NOTES ON A BRIDGE FROM CLASSICAL TO MODERN ANALYSIS.

Fréchet has pointed out that it is impossible to define a metric appropriate to the function space in which $f_n \to f$ when $f_n(x) \to f(x)$ for each $x$. The reason is that in any metric space "the limit of a limit is a limit", that is, if $a_1, a_2, a_3, \ldots$ are limit points of a set $S$ and $a_n \to a$, then $a$ is a limit point of $S$. But Baire showed that this does not hold for functions with limit defined as above. Let $S$ be the set of continuous functions.

Functions in Baire Class 1 are limits of continuous functions. Functions in Baire Class 2 are limits of functions in Baire Class 1 but are not limits of any sequence of continuous functions.

For example let $f_n(x)$ be 1 when $2^n x$ is an integer, 0 otherwise. It is possible to find a sequence of continuous functions tending to $f$. Let $f(x)$ be 1 when $x$ multiplied by any positive integral power of 2, gives an integer, 0 otherwise. Then $f_n \to f$ but (as will be proved below) no sequence of continuous functions tends to $f$.

The idea of Baire category, of a "sparse set", appears for the first time in Baire's demonstration of this result. The idea of category generalizes without any difficulty at all, from the real line to any metric space, and plays an important role in metric space theory.

The present notes give an account of Baire's proof, and may serve the following purposes: (1) to show whence the concept of Baire category came, (2) to explain Fréchet's remark that there are topological spaces (i.e. spaces where limit is defined) for which no metric can be found, (3) to show a way of approaching and visualizing proofs in analysis that may sometimes be found useful.

Since Baire was working on the classification of discontinuous functions, we begin by recalling some terminology related to discontinuities.

**Terminology.**

Consider a particular example of a discontinuous function, $f$, where $f(a) = a^2$ for rational $a$ and $-a^2$ for irrational $a$. The graph, as shown here, consists of 2 dotted parabolas. Consider the values of $f(x)$ when $x$ is in an interval $[a-b, a+b]$ and $b$ is small. Whether $a$ is rational or irrational, the upper bound will be $(a+b)^2$ and the lower bound $-(a+b)^2$, if $a$ and $b$ are positive. As $b \to 0$, the bounds tend to $a^2$ and $-a^2$; for any function, the corresponding limits are denoted by
M(a) and m(a). The difference, \( w(a) = M(a) - m(a) \), which in
our example is \( 2a^2 \), may be called the jump at \( a \).

If \( a \) is rational, the corresponding point on the graph is
on the upper parabola. If we take values of \( x \) near to \( a \), we
may find \( f(x) \) much less than \( f(a) \), because we have gone to
the lower parabola, but it is impossible for \( f(x) \) to be much
larger than \( f(a) \). Here \( f \) is called upper semi-continuous
and \( f(a) = M(a) \).

In the same way, if \( a \) is irrational, it is impossible for
a small variation to produce a large decrease in the value of
the function. In this case \( f(a) = m(a) \) and the function is
lower semi-continuous at this point.

At the origin, \( f(0) = N(0) = m(0) \) and the jump \( w(0) = 0 \).
Here the two halves of continuity have got together to
produce continuity.

Fréchet has pointed out, in Les Espaces Abstraits, that
the length of a curve provides a simple and natural example
of semi-continuity. In Figure 2,
graph \( A \) is a smooth curve, which
we may think of as a thread. In
order to make sure that the curve
is only altered "a little" we
enclose the thread in a narrow
tube. By wrinkling the thread
I can obtain a much longer curve,
\( x \), without going outside the
tube, but by tightening the thread
I can only make the length inside the tube a little shorter.
So the function curve \( \Rightarrow \) length is lower semi-continuous,
but not continuous.

