Examples

1. Find the standard matrix A for the dilation transformation

T (x) = 3x, for x in R^2

2. Assume that T is a linear transformation. Find the standard matrix of T. T : $R^2 \to R^2$ is a vertical shear transformation that maps e_1 into $e_1 - 2e_2$ but leaves the vector e_2 unchanged.

3. Assume that T is a linear transformation. Find the standard matrix of T. T : $R^2 \to R^2$ is a horizontal shear transformation that leaves e_1 unchanged and maps e_2 into $e_2 + 3e_1$.

4. Assume that T is a linear transformation. Find the standard matrix of T. T : $R^2 \to R^2$ rotates points (about the origin) through $3\pi/2$ radians (counter clockwise).

5. Assume that T is a linear transformation. Find the standard matrix of T . T : $R^2 \to R^2$ rotates points (about the origin) through $-\pi/4$ radians (clockwise).

6. Assume that T is a linear transformation. Find the standard matrix of T . T : $R^2 \to R^2$ first reflects points through the horizontal x_1- axis and then reflects points through the line $x_2 = x_1$.

Example 1

- Example Find the standard matrix A for the dilation transformation

T (**x**) = 3**x**, for **x** in R^2

Solution

Standard Matrix is given by $A = [\mathrm{T}(e_1) \quad \mathrm{T}(e_2)]$

$\mathrm{T}(e_1) = 3e_1 = 3\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}3\\0\end{bmatrix}$,

$\mathrm{T}(e_2) = 3e_2 = 3\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}0\\3\end{bmatrix}$

$$A = \begin{bmatrix}3 & 0\\0 & 3\end{bmatrix}$$

Example 2

Assume that T is a linear transformation. Find the standard matrix of T .
$T : R^2 \to R^2$ is a vertical shear transformation that maps e_1 into $e_1 - 2e_2$ but leaves the vector e_2 unchanged.

Solution

Standard Matrix is given by $A = [T(e_1) \quad T(e_2)]$

$T(e_1) = e_1 - 2e_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} - 2\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$

$T(e_2) = e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

$A = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}$

Example 3

Assume that T is a linear transformation. Find the standard matrix of T .
$T : R^2 \rightarrow R^2$ is a horizontal shear transformation that leaves e_1 unchanged and maps e_2 into $e_2 + 3e_1$.

Example 3

Assume that T is a linear transformation. Find the standard matrix of T .
$T : R^2 \rightarrow R^2$ is a horizontal shear transformation that leaves e_1 unchanged and maps e_2 into $e_2 + 3e_1$.

Solution

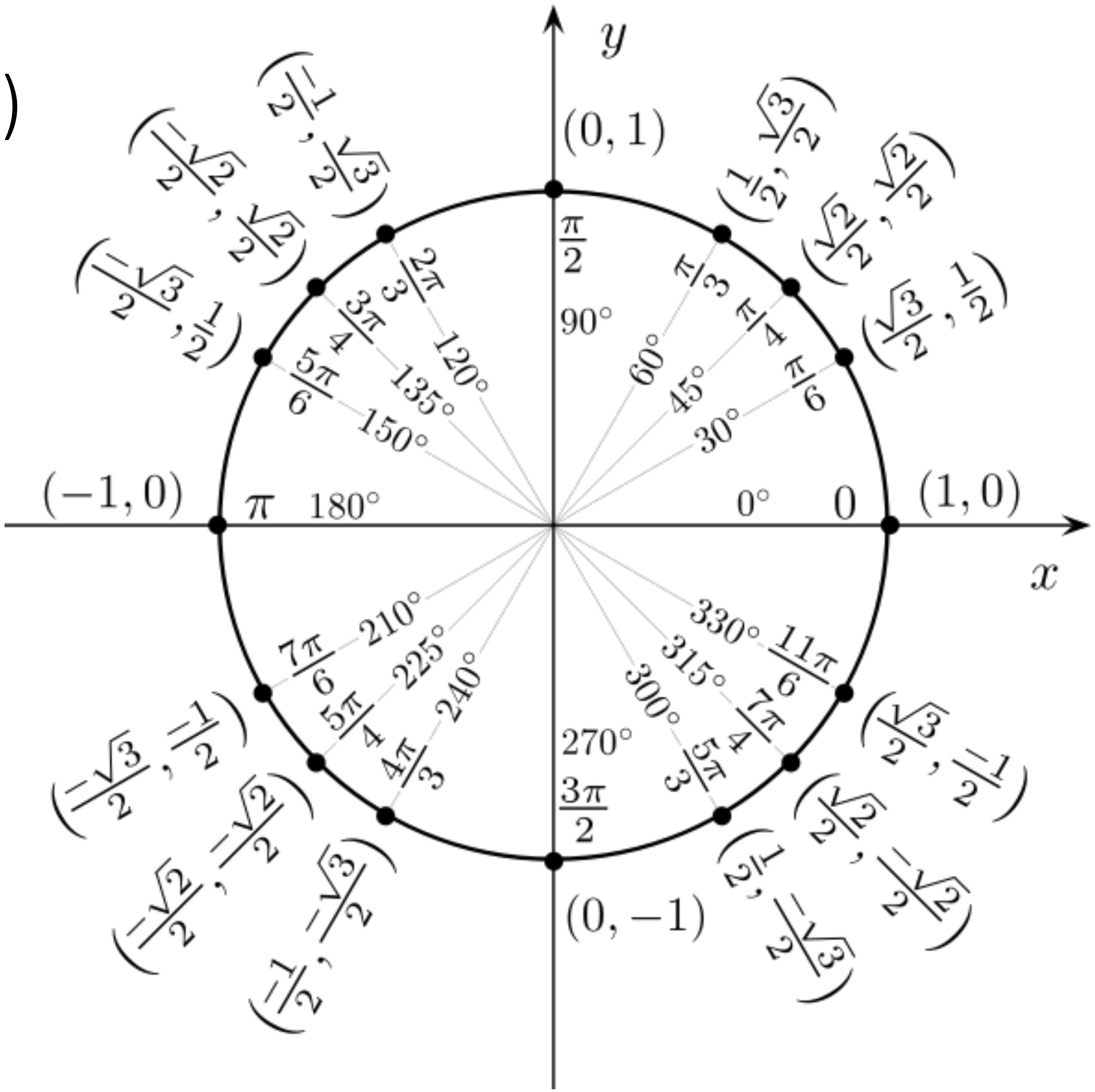
Standard Matrix is given by $A = [\mathrm{T}(e_1) \quad \mathrm{T}(e_2)]$

