COMPACTNESS AND COMPACTIFICATION

TERENCE TAO

In mathematics, it is well known that the behaviour of finite sets and the behaviour of infinite sets can be rather different. For instance, each of the following statements is easily seen to be true whenever X is a finite set, but is false whenever X is an infinite set:

- (All functions are bounded) If $f : X \to \mathbf{R}$ is a real-valued function on X, then f must be bounded (i.e. there exists a finite number M such that $|f(x)| \leq M$ for all $x \in X$).
- (All functions attain a maximum) If $f : X \to \mathbf{R}$ is a real-valued function on X, then there must exist at least one point $x_0 \in X$ such that $f(x_0) \geq f(x)$ for all $x \in X$.
- (All sequences have constant subsequences) If $x_1, x_2, x_3, \ldots \in X$ is a sequence of points in X, then there must exist a subsequence $x_{n_1}, x_{n_2}, x_{n_3}, \ldots$ which is constant, thus $x_{n_1} = x_{n_2} = \ldots = c$ for some $c \in X$. (This fact is sometimes known as the *infinite pigeonhole principle*.)
- (All covers have finite subcovers) If $V_1, V_2, V_3, \ldots \subset X$ are any collection of sets which cover X (thus $\bigcup_{n=1}^\infty V_n = X$, then there must exist a finite sub-collection $V_{n_1}, V_{n_2}, \ldots, V_{n_k}$ of these sets which still cover X (thus $V_{n_1} \cup \ldots \cup V_{n_k} = X$). (We stated this principle for countably infinite collections of sets, but the same statement is in fact true for arbitrary collections of sets, including uncountably infinite ones.)

The first statement - that all functions on a finite set are bounded - can be viewed as a very simple example of a *local-to-global principle*. Namely, the hypothesis is an assertion of "local" boundedness - it asserts that $|f(x)|$ is bounded for each point $x \in X$ separately, but with a bound that can depend on x. The conclusion is that of "global" boundedness - that $|f(x)|$ is bounded by a *single* bound M for all $x \in X$. This local-to-global principle is only valid as stated when the domain X is finite; it is easily seen to fail when X is infinite.

So far we have only viewed the object X as a set. However, in many areas of mathematics we would like to endow our objects with additional structures, such as a topology, a metric, a group structure, and so forth. Once we do so, it turns out that some objects exhibit properties similar to finite sets (and in particular, they enjoy local-to-global principles), even though they may technically be infinite. In the categories of topological spaces and metric spaces, these "almost finite" objects are known as *compact spaces*. (In the category of groups, the analogous notion of an "almost finite" object is that of a *pro-finite group*; for linear transforms between normed vector spaces, the analogous notion of an "almost finite" (or more precisely, "almost finite-rank") object is that of a *compact operator*; and so forth.)

A good example of a compact set is the closed unit interval $X = [0, 1]$. This is an infinite set, so the previous four assertions are all false as stated for X. But if we modify each of these assertions by inserting topological concepts such as continuity, convergence, and open-ness, then we can restore these assertions for $[0, 1]$:

- (All continuous functions are bounded) If $f : X \to \mathbf{R}$ is a real-valued continuous function on X, then f must be bounded. (This is again a type of local-to-global principle: if a function is stable with respect to local perturbations, then it is stable with respect to global perturbations.)
- (All continuous functions attain a maximum) If $f : X \to \mathbf{R}$ is a real-valued continuous function on X, then there must exist at least one point $x_0 \in X$ such that $f(x_0) \geq f(x)$ for all $x \in X$.
- (All sequences have convergent subsequences) If $x_1, x_2, x_3, \ldots \in X$ is a sequence of points in X, then there must exist a subsequence $x_{n_1}, x_{n_2}, x_{n_3}, \ldots$ which is convergent to some limit $c \in X$. (This statement is known as the *Bolzano-Weierstrass theorem.*)
- (All open covers have finite subcovers) If $V_1, V_2, V_3, \ldots \subset X$ are any collection of open sets which cover X, then there must exist a finite sub-collection $V_{n_1}, V_{n_2}, \ldots, V_{n_k}$ of these sets which still cover X. (Again, this statement is true for arbitrary collections of sets.)

In contrast, all four of these topological statements continue to be false for sets such as the open unit interval $(0, 1)$ or the real line $\mathbf{R}$, as one can easily check by constructing simple counterexamples. In fact, the *Heine-Borel theorem* asserts that when X is a subset of a Euclidean space $\mathbf{R}^n$, then the above statements are all true when X is topologically closed and bounded, and all false otherwise.

The above four assertions are closely related to each other. For instance, if you know that all sequences in X contain convergent subsequences, then you can quickly deduce that all continuous functions have a maximum. This is done by first constructing a *maximising sequence* - a sequence of points x_n in X such that $f(x_n)$ approaches the maximal value (or more precisely, the supremum) of f, and then investigating a convergent subsequence of that sequence. In fact, given some fairly mild assumptions on the space X (e.g. that X is a metric space), one can deduce any of these four statements from any of the others.

