Cantor’s Proof of the Nondenumerability of Perfect Sets

Laila Awadalla

University of Missouri — Kansas City

Follow this and additional works at: https://scholar.rose-hulman.edu/rhumj

Recommended Citation
Available at: https://scholar.rose-hulman.edu/rhumj/vol17/iss1/8
Cantor’s Proof of the Nondenumerability of Perfect Sets

Laila Awadalla

Volume 17, No. 1, Spring 2016
Abstract. This paper provides an explication of mathematician Georg Cantor’s 1883 proof of the nondenumerability of perfect sets of real numbers. A set of real numbers is denumerable if it has the same (infinite) cardinality as the set of natural numbers \( \{1, 2, 3, \ldots\} \), and it is perfect if it consists only of so-called limit points (none of its points are isolated from the rest of the set). Directly from this proof, Cantor deduced that every infinite closed set of real numbers has only two choices for cardinality: the cardinality of the set of natural numbers, or the cardinality of the set of real numbers. This result strengthened his belief in his famous continuum hypothesis that every infinite subset of real numbers had one of those two cardinalities and no other. This paper also traces Cantor’s realization that understanding perfect sets was key to understanding the structure of the continuum (the set of real numbers) back through some of his results from the 1874–1883 period: his 1874 proof that the set of real numbers is nondenumerable, which confirmed Cantor’s intuitive belief in the richness of the continuum compared to discrete subsets (such as the set of natural numbers) and proved that there was more than one “size” of infinite cardinal; his 1878 proof that continuous domains of different dimensions (such as a one-dimensional line and a two-dimensional surface) surprisingly have the same cardinality; and his 1883 definition of a continuum as a set that is connected (all of one piece) and perfect.

Acknowledgements: I would like to thank my history of mathematics teacher Dr. Richard Delaware for his generous support and guidance throughout the writing of this paper. I would also like to thank Dr. Delaware and my French teacher Tamara Carrell for helping me translate a tricky section of reference [3].
1 Introduction

Throughout his career, Georg Cantor (1845-1918), the mathematician often credited with founding set theory, investigated and classified a wide variety of sets of points in an attempt to characterize continuous domains, such as the set of real numbers or intervals of that set. One way in which he studied sets involved studying their limit points. A limit point associated with a given set is a point such that every open interval of that point contains infinitely many points of the given set. To study a set, he collected its limit points into what he called a derived set and took the intersection of the initial set with its derived set. This process, called derivation, led him to the notion of a perfect set [6, p. 210], which he defined as a set that equals its derived set and is thus unaffected by derivation; according to this definition, then, perfect sets consist entirely of limit points, so it is equivalent to the modern definition.

As an example of a set that is neither continuous nor perfect consider the set \( \{0, 1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}, \ldots \} \). This set’s only limit point is 0, so its derived set is \{0\}; this set, then, contains, but does not equal, its derived set. The open interval \((0, 1)\) is not perfect either, because its derived set contains not only all the points in \((0, 1)\), but also the points 0 and 1, which are not elements of \((0,1)\). The derived set of \((0,1)\) is the closed interval \([0,1]\), which is, in fact, a perfect set. Later we will see that perfect sets of real numbers need not be intervals.

Since the derived set of a finite set is empty, and hence its intersection with that derived set is also empty, Cantor deduced that perfect sets must be infinite, and since he had already introduced the concept of power (or cardinality) of a set, he wondered whether perfect sets were denumerable, having the cardinality of the set of positive integers \(\{1,2,3,\ldots\}\), or nondenumerable, having the cardinality of the set of real numbers. In 1883 he sent to his friend Gösta Mittag-Leffler (1846-1927), editor of the journal *Acta Mathematica*, a letter containing his proof of the nondenumerability of perfect sets, noting that this proof had also led him to an important insight about closed sets.

Before examining Cantor’s letter, we will discuss in Section 2 how Cantor’s interest in perfect sets can be traced back to his interest in the set of real numbers, traditionally called the continuum. In Sections 3, 4, and 5, we examine Cantor’s letter. Section 3, the longest section of this paper, consists of an explication of Cantor’s proof of the nondenumerability of perfect sets. In Section 4, we explicate his much shorter proof that dense perfect sets are nondenumerable. Finally, in Section 5 we explicate his discussion of closed sets and summarize his thoughts on the continuum hypothesis (defined at the end of Section 2).

