# ROOT FINDING

Rootfinding problems are generall of two categories

1. Simple enclosure method
2. Fixed point method

Simple enclosure method:

- Based on Intermediate Value Theorem
- Initially finds an interval, which is guaranteed to contain a root

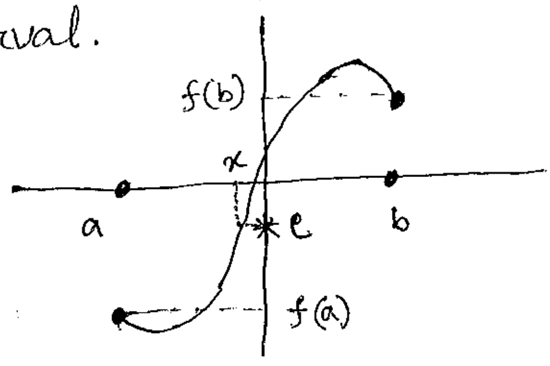
$$\downarrow \text{gradually}$$

Interval size is shrinked

Intermediate Value Theorem:

If $f$ is a continuous fn$^c$ over an interval $[a, b]$ and $c$ is a number between $f(a)$ and $f(b)$, then there's at least one number $x$, where $a < x < b$, such that $f(x) = c$

As a continous fn$^c$ over a close interval, it can assume every value between the values achieved at the endpoints of the interval.




we at least one value that gives $f(x) = c$

# Intermediate Value Theorem

- This theorem states that—

A fn$^c$ that is continuous on a closed interval is guranteed to assume every value between the values at the endpoints of the interval.

↓ Relation with rootfinding

This theorem provides a way to identify intervals which enclose the real zeros of continuous fn$^c$.

↓

All we need is to find an interval such that the values of the fn$^c$ at the endpoints of that interval are of opposite sign.

So, if we can find an interval, where the fn$^c$ values at the end points of that interval are of opposite sign,

We can say that zero value of the fn$^c$ is some where between the values.

this answers

At least one zero of the fn$^c$ on that interval.

Example: Let's assume $f(x) = x^3 + 2x^2 - 3x - 1$

and consider three intervals:

$(-3, -2) \Rightarrow f(-3) = -27 + 18 + 9 - 1 = -1$
$f(-2) = -8 + 8 + 6 - 1 = 5$ } So, zero crossing is there

$(-1, 0) \Rightarrow f(-1) = -1 + 2 + 3 - 1 = 3$
$f(0) = -1 = -1$ } Zero crossing

$(1, 2) \Rightarrow f(1) = 1 + 2 - 3 - 1 = -1$
$f(2) = 8 + 8 - 6 - 1 = 9$ } Zero crossing

Intermediate Value Theorem gives us an interval that contains a zero.

So, What do we do next?

Our goal is to shrink the size of the interval gradually.

Precisely, the root enclosing interval is systematically shrinked.

What would be the simplest way to reduce the interval size?

- "Reduced by half" (Bisecting)

  ↓ determines which half contains a root

  Use Intermediate Value Theorem

  ↓

  New Interval Found → continue

  ↓ if stopping criteria is satisfied

  STOP

This is bisection method.

Bisection method: As seen above, we will have series of Intervals calculated in this process.

- So, let's assume $(a_n, b_n)$ is the interval after $n^{th}$ iteration.

- $P_n$ be the midpoint $P_n = \dfrac{a_n + b_n}{2}$

  → Used as one of the endpoints for next enclosing interval.

  → It is also used as an approximation of the true root "P"

- In next iteration the interval becomes $(a_n, P_n)$ or $(P_n, b_n)$

That is, the biseition method proceeds as :

