Just an idea

The following contains a mix of some very formal mathematics-mainly encountered in theoretical computing- with physics. Further, a lot of statements can be ignored if one is not altogether interested in arguments involving the level of formalism associated with the cardinal numbers. Roger Penrose says that Cantor's abstraction of infinity and the foundations of data/information is historically as interesting as the invention of complex numbers, and (unlike complex numbers) is just waiting for a serious application in physics. Perhaps someday. For now, we split our discussion into two main bodies: if a series of paragraphs begin with a bold **(F:)**, it indicates that there are some elements of these *formal* notions therein. Other arguments will be preceded by a bold **(E:)** and will be *easier* on the formalism- even neglectful.

1. Intro

(E:)

About 3 years ago I began to explore the properties of an imaginary particle which moves through space according to a trajectory inspired by Cantor's bijection from $\mathbf{R}^n \to \mathbf{R}$. The complete lack of continuity apparent in the trajectory of such a particle leads to many interesting phenomena- not the least of which is that by wich the one-dimensional trajectory can cover all of an n-dimensional space in *finite time*. Perhaps later I shall write about these, but for now, one phenomena will be expanded below (briefly.)

If we take $\mathbf{r}(t) \in \mathbf{R}^2$ to be the bijection, with $t \in (0,1)$ and $\mathbf{r}$ the trajectory of the particle, then

$$\boldsymbol{r}(0.m_1 m_2 m_3 m_4......) = (0.m_1 m_3 m_5...., 0.m_2 m_4 m_6......)$$

and its inverse:

$$(\boldsymbol{r})^{-1}(0.a_1 a_2 a_3...., 0.b_1 b_2 b_3......) = 0.a_1 b_1 a_2 b_2......$$

where the m_i's, a_i's and b_i's are digits in the usual weighted numeric representation of numbers between 0 and 1 in some base. **(F:)** [Actually, as far as I know, this bijection is only referred to by Cantor in terms of the cardinal numbers in his original paper, where he wrote: $\aleph_1^n = (2^{\aleph_0})^n = 2^{n\aleph_0} = 2^{\aleph_0} = \aleph_1$. But he most certainly studied this function in this form, and referred to this and other similar ones as *nowhere-differentiable* functions in a study that later led to the invention of the Peano curve.] **(E:)**We consider 2-dimensions because it is very easily generalised to $\mathbf{R}^n$. A particle moving according to $\boldsymbol{r}(t)$ covers the entire plane $(0,1) \times (0,1)$ in 1 unit of time.

The phenomena I wanted to address is as follows. We already know that $\mathbf{r}$ is not differentiable w.r.t. time, but what is the nature of this non-differentiability? Let us define the "velocity" of this particle at time τ as:

$$\lim_{n \to \infty} \frac{\boldsymbol{r}(\tau + 2^{-n}) - \boldsymbol{r}(\tau)}{2^{-n}}$$

If we take the definition of $\mathbf{r}$ above to be in base 2, and take some odd integer $N > 0$, then $\boldsymbol{r}(\tau + 2^{-N}) - \boldsymbol{r}(\tau) = (2^{(-N-1)/2}, 0)$ and $\boldsymbol{r}(\tau + 2^{-(N+1)}) - \boldsymbol{r}(\tau) = (0, 2^{-N/2})$ as can be gleaned from the definiton. This means that the expression argued by the limit here above *alternates* in direction as n goes to infinty and so the limit has no meaning.

