Chapter 1

The Rate of Change of a Function

1.1 Introduction

7/3:
- Discusses the importance of calculus, when it should be used, and why one should study it.
- **Analytic geometry**: "Uses algebraic methods and equations to study geometric problems. Conversely, it permits us to visualize algebraic equations in terms of geometric curves" (Thomas, 1972, p. 2).

1.2 Coordinates

- "The basic idea in analytic geometry is the establishment of a one-to-one correspondence between the points of a plane on the one hand and pairs of numbers (x, y) on the other hand" (Thomas, 1972, p. 2).
- Such a correspondence is most commonly established as follows.

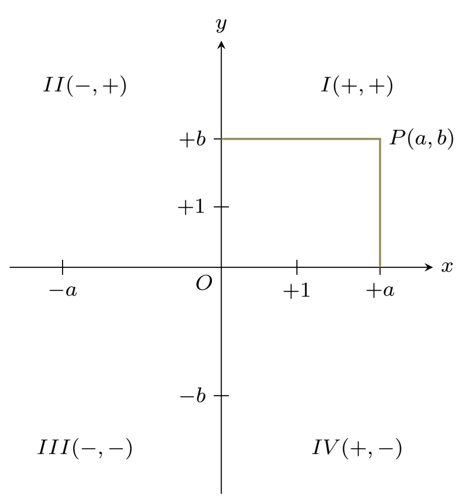


Figure 1.1: Cartesian coordinates.

  - "A horizontal line in the plane, extending indefinitely to the left and to the right, is chosen as the x-axis or axis of **abscissas**. A reference point O on this line and a unit of length are then chosen. The axis is scaled off in terms of this unit of length in such a way that the number zero is attached to O, the number $+a$ is attached to the point which is a units to the right of O, and $-a$ is attached to the symmetrically located point to the left of O. In this way, a one-to-one correspondence is established between points of the x-axis and the set of all **real numbers**" (Thomas, 1972, pp. 2–3).

– "Now through O take a second, vertical line in the plane, extending indefinitely up and down. This line becomes the y-axis, or axis of **ordinates**. The unit of length used to represent $+1$ on the y-axis need not be the same as the unit of length used to represent $+1$ on the x-axis. The y-axis is scaled off in terms of the unit of length adopted for it, with the positive number $+b$ attached to the point b units above O and negative number $-b$ attached to the symmetrically located point b units below O" (Thomas, 1972, p. 3).

– "If a line parallel to the y-axis is drawn through the point marked a on the x-axis, and a line parallel to the x-axis is drawn through the point marked b on the y-axis, their point of intersection P is to be labeled $P(a, b)$. Thus, given the pair of real numbers a and b, we find one and only one point with abscissa a and ordinate b, and this point we denote by $P(a, b)$" (Thomas, 1972, p. 3).

– "Conversely, if we start with any point P in the plane, we may draw lines through it parallel to the coordinate axes. If these lines intersect the x-axis at a and the y-axis at b, we then regard the pair of numbers (a, b) as corresponding to the point P. We say that the coordinates of P are (a, b)" (Thomas, 1972, p. 3).

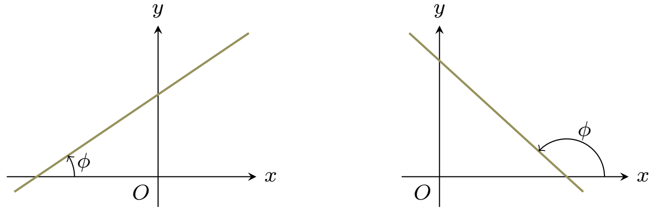
– "The two axes divide the plane into four quadrants, called the first quadrant, second quadrant, and so on, and labeled I, II, III, IV in [Figure 1.1]. Points in the first quadrant have both coordinates positive, and in the second quadrant the x-coordinate (abscissa) is negative and the y-coordinate (ordinate) is positive. The notations $(-, -)$ and $(+, -)$ in quadrants III and IV of [Figure 1.1] represent the signs of the coordinates of points in these quadrants" (Thomas, 1972, p. 3).