The union of certain figure sets.
At a point \( x \) where \( f \) is continuous, \( w(x) = 0 \). Points
of discontinuity have \( w(x) > 0 \). Baire first uses the
concept of sparse set ("set of first category") in
discussing the discontinuities of a function. We will not
go into the details of this theorem, but will merely indicate
how the concept of sparse set enters the argument. As just
mentioned, at a point of discontinuity, \( w(x) > 0 \). We may use
\( S_1 \) to describe the set of points where \( w(x) > 1 \), \( S_2 \) for the
set where \( w(x) > 1/2 \) and generally \( S_n \) for the set of
points with \( w(x) > 1/n \). Each set includes the earlier
ones but that is not particularly relevant. The set of all
discontinuities is \( U S_n \), the union of all these sets as
\( n \) runs through the natural numbers. In certain circumstances
each of the sets \( S_n \) is nowhere dense. \( U S_n \) is thus the union
of \( \frac{1}{n} \) nowhere dense sets ; we condense this by saying
that \( U S_n \) is a sparse set. Baire proves that no sparse set
can contain all the points in an interval of the real line.

The proof is very easy. \( S_1 \) is nowhere dense. This
means that in any interval \( I_0 \) we can find a closed interval \( I_1 \)
free from points of \( S_1 \).
We now operate in $I_1$ instead of $I_0$; the effect is that we have got rid of $S_n$ and it has not cost us anything, for in this work one interval is as good as another. No doubt $I_1$ is shorter than $I_0$, but none of the properties we are dealing with is altered by a change of scale. So we continue; within $I_1$ we find an interval $I_2$ free from points of $S_n$, and so on. The nested closed intervals $I_1, I_2, I_3, \ldots$ are bound to have at least one common point, which will be in none of the sets $S_n$, hence not in $US_n$. So $US_n$ does not contain all the points of $I_0$. Q.E.D.

This proof appears to show that there is one point not in $US_n$, but in fact there will be very many. For instance, we could equally well carry out the above construction within any sub-interval of $I_0$ and find a point in it not in $US_n$.

The concept of sparse set will be seen in action in the proof below of the main theorem. This theorem depends on the concept of total discontinuity. A function $f$ is called totally discontinuous if there is an interval at every point of which the jump $|f(x) - f|$, where of course $K > 0$. (There are other, equivalent definitions.) The function graphed in Figure 1 for example is totally discontinuous. So is the function $f$ mentioned in the second paragraph of page 1 above. So is the well known function with value 1 for $x$ rational, 0 for $x$ irrational.

Baire proved that, if $f_1, f_2, f_3, \ldots$ is a sequence of continuous functions, and $f$ a bounded function such that $f_n(x) \to f(x)$ for each $x$ in some interval, then $f$ cannot be totally discontinuous.

A Lemma: Appearance and Reality.

Baire starts his proof with a lemma, which may serve to illustrate the general principle that proofs in analysis usually look much harder than they are. When you glance at the proof of this lemma, you see 9 inequations. Your mind is perhaps already tired with what you have read, and the lemma may loom in your way as a substantial obstacle. In fact, once you have seen what the lemma means, its proof becomes almost a single mental act. The lemma assumes that there is some set $S$ of real numbers, for which $\sup S = \inf S > 2k$. It states that if $a$ is any real number whatever, there exists a member of $S$ whose distance from $a$ exceeds $k$. 

3.16
Figure 3 shows the situation. The only information we have about the set relates to \( \sup S \) and \( \inf S \). For all we know, the set may consist of only 2 members, one at \( \sup S \) and one at \( \inf S \). If \( \sup S \) is not in \( S \), there must be members of the set just below it. If we mark on our figure a narrow interval, with its ceiling at \( \sup S \), we can be sure this interval is inhabited by elements of \( S \). In the same way we can mark an inhabited interval rising from \( \inf S \). The fact that both intervals are inhabited is all the information at our disposal.

The assertion in the lemma is now obvious. In the figure, if \( a \) is below \( Q \), its distance from an inhabitant of the upper interval must exceed \( k \); if it is at or above \( Q \), its distance from the lower interval exceeds \( k \).

Clearly, Baire did not regard this lemma as expressing any great or surprising truth. It was just a detail he wanted out of the way, so the reader would not be distracted by it in the middle of the main proof. Personally, having seen what this lemma amounted to, I would not even bother to check Baire's 9 inequalities. These cannot amount to anything more than putting into formal shape the visual argument just outlined.

