6336

ETA-LORE

by Douglas R. Hofstadter

Introduction.

The sequences with which this paper deals are infinite sequences composed of a finite number of distinct integers; they have the property of being quasi-periodic sequences, by which I mean that any finite "chunk" which occurs somewhere in a particular sequence will actually occur infinitely often in that sequence. Probably the most important consequence of this is that the sequence can be thought of as a sequence of "chunks", as well as a sequence of integers; now if each distinct "chunk" has a name, then one can specify the entire sequence uniquely, simply by stating the names of the chunks which compose it, in the order in which they occur in the sequence. If integers are chosen to be the "names" of chunks, then the "chunk-description" of the original sequence is itself a new sequence of integers, and it is called the "derivative" of the original sequence (nothing to do with calculus). The sequences that are most interesting are those whose derivatives also have derivatives, which also have derivatives, which also....... Eta-sequences constitute a special case of this kind of "infinitely-differentiable" sequence; in fact, the derivative of an eta-sequence is another eta-sequence. The feature which characterizes eta-sequences is that they only contain occurrences of two distinct integers -- in fact, consecutive integers. In a certain sense, there is a one-to-one correspondence between the set of all eta-sequences and the set of all real numbers -- but at this point, instead of continuing with general results about eta-sequences, I will present eta-sequences in the order in which I developed an acquaintance with them, and the general results will fit smoothly into that context. (I was not the first person to discover eta-sequences; to the best of my knowledge, A. Markov and G. Christoffel were the first to investigate them, towards the end of the last century. But since then, apparently no new work has been done on eta-sequences -- at least I don't know of any published work on eta-sequences after those articles.)

An eta-sequence crops up.

I first came across an eta-sequence as I was working on a problem having to do with squares and triangular numbers. (A triangular number is a sum of consecutive integers, beginning with 1 -- such as 1 + 2 + 3 + 4 + 5 = 15; a square can be similarly described as a sum of consecutive odd integers, beginning with 1 -- for instance 1 + 3 + 5 + 7 = 16.) In the course of this problem, I asked myself how many triangular numbers there are, on the average, between successive squares. Below, I show the result of a simple empirical investigation, with small numbers:

```
 1      4      9      16      25      36      49      64      81      100 ...
        |      |       |       |       |       |       |       |       |
 1    3 |  6   | 10 15 |  21   |  28   |36  45 |  55   | 66 78 |  91.. |...
        |      |       |       |       |       |       |       |       |

   2.....1.....2.......1.....1.......2.......1.......2.......1........
```

As you can see, there seem always to be either 1 or 2 triangular numbers between successive squares. (Where a triangular number and a square coincide (e.g. 36), I have treated the triangular number as if it were greater than the square. Had I treated it as if it were smaller than the square, a similar sequence of 1's and 2's would have resulted -- more on that later.) Below, I exhibit many more terms of the sequence shown above:

```
212112121211212112121211212112121121212112121121212112121121212112...
```

Certainly this sequence gives a strong visual impression of being quasi-periodic; in fact, one might naively guess that it is actually a periodic sequence. As it turns out, such is not the case. A natural thing to do in looking at this sequence is to break it into "chunks" -- probably "21" and "211" are the simplest choice. Below, the same sequence is given, broken into these "chunks":

```
21 211 21 21 211 21 211 21 21 211 21 211 21 211 21 21 211 21 211 ...
```

After a while of staring at this segmented sequence, it would occur to most people that the 211's always occur singly, whereas the 21's sometimes occur singly, sometimes in pairs. So why not write down the number of 21's between successive 211's? This is done below. (Note that this operation, although based on "chunks", is different from taking the derivative. However, the two operations are actually very closely related; their relation will be covered soon.)

```
21 211 21 21 211 21 211 21 21 211 21 211 21 211 21 21 211 21 211 ...
 *  |   *  *  |   *  |   *  *  |   *  |   *  |   *  *  |   *  |
(1)       2          1      2          1       1       2        1   ....
```

We will ignore the parenthesized "1" at the front of the lower sequence, because it occurs before the first "211". What do we have? It is a sequence which reads "2121121...". It occurs to us that this new sequence may actually be just the old sequence! Of course such a hypothesis needs to be checked further (it checks)...and then proved, if possible. How to prove it is not obvious. We postpone the proof to its correct chronological place in my personal development of eta-sequences, and instead proceed to a second place where an eta-sequence cropped up.

Another eta-sequence crops up

My curiosity was piqued by this (empirical) discovery, so I tried to invent similar problems. One of these was the following: how many powers of 2 are there between successive powers of 3? We can construct the first few terms of the sequence as before:

```
 1        3       9       27       81       243       729      2187...
          |       |       |        |        |         |         |
 1    2   |4    8 |  16   | 32  64 |   128  | 256  512 |1024  2048|....
          |       |       |        |        |         |         |

     2........2.......1.........2.........1.........2..........2.........
```

Again, only 1's and 2's appear; it is natural to try the same trick of looking at chunks, only here the chunks are not "21" and "211", but "21" and "221". Below, we give the sequence, in "pre-chunked" form:

221 21 221 21 221 21 21 221 21 221 21 21 221 21 221 21 21 221 21 221

Now we could count the 21's (which occur singly and doubly) between 221's -- but this time let us form the true derivative, by assigning integers as names of the two distinct chunks. We can give to chunk 21 the name "1", and to chunk 221 the name "2". The chunk-sequence (the derivative) can be simply read off of the above line, and we get:

2 1 2 1 2 1 1 2 1 2 1 1 2 1 2 1 1 2 1 2

It's another quasi-periodic sequence, composed of just two integers! (I put an exclamation mark, because I think it's surprising. However, it is probably not as much of a surprise to you as it was to me, because a few lines back you were told to expect it.) Now this sequence is not the same as any of the previous ones we have seen, even though it looks very similar to them. But it shares with the other sequences the fact that it breaks up into natural chunks (21 and 211) so we can form its derivative (which will be the second derivative of the original sequence). And as was said earlier, this process can continue indefinitely.

It is easy to derive a formula for the kth term of the powers-of-2-between-powers-of-3 sequence. First observe that the number of powers of 2 up to (and including) a number N is

$$1 + [\log_2 N]$$

where "[x]" stands for "the greatest integer less than or equal to x".

