Jad Nohra, Private Note.
Revised: November 14, 2016

Analysis Treatments

Volume I

I Taylor and Non-analyticity

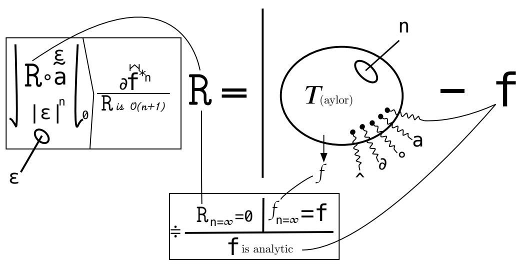

Figure 1: Power Expansion Machine

1. Figure 1 is incomplete result of the hand written notes [@c1]. It needs the following improvements:

1. One surprising source of confusion seems to be the fact that the mind intermingles the multiple stengthenings of the asymptotic properties of the remainder relative to x^n. These strengthenings are not properly treated in the figure.
2. In [@c2, p.8] we see how we have multiple notational versions, again none of which having the 'proof power' of standard notation (an attempt at a proof using them is at p.7) but now we notice the following:
   (a) It may be best to always keep the two extremes of a certain notation: first the 'unpacked version' which does not hide any of the relations between the individual components (this would be the complicated one at 2 o'clock in the page) and second the toolified, mostly packed version.
   (b) There is a tension between the two. The second is necessary for building larger and larger structures, the first is necessary for weak brain with weak memories, or for coming back to low detail when necessary.
   (c) It is pointless to use the packed version if 'properties of it' are not internalized.

2. One is to think of the highlighted figure at [@c2, p.3] as the 'torch' passed to us, an implicit relation pinpointing the discovery of 'd'. The hint guiding to this form are first and foremost instantaneous velocity, secants, tangents. These lead to the limit definition of 'd' which leads to this 'best linear approximation' (BLA) implicit property, equivalent to the definition. The definition is an arithmetic calculation of 'd' such that it is the BLA at 'a', but the torch to accept is 'the whole thing', the existence of 'd' the choice of modeling of BLA (***not as an optimization formulation, but as an asymptotic one***), the calculation of 'd' at a point, then on an interval.

3. Contemplating [@c2, p.3], we have half-explicit answers to the appearance of the division by $(x-a)$ in the definition of derivative.

4. Why do we have $|x-a|$ on the right side, can't we have a constant there? A constant would indeed work in the sense that we can always find how close we need to get to a for the 'BLA' to be as close to $f(x)$ as we wish (ϵ), this is done by simply replacing $|x-a|$ in the right side by the constant. The problem is that:

5. This is not fruitful, and actually $|x-a|$ is a stronger result.

6. We wish for our linear approximation to match $f(x)$ exactly at a, this dictates that given a from for it: $l(x) = d(x-a) + c$, it is necessary that $l(a) = 0$, that is, that $c = f(a)$.

7. Given the above, any d will work at $a = 0$, but we want more than that, and this is where we receive the torch of 'asymptotics'. Since we can easily find $l(x)$

such that it matches $f(x)$ at a, the BLA we seek must also have that property. But now we need to look at bit outside of a since that point alone does not give us any more information than that $c = f(a)$. If we asymptotically look at the error $e(x) = |f(x) - l(x)|$ we see it is useless to compare it to any constant error since we will obviously beat it by continuity. We need to compare it to something. The fact that we compare it to $|x - a|$ is half-common-sensical (HCS) and also provided by the secant hints. We need to take 'a small part of some curve' that extends outside of only a and bounce our 'approximations' of it.

8. There is another HCS answer, where $|x-a|$ shows itself, an arithmetic version of the secant hint, but one that exposes a meta-method of analysis in general, maybe its most essential one. Let us have a desired constant ϵ to which we want to approximate $f(x) - f(0)$ by $d.x$. We need to find how close we need x to be to 0, viz. find $\delta = |x - 0| =?(\epsilon)$, where $|[f(x) - f(0)] - [d.x]| < \epsilon$. Now observe the meta-method and the secant hint. We can easily make this hold irrespective of x by setting $d = [f(x) - f(0)]/x$ (x in the denominator appears). The meta-method is that this property might be also true at $x = a$ (instantaneous velocity), it is to view all the x-dependent values of such d as approximations to d itself, since they all satisfy the property, it is that we have a 'point d' which we know properties of but not a construction but that we can turn the properties into a construction if the properties 'point to a single point', if they converge (and the property still holds at the point where they do), it is that there are alternative ways of construction of points which pass by approximations when they are 'valid approximations', when they can be called so.

9. TODO: what happens if we require a 'best parabolic approximation' (BPA) instead? Does this nicely expand to BLA and the 'second order derivative'? This could be an alternative way to introduce derivatives. Note that BLA seems quite a misleading misnomer and the source of a lot of our misconceptions, we have to keep in mind that it is supposed to mean the highlighted figure at [@c2, p.3].