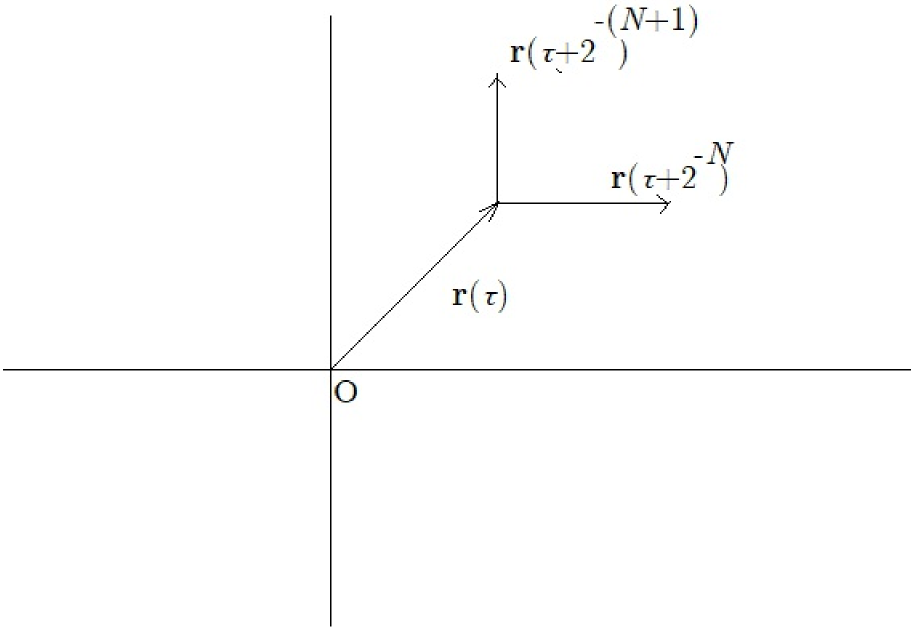

Figure 1.

Furthermore, I suspect that a similar ting happens when we try to define

$$\lim_{n\to\infty} \sum_{i=0}^{2^n\tau} \boldsymbol{v}(2^{-n}i)2^{-n}$$

where $\mathbf{v}(t)$ has the same definition that $\mathbf{r}$ had above, but here it represents the *velocity* of the particle. A new formalism for complementarity, perhaps?

Anyway- it is just a suggestion.

The notion at the base of my original studies, I later discovered, has many precedents. You can skip to the next section, if you are already familiar with these.

i) Feynman and Hibbs spoke about the philosopy of the path-integral approach in their book "Quantum mechanics and path integrals" ; they spoke about successive measurements of an electron's position as forming a Markov chain, and suggested that one can define a *unique* path-history for the eletron more accurately, if you take a path of fractal dimension.

ii) In Landau and Lifshitz's "Quantum Mechanics" (Vol.3), they write about successive measurements of an electron's position, somewhat inaccurately:

> If now, leaving the accuracy of the measurements unchanged, we diminish the intervals Δt between measurements, then adjacent measurements, of course, give neighbouring values of the co-ordinates. However, the results of a series of successive measurements, though they lie in a small region of space, will be distributed in this region in a wholly irregular manner, lying on no smooth curve. In particular, as Δt tends to zero, the results of adjacent

The inaccuracy lies in the fact that, for any fixed Δt, the set of positions resulting from the measurements is discrete, and so can be fitted on to a smooth curve. The point Landau and Lifshitz were trying to make is that the maximum curvature of the resultant curve goes to infinity as Δt goes to zero.

iii) Wheeler studied the notion that some radically new ideas in geometry were needed to correctly describe the phenomena that Landau and Lifshitz are referring to. He coined the term "pregeometry" to refer to the study of a different geometric foundational framework to facilitate this.

iv) Hawking (1995):

"It was John Wheeler who first pointed out that quantum fluctuations in the metric should be of order one at the Planck length. This would give spacetime a foam-like structure that looked smooth on scales large compared to the Planck length. One might expect this spacetime foam to have a very complicated structure, with an involved topology. Indeed, whether spacetime has a manifold structure on these scales is open to question. It might be a fractal. But manifolds are what we know how to deal with, whereas we have no idea how to formulate physical laws on a fractal."

Physicists have generally considered mysterious functions of this type- non-Riemann-integrable functions, (Dirchlet's rational function), Cantor's map, fractals of all shapes etc. One can employ the language of "denseness" of the range of these functions to be rigorous, but this may not be necessary. We have chosen a very different starting point.

2. The Problem

The real problem is manifold, but we will tackle a small one. "Conventional" analysis of dynamical systems in foundational mechanics have their basic facts predicated on the proposition that we will be dealing with continuous/connected spaces, or that the main problems can be reduced to problems concerning these.

We here present a rule of thumb to define the derivative of a function belonging to a much larger lass of functions than $\mathrm{C}^1(\mathbf{R}^n)$, and so we naturally extend the notion of differentiation to a much larger class of functions thereby. This rule of thumb gives us instantly the derivatives of a functions from a certain subclass of this very large class (which is still nevertheless larger than C^1) whereas for certain other functions, there are some difficulties (and our notion of differentiation is not applied to concrete examples of that type here.) We will apply this rule of thumb to some functions- including Cantor's map (as we shall refer to it from here.) It is really quite simple once finished- so many of these functions, as it turns out, can be expressed as infinite sums of *piecewise continuous*, *periodic* functions, which give up continuity at their limit.

