Chapter 2 : How To Solve An Equation Numerically

There are six(actually five, but one of them is a special case of another) main methods we will be using to solve equations numerically :

1. Bisection Method
2. False-Position(Regula-Falsi) Method
3. Fixed Point Iteration
4. Newtons(Newton-Raphson) Method
5. The Seacant Method
6. Accelarated Newton Method

Bisection Method

Recall

> if f ∈ c[a,b] and f(a).f(b) < 0, ∃ at least one r ∈ c(a,b) such that f(r) = 0.

By this principle, we know that if the sign of f changes over the domain, then we know that there is a root. We can find this root if we keep halving the interval, until we eventually zero in on the root.

$$f \in [a, b]; [a, b] = [a_0, b_0]$$

$$f(a_0) * f(b_0) < 0$$

Then the first iteration is $C_0 = \frac{a_0+b_0}{2} \to f(C_0)$

if $f(C_0) < 0 \to [a_1, b_1] = [a_0, C_0]$

and the second iteration is $C_1 = \frac{a_1, b_1}{2}$

The general formula for calculating the nth iteration using the bisection method is

$$C_n = \frac{a_n, b_n}{2}$$

And the **upper bound of error** for the bisection method is given by :

$$\frac{b-a}{2^{n+1}}$$

There are 4 ways we stop when using the bisection method :

1. We reach the desired number of iterations.
2. We have reached a certain **accuracy**.

   The accuracy in the bisection method is given by :

   $$|C_n - C_{n_1}| < \epsilon$$

   That is to say, the accuracy is the difference between the last 2 successive iterations.
3. Stop when $f(C_n) < \epsilon$.
4. Stop when $\frac{|C_n - C_{n-1}|}{C_n} \leq \epsilon$.

The main advantage of the Bisection method is that C always converges, and it always converges to the true root. The main disadvantage is that it is really slow, ie, it takes a lot of time and iterations to get within a reasonable degree of error of the root.

Bisection Method Proof

Prove that $\lim_{n\to\infty} C_n = r$

$b_1 - a_1 = \frac{b-a}{2}$ --> their length is equal when we bisect

$b_2 - a_2 = \frac{b_1-a_1}{2} = \frac{b-a}{2^2}$

$b_3 - a_3 = \frac{b_2-a_2}{2} = \frac{b-a}{2^3}$

Then we can say that

$b_n - a_n = \frac{b-a}{2^n}$

and

$|r - C_n| < \frac{b_n-a_n}{2}$ --> the difference between the real root and the iteration is less than the length.

and

$0 < |r - C_n| < \frac{b-a}{2^{n+1}}$ --> the diffrence is less than the upper bound of error and bigger than 0.

then by the sandwich theorem, since $\lim_{n\to\infty} \frac{b-a}{2^{n+1}} = 0$

and $\lim_{n\to\infty} = 0$,

then $\lim_{n\to\infty} r - C_n = 0$,

therefore, $r = \lim_{n\to\infty} C_n$.

Calculating The Number Of Iterations

since we know that the upper bound for the error is given by :

$$\frac{b-a}{2^{n+1}}$$

and that the upper bound of the error is higher than the real error, then we can say $\frac{b-a}{2^{n+1}} < \frac{\epsilon}{1}$.

$\frac{2^{n+1}}{b-a} > \frac{1}{\epsilon} = 2^{n+1} > \frac{b-a}{\epsilon}$.

take ln of both sides $ln(2^{n+1}) > ln(\frac{b-a}{\epsilon})$

--> $(n+1)ln(2) > ln(\frac{b-a}{\epsilon})$ --> $n+1 > \frac{ln(\frac{b-a}{\epsilon})}{\ln(2)}$ --> $n > \frac{ln(\frac{b-a}{\epsilon})}{\ln(2)} - 1$

Generally then, to find the **minimum** number of iterations for the bisection method to reach a certain accuracy, we use the inequality

$$n > \frac{ln(\frac{b-a}{\epsilon})}{\ln(2)} - 1$$

or

$$n > \log_2(\frac{b-a}{\epsilon}) - 1$$

Converting To f(x) = 0

Eg Estimate $\sqrt[4]{7}$ in the range `[1,2]`

let $x = \sqrt[4]{7}$

$x^4 = 7$

$x^4 - 7 = 0$

$f(x) = x^4 - 7$

Always convert to the form f(x) `= 0` when attempting to solve numerical methods.

