq 0 \iff \forall \lambda \in [0,1], \|x\| \leq \|x-\lambda y\|$$

Proof. 1. $\Longrightarrow$: suppose $\langle x,y\rangle \leq 0$. Then

$$\|x-\lambda y\|^2 - \|x\|^2 = \|x\|^2 - 2\lambda\langle x,y\rangle + \lambda^2\|y\|^2 - \|x\|^2 = \lambda(\lambda\|y\|^2 - 2\langle x,y\rangle) \geq 0$$

2. $\Longleftarrow$: suppose that for every $\lambda \in (0,1]$,

$$\|x-\lambda y\| \geq \|x\|$$

then,

$$\langle x,y\rangle \leq \frac{\lambda}{2}\|y\|^2$$

We can take $\lambda \downarrow 0$ to yield the desired result.

■

Theorem 4 — The projection theorem. Let C be a non-empty, closed, and convex subset of $\mathbb{R}^n$. Then,

1. For all $x \in \mathbb{R}^n$ $P_C(x)$ exists and is unique
2. For all $x \in \mathbb{R}^n$ and every $p \in \mathbb{R}^n$,

$$p = P_C(x) \iff p \in C \wedge \forall y \in C, \langle y - p, x - p \rangle \leq 0$$

Proof. Let $x \in \mathbb{R}^n$,

1. Our goal is to show that x has a unique projection onto C.
   (a) **Existence:** there exists a sequence $(c_n)_n$ in C such that $d_C(x) = \lim_n \|c_n - x\|$. Let $m, n \in \mathbb{N}$, by the convexity of C, we know that

$$\frac{1}{2}(c_m + c_n) \in C$$

   then,

$$d_C(x) = \inf_{c \in C} \|x - c\| \leq \left\| x - \frac{1}{2}(c_m + c_n) \right\|$$

   by Auxiliary I lemma,

$$\begin{aligned} \|c_n - c_m\|^2 &= 2\|c_n - x\|^2 + 2\|c_m - x\| - 4\left\| x - \frac{c_n + c_m}{2} \right\|^2 \\ &\leq 2\|c_n - x\|^2 + 2\|c_m - x\| - 4d_C^2(x) \end{aligned}$$

   let $m, n \to \infty$, we have

$$0 \leq \|c_n - c_m\|^2 \to 2d_C^2(x) + 2d_C^2(x) - 4d_C^2(x) = 0$$

   Thus, $\|c_n - c_m\|^2 \to 0$. Hence, $(c_n)_n$ is Cauchy in C and it converges to some point $p \in C$ by the closedness of C. By sequential continuity, we have

$$d_C(x) = \|x - p\|$$

   (b) **Uniqueness:** suppose that $q \in C$ satisfies that $d_C(x) = \|q - x\|$. By the convexity of C, $\frac{1}{2}(p+q) \in C$. By Auxiliary I,

$$\begin{aligned} 0 &\leq \|p - q\|^2 \\ &= 2\|p - x\|^2 + 2\|q - x\|^2 - 4\left\| x - \frac{p+q}{2} \right\|^2 \\ &\leq 2d_C^2(x) + 2d_C^2(x) - 4d_C^2(x) = 0 \end{aligned}$$

