Ec 181 | KC Border
Convex Analysis and Economic Theory | AY 2019–2020

Supplement A: Mathematical background

A.1 Extended real numbers

The **extended real number system** $\boldsymbol{R}^\sharp$ consists of the real numbers plus two additional entities ∞ (sometimes known as $+\infty$) and $-\infty$. The ordering of the real numbers is extended so that for any real number α,

$$-\infty < \alpha < \infty.$$

Furthermore we extend the definitions of addition and multiplication as follows. For any real number α,

$$\alpha + \infty = \infty \quad \text{and} \quad \alpha - \infty = -\infty;$$

$$\infty \cdot \alpha = \infty, \text{ if } \alpha > 0 \quad \text{and} \quad \infty \cdot \alpha = -\infty, \text{ if } \alpha < 0;$$

$$-\infty \cdot \alpha = -\infty, \text{ if } \alpha > 0 \quad \text{and} \quad -\infty \cdot \alpha = \infty, \text{ if } \alpha < 0;$$

$$\infty \cdot 0 = 0 = -\infty \cdot 0.$$

The expressions $\infty - \infty$ or $-\infty + \infty$ are undefined, much as division by zero is undefined.

The set $\boldsymbol{R}^\sharp$ of extended reals is equipped with a topology (see Section A.7 below) that makes it a compact set, sometimes referred to as the **two-point compactification** of the reals $\boldsymbol{R}$. The collection of intervals of the form $(n, \infty]$, for $n = 1, 2, \ldots$ constitutes a neighborhood base for ∞. That is, ∞ belongs to the interior of a set A if and only if A includes some interval $(n, \infty]$. Thus

$$x_n \to \infty \quad \text{if} \quad (\forall \alpha \in \boldsymbol{R})\ (\exists N \in \mathbb{N})\ (\forall n \geqslant N)\ [\, x_n > \alpha\,].$$

Likewise, neighborhoods of $-\infty$ include an interval of the form $[-\infty, -n)$, $n = 1, 2, \ldots$. This topology is metrizable, see [1, Example 2.75, p. 57].

A.2 Infimum and supremum

A set A of real numbers is **bounded above** if there exists some real number α, satisfying $\alpha \geqslant x$ for all $x \in A$. In this case we say that α is an **upper bound** for A. If a set has one upper bound it has infinitely many. Indeed, the set of upper bounds is an interval.

Similar definitions apply for lower bounds.

A number is the **greatest element** of A if it belongs to A and is an upper bound for A. A lower bound for A that belongs to A is the **least element** of A. Note that greatest and least elements are unique, for if x and y are both upper bounds that belong to A, then $x \geqslant y$ and $y \geqslant x$, so $x = y$.

The **infimum** of a set A of real numbers, denoted $\inf A$, is the greatest lower bound of A in the set of extended real numbers. That is,

$$(\forall \alpha \in A)\ [\inf A \leqslant \alpha],$$

and for any other extended real β,

$$\Big[(\forall \alpha \in A)\ [\beta \leqslant \alpha]\Big] \implies \beta \leqslant \inf A.$$

The **supremum** of A is the least upper bound of A in the extended real numbers. Note that the definitions (vacuously) imply the following.

$$\inf \varnothing = \infty \quad \text{and} \quad \sup \varnothing = -\infty.$$

The empty set is the only set for which the infimum exceeds the supremum.

N.B. When referring to an upper or lower bound for a set of real numbers, we restrict ourselves to the real numbers. When we discuss suprema and infima we allow the use of extended real numbers.

The real numbers are constructed to satisfy the following:

A.2.1 Fact (The real numbers are complete.) *If a nonempty set of real numbers is bounded above, then it has a supremum (that must necessarily be a real number and not ∞). If a nonempty set of real numbers is bounded below, then it has an infimum (that is a real number and not $-\infty$).*

A.3 Sets associated with functions

The **graph** of a function $f \colon X \to \boldsymbol{R}^\sharp$ is just

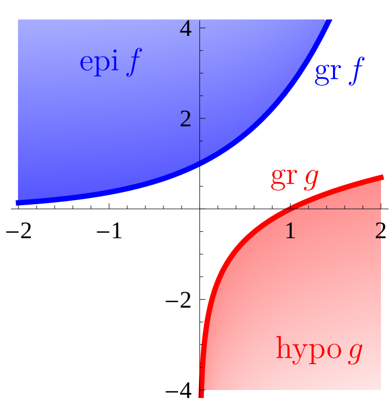
$$\operatorname{gr} f = \{(x, \alpha) \in X \times \boldsymbol{R} : \alpha = f(x)\}.$$

The **epigraph** is

$$\operatorname{epi} f = \{(x, \alpha) \in X \times \boldsymbol{R} : \alpha \geqslant f(x)\},$$

and the **hypograph**[1] is

$$\operatorname{hypo} f = \{(x, \alpha) \in X \times \boldsymbol{R} : \alpha \leqslant f(x)\}.$$



[1] Some authors use the term **subgraph** instead of hypograph, but epi- and hypo- are Greek prefixes, while super- and sub- come from Latin. I will stick with Greek here, since no one says "supergraph."

N.B. The epigraph and hypograph of an extended real-valued function are subsets of $X \times \boldsymbol{R}$, not $X \times \boldsymbol{R}^\sharp$. As a result, the epigraph of the constant function that is identically ∞ is the empty set.

