



SHARE

LATEST

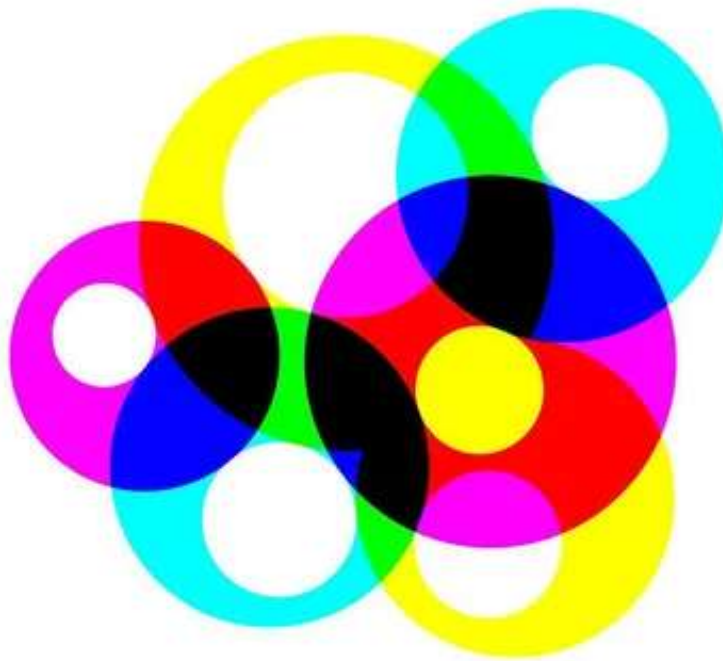


Roots of Unity

What Does Compactness Really Mean?

It took me a long time to understand the mysterious mathematical property of compactness

By Evelyn Lamb on May 25, 2017



Credit: Nils Peters Flickr (CC BY 2.0)

I don't think I've spent more time with a mathematical definition than I did with compactness. It is an important mathematical property and one that initially left me entirely bewildered.

There are two definitions of compactness. One is the real definition, and one is a "definition" that is equivalent in some popular settings, namely the number line, the plane, and other Euclidean spaces. (The fact that the two definitions are equivalent is called the Heine-Borel theorem.)

The real definition of compactness is that a space is compact if every open cover of the space has a finite subcover. I don't know how many times I repeated that definition to myself in my undergraduate topology class, wondering if my incantations would eventually help me understand what in the world compactness was.

Almost simultaneously, I learned the practical definition of compactness in Euclidean spaces: a set is compact if it is closed and bounded. A set is *closed* if it contains all points that are extremal in some sense; for example, a filled-in circle including the outer boundary is closed, while a filled-in circle that doesn't include the outer boundary is not closed. *Bounded* is a little more like what it sounds like: points in a bounded space are all within some fixed distance of each other.

It took me a long time to connect these two ways of looking at compactness, and I'm not going to do that in this post. (If you're taking an introduction to analysis or topology class, you might have the delightful opportunity to learn the Heine-Borel theorem for yourself. Hooray!) But I will unpack the first definition a bit. An *open cover* is a collection of open sets (read more about those here) that covers a space. An example would be the set of all open intervals, which covers the real number line.



A collection of many open intervals on the real number line. Credit: Evelyn Lamb

Of course, the collection of all open intervals in the number line contains a heck of a lot of intervals! Compactness asks if there is a way to whittle down that collection to a finite

number of intervals and still cover the entire number line. That is, could we find a finite number of open intervals so that every point on the number line is in at least one of them? We could eliminate a lot of the intervals and still cover the line — we could, for example, only permit unit-length intervals whose endpoints were at integers or integers-and-a-half — but we could never pare our collection down to a finite number of intervals and still span the entire number line. If we reduced it to 100 unit intervals, for example, we could only cover a maximum of 100 units of length on the infinite number line, and that's if none of the intervals overlapped! So the number line is not compact because we have found an open cover that does not have a finite subcover.

A set does not have to be infinite in length or area to be non-compact. A closed interval and an open interval make a good case study for how we can think about compactness. For convenience, we might as well look at the intervals $(0,1)$ and $[0,1]$. (The first is all the real numbers between 0 and 1 not including the endpoints, the second is all the real numbers between 0 and 1 including 0 and 1.) The open interval $(0,1)$ is not compact because we can build a covering of the interval that doesn't have a finite subcover. We can do that by looking at all intervals of the form $(1/n,1)$. Each one of those intervals lies within $(0,1)$, and put together, any number in the interval $(0,1)$ is in at least one interval of the form $(1/n,1)$. For example, the point .0001 is in the interval $(1/10001,1)$, even though it's not in the intervals $(1/2,1)$, $(1/3,1)$, and so on up to $(1/10000,1)$. But if we want to cover the entire interval $(0,1)$ with only a finite subcollection, we will fail. Any finite subcollection will have a largest interval in it, whether it's $(1/10,1)$ or $(1/10000,1)$ or $(1/\text{Graham's number},1)$. In any case, we can find numbers between 0 and the left endpoint of the largest interval that won't be covered by our finite subcollection.

When we add the endpoints 0 and 1, the interval becomes compact. Now the weird open cover we had no longer covers the whole interval because the points 0 and 1 aren't any of the intervals. It's harder to show that we couldn't cook up a different pathological open cover, so you'll have to take my word for it for now.

Showing that something is compact can be trickier. Proving noncompactness only requires producing one counterexample, while proving compactness requires showing that every single open cover of a space, no matter how oddly constructed, has a finite subcover. But eventually I came to a rigorous understanding of compactness and how both definitions fit together, and I lived happily ever after.

Now, years after wrestling with it for the first time, I've come to what Terry Tao might describe as a post-rigorous understanding of compactness. Compact means small. It is a peculiar kind of small, but at its heart, compactness is a precise way of being small in the mathematical world. The smallness is peculiar because, as in the example of the open and closed intervals $(0,1)$ and $[0,1]$, a set can be made "smaller" (that is, compact) by adding points to it, and it can be made "larger" (non-compact) by taking points away.

As a notion of smallness, then, compactness is a bit fraught. It's a bit unsettling to say that a set can be "smaller" than a set that lies entirely inside it! But I think smallness is a valuable way to see compactness. A set that is compact may be large in area and complicated, but the fact that it is compact means we can interact with it in a finite way using open sets, the building blocks of topology. (For more on open sets, check out my post Change your open sets, change your life.) That's the point of the finite subcover in the definition of compactness. That finite collection of open sets makes it possible to account for all the points in a set in a finite way. That comes up in, for example, the proof of the Heine-Borel theorem.



Sign up for *Scientific American's* free newsletters.

Sign Up

Before I realized compact meant small, I saw that compact sets were often easier to deal with. Continuous functions defined on compact sets have more controlled behavior than functions on non-compact sets. Compact two-dimensional surfaces have a nice classification theorem. Classifying non-compact surfaces is more difficult and less satisfying. Compact surfaces are more constrained. Non-compact ones can squirm out of your hands like blobs of rice pudding. Compact ones are more like jello: they might wobble a bit, but you can hold on to them if you don't mind getting your hands a little dirty.

The post-rigorous understanding of compactness allows the word "compact" to circle around from something that feels like robot speak to something that aligns very closely with an English meaning of the word. I don't know the history of the mathematical use of the word compact, so I don't know how intentional that is. I like to think of it as a delightful accident of mathematical-linguistic convergence.
