

Math 151A - Notes

Jingxuan Shan

Jan 5, 2026

Theorem: The intermediate value theorem (IVT):

Let $f \in C([a, b])$. That is, f is continuous on $[a, b]$. Let $k \in \mathbb{R}$ such that k is between $f(a)$ and $f(b)$.

Then there exists $c \in (a, b)$ such that $f(c) = k$.

As an example, let $f(x) = 4x^2 - e^x$. We have that $f(0) = -1$ and $f(1) = 4 - e > 0$.

By IVT, there exists $x \in (0, 1)$ such that $f(x) = 0$.

Theorem: The mean value theorem:

Let $f \in C([a, b])$ and let f be differentiable on (a, b) .

Then there exists $c \in (a, b)$ such that $f'(c) = \frac{f(b) - f(a)}{b - a}$.

Theorem: Taylor's theorem:

Let $f \in C^n([a, b])$. That is, f is n -times continuously differentiable. Also, let $f^{(n+1)}$ exist on (a, b) .

Then for all $x \in [a, b]$, there exists $c(x) \in \mathbb{R}$ such that c is between x and $x_0 \in [a, b]$ and $f(x) = P_n(x) + R(x)$, where:

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

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

Jan. 7:

Notice that saying $f \in C^1$ is different from saying that f is differentiable.

One important example is $f(x)$ defined to be $x^2 \sin(\frac{1}{x})$ when $x \neq 0$, and 0 otherwise. This function is differentiable, yet its derivative is not continuous.

Convention states that a function $f : [a, b] \rightarrow \mathbb{R}$ can be said to be differentiable at a (or b) if it can be extended into a function $g : (c, d) \rightarrow \mathbb{R}$ with (c, d) containing $[a, b]$, and g is differentiable at a (or b).

Absolute and Relative errors: real numbers may have infinite number of digits and therefore cannot be represented exactly using floating point numbers in a computer.

Definition: Absolute error: $e_{\text{abs}} = |p - \tilde{p}|$.

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

Where p is the value and \tilde{p} is an approximation to it.

Finite Digit Arithmetic: chop or round numbers.

We consider things in decimal (instead of binary) for simplicity.

Every operation introduces error in the process.

We can reduce the number of floating point operations (FLOPs) by nesting our calculations.

For example, $x^3 - 6.1x^2 + 3.2x + 1.5 \Rightarrow ((x - 6.1)x + 3.2)x + 1.5$ (nested form)

Order of convergence for sequences: For a convergent sequence $(p_n)_{n \in \mathbb{N}}$, if $\exists \lambda, \alpha$ with $0 < \lambda < \infty$ and $\alpha > 0$ s.t. $\lim_{n \rightarrow \infty} \frac{|p_{n+1} - p|}{|p_n - p|^\alpha} = \lambda$, then we say p_n converges to p with order α .

Definition: (Big O Notation) If $\exists c > 0$ s.t. $|a(t)| \leq cb(t)$ for $t \rightarrow 0$ or ∞ , then we say $a(t) = O(b(t))$.

Jan. 9

Remark: $a(t) = O(b(t))$ is equivalent to saying $\lim_{t \rightarrow 0 \text{ or } \infty} \frac{|a(t)|}{b(t)}$ is bounded by a positive number.

We have that $f(x_0 + \Delta x) = f(x_0) + f'(x_0)\Delta x + \frac{1}{2}f''(c(\Delta x))\Delta x^2$.

The error is $e(\Delta x) = \frac{1}{2}f''(c(\Delta x))\Delta x$.

As Δx gets very small, subtracting two numbers very close to each other leads to a significant loss due to rounding off errors.

Root-finding problem: find p s.t. $f(p) = 0$.

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

Therefore by IVP, $\exists p$ s.t. $f(p) = 0$. The goal of numerical root finding is to find p s.t. $f(p) = 0$ or is close to 0, but the IVP does not give us the root.

The bisection method: set $a_1 = a, b_1 = b$. Set $p_1 = \frac{a_1 + b_1}{2}$.

If $f(p_1) = 0$ then we are done. Else, if $f(p_1)$ has same sign as $f(a_1)$, then $p \in (p_1, b_1)$. Set $a_2 = p_1, b_2 = b$.

Else, if $f(p_1)$ has same sign as $f(b_1)$, then $p \in (a_1, p_1)$. Set $a_2 = a_1, b_2 = p_1$.

