Solution of Nonlinear Equations

The Basic Problem

Find a root $x \in \mathbb{R}$ of an equation of the form $f(x) = 0$ for a given continuous function f.

In most cases it is not possible to solve analytically.

We will consider iterative methods to approximate the solution.

Our first method will be the <u>Bisection Method</u>.

We must start with an interval $[a, b]$ with $f(a)$ and $f(b)$ of opposite sign.

There must be a number p in (a, b) with $f(p) = 0$ by the :

Intermediate Value Theorem

If $f \in C[a,b]$ and K is any number between $f(a)$ and $f(b)$ then there exists a number c in (a,b) for which $f(c) = K$.

Now set $a_1 = a$ and $b_1 = b$ and let p_1 be the midpoint of $[a_1, b_1]$

$$p_1 = \frac{1}{2}(a_1 + b_1)$$

If $f(p_1) = 0$ then we are done. $p = p_1$

If $f(p_1)$ and $f(a_1)$ are of opposite signs then there must be a $p \in (a_1, p_1)$ so that $f(p) = 0$.

Set $a_2 = a_1$ and $b_2 = p_1$.

Otherwise $f(p_1)$ and $f(b_1)$ are of opposite signs, so there must be a $p \in (p_1, b_1)$ so that $f(p) = 0$.

Set $a_2 = p_1$ and $b_2 = b_1$

Reapply the process to $[a_2, b_2]$

Once appropriate stopping criteria are satisfied we set the midpoint of the interval equal to the estimate for the root.

Possible stopping criteria:

① $\frac{b_n - a_n}{2} < TOL$

or $|p_n - p_{n-1}| < TOL$

GOOD: • Ensures that the returned root value p_n is within TOL of the exact value p
• Easy error analysis

BAD: • Does not ensure that $f(p_n)$ is small.
• An absolute rather than a relative error.

(2) $\dfrac{|p_n - p_{n-1}|}{|p_n|} < TOL_2$, $p_n \neq 0$

Usually preferred over (1) if nothing is known about $f(\cdot)$ or p.

(3) $|f(p_n)| < TOL_3$

Ensures that $f(p_n)$ is small, but p_n may differ significantly from the true root p.

(4) We can also carry out a fixed number of iterations N — this is closely related to (1)

The best stopping criteria will depend on what is known about f and p and on the type of problem.

It is ~~s~~ often useful to use the relative error test (2) with a fixed, maximum number of steps (4).

Now let's consider bisection method with stopping criteria ①

(FOR EASY ANALYSIS)

Choose endpoints a, b, Tolerance TOL & max iterations N_0

Step 1: Set $n = 1$

$FA = f(a)$

Step 2: While $i \leq N_0$ do Steps 3-6

Step 3: Set $p = a + \frac{(b-a)}{2}$

$Fp = f(p)$

Step 4: If $FP = 0$ or $\frac{b-a}{2} < TOL$

then return (p)

Step 5: Set $i = i + 1$

Step 6: If $FA \cdot FP > 0$ then set

$a = p$

$FA = FP$

else set $b = p$

STEP 7 OUTPUT ('Method failed after N_0 iterations, $N_0 =$', N_0)

FINE POINTS:

- $$p = a + \frac{b-a}{2}$$

  is preferred over

  $$p = \frac{a+b}{2}$$

  It is usually best in bisection to add a small correction to a previous approximation. Could otherwise lead to $p \notin [a, b]$.

- To avoid underflows and overflows it is sometimes preferred to use

  $$\text{sign}(FA) \cdot \text{sign}(FP) > 0$$

  ~~rather~~ rather than $FA \cdot FP > 0$.

- We want to choose the initial interval as small as possible to minimize the number of iterations.

From the previous example we see that bisection is slow to converge. However the method will always converge (by the intermediate value theorem) which makes it an excellent choice to start other methods, or to use when other methods fail.

How accurate is bisection?