Given a real function $f\colon X \to \boldsymbol{R}$ (or $\boldsymbol{R}^\sharp$), we may use the statisticians' convention where

$$\begin{aligned} \{f = \alpha\} &\quad \text{means} \quad \{x \in X : f(x) = \alpha\}, \\ \{f > \alpha\} &\quad \text{means} \quad \{x \in X : f(x) > \alpha\}, \\ &\qquad \text{etc.} \end{aligned}$$

In particular, in an inner product space, since a vector p also defines a function $p\colon x \mapsto p \cdot x$, we may write

$$\{p = \alpha\} \quad \text{for} \quad \{x \in X : p \cdot x = \alpha\},$$

etc.

A set of the form $\{f = \alpha\}$ is a **level set** of f, $\{f \geqslant \alpha\}$ is a **superlevel set** or an **upper contour set**, and $\{f > \alpha\}$ is a **strict upper contour set** of f. A set of the form $\{f \leqslant \alpha\}$ is a **sublevel set** or an **lower contour set**, and $\{f > \alpha\}$ is a **strict lower contour set** of f.

A.4 Metric spaces

A.4.1 Definition *A **metric** on a nonempty set X is a function $d\colon X \times X \to \boldsymbol{R}$ satisfying the following four properties that are designed to capture our intuitive notion of distance.*

M.1: $d(x,y) \geqslant 0 \quad \text{and} \quad d(x,x) = 0.$

M.2: $d(x,y) = 0 \implies x = y.$

M.3: $d(x,y) = d(y,x).$

M.4: $d(x,y) \leq d(x,z) + d(z,y).$

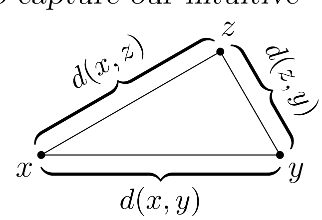


Figure A.4.1. Triangle inequality.

*The pair (X,d) is called a **metric space**.*

A.5 Complete metric spaces

The **diameter** of a set A in a metric space, denoted $\operatorname{diam} A$, is defined to be

$$\operatorname{diam} A = \sup\{d(x,y) : x, y \in Z\}.$$

A sequence $x_1, x_2, \dots$ in the metric space (X,d) is a **Cauchy sequence** if

$$\lim_{n\to\infty} \operatorname{diam}\{x_n, x_{n+1}, \dots\} = 0.$$

A metric space (X,d) is **complete** if every Cauchy sequence converges to a point in X. It is easy to see that any closed subset of a complete metric space is itself complete.

A.5.1 Fact *Each Euclidean space $\boldsymbol{R}^{\mathrm{m}}$ is a complete metric space under the Euclidean metric.*

This is because the real numbers were constructed to be complete. But I don't want to talk about that here.

The next result is a profoundly useful fact about complete metric spaces. Let us say that a sequence $\{A_n\}$ of sets has **v**anishing diameter if

$$\lim_{n\to\infty} \operatorname{diam} A_n = 0.$$

A.5.2 Cantor Intersection Theorem *In a complete metric space, if a decreasing sequence of nonempty closed subsets has vanishing diameter, then the intersection of the sequence is a singleton.*

Proof: Let $\{F_n\}$ be a decreasing sequence (that is, $F_1 \supset F_2 \supset \cdots$) of nonempty closed subsets of the complete metric space (X,d), and assume $\lim_{n\to\infty} \operatorname{diam} F_n = 0$. Then the intersection $F = \bigcap_{n=1}^{\infty} F_n$ cannot have more that one point, for if $a, b \in F$, then $d(a,b) \leqslant \operatorname{diam} F_n$ for each n, so $d(a,b) = 0$, which implies $a = b$.

To see that F is nonempty, for each n pick some $x_n \in F_n$. So for any n, we have

$$\{x_n, x_{n+1}, \dots\} \subset F_n, \text{ so } \operatorname{diam}\{x_n, x_{n+1}, \dots\} \to 0.$$

That is, the sequence (x_n) is Cauchy. Since X is complete there is some $x \in X$ with $x_n \to x$. But x_n belongs to F_m for $n \geqslant m$, and each F_m is closed, so $x = \lim_{n\to\infty} x_n$ belongs to F_m for each m. ∎

A.6 Distance functions

For a nonempty set A in a metric space (X,d), the **distance function** $d(\cdot, A)$ on X is defined by

$$d(x,A) = \inf\Big\{d(x,y) : y \in A\Big\}.$$

Observe that

$$\overline{A} = \{x \in X : d(x,A) = 0\}.$$

A function f from a metric space (X,d) to another metric space (Z,ρ) is **Lipschitz continuous** (or satisfies a **Lipschitz condition**) if there is a **Lipschitz constant** M such that for all $x, y \in X$,

$$\rho\Big(f(x), f(y)\Big) \leqslant M d(x,y).$$

A function is **locally Lipschitz continuous** (or **locally Lipschitzian** if for each $x \in X$, there is a neighborhood of x on which f is Lipschitz continuous.

A.6.1 Theorem *Distance functions are Lipschitz continuous. In particular,*

$$|d(x, A) - d(y, A)| \leqslant d(x, y).$$

Proof: If $x, y \in X$ and $z \in A$, then $d(x, A) \leqslant d(x, z) \leqslant d(x, y) + d(y, z)$. This implies $d(x, A) \leqslant d(x, y) + d(y, A)$, or $d(x, A) - d(y, A) \leqslant d(x, y)$. By symmetry, we have $d(y, A) - d(x, A) \leqslant d(x, y)$, so $|d(x, A) - d(y, A)| \leqslant d(x, y)$. ■

A.7 Topological spaces

You should know that the collection of open subsets of $\boldsymbol{R}^{\mathrm{m}}$ is closed under finite intersections and arbitrary unions. Use that as the motivation for the following definition.

