Calculus One

1 From Counting to measuring

1.1 Natural Numbers

Number is the **name** given to a quantity. Not every aggregate of individuals can have a name all to itself since there would be no end to the vocabulary. We must resort to some system of compounding for the names of large collections. Ancient Hebrew has preserved the descriptive meaning of all those number-names. Biblical Hebrew calls one 'echad' – united, it calls four 'arba' – varied, it calls seven 'sheva' – plentiful or abundant, it calls eight 'shmone' – full or big or fat; and it calls ten 'eser' – rich. Possibly because the the ancient numbers, at least the large ones like ten, were imprecise. The only Hebrew number-name that can be tentatively traced to the human hand is five which it calls 'chamesh' possibly meaning the cluster of fingers in the clenched fist, lending credence to the belief that the preeminence of the decimal system has to do more with our innate inability to distinguish and correctly name on sight quantities larger than ten, rather than the number of fingers on both our hands.

Hebrew calls one hundred 'mea' – collected, one thousand 'elef' – superior, elevated, or elephant-like, and ten thousand 'revava' – a riven multitude. These names were in spoken and literary use thousands of years before the modern introduction of the arithmetical symbols 1,2,3,4,5,6,7,8,9, which together with the zero, written as 0, make up the present **positional** method for writing any number.

Algebra expresses universal and generally truthful facts about the combination of quantities, or numbers, by assigning a letter of the alphabet to designate an arbitrary number. Different letters refer to different numbers. We recognize that the arithmetic of numbers obeys the following intuitively correct rules of reckoning:

(1) Addition is **commutative**, namely $a + b = b + a$ for any pair of numbers a, and b.

(2) Addition is **associative**, namely $(a + b) + c = a + (b + c)$ for any numbers a, b, c.

(3) Multiplication is **commutative**, namely $ab = ba$ for any pair of numbers a, and b.

(4) Multiplication is **associative**, namely $(ab)c = a(bc)$ for any numbers a, b, c.

(5) Multiplication is **distributive** with respect to addition, namely $a(b + c) = ab + ac$.

(6) For any nummber a it is true that $1a = a$.

Notice that we may write ab instead of the full $a \cdot b$ since algebra, contrary to English, refrains from using more than one letter to denote a number.

English has proper names for all numbers from 1 to 10. For 20 it reserved the special name score, for 100 the special name hundred, namely ten pairs of hands, and for 1000 the special name thousand. English does not have a special name for ten thousand but it has a million and a billion.To this day the word-number and the symbol-number coexist in separate peaceful lives. The assertion that God is one, 'echad' or unified in Hebrew, is never God is 1; possibly the first but never the 1st. Similarly William Shakespeare's Twelfth Night is never 12th Night and Alexandre Dumas' The Three Musketeers is never The 3 Musketeers.

1.2 The Zero

Formally, zero is induced into the family of numbers by the definition that for any number a, $a + 0 = a$. It results that zero has this singular property that $a0 = 0$ for any a. Indeed, since $0 + 1 = 1$, and since $a(0 + 1) = a0 + a1$, then $a1 = a0 + a1$, or $a = a0 + a$ implying that $a0 = 0$.

1.3 Fractions or Rational Nubmers

It is also evident that from the earliest of times man used to look upon certain collections or sets of objects as being a unit – again the Hebrew 'echad' – and to speak of portions thereof as a half, a third, a quarter and so on, and these words existed in the spoken and written language long before they were rendered the modern arithmetical form of one-over-two, 1/2, one-over-three, 1/3, etc. Hebrew calls one-half 'chetzi' – a divide or a cut.

A lump of meat can be cut into arbitrary portions, a stick can be chopped into arbitrary sections, a string can be marked at irregular intervals, a stone can be fractured into random bricks, and a heap of flour can be scooped by willful fills. All occasioning measures of an optional division of one whole.

For the sake of arithmetical universality modern mathematics looks upon fractions as numbers, rational numbers, and it formerly includes them in the family of numbers by granting that they can be added and multiplied and that the composed number $1/a,\ a \neq 0$ is by definition such that $a(1/a) = 1$, and $b/a = b(1/a)$, so that $a/a = 1$ and $a/1 = a(1/1) = a$. Algebraically speaking 2/3 is the contrived solution of $3x = 2$. Granting further that rationals add and multiply commutatively and associatively, and that their multiplication is distributive with respect to addition it readily results that

$$\frac{1}{a} \cdot \frac{1}{b} = \frac{1}{b} \cdot \frac{1}{a} = \frac{1}{ab}$$

and that

$$\frac{1}{a} + \frac{1}{b} = \frac{1}{b} + \frac{1}{a} = \frac{a}{ab} + \frac{b}{ab} = \frac{a+b}{ab}.$$

The fact that $a/a = 1$ implies that 40320/362880 is mearly 1/9, but the existence of an unlimited number of primes raises the possibility of fractions-rational numbers-with vast numerators and denominators having no common factor. Such fractions may be hard to record, difficult to handle, and be of an exaggerated fractional practical interest, suggesting their decimal approximation to a prescribed degree of accuracy or resolution. For instance

$$\frac{3 \cdot 7 \cdot 13 \cdot 19 \cdot 29 \cdot 37 \cdot 43 \cdot 53 \cdot 61}{2 \cdot 5 \cdot 11 \cdot 11 \cdot 23 \cdot 31 \cdot 41 \cdot 47 \cdot 59} = 4 + \frac{3}{10} + \frac{1}{100} + \frac{8}{1000} + \frac{9}{10000} = 4.3189.$$

if we are prepared to ignore few 1/100000 parts.

Every analog measuring device has a limit to the fineness of its dial gradation and every digital device can display so many digits and no more. At some point we need to come to terms with the distinction between crude reality and the unlimited theoretical refinement of mathematics.

1.4 Negative Numbers

To avoid having to say no to the solution of $x + 1 = 0$, mathematics has further created another compound number of added meaning, -1, so that $1 + (-1) = 0$. In fact, it proposes that to every number a we introduce its **additive** inverse $-a$ so that $a + (-a) = 0$. Granting these new negative numbers the ability to add and multiply according to the rules of the **positive** numbers one readily obtains, for instance, that $1 + (-2) = 1 + 1 + (-1) + (-2) = 2 + (-2) + (-1) = -1$. From $1 \cdot (-1) = -1$, $(-1) \cdot 0 = 0$, $-1 \cdot (1 + (-1)) = 0$ it results that $(-1)(-1) = 1$, then that $-a = (-1) \cdot a$.

1.5 Irrational numbers

But if there are numbers of no rational practical existence could there be numbers that do not exist rationally even theoretically in spite of the fact that they may have an implicit algebraic description and represent some measurement of concrete nature? One readily encounters such instances. To find the side x of a square field of two area units we need to come up with x so that $x^2 = 2$, which we formally write as $x = \sqrt{2}$, with the $\surd$ being a stylized r standing for radical. This x cannot be measured with a yardstick of finite gradation no matter how fine, since no $x = a/b$ exists being the square root of 2. Indeed, x cannot be odd since then x^2 is odd contradicting the assertion that $x^2 = 2$, and x cannot be even, since any factor 2 in x is doubled in x^2.

In spite of the fact that $\sqrt{2}$ is not rational we still want to think of it as existing, as an irrational number, by dint of the fact that $(\sqrt{2})^2$ is the perfect 2.

Theorem : $\sqrt{2}$ is unique.

proof : Assume there are two positive roots to $x^2 = 2$ so that ${x_1}^2 = 2$ and ${x_2}^2 = 2$. Subtraction yields ${x_2}^2 - {x_1}^2 = (x_2 - x_1)(x_2 + x_1) = 0$, and since $x_1 > 0$ and $x_2 > 0$ it results that $x_1 = x_2$. End of proof.

Even though $\sqrt{2}$ is not rational still rational numbers can be found that are as close to $\sqrt{2}$ as desired or required. A simple routine for constructing such numbers consists of starting with any good rational approximation a/b to $\sqrt{2}$, then adding one to a if $(a/b)^2 < 2$ and adding one to b if $(a/b)^2 > 2$. Starting with 3/2 we obtain the sequence

$$\frac{3}{2}, \frac{3}{3}, \frac{4}{3}, \frac{5}{3}, \frac{5}{4}, \frac{6}{4}, \frac{6}{5}, \frac{7}{5}, \frac{8}{5}, \frac{8}{6}, \frac{9}{6}, \frac{9}{7}, \frac{10}{7} \quad \ldots$$

where $(10/7)^2 = 2(1 + 1/49)$.

1.5 Bisections

This procedure for producing rational approximations to $\sqrt{2}$ is utterly simple but slow. We can do better with the **bisection** method that is still simple yet is of great generality. We start with the two rational approximations x_1 and x_2 to $\sqrt{2}$ so that $x_1^2 < 2$ and $x_2^2 > 2$. this guarantees that $x_1 < \sqrt{2} < x_2$. We consequently propose the next approximation $x_3 = (x_1 + x_2)/2$ that is half way between x_1 and x_2. If it turns out that $x_3^2 > 2$, then now $x_1 < \sqrt{2} < x_3$ and we proceed to the next approximation $x_4 = (x_1 + x_3)/2$, but if $x_3^2 < 2$ then $x_3 < \sqrt{2} < x_2$ and we proceed instead to $x_4 = (x_2 + x_3)/2$ and so on, trapping thus $\sqrt{2}$ in ever narrower straits. After each bisection the interval of certainty for the location of $\sqrt{2}$ is halved. In fact, after n bisections $\sqrt{2}$ is hemmed to the confines of an interval of size $2^{-n}|x_2 - x_1|$ which can be reduced, with a sufficiently large n, to any degree of refinement.

For methodical reasons we prefer to look upon the computation of $\sqrt{2}$ as the search for the positive root of $r(x) = x^2 - 2 = 0$. We denote by n the bisection index so that after n bisections $x_n < x < x_{n+1}$, x_n being the current left or lower bound on $\sqrt{2}$ and x_{n+1} the current right or upper bound on $\sqrt{2}$. For $r(x) = x^2 - 2$ it happens that $r_n = r(x_n) < 0$ and $r_{n+1} = r(x_{n+1}) > 0$, but the reverse could happen. Starting with $x_1 = 9/7$ and $x_2 = 10/7$ we compile by successive bisections the following upper and lower estimates for $\sqrt{2}$:

n	$r_n < 0$	x_n x_{n+1}	$r_{n+1} > 0$	$x_{n+1} - x_n$
1	-0.347	$9/7 < x < 10/7$	0.0408	1/7

$$\text{next x} = \frac{1}{2}\left(\frac{9}{7} + \frac{10}{7}\right) = \frac{19}{14}, \quad \text{next r} = -0.158$$

2	-0.158	$19/14 < x < 20/14$	0.0408	1/14

$$\text{next} \quad x = \frac{1}{2}\left(\frac{19}{14} + \frac{20}{14}\right) = \frac{39}{28}, \quad \text{next r} = -0.0599$$

3	-0.0599	$39/28 < x < 40/28$	0.0408	1/28

$$\text{next} \quad x = \frac{1}{2}\left(\frac{39}{28} + \frac{40}{28}\right) = \frac{79}{56}, \quad \text{next r} = -0.00989$$

4	-0.00989	$79/56 < x < 80/56$	0.0408	1/56

$$\text{next} \quad x = \frac{1}{2}\left(\frac{79}{56} + \frac{80}{56}\right) = \frac{159}{112}, \quad \text{next r} = 0.0154$$

5	-0.00989	$158/112 < x < 159/112$	0.0154	1/112

$$\text{next} \quad x = \frac{1}{2}\left(\frac{158}{112} + \frac{159}{112}\right) = \frac{317}{224}, \quad \text{next r} = 0.00273$$

6	-0.00989	$316/224 < x < 317/224$	0.00273	1/224

where all approximations are of the form

$$-\frac{1}{b} < \sqrt{2} - \frac{a}{b} < \frac{1}{b}.$$

Incidentally, continued fractions

$$\sqrt{2} = 1 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \cdots}}}}}$$

furnishes the compacter approximations

$$\begin{aligned}
-\frac{1}{4} &< \sqrt{2} - \frac{3}{2} < \frac{1}{4} \\
-\frac{1}{25} &< \sqrt{2} - \frac{7}{5} < \frac{1}{25} \\
-\frac{1}{144} &< \sqrt{2} - \frac{17}{12} < \frac{1}{144} \\
-\frac{1}{841} &< \sqrt{2} - \frac{41}{29} < \frac{1}{841} \\
-\frac{1}{4900} &< \sqrt{2} - \frac{99}{70} < \frac{1}{4900}
\end{aligned}$$

which are all of the form

$$-\frac{1}{b^2} < \sqrt{2} - \frac{a}{b} < \frac{1}{b^2}.$$

In the language of calculus $x_{n+1} - x_n \to 0$, tends to zero, as $n \to \infty$, tends to infinity, so that both x_n and x_{n+1} tend to what we denote by $\sqrt{2}$, x_n coming up from below and x_{n+1} coming down from above. It appears that also $r_n \to 0$ and $r_{n+1} \to 0$ as $n \to \infty$, which we formally show to be true by writing $r_n = x_n^2 - 2,\ r_{n+1} = x_{n+1}^2 - 2$ so that

$$r_{n+1} - r_n = x_{n+1}^2 - x_n^2 = (x_{n+1} - x_n)(x_{n+1} + x_n).$$

For large n it is nearly true that $x_{n+1} = x_n = \sqrt{2}$ so that we may write the previous equation as

$$|r_{n+1}| + |r_n| = 2\sqrt{2}(x_{n+1} - x_n)$$

but $-r_n > 0$ and $r_{n+1} > 0$, and indeed both r_n and r_{n+1} tend to zero as $x_{n+1} - x_n$ tends to zero.

Now we may look upon the existence of $\sqrt{2}$ in the sense of a limit—in the sense of the possibility to construct rational approximations to the solution of $x^2 = 2$ of any degree of accuracy.

Application of bisection to the solution $r = x^3 + x - 4 = 0$ produces

n	$r_n < 0$	x_n	x_{n+1}	$r_{n+1} > 0$	$x_{n+1} - x_n$
1	-0.0254	$11/8 < x < 12/8$		0.875	1/8

$$\text{next } \ x = \frac{1}{12}\left(\frac{11}{8} + \frac{12}{8}\right) = \frac{23}{16}, \quad \text{next r} = 0.409$$

2	-0.0254	$22/16 < x < 23/16$	0.408	1/16

$$\text{next } \ x = \frac{1}{2}\left(\frac{22}{16} + \frac{23}{16}\right) = \frac{45}{32}, \quad \text{next r} = 0.187$$

3	-0.0254	$44/32 < x < 45/32$	0.187	1/32

$$\text{next } \ x = \frac{1}{2}\left(\frac{44}{32} + \frac{45}{32}\right) = \frac{89}{64}, \quad \text{next r} = 0.0799$$

4	-0.0254	$88/64 < x < 89/64$	0.0799	1/64

$$\text{next } \ x = \frac{1}{2}\left(\frac{88}{64} + \frac{89}{64}\right) = \frac{177}{128}, \quad \text{next r} = 0.0270$$

5	-0.0254	$176/128 < x < 177/128$	0.0270	1/128

$$\text{next } \ x = \frac{1}{2}\left(\frac{176}{128} + \frac{177}{128}\right) = \frac{353}{256}, \quad \text{next r} = 0.000734$$

6	-0.0254	$352/256 < x < 353/256$	0.000734	1/256

$$\text{next } \ x = \frac{1}{2}\left(\frac{352}{256} + \frac{353}{256}\right) = \frac{705}{512}, \quad \text{next r} = -0.0123$$

7	-0.0123	$705/512 < x < 706/512$	0.000734	1/512

Here $x_{n+1} - x_n = 2^{-n-2}$ and certainly $x_{n+1} - x_n \to 0$ as $n \to \infty$. To theoretically confirm the observation that also both $r_n \to 0$ and $r_{n+1} \to 0$ as $n \to \infty$ we write

$$r_n = x_n^3 + x_n - 4, \quad r_{n+1} = x_{n+1}^3 + x_n - 4$$

to have by subtraction

$$r_{n+1} - r_n = x_{n+1}^3 - x_n^3 + x_{n+1} - x_n$$

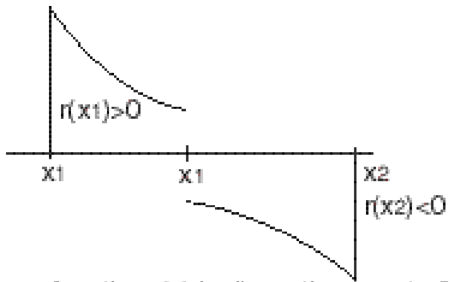
or

$$|r_{n+1}| + |r_n| = (x_{n+1} - x_n)(x_{n+1}^2 + x_{n+1}x_n + x_n^2 + 1)$$

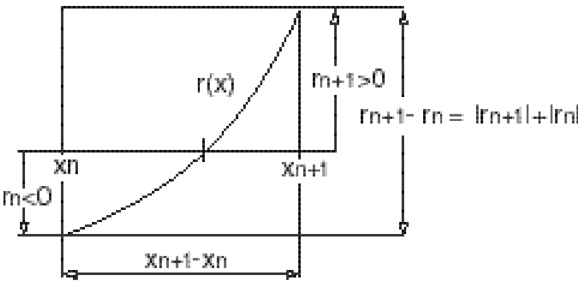
and it so happens.

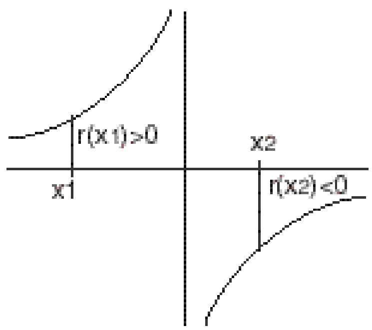
The fact that both $r_n \to 0$ and $r_{n+1} \to 0$ as $n \to \infty$, implies that x_n and x_{n+1} tend to an actual root of $r(x) = x^3 + x - 4 = 0$, and that the function $r(x)$ is **continuous** at x

where $r(x) = 0$. Application of the bisection method to a function exhibiting a discontinuity at some x_0 as in the figure below would lead to a sequence $x_1, x_2, x_3, x_4, \ldots, x_n, x_{n+1}$ such that $x_{n+1} - x_n \to 0$ as $n \to \infty$ but not $r(x_n)$ and not $r(x_{n+1})$ would tend to zero. Such a situation would arise also for a function that diverges up and down at x_0 as in the figure below.



function r(x) is discontinuous at x0





Theorem: *Every polynomial equation of odd degree has a real root.*

Proof: x_1 and x_2 can be found so that $r(x_1) < 0$ and $r(x_2) > 0$. Bisection does the rest.

Inequality Theorems:

(1) *If* $a > b$ *then for any* c, $a \pm c > b \pm c$.

