1 Topic 1.

1.1 *Lecture 1. Affine Sets and Convex Sets.*

1.1.1 Overview

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable and consider the following problem:

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

We call f the **objective function** and C the **constraint set**.

In the special case where $C = \mathbb{R}^n$, the minimizers of f (if any) will occur at the *critical points* of f, namely the points $x \in \mathbb{R}^n$ such that $\nabla f(x) = 0$. This is known as the **Fermat's rule**, which we will learn about more later.

In this course, 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 function f, where f is convex, and/or
2. $\varnothing \neq C \subsetneq \mathbb{R}^n$, where C is convex.

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

Geometrically speaking, a non-empty subset $S \subseteq \mathbb{R}^n$ is **affine** if the *line* [1] connecting any two points in the set lies entirely in the set. The intuitive picture is that of an endless uncurved structure, like a line or a plane in space. [2]

Definition 1.1. Let $S \subseteq \mathbb{R}^n$. Then

- S is an **affine set** if $\lambda x + (1-\lambda)y \in S$ for all $x, y \in S$ and $\lambda \in \mathbb{R}$.
- S is an **affine subspace** if it is a *non-empty* affine set.
- Let $S \subseteq \mathbb{R}^n$. Then **affine hull** of S, denoted by $\operatorname{aff}(S)$, is the intersection of all affine sets containing S. Equivalently, it is the smallest affine set containing S.

Some elementary examples of affine sets of $\mathbb{R}^n$:

1. L, where $L \subseteq \mathbb{R}^n$ is a linear subspace.
2. $a + L$, where $a \in \mathbb{R}^n$ and $L \subseteq \mathbb{R}^n$ is a linear subspace.
3. $\varnothing$ and $\mathbb{R}^n$ are extreme affine sets of $\mathbb{R}^n$.

[1] Not line segment! That is for convex sets.

[2] Another perspective: An affine space is what is left of a vector space after you've forgotten which point is the origin. Marcel Berger: "An affine space is nothing more than a vector space whose origin we try to forget about, by adding translations to the linear maps." Connect this with examples 1 and 2.

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

A subset $C \subseteq \mathbb{R}^n$ is **convex** if given any two points x and y in C, the *line segment* joining x and y lies entirely in C.

Definition 1.2. A subset C of $\mathbb{R}^n$ is **convex** if

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

for all $x, y \in C$ and $0 < \lambda < 1$.

Theorem 1.3. *The intersection of an arbitrary collection of convex sets is convex.*

Proof. Let $(C_i)_{i \in I}$ be a collection of convex subsets of $\mathbb{R}^n$ indexed by I. Define $C := \bigcap_{i \in I} C_i$. Fix $x, y \in C$ and $\lambda \in (0,1)$. Since C_i is convex for all $i \in I$, i.e.,

$$\forall i \in I : \lambda x + (1-\lambda)y \in C_i,$$

we get $\lambda x + (1-\lambda)y \in \bigcap_{i \in I} C_i = C$. Hence, C is convex. □

Corollary 1.4. *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*

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

is convex.

Proof. For each $i \in I$, define

$$C_i := \{x \in \mathbb{R}^n \mid \langle x, b_i \rangle \leq \beta_i\}.$$

We claim that all such C_i's are convex. Indeed, let $i \in I$ and fix $x, y \in C_i$ and $\lambda \in (0,1)$. Set $z := \lambda x + (1-\lambda)y$. Then

$$\begin{aligned}
\langle z, b_i \rangle &= \langle \lambda x + (1-\lambda y), b_i \rangle \\
&= \lambda \langle x, b_i \rangle + (1-\lambda) \langle y, b_i \rangle && \text{linearity of } \langle \cdot, \cdot \rangle \\
&\leq \lambda \beta_i + (1-\lambda)\beta_i && \forall x \in C_i : \langle x, b_i \rangle \leq \beta_i \\
&= \beta_i.
\end{aligned}$$

Thus, $z \in C_i$ and C_i is convex. The result follows from Theorem 1.3. □

1.1.4 Convex Combination of Vectors

Definition 1.5. A vector sum $\lambda_1 x_1 + \cdots + \lambda_m x_m$ is called a **convex combination** of vectors $x_1, \ldots, x_m$ if

$$\forall i \in [m] : \lambda_i \geq 0 \quad \text{and} \quad \sum_{i=1}^{m} \lambda_i = 1.$$