Eg Find the intersection point between $y = e^x, y = x^2 + 5$

let $y = y$

$e^x = x^2 + 5$

$e^x - x^2 - 5 = 0$

$f(x) = e^x - x^2 - 5$

Eg $p(x) = e^x + x^2 - \sqrt{x} + 1$ is a profit function. Estimate the number of units that leads to maximum profit in [a,b]

$p'(x) = e^x + 2x - \frac{1}{(2)(\sqrt{x})}$

and find $p'(x) = 0$

False Position Method

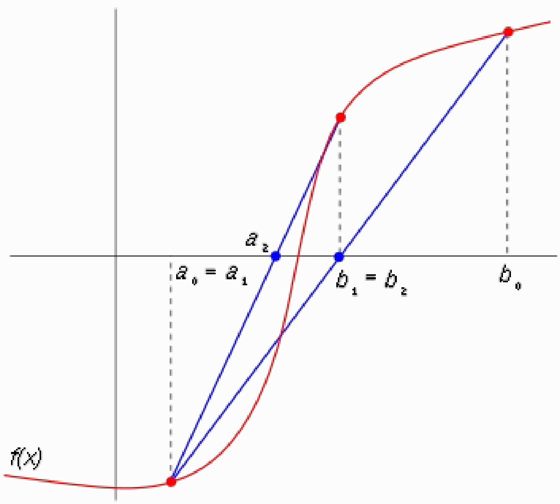
It is similar to bisection, and needs

$f(x) = 0, [a, b]$

where $f(a) * f(b) < 0$

But the value of C_n is not of $f(\frac{b_n - a_n}{2})$

It relies on a geometric method using the **seacant line** between a_n, b_n



Mathemtically, this is expressed as :

$$C_n = b_n - \frac{f(b_n)(b_n - a_n)}{f(b_n) - f(a_n)}$$

then $C_0 = b_0 - \frac{f(b_0)(b_0 - a_0)}{f(b_0) - f(a_0)}, f(C_0)$

and $C_1 = b_1 - \frac{f(b_1)(b_1 - a_1)}{f(b_1) - f(a_1)}, f(C_1)$

so on and so forth.

The slope of the resultant line is given by

$$S = \frac{f(b_0) - f(a_0)}{b_0 - a_0}$$

then $S = \frac{f(b_0) - f(a_0)}{b_0 - a_0} = \frac{0 - f(b_0)}{C_0 - b_0}$

$\frac{C_0 - b_0}{-f(b_0)} = \frac{b_0 - a_0}{f(b_0) - f(a_0)}$

$C_0 - b_0 = \frac{f(b_0)(b_0 - a_0)}{f(b_0) - f(a_0)}$

$C_0 = b_0 - \frac{f(b_0)(b_0 - a_0)}{f(b_0) - f(a_0)}$

We determine the next [a,b] depending on the sign of C_n

if $C_n < 0$, then $[a_{n+1}, b_{n+1}] = [C_0, b_n]$, given that $f(a) < 0$and $f(b) > 0$

This method always converges to the true root, but it is very slow as well.

Fixed Point Iteration

Say we have a function $f(x)$ with root p. The general idea is that we have some function taken from $f(x)$, called $g(x)$. We used the **fixed point** of $g(x)$, which will be the roots of $f(x)$. Not all functions derived are suitable however.

What is a Fixed Point?

$x = p$ is a fixed point of $g(x)$ if $g(p) = p$

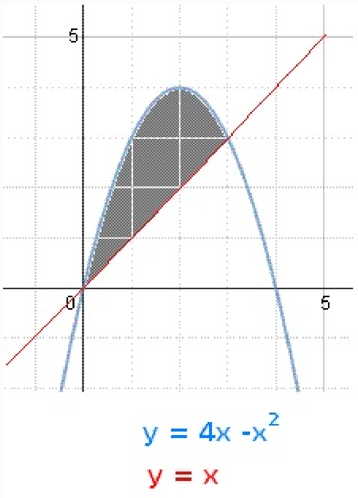
For example, given the function $g(x) = x^2$, then $0, 1$ are fixed points of $g(x)$. For more complex functions, we solve $g(x) = x$, since we want all images of x that are equal to x. so in this case :

$x^2 = x$

$x(1 - x) = 0$

$x = 0, 1$

Geomertrically speaking, the fixed points of $g(x)$ are the intersection points between $g(x)$ and $y = x$, for example,for $g(x) = 4x - x^2$:



In this case, $0, 3$ are fixed points of $4x - x^2$

Where do we get g(x) from?

