Math 151A – Applied Numerical Methods I

University of California, Los Angeles

Duc Vu

Fall 2021

This is math 151A – Applied Numerical Methods taught by Professor Jiang. We meet weekly on MWF from 1:00 pm to 1:50 pm for lecture. The recommended textbook for the class is *Numerical Analysis* 10^{th} by *Burden*, *Faires* and *Burden*. Other course notes can be found at my blog site. Please let me know through my email if you spot any typos in the note.

Contents

List of Theorems

List of Definitions

§1 Lec 1: Sep 24, 2021

§1.1 Calculus Review

- Intermediate Value Theorem (IVT): For continuous function $C([a,b])$, let $f \in C([a,b])$. Let $k \in \mathbb{R}$ s.t. k is strictly between $f(a)$ and $f(b)$. Then, $\exists$ some $c \in (a,b)$ s.t. $f(c) = k$.

  Question 1.1. Why is IVT useful?

  It guarantees the existence of solution to some nonlinear equations.

Example 1.1

Let $f(x) = 4x^2 - e^x$. IVT tells us $\exists x^*$ s.t. $f(x^*) = 0$.

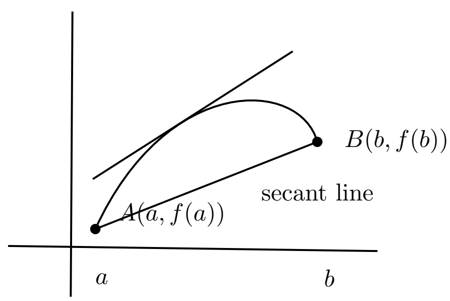
$$f(0) = 0 - e^0 = -1 < 0$$
$$f(1) = 4 - e > 0$$

With $k = 0$, by IVT, $\exists c \in (0,1)$ s.t. $f(c) = 0$.

- Mean Value Theorem (MVT): If $f \in C([a,b])$ and f is differentiable in (a,b), then $\exists c \in (a,b)$ s.t.

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

  in which $f'(c)$ is essentially the slope of the tangent line at $(c, f(c))$.



- Taylor's Theorem: Apply for a differentiable function, $f \in C^m([a,b])$ – f is m times continuously differentiable.

Theorem 1.2 (Taylor)

Let $f \in C^n([a,b])$. Let $x_0 \in [a,b]$. Assume $f^{(n+1)}$ exists on $[a,b]$. Then $\forall x \in [a,b]$, $\exists \xi(x) \in \mathbb{R}$ s.t. $x_0 < \xi < x$ or $x < \xi < x_0$. Then, we can express f as

$$f(x) = P_n(x) + R_n(x)$$

where

$$P_n(x) = f(x_0) + f'(x_0)(x - x_0) + f''(x_0)\frac{(x-x_0)^2}{2!} + \ldots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n$$

and

$$R_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x - x_0)^{n+1}$$

Example 1.3

$f(x) = \cos(x)$, $x_0 = 0$

$$\begin{aligned} f(x) = \cos(x) &= f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f'''(\xi(x))}{3!}x^3 \\ &= 1 + 0 - \frac{1}{2}x^2 + \frac{1}{6}x^3 \sin(\xi(x)) \end{aligned}$$

<u>*Note*</u>: Saying $f \in C^1$ is different from saying $f'(x)$ exists.

Example 1.4

Consider

$$f(x) = \begin{cases} x^2 \sin\left(\frac{1}{x}\right), & x \neq 0 \\ 0, & x = 0 \end{cases}$$

Have

$$\begin{aligned} f'(0) &= \lim_{h\to 0} \frac{f(0+h) - f(0)}{h} \\ &= \lim_{h\to 0} h \sin\left(\frac{1}{h}\right) \\ &= 0 \end{aligned}$$

But $f'(x)$ is not continuous. Specifically,

$$f'(x) = \begin{cases} 2x \sin\left(\frac{1}{x}\right) - \cos\left(\frac{1}{x}\right), & x \neq 0 \\ 0, & x = 0 \end{cases}$$

Take sequence $\frac{1}{2k\pi}$, $f' \to -1$ and $\frac{1}{(2k+1)\pi}$, $f' \to 1$. Thus, the function is not continuous as it converges to two different values.

§2 Lec 2: Sep 27, 2021

§2.1 Errors and Convergence Rate

Fact 2.1.

1. Computers have finite memory
2. Only a subset of rational numbers $\mathbb{Q}$ can be exactly represented/stored.
3. Working with floating numbers instead of reals produces round-off error

Definition 2.1 (Error) — Let $p \in \mathbb{R}$, $\tilde{p}$ approximate to p. We define absolute error as

$$e_{\text{abs}} := |p - \tilde{p}|$$

We define relative error as

$$e_{\text{rel}} := \left| \frac{p - \tilde{p}}{p} \right|$$

Example 2.2

- $p = 1$, $\tilde{p} = 0.9$. In this case,

$$e_{\text{abs}} = 0.1$$
$$e_{\text{rel}} = 0.1$$

- $p = 1000$, $\tilde{p} = 900$

$$e_{\text{abs}} = 100$$
$$e_{\text{rel}} = 0.1$$

Finite Digit Arithmetic

Example 2.3

- π is rounded/chopped by computers
- $x = \frac{5}{7} = 0.\overline{714285}$, $y = \frac{1}{3} = 0.\overline{3}$
  Let $fl(x)$ is the floating point approx. to x. For example, we assume 5 digit rounding.

