2. The concept of derivative

T: In Section 1, we made our first step in the infinite world. Our discussion mainly focused on the sequence with infinite many of items, the first animal in calculus. The second animal in calculus is function and we touch a little in previous section about function. We look at this animal in more details now. Let me start with

Statement 2.1. *The calculus is a* powerful *tool to study functions.*

S: I know very little about function except something like parabolic function $y = x^2$.

T: That is enough for us to start. Conceptually,

Definition 2.1. *A* **function** *defines a mapping rule f that maps each number x in a given set of numbers, denoted by D_f and named as the domain of the function, to a value $f(x)$. To treat a function as a curve in xy-plane, we often write $y = f(x), x \in D_f$ or simply $y = f(x)$ if D_f is clear in the context.*

Remark 3. *If D_f is not specified, the default domain is set to include all x such that $f(x)$ is well-defined. For examples, the default domain for*

1. *$y = x^2$ is the set of all real numbers;*
2. *$y = \sqrt{x}$ is the set of all non-negative numbers;*
3. *$y = \frac{x}{1+x}$ is the set of all number but -1.*

Notice that the domain of the function

$$y = x^2, \quad x \in [0, 1]$$

is specified as $[0, 1]$ and therefore it is different from the standard parabolic function mentioned in Remark 3. On the other hand, the following two functions are identical although they looks different

$$y = |x|, \quad y = \sqrt{x^2}$$

since the two mapping rules produce same numbers and are both defined everywhere, i.e. they have same domain and same mapping results and as such they represent the same function.

S: The definition seems to me pretty arbitrary on both mapping rule and domain.

T: One can certainly make some blizzard function:

Example 3. *(Dirichlet Function)*

$$f(x) = \begin{cases} 1, & \text{if } x \text{ is a rational number in } [0, 1]; \\ 0, & \text{if } x \text{ is an irrational number in } [0, 1]. \end{cases} \tag{2.1}$$

Notice that the mapping rule and the domain ($[0, 1]$) is clearly defined and it is indeed a legitimate function. Functions can be constructed piecewisely as follows.

Example 4.

$$f(x) = \begin{cases} x+2, & -2 \le x \le -1 \\ x^2, & -1 < x < 1 \\ 0.5x^3 - 2x + 3, & 1 \le x \le 2 \\ -3x + 10, & 2 < x \le 4 \end{cases} \tag{2.2}$$

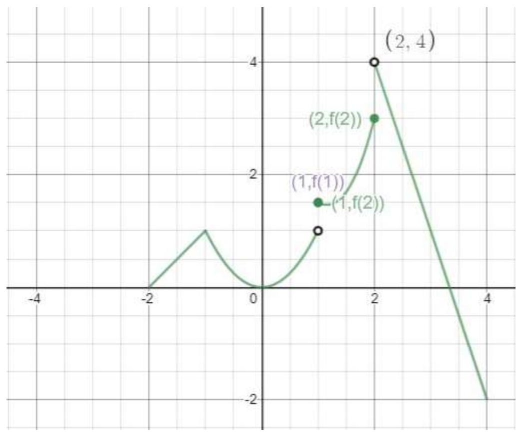

FIG 4. *The graph of the function defined in 2.2 with the domain $D_f = [-1, 4]$*

Remark 4. *Notice that Function (2.2) is defined using different expressions in several* **non-overlapping** *intervals. Function value at each point in the domain $D_f = [-2, 4]$ has to be defined exactly one time. In the graph 4, circles are used to emphasize that those points are excluded. For example, $f(2) = 3$ and $(2, 4)$ is circled out to empathize it is not on the curve.*

We need to restrict our discussion to certain regular functions later. In fact,

Task 2.1. *One of the main tasks in calculus is to construct varieties of functions that can be effectively handled and sufficient for a wide range of applications.*

We shall discuss how to achieve that goal from certain basic functions in future study. At this moment, let us be more specific and restrict $f(x)$ to be a polynomial with the following format:

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 \tag{2.3}$$

where $\{a_i, 0 \le i \le n\}$ are coefficients of the polynomial. $y = x^2$ is a special case with

$$n = 2, \quad a_2 = 1, \quad a_1 = a_0 = 0.$$

S: I prefer to view the function as a curve in xy plane. Could we also use the geometric approach to study general functions?

T: Why not. Let us start with the following geometric problem.