(2) If $a > b$, *then for any* $c > 0$, $ac > bc$ *and* $a/c > b/c$,

(3) If $a > 0$, $b > 0$ *and* $a > b$, *then* $a^2 > b^2$ *but* $1/a < 1/b$.

(4) If $a > b$ *and* $b > c$, *then* $a > c$.

(5) All these statements remain true if $>$ *and* $<$ *are replaced by* $\geq$ *and* $\leq$ *throughout.*

Absolute Value Theorems:

(1) *If* $a = b$, *then* $|a| = |b|$.

(2) $|a| = |-a|$.

(3) $|ab| = |a| \cdot |b|$.

(4) $|a + b| \leq |a| + |b|$.

Triangular inequality 4 turns into an equality if a and b have the same sign.

2 The shape of things

2.1 the point

The **point**, called more emphatically **Punkt** by the Germans, is produced by a **pick** or a **pang** of the **pencil**, is a mere **dot** made notable as a **dent** in matter. Yet it is the primary building block of all shapes and form. Congregating in strings and rows points produce lines and curves, while spreading out they produce shapes, figures and forms.

From computer screen pixels to photographic plate crystals to miniature magnetic poles, to laser produced tiny ink droplets points have a definite size and produce figures of limited resolution. When the eye ceases to discriminate between near points we come to regard the view as visually crisp and clear.

What we colloqoquially mean and grossly think of as a point is a discernible chunk of space small relative to the expanse in which it is immersed. A microscopic bead of ink on this page in a point, so is a pin stuck on a board, so is a stake driven into the ground in a field, so is a house upon the landscape, and so is a city on earth. Earth itself is a mere point in the solar system, which in turn is only a **speck**, produced so to speak by a **pick** of a **spike**, in the universe extending around it. It is all a question of **scale**.

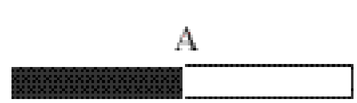
To distinguish between points we label them, usually by capital Roman letters, or in reality describe them or point them out as being the nail head, the tree, the house, etc.

We associate with any two points of space an intuitive **measure** that we call the **distance** between them. The distance between points A and B is denoted by $d(AB)$ and is such that $d(AB) = d(BA)$ and $d(AA) = 0$. Here is where **numbers** come into **geometry**. With any three points of space we associate another measure that we call **area**, and with any four points of space we associate a measure that we call **volume**.

The idea of distance is related to the idea of **rigidity** or **solidity**. The distance between two points marked on a rigid body such as a slab of stone, a board of wood or a strip of metal, remains the same as the body is moved in space and this is how we actually compare and measure distance.

A true geometrical point is an ideal object conceived by a mental process shrinking. We subtrefugely speculate about the geometrical point in the process of diminution of a whale to an elephant to a horse to a dog to a cat to a mouse to flea to an amoeba to something that is small beyond belief, yet existing.

A cleaner way to visualize, and actually **mark**, the ideal point is to think about it as a **margin** or a border of a cut. Point A lies on the border between the dark and the light sections of the strip below.



It is as English aptly puts it a shade to the right of the dark region and a shade to the left of the bright region. It is at once the last point and the first point of the cut.

By admitting a measure for the distance between any two points we grant that there is a number corresponding to any distance and vice versa. If this distance is rational, then it can be numerically expressed as a fraction of some agreed unit of length. But if the distance is irrational, say $d(AB) = \sqrt{2}$ as with a segment AC that is the hypothenous of right triangle ABC with $d(AB) = d(BC) = 1$, then no rigid measuring tape, with no matter how fine gradation, will settle on a fractional mark for the length of this segment. Repeated subdivision, as far as practice will allow, of the markings on the measuring tape will allow a closer fractional approximation for $d(AC)$ up to the limit of practicality.

Transmitting $d(AC)$ to another body without attaching to it a numerical value can be quite accurately done by marking points A and C on a movable clean and flat bar, but again, to a degree of accuracy dictated by the practical considerations of the tools used and the limitation of our power of vision.

To avoid complicated and cumbersome fractions and to ease the sizing of industrial goods industry maintains few standard sizes represented by whole numbers in English or Metric units.

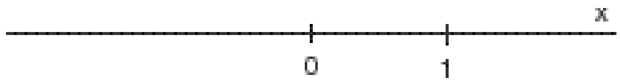
Accurate mechanical sliding calipers have the ability to measure distance to an accuracy of $\pm\frac{1}{10}$ of a millimeter, while accurate screw micrometers have an inherent accuracy of $\pm\frac{1}{50}$ of a millimeter, with limits on the attainable accuracy imposed by lack of absolute rigidity in metals, by the width of the hairline markings, and the restrictions of our sight.

2.2 The real axis

Imagine traveling by car on a long road. Reference to points along the road can be made descriptively by alluding to any prominent marking such as a sign, a tree or a house of some special shape or form. The availability of a speedometer allows us to fix even an unmarked point along the road by its **distance** from some entry point of reference. The access or entry point to the road we call the origin point and mark it by 0. To reach a distant point on the road we must be instructed to turn left or right at the point of entry, then be given the distance of travel. Having carried out these instructions we stop according to the reading of the speedometer and are at the agreed point, more or less.

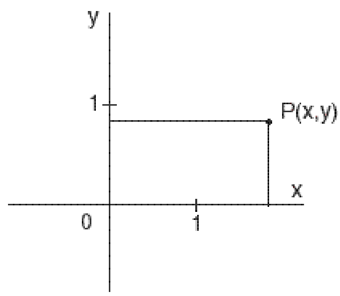
Geometrically speaking, to construct a numerical addressing system for any point on a line we first mark an origin point labeled 0 on the line, then we mark another point labeled 1 at a unit distance to the right of the origin. The line is turned thereby into an **axis** – the real axis – endowed with an entry point and a sense of travel. Travel from point 0 to 1 is said to be travel in the positive direction, while travel backwards is said to be travel in the negative direction. The address of any point on the line is a real number d such that $|d|$ is its distance from the origin, and the sign of d is the sense of travel, positive for travel from 0 to 1, and negative backwards.

To any point on the line there corresponds a real number, and any real number fixes a point on the line, with no gaps.



2.3 The Cartesian coordinate system

Matters become more interesting in the plane where a point is addressed by means of a pair of numbers.



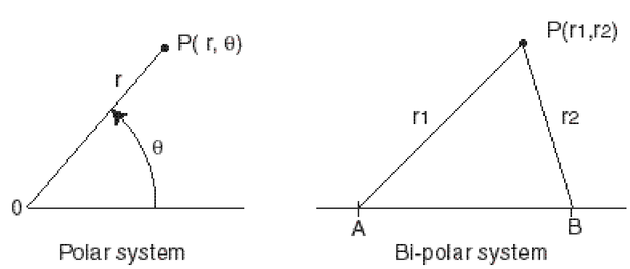
The Cartesian (after Descartes) coordinate system consists of two orthogonal axes, labeled x and y, sharing a common origin. Point $P(x, y)$ is reached by traveling distance $|x|$ along the x axis, to the right if $x > 0$ and to the left if $x < 0$, then turning 90^o and traveling distance $|y|$ parallel to the y axis, up if $y > 0$ and down if $y < 0$. Having accomplished this we are at point P, even if the landscape is devoid of any marks and features. Point P is fixed relative to the Cartesian coordinate system by the intersection of two perpendicular lines, one parallel to the x-axis and passing through point y on the y-axis, and the other parallel to the y-axis and passing through x on the x-axis.

By the orthogonality of the real axes comprising of the Cartesian system we have that the distance between points $P(x_1, y_1)$ and $Q(x_2, y_2)$ is, according to the theorem of Pythagoras

$$d(PQ) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

which is said to be the **metric** of the system.

2.4 Other coordinate systems



2.5 The equations of the circle

Using the metric of the Cartesian coordinate system we may write the functional relationship between the x and y coordinates of arbitrary point $P(x, y)$ on a curve defined in terms of distance.

A circle of center $C(x_0, y_0)$ and radius $r > 0$ consists of the collection of all points $P(x, y)$ such that $d(PC) = r$. Point Q for which $d(QC) < r$ is said to be **inside** the circle, and point R for which $d(RC) > r$ is said to be **outside** the circle.

The declamatoric definition of the circle is translated into an analytical formulation by noting that for any point $P(x, y)$ on the circle it happens that

$$(x - x_0)^2 + (y - y_0)^2 = r^2$$

which we call the equation of the circle. Every property of the circle, from the plainly obvious to the deeply hidden, is embodied in this equation and can be elicited from it by purely analytical means.

The equation of the circle can be written **parametrically** as

$$x = x_0 + r \cos\theta, \ \ y = y_0 + r \sin\theta.$$

The equation of the circle contains one intrinsic parameter, namely the radius r. The coordinates of the center $C(x_0, y_0)$ merely fix the position of the circle relative to the coordinate system.

Elementary properties of the circle:

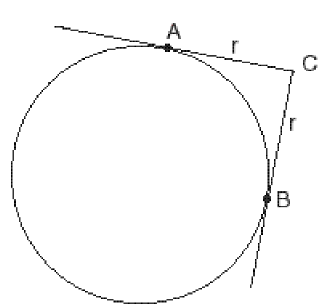
5.1 The circle is fixed by three points.

5.2 The circle is symmetric with respect to any line through its center.

5.3 The line intersects the circle at most at two points.

5.4 The tangent line to the circle is orthogonal to the line through the center and the point of tangency.

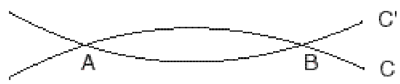
5.5 if $AC \perp BC$, then $d(AC) = d(BC)$ =radius of circle. See the figure below.



2.6 The equation of the line

The **line** is produced by stretching a long string of **linen**, or a **lint**, between two fixed points. Generally two points fix a line and two lines fix an intersection point.

Curve C is not a line since it can be rotated to produce different curve C'.

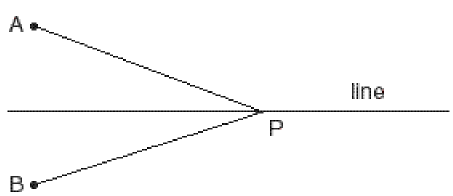


Stretching of a limp string between two points is a legally accepted mode of producing lines in construction and surveying. On a smaller scale we reproduce lines by using carefully manufactured rulers.

Three straightedges A, B, C, can be tested for correctness by holding A on B, then A on C, then B on C, and verifying that there are no light letting gaps between any pair of matched edges.

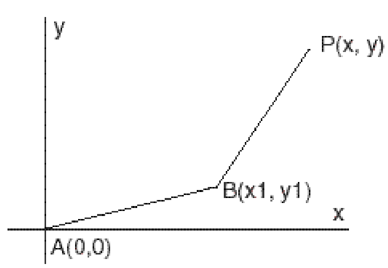
The only known mechanical tool for producing lines is Peaucellier's linkage, invented in 1864, that converts rotation into rectilinear motion.

A line is the collection of all points that are situated symmetrically with respect to two points so that for any point P on the line $d(AP) = d(BP)$. This is how we actually judge a curve to be straight by its symmetric impression on our eyes, or by its image falling evenly on the retina of one eye. See the figure below.



Consider the line fixed by points $A(0,0)$ and $B(x_1, y_1)$ relative to a Cartesian coordinate system. Point $P(x,y)$ with is on the line if and only if $d(AB) + d(BP) = d(AP)$. Using the metric of the system we write its equation in terms of x and y as

$$\sqrt{x_1^2 + y_1^2} + \sqrt{(x - x_1)^2 + (y - y_1)^2} = \sqrt{x^2 + y^2}.$$



Repeated squaring of both sides of this equation leads to the much simplified equation

$$x_1 y = y_1 x$$

or

$$y = \frac{y_1}{x_1} x$$

if $x_1 \neq 0$. The typical quantity $m = y_1/x_1$ is called the **slope** of the line, and if the line is translated up by amount b, then its equation becomes

$$y = mx + b, \quad m = (y_2 - y_1)/(x_2 - x_1).$$

This very simple form of the equation of the line relative to the Cartesian coordinate system is due to the fact that the system itself is composed of a pair of orthogonal lines. In case the line is fixed by the two points $A(x_0, y_0), B(x_0, y_1)$, so that it is parallel to the y-axis or orthogonal to the x-axis, then its equation becomes $x = x_1$ for any y.

The equation of the line contains no intrinsic parameters. Slope m and y-intercept b merely indicate the position of the line relative to the coordinate system.

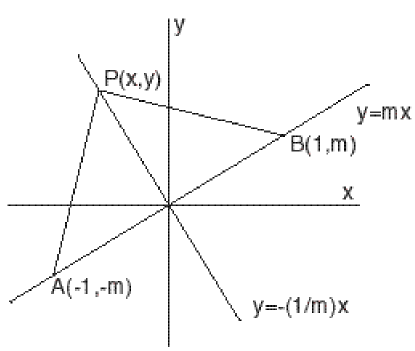
Consider line $y = mx$. Points $A(-1, -m)$ and $B(1, m)$ are situated on it symmetrically with respect to the origin 0 in the sense that $d(0A) = d(0B)$. Another line passing through origin 0 is said to be **orthogonal** or **perpendicular** to line AB if $d(AP) = d(BP)$ for any point P on it. Using the metric of the system we write the equation of the line OP as

$$\sqrt{(x+1)^2 + (y+m)^2} = \sqrt{(x-1)^2 + (y-m)^2}$$

or after squaring both sides as

$$y = -\frac{1}{m} x.$$

Two lines with slopes m_1 and m_2 are mutually orthogonal if $m_1 m_2 = -1$.



Orthogonal lines

2.7 The tangent line

Calculus is greatly interested in the relationship between lines and curves, particularly the relationship of tangency.

To write the equation of the tangent line to $y = x^2$ at $P(1, 1)$ on the curve we write the equation of any line through P as $y-1 = m(x-1)$ and find the intersections of the parabola and the line through the solution of $y = mx + (1 - m)$ and $y = x^2$, leading to the quadratic equation $x^2 - mx + (m - 1) = 0$. One solution of this equation is known to be $x_1 = 1$, and according to Descartes, the other solution is $x_2 = m-1$. A single intersection happens when $m = 2$ and the equation of the tangent line is $y = 2x - 1$.

To write the equation of the tangent line to $y = 2x^2 - x$ through point $P(1, -1)$ not on the curve, we write the equation of any line through $P(1, -1)$ as $y + 1 = m(x - 1)$ and seek the intersections of $y = 2x^2 - x$ and $y = mx + (-m - 1)$ from the quadratic equation $2x^2 + (-1 - m)x + (m + 1) = 0$. Since P is not on the parabola we may not use Descartes and require instead that the discriminant of the quadratic equation be zero;

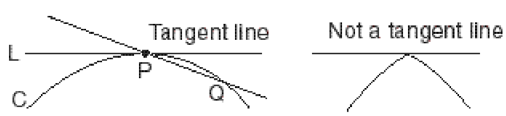
$$(-1 - m)^2 - 4 \cdot 2(m + 1) = 0 \qquad \text{or} \qquad m^2 - 6m - 7 = 0$$

that is satisfied by $m_1 = -1$ and $m_2 = 7$. Now there are two tangent lines, $y = -x$ and $y = 7x - 8$.

The tangency relationship is local and does not depend on what takes place far from the point at which the curve and the line come together.

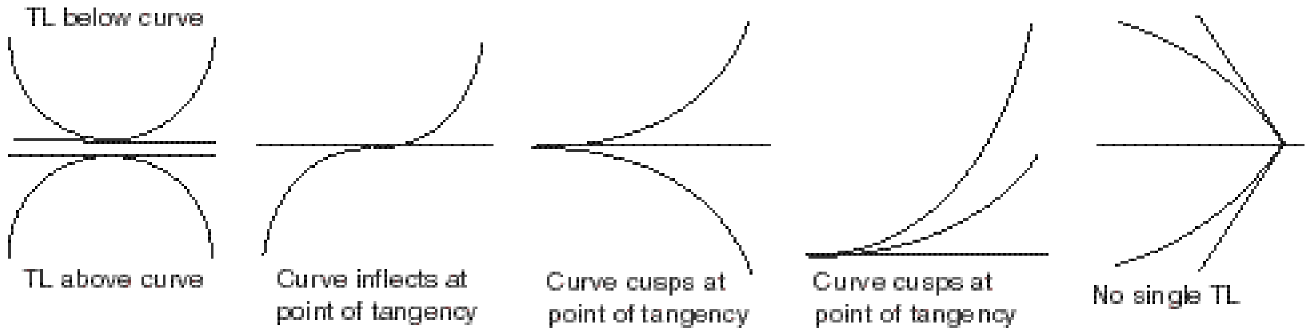
In fact, line L is tangent to curve C at point P if:

1. P is common to both L and C.
2. P is isolated.
3. Upon rotation, no matter how small, of line L around the point of tangency P and towards the curve, line L immediately reintersects curve C at point Q.

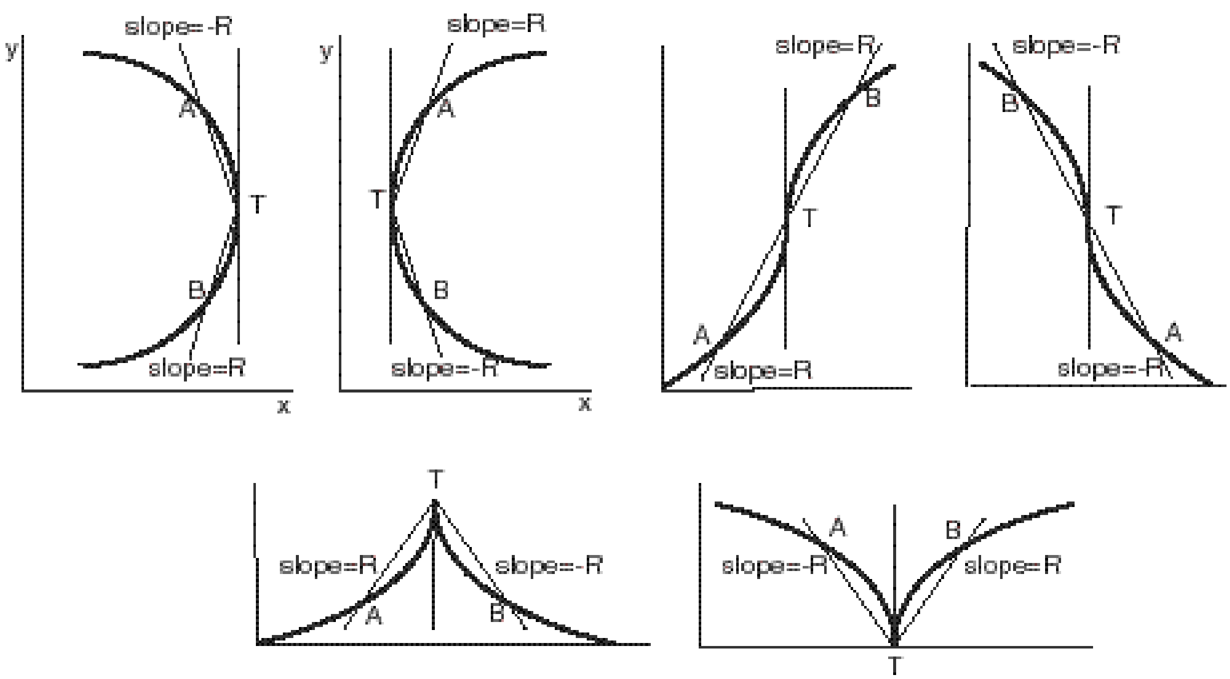


To prove that line $y = 2x - 1$ is tangent to $y = x^2$ we first find their intersection by solving $x^2 = 2x - 1$ or $x^2 - 2x + 1 = 0$ that is satisfied by $x_1 = x_2 = 1$. The line and the parabola intersect at point $P(1, 1)$ and their intersection is surely isolated. Next we rotate line $y = 2x - 1$ around point P by changing its slope from $m = 2$ to $m = 2 + \epsilon$ so that the equation of the rotated line becomes $y - 1 = (2 + \epsilon)(x - 1)$. To find the intersections of this line and the parabola $y = x^2$ we solve $x^2 = 1 + (2 + \epsilon)(x - 1)$ or $x^2 + (-2 - \epsilon)x + (1 + \epsilon) = 0$. One intersection point is $P(1, 1)$ so that one solution of the quadratic equation is $x_1 = 1$. With the aid of Descartes we get the next intersection point $x_2 = 1 + \epsilon$.