$$fl(x) = 0.71428, \qquad fl(y) = 0.33333$$

  Say if we want to add $x + y$ on computer

$$fl\left(fl(x) + fl(y)\right) = fl\left(1.04761\right) = 1.0476$$

Example 2.4

$f(x) = x^3 - 6.1x^2 + 3.2x + 1.5$ where $x = 4.71$. The exact value of $f(x)$ at $x = 4.71$ is -14.263899. Let's assume 3 digit rounding

$$\begin{aligned} fl(x^2) &= fl(4.71 \cdot 4.71) = fl(22.1841) = 22.2 \\ fl(x^3) &= 105 \\ fl(3.2x) &= 15.1 \\ fl\left(f(4.71)\right) &= -13.4 \end{aligned}$$

The relative error here is approximately 6% which is huge. Our example has 7 floating point operations (FLOPs). In order to reduce the floating point error, we want to nest the function

$$f(x) = ((x - 6.1)x + 3.2)x + 1.5 \quad - 5 \text{ FLOPs}$$

So $fl\left(f(4.71)\right) = -14.3$ and the $e_{\text{rel}} = 0.25\%$.

Remark 2.5. Every operation introduces error.

Order of convergence for Sequences:

Definition 2.6 (Order of Convergence for Sequences) — For a convergent sequence $(p_n) = (p_1, p_2, p_3, \ldots)$. Let $p_n \to p$ as $n \to \infty$. Assume $p_n \neq p$. Then, if $\exists \lambda,\ \alpha$ with $0 < \lambda < \infty$ and $\alpha > 0$ s.t.

$$\lim_{n\to\infty} \frac{|p_{n+1} - p|}{|p_n - p|^\alpha} = \lambda$$

Then we say p_n converges to p with order α.

Example 2.7

$p_1 = 1, p_2 = \frac{1}{5}, \ldots, p_n = \frac{1}{5}p_{n-1}$ or $p_n = \frac{1}{5^{n-1}}$ where $p_n \to 0$ as $n \to \infty$.

$$\frac{|p_{n+1} - 0|}{|p_n - 0|^1} = \frac{\left(\frac{1}{5}\right)^n}{\left(\frac{1}{5}\right)^{n-1}} = \frac{1}{5}$$

So p_n converges with $\alpha = 1$

Problem 2.1. Test with $\alpha = 2$.

Definition 2.8 (Big O Notation) — We have $a(t) = \mathcal{O}\left(b(t)\right)$ where a is on the order of $b \iff$

$$\exists C > 0 \quad \ni \quad |a(t)| \leq Cb(t) \quad \text{for } t \to 0 \text{ or } t \to \infty$$

In practice, the definition is equivalent to

$$\lim_{t\to 0} \frac{|a(t)|}{b(t)} \text{ is bounded by a positive number}$$

§3 Lec 3: Sep 29, 2021

§3.1 Lec 2 (Cont'd)

Example 3.1
The Taylor's theorem for $\cos(h)$ about 0 is

$$\cos(h) = 1 - \frac{1}{2}h^2 + \frac{1}{24}h^4 \cos(\xi(h)) \text{ with some } 0 < \xi(h) < h$$

Denote $f(h) = \cos(h) + \frac{1}{2}h^2 - 1 = \frac{1}{24}h^4 \cos(\xi(h))$

$$\lim_{h \to 0} \frac{|f(h)|}{h^4} = \lim_{h \to 0} \frac{1}{24} |\cos(\xi(h))| \leq \frac{1}{24}$$

Thus y definition of big $\mathcal{O}$ notation,

$$f(h) = \mathcal{O}(h^4)$$

§3.2 Root Finding with Bisection

The goal is to find a root, or a zero, of a function f, i.e., find p s.t. $f(p) = 0$. First, let's assume

1. $f \in C([a,b])$
2. $f(a)f(b) < 0$

Then, $\exists p$ s.t. $f(p) = 0$ (by IVT).

Example 3.2
Consider:

$$f(x) = \sqrt{x} - \cos x, \qquad [a,b] = [0,1]$$

Then,

$$f(0) = -1, \qquad f(1) = 1 - \cos 1 > 0$$

Therefore, by IVT, $\exists p \in (0,1)$ s.t. $\sqrt{p} - \cos p = 0$.

Bisection Method (B.M): is an algorithm to approximate p s.t. $f(p) = 0$ on an interval $[a,b]$.
Algorithm 1: Bisection method (given $f(x) \in C([a,b])$, with $f(a)f(b) < 0$)

1. Set $a_1 = a$, $b_1 = b$
2. Set $p_1 = \frac{a_1+b_1}{2}$
3. *if* $f(p_1) == 0$ *then* we are done!
4. *else if* $f(p_1)$ has same sign as $f(a_1)$ *then* $p \in (p_1, b_1)$
   - Set $a_2 = p_1$, $b_2 = b_1$
5. *else if* $f(p_1)$ has same sign as $f(b_1)$ *then* $p \in (a_1, p_1)$
   - Set $a_2 = a_1$, $b_2 = p_1$
6. *end*

7. Set $p_2 = \frac{a_2+b_2}{2}$
8. Reset the entire if/else process.

> **Remark 3.3.** B.M. is similar to binary search in computer algorithms. If there exists multiple roots, e.g., $\{p, q, r\} \in [a, b]$, then the B.M. is guaranteed to find exactly one root, not all of them (but no guarantee exists for which one the method will find).