1.3 Increments

- **Increments**: The values $\Delta x = x_2 - x_1$ and $\Delta y = y_2 - y_1$ concerning a particle, the initial position of which is $P_1(x_1, y_1)$ and the terminal position of which is $P_2(x_2, y_2)$.
- If the unit of measurement for both axes is the same, then we may express distances in the plane in terms of this unit using the Pythagorean theorem.

1.4 Slope of a Straight Line

- Let L be a straight line not parallel to the y-axis intersecting distinct points $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$. Then L has a **rise**, **run**, and **slope**.
- **Rise**: The increment Δy.
- **Run**: The increment Δx.
- **Slope**: The rate of rise per run $m = \frac{\text{rise}}{\text{run}} = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}$. *Also known as* **inclination**.



(a) Negative x-intercept. (b) Positive x-intercept.

Figure 1.2: The slope and the angle of inclination.

  - If we chose different distinct points, the slope would be same because the triangles in the Cartesian plane would be similar.
  - Δy is proportional to Δx with m as the proportionality factor.
  - On interpolation: If we're given the values of a function at (x_1, y_1) and (x_2, y_2), then we may approximate the function by a straight line L passing through those two points and approximate the value $f(x)$ for any $x_1 \leq x \leq x_2$.
  - If the scales on both axes are equal, then the slope of L is equal to the tangent of the **angle of inclination** that L makes with the positive x-axis. That is, $m = \tan\phi$ (see Figure 1.2).
- **Parallel** (lines): Two lines with equal inclinations ($m_1 = m_2$).
- **Perpendicular** (lines): Two lines with inclinations that differ by 90° ($m_1 = -\frac{1}{m_2}$).
  - Note that we can prove the relation between the slopes using the angles of inclination as follows.

$$\begin{aligned}
m_1 &= \tan\phi_1 \\
&= \tan(\phi_2 + 90^\circ) \\
&= -\cot\phi_2 \\
&= -\frac{1}{\tan\phi_2} \\
&= -\frac{1}{m_2}
\end{aligned}$$

1.5 Equations of a Straight Line

7/5:
- How do you know if $P(x, y)$ is a point on the line P_1P_2 through distinct points $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$?
  - If $x_1 = x_2$, then P_1P_2 is vertical and P lies on P_1P_2 iff $x = x_1$.
  - If $x_1 \neq x_2$, then the slope of P_1P_2 $m_{P_1P_2} = \frac{y_2 - y_1}{x_2 - x_1}$. Thus, P lies on P_1P_2 iff $P = P_1$ or, for the line PP_1 through P and P_1, $m_{P_1P_2} = m_{PP_1} = \frac{y - y_1}{x - x_1}$. In other words, the coordinates x, y of P must satisfy $y - y_1 = m_{P_1P_2}(x - x_1)$.
- Thomas (1972) calls the above equation the **point-slope form**.
- **Variable**: "A symbol, such as x, which may take on any value in some set of numbers" (Thomas, 1972, p. 10).
- **Slope-intercept form**: $y = mx + b$.
- **General form**: $Ax + By + C = 0$.
  - Such an equation (one that contains only first powers of x and y and constants) is said to be **linear in x and y**.
  - "Every straight line in the plane is represented by a linear equation and, conversely, every linear equation represents a straight line" (Thomas, 1972, p. 10).
- **y-intercept**: The constant b in the above equation.
- Let L be a line with the equation $Ax + By + C = 0$. The shortest distance d from a point $P_1(x_1, y_1)$ not on L to L is given by

$$d = \frac{|Ax_1 + By_1 + C|}{\sqrt{A^2 + B^2}}$$

  - Derive by finding a line perpendicular to L through P_1.

1.6 Functions and Graphs

- **Domain** (of a variable x): "The set of numbers over which x may vary" (Thomas, 1972, p. 12).
- Defines **open intervals**, **half-open intervals**, and **closed intervals**.