$\mathrm{T}(e_1) = e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$

$\mathrm{T}(e_2) = e_2 + 3e_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} + 3\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$

$$A = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$$

(cos, sin)



y
x
$(0, 1)$
$(1, 0)$
$(-1, 0)$
$(0, -1)$
$\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right)$
$\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$
$\left(\frac{\sqrt{3}}{2}, \frac{1}{2}\right)$
$\left(\frac{-1}{2}, \frac{\sqrt{3}}{2}\right)$
$\left(\frac{-\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$
$\left(\frac{-\sqrt{3}}{2}, \frac{1}{2}\right)$
$\left(\frac{-\sqrt{3}}{2}, \frac{-1}{2}\right)$
$\left(\frac{-\sqrt{2}}{2}, \frac{-\sqrt{2}}{2}\right)$
$\left(\frac{-1}{2}, \frac{-\sqrt{3}}{2}\right)$
$\left(\frac{\sqrt{3}}{2}, \frac{-1}{2}\right)$
$\left(\frac{\sqrt{2}}{2}, \frac{-\sqrt{2}}{2}\right)$
$\left(\frac{1}{2}, \frac{-\sqrt{3}}{2}\right)$
0 $0°$
$\frac{\pi}{6}$ $30°$
$\frac{\pi}{4}$ $45°$
$\frac{\pi}{3}$ $60°$
$\frac{\pi}{2}$ $90°$
$\frac{2\pi}{3}$ $120°$
$\frac{3\pi}{4}$ $135°$
$\frac{5\pi}{6}$ $150°$
π $180°$
$\frac{7\pi}{6}$ $210°$
$\frac{5\pi}{4}$ $225°$
$\frac{4\pi}{3}$ $240°$
$\frac{3\pi}{2}$ $270°$
$\frac{5\pi}{3}$ $300°$
$\frac{7\pi}{4}$ $315°$
$\frac{11\pi}{6}$ $330°$

Example 4

Assume that T is a linear transformation. Find the standard matrix of T.
$T : R^2 \to R^2$ rotates points (about the origin) through $3\pi/2$ radians (counter clockwise).

Example 4

Assume that T is a linear transformation. Find the standard matrix of T .
T : $R^2 \rightarrow R^2$ rotates points (about the origin) through $3\pi/2$ radians (counter clockwise).

$$A = \begin{bmatrix} cos\phi & -sin\phi \\ sin\phi & cos\phi \end{bmatrix}$$

$$T(e_1) = -e_2 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}, T(e_2) = e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

Example 5

- Assume that T is a linear transformation. Find the standard matrix of T . $T : R^2 \rightarrow R^2$ rotates points (about the origin) through $-\pi/4$ radians (clockwise).

Example 5

- Assume that T is a linear transformation. Find the standard matrix of T . T : $R^2 \rightarrow R^2$ rotates points (about the origin) through -π/4 radians (clockwise).
- $A = \begin{bmatrix} cos\phi & -sin\phi \\ sin\phi & cos\phi \end{bmatrix}$

T (e_1) $= \begin{bmatrix} {}^{1}/_{\sqrt{2}} \\ -{}^{1}/_{\sqrt{2}} \end{bmatrix}$, T (e_2) = $\begin{bmatrix} {}^{1}/_{\sqrt{2}} \\ {}^{1}/_{\sqrt{2}} \end{bmatrix}$

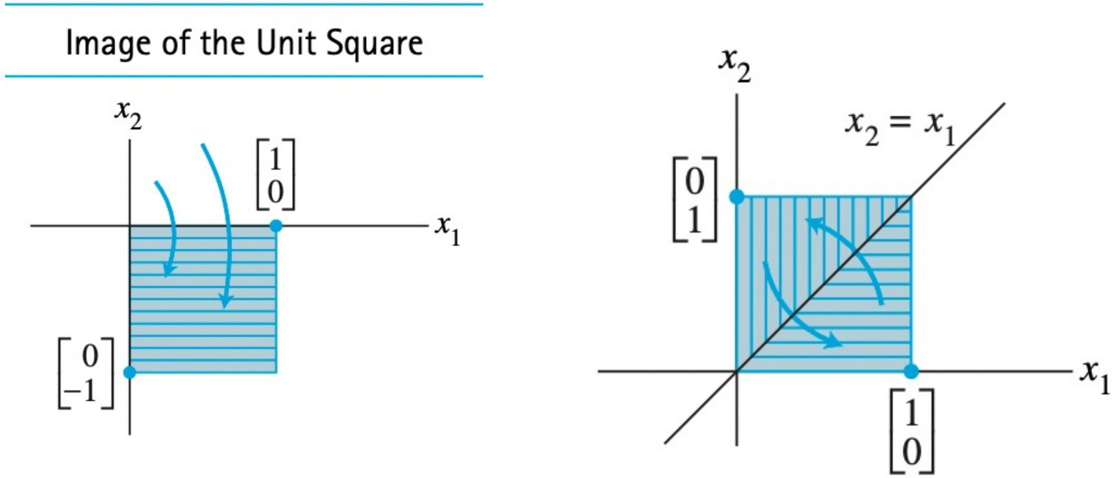
A = $\begin{bmatrix} {}^{1}/_{\sqrt{2}} & {}^{1}/_{\sqrt{2}} \\ -{}^{1}/_{\sqrt{2}} & {}^{1}/_{\sqrt{2}} \end{bmatrix}$

Example 6

- Assume that T is a linear transformation. Find the standard matrix of T . $T : R^2 \rightarrow R^2$ first reflects points through the horizontal x_1- axis and then reflects points through the line $x_2 = x_1$.

Example 6

- Assume that T is a linear transformation. Find the standard matrix of T . T : $R^2 \rightarrow R^2$ first reflects points through the horizontal x_1- axis and then reflects points through the line $x_2 = x_1$.



$\mathbf{e}_1 \rightarrow \mathbf{e}_1 \rightarrow \mathbf{e}_2$ and $\mathbf{e}_2 \rightarrow -\mathbf{e}_2 \rightarrow -\mathbf{e}_1$,

- A = $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$

Numerical Methods

Numerical methods

- A numerical method is an approximate computer method for solving a mathematical problem which often has no analytical solution.

Analytical Solutions

- Many problems have well-defined solutions that are obvious once the problem has been defined.
- A set of logical steps that we can follow to calculate an exact outcome.
- In linear algebra, there are a suite of methods that you can use to factorize a matrix, depending on if the properties of your matrix are square, rectangular, contain real or imaginary values, and so on.
- Some problems in applied machine learning are well defined and have an analytical solution.
- For example, the method for transforming a categorical variable into a one hot encoding is simple, repeatable and always the same methodology regardless of the number of integer values in the set.
- Unfortunately, most of the problems that we care about solving in machine learning do not have analytical solutions.

Numerical Solutions

- There are many problems that we are interested in that do not have exact solutions.
- We have to make guesses at solutions and test them to see how good the solution is. This involves framing the problem and using trial and error across a set of candidate solutions.
- In essence, the process of finding a numerical solution can be described as a search

Numerical Solutions

These types of solutions have some interesting properties:

- We often easily can tell a good solution from a bad solution.
- We often don’t objectively know what a “*good*” solution looks like; we can only compare the goodness between candidate solutions that we have tested.
- We are often satisfied with an approximate or “*good enough*” solution rather than the single best solution.
- Often the problems that we are trying to solve with numerical solutions are challenging, where any “*good enough*” solution would be useful. It also highlights that there are many solutions to a given problem and even that many of them may be good enough to be usable.
- Most of the problems that we are interested in solving in applied machine learning require a numerical solution.

