VIII. Analysis: Sets and Spaces

Real analysis is all about sets in Euclidean space, which I define as “the real space as we know it.” Here are some refreshers on Euclidean space. It is an \( \mathbb{R}^n \)-space. The **length** of a vector \( x \) is defined as:

\[
\|x\| = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}
\]

The **distance** between two vectors \( x \) and \( y \) defined as:

\[
d(x, y) = \|x - y\| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2}
\]

Finally, the **dot product** of two vectors \( x \) and \( y \) is:

\[
x \cdot y = x_1y_1 + x_2y_2 + \ldots + x_ny_n
\]

The remaining topic to discuss is how we build up sets in Euclidean space. First of all, the **complement** of a set \( S \subseteq \mathbb{R}^n \) consists of all points not in that set, \( \{x \in \mathbb{R}^n : x \notin S\} \), and is denoted by \( \mathbb{R}^n \setminus S \) or \( S^c \). The **open ball of radius** \( r \) **centered at** \( x \) is defined as:

\[
B_r(x) = \{y \in \mathbb{R}^n : \|x - y\| < r\}
\]

An open ball is also called an **open neighborhood**. When talking about what goes on really close to a point, we talk about balls of radius \( \varepsilon \), also called \( \varepsilon \)-**balls**. Terms that include the word “locally” usually signal that an \( \varepsilon \)-ball is involved.

**Example:** The preference relation \( \succeq \) on the consumption set \( X \subseteq \mathbb{R}^n_+ \) is called **locally non-satiated** if \( \forall x_i \in X_i \) and \( \forall \varepsilon > 0 \), \( \exists x_i' \in X_i \cap B_\varepsilon(x_i) \) s.t.: \( x_i' \succeq_i x_i \).

Given some set \( S \subseteq \mathbb{R}^n \), a point \( x \in S \) is called an **interior point** of \( S \) if there exists a \( B_{\varepsilon}(x) \), \( \varepsilon > 0 \), such that \( B_{\varepsilon}(x) \subseteq S \). A point \( y \in \mathbb{R}^n \) is a **boundary point** of \( S \) if every \( B_{\varepsilon}(y) \), \( \varepsilon > 0 \)—no matter how small \( \varepsilon \) is—contains points in \( S \) and points outside \( S \).

These are used to define the two major classifications of sets. A set \( S \subseteq \mathbb{R}^n \) is called an **open set** if all its points are interior points. That is, for any \( x \in S \) there exists some \( B_{\varepsilon}(x) \), \( \varepsilon > 0 \) such that \( B_{\varepsilon}(x) \subseteq S \).
A set $T \subseteq \mathbb{R}^n$ is called a **closed set** if the set $\mathbb{R}^n \setminus T$ is open. Equivalently, $T \subseteq \mathbb{R}^n$ is closed if and only if it contains all its boundary points. When I was a little kid, my brother and I would torture our little sister. She could go screaming to our parents if we touched her, so the goal of the game we played was to try to hover our hands as close to her as we possibly could without actually touching her skin. (Then we could defend ourselves by saying, “but I didn’t even touch her!”) She was like a closed set—her skin, the boundary, was part of her. The rest of the universe was an open set. It was always possible to get just a little closer to the skin without actually touching her. If you ever did this to your younger sibling, you have a good idea of what mathematicians mean by “epsilon arbitrarily greater than zero,” since that was the ideal distance you’d keep between you and your sibling.

Moving on from the games that Rotten Children play, here’s a piece of trivia. Some sets, like $\emptyset$ and $\mathbb{R}$, fit both the definition of closed and open. The empty set contains no points, so all zero of these points are interior—it’s trivially an open set. For any point $x \in \mathbb{R}$, the entire interval $(x-\varepsilon, x+\varepsilon)$ is contained within $\mathbb{R}$, so $x$ is interior. The complement of each must be closed then. Of course, $\mathbb{R}$ and $\emptyset$ are each other’s complements. (Sets that are both closed and open are given the strange name of **clopen**.) In particular, this is useful to keep in mind for this reason: when a set goes off to infinity in some direction, this can still meet the definition of either kind of set.