See the figure below



See the figure below



3.1 Linearization

Consider the function $y = f(x) = x^2$ with the reference point $x_0 = 1$ for which $y_0 = f(x_0) = x_0^2 = 1$. Having interest in the change in y due to a change in x near $x_0 = 1$ we introduce the new variables δx and δy so that

$$x = x_0 + \delta x \qquad \text{and} \qquad y = y_0 + \delta y$$

in which δx is independent and δy depends on it. The Greek letter δ, Delta, is used to indicate that δx and δy are *differences* in x and y, respectively. This is the only instance in mathematics where two letters are used to denote a variable.

Here $x = 1 + \delta x$ and $y = 1 + \delta y$ with which that $y = f(x) = x^2$ becomes

$$1 + \delta y = (1 + \delta x)^2$$

or

$$\delta y = (2 + \delta x)\delta x.$$

It turns out that δy is proportional to δx with the proportionality factor being here $2 + \delta x$.

At this point enters calculus with the assumption or restriction that $|\delta x| << 1$. Under this assumption $2 + \delta x$ is nearly equal to 2 and the relationship $\delta y = (2 + \delta x)\delta x$ may be simplified to the linear approximation $\delta y = 2\delta x$. To point out that the change in y in the linearized relationship is not the original one Greek δ is replaced by the Roman d so that for *differences* δx and δy we write now *differentials* dx and dy,

$$dy = 2dx, \quad y - 1 = 2(x - 1).$$

Notice that even though dy is not δy, yet $\delta x = dx$.

The linearized relationship $dy = 2dx$ holds for any dx and dy, but if $|\delta x| = |dx| << 1$, then δy should differs little from dy. Indeed,

$$\delta y - dy = (2 + \delta x)\delta x - 2\delta x = (\delta x)^2$$

and if $|\delta x| << 1$, or $|\delta x|$ is near zero, then $|\delta x|^2$ is much smaller than $|\delta x|$.

Consider now the more general

$$y = 3x^2 - 2x + 1, \quad x_0 = 1, \quad y_0 = 2$$

which with

$$x = x_0 + \delta x = 1 + \delta x, \; y = y_0 + \delta y = 2 + \delta y$$

becomes

$$2 + \delta y = 3(1 + \delta x)^2 - 2(1 + \delta x) + 1$$

or

$$\delta y = (4 + 3\delta x)\delta x.$$

In the process of linearization we always strive to write the relationship between δx and dy as $\delta y = (\quad)\delta x$. We expect this form since $\delta x = 0$ implies $\delta y = 0$.

At this point algebra ends and calculus takes over with its universal assumption that $|\delta x| << 1$. The factor $4 + 3\delta x$ is now seen to be nearly 4 for δx that is nearly zero. We express this fact by saying that

$$\lim_{\delta x \to 0} (4 + 3\delta x) = 4$$

and have that

$$dy = 4dx, \; y - 2 = 4(x - 1).$$

Notice that in the process of linearization δx is not set equal to zero. Doing this would lead a constantization in which $\delta y = 0$. In the process of linearization independent variable δx is allowed to vary albeit in a relatively small range about $x = x_0$.

The linear relationship $dy = 4dx$ between differentials dx and y is valid for any dx, but if $|dx| = |\delta x| << 1$, then δy should be close to dy, with ultimately $dy = \delta y = 0$ for $\delta x = dx = 0$. In fact,

$$\delta y - dy = (4 + 3\delta x)\delta x - 4\delta x = 3(\delta x)^2$$

and if $|\delta x| << 1$, then $|\delta x|^2$ is considerably smaller than $|\delta x|$.

Consider next the rational function

$$y = \frac{1}{3x - 5}, \; x_0 = 2, \; y_0 = 1$$

that with

$$x = 2 + \delta x, \quad y = 1 + \delta y$$

becomes

$$1 + \delta y = \frac{1}{3(2 + \delta x) - 5}$$

or

$$\delta y = \frac{1}{1 + 3\delta x} - 1.$$

Common denominator brings the relationship between differences δx and δy to the desireed form

$$\delta y = \frac{-3}{1 + 3\delta x}\delta x$$

or $\delta y = (\quad)\delta x$ required for the linearization.

Now

$$\lim_{\delta x \to 0} \frac{-3}{1 + 3\delta x} = -3$$

and linearization is accomplished with

$$dy = -3dx, \; y - 1 = -3(x - 2).$$

Here

$$\delta y - dy = \frac{-3}{1 +_3 \delta x}\delta x - (-3\delta x) = \frac{9}{1+3\delta x}(\delta x)^2$$

in which

$$\lim_{\delta x \to 0} \frac{9}{1+3\delta x} = 9$$

so that for $|\delta x| << 1$

$$\delta y - dy = 9(\delta x)^2$$

approximately.

Consider now the radical function

$$y = \sqrt{-3x+1}, \quad x_0 = -1, \ y_0 = 2$$

that with

$$x = x_0 + \delta x = -1 + \delta x, \ y = y_0 + \delta y = 2 + \delta y$$

becomes

$$2 + \delta y = \sqrt{-3(-1+\delta x)+1}$$

or

$$\delta y = \sqrt{4 - 3\delta x} - 2$$

for variables $\delta x, \delta y$.

To bring the relationship between δy and δx to the desired form $\delta y = (\quad)\delta x$, we multiply and divide δy by its conjugate to have

$$\delta y = \frac{\sqrt{4-3\delta x} - 2}{1} \cdot \frac{\sqrt{4-3\delta x}+2}{\sqrt{4-3\delta x}+2}$$

resulting in

$$\delta y = \frac{-3}{\sqrt{4-3\delta x}+2}\delta x.$$

With

$$\lim_{\delta x \to 0} \frac{-3}{\sqrt{4-3\delta x}+2} = -\frac{3}{4}$$

linearization of $y = \sqrt{-3x+1}$ at $x_0 = -1$ is accomplished as

$$dy = -\frac{3}{4}dx, \ y - 2 = -\frac{3}{4}(x+1).$$

Here

$$\delta y - dy = \frac{-3}{\sqrt{4-3\delta x}+2}\delta x - \left(-\frac{3}{4}\right)\delta x$$

or

$$\delta y - dy = \frac{\sqrt{4-3\delta x}-2}{\sqrt{4-3\delta x}+2} \cdot \frac{3}{4}\delta x$$

or

$$\delta y - dy = -\frac{9}{4} \cdot \frac{1}{(\sqrt{4 - 3\delta x} + 2)^2} (\delta x)^2.$$

But

$$\lim_{\delta x \to 9} \frac{1}{(\sqrt{4 - 3\delta x} + 2)^2} = \frac{1}{16}$$

so that

$$\delta y - dy = -\frac{9}{64} (\delta x)^2$$

approximately, if $|\delta x| << 1$.

3.2 The derivative function

In the first example of the previous section we considered the linearization of $y = x^2$ at $x_0 = 1$, $y_0 = 1$. We shall now leave the value of x_0 open and write $x_0 = x_0$ for arbitrary x_0. Corresponding to this x_0 we have $y_0 = x_0^2$ and the shift of variables becomes here

$$x = x_0 + \delta x, \;\; y_0 = y_0 + \delta y = x_0^2 + \delta y$$

with which $y = x^2$ becomes

$$x_0^2 + \delta y = (x_0 + \delta x)^2 = x_0^2 + 2x_0 \delta x + \delta x^2$$

or

$$\delta y = (2x_0 + \delta x)\delta x$$

which is ready in this form for linearization.
In fact,

$$\lim_{\delta x \to 0} (2x_0 + \delta x) = 2x_0$$

and linearization is accomplished with

$$dy = (2x_0)dx, \;\; y - y_0 = (2x_0)(x - x_0).$$

Factor $2x_0$ in $dy = (\;\;)dx$ is called the *derivative* function of x^2, $y'(x) = (x^2)' = 2x$. The derivative function allows the immediate linearization of x^2 at any x_0.

Consider next the rational function

$$y = \frac{1}{2x + 1}, \;\; x_0 = x_0, \;\; y_0 = \frac{1}{2x_0 + 1}$$

for which

$$x = x_0 + \delta x, \;\; y = y_0 + \delta y = \frac{1}{2x_0 + 1} + \delta y.$$

Now

$$\frac{1}{2x_0 + 1} + \delta y = \frac{1}{2(x_0 + \delta x) + 1}$$

or

$$\delta y = \frac{1}{2(x_0 + \delta x) + 1} - \frac{1}{2x_0 + 1}.$$

A common denominator reforms δy into the form

$$\delta y = \frac{-2}{(2x_0 + 1)(2x_0 + 1 + 2\delta x)}\delta x$$

that we seek before linearization. We have that

$$\lim_{\delta x \to 0} \frac{-2}{(2x_0 + 1)(2x_0 + 1 + 2\delta x)} = \frac{-2}{(2x_0 + 1)^2}$$

and hence

$$dy = \frac{-2}{(2x_0 + 1)^2}dx, \quad y'(x) = -\frac{2}{(2x + 1)^2}.$$

We shall now linearize the radical function

$$y = \sqrt{-3x + 2}, \quad x_0 = x_0, \ y_0 = \sqrt{-3x_0 + 2}.$$

Substitution of

$$x = x_0 + \delta x, \ y = y_0 + \delta y = \sqrt{-3x_0 + 2} + \delta y$$

into the function results in

$$\sqrt{-3x_0 + 2} + \delta y = \sqrt{-3(x_0 + \delta x) + 2}$$

or

$$\delta y = \sqrt{-3x_0 + 2 - 3\delta x} - \sqrt{-3x_0 + 2}.$$

Multiplication and division of δy by its conjugate reforms δy into

$$\delta y = \frac{-3}{\sqrt{-3x_0 + 2 - 3\delta x} + \sqrt{-3x_0 + 2}}\delta x$$

and with

$$\lim_{\delta x \to 0} \frac{-3}{\sqrt{-3x_0 + 2 - 3\delta x} + \sqrt{-3x_0 + 2}} = -\frac{3}{2}\frac{1}{\sqrt{-3x_0 + 2}}$$

we have the linearization

$$dy = \frac{-3}{2\sqrt{-3x_0 + 2}}dx, \quad y'(x) = \frac{-3}{2\sqrt{-3x + 2}}.$$

3.3 Quadratization

We shall linearize and then quadratize

$$y = \frac{1}{3x - 2}, \ x_0 = 1, \ y_0 = 1.$$

With the shift of variables

$$x = x_0 + \delta x = 1 + \delta x,\ y = y_0 + \delta y = 1 + \delta y$$

the function is transformed into

$$1 + \delta y = \frac{1}{3(1+\delta x) - 2}$$

or

$$\delta y = \frac{1}{3\delta x + 1} - 1$$

which by a common denominator changes into

$$\delta y = \frac{-3}{1 + 3\delta x}\delta x$$

that is ready for linearization. In fact

$$\lim_{\delta x \to 0} \frac{-3}{1 + 3\delta x} = -3$$

and the linearization becomes

$$dy = -3dx,\ y - 1 = -3(x - 1).$$

Quadrization starts by recalling that $dx = \delta x$ so that here $dy = -3\delta x$. The general form of δy is $\delta y = (\ \ )\delta x$. To this we add and subtract dy so as to have here $\delta y = (\ )\delta x + 3\delta x - 3\delta x$. Factoring out δx we have here

$$\delta y = \left(\frac{-3}{1 + 3\delta x} + 3 - 3\right)\delta x$$

or

$$\delta y = -3\delta x + \left(\frac{-3}{1 + 3\delta x} + 3\right)\delta x$$

where $-3\delta x$ is dy. A common denominator in the second term of δy brings it to the form

$$\delta y = -3\delta x + \left(\frac{9}{1 + 3\delta x}\right)\delta x^2$$

which is the appropriate form for quadratization. Notice that this form of δy is merely $\delta y = dy + (\delta y - dy)$ that we prefer to write as $\delta y = dy + (\ \ )\delta x^2$.

Now

$$\lim_{\delta x \to 0} \frac{9}{1 + 3\delta x} = 9$$

and the quadratization of $y = 1/(3x - 2)$ at $x_0 = 1$ $y_0 = 1$ is accomplished with

$$\delta y = -3\delta x + 9(\delta x)^2,\ y - 1 = -3(x - 1) + 9(x - 1)^2$$

which is an equation of a *parabola*. We do not replace δx and δy in the quadratization by dx and dy. This differential notation is reserved for the linearization only.

The parabola produced by quadratizaation provides a better approximation to the function near $x_0 = 1$ than the line produced by the linearization.

In fact, here

$$\begin{aligned}
\delta y - (-3\delta x + 9\delta x^2) &= \frac{-3}{1+2\delta x}\delta x - (-3\delta x + 9\delta x^2) \\
&= 9\left(\frac{1}{1+3\delta x} - 1\right)\delta x^2 \\
&= \frac{-27}{1+3\delta x}(\delta x)^3 \\
&= -27(\delta x)^3
\end{aligned}$$

approximately if $|\delta x| << 1$. The difference between the true δy and the quadratized δy is nearly proportional to $(\delta x)^2$ and if $|\delta x| << 1$, then $(\delta x)^2$ is smaller than $|\delta x|$, and $|\delta x|^3$ is still smaller than $(\delta x)^2$.

3.4 Cubization

The processes of linearization and quadratization can be continued into cubization by recalling that quadratization produced

$$\delta y = -3\delta x + \frac{9}{1+3\delta x}\delta x^2, \quad \delta y = -3\delta x + 9(\delta x)^2$$

where δy to the left is the exact one and δy to the right is the quadratic one. The exact δy is now rewritten as

$$\delta y = -3\delta x + \left(\frac{9}{1+3\delta x} + 9 - 9\right)\delta x^2$$

or

$$\delta y = -3\delta x + 9(\delta x)^2 + 9\left(\frac{1}{1+3\delta x} - 1\right)\delta x^2.$$

A common denominator for the factor of $(\delta x)^2$ transform δy into

$$\delta y = -3\delta x + 9(\delta x)^2 + \frac{-27}{1+3\delta x}(\delta x)^3.$$

Cubization of the true δy is achieved with

$$\lim_{\delta x \to 0} \frac{-27}{1+3\delta x} = -27$$

so that the cubic δy is

$$\delta y = -3\delta x + 9(\delta x)^2 - 27(\delta x)^3$$

or

$$y - 1 = -3(x-1) + 9(x-1)^2 - 27(x-1)^3.$$

The difference between the true δy and the cubic δy is equal to

$$\left(\frac{-27}{1+3\delta x} - (-27)\right)(\delta x)^3 = \frac{81}{1+3\delta x}(\delta x)^4$$

or nearly $81(\delta x)^4$ if $|\delta x| << 1$. Say $\delta x = 10^{-2}$, then $81(\delta x)^4 = 0.71\ 10^{-6}$, nearly.

3 **Limits**

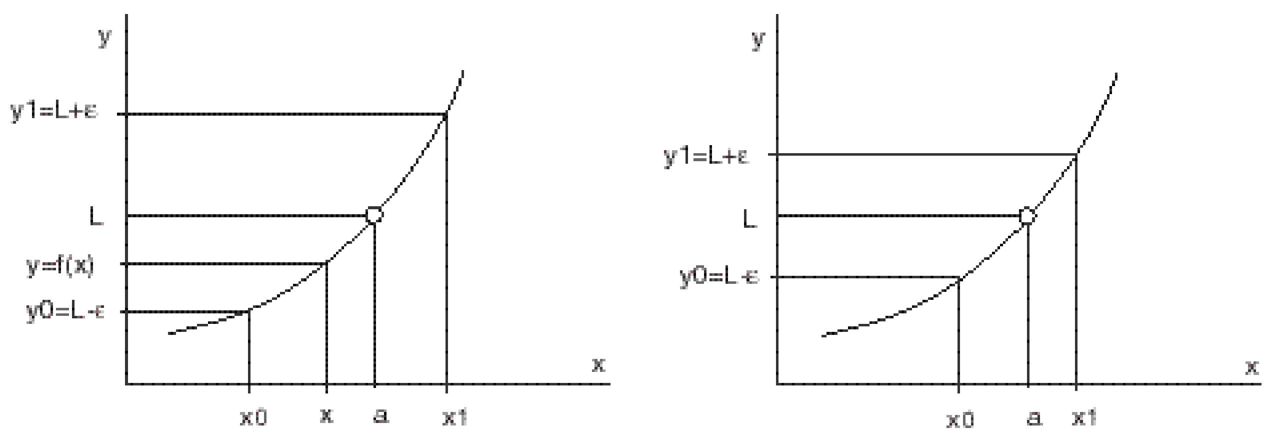
3.1 Intervals and Neighborhoods

Open interval (a, b) contains all points x such that $a < x < b$. **closed** interval $[a, b]$ contains all points x such that $a \leq x \leq b$. A **neighborhood** of x_0 includes all point x in an open interval containing x_0. Symmetric interval $|x - x_0| < \epsilon$, $\epsilon > 0$, contains all points x such that $x_0 - \epsilon < x < x_0 + \epsilon$.

Interval $a < x < b$ is open ended in the sense that, in theory, number x can be selected willfully close to the end points a and b yet no x values exist that are the least and the most in the interval. The set of numbers $a < x < b$ is obviously bound from below and from above. Its **least upper bound(lub)** or **supremum** is b; $\sup(a < x < b) = b$. Its **greatest lower bound(glb)** or **infimum** is a; $\inf(a < x < b) = a$. The interval $a < x < b$ does not contain a minimal value nor a maximal value since both its supremum and infimum are not in the interval. On the other hand the set of numbers $a \leq x \leq b$ attains the **minimum** a; $\min(a \leq x \leq b) = a$ and the **maximum** b; $\max(a \leq x \leq b) = b$.

