This is an Accepted Manuscript of an article published by Taylor & Francis in International Journal of Mathematical Education in Science and Technology on 10 March 2021, available online: https://www.tandfonline.com/10.1080/0020739X.2021.1896812

A New Perspective on Teaching the Natural Exponential to Engineering Students

Mukhtar Ullah ·
Muhammad Naveed Aman ·
Olaf Wolkenhauer · **Jamshed Iqbal**

Received: date / Accepted: date

Abstract The natural exponential and logarithm are typically introduced to undergraduate engineering students in a calculus course using the notion of limits. We here present an approach to introducing the natural exponential (logarithm) through a novel interpretation of derivatives. This approach does not rely on limits, allowing an early and intuitive introduction of these functions. The question behind our contribution is whether one can introduce derivatives using only polynomials and power series? Motivated by an earlier exposure of engineering students to differential equations, we demonstrate that the natural exponential/logarithm can arise from two common differential equations. Our limit-free approach to derivatives provides an intuitive interpretation of e, the Euler number, and an intuitive introduction of time constants in first-order dynamical systems.

M. Ullah
Deptt. of Electrical Engineering, National University of Computer and Emerging Sciences, Islamabad, Pakistan.
Tel.: +92-346-9546289
E-mail: mukhtar.ullah@nu.edu.pk

M. N. Aman
School of Computing, National University of Singapore, Singapore.
Tel.: +65-8510 9201
E-mail: naveed@comp.nus.edu.sg

O. Wolkenhauer
Institute of Computer Science, University of Rostock, Rostock, Germany.
Tel.:+49-171-7410731
E-mail: olaf.wolkenhauer@uni-rostock.de

J. Iqbal (Correspondin Author)
Department of Computer Science and Technology, Faculty of Science and Engineering, University of Hull, HU6 7RX, UK
Tel.:+44-7459017229
E-mail: j.iqbal@hull.ac.uk

Keywords Exponential Function · Logarithmic Function · Power Series · Ordinary Differential Equations

1 Introduction

The natural exponential and logarithm are typically introduced to undergraduate engineering students in a calculus course using the notion of limits [1]. These functions play an important role in ordinary differential equations (ODEs), a topic introduced later. These two tightly linked concepts are thus introduced in separate classes, often with different textbooks and by different lecturers. Therefore, we found it useful to introduce the natural exponential earlier, as a solution to an ODE, but then we also require a simplified introduction to ODEs, without relying on limits. We found that students often struggle with the limit-based introduction to calculus. The limits-based introduction to calculus has been enriched in [2] with computational and graphical aids. More intuitive, it seems, is the introduction of the derivative as the slope of a tangent to a polynomial curve [3]. We extend the approach in [3] to include non-polynomials of a wider class. While this approach is not a replacement of a rigorous limits-based treatment of calculus [1,4], teaching the natural exponential together with differential equations links the concepts early to real world problems and provides an intuitive introduction to time constants in first-order dynamical systems.

The rest of the paper is organized as follows. Section 2 introduces the concept of tangent for polynomial curves. Section 3 presents examples of how ODEs arise in engineering. Section 4 discusses a commonly arising ODE to introduce the exponential and logarithmic functions as a power series. Section 5 concludes the paper.

2 From Tangent to Derivative

Experience with curves shows that the linear trend of a curve at a point is reasonably captured by a tangent line to the curve at that point. The "linear trend" answers two questions at a point: (i) what is the ordinate of the curve? and (ii) how is the curve directed? The tangent to the curve at a point is unique among all the lines passing through the point in that it is directed along the curve by "just touching" it. Procedures to draw tangents to a circle (and other simple curves) are taught in basic geometry courses. One such example is a compass-and-straightedge construction of a tangent to a circle illustrated in Figure 1. To construct a line `t` tangent to a circle at a point `T` on the circle, a line `a` is drawn from the center `O` through `T` and a line perpendicular to `a` through `T` [5] is drawn. Neither the informal description "just touching", nor any geometrical procedure helps understand the relationship between the tangent and the derivative. Students could answer questions about the conditions for a line to be tangent to a curve at a point. Curves of polynomials are familiar to freshman students. As our first example, consider a curve represented by a quadratic polynomial

$$P(x) = a_0 + a_1 x + a_2 x^2. \tag{1}$$

We are seeking the linear trend of the curve near $x = r$, i.e., a linear polynomial $L(x) = P(r) + m(x - r)$ whose graph is a line that passes through, and has a slope m that truly represents the direction of the curve at, the point $(r, P(r))$. Towards that end, let us express $P(x)$ in powers of $(x - r)$

$$\begin{aligned} P(x) &= a_0 + a_1 (r + x - r) + a_2 (r + x - r)^2 \\ &= P(r) + (a_1 + 2a_2 r)(x - r) + a_2 (x - r)^2 \end{aligned} \tag{2}$$

and examine the quadratic remainder

$$\begin{aligned} R(x) &= P(x) - L(x) \\ &= (a_1 + 2a_2 r - m)(x - r) + a_2 (x - r)^2 \ . \end{aligned} \tag{3}$$