A.7.1 Open sets

A.7.1 Definition *A **topology** τ on a nonempty set X is a family of subsets of X, called **open sets** satisfying*

1. $\varnothing \in \tau$ *and* $X \in \tau$.
2. *The family τ is closed under finite intersections. That is, if $U_1, \dots, U_m$ belong to τ, then $\bigcap_{i=1}^m U_i$ belongs to τ.*
3. *The family τ is closed under arbitrary unions. That is, if U_α, $\alpha \in A$, belong to τ, then $\bigcup_{\alpha \in A} U_\alpha$ belongs to τ.*

*The pair (X, τ) is a **topological space**.*

*The topology τ has the **Hausdorff property** or is a **Hausdorff topology** if for every two distinct points x, y in X there are disjoint open sets U, V with $x \in U$ and $y \in V$.*

*The set A is a **neighborhood** of x if there is an open set U satisfying $x \in U \subset A$. We also say that x is an **interior point** of A.*

*The **interior of A**, denoted $\operatorname{int} A$, is the set of interior points of A.*

*A set is **closed** if its complement is open.*

A.7.2 Fact *It can be shown that the interior of any set A is open (possibly empty), and is indeed the largest open set included in A.*

A.7.3 Lemma *A set is open if and only it is a neighborhood of each of it points.*

The collection of open sets in $\boldsymbol{R}^{\mathrm{m}}$ is a Hausdorff topology. A property of a topological space X that can be expressed in terms of its topology is called a **topological property**.

A.7.2 Closed sets

A set F in a topological space is **closed** if its complement is open. The empty set and the space X are thus closed. Also finite unions and arbitrary intersections of closed sets are closed. Every set A is included in a smallest closed set, called the **closure** of A, denoted $\overline{A}$.

A.7.4 Fact (Properties of closures) *For any sets A and B:*

- $A \subset \overline{A}$.
- *A is closed if and only if $A = \overline{A}$.*
- $\overline{\overline{A}} = \overline{A}$.
- *$A \subset B$ implies $\overline{A} \subset \overline{B}$.*
- $\overline{(A \cup B)} = \overline{A} \cup \overline{B}$.

A.7.3 Continuous functions

A.7.5 Definition *Let X and Y be topological spaces and let $f \colon X \to Y$. Then f is **continuous** if the inverse image of open sets are open. That is, if U is an open subset of Y, then $f^{-1}(U)$ is an open subset of X.*

*The function f is **continuous at x** if the inverse image of every neighborhood of $f(x)$ is a neighborhood of x.*

This corresponds to the usual ε-δ definition of continuity that you are familiar with. Clearly a function is continuous if and only it is continuous at each point. The following lemma is immediate from the definitions.

A.7.6 Lemma *A function $f \colon X \to Y$ is continuous if and only if the inverse image of every closed set is closed.*

A.7.7 Lemma *If $f \colon X \to Y$ is continuous, then for every $A \subset X$, we have $f(\overline{A}) \subset \overline{f(A)}$.*

Proof: Since f is continuous and $\overline{f(A)}$ is closed, $f^{-1}\big(\overline{f(A)}\big)$ is a closed set that clearly includes A, and so includes its closure $\overline{A}$. That is, $\overline{A} \subset f^{-1}\big(\overline{f(A)}\big)$, so $f(\overline{A}) \subset f\Big(f^{-1}\big(\overline{f(A)}\big)\Big) = \overline{f(A)}$. ∎

Base for a topology

A.7.8 Definition *A family $\mathcal{G}$ of open sets is a* ***base*** *(or basis) for the topology τ if every open set in τ is a union of sets from $\mathcal{G}$. A* ***neighborhood base at x*** *is a collection $\mathcal{N}$ of neighborhoods of x such that for every neighborhood G of x there is a neighborhood U of x belong to $\mathcal{N}$ satisfying $x \in U \subset G$.*

In a metric space, the collection of open balls $\{B_\varepsilon(x) : \varepsilon > 0,\ x \in X\}$ is base for the metric topology, and $\{B_{1/n}(x) : n > 0\}$ is a neighborhood base at x.

Given a nonempty family $\mathcal{A}$ of subsets of X there is a smallest topology $\tau_\mathcal{A}$ on X that includes $\mathcal{A}$, called the **topology generated by $\mathcal{A}$**. It consists of arbitrary unions of finite intersections of members of $\mathcal{A}$. If $\mathcal{A}$ is closed under finite intersections, then $\mathcal{A}$ is a base for the topology $\tau_\mathcal{A}$.

Product topology

A.7.9 Definition *If X and Y are topological spaces, the collection sets of the form $U \times V$, where U is an open set in X and V is an open set in Y, is closed under finite intersections, so it is a base for the topology it generates on $X \times Y$, called the* ***product topology****.*

A.7.10 Fact *For the product topology, $\overline{A \times B} = \overline{A} \times \overline{B}$.*

Homeomorphism

A.7.11 Definition *Let X and Y be topological spaces. A function $f : X \to Y$ is a* ***homeomorphism*** *if it is a bijection (one-to-one and onto), is continuous, and its inverse is continuous.*

If f is homeomorphism $U \leftrightarrow f(U)$ is a one-to-one correspondence between the topologies of X and Y. Thus X and Y have the same topological properties. They can in effect be viewed as the same topological space, where f simply renames the points.

Let K be a subset of a topological space. A family $\mathcal{A}$ of sets is a **cover** of K if

$$K \subset \bigcup_{A \in \mathcal{A}} A.$$

If each set in the cover $\mathcal{A}$ is open, then $\mathcal{A}$ is an **open cover** of K. A family $\mathcal{B}$ of sets is a **subcover** of $\mathcal{A}$ if $\mathcal{B} \subset \mathcal{A}$ and $K \subset \cup_{A \in \mathcal{B}} A$.