We need to bring $f(x) = 0$ to the form $x - g(x) = 0$, for example $f(x) = x^2 - 5x + 6 = 0$ becomes $x^2 = 5x - 6 \to x = \sqrt{5x - 6} \to x - \sqrt{5x - 6}$, so in this case $g(x) = \sqrt{5x - 6}$. We can extract more than one $g(x)$ from $f(x)$, for example, we can extract $x = \frac{x^2 + 6}{5}$ by reearanging as well.

We can do this reerangement because we assumed that $f(x) = 0$

Why are Fixed Points of g(x) Roots of f(x)?

let $g(p) = p$

then $p - g(p) = 0$

but $p - g(p) = f(p)$

and since $p - g(p) = 0$, then $p - g(p) = f(p) = 0$ and p is a root of $f(x)$

FPI is a method where we find approximations of the fixed points of $g(x)$. we need :

1. $g(x)$.
2. p_0, or the initial value/guess value.

The formula of FPI is given by :

$$p_{n+1} = g(p_n)$$

then, $p_1 = g(p_0)$, $p_2 = g(p_1)$, so on and so forth.

If the sequence $p_0, p_1, p_2, \ldots p_n$ convrges to some p, then p is a fixed point of $g(x)$, and a root of $f(x)$

This can be proven as follows :

$\lim_{n\to\infty} p_n = g(p_{n-1})$

$p_\infty = g(p_{\infty-1}) = g(p_\infty)$

Existance, Uniqueness, and Convergance of Fixed Points

Given a function $g(x)$ and an interval $[a, b]$, then if

1. $g(x)$ continious on $[a, b]$
2. $a \leq g(x) \leq b \forall x \in [a, b]$

   then $g(x)$ has **at least one** fixed point in $[a, b]$, that is $\exists p \in [a, b]$ where $g(p) = p$

   furethermore, if
3. $|g'(x)| \leq k \leq 1 \forall x \in [a, b]$, then
   1. The fixed point is unique
   2. The FPI of $g(x)$ will converge to p for any $p_0 \in [a, b]$, where k is the maximum of $|g'(x)|$.

Proof

If 1 and 2 are satisifed, then we need to to show that g has a fixed point in $[a, b]$

Let us assume 3 cases :

1. if $g(a) = a$, then we are done.
2. if $g(b) = b$, then we are done.
3. if $g(a) \neq a$ and $g(b) \neq b$:

   We know that $a \leq g(x) \leq b$, and since we know that $g(a) \neq a$ and $g(b) \neq b$, then we know that $g(a) > a$ and $g(b) < b$, then we apply bolazano on $h(x) = g(x) - x$.

   We know that h is continous and also $h(a) = g(a) - a > 0$ and $h(b) = g(b) - b < 0$, Then $\exists p \in [a, b]$ such that $h(p) = 0$. Since $h(x) = g(x) - x$, and $h(x) = 0$, then $g(x) - x = 0$ so $g(x) = x$ exists.

Secondly, we need to prove uniqness and convergence

1. Uniquness

   From the previous proof, we have proved that p exists. Let us assume that there are other fixed points for g in $[a, b]$, called q, then

   $g(p) = p$

   $g(q) = q$

   lets apply the mean value theorem on $[p, q] \leq [a, b]$. We know that g is continous on $[p, q]$ because it is continous on $[a, b]$, Therefore, $\exists c \in (p, q)$ such that $g'(c) = \frac{g(q)-g(p)}{q-p} = 1$, therefore $|g'(c)| = 1$ which is a contradiction, since $max(g'(x)) < 1$, so no other point exists.