Following observations can be made:

- $R(r) = 0$ as the line passes through $(r, P(r))$ regardless of the slope m.
- If $m \neq a_1 + 2a_2 r$, the $R(x)$ has a single root at $x = r$, that is, the line intersects the curve once at $(r, P(r))$ and once elsewhere. With intersections at different points, the slope m cannot represent the direction of the curve.
- If $m = a_1 + 2a_2 r$, $R(x)$ has a double root at $x = r$, i.e., the line intersects the curve twice at $(r, P(r))$. With repeated intersections at the point, the slope $a_1 + 2a_2 r$ truly represents the direction of the curve at the point.

Figure 2 illustrates the last two points and what is meant by a line *tangent* to a curve at a point: it truly represents the ordinate and direction of the curve because it intersects the curve *twice* at the point in question. The repeated intersection at the point is the precise condition replacing the informal phrase "just touch". Lines intersecting a curve at distinct points are said to be secant to the curve. What we essentially learned from the above is the following:

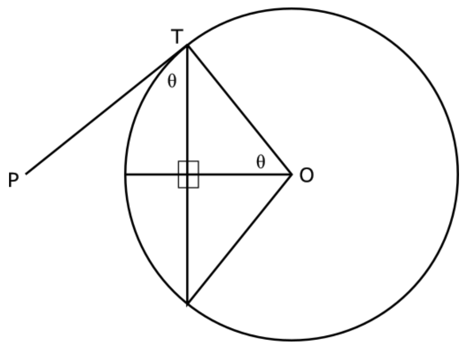

Fig. 1 Geometric construction of a tangent to a circle modified from [5]

.

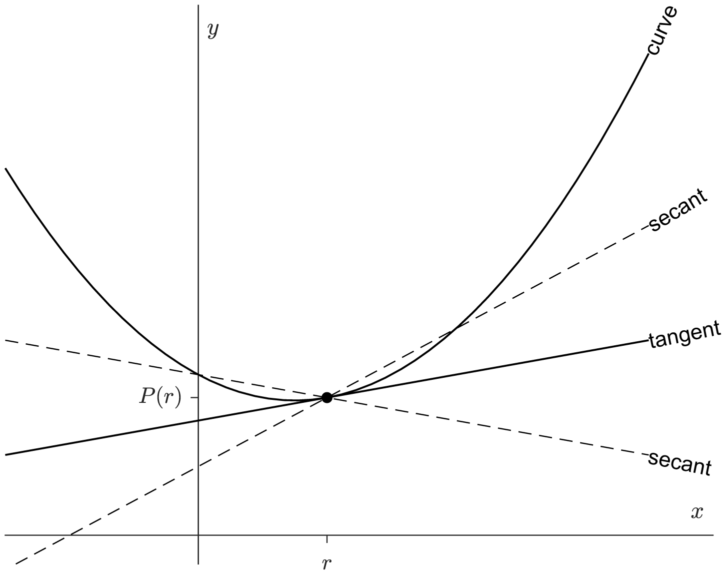

Fig. 2 The tangent "just touches" the curve whereas the secants cross it.

Lemma 1 *In the vicinity of a point* $x = r$*, a quadratic polynomial* $P(x) = a_0 + a_1 x + a_2 x^2$ *has a linear trend* $P(r) + (a_1 + 2a_2 r)(x - r)$ *and a quadratic remainder* $a_2 (x - r)^2$*. The graph of the linear trend is the tangent line to the curve at* $(r, P(r))$ *and has slope* $a_1 + 2a_2 r$*.*

As a specific example, in the vicinity of $x = 2$, the quadratic polynomial $14 - 3x + x^2 = 12 + (x - 2) + (x - 2)^2$ has a linear trend $12 + (x - 2)$ and a quadratic remainder $(x-2)^2$. We now extend the above approach to work out the linear trend of a degree-n polynomial,

$$P(x) = \sum_{k=0}^{n} a_k x^k = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n \tag{4}$$

at $x = r$. As before, $P(x)$ needs to be expressed in powers of $x - r$. A good starting point is the expansion

$$\begin{aligned} x^k &= (r + x - r)^k = \sum_{i=0}^{k} \binom{k}{i} r^{k-i} (x - r)^{i-2} \\ &= r^k + (x - r) k r^{k-1} + (x - r)^2 Q_{k-2}(r, x - r) \end{aligned} \tag{5}$$

where

$$Q_{k-2}(r, x - r) = \sum_{i=2}^{k} \binom{k}{i} r^{k-i} (x - r)^{i-2} \tag{6}$$

is a polynomial of degree $k-2$ in $x-r$. Termwise substitution of this expansion allows to expand $P(x)$ too:

$$\begin{aligned}
P(x) &= \sum_{k=0}^{n} a_k x^k \\
&= \sum_{k=0}^{n} a_k \left[r^k + (x-r) k r^{k-1} + (x-r)^2 Q_{k-2}(r, x-r) \right] \\
&= P(r) + (x-r) P'(r) + (x-r)^2 \tilde{P}(r, x-r)
\end{aligned} \tag{7}$$

where

$$P'(r) = \sum_{k=1}^{n} k a_k r^{k-1} \tag{8}$$

$$\tilde{P}(r, x-r) = \sum_{k=2}^{n} a_k Q_{k-2}(r, x-r) \tag{9}$$

Examine the polynomial remainder

$$\begin{aligned}
R(x) &= P(x) - L(x) \\
&= (x-r)(P'(r) - m) + (x-r)^2 \tilde{P}(r, x-r) \ .
\end{aligned} \tag{10}$$

Observations similar to those in the quadratic case can be made.

- $R(r) = 0$ because the line passes through $(r, P(r))$ regardless of the slope m.
- If $m \neq P'(r)$, the remainder has only a single root at $x = r$, the remaining roots being elsewhere.
- If $m = P'(r)$, the remainder has at least two roots at $x = r$, that is, the line intersects the curve at least twice at $(r, P(r))$.

Thus, the result previously obtained for a quadratic polynomial holds for all polynomials:

Lemma 2 *In the vicinity of a point $x = r$, a degree-n polynomial can be decomposed into a linear trend and and a nonlinear remainder, namely*

$$\begin{aligned}
P(x) &= \sum_{k=0}^{n} a_k x^k \\
&= P(r) + (x-r) P'(r) + (x-r)^2 \tilde{P}(r, x-r)
\end{aligned} \tag{11}$$

where $P'(r)$, given by (8), is the slope of the linear trend $P(r) + (x-r) P'(r)$ and $\tilde{P}(r, x-r)$, given by (9), is a polynomial of degree $n-2$ in $x-r$. The last equation can be re-written in an alternative form to be found useful later:

$$\begin{aligned}
P(r+h) &= \sum_{k=1}^{n} a_k (r+h)^k \\
&= P(r) + hP'(r) + h^2 \tilde{P}(r, h)
\end{aligned} \tag{12}$$

As a specific example, in the vicinity of $x = 2$, the degree-3 polynomial $4+11x-5x^2+x^3 = 12+3(x-2)+(x-2)^2+(x-2)^3$ has a linear trend $12+3(x-2)$ and a nonlinear remainder $(x-2)^2(x-1)$. Starting from qualitative statements about polynomial curves, we have derived a quantitative and precision notion of the tangent. At this point an educator may rise the question of what happens to non-polynomials. The prelude in [3] provides a starting point answering this question but crucially, this argument eventually requires again limits-based calculus. What we contribute here is an intuitive extension of the approach in [3] to non-polynomials.

2.1 Functions beyond polynomials

Not every function is a polynomial and the decomposition in (12) needs extension to accommodate non-polynomials. Examples include results of taking reciprocal or square root of a polynomial and dividing a polynomial by another of a higher degree:

$$\frac{1}{r+h} = \frac{1}{r} - \frac{h}{r^2} + \frac{h^2}{r^3}\left(1 - \frac{h}{r} + \frac{h^2}{r^2}\cdots\right) \tag{13}$$

$$\sqrt{r+h} = \sqrt{r} + \frac{1}{2}\frac{h}{\sqrt{r}} - \frac{h^2}{8r\sqrt{r}}\left[1 - \frac{1}{2}\frac{h}{r} + \frac{5}{16}\frac{h^2}{r^2} + \cdots\right] \tag{14}$$

In both cases, the expression on the right can be viewed as a decomposition of the form

$$f(r+h) = f(r) + hf'(r) + h^2\tilde{f}(r,h) \tag{15}$$

where the factor $\tilde{f}(r,h)$ appearing in the nonlinear remainder is a power series convergent over the interval $|h| < r$ around $r \neq 0$. Following extension of the foregoing decomposition into linear trend and nonlinear remainder is proposed so as to accommodate a wide class of non-polynomials.

Lemma 3 *In the vicinity of a point $x = r$, the linear trend of a function $f(x)$ can be defined if it is possible to decompose the function around $x = r$ in the form*

$$f(x) = f(r) + (x-r)f'(r) + (x-r)^2\tilde{f}(r, x-r) \tag{16}$$

where $\tilde{f}(r, x-r)$ is either a polynomial or a power series in $x-r$ convergent over an approriate region of interest around $x = r$. The slope $f'(r)$ of the linear trend is called the derivative of $f(x)$ at r[1].

