EE 202 Numerical Methods for Engineers

COURSE CONTENTS

1. The Newton's Method for Root Approximations.
2. Interpolation by Newton's Divided Differences.
3. Interpolation by Trigonometric Functions.
4. Curve fitting by the Least Squares Algorithm.
5. Series Representation of Functions (Fourier Series)
6. Solution of Differential Equations; Euler's Method.
7. The Runge Kutta 4th Degree Method.
8. The Method of Finite Differences (FD).
9. The Finite Element Method (FEM).
10. Solution of Integral Equations; Method of Moments.
11. Optimization; Convexity and Convergence.
12. The Steepest Descent Method.
13. The Gauss-Newton Method.
14. Other Algorithms than Gradients such as the Genetic Algorithm.

GRADING

LABS	20%
MID TERM	40%
FINAL EXAM	40%

1. The Newton's Method for Root approximations.

The roots of a function are the points where the function is equal to zero.

At these points; the graph of the function touches the x-axis.

Newton's method estimates these roots using tangent lines.

Example: Find the roots of $f(x) = x^2 - 3$ or solve $x^2 - 3 = 0$ by using Newton's method.

Use of the Intermediate Value Theorem for locating an approximate root.

Make an *x*-*y* chart to find the changes of sign in the function.

x	*y*
0	-3
1	-2
2	1

There is a change in the sign between $f(1)$ and $f(2)$.

Intermediate Value Theorem says that there is a root between $x=1$ and $x=2$.

Next step is to find a point of tangency on the function in order to have an initial guess for the root.

Let's have a guess for the root to be x=1.5, i.e., x_1=1.5.

The corresponding y for x_1=1.5 is $f(x)=(1.5)^2-3=2.25-3=-0.75$ which is not very close to 0

We conclude that the point of tangency is (1.5, –.75).

Then we write the equation of the tangent line at the point of tangency which is (1.5, –.75).

For this purpose we first find the derivative at the point of tangency

$f(x)=x^2-3$
$f'(x)=2x$
$f'(1.5)=2\times1.5=3$

i.e., the slope of the tangent line is m=3

To find the tangent line we use the equation $y=mx+b$

The point of tangency was (1.5, –0.75) so

$-0.75=3\times1.5+b \Rightarrow b=-5.25$

So the tangent line equation is $y = 3x – 5.25$

Next we find the point at which the tangent line crosses the x-axis

So we equate the tangent line equation to zero, i.e., $0 = 3x – 5.25 \rightarrow x = 1.75$

We choose this point of x = 1.75 as the new approximate of the root.

The question: How close is this approximation?

To answer this, find the value of the function y at this point of approximation, i.e., at x=1.75. The closer the function is to 0, the more accurate our approximate is.

If the function equals exactly 0, then the root is exact.

For x=1.75, $f(x) = x^2–3$ becomes $f(1.75) = (1.75)^2–3 =.0625$

Not too far from 0 but also not very close to 0.

We conclude that the point of tangency is (1.75,0.0625).

Now with this new approximate root of 1.75 we repeat the above steps, i.e.,

Then we write the equation of the tangent line at the point of tangency which is (1.75,0.0625).

For this purpose we first find the derivative at the point of tangency

$$f(x) = x^2 - 3$$
$$f'(x) = 2x$$
$$f'(1.75) = 2 \times 1.75 = 3.5$$

i.e., the slope of the tangent line is m=3.5

To find the tangent line we use the equation $y = mx + b$

The point of tangency was (1.75,0.0625) so

$$0.0625 = 3.5 \times 1.75 + b \Rightarrow b = -6.0625$$

So the tangent line equation is y=3.5x –6.0625

Next we find the point at which the tangent line crosses the x-axis

So we equate the tangent line equation to zero, i.e., 0=3.5x-6.0625 → x=1.732143

We choose this point of x=1.732143 as the third approximate of the root.

The question: How close is this approximation?

To answer this, find the value of the function y at this point of approximation, i.e., at x=1.732143. The closer the function is to 0, the more accurate our approximate is.

If the function equals exactly 0, then the root is exact.

For x=1.732143, $f(x) = x^2$–3 becomes f(1.732143) = $(1.732143)^2$–3 =0.000319

Not too far from 0 so 1.732143 can be taken as the root.

But if the accuracy is not satisfactory, we can do one or more extra approximations.

Using Newton's method, only one root at a time can be found.

In order to find all the roots of a function using Newton's method, we need to do the same steps for every sign change in the function

So to repeat for the other root

Use of the Intermediate Value Theorem for locating an approximate root.

Make an *x*-*y* chart to find the changes of sign in the function.
$f(x) = x^2 - 3$

x	*y*
0	-3
-1	-2
-2	1

There is a change in the sign between $f(-1)$ and $f(-2)$.

Intermediate Value Theorem says that there is a root between $x = -1$ and $x = -2$.

THE FOLLOWING IS HW-1

Next step is to find a point of tangency on the function in order to have an initial guess for the root.

Let's have a guess for the root to be $x = -1.5$, i.e., $x_1 = -1.5$.

The corresponding y for $x_1 = -1.5$ is $f(x) = (-1.5)^2 - 3 = 2.25 - 3 = -0.75$ which is not very close to 0

We conclude that the point of tangency is (–1.5, –0.75).

Then we write the equation of the tangent line at the point of tangency which is (–1.5, –0.75).

For this purpose we first find the derivative at the point of tangency

$f(x) = x^2 - 3$
$f'(x) = 2x$
$f'(1.5) = 2 \times (-1.5) = -3$

i.e., the slope of the tangent line is m = –3

To find the tangent line we use the equation $y = mx + b$

The point of tangency was (–1.5, –0.75) so

$-0.75 = -3 \times (-1.5) + b \Rightarrow b = -5.25$

So the tangent line equation is $y = -3x - 5.25$

Next we find the point at which the tangent line crosses the *x*-axis

So we equate the tangent line equation to zero, i.e., $0 = -3x - 5.25 \rightarrow x = -1.75$

We choose this point of $x = -1.75$ as the new approximate of the root.

The question: How close is this approximation?

To answer this, find the value of the function y at this point of approximation, i.e., at $x=-1.75$. The closer the function is to 0, the more accurate our approximate is.

If the function equals exactly 0, then the root is exact.

For $x = -1.75$, $f(x) = x^2 - 3$ becomes $f(-1.75) = (-1.75)^2 - 3 = .0625$

Not too far from 0 but also not very close to 0.

Now with this new approximate root of -1.75 we repeat the above steps, i.e.,

The corresponding y for $x_2 = -1.75$ is $f(x) = (-1.75)^2 - 3 = 3.0625 - 3 = 0.0625$ which is not very close to 0

We conclude that the point of tangency is $(-1.75, 0.0625)$.

Then we write the equation of the tangent line at the point of tangency which is $(-1.75, 0.0625)$.

For this purpose we first find the derivative at the point of tangency

$$f(x) = x^2 - 3$$
$$f'(x) = 2x$$
$$f'(1.75) = 2 \times -1.75 = -3.5$$

i.e., the slope of the tangent line is $m = -3.5$

To find the tangent line we use the equation $y = mx + b$

The point of tangency was $(-1.75, 0.0625)$ so

$$0.0625 = -3.5 \times (-1.75) + b \Rightarrow b = -6.0625$$

So the tangent line equation is $y = -3.5x - 6.0625$

Next we find the point at which the tangent line crosses the x-axis

So we equate the tangent line equation to zero, i.e., $0 = -3.5x - 6.0625 \rightarrow x = -1.732143$

We choose this point of $x = -1.732143$ as the third approximate of the root.

The question: How close is this approximation?

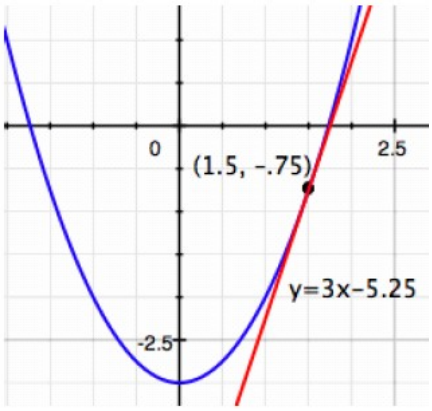
To answer this, find the value of the function y at this point of approximation, i.e., at $x = -1.732143$. The closer the function is to 0, the more accurate our approximate is.

If the function equals exactly 0, then the root is exact.

For $x = -1.732143$, $f(x) = x^2 - 3$ becomes $f(-1.732143) = (-1.732143)^2 - 3 = 0.000319$

Not too far from 0 so –1.732143 can be taken as the root.

But if the accuracy is not satisfactory, we can do one or more extra approximations.



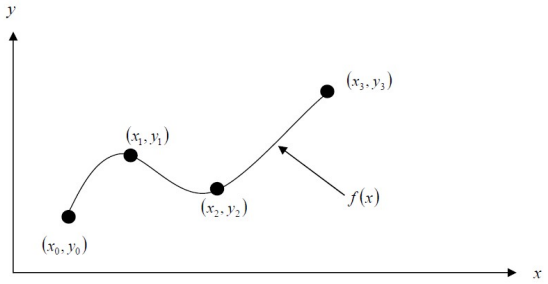
RESULT: By using Newton's Method Roots of $f(x) = x^2 - 3$ are found to be 1.732143 and –1.732143

2. Interpolation by Newton's divided differences.

In many situations, the function $y = f(x)$ may not be known but we may have several data only at discrete points (x_0, y_0), (x_1, y_1), (x_2, y_2), ..., (x_{n-1}, y_{n-1}), (x_n, y_n).

Question: How can we find the value of the function, i.e. y at other values of x?

Using the known $n+1$ discrete data points, we can plot a continuous function $f(x)$.



From this plot we can find the value of y at any x . This is known as interpolation.

If x at which y is to be found is outside the range of x for which the data is given, then it is known as extrapolation.

Question: How is $f(x)$ chosen?

Polynomial is a common choice for interpolation since compared to trigonometric and exponential series, polynomials are easily evaluated, differentiated and integrated.

In polynomial interpolation, finding a polynomial of order n that passes through the n +1 points is needed.

One of the methods of interpolation is known as Newton's divided difference polynomial method (other methods are direct method and the Lagrangian interpolation method).

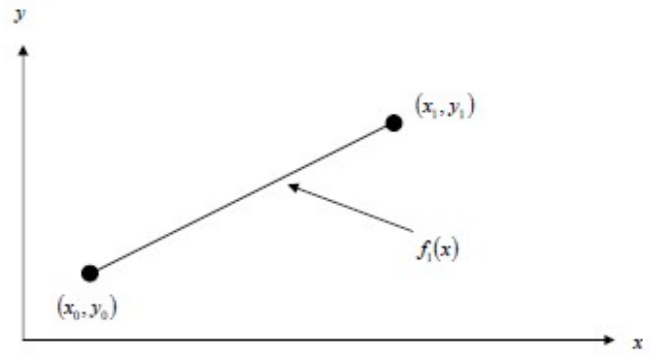
Newton's Divided Difference Polynomial Method

Linear Interpolation (First order polynomial interpolation by Newton's divided difference polynomial method)

(x_0 , y_0) and (x_1 , y_1) are given.

Function is $y = f(x) \rightarrow y_0 = f(x_0)$, $y_1 = f(x_1)$

Fit a linear interpolant $f_1(x)$ passing through the data which is linear.



The equation is: $\dfrac{y - y_0}{y_1 - y_0} = \dfrac{x - x_0}{x_1 - x_0}$

$$y = (y_1 - y_0)\frac{x - x_0}{x_1 - x_0} + y_0 \quad \Rightarrow \quad \boxed{f(x) = [f(x_1) - f(x_0)]\frac{x - x_0}{x_1 - x_0} + f(x_0)}$$

Example: For a function, the following data is given. By using Linear Interpolation (First order polynomial interpolation by Newton's divided difference polynomial method) find y at $x = 16$.

x	y
0	0
10	227.04
15	362.78
20	517.35
22.5	602.97
30	901.67

Two data points that are closest to $x = 16$ are chosen which are $x_0 = 15$ and $x_1 = 20$.

Using $f(x) = [f(x_1) - f(x_0)]\frac{x - x_0}{x_1 - x_0} + f(x_0)$ we have

$$f(x) = [517.35 - 362.78]\frac{x - 15}{20 - 15} + 362.78$$

At the required point of $x = 16$

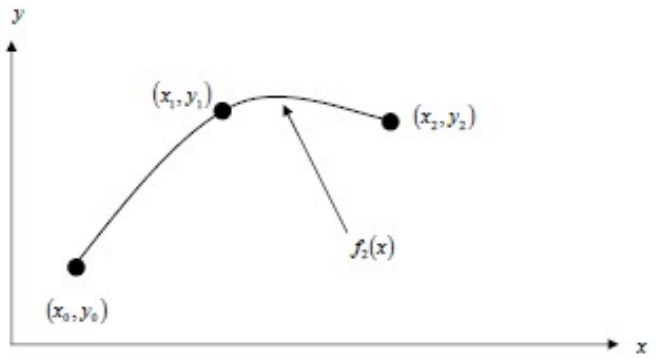
$$f(x) = [517.35 - 362.78]\frac{16 - 15}{20 - 15} + 362.78 = 393.694$$

Quadratic Interpolation (Second order polynomial interpolation by Newton's divided difference polynomial method)

(x_0, y_0) and (x_1, y_1) and (x_2, y_2) are given.

Function is $y = f(x) \rightarrow y_0 = f(x_0)$, $y_1 = f(x_1)$, $y_2 = f(x_2)$

Fit a quadratic interpolant $f_2(x)$ passing through the data which is quadratic.



The equation is: $f_2(x) = b_0 + b_1(x - x_0) + b_2(x - x_0)(x - x_1)$

When $x = x_0$,

$$f_2(x_0) = b_0 + b_1(x_0 - x_0) + b_2(x_0 - x_0)(x_0 - x_1) = b_0 \Rightarrow b_0 = f_2(x_0)$$

When $x = x_1$,

$$f_2(x_1) = b_0 + b_1(x_1 - x_0) + b_2(x_1 - x_0)(x_1 - x_1) = b_0 + b_1(x_1 - x_0) \Rightarrow f_2(x_1) = b_0 + b_1(x_1 - x_0)$$

$$f_2(x_1) = f_2(x_0) + b_1(x_1 - x_0) \Rightarrow b_1 = \frac{f_2(x_1) - f_2(x_0)}{x_1 - x_0}$$

When $x = x_2$,

$$f_2(x_2) = b_0 + b_1(x_2 - x_0) + b_2(x_2 - x_0)(x_2 - x_1)$$

$$f_2(x_2) = f_2(x_0) + \frac{f_2(x_1) - f_2(x_0)}{x_1 - x_0}(x_2 - x_0) + b_2(x_2 - x_0)(x_2 - x_1)$$

Taking

$$b_2 = \frac{\dfrac{f_2(x_2) - f_2(x_1)}{x_2 - x_1} - \dfrac{f_2(x_1) - f_2(x_0)}{x_1 - x_0}}{x_2 - x_0}$$

Inserting b_0, b_1 and b_2, the quadratic interpolant function becomes

$$f_2(x) = f_2(x_0) + \frac{f_2(x_1) - f_2(x_0)}{x_1 - x_0}(x - x_0) + \frac{\dfrac{f_2(x_2) - f_2(x_1)}{x_2 - x_1} - \dfrac{f_2(x_1) - f_2(x_0)}{x_1 - x_0}}{x_2 - x_0}(x - x_0)(x - x_1)$$

To check: When $x = x_0$, $f_2(x) = f_2(x_0)$, when $x = x_1$, $f_2(x) = f_2(x_1)$, when $x = x_2$

$$f_2(x) = f_2(x_0) + \frac{f_2(x_1) - f_2(x_0)}{x_1 - x_0}(x - x_0) + \frac{\dfrac{f_2(x_2) - f_2(x_1)}{x_2 - x_1} - \dfrac{f_2(x_1) - f_2(x_0)}{x_1 - x_0}}{x_2 - x_0}(x - x_0)(x - x_1) = f_2(x)$$

So OK.

Example: For a function, the following data is given. By using Quadratic Interpolation (Second order polynomial interpolation by Newton's divided difference polynomial method) find y at $x = 16$.

x	y
0	0
10	227.04

15	362.78
20	517.35
22.5	602.97
30	901.67

Three data points that are closest to $x = 16$ are chosen which are $x_0 = 10$, $x_1 = 15$ and $x_2 = 20$. Using

$$f_2(x) = f_2(x_0) + \frac{f_2(x_1) - f_2(x_0)}{x_1 - x_0}(x - x_0) + \frac{\dfrac{f_2(x_2) - f_2(x_1)}{x_2 - x_1} - \dfrac{f_2(x_1) - f_2(x_0)}{x_1 - x_0}}{x_2 - x_0}(x - x_0)(x - x_1)$$

we have the quadratic interpolation equation as

$$f_2(x) = 227.04 + \frac{362.78 - 227.04}{15 - 10}(x - 10) + \frac{\dfrac{517.35 - 362.78}{20 - 15} - \dfrac{362.78 - 227.04}{15 - 10}}{20 - 10}(x - 10)(x - 15)$$

Evaluating the above quadratic interpolation equation at the required point, i.e., at $x = 16$, we have

$$f_2(16) = 227.04 + \frac{362.78 - 227.04}{15 - 10}(16 - 10) + \frac{\dfrac{517.35 - 362.78}{20 - 15} - \dfrac{362.78 - 227.04}{15 - 10}}{20 - 10}(16 - 10)(16 - 15)$$

$$= 227.04 + \frac{362.78 - 227.04}{5}6 + \frac{\dfrac{517.35 - 362.78}{5} - \dfrac{362.78 - 227.04}{5}}{10}6 = 392.1876$$

General Form of Newton's Divided Difference Polynomial

We have found linear and quadratic interpolants for Newton's divided difference method. Recalling the quadratic polynomial interpolant formula which is

$$f_2(x) = b_0 + b_1(x - x_0) + b_2(x - x_0)(x - x_1)$$

where

$$b_0 = f_2(x_0),\ b_1 = \frac{f_2(x_1) - f_2(x_0)}{x_1 - x_0},\ b_2 = \frac{\dfrac{f_2(x_2) - f_2(x_1)}{x_2 - x_1} - \dfrac{f_2(x_1) - f_2(x_0)}{x_1 - x_0}}{x_2 - x_0}$$

Note that b_0 , b_1 and b_2 are finite divided differences.

b_0 , b_1 and b_2 are the first, second and third finite divided differences, respectively. Denoting the first divided difference by $f[x_0] = f_2(x_0)$,

the second divided difference by $f[x_1, x_0] = \dfrac{f_2(x_1) - f_2(x_0)}{x_1 - x_0}$,

the third divided difference by $f[x_2, x_1, x_0] = \dfrac{f[x_2, x_1] - f[x_1, x_0]}{x_2 - x_0}$

Here $f[x_0]$, $f[x_1, x_0]$, $f[x_2, x_1, x_0]$ are known as the bracketed functions of their variables enclosed in the square brackets.

Writing the quadratic interpolation function by the bracketed functions, we have

$$f_2(x) = b_0 + b_1(x - x_0) + b_2(x - x_0)(x - x_1)$$
$$= f[x_0] + f[x_1, x_0](x - x_0) + f[x_2, x_1, x_0](x - x_0)(x - x_1)$$

We can generalize this equation and write the general form of the Newton's divided difference polynomial for $n + 1$ data points
(x_0, y_0), (x_1, y_1), (x_2, y_2), ..., (x_{n-1}, y_{n-1}), (x_n, y_n) as

$$f_n(x) = b_0 + b_1(x - x_0) + \ldots + b_n(x - x_0)(x - x_1)\ldots(x - x_{n-1})$$

where

$$b_0 = f[x_0]$$
$$b_1 = f[x_1, x_0]$$
$$b_2 = f[x_2, x_1, x_0]$$

.