2. Convergence

   Apply MVT on $[p_0, p]$ therefore $\exists C \in (p_0, p)$ such that

   $g'(c) = \frac{g(p)-g(p_0)}{p-p_0}$

   $g'(c) = \frac{p-p_1)}{p-p_0}$

   $|p - p_1| = |g'(c)||p - p_0|$

   $|p - p_1| \leq k.|p - p_0|$

   This means that p_1 is closer to p that p_0.

   Apply the MVT on $[p_1, p]$ --> $|p - p_2| \leq k.|p - p_1|$ in a similar way.

   So, $|p - p_2| \leq k^2.|p - p_0|$

   $|p - p_3| \leq k^3.|p - p_0|$

   Generally,

   $$|p - p_n| \leq k^n.|p - p_0|$$

   Which is also an upper bound for the error.

Error in Fixed Point Iteration

Error in FPI is given by the difference between each 2 successive iterations. So we express this as

$$\epsilon \leq |p_{n+1} - pn|$$

Another way to express error is

$$\frac{k^n |p_1 - p_0|}{1-k} < \epsilon$$

So we can theoretically calculate the number of iterations using

$$n > ln(\frac{\epsilon(1-k)}{p_1 - p_0})$$

Convergence To Fixed Points

Let p be the fixed point of $g(x)$

Can i prove that the FPI of g will go to p before solving?

1. if $|g'(p)| < 1$, then the FPI of $g(p)$ will converge to p for any p_0 close to p. We call this an **attractive fixed point**.
2. if $|g'(p)| > 1$, then the FPI of $g(p)$ will not converge to p. We call this a **repulsive fixed point**.
3. if $|g'(p)| = 1$, then we cannot guarantee any outcome.

It is not nesecarry for p_0 approaches p from one side. It could be approached from both sides, which is called **oscilating convergence**. If it is however, only approached from one side, then that is called **monotonic convergence**.

In fact, if $-1 < g'(p) < 1$, Then, if $-1 < g'(p)$, then it is oscilating, if $g'(p) > 1$, then it is monotonic. Otherwise, we can make no guarantees.