2 Cantor’s interest in continua and perfect sets

Cantor was already interested in the properties of continuous sets in 1872, when he and the German mathematician Richard Dedekind (1831-1816) were both, independently, studying the real number line and irrational numbers; both knew that “the real numbers were far richer in number and in properties than the rationals” [5, p. 50], but neither knew how to explain this more precisely. Dedekind had developed his theory of the real number line and continuity much earlier.
than Cantor, but he published it shortly after Cantor sent him a copy of his paper outlining his theory of real numbers. Dedekind mailed Cantor his own paper too, and thus their “friendly exchange of ideas” began [5, p. 48]. In his paper, Dedekind compares the rational numbers with a straight line, the geometric representation of the set of real numbers, and concludes that “the line $L$ is infinitely richer in point-individuals than is the domain $R$ of rational numbers in number-individuals” [5, p. 48]. In the following years Cantor developed the concept of the power, or cardinality, of a set, which allowed him to confirm this comparative richness. Under Cantor’s definition, two sets have the same power if there exists a one-to-one correspondence between them [5, p. 59]. Cantor’s exploration of cardinality culminated in his first proof (in 1874) of the nondenumerability of the set of real numbers, or the continuum.

In 1873 Cantor sent a letter to Dedekind asking whether he thought it was possible to establish a one-to-one correspondence between the set of integers and the set of real numbers; he believed it was not, because the first was discrete while the second was continuous, but he did not want to rely on his intuition because, after all, there were clearly more rational numbers than integers, yet both had the same power, as they were denumerable [5, p. 49]. Dedekind replied that he did not have an answer. A few months later, in 1874, Cantor published his answer to the question; he found such a correspondence was not possible and concluded that the real numbers were nondenumerable. Cantor’s 1874 paper was significant because in addition to establishing an important difference between discrete and continuous aggregates of numbers, it presented a method of comparing infinite sets in a precise, quantitative way.

Between 1874 and 1877, Cantor made a surprising discovery about continuity: continuous sets of different dimensions have the same power. In 1874, shortly after his first proof that the real numbers were nondenumerable, he sent a letter to Dedekind asking whether he thought it was possible to establish a one-to-one correspondence between a surface and a line, a question many of his colleagues considered absurd because of the assumed independence of the coordinates that determined points in two-dimensional surfaces and one-dimensional lines. Dedekind did not answer Cantor’s question, but three years later, in 1877, Cantor proved that even though it seemed absurd it was possible and sent the proof in a letter to Dedekind [5, p. 54]. In his letter, Cantor marveled at the “wonderful power there is in the ordinary real (rational and irrational) numbers, since one is in a position to determine uniquely, with a single coordinate, the elements of a $p$-dimensional continuous space” [5, p. 55]. This theorem showed Cantor that to understand the nature of continuous sets it would suffice to understand the set of real numbers, the continuum, so Cantor turned his attention to continuous linear sets [5, p. 65]. To study these, he needed to formalize the concept of continuum, since before him “different features had always been stressed, but no exact or complete definition had ever been given” [5, p. 107].

In 1883 Cantor published his first definition of continua [5, p. 110]. In his paper, leading up to his definition, he states what is now called the Cantor-Bendixson Theorem: a nondenumerable derived set $P'$ (the set of limit points of a set $P$) can be written in a unique way as a disjoint union of a denumerable set and a perfect set (recall perfect sets are those unaffected by derivation: $P = P'$). This theorem allowed him to conclude that continua must be perfect [5, p. 109]. Cantor notes that since a perfect set is not necessarily dense in any interval (meaning it does not necessarily appear in every open subinterval of any interval), a perfect set is not necessarily a continuum [4, p. 406], and as an example of a perfect set that is nowhere dense he
offers the set consisting of numbers of the form \( \frac{c_1}{3} + \frac{c_2}{3^2} + \frac{c_3}{3^3} + \ldots + \frac{c_v}{3^v} + \ldots \), where the \( c_v \) are 0 or 2 [4 p. 407]; we now know this set as the Cantor ternary set. The existence of perfect sets that are not continua shows the definition of continua needs another condition, connectedness, which Cantor defines as follows: a set \( T \) is connected “if for any two points \( t \) and \( t' \) of \( T \), there always exists a finite number of points \( t_1, t_2, \ldots, t_v \) of \( T \) such that the distances \( \ell t_1, \ell t_2, \ldots, \ell t_v t' \) are all less than any given arbitrarily small number \( \epsilon \)” [5, p. 109]; the tiny line segments \( tt_1, t_1t_2, \ldots, t_v t' \) with these points as endpoints form a path between \( t \) and \( t' \) that does not need to be contained in the set. A set that is connected under Cantor’s definition but is not a continuum is the set of rational numbers [6, p. 209]; note that it is connected because it is dense. Correspondingly, the fact that the Cantor ternary set is not connected follows from its being nowhere dense: if we take two distinct points \( t \) and \( t' \) of the Cantor set, we are guaranteed to find an open interval between them that contains no points of our set, and the (positive) length of this interval is a lower bound for one of the distances in the definition of connectedness; since this distance cannot be made arbitrarily small, the set is not connected.