.

.

$$b_{n-1} = f[x_{n-1}, x_{n-2}, \ldots, x_0]$$
$$b_n = f[x_n, x_{n-1}, \ldots, x_0]$$

Here the definition of the m^{th} divided difference is

$$b_m = f[x_m, x_{m-1}, \ldots, x_0] = \frac{f[x_m, x_{m-1}, \ldots, x_1] - f[x_{m-1}, x_{m-2}, \ldots, x_0]}{x_m - x_0}$$

For example, for the third order polynomial where the data points are
(x_0, y_0), (x_1, y_1), (x_2, y_2) and (x_3, y_3), the interpolation function is

$$f_3(x) = f[x_0] + f[x_1, x_0](x - x_0) + f[x_2, x_1, x_0](x - x_0)(x - x_1)$$
$$+ f[x_3, x_2, x_1, x_0](x - x_0)(x - x_1)(x - x_2)$$

Example: **HW:** For a function, the following data is given. By using third order polynomial interpolation by Newton's divided difference polynomial method, find y at $x = 16$.

x	y

0	0
10	227.04
15	362.78
20	517.35
22.5	602.97
30	901.67

3. Interpolation by Trigonometric Functions.

Consider periodic functions, i.e., functions that satisfy

$f(t) = f(t+T) \quad \forall t \ \varepsilon(-\infty, \infty)$

where T is the period. Taking $T = 2\pi$

$f(t) = f(t+2\pi)$

$f(t)$ can be approximated by a trigonometric polynomial, as:

$$\boxed{p_n(t) = a_0 + \sum_{j=1}^{n} \left[a_j \cos(jt) + b_j \sin(jt) \right]}$$

which is an n^{th} degree polynomial with $|a_n| + |b_n| \neq 0$.

We have $f(t_i) = p_n(t_i), \quad i = 0,1,2,....,2n$

$0 \leq t_0 < t_1 < t_2 < ... < t_{2n} < 2\pi$

We want $2n+1$ points in t because we want $2n+1$ coefficients.

Noting that $e^{i\theta} = \cos\theta + i\sin\theta$, $p_n(t)$ can be written as

$$p_n(t) = \sum_{j=-n}^{n} c_j e^{ijt}$$

where $c_0 = a_0$, $c_j = 0.5(a_j - ib_j)$, $c_{-j} = 0.5(a_j + ib_j)$, $1 \leq j \leq n$

To determine $\{c_j\}$ we find $\{a_j\}$ and $\{b_j\}$.

Let $z = e^{it}$, then $p_n(t) = \sum_{j=-n}^{n} c_j z^j$

$\therefore \quad z^n p_n(t)$ is a polynomial of degree $\leq 2n$.

Thus, interpolation means that

$$p_n\left(z_j\right)=f\left(t_j\right) \qquad j=0,1,...,2n$$

If $2n+1$ distinct points are taken on the unit circle $|z|=1$

Interpolation at Evenly Spaced Points

Consider $t_j=\frac{2\pi}{2n}j, \qquad j=0,1,2,...,n$

Theorem: The coefficients of $p_n\left(t_j\right)=\sum_{k=-n}^{n}c_k e^{ikt_j}$ for $j=0,1,...,2n$ are given by

$$c_k=\frac{1}{2n+1}\sum_{j=0}^{2n}e^{-ikt_j}f\left(t_j\right), \quad k=-n,...,0,....n$$

Then the coefficients are

$$a_0=c_0,\ a_j=c_j+c_{-j},\ b_j=i\left(c_j-c_{-j}\right)$$

Example: Construct trigonometric polynomial interpolation of degree 2 to $f\left(t\right)=e^{\sin t+\cos t}$ on $\left[0,2\pi\right]$,

$$p_2\left(t_j\right)=e^{\sin t_j+\cos t_j}$$

$$t_j=\frac{2\pi}{2n}j=\frac{2\pi}{4}j=\frac{\pi}{2}j\Rightarrow t_0=0,t_1=\frac{\pi}{2},t_2=\pi,t_3=\frac{3\pi}{2},t_4=2\pi$$

i.e., At $j=0,\ t_0=0,\ f\left(t_0=0\right)=e^{\sin(0)+\cos(0)}=e^1=e$,

at $j=1,\ t_1=\pi/2,\ f\left(t_1=\pi/2\right)=e^{\sin(\pi/2)+\cos(\pi/2)}=e^1=e$

at $j=2,\ t_2=\pi,\ f\left(t_2=\pi\right)=e^{\sin(\pi)+\cos(\pi)}=e^{-1}$

at $j=3,\ t_3=3\pi/2,\ f\left(t_3=3\pi/2\right)=e^{\sin\left(3\pi/2\right)+\cos\left(3\pi/2\right)}=e^{-1}$

at $j=4,\ t_4=2\pi,\ f\left(t_4=2\pi\right)=e^{\sin(2\pi)+\cos(2\pi)}=e^1=e$

$$c_k = \frac{1}{2n+1}\sum_{j=0}^{2n} e^{-ikt_j} f\left(t_j\right), \qquad k = -n,...,0,....n$$

$$c_k = \frac{1}{2n+1}\sum_{j=0}^{4} e^{-ikt_j} f\left(t_j\right) = \frac{1}{4+1}\sum_{j=0}^{4} e^{-ik\frac{\pi}{2}j} e^{\sin\left(\frac{\pi}{2}j\right)+\cos\left(\frac{\pi}{2}j\right)}$$

$$= \frac{1}{4+1}\left[e^{0}e^{\sin(0)+\cos(0)} + e^{-ik\frac{\pi}{2}}e^{\sin\left(\frac{\pi}{2}\right)+\cos\left(\frac{\pi}{2}\right)} + e^{-ik\pi}e^{\sin(\pi)+\cos(\pi)} + e^{-ik\frac{\pi}{2}3}e^{\sin\left(\frac{\pi}{2}3\right)+\cos\left(\frac{\pi}{2}3\right)} + e^{-ik2\pi}e^{\sin(2\pi)+\cos(2\pi)} \right.$$

$$= 0.2\left[e + e^{-ik\frac{\pi}{2}}e + e^{-ik\pi}e^{-1} + e^{-ik\frac{\pi}{2}3}e^{-1} + e^{-ik2\pi}e \right]$$

$$= 0.2\left[e\left(1 + e^{-ik\frac{\pi}{2}} + e^{-ik2\pi}\right) + e^{-1}\left(e^{-ik\pi} + e^{-ik\frac{\pi}{2}3}\right) \right]$$

Using $a_0 = c_0,\ a_k = c_k + c_{-k},\ b_k = i\left(c_k - c_{-k}\right)$

$$a_0 = 0.2\left[e\left(1+1+1\right) + e^{-1}\left(1+1\right) \right] = 0.2\left[3e + 2e^{-1} \right]$$

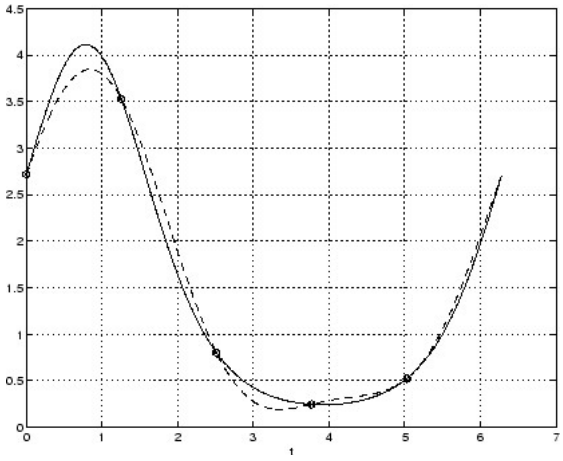
$$a_1 = c_1 + c_{-1} = 0.2\left[e\left(1 + e^{-i\frac{\pi}{2}} + e^{-i2\pi}\right) + e^{-1}\left(e^{-i\pi} + e^{-i\frac{\pi}{2}3}\right) \right]$$
$$+ 0.2\left[e\left(1 + e^{i\frac{\pi}{2}} + e^{i2\pi}\right) + e^{-1}\left(e^{i\pi} + e^{i\frac{\pi}{2}3}\right) \right]$$

$$a_2 = c_2 + c_{-2} = 0.2\left[e\left(1 + e^{-i\pi} + e^{-i4\pi}\right) + e^{-1}\left(e^{-i2\pi} + e^{-i3\pi}\right) \right]$$
$$+ 0.2\left[e\left(1 + e^{i\pi} + e^{i4\pi}\right) + e^{-1}\left(e^{2i\pi} + e^{i3\pi}\right) \right]$$

Similarly b_1 and b_2 is found and inserting the coefficients into

$p_n(t) = a_0 + \sum_{j=1}^{n}\left[a_j \cos\left(jt\right) + b_j \sin\left(jt\right) \right]$, trigonometric polynomial interpolation of degree 2 is found. Then at any t value, e.g. at $t = 0.3\ \pi$, the interpolation value can be found from

$$p_n\left(0.3\pi\right) = a_0 + \sum_{j=1}^{n}\left[a_j \cos\left(j0.3\pi\right) + b_j \sin\left(j0.3\pi\right) \right]$$

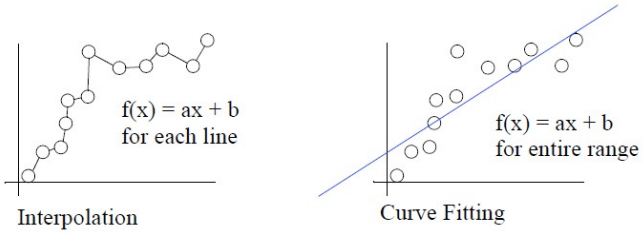


4. Curve fitting by the Least Squares Algorithm

Curve fitting is capturing the trend in the data by assigning a single function across the entire range.

A straight line is described by $f(x) = ax + b$

We want to find the coefficients 'a' and 'b' such that $f(x)$ fits the data well

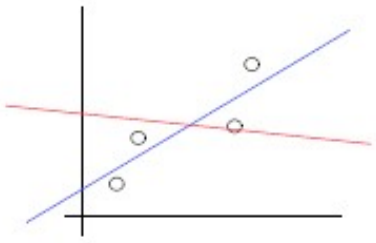


Linear curve fitting (linear regression)

Given a straight line $f(x) = ax + b$

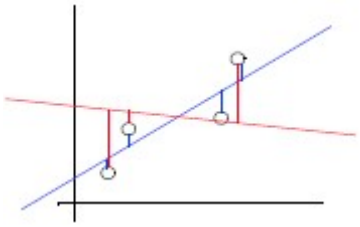
How to choose the coefficients which best fits the line to the data?

What makes a certain straight line a good fit?



Which one of these two lines fits better?

Consider the distance between the data and points on the line



For each straight line add the length of all the verticle lines

This can express the error between data and fitted line

The straight line that gives the minimum error is the best fit.

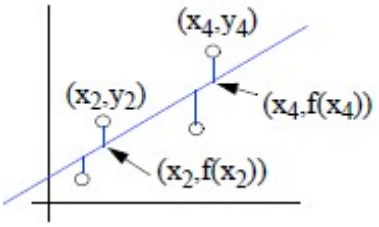
Another way in determining the error can be obtained by squaring the distance.In this way:

1) positive or negative error have the same value (data point is above or below the line)
2) Weight greater errors more heavily

For this purpose:

Denote the data values by (x, y) and denote the points on the fitted line as $(x, f(x))$.

Then sum the error at the data points.



For 4 data points

$$Error = \sum (d_i)^2 = \left[y_1 - f(x_1)\right]^2 + \left[y_2 - f(x_2)\right]^2 + \left[y_3 - f(x_3)\right]^2 + \left[y_4 - f(x_4)\right]^2$$

Since the fit is a straight line, substituting $f(x) = ax + b$

$$Error = \sum_{i=1}^{\#\text{ data points}} \left[y_i - f(x_i)\right]^2 = \sum_{i=1}^{\#\text{ data points}} \left[y_i - (ax_i + b)\right]^2$$

The best line is the one which has the minimum error between the line and data points.

This is called the least squares method, since we minimize the square of the error.

So the problem is reduced to minimizing the "*Error*" expression. Thus taking the derivative and equating the derivative to zero

$$\frac{\partial(Error)}{\partial a} = -2\sum_{i=1}^{n} x_i (y_i - ax_i - b) = 0$$

$$\frac{\partial(Error)}{\partial b} = -2\sum_{i=1}^{n}(y_i - ax_i - b) = 0$$

To solve for a and b, we write

$$a\sum_{i=1}^{n}x_i^2 + b\sum_{i=1}^{n}x_i = \sum_{i=1}^{n}(x_i y_i)$$

$$a\sum_{i=1}^{n}x_i + bn = \sum_{i=1}^{n}y_i$$

Converting these 2 equations to matrix form

$$\begin{bmatrix} n & \sum_{i=1}^{n}x_i \\ \sum_{i=1}^{n}x_i & \sum_{i=1}^{n}x_i^2 \end{bmatrix}\begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n}y_i \\ \sum_{i=1}^{n}(x_i y_i) \end{bmatrix}$$

The solution is $\begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{n}x_i \\ \sum_{i=1}^{n}x_i & \sum_{i=1}^{n}x_i^2 \end{bmatrix}^{-1}\begin{bmatrix} \sum_{i=1}^{n}y_i \\ \sum_{i=1}^{n}(x_i y_i) \end{bmatrix}$

For a 2x2 matrix

If $A = \begin{bmatrix} e & f \\ g & h \end{bmatrix} \Rightarrow A^{-1} = \begin{bmatrix} e & f \\ g & h \end{bmatrix}^{-1} \Rightarrow$ Taking the transpose $A^T = \begin{bmatrix} e & f \\ g & h \end{bmatrix}^T = \begin{bmatrix} e & g \\ f & h \end{bmatrix}$

Adjugate matrix $A^{Adj} = \begin{bmatrix} h & -f \\ -g & e \end{bmatrix}$, Determinant $A = |A| = eh - fg$, $A^{-1} = \frac{A^{Adj}}{|A|} = \frac{1}{(eh-fg)}\begin{bmatrix} h & -f \\ -g & e \end{bmatrix}$

So $\begin{bmatrix} n & \sum_{i=1}^{n}x_i \\ \sum_{i=1}^{n}x_i & \sum_{i=1}^{n}x_i^2 \end{bmatrix}^{-1} = \frac{1}{\left[n\sum_{i=1}^{n}x_i^2 - \left(\sum_{i=1}^{n}x_i\right)^2\right]}\begin{bmatrix} \sum_{i=1}^{n}x_i^2 & -\sum_{i=1}^{n}x_i \\ -\sum_{i=1}^{n}x_i & n \end{bmatrix}$

So using $\begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{n}x_i \\ \sum_{i=1}^{n}x_i & \sum_{i=1}^{n}x_i^2 \end{bmatrix}^{-1}\begin{bmatrix} \sum_{i=1}^{n}y_i \\ \sum_{i=1}^{n}(x_i y_i) \end{bmatrix} = \frac{1}{\left[n\sum_{i=1}^{n}x_i^2 - \left(\sum_{i=1}^{n}x_i\right)^2\right]}\begin{bmatrix} \sum_{i=1}^{n}x_i^2 & -\sum_{i=1}^{n}x_i \\ -\sum_{i=1}^{n}x_i & n \end{bmatrix}\begin{bmatrix} \sum_{i=1}^{n}y_i \\ \sum_{i=1}^{n}(x_i y_i) \end{bmatrix}$

$$a = \frac{-\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right) + n\sum_{i=1}^{n}(x_i y_i)}{\left[n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right]}$$

Thus

$$b = \frac{\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left[\sum_{i=1}^{n}(x_i y_i)\right]}{\left[n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right]}$$

Example: If the data is

i	1	2	3	4	5	6
x	0	0.5	1	1.5	2	2.5
y	0	1.5	3	4.5	6	7.5

$a = \frac{78.75}{26.25} = 3, \quad b = 0/26.25 = 0 \Rightarrow f(x) = 3x + 0 = 3x$ is the curve fitted by least squares.

To find the error

$$Error = \sum_{i=1}^{6}\left[y_i - (ax_i + b)\right]^2 = \sum_{i=1}^{6}\left[y_i - (3x_i)\right]^2$$

$$= (0 - 3\times 0)^2 + (1.5 - 3\times 0.5)^2 + (3 - 3\times 1)^2 + (4.5 - 3\times 1.5)^2 + (6 - 3\times 2)^2 + (7.5 - 3\times 2.5)^2$$

$$= 0 + 0 + 0 + 0 + 0 + 0 = 0$$

This curve fits the data exactly, i.e., the error is zero.
Usually this is not the case. More commonly, we have noisy data that does not exactly fit a straight line.

HW-4: Given the data as

x = [0 0.5 1 1.5 2 2.5], y = [-0.4326 -0.1656 3.1253 4.7877 4.8535 8.6909]

Find the curve fitted by a straight line by using least squares. Find the error.

Curve fitting by higher order polynomials

We worked on the linear curve fit by choosing a straight line $f(x) = ax + b$
This is just one kind of function. There are an infinite number function forms we can choose.

Polynomial Curve Fitting

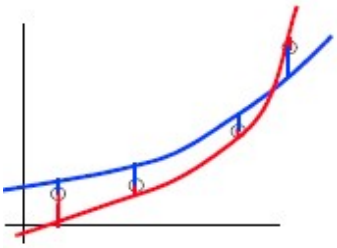
Polynomial of order j

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \ldots + a_j x^j = a_0 + \sum_{k=1}^{j} a_k x^k$$

Question: How can we find the coefficients that best fits the curve to the data?

We can say that the curve that gives minimum error between the data and the fitted curve is the best fit.

Let's have two curve fits for the given data set



To find the error for these two different polynomials,

First for the first polynomial, we add the vertical lengths between the data values and the values of the first polynomial fit.

Then for the second polynomial, we add the vertical lengths between the data values and the values of the second polynomial fit.
Finally choose the curve with minimum total error as the best fit.

Polynomial Curve Fitting by Least Squares Approach

$$Error = \sum_{i=1}^{\#\text{ data points}} \left[y_i - f(x_i) \right]^2$$

data points = n. Substituting $f(x)$, we have

$$Error = \sum_{i=1}^{n} \left[y_i - \left(a_0 + \sum_{k=1}^{j} a_k x^k \right) \right]^2$$

To find the best fit, find the coefficients of the polynomial a_0 and a_k for $k = 1, 2, \ldots j$ so that "*Error*" is minimized.

Thus, taking the derivatives and equating to zero, we have

$$\frac{\partial (Error)}{\partial a_0} = -2 \sum_{i=1}^{n} \left[y_i - \left(a_0 + \sum_{k=1}^{j} a_k x^k \right) \right] = 0$$

$$\frac{\partial (Error)}{\partial a_1} = -2 \sum_{i=1}^{n} \left[y_i - \left(a_0 + \sum_{k=1}^{j} a_k x^k \right) \right] x = 0$$

$$\frac{\partial(Error)}{\partial a_2} = -2\sum_{i=1}^{n}\left[y_i - \left(a_0 + \sum_{k=1}^{j} a_k x^k \right) \right] x^2 = 0$$

.
.

$$\frac{\partial(Error)}{\partial a_j} = -2\sum_{i=1}^{n}\left[y_i - \left(a_0 + \sum_{k=1}^{j} a_k x^k \right) \right] x^j = 0$$

Rewriting in matrix form

$$\begin{bmatrix} n & \sum x_i & \sum x_i^2 & \cdots & \sum x_i^j \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \cdots & \sum x_i^{j+1} \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 & \cdots & \sum x_i^{j+2} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum x_i^j & \sum x_i^{j+1} & \sum x_i^{j+2} & \cdots & \sum x_i^{j+j} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_j \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum (x_i y_i) \\ \sum \left(x_i^2 y_i \right) \\ \vdots \\ \sum \left(x_i^j y_i \right) \end{bmatrix}$$

Solve for the coefficients of the polynomial a_0 and a_k for $k = 1, 2, \ldots j$ for minimum "*Error*".