Theorem 1.6. *A subset $C \subseteq \mathbb{R}^n$ is convex iff it contains all the convex combinations of its elements.*

Proof. $\Leftarrow$: Trivial. $\Rightarrow$: Suppose C is convex. We proceed by induction on m, the number of vectors in the convex combination. For $m = 2$, the conclusion is clear as C is convex. Now suppose that for some $m > 2$, it holds that any convex combination of m vectors lies in C. Let $\{x_1, \ldots, x_m, x_{m+1}\} \subseteq C$ and $\lambda_1, \ldots, \lambda_m, \lambda_{m+1} \geq 0$ such that $\sum_{i=1}^{m+1} \lambda_i = 1$. Our goal is to show that

$$z := \lambda_1 x_1 + \cdots + \lambda_m x_m + \lambda_{m+1} x_{m+1} \in C.$$

Observe there must exist at least on $\lambda \in [0, 1)$ or else if all $\lambda_i = 1$ the sum would be greater than 1. WLOG, assume that $\lambda_{m+1} \in [0, 1)$. Now

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

Observe that $\lambda_i' := \dfrac{\lambda_i}{1 - \lambda_{m+1}} \geq 0$ and $\displaystyle\sum_{i=1}^{m} \lambda_i' = \frac{\lambda_1 + \cdots + \lambda_m}{1 - \lambda_{m+1}} = \frac{1 - \lambda_{m+1}}{1 - \lambda_{m+1}} = 1$ as $\displaystyle\sum_{i=1}^{m+1} \lambda_i = 1$. Then z is a convex combination of two vectors in C, so it also lies in C i.e.,

$$z = \underbrace{(1 - \lambda_{m+1}) \left(\sum_{i=1}^{m} \lambda_i' x_i \right)}_{\in C \text{ by IH}} + \lambda_{m+1} \underbrace{x_{m+1}}_{\in C} \in C.$$

It follows that C is convex as desired. $\square$

Definition 1.7. 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 $\operatorname{conv}(S)$. It is the smallest convex set containing S.

Theorem 1.8. *Let $S \subseteq \mathbb{R}^n$. Then $conv(S)$ consists of all convex combinations of the elements of S, i.e.,*

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

Proof. Define

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

We first show $\operatorname{conv}(S) \subseteq D$. Clearly, $S \subseteq D$. Moreover, D is convex. Indeed, let $d_1, d_2 \in D$ and $\lambda \in [0, 1]$. Then we can write

$$d_1 = \sum_{i=1}^{k} \lambda_i x_i \quad \text{where} \quad \lambda_1, \ldots, \lambda_k \geq 0, \sum_{i=1}^{k} \lambda_i = 1, \{x_1, \ldots, x_k\} \subseteq S,$$
$$d_2 = \sum_{j=1}^{r} \mu_j y_j \quad \text{where} \quad \mu_1, \ldots, \mu_r \geq 0, \sum_{j=1}^{r} \mu_j = 1, \{y_1, \ldots, y_r\} \subseteq S.$$

Therefore,

$$\lambda d_1 + (1-\lambda) d_2 = [\lambda\lambda_1 x_1 + \cdots + \lambda\lambda_k x_k] + [(1-\lambda)\mu_1 y_1 + \cdots + (1-\lambda)\mu_r y_r].$$

Observe that $\lambda\lambda_1$ and $(1-\lambda)\mu_j$ are non-negative for all $i \in [k]$ and $j \in [r]$ and that

$$\begin{aligned} \lambda\lambda_1 + \cdots + \lambda\lambda_k + (1-\lambda)\mu_1 + \cdots + (1-\lambda)\mu_r &= \lambda \sum_{i=1}^{k} \lambda_i + (1-\lambda) \sum_{j=1}^{r} \mu_j \\ &= \lambda \cdot 1 + (1-\lambda) \cdot 1 = 1. \end{aligned}$$

Altogether, we conclude that D is a convex set $\supseteq S$ and thus $\operatorname{conv}(S) \subseteq D$.

We now show that $D \subseteq \operatorname{conv}(S)$. Observe that $S \subseteq \operatorname{conv}(S)$. Now combine with Theorem 1.6 to learn that the convex combinations of elements of S lie in $\operatorname{conv}(S)$. □

1.2 Lecture 2. Convex Sets: Best Approximations.

1.2.1 Review

Recall the following facts from real analysis.