For instance, up to 9 there are 4 powers of 2 (1,2,4,8), and log (base 2) of 9 is a shade over 3, so that our expression gives the right answer. Between the kth power of 3 and the k+1st power of 3, therefore, the number of powers of 2 is

$$1 + [\log_2 (3^{k+1})] - 1 - [\log_2 (3^k)])$$

$$= [(k+1) \log_2 3] - [k \log_2 3]$$

A general formula for eta-sequences

This expression provides us with a model that we can easily generalize, as follows:

$$eta_k (alpha) = [(k+1) \; alpha] - [k \; alpha]$$

Here, we can take alpha to be any real number. For each value of alpha, we get a characteristic sequence, eta(alpha). (In fact, the name which Christoffel gave to eta(alpha) was "characteristic sequence of alpha".) It is straightforward to show that always, eta(alpha) contains the two integers between which alpha lies, and only those two integers. (If alpha is itself an integer, eta(alpha) is totally trivial, consisting merely of an endless sequence of alpha's.) So suppose alpha is pi; then eta(alpha) must be composed exclusively of 3's and 4's. In fact, eta(pi) runs like this:

33333343333334333333433333343333334333333433333343333334333333433...

The sequence itself is composed of 3's and 4's; the "natural" chunks which the eye breaks it up into are "3333334" and "33333334" (the first occurrence of the latter chunk is a bit further out than what is shown above.) To form a derivative of eta(pi), we must give integer-names to these two chunks; one possibility would be

```
"Old            3333334 --> 6
  Style"       33333334 --> 7,
```

where the name tells the number of 3's in the chunk; or else we could simply tell the total length of the chunk, like this:

```
"New            3333334 --> 7
  Style"       33333334 --> 8.
```

Both styles are quite natural. If we take the former choice, our derivative sequence will count 3's between successive 4's. This was the original notion of derivative, and from it sprang the terms "coun" -- here 3 -- and "sep" (for "separator") -- which is 4 here. In general, the coun will be the closest integer to alpha, while the sep is the second-closest integer to alpha. The eta-sequence of alpha will always have seps occurring singly, and couns variably. To make the derivative, you count couns between seps. Notice that if you literally mean "between seps", you have to disregard the very first group of couns, since they precede the first sep. This was the definition of derivative for a long time -- I will call it the "old-style" derivative.

The new-style derivative is almost the same -- it's just that (1) every term is one bigger than in the old-style derivative, and (2) there is an extra term in the new-style derivative, corresponding to the first group of couns and sep.

Now suppose we take the old-style derivative of eta(pi). I will leave out the actual sequence, because you have seen enough of them to get the picture; it is composed of 6's and 7's and has that typical appearance of quasi-periodicity which is so characteristic of eta-sequences. It is itself an eta-sequence -- but to what value of alpha does it belong? Clearly this is a key question.

The Fundamental Theorem of Eta-sequences

The answer is given by this simple, central result:

Fundamental Theorem of Eta-Sequences. eta'(alpha) = eta(alpha'), where by eta'(alpha) is meant the old-style derivative of eta(alpha), and by alpha' is meant the quantity

$$\frac{s - alpha}{alpha - c},$$

where "s" stands for the sep of alpha, and "c" for the coun of alpha.

In the case of pi, this tells us that eta'(pi) is the eta-sequence belonging to (4-pi)/(pi-3), which comes out to about 6.0625. Our knowledge of eta-sequences so far tells us that we should expect the eta-sequence of 6.0625 to consist mostly of 6's, with sparsely-spaced 7's, which is just what eta'(pi) looks like. What about a proof for the above result? The proof is given below. Once it is proven for values of alpha between 0 and 1/2, it follows quickly for all values of alpha.

When 0 < alpha < 1/2, coun(alpha) is zero, and sep(alpha) is one, so eta(alpha) consists of a row of zeros and ones. By referring to the figure below, you can visualize where seps ("1") occur, and where couns ("0") occur. The real axis is plotted horizontally, and three integers (N-1,N,N+1) are shown. Also, multiples of alpha are indicated by the letter "a". (Incidentally, this figure shows why I sometimes call eta-sequences "sidewalk-sequences". If you take steps of length alpha down a sidewalk whose cracks are 1 unit apart, the number of cracks you cross on the kth step is the kth member of eta(alpha), provided the zeroth step begins exactly on a crack.)

```
            |                   |                   |
.....a..|a...a...a...a..|a...a...a...a..|a...a......
            |                   |                   |

           N-1                  N                  N+1
```

Stepping from one "a" to the next, you can cross either one integer, or none. When you cross one, a "1" appears in the eta-sequence; when you cross none, a "0" appears. Sooner or later, each integer -- say N -- gets straddled by two successive multiples of alpha. When that happens, a "1" appears in the eta-sequence; in fact, it must be exactly the Nth "1". Our goal is to count the number of zeros until the next "1".

Suppose that the multiples of alpha which straddle N are p alpha and (p+1)alpha; and that the multiples of alpha which straddle N+1 are q alpha and (q+1)alpha. Then

$$\text{eta}_p(\text{alpha}) = \text{eta}_q(\text{alpha}) = 1$$

and all terms of eta(alpha) between the pth and the qth are zeros. So how many zeros does that make? Exactly q-p-1. And this will be, by definition, the Nth term of the derivative of eta(alpha). Now we can specify both p and q in terms of N; there are exactly p multiples of alpha up to N, which means

p = [N/alpha];

similarly, q = [(N+1)/alpha].

Putting our pieces of knowledge together, we know that the Nth term of eta'(alpha) is equal to q-p-1; and this is

[(N+1) beta] - [N beta],

where beta = 1/alpha - 1. What we have is the expression for the Nth term of eta(beta); moreover,

$$\text{beta} = \frac{1 - \text{alpha}}{\text{alpha} - 0} = \text{alpha}'$$

which proves our theorem for 0 < alpha < 1/2.