Discussion Question 2.1. *Find the equation of the line that is tangent to the curve $y = x^2$ at $p_0 = (x_0, y_0) = (2, 2^2)$* [15] *as shown in Figure 5.*

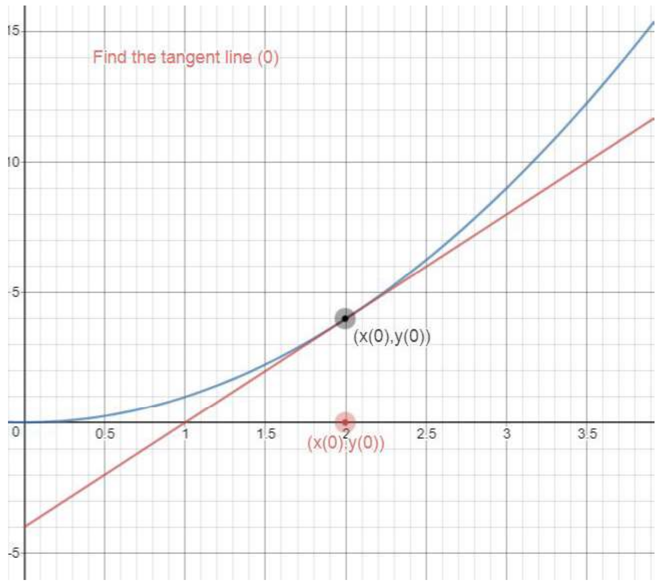

FIG 5. *The tangent line of $y = x^2$ at $p_0 = (x_0, y_0)$*

S: Given with two points in the line, I know how to find the equation. but ...

T: Recall that a line can also be determined by a point if the direction (i.e. the slope) is specified.

S: Wait! it seems that the direction is given implicitly in this case. I mean if you ask me to draw the line, I can adjust my ruler to find the direction. But how can I get the exact value of the slope?

T: There is no direct solution for the slope k. Let us follow the statement 1.3 and start with estimating it by the slope of the segment p_0p_2 that connects p_0 and a nearby point $p_2 = (x_2, y_2)$ on the curve as shown in Figure 6.

$$k \approx \frac{y_2 - y_0}{x_2 - x_0} = \frac{x_2^2 - x_0^2}{x_2 - x_0} = \frac{(x_2 - x_0)(x_2 + x_0)}{x_2 - x_0} =: g(x_2) \tag{2.4}$$

S: Should we cancel the common factor $(x_2 - x_0)$ in Equation (2.4) to simplify the function $g(x_2)$?

T: Before canceling out the common factor, I like to remind you that $g(x_2)$ is not well-defined at x_0. Remember that we need two points to draw a line and p_2 can not be reduced to p_0, i.e. $x_2 \neq x_0$. Now, canceling the common factor, we have

$$g(x_2) = x_2 + x_0 = x_2 + 2, \qquad x_2 \neq x_0 \tag{2.5}$$

[15] we write the point $p_0 = (2, 2^2)$ rather than $(2, 4)$ to empathize that the point is on the curve, i.e. it meets the curve equation $y = x^2$.

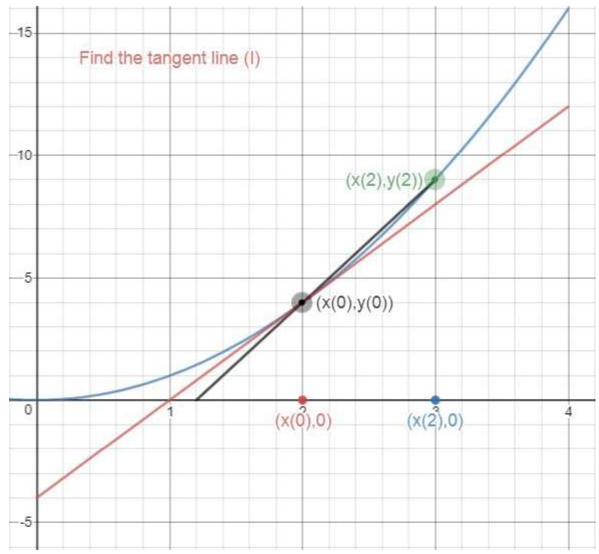

FIG 6. *Approximate the slope of tangent line by the slope of the line p_0p_2. The red line is the tangent line. The black line is determined by $p_0 = (x_0, y_0)$ and $p_2 = (x_2, y_2)$*

S: After the cancellation, it looks that $g(x_2)$ by (2.5) can be evaluated at $x_2 = 2$ and take the value 4. Does it imply that the slope of the tangent line is 4?