Set $p_2 = \frac{a_2 + b_2}{2}$, and repeat.

If there are multiple roots, then this algorithm is guaranteed to find exactly one root.

Stopping criteria: we need a sequence (p_1, p_2, \dots) and a specified tolerance ϵ .

Some choices for when to stop an algorithm:

(1) $|p_n - p_{n-1}| < \epsilon$, absolute difference

(2) $\frac{|p_n - p_{n-1}|}{|p_n|} < \epsilon$ (assuming $p_n \neq 0$), relative difference

(3) $|f(p_n)| < \epsilon$, a "residual"

Note that the first two can only be used if the sequence is guaranteed to converge.

Jan. 12

Remark: Bisection is a global method as long as the assumptions are satisfied. It will converge (approximately) to some p such that $f(p) = 0$.

Remark: Bisection can only find one of the roots.

Remark: Bisection would not work for functions like $f(x) = x^2$, since it is impossible to find $[a, b]$ s.t. $f(a)f(b) < 0$.

Theorem: The sequence $(p_n)_{n \in \mathbb{N}}$ provided by bisection satisfies $|p_n - p| \leq \frac{b-a}{2^n}$. Thus, as $n \rightarrow \infty$, the error $\rightarrow 0$.

Remark: This error bound, $\frac{b-a}{2^n}$, converges linearly: $\lim_{n \rightarrow \infty} \frac{|p_{n+1} - p|}{|p_n - p|} = \lim_{n \rightarrow \infty} \frac{2^n}{2^{n+1}} = \frac{1}{2}$.

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

Definition: Let function $g : [a, b] \rightarrow \mathbb{R}$. If $g(p) = p$, then p is a fixed point of g .

For example, $f(x) = x^2$ has two fixed points, 0 and 1.

Theorem: Let p be a fixed point of g . Then p is a root of $G(x) = g(x) - x$.

Given a root-finding problem $f(p) = 0$, we can define functions g with p as a fixed point in a number of ways.

A fixed point for $g \Rightarrow y = g(x)$ intersects $y = x$.

Fixed Point Iteration: for $g \in C([a, b])$, let $p_0 \in [a, b]$ and set $p_{n+1} = g(p_n)$.

We also need that $g(x) \in [a, b]$. Notice that the initial selection of p_0 is arbitrary.

The stopping criteria of this algorithm is identical to that of the bisection.

One may wonder when does FPI converges and when does it fail.

For example, when trying to solve $x^2 - 7 = 0$, the function $g(x) = \frac{7}{x}$ jumps between 3 and $\frac{7}{3}$ for $p_0 = 3$.

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

Jan. 14

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

Proof: If $g(a) = a$ or $g(b) = b$, then p is already found.

Otherwise, define $G(x) = g(x) - x$. Then since $g(x) \in [a, b] \forall x \in [a, b]$, we know $G(a) = g(a) - a > 0$ and $G(b) = g(b) - b < 0$. We also have that $G(x) \in C([a, b])$ since the sum of continuous functions is continuous.

Therefore by IVT, $\exists p \in [a, b]$ s.t. $G(p) = 0 \Rightarrow g(p) = p$.

Definition: A function f satisfies the Lipschitz condition on an interval $I \subseteq \mathbb{R}$ if $\exists L > 0$ s.t. $|f(x) - f(y)| \leq L|x - y| \forall x, y \in I$. L is called the Lipschitz constant.

Proposition: Assume $g \in C([a, b])$ and $g(x) \in [a, b] \forall x \in [a, b]$. Furthermore, assume the Lipschitz constant of g on $[a, b]$, k , satisfies $0 < k < 1$. Then the following all holds:

- (1) There exists a unique $p \in [a, b]$ s.t. $g(p) = p$.
- (2) The fixed point iteration will converge to p .
- (3) Error estimate: $|p_n - p| \leq k^n \max\{b - p_0, p_0 - a\}$.

Proof: (1) We have already proved the existence of p in the previous proposition. Now, assume $\exists p, q \in [a, b]$ s.t. $g(p) = p$ and $g(q) = q$, yet $p \neq q$. So we have that $p - q \neq 0$.

Then since $k \in (0, 1)$, we have that $|g(p) - g(q)| \leq |g(p) - g(q)| < |p - q| \Rightarrow |p - q| < |p - q|$, contradiction.