Suppose alpha lies between 0 and 1/2, and alpha + gamma = 1. Then one can easily show that eta(alpha) and eta(gamma) are complementary to each other, in the sense that where "0" occurs in one, "1" occurs in the other, and vice versa. Consequently their derivatives are the same sequence. That is,

$$\text{eta}'(\text{gamma}) = \text{eta}'(\text{alpha})$$

But eta'(alpha) is known, from above: eta'(alpha) = eta(1/alpha - 1). And

$$1/\text{alpha} - 1 = \frac{1 - \text{alpha}}{\text{alpha}}$$

$$= \frac{\text{gamma}}{1 - \text{gamma}} = \text{gamma}'$$

So we have shown that eta'(gamma) = eta(gamma'), for any gamma between 1/2 and 1. The only remaining values for which the theorem needs to be proven are those of the form N + alpha, where 0 < alpha < 1. It is trivial to show that eta(N + alpha) = N + eta(alpha), and from this it follows that eta'(N + alpha) = eta'(alpha). It is just algebra to show that (N + alpha)' = alpha', and this completes the proof.

I now mention two other results whose proofs are extremely simple:

Theorem. As N approaches infinity, the average of the first N terms of eta(alpha) approaches alpha.

Theorem. If alpha = p/q (a rational number in lowest terms) then eta(alpha) is a periodic sequence, with period q. If alpha is irrational, then eta(alpha) is not periodic.

Triangles-between-squares seen in a new light.

Let us now go back to the triangles-between-squares example. As I pointed out, the operation we performed on the sequence was not exactly taking the derivative. The chunks we perceived were 21 and 211; to form the derivative, we should replace 21 by some integer, and 211 by a different integer. Let's do it in the "old style":

```
 21 --> 1
211 --> 2
```

The sequence and its derivative are shown below.

```
21 211 21 21 211 21 211 21 21 211 21 211 21 211 21 21 211 21 211 ...
 1   2  1  1   2  1   2  1  1   2  1   2  1   2  1  1   2  1   2 ...
    -------  ----  -------  ----  ----  -------  ----
```

The underlining highlights the fact that the derivative is also composed of 211's and 21's. It is, however, not identical to any of the sequences we have exhibited so far. But we can take another derivative, thus getting the second derivative of the original sequence. It is

```
        2       1       2       1       1       2       1 ....
```

Now this sequence appears familiar -- it looks like the original sequence! It is indeed the original sequence. So, for the second time, we have found the original sequence coded in itself. The first time, we got it by counting 21's between 211's; the second time, we got it by taking the second derivative. It turns out that the two processes are really one and the same process, presented in two superficially different ways. Consider the following idea: when we took the second derivative, we broke the first derivative into chunks; now each single term in the first derivative represented a chunk in the top-level sequence, so that to each chunk in the first derivative, there is a "chunk of chunks" in the top-level sequence. Here it is visually:

```
21  211-21-21  211-21  211-21-21  211-21  211-21  211-21-21  211-21 ...
     2--1--1   2---1   2---1--1   2---1   2---1   2---1--1   2---1 ...
       2         1        2         1       1        2         1 ...
```

Each number in the bottom sequence reflects the occurrence in the top sequence of a "second-order chunk" (chunk of chunks) -- either 2112121, or 21121. Now when we counted 21's between 211's, we were in effect assigning "2" as a name to the superchunk 2112121, and "1" to the superchunk 21121. And that is exactly what taking the second derivative does, too. So the two processes come down to the same thing. This points out the important fact that what you get for your first derivative depends on what you choose for "chunks", in the top-level sequence . If we'd chosen 2112121 and 21121, then the first derivative would have given us back the original sequence, but with 211 and 21 as chunks, you have to wait until the second derivative to return the top-level sequence.

In eta-sequences, the usual choice for a chunk just contains a single sep, preceded or followed by some couns. Such chunks are called "first order chunks". The natural way to make "second-order chunks" is to make superchunks in the top-level sequence which reflect the first-order chunks in the first derivative. Third-order and higher-order chunks are defined analogously. There are always just two Nth-order chunks, no matter what N is. The higher the order of a chunk, the longer it is bound to be. As N goes to infinity, the length of both Nth-order chunks goes to infinity. A 17-th order chunk is a giant segment of the eta-sequence of alpha, and the arithmetic average of its terms provides a good approximation to alpha. We will go into this in more detail shortly.

Eta-sequences in full generality

Going back to the sidewalk-image, recall that I pointed out that the sidewalk-sequence gives eta(alpha) -- provided you start on a crack. What happens if you don't start on a crack? Suppose that the starting-point of your zeroth step is displaced by a distance delta from a crack. Then you get something very much like an eta-sequence. In fact, I call it eta(alpha;delta). What we have previously called eta(alpha) is the same as eta(alpha;0). The formula for the kth term of eta(alpha;delta) is easy to derive, and is:

$$[(k + 1)\ \text{alpha} + \text{delta}] - [k\ \text{alpha} + \text{delta}]$$

From now on, by "eta-sequence", I mean something of the above form. Now suppose delta equals minus alpha. Then we have

$$\begin{aligned} \text{eta}_k(\text{alpha};-\text{alpha}) &= [(k + 1)\text{alpha} - \text{alpha}] - [k\ \text{alpha} - \text{alpha}] \\ &= [k\ \text{alpha}] - [(k - 1)\ \text{alpha}] \\ &= \text{eta}_{k-1}(\text{alpha};0) \end{aligned}$$

As a matter of fact, this is intuitively obvious: shifting every step to the left by alpha only postpones arriving at a given spot by exactly one step. In general, shifting every step to the left by m times alpha has the effect of postponing the moment of arrival at a given spot by m steps:

$$\text{eta}_k(\text{alpha};-m\ \text{alpha}) = \text{eta}_{k-m}(\text{alpha};0)$$

You can move the whole sidewalk to the right or left by one square and nobody will know the difference. This is saying that you can add or subtract any integer to delta and the eta-sequence won't change. In other words, only the fractional part of delta -- which is denoted as "{delta}" -- matters. In symbols,

$$\text{eta}_k(\text{alpha};\text{delta}) = \text{eta}_k(\text{alpha};\{\text{delta}\})$$

There is a generalization of the Fundamental Theorem of Eta-Sequences, which holds for all eta-sequences. I state it without proof, since the proof follows the lines of the earlier theorem.

Generalized Fundamental Theorem of Eta-Sequences.
eta' (alpha; delta) = eta (alpha'; delta'), where alpha' is as before, and delta' can be defined the following way:

```
                  /  f({alpha},{delta}),   if {alpha} < 1/2
     delta'   =   |
                  \  f({-alpha},{-delta}),   if 1/2 < {alpha}
```

where

```
                  /  -y/x   for x+y < 1
     f(x,y)   =   |
                  \  (1-y)/x   for x+y > 1.
                                       -
```

So that you are not deprived of the experience of seeing an eta-sequence whose delta is non-zero, I now exhibit eta (sqrt 2; 1/2):

2121121212112121121212112121121211211212121121211212121121211212121 1...