T: strictly speaking, we don't evaluate $g(x_2)$ directly at $x_2 = 2$ since $g(2)$ is not defined. But we see that $g(x_2)$ approaches to 4 as a "limit value" when x_2 "goes" to 2 and we take that "limit value" as the slope of the tangent line at p_0. The tangent line equation is

$$y - 2 = 4(x - 2)$$

S: I can see intuitively that as p_2 approaches to p_0, the slope of the line p_0p_2 should approach to the slope of the tangent line. But is it sufficient for us to depend on intuitive observation?

T: Let us try for a rigorous proof. In Figure 6, x_2 is on the right side of x_0. Do you think it is self-evidence that $g(x_2)$, the slope of line p_0p_2, is larger than k.

S: It looks obvious to me.

T: Let us take $g(x_2) > k$ for grant. You should also agree that, estimating from other side as shown in Figure 7, the slope $g(x_1)$ of the line p_1p_0 is less than k. As such, we have

$$x_1 + 2 < k < x_2 + 2,$$

for ANY x_1 and x_2 that are close to x_0 from left and right side respectively. Let us assume that $x_2 = 2 + \epsilon$ and $x_1 = 2 - \epsilon$ for ANY small $\epsilon > 0$, then we have

$$k < 4 + \epsilon, \tag{2.6}$$
$$k > 4 - \epsilon, \tag{2.7}$$

or equivalently

$$|k - 4| < \epsilon, \qquad \forall \epsilon > 0 \tag{2.8}$$

i.e. k can be in ANY ϵ neighborhood of 4.

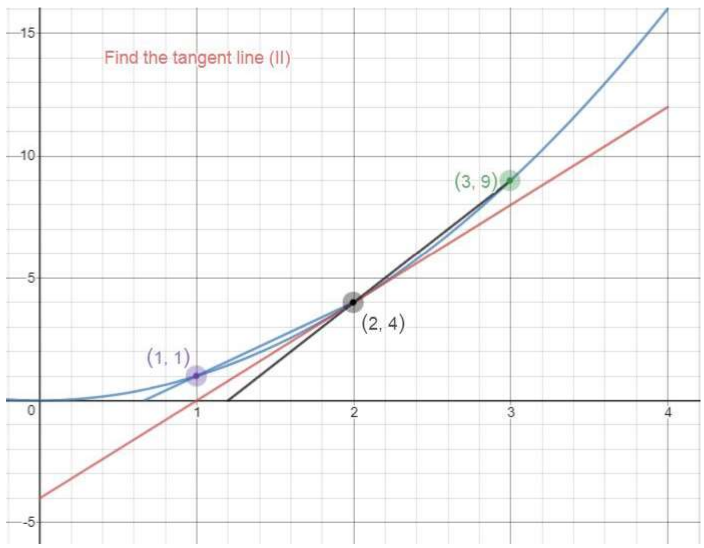

FIG 7. *Approximate the slope of tangent. The red line is the tangent line. the black line is determined by* $p_0 = (x_0, y(0))$ *and* $p_2 = (x_2, y_2)$ *and blue line is determined by* $p_0 = (x_0, y(0))$ *and* $p_1 = (x_1, y_1)$

S: Then k can not be any other number but 4. Otherwise, it would cause contradiction since ϵ can be arbitrarily small.

T: The above example demonstrates the idea summarized in Statement 1.3. Since there is no direct way to get the slope, we do the approximation in two directions: one from above and the other from below and obtains some inequity like (2.8), which can be used for us to find the target value. We summarize this typical sandwich method to show a quantity A (as the slope of the tangent line in above example) to be equal to B (the value 4 in above example).

Statement 2.2. *In stead of proving $A = B$ directly, we often show both $A \geq B$ and $A \leq B$ hold. To that goal, we can add a buffer and show that $A \geq B - \epsilon$ and $A \leq B + \epsilon$ for any $\epsilon > 0$. The use of buffer ϵ is a typical trick in further analysis.*

Applying the same idea, we find the slope of $y = x^n$ at $p_0 = (x_0, y_0)$.