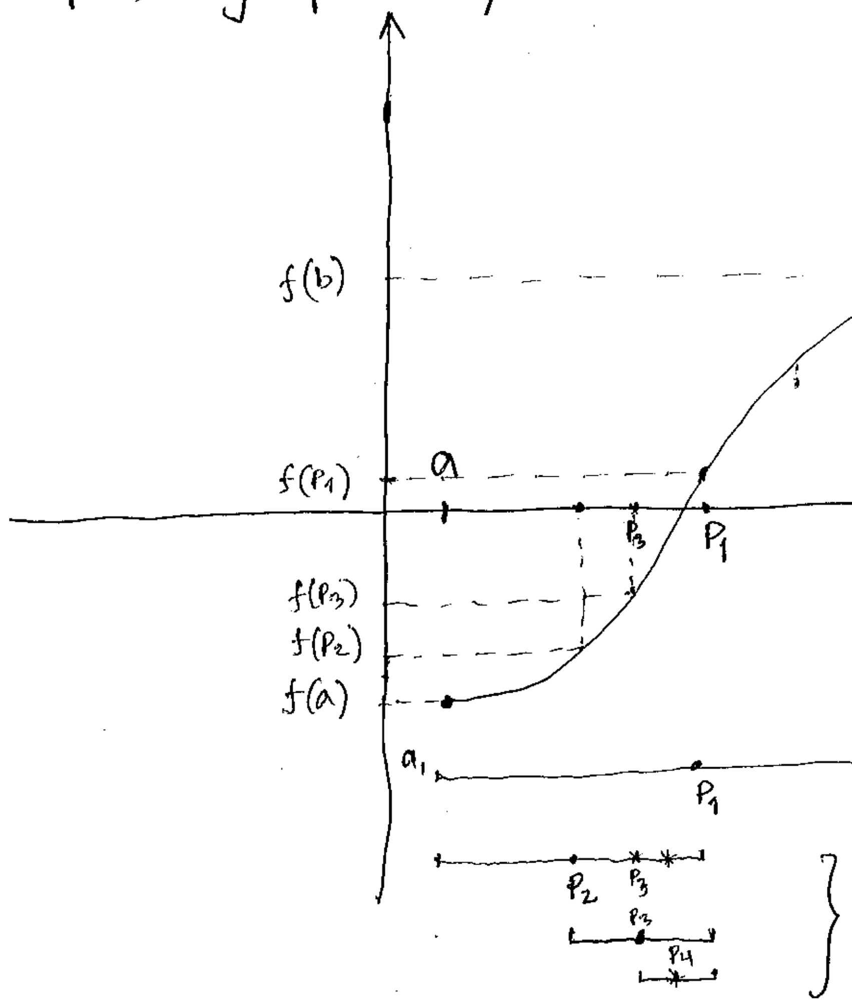
- if $f(P_n) = 0$, then $P = P_n$. We get the root.
- If $f(P_n) \neq 0$, then $f(P_n)$ has the same sign as either $f(a_n)$ or $f(b_n)$
- → if $f(P_n)$ and $f(a_n)$ are of same sign, then $a_{n+1} = P_n$
  → if $f(P_n)$ and $f(b_n)$ are of same sign, then $b_{n+1} = P_n$

↓ This gives new interval

$[a_{n+1} \quad b_{n+1}]$

Repeat.

For example, graphically




interval is shrinking

# Algorithm for Bisection Method

Given
fn$^c$ $f$ whose zero to be found out
left end point of interval $a$
Right end point of Interval $b$
Convergence tolerance $\epsilon$
Maximum number of iterations $N_{max}$

STEP 1 save $sfa = \text{sign}(f(a))$

STEP 2 for $i$ from 1 to $N_{max}$ do

STEP 3 $p = \frac{a+b}{2}$

STEP 4 if $((b-a) < 2\epsilon)$ then OUTPUT $p$

STEP 5 save $sfp = \text{sign}(f(p))$

STEP 6 if $(sfa * sfp < 0)$ then

assign the value of $p$ to b

else

assign the value of $p$ to a

assign the value of $sfp$ to $sfa$

end

end

STEP 7 Output A message that the maximum number of iterations has been exceeded prior to achieving convergence.

STEP6 : We don't use $f(a) * f(p)$, as that may result in underflow. very small, may be.

both $a$ and $p$ will be converging toward a zero of $f$

Example : $f(x) = x^3 + 2x^2 - 3x - 1$ , Given interval $[1, 2]$

So, $f(1) = -1$, $f(2) = 9$ $\Big\}$ opposite sign shows that there's a root in between the interval

First approximation: $P_1 = \frac{a_1+b_1}{2} = \frac{1+2}{2} = 1.5$

So, the fn value $f(P_1) = 2.375 > 0$ root not found

Now, As $f(P_1)$ is +ve, $f(b)$ is +ve | So, new "$b = P_1$"

The new interval $[a_2\ b_2] \equiv [a_1\ P_1]$

Second approximation:

$$P_2 = \frac{a_2+b_2}{2} = \frac{a_1+P_1}{2} = \frac{1+1.5}{2} = 1.25$$

So, $f(P_2) \approx 0.328 > 0$ , this is of opposite sign from $f(a_1)$. ~~So, new~~