Analytical vs Numerical Solutions

- An analytical solution involves framing the problem in a well-understood form and calculating the exact solution.
- A numerical solution means making guesses at the solution and testing whether the problem is solved well enough to stop.
- An example is the square root that can be solved both ways.
- We prefer the analytical method in general because it is faster and because the solution is exact. Nevertheless, sometimes we must resort to a numerical method due to limitations of time or hardware capacity.
- A good example is in finding the coefficients in a linear regression equation that can be calculated analytically (e.g. using linear algebra), but can be solved numerically when we cannot fit all the data into the memory of a single computer in order to perform the analytical calculation (e.g. via gradient descent).
- Sometimes, the analytical solution is unknown and all we have to work with is the numerical approach.

Numerical Solutions in Machine Learning

- Applied machine learning is a numerical discipline.
- The core of a given machine learning model is an optimization problem, which is really a search for a set of terms with unknown values needed to fill an equation. Each algorithm has a different "*equation*" and "*terms*", using this terminology loosely.
- The equation is easy to calculate in order to make a prediction for a given set of terms, but we don't know the terms to use in order to get a "*good*" or even "*best*" set of predictions on a given set of data.
- This is the numerical optimization problem that we always seek to solve.
- It's numerical, because we are trying to solve the optimization problem with noisy, incomplete, and error-prone limited samples of observations from our domain. The model is trying hard to interpret the data and create a map between the inputs and the outputs of these observations.

Types of Errors

Numerically computed solutions are subject to certain errors. Mainly there are three types of errors. They are inherent errors, truncation errors and errors due to rounding.

1. Inherent errors or experimental errors arise due to the assumptions made in the mathematical modelling of problem. It can also arise when the data is obtained from certain physical measurements of the parameters of the problem. i.e., errors arising from measurements.

Types of Errors

2. Truncation errors are those errors corresponding to the fact that a finite (or infinite) sequence of computational steps necessary to produce an exact result is “truncated” prematurely after a certain number of steps.

Types of Errors

3. Round of errors are errors arising from the process of rounding off during computation. These are also called chopping, i.e. discarding all decimals from some decimals on.

Errors in Numerical Procedures

The relationship between the exact, or true, result and the approximation can be formulated as

True value = approximation + error

We get,

E_t = | True value − approximation |

where E_t is used to designate the exact value of the absolute error. The subscript *t* is included to designate that this is the “absolute true” error.

Errors in Numerical Procedures

- A shortcoming of this definition is that it takes no account of the order of magnitude of the value under examination.
- $Absolute\ True\ relative\ error = \left|\frac{true\ error}{true\ value}\right|$
- $Absolute\ percent\ relative\ error\ \varepsilon_T = \left|\frac{true\ error}{true\ value}\right| \times 100$

Errors in Numerical Procedures

- However, in machine learning applications, we will obviously not know the true answer beforehand. For these situations, an alternative is to normalize the error using the best available estimate of the true value, that is, to the approximation itself,

$$\varepsilon_a = \frac{approximate\ error}{approximation}\ 100\%$$

Errors in Numerical Procedures

- Certain numerical methods use an *iterative approach* to compute answers. In such an approach, a present approximation is made on the basis of a previous approximation. This process is performed repeatedly, or iteratively, to successively compute better and better approximations. For such cases, the error is often estimated as the difference between previous and current approximations. Thus, percent relative error is determined according to
- $\varepsilon_a = \left| \frac{current\ approximation - previous\ approximation}{current\ approximation} \right| \times 100$

Errors in Numerical Procedures

- When performing computations, we may not be concerned with the sign of the error, but we are interested in whether the percent absolute value is lower than a prespecified percent tolerance ε.

$$|\varepsilon_a| < \varepsilon$$

We say that the estimate is correct to *n* decimal digits if:

$$|\,\mathrm{Error}\,| \le 10^{-n}$$

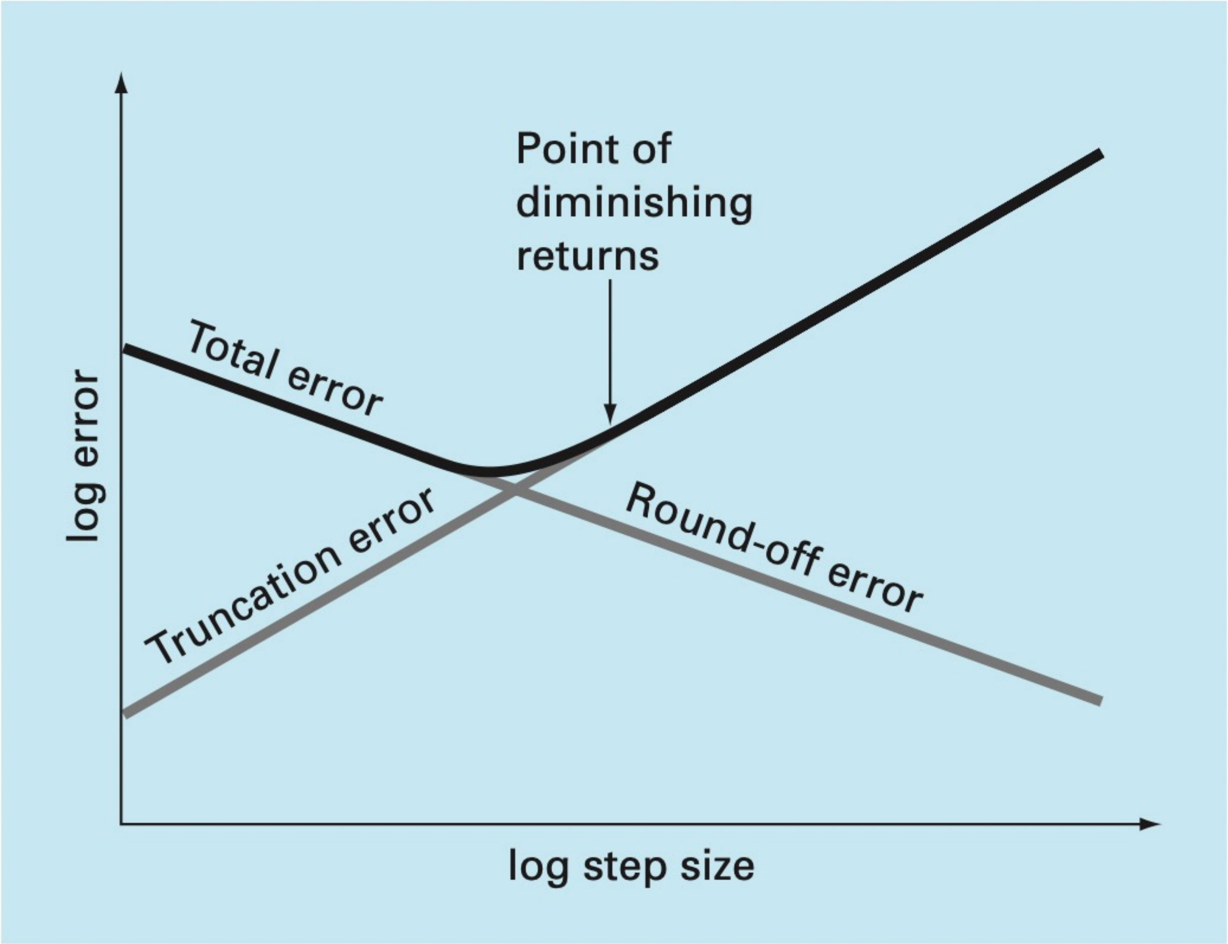
We say that the estimate is correct to *n* decimal digits **rounded** if:

$$|\,\mathrm{Error}\,| \le \frac{1}{2} \times 10^{-n}$$

Errors in Numerical Procedures

- The *total numerical error* is the summation of the truncation and round-off errors. In general, the only way to minimize round-off errors is to increase the number of significant figures of the computer. Further, we have noted that round-off error will *increase* due to subtractive cancellation or due to an increase in the number of computations in an analysis. In contrast, the truncation error can be reduced by decreasing the step size. Because a decrease in step size can lead to subtractive cancellation or to an increase in computations, the truncation errors are *decreased* as the round-off errors are *increased.*

Errors in Numerical Procedures



Root Finding Methods

Root Finding Methods

- An important problem in applied mathematics is to "solve $f(x) = 0$" where $f(x)$ is a function of x.
- The values of x that make $f(x) = 0$ are called the *roots* of the equation. They are also called the *zeros* of $f(x)$.