First of all, just to get this out of the way, I imagine that the historical route via which integration was extended beyond Riemann-integrable functions, by using Lebesgue integration, *already solves the problem* of differentiating many non-differentiable functions as below, but I don't know what this technique is called, what mathematical circumstances it is generally associated with, and I haven't read it anywhere either.

We may start with the assumption that, if a (truly) non-holomorphic function $f\colon \mathbf{C}\to\mathbf{C}$ can be "differentiated" once, it is "analytic"- so that Cauchy's formula "holds"- because the ability to compute a derivative makes it "holomorphic". [Of course, the function is not holomorphic/analytic, but let's just go on.]

Then,

$$f'(\alpha) = \frac{2!}{2\pi i}\oint_{S\leftarrow\ni\alpha}\frac{f(z)}{(z-\alpha)^2}\,dz$$

(since f is "analytic".) This way we can compute the integral using Lebesgue integration and thereby obtain a unique $f'(\alpha)$.

But this entire line of reasoning is too convoluted for my liking- we have to assume that a non-analytic function is analytic for the derivative to be calculable- and it is just a little messy. But still it seems interesting.

Our solution has nothing to do with this line of reasoning (as far as I can see.)

3. Our Solution

We will construct a rule of thumb for differentiating a function $f\colon \mathbf{R}\to\mathbf{R}$ that is not differentiable by the usual definition.

To what class of functions are you extending the notion of differentiability exactly?

Let me start backward. The largest possible class is the set of all functions f: $\mathbf{R}\to\mathbf{R}$, usually denoted $\mathbf{R}^{\boldsymbol{R}}$. We will now restrict ourselves to a certian subclass of this and explain why. We will (first) restrict ourselves to the set of all functions that have a clear-cut, finitely expressible definition, i.e. a rule of thumb, or formula, that unambiguously describes how exactly $f(x)$ is manufactured from x (more details below.)

The reasons: (Skip if you want.)

1) Normally in physics, when we are given a function without a closed-form expression or some formula of some type to work with, like a potential $V(x)$, we just write $V'(x)$ and assume that in the concrete case, either: i) $V'(x)$ has a derivative we can compute by looking up a few derivatives and applying the usual rules, or ii) Information about $V(x)$ exists as a sequence $\{V(a_0+n\delta)\}_{n=1}^{N}$ and we have to substitue $\frac{[V(a_0+n\delta)-V(a_0+(n-1)\delta)]}{\delta}$ werever $V'(x)$ occurs, for some n; or iii)We can use perturbation theory, or iv) We can use some declarative statements (some assumptions) about the *form* of V- like analyticity, or the statement that V satisfies some differential equation,- that restricts the *form* of V', and we can substitue that expression- a function of x and some unknowns- for $V'(x)$.

2) In all of the above cases, with the exception of ii), we are differentiating at least one function which has a representative formula, which means that we are looking up[1] the derivative of at least one function in these cases. In the case of ii), we are not really carrying out differentiation. The sequence described can be fitted onto any smooth curve, (as long as it doesn't include $\pm\infty$), and we can indefinitely increase the error in the computation of V' according to the given formula by increasing the curvature of the function that this sequence was derived from (in a process that communication engineers refer to as *sampling*.)

3) The functions that interest us all fall in this class- the Cantor map, the Dirchlet function etc.

How, then, do we explicitly represent this class?

The answer to that question lies in a generalisation of the Church-Turing thesis.

(F:) If we summarise the Church-Turing thesis by saying that all of mathematics consists in operations on symbols that can be simulated by a Turing machine, then we may be inspired towards the appropriate generalisation[2]. Such a train of thought appears to lead us to this: the kernel of the notion of "real computing". I have not reviewed a lot of the source material on that subject,but I imagine that this is the needed generalisation-as below.

But, first, we introduce a model of computation that is equivalent to the Turing one, in which the Turing machine is replaced by a function on real numbers. I introduced this idea in a conference on computing last year, but had not found any significant (non-recreational) application till now[3]. Let the total number of states of the machine be N. In this model, the 'state' of the machine is encoded in base 2 and the result is stored in the first $[\log_2 N]+1$ digits after the radix point (where $[y]$ is the *integer part* of y.) The following digits represent the contents of the cells of the tape (the tape configuration)- since our *tape alphabet* is just $\{0,1\}$. The integer part of x contains the *cell-number*, or the *address* of the cell currently under the head.