Stopping Criteria: We need a sequence $(p_1, p_2, \dots)$ and need specified tolerance ε. Choices for when to stop an algorithm:

- $|p_n - p_{n-1}| < \varepsilon$ – absolute difference between successive elements of the sequence
- $\frac{|p_n - p_{n-1}|}{|p_n|} < \varepsilon$ (assume $p_n \neq 0$) – relative difference
- $|f(p_n)| < \varepsilon$ – sometimes called a residual (how close are we to the answer).

§4 Lec 4: Oct 1, 2021

§4.1 Bisection Method (Cont'd)

Remark 4.1. B.M. is a global method, $f \in C([a, b])$ as long as the assumptions are satisfied, $f(a)f(b) < 0$, the B.M. will converge. In particular, it will converge to some point p s.t. $f(p) = 0$. Here "global" means the algorithm doesn't need a good initial guess p_0 unlike some "local" methods that we will cover later.

Example 4.2

The B.M. won't work for functions like $f(x) = x^2$ even though it has a root at $p = 0$ because we can't find any a, b that satisfies $f(a)f(b) < 0$.

Theorem 4.3 (Convergence Order of B.M.)

The sequence provided by B.M. satisfies

$$|p_n - p| \le \frac{b-a}{2^n}$$

which approaches to 0 as $n \to \infty$.

This further tells us that the error bound of B.M. converges linearly. Recall from previous lectures that linear convergence for a convergent sequence (p_n) means that

$$\lim_{n\to\infty} \frac{|p_{n+1} - p|}{|p_n - p|^1} = \lambda \quad \text{for some finite positive } \lambda$$

and

$$p_n = \frac{b-a}{2^n}, \quad p = 0$$

We can easily show that $\lambda = \frac{1}{2}$.

Remark 4.4. The B.M. converges slowly compared to other methods. We will soon see that Newton's method has quadratic order of convergence.

§4.2 Fixed Points

Definition 4.5 (Fixed Point) — Let function g be $g : [a, b] \to \mathbb{R}$ and $p \in [a, b]$ s.t. $g(p) = p$. Then p is a fixed point of g.

Theorem 4.6

Let p be a fixed point of g, then p is also a root of $G(x) := g(x) - x$.

Proof. Obvious by definition. □

Given a root-finding problem $f(p) = 0$, we can define functions g with a fixed point at p in a number of ways. For example, as $g(x) = x - f(x)$ or as $g(x) = x + 3f(x)$.
A fixed point for g just corresponds to the intersection between $y = g(x)$ and $y = x$.
Fixed Point Iteration (F.P.I): the F.P.I method is quite simple. For $g \in C([a, b])$ and $p_0 \in [a, b]$ we set $p_{n+1} = g(p_n)$. We also need $g(x) \in [a, b]$, otherwise at some point of the algorithm we won't be able to proceed to evaluate g. Also note that the initial guess p_0 is arbitrary.

$$p_1 = g(p_0), \quad p_2 = g(p_1), \ldots, p_{n+1} = g(p_n)$$

We use the same stopping criteria as in B.M.

Example 4.7 (F.P.I Failure Case)
To solve $x^2 - 7 = 0$, it is equivalent to $x = \frac{7}{x}$. We want to use F.P.I to find $p = \sqrt{7} \approx 2.64575\ldots$, so we can set

$$g_1(x) = \frac{7}{x}$$

then the goal is to find p s.t. $p = g_1(p)$. Another option is to use

$$g_2(x) := \frac{x + \frac{7}{x}}{2} = x$$

Let $p_0 = 3$ we can show that

- $g_1(x)$: $p_0 = 3$, $p_1 = \frac{7}{3}$, $p_2 = 3, \ldots$, oscillates between 2 numbers
- $g_2(x)$: $p_0 = 3$, $p_1 = 2.666\ldots$, $p_2 = 2.645833\ldots, \ldots$

Example 4.8
$x^3 + 4x^2 - 10 = 0$ has a unique root in $[1, 2]$, i.e. $p = 1.365230013$.

a) $x = g_1(x) = x - x^3 - 4x^2 + 10$ – does not converge

b) $x = g_2(x) = \left(\frac{10}{x} - 4x\right)^{\frac{1}{2}}$ – does not converge

c) $x = g_3(x) = \frac{1}{2}(10 - x^3)^{\frac{1}{2}}$ – converge

d) $x = g_4(x) = \left(\frac{10}{x+4}\right)^{\frac{1}{2}}$ – converge

e) $x = g_5(x) = x - \frac{x^3+4x^2-10}{3x^2+8x}$ – converge

We can see that the choice of $g(x)$ is critical to determine whether the algorithm converges. Before delving into that problem, let's first establish a theorem about the existence of a fixed point.

Theorem 4.9 (Existence of a Fixed Point)
Let $g \in C([a, b])$ with $a \leq g(x) \leq b$. Then, $\forall x \in [a, b]$, $\exists$ at least one fixed point p s.t. $g(p) = p$.