__Thm__ Suppose that $f \in C[a,b]$ and $f(a) \cdot f(b) < 0$. ~~Thus~~ Then the bisection algorithm given above generates a sequence $\{p_n\}$ approximating a zero p of f with

$$|p_n - p| \leq \frac{b-a}{2^n} \qquad n \geq 1.$$

__Pf__: For each $n \geq 1$ we have

$$b_n - a_n = (b-a)\frac{1}{2^{n-1}} \quad \& \quad p \in (a_n, b_n)$$

Since $p_n = \frac{1}{2}(a_n + b_n)$ for all $n \geq 1$

$$|p_n - p| \leq \frac{1}{2}(b_n - a_n) = \frac{b-a}{2^n}$$

Notice $p_n = p + O\left(\frac{1}{2^n}\right)$

Ex: Use the theorem to find a bound for the number of iterations needed to approximated a solution to the equation $x^3 + x - 4 = 0$ on the interval $[1, 4]$ to an accuracy of 10^{-3}.

Fixed Point Iteration

We wish to find the roots of an equation $f(p) = 0$.

We will be focusing on methods that iterate to find the root:

$$p^{n+1} = g(p^n)$$

We start by considering the fixed point problem.

Def: A fixed point p is the value p s.t. $g(p) = p$.

Notice that fixed point problems and root finding problems are equivalent.

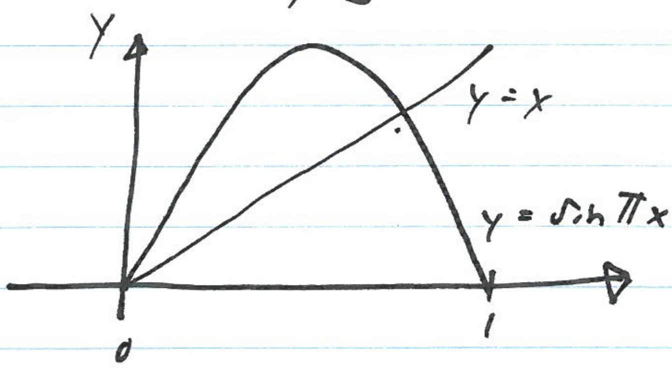
Example Let $g(\cancel{x}) - \cancel{x} = f(\cancel{x})$

f has a root p $\Longrightarrow$ g has a fixed pt p
$f(p) = 0$ $\qquad$ $g(p) = p$

g has a fixed pt p $\Longrightarrow$ f has a root p
$g(p) = p$ $\qquad$ $f(p) = 0$

Note: many other possible choices for g: example $g(x) - x = (f(x))^3$

Ex The function $g(x) = \sin \pi x$ has _ fixed points on $[0, 1]$.



Ex The function $g(x) = x$ has an infinite number of fixed points over $[a, b]$.

(each $x \in [a, b]$ is a fixed pt)

We now consider what functions have fixed points and if the fixed point is unique.

Preliminaries: ① Intermediate Value Theorem

② Mean Value theorem

If $f \in C[a,b]$ and f is differentiable on (a, b) then there is a number c in (a, b) with

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

Thm (existence and uniqueness)

* If $g \in C[a,b]$ and $g(x) \in [a,b]$ for all $x \in [a,b]$ then $g(x)$ has a fixed point in $[a,b]$

** Suppose, in addition, that $g'(x)$ exists on (a,b) a that a positive constant $K<1$ exists with

$$|g'(x)| \leq K < 1 \quad \text{for all } x \in (a,b)$$

then the fixed point in $[a,b]$ is unique.

Proof (*): If $g(a)=a$ or $g(b)=b$ the g has a fixed point at an end point.

Suppose not, then it must be true that $g(a)>a$ and $g(b)<b$.

Define $h(x)=g(x)-x$. Then h is continuous on $[a,b]$ and

$$h(a)=g(a)-a>0 \quad \& \quad h(b)=g(b)-b<0$$

The Intermediate Value Theorem implies that there exists a $p \in (a,b)$ for which $h(p)=0$.