A specific example is the expansion on the right of (13) of $f(x) = 1/x$ with $r = 2$ and $f'(r) = -1/r^2$. Since r can be chosen anywhere on the real line, it is legitimate to replace r in $f'(r)$ with x so as to get a function in its own

[1] Notice the subtle difference between the usage of the two terms: slope of a line versus derivative of a function

right. Viewed as a function of x, $f'(x)$ is called the *derivative* of $f(x)$ and must be interpreted as appearing in the following decomposition:

$$f(x+\delta x) = f(x) + f'(x)\,\delta x + \tilde{f}(x,\delta x)\,(\delta x)^2 \tag{17}$$

For any *increment* δx in x, the increment in $f(x)$ is

$$\delta f(x) = f(x+\delta x) - f(x)\ . \tag{18}$$

The linear trend $f'(x)\,\delta x$ of the increment is called the differential of $f(x)$ and denoted by $\mathrm{d}f(x)$. Commonly, the increment in x is replaced by the differential $\mathrm{d}x$ so as to rewrite foregoing decomposition as[2]

$$f(x+\mathrm{d}x) = f(x) + f'(x)\,\mathrm{d}x + \tilde{f}(x,\mathrm{d}x)\,(\mathrm{d}x)^2\ . \tag{19}$$

The increment and differential of $f(x)$ are then written as

$$\delta f(x) = f(x+\mathrm{d}x) - f(x) \tag{20}$$

$$\mathrm{d}f(x) = f'(x)\,\mathrm{d}x \tag{21}$$

It is very important to realize that the derivative $\mathrm{d}f/\,\mathrm{d}x$ is a valid quotient of two differentials, a fact often reminded by calling it differential quotient. The differentials are widely mistaken as infinitesimals to the extent that some criticize treating the derivative as a quotient. With equation (19), we now have a scheme by which we can differentiate functions without limits.

2.2 Properties of differentiation

In case someone wonders how to derive the various rules of differentiation from (19), we include a few derivations here. Let us consider two functions $f(x)$ and $g(x)$ satisfying

$$f(x+\mathrm{d}x) = f(x) + f'(x)\,\mathrm{d}x + \tilde{f}(x,\mathrm{d}x)\,(\mathrm{d}x)^2 \tag{22}$$

$$g(x+\mathrm{d}x) = g(x) + g'(x)\,\mathrm{d}x + \tilde{g}(x,\mathrm{d}x)\,(\mathrm{d}x)^2 \tag{23}$$

1. **Constant:** c is a special case of (11) when $a_0 = c$ and $a_k = 0$ for all $k > 0$. Therefore $\mathrm{d}c/\,\mathrm{d}x = 0$.
2. **Power rule:** x^n is a special case of (11) when $a_n = 1$ and $a_k = 0$ for all $k < n$. Therefore $\mathrm{d}x^n/\,\mathrm{d}x = 0$
3. **Product rule:**

$$f(x+\mathrm{d}x)\,g(x+\mathrm{d}x) =$$

$$f(x)g(x) + (f(x)\,g'(x) + f'(x)\,g(x))\,\mathrm{d}x + \widetilde{fg}(x,\mathrm{d}x)\,(\mathrm{d}x)^2 \tag{24}$$

$$\Rightarrow (fg)'(x) = f(x)\,g'(x) + f'(x)\,g(x)\,. \tag{25}$$

[2] Since $\mathrm{d}x$ can be chosen arbitrarily (large or small) just like δx, there is no point in distinguishing between the two and we can take $\mathrm{d}x = \delta x$.

4. **Scaling:** Set $f(x) = c$ in the product rule

$$(cg)'(x) = cg'(x). \tag{26}$$

5. **Addition of n functions**

$$\sum_{i=1}^{n} f_i(x + \mathrm{d}x) =$$
$$\sum_{i=1}^{n} f_i(x) + \left[\sum_{i=1}^{n} f_i'(x)\right]\mathrm{d}x + \left[\sum_{i=1}^{n} \widetilde{f}_i(x, \mathrm{d}x)\right](\mathrm{d}x)^2 \tag{27}$$
$$\Rightarrow \frac{\mathrm{d}}{\mathrm{d}x}\sum_{i=1}^{n} f_i(x) = \sum_{i=1}^{n} f_i'(x). \tag{28}$$

6. **Power series:** We expect the addition rule to be true for an infinite power series, though the proof is beyond the scope of this paper. This is one example of why rigorous treatments of calculus cannot be replaced by intuitive ones like ours.

$$\frac{\mathrm{d}}{\mathrm{d}x}\sum_{i=1}^{\infty} f_i(x) = \sum_{i=1}^{\infty} \frac{\mathrm{d}}{\mathrm{d}x} f_i(x) \tag{29}$$
$$\frac{\mathrm{d}}{\mathrm{d}x}\sum_{k=1}^{\infty} a_k (x - c)^k = \sum_{k=1}^{\infty} a_k k (x - c)^{k-1}. \tag{30}$$

7. **Linearity:** Combining scaling and addition

$$(af + bg)'(x) = af'(x) + bg'(x). \tag{31}$$

8. **Reciprocal:** Set $g = 1/f$ in the product rule

$$0 = f(x)(1/f)'(x) + f'(x)\frac{1}{f(x)} \tag{32}$$
$$\Rightarrow (1/f)'(x) = -f'(x)/[f(x)]^2. \tag{33}$$