§5 Lec 5: Oct 4, 2021

§5.1 Fixed Point Iteration (Cont'd)

Recall

Theorem 5.1
Let $g \in C[a,b]$ with $a \le g(x) \le b \, \forall x \in [a,b]$, then $\exists$ at least one fixed point p s.t. $g(p) = p$.

Let's prove it.

Proof. First, we need to check if an end point is a fixed point, i.e., if

$$g(a) = a \quad \text{or} \quad g(b) = b$$

then we're done. Otherwise, let's define $G(x) \coloneqq g(x) - x$. Our goal is to use IVT to prove that G has a root. Then since $g \in [a,b]$, we know

$$G(a) = g(a) - a > 0, \quad G(b) = g(b) - b < 0$$
$$\implies G(a)G(b) < 0$$

Also, $G \in C([a,b])$. Therefore, by IVT, $\exists p$ s.t. $G(p) = 0$, i.e., $\exists p \in [a,b]$ s.t. $g(p) = p$. □

Remark 5.2. The theorem is just a sufficient condition for existence.

Theorem 5.3 (FPI Convergence with Lipschitz Continuity)
Assume $g \in C([a,b])$, $g \in [a,b]$ (*) and $\exists k \in (0,1)$ s.t.

$$|g(x) - g(y)| \le k\,|x - y|, \quad \forall x, y \in [a,b] \tag{**}$$

Then,

1. $\exists$ unique p s.t. $g(p) = p$.
2. The F.P.I $(p_{n+1} = g(p_n))$ will converge to p.
3. Error estimate: $|p_n - p| < k^n \max\{b - p_0, p_0 - a\}$.

Proof. 1. Let's prove by contradiction. Assume $\exists$ two different fixed points p and q, then

$$|g(p) - g(q)| = |p - q|$$

But by (**) we know

$$|g(p) - g(q)| \le k\,|p - q|$$

This implies that

$$|p - q| \le k\,|p - q|$$

which cannot be true since $p \ne q$ and $k \in (0,1)$ – contradiction!

2. + 3. By (**) we know that differences of g values are bounded. So let's try to convert this to something with g values. We know F.P.I $g(p_n) = p_{n+1}$ and also let p be the solution $g(p) = p$. So

$$|p_n - p| = |g(p_{n-1}) - g(p)|$$

By (**), we know

$$|p_n - p| = |g(p_{n-1}) - g(p)| \leq k\,|p_{n-1} - p|$$

Similarly,

$$k\,|p_{n-1} - p| = k\,|g(p_{n-2}) - g(p)| \leq k^2\,|p_{n-2} - p|$$

Recursively apply this until $n = 0$. Then,

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

Notice that

$$|p - p_0| \leq \max\{b - p_0, p_0 - a\}$$

Thus,

$$|p_n - p| \leq k^n \max\{b - p_0, p_0 - a\}$$

Since $k \in (0,1)$, this goes to 0.

□

Remark 5.4. Speed of convergence depends on k. The closer to 0 k is, the faster it converges.

In practice, to use the theorem, it is sometimes more useful to look at the derivatives instead of Lipschitz condition.

Theorem 5.5 (FPI Convergence with Bounded Derivative)
Assume (*) and $g \in C^1[a,b]$ and that $\forall x \in [a,b]$, $\exists k \in (0,1)$ s.t. $|g'(x)| \leq k$. Then, following the above theorem, F.P.I converges to the unique solution.

Proof. Here we need to prove that bounded derivative gives Lipschitz. Let's use MVT, $\exists c \in (a,b)$ s.t. $\forall x, y \in [a,b]$

$$g'(c) = \frac{g(x) - g(y)}{x - y}$$

Thus,

$$|g(x) - g(y)| = |g'(c)|\,|x - y| \leq k\,|x - y|$$ □

§6 Lec 6: Oct 6, 2021

§6.1 Newton's Method

Newton's Method (N.M.) is a classic technique used in science and engineering, research and industry all the time. There are many ways to derive it, and we will go over 3 today.
Analytic derivation with Taylor's polynomial:

Let $f \in C^2([a,b])$, p is a root ($f(p) = 0$). Suppose p_n is "close to" p, i.e., $|p_n - p|$ is "small".

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

If $|p - p_n|$ is "small", then $|p - p_n|^2$ is "really small". Up to an error of size $\approx (p - p_n)^2$,

$$0 = f(p) \approx f(p_n) + f'(p_n)(p - p_n)$$

So

$$p = p_n - \frac{f(p_n)}{f'(p_n)}$$

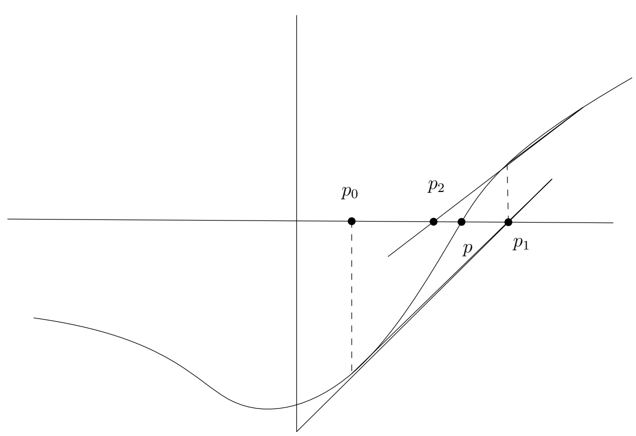
This can be used to "invent" Newton's method.