Newton-Raphson(or Newtons) Method

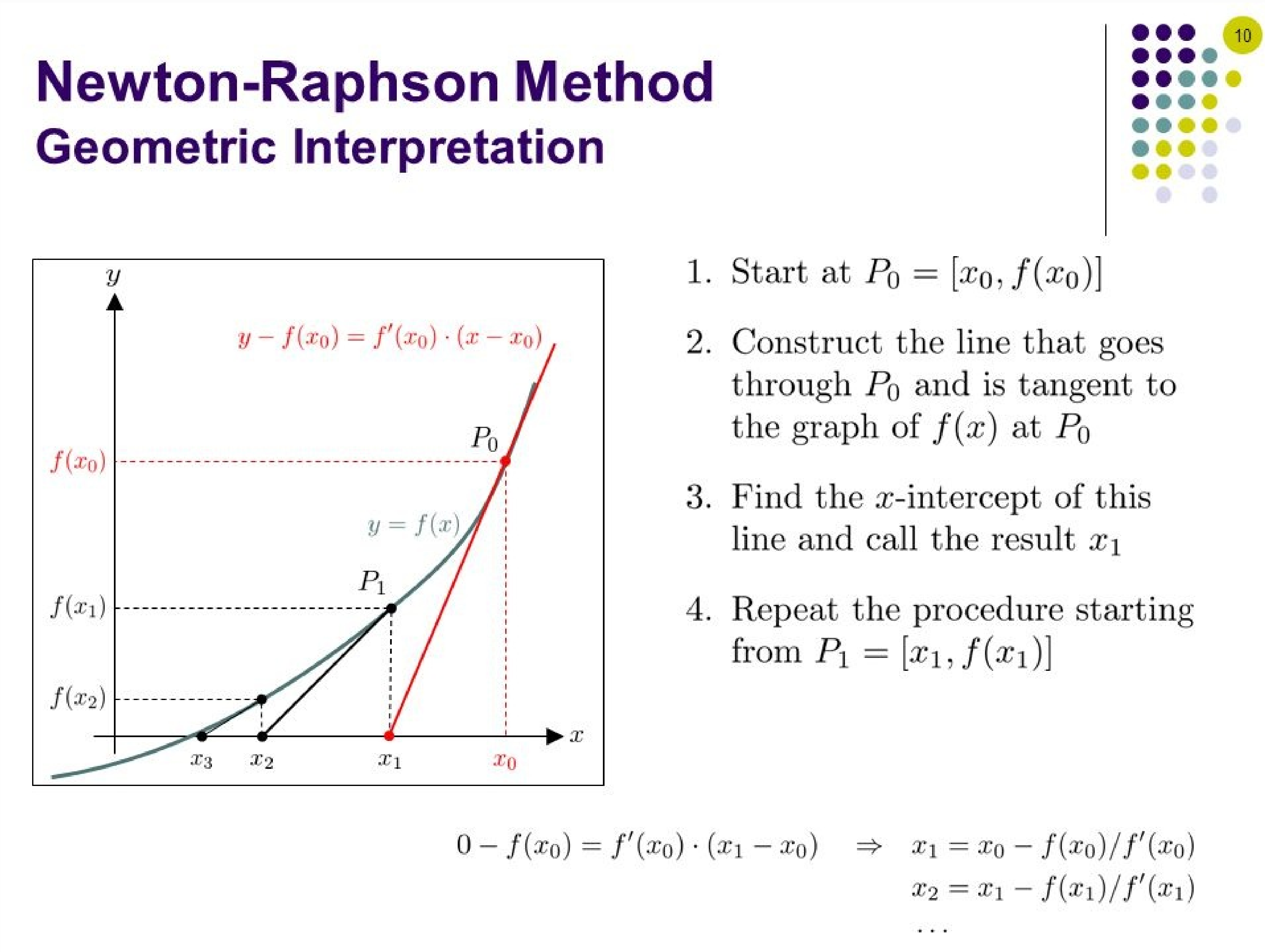
Say we have $f(x) = 0$, and some P_0

Then we can say

$$P_{n+1} = P_n - \frac{f(P_n)}{f'(P_n)}$$

Newtons-Raphson is a special case/example of Fixed Point Iteration. Threfore, any theorem that applies to FPI applies to the Newton-Raphson method(including error, convergence, and attractiveness)

Geometrically, We can express it as follows

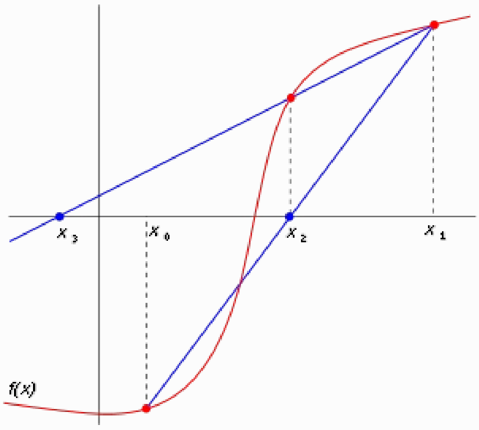


Secant Method

Say we have $f(x) = 0$, some P_0, and some P_1, then

$$P_{n+1} = P_n - \frac{f(P_n)(P_n - P_{n-1})}{f(P_n) - f(P_{n-1})}$$

Geometrically, this can be expressed as



Where P_{n+1} is the point where the secant line of P_n, P_{n-1} intersects the x-axis.

Measuring Speed (Order Of Convergence)

The order of convergence R, is a measure of how fast a method converges to the root. It is a positive number, the higher it is, the faster the method is, that is, the error between subsequent interations decreases faster.

The value of R depends on the type of root.

Multiplicty of Roots

Multiplicity is the number of times a root is repeated. That is, if we have a function that has the roots $(1, -2, 1)$, then 1 has a multiplicity of 2, and 2 has a multiplicity of 1.

Formally, let p be a root of $f(x)$. if

$$f(p) = f'(p) = f"(p) \ldots f^{(M-1)}(p) = 0$$

but

$$f^{(M)}(p) \neq 0$$

, then we say that the root p has multiplicity M. The smallest value of M is 0. A root with $M = 1$ is called a **simple root**. if $M > 1$, then the root is called a **multiple root**. if $M = 2$, it is called a **double root**, so on, and so forth.