To summarise:

$$x=\underline{101011010}\ .\ \underline{0101001110}\ 0101010110100111.....\ \text{in base 2}$$

Here, the first and second underlined sequence of 0's and 1's, when taken *out* of x, represent the *address* of the cell currently under the head, and the current *state* of the machine respectively.

Now consider the expression:

$$[2^n x]-2^{n-m}[2^m x]$$

1. The phrase "looking up" is formal and is used in computer science; in practice, we have commited them to memory, and we are "looking up" a table in our memory

2. There is a somewhat significant deal of thought that has historically gone into this.

3. Don't bother checking out that paper- it was only a rough theoretical notion; if you do, kindly ignore the last theorem

where we again use the [ ] notation for the integer-part of its argument[4]. This expression represents the integer formed by taking te digits of x from te m^{th} to te n^{th} place[5] (both inclusive).

Thus the function:

$$S(x) = [2^{[\log_2 N]+1}x] - 2^{[\log_2 N]+1}[x]$$

represents the "state" of the Turing machine, and the function:

$$[x]$$

is the address/cell-number of the cell currently under the head (i.e. the one currently being scanned.) The expression:

$$\delta_{S(x)I}([2^{[x]}x] - 2[2^{[x]-1}x])$$

is 1 iff $S(x)$ is I and the $[x]^{\text{th}}$ digit of x is 1; it is 0 otherwise. Similarly,

$$\delta_{S(x)I}(1 - [2^{[x]}x] + 2[2^{[x]-1}x])$$

is 1 iff $S(x)$ is I and the $[x]^{\text{th}}$digit of x is 0; it is 0 otherwise. Thus, the operation:

$$x + p\delta_{S(x)I}([2^{[x]}x] - 2[2^{[x]-1}x]) + q\delta_{S(x)I}(1 - [2^{[x]}x] + 2[2^{[x]-1}x])$$

gives us $x + p$ if the $[x]^{\text{th}}$ digit is 1, $x + q$ if the $[x]^{\text{th}}$ digit is 0- iff $S(x) = I$. Thus, by suitably choosing a p, q for various I's, one can modify any part/digit in x with the help of the above operation- thus enabling us to modify the state/cell number and the corresponding contents.

Take the function:

$$f(x) = x + \sum_{I=0}^{N} p_I\delta_{S(x)I}([2^{[x]}x] - 2[2^{[x]-1}x]) + q_I\delta_{S(x)I}(1 - [2^{[x]}x] + 2[2^{[x]-1}x])$$

From the above arguments, it can be seen that the run of a Turing machine is equivaent to the repeated application (compsition) of f; that is, if the transition function (δ) is evaluated n times, the values of the state, cell-number and tape configuration are contained in:

$$f^{(n)}(x) = (f \circ f \circ f \circ \ldots. \circ f)(x), \text{ where } f \text{ is applied to } x\ n \text{ times}$$

and the result - i.e. the *output* of the algorithm/the complete run of the Turing machine is equal to $f^{(\infty)}(x)$.

So this is equivalent to the traditional Turing model. Note:

1) For "traditional" functions for which the Turing machine's run terminates in finite time, there exists some μ such that $f(f^{(\mu)}(x)) = f^{(\mu)}(x) \Rightarrow f^{(n)}(x) = f^{(\mu)}(x)\ \forall n \geq \mu$, i.e. the application of f becomes redundant after some number of iterations (Knuth.)

2) For these functions, the initial configuration of the tape can either be such that:

i) There exists a finite number of 1's, or

ii) there exists some Turing machine (or algorithm) to generate the n^{th} digit of x, given n.

4. The floor function was originally considered, but was ruled out after the considerations in section 3, later

5. Where we use a slightly different "place-value" system:

…-5 -4 -3 -2 -1 0   1 2 3 4 5 …

… 1 0 1 1 1 0 . 1 0 0 1 1 …

Real numbers which are of type ii) constitute what Turing refers to as "computable numbers", when he proved that the set of all of them is only $\aleph_0$- cardinal.

The Church-Turing thesis contends that all mathematical operations boil down to things that can be simulated by the operation of a Turing machine. Turing further believed that this has to do with the inner working of the mind. As it happens, this hypothesis is also applicable to algebraic manipulation[6] and algorithms that take a function g:$\mathbf{R}\rightarrow\mathbf{R}$, and determine wheter it is continuous or not, and therefore, is generally referred to these.