Thus $g(p)-p=0 \Rightarrow p$ is a fixed point of g. $\square$

pf (**)

Suppose, in addition $|g'(x)| \le K < 1$ for all $x \in (a, b)$.

and that p and q are both fixed points in $[a, b]$ with $p \neq q$.

By the Mean Value Theorem a number ξ exists between p and q such that

$$\frac{g(p) - g(q)}{p - q} = g'(\xi)$$

then
$$\begin{aligned} |p-q| &= |g(p) - g(q)| \\ &= |g'(\xi)|\,|p-q| \\ &\le K\,|p-q| \\ &< |p-q| \end{aligned}$$

which is a contradiction

This contradiction must come from the assumption that $p \neq q$.

$\therefore$ $p = q$ & the fixed point is unique $\square$

We want to approximate the fixed point of a function g.

IDEA
- choose an initial approximation p_0
- generate a sequence $\{p_n\}_{n=0}^{\infty}$ such that $p_n = g(p_{n-1})$ $n \geq 1$.

If the sequence converges to p and g is continuous

$$p \equiv \lim_{n\to\infty} p_n$$

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

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

$$= g(p)$$

This gives the Fixed Point Algorithm:

Fixed Point Algorithm

Input p_0, TOL, N_0

Step 1: set $i=1$

Step 2: While $i \le N_0$ do Steps 3-6

Step 3: Set $p = g(p_0)$

Step 4: If $|p-p_0| < TOL$ then
output (p).
STOP.

Step 5: Set $i = i+1$

Step 6: Set $p_0 = p$

Step 7: Output (Iteration exceeded. $N > N_0$)

Example (Example 3 of Section 2.2)

The equation $x^3 + 4x^2 - 10 = 0$ has a unique root in $[1, 2]$.

There are many ways to change the equation to the form $x = g(x)$.

We can the apply the Fixed Point algorithm to try to find the root.

(see text for results)

why do some converge & some diverge?
why do they converge with different rates?

Consider a simple illustrative example.

$$g(x) = ax+b.$$

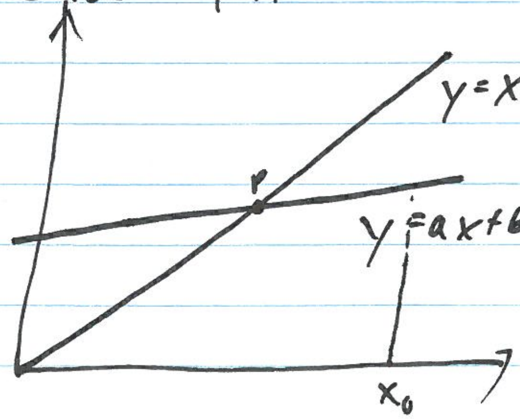
We have
$$x_1 = ax_0 + b$$
$$x_2 = ax_1 + b = a(ax_0+b)+b = a^2x_0 + (1+a)b$$
$$x_3 = ax_2 + b = \cdots = a^3x_0 + (1+a+a^2)b$$

and by induction,
$$x_n = \begin{cases} a^n x_0 + \frac{1-a^n}{1-a} b & \text{for } a \neq 1 \\ x_0 + nb & \text{for } a = 1. \end{cases}$$

$$\therefore \lim_{n\to\infty} x_n = \begin{cases} \frac{1}{1-a} b & \text{for } |a| < 1 \\ x_0 & \text{for } a = 1,\ b = 0 \end{cases}$$

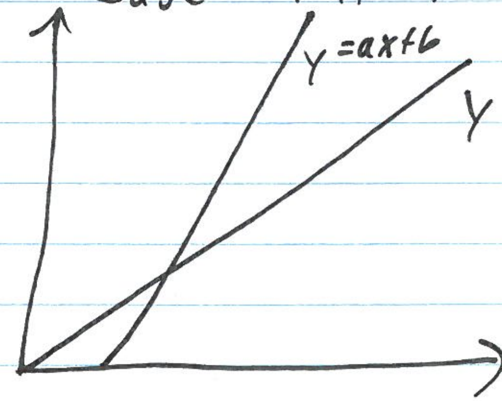
& no proper limit exists for all other values of a, b.