3.2 **Limits of Functions**

Definition: *Let $f(x)$ be defined for any x except for possibly $x = a$. $\lim_{x\to a} f(x) = L$ means that $f(x)$ can be restricted to a willfully small neighborhood of L with x restricted a sufficiently small neighborhood of a.* See the figure below.



Figuratively speaking the limit of function $f(x)$ as $x \to a$ is the value of $f(x)$ immediately to the right of a and equally the value of $f(x)$ immediately to the left of a. What English means by immediately is the next moment in time with no gap between the first moment and the one that follows. But quantitatively immediately is paradoxical, for if the distance between the two moments is zero, then they are the same, and if there is a stretch of time between them, then there is no immediacy. It has all to do with the basic arithmetical fact that numbers exist that while being different from zero may be arbitrarily small. There exists an arithmetical intermediate state between to be and not to be, namely to be no end tiny. In calculus we are in constant need to move but in a movement that is imperceptibly small. Calculus decides then that the very small is the ultimately small, and that the very large is the ultimately large, so as to remove the uncertainty of size.

Say we wish to compute the value of $y = x^2$ immediately to the right and left of $x = 2$. To do this we undertake the computation of $y = f(x)$ at $x = 2+\epsilon$. We are now confronted by the apparently contradicting demands that ϵ be different from zero yet be practically zero. We may not select any **definite** numerical value for ϵ lest it may not be small enough for $x = 2+\epsilon$ to qualify for being immediate to 2. So instead we compute $y(2+\epsilon) = (2+\epsilon)^2 = 4+4\epsilon+\epsilon^2$ and **reason** that that as $\epsilon \to 0$, $f(2+\epsilon) \to 4$. The values of $f(x) = x^2$ immediately to the right and to the left of $x = 2$ are $y = 4$. It also happens that $y(2)$ is defined here and at exactly $x = 2, y = 4$.

Examples:

1. We shall verify that $\lim_{\delta x \to 0}(4 + 3\delta x) = 4$. Say we wish to limit $f(\delta x) = 4 + 3\delta x$ to the interval $4 - \epsilon \leq f(\delta x) \leq 4 + \epsilon$, where $\epsilon > 0$. Then

$$4 - \epsilon \leq 4 + 3\delta x \leq 4 + \epsilon$$

and it readily results that $-\epsilon/3 \leq \delta x \leq \epsilon/3$.

2. We shall verify that $\lim_{\delta x \to 0} 1/(2-3\delta x) = 1/2$. Say we wish to limit $f(\delta x) = 1/(2-3\delta x)$ to the interval $1/2 - \epsilon \leq f(\delta x) \leq 1/2 + \epsilon$, where $\epsilon > 0$. Then

$$\frac{1}{2} - \epsilon \leq \frac{1}{2 - 3\delta x} \leq \frac{1}{2} + \epsilon$$

or

$$-\epsilon \leq \frac{1}{2 - 3\delta x} - \frac{1}{2} \leq \epsilon$$

which by common denominator becomes

$$-\epsilon \leq \frac{3\delta x}{4 - 6\delta x} \leq \epsilon.$$

Restricting δx to be such that $4 = 6\delta x > 0$ we obtain, through multiplication by $4 - 6\delta x$, that

$$\frac{-4\epsilon}{3 - 6\epsilon} \leq \delta x \leq \frac{4\epsilon}{3 + 6\epsilon}$$

provided that $\epsilon < 1/2$.

3. We shall verify that $\lim_{\delta x \to 0} \sqrt{4 - 3\delta x} = 2$. Say we wish to limit $f(\delta x) = \sqrt{4 - 3\delta x}$ to the interval $2 - \epsilon \leq f(\delta x) \leq 2 + \epsilon$, where $\epsilon > 0$. For this, δx need be such that

$$2 - \epsilon \leq \sqrt{4 - 3\delta x} \leq 2 + \epsilon.$$

If $\epsilon < 2$, then we may square the inequalities so as to have

$$4 - 4\epsilon + \epsilon^2 \leq 4 - 3\delta x \leq 4 + 4\epsilon + \epsilon^2$$

or

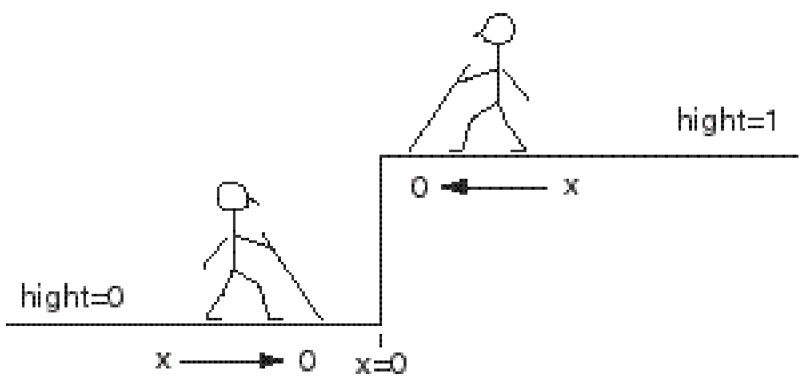
$$\frac{1}{3}(4\epsilon - \epsilon^2) \geq \delta x \geq \frac{1}{3}(-4\epsilon - \epsilon^2).$$

Remark: Notice that the statement $\lim_{x\to x_0} f(x) = L$ contains the implication that $f(x)$ exists in some neighborhood of $x = x_0$.

Theorem: *The limit of a function is unique.*

Proof: Say L_1 and L_2 are two different limits of $f(x)$ as $x \to x_0$. By restricting x to a proper neighborhood of $x_0, f(x)$ can be restricted to a willfully small neighborhood of L_1 that excludes L_2, and vice versa. End of proof.

Finding the limit of $f(x)$ as $x \to x_0$ can be looked upon as a **process** by which x is taken closer and closer to x_0 with the intention of discovering the ultimate value of $f(x)$. But $\lim_{x\to x_0} f(x)$ can be different than $f(x_0)$ even if $f(x_0)$ is defined at x_0.



examples:

1. Let $f(x) = |x|/x$. At $x = 0, f(0) = 0/0$ is ill fined — $0/0$ being an ambiguity. Trying to find $\lim_{x\to 0} |x|/x$ we start with the sequence $x = -1, -1/2, -1/3, -1/4 \ldots$, which tends to zero from below, and for which we always compute $f(x) = -1$, tempting us to believe in the limit -1. But the sequence $x = 1, 1/2, 1/3, 1/4 \ldots$, which tends to zero from above, yields $f(x) = 1$, tempting us to believe that the limit is 1. Since the limit may not depend on the approach, we conclude that there is no limit to $|x|/x$ as $x \to 0$.

2. Let $f(x) = (-1)^x$ if $x \neq 0$ and $f(x) = 1$, if $x = 0$. To try and find $\lim_{x\to 0} f(x)$ we start with the sequence $x = 1/3,\ 1/5,\ 1/7,\ 1/9 \ldots$ that clearly tends to zero. Correspondingly we always find $f(x) = -1$ tempting us to believe that the limit is -1. But the sequence $x = 2/3,\ 2/5,\ 2/7,\ 2/9 \ldots$ that also tends to zero predicts a limit of 1. Since the limit may not depend on the way $x \to x_0$ we conclude that $\lim_{x\to 0}(-1)^x$ does not exist, even though $(-1)^0 = 1$.

3. Let $f(x) = \sin x$. Trying to find $\lim_{x\to\infty} f(x)$ we select the sequence $x = 0,\ \pi,\ 2\pi,\ 3\pi \ldots$ for which we compute $f(x) = 0$, tempting us to believe that the limit is zero. But the sequence $x = \pi/2,\ 5\pi/2,\ 9\pi/2 \ldots$ for which we compute $f(x) = 1$ tempts us to believe that the limit is one. Since the limit must be independent of the way $x \to x_0$ we conclude that $\lim_{x\to\infty} \sin x$ does not exist.

4. That $\lim_{x\to 0}(1/x^2)$ does not exist results from the fact that $1/x^2$ becomes ever bigger as $|x|$ becomes ever smaller.

5. Here

$$\lim_{x\to 2}\frac{x^2-4}{x-2}=\lim_{x\to 2}\frac{(x-2)(x+2)}{x-2}=\lim_{x\to 2}(x+2)=4,$$

even though $(x^2+4)/(x-2)$ is not defined at $x=2$.

3.3 **Limits of Multiplied Functions**

If $\lim_{x\to a} f(x)=L$, then $\lim_{x\to a}(L+(f(x)-L))=L$. Hence $\lim_{x\to a} f(x)=\lim_{x\to a}(L+\epsilon(x))$ where $\epsilon(x)=f(x)-L\to 0$ as $x\to a$.

Lemma: *If* $\lim_{x\to a}\epsilon(x)=0$, *then also* $\lim_{x\to a}(c\epsilon(x))=0$ *for any constant* c.

Proof: Since $\epsilon(x)\to 0$ as $x\to a$, neighborhoods of $x=a$ exist for which $|\epsilon(x)|$ can be made willfully small. Consequently $|c\epsilon(x)|$ can also be made willfully small. End of proof.

Lemma: *Let* $f(x)$ *be* **bounded** *in some neighborhood of* $x=a$, *namely* $|f(x)|<B$, *for some interval around* $x=a$. *Then* $\lim_{x\to a} f(x)\epsilon(x)=0$ *if* $\lim_{x\to a}\epsilon(x)=0$.

Proof: Since $|f(x)\epsilon(x)|=|f(x)||\epsilon(x)|<B|\epsilon(x)|$, according to the previous lemma if $\epsilon(x)\to 0$ as $x\to 0$, then so does $B|\epsilon(x)|$ and consequently so does $f(x)\epsilon(x)$.

Lemma: $1\le 1/(1-\epsilon)\le 1+n\epsilon$ *if* $0\le\epsilon\le(n-1)/n,\ n>1$.

Theorem *If* $\lim_{x\to a} f(x)=L_1$ *and* $\lim_{x\to a} g(x)=L_2$, *then*

1. $\lim_{x\to a}(f+g)=L_1+L_2$.
2. $\lim_{x\to a}(fg)=L_1L_2$.
3. $\lim_{x\to a}(1/g)=1/L_2$, *provided* $L_2\ne 0$.
4. $\lim_{x\to a}(f/g)=L_1/L_2$, provided $L_2\ne 0$.

Proof: Write $f(x)=L_1+\epsilon_1(x), g(x)=L_2+\epsilon_2(x)$ where $\epsilon_1(x)\to 0$ and $\epsilon_2(x)\to 0$ as $x\to a$. Then

1. $f+g=(L_1+L_2)+(\epsilon_1+\epsilon_2)$
2. $fg=(L_1L_2)+(L_1\epsilon_2+L_2\epsilon_1+\epsilon_1\epsilon_2)$.
3. $\dfrac{1}{g}=\dfrac{1}{L_2+\epsilon_2}=\dfrac{1}{L_2}-\dfrac{\epsilon_2}{L_2(L_2+\epsilon_2)}=\dfrac{1}{L_2}-\dfrac{\epsilon_2}{L_2^2(1+\epsilon_2/L_2)}$

and since $L_2\ne 0$ and $\epsilon_2\to 0$ as $x\to a, 1/(1+\epsilon_2/L_2)$ is, according to the lemma, bounded near $x=a$.

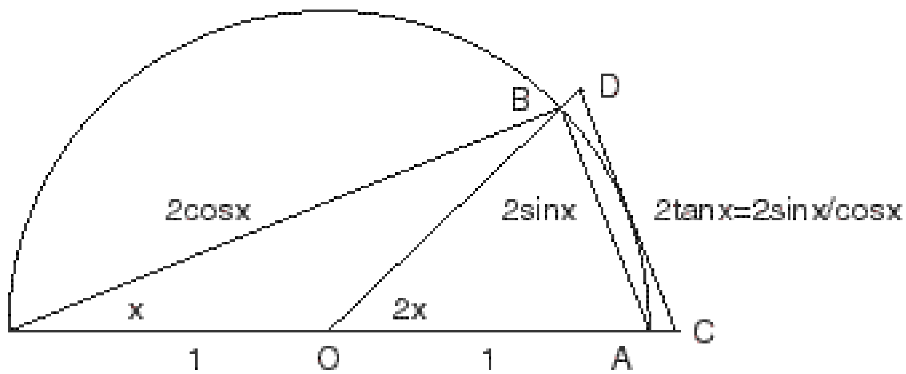
4. $\dfrac{f}{g}=f\dfrac{1}{g}$.

End of proof.

Theorem: *Let $f(x)$ be defined in some neighborhood of $x = a$, possibly excluding a itself.If $\lim_{x\to a} f(x) = L > 0$, then a neighborhood of a, excluding $x = a$ exists at which $f(x) > 0$.*

proof: A neighborhood of a, $x \neq a$, exists at which $L - \epsilon \leq f(x) \leq L + \epsilon$, with a willfully small $\epsilon > 0$.

3.4 Limits of Trigonometric Functions



Theorem: $\lim_{x\to 0} \sin x = 0$. *and* $\lim_{x\to 0} \cos x = 1$

proof: Looking at the figure above we see that, for $x > 0$, $0 < \sin x < x$. Also, $\cos x = \sqrt{1 - \sin^2 x}$.

Theorem: $\lim_{x\to 0} \sin x/x = 1$.

proof: Looking at the figure above we see that the area of sector OAB is greater then the area of triangle OAB but less than triangle OCD. Or $\sin x \cos x < x < \sin x/ \cos x$. so that

$$\cos x < x/ \sin x < 1/ \cos x.$$

But $\cos x \to 1$ as $x \to 0$. End of proof.

3.5 One Sided Limits

Definition: $\lim_{x\to a^+} f(x) = R$ *is the* **right** *limit of $f(x)$ if a neighborhood $a < x < a + \delta$, $\delta(\epsilon) > 0$ esists to the right of a so that for every x in this neighborhood $L-\epsilon < f(x) < L+\epsilon$, for any willfully small ϵ. Similarly,* $\lim_{x\to a^-} f(x) = L$ *is the* **left** *limit of $f(x)$ as x tends to a from the left.*

Example: $\lim_{x\to 0^+} \sqrt{x} = 0$.

theorem: $\lim_{x\to a} f(x) = L$ *if and only if* $\lim_{x\to a^-} f(x) = \lim_{x\to a^+} f(x) = L$.

theorem: *If function $f(x) \geq 0$ for all $x > a$, then* $\lim_{x\to a^+} f(x) \geq 0$.

3.6 Limits when $x \to \infty$ or when $x \to -\infty$

Notice that a^+ and a^- are not numbers but mere abbreviations. Hence if variable x of function $f(x)$ is restricted to open interval $a < x < b$, then $f(a^+)$ and $f(b^-)$, the values of $f(x)$ for x immediately to the right of $x = a$ and the left of $x = b$, namely $f(a^+) = \lim_{x \to a^+} f(x)$ and $f(b^-) = \lim_{x \to b^-} f(x)$, are conventionally excluded from the range of $f(x)$. Say $f(x) = x,\ 0 < x < 1$. Since x must be definitely numerical, $0 < f(x) < 1$ even though number x may be theoretically selected to be as close zero or as close to one as desired.

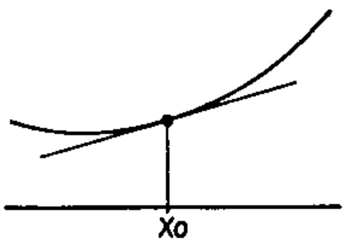
$\lim_{x \to 0^+} 1/x = \infty$ means that $1/x$ can be made willfully large by selecting a sufficiently small $x > 0$. Silarly, $\lim_{x \to 0^-} 1/x = -\infty$ means that $|1/x|$ can be made willfully large by selecting a sufficiently small $|x|,\ x < 0$.

Definition: $\lim_{x \to \infty} f(x) = L$ *means that positive number* $N(\epsilon)$ *exists so that for any* $x > N$ $L - \epsilon < f(x) < L + \epsilon$ *for any willfully small* $\epsilon > 0$.

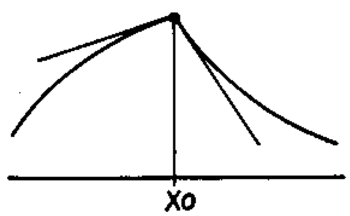
Examples:

1. $\lim_{x \to \infty} 1/x = 0$ since for any $x > 1/\epsilon,\ 1/x < \epsilon$. In other words, $1/x$ can be made willfully small with a sufficiently large x.

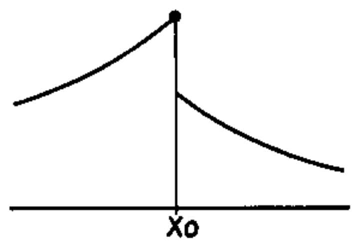
2 $\lim_{x \to \infty}(2x + 1)/(3x - 2) = \lim_{x \to \infty}(2 + 1/x)/(3 - 2/x) = 2/3$.



function continuous and differentiable at x_0



function continuous but not differentiable at x_0



function discontinuous and therfore also not differentiabe at x_0

4. Continuity

4.1 Continuous functions:

Changes can occur gradually, or they may happen suddenly – for better or worse.

Definition: *Function* $f(x)$ *is continuous at* $x = x_0$ *if the values of the function immediately to the right and immediately to the left of* x_0 *are both equal to* $f(x_0)$. *Otherwise the function is discontinuous. Or, function* $f(x)$ *is continuous at* x_0 *if and only if* $\lim_{x \to x_0} f(x) = f(x_0)$.

Examples:

1. Function $y = x^2$ is continuous everywhere since $\lim_{x \to x_0} x^2 = {x_0}^2$ whatever x_0.

2. Function $y = \sqrt{x}$ is continuous at all $x > 0$, since $\lim_{x \to x_0} \sqrt{x} = \sqrt{x_0}$ if $x_0 > 0$. Notice that at $x = 0$, $\sqrt{x}$ has a right limit only, $\lim_{x \to 0^+} \sqrt{x} = 0$.

3. Function $y = 1/x$ is continuous at all $x \neq 0$ since $\lim_{x \to x_0} 1/x = 1/x_0$ as long as $x \neq 0$.

4. Function

$$f(x) = \begin{cases} x, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$$

is continuous everywhere, including $x = 1$, since $\lim_{x \to 1^-} f(x) = \lim_{x \to 1^+} f(x) = f(1) = 1$.

4.2 Properties of continuous functions

Theorem (continuity of product functions): *If functions f and g are continuous at x_0, then*

1. *cf is continuous at x_0 for any value of constant c.*
2. *$f + g$ is continuous at x_0.*
3. *$f \cdot g$ is continuous at x_0.*
4. *f/g is continuous at x_0 if $g(x_0) \neq 0$.*

Proof: It all results from the theorem on the limits of product functions.