However, we know how to compute x^2 for some $x\in(0,1)$ if we were given its weighted numeric representation ($x=0.9521....$), even if the digits of that representation contain infinitely many nonzero digits, and do not follow a pattern (thus transcending the descriptions in 2)i) and 2)ii).) That is because we have a *defintion* of this function that allows us to imagine computation of this funcion. Turing's contention was not that we cannot define functions on non-computable numbers, it was simply that there is no way of *specifying* and/or comprehending a non-computatble number; but that does not disallow us from defining the computation of functions on them- theoretically.

Thus we have one feature of the necessary generalisation of the Turing model: the initial tape configuration need not be a computable number. This, coupled with 2)i) means that there is no way to compute the expression:

$$[2^n x]-2[2^{n-1}x]$$

given n. So the similar expression after the summation sign, which is the most general representation of the $[x^{\text{th}}]$ of x cannot be replaced by some simpler function of $[x]$ for non-computable numbers- and is thus the only representation of this gives the correct *for all*$\in$ $\mathbf{R}$.

The second feature of the generalisation comes from 1) coupled with 2)i): there need not exist an integer μ such that $f^{(n)}(x)=f^{(\mu)}(x)\,\forall n\geq\mu$, meaning that the computation may go on endlessly. In fact, this is the kind of computation that leads us to more and more decimal places of π, e etc. Knuth states that is (even in the traditional setup) is distinguished from those that terminate by calling the non-terminating run a "computational method" and the terminating one, an "algorithm". Our generalisation means that the best representation of the final answer is $f^{(\infty)}(x)$ as written earlier.

And therefore, it may be noted that, in generalising the Turing model, nothing has changed from our original definition, but some assumptions about the concrete cases (inspired by the mathematical paradigm of *finitism*) are violated.

(E:) To summarise, all "well-defined" functions g: $\mathbf{R}\rightarrow\mathbf{R}$ can thus be expressed in terms of a function:

$$f(x)=x+\sum_{I=0}^{N} p_I\delta_{S(x)I}([2^{[x]}x]-2[2^{[x]-1}x])+q_I\delta_{S(x)I}(1-[2^{[x]}x]+2[2^{[x]-1}x])$$

for some N, p_I's and q_I's, and where $g(x)=f^{(\infty)}(x)$.

4. Subclass

And so we have a large subclass of $\mathbf{R}^{\boldsymbol{R}}$, which is promised to cover all "interesting" cases by the Church-Turing thesis. But the form of $f^{(\infty)}(x)$ was intimidating enough for me to assume I will not be able to consider differentiaiting it, even after the critical result we will introduce next.

I attempted to replace the model with something simpler.

(F:) The best simplifications came from considering, rather than repeated application of f, a sequence f_i and inventing a recurrence relation with $f_0=x$. Further, the functions $S(x)$ and $[x]$ were replaced by other sequences s_i and m_i, with $s_0=m_0=x$. Further, the idea that instead of the tape head moving, the entire tape may be moved after each iteration was also helpful. This translated to a multiplicative power-of-2 factor in the recurrence relation, and was an attractive idea.

6. The reason is that algebra is rooted in symbolic manipulation

With each model, however, the same complication appeared- and was associated with the fact that the cell-number itself varies, as in $[x]$ in our original representation. This complication disappears when we take a less general model, as with our analogue of a finite automaton. Thus, with the result below, the *real* generalisation of some finite automata will surely produce (traditionally) non-differentiable functions, which become "differentiable" in our scheme.

(E:)[The reader who is familiar with power series, Feynman diagrams etc. may feel that there is undue effort here, but it must be remembered that we have started at a different place and obtained our result. I have conscientiously avoided power series.]

The following representation of $[x]$ is the key notion that enables the differentiation of those "nowhere-differentiable" functions that Cantor played with. It may have already been guessed straightaway from the fact that the function that yields the n^{th} digit must be periodic in x. But we will derive this *apriori*.