So, new $b = P_2$

New interval $[a_3\ b_3] = [a_1\ P_2]$

Third approximation:

$$P_3 = \frac{a_3+b_3}{2} = \frac{a_1+P_2}{2} = \frac{1+1.25}{2} = 1.125$$

and $f(P_3) = -0.420 < 0$

So. new $a = P_3$

New interval $[a_4\ b_4] = [P_3\ b_3]$

# Stopping Criteria

Thr$^m$:

Let's consider $f$ be continuous on the closed interval $[a, b]$ and suppose $f(a)f(b) < 0$.

Bisection method generates a sequence of approximations $\{P_n\}$ which converges to a root $p \in (a, b)$ with the property.

$$|P_n - P| \leqslant \frac{b-a}{2^n}$$

error (Absolute error)

Construction of Bisection algorithm shows that for each 'n', $p \in (a_n, b_n)$ and $P_n$ is the midpoint of $(a_n, b_n)$. This implies that

$P_n$ can differ from "p" by no more than half of the length of $(a_n, b_n)$. So,

$$|P_n - P| \leqslant \frac{1}{2}(b_n - a_n)$$

We can terminate the bisection method when

$$\frac{b_n - a_n}{2} < \varepsilon$$

Fourth Approximation:

$$P_4 = \frac{a_4 + b_4}{2} = 1.1875$$

Now, the true value of the root $p$ is $1.1986912435$

That is, Bisection method gives us the root $p$ with some amount of error.

Here, Absolute error $= \left| 1.1986912435 - 1.1875 \right|$

$= 1.119 \times 10^{-2}$

Precisely this is Bisection Method.

However, Since Bisection method is iterative in nature, we need a stopping criteria

↓ stopping criteria: any of the below three

1. Absolute error in the location of the root
   Terminate the iteration when $|P_n - P| < \varepsilon$

2. Relative error in the location of the root
   Terminate the iteration when $|P_n - P| < \varepsilon P$

3. Test for a root
   Terminate the iteration when $|f(P_n)| < \varepsilon$

# Method of False Position

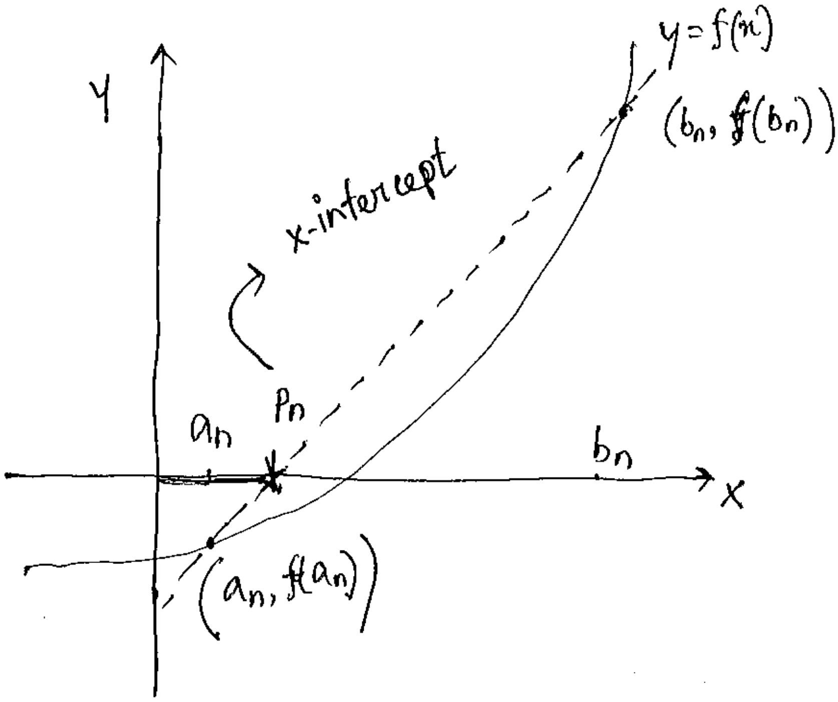
- Unlike the Bisection method, this method is able to compute a fairly accurate estimate for the error.