9. **Quotient rule:** Set $g = 1/q$ in the product rule

$$\begin{aligned}(f/q)'(x) &= -f(x)q'(x)/[q(x)]^2 + f'(x)/q(x) \\ &= \frac{f'(x)q(x) - f(x)q'(x)}{[q(x)]^2}.\end{aligned} \tag{34}$$

10. **Chain rule:**

$$\mathrm{d}f(g(x)) = f'(g(x))\,\mathrm{d}g(x) = f'(g(x))\,g'(x)\,\mathrm{d}x. \tag{35}$$

$$\begin{aligned}(f \circ g)'(x) &= \frac{\mathrm{d}f(g(x))}{\mathrm{d}x} = \frac{\mathrm{d}f(g(x))}{\mathrm{d}g(x)}\frac{\mathrm{d}g(x)}{\mathrm{d}x} \\ &= f'(g(x))\,g'(x).\end{aligned} \tag{36}$$

$$\frac{\mathrm{d}\left[f\left(x\right)\right]^{n}}{\mathrm{d}x}=\frac{\mathrm{d}\left[f\left(x\right)\right]^{n}}{\mathrm{d}f\left(x\right)}\frac{\mathrm{d}f\left(x\right)}{\mathrm{d}x}=n\left[f\left(x\right)\right]^{n-1}f'\left(x\right). \tag{37}$$

We have reached our first milestone, introducing differential calculus without limits, preparing us to understand ODEs. As a bonus, working out differenia-tion from first priciples gets self-explanatory and compact.

2.3 The D-operator and its inverse

The notation used in our introduction to differentiation is not helpful in our journey to the exponential via ODEs. Towards that end, it is neater to introduce a so called D-operator defined by its action $\mathrm{D}\,f\left(x\right)=f'\left(x\right)=\mathrm{d}f/\,\mathrm{d}x$ on a function $f\left(x\right)$.

The question then is whether we can recover a function $f\left(x\right)$ from its derivative $g\left(x\right)$. In other words, can we solve the ODE

$$\mathrm{D}\,f\left(x\right)=g\left(x\right). \tag{38}$$

Say we identify, after a few tries, a function $G\left(x\right)$ with derivative $g\left(x\right)$. We say that $G\left(x\right)$ is an anti-derivative of $g\left(x\right)$ and the general solution to the ODE is the family of all anti-derivatives

$$f\left(x\right)=\mathrm{D}^{-1}\,g\left(x\right)+c=G\left(x\right)+c,\quad c\in\mathbb{R} \tag{39}$$

because the derivative of every constant c is zero. The D^{-1}-operator here gives a particular anti-derivative, also called a particular integral, which ignores constants. At times, we may require the solution to satisfy what is called an initial condition. The ODE together with such an initial condition is called an initial value problem (IVP). If we require the solution to pass through (x_0, f_0), the IVP takes the form

$$\mathrm{D}\,f\left(x\right)=g\left(x\right),\quad f\left(x_0\right)=f_0\,. \tag{40}$$

The particular constant needed to satisfy the initial condition can be determined by choosing $x=x_0$ in the general solution, namely

$$f_0=G\left(x_0\right)+c \tag{41}$$

which gives $c=f_0-G\left(x_0\right)$. The solution to the IVP then reads

$$f\left(x\right)=f_0+\left[\mathrm{D}^{-1}\,g\left(t\right)\right]_{t=x_0}^{t=x}=f_0+G\left(x\right)-G\left(x_0\right)\,. \tag{42}$$

What follows are a few examples of how ODEs arise in practice.

3 How do ODEs arise?

As our goal is to show that the exponential function arises naturally as the solution to an ODE, one must wonder how do ODEs arise in the first place? It turns out that, for many physical processes, derivatives come naturally and one has to find the anti-derivative. Many functions are abstractions of physical quantities generated by some physical process. Attempts of understanding a process based on measured values of these functions often suggest that the process can be modeled by a first-order ODE, similar to that discussed in the closing remarks of the previous section.

3.1 Population growth

According to the Malthusian model, the instantaneous growth rate $\mathrm{d}n/\,\mathrm{d}t$ of a population of bacteria in a resource-rich environment, is to proportional to the number $n\,(t)$ of bacteria present at that instant. The population growth obeys the IVP

$$\mathrm{d}n/\,\mathrm{d}t = rn\,(t)\,, \quad n(0) = n_0, \tag{43}$$

where r is the rate constant and n_0 is the initial number of bacteria. Other examples of unrestricted growth are tumor growth, nuclear chain reaction, and the avalanche breakdown.