We know that:

$$[x] = x - \{x\}$$

where $\{x\}$is the fractional part of x and is periodic in x, with period 1. By Fourier-expanding it,

$$\{x\} = \sum_{\nu=0}^{\infty} \alpha_\nu \cos 2\pi\nu x + \beta_\nu \sin 2\pi\nu x$$

we get $\alpha_\nu = 0$, $\beta_\nu = \frac{1}{\pi\nu}$.

$$\Rightarrow [x] = x - \{x\} = x - \sum_{\nu=0}^{\infty} \frac{1}{\pi\nu} \sin 2\pi\nu x$$

Therefore, the n^{th} digit of x after the radix point- in base 2,

$$[2^n x] - 2[2^{n-1}x] = 2^n x - \{2^n x\} - 2^n x + 2\{2^{n-1}x\}$$

$$= 2\{2^{n-1}x\} - \{2^n x\}$$

$$= 2\sum_{\nu=0}^{\infty} \frac{1}{\pi\nu} \sin 2^n \pi\nu x - \sum_{\nu=0}^{\infty} \frac{1}{\pi\nu} \sin 2^{n+1}\pi\nu x$$

$$= \sum_{\nu=0}^{\infty} 2\frac{1}{\pi\nu} \sin 2^n\pi\nu x - \frac{1}{\pi\nu} \sin 2^{n+1}\pi\nu x$$

$$= \sum_{\nu=0}^{\infty} a_\nu \sin 2^n \pi\nu x$$

where $a_\nu = \frac{1 + \nu \bmod 2}{\pi\nu}$.

It should now be apparent why I suspect it it may be easy to "differentiate" a function which generalises the finite automaton to **R**. But we will not expand on this. With this tool, another abstraction came to mind, that, while less explicitly Turing-like, is indeed powerful. It is clear that any function $f(x)$ in this subclass we have been discussing, can be written as a function that can *separately argue every digit of* x:

$$h\left(\left\{\sum_{\nu=0}^{\infty} a_\nu \sin 2^n \pi\nu x\right\}_{n=0}^{\infty}\right)$$

though we do not write about the Turing machine formalism for this. The advantage of this notation is i) that it frees up our mind to easily construct functions like Cantor's map, and yet appears completely differentiable, and ii) the earlier model must be further modified to get us the final tape configuration: $\{2^{[\log_2 N]+1}x\}$.

[Although I suspect that a suitable choice of linear operators on the Hilbert space of $\bigotimes_{m=0}^{\infty}\sum_{\nu=0}^{\infty}a_\nu \sin 2^n\pi\nu x$ will form an equivalent Turing machine model. I myself have been experimenting with these, and have found some operators that can fulfill some of the requirements of a Turing machine.]

5. Finally

Now how does this help with more concrete examples? The representations which have the states and memory address encoded in the digits of x must be further processed to eliminate these from our final answer, as indicated above. This is too messy, and here, again, the above notation, for h is nicer.

In order to differentiate a *bounded* function g on a bounded set $S \subset \mathbf{R}^n$, we write it as:

$$g(x_1, \ldots., x_n) = \max(g) \times h\left(\left\{\sum_{\nu=0}^{\infty} a_\nu \sin 2^n\pi\nu x\right\}_{n=0}^{\infty}\right)$$

and proceed as in the general plan apparent in the applications below:

I. Cantor's map

$$g(x_1, \ldots. x_n) = \sum_{m=1}^{\infty}\sum_{j=1}^{\infty} 2^{-j\times m}\sum_{\nu=0}^{\infty} a_\nu \sin 2^m\pi\nu x$$

Then,

$$\frac{\partial g}{\partial x_k} = \sum_{m=1}^{\infty} 2^{-k\times m}\sum_{\nu=0}^{\infty} 2^m\pi\nu a_\nu \cos 2\pi\nu x$$

II. x^2 -just because we mentioned it before

$$2x = \sum_{i=1}^{\infty} 2^{1-i}\sum_{\nu=0}^{\infty} a_\nu \sin 2^i\pi\nu x$$

$$\Rightarrow x^2 + c = \int 2x\,dx = \sum_{i=1}^{\infty} 2^{1-i}\sum_{\nu=0}^{\infty} \frac{-a_\nu}{2^i\pi\nu}\cos 2^i\pi\nu x$$

I have a further motive in trying out this function- I have also constructed a Turing machine for x^2 as below:

$$x^2 = \sum_{i=1}^{\infty} 2^{-i} \times \left(\sum_{\nu=0}^{\infty} a_\nu \sin 2^i\pi\nu x\right) \times \left(\sum_{\zeta=1}^{\infty} 2^{-\zeta}\sum_{\mu=1}^{\infty} a_\mu \sin 2^\zeta\pi\nu x\right)$$

Therefore, we may check how crazy our formalism is exactly by setting the first expression to the second, and checking if it forms an identity.