Case $|a|<1$



Convergent iteration

Case $|a|>1$



Divergent Iteration

When does a fixed point iteration converge? How quickly does it converge?

For this we turn to the

Fixed Point Theorem:

Let $g \in C[a,b]$ and suppose $g(x) \in [a,b]$ for all x in $[a,b]$.

Suppose, in addition, that g' exists on (a,b) and a positive constant $K<1$ exists with

$$|g'(x)| \leq K \quad \text{for all } x \in (a,b)$$

Then for any number p_0 in $[a,b]$ the sequence defined by

$$p_n = g(p_{n-1}) \qquad n \geq 1$$

converges to the unique fixed point p in $[a,b]$.

Pf. Our earlier Thm (existence & uniqueness) tells us that a unique fixed point exists in $[a, b]$.

Notice that g maps $[a, b]$ into itself, so the sequence $\{p_n\}_{n=0}^{\infty}$ is defined for all $n \geq 0$ and $p_n \in [a, b]$ for all n.

We may apply the mean value theorem to g to show that for any n

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

where $\xi \in (a, b)$. Applying the inequality inductively gives

$$|p_n - p| \leq K|p_{n-1} - p| \leq K^2|p_{n-2} - p| \leq \cdots \leq K^n|p_0 - p|$$

Since $K < 1$, $\lim_{n \to \infty} |p_n - p| \leq \lim_{n \to \infty} K^n |p_0 - p| = 0$

$\therefore \{p_n\}_{n=0}^{\infty}$ converges to p.

$\square$

This proof also gives us a natural bound for the error.

<u>Corollary</u>: If g satisfies the hypothesis of the Fixed Point Thm,

$$\underbrace{|p_n - p|}_{\text{absolute error}} \le K^n \max\{p_0 - a,\ b - p_0\}$$

and

$$|p_n - p| \le \frac{K^n}{1-K} |p_1 - p_0| \quad \text{for all } n \ge 1.$$

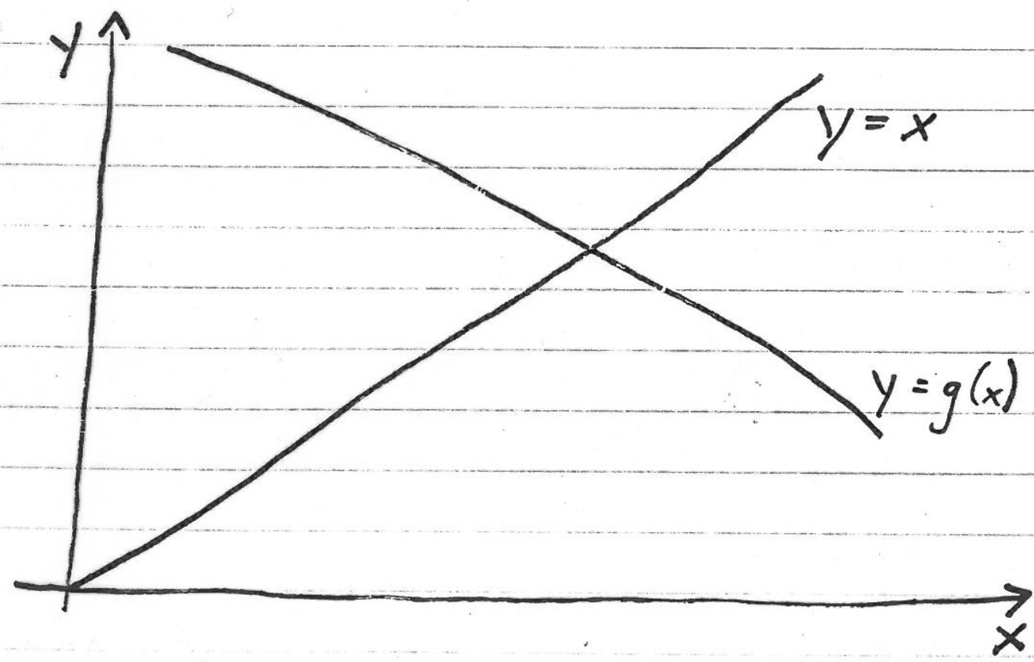
<u>Pf</u>. see text.