Definition 6.1 (Newton's Method) — Start with p_0 close to p, then do

$$p_{n+1} = p_n - \frac{f(p_n)}{f'(p_n)}$$

Remark 6.2. The initial guess p_0 must be close to p, otherwise the analytic derivation breaks down.

Graphical Derivation:



Tangent line: $y = ax + b$ and the intersection with the x-axis is p_{n+1}. We know

$$\begin{aligned} f(p_n) &= ap_n + b \\ 0 &= ap_{n+1} + b \\ a &= f'(p_n) \end{aligned}$$

The unknowns are a, b, p_{n+1}. Solving them we obtain

$$p_{n+1} = p_n - \frac{f(p_n)}{f'(p_n)}$$

Fixed Point Derivation Method:

Theorem 6.3

Let $g(x) := x - \frac{f(x)}{f'(x)}$ for some $f \in C^1([a,b])$ where $f'(x) \neq 0$ for $x \in [a,b]$. Then $g(p) = p$ if and only if $f(p) = 0$.

Proof. Basic algebra :D ☐

Define a fixed point iteration from g

$$p_{n+1} = g(p_n) = p_n - \frac{f(p_n)}{f'(p_n)}$$

Remark 6.4. We must have $f'(p_n) \neq 0 \ \forall n$, otherwise, N.M will fail.

Pros of N.M:

- It will converge faster than the B.M. to the root p of function $f(x)$ (when it does converge).

Cons of N.M.:

- Unlike the B.M., N.M. is a local method, not global. That means p_0 must be sufficiently close to p for success.
- N.M. requires knowledge of $f'(x)$ and evaluation of $f'(x)$ (especially when f is $\mathbb{R}^n \to \mathbb{R}^m$).

In higher dimension, if f is $\mathbb{R}^n \to \mathbb{R}^n$, then N.M. is

$$\mathbf{x}_{n+1} = \mathbf{x}_n - (\mathbf{J}(\mathbf{x}_n))^{-1} \mathbf{f}(\mathbf{x}_n)$$

where $\mathbf{J}(x)$ is the Jacobian matrix and $J_{ii} = \frac{\partial f_i}{\partial x_j}(\mathbf{x})$.

§6.2 Secant Method

N.M. requires the knowledge of $f'(x)$ and evaluation of $f'(x)$, so we can approximate it as follows

$$f'(p_n) \approx \frac{f(p_n) - f(p_{n-1})}{p_n - p_{n-1}}$$

This defines the **Secant Method**.

Definition 6.5 (Secant Method) — Given some p_0 and p_1, define

$$p_{n+1} = p_n - f(p_n)\frac{p_n - p_{n-1}}{f(p_n) - f(p_{n-1})}$$

Secant method is useful when you don't have access to $f'(x)$, e.g., when we don't have access to f.

§7 Lec 7: Oct 8, 2021

§7.1 Secant Method (Cont'd)

Recall Newton's Method (N.M.) is defined as follows

$$\text{Given } p_0, \quad p_{n+1} = p_n + \frac{f(p_n)}{f'(p_n)}$$

This requires evaluation of f'. In general, this could be expensive or unknown, e.g., in higher dimension or $f(x)$ comes from experimental data. The definition of derivative is

$$f'(x) := \lim_{h\to 0} \frac{f(x) - f(x-h)}{h}$$

So when h is small, the derivative can be approximated by "finite difference",

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

So if we let $x = p_n$, and $x - h = p_{n-1}$, then this becomes

$$f'(p_n) \approx \frac{f(p_n) - f(p_{n-1})}{p_n - p_{n-1}}$$

which holds true when $p_n - p_{n-1}$ is small. This leads us to the definition of secant method.

Definition 7.1 — Given p_0, p_1, secant method is defined as

$$p_{n+1} = p_n - f(p_n)\frac{p_n - p_{n-1}}{f(p_n) - f(p_{n-1})}$$

where the fraction is approximating $(f'(p_n))^{-1}$.

Question 7.1. How to get p_1?

e.g., running one iteration of bisection method.

§7.2 Local Convergence of Newton's Method

Theorem 7.2 (Newton Convergence)
Let $f \in C^2([a,b])$ and $p \in (a,b)$ s.t.

i) $f(p) = 0$

ii) $f'(p) \neq 0$

Then $\exists \delta > 0$ s.t. N.M. will converge for $\forall p_0 \in [p-\delta, p+\delta]$.

There is no guideline to find the exact δ – which means we don't know what close-enough means in practice unfortunately.

Proof. The idea here is to apply the F.P.I. theorem from previous lectures to some to-be-defined function g. What is g?
Key conditions to satisfy:

1. $\left[\hat{a}, \hat{b}\right] \to \left[\hat{a}, \hat{b}\right]$
2. g is C^1
3. g has bounded derivative with bound in $(0,1)$.

Define $g(x) \coloneqq x - \frac{f(x)}{f'(x)}$. N.M. on $f(x)$ is the same as F.P.I. on $g(x)$

$$p_{n+1} = p_n - \frac{f(p_n)}{f'(p_n)} \iff g(p_n) = p_{n+1}$$

Thus, we just need to show the three postulates about g.