10. This meta-method is the same for defining the 'continuity points' of $f(x)$ at a. It appears in more clarity with non-analytic functions as seen at [@c3] where we do know that the multi-vec LBA must be independent of any specific direction v, but we can see that if we add tv where t is a scalar, we can by pinpoint what we need by going through a limit roundabout, through 'approximations'. Limit provides a way to enable hints and heuristics in this manner, allowing to turn properties to constructions in case (sometimes) checkable conditions of the construction reached by the limit hold. This works also for 'definitions', in the sense that these definitions are but (universally holding) calculations, constructions. In a way, 'limits' allow to break the incommensurability between the explicit (constructions by default non-limits) and the implicit (constructions by explicitization of implicit properties nearby using limits). Limits are (conditional) shortcuts from the nearby implicit to the explicit at the 'projection to the center of the nearby'.

11. Finally, it is strange that we never noticed but one can take in analysis the limit of anything, this shows the power of the above. The binary nature of 'converges/diverges' seems philosophically meaningful. For all things that 'converge' there is an 'approximation interpretation' nearby. In a way, this brings us back to 'continuity through nearness spaces' and the 'interpretation of topology using logic' that we never continued.

12. Having analyzed this, we force ourself to stop delving into the details of [@c2] and focus on a similarly essentializing analysis of the notation of vec-calc LBA to internalize it [@c3].

13. Note that in the Peano form of the remainder (the weakest form) (https://en.wikipedia.org/wiki/Taylor%27s_theorem) which requires only pointillistic differentiability as opposed to intervallic one (for Lagrange's form http://math.stackexchange.com/questions/1312577), we have an explicit function $h(x)$ such that $\lim_{x\to 0} h(x) = 0$. This function does not appear in our notes, or in their inspirational source [@c4]. The relation between them is that in [@c4] we show that the remainder is $o(|x - a|)^n$, which is also true for $h(x)(x - a)^n$. Is the former always expressible as the latter? We don't know, but one can prove the form of the latter (I cannot find the reference, but I think one simply extracts $h(x)$ to one side and proves it tends to zero, the find, postulate prove instead of derive annoyance). Actually, it is obvious that the two forms are equivalent. If we have some $g(x)$ such that $\lim_{x\to 0} g(x)/|x|^n = 0$, then we can always find a function $h(x)$ such that $h(x)x^n = g(x)$ simply by setting $h(x) = g(x)/x^n$. So in conclusion the more specific form of the Peano remainder version is just a canonical form of a function that is $o(x^n)$. Notice that in Gower's note below, he refers to Peano's version also without explicit mention of $h(x)$.

14. Gower says : "It is also the one result that I was dreading lecturing, at least with the Lagrange form of the remainder, because in the past I have always found that the proof is one that I have not been able to understand properly. I don't mean by that that I couldn't follow the arguments I read. What I mean is that I couldn't reproduce the proof without committing a couple of things to memory". The generalized MVT (and Rolle) are interesting, they are treated in more detail in [@c5] and they are good vehicles to prove Taylor theorems through MVT. Also note that one comment on the post says that the integral remainder is better because it generalizes to vector calculus while the others do not. It is sad but the fact that almost everything in analysis is expressed through limits is so easily explainable. Since the rationals are incomplete, we complete them using limits. All the completion numbers (irrationals) will appear in any construct we devise, so per example, if we devise a construction for derivatives, it must forcefully contain limits for irrationals can very well be derivatives, so if the construction of derivatives did not contain limits, we would have a way to construction irrationals bypassing limits, which makes no sense.

II Asymptotic Notation

15. It is important to internalize asymptotic notation such that it becomes second nature. In our opinion, there are obstacles to this happening naturally which can in fact be remedied. We try to do this here.

16. Consider the relation between x^2 and x^3. It is both true that x^3 'grows faster' than x^2 and that x^3 'vanishes faster' than x^2. Of course, we 'mean' the former in terms of x increasing to infinity, and the latter in terms of x decreasing to zero, 'vanishing'.

17. The difference between the cases above is captured by having to specify what x tends to when using the little-o notation but this does not change the fact that it disallows us to write what we would like to, a simple 'asymptotic' inequality, since then we could write $x^3 > x^2$ for the first case, contradicting $x^3 < x^2$ for the second. This is not a problem because we have to specify what x tends to, but then we loose the simple notation. Is there a better way? None of the known notations solve this.

18. The problem is that the definitions are too general. Per example,

$$f(x) \in o(g(x)) \text{ as } x \to x_0$$

is defined as

$$\lim_{x\to x_0} \frac{f(x)}{g(x)} = 0.$$

For the limit of this ratio (assuming $g(x)$ eventually either keeps 'growing' or 'shrinking' as x tends to x_0) to go to zero we have two possibilities

1. If the denominator is 'growing', the numerator is either 'shrinking' or 'growing slower' than it. In short, the denominator is 'growing faster' than the numerator.
2. If the denominator is 'shrinking', the numerator is 'shrinking faster' than the denominator.