For example, let K be a subset of $\boldsymbol{R}$, and for each $x \in K$, let $\varepsilon_x > 0$. Then the family $\mathcal{A} = \{(x - \varepsilon_x, x + \varepsilon_x) : x \in K\}$ of open intervals is a open cover of K.

A.7.12 Definition *A set K in a topological space X is* ***compact*** *if for every family $\mathcal{G}$ of open sets satisfying $K \subset \cup \mathcal{G}$ (an* ***open cover*** *of K), there is a finite subfamily $\{G_1, \dots, G_k\} \subset \mathcal{G}$ with $K \subset \bigcup_{i=1}^k G_i$ (a* ***finite subcover*** *of K).*

A.7.13 Lemma *A closed subset of a compact set is compact.*

Proof: Let K be compact and $F \subset K$ be closed. Let $\mathcal{G}$ be an open cover of F. Then $\mathcal{G} \cup \{F^c\}$ is an open cover of K. Let $\{G_1, \ldots, G_k, F^c\}$ be a finite subcover of K. Then $\{G_1, \ldots, G_k\}$ is a finite subcover of F. ∎

A.7.14 Lemma *A compact subset of a Hausdorff space is closed.*

Proof: Let K be compact, and let $x \notin K$. Then by the Hausdorff property, for each $y \in K$ there are disjoint open sets U_y and V_y with $y \in U_y$ and $x \in V_y$. By compactness there are $y_1, \ldots, y_k$ with $K \subset \bigcup_{i=1}^k U_{y_i} = U$. Then $V = \bigcap_{i=1}^k V_{y_i}$ is an open set satisfying $x \in V \subset U^c \subset K^c$. That is, K^c is a neighborhood of x. Since x is an arbitrary member of K^c, we see that K^c is open (Lemma A.7.3), so K is closed. ∎

A.7.15 Lemma *Let $f\colon X \to Y$ be continuous. If K is a compact subset of X, then $f(K)$ is a compact subset of Y.*

Proof: Let $\mathcal{G}$ be an open cover of $f(K)$. Then $\{f^{-1}(G) : G \in \mathcal{G}\}$ is an open cover of K. Let $\{f^{-1}(G_1), \ldots, f^{-1}(G_k)\}$ be a finite subcover of K. Then $\{G_1, \ldots, G_k\}$ is a finite subcover of $f(K)$. ∎

Relative topology

A.7.16 Definition (Relative topology) *If (X, τ) is a topological space and $A \subset X$, then (A, τ_A) is a topological space with its* ***relative topology****, where $\tau_A = \{G \cap A : G \in \tau\}$.*

Not that if τ is a Hausdorff topology, then τ_A is also a Hausdorff topology.

A.7.17 Lemma *If (X, τ) is a topological space and $K \subset A \subset X$, then K is a compact subset of (A, τ_A) if and only if it is a compact subset of (X, τ).*

Proof: Assume K is a compact subset of (X, τ). Let $\mathcal{G}$ be a τ_A-open cover of K in A. For each $G \in \mathcal{G}$ there is some $U_G \in \tau$ with $G = U_G \cap A$. Then $\{U_G : G \in \mathcal{G}\}$ is a τ-open cover of K in X, so it has a finite subcover $U_{G_1}, \ldots, U_{G_k}$. But then $G_1, \ldots, G_k$ is a finite subcover of K in A.

The converse is similar. ∎

A.7.18 Lemma *Let $f\colon X \to Y$ be one-to-one and continuous, where Y is a Hausdorff space and X is compact. The $f\colon X \to f(X)$ is a homeomorphism, where $f(X)$ has its relative topology as a subset of Y.*

Proof: We need to show that the function $f^{-1}: f(X) \to X$ is continuous. So let G be any open subset of X. We must show that $(f^{-1})^{-1}(G) = f(G)$ is open in $f(X)$. Now G^c is a closed subset of X, and thus compact. Therefore $f(G^c)$ is compact, and since Y is Hausdorff, so is $f(X)$, so $f(G^c)$ is a closed subset of Y. Now $f(X) \cap f(G^c)^c = f(G)$, so $f(G)$ is open in $f(X)$. ■

There is an equivalent characterization of compact sets that is perhaps more useful. A family $\mathcal{A}$ of sets has the **finite intersection property** if every finite subset $\{A_1, \dots, A_n\}$ of $\mathcal{A}$ has a nonempty intersection, $\bigcap_{i=1}^n A_i \neq \varnothing$.

A.7.19 Theorem *A set K is compact if and only if every family of closed subsets of K having the finite intersection property has a nonempty intersection.*

Proof: Start with this observation: Let $\mathcal{A}$ be an arbitrary family of subsets of K, and define $\overline{\mathcal{A}} = \{K \setminus A : A \in \mathcal{A}\}$. By de Morgan's Laws $\bigcap_{A \in \mathcal{A}} A = \varnothing$ if and only if $K = \bigcup_{B \in \overline{\mathcal{A}}} B$. That is, $\mathcal{A}$ has an empty intersection if and only if $\overline{\mathcal{A}}$ covers K.

($\Longrightarrow$) Assume K is compact and let $\mathcal{F}$ be a family of closed subsets of K. Then $\overline{\mathcal{F}}$ is a family of relatively open sets of K. If $\mathcal{F}$ has the finite intersection property, by the above observation, no finite subset of $\overline{\mathcal{F}}$ can cover K. Since K is compact, this implies that $\overline{\mathcal{F}}$ itself cannot cover K. But then by the observation $\mathcal{F}$ has nonempty intersection.