2. $f \in C^2([a,b])$ so $f \in C([a,b])$ and $f' \in C([a,b])$ and $f'' \in C([a,b])$. Let's compute $g'(x)$

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

There exists a region $[p-\delta_1, p+\delta_1]$ in $[a,b]$ s.t. $f'(x) \neq 0$, so g' is continuous in $[p-\delta_1, p+\delta_1]$. This proves 2.

3. WTS: bounded derivative

$$\begin{aligned} g'(x) &= \frac{f(x)f''(x)}{(f'(x))^2} \\ g'(p) &= 0 \end{aligned}$$

Due to continuity of g' in $[p-\delta_1, p+\delta_1]$, there exists a region (with $0 < \delta < \delta_1$) s.t. $|g'(x)| \leq k$ in $[p-\delta, p+\delta]$ for any $k \in (0,1)$. This proves 3.

Lastly, let's show 1.

1. Need to prove g maps $[p-\delta, p+\delta]$ to $[p-\delta, p+\delta]$

$$|g(x) - p| = |g(x) - g(p)| = |g'(\xi)|\,|x-p| \leq k\,|x-p| < |x-p|$$

By M.V.T, $\exists \xi \in (x,p)$. N.M. on $f(x)$ is the same as F.P.I on $g(x)$.

Now, we proved that F.P.I converges to p for any $p_0 \in [p-\delta, p+\delta]$. Equivalently, N.M. converges for f at p. $\square$

Remark 7.3. δ cannot be a priori. In practice, we can

- begin with some $p_0 \in [a,b]$
- run several iterations of B.M. (a global method)
- switch to N.M.

§8 Lec 8: Oct 11, 2021

§8.1 Convergence Order Theorem

Let's begin with a fact.

Fact 8.1. Let $p = g(p)$ e a fixed point, F.P.I $g(p_n) = p_{n+1}$

- If $g'(p) \neq 0$, we get linear convergence (order of convergence $\alpha = 1$)
- If $g'(p) = 0$, we get quadratic convergence ($\alpha = 2$)

Theorem 8.1

Let $g \in C^1([a, b])$ with $|g'(x)| \leq k$ for some $0 < k < 1$. If $g'(p) \neq 0$, then F.P.I. converges to p linearly.

Proof. From F.P.I convergence theorem, we know that F.P.I converges in this case. So we just need to prove the linear order. Use M.V.T.:

$$p_{n+1} - p = g(p_n) - g(p) = g'(\xi)(p_n - p)$$

where ξ is between p_n and p.

$$\lim_{n\to\infty} \frac{|p_{n+1} - p|}{|p_n - p|} = \lim_{n\to\infty} |g'(\xi)| = |g'(p)| = k$$

in which k is a positive number that is smaller than 1. It's also easy to see that it only has linear convergence, e.g.,

$$\lim_{n\to\infty} \frac{|p_{n+1} - p|}{|p_n - p|^2} = \lim_{n\to\infty} |g'(\xi)| \frac{1}{|p_n - p|} = \infty$$ □

Theorem 8.2 (Convergence Order Theorem of FPI)

Let $g \in C^\alpha([a, b])$, $\alpha \geq 2$ is an integer. If

i) $g(p) = p$

ii) $g'(p) = g''(p) = \ldots = g^{(\alpha-1)}(p) = 0$

iii) $g^{(\alpha)}(p) \neq 0$

Then, F.P.I. converges $\forall p_0$ sufficiently close to p with order α.

Proof. First let's prove that $p_n \to p$. We can follow the procedure in the proof in lecture 7. Sketch: $g'(p) = 0$ and $g' \in C([a, b])$, $\exists \delta$ s.t.

$$|g'(x)| \leq k \in [p - \delta, p + \delta] \text{ for any } k \in (0, 1)$$

Also,

$$|g(x) - p| = |g(x) - g(p)| = |g'(\xi)| |x - p| \leq k |x - p| < |x - p|$$
$$p - \delta \leq g(x) \leq p + \delta$$

These conditions guarantee convergence for $p_n \to p$ by F.P.I Theorem. Next, let's prove that the order is α. Let $n = \alpha - 1$,

$$g(x) = g(x_0) + g'(x_0)(x - x_0) + g''(x_0)\frac{(x-x_0)^2}{2!} + \ldots + g^{(\alpha-1)}(x_0)\frac{(x-x_0)^{\alpha-1}}{(\alpha-1)!} + g^{(\alpha)}\left(\xi(x)\right)\frac{(x-x_0)^{\alpha}}{\alpha!}$$

where $\xi(x)$ is between x_0 and x is a general unknown. Let $x = p_n$ and $x_0 = p$,

$$g(p_n) = p + g^{(\alpha)}(\xi_n)\frac{(p_n - p)^{\alpha}}{\alpha!}$$

where $\xi_n := \xi(p_n)$ is between p_n and p. So

$$p_{n+1} = p + g^{(\alpha)}(\xi_n)\frac{(p_n - p)^{\alpha}}{\alpha!}$$

After some manipulation we get

$$\lim_{n\to\infty} \frac{|p_{n+1} - p|}{|p_n - p|^{\alpha}} = \lim_{n\to\infty} \left| \frac{g^{(\alpha)}(\xi_n)}{\alpha!} \right|$$

We know $g \in C^{\alpha}\left([a,b]\right)$, then

$$\lim_{n\to\infty} \frac{|p_{n+1} - p|}{|p_- p|^{\alpha}} = \frac{1}{\alpha!}\left| g^{(\alpha)}\left(\lim_{n\to\infty} \xi_n\right) \right|$$

Recall $\xi_n \in [p_n, p]$ or $[p, p_n]$, so $p_n \to p \implies \xi_n$ converges to p.

$$\lim_{n\to\infty} \frac{|p_{n+1} - p|}{|p_n - p|^{\alpha}} = \frac{1}{\alpha!}\left| g^{(\alpha)}(p) \right| := \alpha \in (0, \infty)$$

Note that from Extreme Value Theorem we know that continuous function in a bounded interval is bounded. □

$g(p_n) = p_{n+1}$ converges with order 2 (or better, if $g''(p) = 0$).