To oversimplify a little bit, we say that a topological space X is *compact* if one (and hence all) of the above four assertions holds for this space. (Actually, because these four assertions are not quite equivalent in general, the formal definition of compactness uses only the fourth version: that every open cover has a finite subcover. There are other notions of compactness, such as *sequential compactness*, which is based on the third version. The distinctions between these notions are technical and we shall gloss over them here.)

Compactness is a powerful property of spaces, and is used in many ways in many different areas of mathematics. One is via appeal to local-to-global principles; one establishes local control on some function or other quantity, and then uses compactness to boost the local control to global control. Another is to locate maxima or minima of a function, which is particularly useful in calculus of variations. A third is to partially recover the notion of a limit when dealing with non-convergent sequences, by accepting the need to pass to a subsequence of the original sequence. (Note however that different subsequences may converge to different limits; compactness guarantess the existence of a limit point, but not uniqueness.) Compactness of one object also tends to beget compactness of other objects; for instance, the image of a compact set under a continuous map is still compact, and the product of finitely many or even infinitely many compact sets continues to be compact (this is known as *Tychonoff's theorem*).

Of course, many spaces of interest are not compact. For instance, the real line $\mathbf{R}$ is not compact because it contains sequences such as $1, 2, 3, \ldots$ which are "trying to escape" the real line, and are not leaving behind any convergent subsequences. However, one can often recover compactness by adding a few more points to the space; this process is known as *compactification*. For instance, one can compactify the real line by adding one point at either end of the real line, $+\infty$ and $-\infty$. The resulting object, known as the *extended real line* $[-\infty, +\infty]$, can be given a topology (which basically defines what it means to converge to $+\infty$ or to $-\infty$). The extended real line is compact: any sequence x_n of extended real numbers will have a subsequence that either converges to $+\infty$, converges to $-\infty$, or converges to a finite number. Thus by using this compactification of the real line, we can generalise the notion of a limit, by no longer requiring that the limit has to be a real number. While there are some drawbacks to dealing with extended reals instead of the ordinary reals (for instance, one can always add two real numbers together, but the sum of $+\infty$ and $-\infty$ is undefined), the ability to take limits of what would otherwise be divergent sequences can be very useful, particularly in the theory of infinite series and improper integrals.

It turns out that a single non-compact space can have many different compactifications. For instance, by the device of *stereographic projection*, one can topologically identify the real line with a circle with a single point removed (e.g. by mapping the real number x to the point $(\frac{x}{1+x^2}, \frac{x^2}{1+x^2})$, one maps $\mathbf{R}$ to the circle of radius $1/2$ and centre $(0, 1/2)$ with the "north pole" $(0, 1)$ removed), and then by inserting this point we obtain the *one-point compactification* $\mathbf{R} \cup \{\infty\}$ of the real line. More generally, any reasonable non-compact space (in particular, *locally compact spaces*) has a number of compactifications, ranging from the one-point compactification $X \cup \{\infty\}$ (which is the "minimal" compactification, adding only one point) to the *Stone-Cech compactification* βX, (which is the "maximal" compactification, and adds an enormous number of points). The Stone-Cech compactification $\beta\mathbf{N}$ of the natural numbers $\mathbf{N}$ is the space of *ultrafilters*, which are very useful tools in the more infinitary parts of mathematics.

One can use compactifications to distinguish between different types of divergence in a space. For instance, the extended real line $[-\infty, \infty]$ distinguishes between divergence to $+\infty$ and divergence to $-\infty$. In a similar spirit, by using compactifications of the plane $\mathbf{R}^2$ such as the *projective plane*, one can distinguish a sequence diverging along (or near) the x-axis from a sequence diverging along (or near) the y-axis. Such compactifications thus arise naturally in situations in which sequences that diverge in different ways exhibit markedly different behaviour.

Another use of compactifications is to allow one to rigorously view one type of mathematical object as a limit of others. For instance, one can view a straight line in the plane as the limit of increasingly large circles, by describing a suitable compactification of the space of circles which includes lines; this perspective allows us to deduce certain theorems about lines from analogous theorems about circles, and conversely to deduce certain theorems about very large circles from theorems about lines. In a rather different area of mathematics, the Dirac delta function is not, strictly speaking, a function, but exists in a certain (local) compactification of spaces of functions, such as spaces of measures or distributions. Thus one can view the Dirac delta function as a limit (in a suitably weak topology) of classical functions, which can be very useful for manipulating that function. One can also use compactifications to view the continuous as the limit of the discrete; for instance, it is possible to compactify the sequence $\mathbf{Z}/2\mathbf{Z}$, $\mathbf{Z}/3\mathbf{Z}$, $\mathbf{Z}/4\mathbf{Z}$, etc. of cyclic groups, so that their limit is the circle group $\mathbf{T} = \mathbf{R}/\mathbf{Z}$. These simple examples can be generalised into much more sophisticated examples of compactifications (and to the closely related concept of *completions*), which have many applications in geometry, analysis, and algebra.

Department of Mathematics, UCLA, Los Angeles CA 90095-1555

E-mail address: `tao@math.ucla.edu`