Fixed Point Iteration

Recall that a fixed point p is the value p s.t. $g(p) = p$

Consider the iteration for finding fixed points

$$p_0 \text{ given}$$
$$p_n = g(p_{n-1})$$



Newton's Method

(or Newton-Raphson Method)

One of the most powerful & well-known methods for solving a root-finding problem

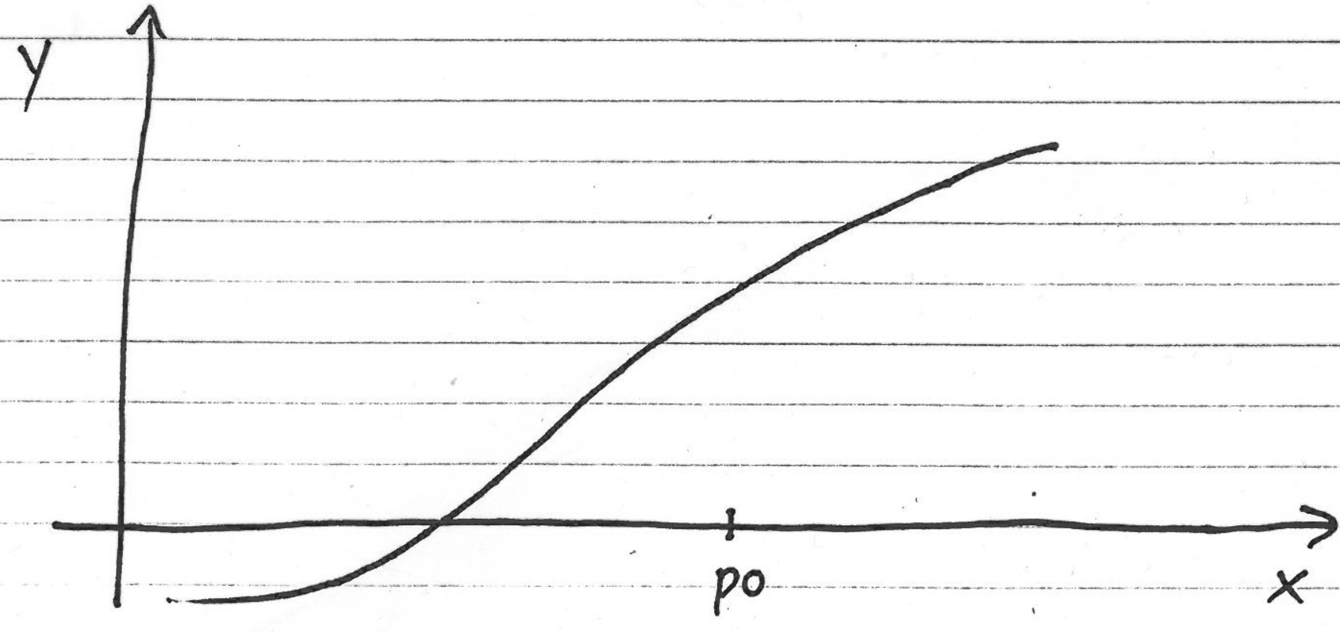
$$f(x) = 0.$$

Pros: Much faster than Bisection

Cons: Need $f'(x)$
Not guaranteed to always converge.

Want $x = p$ such that $f(x) = 0$.

Idea: Use slope as well as function values:



Derivation (by Taylor's thm)

__Want__ $x = p$ s.t. $f(x) = 0$

Suppose $f \in C^2[a,b]$. Let $\bar{x} \in [a,b]$ be an approximation to p s.t. $f'(\bar{x}) \neq 0$ and $|\bar{x} - p|$ is sufficiently small. Then

$$f(x) = f(\bar{x}) + f'(\bar{x})(x-\bar{x}) + \tfrac{1}{2} f''(\xi(x))(x-\bar{x})^2$$

where $\xi(x)$ lies between x and $\bar{x}$.