$(-\infty, 2)$ is an open set, and $(-\infty, 2]$ is a closed set.

The **interior** of a set $S \subseteq \mathbb{R}^n$ consists of all points that are interior points of $S$, and is denoted by $\text{int}(S)$. From the definition, a set $S$ is an open set if and only if $S = \text{int}(S)$. The **boundary** of a set $T \subseteq \mathbb{R}^n$ consists of all boundary points of $T$ and is usually denote by $\partial T$. The **closure** of $T$ is $T \cup \partial T$. This is denoted by $\overline{T}$ or $\text{clos}(T)$. A set $T$ is a closed set if and only if $T = \text{clos}(T)$.

Notes on unions and intersections of open and closed sets:

**Theorem:** The union of open sets is open.

**Theorem:** The intersection of a finite number of open sets is open.

**Theorem:** The union of a finite number of closed sets is closed.
Theorem: The intersection of closed sets is closed.

A set \( S \subseteq \mathbb{R}^n \) is called **bounded** if there exists some \( M > 0 \) such that \( S \subseteq B_M(x) \). In other words, it doesn’t go off forever in any direction. It is **compact** if it is both closed and bounded. Compact sets are very nice and neat. One nice property is that all continuous functions achieve a maximum on closed set. Consider the utility maximization problem with:

\[
U(x_1, x_2) = x_1^{1/2} x_2^{1/2}
\]

With one of these budget sets:

\[
\{(x_1, x_2) \in \mathbb{R}^n : x_2 \leq 12\}
\]

\[
\{(x_1, x_2) \in \mathbb{R}^n : 2x_1 + 3x_2 < 12\}
\]

The first of these isn’t compact because it isn’t bounded. The utility maximizing agent would say, “I want a million units of \( x_1 \)! No, a million and one units! No, even more!” He’ll never be satisfied. The second budget set isn’t compact because it isn’t closed. Here, the individual is told that he can’t spent all of his twelve dollars. He sits down and works out the utility maximization problem, leaving a penny in change, and finds that he demands 2.995 units of \( x_1 \). But he can do better than that if he leaves only a tenth of a penny—in this case, he would demand 2.9995 units. But leaving only a hundredth of a penny could give an even better outcome! And so it goes, as he tries to get really close to spending all his money without spending everything.

A **sequence** in \( \mathbb{R}^n \) assigns to every natural number \( m = 1, 2, 3, \ldots \) a vector \( x^m \in \mathbb{R}^n \). The superscript represents what number this is in the sequence. Sequences are usually denoted with curly brackets, in one of these ways:

\[
\{x^m\}_{m=1}^{\infty}, \{x^m\}_{m \in \mathbb{N}}, \{x^m\}
\]

Some possible sequences might begin like these:

\[
\{1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \ldots\}
\]

\[
\{-1, 1, -1, 1, -1, 1, \ldots\}
\]

\[
\{1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \ldots\}
\]
The sequence might follow a rule (the first four can be written in terms of a rule) and there might be absolutely no sense to it.

The sequence \( \{x^m\} \) converges to some point \( \bar{x} \in \mathbb{R}^n \) if for all \( \varepsilon > 0 \) there exists some number \( M \) that the distance \( d(x^m, \bar{x}) < \varepsilon \) for any \( n > M \). This is expressed \( \lim_{m \to \infty} x^m = \hat{x} \) or \( x^m \to \bar{x} \). The point \( \bar{x} \) is called the limit of the sequence. What this definition says is that no matter how small you pick \( \varepsilon \), even if it’s a millionth of a millionth, at some point the sequence gets (and stays) at least that close to the limit. Which of the sequences above converges?