Theorem (the continuity of compound functions): *Let $f(x)$ and $g(x)$ be two functions such that $\lim_{x \to x_0} g(x) = y_0$, and $f(x)$ being continuous at y_0, then $\lim_{x \to x_0} f(g(x)) = f(y_0)$.*

Proof: $x \to x_0$ implies $y \to y_0$ which in turn implies $f(y) \to f(y_0)$. End of proof.

Corollary: *If function $g(x)$ is continuous at x_0 and $f(x)$ is continuous at $y_0 = g(x_0)$, then the compound function $f(g(x))$ is continuous at x_0.*

Theorem (boundness of continuous functions): *If function $f(x)$ is continuous at every point of closed interval $a \leq x \leq b$, then $f(x)$ is bounded in this interval.*

Proof: Function $f(x)$ is defined for any x in the interval. End of proof.

To observe the necessity of the closeness of the interval in the above theorem consider the function $f(x) = x^{-1}$, $0 < x \leq 1$, which is continuous at every $x > 0$. A value for x^{-1} can be theoretically computed for any x no matter how small. However, as $x \to 0$, $f \to \infty$. The condition $x > 0$ is open ended yet excludes $x = 0^+$ at which $f = \infty$.

Theorem: *If $f(x)$ is continuous at x_0, and $f(x_0) > 0$, then $f(x) > 0$ in some neighborhood of x_0*

The intermediate value Theorem: *If function $y = f(x)$ is continuous in the closed interval $a \leq x \leq b$ and is such that $f(a) < 0$ and $f(b) > 0$, then $f(x) = 0$ at least once in the interval $a < x < b$.*

Proof: The method of bisection will locate x so that $f(x) \to 0$ as $x \to x_0$.

The extremal values theorem: *If function $y = f(x)$ is continuous in the closed interval $a \leq x \leq b$, then there exists at leasat one value of x in this interval at which $f(x)$ attains its highest value, and at least one value of x in the interval at which $f(x)$ attains its lowest value.*

Proof: Function $f(x)$ is defined for any x in the interval. End of proof.

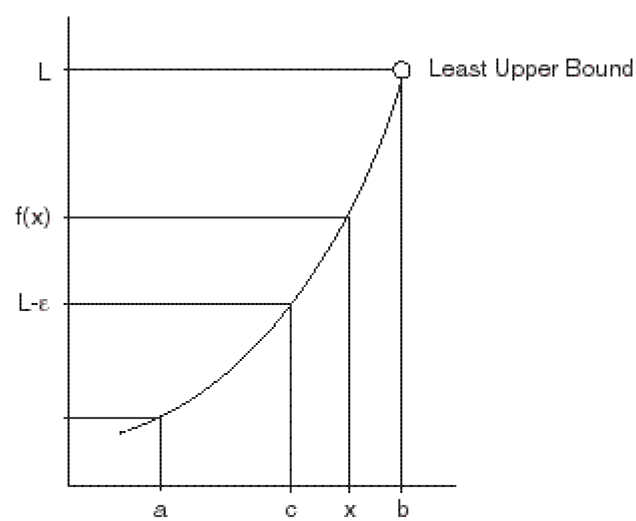
To understand the point of this theorem consider the counter example of the discontinuous function.

$$f(x) = \begin{cases} 2x - x^2 & x \neq 1 \\ 0, & x = 1. \end{cases}$$

The curve of this function is perforated at $x = 1$ and it reaches its peak of 1 at $x = 1^-$ and $x = 1^+$ which are excluded from $x \neq 1$. Hence no x exists for which $f(x) = 1$.

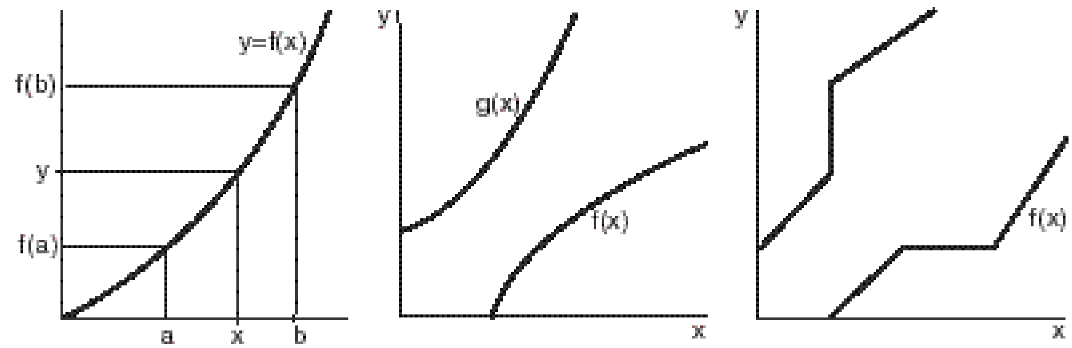
Theorem: *Let $f(x)$ be an increasing function in the open interval $a < x < b$, namely $f(x_1) > f(x_0)$ for any $x_1 > x_0$ in the interval. If $f(x)$ is bounded from above, with L being its least upper bound, then $\lim_{x \to b^-} f(x) = L$.*

Proof: If L is the least upper bound on $f(x)$, then $f(x) \leq L$ for any x in the interval. Let $\epsilon > 0$ be willfully small. For any such ϵ, $L - \epsilon$ is not an upper bound on $f(x)$. Hence for any $c < x < b$, $L - \epsilon < f(x) < L$. End of proof. See the figure below.



The invertibility theorem: *If continuous function $f(x)$ is increasing in $a \leq x \leq b$, then $f(x)$ is invertible in $a \leq x \leq b$, and the inverse function $y = g(x)$ is both continuous and increasing in the interval $f(a) \leq x \leq f(b)$.*

See the figure below.



For one x there is one y and conversely

Function f(x) and its inverse function g(x)

Lack of an inverse on flat section of f(x)

5. Differentiation

5.1 The derivative function

Function $f'(x)$ is derived from function $f(x)$ by a process of differentiation.

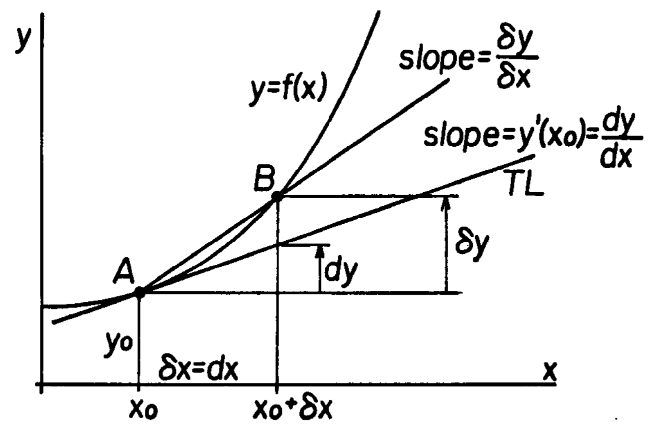
Definition: *Let function $y = f(x)$ be defined in some neighborhood of $x = x_0$. Introduce the new variables δx and δy so that $x = x_0 + \delta x$, $y = y_0 + \delta y$, $y = f(x_0)$. Then*

$$\delta y = \left(\frac{\delta y}{\delta x}\right)\delta x = \frac{f(x_0+\delta x) - f(x_0)}{\delta x}\delta x,$$

and the **derivative function** *f' of f at x_0 is defined as*

$$f'(x_0) = \lim_{\delta x\to 0}\frac{\delta y}{\delta x} = \lim_{\delta x\to 0}\frac{f(x_0+\delta x) - f(x_0)}{\delta x}$$

provided that the limit exists (and is independent of the way δx tends to zero). When $f'(x_0)$ exists function f is said to be **differentiable** *at $x = x_0$. If $f(x)$ is differentiable at $x = x_0$, then $f(x)$ is also continuous at $x = x_0$.*



5.2 The fundamental statement of calculus

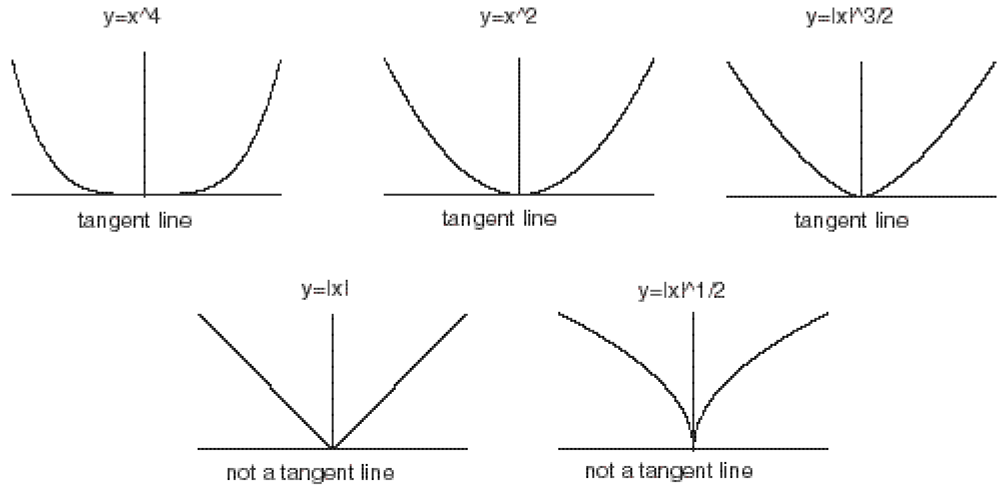
The linear relationship between **differentials** dx and dy is

$$dy = y'(x_0)dx,\ \ y - y_0 = y'(x_0)(x - x_0),\ \ \delta x = dx,\ \ \delta y = dy + e\delta x$$

where

$$e(x_0, \delta x) = \frac{\delta y}{\delta x} - f'(x_0), \quad \text{and} \quad \lim_{\delta x\to 0} e(x_0, \delta x) = 0.$$

Geometrically speaking, $y - y_0 = y'(x_0)(x - x_0)$ is the **tangent**, or **osculating** line to $y = f(x)$ at $x = x_0$. This line constitutes an optimal linear approximation to the function in the sense that $\delta y - dy = e(x_0, \delta x)\delta x$ with $e \to 0$ as $\delta x \to 0$. However, the tendency of $e(x_0, \delta x)$ to zero may be slow. Consider $y = x^{1+\epsilon}$, $\epsilon = 2n/(2n-1)$ for some natural n. At $x_0 = 0$, $\delta y = y = \delta x \cdot \delta x^{\epsilon}$ and $e(0, \delta x) = \delta x^{\epsilon}$. See the figure.



The differential relationship $dy = f'(x_0)dx$ or $dy/dx = f'(x_0)$ is often used to represent the derivative function. Writing dy/dx for $f'(x)$ makes clear that the reference is to a derivative function of $f(x)$ with respect to independent variable x. In this notation $(f')' = f''$, the **second order** derivative of $f(x)$ is represented by $dy'/dx = d^2y/dx^2$, and so on.

5.3 **Theorem (the derivatives of compound functions):**

1. $(cf)' = cf'$ for any value of constant c.

2. $(f+g)' = f' + g'$.

3. $(fg)' = f'g + fg'$.

4. $(1/f)' = -f'/f^2$.

5. $(f/g)' = (f'g - fg')/g^2$

Proof: We use the concise notation $f_1 = f(x)$, $f_2 = f(x + \delta x)$. By the assumption that f is continuous at x, $\lim_{\delta x \to 0} f_2 = f_1$, and

1. $$(cf)' = \lim_{\delta x \to 0} \frac{cf_2 - cf_1}{\delta x} = c \lim_{\delta x \to 0} \frac{f_2 - f_1}{\delta x}.$$

2. $$(f+g)' = \lim_{\delta x \to 0} \frac{(f_2 + g_2) - (f_1 + g_1)}{\delta x} = \lim_{\delta x} \frac{f_2 - f_1}{\delta x} + \frac{g_2 - g_1}{\delta x}.$$

3. $$(fg)' = \lim_{\delta x \to 0} \frac{f_2 g_2 - f_1 g_1}{\delta x} = \lim_{\delta x \to 0} \frac{f_2 g_2 - f_1 g_1 + f_1 g_2 - f_1 g_2}{\delta x}$$
$$= \lim_{\delta x \to 0} \frac{g_2(f_2 - f_1)}{\delta x} + \frac{f_1(g_2 - g_1)}{\delta x}.$$

4. $$(1/f)' = \lim_{\delta x \to 0} \frac{1/f_2 - 1/f_1}{\delta x} = \lim_{\delta x \to 0} -\frac{1}{f_1 f_2} \frac{f_2 - f_1}{\delta x}.$$

Theorem (the chain rule:) *Consider the functional chain* $y = f(u),\ u = g(x)$. *If* $g(x)$ *is differentiable at* x *and* $f(u)$ *is differentiable at* $u = g(x)$, *then* $y'(x) = f'(u)g'(x)$.

Proof: Let $u_0 = g(x_0)$ and $y_0 = f(u_0)$. Write $x = x_0 + \delta x,\ u = u_0 + \delta u,\ y = y_0 + \delta y$ in which only δx is independent. Now $u_0 = g(x_0)$ and $y_0 = f(u_0)$. Write $x = x_0 + \delta x$, $u = u_0 + \delta u,\ y = y_0 + \delta y$ in which only δx is independent. Now

$$\delta u = \frac{g(x_0 + \delta x) - g(x_0)}{\delta x}\delta x, \quad \delta y = \frac{f(u_0 + \delta u) - f(u_0)}{\delta u}\delta u$$

and since $g(x)$ is differentiable at $x = x_0$, $\delta u \to 0$ as $\delta x \to 0$. Hence

$$\delta y = \frac{f(u_0 + \delta u) - f(u_0)}{\delta u}\frac{g(x_0 + \delta x) - g(x_0)}{\delta x}\delta x$$

and

$$dy = f'(u_0)g'(x_0)dx, \quad dy = y'(x_0)dx.$$

End of proof.

5.4 **Inverse functions**

Let $y = f(u)$ and $u = g(x)$ be such that $y = x$. Then g is the inverse function of f, $f(g(x)) = x$. For instance $y = u^2$, $u = \sqrt{x}$. From the chain rule we have that $y'(x) = f'(u),\ g'(x)$, or with $y' = 1$, $1 = f'(u)g'(x)$.

Examples:

1. For $y = u^2, u = \sqrt{x}$ we have that $1 = 2u(\sqrt{x})'$ or $(\sqrt{x})' = 1/(2u) = 1/(2\sqrt{x})$.

2. For $y = u^n, u = x^{\frac{1}{n}}$ we have that $1 = nu^{n-1}(x^{\frac{1}{n}})'$ and $(x^{\frac{1}{n}})' = (1/n)u^{1-n} = (1/n)x^{(1-n)/n}$.

3. For $y = \sin u,\ u = \sin^{-1} x$ we have that $1 = \cos u(\sin^{-1} x)'$ and $(\sin^{-1} x)' = 1/\cos u = 1/\sqrt{1 - x^2}$.

5.5 **Derivatives of some elementary functions**

1. $(1)' = 0$	2. $(x)' = 1$
3. $(x^2)' = 2x$	4. $(x^3)' = 3x^2$
5. $(\sqrt{x})' = 1/(2\sqrt{x})$	6. $(1/x)' = -1/x^2$
7. $(x^n)' = nx^{n-1}$	8. $(\ln x)' = 1/x$
9. $(e^x)' = e^x$	10. $(\cos x)' = -\sin x$
	11. $(\sin x)' = \cos x$

6. Properties of differentiable functions

Theorem: *Let $f(x)$ be differentiable at x_0. If $f'(x_0) > 0$, then there is a neighborhood of x_0 in which $f(x_0 + \delta x) > f(x_0)$ if $\delta x > 0$, and $f(x_0 + \delta x) < f(x_0)$ if $\delta x < 0$.*

Proof: Since $f(x)$ is differentiable at x_0

$$f'(x_0) - \epsilon < \frac{f(x_0 + \delta x) - f(x_0)}{\delta x} < f'(x_0) + \epsilon, \ \epsilon > 0.$$

We may select, willfully small, ϵ so that $\epsilon_1 = f'(x_0) - \epsilon > 0$. Then $f(x_0 + \delta x) > \epsilon_1 \delta x + f(x_0)$ if $\delta x > 0$, and $f(x_0 + \delta x) < \epsilon_1 \delta x + f(x_0)$ if $\delta x < 0$. End of proof

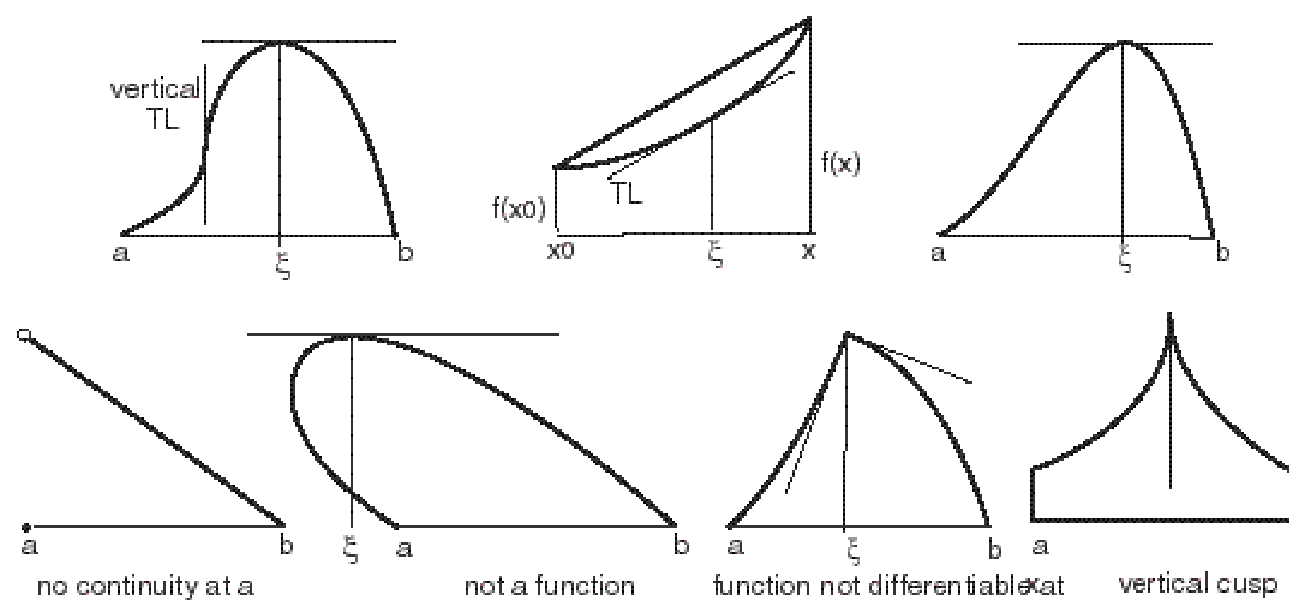
Theorem (Fermat): *Let $f(x_0)$ be maximal at $x = x_0$, $f(x) < f(x_0)$, $x \neq x_0$, in some neighborhood of x_0. If $f(x)$ is differentiable at $x = x_0$, then $f'(x_0) = 0$.*

Proof: It can not be that $f(x_0) > 0$ because this would imply that $f(x) > f(x_0)$ for some $x > x_0$, and it can also not happen that $f(x_0) < 0$, since this would imply that $f(x) > f(x_0)$ for some $x < x_0$. End of proof.