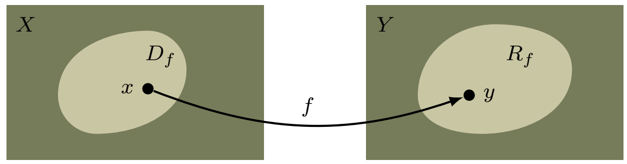


Figure 1.3: A function f maps the domain D_f onto the range R_f. The image of x is $y = f(x)$.

- **Function**: For two nonempty sets X, Y, the collection f of ordered pairs (x, y) with $x \in X$ and $y \in Y$ that assigns to every $x \in X$ a unique $y \in Y$. *Also known as* **mapping** (from X to Y), $\boldsymbol{y = f(x)}$, $\boldsymbol{f : x \to y}$[1].
  - When using the latter notation, it is understood that the domain is $\mathbb{R}$ unless this is impossible (e.g., $f : x \to \frac{1}{x}$ must exclude 0 from the domain).
- **Domain** (of a function f): "The collection of all first elements x of the pairs (x, y) in f" (Thomas, 1972, p. 13). *Also known as* $\boldsymbol{D_f}$.
- **Range** (of a function f): "The set of all second elements y of the pairs (x, y) in f" (Thomas, 1972, p. 13). *Also known as* $\boldsymbol{R_f}$.
- **Image** (of x): The value y to which a function maps x.
- Thomas (1972) considers functions from the reals to the reals, but also more abstract functions.
  - For example, it considers the function from all triangles (a set of decidedly nonnumerical objects) to their enclosed areas (the set of positive real numbers).

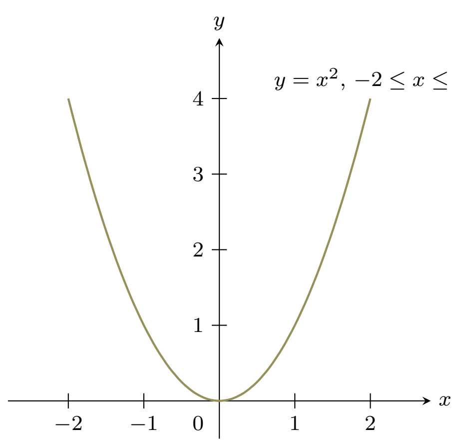


Figure 1.4: Graph of a function.

- **Graph** (of a function): "The set of points which correspond to members of the function" (Thomas, 1972, p. 14).

[1] "eff sends ex into wy"

  - For example, let X be the closed interval $[-2,2]$. To each $x \in X$, assign the number x^2. This describes the function
$$f = \left\{(x,y) : -2 \leq x \leq 2,\ y = x^2\right\}$$
    The graph of f can be seen in Figure 1.4.

- **Independent variable**: The first variable x in the ordered pair (x,y). *Also known as* **argument**.
- **Dependent variable**: The second variable y in the ordered pair (x,y).
- **Real-valued function of a real variable**: "A function f whose domain and range are sets of real numbers" (Thomas, 1972, p. 14).
  - As a general rule, *function* indicates a real-valued function of a real variable for the first seven chapters of Thomas (1972).
- f can be represented by...
  - A table of corresponding values (this will be incomplete, though).
  - Corresponding numerical scales, as on a slide rule (this will be incomplete, though).
  - A simple formula, such as $f(x) = x^2$ (this may be less exact than ordered pairs, but it is more easily understood/applicable/complete).
  - A graph (for any value x in the domain, begin x units from the origin along the x-axis, move vertically until intersecting the curve, and then move horizontally until intersecting the image y on the y-axis).
- Some mappings cannot be expressed in terms of algebraic operations on x.
  - For example, the **greatest-integer function** "maps any real number x onto that unique integer which is the largest among all integers that are less than or equal to x" (Thomas, 1972, p. 15).
    - The image of x is represented by $[x]$, and the function by $f : x \to [x]$.
    - An example of a **step function**.
    - It exhibits points of **discontinuity**.
- Note: The fact that a one-to-one mapping exists between the points in the interval $(0,1]$ and $[1,\infty)$ (namely, $f : x \to \frac{1}{x}$) proves that there are equally many points in both intervals.
- The absolute value function can be geometrically interpreted in the context of distance from a point. As such, it is useful in describing **neighborhoods**.