A very similar concept to a convergent sequence is a Cauchy sequence. We say that the sequence \( \{x^m\} \) is Cauchy if for any \( \varepsilon > 0 \) there exists some number \( M \) that \( d(x^m, x^{m+h}) < \varepsilon \) for any \( m > M, h \geq 1 \). This says that each element of the sequence gets as close as you like to the ones following it in the sequence. In contrast, the convergent sequence gets arbitrarily close to some limit. They concepts are very similar. I think the rules are that every convergent sequence is Cauchy; and that in so-called complete spaces (such as the Euclidean space with which we are familiar) convergent is equivalent to Cauchy.

**Theorem:** A set \( T \subseteq \mathbb{R}^n \) is closed if and only if every for every convergent sequence \( x^m \to \bar{x} \) with \( x^m \in T \ \forall m \), then \( \bar{x} \) is also in \( T \).

Look back to the example of the guy maximizing utility subject to
\[
\{ (x_1, x_2) \in \mathbb{R}_+^2 : 2x_1 + 3x_2 < 12 \}.
\]

His choices for \( (x_1, x_2) \) are following a sequence \( \{(2.95,1.95), (2.995,1.995), (2.9995,1.9995), \ldots \} \) that converging to the point \( (3,2) \). That point isn’t in the budget set, though all the numbers in the sequence are. This verifies that the budget set isn’t closed and thus not compact.

Why is compactness such a nice property? One of the reasons is known as **Weierstraß’ theorem** or the **extreme value theorem**:

**Theorem:** Let \( f : X \to \mathbb{R} \) be a continuous function on a compact set \( X \subseteq \mathbb{R}^k \). Then there exist maximum and minimum points for \( f \) in \( X \).
This is sometimes stated as “a continuous function achieves its maximum on a compact set.” Provided the budget set is compact, there does exist a utility maximizing bundle.

Occasionally you'll see theorems and definitions referring to strange abstract spaces. So why do we introduce these weird things, and why do we care about them?

Well, one reason is that math is like a bunch of sports. In order that everyone can play together even when they’re not really trained in the other areas, it’s useful to consider characteristics that define broad categories of sports. For instance, think about the sports characterized by the phrase “the objective is to move a relatively large ball violently toward a goal.” Soccer, rugby, and American football all fall into this category. If you think about the general objective, you come very quickly to some principles that apply to all of them: “it would probably be a good idea to put my defensive players between the opponents and the goal when they have the ball” or “it would probably be a good idea to get the ball out of my opponents’ control,” for instance. These statements apply to all games that share that characteristic, but would be nonsensical in other games (like baseball: you want to keep the player, not the ball, from getting to the homeplate; also, the runner wants to avoid contacting the ball when opponents have it). The immediate advantage of knowing the general principles of all the games in a category is that if you were ever asked to play a different game that also has this main characteristic (like Aussie Rules football), you’d have a good idea what you should be doing. Once you know the general technique for optimizing in a vector space, you can solve an optimization problem in any vector space.

The second reason is to be able to see analogies. If you see a clever, new strategy in rugby, you try to convert it into a new strategy for football. This will work if the strategy relies on the general structure and objective of the game, rather than some details that are specific to rugby. Similarly, if you see an exciting, new theorem proven for a problem in one Banach space, you might try to see if this can be applied to another problem in another Banach space—even though the spaces may look very different, the idea may be imported if it relies mostly on the general structure instead of the specifics.

Finally, you'll want to know about these abstract spaces because your teachers will talk about them. They'll state a result that holds “for all Hilbert spaces” rather than
the one that they’re talking about now, which may be \( \mathbb{R}^n \). The teacher’s hope is that the nerdy people will then see applications to spaces other than \( \mathbb{R}^n \). Rest assured that most grad students survive the program without knowing the difference between a Hausdorff space and a Banach space, but you can avoid panic by at least being aware that they exist.

Euclidean space. This is the specific space that you’re most familiar with. It fits into all the categories below. Distance between two points is a “metric”; the length of a vector is a “norm”; the dot product of vectors is an “inner product.” Using \( \varepsilon \)-balls, we said today what constituted an open set in Euclidean space—this defines its “topology.”