Theorem (Rolle): *Let function $f(x)$ be continuous on $a \leq x \leq b$ and differentiable on $a < x < b$. If, in addition, $f(a) = f(b) = 0$, then $f'(\xi) = 0$ for some $a < \xi < b$.*

Proof: Discount the trivial case of $f(x) = 0$. Because it is continuous, function $f(x)$ reaches a maximum (minimum) inside the interval. At this point $f' = 0$. End of proof.

see the figure below



The Mean Value Theorem (MVT): *If $f(x)$ is continuous in the closed interval $a \leq x \leq b$ and is differentiable in the open interval $a < x < b$, then there is at least one point ξ inside* *the interval so that*

$$\frac{f(b)-f(a)}{b-a} = f'(\xi), \quad a < \xi < b$$

or

$$f(x) = f(x_0) + (x-x_0)f'(\xi), \quad x_0 < \xi < x$$

or

$$\delta y = f'(\xi)\delta x, \quad x_0 < \xi < x_0 + \delta x$$

as compared to $dy = f'(x_0)dx$.

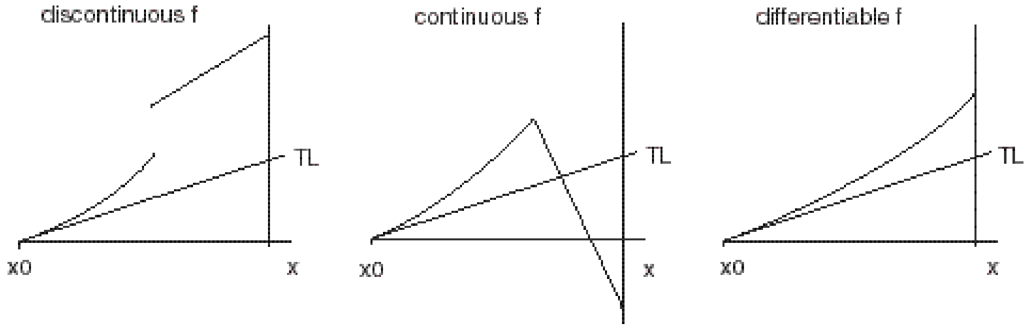
Proof: Define

$$F(x) = f(x) - f(a) - (x-a)R,\ 0 = f(b) - f(a) - (b-a)R$$

so that $F(a) = F(b) = 0$. Also $F'(x) = f'(x) - R$, and hence by Rolle's theorem $R = f(\xi)$. End of proof.

The linear approximation $y = f(x_0) + f'(x_0)(x - x_0)$ can be looked upon as a **predictor** for the function $f(x)$. This prediction can be of low quality if function f suffers a discontinuity between x_0 and x. The prediction is more trustworthy if the function is continuous between x_0 and x, and is still more reliable if the function is not only continuous but also differentiable between x_0 and x.

See the figure below



The constant function theorem: *If $f'(x) = 0$ for all x in an interval, then f is constant in the interval.*

Proof: The MVT assumes that $f(b) = f(a)$ for any a and b in the interval. End of proof.

Corollary: If $(f - g)' = 0$, then $f - g =$ constant.

The Cauchy Mean Value Theorem (CMVT): *Let $f(x)$ and $g(x)$ be continuous on $a \leq x \leq b$ and differentiable in $a < x < b$ with $f'(x)$ and $g'(x)$ not being zero at once. Then*

$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(\xi)}{g'(\xi)}, \quad a < \xi < b.$$

Proof: Write

$$F(x) = f(x)(g(b) - g(a)) - g(x)(f(b) - f(a))$$

and observe that $F(a) = F(b) = 0$. By Rolle's theorem $F'(\xi) = 0,\ a < \xi < b$, or

$$f'(\xi)(g(b) - g(a)) - g'(\xi)(f(b) - f(a)) = 0.$$

End of proof.

Theorem: *If function f is continuous and differentiable in some neighborhood of x_0, then $\lim_{x \to x_0} f'(x) = f'(x_0)$.*

Proof: By the MVT

$$\frac{f(x_0 + \delta x) - f(x_0)}{\delta x} = f'(\xi),\ x_0 < \xi < x_0 + \delta x$$

and hence

$$\lim_{\delta x \to 0} f'(\xi) = f'(x_0).$$

End of proof.

The derivative function can not experience a jump discontinuity since $f'(x) = \lim_{\delta x \to 0} (\delta y / \delta x)$ independently of the way $\delta x \to 0$. The derivative function may yet be discontinuous at $x = x_0$ in the sense that $\lim_{x \to x_0} f(x) \neq f'(x_0)$. For instance

$$f(0) = 0,\ f(x) = x^2 \sin(1/x),\ x \neq 0$$

for which $f'(0) = 0,\ f'(x) = x\sin(1/x) - \cos(1/x),\ x \neq 0.$

Theorem (Darboux): *If $f(x)$ is differentiable on $a \leq x \leq b$, and if $f'(a) > 0$ and $f'(b) < 0$, or vice versa, then $f'(\xi) = 0,\ a < \xi < b$.*

The interesting aspect of this theorem is that it does not require the continuity of f', merely its existence.

Proof: We shall assume, for the sake of simplicity, that $f(a) = f(b) = 0$. Since f is differentiable on the interval it is also continuous on it. Because $f'(a) > 0,\ f(x) > 0$ for some $x > a$, and because $f'(b) < 0,\ f(x) > 0$ for some $x < b$. The function reaches its maximum at $a < \xi < b$ and $f'(\xi) = 0$.

L' Hôpital's rule: If f and g are differentiable functions on $a < x < b$ except possibly at point $x = x_0$, $a < x_0 < b$, and if

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

and if $g'(x) \neq 0$ for all $x \neq x_0$ in $a < x < b$, then

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

Proof: We may assume that $f(x_0) = g(x_0) = 0$ since the values of the functions at x_0 has no effect on the limit. Then by the MVT

$$f(x) = (x - x_0)f'(\xi),\ g(x) = (x - x_0)g'(\eta),\ x_0 < \xi,\ \eta < x$$

and

$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = \lim_{x \to x_0} \frac{f'(\xi)}{g'(\eta)}.$$

End of proof.

The Extended Mean Value theorem (EMVT): *Let f be continuous in the closed interval between x_0 and x. Let also f' and f'' be defined in the open interval (actually on $a^+ < x < b^-$) between x_0 and x, then*

$$f(x) = f(x_0) + (x - x_0)f'(x_0) + \frac{1}{2}(x - x_0)^2 f''(\xi),\ x_0 < \xi < x$$

or

$$\delta y = f'(x_0)\delta x + \frac{1}{2}f''(\xi)(\delta x)^2,\ x_0 < \xi < x_0 + \delta x.$$

Proof: The proof to the EMVT is based on a double application of Rolle's theorem to

$$F(x) = f(x) - f(a) - (x - a)f'(a) - \frac{1}{2}(x - a)^2 R$$

$$0 = f(b) - f(b) - (b-a)f'(a) - \frac{1}{2}(b-a)^2 R$$

and the observation that $F(a) = F(b)$, and $F'(a) = 0,\ F'(\xi) = 0, a < \xi < b$.

7. The osculating parabola and circle

Osculating line $y - y_0 = f'(x_0)(x - x_0)$ has the property that at point $x_0, y = y_0 = f(x_0)$, and $y' = f'(x_0)$. The extended mean value theorem shows that the osculating line is an excellent approximation to $f(x)$ near $x = x_0$ if $f''(x)$ is bounded there since under these circumstances $f(x) - y = \frac{1}{2}(x - x_0)^2 f''(\xi),\ x_0 < \xi < x$.

7.1 The osculating parabola

The general equation of the parabola is $y = a_2x^2 + a_1x + a_0$ and coefficients a_0, a_1, a_2 can be used to enforce the higher degree of osculation, $y(x_0) = f(x_0),\ y'(x_0) = f'(x_0)$, $y''(x_0) = f''(x_0)$ leading to

$$y - y_0 = y'(x_0)(x - x_0) + \frac{1}{2}y''(x_0)(x - x_0)^2$$

in which $y_0 = f(x_0)$.

Because the parabola can be variously bent it can be made to hug the graph of $f(x)$ to a higher degree of closeness near $x = x_0$ than the tangent line. We will not theoretically pursue this question here but will illustrate it on an example. The osculating (tangent) line to $f(x) = x^{-1}$ at $x = 1$ is $y = -x + 2$, and $y - x^{-1} = -(x-1)^2/x$ so that near $x = 1$ the error in the linear approximation is nearly proportional to $\delta x^2, \delta x = x - 1$. The equation of osculating parabola to $y = x^{-1}$ at $x_0 = 1$ is $y = x^2 - 3x + 3$ so that $y - x^{-1} = (x-1)^3/x$, and the error in the quadratic approximation is now nearly proportional to δx^3, $\delta x = x - 1$, if x is close to 1.

Notice that function $f(x) = x^4$ is so flat near $x = 0$ that its osculating parabola degenerates at this point to the line $y = 0$. On the other hand $f(x) = x^{8/5}$ is too pointed at $x = 0$, $f'' = (24/25)x^{-2/5}$, for it to have an osculating parabola at this point.

The process of quadratization produces an osculating parabola. To see this we recall that

$$\delta y = \frac{f(x_0 + \delta x) - f(x_0)}{\delta x}\delta x,\ dy = f'(x_0)dx.$$

We write

$$\delta y = \left[\frac{f(x_0 + \delta x) - f(x_0)}{\delta x} + f'(x_0) - f'(x_0)\right]\delta x$$

to have

$$\delta y = f'(x_0)\delta x + \left[\frac{f(x_0 + \delta x) - f(x_0)}{\delta x} - f'(x_0)\right]\delta x$$

or

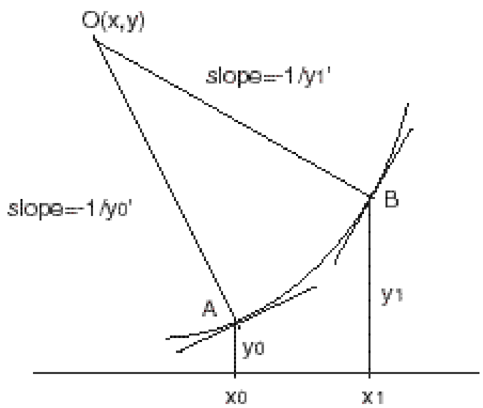
$$\delta y = f'(x_0)\delta x + \frac{f(x_0 + \delta x) - f(x_0) - f'(x_0)\delta x}{\delta x^2}\delta x^2$$

and by L'Hôpital's rule

$$\lim_{\delta x \to 0} \frac{f(x_0+\delta x) - f(x_0) - f'(x_0)\delta x}{\delta x^2} = \lim_{\delta x \to 0} \frac{f'(x_0+\delta x) - f'(x_0)}{2\delta x} = \frac{1}{2} f''(x_0).$$

7.2 The osculating circle

The center of a circle, and consequently its radius is found at the intersection of two normals. Let the arc in the figure be smooth, being the graph of function f with f'' being defined in some neighborhood of x_0. A small portion of the arc around point A appears circular. The two normals raised at point A and B intersect at point O. If point O reaches a limit position as $B \to A$, then point O is the center of the osculating circle to the curve at point A.
See the figure below



The coordinates of point O are found from the intersection of the two normals

$$y - y_0 = -\frac{1}{y_0'}(x - x_0), \; y - y_1 = -\frac{1}{y_1'}(x - x_1)$$

and

$$x = \frac{-y_1 y_1' y_0' + y_0 y_1' y_0' - y_0' x_1 + y_1' x_0}{y_1' - y_0}$$

Application of L'Hôpital's rule renders the limit values

$$x = x_0 - \frac{y_0'}{y_0''}(1 + y_0'^2), \; y = y_0 + \frac{1}{y_0''}(1 + y_0'^2)$$

so that radius r of the limit circle is

$$r = [(x - x_0)^2 + (y - y_0)^2]^{1/2} = \frac{1}{|y_0''|}(1 + y'^2)^{3/2}.$$

The inverse $\kappa = 1/r$ is called the **curvature** of the curve at point A. For a line $\kappa = 0$, and at a corner $\kappa = \infty$.

8. Maxima and Minima

8.1 Increasing (decreasing) function theorem

If function $y = f(x)$ is continuous in the closed interval $a \leq x \leq b$, and $f'(x) > 0$ ($f'(x) < 0$) in the open interval $a < x < b$, then function f is increasing (decreasing) in $a \leq x \leq b$.

Proof: Let x_0 and x, $x < x_0$ be in the interval. By the MVT $f(x) = f(x_0) + (x - x_0)f'(\xi),\ x_0 < \xi < x$. Since $x - x_0 > 0$ and $f'(\xi) > 0$, then $f(x) > f(x_0)$. End of proof.

8.2 Pick-up and Slow-Down; Convexity and Concavity

Function f is increasing around $x = x_0$ if $f(x_0) > 0$ and is decreasing around $x = x_0$ if $f(x_0) < 0$. In case $y''(x_0) > 0$ the slope f' of the tangent line to f increases around $x = x_0$ and the growth of the function is picking up or **accelerating** around $x = x_0$. The function is then **convex** near $x = x_0$. In case $y''(x_0) < 0$ the slope f' of the tangent line to f decreases around $x = x_0$, and the function is then **concave** near $x = x_0$.

Indeed, the EMVT asserts that

$$\delta y = dy + \frac{1}{2}(x - x_0)^2 y''(\xi),\ x_0 < \xi < x$$

and if $y''(\xi) > 0$, then $\delta y > dy$, and if $y'(\xi) < 0$, then $\delta y < dy$. Notice that $y''(x_0)$ need not exist for this to happen.

Parabola $y = ax^2$ is convex if $y''(x) = a > 0$, and is concave if $y''(x) = a < 0$. The convex parabola lies entirely above any of its tangent lines, and the concave parabola lies entirely below any of its tangent lines. Any arc of the convex parabola lies entirely below its cord, and any arc of the concave parabola lies entirely above its cord.

A point on a curve at which the curve changes from being convex to being concave, or vice versa, is a point of **inflection**. Function $f(x) = x^3$ is concave at $x < 0$ and convex at $x > 0$. The function inflects at $x = 0$ where f'' changes sign. Point $x = 0$ is not an inflection point of $f(x) = x^4$, even though $f''(0) = 0$, since $f''(x)$ does not change its sign at $x = 0$.

8.3 Critical points

Point $x = x_0$ is said to be a **critical** point of $f(x)$ if $f'(x_0) = 0$. Around a critical point the MVT and EMVT reduce to

$$f(x) = f(x_0) + (x - x_0)f'(\xi), \quad f(x) = f(x_0) + \frac{1}{2}(x - x_0)^2 f''(\xi) x_0 < \xi << x$$

that can be used to determine the nature (min, max, inflection) of the critical point.

16. The critical points of a function and their nature

Let $x = x_0$ be a critical point of $y = f(x)$, $f'(x_0) = 0$.

Function $y = f(x)$ attains a **local minimum** at $x = x_0$ if in some neighborhood $|x - x_0| < \epsilon$, $\epsilon > 0$ of x_0:

1. $f(x < x_0) > f(x_0)$ and $f(x > x_0) > f(x_0)$ or

2. $f'(x < x_0) < 0$ and $f'(x > x_0) > 0$, since by the MVT

$y(x) = y(x_0) + (x - x_0)f'(\xi)$, $x_0 < \xi < x$, or

3. $f^{''}(x < x_0) > 0$ and $f^{''}(x > x_0) > 0$, since by the EVMT

$y(x) = y(x_0) + \frac{1}{2}(x - x_0)^2 f^{''}(\xi)$, $x_0 < \xi < x$, or

4. $y^{''}(x_0) > 0$ provided that $y^{''}(x)$ is continuous at x_0, since then $y^{''}(x) > 0$ is positive in some neighborhood of x_0.

condition 2 is sufficient for a local minimum even if $f(x)$ is not differentiable at x_0.

Function $y = f(x)$ attains a **local maximum** at $x = x_0$ if in some neighborhood $|x - x_0| < \epsilon$, $\epsilon > 0$, of x_0:

1. $f(x < x_0) < f(x_0)$ and $f(x > x_0) < f(x_0)$, or

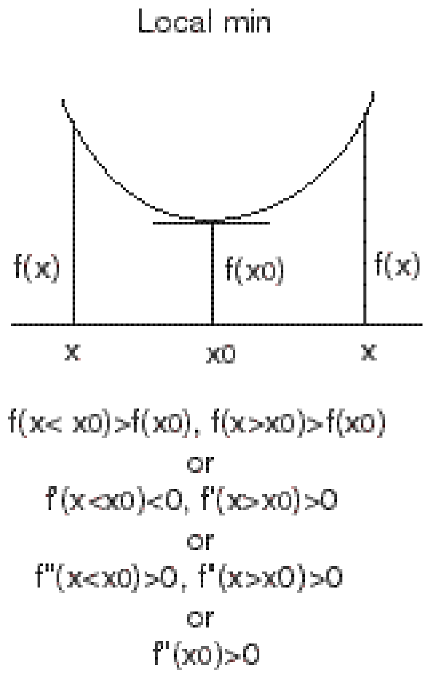
2. $f'(x < x_0) > 0$ and $f'(x > x_0) < 0$, or

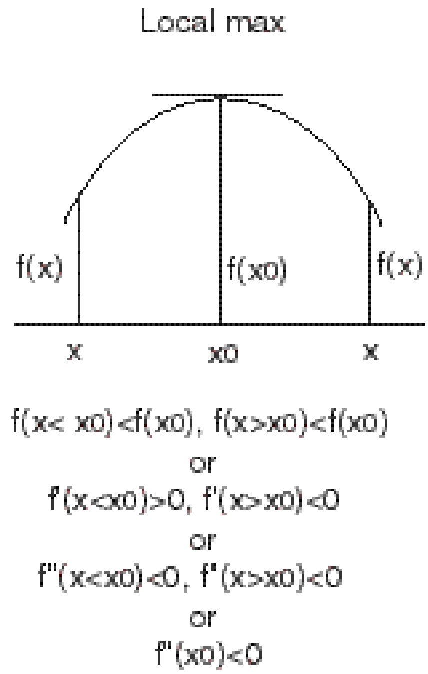
3. $f^{''}(x < x_0) < 0$ and $f^{''}(x > x_0) < 0$, or

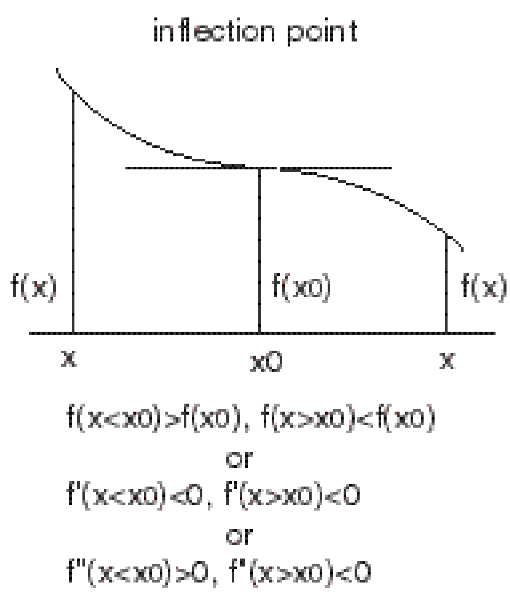
4. $f^{''}(x) < 0$ and $f^{''}(x)$ is continuous at x_0.