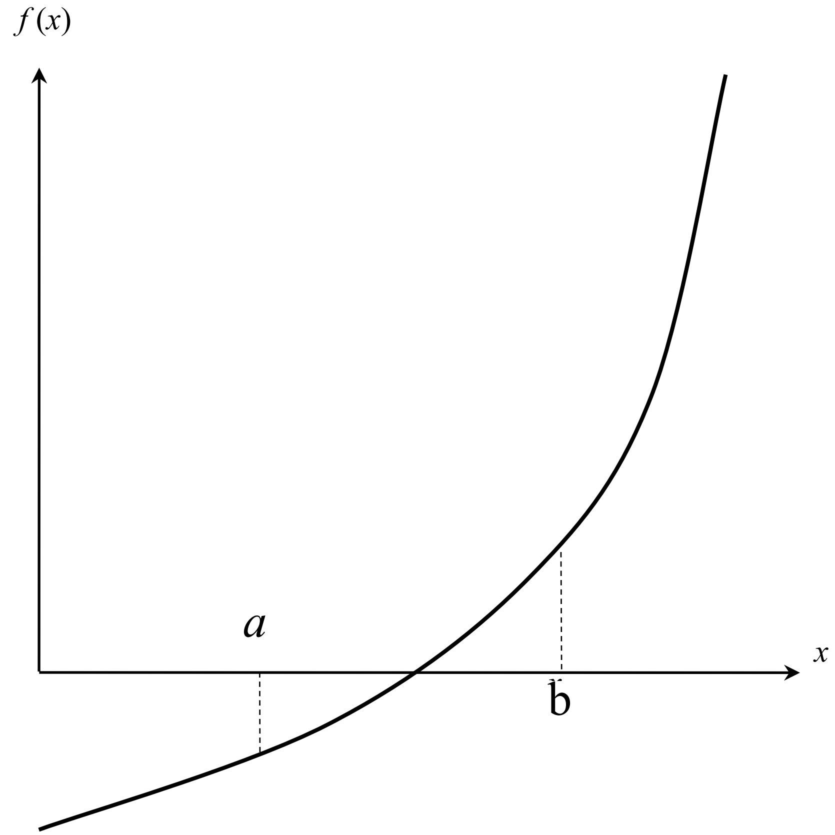
Root Finding Methods

- The root finding methods are divided into two categories: bracketing and open methods.
- The bracketing methods require the limits between which the root lies
- Bisection and False position methods are two known examples of the bracketing methods.
- Open methods require the initial estimation of the solution. Among the open methods, the Newton-Raphson and Secant is most commonly used.
- The most popular method for solving a non-linear equation is the Newton-Raphson method and this method has a high rate of convergence to a solution.

Bisection Method

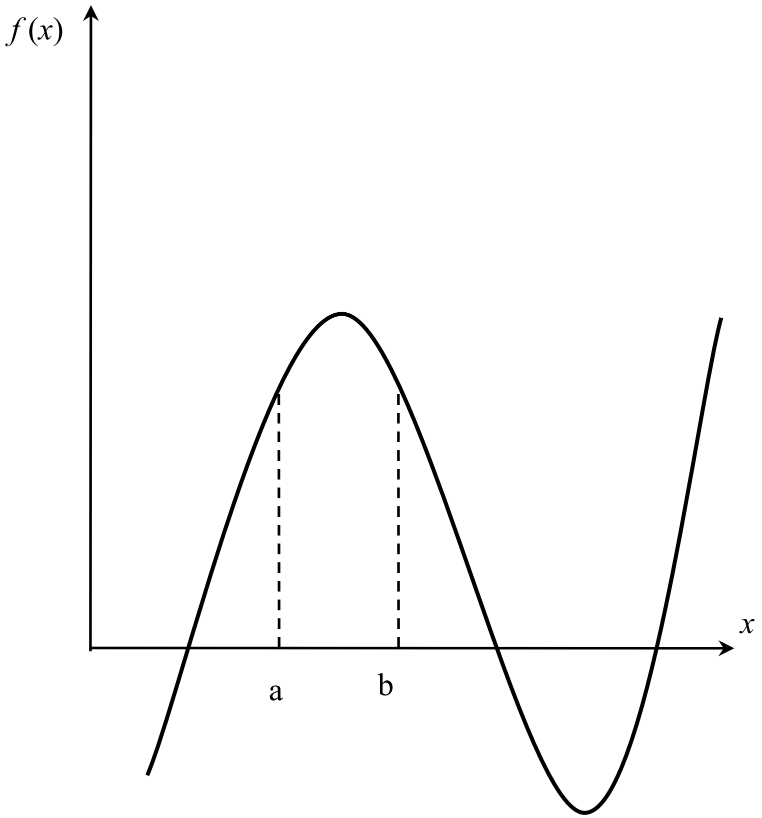
Bisection Method

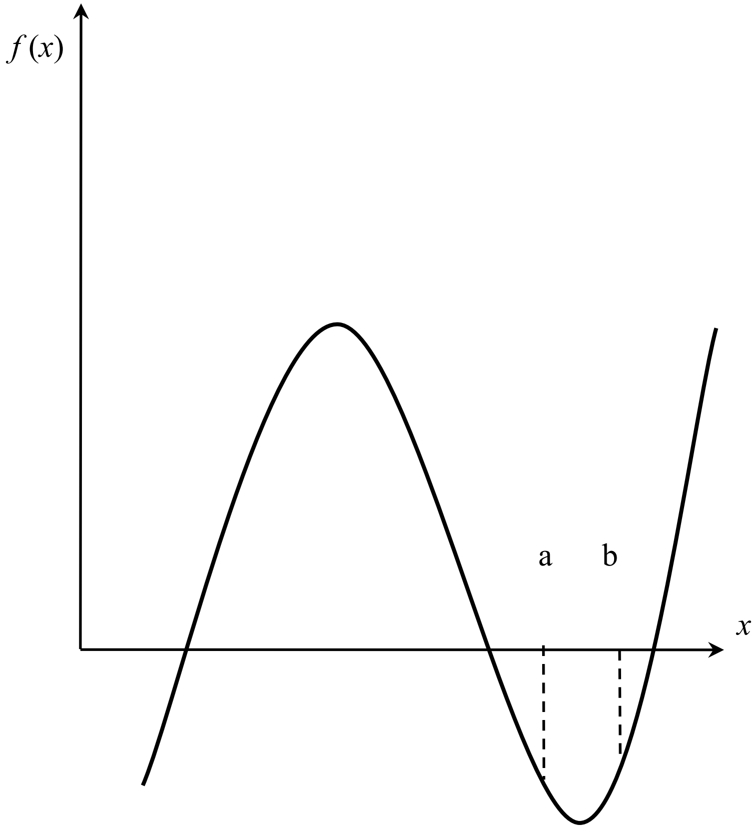
- This method is based on the repeated application of intermediate value property.
- Let the function $f(x)$ be continuous between a and b.
- If the function $f(x)$ satisfies $f(a) \cdot f(b) < 0$, then the equation $f(x) = 0$ has atleast one real root or an odd number of real roots in the interval (a, b).