If you delve into econometrics, you'll start dealing with spaces of random variables. Two random variables, like “the color of the clouds in the sky today” and “the amount of rainfall today” (or “how much you get paid” and “how much education you have”) might seem to be closely related to each other. How do we talk about how “close” random variables are to each other? One random variable (like “how much a soda costs today”) might seem very small in the sense that it’s not very random at all (the machine in the hall has been selling them for sixty cents for as long as I’ve been here) but another (like “how much a share of IBM stock costs today”) seems very big. How do we define how “big” a random variable is?

A third space is the set of functions which are continuous on some interval. The functions \( \sin(x) \) and \( 0.9 \sin(x) \) have graphs which are very similar. It makes sense to say that this functions are “closer” to being the same than, say, \( \sin(x) \) and \(-\sin(x)\), but how can you define distance between functions? The function \( 4 \sin(x) \) seems a lot “bigger” than \( \frac{1}{2} \sin(x) \), but how do you define the size of a function?

Other spaces will show up in game theory, econometrics, even in macro. For reference, here’s a list of definitions of the general abstract spaces.

A **metric space** is a pair \((X, \rho)\) consisting of a set \(X\) and a **distance** or **metric** \(\rho\); that is, a \(\mathbb{R}^+\)-valued function \(\rho(x, y)\) that has three properties:

1. **identity**: \(\rho(x, y) = 0 \iff x = y\)
2. **symmetry**: \(\rho(x, y) = \rho(y, x) \quad \forall x, y \in X\)
3. **triangle inequality**: \(\rho(x, z) \leq \rho(x, y) + \rho(y, z) \quad \forall x, y, z \in X\)
In short, in a metric space we have defined some notion of the distance between elements of the set, whether they are vectors, functions, or whatever. One thing that a metric allows us to do is to talk about convergence—we can say that a sequence of points get really close to each other, or that a sequence of functions get really close to each other.

A vector space or linear space is a set $X$ of elements, called vectors, equipped with two operations: addition, whereby vectors $x, y \in X$ uniquely determine $z = x + y$, with $z \in X$; and scalar multiplication, whereby $x \in X$ and a scalar $\alpha \in \mathbb{R}$ uniquely determine $w = \alpha x$, with $w \in X$. There also exist vectors determined by:

1. the null vector, zero vector, or origin $\theta$ of $X$: $\forall x \in X, \ x + \theta = x$
2. the negative $-x$ of each $x \in X$: $-x \equiv -1x$

Addition satisfies four properties:

1. identity: $x + \theta = x$
2. commutativity: $x + y = y + x$
3. associativity: $(x + y) + z = x + (y + z)$
4. cancellation: $x + (-x) = \theta$

Scalar multiplication satisfies five properties:

1. identity: $1x = x$
2. cancellation: $0x = \theta$
3. associativity: $(\alpha \beta)x = \alpha (\beta x)$
4. distributivity: $(\alpha + \beta)x = \alpha x + \beta x$
5. distributivity: $\alpha (x + y) = \alpha x + \alpha y$

A normed linear vector space is a pair $(X, \|\|)$, consisting of a vector space $X$ and a norm $\|\|$; that is, a $\mathbb{R}_+^1$-valued function that maps each $x$ into some $\|x\|$. The norm satisfies three properties:

1. $\|x\| \geq 0$, and $\|x\| = 0 \iff x = \theta$
2. $\|\alpha x\| = |\alpha| \|x\|$
3. $\|x + y\| \leq \|x\| + \|y\|
Any normed linear vector space can become a metric space by defining the metric as the norm of the difference of vectors: \( \rho(x, y) = \|x - y\| \). A normed linear vector space \( X \) is **complete** if every Cauchy sequence from \( X \) has a limit in \( X \). A complete normed linear vector space is called a **Banach space**.

In normed spaces, we are defining the *size of the elements* of the space. This allows us to answer questions about whether one vector or function is bigger than another, or talk about the magnitude of an effect.