($\Longleftarrow$) Assume that every family of closed subsets of K having the finite intersection property has a nonempty intersection, and let $\mathcal{G}$ be an open cover of K. Then $\overline{\mathcal{G}}$ is a family of closed having an empty intersection. Thus $\overline{\mathcal{G}}$ cannot have the fintiie intersection property, so there is a finite subfamily $\overline{\mathcal{G}_0}$ of $\overline{\mathcal{G}}$ with empty intersection. But then $\overline{\overline{\mathcal{G}_0}}$ is a finite subfamily of $\mathcal{G}$ that covers K. Thus K is compact. ■

A.7.20 Corollary *Let K_n be a decreasing sequence of nonempty compact sets. That is, $K_1 \supset K_2 \supset \cdots$. Then $\bigcap\limits_{n=1}^{\infty} K_n \neq \varnothing$.*

Proof: Cleary the sequence has the finite intersection property. ■

A.8 Semicontinuity

A real function $f: X \to \boldsymbol{R}$ is **upper semicontinuous** if for each $\alpha \in \boldsymbol{R}$, the **superlevel set** $\{f \geqslant \alpha\}$ is closed. It is **lower semicontinuous** if every **sublevel set** $\{f \leqslant \alpha\}$ is closed. Combine this with Section 13.4 and move to a self-contained location.

The **epigraph** of a real-valued function f is defined by

$$\operatorname{epi} f = \{(x, \alpha) \in X \times \boldsymbol{R} : f(x) \leqslant \alpha\}$$

and its **hypograph** is defined by

$$\operatorname{hypo} f = \{(x, \alpha) \in X \times \boldsymbol{R} : f(x) \geqslant \alpha\}$$

Note that

$$(x, \alpha) \in \operatorname{epi} f \iff x \in \{f \leqslant \alpha\}.$$

Likewise $(x, \alpha) \in \operatorname{hypo} f \iff x \in \{f \geqslant \alpha\}$.

A.8.1 Proposition *The real-valued function $f \colon X \to \boldsymbol{R}$ is lower semicontinuous if and only its epigraph is a closed subset of $X \times \boldsymbol{R}$. Similarly, f is upper semicontinuous if and only if its hypograph is a closed subset of $X \times \boldsymbol{R}$.*

Proof: Assume $\operatorname{epi} f$ is closed and let x_n be a sequence in the sublevel set $\{f \leqslant \alpha\}$ with $x_n \to x$. Then (x_n, α) is a sequence in $\operatorname{epi} f$ converging to (x, α). Since $\operatorname{epi} f$ is closed, (x, α) belongs to $\operatorname{epi} f$, so $x \in \{f \leqslant \alpha\}$. Thus $\{f \leqslant \alpha\}$ is closed.

Conversely, assume that each sublevel set is closed, and let (x_n, α_n) be a sequence in $\operatorname{epi} f$ converging to (x, α). Let $\varepsilon > 0$. Then for large enough n, $\alpha_n < \alpha + \varepsilon$, so for large n, $(x_n, \alpha_n) \in \{f \leqslant \alpha + \varepsilon\}$, which is closed. Thus $(x, \alpha) \in [f \leqslant \alpha + \varepsilon]$. Since this must be true for every $\varepsilon > 0$, we have $(x, \alpha) \in \{f \leqslant \alpha\}$, so $(x, \alpha) \in \operatorname{epi} f$. Thus $\operatorname{epi} f$ is closed. ∎

Note that f is upper semicontinuous if and only if $-f$ is lower semicontinuous.

We can also talk about semicontinuity at a point. The real-valued function f is **upper semicontinuous at the point $\boldsymbol{x}$** if

$$(\forall \varepsilon > 0)\ (\exists \delta > 0)\ [\, d(y, x) < \delta \implies f(y) < f(x) + \varepsilon \,].$$

Similarly, f is **lower semicontinuous at the point $\boldsymbol{x}$** if

$$(\forall \varepsilon > 0)\ (\exists \delta > 0)\ [\, d(y, x) < \delta \implies f(y) > f(x) - \varepsilon \,].$$

Equivalently, f is upper semicontinuous at x if

$$f(x) \geqslant \limsup_{y \to x} f(y) = \inf_{\varepsilon > 0} \sup_{0 < d(y,x) < \varepsilon} f(y).$$

Similarly, f is lower semicontinuous at x if

$$f(x) \leqslant \liminf_{y \to x} f(y) = \sup_{\varepsilon > 0} \inf_{0 < d(y,x) < \varepsilon} f(y).$$

A.8.2 Proposition *A real valued function is continuous at x if and only if is both upper and lower semicontinuous at x.*

A.8.3 Proposition *A function is lower semicontinuous if and only if it is lower semicontinuous at every point. Likewise, a function is upper semicontinuous if and only if it is upper semicontinuous at every point.*

A.8.4 Proposition *Let $\{f_i : i \in I\}$ be a family of lower semicontinuous functions. Then g defined by pointwise by*

$$g(x) = \sup_i f_i(x)$$

is lower semicontinuous.

Similarly, if $\{f_i : i \in I\}$ is a family of upper semicontinuous functions, then g defined by pointwise by $h(x) = \inf_i f_i(x)$ is upper semicontinuous.