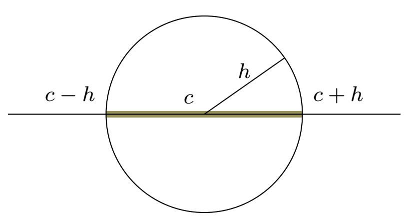


Figure 1.5: The symmetric neighborhood $N_h(c)$, centered at c, with radius h.

- **Symmetric neighborhood** (of a point c): "The open interval $(c-h, c+h)$, where h may be any positive number" (Thomas, 1972, p. 17). *Also known as* $\boldsymbol{N_h(c)}$.
- **Radius** (of a symmetric neighborhood): The value h (see Figure 1.5).
- **Neighborhood** (of a point c): The open interval $(c-h, c+k)$, where h, k may be any positive numbers.
  - Like requiring that $|x-c|$ is small.

- **Deleted neighborhood** (of a point c): "A neighborhood of c from which c itself has been removed" (Thomas, 1972, p. 17).
  - Like requiring that $|x-c|>0$.
- **Intersection** (of the neighborhoods $(c-h_1, c+k_1)$ and $(c-h_2, c+k_2)$): The neighborhood $(c-h, c+k)$, where $h=\min(h_1,h_2)$ and $k=\min(k_1,k_2)$.
  - "The intersection of two neighborhoods of c is a neighborhood of c, and the intersection of two deleted neighborhoods of c is a deleted neighborhood of c" (Thomas, 1972, p. 18).
- Let A be a neighborhood of c. Then denote the deleted neighborhood equivalent to A with c removed by A^-.

7/6:

- "A function is determined by the domain and by any rule that tells what image in the range is to be associated with each element of the domain" (Thomas, 1972, p. 18).
  - Thus, we can think of a "function machine" that takes in elements of the domain and computes the image based on the rule.
  - Thomas (1972) visualizes function machines as flow charts.
  - In theory, a function machine could store every pair (x,y) in its memory to be recalled later. Since machines have limited memory, a fractional set of pairs could also be stored and the values in between calculated by interpolation.
  - In practice, though, calculating as we go is usually best.
- Two restrictions should be inferred to apply to the domain of a function, even if they are unstated: First, never divide by 0. Second, do not consider complex outputs (yet).
- Sometimes we have functions of more than one independent variable.
  - For example, the volume $v=\frac{1}{3}\pi r^2 h$ of a right circular cone is uniquely determined only when r,h are given definite, positive, nonzero values.
  - "Its domain is the set of all pairs (r,h) with $r>0$, $h>0$. Its range is the set of positve numbers $v>0$" (Thomas, 1972, p. 20).
  - r,h are independent variables. v is a dependent variable.
- "More generally, suppose that some quantity y is uniquely determined by n other quantities $x_1,x_2,x_3,\dots,x_n$. The set of all ordered $(n+1)$-tuples $(x_1,x_2,x_3,\dots,x_n,y)$ that can be obtained by substituting permissible values of the variables $x_1,x_2,\dots,x_n$ and the corresponding values of y is a function whose domain is the set of all allowable n-tuples $(x_1,x_2,x_3,\dots,x_n)$ and whose range is the set of all possible values of y corresponding to this domain. If values can be assigned independently to each of the x's, we call them independent variables and say that y is a function of the x's. We also write $y=f(x_1,x_2,x_3,\dots,x_n)$ to indicate that y is a function of the n x's, just as we write $y=f(x)$ to indicate that y is a function of one independent variable x" (Thomas, 1972, p. 20).
  - Note, though, that functions of a single variable will be the primary concern of this book.
- **Signum function**: The function

$$\operatorname{sgn} x = \begin{cases} -1 & x<0 \\ 0 & x=0 \\ 1 & x>0 \end{cases}$$

1.7 Ways of Combining Functions

7/7:

- The domains of the sum, product, difference, and quotient of two functions f and g are the intersections of D_f and D_g.
  - Note that for the quotient, we must also exclude points where $g(x) = 0$.
- There is a distinction between $x \cdot \frac{1}{x}$ and 1 (namely, the fact that the latter includes 0 in its domain while the former excludes it).
- Translation.