A **pre-Hilbert space** or an **inner product space** is a pair \((X, \langle \cdot, \cdot \rangle)\) consisting of a linear vector space \( X \) with an **inner product** \( \langle \cdot, \cdot \rangle \); that is a \( \mathbb{R}^1 \)-valued function defined on \( X \times X \). The inner product satisfies four properties:

1. \( \langle x, y \rangle = \langle y, x \rangle \) (actually, the complex conjugate of the right-hand side)
2. \( \langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle \)
3. \( \langle \alpha x, y \rangle = \alpha \langle x, y \rangle \)
4. \( \langle x, x \rangle \geq 0 \), and \( \langle x, x \rangle = 0 \iff x = \theta \)

Any pre-Hilbert space can become a normed space by defining the norm as the square root of the inner product: \( \|x\| = \sqrt{\langle x, x \rangle} \). A **Hilbert space** is a complete pre-Hilbert space.

The inner product is a concept of *how similar two vectors are in orientation*—not size. While the vectors \((1,1)\) and \((99,101)\) are very different in size, they go in almost the same direction—they are much closer than, for example, \((1,1)\) and \((1,-1)\), which head in completely unrelated directions, or the vectors \((1,1)\) and \((-1,-1)\), which go in exactly opposite directions.

A **topological space** is a pair \((X, \tau)\) consisting of a set \( X \) and a **topology** \( \tau \); that is, a system of subsets \( G \subseteq X \), called **open sets relative to** \( \tau \), that has two properties:

1. the set \( X \) itself and \( \emptyset \) belong to \( \tau \);
2. arbitrary (finite or infinite) unions \( \bigcup_{\alpha} G_{\alpha} \) and finite intersections \( \bigcap_{i=1}^{n} G_{i} \) of open sets belong to \( \tau \).

In a topological space, we have defined some concept of what constitutes an open set. While the definition of an open set might seem clear-cut in Euclidean space, what is an open set of the integers? What is an open set of functions? Some special types of
topological spaces include a Tychonoff space, a Hausdorff space, a regular space, a normal space, and a completely regular space. For more information about these fun and exciting worlds, consult Kolmogorov and Fomin or Royden.

**Example:** Consider the space of zero-mean random variables (if you’re familiar with random variables; otherwise, disregard this until later in the class) \( \mathcal{X} := \{ (X : \Omega \to \mathbb{R}) : \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\omega) = 0 \} \), where \( (\Omega, \mathbb{P}) \) is some probability space. First, we will show that this space is vector space. (Just assume that for all \( \omega \), \( \mathbb{P}(\omega) > 0 \).) A natural candidate for the “zero vector” is the “degenerate” random variable (that is, not really random at all) that always comes up zero: \( \theta(\omega) = 0 \). It is easy for us to show that if we take two random variables \( X, Y \in \mathcal{X} \), then \( X + \theta = X \) (just adding zero all the time doesn’t change \( X \)), \( X + Y = Y + X \) (random variables take on values of real numbers, and addition of reals is commutative), and that \( X - X = \theta \) (subtracting a real number from itself all the time always gives you zero). The scalar multiplication properties similarly work out.

Next, let’s show that this can also be a normed space—one for which we have some concept of the “size” of different elements. A sensible way to think about one random variable being “bigger” than another, given that they have the same mean, is that it has more variance—or equivalently, a larger standard deviation. Let us define the norm on \( \mathcal{X} \) to be
\[
\|X\| := \sqrt{\sum_{\omega \in \Omega} X(\omega)^2 \mathbb{P}(\omega)} = \sqrt{\text{Var}(X)}.
\]
According to this definition, \( \|\theta\| = 0 \). Good! Since the variance of any random variable is positive or zero, we also have that \( \|X\| \geq 0 \), another requirement for a norm. Additionally, if we scale a random variable up or down by some \( \alpha \), we get that \( \|\alpha X\| = \sqrt{\text{Var}(\alpha X)} = \sqrt{\alpha^2 \text{Var}(X)} = |\alpha| \cdot \|X\| \). Finally, if we take two random variables, \( \|X + Y\| = \sqrt{\text{Var}(X + Y)} = \sqrt{\text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)} \). Since \( \text{Cov}(X,Y) \leq \sqrt{\text{Var}(X)\cdot \text{Var}(Y)} \) (an application of the Cauchy-Schwarz inequality), then
\[
\|X + Y\| \leq \sqrt{\text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)} = \sqrt{\text{Var}(X) + \text{Var}(Y) + 2\text{Var}(X)\text{Var}(Y)}^{1/2}.
\]
Standard deviation satisfies the “triangle inequality,” and so it meets all the requirements of a norm.