  - This method also iteratively determines a series of root enclosing intervals. $(a_n, b_n)$

  - Method of false position differs in the process of selecting $P_n$

- Selection of $P_n$

This method uses the x-intercept of the line that passes through points $(a_n, f(a_n))$ and $(b_n, f(b_n))$




Exact equation of the line is:

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

For x-intercept, we always require $y = 0$,

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

$$\Rightarrow \quad P_n = b_n - f(b_n) \cdot \frac{b_n - a_n}{f(b_n) - f(a_n)}$$

Example

Given $f(x) = x^3 + 2x^2 - 3x - 1$

interval $[a, b] \equiv [1, 2]$

First iteration:

$a_1, b_1 = 1, 2$ | $f(a_1) = -1, < 0$

$f(b_1) = 9, > 0$

So,

$$P_1 = b_1 - f(b_1) \cdot \frac{b_1 - a_1}{f(b_1) - f(a_1)}$$

$$= 2 - 9 \cdot \frac{(2-1)}{9+1} = 2 - 9 \cdot \frac{1}{10}$$

$$= 1 \cdot 1$$

Now, check for sign $f(P_1) \equiv f(1\cdot1) \equiv -0\cdot549 < 0$

Since $f(a_1)$ and $f(P_1)$ are of same sign, the intermediate value theorem tells us the zero between $P_1$ and $b_1$. So,

$$(a_2, b_2) \equiv (P_1, b_1)$$

unchanged

replaced by $P_1$.

Now, Second iteration

$$P_2 = 2 - 9 \frac{2 - 1\cdot1}{9 - (-0\cdot549)}$$

$$= 1\cdot151743638$$

□ Now, because one endpoint of the closing interval remain fixed while the other is just $P_{n-1}$, Eq. 2 becomes:

$$e_n = P_n - P$$
$$\approx \lambda \, e_{n-1} \longrightarrow (P_{n-1} - P)$$

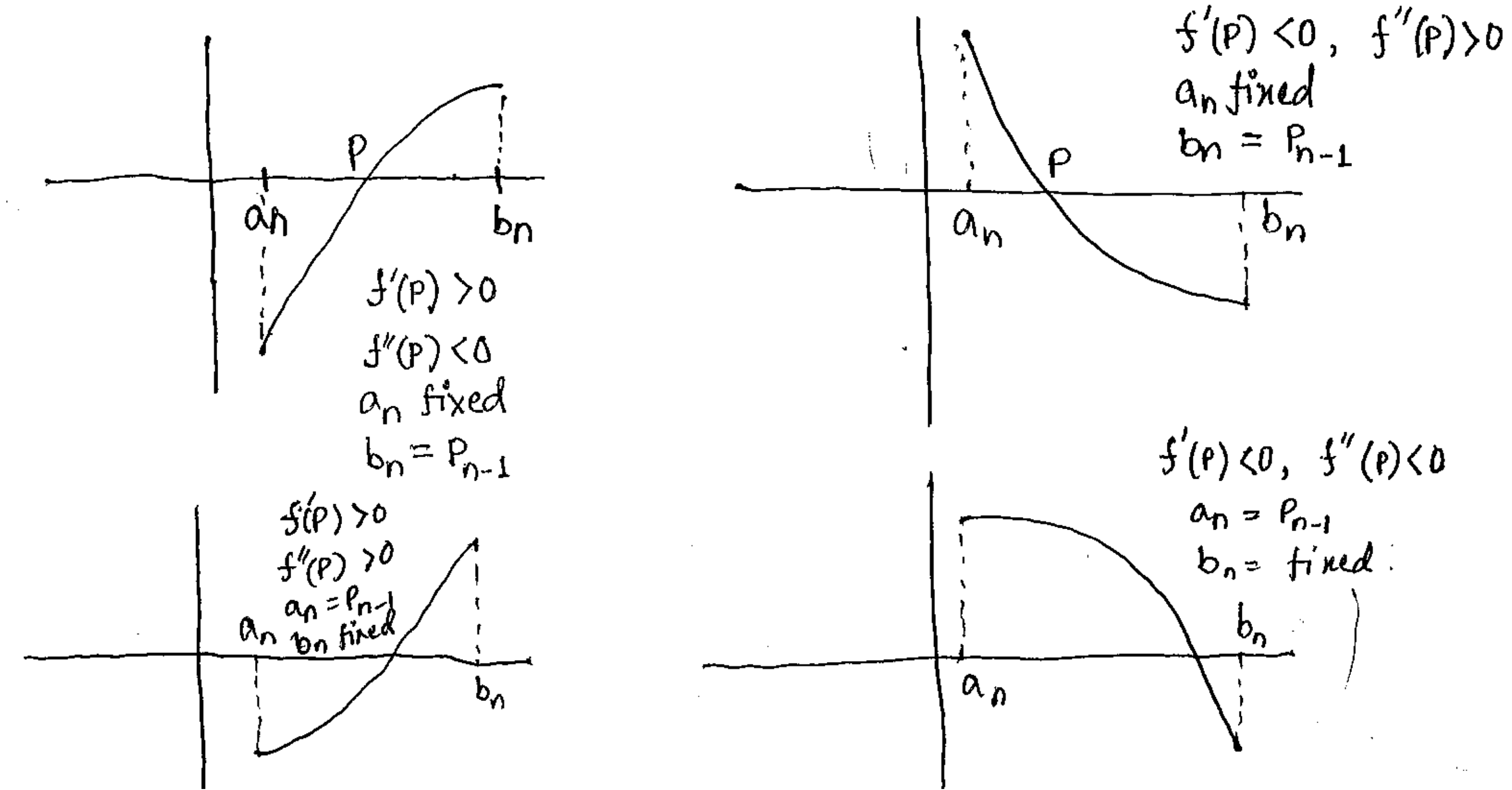
with
$$\lambda = \frac{\ell f''(P)}{2f'(P) + \ell f''(P)}$$