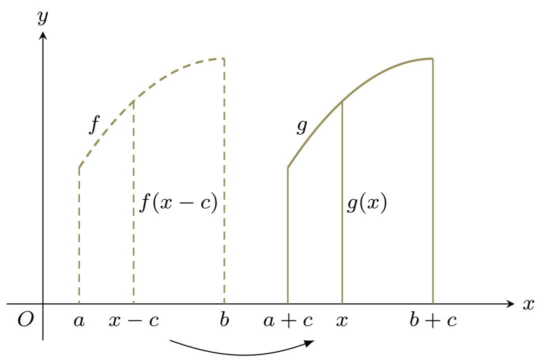
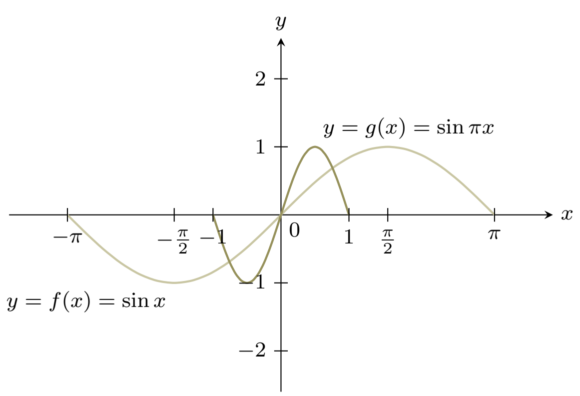
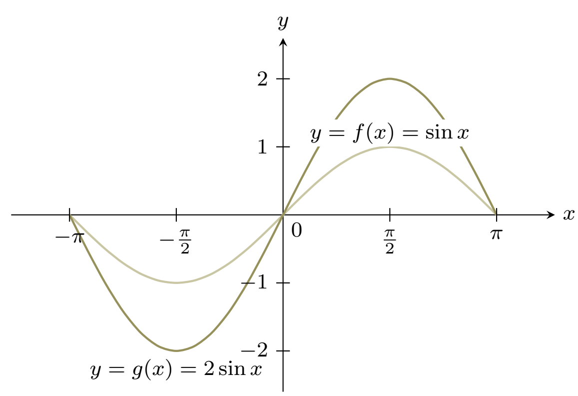


Figure 1.6: Translation.

  - "Take the ordinate of f at $x-c$ and shift it to the right c units to get the ordinate of g at x" (Thomas, 1972, p. 23).
  - $g(x) = f(x-c)$ implies that the graph of g is that of f translated c units to the right.
  - If $D_f = [a,b]$ ($f(x)$ is only defined when $a \leq x \leq b$), then $f(x-c)$ is only defined when $a \leq x-c \leq b$. This implies that $g(x)$ is only defined when $a+c \leq x \leq b+c$; hence, $D_g = [a+c, b+c]$.
- Change of scale.



(a) Horizontal scaling.



(b) Vertical scaling.

Figure 1.7: Change of scale.

  - Suppose $g(x) = f(kx)$. Then to transform the graph of f into that of g, "compress or stretch the x-axis by shrinking (if $k > 1$) or stretching (if $k < 1$) every interval of length k on the x-axis into an interval of length 1" (Thomas, 1972, p. 24).

    - Let $g(x) = f(kx)$. If $D_f = [a, b]$ ($f(x)$ is only defined when $a \leq x \leq b$), then $f(kx)$ is only defined when $a \leq kx \leq b$. This implies that $g(x)$ is only defined when $a/k \leq x \leq b/k$; hence, $D_g = [a/k, b/k]$.
    - Be wary when $k = 0$.
  - Suppose $g(x) = k \cdot f(x)$. Then to transform the graph of f into that of g, "stretch the f curve vertically (if $k > 1$), or compress it (if $k < 1$) [or] change the scale on the y-axis so that points labeled $1, 2, 3, \ldots$ for the graph of f are relabled $k, 2k, 3k, \ldots$ for the graph of [g]" (Thomas, 1972, p. 25).
    - This kind of stretching may cause R_g to differ from R_f by some factor k, but it will not affect D_f and D_g.