In his study of the continuum, Cantor encountered many different types of infinite subsets of the real numbers, and for each type of set he was able to establish a one-to-one correspondence between either the set and the natural numbers or the set and the real numbers. This, along with the fact that the Cantor-Bendixson theorem showed him that the process of derivation “yields only sets that are denumerable or perfect, or composed of both” [6, p. 210], led Cantor to conjecture that infinite subsets of the real numbers could only have the power of the natural numbers or of the real numbers. This conjecture, which he called the continuum hypothesis (CH), would guide his research for the following years.

3 Proof of the nondenumerability of perfect sets

In 1883 Cantor realized that perfect sets were instrumental in the study of the continuum: he proved that the CH held for closed sets by proving the nondenumerability of perfect sets. This was, recall, in the letter that Cantor sent in 1883 to Mittag-Leffler, who published excerpts of it in 1884 in the journal Acta Mathematica.

A few of the notational conventions in Cantor’s proof of the nondenumerability of perfect sets need explanation: Cantor writes the open interval \( (a, b) \) as \( (a \ldots b) \); if \( A \) is a subset of \( B \), he says that \( A \) is a divisor of \( B \), \( B \) is a multiple of \( A \), or \( A \) is a partial system of \( B \); if two sets \( A, B \) have the same power (cardinality) he writes \( A \sim B \); and if a set \( A \) is the disjoint union of other sets \( B, C, \) and \( D \), he writes \( A \equiv B + C + D \) [5, p. 88]. Also note that in the Acta Mathematica publication, the excerpts are not titled or numbered; I numbered them here for convenience. All comments in square brackets and a smaller font are mine.

In Excerpt 1 Cantor proves that an arbitrary perfect nowhere dense set \( S \) has the same power as the closed interval \([0, 1]\) by decomposing each into a disjoint union and establishing a one-to-one correspondence between all their parts. Recall that in 1874 Cantor had shown that the real number continuum was nondenumerable, and in 1878 Cantor showed intervals of real numbers
were nondenumerable too [5, p. 79], so he knew that to show the nondenumerability of a set it sufficed to establish a one-to-one correspondence between it and an interval, in this case [0, 1].

**Excerpt 1**

My theorem says that perfect sets of points all have the same power, the power of the continuum.

As a first step in the proof, I will show it for perfect linear sets [meaning perfect subsets of the real number line]. Let $S$ be an arbitrary perfect set of points which is not condensed in the span of any interval [meaning $S$ is nowhere dense], no matter how small. We assume that $S$ is contained in the interval $(0 \ldots 1)$ whose endpoints 0 and 1 belong to $S$ [so $S$ is actually contained in the closed interval $[0, 1]$] and it is clear that all other cases in which the perfect set is not condensed in the span of any interval can be projected into this one [with a first-degree polynomial function mapping endpoints to endpoints].

However, by considerations in Acta Mathematica 2, p. 378, [1883, Cantor, reference [3]] there exists an infinite number of distinct intervals, each totally separated [disjoint] from the others, that can be represented by ordering them according to their size in such a way that smaller intervals come after larger ones. [In that article, and in this one, the length of an interval is the difference of the endpoints: for $a < b$, the length of $(a, b)$ is $b - a$.] We denote this ordering by

$$(a_1 \ldots b_1), (a_2 \ldots b_2), \ldots, (a_v \ldots b_v), \ldots$$

(1)

With respect to the set $S$, these intervals have the property that no point of $S$ lies in any of their interiors while their endpoints $a_v$ and $b_v$ in conjunction with the other limit points of the set of endpoints $\{a_v, b_v\}$ \[= \{a_1, a_2, a_3, \ldots\} \cup \{b_1, b_2, b_3, \ldots\}\] belong to $S$ and determine it [because $S$ is perfect, so it consists only of limit points]. We denote by $g$ a limit point of $\{a_v, b_v\}$ and by $\{g\}$ the set of these. We then have

$$S \equiv \{a_v\} + \{b_v\} + \{g\}$$

(2)

[S is the disjoint union of $\{a_v\}, \{b_v\},$ and $\{g\}$]

Moreover, the sequence (1) of disjoint intervals is such that the space between any two $(a_v \ldots b_v)$ and $(a_\mu \ldots b_\mu)$ always contains an infinite number of others [other such intervals] and that for any interval $(a_q \ldots b_q)$ there are others in the same sequence (1) that approach arbitrarily closely either the point $a_q$ or the point $b_q$ [because the sequence is ordered so that the length, or the diameter, of the intervals shrinks as $v$ gets large], since $a_v$ and $b_v$ are points of the perfect set $S$ and so are limit points [of $S$, because $S$ consists entirely of limit points].