Bisection Method

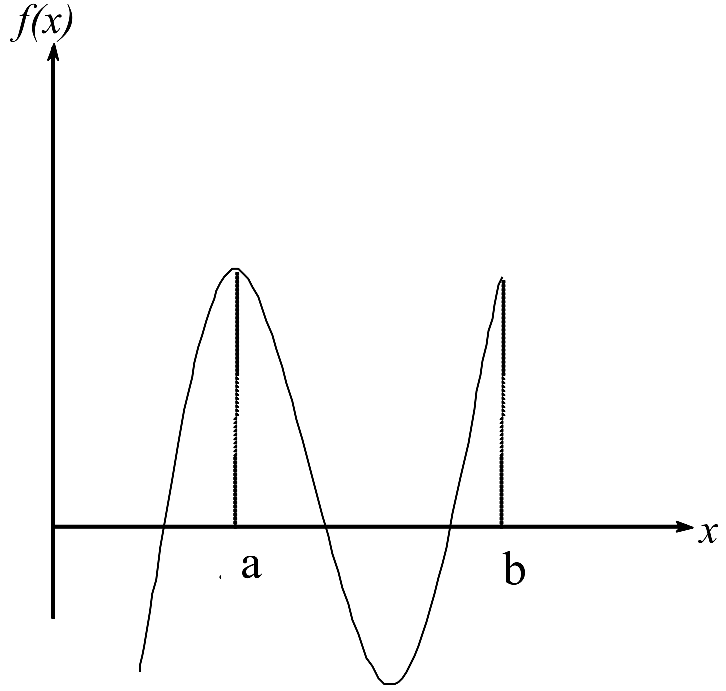
- If the function f(x) does not change sign between two points, there may not be any root of the equation f(x)=0 between the two points.





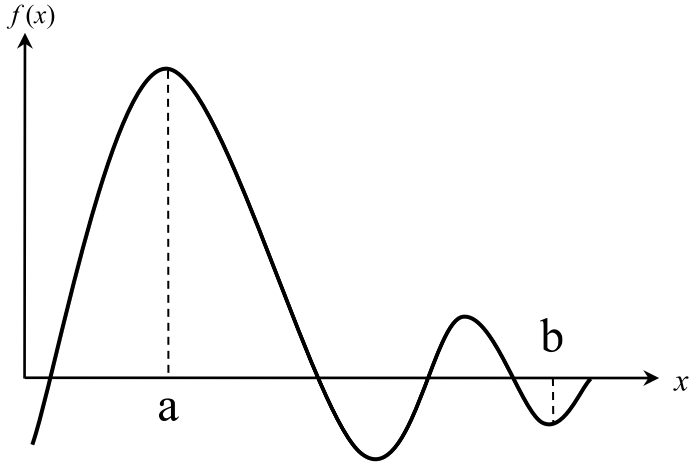
Bisection Method

- If the function f(x) does not change sign between two points, roots of the equation f(x)=0 may still exist between the two points.



Bisection Method

- If the function f(x) changes sign between two points, more than one roots of the equation f(x)=0 may exist between the two points.



Bisection Method

The first approximation to the root is

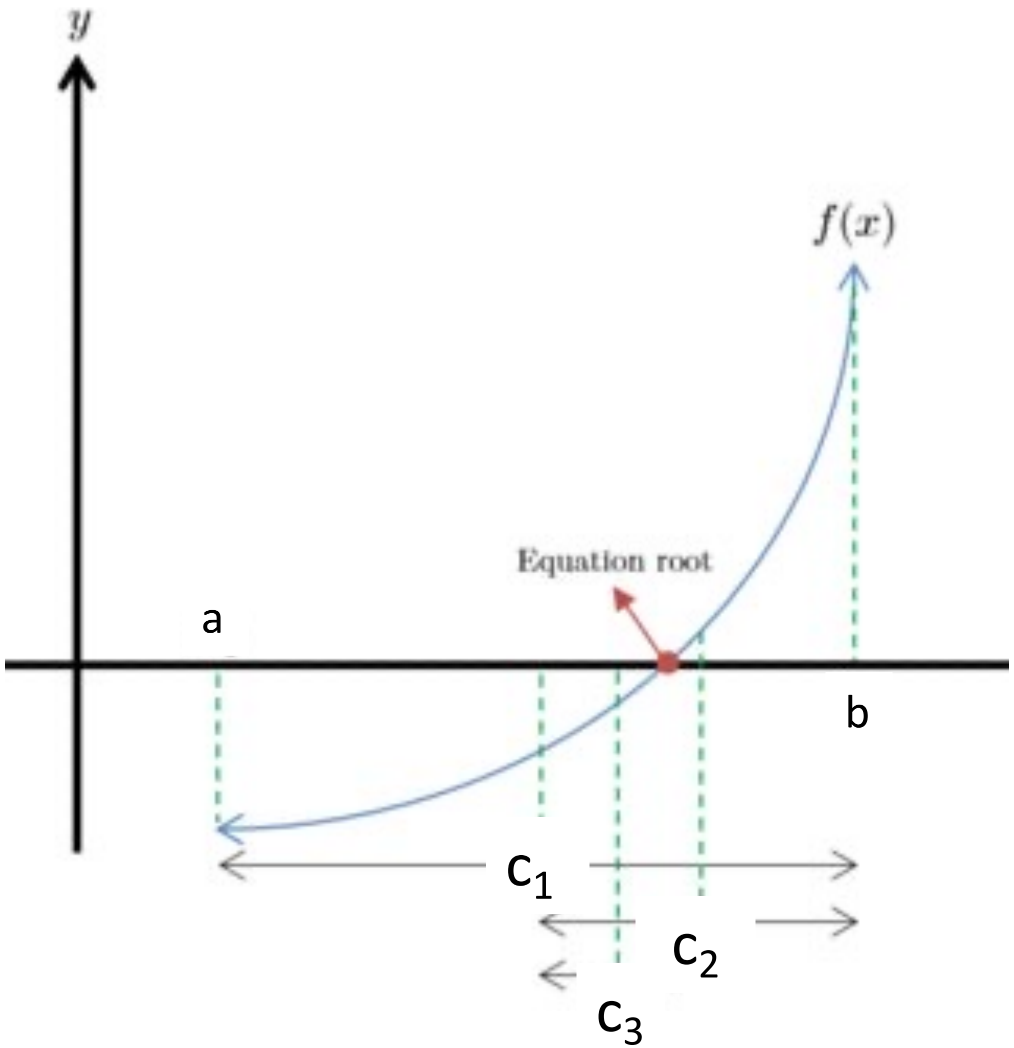
$$c_1 = \frac{(a+b)}{2}.$$

If $f(c_1) = 0$, then c_1 is a root of $f(x) = 0$,

otherwise, the root lies in

(a, c_1) or (c_1, b) according to $f(c_1)$ is (+)ve or (–)ve.