Do you recognize this sequence? It is our old friend, Triangles-between-squares! When I discovered this, I was really amazed. How is it possible to prove this? Actually, it is quite easy. First, let us derive a formula for Triangles-between-squares. The nth triangular number is equal to n(n+1)/2; if we invert this function, we will get a function that tells us how many triangular numbers there are up to a given size. In other words,

[sqrt(2N - 1/4) -1/2]

is the number of triangular numbers less than or equal to N. Therefore we should evaluate this quantity using the square of k+1 as N, and then the square of k, and then take the difference:

$$[\mathrm{sqrt}(2(k+1)^2 - 1/4) -1/2] - [\mathrm{sqrt}(2k^2 - 1/4) - 1/2]$$

This expression gives the kth term of triangles-between-squares. It can be simplified, with the help of the following identity (whose not-too-tricky proof is omitted, since it is not central to eta-theory):

$$[\mathrm{sqrt}(2k^2 -1/4) - 1/2] = [k\ \mathrm{sqrt}\ 2 - 1/2]$$

With it, we get a revised expression for the kth term of the triangles-between-squares sequence:

```
[(k+1) sqrt 2 - 1/2] - [k sqrt 2 - 1/2]

              =  eta (sqrt 2; -1/2)
                    k

              =  eta (sqrt 2; 1/2)
                    k
```

which is the desired amazing result -- perhaps less amazing for its scrutability. With it, we can at last prove the observation that when you count 21's between 211's in Triangles-between-squares, you get the same sequence back. We now know that counting 21's between 211's is the same as taking the second derivative; and so the question amounts to whether or not eta(sqrt 2; 1/2) equals its own second derivative. A few manipulations show that alpha' and alpha'' are both equal to sqrt 2, and that delta' equals (3 - sqrt 2)/2, and delta'' equals 1/2; and that is that.

Now to round out our discussion on triangular numbers between squares, we can take a look at what happens when we handle coincidences of triangles and squares (such as 36) in the other way than before. This means counting the triangular number as if it were less than the square of the same magnitude:

```
 1     4     9     16     25     36     49     64     81     100 ...
 |     |     |      |      |      |      |      |      |      |
1|   3 |  6  |10  15|  21  |28  36|   45 |  55  |66  78 |  91  |
 |     |     |      |      |      |      |      |      |      |

(1)...1....1......2.....1......2......1......1......2......1.........
```

We have a new quasi-periodic sequence of 1's and 2's which differs from the original one only in a few scattered places. (Actually, saying "a few" is a distortion; there are in fact infinitely many squares which coincide with triangular numbers, but such coincidences are quite sparsely spaced -- or, to justify my earlier terminology, they are "few and far between"!) If you count the parenthesized "1", this sequence begins with a trio of 1's. It's displayed more fully below, together with its first and second derivatives, using 12 and 112 as "chunks".

```
(1)1121211212112121121212112121121212112121121211212121121211212121121.
     2  1 2  1 2  1 2  1 1 2  1 2  1 1 2  1 2  1 2  1 1 2  1 2  1 1 2...
         1    1    1     2     1     2     1    1     2     1     2
```

Well, not surprisingly by now, it equals its second derivative (but is nevertheless different from the other version of triangles-between-squares).

Finite segments of eta-sequences

Suppose you had a window six units wide, through which you could look at some eta-sequence. What I mean by this is that you could see exactly six consecutive terms of that eta-sequence. How many different scenes could you be entertained by? Naturally, it ought to depend on which eta-sequence you have got, so let us try it with eta (sqrt 2; 0). Here are the possible views through a 6-unit window:

112121
121121
121211
121212
211212
212112
212121

Seven, they are seven. What may surprise you is that this result holds whatever the eta-sequence -- there are always seven different views through a 6-unit window. And there is nothing special about 6; if you have a window of width n, there are always n+1 distinct views to be savored. Actually, the claim is not quite true; it requires alpha to be irrational. Thus the theorem can be stated formally this way: Given an eta-sequence belonging to an irrational alpha, there are exactly n+1 distinct segments of length n. This may seem like a remarkably simple answer to a complex combinatoric problem; but although it can be looked on combinatorically, there is an easy route to the answer which avoids any combinatorial analysis. The proof is as follows.

Each of the distinct segments of length n can be produced by using an appropriate value of delta, and generating the first n terms of eta(alpha; delta). Since there are only a finite number of distinct segments, but an uncountable number of delta's between 0 and 1, many delta's yield the same segment. A good guess is that for each distinct segment there is a little interval inside [0,1] where all the delta's produce that segment. This idea is pictured below, using segments of length 2 and alpha = sqrt 2.

```
|all delta's|   all delta's     |    all delta's    |
|here yield |   here yield      |    here yield     |
|   "11"    |      "12"         |       "21"        |

0          0.171...           0.587...              1
```

This is exactly the way things work. The internal demarcation-lines are determined by the first n multiples of alpha (as it turns out). To see this, imagine that the first n multiples of alpha, and zero, have been marked on a transparent plastic sheet which we can slide above the "sidewalk" defined by the integers. Set the plastic sheet on the sidewalk so that their zeros coincide, which sets delta to zero, and gives a segment of length n. Now slide the sheet to the right. Until one of the marks on the sheet crosses an integer-line (crack in the sidewalk), none of the terms in the segment will change. When a mark does eventually cross a line, the segment will change. Remember the kth term of the segment is given by the number of lines crossed between the k-1st and kth marks. Therefore, when the kth mark crosses a line, the kth term in the segment increases by 1, and the k+1st term decreases by 1. (Exceptions: when the zero-mark crosses a line, only the first term of the segment changes, decreasing by 1; and when the rightmost mark crosses a line, only the last (i.e. the nth) term of the segment changes, increasing by 1.) As the sheet continues sliding to the right, one by one, the marks will cross integer-lines. No two will do so simultaneously, because alpha is irrational. Now eventually, the zero on the plastic sheet will reach the integer 1 on the sidewalk. Once that has happened, the whole thing starts over again. But in the meantime, each mark on the sheet will have crossed exactly one integer. (This must be so, because the sheet has moved one integer unit to the right.) Since there are n+1 marks on the sheet, there have been n+1 distinct segments generated. That's the proof, and it corroborates the picture we had of n+1 little intervals in [0,1].