5. Series Representation of Functions (Fourier Series)

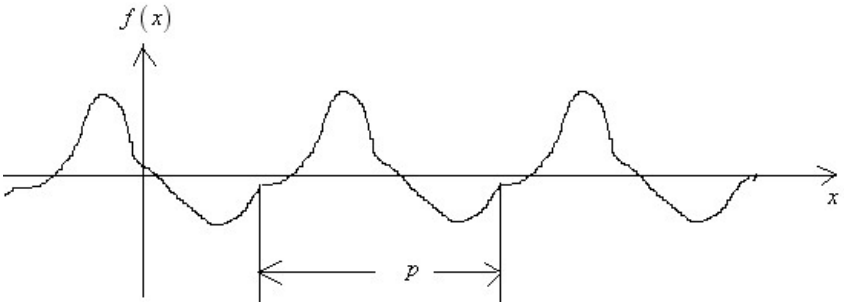
Periodic Functions, Trigonometric Series

A function $f(x)$ is called periodic if it is defined for all real x and if there is some positive number p such that

$$f(x+p) = f(x)$$

p is called the period of $f(x)$.

Graph of a periodic function is drawn by periodically repeating in any interval of length p.



Periodic function

Examples of periodic functions:

1. Sine and cosine functions
2. $f = c = \text{constant}$. It satisfies $f(x+p) = f(x)$ for every positive p

Examples of non-periodic functions: $x,\ x^2, e^x,\ \ln x$

From $f(x+p) = f(x) \ \Rightarrow f(x+2p) = f\left[(x+p)+p\right] = f(x+p) = f(x)$ etc., thus for any integer n

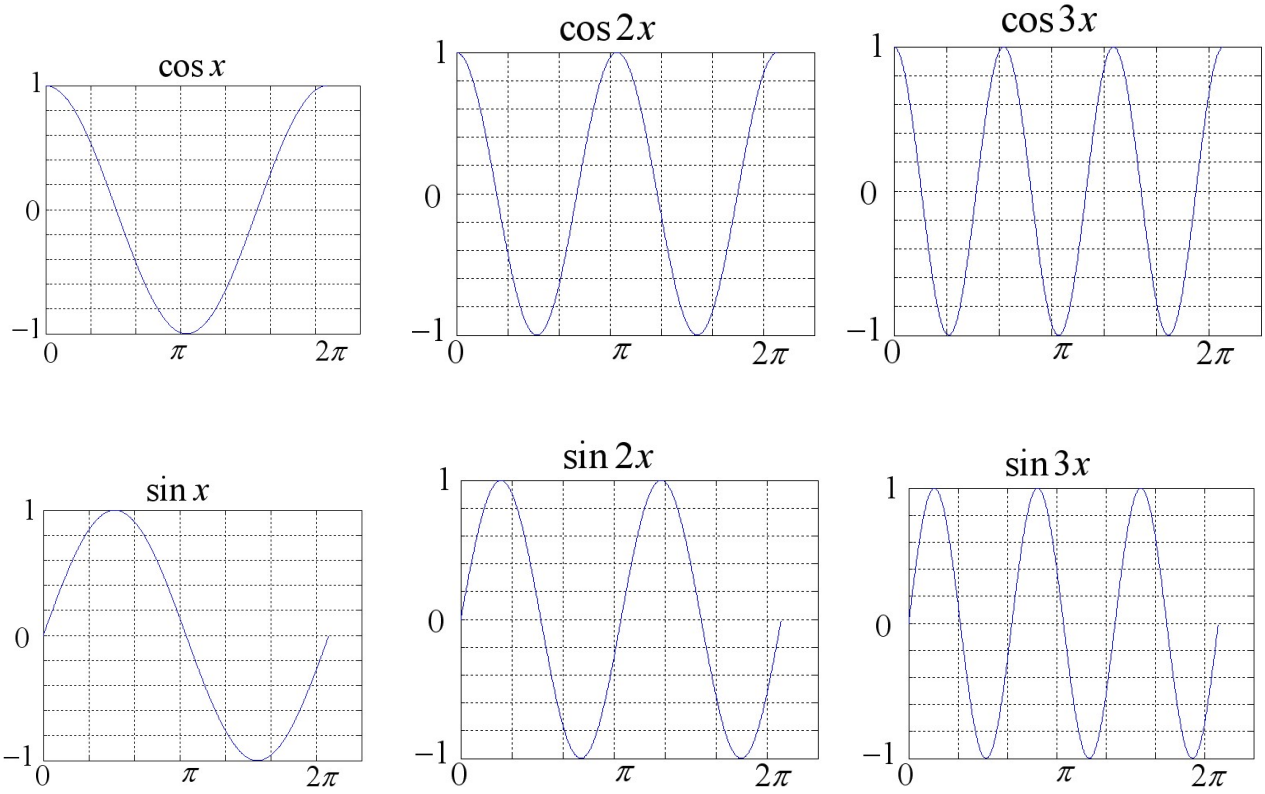
$f(x+np) = f(x)$ for all x.

So $2p,\ 3p,\ 4p$ are also periods of $f(x)$.

Furthermore, if $f(x)$ and $g(x)$ have period p, then $h(x) = af(x) + bg(x)$ (with $a,\ b$ constants) also has the period p.

Representation of various functions of period $p = 2\pi$ in terms of simple functions

1, $\cos x,\ \sin x,\ \cos 2x,\ \sin 2x,\ \ldots\ldots,\ \cos nx,\ \sin nx,\ \ldots\ldots$ which have the period 2π.



The series to be formed in this manner will be

$a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \ \ldots\ldots$

where $a_0,\ a_1,\ a_2,\ .....,\ b_1,\ b_2,\ .....$ are real constants.

This is called trigonometric series, a_n and b_n are the coefficients of the series.

This series can be written as

$$a_0 + \sum_{n=1}^{\infty} \left(a_n \cos nx + b_n \sin nx \right)$$

Each term of the above series has period 2π. So, if the above series converges, its sum will be a function of period 2π.

Fourier Series

Euler formulas for the Fourier Coefficients

Let $f(x)$ is a periodic function with period 2π which can be represented by the trigonometric series

$$f(x) = a_0 + \sum_{n=1}^{\infty} \left(a_n \cos nx + b_n \sin nx \right)$$

Here we assume that this series converges and $f(x)$ is its sum.

Problem: Given such an $f(x)$. Determine the coefficients a_n and b_n of the corresponding series.

To determine a_0, integrate both sides of the series from $-\pi$ to π

$$\int_{-\pi}^{\pi} f(x)\, dx = \int_{-\pi}^{\pi} \left[a_0 + \sum_{n=1}^{\infty} \left(a_n \cos nx + b_n \sin nx \right) \right] dx \qquad \Rightarrow$$

$$\int_{-\pi}^{\pi} f(x)\, dx = a_0 \int_{-\pi}^{\pi} dx + \sum_{n=1}^{\infty} \left(a_n \int_{-\pi}^{\pi} \cos nx\ dx + b_n \int_{-\pi}^{\pi} \sin nx\ dx \right) \qquad \Rightarrow$$

$$\int_{-\pi}^{\pi} f(x)\, dx = 2\pi a_0 + 0 + 0 \qquad \Rightarrow \qquad a_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)\, dx$$

To determine a_n, multiply the series by $\cos mx$ (m being any fixed positive integer), and integrate both sides of the series from $-\pi$ to π

$$\int_{-\pi}^{\pi} f(x)\ \cos mx\ dx = \int_{-\pi}^{\pi}\left[a_0 + \sum_{n=1}^{\infty}\left(a_n \cos nx + b_n \sin nx\right)\right]\ \cos mx\ dx \qquad \Rightarrow$$

$$\int_{-\pi}^{\pi} f(x)\ \cos mx\ dx = a_0 \int_{-\pi}^{\pi} \cos mx\ dx + \sum_{n=1}^{\infty}\left(a_n \int_{-\pi}^{\pi} \cos nx \cos mx\ dx + b_n \int_{-\pi}^{\pi} \sin nx \cos mx\ dx\right) \quad (\textbf{\textit{Eq.1}})$$

Using

$$\int_{-\pi}^{\pi} \cos nx \cos mx\ dx = \frac{1}{2}\int_{-\pi}^{\pi} \cos(n+m)x\ dx + \frac{1}{2}\int_{-\pi}^{\pi} \cos(n-m)x\ dx = 0 + \begin{cases}\pi \text{ when } n=m \\ 0, \text{ otherwise}\end{cases} = \begin{cases}\pi \text{ when } n=m \\ 0, \text{ otherwise}\end{cases}$$

and

$$\int_{-\pi}^{\pi} \sin nx \cos mx\ dx = \frac{1}{2}\int_{-\pi}^{\pi} \sin(n+m)x\ dx + \frac{1}{2}\int_{-\pi}^{\pi} \sin(n-m)x\ dx = 0 + 0 = 0 \quad , \qquad (\textbf{\textit{Eq.1}})$$

becomes

$$\int_{-\pi}^{\pi} f(x)\ \cos mx\ dx = 0 + 0 + a_n . \begin{cases}\pi \text{ when } n=m \\ 0, \text{ otherwise}\end{cases} + 0 + 0 = \pi a_m$$

Replacing m with n

$$\int_{-\pi}^{\pi} f(x)\ \cos nx\ dx = \pi a_n \qquad \Rightarrow \qquad a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\ \cos nx\ dx$$

In a similar manner, b_n can be found as

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x) \sin nx\ dx$$

In summary

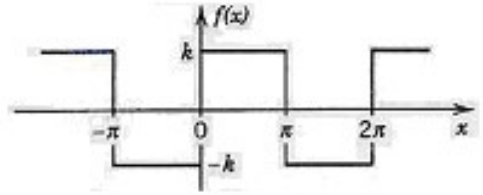
$f(x) = a_0 + \sum_{n=1}^{\infty}\left(a_n \cos nx + b_n \sin nx\right)$ is called the Fourier Series with Fourier Coefficients of

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\,dx, \quad a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\ \cos nx\ dx, \quad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x) \sin nx\ dx$$

- Example

Find the Fourier coefficients of the periodic square wave

$$f(x)=\begin{cases}-k & \text{if } -\pi<x<0\\ k & \text{if } 0<x<\pi\end{cases} \quad \text{and} \quad f(x+2\pi)=f(x)$$



Solution: $a_0=\dfrac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\,dx=0,$

$$a_n=\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,\cos nx\,dx=\frac{1}{\pi}\left[\int_{-\pi}^{0}-k\,\cos nx\,dx+\int_{0}^{\pi}k\,\cos nx\,dx\right]$$

$$=\frac{1}{\pi}\left[-k\frac{\sin nx}{n}\,\Big|_{-\pi}^{0}+k\frac{\sin nx}{n}\,\Big|_{0}^{\pi}\right]=0$$

$$b_n=\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,\sin nx\,dx=\frac{1}{\pi}\left[\int_{-\pi}^{0}-k\,\sin nx\,dx+\int_{0}^{\pi}k\,\sin nx\,dx\right]$$

$$=\frac{1}{\pi}\left[k\frac{\cos nx}{n}\,\Big|_{-\pi}^{0}-k\frac{\cos nx}{n}\,\Big|_{0}^{\pi}\right]$$

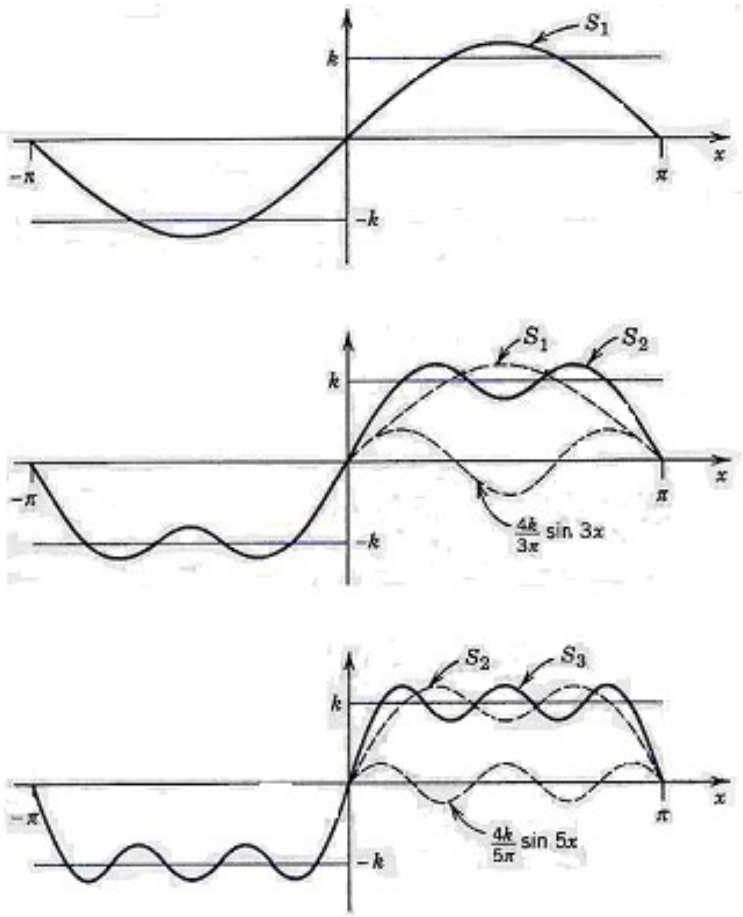
$$=\frac{k}{n\pi}\left[\cos 0-\cos(-n\pi)-\cos n\pi+\cos 0\right]=\frac{2k}{n\pi}(1-\cos n\pi)=\begin{cases}\frac{4k}{n\pi} & \text{for odd } n\\ 0 & \text{for even } n\end{cases}$$

Thus the Fourier series can be represented by

$$f(x)=\frac{4k}{\pi}\left(\sin x+\frac{1}{3}\sin 3x+\frac{1}{5}\sin 5x+.....\right)$$

Partial sums are: $S_1=\dfrac{4k}{\pi}\sin x,\quad S_2=\dfrac{4k}{\pi}\left(\sin x+\dfrac{1}{3}\sin 3x\right),\quad S_3=\dfrac{4k}{\pi}\left(\sin x+\dfrac{1}{3}\sin 3x+\dfrac{1}{5}\sin 5x\right)$

$S_1,\ S_2,$ and S_3 are plotted below.



The first three partial sums of the corresponding Fourier series

It is observed that as more terms are included in the sum Fourier series approaches the original periodic square wave $f(x)$.

Orthogonality

We have a system formed by functions of

$1,\ \cos x,\ \sin x,\ \cos 2x,\ \sin 2x,\ .....,\ \cos nx,\ \sin nx,\ .....$

This system is said to be orthogonal on the interval $-\pi \le x \le \pi$ if the integral of the product of any two different of these functions over that interval is zero.i.e., for any integers m and $n \neq m$ we have

$$\int_{-\pi}^{\pi} \cos mx\ \cos nx\ dx = 0 \quad \text{for} \quad m \neq n \quad \text{and} \quad \int_{-\pi}^{\pi} \sin mx\ \sin nx\ dx = 0 \quad \text{for} \quad m \neq n$$

and for any integer m and n (including $m = n$) we have

$$\int_{-\pi}^{\pi} \cos mx\ \sin nx\ dx = 0$$

Functions of Any Period $p = 2L$

Up to here period was taken as 2π. However, in many applications, periodic functions have periods other than 2π.

We make a change of scale by setting $\nu = \dfrac{\pi x}{L} \quad \Rightarrow \quad x = \dfrac{L\nu}{\pi}$. Then $x = \pm L$ corresponds to $\nu = \pm\pi$. Thus $f(x) = g(\nu)$ has period 2π.

Using ν instead of x, this new function $g(\nu)$, having period 2π, has the Fourier series

$g(\nu) = a_0 + \sum_{n=1}^{\infty} (a_n \cos n\nu + b_n \sin n\nu)$ with coefficients

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} g(\nu)\, d\nu, \quad a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} g(\nu)\cos n\nu\, d\nu, \quad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} g(\nu)\sin n\nu\, d\nu$$

Transforming the above Fourier series by applying $\nu = \dfrac{\pi x}{L}$, it is found that if a function $f(x)$ has period $p = 2L$, then it is Fourier series is expressed as.

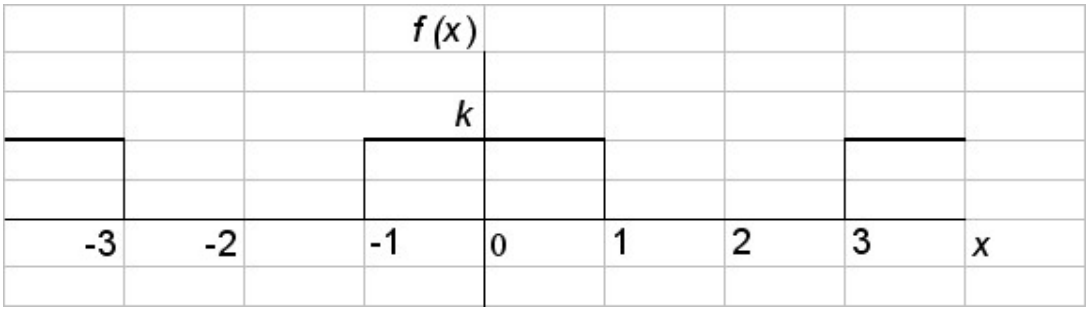
$$f(x) = a_0 + \sum_{n=1}^{\infty}\left(a_n \cos\frac{n\pi}{L}x + b_n \sin\frac{n\pi}{L}x\right)$$ with Fourier Coefficients of

$$a_0 = \frac{1}{2L}\int_{-L}^{L} f(x)\, dx, \quad a_n = \frac{1}{L}\int_{-L}^{L} f(x)\cos\frac{n\pi x}{L}\, dx, \quad b_n = \frac{1}{L}\int_{-L}^{L} f(x)\sin\frac{n\pi x}{L}\, dx$$

- Example

Find the Fourier series of periodic square wave given by

$$f(x) = \begin{cases} 0 & \text{if} \quad -2 < x < -1 \\ k & \text{if} \quad -1 < x < 1 \\ 0 & \text{if} \quad 1 < x < 2 \end{cases}$$



Solution

$p = 2L = 4 \quad \Rightarrow \quad L = 2,$

$$a_0 = \frac{1}{2L}\int_{-L}^{L} f(x)\,dx = \frac{1}{4}\int_{-2}^{2} f(x)\,dx = \frac{1}{4}\int_{-1}^{1} k\,dx = \frac{k}{2},$$

$$a_n = \frac{1}{L}\int_{-L}^{L} f(x)\,\cos\frac{n\pi x}{L}\,dx = \frac{1}{2}\int_{-2}^{2} f(x)\,\cos\frac{n\pi x}{2}\,dx = \frac{1}{2}\int_{-1}^{1} k\,\cos\frac{n\pi x}{2}\,dx = \frac{2k}{n\pi}\sin\frac{n\pi}{2},$$

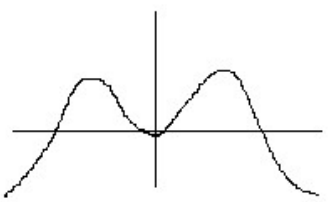
$$b_n = \frac{1}{L}\int_{-L}^{L} f(x)\,\sin\frac{n\pi x}{L}\,dx = \frac{1}{2}\int_{-2}^{2} f(x)\,\sin\frac{n\pi x}{2}\,dx = \frac{1}{2}\int_{-1}^{1} k\,\sin\frac{n\pi x}{2}\,dx = 0$$

The Fourier series becomes

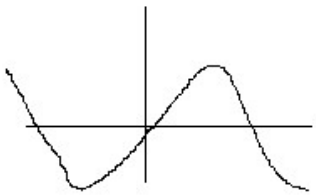
$$f(x) = a_0 + \sum_{n=1}^{\infty}\left(a_n\cos\frac{n\pi}{L}x + b_n\sin\frac{n\pi}{L}x\right) = \frac{k}{2} + \frac{2k}{\pi}\left[\cos\left(\frac{\pi}{2}x\right) - \frac{1}{3}\cos\left(\frac{3\pi}{2}x\right) + \frac{1}{5}\cos\left(\frac{5\pi}{2}x\right) - .. + ..\right]$$

Even and Odd Functions

A function $y = g(x)$ is even if $g(-x) = g(x)$ for all x, e.g., $\cos nx$.
A function $y = h(x)$ is odd if $h(-x) = -h(x)$ for all x, e.g., $\sin nx$.