- (2) This is immediate from (3).
- (3) Let p be the unique fixed point s.t. $g(p) = p$, and let $(p_n)_{n \geq 0}$ be the fixed point iteration of g , with $p_0 \in [a, b]$. Then $\forall n \in \mathbb{N}$, we have that:

$$\begin{aligned} |p_n - p| &= |g(p_{n-1}) - g(p)| \leq k|p_{n-1} - p| \\ k|p_{n-1} - p| &= k|g(p_{n-2}) - g(p)| \leq k^2|p_{n-2} - p| \\ &\vdots \\ k^{n-1}|p_1 - p| &= k^{n-1}|g(p_1) - p| \leq k^n|p_0 - p|. \end{aligned}$$

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

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

The theorem gives us a way to guarantee fixed point iteration convergence by looking at the Lipschitz condition.

In practice, it is often more convenient to look at the derivative.

Proposition: Assume $g \in C([a, b])$ and $g \in [a, b]$. Further assume that $g \in C^1([a, b])$, and that $\exists k > 0$ s.t. $\forall x \in [a, b], |g'(x)| \leq k$.

Then $g(x)$ is Lipschitz on $[a, b]$ with constant k .

Proof: By the mean value theorem, $\forall x, y \in [a, b], |g(x) - g(y)| = |g'(c)||x - y| \leq k|x - y|$.

Next lecture we will talk about Newton's method. **Jan. 14**

Newton's method and Secant method are root-finding methods.

There are three ways to derive Newton's method, which is a classic technique in root finding.

Let $f \in C^2([a, b])$, and p be a root. That is, $f(p) = 0$.

Suppose p_n is "close" to p . That is, $|p - p_n|$ is "small".

By Taylor's theorem, $0 = f(p) = f(p_n) + f'(p_n)(p - p_n) + f''(c)\frac{(p-p_n)^2}{2}$.

If $|p - p_n|$ is "small", then $|p - p_n|^2$ is "really small".

Then $0 \approx f(p_n) + f'(p_n)(p - p_n) \Rightarrow p \approx p_n - \frac{f(p_n)}{f'(p_n)}$.

Definition: (Newton's method) Suppose we start with p_n close to p , then $p_{n+1} = p_n - \frac{f(p_n)}{f'(p_n)}$.

Graphical derivation of Newton's method: If p_n is close to p , the x -intercept of the tangent line of f at $x = p_n$ would be closer to p than p_n .

Let the tangent line be $y = ax + b$ and the intercept be p_{n+1} . So $f(p_n) = ap_n + b$, $0 = ap_n + b$ and $a = f'(p_n)$.

Solving for a, b and p_{n+1} , we get $p_{n+1} = p_n - \frac{f(p_n)}{f'(p_n)}$.

If we picked a p_0 that is far away from p , it could lead us further away.

Fixed Point Iteration of Newton's method

Theorem: Let $g(x) = x - \frac{f(x)}{f'(x)}$ for some $f \in C^1([a, b])$.

Also assume that $f'(x) \neq 0$ for $x \in [a, b]$. Then $f(p) = p$ if and only if $f(p) = 0$.

We can define a fixed point iteration for g : $p_{n+1} = g(p_n) - \frac{f(p_n)}{f'(p_n)}$.

Remark: We must have $f'(p_n) \neq 0 \forall n$, otherwise Newton's method will fail.

Remark: If it actually converges, Newton's method will converge faster than bisection to the root.

Remark: Unlike the bisection method, Newton's method is a local method, rather than global.

Remark: Newton's method requires knowledge and evaluation of $f'(x)$. This may be costly, especially if $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$.

In that case we have $x_{n+1} = x_n - (J(x_n))^{-1}f(x_n)$, where $J(x)$ is the Jacobian matrix, $J_{i,j} = \frac{\partial f_i}{\partial x_j}(x)$.

Secant Method

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

Definition: We define the secant method as: $p_{n+1} = p_n - f(p_n) \frac{p_n - p_{n-1}}{f(p_n) - f(p_{n-1})}$.

Remark: Secant method is useful when we do not have information about $f'(x)$, e.g. when $f(x)$ comes from experiment data.

Next time we will talk about how Newton's method is guaranteed to converge for sufficiently close guesses.