It is the (powerful and arguably elegant) mixing of these two possibilities into the same relation that causes latent confusion and refusal to assimilate it as 'second nature' by the cerebra of simple-minded humans such as the author, despite the differentiation depending on whether x tends to a number or to infinity. Admittedly, the generality allows talking not only about x tending to something, but could also about $h(x)$ tending to $k(x)$. Haivng said that, we have never seen such uses. Similarly, what $f(x) \in o(g(x))$ means would be quite convoluted if we chose $g(x)$ to be $1/x$ or $-x$. Such uses we have never seen either. $g(x)$ is usually assumed to be 'eventually' monotonically increasing in x.

19. Given the above, we propose to only ever consider two cases $x \to 0$ and $x \to \infty$ and recommend eventual increasing monotonicity of $g(x)$ short of requiring it. We relate the case $x \to 0$ to the symbol

$$\urcorner$$

where the exaggerated accent suggests that '$(x - x_0)$ is going down (to zero)'. In similar vain, we use for $x \to \infty$ the symbol

$$\llcorner$$

noting that we are able to reverse the symbols in the same way we do for ordinary inequalities. Furthermore, we use non-strict inequalities to differentiate between little-o and big-o.

20. Given appropriate choice of typography, we can superimpose both symbols as shown in the asymptotic inequalities (1) and (2). This can be considered as both the 'highlight' of this notation and its usefulness in clearly exposing the 'issues' we mentioned.

21. This notation translates in the following manner to Landau notation:

$(x - x_0) \to 0,$	$f(x) \in o(g(x))$	$f \urcorner g$	$g \ulcorner f$
$(x - x_0) \to 0,$	$f(x) \in O(g(x))$	$f \overline{\urcorner} g$	$g \overline{\ulcorner} f$
$x \to \infty,$	$f(x) \in o(g(x))$	$f \llcorner g$	$g \lrcorner f$
$x \to \infty,$	$f(x) \in O(g(x))$	$f \underline{\llcorner} g$	$g \underline{\lrcorner} f$

and can be faithfully read this way:

$f \urcorner g$	f vanishes faster than g
$f \ulcorner g$	f vanishes slower than g
$f \lrcorner g$	f grows faster than g
$f \llcorner g$	f grows slower than g

22. In this notation we have:

$$x^3 \lrcorner x^2$$
$$x^2 \llcorner x^3$$
$$x^3 \urcorner x^2$$
$$x^2 \ulcorner x^3$$

and we can go (mischievously) further writing:

$$x^3 \mathrel{\lrcorner\!\!\!\urcorner} x^2 \tag{1}$$

$$x^2 \mathrel{\ulcorner\!\!\!\llcorner} x^3. \tag{2}$$

III Vector Calculus with Peter Olver

III.1 In Two Dimensions

23. Why is the gradient an 'operator'? It obviously is, but there are comments to be made on this. The gradient is an 'operator' in the following sense: Take a vector field, which is a vector valued function $R^m \to R^n$ (per example take $f(x, y) : R^2 \to R$). The (formal) gradient $(\partial/d_x, \partial/d_y)$ is indeed an operator, since it maps R^2 to R^2, but once one has seen tensors, duals, etc. One cannot leave it at that anymore, since the gradient is actually mapping 'points' to 'vectors'. It is a tensor onto the vector space R^2 only up to 'erasure of type', and this has to be noted. This actually probably has lead to implicit misconceptions disallowing internalization. For a quick definition see http://www.math.uiuc.edu/~song74/teaching/m241summer14/vector_field/Geometry%20of%20Vector%20Field.pdf

24. Remember: Tangent to a curve:$(x(t), y(t))$and gradient of a 'vector field', of a functional, of a vector valued function:$f(x, y)$

25. Can one treat tangents and gradients in a unified manner? For a curve, what would the one-dimensional number indicate as a kind of gradient separately for each coordinate?

26. A nice justification (other than cusps) for a particle's trajectory being smooth requiring that its speed (and hence its tangent vector) never being zero is that because this implies it changes its 'direction of motion' instantaneously, from positive to negative in the zero sized (infinitesimal) interval where it passes zero.

Bibliography

Burn, Robert Pemberton. 2015. ***Numbers and Functions: Steps into Analysis***. Cambridge University Press.

Hildebrand, A.J. "Asymptotic Analysis Lecture Notes, Math 595, Fall 2009."

Nohra, Jad. “Hand_mvt_ivt_taylor_onot.”

———. “Hand_absorb_dx.”

Tucker, Thomas W. 1997. “Rethinking Rigor in Calculus: The Role of the Mean Value Theorem.” ***The American Mathematical Monthly*** 104 (3). JSTOR: 231–40.