Even function



Odd function

If $g(x)$ is an even function, then $\int_{-L}^{L} g(x)\,dx = 2\int_{0}^{L} g(x)\,dx$

If $h(x)$ is an odd function, then $\int_{-L}^{L} h(x)\,dx = 0$

If $g(x)$ is even and $h(x)$ is odd, then the product $Q(x) = g(x)h(x)$ is odd since

$Q(-x) = g(-x)h(-x) = g(x)\left[-h(x)\right] = -Q(x) \quad \Rightarrow \quad Q(-x) = -Q(x)$, i.e., $Q(x)$ is odd

Using the above result, if $f(x)$ is even, then

the integrand of $b_n = \frac{1}{L}\int_{-L}^{L} f(x)\,\sin\frac{n\pi x}{L}\,dx$, i.e., $f(x)\sin\frac{n\pi x}{L}$ is odd $\quad \Rightarrow \quad b_n = 0$

Similarly, if $f(x)$ is odd, then

the integrand of $a_n = \frac{1}{L}\int_{-L}^{L} f(x)\,\cos\frac{n\pi x}{L}\,dx$, i.e., $f(x)\cos\frac{n\pi x}{L}$ is odd $\quad \Rightarrow \quad a_n = 0$

Theorem: Fourier series of even and odd functions

Fourier series of an even function of period $2L$ is a "Fourier cosine series"

$$f(x) = a_0 + \sum_{n=1}^{\infty} a_n \cos\frac{n\pi}{L}x \text{ with coefficients } a_0 = \frac{1}{L}\int_0^L f(x)\,dx, \quad a_n = \frac{2}{L}\int_0^L f(x)\,\cos\frac{n\pi x}{L}\,dx,$$

Similarly, Fourier series of an odd function of period $2L$ is a "Fourier sine series"

$$f(x) = \sum_{n=1}^{\infty} b_n \sin\frac{n\pi}{L}x \quad \text{with coefficients} \quad b_n = \frac{2}{L}\int_0^L f(x)\,\sin\frac{n\pi x}{L}\,dx$$

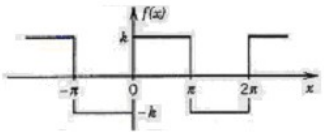
Theorem:

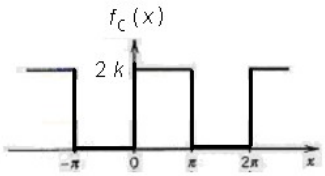
- Fourier coefficients of the sum function $f_s(x) = f_1(x) + f_2(x)$ are the sums of the corresponding Fourier coefficients of $f_1(x)$ and $f_2(x)$.

- Fourier coefficients of the function $f_c(x) = c\,f(x)$ are c times the corresponding Fourier coefficients of $f(x)$.

- Example: Find the Fourier series of

$$f_s(x) = k + f(x) \text{ where } f(x) = \begin{cases} -k & \text{if } -\pi < x < 0 \\ k & \text{if } 0 < x < \pi \end{cases} \text{ and } f(x+2\pi) = f(x)$$





Using the result for $f(x)$ found in a previous example and the above theorem for the sum

$$f_c(x) = k + \frac{4k}{\pi}\left(\sin x + \frac{1}{3}\sin 3x + \frac{1}{5}\sin 5x + ...\right)$$

6. Solution of Differential Equations; Euler's Method

Many differential equations do not have algebraic solutions so we solve them numerically.

General Initial Value Problem

Differential equation to be solved is:

$\frac{dy}{dx} = f(x,y)$ and initial value $y(a)$ is known.

e.g. $\frac{dy}{dx} = \frac{10x^2 - y}{2}, \quad y(0) = 0$

Euler's Method assumes that the solution is written in the form of Taylor's Series, i.e.,

$$y(x+h) \approx y(x) + hy'(x) + \frac{h^2 y''(x)}{2!} + \frac{h^3 y'''(x)}{3!} + \frac{h^4 y^{iv}(x)}{4!} + \ldots$$

In Euler's Method, only the first two terms are taken, i.e.,

$$y(x+h) \approx y(x) + hy'(x) \approx y(x) + h\frac{dy}{dx} \approx y(x) + hf(x,y)$$

Lets assume that we have a known value of y as y_0 when $x=x_0$, i.e., the initial value is (x_0, y_0).

Let's call y_1 as the value of y, one h step to the right of the current value.

$$y_1 \approx y_0 + hf(x_0, y_0)$$

where y_0 is the current value, h is the interval between steps, y_1 is the next estimated solution value, $f(x_0, y_0)$ is the value of the derivative at the starting point (x_0, y_0).

Next value: Making use of y_1, we find the next value y_2 as

$$y_2 \approx y_1 + hf(x_1, y_1)$$

where y_1 is the current value, h is the interval between steps, y_2 is the next estimated solution value, $x_1 = x_0 + h$, $f(x_1, y_1)$ is the value of the derivative at the current point (x_1, y_1).

This process is continued for as many steps as required.

What is the meaning of this solution? $y_2 \approx y_1 + hf(x_1, y_1)$

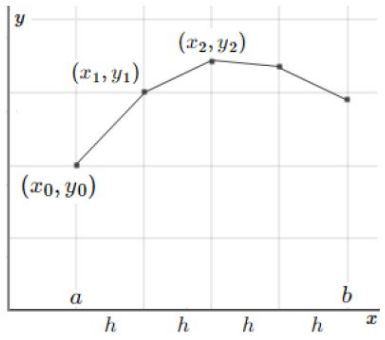
The RHS means:

Start at the known y value

Then move one step h units to the right in the direction of the slope at that point, which is $\frac{dy}{dx} = f(x,y)$

A good approximation to the curve's y point will be arrived at that new point.

This is done for each sub-point, h apart, from the starting value $x=a$ to the end value $x=b$.



Example for Euler's Method

Initial value problem is $\frac{dy}{dx} = \frac{y \ln y}{x}$ and $y(2) = e$

Solution: Step 1: Start at the point $(x_0, y_0) = (2, e)$ and use step size h=0.1 and use 10 steps, i.e., the solution of the differential equation will be approximated from x=2 to x=3.

Calculating the value of the derivative $\frac{dy}{dx} = f(x, y)$ at the initial point $x_0 = 2, y_0 = e$

$\frac{dy}{dx} = \frac{y \ln y}{x} \Rightarrow f(2, e) = \frac{e \ln e}{2} = \frac{e}{2} \approx 1.3591409$, i.e., slope of the line from x=2 to x=2.1 is

approximately 1.3591409

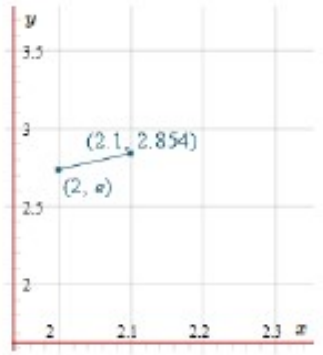
Step 2: Next point is $x+h$=2+0.1=2.1

Substituting in the Euler Method's formula $y(x+h) \approx y(x) + hf(x, y)$

$y_1 = y(x_0 + h) \approx y(x_0) + hf(x_0, y_0) \Rightarrow y_1 = y(x_0 + h) \approx y_0 + hf(x_0, y_0)$

$y_1 = y(2 + 0.1) \approx e + 0.1\frac{e}{2} = 2.8540959$

i.e., the approximate value of the solution at x=2.1 is 2.8540959 .



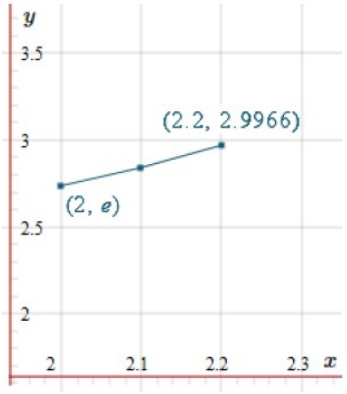
Now, we find the new slope at $x_1 = 2.1, y_1 = 2.8540959$

$$\frac{dy}{dx} = \frac{y \ln y}{x} = f\left(2.1, 2.8540959\right) = \frac{2.8540959 \ln\left(2.8540959\right)}{2.1} = 1.4254536$$

i.e., slope of the line from x=2.1 to x=2.2 is approximately 1.4254536

Step 3: We find the solution value when x=2.2.

$$y_2 = y\left(x_1 + h\right) = y\left(2.2\right) \approx y\left(x_1\right) + hf\left(x_1, y_1\right) \Rightarrow y_2 = y\left(x_1 + h\right) \approx y_1 + hf\left(x_1, y_1\right)$$
$$= 2.8540959 + 0.1 \times f\left(2.1, 2.8540959\right)$$
$$= 2.8540959 + 0.1\frac{2.8540959 \ln\left(2.8540959\right)}{2.1} = 2.8540959 + 0.1 \times 1.4254536 = 2.99664126$$



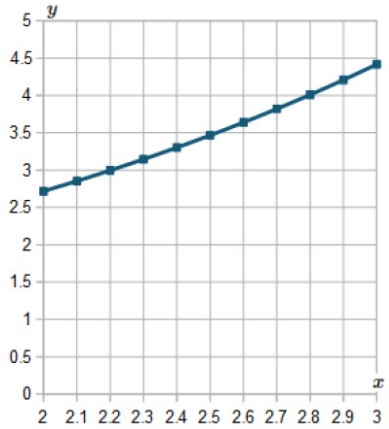
The slope at this new point (x_2=2.2, y_2=2.99664126) is

$$\frac{dy}{dx} = \frac{y \ln y}{x} = f\left(2.2, 2.99664126\right) = \frac{2.99664126 \ln\left(2.99664126\right)}{2.2} = 1.49490457$$

i.e., slope of the line from x=2.2 to x=2.3 is approximately 1.49490457

Continuing with Steps 4, 5, ... , 10, we find

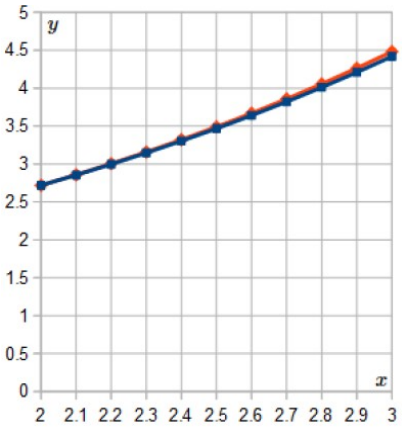
x	y	$\frac{dy}{dx}$
2.0	$e = 2.7182818285$	$(e \ln e)/2 = 1.3591409142$
2.1	$e+0.1(e/2) = 2.8541959199$	$(2.8541959199 \ln 2.8541959199)/2 = 1.4254536226$
2.2	2.9967412821	1.4949999323
2.3	3.1462412754	1.5679341197
2.4	3.3030346873	1.6444180873
2.5	3.4674764961	1.7246216904
2.6	3.6399386651	1.8087230858
2.7	3.8208109737	1.8969091045
2.8	4.0105018841	1.9893756448
2.9	4.2094394486	2.08632809
3.0	4.4180722576	



To check how close is the Euler's Method solution to the exact solution?

The same initial value problem $\frac{dy}{dx} = \frac{y \ln y}{x}$ and $y(2) = e$ can be solved algebraically, e.g. by Separation of Variables and the exact solution is $y = e^{0.5x}$.

To compare the Euler's Method solution with the exact solution



At x=3, the exact solution is y=4.4816890703 and Euler's Method solution is y=4.4180722576

i.e., Error=(4.4816890703-4.4180722576)/ 4.4816890703x100=1.42 %

Thus, the two solutions are very close.

HW-6

Solve $\frac{dy}{dx}=\sin(x+y)-e^{x}$ and $y(0)=4$ from x=0 to x=0.2 with steps h=0.1

7. The Runge Kutta 4th Degree Method

Euler's Method gives reasonably accurate result but it is not very efficient. The Runge-Kutta 4th Degree Method produces a better result in fewer steps.

$$y(x+h)=y(x)+\frac{1}{6}(F_1+2F_2+2F_3+F_4)$$

where

$$F_1=hf(x,y),\quad F_2=hf\left(x+\frac{h}{2},y+\frac{F_1}{2}\right),\quad F_3=hf\left(x+\frac{h}{2},y+\frac{F_2}{2}\right),\quad F_4=hf(x+h,y+F_3)$$

Example

Using Runge-Kutta Method of Order 4, solve $\frac{dy}{dx}=\frac{5x^2-y}{e^{x+y}}$ with $y(0)$=1 by using step size of h=0.1 for $0\le x\le 1$.

Step 1: Start with x=0 and y=1 and then find the F values.

$$F_1=hf(x,y)=0.1\frac{5(0)^2-1}{e^{0+1}}=-0.03678794411$$

$$F_2=hf\left(x+\frac{h}{2},y+\frac{F_1}{2}\right)=0.1f\left(0+\frac{0.1}{2},1+\frac{-0.03678794411}{2}\right)=0.1\left[\frac{5(0.05)^2-0.98160602794}{e^{0.05+0.98160602794}}\right]$$

$$=-0.03454223937$$

$$F_3=hf\left(x+\frac{h}{2},y+\frac{F_2}{2}\right)=0.1f\left(0+\frac{0.1}{2},1+\frac{-0.03454223937}{2}\right)=0.1\left[\frac{5(0.05)^2-0.98272888031}{e^{0.05+0.98272888031}}\right]$$

$$=-0.03454345267$$

$$F_4=hf(x+h,y+F_3)=0.1f(0+0.1,1-0.03454345267)=0.1\left[\frac{5(0.1)^2-0.96545654732}{e^{0.1+0.96545654732}}\right]$$

$$=-0.03154393258$$

Step 2: Substitute F_1, F_2, F_3, F_4 into the Runge-Kutta RK4 formula

$$y(x+h) = y(x) + \frac{1}{6}(F_1 + 2F_2 + 2F_3 + F_4)$$

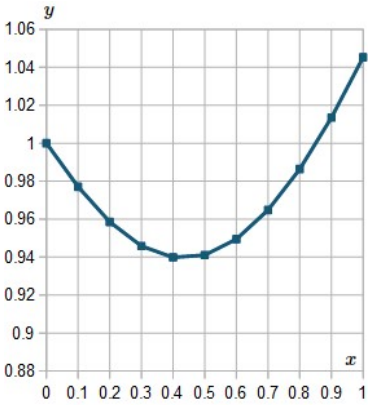
$$= y(0) + \frac{1}{6}(-0.03678794411 - 2\times 0.03454223937 - 2\times 0.03454345267 - 0.03154393258)$$

$$= 1 + \frac{1}{6}(-0.03678794411 - 2\times 0.03454223937 - 2\times 0.03454345267 - 0.03154393258)$$

$$= 0.9655827899$$

Using this new y value of 0.9655827899, new F_1, F_2, F_3, F_4 values are found at x=0.01+0.01=0.02 and substituted into the Runge-Kutta 4 formula. This process is continued and the following table is found

x	y	$F_1 = h\ dy/dx$	x+h/2	$y+F_1/2$	F_2	$y+F_2/2$	F_3	x+h	$y+F_3$	F_4
0	1	-0.0367879441	0.05	0.9816060279	-0.0345422394	0.9827288803	-0.0345434527	0.1	0.9654565473	-0.0315439326
0.1	0.9655827899	-0.0315443	0.15	0.9498106398	-0.0278769283	0.9516443257	-0.0278867954	0.2	0.9376959945	-0.023647342
0.2	0.937796275	-0.023648185	0.25	0.9259721824	-0.0189267761	0.9283328869	-0.0189548088	0.3	0.9188414662	-0.0138576597
0.3	0.9189181059	-0.0138588628	0.35	0.9119886745	-0.0084782396	0.9146789861	-0.0085314167	0.4	0.9103866892	-0.0029773028
0.4	0.9104421929	-0.0029786344	0.45	0.9089528756	0.0026604329	0.9117724093	0.002580704	0.5	0.9130228969	0.0082022376
0.5	0.913059839	0.0082010354	0.55	0.9171603567	0.013727301	0.9199234895	0.0136258867	0.6	0.9266857257	0.018973147
0.6	0.9267065986	0.0189722976	0.65	0.9361927474	0.0240794197	0.9387463085	0.0239658709	0.7	0.9506724696	0.0287752146
0.7	0.9506796142	0.0287748718	0.75	0.9650670501	0.0332448616	0.967302045	0.0331305132	0.8	0.9838101274	0.0372312889
0.8	0.9838057659	0.0372315245	0.85	1.0024215282	0.0409408747	1.0042762033	0.0408359751	0.9	1.024641741	0.0441484563
0.9	1.024628046	0.0441492608	0.95	1.0467026764	0.0470593807	1.0481577363	0.0469712279	1	1.0715992739	0.0494916177
1	1.0715783953									



Extending the result up to x=10

HW-7

Using Runge-Kutta Method of Order 4, solve $\frac{dy}{dx} = (x+y)\sin xy$ with $y(0)=5$ by using step size of $h=0.2$ for $0 \le x \le 2$.

8. The Method of Finite Differences (FD).

FD methods are used to solve differential equations numerically. The solution is obtained by approximating the differential equation with difference equation. Derivatives are approximated by the differences.

Derivation using Taylor's polynomial:

$$f(x_0+h) = f(x_0) + \frac{hf'(x_0)}{1!} + \frac{h^2 f''(x_0)}{2!} + \frac{h^3 f'''(x_0)}{3!} + ... + \frac{h^n f^{(n)}(x_0)}{n!} + R_n(x_0),$$

$R_n(x)$ is the remainder term which gives the difference between the original function and the Taylor polynomial of degree n.