1.8 Behavior of Functions

- Linear functions (refer to Figure 1.8 throughout the following).

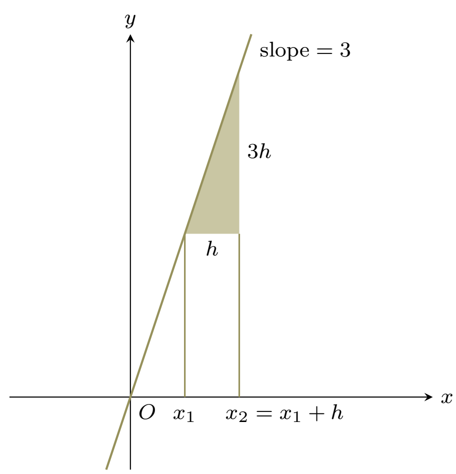


Figure 1.8: Behavior of linear functions.

  - For example, let $f(x) = 3x$. The graph of f is a straight line through the origin with slope $+3$.
  - Let x_1 be an initial value of x, and let $x_2 = x_1 + h$ be a new value of x obtained by increasing x_1 by h units[2]. The corresponding increase in $f(x)$ is given by

$$f(x_1 + h) - f(x_1) = 3(x_1 + h) - 3x_1 = 3h$$

    Thus, we see that f everywhere changes three times as fast as x.

- Quadratic functions (refer to Figure 1.9 throughout the following).
  - Unlike linear functions, quadratic functions do not everywhere have a constant rate of change.
  - Let $f(x) = x^2$. In an analogous manner to the previous example, we find that the change between x_1 and $x_2 = x_1 + h$ is

$$f(x_1 + h) - f(x_1) = (x_1 + h)^2 - x_1^2 = 2x_1h + h^2$$

    Thus, we see that the rate of change of f is dependent on both the initial value of x and the amount of increase in x.

[2]Note that h is an alternative notation for Δx.

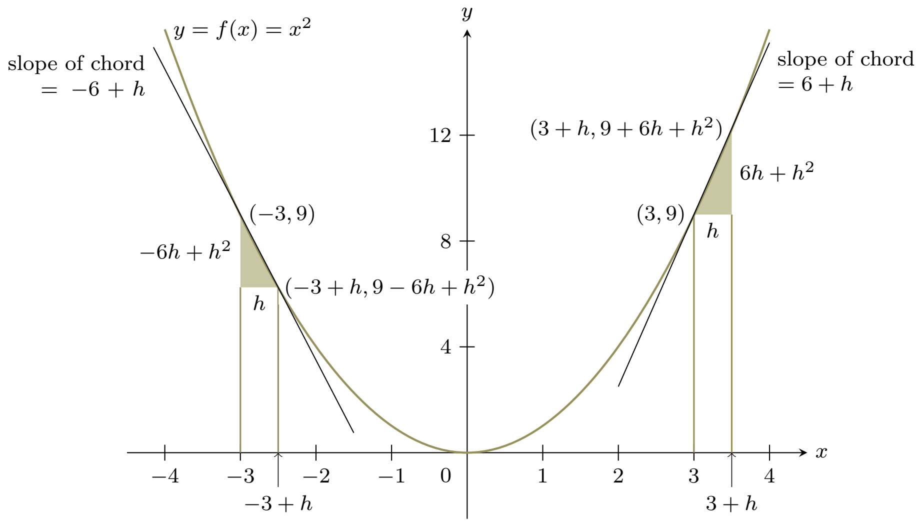


Figure 1.9: Behavior of quadratic functions.