Next, let’s show that this is an inner product space by defining \( \langle X, Y \rangle := \sum_{\omega \in \Omega} X(\omega)Y(\omega) \mathbb{P}(\omega) = \text{Cov}(X,Y) \). First, we have that
\[
\langle X, Y \rangle = \sum_{\omega \in \Omega} X(\omega)Y(\omega) \mathbb{P}(\omega) = \sum_{\omega \in \Omega} Y(\omega)X(\omega) \mathbb{P}(\omega) = \langle Y, X \rangle \text{ since multiplication of reals is commutative.}
\]
Second, \( \langle X + Y, Z \rangle = \sum_{\omega \in \Omega} (X(\omega) + Y(\omega))Z(\omega) \mathbb{P}(\omega) = \sum_{\omega \in \Omega} X(\omega)Z(\omega) \mathbb{P}(\omega) + \sum_{\omega \in \Omega} Y(\omega)Z(\omega) \mathbb{P}(\omega) = \langle X, Z \rangle + \langle Y, Z \rangle \). Next, \( \langle \alpha X, Y \rangle = \sum_{\omega \in \Omega} \alpha X(\omega)Y(\omega) \mathbb{P}(\omega) = \alpha \sum_{\omega \in \Omega} X(\omega)Y(\omega) \mathbb{P}(\omega) = \alpha \langle X, Y \rangle \). Finally, \( \langle X, X \rangle = \text{Var}(X) \geq 0 \); and \( \langle X, X \rangle = \text{Var}(X) = 0 \) means that \( X \) is that degenerate random variable that we called \( \theta \). Covariance satisfies the definition of an inner product for random variables.

Fall 2007 math class notes, page 64
So I promised you that one of the neat things about thinking abstractly is that you start to see analogies between things that seem very different. One of these is that the cosine of the angle between two vectors (in $\mathbb{R}^n$) is the same concept as the correlation between two random variables. If you remember back to trigonometry class, you'll recall that if $x, y \in \mathbb{R}^n$ have an angle of $\alpha$ between then, then the cosine of $\alpha$ equals their dot product divided by their lengths. To use the more general notation for norms and inner products, $\cos \alpha = \langle x, y \rangle / \left( \| x \| \cdot \| y \| \right)$. In the very different world of random variables, the correlation between $X, Y \in \mathcal{X}$ is defined as their covariance divided by their standard deviations. Again, in the general notation, $\text{Corr}(X,Y) = \langle X, Y \rangle / \left( \| X \| \cdot \| Y \| \right)$.

Whether this analogy serves any useful purpose, well, that’s a question that I can’t answer. Maybe you can find an application for it.

**Example:** Consider the space of continuous, bounded functions $f : \mathbb{R} \to \mathbb{R}$. The norm on this space is often defined as the largest value $f$ takes:

$$\| f \| = \max_{x \in \mathbb{R}} | f(x) |$$

The metric is usually defined as the greatest distance between functions at any point:

$$\rho(f, g) = \| f - g \| = \max_{x \in \mathbb{R}} | f(x) - g(x) |$$

Consider the sequence of functions

$$\left\{ (1 - \gamma) \sin(x) \right\}$$

What’s your intuition on the limit of this sequence? Can you show that it converges, and if so, to what? What about the sequence:

$$\left\{ \sin((1 - \gamma) x) \right\}$$