Making an approximation with the first derivative

$$f(x_0+h) = f(x_0) + \frac{hf'(x_0)}{1!} + R_1(x_0) = f(x_0) + hf'(x_0) + R_1(x_0)$$

Dividing by h

$$\frac{f(x_0+h)}{h} = \frac{f(x_0)}{h} + f'(x_0) + \frac{R_1(x_0)}{h}$$

Solving for $f'(x_0)$

$$f'(x_0) = \frac{f(x_0+h) - f(x_0)}{h} - \frac{R_1(x_0)}{h}$$

Assuming that $R_1(x_0)$ is sufficiently small, the approximation of the first derivative becomes

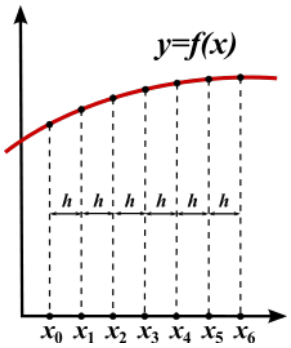
$$f'(x_0) \approx \frac{f(x_0+h) - f(x_0)}{h}$$

Error=Approximate solution – exact solution.

Two sources of error: Round-off error which is due to computer rounding of decimal quantities.

Truncation error which is the difference between the exact solution of the finite difference equation and the quantity with no round-off.

In order to use finite difference method for approximation to the solution, first the problem's domain is discretized so divide the domain into a uniform grid



Local truncation error $= f'(x_i) - f_i'$

$f'(x_i)$ is the exact value and f_i' is the numerical approximation.

To analyze the local truncation error, the remainder term of a Taylor polynomial is used.

Lagrange form of the remainder from the Taylor polynomial for $f(x_0 + h)$ is

$$R_1(x_0 + h) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(h)^{(n+1)} \quad \text{where} \quad x_0 < \xi < x_0 + h$$

From this, dominant term of the local truncation error is found.

e.g. using $f(x_i) = f(x_0 + ih) \Rightarrow$

$$\frac{f(x_0 + ih) - f(x_0)}{ih} = f'(x_0) + \frac{f''(\xi)}{2!}ih$$

Using the above expression, the dominant term of the local truncation error can be found.

LHS is the approximation from the finite difference method

RHS is the exact quantity plus a remainder where the remainder is the local truncation error.

Thus, it can be written as

$$\frac{f(x_0 + ih) - f(x_0)}{ih} = f'(x_0) + O(h)$$

where $O(h)$ is the order.

The local truncation error is proportional to the step sizes.

The quality and duration of simulated FDM solution depends on the discretization equation selection and the step sizes (time and space steps).

The data quality and simulation duration increase significantly with smaller step size

Example: The ordinary differential equation

$$u'(x) = 2u(x) + 1$$

Using the finite difference quotient in Euler method to solve this equation

$$\frac{u(x+h) - u(x)}{h} \approx u'(x) \Rightarrow \frac{u(x+h) - u(x)}{h} \approx 2u(x) + 1$$

$$u(x+h) - u(x) \approx 2hu(x) + h \Rightarrow u(x+h) \approx u(x) + h\left[2u(x) + 1\right]$$

which is the finite-difference equation, the solution of which will give an approximate solution to the differential equation.

Example

$$u(x+h) = 0.3u(x) + 1000, \quad u(x_0) = y_0 = 1000$$

which is a linear finite difference equation

$$u(x_0 + h) = y_1 = 0.3u(x_0) + 1000 = 0.3y_0 + 1000$$

$$y_2 = 0.3u(x_1) + 1000 = 0.3y_1 + 1000 = 0.3\left(0.3y_0 + 1000\right) + 1000$$

$$y_3 = 0.3u(x_2) + 1000 = 0.3y_2 + 1000 = 0.3\left[0.3\left(0.3y_0 + 1000\right) + 1000\right] + 1000$$
$$= 1000 + 0.3(1000) + \left(0.3\right)^2\left(1000\right) + \left(0.3\right)^3 y_0$$

In general,

$$y_n = 1000\left[1 + 0.3 + \left(0.3\right)^2 + \left(0.3\right)^3 ... + \left(0.3\right)^{n-1}\right] + \left(0.3\right)^n y_0$$

HW-8

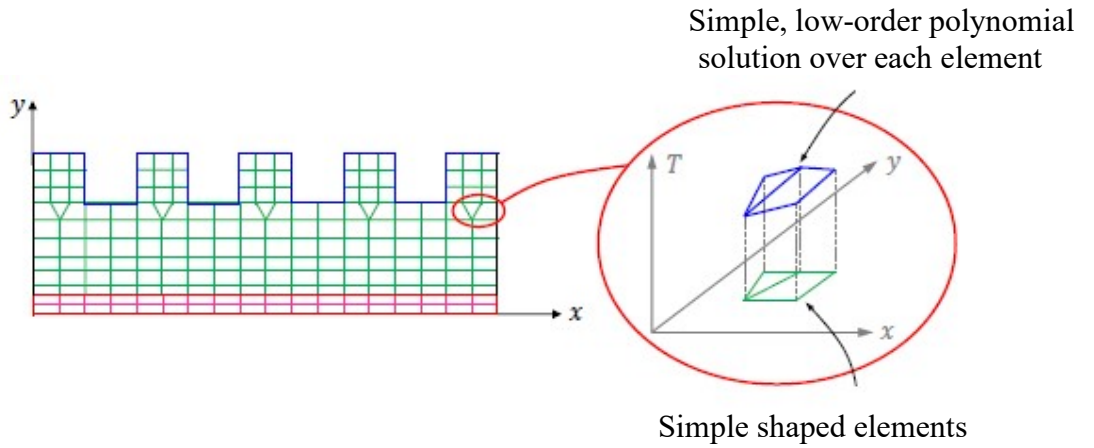
Find y_n if $u(x+h) = 2u(x) + 2, \quad u(x_0) = y_0 = 2$

9. The Finite Element Method (FEM) (Ref: Chapter 2 of the course notes are prepared by

Dr. Cüneyt Sert, METU, http://www.me.metu.edu.tr/people/cuneyt)

- Does not seek a global solution
- Divides the problem domain into elements of simple shapes
- Works with simple polynomial type approximate solutions over each element

- Weight function is selected

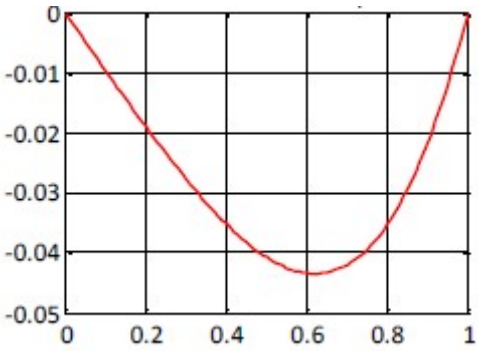


Example: Solve $-\frac{d^2u}{dx^2}-u=-x^2, \quad 0<x<1, \quad u(0)=0, \quad u(1)=0$ by using FEM

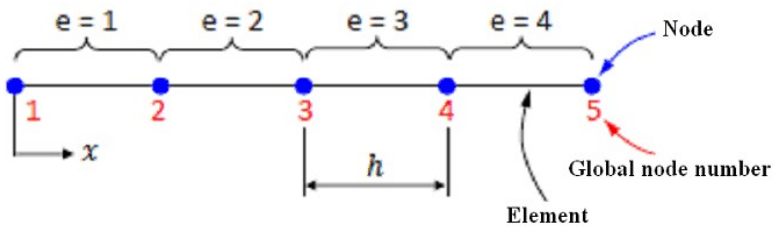
Exact solution is

$$u_{exact}=\frac{\sin(x)+2\sin(1-x)}{\sin(1)}+x^2-2$$

Exact solution

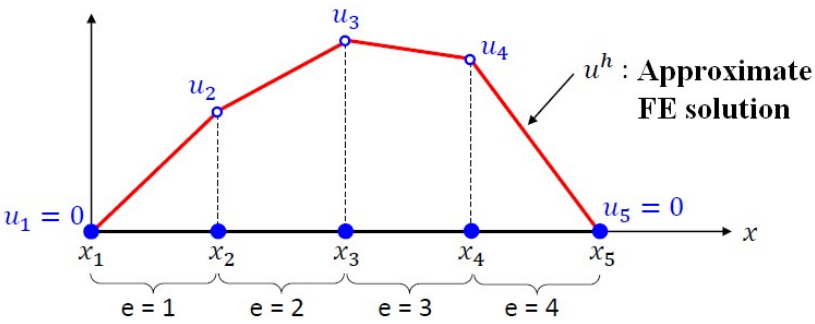


5 node ($NN = 5$) and 4 element ($NE = 4$) mesh (grid) will be used as shown below:



This is a mesh of linear elements (elements that are defined by 2 nodes).
This mesh is uniform, i.e. element length $h = 0.25$ is constant.

Using linear elements, a piecewise linear solution is obtained.



Solution is linear in every linear piecewise segment and continuous at element interfaces, however, 1st derivative of the solution is not continuous.

u_j's are the nodal unknown values which will be calculated.

u_1 and u_5 are the element boundary conditions which are known.

The solution is $u^h = \sum_{j=1}^{NN} u_j \phi_j$

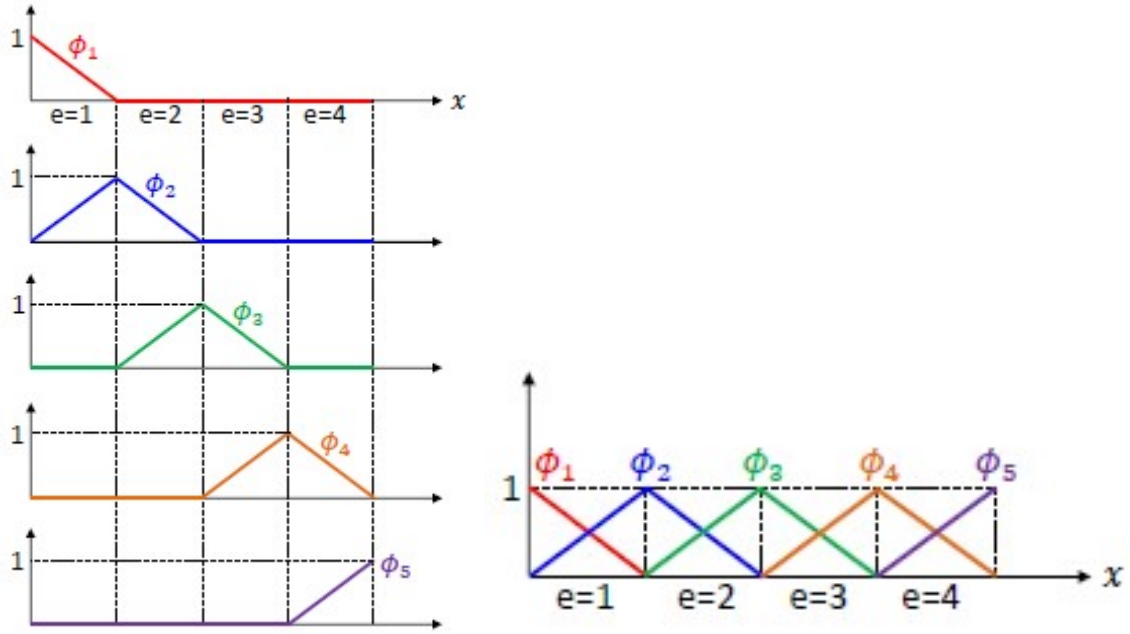
where $\phi_j(x)$ are the approximate solutions, NN is the number of nodes and u_j's are the nodal unknowns.

To have a piecewise linear u^h each $\phi_j(x)$ should be linear.

$u^h = \sum_{j=1}^{NN} u_j \phi_j$ should provide nodal unknown values at the nodes which is satisfied if the Kronecker-Delta property

$$\phi_j(x_i) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}, \quad i, j = 1, 2, ..., NN \quad \text{holds.}$$

Below approximation functions will be OK

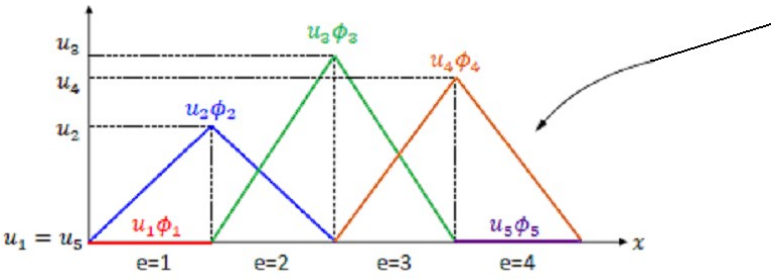


These are Lagrange type approximation functions that make sure that the solution is continuous across elements, but not its first derivative.

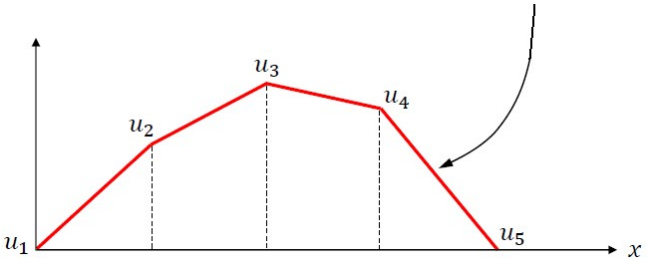
They fulfill the Kronecker-Delta property.

They are nonzero only over at most two elements.

Each $\phi_j(x)$ by u_j



$$u^h = \sum_{j=1}^{NN} u_j \phi_j$$



Using u instead of u^h

$$R = -\frac{d^2u}{dx^2} - u + x^2$$

$$\int_0^1 wR\ dx = 0 \rightarrow \int_0^1 w\left(-\frac{d^2u}{dx^2} - u + x^2 \right) dx = 0 \rightarrow \int_0^1 \left(-w\frac{d^2u}{dx^2} - wu + wx^2 \right) dx = 0$$

Applying integration by parts, i.e., $\int_a^b u\frac{dv}{dx} = uv\Big|_a^b - \int v\frac{du}{dx}$ to the first term $\int_0^1 -w\frac{d^2u}{dx^2}dx$

where $u = w$, $\frac{dv}{dx} = \frac{d^2u}{dx^2} \rightarrow \frac{du}{dx} = \frac{dw}{dx}$, $\mathrm{v} = \frac{du}{dx}$

The first term becomes $\int_0^1 -w\frac{d^2u}{dx^2}dx = -\left(w\frac{du}{dx}\Big|_0^1 - \int_0^1 \frac{dw}{dx}\frac{du}{dx}dx \right)$

Thus the equation becomes

$$\int_0^1 \left(-w\frac{d^2u}{dx^2} - wu + wx^2 \right) dx = \int_0^1 \left(\frac{dw}{dx}\frac{du}{dx} - wu + wx^2 \right) dx - w\frac{du}{dx}\Big|_0^1 = 0$$

Writing this equation *NN* times with different *w*'s where $w_i = \phi_i, \quad i = 1,2,...NN$

1^{st} Eq. with $w = \phi_1 \rightarrow \int_0^1 \left(\frac{d\phi_1}{dx}\frac{du}{dx} - \phi_1 u + \phi_1 x^2 \right) dx - \phi_1\frac{du}{dx}\Big|_1 + \phi_1\frac{du}{dx}\Big|_0 = 0$

2^{nd} Eq. with $w = \phi_2 \rightarrow \int_0^1 \left(\frac{d\phi_2}{dx}\frac{du}{dx} - \phi_2 u + \phi_2 x^2 \right) dx - \phi_2\frac{du}{dx}\Big|_1 + \phi_2\frac{du}{dx}\Big|_0 = 0$

3^{rd} Eq. with $w = \phi_3 \rightarrow \int_0^1 \left(\frac{d\phi_3}{dx}\frac{du}{dx} - \phi_3 u + \phi_3 x^2 \right) dx - \phi_3\frac{du}{dx}\Big|_1 + \phi_3\frac{du}{dx}\Big|_0 = 0$

4^{th} Eq. with $w = \phi_4 \rightarrow \int_0^1 \left(\frac{d\phi_4}{dx}\frac{du}{dx} - \phi_4 u + \phi_4 x^2 \right) dx - \phi_4\frac{du}{dx}\Big|_1 + \phi_4\frac{du}{dx}\Big|_0 = 0$

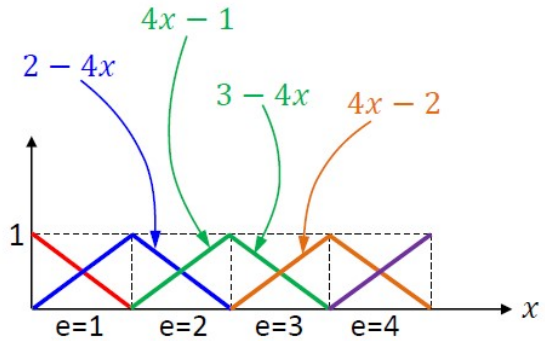
5^{th} Eq. with $w = \phi_5 \rightarrow \int_0^1 \left(\frac{d\phi_5}{dx}\frac{du}{dx} - \phi_5 u + \phi_5 x^2 \right) dx - \phi_5\frac{du}{dx}\Big|_1 + \phi_5\frac{du}{dx}\Big|_0 = 0$

Noting that

$$\phi_1(x)\Big|_1 = \phi_2(x)\Big|_{x=1} = \phi_2(x)\Big|_{x=0} = \phi_3(x)\Big|_1 = \phi_3(x)\Big|_0 = \phi_4(x)\Big|_1 = \phi_4(x)\Big|_0 = \phi_5(x)\Big|_0 = 0$$

and $\phi_1(x)\Big|_{x=0} = \phi_5(x)\Big|_{x=1} = 1$ and noting that the integrals are non zero only over certain elements

e.g. integral of the 3rd Eq. is $I_3 = \int_0^1 \left(\frac{d\phi_3}{dx}\frac{du}{dx} - \phi_3 u + \phi_3 x^2 \right) dx$ non zero only over *e*=2 and *e*=3 because ϕ_3 is non zero only over *e*=2 and *e*=3.



I_3 can be found as $I_3 = -\frac{97}{24}u_2 + \frac{47}{6}u_3 - \frac{97}{24}u_4 + \frac{25}{384}$

Evaluating the other integrals we find

$$\begin{bmatrix} \frac{47}{12} & -\frac{97}{24} & & & \\ -\frac{97}{24} & \frac{47}{6} & -\frac{97}{24} & & \\ & -\frac{97}{24} & \frac{47}{6} & -\frac{97}{24} & \\ & & -\frac{97}{24} & \frac{47}{6} & -\frac{97}{24} \\ & & & -\frac{97}{24} & \frac{47}{12} \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{Bmatrix} = \begin{Bmatrix} -\frac{1}{768} \\ -\frac{7}{384} \\ -\frac{25}{384} \\ -\frac{55}{384} \\ -\frac{27}{256} \end{Bmatrix} + \begin{Bmatrix} \left(-\frac{du}{dx}\right)\Big|_{x=0} \\ 0 \\ 0 \\ 0 \\ \left(\frac{du}{dx}\right)\Big|_{x=1} \end{Bmatrix} \begin{matrix} Q_1 \\ \\ \\ \\ Q_5 \end{matrix}$$

$[K]$ $\{u\}$ $\{F\}$ $\{Q\}$

i.e., this system has 5 equations for 5 unknowns.

u_1 and u_5 are known but Q_1 and Q_5 are unknown.

$$\begin{bmatrix} K_{11} & K_{12} & K_{13} & K_{14} & K_{15} \\ K_{21} & K_{22} & K_{23} & K_{24} & K_{25} \\ K_{31} & K_{32} & K_{33} & K_{34} & K_{35} \\ K_{41} & K_{42} & K_{43} & K_{44} & K_{44} \\ K_{51} & K_{52} & K_{53} & K_{54} & K_{55} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{bmatrix} = \begin{bmatrix} F_1 \\ F_2 \\ F_3 \\ F_4 \\ F_5 \end{bmatrix} + \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \\ Q_4 \\ Q_5 \end{bmatrix}$$

We want to solve $u's$. For this, we apply reduction to the *NN* x *NN* system and drop the 1st and 5th equations, because u_1 and u_5 are known.