- While a linear function increases at a rate directly proportional to the increase in x, the above demonstrates that "the increase in x^2, as x increases from x_1 to $x_1 + h$, is $2x_1 + h$ times the increase in x" (Thomas, 1972, p. 27).
- We can now define the **average rate of increase**.
- Thus, the average rate of increase of x^2 is $2x_1 + h$.
- Now let's see what happens as h shrinks.

		x_1			
		2	3	-2	-3
h	1	5	7	-3	-5
	0.5	4.5	6.5	-3.5	-5.5
	0.25	4.25	6.25	-3.75	-5.75
	0.1	4.1	6.1	-3.9	-5.9
	0.01	4.01	6.01	-3.99	-5.99
	0.001	4.001	6.001	-3.999	-5.999

Table 1.1: Average rate of change of x^2 versus h.

- From Table 1.1, we see that smaller values of h cause the average rate of change to tend toward $2x_1$. This is the beginning of **differential calculus**.

- **Average rate of increase** (of $f(x)$, per unit of increase in x, from x_1 to $x_1 + h$): The ratio,

$$\frac{f(x_1 + h) - f(x_1)}{(x_1 + h) - x_1} = \frac{f(x_1 + h) - f(x_1)}{h} = \frac{\text{change in } f(x)}{\text{change in } x}$$

- **Differential calculus**: The branch of calculus concerned with the *instantaneous* rate of increase of a function, as opposed to the *average* rate of increase.
- So constant and linear functions are easy to analyze. But for more complicated functions, we need more advanced tools.
- Let's begin exploring the instantaneous rate of change, continuing with the parabola example.

  - For x^2, the average rate of change is given by $\frac{f(x_1+h)-f(x_1)}{h} = 2x_1 + h,\ h \neq 0$.
  - The $h \neq 0$ is critical — we wish to consider the case where $h = 0$, but we cannot. However, we can consider values of the slope function $m(h) = 2x_1 + h$ in a deleted neighborhood of $h = 0$. By decreasing the radius of the neighborhood, we can get progressively closer to analytically approximating $m(0)$.

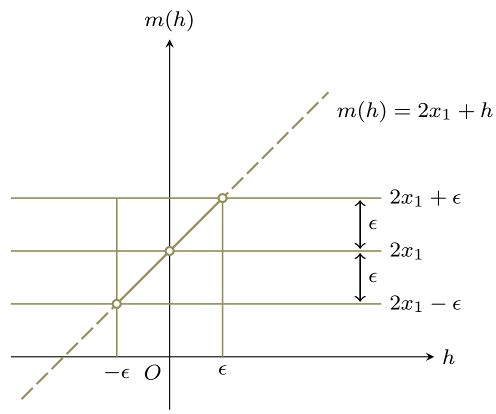


Figure 1.10: Deleted neighborhood of a slope function.

  - From Figure 1.10, we can see that $m(h)$ is bounded between $2x_1 + \epsilon$ and $2x_1 - \epsilon$ when $|h|$ is less than ϵ (or any positive number smaller than ϵ)[3].
  - In fact, $0 < |h| < \epsilon$, or $-\epsilon < h < \epsilon$, directly implies $2x_1 - \epsilon < 2x_1 + h < 2x_1 + \epsilon$.
  - We can now formally define **approximation**, such as what was just described.

- **Approximation** (of $f(x)$ by L to within ϵ on (a, b)): The value L approximates $f(x)$ to within ϵ on the interval (a, b) if $L - \epsilon < f(x) < L + \epsilon$ when $a < x < b$.

1.9 Slope of a Curve

- **Slope of the curve** (at P): The limiting value of the slope of the secant between distinct points P, Q on the curve $y = f(x)$ as Q moves along the curve progressively closer to P. *Also known as* **slope of the tangent to the curve** (at P).
  - A purely geometric definition also exists: "Let C be a curve and P a point on C. If there exists a line L through P such that the measure of one of the angles between L and the secand line PQ approaches zero as Q approaches P along C, then L is said to be tangent to C at P" (Thomas, 1972, p. 30).
  - An advantage of the geometric definition is that it does not depend on the coordinate axes and allows vertical lines.
  - However, in most cases, we will stick with the algebraic definition.
- Thomas (1972) considers the average rate of increase equation for a cubic function, informally allowing Δx to tend towards 0 to derive a slope function.