Definition 1.9. A sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathbb{R}^n$ is said to be **Cauchy** if $\|x_m - x_n\| \to 0$ as $\min\{m, n\} \to \infty$.

Proposition 1.10. *$\mathbb{R}^n$ is **complete**: every Cauchy sequence in $\mathbb{R}^n$ converges.*

Proposition 1.11. *Let $y \in \mathbb{R}^n$ and $\|\cdot\|$ be the Euclidean norm on $\mathbb{R}^n$. Then the function $f : \mathbb{R}^n \to \mathbb{R}$ defined by $x \mapsto \|x - y\|$ is continuous.*

Lemma 1.12. *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. Elementary algebra. □

Lemma 1.13. *Let $x, y \in \mathbb{R}^n$. Then*

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

Proof. Observe that

$$\|x - \lambda y\|^2 - \|x\|^2 = \cancel{\|x\|^2} - 2\lambda\langle x, y\rangle + \lambda^2\|y\|^2 - \cancel{\|x\|^2} = \lambda(\lambda\|y\|^2 - 2\langle x, y\rangle). \tag{1.1}$$

Suppose $\langle x, y\rangle \le 0$. Then

$$\lambda\|y\|^2 \ge 0 \wedge -2\langle x, y\rangle \ge 0 \implies \|x - \lambda y\|^2 - \|x\|^2 = \lambda(\lambda\|y\|^2 - 2\langle x, y\rangle) \ge 0.$$

Conversely, suppose that for every $\lambda \in (0,1]$, $\|x - \lambda y\| \ge \|x\|$. Then (1.1) implies that

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

Taking the limit as $\lambda \to 0$ yields the desired result. □

1.2.2 Projection

The distance between a point x and a set $S \subseteq \mathbb{R}^n$ is the infimum of the distances between the point and those in the set. Intuitively, we find the element $s \in S$ "closest" to x and then measure the distance between them.

Definition 1.14. Let $S \subseteq \mathbb{R}^n$. The **distance function** to S is the function

$$\begin{aligned} d_S : \mathbb{R}^n &\to [0, \infty] \\ x &\mapsto \inf_{s \in S} \|x - s\|. \end{aligned}$$

The projection of x onto C is the element in C that attains the infimum given by $d_C(x)$. If $p = P_C(x)$, then it is the closest one (to x) among all elements in C.

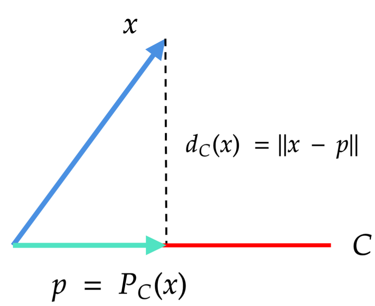


Figure 1.1: Projecting x onto C.

Definition 1.15. Let $\varnothing \neq C \subseteq \mathbb{R}^n$, $x \in \mathbb{R}^n$, and $p \in C$. Then p is the **projection** of x onto C, denoted by $P_C(x)$, if

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

Theorem 1.16 (The Projection Theorem). *Let C be a non-empty, closed, convex subset of $\mathbb{R}^n$. Then the following holds:*

1. $\forall x \in \mathbb{R}^n : P_C(x)$ *exists and is unique.*
2. *For every $x \in \mathbb{R}^n$ 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 of Claim 1. Let $x \in \mathbb{R}^n$. Our goal is to show that x has a unique projection onto C.

Existence. By definition, the distance from x to C is given by $d_C(x) = \inf_{c \in C} \|x - c\|$. Therefore, there exists a sequence $(c_n)_{n \in \mathbb{N}}$ in C such that

$$d_C(x) = \lim_{n \to \infty} \|c_n - x\|. \tag{1.2}$$

Now let $m, n \in \mathbb{N}$. By convexity of C, we know that $(c_m + c_n)/2 \in C$. Hence,

$$d_C(x) = \inf_{c \in C} \|x - c\| \leq \left\| x - \frac{1}{2}(c_m + c_n) \right\|. \tag{1.3}$$