   This means $p = q$.
2. From part 1, we note that

$$p = P_C(x) \iff p \in C \wedge \|x - p\|^2 = d_C^2(x)$$

   Note that for every $y \in C$,

$$\alpha \in [0,1], y_\alpha := \alpha y + (1 - \alpha)p \in C$$

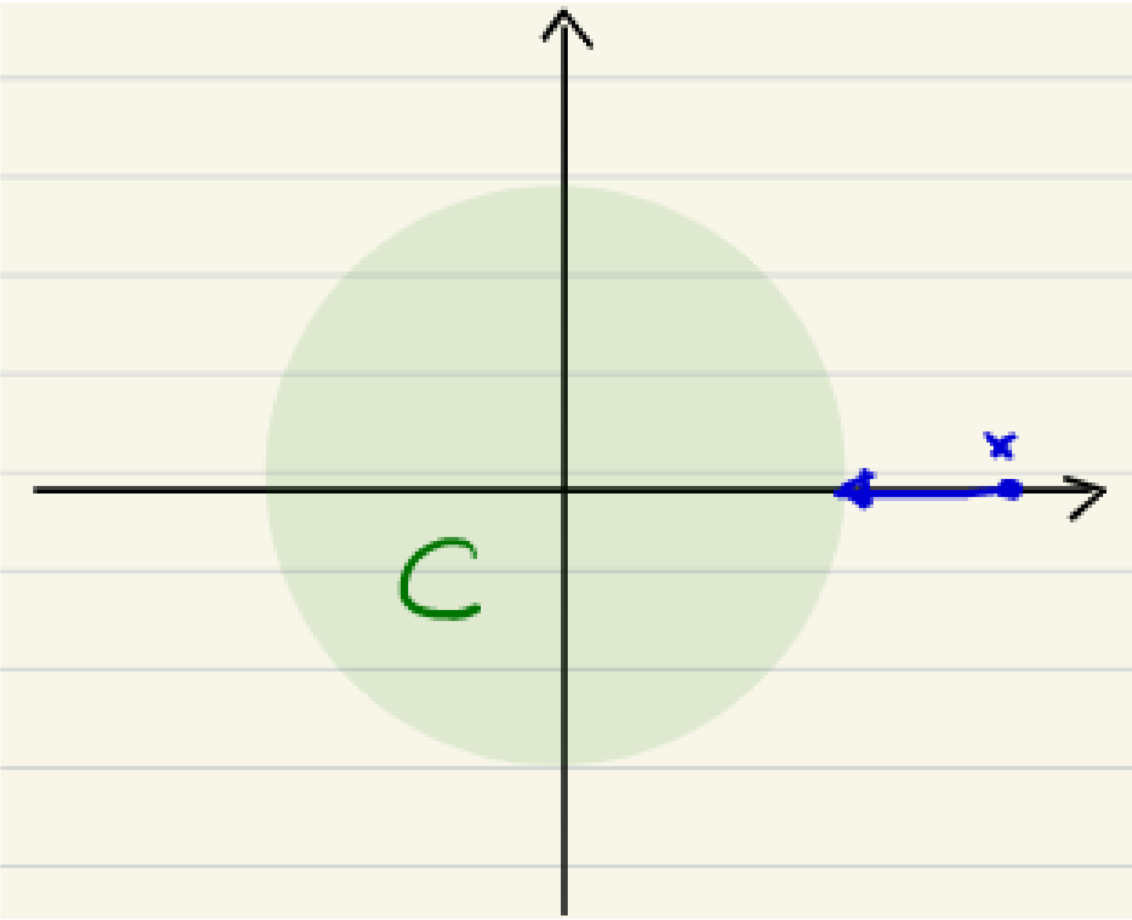
Therefore,

$$\begin{aligned} &\|x-p\|^2 = d_C^2(x) \\ \iff &\forall y \in C, \forall \alpha \in [0,1], \|x-p\|^2 \leq \|x-y_\alpha\|^2 \\ \iff &\forall y \in C, \forall \alpha \in [0,1], \|x-p\|^2 \leq \|x-p-\alpha(y-p)\|^2 \\ \iff &\forall y \in C \, \langle x-p, y-p \rangle \leq 0 \end{aligned}$$

last $\iff$ is by Auxiliary II.

■

■ **Example 1.3 — Absence of closedness.** For any $x \in \mathbb{R}^n \backslash C$, the projection of x onto C does not exist.

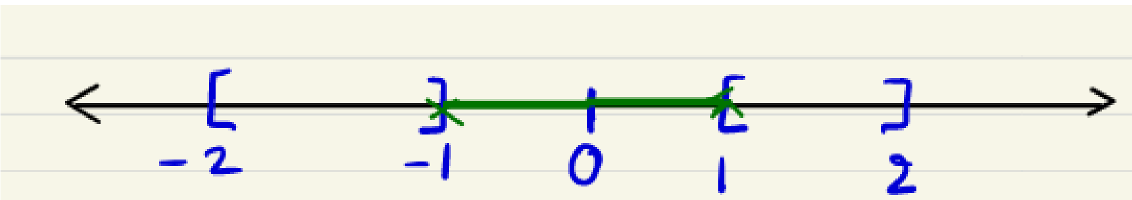


■

■ **Example 1.4 — Absence of convexity.** On the real line $\mathbb{R}$, consider

$$C = [-2,-1] \cup [1,2]$$

which is not convex. Both $1, -1$ are projections of 0 onto C.



■

Exercise 1.1 Let $\varepsilon > 0$, and let $C = B(0,\varepsilon) = \left\{x \in \mathbb{R}^n : \|x\|^2 \leq \varepsilon^2\right\}$. Show that

$$\forall x \in \mathbb{R}^n, P_C(x) = \frac{\varepsilon}{\max\{\|x\|,\varepsilon\}}x$$

Proof. Let $x \in \mathbb{R}^n$ and let $p := \frac{\varepsilon}{\max\{\|x\|,\varepsilon\}}x$. By the projection theorem, it suffices to show that