> **Remark 8.3.** Suppose that derivative vanishes at p, i.e., $f'(p) = 0$, then N.M. may
>
> 1. not converge at all
> 2. or converge very slowly (only linearly) depending on initial guess

If $p_n \to p$ and $f'(p) = 0$, then that timplies $f'(p_n) \approx 0$ for n large.

Example 8.4

Consider: $f(x) = x^2 \implies f'(x) = 2x$

$$f(0) = 0, \quad f'(0) = 0$$

0 is a double root of f. Have

$$f'(p) = f(p) = 0$$

and

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

which converges linearly (bad case).

§9 Lec 9: Oct 13, 2021

§9.1 Multiple Roots

Definition 9.1 (Multiple Root) — A root of $f(x) = 0$, p, is called a root of multiplicity m of $f \iff$ for $x \neq p$, there exists decomposition

$$f(x) = (x-p)^m q(x) \text{ where } \lim_{x\to p} q(x) \neq 0$$

If the multiplicity of a root p is 1, then p is called a simple root/zero.

Theorem 9.2
Let $f \in C^m([a,b])$, $p \in [a,b]$. Then p is a root of multiplicity $m \iff$

$$f(p) = f'(p) = f''(p) = \ldots = f^{m-1}(p) = 0 \text{ but } f^{(m)}(p) \neq 0$$

Example 9.3
Consider $f(x) = x^2$, $f'(x) = 2x$, $f''(x) = 2$. So $p = 0$ and $m = 2$.

$$f(x) = (x-0)^2 \cdot 1, \quad q(x) = 1$$

Example 9.4
Consider $f(x) = e^{x^2} - 1$

$$\begin{aligned} f(0) &= 0 \\ f'(x) &= 2xe^{x^2} \\ f'(0) &= 0 \\ f''(x) &= 2e^{x^2} + 4x^2 e^{x^2} \\ f''(0) &= 2 \end{aligned}$$

Have $p = 0$, $m = 2$

$$f(x) = (x-0)^2 \frac{e^{x^2}-1}{x^2}, \quad q(x) = \frac{e^{x^2}-1}{x^2}$$

So

$$\lim_{x\to 0} q(x) = \lim_{x\to 0} \frac{1 + x^2 + \frac{1}{2}x^4 + \frac{1}{6}x^6 + \mathcal{O}(x^8) - 1}{x^2} = 1$$

Question 9.1. How does this relate to N.M?

We know that N.M. fails when $f(p) = 0$ and $f'(p) = 0$. To resolve this, let's introduce $\mu(x) = \frac{f(x)}{f'(x)}$. We have

$$f'(x) = m(x-p)^{m-1}q(x) + (x-p)^m q'(x)$$

So

$$\mu(x) = x - p\frac{q(x)}{mq(x) + (x-p)q'(x)}$$

and $\mu(p) = 0$

$$\frac{q(p)}{mq(p) + (p-p)q'(p)} = \frac{1}{m} \neq 0$$

$\mu(x)$ has root p with multiplicity 1 ($\mu'(p) \neq 0$).
Modified Newton's Method: Given p_0, define

$$\mu(x) \coloneqq \frac{f(x)}{f'(x)}$$

$$p_{n+1} = p_n - \frac{\mu(p_n)}{\mu'(p_n)}$$

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

This allows us to find p without worrying about division by zero. However, the drawback here is we have to compute second derivative...

§9.2 Interpolation

Given n discrete points $(x_0, f(x_0)), (x_1, f(x_1)), \dots, (x_n, f(x_n))$. We want to find polynomial $P(x)$

$$P(x) = f(x), \quad \text{at } x = x_i, \quad \forall 0 \leq i \leq n$$

Lagrangian polynomials is our solution here. Given $n+1$ data points, these will produce a polynomial of degree n.

Example 9.5
- 1 data point gives a constant function
- 2 data points give a linear function

$$P(x) = f(x_0)\frac{x-x_1}{x_0-x_1} + f(x_1)\frac{x-x_0}{x_1-x_0}$$

Clearly, $P(x_0) = f(x_0)$, $P(x_1) = f(x_1)$.