This having been established, I will take an arbitrary set of the first power [a denumerable set]

$$\varphi_1, \varphi_2, \ldots, \varphi_v, \ldots$$

(3)
a set of distinct points all lying on the interval (0...1), in the whole span of which they are condensed [dense] and I will suppose that the endpoints 0 and 1 are terms of the sequence \( \varphi_v \).

In order to give an appropriate example I recall the type of sequence that I used to enumerate all rational numbers \( \geq 0 \) and \( \leq 1 \) in Acta Mathematica 2, page 319, where, for our purposes, one must remove the first two terms which are 0 and 1.

In that article, which is reference [2], Cantor takes a rational number \( \frac{p}{q} \) in [0, 1], in lowest terms, (meaning that the greatest common factor of \( p \) and \( q \) is 1) and setting \( p + q = N \) notes that to each \( \frac{p}{q} \) corresponds one value of \( N \) and, since both \( p \) and \( q \) are positive, to each value of \( N \) correspond finitely many \( \frac{p}{q} \). He forms a sequence by ordering the \( \frac{p}{q} \) as follows:

1. The \( \frac{p}{q} \) corresponding to smaller values of \( N \) precede the ones corresponding to larger values of \( N \).
2. Within each set of \( \frac{p}{q} \) corresponding to the same \( N \), the numbers \( \frac{p}{q} \) can be ordered according to their magnitude.

With these rules, the first ten terms of this sequence are

\[
0, \frac{1}{1} (\text{so } N = 1 + 1 = 2); \frac{1}{2} (N = 3); \frac{1}{3} (N = 4); \frac{1}{4}; \frac{2}{3} (N = 5); \frac{1}{5} (N = 6); \frac{1}{6}; \frac{2}{5}, \text{ and } \frac{3}{4} (N = 7).
\]

To consider rational numbers strictly between 0 and 1, “one must remove the first two terms,” so the sequence would start at \( \frac{1}{2} \).

**Excerpt 1 continued**

But I insist that sequence (3) \([\varphi_1, \varphi_2, \ldots, \varphi_v, \ldots]\) be left in full generality.

This is what I am claiming: the set of points \( \{\varphi_v\} \) and the set of intervals \( \{(a_v \ldots b_v)\} \) can be associated to each other uniquely in such a way that for any pair of intervals \( (a_v \ldots b_v), (a_\mu \ldots b_\mu) \) belonging to sequence (1) and \( \varphi_{k_v}, \varphi_{k_\mu} \) the corresponding points of sequence (3) one always has that the number \( \varphi_{k_v} \) is smaller or larger than \( \varphi_{k_\mu} \) depending on whether in the segment (0...1) the interval \( (a_v \ldots b_v) \) comes before or after the interval \( (a_\mu \ldots b_\mu) \). [As Edgar notes, “this is not the order of sequence (1),” which, recall, is ordered by the length of the intervals, not by their position in (0...1).]

Such a correspondence between the two sets \( \{\varphi_v\} \) and \( \{(a_v \ldots b_v)\} \) can be realized, for example, by the following rule:

We associate to the interval \( (a_1 \ldots b_1) \) the point \( \varphi_1 \), to the interval \( (a_2 \ldots b_2) \) the term \( \varphi_{k_2} \) of sequence (3) having the smallest index and the same order relation to \( \varphi_1 \) that the interval \( (a_2 \ldots b_2) \) has to \( (a_1 \ldots b_1) \) with respect to their location in the segment (0...1). Moreover, we associate to the interval \( (a_3 \ldots b_3) \) the
term with smallest index that has the same order relation with respect to \( \varphi_1 \) and \( \varphi_{k_2} \) that the interval \((a_3 \ldots b_3)\) has to the intervals \((a_1 \ldots b_1)\) and \((a_2 \ldots b_2)\) with respect to their locations in the interval \((0 \ldots 1)\).

In general, we associate to the interval \((a_v \ldots b_v)\) the term \( \varphi_{k_v} \) of sequence \((3)\) having the smallest index and the same order relation with respect to the points \( \varphi_1, \varphi_{k_2}, \ldots, \varphi_{k_{v-1}} \) already constructed that the interval \((a_v \ldots b_v)\) has to the corresponding intervals \((a_1 \ldots b_1), (a_2 \ldots b_2), \ldots, (a_{v-1} \ldots b_{v-1})\) with respect to their locations in the interval \((0 \ldots 1)\).