Then we bisect the interval as before and continue the process until the root is found to the desired accuracy.



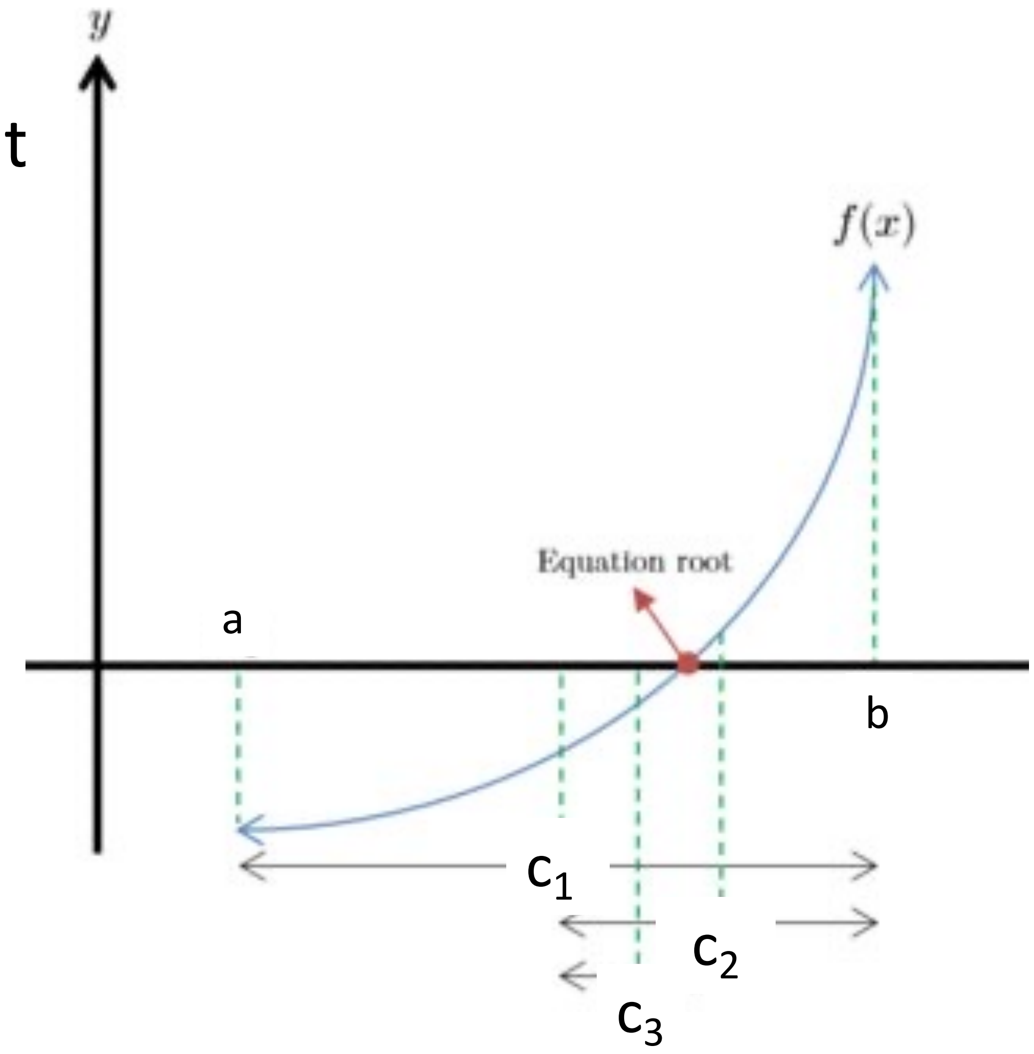
Bisection Method

In the adjoining figure, $f(c_1)$ is (-)ve so that the root lies between b and c_1.

$f(c_1) \,.\, f(b) < 0$

The second approximation to the root is

$$c_2 = \frac{(c_1 + b)}{2}.$$



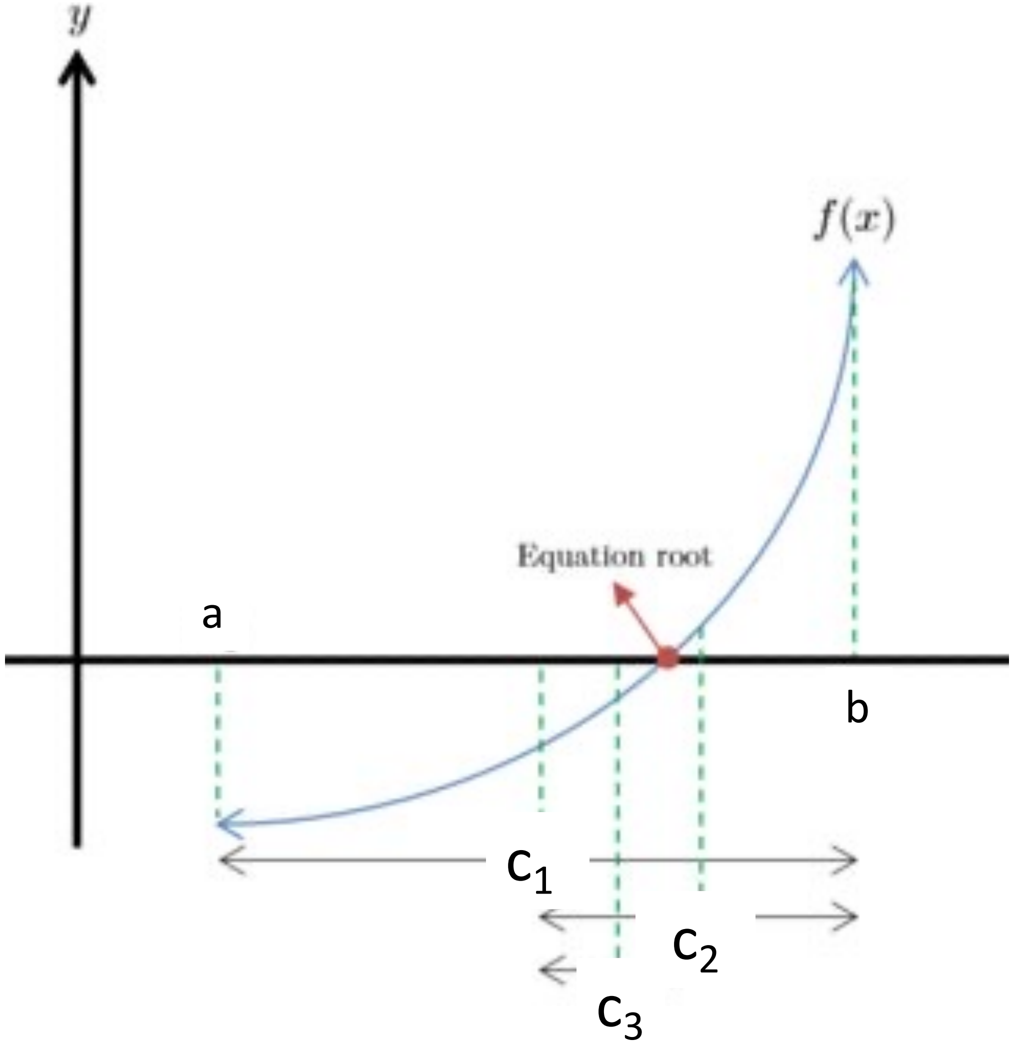
Bisection Method

$f\,(c_2)$ is (+)ve the root lies between c_1 and c_2.