Proof: For the lower semicontinuous case this follows from the fact that $\operatorname{epi} g = \bigcap_i \operatorname{epi} f_i$. For the upper semicontinuous case use $\operatorname{hypo} h = \bigcap_i \operatorname{hypo} f_i$. ∎

A.9 Weierstrass's Theorem

A.9.1 Weierstrass' Theorem *Let K be a compact set, and let $f\colon K \to \boldsymbol{R}$ be continuous. Then f achieves both a maximum and a minimum in K.*

Proof: I will prove that f achieves a maximum, the proof for a minimum is similar. For each $\alpha \in \boldsymbol{R}$, let

$$F_\alpha = \{x \in K : f(x) \geqslant \alpha\}.$$

Since f is continuous, each set F_α is a closed subset of K, and if $\alpha \in \operatorname{range} f$, then F_α is nonempty. Observe that the family

$$\mathcal{F} = \{F_\alpha : \alpha \in \operatorname{range} f\} \text{ has the finite intersection property.}$$

For if $\alpha^* = \max_{i=1,\dots,m} \alpha_i$, [2] then $\bigcap_{i=1}^{n} F_{\alpha_i} = F_{\alpha^*}$. Thus by Theorem A.7.19 $\mathcal{F}$ has a nonempty intersection. Now if $x \in \bigcap \mathcal{F} = \bigcap_{\alpha \in \operatorname{range} f} F_\alpha$, then $f(x) \geqslant \alpha$ for every $\alpha \in \operatorname{range} f$. In other words, x maximizes f over K. ∎

Note that we only used continuity to show that the sets F_α are closed. This requires only **upper semicontinuity** of f. Also, the set K itself need not be compact, as long as there exists some α such that F_α is compact.

A.10 Compactness in metric spaces

A nonempty subset of a metric space is **totally bounded** if for every $\varepsilon > 0$, it can be covered by finitely many ε-balls. Total boundedness is not a topological property—it depends on the particular metric. However compactness, which is a topological property, is characterized by the conjunction of two metric-dependent properties.

[2] Prove that such an α^* exists. That is, prove that any finite set of real numbers has a greatest element.

A.10.1 Theorem *A metric space is compact if and only if it is complete and totally bounded.*

A proof may be found, for instance, in [1, Theorem 3.28, p. 86].

It is easy to see that any bounded subset of real numbers (with the usual metric) is totally bounded. Consequently, every bounded subset of $\boldsymbol{R}^{\mathrm{m}}$ (with the Euclidean metric) is totally bounded. Thus we have as a corollary, the following well-known result, see, e.g., Rudin [5, Theorem 2.41, p. 40].

A.10.2 Heine–Borel–Lebesgue Theorem *A subset of $\boldsymbol{R}^{\mathrm{m}}$ is compact if and only if it is both closed and bounded in the Euclidean metric.*

This result is special. In general, a subset of a metric space may be closed and bounded without being compact. (Consider the coordinate vectors in ℓ_∞.)

A.11 Topological vector spaces

A.11.1 Definition *A (real)* ***topological vector space*** *(abbreviated* ***tvs****) is a vector space X together with a topology τ where τ has the property that scalar multiplication and vector addition are continuous functions. That is, the mappings*

$$(x, \alpha) \mapsto \alpha x$$

from $X \times \boldsymbol{R}$ to X and

$$(x, y) \mapsto x + y$$

from $X \times X$ to X are continuous. (Where, of course, $\boldsymbol{R}$ has its usual topology, and $\boldsymbol{R} \times X$ and $X \times X$ have their product topologies.)

For a detailed discussion of topological vector spaces, see chapter five of the *Hitchhiker's Guide* [1]. But here are some of the results we will need.

A.11.2 Definition *A set A in the vector space X is* ***circled*** *if for each $x \in A$ the line segment joining the points x and $-x$ lies in A.*

Let X be a topological vector space. The for each $\alpha \neq 0$ the mapping $x \mapsto \alpha x$ is continuous, and so is its inverse $x \mapsto (1/\alpha)x$. This means that each is a homeomorphism, so that if W is an open set, then so is αW.

Let $f \colon X \times \boldsymbol{R} \to X$ be scalar multiplication, $f(x, \alpha) = \alpha x$, so that $f(0, 0) = 0$. The continuity of f guarantees that if V is a neighborhood of zero in X, then $f^{-1}[V]$ is a neighborhood of $(0, 0)$ in $X \times \boldsymbol{R}$. Thus there is an open neighborhood W of zero in X and an open neighborhood $(-\delta, \delta)$ of zero in $\boldsymbol{R}$ such that $x \in W$ and $|\alpha| \leqslant \delta$ imply $\alpha x \in V$. Now setting $U = \bigcup_{0<|\alpha|\leqslant\delta} \alpha W$, we see that U is open and contains 0 (since each αW is open and contains 0), $U \subset V$, and U is circled. Therefore we have proven the following.

A.11.3 Lemma *In any tvs, every neighborhood of zero includes an open circled neighborhood of zero.*

(In $\boldsymbol{R}^{\mathrm{m}}$, a ball of radius ε centered at zero is a circled neighborhood of zero.)

For each vector y the maps $x \mapsto x + y$ and $x \mapsto x - y$ are continuous and mutual inverses, and so homeomorphisms. Thus a set G is open if and only its translation $y + G$ is open. Therefore the topology on X is completely determined by its neighborhoods of zero and *a linear mapping between topological vector spaces is continuous if and only if it is continuous at zero.*

A.12 Continuity of the coordinate map

A.12.1 Lemma *Let X be a Hausdorff tvs, and let $\{x_1, \ldots, x_n\}$ be a linearly independent subset of X. Let α_m be a sequence in $\boldsymbol{R}^{\mathrm{n}}$. Then*

$$\sum_{i=1}^{n} \alpha_{mi}\, x_i \xrightarrow[m\to\infty]{} \sum_{i=1}^{n} \alpha_i x_i \implies \alpha_{mi} \xrightarrow[m\to\infty]{} \alpha_i,\ i = 1, \ldots, n.$$

When X is already some $\boldsymbol{R}^{\mathrm{m}}$, there is a simple proof of the lemma.