Extraction

We are about to wind up the "first phase" of our discussion of eta-sequences; this phase has consisted largely of explorations of the horizontal aspect of eta-sequences. "Horizontal" properties are those which involve a single eta-sequence, and which make little or no explicit reference to its derivatives. They are horizontal because, obviously, a single eta-sequence is thought of as extending out to infinity horizontally. "Vertical" properties are coming up soon. Now I do not mean to imply that there is a clean separation between horizontal and vertical properties; in fact they are very tangled up together and probably it is a silly distinction -- but the distinction perhaps can aid one's intuition, as one grows used to it. When we come to vertical properties, I am sure that you will get a clearer idea of this distinction.

But now I would like to give an example par excellence of horizontal properties, a property which I call "extraction". The idea is this. To begin with, write down eta(alpha;0). Now choose some arbitrary term in it, called the "starting point". Beginning at the starting point, try to match eta(alpha;0) term by term. Every time you find a match, circle that term. Soon you will come to a term which differs from eta(alpha;0). When this happens, just skip over it without circling it, and look for the earliest match to the term of eta(alpha;0) you are seeking. Continue this process indefinitely. In the end you have circled a great number of terms after the starting point, and left some uncircled. We are interested in the uncircled terms, which are now "extracted" from eta(alpha;0). The first interesting fact is that the extracted sequence is itself an eta-sequence; but what's more, it is the subsequence of eta(alpha;0) which begins two terms earlier than the starting-point! To decrease confusion, I now show an example, where instead of circling I underline the terms which match eta(alpha;0). In this example, alpha equals log (base 2) of 3.

```
               I have chosen this "2" as the starting point.
              |
              |
2 1 2 1 2 2 1 2 1 2 2 1 2 1 2 1 2 2 1 2 1 2 2 1 2 1 2 1 2 2 1 2 1 2 2
              - - -   - -   - - -   - -   -   - - - - - - - - - - - -

1 2 1 2 1 2 2 1 2 1 2 2 1 2 1 2 2 1 2 1 2 1 2 2 1 2 1 2 2 1 2 1 2 1 2
- - - - - - - - - - - - - - - -   - -   - - -   - -   -   - - - -   -
```

The underlined sequence matches the full eta-sequence, term by term. Now what is the extracted sequence? It is:

2 1 2 1 2 2 1 2 1 2 1 .....

And you will find that this matches with the sequence which begins two places earlier than the starting-point. Carrying it further is tedious, and does nothing but confirm our observation. Why does this extraction-property hold? At this point, I must admit that I don't know. It is a curious property which needs further investigation. For example, who knows what happens if, instead of using eta(alpha;0) as the sequence to be matched, you use eta(alpha;delta)?

"Altsum"

As a transition into vertical properties, I cite one last problem which gives the appearance of being a "horizontal" problem, but whose answer turns out to be very intimately related to vertical properties. This is the question of "altsum" -- an abbreviation for "alternating sum". The definition is:

$$\text{altsum}(k) = \text{eta}_1 - \text{eta}_2 + \text{eta}_3 - \ldots + (-1)^{k+1}\,\text{eta}_k$$

(In the above, alpha and delta are assumed known and fixed.) Now if alpha is irrational, it makes sense to guess that the terms of such a sum tend to cancel each other out, more or less, over a long span.

Therefore, the expected approximate behavior of altsum is that it will hover near zero, straying away occasionally, somewhat like the "random walk" of a drunk away from his lamppost. But one expects the fluctuations to be small, in the following sense:

$$\lim \frac{\text{altsum }(k)}{k} = 0.$$

This follows from a general theorem which one can assert about the eta-sequence of any irrational number:

Theorem: The arithmetic average of terms in eta(alpha; delta) whose subscripts form an arithmetic progression is alpha.

The proof of the theorem depends on a famous property of the multiples of any irrational -- namely, that they are uniformly distributed, modulo 1. This means that out of the first N multiples of alpha, the proportion whose fractional parts lie in any interval inside [0,1] is asymptotically equal to the length of that interval. Now from the uniform distribution of the numbers {n alpha}, an immediate corollary is the uniform distribution of the numbers {n alpha + delta}, whatever delta is. We can now apply this fact to prove the theorem. First we make the remark that there is a dividing-line inside the interval [0,1] such that the position of {n alpha} with respect to that line determines whether the nth term of the eta-squence is a coun or a sep. (This just says for segments of length 1 what was said a couple of pages earlier, for segments of arbitrary length.) It is illustrated below.

```
|                         |           |
|     "coun country"      | "septalia"|
|                         |           |
0 . . . . . . . . . . . . . . . . . . 1
```

Suppose the arithmetic progression of subscripts is qk+p, with k varying. Then what matters in the above diagram is where the number {(qk+p)alpha + delta} falls. Numbers of this form are, however, also of the form {k beta + gamma} where beta = q alpha, and gamma = p alpha + delta. Now beta is irrational whenever alpha is, which allows us to say that the multiples of beta, shifted by gamma, are uniformly distributed in [0,1]. Therefore, the proportion of such numbers which land in "septalia" is asymptotically equal to the length of septalia. But the proportion of numbers of the form {k alpha + delta} in septalia is also the length of septalia -- which means that the proportion of seps in the subsequence defined by the arithmetic progression qk+p is asymptotically the same as in the eta-sequence itself. (Of course the proportion of couns is the same too.) Consequently, the average of the subsequence must be equal to the average of the sequence itself.

Finally we can prove that altsum(k) becomes negligible in comparison to k. To get altsum(k), you add up k/2 terms of eta whose subscripts are odd, then you subtract from that the sum of k/2 terms of eta whose subscripts are even. From the just-proven theorem, both the subsequence whose subscripts are odd, and the subsequence whose subscripts are even, have average alpha. Therefore both sums will be of magnitude (k/2) alpha, with correction terms which necessarily become small compared to k, as k approaches infinity. When the even sum is subtracted from the odd, all that is left is a quantity which is small compared to k -- so the limiting value of altsum(k)/k is zero, as we set out to prove.

But how does altsum act, in more detail? When does it have large fluctuations? Below are the first 100 terms of the altsum belonging to eta (sqrt 2; 0), so that you can see for yourself.