Note that the sequence \( \varphi_1, \varphi_{k_2}, \ldots, \varphi_{k_v}, \ldots \) that he is constructing is not a subsequence of \( \{\varphi_v\} \), sequence \((3)\), because when he looks for “the term \( \varphi_{k_v} \) of sequence \((3)\) having the smallest index,” he starts at the beginning of the sequence, so \( k_v \) is not necessarily larger than \( k_{v-1} \), meaning \( \varphi_{k_v} \) is not necessarily later in the sequence than \( \varphi_{k_{v-1}} \). Note also that the order relation he is using for the sequence \( \{\varphi_v\} \) is the usual order relation on the real numbers, so if, for example, we seek \( \varphi_{k_3} \) and know that \((a_3 \ldots b_3)\) comes before \((a_1 \ldots b_1)\) and after \((a_2 \ldots b_2)\), we will choose the \( \varphi_{k_3} \) with the smallest index so that \( \varphi_1 > \varphi_{k_3} > \varphi_{k_2} \) (the inequalities are strict because the intervals are distinct and disjoint).

**Excerpt 1 continued**

I claim that, by this rule, the points \( \varphi_1, \varphi_2, \ldots, \varphi_v, \ldots \) of sequence \((3)\) will be successively, even with a different ordering than the one given by sequence \((3)\), all associated to distinct intervals of sequence \((1)\); since for each order relation between a finite number of points in sequence \((3)\) there are, for the same number of intervals, relations with respect to location in the interval \((0 \ldots 1)\) that conform to this ordering. This is due to the fact that the set \( S \) is a perfect set that is not condensed in any interval, no matter how small. [Since \( S \) is “not condensed in any interval,” meaning nowhere dense, there are infinitely many intervals (all disjoint) to the left and to the right of each interval in the sequence.]

To simplify matters, we put

\[
\varphi_1 = \psi_1, \varphi_2 = \psi_2, \ldots, \varphi_{k_v} = \psi_v, \ldots
\]

As a consequence the following sequence

\[
\psi_1, \psi_2, \ldots, \psi_v, \ldots
\]

(4) consists of exactly the same numbers as sequence \((3)\) \([\{\varphi_v\}\)]. The two sequences \((3)\) and \((4)\) differ only by the ordering of their terms.

The sequence \( \psi_v \) of points in \((4)\) thus has the remarkable relation with the sequence of intervals \((1)\): whenever \( \psi_v \) is larger or smaller than \( \psi_\mu \) one has that \( a_v \) and \( b_v \) are respectively larger or smaller than \( a_\mu \) and \( b_\mu \) [or, equivalently because the intervals are disjoint, one has that \((a_v \ldots b_v)\) is earlier or later than \((a_\mu \ldots b_\mu)\) in the interval \((0 \ldots 1)\)]. Once more I recall that, since the set \( \{\psi_v\} \) corresponds to the given set
\( \{ \varphi_v \} \) except for order, it is condensed \([\text{dense}]\) in the entire segment \((0 \ldots 1)\) and the endpoints 0 and 1 do not belong to this set.

The consequences of such a correspondence between the two sets \( \{ \psi_v \} \) and \( \{(a_v \ldots b_v)\}\) are, as it is easy to show, the following:

If \((a_{\lambda_1} \ldots b_{\lambda_1}), (a_{\lambda_2} \ldots b_{\lambda_2}), \ldots, (a_{\lambda_v} \ldots b_{\lambda_v}), \ldots\) is an arbitrary sequence of intervals belonging to the series \([\text{sequence}]\) (1) that converges either to the point \(a_q\) or the point \(b_q\) then the corresponding series \([\text{sequence}]\) of points \(\psi_{\lambda_1}, \psi_{\lambda_2}, \ldots, \psi_{\lambda_v}, \ldots\) belonging to sequence (4) converges to the point \(\psi_q\) and vice-versa.

Cantor had already mentioned the first part of this first consequence: on the first part of excerpt 1, when describing the properties of the sequence of intervals \((a_v \ldots b_v)\), Cantor says, “for any interval \((a_q \ldots b_q)\) there are others in the same sequence (1) that approach arbitrarily closely either the point \(a_q\) or the point \(b_q\).” We will now show how this follows from the ordering of the sequence (“smaller intervals come after larger ones”) and from the fact that \(S\) is nowhere dense (which ensures that the intervals are disjoint and the space between any two intervals “always contains an infinite number of others”).