1. $p \in C$:
   (a) **Case 1:** when $\|x\| \leq \varepsilon$. Then, $x \in C$, $p = \frac{\varepsilon}{\varepsilon}x = x \in C$.
   (b) **Case: 2** when $\|x\| > \varepsilon$. Then, $p = \frac{\varepsilon}{\|x\|}x$, and $\|p\| = \varepsilon \implies p \in C$.
2. $\forall y \in C, \langle x-p, y-p\rangle \leq 0$: let $y \in C$,
   (a) **Case 1:** when $\|x\| \leq \varepsilon$, $p = x$ and

$$0 \leq \langle x-p, y-p\rangle \leq 0$$

   (b) **Case 2:** when $\|x\| > \varepsilon$, then $p = \frac{\varepsilon}{\|x\|}x$. We check

$$\begin{aligned}
\langle x-p, y-p\rangle &= \left\langle x - \frac{\varepsilon}{\|x\|}x, y - \frac{\varepsilon}{\|x\|}x\right\rangle \\
&= \left(1 - \frac{\varepsilon}{\|x\|}\right)\left\langle x, y - \frac{\varepsilon}{\|x\|}x\right\rangle \\
&= \left(1 - \frac{\varepsilon}{\|x\|}\right)\left(\langle x, y\rangle - \frac{\varepsilon}{\|x\|}\|x\|^2\right) \\
&\overset{\text{C-S Inequality}}{\leq} \left(1 - \frac{\varepsilon}{\|x\|}\right)(\|x\|\|y\| - \varepsilon\|x\|) \\
&\overset{\|y\|\leq\varepsilon}{\leq} \left(1 - \frac{\varepsilon}{\|x\|}\right)(\|x\|\varepsilon - \varepsilon\|x\|) \\
&= 0
\end{aligned}$$

Thus, we are done. ■

■

Definition 1.5.1 — Minkowski sum of two sets. Let C, D be two subsets of $\mathbb{R}^n$. Th e Minkowski sum of C, D, denoted by $C + D$, is

$$C + D := \{c + d : c \in C, d \in D\}$$

Theorem 5 Let $C_1, C_2 \subseteq \mathbb{R}^n$ be convex. Then, $C_1 + C_2$ is also convex.

Proof. Left as an exercise to the reader. ■

Proposition 1.5.1 Let C, D be non-empty, closed, convex subsets of $\mathbb{R}^n$ such that D is bounded. Then, $C + D$ is non-empty, closed, and convex.

Proof. It is clear that $C + D$ is non-empty when C, D are non-empty and by Theorem 5, $C + D$ is convex. It remains to check $C + D$ is closed. Let $(x_n + y_n)_n$ be a sequence in $C + D$ such that $(x_n)_n$ is in C and $(y_n)_n$ is in D. Moreover, $x_n + y_n \to z$. We want to show $z \in C + D$. Since D is bounded, we have $(y_n)_n$ bounded. Then, by Bolzano-Weierstrass Theorem, we know that there exists a subsequence $(y_{n_k})_k$ such that $y_{n_k} \to y \in D$. Then, $x_{n_k} \to \bar{x} = z - y \in C$ by closedness. Thus, $z \in C + y \subseteq C + D$. ■

R What happens if we drop the assumption that D is bounded?

Exercise 1.2 Given an example of two closed convex cones $K_1, K_2 \subseteq \mathbb{R}^n$ such that $K_1 + K_2$ is not closed.

Proof. Consider $n = 3$, $S_1 = \left\{(x,y,z) \in \mathbb{R}^3 : x^2 \geq y^2 + z^2, x \geq 0\right\}$ and $S_2 = \{t(-1,0,1) : t \geq 0\}$. It is clear that S_2 is a closed cone. We note that S_1 is a closed set and we proceed to show S_1 is indeed a cone. Note that S_1 is the cone constructed by lifting a convex disk $S = \left\{(x,y) : x^2 + y^2 \leq 1\right\}$. Thus, by the lifting lemma proved in class, S_1 is a convex cone. Now, for each $n \in \mathbb{N}$, consider

$$n(-1,0,1) + \left(\sqrt{n^2 + \left(1 + \frac{1}{n}\right)^2}, 1 + \frac{1}{n}, -n\right) \in S_1 + S_2$$

Then, as $n \to \infty$, this sequence of points converges to $(0,1,0)$. The sequence converges but $(0,1,0)$ is not in S_1 nor S_2. We claim $(0,1,0) \notin S_1 + S_2$. For the sake contradiction, say $(0,1,0) \in S_1 + S_2$. Then, we can write $(0,1,0) = s_1 + s_2$ with $s_1 = (-\lambda, 0, \lambda)$ for some $\lambda \geq 0$ and $s_2 \in S_2$. Note that this forces

$$s_2 = (\lambda, 1, -\lambda)$$

but $\lambda^2 < 1 + \lambda^2 \implies s_2 \notin S_2$. This yields a contradiction and $(0,1,0) \notin S_1 + S_2$ gives us a valid counterexample. ■