1 -1 0 -2 -1 -2 0 -1 1 0 1 -1 0 -2 -1 -3 -2 -3 -1 -2

0 -1 0 -2 -1 -3 -2 -4 -3 -4 -2 -3 -1 -2 -1 -3 -2 -4 -3 -4

-2 -3 -1 -2 0 -1 0 -2 -1 -3 -2 -3 -1 -2 0 -1 1 0 1 -1

0 -2 -1 -2 0 -1 1 0 1 -1 0 -2 -1 -3 -2 -3 -1 -2 0 -1

0 -2 -1 -3 -2 -4 -3 -4 -2 -3 -1 -2 -1 -3 -2 -4 -3 -5 -4 -5

Aside from the initial term of 1, all the terms are non-positive. And successive minima are reached at the 2nd, 4th, 16th, 28th, and 98th terms. What are these numbers, and how do they continue? As a matter of fact, they continue as follows:

2, 4, 16, 28, 98, 168, 576, 984, 3362, 5740, ...

You will notice that the differences between elements occur twice:

2 12 12 70 70 408 408 2378 2378 ....

This is another curious effect, whose explanation will be slightly postponed, until we have built up a repertoire of "vertical" concepts, in terms of which an explanation becomes very natural.

"Vertical" concepts

Vertical concepts are those related to the results of repeated differentiation of an eta-sequence. The first vertical concepts we describe are the "Vertical Coun and Sep-sequences" (VC and VS sequences). The nth term of VC is the coun of the nth derivative of eta(alpha). The analogous definition goes for VS. The square root of two provides us with an easy (but needless to say, atypical) example:

	VC	VS
n=0	1	2
n=1	1	2
n=2	1	2
n=3	1	2

etc.

There is no reason that we should expect such simple behavior in the VC and VS sequences belonging to randomly chosen values of alpha. In fact, the VC and VS of pi exude an utterly different aroma:

	VC	VS
n=0	3	4
n=2	6	7
n=3		
n=4		
n=5		

"and so on" -- if you can find any rhyme or reason to the sequences above! On the other hand, e has very regular VC and VS sequences:

VC	VS
3	2
3	2
1	2
4	3
1	2
6	5
1	2
8	7

Look at the VS-sequence. What would you say is its pattern? If you are slightly naive in number theory, you might guess, optimistically, that the pattern of the VS-sequence is: primes alternating with 2's. However, that would be too spectacular. Nature never hands you the primes on a platter. The actual pattern is more humble, but the very fact of a pattern being there at all is remarkable, when you think of pi! It consists of the successive odd numbers (but with "2" replacing "1"), alternating with 2's. And the VC-sequence, after a shaky start, consists of an alternation between the even numbers and 1's.

If you have ever seen the simple continued fraction for e, you may have noticed a similarity. Here it is:

$$e = 2 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{4 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{6 + \ddots}}}}}}}}$$

The denominators go: 2, 1, 4, 1, 1, 6, 1, 1, 8, 1, 1, 10, ... The similarity is so striking that one wonders if these aren't two representations of one thing.

Eta-sequences and Continued Fractions

There is a definite relation between eta-sequences and continued fractions. To show it, let us go back to the definition of the derivative. Recall that I spoke of two "styles" of assigning names to chunks, the old style and the new. In the old style, a chunk like "211" would get named "2" because of the two 1's, but in the new style it gets named "3" because that is its length. Let us consider what happens if we take derivatives in the new style. Every term of the derivative is increased by 1; hence the derivative is no longer the eta-sequence belonging to alpha', but to 1+alpha'. We could make matters confusing, by calling 1+alpha' the "new-style" alpha'...but instead of that we will write "D(alpha)" for it:

$$D(\text{alpha}) = \frac{s - \text{alpha}}{\text{alpha} - c} + 1$$

c and s being the coun and sep of alpha, as before. Let us put

$$x = s - c$$

Naturally, x is a function of alpha, and equals plus or minus one. Then we can write

$$D(\text{alpha}) = \frac{x}{\text{alpha} - c}$$

Inverting this equation, we get

$$\text{alpha} = c(\text{alpha}) + \frac{x(\text{alpha})}{D(\text{alpha})}$$

Here, alpha is expressed as an integer plus (or minus) a fraction with numerator 1. The trick is to express D(alpha) likewise:

$$D(\text{alpha}) = c(D(\text{alpha})) + \frac{x(D(\text{alpha}))}{D(D(\text{alpha}))}$$

This trick can be repeated for D(D(alpha)), then D(D(D(alpha))), etc.. Each time the trick is done, it corresponds to one more level of differentiation of the eta-sequence. So we are looking at the vertical structure of an eta-sequence this way. Now the whole thing can be summed up in one grand continued fraction; but before we write that down let us make some changes in convention. I just introduced the VC and VS sequences and so, presumably, you are not so used to them that you will vigorously protest if I change my definition of them... All I propose is that everything should be as before, except that all derivatives should be taken according to the new style. That has only one effect: it raises each term of VC and VS (except the zeroth) by one. Therefore, our revised VC and VS for the square root of two go:

	VC	VS	VX
n=0	1	2	+1
n=1	2	3	+1
n=2	2	3	+1
n=3	2	3	+1
	etc.		

I have put in an extra column for the X-sequence. The nth element of VC (or VS) is now equal to the coun (or sep) of D(D(...D(alpha)...)), where the number of D's is n.

Now we can write down the continued fraction for alpha, using the redefined VC-sequence, and the VX-sequence:

$$\text{alpha} = VC(0) + \cfrac{VX(0)}{VC(1) + \cfrac{VX(1)}{VC(2) + \cdots}}$$

Here, "VC(n)" stands for the nth element of the VC belonging to alpha, of course.

For instance,

$$\text{sqrt } 2 = 1 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \cdots}}}$$

← "e" instead

Now I must hasten to mention that this kind of continued fraction is slightly different from the most commonly exhibited kind, the difference being that the usual ones only have +1's in their numerators, never -1's (which will happen to our fractions whenever VX(n) = -1). The usual type of continued fraction is called "simple"; our fractions also have a name: "nearest-integer continued fractions" (this is the best translation I can give for the German phrase "Kettenbrueche nach naechsten Ganzen", which is what Oskar Perron calls them in his classic work on continued fractions, "Die Lehre von den Kettenbruechen", which is, unfortunately, out of print).