3.2 R-C circuit

The voltage $V\,(t)$ across the capacitor in a source-free R-C circuit is dropped across the resistor as $-RC\,\mathrm{d}V/\,\mathrm{d}t$. Starting from an initial voltage V_0, the decaying capacitor voltage obeys the IVP

$$\mathrm{d}V/\,\mathrm{d}t = (-1/RC)\,V\,(t)\,, \quad V\,(0) = V_0\,. \tag{44}$$

3.3 R-L circuit

The current $I(t)$ through an inductor in a source-free R-L circuit establishes a potential difference $L\,\mathrm{d}I/\,\mathrm{d}t$ across the inductor that is dropped across the resistor as $-RI(t)$. Starting from an initial current I_0, the decaying inductor current obeys the IVP

$$\mathrm{d}I/\,\mathrm{d}t = (-R/L)\,I(t), \quad I\,(0) = I_0\,. \tag{45}$$

Table 1

	R-C	R-L	Dashpot-mass	Dashpot-spring	Population growth
q	V	i	v	x	n
a	$-1/RC$	$-R/L$	$-b/m$	$-k/b$	r

3.4 Dashpot-mass model

The rate $-m\,\mathrm{d}v/\,\mathrm{d}t$ of decrease of momentum of a point mass is caused by a viscous force $bv(t)$ in the absence of a driving force. Starting from an initial velocity v_0, the decaying velocity v_t obeys the IVP

$$\mathrm{d}v/\,\mathrm{d}t = (-b/m)\,v\,(t)\,, \quad v\,(0) = v_0\,. \tag{46}$$

3.5 Dashpot-spring model

The deformation $x(t)$ present in a spring establishes a restoring force $-kx(t)$ that is balanced by a viscous force $b\,\mathrm{d}x/\,\mathrm{d}t$ the absence of a driving force. Starting from an initial deformation x_0, the decaying deformation obeys the IVP

$$\mathrm{d}x/\,\mathrm{d}t = (-k/b)\,x\,(t)\,, \quad x\,(0) = x_0\,. \tag{47}$$

3.6 One ODE for all

One notices that the different IVPs can all be represented by a single IVP of the form

$$\mathrm{d}q/\,\mathrm{d}t = aq\,(t)\,, \quad q\,(0) = q_0 \tag{48}$$

with rate parameter a and intial value q_0. The rate parameter $a = (\mathrm{d}q/\,\mathrm{d}t)\,/q$, also called the relative growth rate, helps to distinguish among the different processes, as illustrated in Table 1.

The initial value q_0 helps to distinguish among the different solutions for a particular process. Can we simplify (48) so that we do not have to worry about the values of a and q_0? Towards, that end, define two dimensionless quantities

$$\bar{t} = at \tag{49}$$

$$\bar{q}\,(\bar{t}) = q\,(t)\,/q_0 = q\,(\bar{t}/a)\,/q_0 \tag{50}$$

and employ the chain rule

$$\frac{\mathrm{d}\bar{q}}{\mathrm{d}\bar{t}} = \frac{\mathrm{d}\bar{q}}{\mathrm{d}t}\frac{\mathrm{d}t}{\mathrm{d}\bar{t}} = \frac{1}{q_0}\frac{\mathrm{d}q}{\mathrm{d}t}\frac{1}{a} = \frac{q\,(t)}{q_0} = \bar{q}\,(\bar{t}) \tag{51}$$

to rewrite the IVP in the form

$$\mathrm{d}\bar{q}/\,\mathrm{d}\bar{t} = \bar{q}\,(\bar{t})\,, \quad \bar{q}\,(0) = 1\,. \tag{52}$$

Assume that this IVP could be solved for the dimensionless $\bar{q}$, the associated physical quantity can then be recovered as

$$q\left(t\right) = q_0 \bar{q}\left(at\right) \,. \tag{53}$$

This relation highlights an important fact about time scales. Each first-order process has an intrinsic time scale, a characteristic time $\tau = 1/\left|a\right|$ called the time constant, that is the shortest time for discernible changes to be observed in a physical quantity [6]. In terms of the time constant, (53) can be rewritten as

$$q\left(t\right) = q_0 \bar{q}\left(\pm t/\tau\right), \tag{54}$$

where $\pm$ translates to $+$ for growth and $-$ for decay. The solution of the IVP in the dimensionless form is attempted in the following section. The value of the dimensionless representation is in its generalized nature that allows one to study a range of systems regardless of their domain or specifc parameters.

We have now reached our second milestone: we have completed our introduction to ODEs necessary for our final endeavor.