Consider an interval \((c, d)\) from the sequence \(\{(a_v \ldots b_v)\}\). Traversing the sequence, we will find an interval \((c_1, d_1)\) that is to the left of \((c, d)\). Since there are infinitely many intervals between these two intervals, we can find in the sequence an interval \((c_2, d_2)\) that lies between \((c, d)\) and \((c_1, d_1)\); then we can find another interval \((c_3, d_3)\) in the sequence \(\{(a_v \ldots b_v)\}\) that lies between \((c, d)\) and \((c_2, d_2)\). In this way we create a sequence of intervals from \(\{(a_v \ldots b_v)\}\), not necessarily a subsequence, with the property that, for all \(k\), \((c_k, d_k)\) lies between \((c, d)\), and \((c_{k-1}, d_{k-1})\).

We can say that this sequence of intervals converges to \(c\): for all \(k\), \((c_k, d_k)\) is to the right of \((c_{k-1}, d_{k-1})\) and to the left of \((c, d)\), so as \(k\) increases the intervals get closer to \((c, d)\); we can say, then, that our sequence \(\{(c_v, d_v)\}\) converges to \((c, d)\), but since all the sequence intervals are to the left of \((c, d)\), all of their points are closer to \(c\) than they are to \(d\), so we can also say that the sequence converges to \(c\). We can construct a sequence of intervals from \(\{(a_v \ldots b_v)\}\) that converges to \(d\) instead in a similar way, finding intervals in the sequence that are always to the right, instead of to the left, of \((c, d)\).

The second part of the first consequence, that convergence of the sequence of intervals is equivalent to convergence of the sequence of numbers, follows directly from the correspondence between the sequences \(\{(a_v \ldots b_v)\}\) and \(\{\psi_v\}\). The fact that the intervals \((c_k, d_k)\), from the sequence \(\{(a_v \ldots b_v)\}\), are getting closer to each other implies that the numbers \(\psi_k\) are also getting closer to each other. Furthermore, since \((c, d)\) is an element of \(\{(a_v \ldots b_v)\}\), there is a \(\psi\) in \(\{\psi_v\}\) that corresponds to \((c, d)\), and the subsequence of numbers \(\{\psi_v\}\) corresponding to our subsequence of intervals \(\{(c_k, d_k)\}\) must converge to it.

We return to Cantor’s letter, which now states the second consequence of the correspondence of sequences.
Excerpt 1 continued

[Second consequence:] If \((a_{\lambda_1} \ldots b_{\lambda_1}), \ldots, (a_{\lambda_2} \ldots b_{\lambda_2}), \ldots, (a_{\lambda_v} \ldots b_{\lambda_v}), \ldots\) is an arbitrary sequence of the same type [from the sequence \(\{(a_v \ldots b_v)\}\)] but its terms converge to a point \(g\) of the set \(S\) (see formula (2) and the meaning of \(g\) [recall \(g\) is a limit point of the set of endpoints of the intervals and \(S\) is the disjoint union of those endpoints and their limit points]) then the corresponding sequence \(\psi_{\lambda_1}, \psi_{\lambda_2}, \ldots, \psi_{\lambda_v}, \ldots\) in turn converges to a well defined point of the interval \((0 \ldots 1)\) not occurring in sequence (3) or (4) \([\{\varphi_v\} \text{ or } \{\psi_v\}\)] and completely determined by \(g\). We denote this point corresponding to \(g\) by \(h\). Conversely, an arbitrary point on the interval \((0 \ldots 1)\) not belonging to sequences (3) or (4) determines a point \(g\) in the set \(S\) unequal to the points \(a_v\) and \(b_v\). The variables \(g\) and \(h\) are single valued functions of each other [each \(g\) corresponds to exactly one \(h\) so the sets \(\{g\}\) and \(\{h\}\) certainly have the same power [by definition].

Our theorem now follows.

For we have, by formula (2),

\[S \equiv \{a_v\} + \{b_v\} + \{g\}\]

and it is clear that

\[(0 \ldots 1) \equiv \{\varphi_{2v}\} + \{\varphi_{2v-1}\} + \{h\}.

[Recall that the sequence \(\{\varphi_v\}\) is dense in \((0, 1)\), so every point of \((0, 1)\) that is not in \(\{\varphi_v\}\) is a limit point of a sequence of elements of \(\{\varphi_v\}\), which means that it is an element of \(\{h\}\); this is analogous to Cantor’s definition of irrational numbers as the non-rational limit points of sequences of rational numbers.]