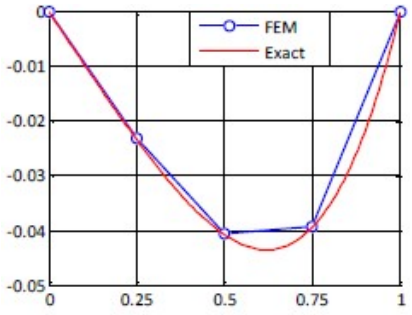
Reduced system is 3 x 3

$$\begin{bmatrix} K_{22} & K_{23} & K_{24} \\ K_{32} & K_{33} & K_{34} \\ K_{42} & K_{43} & K_{44} \end{bmatrix} \begin{bmatrix} u_2 \\ u_3 \\ u_4 \end{bmatrix} = \begin{bmatrix} F_2 - K_{21}u_1 - K_{25}u_5 \\ F_3 - K_{31}u_1 - K_{35}u_5 \\ F_4 - K_{41}u_1 - K_{45}u_5 \end{bmatrix} + \begin{bmatrix} Q_2 \\ Q_3 \\ Q_4 \end{bmatrix}$$

Since $u_1 = u_5 = 0$

$$\begin{bmatrix} \frac{47}{6} & -\frac{97}{24} & 0 \\ -\frac{97}{24} & \frac{47}{6} & -\frac{97}{24} \\ 0 & -\frac{97}{24} & \frac{47}{6} \end{bmatrix} \begin{bmatrix} u_2 \\ u_3 \\ u_4 \end{bmatrix} = \begin{bmatrix} -\frac{7}{384} - 0 - 0 \\ -\frac{25}{384} - 0 - 0 \\ -\frac{55}{384} - 0 - 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} u_2 \\ u_3 \\ u_4 \end{bmatrix} = \begin{bmatrix} -0.0232 \\ -0.0405 \\ -0.0392 \end{bmatrix}$$

where exact solution is $\begin{bmatrix} u_2 \\ u_3 \\ u_4 \end{bmatrix}_{\text{exact}} = \begin{bmatrix} -0.0234 \\ -0.0408 \\ -0.0394 \end{bmatrix}$



HW-9: Solve $-\frac{d^2u}{dx^2} - 2u = -x, \quad 0 < x < 1, \quad u(0) = 0, \quad u(1) = 0$ by using FEM

10. Solution of Integral Equations; Method of Moments.

Method of Moments (MoM) transforms integro-differential equations into matrix systems of linear equations which can be solved numerically.

Consider the inhomogeneous equation

$$L(u) = k \Rightarrow L(u) - k = 0$$

where L is a linear integro-differential operator, u is an unknown function (to be solved) and k is a known function (excitation).

For example,

(a) consider the integral equation for a line charge density $V_0 = \int \frac{\lambda(x')dx'}{4\pi\varepsilon_0 r(x,x')}$ where

$$u = \lambda(x), k = V_0, L = \int \frac{dx'}{4\pi\varepsilon_0 r(x,x')}$$

(b) $-\frac{d^2 f(x)}{dx^2} = 3 + 2x^2$ where $u = f(x), k = 3 + 2x^2, L = -\frac{d^2}{dx^2}$

To solve u approximate it by sum of weighted known basis functions or expansion functions such as

$$u = \sum_{n=1}^{N} u_n = \sum_{n=1}^{N} I_n b_n, \qquad n = 1,2,...N$$

where b_n is the expansion function, I_n is its unknown complex coefficients to be found, N is the total number of expansion function.

Since L is linear, substitution of the above equation in the integro-differential equation, we obtain

$L\left(\sum_{n=1}^{N} I_n b_n\right) \approx k$ where the error or residual is $R = k - L\left(\sum_{n=1}^{N} I_n b_n\right)$

Replacing u by u_n where $n = 1,2,...N$

Taking inner product with a set of w_m weighting or testing functions, making the residual R=0

In the range of $L \rightarrow \langle w_m, \left(L\left(u_n\right) - k\right)\rangle = 0, \qquad m = 1,2,...,M$

Since I_n is constant, we can write

$\sum_{n=1}^{N} I_n \langle w_m, L\left(b_n\right)\rangle = \langle w_m, k\rangle, \qquad m = 1,2,...,M$ where M and N are theoretically infinite but in practice they are finite numbers.

Inner product $\langle w, g\rangle$ which is a scalar is defined by $\langle w, g\rangle = \langle g, w\rangle = \int g\left(x\right) w\left(x\right) dx$

$\langle bf + cg, w\rangle = b\langle f, w\rangle + c\langle g, w\rangle$, $\langle g^*, g\rangle > 0$ if $g \neq 0$ and $\langle g^*, g\rangle = 0$ if $g = 0$

where b and c are scalar and * is the complex conjugate.

Writing $\sum_{n=1}^{N} I_n \langle w_m, L\left(b_n\right)\rangle = \langle w_m, k\rangle, \qquad m = 1,2,...,M$ in matrix form

$$[Z][I] = [V]$$

where

$[I] = \begin{bmatrix} I_1 & I_2 & \dots & I_N \end{bmatrix}^T$, $[V] = \begin{bmatrix} \langle k, w_1\rangle & \langle k, w_2\rangle & \dots & \langle k, w_M\rangle \end{bmatrix}^T$,

$$[Z]=\begin{bmatrix} \langle w_1, L(b_1)\rangle & \langle w_1, L(b_2)\rangle & \dots & \langle w_1, L(b_N)\rangle \\ \langle w_2, L(b_1)\rangle & \langle w_2, L(b_2)\rangle & \dots & \langle w_2, L(b_N)\rangle \\ \langle w_3, L(b_1)\rangle & \langle w_3, L(b_2)\rangle & \dots & \langle w_3, L(b_N)\rangle \\ . & . & \dots & . \\ . & . & \dots & . \\ . & . & \dots & . \\ \langle w_M, L(b_1)\rangle & \langle w_M, L(b_2)\rangle & \dots & \langle w_M, L(b_N)\rangle \end{bmatrix}$$

Solving for the unknown $[I]$

$$[I]=[Z]^{-1}[V]$$

Example: Consider a 1-D differential equation $-\frac{d^2 f(x)}{dx^2}=3+2x^2$ with boundary conditions $f(0)=f(1)=0$. Solve this equation using Galerkin's Method of Moments (MoM).

Solution: $u=f(x)$

$k=3+2x^2,\ \ L=-\frac{d^2}{dx^2}$

Due to the nature of $k=3+2x^2,$ choose the basis function to be $b_n(x)=x^n$ but the boundary condition $f(1)=0$ cannot be satisfied with this basis function so better to choose $b_n(x)=x-x^{n+1},\qquad n=1,2,...,N.$

Assuming $N=2$ which is the total number of subsections in the interval $[0,1]$

Approximation of the unknown function

$$f(x)=I_1 b_1(x)+I_2 b_2(x)=I_1\left(x-x^2\right)+I_2\left(x-x^3\right)$$

In Galerkin's MoM, the weighting functions are

$w(x)=x-x^{m+1},\qquad m=1,2,...,M$

Writing $[Z]$ with $M=N=2$ where

$$Z_{11}=\langle w_1, L(b_1)\rangle=\int_0^1 w_1(x)\,L(b_1(x))\,dx=\int_0^1\left(x-x^2\right)(2)\,dx=\frac{1}{3}$$

$$Z_{12} = \langle w_1, L(b_2) \rangle = \int_0^1 w_1(x) L(b_2(x)) dx = \int_0^1 (x - x^2)(6x) dx = \frac{1}{2}$$

$$Z_{21} = \langle w_2, L(b_1) \rangle = \int_0^1 w_2(x) L(b_1(x)) dx = \int_0^1 (x - x^3)(2) dx = \frac{1}{2}$$

$$Z_{22} = \langle w_2, L(b_2) \rangle = \int_0^1 w_2(x) L(b_2(x)) dx = \int_0^1 (x - x^3)(6x) dx = \frac{4}{5}$$

Writing $[V]$ where

$$V_1 = \langle k, w_1 \rangle = \int_0^1 k(x) w_1(x) dx = \int_0^1 (3 + 2x^2)(x - x^2) dx = \frac{3}{5}$$

$$V_2 = \langle k, w_2 \rangle = \int_0^1 k(x) w_2(x) dx = \int_0^1 (3 + 2x^2)(x - x^3) dx = \frac{11}{12}$$

So $[Z][I] = [V] \Rightarrow \begin{bmatrix} \frac{1}{3} & \frac{1}{2} \\ \frac{1}{2} & \frac{4}{5} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} \frac{3}{5} \\ \frac{11}{12} \end{bmatrix} \Rightarrow [I] = \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} \frac{13}{10} \\ \frac{1}{3} \end{bmatrix}$

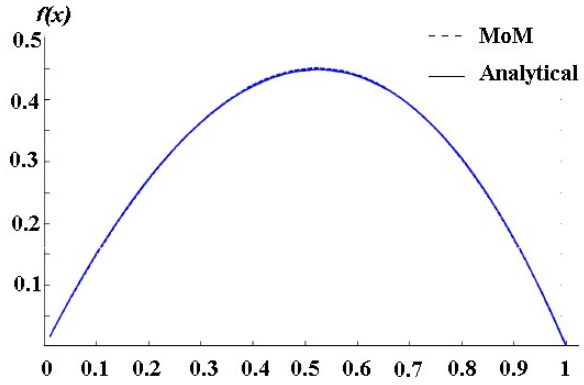
The unknown function $f(x)$ is

$$f(x) = I_1(x - x^2) + I_2(x - x^3) = \frac{13}{10}(x - x^2) + \frac{1}{3}(x - x^3)$$

Which satisfies the boundary conditions of $f(0) = f(1) = 0$

On the other hand, the analytical solution of this differential equation is $f(x) = \frac{5}{3}x - \frac{3}{2}x^2 - \frac{1}{6}x^4$

Comparison of the exact solution (analytical) and the approximate solution (MoM) of this example is as follows:



HW-10: Solve the differential equation $-\frac{d^2 f(x)}{dx^2} = 1 + x^2$ with boundary conditions $f(0) = f(1) = 0$ by using Galerkin's Method of Moments (MoM).

11. Optimization; Convexity and Convergence.

Optimization problem: Standard form

Minimize $f_0(x)$ subject to $f_i(x) \le 0$, $i = 1, \ldots, m$ and $h_i(x) = 0$, $i = 1, \ldots, p$

where x is optimization variable, f_0 is objective or cost function,

$f_i(x) \le 0$ are the inequality constraints, $h_i(x) = 0$ are the equality constraints

x is feasible if it satisfies the constraints

The feasible set C is the set of all feasible points

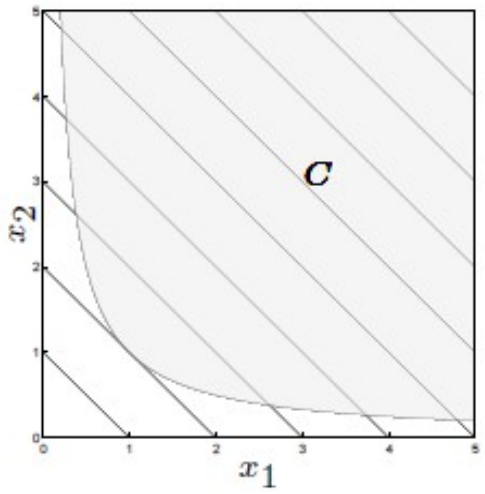
Problem is feasible if there are feasible points

Problem is unconstrained if $m = p = 0$

Optimal point: $x \in C$ such that $f(x) = f^*$, optimal set: $X_{opt} = \{ x \in C \mid f(x) = f^* \}$

Example: Minimize $x_1 + x_2$ subject to $-x_1 \le 0$, $-x_2 \le 0$, $1 - x_1 x_2 \le 0$

Feasible set C is half-hyperboloid



Optimal point is $x^* = (1, 1)$

Optimal value is $f^* = 2$

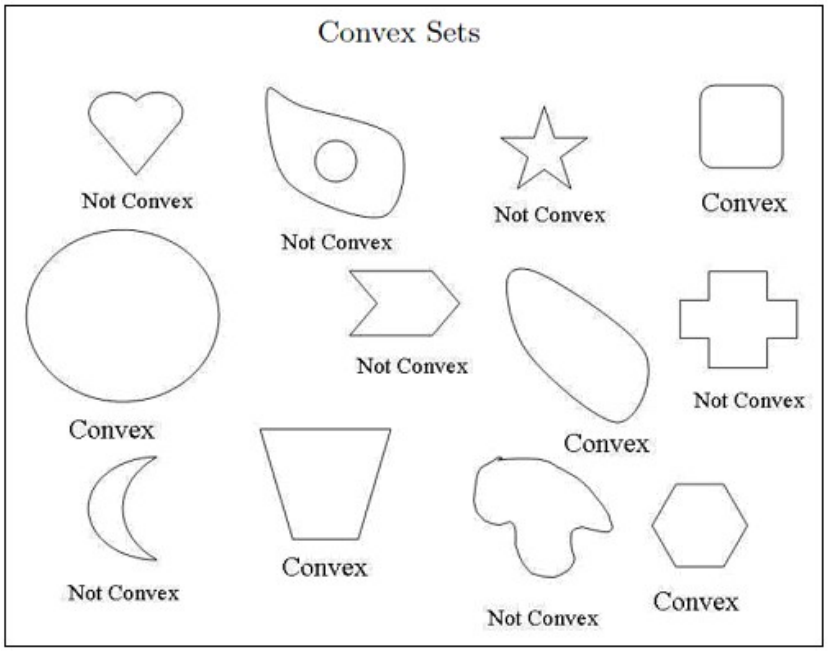
Definition of a convex set:

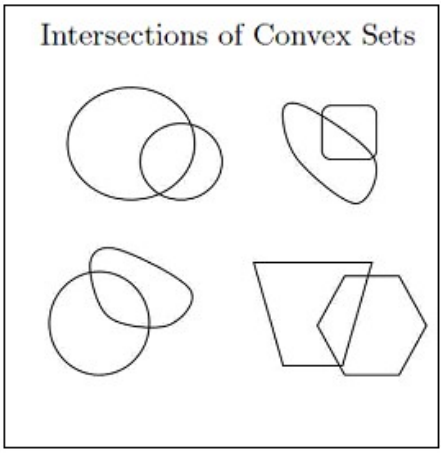
A set S is said to be convex if for each $x_1, x_2 \in S$, the line segment $\lambda x_1 + (1-\lambda) x_2$ for $\lambda \in (0,1)$ belongs to S.

This says that all points on a line connecting two points in the set are in the set.

The intersection of a finite or infinite number of convex sets is also convex.

Examples of Convex Sets





Convex optimization problems can be solved quickly and reliably up to very large size (hundreds of thousands of variables and constraints).

The issue is that, unless our objective and constraints are linear, it is difficult to determine whether or not they were convex.

Convex optimization problem

In standard form

minimize $f_0(x)$ subject to $f_i(x) \le 0, \quad i = 1,.....,m, \quad a_i^T x - b_i = 0, \quad i = 1,....., p$

where $f_0(x), f_1(x),...., f_m(x)$ are convex.

Example: minimize $x_1 + x_2$ subject to $-x_1 \le 0, \quad -x_2 \le 0, \quad 1 - \sqrt{x_1 x_2} \le 0$ where $1 - \sqrt{x_1 x_2}$ is convex.

or: minimize $x_1 + x_2$ subject to $-x_1 \le 0, \quad -x_2 \le 0, \quad -\log(x_1) - \log(x_2) \le 0$

Example: A farmer has 2400 m of fencing and wants to fence off a rectangular field that borders a straight river. He needs no fence along the river. What are the dimensions of the field that has the largest area?

Answer: Maximize $A = xy$

Constraint: $2x + y = 2400$

Solving the second equation for y

$2x + y = 2400 \Rightarrow y = 2400 - 2x$

Substituting the result into the first equation

$A = xy \Rightarrow x(2400 - 2x) = 2400x - 2x^2$

To find the absolute maximum value of $A = 2400x - 2x^2$,

Closed Interval Method is used to find the absolute maximum and minimum values of a continuous function f on a closed interval $[a, b]$ by using the below steps:

1. Find the values of f at the critical numbers of f in $[a, b]$
2. Find the values of f at the end points of the interval.
3. The largest of the values from Step 1 and 2 is the absolute maximum value; the smallest value of these values is the absolute minimum value.

$y \ge 0 \Rightarrow 2400 - 2x \ge 0 \Rightarrow 2400 \ge 2x \Rightarrow 1200 \ge x$

On the other hand $x \geq 0$

Combining these two inequalities gives $0 \leq x \leq 1200$

The derivative of $A(x)$ is

$$A'(x) = \left(2400x - 2x^2\right)' = 2400x' - 2\left(x^2\right)' = 2400 - 4x$$

To find the critical numbers we solve the equation

$$2400 - 4x = 0 \Rightarrow x = 600$$

To find the maximum value of $A(x)$ we evaluate it at the end points and critical number:

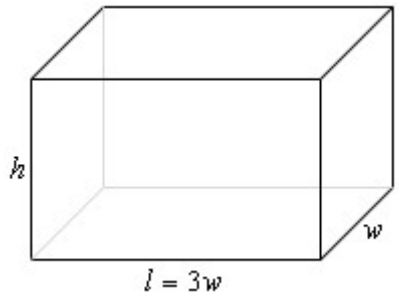
$$A(0) = 0, A(600) = 2400 \times 600 - 2(600)^2 = 720000,\ A(1200) = 0$$

The Closed Interval Method gives the maximum value as $A(600) = 720000$ m^2

The dimenions are $x = 600$ m, $y = 2400 - 2 \times 600 = 1200$ m $= 1200$ m.

Example: We want to construct a box whose base length is 3 times the base width. The material used to build the top and bottom costs 10 TL/ m^2 and the material used to build the sides costs 6 TL/m^2. If the box must have a volume of 50 m^3, find the dimensions that will minimize the cost to build the box.

Answer:



We want to minimize the cost of the materials subject to the constraint that the volume must be 50 m^3. Note as well that the cost for each side is just the area of that side times the appropriate cost.

The two functions are

$$\text{Minimize: } C = 10(2lw) + 6(2wh + 2lh) = 10\left[2(3w)w\right] + 6\left[2wh + 2(3w)h\right] = 60w^2 + 48wh$$

$$\text{Constraint: } 50 = lwh = 3wwh = 3w^2h$$

Solving the constraint for one of the variables

$$h = \frac{50}{3w^2}$$

Inserting into the cost

$$C(w) = 60w^2 + 48wh = 60w^2 + 48w\frac{50}{3w^2} = 60w^2 + 48\frac{50}{3w} = 60w^2 + \frac{800}{w}$$

We can't use the Closed Interval Method because the domain of $C(w)$ is $(0,\infty)$ which is not a finite interval. Instead, we will use

FIRST DERIVATIVE TEST FOR ABSOLUTE EXTREME VALUES: Suppose that c is a critical number of a continuous function f defined on an interval.

(a) If $f'(x) > 0$ for all $x < c$ and $f'(x) < 0$ for all $x > c$, then $f(c)$ is the absolute maximum value of f.

(b) If $f'(x) < 0$ for all $x < c$ and $f'(x) > 0$ for all $x > c$, then $f(c)$ is the absolute minimum value of f.