$$\begin{aligned} k &\approx \frac{y_1 - y_0}{x_1 - x_0} \\ &= \frac{x_1^n - x_0^n}{x_1 - x_0} = \frac{(x_1 - x_0)(x_1^{n-1} + x_1^{n-2}x_0 + ... + x_1 x_0^{n-2} + x_0^{n-1})}{x_1 - x_0} \\ &= x_1^{n-1} + x_1^{n-2}x_0 + ... + x_1 x_0^{n-2} + x_0^{n-1}, \qquad x_1 \neq x_0, \end{aligned}$$

and by letting x_1 approach to x_0, the estimation approaches to

$$k = n x_0^{n-1}, \quad n \geq 1 \tag{2.9}$$

Remark 5. *As a special case, the slope of $y = x$ is 1 everywhere as expected. For $n = 0$, the function is reduced to a constant function and the slope is equal to 0.*

Example 5. *Applying the same argument as above example, the slope $y = (x - a)^n$ at x_0 is*

$$k = n(x_0 - a)^{n-1} \tag{2.10}$$

For a general function $y = f(x)$, the slope at a point (x_0, y_0) on the curve can be similarly estimated by

$$k \approx \frac{y_1 - y_0}{x_1 - x_0} = \frac{f(x_1) - f(x_0)}{x_1 - x_0}. \tag{2.11}$$

where x_1 is close to x_0. If the ratio (2.11) approaches to certain value by letting x_1 approach to x_0, we then obtain the exact value of the slope. Notice that k depends on x_0 and it is function of x_0. We write $f'(x)$ [16] to denote the slope function associated to the curve $y = f(x)$, i.e. $f'(x)$ denotes the slope of the curve at the point $(x, f(x))$. As one of the important features of the slope function, we have

Proposition 2.2. *if $f(x)$ and $g(x)$ are two functions such that their slope function can be found and a, b are two constant. Define $h(x) = af(x) + bg(x)$. Then*

$$(af(x) + bg(x))' = h'(x) = af'(x) + bg'(x) \tag{2.12}$$

Proof. The rigorous proof will be given in Section 6. It is based on the following identity.

$$\frac{h(x_1) - h(x)}{x_1 - x} = a\frac{f(x_1) - f(x)}{x_1 - x} + b\frac{g(x_1) - g(x)}{x_1 - x}.$$

As x_1 "approaches" to x, the left side "approach" to $h'(x)$ while right side to $af'(x) + bg'(x)$. □

[16]Usually, we use the notation x_0 when we want to emphasize that it is fixed during certain analysis. In above discussion, x_0 in the estimation process for $f'(x_0)$ is fixed and x_1 varies around x_0. Once $f'(x_0)$ is calculated, the value itself depends on the selection of x_0 and therefore it is a function of x_0. By the convention, we change the notation $f'(x_0)$ to $f'(x)$ to emphasize that x is the variable of the function $f'(x)$

S: What motivates us to look at the slope function?

T: It turns out that the slope function $f'(x)$ characterizes important features of the original function $f(x)$. Figure 8 shows a general curve with some tangent lines at a few points on the curve. Pay attention on how the movement of the slope function $f'(x)$ is associated to the movement of the curve itself.

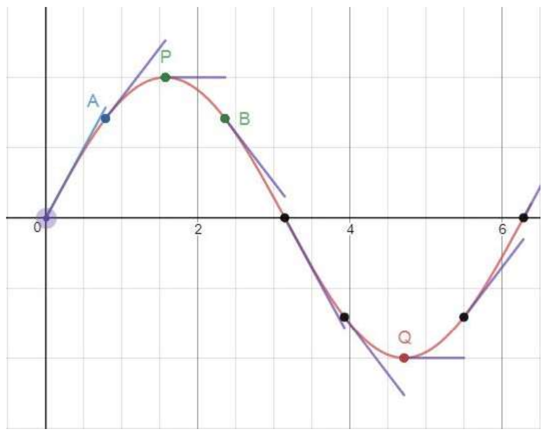

FIG 8. *The movement of the tangent lines along the curve*

S: I can see that the direction of tangent line keeps changes. There are two special points P, Q where the tangent line become horizontal. Other than that, the tangent lines can be upward or downward depending on whether $f(x)$ is increasing or decreasing. seems that the slopes tell how the curve moves in some way.