$f(c_1)\;.\;f(c_2) < 0$

Similarly, the third approximation to the root is c_3 and so on.

- Then $c_3 = \frac{(c_1 + c_2)}{2}$.



Advantages and Disadvantages

Advantages of bisection method

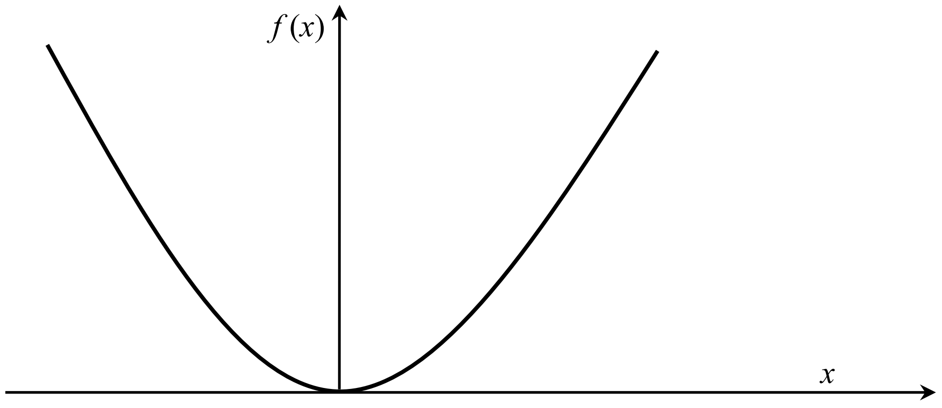
- The bisection method is always convergent. Since the method brackets the root, the method is guaranteed to converge.
- As iterations are conducted, the interval gets halved. So one can guarantee the error in the solution of the equation.

Drawbacks of bisection method

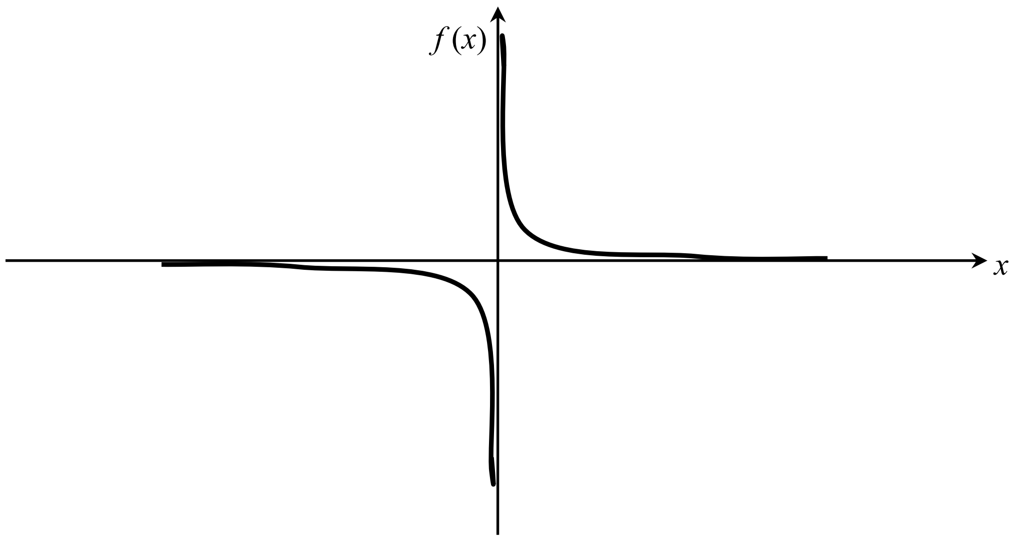
- The convergence of the bisection method is slow as it is simply based on halving the interval.
- If one of the initial guesses is closer to the root, it will take larger number of iterations to reach the root.

Disadvantages

- If a function f(x) is such that it just touches the x-axis such as f(x) = $x^2 = 0$, it will be unable t0 find a, b such that f(a).f(b) < 0



- For functions f(x) where there is a singularity (A singularity in a function is defined as a point where the function becomes infinite) and it reverses sign at the singularity, the bisection method may converge on the singularity
- $f(x) = \frac{1}{x}$



Different Stopping Criterion

- The sequence of $c_k's$ approaches the root. That is why, in case of the Bisection section algorithm, iterative process was stopped, when
- $|c_k - c_{k-1}| < \epsilon$
- It is called x – tolerance criterion (X-TOL).
- The recent most c_k is the estimate of the root.
- There is another stopping criterion called function tolerance (F–TOL). At root, function value is zero, so if estimate is quite near the root then function value would be small. Hence, many times, one would like to stop when
- $|f(c_k)| < \epsilon$

Different Stopping Criterion

- If slope of the curve is small near the root, curve is almost horizontal, and then function tolerance may not be appropriate stopping criterion, because curve is rising slowly, function values in neighbourhood of the root are going to be small, so even if estimate c is not sufficiently near the root, one may stop.

- On the other hand, if it is expected that slope is high near the root or sequence of approximations c k's may converge to the root, like in case of almost vertical graph, then function tolerance ensures that estimate is good approximation to the root.

- In simple words, for f (x) , if f `(x) is high near the root, FTOL would be better to use as stopping criterion

Different Stopping Criterion

- Many times instead of absolute error $|c_k - c_{k-1}|$ bound on Relative error $\frac{|c_k - c_{k-1}|}{|c_k|} < \epsilon$ is used .This needs to be applied when looking answer correct to certain number of significant digits.
- For example, the root could be like 1.2749×10^{-12}
- So here, if we go for $|c_k - c_{k-1}| < 10^{-5}$ we shall get answer zero, but if we go for $\frac{|c_k - c_{k-1}|}{|c_k|} < 10^{-5}$,we would get root correct to 5 significant digits.
- It root is required to be correct to N significant digits, then one should apply
- $\frac{|c_k - c_{k-1}|}{|c_k|} < 10^{-N}$

Bisection Method

$$x^4 - 3x^2 + x - 10 = 0$$

Find an interval of unit length which contains this root.