Taking the derivative

$$C'(w) = 120w - \frac{800}{w^2} = \frac{120w^3 - 800}{w^2}$$

Since $w > 0$, the only critical number is found from

$$C'(w) = \frac{120w^3 - 800}{w^2} = 0 \Rightarrow w = \left(\frac{800}{120}\right)^{1/3} = \left(\frac{20}{3}\right)^{1/3} \approx 1.8821$$

It can be seen that $C'(w) < 0$ for all $0 < w < \left(\frac{20}{3}\right)^{1/3}$ and $C'(w) > 0$ for all $w > \left(\frac{20}{3}\right)^{1/3}$

Thus, the minimum value of the cost should occur at $w = \left(\frac{20}{3}\right)^{1/3}$

The dimensions are: $w \approx 1.8821$ m, $l = 3w = 3 \times 1.8821 \approx 5.6463$ m, $h = \frac{50}{3w^2} \approx 4.7050$ m

Minimum cost is $C\left[w = \left(\frac{20}{3}\right)^{1/3}\right] = 60\left(\frac{20}{3}\right)^{2/3} + \frac{800}{\left(\frac{20}{3}\right)^{1/3}} \approx 637.60$ TL

Convergence

$$\text{Norm} = \|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}$$

Types of convergence definitions:

1. If $x_n \to x^*$ and there is $K > 0$ such that $\|x_{n+1} - x^*\| \le K \|x_n - x^*\|^2$

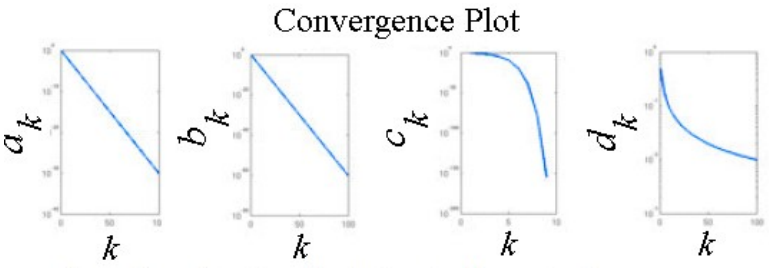
2. If $x_n \to x^*$ and there is $K > 0$ such that $\|x_{n+1} - x^*\| \le K \|x_n - x^*\|^\alpha$ where $\alpha > 1$

3. If $x_n \to x^* \Rightarrow \lim_{n \to \infty} \dfrac{\|x_{n+1} - x^*\|}{\|x_n - x^*\|} = 0$

4. If $x_n \to x^* \Rightarrow \|x_{n+1} - x^*\| \le \alpha \|x_n - x^*\|$ where $\alpha \in (0,1)$ for n sufficiently large.

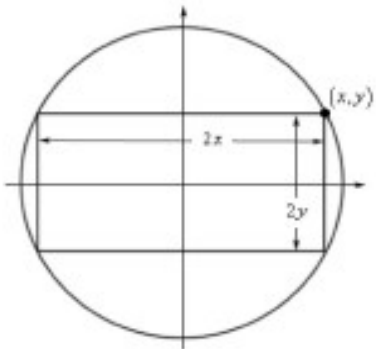
Examples: Consider the following sequences:

$$\begin{array}{llllllll}
a_0 = 1, & a_1 = \frac{1}{2}, & a_2 = \frac{1}{4}, & a_3 = \frac{1}{8}, & a_4 = \frac{1}{16}, & a_5 = \frac{1}{32}, & \ldots, & a_k = \frac{1}{2^k}, \quad \ldots \\
b_0 = 1, & b_1 = 1, & b_2 = \frac{1}{4}, & b_3 = \frac{1}{4}, & b_4 = \frac{1}{16}, & b_5 = \frac{1}{16}, & \ldots, & b_k = \frac{1}{4^{\left\lfloor \frac{k}{2} \right\rfloor}}, \quad \ldots \\
c_0 = \frac{1}{2}, & c_1 = \frac{1}{4}, & c_2 = \frac{1}{16}, & c_3 = \frac{1}{256}, & c_4 = \frac{1}{65\,536}, & & \ldots, & c_k = \frac{1}{2^{2^k}}, \quad \ldots \\
d_0 = 1, & d_1 = \frac{1}{2}, & d_2 = \frac{1}{3}, & d_3 = \frac{1}{4}, & d_4 = \frac{1}{5}, & d_5 = \frac{1}{6}, & \ldots, & d_k = \frac{1}{k+1}, \quad \ldots
\end{array}$$



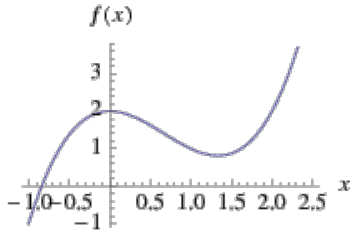
Linear, Linear, Superlinear/Quadratic and sublinear rate of convergence.

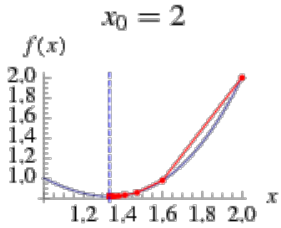
HW-11: Determine the area of the largest rectangle that can be inscribed in a circle of radius 4.

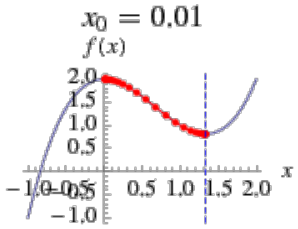


12. The Steepest Descent Method.

The Steepest Descent Method is an algorithm for finding the nearest local minimum of a function which presupposes that the gradient of the function can be computed. The method of steepest descent starts at a point P_0 and, as many times as needed, moves from P_i to P_{i+1} by minimizing along the line extending from P_i in the direction of $-\nabla f(\boldsymbol{P_i})$, the local downhill gradient.







The gradient vector of a scalar function $f(x_1, x_2 ..., x_n)$ is defined as a column vector

$$\nabla f = \left[\frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2} \quad \ldots \quad \frac{\partial f}{\partial x_n} \right]^T = \boldsymbol{c}$$

e.g. For $f(x_1, x_2) = 25x_1^2 + x_2^2 \Rightarrow \nabla f = \left[\frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2} \right]^T = \left[2(25)x_1 \quad 2x_2 \right]^T$

At the point $x_1 = 0.6, x_2 = 4 \Rightarrow f(x_1, x_2) = 25x_1^2 + x_2^2 \Rightarrow \boldsymbol{c} = \nabla f = \left[2(25)0.6 \quad 2(4) \right]^T = \left[30 \quad 8 \right]^T$

Defining the normalized gradient vector $\bar{\boldsymbol{c}} = \dfrac{\boldsymbol{c}}{\sqrt{\boldsymbol{c}^T \boldsymbol{c}}}$

At the point $x_1 = 0.6, x_2 = 4 \Rightarrow \bar{\boldsymbol{c}} = \dfrac{\boldsymbol{c}}{\sqrt{\boldsymbol{c}^T \boldsymbol{c}}} = \dfrac{1}{\sqrt{\begin{bmatrix} 30 & 8 \end{bmatrix} \begin{bmatrix} 30 \\ 8 \end{bmatrix}}} \begin{bmatrix} 30 \\ 8 \end{bmatrix} = \dfrac{1}{\sqrt{(30)^2 + 8^2}} \begin{bmatrix} 30 \\ 8 \end{bmatrix} = \begin{bmatrix} 0.96625 \\ 0.2577 \end{bmatrix}$

Remark: The gradient vector represents a direction of maximum rate of increase for the function at the point of evaluation, i.e., at the point $x_1 = 0.6, x_2 = 4$ in the above example, i.e.,

$$f(0.6, 4) = 25(0.6)^2 + 4^2 = 25$$

If x is increased in the direction $\bar{\boldsymbol{c}}$ by a step $\alpha = 0.5$

$$\boldsymbol{x}^{(1)} = \boldsymbol{x}^{(0)} + \alpha \bar{\boldsymbol{c}} = \begin{bmatrix} 0.6 \\ 4 \end{bmatrix} + 0.5 \begin{bmatrix} 0.96625 \\ 0.2577 \end{bmatrix} = \begin{bmatrix} 1.083125 \\ 4.12885 \end{bmatrix}$$

The value of the function becomes

$$\boldsymbol{f}\left(\boldsymbol{x}^{(1)}\right) = 25(1.083125)^2 + (4.12885)^2 = 46.327$$

If we move in a direction $\begin{bmatrix} 1 & 0 \end{bmatrix}^T$

$$\boldsymbol{x}^{(1)} = \boldsymbol{x}^{(0)} + \alpha \bar{\boldsymbol{c}} = \begin{bmatrix} 0.6 \\ 4 \end{bmatrix} + 0.5 \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1.1 \\ 4 \end{bmatrix}$$

The value of the function becomes

$$\boldsymbol{f}\left(\boldsymbol{x}^{(1)}\right) = 25(1.1)^2 + (4)^2 = 46.25$$

If we move in a direction $\begin{bmatrix} 0 & 1 \end{bmatrix}^T$

$$\boldsymbol{x}^{(1)} = \boldsymbol{x}^{(0)} + \alpha \bar{\boldsymbol{c}} = \begin{bmatrix} 0.6 \\ 4 \end{bmatrix} + 0.5 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.6 \\ 4.5 \end{bmatrix}$$

The value of the function becomes

$$\boldsymbol{f}\left(\boldsymbol{x}^{(1)}\right) = 25(0.6)^2 + (4.5)^2 = 29.25$$

The result is that moving along the gradient direction results in the maximum increase in the function.

Remark: The gradient vector $\boldsymbol{c}$ of $f(x_1, x_2 ..., x_n)$ at the point of evaluation x^* is orthogonal (normal) to the tangent plane for the surface $f(x^*) = \text{constant}$. For example, at the point x^* where $x_1 = 0.6, x_2 = 4 \Rightarrow f(x_1, x_2) = 25x_1^2 + x_2^2 = 25(0.6)^2 + (4)^2 = 25$, the slope at the point x^* is

$$\text{slope} = \left.\frac{dx_2}{dx_1}\right|_{x_1 = 0.6, x_2 = 4}$$

Equating the differentiation of $f(x_1, x_2) = 25x_1^2 + x_2^2$ to zero

$$\Rightarrow 25(2)x_1 dx_1 + 2x_2 dx_2 = 0 \Rightarrow \frac{dx_2}{dx_1} = -\frac{25x_1}{x_2}$$

At the point x^*, slope $= \left.\dfrac{dx_2}{dx_1}\right|_{x_1=0.6,\,x_2=4} = -\dfrac{25(0.6)}{4} = -3.75$

The direction of the tangent line is given by $\boldsymbol{t} = \begin{bmatrix} 1 \\ -3.75 \end{bmatrix}$

$\boldsymbol{c}$ and $\boldsymbol{t}$ are normal each other as $\boldsymbol{c}^T\boldsymbol{t} = \begin{bmatrix} 30 & 8 \end{bmatrix}\begin{bmatrix} 1 \\ -3.75 \end{bmatrix} = 30 - 8(3.75) = 0$

Remark: The maximum rate of change of $f(x)$ at any point x^* is the magnitude of the gradient vector given by $\| \boldsymbol{c} \| = \sqrt{\boldsymbol{c}^T\boldsymbol{c}}$

Steepest descent direction. Let $f(x)$ be a differentiable function with respect to *x*. The direction of steepest descent for $f(x)$ at any point is $\boldsymbol{d} = -\boldsymbol{c}$ or $\bar{\boldsymbol{d}} = -\bar{\boldsymbol{c}}$

Example. Use the steepest descent direction to find the minimum of $f(x_1, x_2) = 25x_1^2 + x_2^2$ starting at $x^{(0)} = \begin{bmatrix} 1 & 3 \end{bmatrix}^T$ with a step size of $\alpha = 0.5$. The function value at the starting point is

$$f\left(\boldsymbol{x}^{(0)}\right) = 25x_1^2 + x_2^2 = 25(1)^2 + 3^2 = 34$$

From the analytical solution, the minimum point is at $x^* = \begin{bmatrix} 0 & 0 \end{bmatrix}^T$ and $f(\boldsymbol{x}^*) = 0$.
Starting the process of iterations.

$$\boldsymbol{c} = \nabla f = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} & \dfrac{\partial f}{\partial x_2} \end{bmatrix}^T \Rightarrow \boldsymbol{c}^{(0)} = \begin{bmatrix} 2(25)1 \\ 2(3) \end{bmatrix} = \begin{bmatrix} 50 \\ 6 \end{bmatrix}$$

$$\bar{\boldsymbol{c}} = \frac{\boldsymbol{c}}{\sqrt{\boldsymbol{c}^T\boldsymbol{c}}} \Rightarrow \bar{\boldsymbol{c}}^{(0)} = \frac{1}{\sqrt{(50)^2 + 6^2}}\begin{bmatrix} 50 \\ 6 \end{bmatrix} = \begin{bmatrix} 0.9929 \\ 0.1191 \end{bmatrix}$$

$$\boldsymbol{x}^{(1)} = \boldsymbol{x}^{(0)} - 0.5\bar{\boldsymbol{c}}^{(0)} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} - 0.5\begin{bmatrix} 0.9929 \\ 0.1191 \end{bmatrix} = \begin{bmatrix} 0.50359 \\ 2.9404 \end{bmatrix}$$

$$f\left(\boldsymbol{x}^{(1)}\right) = 25(0.5035)^2 + (2.9404)^2 = 14.984$$

Next iteration

$$\boldsymbol{c}^{(1)} = \begin{bmatrix} 2(25)(0.5035) \\ 2(2.9404) \end{bmatrix} = \begin{bmatrix} 25.175 \\ 5.8808 \end{bmatrix}$$

$$\bar{\boldsymbol{c}} = \frac{\boldsymbol{c}}{\sqrt{\boldsymbol{c}^T\boldsymbol{c}}} \Rightarrow \bar{\boldsymbol{c}}^{(1)} = \frac{1}{\sqrt{(25.175)^2 + (5.8808)^2}} \begin{bmatrix} 25.175 \\ 5.8808 \end{bmatrix} = \begin{bmatrix} 0.9738 \\ 0.2275 \end{bmatrix}$$

$$\boldsymbol{x}^{(2)} = \boldsymbol{x}^{(1)} - 0.5\bar{\boldsymbol{c}}^{(1)} = \begin{bmatrix} 0.50359 \\ 2.9404 \end{bmatrix} - 0.5 \begin{bmatrix} 0.9738 \\ 0.2275 \end{bmatrix} = \begin{bmatrix} 0.0166 \\ 2.8267 \end{bmatrix}$$

$$f\left(\boldsymbol{x}^{(2)}\right) = 25(0.0166)^2 + (2.8267)^2 = 7.997$$

Next iteration

$$\boldsymbol{c}^{(2)} = \begin{bmatrix} 0.83 \\ 5.6534 \end{bmatrix} \Rightarrow \bar{\boldsymbol{c}}^{(2)} = \begin{bmatrix} 0.1453 \\ 0.9894 \end{bmatrix}$$

$$\boldsymbol{x}^{(3)} = \boldsymbol{x}^{(2)} - 0.5\bar{\boldsymbol{c}}^{(2)} = \begin{bmatrix} 0.166 \\ 2.8267 \end{bmatrix} - 0.5 \begin{bmatrix} 0.1453 \\ 0.9894 \end{bmatrix} = \begin{bmatrix} -0.0561 \\ 2.332 \end{bmatrix}$$

$$f\left(\boldsymbol{x}^{(3)}\right) = 25(-0.0561)^2 + (2.332)^2 = 5.5169$$

Next iteration

$$\boldsymbol{c}^{(3)} = \begin{bmatrix} -2.805 \\ 4.664 \end{bmatrix} \Rightarrow \bar{\boldsymbol{c}}^{(3)} = \begin{bmatrix} -0.5154 \\ 0.8570 \end{bmatrix}$$

$$\boldsymbol{x}^{(4)} = \boldsymbol{x}^{(3)} - 0.5\bar{\boldsymbol{c}}^{(3)} = \begin{bmatrix} -0.0561 \\ 2.332 \end{bmatrix} - 0.5 \begin{bmatrix} -0.5154 \\ 0.857 \end{bmatrix} = \begin{bmatrix} 0.2016 \\ 1.9035 \end{bmatrix}$$

$$f\left(\boldsymbol{x}^{(4)}\right) = 25(0.2016)^2 + (1.9035)^2 = 4.6394$$

Next iteration

$$\boldsymbol{c}^{(4)} = \begin{bmatrix} 10.08 \\ 3.807 \end{bmatrix} \Rightarrow \bar{\boldsymbol{c}}^{(4)} = \begin{bmatrix} 0.9355 \\ 0.3533 \end{bmatrix}$$

$$\boldsymbol{x}^{(5)} = \boldsymbol{x}^{(4)} - 0.5\bar{\boldsymbol{c}}^{(4)} = \begin{bmatrix} 0.2016 \\ 1.9035 \end{bmatrix} - 0.5 \begin{bmatrix} 0.9355 \\ 0.3533 \end{bmatrix} = \begin{bmatrix} -0.2662 \\ 1.7269 \end{bmatrix}$$

$$f\left(\boldsymbol{x}^{(5)}\right) = 25(-0.2662)^2 + (1.7269)^2 = 4.7537$$

Note that the function values start to oscillate, i.e., not monotonically reduce. This is caused by the constant step size. When the iteration approaches the minimum, a smaller step size should be used. Otherwise, an "overshoot" will occur which means that we move along the steepest direction more than needed. We must find the best step size at each iteration by conducting a one-D optimization in

the steepest descent direction. For example, the new point can be expressed as a function of step size α, i.e.,

$$\boldsymbol{x}^{(1)} = \boldsymbol{x}^{(0)} - \alpha \bar{\boldsymbol{c}}^{(0)} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} - \alpha \begin{bmatrix} 0.9929 \\ 0.1191 \end{bmatrix} = \begin{bmatrix} 1 - 0.9929\alpha \\ 3 - 0.1191\alpha \end{bmatrix}$$

$f\left(\boldsymbol{x}^{(1)}\right) = 25\left(1 - 0.9929\alpha\right)^2 + \left(3 - 0.1191\alpha\right)^2$ which is a function of α

Taking the derivative of $f\left(\boldsymbol{x}^{(1)}\right)$ with respect to α and equating to zero

$$\frac{df\left(\boldsymbol{x}^{(1)}\right)}{d\alpha} = 2(25)\left(1 - 0.9929\alpha^{(0)}\right)(-0.9929) + 2\left(3 - 0.1191\alpha^{(0)}\right)(-0.1191) = 0$$

$$\alpha^{(0)} = \frac{25(0.9929) + 3(0.1191)}{25(0.9929)^2 + (0.1191)^2} = 1.0211$$

$$\boldsymbol{x}^{(1)} = \boldsymbol{x}^{(0)} - \alpha^{(0)} \bar{\boldsymbol{c}}^{(0)} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} - 1.0211 \begin{bmatrix} 0.9929 \\ 0.1191 \end{bmatrix} = \begin{bmatrix} -0.0139 \\ 2.8784 \end{bmatrix}$$

$$f\left(\boldsymbol{x}^{(1)}\right) = 25(-0.0139)^2 + (2.8784)^2 = 8.29$$

Thus, $f = 8.29$ is the minimum value found at this iteration.

HW-12: Using the steepest descent direction, find the minimum of $f(x_1, x_2) = 3x_1^2 + 2x_2$ starting at $x^{(0)} = \begin{bmatrix} 1 & 2 \end{bmatrix}^T$ with a step size of $\alpha = 0.5$. Use 2 iterations.