Another defenition, that not always works but is useful sometimes, is that let p be a root of $f(x)$. This root has multiplicity M if we can write

$$f(x) = (x-p)^M . h(x); h(p) \neq 0$$

where $h(x)$ is some arbitriary function.

The secant and newton methods are fast for roots with $M = 1$

Given a sequece of iterations $[P_n]_{n=0}^{\infty}$ that converges to p. And $|E_n| = |p - p_n|$. If the convergence is fast, then E decreases faster.

Now, if $\exists$ two positve real numbers A, R such that

$$\lim_{n \to \infty} \frac{|E_{n+1}|}{|E_n|^R} = A$$

Then we say that the sequence converges to p with **order of convergence** R and A is called the **asymptotic error constant**. Usually, $A < 1$.

The above limit means that when n is large, then the value of $\frac{|E_{n+1}|}{|E_n|^R} \approx A$. Rearranging, we get

$$|E_{n+1}| \approx A|E_n|^R$$

That is, when R increases, then it converges faster (error decreases faster).

Note that :

1. If $R = 1$, then the convergence is called linear.
2. If $R > 1$, then the convergence is called quadratic, cubic, etc.
3. if $1 < R < 2$, then the convergence is superlinear.

Secant Method Convergence

Remember that

$$P_{n+1} = P_n - \frac{f(P_n)(P_n - P_{n-1})}{f(P_n) - f(P_{n-1})}$$

If it converges to p, then we have 2 cases :

1. If p is a simple root, then $R = 1.618$, and $A = |\frac{f''(p)}{2f'(p)}|^{0.618}$

2. If p is a multiple root, then $R = 1$, and we cant find A theoretically, only numerically. This is usually done by finding the real root, and finding $|E_n| = |p - p_n|$, and using $A = \frac{|E_{n+1}|}{|E_n|^1}$

Bisection Method Convergence

$R = 1$ in all cases, and $A = 0.5$.

False Positon Method Convergence

$R = 1$ in all cases, and A has no theoretical value.

Fixed Point Iteration Convergence

Let P be a fixed point of $g(x)$. If $g'(p) = g''(p) = \ldots.. = g^{(k-1)}(p) = 0$, but $g^{(k)} \neq 0$ then the fixed point iteration will converge to p with $R = k$, $A = \frac{g^{(k)}(p)}{k!}$

Proof

Based on what is given , we need to show that

$$lim_{n\to\infty} \frac{E_{n+1}}{E_n^k} = |\frac{g^{(k)}(p)}{k!}|$$

Now, take taylor expansion of $g(x)$ about P, which is

$$g(x) = g(p) + g'(p)(x-p) + \frac{g''(p)(x-p)^2}{2!} \ldots . \frac{g^{(k)}(p)(x-p)^k}{k!}$$

We know however, that $g'(p) = g''(p) = \ldots.. = g^{(k-1)}(p) = 0$, so we end up with

$$g(x) = p + \frac{g^{(k)}(c)(x-p)^k}{k!}$$

let $x = P_n$

$$g(x) = p + \frac{g^{(c)}(p)(P_n - p)^k}{k!}$$

and so

$$|P_{n+1} - p| = \frac{|g^{(c)}(p)||(P_n - p)^k|}{k!}$$

which results in

$$|E_{n+1}| = \frac{|g^{(c)}(p)|}{k!}|E_n|^k$$

Therefore,

$$\frac{|E_{n+1}|}{|E_n|^k} = \frac{|g^{(c)}(p)|}{k!}$$

Where $P_n < c < P$

Newton's Method Convergence

Remember that

$$P_{n+1} = P_n - \frac{f(P_n)}{f'(P_n)}$$

If it converges to p, then we have 2 cases :

1. If p is a simple root($M = 1$), then $R = 2$ and $A = |\frac{f''(p)}{2f'(p)}|$. So,

$$\lim_{n\to\infty} \frac{|E_{n+1}|}{|E_n|^R} = |\frac{f''(p)}{2f'(p)}|$$

1. If p is a multiple root($M > 1$), then $R = 1$ and $A = \frac{M-1}{M}$

Proof

We need to prove that

$$lim_{n\to\infty} \frac{E_{n+1}}{E_n^k} = |\frac{f''(p)}{2f'(p)}|$$

Method 1

We can consider newton as a special case of FPI where $g(x) = x - \frac{f(x)}{f'(x)}$