But, since we have the following formulas,

\[\{a_v\} \sim \{\varphi_{2v}\}, \{b_v\} \sim \{\varphi_{2v-1}\}, \text{ and } \{g\} \sim \{h\},\]

[The first two follow from the fact that each \(\varphi_v\) corresponds to exactly one \((a_v \ldots b_v)\), which is determined by its endpoints, so \(\{a_v\}, \{b_v\} \sim \{\varphi_v\}\), meaning both are denumerable. The sets \(\{a_v\}\) and \(\{b_v\}\) are denumerable subsets of \(\{a_v, b_v\}\), and the sets \(\{\varphi_{2v}\}\) and \(\{\varphi_{2v-1}\}\) are denumerable subsets of \(\{\varphi_v\}\), so they all have the same power.]

one concludes, by Theorem (E) [written by Cantor, stating that if in two sequences, both finite or both infinite, both made of constants or both made of variables, to each element of one corresponds exactly one element of the other, then the two sequences have the same power; note this theorem applies only because the formulas show that \(S\) and \((0 \ldots 1)\) are disjoint unions of the elements used to establish the correspondence] of Acta Mathematica 2, page 318, [reference [2]] the formula:

\[S \sim (0 \ldots 1),\]

i. e., the perfect set \(S\) has the same power as the continuous segment \((0 \ldots 1)\), which is what was claimed.

It is amazing how effortlessly the conclusion follows from the careful establishment of the decompositions and the correspondences. We can now summarize Cantor’s approach:
First, the decomposition of \( S \). By assumption, \( S \) is nowhere dense, so it determines infinitely many intervals; these intervals, since they are disjoint, are denumerable, so they “can be represented by ordering them according to their size in such a way that smaller intervals come after larger ones.” Thus Cantor creates the sequence \( \{(a_v \ldots b_v)\} \). Each interval is determined by its endpoints, and since \( S \) is perfect, it consists exactly of the endpoints and their limit points: \( S \) is a disjoint union of left endpoints, right endpoints, and limit points. (This is formula 2: \( S \equiv \{a_v\} + \{b_v\} + \{g\} \).)

Second, the decomposition of \((0, 1)\) actually follows from the correspondences. Cantor takes an arbitrary sequence of points \( \varphi_v \) from \((0, 1)\) and reorders the sequence so its elements follow the same relation as the intervals in \( \{(a_v \ldots b_v)\} \); he calls this reordered sequence \( \psi_v \). This reordering is only necessary to explain the correspondence of the sequence of intervals and the sequence of numbers, so he uses \( \varphi_v \) in the rest of his proof. Then he finds \( \{h\} \), the set of points in \([0, 1]\) corresponding to the set of limit points \( \{g\} \) of the endpoints of the intervals, and he notes that this is exactly the relative complement of \( \varphi_v \) with respect to \((0, 1)\), so \((0, 1)\) decomposes into \( \varphi_v \) and \( \{h\} \). He decomposes it further using the subsequences of \( \varphi_v \) determined by whether of their indices are even or odd and gets the formula \((0 \ldots 1) \equiv \varphi_{2v} + \varphi_{2v-1} + \{h\} \).

Cantor then establishes the correspondence of each subset in the decompositions of \( S \) and \((0, 1)\). He notes that the subsequences \( \{a_v\} \) and \( \{b_v\} \) of \( \{a_v, b_v\} \) correspond to the subsequence \( \varphi_{2v} \) and \( \{\varphi_{2v-1}\} \) of \( \varphi_v \), so they are all denumerable, and \( \{g\} \) corresponds to \( \{h\} \) by construction. Since the unions are disjoint, correspondence of the subsets translates into correspondence of their unions, and Cantor concludes that \( S \) is in one-to-one correspondence to \((0, 1)\), so \( S \sim (0 \ldots 1) \).

4 Proof of the nondenumerability of dense perfect sets

The proof Cantor presents in Excerpt 3 that dense perfect sets are nondenumerable is much shorter.

Excerpt 3

I showed above that all perfect linear sets of points that are not condensed in any part of the segment on which they lie, no matter how small, have the same power as the linear continuum. Let us now take an arbitrary [not necessarily nowhere dense] perfect linear set of points \( S \) lying on the interval \((-\omega \ldots +\omega)\). I assert that \( S \) has the power of the continuum \((0 \ldots 1)\).