Set $x = p$ in the expansion $(f(p) = 0)$ so

$$0 = f(\bar{x}) + f'(\bar{x})(p-\bar{x}) + \tfrac{1}{2} f''(\xi(p))(p-\bar{x})^2$$

Now if $|p-\bar{x}|$ is small, $|p-\bar{x}|^2$ smaller

$$0 \approx f(\bar{x}) + f'(\bar{x})(p-\bar{x})$$

$$\Longrightarrow \quad f'(\bar{x})(p-\bar{x}) \doteq -f(\bar{x})$$

$$(p-\bar{x}) \approx -f(\bar{x}) / f'(\bar{x})$$

$$p \approx \bar{x} - f(\bar{x}) / f'(\bar{x})$$

So Newton's method begins with an estimate p_0 and generates a sequence $\{p_n\}$

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

Stopping the iteration is similar to the bisection method:

We can use $|p_N - p_{N-1}| < \varepsilon$

$$\frac{|p_N - p_{N-1}|}{|p_N|} < \varepsilon \ , \ p_N \neq 0$$

or $|f(p_N)| < \varepsilon$

Notice that Newton's method fails if $f'(p_{n-1}) = 0$.

The method is most effective when f' is bounded away from zero near p.

Also we saw in the derivation that $|p - \bar{x}|$ had to be small $\Rightarrow$ we need a good initial guess.

Example:

Use Newton's Method to compute the square root of a number R.

Want to find the roots of $p^2 - R = 0$.

Let $f(x) = x^2 - R$
$f'(x) = 2x$

Newton's method takes the form

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

$$= p_{n-1} - \frac{p_{n-1}^2 - R}{2p_{n-1}}$$

$$= \frac{1}{2}\left(p_{n-1} + R/p_{n-1}\right)$$

Method credited to Heron, Greek Engineer Circa 100 BC → 100 AD

Try $R=2$: $p_0 = 2$
$p_1 = 1.5$
$p_2 = 1.416666\cdots$
$p_3 = 1.41421562$
$p_4 = 1.41421356$
(12 digits correct)

Newton's method can be shown to converge under reasonable assumptions
(smoothness of $f(\cdot)$ ~~and~~, a good initial guess, $f'(p) \neq 0$)

Thm (Convergence):

Let $f \in C^2[a,b]$. If $p \in [a,b]$ is such that $f(p) = 0$ and $f'(p) \neq 0$, then there exists a $\delta > 0$ such that Newton's method generates a sequence $\{p_n\}_{n=1}^{\infty}$ converging to p for any initial approximation

$$p_0 \in [p-\delta, p+\delta]$$

<u>Pf</u>. see text. Idea is to analyze Newton's method as the functional iteration scheme

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

with $g(x) = x - f(x)/f'(x)$

Proof finds an interval $[p-\delta, p+\delta]$ that g maps into itself, and $|g'(x)| \leq k_{<1}$ for all $x \in (p-\delta, p+\delta)$

Proof then follows by the fixed point thm.

Notice that Newton's Method has the major difficulty that the derivative of f is needed at each approximation.

Since $f'(p_{n-1}) = \lim_{x \to p_{n-1}} \dfrac{f(x) - f(p_{n-1})}{x - p_{n-1}}$