Many of the theorems which hold for simple continued fractions carry over to nearest-integer continued fractions. For instance, a famous theorem on simple continued fractions asserts that the sequence of denominators in the continued fraction for a real number will be periodic if and only if that number is a quadratic irrationality. (Here, the word "periodic" means that after a while, the sequence repeats over and over again; but the block which is repeated need not start with the very first term.) A slight modification of this statement holds for VC and VX sequences, namely:

> If the VC and VX sequences belonging to alpha are both periodic (in the above sense), then alpha is a quadratic irrationality; and conversely, if alpha is a quadratic irrationality, then its VC and VX sequences are necessarily periodic.

This theorem can be equivalently restated, using "VS" in place of "VX". A corollary is this theorem:

> Alpha is a quadratic irrationality if and only if there are unequal integers m and n such that the mth and nth derivatives of eta (alpha; 0) are identical.

The function INT

There is no criterion by which one can distinguish a VC-sequence from a VS-sequence. For instance, the VC-sequence of pi could very easily be the VS-sequence of some other number (in fact it is the VS-sequence of many other numbers!). Because of the indistinguishability of VC and VS-sequences, one is naturally led to ask, what if I interchange the VC and VS-sequences of alpha -- what number has for its VC-sequence the VS-sequence of alpha, and for its VS-sequence the VC-sequence of alpha? What number beta has the following VC and VS?

	VC	VS
n=0	2	1
n=1	3	2
n=2	3	2
n=3	3	2
n=4	3	2
.	.	.
.	.	.
.	.	.

This number, whatever it is, will be called "INT(sqrt 2)", because its VC and VS are those of sqrt 2, interchanged. Generally, for beta to be equal to INT(alpha) means that for all non-negative n,

$$VC_n(\text{beta}) = VS_n(\text{alpha}); \quad VS_n(\text{beta}) = VC_n(\text{alpha}).$$

Obviously, if beta = INT(alpha), then alpha = INT(beta), as well. It so happens that

INT(sqrt 2) = phi,

where "phi" stands for the "golden ratio",

phi = (1 + sqrt 5)/2.

(The proof of this comes from the fact that D(phi) = 1 + phi.) One can easily see that INT of any quadratic must be another quadratic, simply because the new VC and VS sequences are periodic. You may have noticed that INT(alpha) is not yet well-defined for rational values of alpha. This is because the VS-sequence for any rational number hits a snag after a finite number of steps -- in fact, on the mth step, where m is defined by:

```
D(D(...D(alpha)...)) = an integer
\         /
 \       /
   m D's
```

Notice that such an m is guaranteed to exist, for only irrationals have infinitely differentiable eta-sequences. On the mth step, then, VC is well-defined (being the integer itself), but VS is not, since the second-closest integer to an integer is not well-defined: there are two vying candidates. The most esthetically pleasing solution, perhaps, is to define INT at a rational value as being the arithmetic average of two values, one calculated by taking the mth VS to be one less than the mth VC, the other by taking the mth VS to be one greater than the mth VC. Under this definition of INT at rationals, it is easy to see that INT of a rational is also a rational. So we may now say:

> If alpha is algebraic of degree 0 (an integer), then INT(alpha) is likewise algebraic of degree 0.
>
> If alpha is algebraic of degree 1 (a rational number), then INT(alpha) is likewise algebraic of degree 1.
>
> If alpha is algebraic of degree 2 (a quadratic), then INT(alpha) is likewise algebraic of degree 2.

The pattern is suggestive, is it not? However, I am not sure if the extension of this pattern is valid or not. A most interesting question. Incidentally, if it were valid, then one would have as a corollary the following statement:

> If alpha is transcendental (not algebraic of any degree), then INT(alpha) is likewise transcendental.

Certainly it is provocative to ask what mathematical significance a constant such as INT(e) or INT(pi) has. I have not been able to find any for either. Their values are, roughly:

INT(e) = 2..........;   INT(pi) = 3.86...........

Now before going any further in the description of INT, it is vital to exhibit a plot of it. All the information about INT is contained in a plot where alpha runs only from 0 to 1. To get the value of INT for any other value of alpha, subtract the integer part of alpha, consult the graph between zero and one, and then add back the integer part:

INT(alpha) = INT (alpha-N) + N

This means that the graph of INT consists of infinitely many copies of the contents of a single "box", touching each other at their corners, as shown on the next page. What happens inside each of the boxes is then shown on the page after that.

The most striking fact, at first glance anyway, is how the graph inside each box -- henceforth called a "box-graph" -- consists of scads of little "subgraphs". All the subgraphs seem to resemble each other, and are aligned more or less parallel to each other, the only difference being that as they recede into the corners of the box, they get smaller, and smaller, and smaller... But the next level of observation brings a yet greater surprise: all the little subgraphs themselves seem also to be composed of subgraphs of their own. And, to the extent that the graph allows you to peer into the level beyond that, the same thing seems to be happening. It never comes to an end. It may seem mind-boggling at first, but after it has settled, it makes a little sense, because it is reminiscent of the way that an eta-sequence yields another eta-sequence, on being differentiated, which yields another one, and so on. Of course, infinite differentiability requires alpha to be irrational -- but that brings us back to INT, which seemed to be more natural to define at irrationals, anyway.

And then the idea of INT hits you: the little subgraphs are "copies" of the box-graph. That poses a question, however: "How can the subgraphs be copies of the box-graph when the box-graph is straight, and the subgraphs are all curved?" The answer is that they are copies in an extended sense of the word -- they are not only of a different size than the original, but they are also a little distorted. The distortion is not chaotic or random, though, but quite neat and systematic.

If you look back at all the boxes touching each other along a diagonal, you will see that the large-scale structure of INT is just like the structure in each box: many repeated parallel copies of one item. The total graph contains identical copies, and is therefore of infinite extension; a box-graph, on the other hand, involves the squeezing of an infinite number of copies into a finite space, and therefore causes shrinking to occur, near the corners. You can probably imagine a giant with infinite reach picking up the whole INT-graph, rotating it 90 degrees, and then compressing it so that it will fit inside a 1x1 box (and in the process, slightly distorting the pieces composing it). Such a squeezing-process, if done right, would transform the total graph of INT into one single box-graph! A proof of this would establish all the earlier speculations about the nested structure of each individual box-graph. We will prove that such a giant exists, in the following sense: we will provide a monotone mapping which compresses boxes 2 through infinity down into the upper left half of a box-graph. (The lower right can be taken care of by symmetry.)