Proof of Lemma for $\boldsymbol{R}^{\mathrm{m}}$: Let X be the $m \times n$ matrix whose j^{th} column is x_j. By the theory of ordinary least squares estimation if $x = X\alpha = \sum_{j=1}^{n} \alpha_j x_j$ is a linear combination of $\{x_1, \ldots, x_n\}$, then the coordinate mapping $T(x)$ is given by

$$T(x) = (X'X)^{-1}X'x,$$

which is clearly continuous. ∎

The lemma is rather delicate—it can fail if either X is not Hausdorff or $\{x_1, \ldots, x_n\}$ is dependent.

A.12.2 Example Let $X = \boldsymbol{R}^2$ under the semi-metric $d\big((x,y),(x',y')\big) = |x - x'|$. (This topology is not Hausdorff.) Then X is a topological vector space. Let $x_1 = (1,0)$ and $x_2 = (0,1)$ be the unit coordinate vectors. Then $\frac{1}{m}x_1 + 0x_2 = (1/m, 0) \to (0,1) = 0x_1 + 1x_2$, (since $d\big((1/m,0),(0,1)\big) = 1/m$, but the second coordinates do not converge ($0 \not\to 1$). □

A.12.3 Example Let $X = \boldsymbol{R}^2$ with the Euclidean topology and let $x_1 = (1,0)$ and $x_2 = (-1,0)$. Then $nx_1 + nx_2 = (0,0) \to (0,0) = 0x_1 + 0x_2$, but $n \not\to 0$. □

A.13 $\boldsymbol{R}^{\mathrm{m}}$ is a Hilbert space

A.13.1 Inner product

A.13.1 Definition *A real linear space V has an* ***inner product*** *if for each pair of vectors x and y there is a real number, traditionally denoted (x, y), satisfying the following properties.*

IP.1 $(x, y) = (y, x)$.

IP.2 $(x, y + z) = (x, y) + (x, z)$.

IP.3 $\alpha(x, y) = (\alpha x, y) = (x, \alpha y)$.

IP.4 $(x, x) > 0$ *if* $x \neq 0$.

It is unfortunate that the traditional inner product notation is the same as that for an open line segment. That is one reason I often use the ***dot product*** *notation*

$$x \cdot y.$$

A vector space V *equipped with an inner product is called an* ***inner product space****.*

For a complex vector space, the inner product is complex-valued, and property (1) is replaced by $(x, y) = \overline{(y, x)}$*, where the bar denotes complex conjugation, and (3) is replaced by* $\alpha(x, y) = (\alpha x, y) = (x, \overline{\alpha} y)$*.*

Vectors x *and* y *are said to be* ***orthogonal*** *if* $(x, y) = 0$*.*

The next result may be found, for instance, in [3, Theorem 1.8, p. 16].

A.13.2 Cauchy–Schwartz Inequality *For any real inner product,*

$$(x, y)^2 \leqslant (x, x)(y, y) \tag{1}$$

with equality if and only if x *and* y *are dependent.*

Proof: If either x or y is zero, then we have equality, so assume x, y are both nonzero. Define the quadratic polynomial $Q \colon \boldsymbol{R} \to \boldsymbol{R}$ by

$$Q(\lambda) = (\lambda x + y, \lambda x + y) = (x, x)\lambda^2 + 2(x, y)\lambda + (y, y).$$

By Property IP.4, $Q(\lambda) \geqslant 0$ for each $\lambda \in \boldsymbol{R}$. Therefore the discriminant of Q is nonpositive,[3] that is, $4(x, y)^2 - 4(x, x)(y, y) \leqslant 0$, or $(x, y)^2 \leqslant (x, x)(y, y)$. Equality in (1) can occur only if the discriminant is zero, in which case Q has a real root. That is, there is some λ for which $Q(\lambda) = (\lambda x + y, \lambda x + y) = 0$. But this implies that $\lambda x + y = 0$, which means the vectors x and y are linearly dependent. ∎

A.13.3 Definition *A* ***norm*** $\|x\|$ *is a real function on a vector space that satisfies:*

N.1 $\|0\| = 0$

N.2 $\|x\| > 0$ *if* $x \neq 0$

[3] In case you have forgotten how you derived the quadratic formula in Algebra I, rewrite the polynomial as

$$f(z) = \alpha z^2 + \beta z + \gamma = \tfrac{1}{\alpha}\big(\alpha z + \tfrac{\beta}{2}\big)^2 - (\beta^2 - 4\alpha\gamma)/4\alpha,$$

and note that the only way to guarantee that $f(z) \geqslant 0$ for all z is to have $\alpha > 0$ and $\beta^2 - 4\alpha\gamma \leqslant 0$.

N.3 $\|\alpha x\| = |\alpha| \, \|x\|$.

N.4 $\|x + y\| \leqslant \|x\| + \|y\|$ *with equality if and only if* $x = 0$ *or* $y = 0$ *or* $y = \alpha x$, $\alpha > 0$.

A.13.4 Definition *In a normed space, the unit ball* U *is the set*

$$U = \{x : \|x\| \leqslant 1\}.$$

A.13.5 Proposition *In a normed space, the unit ball is a convex set.*

A.13.6 Proposition *If* $(\cdot,\cdot)$ *is an inner product, then* $\|x\| = (x,x)^{\frac{1}{2}}$ *is a norm.*

A.13.7 Proposition *If* $\|\cdot\|$ *is a norm, then*

$$d(x,y) = \|x - y\|$$

is a metric.