Applying Lemma 1.12 with $(x, y, z) = (c_m, c_n, (c_m + c_n)/2)$, we learn that

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

where the inequality follows from (1.3). By (1.2), letting $m \to \infty$ and $n \to \infty$, we see that

$$0 \leq \|c_n - c_m\|^2 \overset{m,n \to \infty}{\longrightarrow} 2d_C^2(x) + 2d_C^2(x) - 4d_C^2(x) = 0.$$

Hence, $(c_n)_{n \in \mathbb{N}}$ is a Cauchy sequence in C and thus converges to some p:

$$\lim_{n \to \infty} c_n = p.$$

Note $p \in C$ as C is closed. We will now show that $d_C(x) = \|x - p\|$, so by definition p is the desired projection. First, the function $\|x - \cdot\|$ is continuous. Combining with $c_n \to p$ and (1.2), we learn that $\|x - c_n\| \to d_C(x)$ and $\|x - c_n\| \to \|x - p\|$, which gives

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

This concludes the existence of $p = P_C(x)$.

Uniqueness. Suppose that $q \in C$ satisfies $d_C(x) = \|q - x\|$. By convexity of C, $(p + q)/2 \in C$. Using Lemma 1.12 with $(x, y, z) = (p, q, (p + q)/2)$, we see that

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

Hence, $\|p - q\| = 0$ and $p = q$. Therefore the projection is unique.

The proof for part 1 is complete. □

Proof of Claim 2. We want to show that

$$\forall x \in \mathbb{R}^n, \forall p \in \mathbb{R}^n : p = P_C(x) \iff p \in C \wedge \forall y \in C : \langle y - p, x - p \rangle \leq 0.$$

We do so with a series of iffs. Indeed,

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

By convexity of C, $y_\alpha := \alpha y + (1 - \alpha)p \in C$ for every $y \in C$ and $\alpha \in [0, 1]$. Therefore,

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

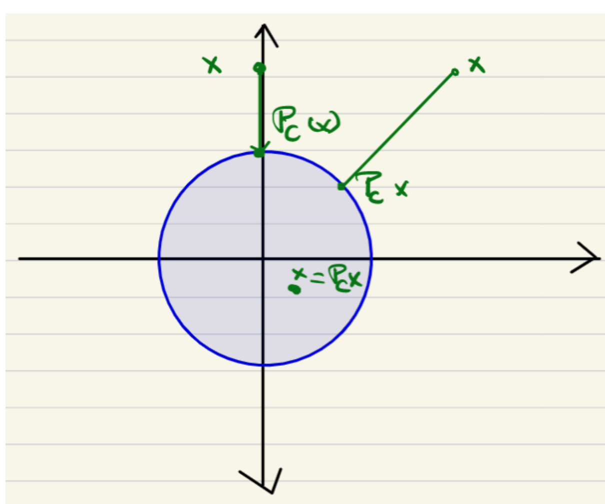
where in the second iff we subtracted p and added αp and $\alpha \in [0, 1]$; the third iff uses Lemma 1.13 with $(x, y) = (x - p, y - p)$. □

Remark. Note both closedness and convexity are necessary.

- In the absence of closedness, $P_C(x)$ doesn't exist for all $x \notin C$ as the limit point of the sequence $(c_n)_{n\in\mathbb{N}}$ in the first proof is not contained by C.
- In the absence of convexity, the projection might not unique. For example, with $C := [-2, 1] \cup [1, 2]$, $x = 0$ has two closest points: -1 and 1, so $P_C(0) = \{-1, 1\}$.

Example. Let $\varepsilon > 0$ and $C = \bar{b}_\varepsilon(0) := \{x \in \mathbb{R}^n : \|x\|^2 \leq \varepsilon^2\}$, i.e., the closed ball in $\mathbb{R}^n$ centered at 0 with radius ε. We claim that

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



Let $x \in \mathbb{R}^n$ and set $p = P_C(x)$ given above. Using the projection theorem, it suffices to show that $p \in C$ and $\langle x - p, y - p \rangle \leq 0$ for all $y \in C$.