In fact, as we have already dealt with the case in which the set \( S \) is not condensed in any part of the segment \((-\omega \ldots +\omega)\), let us take an arbitrary interval \((c \ldots d)\) in whose interior \( S \) is everywhere condensed [no generality is lost because a set that is not nowhere dense must be dense in some interval]. All the points of \((c \ldots d)\) will also belong to \( S \) since \( S \) is a perfect set [because since \( S \) is dense in \((c \ldots d)\), every point of \((c \ldots d)\) is a limit point of \( S \), and is hence, because \( S \) is perfect, contained in \( S \).]
The set of points \((c \ldots d)\) is a partial system \([\text{subset}]\) of \(S\) and \(S\) is a partial system of the segment \((-\omega \ldots +\omega)\) [so we have \((c \ldots d) \subseteq S \subseteq (-\omega \ldots +\omega)\)]. Since the set \((c \ldots d)\) has the same power as the set \((-\omega \ldots +\omega)\) [Cantor showed in 1874 that all intervals are equivalently nondenumerable [5, p. 79]] one concludes that \(S\) has the same power as \((-\omega \ldots +\omega)\), i.e., the power of \((0 \ldots 1)\); for one has the general theorem [now called the Cantor-Bernstein Theorem]:

“Given a well defined set \(M\) of arbitrary power, a partial set \(M'\) contained in \(M\) and a partial set \(M''\) contained in \(M'\), if the last system \(M''\) has the same power as the first \(M\), then the intermediate set \(M'\) also has the same power as \(M\) and \(M''\).” (see Acta Mathematica 2, page 392 [reference [4]]).

Here Cantor ends his proof that a perfect set \(S\) that is dense in some interval \((c, d)\) has the same power as \((0, 1)\), and is hence nondenumerable (though this is not the end of the excerpt).

5 Closed sets and the continuum hypothesis

Next, Cantor turns his attention to closed sets. Note that the definition that Cantor gives for a closed set \(P\) in the first sentence below is equivalent to the modern definition, because “\(P\) contains all its limit points” is equivalent to “\(P^{(1)} \subseteq P\)”, where \(P^{(1)}\) (or \(P'\)) is the set of all limit points of \(P\), which he called the first derived set of \(P\). Recall that every perfect set is closed by definition, but a closed set \(P\) need not be perfect because it can have isolated points, or points \(x\) such that the intersection of some open interval containing \(x\) with \(P\) is \(\{x\}\). Cantor uses what he proved about perfect sets to explain why infinite closed sets must be either denumerable (of “the first power,” meaning the power of the integers) or nondenumerable (of “the power of the arithmetic continuum,” meaning the power of the real numbers).

Excerpt 3 continued

When a set \(P\) is such that its first derived set \([\text{the set of its limit points}]\) \(P^{(1)}\) is a divisor \([\text{is contained in} \ P]\), I will say that \(P\) is a closed set.

Each closed set \(P\) of a larger power than the first \([\text{meaning each closed, nondenumerable set}]\) decomposes, as we know, in a unique way into \([\text{can be expressed as the disjoint union of}]\) a set \(R\) of the first power \([\text{denumerable}]\) and a perfect set \(S\). [This is the Cantor-Bendixson Theorem, and Cantor applies it because at this point he knows that a closed set is the derived set of another set [6, p. 212].] Using our theorems we conclude the following: “All closed sets of points fall into two classes, the ones of the first power, and the ones of the power of the arithmetic continuum” \([\text{the set of real numbers}]\).

The conclusion follows immediately from the Cantor-Bendixson Theorem. On the one hand, if an infinite closed set cannot be written as the union of a denumerable set and a perfect set, then its cardinality is not larger than the first (that of the integers), so it is denumerable. On the other hand, if an infinite closed set is known to have larger power than the first, then it can be
decomposed into a denumerable set and a perfect set, which, since the perfect set has the same power as the real numbers, means that the closed set has the same power as the real numbers too.

**Excerpt 3 continued**

In an upcoming communication I will show that this dichotomy also holds for sets of points that are not closed. By this we will arrive at, with the help of the principles of paragraph 13 of my memoir in Acta Mathematica 2, page 390, [reference [4]] the determination of the power of the arithmetic continuum, by showing that it coincides with that of the second number class.

At this point, Cantor believed perfect sets were the key to proving the CH, which, recall, states that if an infinite subset of the real numbers is not denumerable, then it has the same power as the set of real numbers. Perfect sets elucidated the structure of the continuum [5, p. 111], which motivated him to keep studying them.

Cantor’s proof that perfect sets, even if nowhere dense, had the power of the continuum also strengthened his conviction that the CH was true and, as the end of Excerpt 3 of his letter shows, led him to believe he was closer than ever to proving it. However, no upcoming communication by Cantor proved the CH; in fact, the CH was surprisingly shown to be undecidable, meaning it could not be proved or disproved, by Gödel, who in 1936 proved it was consistent with the standard Zermelo-Fraenkel axioms of set theory, and Cohen, who in 1963 proved its negation was consistent with those axioms [5, p. 268]. Nonetheless, Cantor’s study of the CH was fruitful, in great part because it led him to a deep study of the properties of infinite sets of real numbers.

**References**