Proof of the Nesting-Property of INT

Earlier, I showed a way that INT(alpha) can be calculated from just one box-graph: shift alpha into the relevant region, use the box-graph, then un-shift the result. This has the general form:

INT(alpha) = g(INT(f(alpha)))

where f is a function that shifts any alpha into the relevant region of the x-axis, and g is a function that shifts values of INT obtained from the box-graph back into the correct part of the y-axis. When we did this before, f(X) was X-N and g(y) was y+N; they were inverse functions. The resulting equation told us that one box-graph looks exactly the same as any other box-graph, because the functions f and g are simple translations: they do not shrink or expand, they merely shift. To prove our nesting-property, we will need an equation of the above form, but where f is a "shrinking-function", one which carries values of alpha between 2 and infinity into values lying between 0 and 1/2; and where g is an expanding-function, which carries the small interval [0, 1/2] back onto the half-line from 2 to infinity.

Fortunately, such an equation is not hard to come by; in fact it practically falls in our laps, once we know we are looking for it! It all comes from looking at the VC and VS-sequences of alpha and INT(alpha). Let us exhibit them. First, the table for alpha:

	VC-seq		VS-seq
n=0	A(0)	alpha	B(0)
n=1	A(1)	D(alpha)	B(1)
n=2	A(2)	D(D(alpha))	B(2)
.	.	.	.
.	.	.	.
.	.	.	.

The reason I have included the numbers alpha, D(alpha),... in the middle is the following. The VC and VS-sequences which begin at any given level and continue downwards are VC and VS-sequences in their own right, which belong to the number written at the corresponding level in the middle. For instance, the VC and VS-sequences belonging to the 17th derivative of alpha are, respectively, A(17), A(18), A(19),... and B(17), B(18), B(19),...

Now INT(alpha) -- let us call it "beta" -- has the following analogous table:

	VC-seq		VS-seq
n=0	B(0)	beta	A(0)
n=1	B(1)	D(beta)	A(1)
n=2	B(2)	D(D(beta))	A(2)
.	.	.	.
.	.	.	.
.	.	.	.

Now suppose we chop off the top level in the table for alpha. The remaining sequences are A(1), A(2),... and B(1), B(2),..., which are the VC and VS-sequences of D(alpha). If we now interchange, we have the VC and VS-sequences of INT(D(alpha)) -- by definition of INT. But they are just the same as the VC and VS-sequences of D(beta), as you can see by looking at the table for beta. Therefore, in calculating D(INT(alpha)), we can go either of two routes: first get INT(alpha), then take D of it, or else take D first, and then get INT of that. Symbolically,

INT(D(alpha)) = D(INT(alpha)).

This is close to, but not exactly, what we want. Let us write gamma = D(alpha). Then what is alpha in terms of gamma? From the discussion on continued fractions, we have

$$\text{alpha} = c + \frac{s - c}{\text{gamma}}$$

Now, if gamma alone is given, then alpha still has not been determined; we have the freedom to choose the coun and sep of alpha as we like. Let us denote the value of this expression, with c=a and s=b, by

$$D^{-1}_{a,b}(\text{gamma})$$

(Of course, a and b must differ by one.) If we substitute this into the earlier equation, we get

$$\text{INT(gamma)} = \text{D(INT}(D^{-1}_{a,b}(\text{gamma}))$$

Lo and behold, we have some candidates for the shrinking and expanding functions! For notational simplicity, let us use "X" for the argument of the shrinking-function D-inverse, and "x" for its value; likewise, "y" for the argument of the expanding-function D, and "Y" for its value. Let us take a=0, and b=1 in the shrinking-function:

$$x = D^{-1}_{0,1}(X) = 1/X$$

Now if we let X vary from 2 to infinity, the shrinking-function's range will be the interval [0, 1/2]. Suppose we see what happens when X varies inside box number 2. The values of X are shifted into the range [1/2, 1/3] by the shrinking-function; then INT provides y's between 1/2 and 2/3; finally, these y's are transformed by the outer D-function into Y's between 2 and 3, fitting, as expected, into the original box. Thus we see very directly how a particular box is represented by a particular subgraph. A similar argument holds for any other box to the right of the one just considered. The box between N and N+1 is mapped onto a subgraph located between 1/N and 1/N+1. As promised, this shows how the subgraphs of box 1 (or any other box) are "copies" of the box-graphs from 2 out to infinity. Since each box-graph is symmetric with respect to both of its diagonals, the proof for the lower right half is implicit in what we have done.

An additional fact about the little copies, which is furnished to us by the shifting-equation, is how a box-graph must be compressed horizontally and vertically in order to be brought to coincide with a given subgraph. Consider once again the mapping between box 2 and the subgraph between 1/2 and 1/3. The shrinking-function f(X) is

$$x = 1/X$$

and the expanding-function g(y) is

$$Y = 1/(1-y).$$

Remember that f maps from the box to the subgraph, while g maps from the subgraph back to the box. Consider two nearby points on the X-axis, and what f does to them. If they are X and X+dX (with dX infinitesimal) then f carries them into f(X) and f(X+dX) respectively; and the latter, by Taylor's theorem, equals f(X) + dX f'(X). The separation between the image-points, therefore, is multiplied by the factor f'(X). This is called the "local compression factor" (and notice it is written as a function of the variable Big-X -- the variable belonging to the box-graph, not little-x, of the subgraph). The local expansion-factor due to g is g'(y) -- but we are interested in compression, not expansion, so we must take the reciprocal. Secondly, we want to write it in terms of Big Y, the box-graph variable, not little y. This gives us:

Local horizontal compression factor =

$$f'(X) = -1/X^2$$

Local vertical compression factor =

$$1/g'(y) = (1-y)^2 = 1/Y^2$$

Notice that these compression factors both vary between 1/4 and 1/9 when X and Y range over box 2; but when X and Y vary over box 20, say, then (a) there is much greater compression, and (b) it is much closer to uniform, since both factors remain almost constant, varying between 1/400 and 1/441. Thus, we obtain an answer to how the curvature of the subgraphs comes about, and secondly, we learn why the subgraphs closest to the corners of any box are so much less curved than the ones in the center, so much more faithful as "copies" of the box-graphs themselves.