The strategy here is to sum up polynomials so that each piece vanishes at other data points.

$$L_0(x) \coloneqq \frac{x-x_1}{x_0-x_1}, \quad L_1(x) \coloneqq \frac{x-x_0}{x_1-x_0}$$

So

$$L_0(x_i) = \delta_{i0}$$
$$L_1(x_i) = \delta_{i1}$$
$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Then

$$P(x) = f(x_0)L_0(x) + f(x_1)L_1(x)$$

Suppose we have $n+1$ distinct points. Then we define

$$L_i(x) \coloneqq \frac{(x-x_0)(x-x_1)\dots(x-x_{i-1})(x-x_{i+1})\dots(x-x_n)}{(x_i-x_0)(x_i-x_1)\dots(x_i-x_{i-1})(x_i-x_{i+1})\dots(x_i-x_n)}$$

or more compactly

$$L_i(x) = \prod_{j=0,\, j\neq i}^{n} \frac{x-x_j}{x_i-x_j}, \quad 0 \leq i \leq n, \quad L_i(x_j) = \delta_{ij}$$

Definition 9.6 (Lagrangian Polynomial) — A Lagrangian polynomial of degree n of $f(x)$ is

$$P(x) = \sum_{i=0}^{n} f(x_i) L_i(x)$$

§10 Lec 10: Oct 15, 2021

§10.1 Theoretical Results for Lagrangian Polynomials

Given input data points $\{x_i, f(x_i)\}_{i=0}^n$ we say

$$L_i(x) = \prod_{j=0, j\neq i}^{n} \frac{x - x_j}{x_i - x_j}, \quad P(x) = \sum_{i=0}^{n} f(x_i) L_i(x)$$

where $P(x)$ is a degree n polynomial.

Example 10.1

Let $f(x) = e^x$, $x_0 = 0$, $x_1 = \frac{1}{2}$, $x_2 = 1$. Then,

$$f(x_0) = 1, \quad f(x_1) = \sqrt{e}, \quad f(x_2) = e$$

So,

$$\begin{aligned} P(x) &= 1 \cdot L_0(x) + \sqrt{e} \cdot L_1(x) + e \cdot L_2(x) \\ &= 1 \cdot \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} + \sqrt{e} \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} + e \cdot \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)} \end{aligned}$$

Summing up degree 2 polynomials, the result is a degree 2 polynomial

$$\begin{aligned} P(1/4) &\approx 1.2717 \\ f(1/4) &\approx 1.2840 \end{aligned}$$

which is roughly 1% error.

In the above example, using more points than $n + 1 = 3$ will result in a better approximation.

Question 10.1. How do we measure error?

First, we need two results from calculus

Theorem 10.2 (Generalized Rolle)

Let $f \in C^n([a, b])$. Suppose $\exists n+1$ distinct roots of f on $[a, b]$. Then $\exists \xi \in (a, b)$ s.t. $f^{(n)}(\xi) = 0$.

This basically says zeros in a function implies a zero of the high-order derivative.

Lemma 10.3

Derivative of Multiplied Monomials

$$\frac{d^{n+1}}{dt^{n+1}} (t - t_0)(t - t_1) \dots (t - t_n) = (n + 1)!$$

Example 10.4

Have
$$\frac{d}{dt}(t-x_0) = 1 = 1!, \quad \frac{d^2}{dt^2}(t-x_0)(t-x_1) = 2 = 2!$$
Induction!

Theorem 10.5 (Error of Lagrangian Polynomial Interpolation)

Let $\{x_0, x_1, \ldots, x_n\} \in [a,b]$ be distinct. Let $f \in C^{n+1}([a,b])$, $P(x) = \sum_{i=0}^n f(x_i)L_i(x)$, then $\forall x \in [a,b]$, $\exists \xi(x) \in (a,b)$ s.t.
$$\begin{aligned} f(x) &= P(x) + \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_0)(x-x_1)\ldots(x-x_n) \\ &= P(x) + \frac{f^{(n+1)}(\xi)}{(n+1)!}\prod_{k=0}^n (x-x_k) \end{aligned}$$

Proof. True if $x = x_i$ since $f(x_i) = P(x_i)$ by construction. So we only deal with $x \neq x_i$. Let x be fixed and define
$$g(t) := f(t) - P(t) - (f(x) - P(x)) \cdot \prod_{j=0}^n \left(\frac{t-x_j}{x-x_j}\right)$$
We want to apply Generalized Rolle's Theorem on $g(t)$ to claim $g^{(n+1)}(\xi) = 0$, and we need to show g is $C^{n+1}([a,b])$ and has $n+2$ distinct roots. Generalized Rolle's Theorem says $g^{(n+1)}(\xi) = 0$.
$$\begin{aligned} g^{(n+1)}(t) &= f^{(n+1)}(t) - P^{(n+1)}(t) - (f(x) - P(x))\frac{d^{n+1}}{dt^{n+1}}\prod_{j=0}^n \frac{(t-x_j)}{x-x_j)} \\ &= f^{(n+1)}(t) - (f(x) - P(x))(n+1)!\prod_{j=0}^n \frac{1}{(x-x_j)} \\ 0 = g^{(n+1)}(\xi) &= f^{(n+1)}(\xi) - (f(x) - P(x))(n+1)!\prod_{j=0}^n \frac{1}{(x-x_j)} \\ f(x) &= P(x) + \frac{f^{(n+1)}(\xi)}{(n+1)!}\prod_{k=0}^n (x-x_k) \end{aligned}$$
$\square$

Remark 10.6. The pointwise error
$$f(x) - P(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}\prod_{k=0}^n (x-x_k)$$
In order for it to be useful, we need a bound on $\left|f^{n+1}(\xi)\right|$. And L.P. is unique.