T: A sharp observation! Especially, you see three status of tangent lines: horizontal, upward and downward. More specifically, $k = 0$ for horizon tangent, $k > 0$ for upward tangent line and $k < 0$ for downward tangent line. Notice that the function reaches the maximum value or the minimum value at P, Q with $f'(P) = f'(Q) = 0$. Since $f'(x)$ is quite important for us to study $f(x)$, let us assume that

Assumption 2.1. *For the rest of the section, $f(x)$ is a general function over $[a, b]$ whose tangent line exists at any point on the curve $y = f(x)$.*

Remark 6. *Notice that the slope does not exist for $f(x) = |x|$ at $x = 0$.*

We have following observations:

Theorem 2.3. *Let $f(x)$ be as in Assumption 2.1.*

1. *If $x_0 \in (a, b)$ is a point inside the interval and $f(x)$ reaches maximum (or minimum) value at x_0 in a neighborhood $O(x_0, r) \subset [a, b]$ for some $r > 0$, i.e.*
$$f(x) \leq (\geq) f(x_0), \quad \text{for all } x \in (x_0 - r, x_0 + r), \tag{2.13}$$
then $f'(x_0) = 0$.

2. (**Mean Value Theorem**) *For any two points* $x_1, x_2 \in [a,b]$, *there is a point* x_0 *between* x_1 *and* x_2 *such that*

$$f'(x_0) = \frac{f(x_2) - f(x_1)}{x_2 - x_1}, \tag{2.14}$$

i.e. the line connecting $(x_1, f(x_1))$ *and* $(x_2, f(x_2))$ *is parallel to a tangent line at some point on the curve between those two points. See Figure 9.*
3. *If* $f'(x) \geq (\leq) 0$ *holds on an interval* $[x_1, x_2] \subset [a,b]$, *then* $f(x)$ *is increasing (decreasing) over* $[x_1, x_2]$. *In particular, if* $f'(x) \equiv 0$ [17] *over* $[x_1, x_2]$, *then* $f(x)$ *is constant over* $[x_1, x_2]$.
4. $f(x)$ *is constant if and only if* $f'(x) \equiv 0$. *As a corollary,* $f(x) \equiv g(x) + C$ *for some constant if* $f'(x) \equiv g'(x)$.

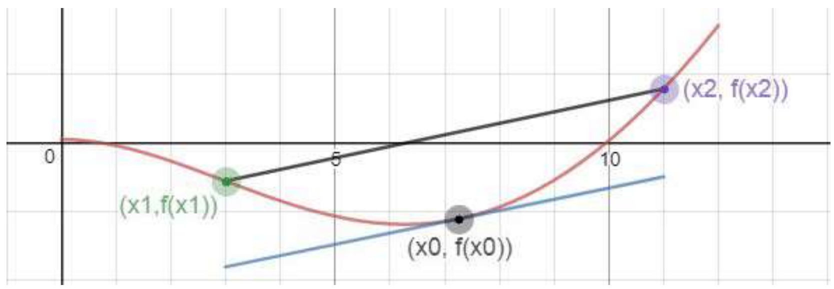

FIG 9. *The mean value*

S: They look right to me. But I guess that is not sufficient for a mathematician.

T: Let us try to prove them since the details are helpful for us to understand the concept.

Proof. 1. If $f(x)$ reaches the maximum value at x_0 over the neighborhood $(x_0 - r, x_0 + r)$, then

$$f(x) - f(x_0) \leq 0, \qquad \forall x \in (x_0 - r, x_0 + r),$$

and therefore

$$\frac{f(x) - f(x_0)}{x - x_0} \geq 0, \qquad x_0 - r < x < x_0, \tag{2.15}$$

$$\frac{f(x) - f(x_0)}{x - x_0} \leq 0, \qquad x_0 < x < x_0 + r. \tag{2.16}$$

Inequity (2.15) implies $f'(x_0) \geq 0$ and Inequity (2.16) implies $f'(x_0) \leq 0$, which implies that $f'(x_0) = 0$[18]. Similarly, one can prove the case where $f(x)$ reaches the minimum value at x_0.

[17] We write $f(x) \equiv g(x)$ if two functions are identical. In general, for two functions $f(x)$ and $g(x)$, $f(x) = g(x)$ is a equation that might have some values as the solutions, but f and g are generally not equal to each other everywhere.

[18] Notice that we aim to show $f'(x_0) = 0$ under the assumption $f'(x_0)$ exists. The conclusion becomes trivial once the derivative is further clarified in Section 6.