Now, since it is a special case of FPI, use the proof of FPI.

$$g'(x) = 1 - \frac{(f'(x))^2 - f(x)f''(x)}{(f'(x))^2}$$

Simplifying, this results in

$$g'(x) = \frac{f(x) * f''(x)}{(f'(x))^2} = 0$$

Deriving again

$$g''(x) = \frac{(f'(x))^2 f(x) f'''(x) - f''(x) f'(x))}{f'(x))^4} THISEQUATIONMIGHTNOTBECOMPLETE$$

Substituting $x = p$, we get $g''(p) = \frac{f''(p)}{f'(p)}$

so by FPI theorems, $R = 2$, A=$|\frac{g''(p)}{2!}| = |\frac{f''(p)}{2f'(p)}|$, Which is what we want to demonstrate.

Method 2

If p is a simple root of $f(x)$ we want to prove that

$$\lim_{n\to\infty} \frac{|E_{n+1}|}{|E_n|^R} = |\frac{f''(p)}{2f'(p)}|$$

Apply taylor series to $f(x)$ about $x = p_n$ $f(x) = f(p_n) + f'(p_n)(x - p_n) + f''(c)\frac{(x-p_n)^2}{2!}$

let $x = p$

$f(p) = f(p_n) + f'(p_n)(p - p_n) + f''(c)\frac{(p-p_n)^2}{2!}$

$0 = f(p_n) + f'(p_n)(p - p_n) + f''(c)\frac{(p-p_n)^2}{2!}$

$0 = \frac{f(p_n)}{f'(p_n)} + p - p_n + \frac{f''(c)}{f'(p_n)}(p - p_n)^2$

$0 = p - (p_n - \frac{f(p_n)}{f'(p_n)}) + \frac{f''(c)}{2f'(p_n)}(p - p_n)^2$

$|p - p_{n+1}| = |\frac{f''(c)}{2f'(p_n)}|(p - p_n)^2$

$\frac{|E_{n+1}|}{|E_n|^2} = |\frac{f''(c)}{2f'(p_n)}|$

Accelarated Newton Method

For multiple roots, newton's method is slow. Given that $M > 1$,then we can change $R = 1$ to $R = 2$ by using the formula

$$P_{n+1} = P_n - \frac{Mf(P_n)}{f'(P_n)}$$

Summary Table

Method Name	Requirments	Iteration	Convergence
Newton	$p_0, f(p), f'(p)$	$p_{n+1} = p_n - \frac{f(p_n)}{f'(p_n)}$	$if \quad M = 1, R = 2, A = \frac{f''(p)}{2f'(p)}$,$if \quad M > 1, R = 1, A = \frac{M-1}{M}$
Seacant	$p_0, p_1, f(p)$	$p_{n+1} = p_n - \frac{f(p_n)(p_n - p_{n-1})}{f(p) - f(p_{n-1})}$	$if \quad M = 1, R = 1.618, A = \|\frac{f''(p)}{2f'(p)}\|^{0.618}, if \quad M > 1, R = 1,$ calculate p numerically and use $\frac{E_{n+1}}{E_n}$ to find A
Accelarated Newton	$p_0, f(p), f'(p), M$	$p_{n+1} = p_n - \frac{Mf(p_n)}{f'(p_n)}$	$R = 2$, A can only be calculated numerically
Bisection	$f(p), (a_0, b_0)$	$c_n = \frac{a_n + b_n}{2}$, choose (a_{n+1}, b_{n+1}) based on sign of $f(c_n)$	$R = 1, A = 0.5$
False Position	$f(p), (a_0, b_0)$	$c_n = b_n - \frac{f(b_n)(b_n - a_n)}{f(b_n) - f(a_n)}$, choose (a_{n+1}, b_{n+1}) based on sign of $f(c_n)$	$R = 1$,A can only be found numerically(as in seacant method)
Fixed Point Iteration	$p_0, g(x) = x$	$p_{n+1} = g(p_n)$	$R = k$, where k is the order of first nonzero derivative of p , $A = \|\frac{g^{(k)}(p)}{k!}\|$