4 How does the exponential arise?

We have all the tools ready to show how the exponential function arises naturally as a solution to an IVP. Let us solve the IVP (52) written here in a more standard notation:

$$\mathrm{d}y/\,\mathrm{d}x = y, \quad y(0) = 1\,. \tag{55}$$

The "aha"-moment now comes when we simply apply successive anti-differentation using the D^{-1} operator,

$$\begin{aligned} y\left(x\right) &= y\left(0\right) + \left[\mathrm{D}^{-1}\, y\left(t\right)\right]_{t=0}^{t=x} = 1 + \mathrm{D}^{-1}\, y\left(x\right) \\ &= 1 + \mathrm{D}^{-1}\left[1 + \mathrm{D}^{-1}\, y\left(x\right)\right] = 1 + x + \mathrm{D}^{-2}\, y\left(x\right) \\ &= 1 + x + \mathrm{D}^{-2}\left[1 + \mathrm{D}^{-1}\, y\left(x\right)\right] = 1 + x + x^2/2 + \mathrm{D}^{-3}\, y\left(x\right) \\ &= 1 + x + x^2/2 + \mathrm{D}^{-3}\left[1 + \mathrm{D}^{-1}\, y\left(x\right)\right] \\ &= 1 + x + x^2/2! + x^3/3! + \cdots\,. \end{aligned} \tag{56}$$

This emergent power series satisfies the constraint in (55) on the process generating $y(t)$. Our next task is to establish that this power series can be viewed as an exponential function, one in which a constant is raised to x. We have enough reason and motivation to define a new function

$$\exp\left(x\right) = \sum_{n=0}^{\infty} x^n/n!, \tag{57}$$

which arises as a solution to a simple ODE. Figure 3 illustrates the construction of this new function from polynomials of increasing degrees.

Following observations can be made about $\exp\left(x\right)$ and its relationship with the ODE $\mathrm{d}y/\,\mathrm{d}x = y$:

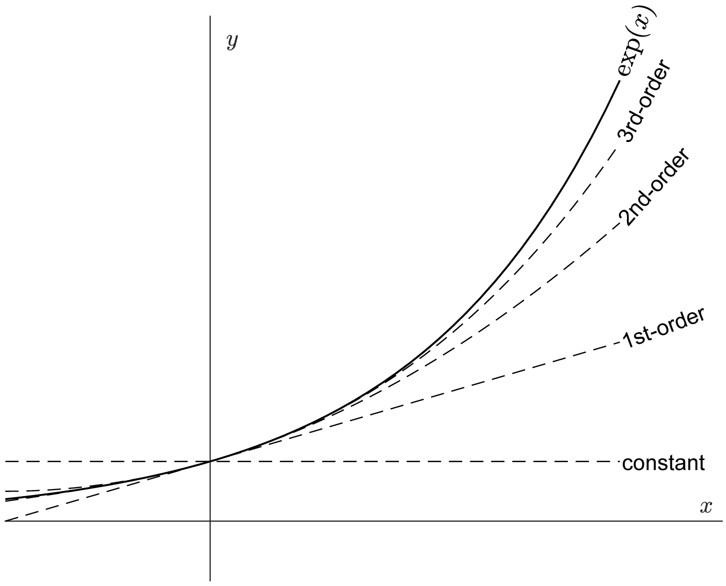

Fig. 3 Construction of the exponential from polynomials

- $\exp(x)$ is a (particular) solution passing through $(0,1)$.
- $y_0 \exp(x)$ is a solution passing through $(0, y_0)$.
- $\exp(x - x_0)$ is a solution passing through $(x_0, 1)$.
- $y_0 \exp(x - x_0)$ is a solution passing through (x_0, y_0).
- $\exp(x_0) \exp(x - x_0) = \exp(x)$ is the same solution that passes through $(x_0, \exp(x_0))$ and $(0,1)$.
- $\exp(x_1) \exp(x_2) = \exp(x_1 + x_2)$
- $\exp(x) \exp(-x) = \exp(0) = 1$ implies that $\exp(x)$ and $\exp(-x)$ are multiplicative inverses of each other.
- $\exp(x_1) / \exp(x_2) = \exp(x_1 - x_2)$
- $[\exp(x)]^m = \exp(mx)$
- $[\exp(x)]^{1/n} = \exp(x/n)$
- $[\exp(x)]^{m/n} = \exp(mx/n)$

All these properties are shared only by exponential functions. We have no choice left but to define the natural exponential function

$$\mathrm{e}^x = \exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots , \tag{58}$$

with the natural base e defined by

$$\mathrm{e} = \exp(1) = \sum_{n=0}^{\infty} \frac{1}{n!} = 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots . \tag{59}$$

The inverse function $\exp^{-1}$ of the natural exponential must be a logarithmic function. That motivates us to define the natural Logarithm by

$$\ln(y) = \exp^{-1}(y)\,. \tag{60}$$

Note that $\ln(y)$ is the length of interval (on the x-axis) in which $\exp(x)$ changes by a factor of y from its intial value of unity. Moreover, $\ln(y/y_0)$ is the length of interval (on the x-axis) in which $y_0 \exp(x)$ changes by a factor of y/y_0 from its intial value of y_0. Calculating $\ln(y)$ requires to solve $\exp(x) = y$ for x and that is difficult algebraically. The good news is that $\ln(y)$ is the solution to the IVP

$$\mathrm{d}x/\,\mathrm{d}y = 1/y, \quad x(1) = 0\,. \tag{61}$$

Similarly, $\ln(1+v)$ is the solution to the IVP

$$\frac{\mathrm{d}u}{\mathrm{d}v} = \frac{1}{1+v} = 1 - v + v^2 - v^3 + \cdots\,,\ |v| < 1,\ u(0) = 0\,. \tag{62}$$