13. The Gauss-Newton Method.

A nonlinear least squares problem is an unconstrained minimization problem of the form

Minimize $f(x_1, x_2, ..., x_n) = \sum_{i=1}^{m} f_i(x_1, x_2, ..., x_n)^2 = F(x_1, x_2, ..., x_n)^T F(x_1, x_2, ..., x_n)$ over x_1 and x_2

where $F(x_1, x_2, ..., x_n) = \left(f_1(x_1, x_2, ..., x_n) \quad f_2(x_1, x_2, ..., x_n) \quad ... \quad f_m(x_1, x_2, ..., x_n)\right)^T$

Jacobian of $F(x_1, x_2, ..., x_n) = J(x_1, x_2, ..., x_n) = \nabla F(x_1, x_2, ..., x_n)^T$

$$\nabla F\left(x_1, x_2, ..., x_n\right)^T = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ . & . & . & . \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & . & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

The gradient is $\nabla f\left(x\right) = \nabla F\left(x\right) F\left(x\right)$

$\nabla^2 f\left(x\right)$ is obtained by differentiating with respect to x_i

$\nabla^2 f\left(x\right) = \nabla F\left(x\right) \nabla F\left(x\right)^T + \sum_{i=1}^{m} f_i\left(x\right) \nabla^2 f_i\left(x\right)$ known as the Hessian of *f*.

Gauss-Newton method computes a search direction using the formula

$\nabla^2 f\left(x\right) p = -\nabla f\left(x\right)$ and replaces the Hessian with the approximation

$\nabla F\left(x\right) \nabla F\left(x\right)^T p = -\nabla F\left(x\right) F\left(x\right)$ where p is a vector

Example: Applying Gauss-Newton method to

$f\left(x_1, x_2\right) = \sum_{i=1}^{5} \left(x_1 e^{x_2 t_i} - y_i \right)$ with data

$t = \begin{pmatrix} 1 & 2 & 4 & 5 & 8 \end{pmatrix}^T$

$y = \begin{pmatrix} 3.2939 & 4.2699 & 7.1749 & 9.3008 & 20.259 \end{pmatrix}^T$

Using an initial guess that is close to the solution

$$x = \begin{pmatrix} 2.50 \\ 0.25 \end{pmatrix}$$

Solution: $F\left(x\right) = \begin{pmatrix} f_1\left(x\right) & f_2\left(x\right) & ... & f_m\left(x\right) \end{pmatrix}^T$

$$F\left(x\right) = \begin{pmatrix} x_1 e^{x_2 t_1} - y_1 \\ x_1 e^{x_2 t_2} - y_2 \\ x_1 e^{x_2 t_3} - y_3 \\ x_1 e^{x_2 t_4} - y_4 \\ x_1 e^{x_2 t_5} - y_5 \end{pmatrix}$$

Evaluating $F(x)$ at the an initial guess $x = \begin{pmatrix} 2.50 \\ 0.25 \end{pmatrix}$ and at

$t = \begin{pmatrix} 1 & 2 & 4 & 5 & 8 \end{pmatrix}^T$

$y = \begin{pmatrix} 3.2939 & 4.2699 & 7.1749 & 9.3008 & 20.259 \end{pmatrix}^T$

$$F(x) = \begin{pmatrix} 2.5e^{(0.25)(1)} - 3.2939 \\ 2.5e^{(0.25)(2)} - 4.2699 \\ 2.5e^{(0.25)(4)} - 7.1749 \\ 2.5e^{(0.25)(5)} - 9.3008 \\ 2.5e^{(0.25)(8)} - 20.259 \end{pmatrix} = \begin{pmatrix} -0.0838 \\ -0.1481 \\ -0.3792 \\ -0.5749 \\ -1.7864 \end{pmatrix}$$

$$\nabla F(x_1, x_2, ..., x_n)^T = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ . & . & . & . \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & . & \frac{\partial f_m}{\partial x_n} \end{bmatrix} = \begin{bmatrix} e^{x_2 t_1} & x_1 t_1 e^{x_2 t_1} \\ e^{x_2 t_2} & x_1 t_2 e^{x_2 t_2} \\ e^{x_2 t_3} & x_1 t_3 e^{x_2 t_3} \\ e^{x_2 t_4} & x_1 t_4 e^{x_2 t_4} \\ e^{x_2 t_5} & x_1 t_5 e^{x_2 t_5} \end{bmatrix} = \begin{bmatrix} e^{(0.25)1} & (2.5)1e^{(0.25)1} \\ e^{(0.25)2} & (2.5)2e^{(0.25)2} \\ e^{(0.25)4} & (2.5)4e^{(0.25)4} \\ e^{(0.25)5} & (2.5)5e^{(0.25)5} \\ e^{(0.25)8} & (2.5)8e^{(0.25)8} \end{bmatrix}$$

$$= \begin{bmatrix} 1.2840 & 3.2101 \\ 1.6487 & 8.2436 \\ 2.7183 & 27.1828 \\ 3.4903 & 43.6293 \\ 7.3891 & 147.7811 \end{bmatrix}$$

$$\nabla f(x) = \nabla F(x) F(x) = \begin{bmatrix} 1.2840 & 1.6487 & 2.7183 & 3.4903 & 7.3891 \\ 3.2101 & 8.2436 & 27.1828 & 43.6293 & 147.7811 \end{bmatrix} \begin{bmatrix} -0.0838 \\ -0.1481 \\ -0.3792 \\ -0.5749 \\ -1.7864 \end{bmatrix} = \begin{bmatrix} -16.5888 \\ -300.8722 \end{bmatrix}$$

For Gauss-Newton

$$\nabla F(x) \nabla F(x)^T \, p = -\nabla F(x) F(x)$$

$$\begin{bmatrix} 1.2840 & 1.6487 & 2.7183 & 3.4903 & 7.3891 \\ 3.2101 & 8.2436 & 27.1828 & 43.6293 & 147.7811 \end{bmatrix} \begin{bmatrix} 1.2840 & 3.2101 \\ 1.6487 & 8.2436 \\ 2.7183 & 27.1828 \\ 3.4903 & 43.6293 \\ 7.3891 & 147.7811 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} 16.5888 \\ 300.8722 \end{bmatrix}$$

i.e., $p = \begin{bmatrix} 0.0381 \\ 0.0102 \end{bmatrix}$. The new estimate of the solution is

$$x \to x + p = \begin{bmatrix} 2.50 \\ 0.25 \end{bmatrix} + \begin{bmatrix} 0.0381 \\ 0.0102 \end{bmatrix} = \begin{bmatrix} 2.5381 \\ 0.2602 \end{bmatrix}$$

The iteration is continued.

HW-13: Apply Gauss-Newton method to $f(x_1, x_2) = \sum_{i=1}^{4} (2x_1 x_2 t_i - y_i)$ with data

$t = \begin{pmatrix} 2 & 3 & 6 & 7 \end{pmatrix}^T$

$y = \begin{pmatrix} 3 & 6 & 8 & 9 \end{pmatrix}^T$

Use an initial guess $x = \begin{pmatrix} 2 \\ 0.5 \end{pmatrix}$

14. Other Algorithms than Gradients such as the Genetic Algorithm.

Genetic Algorithm is used to find solution to a problem called objective function.

Solution generated by genetic algorithm is called a chromosome.

Collection of chromosome is called population.

A chromosome is composed of genes and its value can be either numerical, binary, symbols or characters depending on the problem to be solved.

These chromosomes will undergo a process called fitness function to measure the suitability of solution generated by Genetic Algorithm with problem.

Some chromosomes in population will mate through process called crossover thus producing new chromosomes named offspring which its genes composition are the combination of their parent.

In a generation, a few chromosomes will also go through mutation in their gene.

The number of chromosomes which will undergo crossover and mutation is controlled by crossover rate and mutation rate value.

Chromosome in the population that will maintain for the next generation will be selected based on the rule that the chromosome which has higher fitness value will have greater probability of being selected again in the next generation.

After several generations, the chromosome value will converge to a certain value which is the best solution for the problem.

Thus, the genetic algorithm process is as follows

Step 1. Determine the number of chromosomes, generation, and mutation rate and crossover rate value

Step 2. Generate chromosome-chromosome number of the population, and the initialization value of the genes chromosome-chromosome with a random value

Step 3. Process steps 4-7 until the number of generations is met

Step 4. Evaluation of fitness value of chromosomes by calculating objective function

Step 5. Chromosomes selection

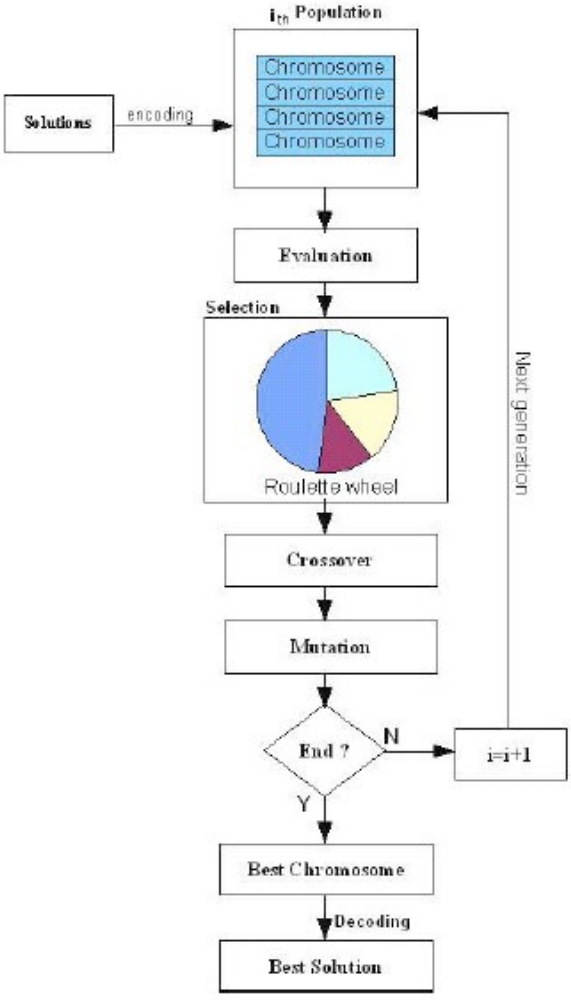
Step 6. Crossover

Step 7. Mutation

Step 8. New Chromosomes (Offspring)

Step 9. Solution (Best Chromosomes)

The flowchart of algorithm is



Example: For the equality $a + 2b + 3c + 4d = 30$, Find a, b, c and d by using genetic algorithm.

Formulate the objective function. The objective is minimizing the value of function $f(x)$ where $f(x) = ((a + 2b + 3c + 4d) - 30)$. Since there are four variables in the equation, namely a, b, c and d, we can compose the chromosome as follows:

To speed up the computation, we can restrict the values of variables a, b, c and d to integers between 0 and 30.

Step (i). Initialization

For example we define the number of chromosomes in population are 6, then we generate random value of gene a, b, c and d for 6 chromosomes

Chromosome[1] = [a;b;c;d] = [12;05;23;08], Chromosome[2] = [02;21;18;03], Chromosome[3] = [10;04;13;14], Chromosome[4] = [20;01;10;06], Chromosome[5] = [01;04;13;19], Chromosome[6] = [20;05;17;01]

Step (ii). Evaluation

We compute the objective function value for each chromosome produced in initialization step:

F_obj[1] = 12 + 2x5 + 3x23 + 4x8 - 30 = 93

F_obj[2] = 2 + 2x21 + 3x18 + 4x3 – 30 = 80

F_obj[3] = 10 + 2x4 + 3x13 + 4x14 - 30 = 83

F_obj[4] = 20 + 2x1 + 3x10 + 4x6 - 30 = 46

F_obj[5] = 1 + 2x4 + 3x13 + 4x19 - 30 = 94

F_obj[6] = 20 + 2x5 + 3x17 + 4x1 - 30 = 55

Step (iii). Selection 1.

The fittest chromosomes have higher probability to be selected for the next generation.

To compute fitness probability we compute the fitness of each chromosome.

To avoid divide by zero problem, the value of F_obj is added by 1.

Fitness[1] = 1 / (1+F_obj[1]) = 1 / 94 = 0.0106

Fitness[2] = 1 / (1+F_obj[2]) = 1 / 81 = 0.0123

Fitness[3] = 1 / (1+F_obj[3]) = 1 / 84 = 0.0119

Fitness[4] = 1 / (1+F_obj[4]) = 1 / 47 = 0.0213

Fitness[5] = 1 / (1+F_obj[5]) = 1 / 95 = 0.0105

Fitness[6] = 1 / (1+F_obj[6]) = 1 / 56 = 0.0179

Total = 0.0106 + 0.0123 + 0.0119 + 0.0213 + 0.0105 + 0.0179 = 0.0845

The probability for each chromosomes is formulated by P[i] = Fitness[i] / Total

P[1] = 0.0106 / 0.0845 = 0.1254

P[2] = 0.0123 / 0.0845 = 0.1456

P[3] = 0.0119 / 0.0845 = 0.1408

P[4] = 0.0213 / 0.0845 = 0.2521

P[5] = 0.0105 / 0.0845 = 0.1243

P[6] = 0.0179 / 0.0845 = 0.2118

From the probabilities above we see that Chromosome 4 that has the highest fitness, this chromosome has highest probability to be selected for next generation chromosomes.

For the selection process we use roulette wheel, for that purpose we compute the cumulative probability values:

C[1] = 0.1254

C[2] = 0.1254 + 0.1456 = 0.2710

C[3] = 0.1254 + 0.1456 + 0.1408 = 0.4118

C[4] = 0.1254 + 0.1456 + 0.1408 + 0.2521 = 0.6639

C[5] = 0.1254 + 0.1456 + 0.1408 + 0.2521 + 0.1243 = 0.7882

C[6] = 0.1254 + 0.1456 + 0.1408 + 0.2521 + 0.1243 + 0.2118 = 1.0

The process now is to generate random number R in the range 0-1 as follows:

R[1] = 0.201, R[2] = 0.284, R[3] = 0.099, R[4] = 0.822, R[5] = 0.398, R[6] = 0.501.

If random number R [1] is greater than C [1] and smaller than P [2] then select Chromosome [2] as a chromosome in the new population for next generation:

NewChromosome[1] = Chromosome[2]
NewChromosome[2] = Chromosome[3]
NewChromosome[3] = Chromosome[1]
NewChromosome[4] = Chromosome[6]
NewChromosome[5] = Chromosome[3]
NewChromosome[6] = Chromosome[4]

Chromosomes in the population thus become:

Chromosome[1] = [02;21;18;03], Chromosome[2] = [10;04;13;14], Chromosome[3] = [12;05;23;08], Chromosome[4] = [20;05;17;01], Chromosome[5] = [10;04;13;14], Chromosome[6] = [20;01;10;06]

Step (iv). Crossover

In this example, we use one-cut point, i.e. randomly select a position in the parent chromosome then exchanging sub-chromosome.

Parent chromosome which will mate is randomly selected and the number of mate Chromosomes is controlled using crossover_rate (ρc) parameters.

Pseudo-code for the crossover process is as follows:

begin k = 0
while (k < ρc)
then select Chromosome[k] as parent
end
k = k + 1
end
end

Chromosome k will be selected as a parent if R [k] < Chromosome[4] Chromosome[4] >< Chromosome[5] Chromosome[5] >< Chromosome[1]

After chromosome selection, the next process is determining the position of the crossover point. This is done by generating random numbers between 1 to (length of Chromosome – 1).

In this case, generated random numbers should be between 1 and 3.

After we get the crossover point, parents Chromosome will be cut at crossover point and its gens will be interchanged.

For example we generated 3 random number and we get: C[1] = 1 C[2] = 1 C[3] = 2

Then for first crossover, second crossover and third crossover, parent's gens will be cut at gen number 1, gen number 1 and gen number 3 respectively, e.g. Chromosome[1] = Chromosome[1] >< Chromosome[4] = [02;21;18;03] >< [20;05;17;01] = [02;05;17;01] Chromosome[4] = Chromosome[4] >< Chromosome[5] = [20;05;17;01] >< [10;04;13;14] = [20;04;13;14] Chromosome[5] = Chromosome[5] >< Chromosome[1] = [10;04;13;14] >< [02;21;18;03] = [10;04;18;03]

Thus Chromosome population after experiencing a crossover process: Chromosome[1] = [02;05;17;01] Chromosome[2] = [10;04;13;14] Chromosome[3] = [12;05;23;08] Chromosome[4] = [20;04;13;14] Chromosome[5] = [10;04;18;03] Chromosome[6] = [20;01;10;06]

Step (v).

Mutation Number of chromosomes that have mutations in a population is determined by the mutation rate parameter.

Mutation process is done by replacing the gen at random position with a new value.

The process is as follows:

First we calculate the total length of gen in the population.

In this case the total length of gen is total_gen = number_of_gen_in_Chromosome x number of population = 4 x 6 = 24

Mutation process is done by generating a random integer between 1 and total_gen (1 to 24).

If generated random number is smaller than mutation_rate(ρm) variable then marked the position of gen in chromosomes.

Assume that we define ρm as 10%, it is expected that 10% (0.1) of total_gen in the population that will be mutated:

number of mutations = 0.1 x 24 = 2.4 ≈ 2

Assume that the generation of random number yield 12 and 18 then the chromosome which have mutation are Chromosome number 3 gen number 4 and Chromosome 5 gen number 2. The value of mutated gens at mutation point is replaced by random number between 0-30.

Assume that the generated random number are 2 and 5 then Chromosome composition after mutation are: Chromosome[1] = [02;05;17;01] Chromosome[2] = [10;04;13;14] Chromosome[3] = [12;05;23;02] Chromosome[4] = [20;04;13;14] Chromosome[5] = [10;05;18;03] Chromosome[6] = [20;01;10;06]

After finishing the mutation process, we then have one iteration or one generation of the genetic algorithm.

We can now evaluate the objective function after one generation:

Chromosome[1] = [02;05;17;01] F_obj[1] = 2 + 2x5 + 3x17 + 4x1 - 30 = 37
Chromosome[2] = [10;04;13;14] F_obj[2] = 10 + 2x4 + 3x13 + 4x14 – 30 = 77
Chromosome[3] = [12;05;23;02] F_obj[3] = 12 + 2x5 + 3x23 + 4x2 – 30 = 47
Chromosome[4] = [20;04;13;14] F_obj[4] = 20 + 2x4 + 3x13 + 4x14 – 30 = 93
Chromosome[5] = [10;05;18;03] F_obj[5] = 10 + 2x5 + 3x18 + 4x3 - 30 = 56
Chromosome[6] = [20;01;10;06] F_obj[6] = 20 + 2x1 + 3x10 + 4x6 - 30 = 46

From the evaluation of new Chromosome we see that the objective function is decreasing, this means that we have better Chromosome or solution compared with previous Chromosome generation.

New Chromosomes for next iteration are: Chromosome[1] = [02;05;17;01], Chromosome[2] = [10;04;13;14], Chromosome[3] = [12;05;23;02], Chromosome[4] = [20;04;13;14], Chromosome[5] = [10;05;18;03], Chromosome[6] = [20;01;10;06].

These new Chromosomes will undergo the same process as the previous generation of Chromosomes such as evaluation, selection, crossover and mutation and at the end it will produce new generation of Chromosomes for the next iteration.

This process will be repeated until a predetermined number of generations.

For this example, after running 50 generations, best chromosome is obtained as Chromosome = [07; 05; 03; 01].

This means that: a = 7, b = 5, c = 3, d = 1.

If we use the number in the problem equation a +2 b +3 c +4 d = 30 7 + (2 x 5) + (3 x 3) + (4 x 1) = 30.

Thus, the values of the variables a, b, c and d generated by genetic algorithm satisfy the equality.