Claim 1. $p \in C$. We examine two cases. First, if $\|x\| \leq \varepsilon$, then $x \in C$ and $p = \frac{\varepsilon}{\varepsilon} x = x \in C$. Now if $\|x\| > \varepsilon$, then

$$p = \frac{\varepsilon}{\|x\|} x \implies \| = \varepsilon \frac{\|x\|}{\|x\|} = \varepsilon \implies p \in C.$$

Claim 2. $\forall y \in C : \langle x - p, y - p \rangle \leq 0$. Let $y \in C$. If $\|x\| \leq \varepsilon$, then $p = x$ and $0 = \langle x - p, y - p \rangle \leq 0$. Now if $\|x\| \geq \varepsilon$, then $p = \frac{\varepsilon}{\|x\|} x$. Moreover,

$$\begin{aligned}
\langle x - p, y - p \rangle &= \left\langle x - \frac{\varepsilon}{\|x\|} x, y - \frac{\varepsilon}{\|x\|} x \right\rangle \\
&= \left(1 - \frac{\varepsilon}{\|x\|}\right) \left\langle x, y - \frac{\varepsilon}{\|x\|} x \right\rangle \\
&= \left(1 - \frac{\varepsilon}{\|x\|}\right) \left(\langle x, y \rangle - \frac{\varepsilon}{\|x\|} \|x\|^2\right) \\
&= \left(1 - \frac{\varepsilon}{\|x\|}\right) (\langle x, y \rangle - \varepsilon \|x\|) \\
&\leq \left(1 - \frac{\varepsilon}{\|x\|}\right) (\|x\| \|y\| - \varepsilon \|x\|) && \text{Cauchy-Schwarz} \\
&\leq \left(1 - \frac{\varepsilon}{\|x\|}\right) (\|x\| \varepsilon - \varepsilon \|x\|) && y \in C \implies \|y\| \leq \varepsilon \\
&= 0.
\end{aligned}$$

□

1.2.3 The Algebra of Convex Sets

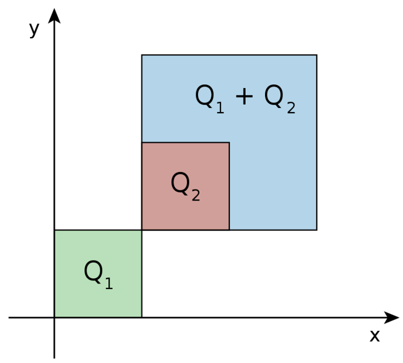


Figure 1.2: $Q_1 + Q_2 = [0,1]^2 + [1,2]^2 = [1,3]^2$.

Definition 1.17 (Minkowski Sum of Sets)**.** Let $C, D \subseteq \mathbb{R}^n$. The **Minkowski sum** of C and D, denoted by $C + D$, is

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

Theorem 1.18. *Let $C_1, C_2 \subseteq \mathbb{R}^n$ be convex. Then $C_1 + C_2$ is convex.*

Proof. If one of them is $\varnothing$ then their Minkowski sum is $\varnothing$ and the conclusion follows. Now suppose both are not empty, so $C + D$ is non-empty. Let $x, y \in C_1 + C_2$ and $\lambda \in (0,1)$. Since $x \in C_1 + C_2$, there exists $x_1 \in C_1, x_2 \in C_2$ such that $x = x_1 + x_2$. Similarly, we can find $y_1 \in C_1$, $y_2 \in C_2$ such that $y = y_1 + y_2$. Now

$$\lambda x + (1-\lambda)y = \lambda(x_1 + x_2) + (1-\lambda)(y_1 + y_2) = \lambda x_1 + (1-\lambda)y_1 + \lambda x_2 + (1-\lambda)y_2 \in C_1 + C_2.$$

The proof is complete. □

Proposition 1.19. *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. $C \neq \varnothing \wedge D \neq \varnothing \implies C + D \neq \varnothing$.
C and D are convex so $C + D$ is convex by Theorem 1.18.
It remains to show that $C + D$ is closed. Take a convergent sequence $(x_n + y_n)_{n\in\mathbb{N}}$ in $C + D$ such that $(x_n)_{n\in\mathbb{N}}$ lies in C, $(y_n)_{n\in\mathbb{N}}$ lies in D, and $x_n + y_n \to z$. Our goal is to show that $z \in C + D$. By assumption, D is bounded, so $(y_n)_{n\in\mathbb{N}}$ is bounded. Using BW, there exists a convergent subsequence $(y_{k_n})_{n\in\mathbb{N}}$ converging to $y \in D$. Since C is closed, $x_{k_n} \to \bar{x} \in C$. Then $x_{k_n} \to z - y$ and $x_{k_n} \to \bar{x} \in C$. In other words, $z \in C + y \subseteq C + D$. □