2. Let us start with the special case $f(x_1) = f(x_2)$. Equation (2.14) is reduced to $f'(x_0) = 0$ at some x_0 over $[a, b]$. Assuming $f(x)$ reaches maximum and minimum at c and d respectively over $[x_1, x_2]$[19]. Unless $f(x)$ is constant over $[x_1, x_2]$, $f(c) > f(d)$ and therefore one of c, d need to be inside the interval $[x_1, x_2]$ since $f(x_1) = f(x_2)$. Without loss of generosity, assuming $x_0 = c \in (x_1, x_2)$, which implies $f'(x_0) = 0$ by part 1. For general case, let $h(x) = k(x - x_1) + f(x_1)$ be the line that connects $(x_1, f(x_1))$ and $(x_2, f(x_2))$ with $k = \frac{f(x_2)-f(x_1)}{x_2-x_1}$; define

$$F(x) := f(x) - h(x) = f(x) - k(x - x_1) - f(x_1).$$

It is clear that $F(x_1) = F(x_2) = 0$ and therefore there exists a point x_0 such that $F'(x_0) = 0$. Applying Eq (2.12) and Remark 5, we have

$$0 = F'(x_0) = f'(x_0) - k = f'(x_0) - \frac{f(x_2) - f(x_1)}{x_2 - x_1}$$

which implies Equation (2.14).
3. Let $f'(x) \geq 0$ hold over $[a, b]$. For any two points $x_1 < x_2$ in $[a, b]$, by Property 2, there exist x_0 in $[x_1, x_0] \subset [a, b]$ such that Eq (2.14) holds, which implies $f(x_2) \geq f(x_1)$ since $f'(x_0) \geq 0$.
4. it is clear that $f'(x) \equiv 0$ if $f(x)$ is constant. On the other hand, if $f'(x) \equiv 0$, then $f(x)$ is increasing and decreasing simultaneously by Part 3, which implies that f is constant.

□

S: The proof for Theorem 2.3 (2) is interesting. The special case $f(x_1) = f(x_2)$ is not hard to prove. The general statement can actually be simplified to the special case by transferring $f(x)$ to $F(x)$.

T: Yes. as a principal, it is good idea to start with simple case to get some idea in solving hard questions. The observation on special situation can often be applied to a general setting by certain transformation. Actually, we can further generalize Eq 2.14 as follows and you are encouraged to figure out the proves before looking the footnote [20].

---

[19] This need to be proved strictly speaking, which is left in Section 4.

[20]

Proof. Define

$$F(x) = f(x) - f(a) - \frac{f(b) - f(a)}{g(b) - g(a)}[g(x) - g(a)]$$

It is clear that $F(a) = F(b) = 0$, hence there exists $c \in (a, b)$ such that $F'(c) = 0$, equivalently

$$f'(c) - \frac{f(b) - f(a)}{g(b) - g(a)} g'(c) = 0$$

implying (2.17) since $g'(c) \neq 0$ by the assumption. □

Theorem 2.4. Cauchy Theorem *If $f'(x)$ and $g'(x)$ exist on $[a, b]$ and $g'(x) \neq 0$, then there exist $c \in (a, b)$ such that*

$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(c)}{g'(c)} \tag{2.17}$$

S: Looks that slope is indeed quite useful.

T: It is so important that people coin a specific term for it.

Definition 2.5. *Let $y = f(x)$ be a curve. The slope $f'(x)$ of the function, if it exists, is called the* derivative *of $f(x)$.*

S: Why not just call it slope?

T: If you are only interested in geometry, slope is the right term. How about if x represents time, $f(x)$ represents the position an object moves along a line?

S: Hmmm...?

T: Remind you that we start with approximation of slope by (2.11).

S: Let me see. If x_0 and x_1 represent two time stamps, $f(x_0)$ and $f(x_1)$ are the positions of the object at those two moments, then $f(x_1) - f(x_0)$ represent the distance it travels over the period from x_0 to x_1, so the ratio represent the average speed. Let x_1 approach x_0, the slope should refer to the instantaneous speed at time x_0.

T: Exactly! You see that slope is not proper word for other applications. However, you can still treat derivative simply as slope if it helps for an intuitive understand.

S: I guess that how to calculate effectively derivatives must be another major task for calculus!

T: You just said it! Let us work on the general polynomial $f(x)$ defined in Eq (2.3). According to (2.12), we have

$$f'(x) = na_n x^{n-1} + \cdots + 2a_2 x + a_1 \tag{2.18}$$

S: Cool. I have to admit that polynomial at degree n is not easy to me. Somehow magically, you makes it pretty trivial.

T: Property (2.12) is the key. It says that if we can decompose a complicated function into the summation of simple ones, then we only need to work on each of the terms separately. We shall develop more tools to find derivatives of sophisticated functions. But you are right, we need

Task 2.2. *Develop tools to calculate the derivatives of wide range of functions that are enough for real world applications.*