Bisection Method

$f(0) = 0 - 0 + 0 - 10 = -10$

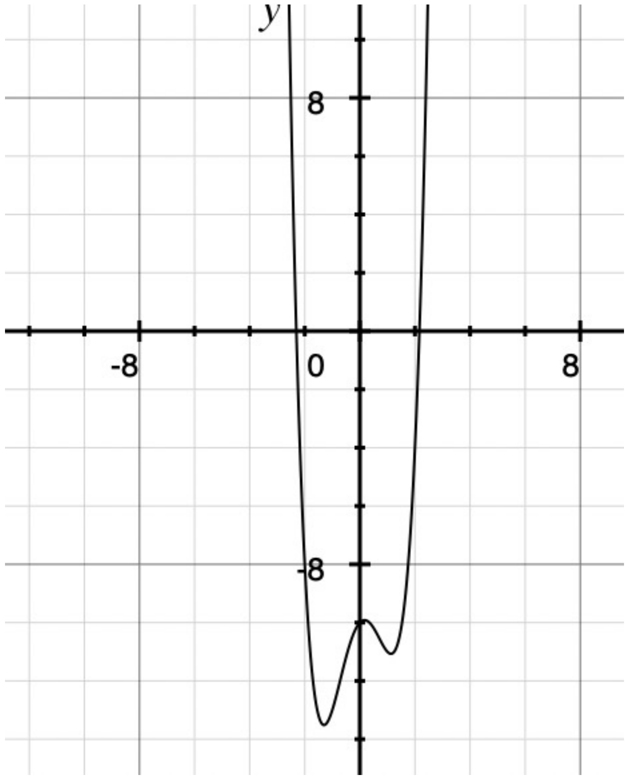
$f(1) = 1 - 3 + 1 - 10 = -11$

$f(2) = 16 - 12 + 2 - 10 = -4$

$f(3) = 81 - 27 + 3 - 10 = 47$

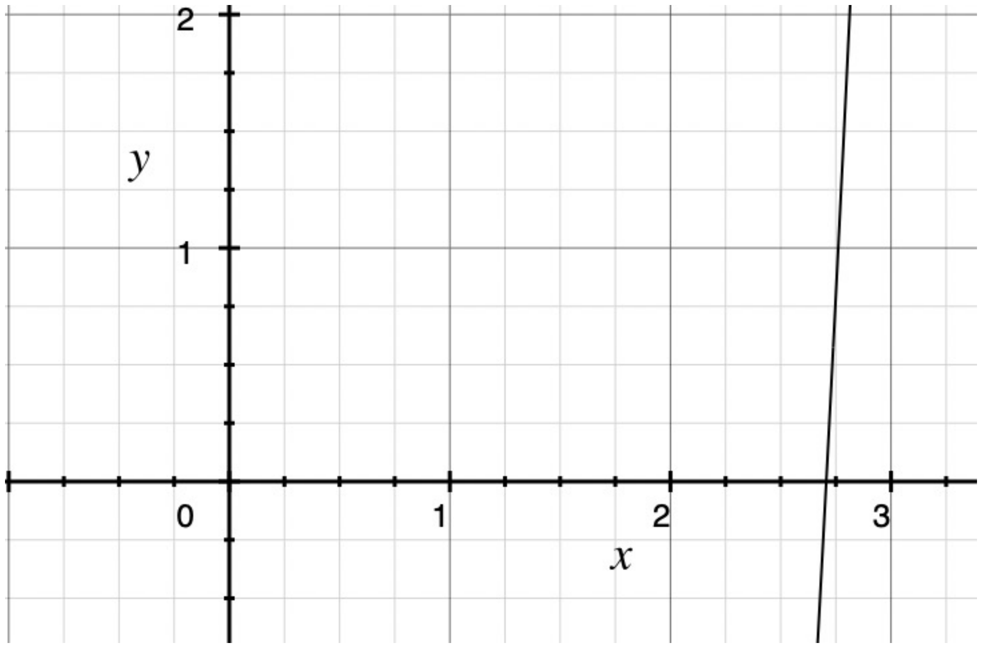
An interval of unit length which contains the root of

$x^4 - 3x^2 + x - 10 = 0$ is (2,3)



Bisection Method

Eg 1. *Find the real root of the equation* $x^3 - 4x - 9 = 0$ *by Bisection method correct. Take* $a = 2.706$, $b = 2.707$, $\epsilon = 0.0001$



Bisection Method

Eg 1. *Find the real root of the equation $x^3 - 4x - 9 = 0$ by Bisection method correct. Take $a = 2.706$, $b = 2.707$,* $\epsilon = 0.0001$

$f(x) = x^3 - 4x - 9$

$f(2.706) = -0.009488$ *i.e.,* (–)ve

and

$f(2.707) = 0.008487$ *i.e.,* (+)ve

Hence, the root lies between 2.706 and 2.707.

x	0	1	2	3
f(x)	-9	-12	-9	6

a	f(a)	b	f(b)
2.706	– 0.009488	2.707	0.008487

Bisection Method

First approximation to the root is

$$c = \frac{(2.706 + 2.707)}{2}$$

c = 2.7065

a	f(a)	b	f(b)	c = (a+b)/2	f(c)	$\lvert c_k - c_{k-1} \rvert$
2.706	– 0.009488	2.707	0.008487	2.7065	– 0.0005025	-

Now $f(c)$ = – 0.0005025 *i.e.,* (–)ve and

f(b) = 0.008487 *i.e.,* (+)ve

Hence, the root lies between 2.7065 and 2.707.

Bisection Method

Second approximation to the root is

$$c = \frac{(2.7065 + 2.707)}{2}$$

= 2.70675

Now $f(c)$ = 0.003992 *i.e.,* (+)ve and $f(a)$ = – 0.0005025 *i.e.,* (–)ve

Hence, the root lies between 2.7065 and 2.70675.

a	f(a)	b	f(b)	c = (a+b)/2	f(c)	$\|c_k - c_{k-1}\|$
2.706	– 0.009488	2.707	0.008487	2.7065	– 0.0005025	-
2.7065	– 0.0005025	2.707	0.008487	2.70675	0.003992	0.00025

Bisection Method

Third approximation to the root is

$$c = \frac{(2.7065 + 2.70675)}{2}$$

= 2.706625

Now $f(c)$ = 0.001744 *i.e.,* (+)ve and $f(a)$ = – 0.0005025 *i.e.,* (–)ve

Hence, the root lies between 2.7065 and 2.706625.

a	f(a)	b	f(b)	c = (a+b)/2	f(c)	$\lvert c_k - c_{k-1} \rvert$
2.706	– 0.009488	2.707	0.008487	2.7065	– 0.0005025	-
2.7065	– 0.0005025	2.707	0.008487	2.70675	0.003992	0.00025
2.7065	– 0.0005025	2.70675	0.003992	2.706625	0.001744	0.000125

Bisection Method

Fourth approximation to the root is

$c = \frac{(2.7065+2.7406625)}{2}$ = 2.7065625

ϵ = 0.0001, $|c_k - c_{k-1}|$ = 0.0000625 < ϵ

Hence, the root is 2.7065625, correct to three decimal places.

a	f(a)	b	f(b)	c = (a+b)/2	f(c)	$\|c_k - c_{k-1}\|$
2.706	– 0.009488	2.707	0.008487	2.7065	– 0.0005025	-
2.7065	– 0.0005025	2.707	0.008487	2.70675	0.003992	0.00025
2.7065	– 0.0005025	2.70675	0.003992	2.706625	0.001744	0.000125
2.7065	– 0.0005025	2.706625	0.001744	2.7065625		0.0000625