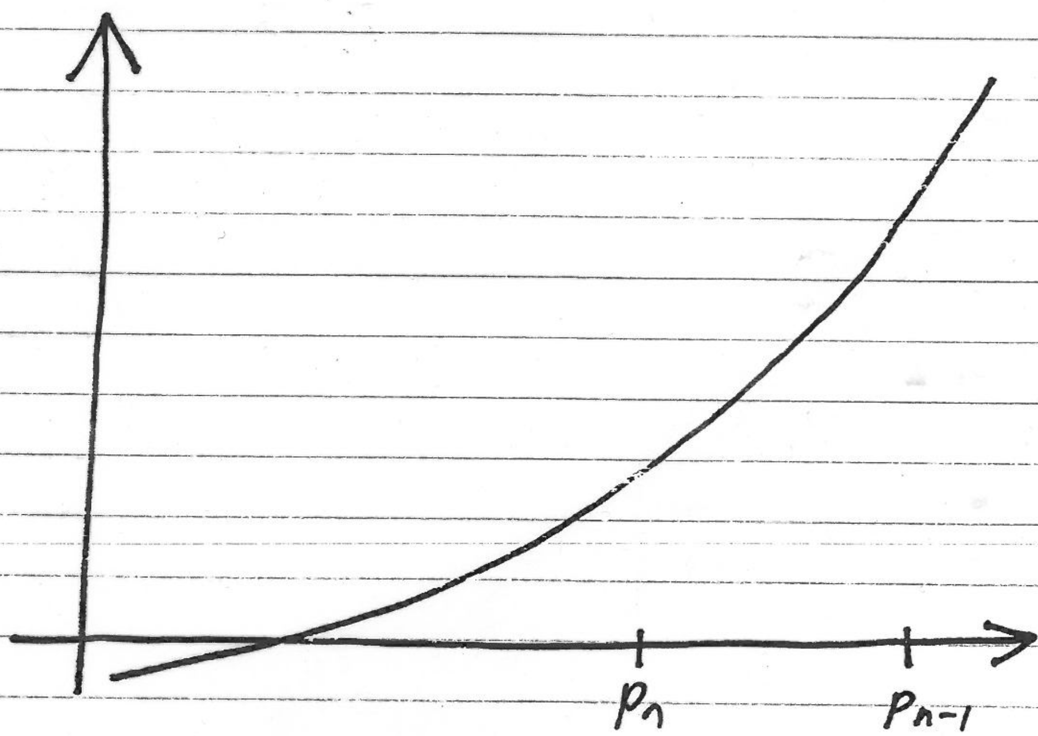
we approximate:

$$f'(p_{n-1}) \approx \frac{f(p_{n-2}) - f(p_{n-1})}{p_{n-2} - p_{n-1}}$$

$$= \frac{f(p_{n-1}) - f(p_{n-2})}{p_{n-1} - p_{n-2}}$$

To give the Secant Method.

Graphically:



Algorithm Secant

Find $f(x) = 0$ given p_0, p_1

Input p_0, p_1, tolerance, N_0

Step 1. Set $i = 2$, $q_0 = f(p_0)$, $q_1 = f(p_1)$

Step 2. While $i \leq N_0$ do 3-6

Step 3. Set $p = p_1 - \dfrac{q_1(p_1 - p_0)}{q_1 - q_0}$

Step 4. If $|p - p_1| < TOL$
then output p
Stop.

Step 5. $i = i + 1$

Step 6. Set
$p_0 = p_1$
$q_0 = q_1$
$p_1 = p$
$q_1 = f(p)$

Step 7. Output ('Method failed after N_0 iterates')

Comparison of Newton's Method vs secant method.

Consider the root finding problem for $R = 2$: $p^2 - \sqrt{R} = 0$

Newton's Method	Secant Method
$p_0 = 2$	$p_0 = 1$
$p_1 = 1.5$	$p_1 = 2$
$p_2 = 1.4166\cdots$	$p_2 = 1.33\cdots$
$p_3 = 1.41421562\cdots$	$p_3 = 1.4$
$p_4 = 1.41421356\cdots$	$p_4 = 1.414634\cdots$
	$p_5 = 1.414214\cdots$
	$p_6 = 1.41421356\cdots$

Notice

- Both Newton's Method and Secant method converge rapidly, but Secant Method is slightly slower.
- convergence is slow at first but improves near the root

$\Rightarrow$ These methods are often used to refine an initial guess that is made using another method such as bisection.