Termwise anti-differentiation gives

$$\ln(1+v) = v - \frac{1}{2}v^2 + \frac{1}{3}v^3 - \frac{1}{4}v^4 + \cdots\,,\ -1 < v \le 1\,. \tag{63}$$

Replacing v by $-v$ gives,

$$\ln(1-v) = -v - \frac{1}{2}v^2 - \frac{1}{3}v^3 - \frac{1}{4}v^4 - \cdots\,,\ -1 \le v < 1\,. \tag{64}$$

Subtracting this from the previous equation gives

$$\begin{aligned} \ln\left(\frac{1+v}{1-v}\right) &= \ln(1+v) - \ln(1-v) \\ &= 2\left(v + \frac{1}{3}v^3 + \frac{1}{5}v^5 + \cdots\right),\ |v| < 1\,. \end{aligned} \tag{65}$$

Setting $v = (y-1)/(y+1)$ gives a power series of $\ln(y)$ valid for all positive reals,

$$\begin{aligned} \ln(y) &= \ln\left(\frac{1 + (y-1)/(y+1)}{1 - (y-1)(y+1)}\right) \\ &= 2\left(\frac{y-1}{y+1} + \frac{1}{3}\left(\frac{y-1}{y+1}\right)^3 + \frac{1}{5}\left(\frac{y-1}{y+1}\right)^5 + \cdots\right), y > 0\,. \end{aligned} \tag{66}$$

Figure 4 illustrates the construction of $\ln(y)$ from polynomials of increasing degrees in $(y-1)/(y+1)$. That completes our exploration: the exponential function arises naturally as the solution to a specific IVP and the logarithmic function arises as the solution to an associated IVP. It is important to notice that the IVPs were solved by successive anti-differentiation. This is very different from standard methods of solving these IVPs which require knowledge of the two functions (exponential and logaritmic). We have turned that around and showed how the two functions arise naturally. One last comment: (53) allows to use our solution to (55) to solve equations of the form (48) in the dimensional form.

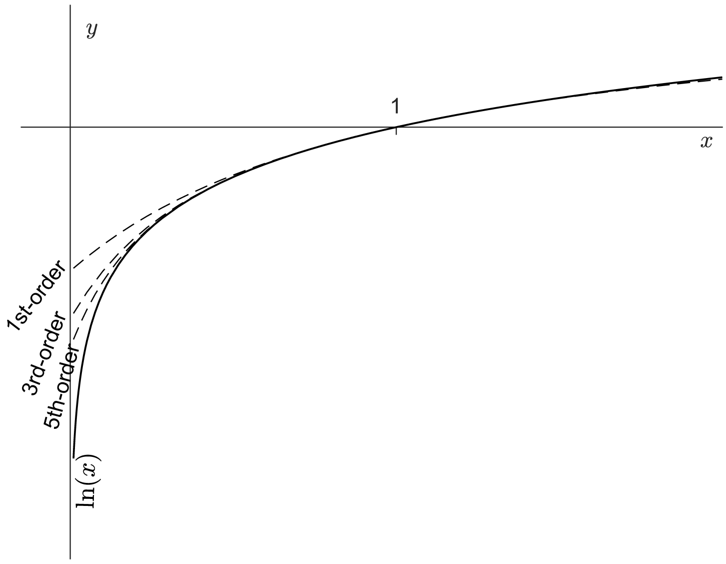

Fig. 4 Construction of the logarithm from polynomials.

5 Conclusion

This paper introduced differentiation and differential equations based on polynomials and power series rather than limits. It is shown that the natural exponential function and its inverse (the natural logrithmic function), arise naturally as solutions to commonly occuring first-order differential equations. We believe this quick introduction will enhance the understanding of undergraduate students before they are exposed to more rigrorous treatements of calculus. While this approach is not a replacement of a rigorous limits-based treatment of calculus, teaching the natural exponential together with differential equations links the concepts early to real world problems and provides an intuitive introduction to time constants in first-order dynamical systems.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Joel R. Hass, Christopher E. Heil, Maurice D. Weir, "Thomas' Calculus (Early Transcendentals)", Pearson 2018.
2. Kenneth Eriksson, Donald Estep, Claes Johnson, "Applied Mathematics: Body and Soul (Volume 1: Derivatives and Geometry in IR3)", Springer-Verlag 2004.
3. R Michael Range, "What is Calculus? From Simple Algebra to Deep Analysis", World Scientific 2015.

4. Philip M. Anselone & John W. Lee, Differentiability of Exponential Functions, The College Mathematics Journal, vol. 36:5, pp. 388-393, 2005.
5. https://en.wikipedia.org/wiki/Tangent_lines_to_circles.
6. J. David Logan, *Applied Mathematics*, John Wiley and Sons, New York (4ed), 2013.