**Proof of the Main Theorem**

If you make a few attempts to construct a sequence of continuous functions with the function graphed in Figure 1 as a limit, you will not find it hard to believe that this task is in fact impossible, as the theorem implies. The functions are required to loop about in a very violent way. Baire's proof sets out to establish a contradiction between the following statements:

1. \( f \) is bounded and totally discontinuous.
2. Each \( f_n \) is continuous.
3. \( f_n(x) \to f(x) \) for each \( x \) in \( I \).

The break eventually comes in statement (3). Baire finds an \( x \) for which \( \{f_n(x)\} \) is not a Cauchy sequence, so it does not converge at all, and hence cannot tend to \( f(x) \).

Accordingly our concern is with the tail of the sequence \( f_p, f_{p+1}, f_{p+2}, \ldots \). We allow an opponent to select the number \( p \).

As \( f \) is totally discontinuous, in some interval \( I \) at every point \( w(x) \to x \).
For convenience of illustration, we choose a particular \( K \), say \( K = 14 \). This means that, in any vertical strip, however thin, we can find points in the graph of \( f \) with vertical separation exceeding 14. Within \( I_o \) we let the opponent choose any interval \( I_0 \) and any \( a_o \) in \( I_o \). By the lemma, we can find \( a_1 \), as near as we like to \( a_o \), with \( |f(a_1) - f(a_o)| > 7 \).

(See Figure 4.)

In drawing diagrams to illustrate arguments in analysis, we must always load the dice against ourselves. The figure must show the situation least favourable to our argument. For instance, in Figure 4, we want the difference between \( f(a_1) \) and \( f(a_o) \) to be more than 7. Accordingly \( f(a_1) \) is shown half-way up the region occupied by the graph, for this is the place it is hardest to be remote from.

We do not proceed immediately to select \( a_1 \). Instead, we first look at the functions \( f_n \) assumed to tend to \( f \). As \( f_n(a_o) \to f(a_o) \) we can find \( N > p \) so that

\[
|f_N(a_o) - f(a_o)| < 1.
\]

As \( f_n \) is continuous, we can choose an interval \( I_1 \), within \( I_o \), throughout which \( |f_N(x) - f_N(a_o)| < 1 \).

(See Figure 5.)
It is now that we choose \( a_1 \), inside \( I_1 \), so that
\[
|f(a_1) - f(a_0)| > 7.
\]
Then we have two steps similar to those done a little earlier. We choose \( M > p \) so that
\[
|f_M(a_1) - f(a_1)| < 1.
\]
As \( f_M \) is continuous, there is an interval \( I_2 \) throughout which
\[
|f_M(x) - f_M(a_1)| < 1.
\]
(See the lower part of Figure 5.)

It is evident that, in \( I_2 \),
\[
|f_M(x) - f_N(x)| > 3.
\]

Now if \( \{y_p\} \) is a convergent sequence, the differences between the terms should become indefinitely small, so certainly less than \( 3 \). We shall say the sequence has "settled down after \( p \) terms" if \( M \geq p, N \geq p \) implies
\[
|y_M - y_N| < 3.
\]

Our work above shows that in any interval \( I_0 \) chosen by the opponent we can find an interval \( I_0 \) in which the sequence \( \{f_n(x)\} \) has not settled down by the \( p \)th term.

If \( S_p \) denotes the set of \( x \) for which \( \{f_n(x)\} \) has settled down after \( p \) terms, this means that the set \( S_p \) is nowhere dense.

If \( x \) is not in the set \( S_p \), it does not follow that the sequence \( \{f_n(x)\} \) diverges. It still has time for a deathbed repentance; it may settle down for some number larger than \( p \). But to converge, it must settle down at some stage. Now the choice of \( p \) was left to the opponent. We have shown that for each \( p, S_p \) is nowhere dense. If the sequence converges, it must \( p \) settle down for some value of \( p \); that is, \( x \) must lie in \( U S_p \).

But \( U S_p \) is a sparse set, and cannot fill the interval \( I \).

At any point not in the sparse set, the sequence \( f_n(x) \) fails to converge, and we have a contradiction.

Note. It can now be seen why Figure 5 was drawn as it was. We are trying to establish that \( f_M(x) \) and \( f_N(x) \) are far apart. All the inequalities are therefore drawn in such a way that \( f_M \) and \( f_N \) are made to approach each other as much as possible—the case most unfavourable to our argument.