and
$$\ell = \begin{cases} a_n - P, & a_n \text{ remains fixed} \\ b_n - P, & b_n \text{ remains fixed} \end{cases}$$

so,
$$e_n \approx \lambda \cdot e_{n-1}$$

↳ Provided $|\lambda| < 1$, error sequence generated by false position method converges.

□ Eventual configurations for the method of false position



# False Position Vs. Bisection

This is highly problem oriented, that is, contex dependent.

For some problems, $f(x) = \tan(\pi x) - x - 6$, $(0.4, 0.48)$ — Bisection method requires less number of fn' evaluation to ensure the level of accuracy sought.

For other problems — Method of false position may work well.

$f(x) = x^3 + 2x^2 - 3x - 1$ — To ensure the absolute error less than $5\times10^{-5}$, bisection takes 15 fn' evaluations, whereas the false position takes about 5 evaluations.

Generally, there's no theory to tell which method is better for a given problem, but

One main advantage the Method of False position has over the bisection method is the existence of a computable error estimate.

NEXT CLASS

# False Position Vs. Bisection Method

✓ False position method uses more information about the fnc.

↓

But it does not always outperform Bisection method.

✓ Performance is context dependent.

→ for some problem Bisection is better

→ for some problem False Position.

* However, the false position method provides a computable error estimate.

Error sequence:

For root $P$, let's assume error sequence $\{e_n\}$ and the approximation $\{P_n\}$.

$$e_n = P_n - P$$

Now, $\{P_n\}$ converges if and only if $|e_n| \to 0$ as $n \to \infty$

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

$$= b_n - P - f(b_n)\frac{b_n - a_n}{f(b_n) - f(a_n)} \quad \text{-----} \ (1)$$

You can approximate $f(b_n)$, $f(a_n)$ using Taylor polynomial.

$$f(a_n) = f(p) + f'(p)(a_n - p) + f''(p)\frac{(a_n-p)^2}{2!} + \text{higher order terms}$$

$$f(b_n) = f(p) + f'(p)(b_n - p) + f''(p)\frac{(b_n-p)^2}{2!} + \text{higher order terms}$$

Now, $f(p) = 0$, as $p$ is the root. so,

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

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

Then, $f(b_n) - f(a_n) \approx$

$$f'(p)(b_n - a_n) + \frac{f''(p)}{2}\left[(b_n-p)^2 - (a_n-p)^2\right]$$

$$= (b_n - a_n)\left[f'(p) + \frac{f''(p)}{2}(b_n + a_n - 2p)\right]$$

If we substitute $f(a_n)$, $f(b_n)$, $f(b_n) - f(a_n)$ in Eq. 1, ~~and~~ and then factoring the term $(b_n - p)$ and dividing out the term $b_n - a_n$, we obtain:

$$p_n - p \approx (b_n - p)\left[1 - \frac{f'(p) + \frac{f''(p)}{2}(b_n - p)}{f'(p) + \frac{f''(p)}{2}(b_n + a_n - 2p)}\right]$$

$$\approx (b_n - p)(a_n - p)\frac{f''(p)}{2f'(p) + f''(p)(b_n + a_n - 2p)}$$

Generally, one of the endpoints ~~is~~ remains fixed after each iterations

SEE EXAMPLES

## Stopping criteria

We would like to terminate false position method when $|e_n|$ falls below a specified limit. Let's say,

$$e_n = P_n - P = P_n - P_{n-1} + P_{n-1} - P$$
$$= P_n - P_{n-1} + e_{n-1} \quad \cdots\cdots (1)$$

We know that

$$e_n \approx \lambda e_{n-1}$$
$$\Rightarrow \quad e_{n-1} = \frac{e_n}{\lambda} \quad \cdots\cdots (2)$$

So, by plugging in (2) into (1)

$$e_n = P_n - P_{n-1} + \frac{e_n}{\lambda}$$
$$\Rightarrow \quad e_n\left(1 - \frac{1}{\lambda}\right) = P_n - P_{n-1}$$
$$\Rightarrow e_n\left(\frac{\lambda-1}{\lambda}\right) = P_n - P_{n-1}$$
$$\Rightarrow e_n = (P_n - P_{n-1})\left(\frac{\lambda}{\lambda-1}\right)$$

Therefore,

$$|e_n| = \left|\frac{\lambda}{\lambda-1}\right| \, |P_n - P_{n-1}| \quad \cdots \cdots (3)$$

We have ways to estimate the value of $\lambda$.

Consider ratio:

$$\frac{P_n - P_{n-1}}{P_{n-1} - P_{n-2}} = \frac{(P_n - P) - (P_{n-1} - P)}{(P_{n-1} - P) - (P_{n-2} - P)}$$
$$= \frac{e_n - e_{n-1}}{e_{n-1} - e_{n-2}}$$

$e_n = \lambda e_{n-1}$

$e_{n-2} = e_{n-1}/\lambda$

Finally, you find after further simplification

$$\frac{P_n - P_{n-1}}{P_{n-1} - P_{n-2}} \approx \lambda \quad \cdots \cdots (4)$$

Using Eq. 3 and Eq. 4

We can constitute a computable estimate for $|e_n|$.

That is,

$$|e_n| \approx \left|\frac{\lambda}{\lambda-1}\right| |P_n - P_{n-1}|$$

We know,
$$\lambda = \frac{P_n - P_{n-1}}{P_{n-1} - P_{n-2}}$$

So, the stopping criteria,

$$\left|\frac{\lambda}{\lambda-1}\right| |P_n - P_{n-1}| < \epsilon$$

# Method of False Position Continues ...

For a fn$^c$ $f(x) = x^3 + 2x^2 - 3x - 1$ | interval $[1, 2]$

so, $f(a_1) = -1$

$f(b_1) = 9$

so, $\underbrace{\text{First approximation}}_{\text{iteration 1}}$ gives

$$P_1 = b_1 - f(b_1)\frac{b_1 - a_1}{f(b_1) - f(a_1)}$$

$$\Rightarrow P_1 = 2 - 9\frac{2-1}{9-(-1)} = 1.1$$

To determine whether zero is contained

$$f(P_1) = -0.549 < 0$$

So, $(a_2, b_2) \equiv (P_1, b_1) \equiv (1.1, 2)$

2$^{nd}$ iteration: $P_2 = 2 - 9\dfrac{2-1.1}{9-(-0.549)} = 1.15174$

$f(P_2) \approx -0.274 < 0 \longrightarrow$ same sign as $f(a_2)$.

So, $(a_3, b_3)$ becomes $(P_2, b_2)$

$b_2 = 2$ (same as $b_1$)

we calculate $P_3$

3$^{rd}$ iteration $P_3 = b_3 - f(b_3) \cdot \dfrac{b_3 - a_3}{f(b_3) - f(a_3)}$

$$= b_2 - f(b_2) \cdot \frac{b_2 - P_2}{f(b_2) - f(P_2)}$$

So, total fn$^c$ evaluation

$$f(a_1),\ f(b_1),\ f(P_2),\ f(P_1)$$

That is $n^{th}$ iteration takes $(n+1)$ fn$^c$ evaluation.