1.10 Derivative of a Function

- We now formalize our notion of a slope function.

[3]Note that ϵ is used to denote an arbitrary (often arbitrarily small) positive number.

  - We know that the slope m_{sec} of the secant from $P(x,y)$ to a point on the curve $y = f(x)$ at $(x+\Delta x, f(x+\Delta x))$ is given by

$$m_{\text{sec}} = \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

  - Now as Δx tends toward 0, m_{sec} tends toward the slope $m_{\tan}$ of the tangent at P. The mathematical symbols which summarize this discussion are

$$m_{\tan} = \lim_{Q \to P} m_{\text{sec}} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

- The number given by the last operation in the above equation is clearly related to f. Thus, to indicate relation, we define

$$\boldsymbol{f'(x)} = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

  Also known as $\boldsymbol{y'}$, $\boldsymbol{\frac{\mathrm{d}y}{\mathrm{d}x}}$, $\boldsymbol{D_x y}$.

- This limit may sometimes fail to exist. However, at each point where it does exist, f is said to have a **derivative**, or to be **differentiable**. Similarly, $f'(x)$ is said to be the **derivative** (of f at x).
- Differential calculus is concerned with two problems.
  1. "Given a function f, determine those values of x (in the domain of f) at which the function possesses a derivative" (Thomas, 1972, p. 32).
  2. "Given a function f and an x at which the derivative exists, find $f'(x)$" (Thomas, 1972, p. 32).
- **Derived function**: "The set of all pairs of numbers $(x, f'(x))$ that can be formed by this process" (Thomas, 1972, p. 33). *Also known as* **derivative** (of f).
  - "The domain of f' is a subset of the domain of f" (Thomas, 1972, p. 33).
  - Symbolically, $D_{f'} \subset D_f$. However, for most functions considered in this book, $D_{f'} = D_f$ with maybe a few exceptions.
- On computing $f'(x)$ by eliminating the division by 0 and then substituting: "We may say that after the division by Δx has been carried out and the expression has been reduced to a form... which 'makes sense' (that is, does not involve division by zero) when Δx is taken equal to zero, then the limit as Δx approaches zero does exist and may be found by simply replacing Δx by zero in this reduced form" (Thomas, 1972, p. 34).
- Essentially, what we are doing when we eliminate the division by zero is we are expanding the domain of the function, the limit of which we are taking, to include a point of interest (Thomas (1972) elaborates quite a bit on this point).

1.11 Velocity and Rates

- Mainly just applies average and instantaneous rates of change to the physical problem of distance and velocity. However...
- "Derivatives are important in economic theory, where they are usually indicated by the adjective **marginal**" (Thomas, 1972, p. 37).
  - "Suppose that in order to produce $x + \Delta x$ tons of steel weekly, it would cost $y + \Delta y$ dollars. The increase in cost per unit increase in output would be $\Delta y / \Delta x$. The limit of this ratio, as Δx tends to zero, is called the **marginal cost**" (Thomas, 1972, p. 37).
  - There also exists **marginal revenue** $\mathrm{d}P/\mathrm{d}x$ and **marginal profit** $\mathrm{d}T/\mathrm{d}x$.

- Note: "The *average* rate of change of y per unit change in x, $\Delta y/\Delta x$, when multiplied by the number of units change in x, Δx, gives the actual change in y:

  $$\Delta y = \frac{\Delta y}{\Delta x}\,\Delta x$$

  The *instantaneous* rate of change of y per unit change in x, $f'(x)$, multiplied by the number of units change in x, Δx, gives the change that would be produced in y if the point (x, y) were to move along the tangent line instead of moving along the curve; that is

  $$\Delta y_{\mathrm{tan}} = f'(x)\,\Delta x$$

  One reason calculus is important is that it enables us to find quantitatively how a change in one of two related variables affects the second variable" (Thomas, 1972, p. 38).