The natural metric on $\boldsymbol{R}$ is

$$d(x,y) = |x - y|.$$

A.13.8 Definition *An inner product space is a* ***Hilbert space*** *if the metric induced by the inner produce is complete.*

(A metric space is **complete** if every Cauchy sequence has a limit point in the space.) Every finite dimensional Euclidean space is a Hilbert space. So is ℓ_2, the space of all sequences $x = (x_1, x_2, \dots)$ such that $\sum_{n=1}^{\infty} x_n^2 < \infty$.

A.14 Parallelogram Law

A.14.1 Parallelogram Law *In an inner product space (such as* $\boldsymbol{R}^{\mathrm{m}}$*), for any vectors* x *and* y *we have*

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

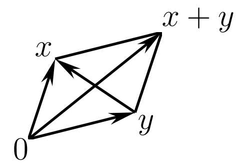
Proof: Note that

$$(x + y) \cdot (x + y) = x \cdot x + 2x \cdot y + y \cdot y$$

and

$$(x - y) \cdot (x - y) = x \cdot x - 2x \cdot y + y \cdot y.$$

Add these two equations and restate in terms of norms. ∎



This is called the Parallelogram Law because it asserts that the sum of the squares of the lengths of the diagonals of a parallelogram is equal to the sum of the squares of the lengths of the sides. Consider the parallelogram with vertices 0, x, y, and $x+y$. Its diagonals are the segments $[0, x+y]$ and $[x, y]$, and their lengths are $\|x+y\|$ and $\|x-y\|$. It has two sides of length $\|x\|$ and two of length $\|y\|$.

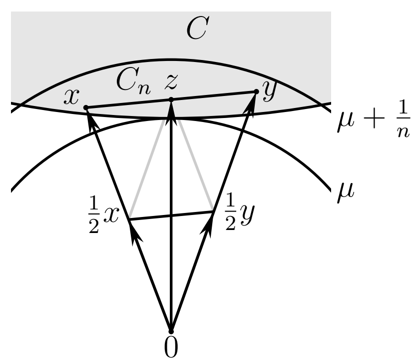
As an aside, a norm on a vector space is induced by an inner product if and only if it satisfies the parallelogram law; see for instance [2, Problem 32.10, p. 303].

A.15 Metric projection in a Hilbert space

A.15.1 Theorem *Let C be a closed convex subset of a Hilbert space H. Then there is a unique point in C of least norm.*

Proof: It is simpler to work with the square of the norm, so let

$$\mu = \inf\Big\{\|x\|^2 : x \in C\Big\}.$$



For each n, define

$$C_n = C \cap \Big\{x \in H : \|x\|^2 \leqslant \mu + \tfrac{1}{n}\Big\}.$$

By the definition of μ each C_n is nonempty, and each is closed and convex as the intersection of two closed convex sets, and finally note that $C_1 \supset C_2 \supset \cdots$. Moreover, for any point $x \in C_n$, we have $\mu \leqslant \|x\|^2 \leqslant \mu + \frac{1}{n}$. Let us now proceed to bound the diameter of C_n: Let x and y belong to C_n, and set $z = \frac{1}{2}x + \frac{1}{2}y$. Apply the Parallelogram Law to the points $\frac{1}{2}x$ and $\frac{1}{2}y$:

$$\|\underbrace{\tfrac{1}{2}x + \tfrac{1}{2}y}_{z}\|^2 + \|\tfrac{1}{2}x - \tfrac{1}{2}y\|^2 = 2\|\tfrac{1}{2}x\|^2 + 2\|\tfrac{1}{2}y\|^2$$

or

$$\tfrac{1}{4}\|x-y\|^2 = \tfrac{1}{2}\|x\|^2 + \tfrac{1}{2}\|y\|^2 - \|z\|^2. \tag{2}$$

Since C_n is convex, $z \in C_n$, too, so $\|x\|^2$, $\|y\|^2$, $\|z\|^2$ all lie in the interval $\Big[\mu, \mu + \frac{1}{n}\Big]$, so the right-hand side of (2) lies in the interval $\Big[0, \frac{1}{n}\Big]$. This implies $\|x-y\|^2 \leqslant 4/n$. Therefore,

$$\operatorname{diam} C_n \leqslant \frac{2}{\sqrt{n}} \xrightarrow[n\to\infty]{} 0.$$

So by the Cantor Intersection Theorem, $\bigcap_n C_n$ consists of a single point, call it c. Then $\|c\|^2 = \mu$ and c is the unique norm minimizer in C. ∎

A.15.2 Corollary *Let C be a closed convex subset of a Hilbert space H. Then for each point $x \in H$ there is a unique point y in C closest to x. The point y is called the* ***metric projection*** *of x on C.*

Proof: Apply Theorem A.15.1 to $C - x$. If y has least norm in $C - x$, then $x + y$ is the point in C closest to x, since distances are not affected by translation. ∎

References

[1] C. D. Aliprantis and K. C. Border. 2006. *Infinite dimensional analysis: A hitchhiker's guide*, 3d. ed. Berlin: Springer–Verlag.

[2] C. D. Aliprantis and O. Burkinshaw. 1999. *Problems in real analysis*, 2d. ed. San Diego: Academic Press. Date of publication, 1998. Copyright 1999.

[3] T. M. Apostol. 1969. *Calculus*, 2d. ed., volume 2. Waltham, Massachusetts: Blaisdell.

[4] R. B. Holmes. 1975. *Geometric functional analysis and its applications*. Number 24 in Graduate Texts in Mathematics. Berlin: Springer–Verlag.

[5] W. Rudin. 1976. *Principles of mathematical analysis*, 3d. ed. International Series in Pure and Applied Mathematics. New York: McGraw Hill.