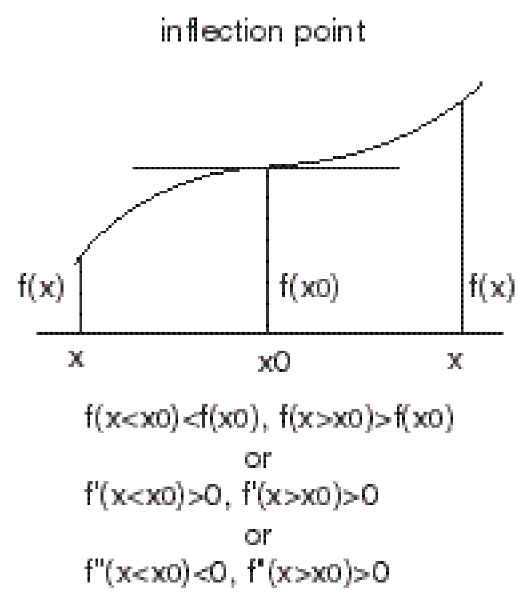
Function $y = f(x)$ **inflects** at $x = x_0$ if $f^{''}(x)$, which may even not exist at x_0, changes sign around x_0.

critical points and their nature









The osculating circle

NL at A and B:

$$y - y_0 = -\frac{1}{y'_o}(x - x_0), \quad y - y_1 = -\frac{1}{y'_1}(x - x_1)$$

Intersection

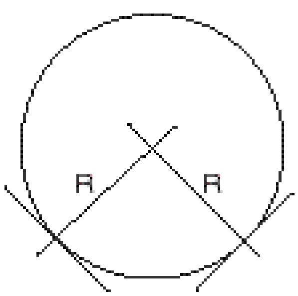
$$x = \frac{1}{y'_0 - y'_1}(y_1 y'_0 y'_1 - y_0 y'_0 y'_1 + y'_0 x_1 - y'_1 x_0)$$

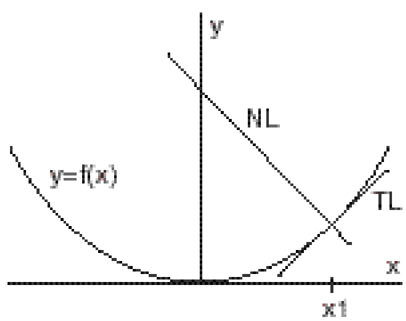
$\lim_{x_1 \to x_0} x(x_1) = \bar{x}$. Using L'Hôpital

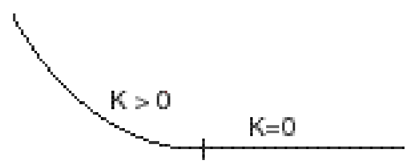
$$\bar{x} = x_0 - \frac{y'_0}{y''_0}[1 + (y'_0)^2], \quad \bar{y} = y_0 + \frac{1}{y''_0}[1 + (y'_0)^2]$$

Radius of osculating circle

$$R^2 = (\bar{x} - x_0)^2 + (\bar{y} - y_0)^2, \quad R(x) = \frac{1}{|y''|}[1 + (y')^2]^{3/2}, \quad K(x) = \frac{1}{R}$$







For example $y = x^2,\ y' = 2x,\ y'' = 2,\ R(x) = \frac{1}{2}(1 + 4x^2)^{3/2}$ and $R(0) = \frac{1}{2}$ but $R(1) = \frac{1}{2}5^{3/2}$ is much larger.

Rise of a Function

$$y(x) = y(x_0) + (x - x_0)y'(x_0) + \frac{1}{2}(x - x_0)^2 y''(\xi), \quad a < \xi < b$$

But $y'(x_0) = 0,\ y(x_0) = \max\limits_{a \le x \le b} y(x)$.

Set $x = b$

$$y(x_0) = -\frac{1}{2}(b - x_0)y''(\xi),\ a < \xi < b$$

But $b - x_0 \le \dfrac{1}{2}(b - a)$. Hence

$$|y(x)| \le \frac{1}{8}(b - a)^2 \max_{a < x < b} |y''(x)|$$

The estimate is sharp. Look at the parabola

$$y = \frac{1}{2}k(x-a)(b-x), \quad y'' = -k, \quad k > 0$$

The vertex of the parabola occurs at $x = \frac{1}{2}(a+b)$ and

$$\max y(x) = \frac{1}{8}k(b-a)^2$$

Linear interpolation

The Weierstrass approximation theorem(WAT): If function $f(x)$ is continuous on $a \le x \le b$, then for any $\epsilon > 0$ there exists a polynomial $p_n(x)$ so that

$$|f(x) - p_n(x)| \le \epsilon$$

for any $a \le x \le b$.

consider the linear interpolant

$$p_1(x) = \frac{y_1 - y_0}{x_1 - x_0}x + \frac{y_0 x_1 - y_1 x_0}{x_1 - x_0}.$$

Construct

$$\phi(z) = f(z) - p_1(z) - \frac{e(x)}{(x-x_0)(x-x_1)}(z-x_0)(z-x_1), \quad e(x) = f(x) - p_1(x).$$

A double application of Rolle's theorem assures us that $\phi''(\xi) = 0, \quad x_0 < \xi < x$.

$$\phi''(z) = f''(z) - p_1''(z) - 2\frac{f(x) - p_1(x)}{(x-x_0)(x-x_1)}.$$

But $p_1''(z) = 0$. So

$$e(x) = \frac{1}{2}(x-x_0)(x-x_1)f''(\xi), \quad x_0 < \xi < x_1.$$

$$|e(x)| \le \frac{1}{2} \max_{x_0 \le x \le x_1} |(x-x_0)(x-x_1)| \max_{x_0 < x < x_1} |f''(x)|$$

But $\max |(x-x_0)(x-x_1)|$ occurs at $x = (x_0+x_1)/2$, and with $x_1 - x_0 = h$

$$|e(x)| \le \frac{1}{8}h^2 \max_{x_0 < x < x_1} |f''(x)|$$

Quadratic interpolation

$$\phi(z) = f(z) - p_2(z) - \frac{e(x)}{(x-x_0)(x-x_1)(x-x_2)}(z-x_0)(z-x_1)(z-x_2), \quad e(x) = f(x) - p_2(x)$$

Function $\phi(z) = 0$ at x_0, x, x_1, x_2 and therefore by Rolle's theorem $\phi'(z) = 0$ at three points inside $x_0 < x < x_2$, $\phi''(z) = 0$ at two points inside $x_0 < x < x_2$, and $\phi'''(z) = 0$ at least once in the interval $x_0 < x < x_2$, $x \neq x_1$. Also because $p_2''(z) = 0$ it results that

$$e(x) = \frac{1}{6}(x-x_0)(x-x_1)(x-x_2)f'''(\xi), \quad x_0 < \xi < x_2.$$

Say $x_0 = -h,\ x_1 = 0,\ x_2 = h$, then

$$e(x) = \frac{1}{6}x(x^2-h^2)f'''(\xi), \quad -h < \xi < h$$

max $|x(x^2-h^2)|$ occurs at $x = \frac{\sqrt{3}}{3}h$, and

$$|e(x)| \leq \frac{\sqrt{3}}{27}h^3 \max_{x_0<x<y_2} |y'''(x)|.$$

17. The indefinite and definite integrals

Let $y = f(x)$ be a continuous function for $a \leq x \leq b$

$$\int f(x)dx = F(x) + c,\ F'(x) = f(x).$$

is its indefinite integral.

$\int cf dx = \int f dx$, $\int (f+g)dx = \int f dx + \int g dx$, $\int u' f(u) dx = \int f(u) du,\ u = u(x)$.
$\int uv' dx = uv - \int vu' dx$.

The definite integral of $f(x)$ is

$$A(x) = \int_a^x f(x)dx = F(x)\Big|_a^x = F(x) - F(a),\ \text{and}\ A(a) = 0.$$

Also

$$\int_a^b f(x)dx = \int_a^m f(x)dx + \int_m^b f(x)dx,\ \int_a^b f dx = -\int_b^a f dx.$$

Since $A'(x) = F'(x) = f(x)$, we have the

Monotonicity theorem for the definite integral: if $f(x) > 0$, except for an occasional $f(x_0) = 0$, then $A(x)$ is increasing.

Consequently

$$\int_a^b |f(x)|dx \geq \int_a^b f(x)dx$$

.

18. Some integrals

$$\int x^n dx = \frac{1}{n+1}x^{n+1} + c,\ n \neq -1,\ \int \frac{1}{x}dx = \ln x + c,\ \int a^x dx = \frac{1}{\ln a}a^x + c,\ \int e^x dx = e^x + c$$

$$\int \sin x dx = -\cos x + c,\ \int \cos x dx = \sin x + c,\ \int \frac{1}{\cos^2 x}dx = \tan x + c$$

$$\int \frac{1}{1+x^2}dx = \tan^{-1} x + c,\ \int \frac{1}{\sqrt{1-x^2}}dx = \sin^{-1} x + c.$$

19. The integral mean value theorems

$$\int_a^b f(x)dx = \frac{F(b)-F(a)}{b-a}(b-a) =$$

$$F'(\xi)(b-a) = f(\xi)(b-a),\ a < \xi < b.$$

<u>The general integral mean-value theorem</u> claims that if $g(x) > 0,\ a \leq x \leq b$, with an occasional $g(x_0) = 0$, and if $m < f(x) < M$, then

$$\int_a^b f(x)g(x)dx = \mu \int_a^b g(x)dx,\ m \leq \mu \leq M.$$

Indeed,

$$\int_a^b (M - f(x))g(x)dx > 0,\ \int_a^b (f(x) - m)g(x)dx > 0$$

so that

$$M \int_a^b g(x)dx > \int_a^b f(x)g(x)dx > m \int g(x)dx$$

and therefore

$$\int_a^b f(x)g(x)dx = \mu \int_a^b g(x)dx.$$

If $f(x)$ is continuous in $a \leq x \leq b$, then $\mu = f(\xi),\ a < \xi < b$, and

$$\int_a^b f(x)g(x)dx = f(\xi) \int_a^b g(x)dx,\ a < \xi < b.$$

20. Accuracy of the mid-point Riemann sum

By the EMVT

$$F(b) = F(a) + (b-a)F'(a) + \frac{1}{2}(b-a)^2 F''(\xi),\ a < \xi < a$$

or since $F'(x) = f(x)$

$$F(b) = F(a) + (b-a)f(a) + \frac{1}{2}(b-a)^2 f'(\xi), \;\; a < \xi < b.$$

Also

$$F(a) = F(b) - (b-a)f(b) + \frac{1}{2}(b-a)^2 f'(\eta), \;\; a < \eta < b.$$

If m is midpoint between a and b, then

$$F(a) = F(m) - (m-a)f(m) + \frac{1}{2}(m-a)^2 f'(\xi), \;\; a < \xi < m$$

$$F(b) = F(m) + (b-m)f(m) + \frac{1}{2}(b-m)^2 f'(\eta), \;\; m < \eta < b$$

Hence if $m - a = b - m = \epsilon$

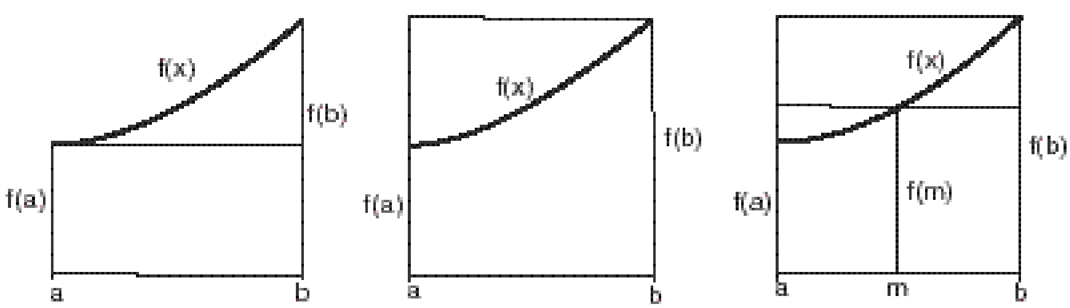
$$\int_a^b f(x)dx = F(b) - F(a) = (b-a)f(m) + \frac{1}{2}\epsilon^2(f'(\eta) - f'(\xi))$$

or by the MVT

$$\int_a^b f(x)dx = (b-a)f(m) + \frac{1}{2}\epsilon^2(\eta - \xi)f^{''}(\zeta), \;\; a < \zeta < b$$

But $|\eta - \xi| < 2\epsilon$, and since $2\epsilon = b - a$

$$\left|\int_a^b f(x)dx - (b-a)f(m)\right| \leq \frac{1}{8}(b-a)^3 \left|\max_{a<x<b} f^{''}(x)\right|.$$



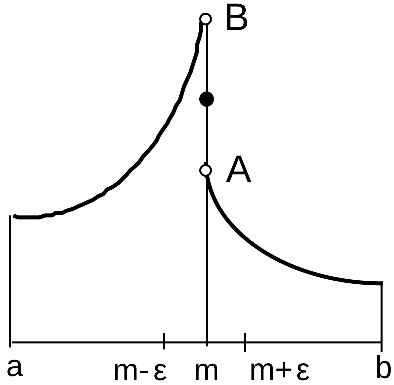
21. Discontinuous integrand

In case $f(x)$ is discontinuous at $a < x = m < b$ we define

$$\int_a^b f(x)dx = \lim_{\epsilon \to 0} \int_a^{m-\epsilon} f(x)dx + \int_{m+\epsilon}^b f(x)dx$$

which is independent of $f(m)$, and amounts to assuming a vertical fence between points A and B. The area function $A(x)$ under $f(x)$ between a and b is then a continuous function of x.

See the figure below



Implicit Functions

Expression

$$x^2 + xy + y^2 - 3 = 0 \tag{1}$$

is an implicit representation of the two functions

$$y_1(x) = \frac{1}{2}(-x + \sqrt{12 - 3x^2}), \quad y_2(x) = \frac{1}{2}(-x + \sqrt{12 - 3x^2}) \tag{2}$$

in which x is restricted to $-2 \leq x \leq 2$. We were able to produce the two explicit functions $y_1(x)$ and $y_2(x)$ out of implicit expression (1) because for any fixed x expression (1) reduces to a mere quadratic equation for y which can be solved by simple and formal algebraic means.

In the higher order expression

$$x^3y + xy^3 + x + y - 4 = 0. \tag{3}$$

selection of $x = 1$ leaves us with $y^3 + 2y - 3 = 0$. An algebraic formula for the solution of the cubic equation exists but is considerably more involved than the formula for the quadratic equation, and we may opt for a numerical, computer driven, procedure such as bisection for the solution of the cubic equation. Here we just observe that $y = 1$ solves the cubic equation so that the pair $x_0 = 1$, $y_0 = 1$ satisfies equation (3). Since the cubic equation infallibly possesses a real root we are confident that for any chosen x we should be able to get from eq.(3) a corresponding y to any desired degree of approximation, implying that eq.(3) actually harbors at least one function $y = f(x)$.

If we choose in the expression

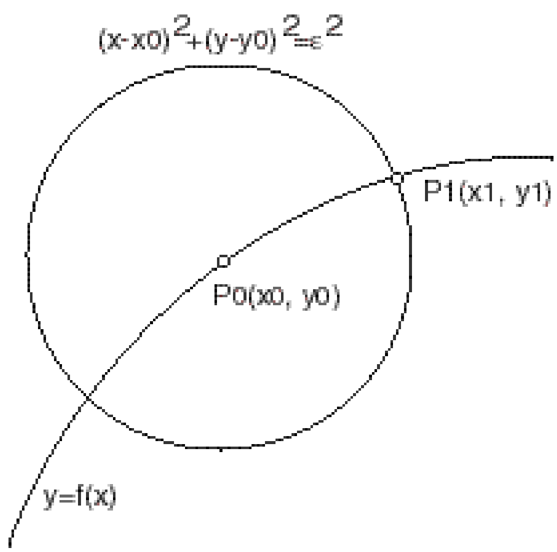
$$xe^y + ye^x + x^2 + y^2 - 9 = 0 \tag{4}$$

$x = 1$, then we are left with the **transcendental** equation $e^y + y + y^2 - 8 = 0$ which can be solved for y but numerically and approximately. Now, even if expression (4) contains $y = f(x)$ there is no way this equation can be written explicitly in terms of elementary functions.

Generally, expression $F(x,y) = 0$ is an implicit representation of function $y = f(x)$ if $F[x, f(x)]$ is identically zero. Even if an explicit $y = f(x)$ can not be extracted from $F(x,y) = 0$ we may still want to trace the graph of this function passing through point $P_0(x_0, y_0)$. to construct this curve we shall need a sequence of points $P_1(x_1, y_1),\ P_2(x_2, y_2),\ P_3(x_3, y_3)$ and so on closely strung on the curve. A way of moving from point P_0 on the curve to a nearby point P_1 consists of numerically solving the pair of equations

$$F(x,y) = 0, \quad (x - x_0)^2 + (y - y_0)^2 = \epsilon^2 \tag{5}$$

in which $\epsilon > 0$ is chosen small enough to produce close points for a smooth looking curve.



But does equation (4), or for that matter the general $F(x,y) = 0$, actually contain a $y = f(x)$ in the neighborhood of some $x = x_0,\ y = f(x_0)$, for which $F(x_0, y_0) = 0$, even if this function can not be written explicitly?

The following theorem gives a decisive answer to this question.

Existence Theorem. If function $F(x,y)$ of the two variables x and y satisfies the three conditions:

(i) $F(x_0, y_0) = 0$ **for** $x = x_0,\ y = y_0$**;**

(ii) Partial derivatives F_x and F_y of F exist and are continuous in some neighborhood of $x = x_0,\ y = y_0$;

(iii) $F_y(x_0, y_0) \neq 0$**,**

then there is one and only one function $y = f(x),\ y_0 = f(x_0)$, assured to be continuous and differentiable, that satisfies $F(x,y) = 0$ for every x in some neighborhood of x_0.

Things that can go wrong with implicit differentiation:

1. For $y^2 - x = 0$ we obtain by implicit differentiation $2yy' - 1 = 0$ leading to the absurdity $0 = -1$ if $x = y = 0$ because the function $y = \sqrt{x}$ is not differentiable at $x = 0$.

2. For $(x-y)^2 = x^2 - 2xy + y^2 = 0$ we obtain by implicit differentiation $(x-y)(1-y') = 0$ and $y' = 1$ if $x - y \neq 0$. Actually $y = x$ and $y' = 1$ for any x.

3. Expression $(x-y)(x+y) = x^2 - y^2 = 0$ represents two orthogonal lines through $x = y = 0$. By implicit differentiation $x - yy' = 0$ and at $x = y = 0$ we get $0 = 0$ because of the ambiguity of the bifurcation.

4. Expression $x^2 + y^2 = 0$ represents the mere point $x = y = 0$. By implicit differentiation $x + yy'$=0, and at $x = y = 0$ this reduces to $0 = 0$.

Even if the implicit $F(x, y) = 0$ does not yield an explicit $y = f(x)$, still $y'(x)$ can be explicitly written in terms of both variables x and y.

We recall that if $y = f(x)$, then by the chain rule $(y^2)' = 2yy'$ and $(y^3)' = 3y^2y'$, where $(\ )'$ means differentiation with respect to x. Applying $(\ )'$ to both side of equation (3) produces

$$(3x^2 + x^3y') + (y^3 + 3xy^2y') + 1 + y' = 0$$

which is linear in y', and is readily solved to produce

$$y' = -\frac{3x^2y + y^3 + 1}{x^3 + 3xy^2 + 1} = -\frac{F_x(x, y)}{F_y(x, y)} \tag{6}$$

which for $x = 1,\ y = 1$, that satisfy equation (3), yields $y'(1) = -1$

From equation (4) we obtain by implicit differentiation that

$$(e^y + xe^yy') + (y'e^x + ye^x) + 2x + 2yy' = 0 \tag{7}$$

and

$$y' = \frac{e^y + ye^x + 2x}{e^x + xe^y + 2y} = -\frac{F_x(x, y)}{F_y(x, y)} \tag{8}$$

so that all we need for the evaluation of $y'(x_0)$ is the pair $x = x_0,\ y = y_0$ that satisfies $F(x, y) = 0$.

Expression $x^2 + y^2 = R^2$ describes both the upper and the lower arc of a circle centered at $C(0, 0)$ and having radius R. Repeated differentiation produces here

$$x + yy' = 0 \qquad \text{and} \qquad 1 + (y')^2 + yy'' = 0 \tag{9}$$

so that the critical points of both functions implicit in the equation of the circle happen to be at $x = 0,\ y = \pm R$. But if $y' = 0$, then $y'' = -1/y$, and $y'' = -1/R$ if $y = +R$, implying a local maximum, and $y'' = 1/R$ if $y = -R$ implying a local minimum.

Exercise: Find all critical points of the folium of Descartes $x^3 + y + 3 - 3xy = 0$. Ans. $y' = -(x^2 - y)/(y^2 - x)$. $x_0 = 2^{1/3}, y_0 = 4^{1/3}$. Max.

For the circle $(x-x_0)^2+(y-y_0)^2=R^2$ we obtain by repeated implicit differentiations that

$$(x-x_0)+(y-y_0)y'=0 \qquad \text{and} \qquad 1+(y')^2+(y-y_0)y''=0. \tag{10}$$

By some algebra we get from equation (10) and the equation of the circle that

$$y-y_0=-\frac{1+(y')^2}{y''},\ x-x_0=-(y-y_0)y',\ R^2=\frac{(1+(y')^2)^3}{(y'')^2}. \tag{11}$$

Tricking equations (11) to believe that any twice differentiable function $y=f(x),\ y''\neq 0$, is a circle we obtain for this function the radius and center of its osculating circle. For example, for $y=x^2,\ x=y=1$, equations (11) produce $R^2=125/4$, or $R=5.59$, and $x_0=-4,\ y_0=7/2$. The curvature $\kappa=1/R$ of $y=x^2$ at $x=1$ is, then, $\kappa=0.179$.

Rational approximations

Consider the rational polynomial approximation

$$y(x)=\frac{1+a_1x}{1+b_1x}$$

to the function

$$f(x)=\sqrt{1+x}$$

at $x_0=0$. We have that

$$f(x)=(1+x)^{\frac{1}{2}},\ f'(x)=\frac{1}{2}(1+x)^{-\frac{1}{2}},\ f''(x)=-\frac{1}{4}(1+x)^{-\frac{3}{2}}$$

so that $f(0)=1,\ f'(0)=1/2,\ f''(0)=-1/4$. On the other hand

$$y'(x)=\frac{a_1-b_1}{(1+b_1x)^2},\quad y''(x)=-2\frac{b_1(a_1-b_1)}{(1+b_1x)^3}$$

so that $y'(0)=a_1-b_1$ and $y''(0)=-2b_1(a_1-b_1)$. From $y(0)=f(0),\ y'(0)=f'(0),\ y''(0)=f''(0)$ we have that $a_1=3/4$ and $b_1=1/4$, or

$$y(x)=\frac{4+3x}{4+x}.$$

The error in this rational approximation of $\sqrt{1+x}$ at $x=0$ is

$$e(x)=\sqrt{1+x}-\frac{4+3x}{4+x}$$

or

$$e(x)=\left(\sqrt{1+x}-\frac{4+3x}{4+x}\right)\left(\sqrt{1+x}+\frac{4+3x}{4+x}\right)\left(\sqrt{1+x}+\frac{4+3x}{4+x}\right)^{-1}$$

that reforms into

$$e(x) = \left[(1+x) - \left(\frac{4+3x}{4+x}\right)^2\right]\left(\sqrt{1+x} + \frac{4+3x}{4+x}\right)^{-1}$$

then into

$$e(x) = x^3 g(x), \quad g(x) = (4+x)^{-2}\left(\sqrt{1+x} + \frac{4+3x}{4+x}\right)^{-1}.$$

We readily have that

$$\lim_{x\to 0} g(x) = \frac{1}{32}$$

so that approximately

$$e(x) = \frac{1}{32}x^3$$

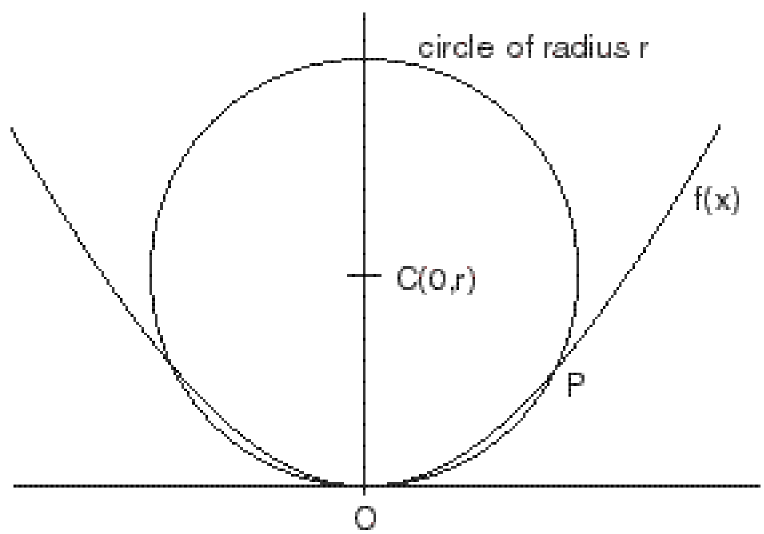
if $|x| << 1$.

Let $x = 0.1$. Then $f(x) = 1.0488088$, $y(x) = 4.3/4.1 = 1.0487805$ and $e(x) = f(x) - y(x) = 2.83 \cdot 10^{-5}$.

Another look at the osculating circle

Let $f(x)$ be twice differentiable at $x = 0$ and such that $f(0) = f'(0) = 0$. The equation of a circle of radius r centered at $C(0, r)$ and passing through $O(0, 0)$ is $x^2 + (y - r)^2 = r^2$. The intersection of this circle and $f(x)$ occurs at points O and P. See the figure. At an intersection point $y = f(x)$ and at such a point $x^2 + (f - r)^2 = r^2$ or

$$r = \frac{x^2 + f^2}{2f}.$$



Let now $x \to 0$ or $P \to O$. By L'Hôpital's rule we have that

$$\lim_{x\to 0} \frac{x^2 + f^2}{2f} = \lim_{x\to 0} \frac{2x + 2ff'}{2f'} = \lim_{x\to 0} \frac{1 + f'^2 + ff''}{f''} = \frac{1}{f''(0)}$$

which is the radius of the osculating circle at P.

Taylor's theorem

Mean Value Theorem(MVT)

Say $f(0) = 0$. Write $F(x) = f(x) - (x/b)f(b)$ so that $F(0) = F(b) = 0$. By Rolle's theorem $F'(\xi) = 0$ for some $0 < \xi < b$. But $F'(x) = f'(x) - (1/b)f(b)$ so that $F'(\xi) = f'(\xi) - (1/b)f(b) = 0$ and $f(b) = bf'(\xi)$. Writing x for b we have that

$$f(x) = xf'(\xi), \quad 0 < \xi < x$$

if $x > 0$. Or

$$f(x) = xf'(0) + x(f'(\xi) - f'(0)), \quad 0 < \xi < x.$$

More generally,

$$f(x) = f(a) + (x-a)f'(\xi), \quad a < \xi < x$$

if $x > a$.

Extended Mean Value Theorem(EMVT)

Say $f(0) = f'(0) = 0$. Write $F(x) = f(x) - (x/b)^2 f(b)$ so that $F(0) = F(b) = 0$. By Rolle's theorem $F'(\xi 1) = 0$ for some $0 < \xi 1 < b$ if $b > 0$. The derivative function of $F(x)$ is $F'(x) = f'(x) - (2x/b^2)f(b)$ and $F'(0) = 0$. Assuming that f, and consequently, F is a 'nice' function we may go on and apply Rolle's theorem to $F'(x),\ 0 \leq x \leq \xi 1$, to be assured of $F''(\xi) = 0,\ 0 < \xi < \xi 1$. Or $F''(\xi) = f''(\xi) - (2/b^2)f(b) = 0$ resulting in $f(b) = (1/2!)b^2 f''(\xi),\ 0 < \xi < b$. Writing x for b we have that

$$f(x) = (1/2!)x^2 f''(\xi),\ 0 < \xi < x$$

if $x > 0$.

More generally

$$f(x) = f(a) + (x-a)f'(a) + (1/2!)f''(\xi),\ a < \xi < x$$

if $x > a$.

Taylor's Theorem(TT)

Say $f(0) = f'(0) = f''(0) = ... = f^{(n-1)}(0) = 0$. Introducing $F(x) = f(x) - (x/b)^n f(b)$ and repeatedly applying Rolle's theorem produces

$$f(x) = (1/n!)x^n f^{(n)}(\xi),\ 0 < \xi < x$$

if $x > 0$. More generally

$$f(x) = f(a) + (x-a)f'(a) + (1/2!)(x-a)^2 f''(a) + (1/3!)(x-a)^3 f'''(a) + ... + R_n$$

where

$$R_n = (1/n!)(x-a)^n f^{(n)}(\xi), \ a < \xi < x$$

if $x > a$.

Solution of nonlinear equations by linearization.

Consider the equation $x^2 = 5$ solved by $\sqrt{5}$ which is irrational. We have seen how to create a *converging sequence* of rational approximations to $\sqrt{5}$ by the method of bisection. Herein we consider another idea attributed to the great Newton himself for the generation of another, faster converging, sequence of rational approximations to $\sqrt{5}$. We start the process by selecting a good rational approximation x_0 to $\sqrt{5}$, say $x_0 = 2$ so that $x_0^2 = 4$. An improved approximation x_1 is chosen in the form $x_1 = x_0 + \epsilon = 2 + \epsilon$ to have

$$(2+\epsilon)^2 = 4 + 4\epsilon + \epsilon^2 = 5$$

or $\epsilon(4+\epsilon) = 1$. Knowing that $|\epsilon| < 1$ we are prepared to ignore ϵ relative to 4 and replace the last equation with the approximate $4\epsilon = 1$ or $\epsilon = 1/4$, so as to have $x_1 = 9/4$ with $x_1^2 = (9/4)^2 = 5.0625$, a certain improvement over $x_0^2 = 4$. We may repeat the process with $x_2 = 9/4+\epsilon$, to get a yet hopefully better rational approximation to $\sqrt{5}$. We shall not pursue this now but instead will look at the theoretical analysis of the convergence properties of this method for the solution of $x^2 = 2$, that we denote by $x = \sqrt{2}$.

Let x_0 be a good initial rational guess for $\sqrt{2}$. We seek an improvement thereof in the form $x_1 = x_0 + \epsilon$ so that

$$(x_0+\epsilon)^2 = x_0^2 + 2x_0\epsilon + \epsilon^2 = 2$$

once more we ignore the supposedly small ϵ, relative to $2x_0$ to have the linear, easy to solve, equation $2x_0\epsilon = 2 - x_0^2$ or $\epsilon = (2-x_0^2)/2x_0$ leading to the *iterative* scheme

$$x_1 = x_0 + \epsilon = \frac{1}{2}x_0 + \frac{1}{x_0}$$

which produces excellent results as we shall see. Before continuing with the theoretical analysis of Newton's method we will pause for a concrete example. Starting with $x_0 = 1$ we obtain by it the convergent sequence $1, \ 3/2, \ 17/12, \ 577/408\ldots$ with $1^2 = 1, \ (3/2)^2 = 2.25, \ (17.12)^2 = 2.00694, \ (577/408)^2 = 2.00006$. The method of bisection produced both upper and lower bounds on $\sqrt{2}$, but it appears that Newton's method produces here mere upper bounds on $\sqrt{2}$, albeit becoming very accurate very fast.

We shall now confirm all this theoretically.

We have that

$$x_1 - \sqrt{2} = \frac{x_0^2+2}{2x_0} - \sqrt{2}$$

or

$$x_1 - \sqrt{2} = \frac{1}{2x_0}(x_0 - \sqrt{2})^2$$

and if x_0 is nearly $\sqrt{2}$, then

$$x_1 - \sqrt{2} = \frac{\sqrt{2}}{4}(x_0 - \sqrt{2})^2.$$

The error is in the $(n+1)$ step of the Newton method is nearly proportional to the *square* of the error in the nth step. Such convergence behavior is termed *quadratic*. We also confirm from the error propagation formula that, since $\sqrt{2}/4 > 0$, $x_n - \sqrt{2} > 0$ if $n > 0$ meaning that convergence to $\sqrt{2}$ takes place from above.

Let us try something else. The equation $x^2 = 2$ can be written as $x^2 - 1 = 1$, as $(x-1)(x+1) = 1$ or as $x = 1 + 1/(1+x)$ which we propose to turn into the iterative scheme

$$x_1 = 1 + \frac{1}{1 + x_0}.$$

To analyze its performance we write it as

$$x_1 - \sqrt{2} = \frac{x_0 + 2}{x_0 + 1} - \sqrt{2}$$

from which we readily obtain that

$$x_1 - \sqrt{2} + \frac{1 - \sqrt{2}}{1 + \sqrt{2}}(x_0 - \sqrt{2})$$

if x_0 is very close to $\sqrt{2}$. Here the error in the $(n+1)$th step is proportional to the error in the nth step with the negative proportionality factor

$$(1 - \sqrt{2})(1 + \sqrt{2}) = -0.1716$$

implying two things:

(1) that convergence takes place but is only linear in nature with a constant *rate*, and,

(2) that the error alternates in signs meaning that the approximations alternate between being an overestimate and an underestimate.

For example, if $x_0 = 1$, then we generate by this iterative scheme the convergent sequence $1,\ 3/2,\ 7/5,\ 17/12,\ 41/29, \ldots$ with

$$(1)^2 = 1,\ (3/2)^2 = 2.25,\ (17/12)^2 = 2.007,\ (41/29)^2 = 1.9988.$$

Substituting $x_1 = (x_0 + 2)/(x_0 + 1)$ into $x_2 = (x_1 + 2)/(x_1 + 1)$ we obtain

$$x_2 = 1 + \cfrac{1}{2 + \cfrac{1}{1 + x_0}}$$

becoming upon repetitive substitutions

$$\sqrt{2} = 1 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \cdots}}}}}$$

which is the *continued fraction* approximation to $\sqrt{2}$.

Here is an instance of mathematics seizing the opportunity to generalize and improve a recognized pattern.

We have just proved that the iterative scheme $x_1 = (x_0 + 2)/(x_0 + 1)$ is such that $x_1 - \sqrt{2} = c(x_0 - \sqrt{2})$, with constant $c = (1 - \sqrt{2})/(1 + \sqrt{2})$. It occurs to us that this scheme is but a special case of the more general

$$x_1 = \frac{a_1 + a_2 x_0}{b_1 + b_2 x_0}$$

and we now want to know if coefficients a_1, a_2, b_1, b_2 can be selected so that $x_1 - \sqrt{2} = c(x_0 - \sqrt{2})$ with a smaller $|c|$, since the smaller $|c|$ is the faster the convergence.

Some simple algebra leads us to

$$x_1 - \sqrt{2} + \frac{(a_1 - 2b_2) + \sqrt{2}(a_2 - b_1)}{b_1 + b_2 x_0} + \frac{a_2 - \sqrt{2}b_2}{b_1 + b_2 x_0}(x_0 - \sqrt{2}).$$

For convergence the first right hand side term of the above equation must be zero and we set $a_1 - 2b_2 = 0,\ a_2 - b_1 = 0$ so as to be left with

$$x_1 - \sqrt{2} = \frac{2b_1 - \sqrt{2}a_1}{2b_1 + \sqrt{2}a_1}(x_0 - \sqrt{2}),\ x_1 - \sqrt{2} = c(x_0 - \sqrt{2})$$

provided that x_0 is near $\sqrt{2}$. The smallest $|c| = 0$ is obtained with $b_1 = (\sqrt{2}/2)a_1$. Wanting to exclude irrational coefficients we resort to an approximation of this relationship between a_1 and b_1. We know that $7/5$ is a good rational approximation to $\sqrt{2}$, and elect to have $b_1 = (7/10)a_1$ so as to have

$$x_1 = \frac{10 + 7x_0}{7 + 5x_0}.$$

Starting with $x_0 = 3/2$ we generate with this formula the convergent sequence

$$3/2,\ 41/29,\ 577/408\ldots$$

for $\sqrt{2}$, being such that

$$(3/2)^2 = 2.25,\ (41/29)^2 = 1.9988,\ (577/408)^2 = 2.000006,$$

with $|577/408 - \sqrt{2}| = (1/408)^{2.42}$.