NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Module 1: Root Finding

Learning Objectives

- Bisection method
- Newton's method
- Learn how to analyze these methods
- Understand some extensions of Newton's method

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Sources

- Textbook (Chapter 1: Solving Equations)

- Wiki: “Abel-Ruffini theorem”, [online]:
  https://en.wikipedia.org/wiki/Abel%E2%80%93Ruffini_theorem
- The “Cubic Formula”, [online]:
  http://www.sosmath.com/algebra/factor/fac11/fac11.html

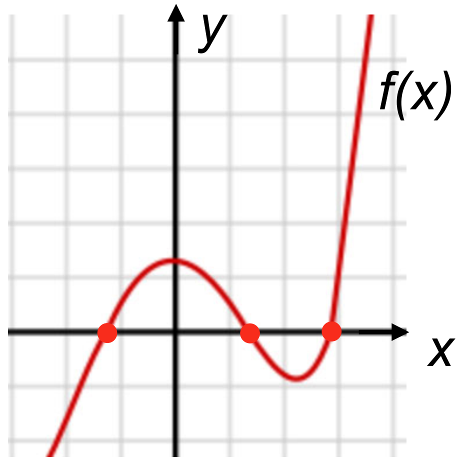
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Outline

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Introduction

- **Basic problem:** given a continuous function $f(x)$, find r such that $f(r) = 0$.
  - r is called a root of $f(x)$.
  - There may exist many roots, but often one of them is of interest in practice.
  - There may not exist a solution.



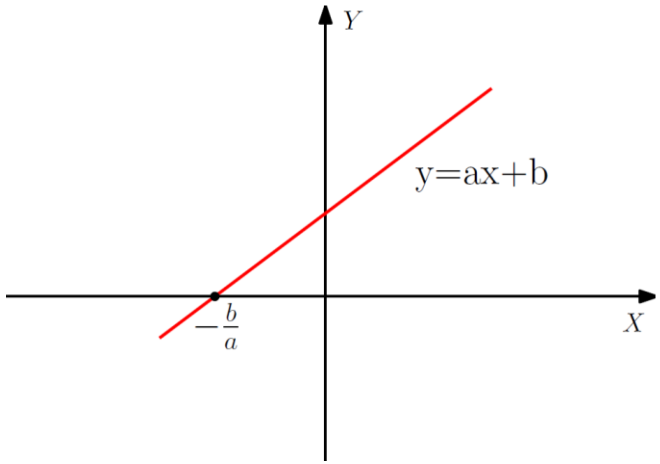
- Serve as a foundation for various applications
  - Many problems can be formulated as an equation.
  - Optimization problems may be converted into root finding

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Simple situations

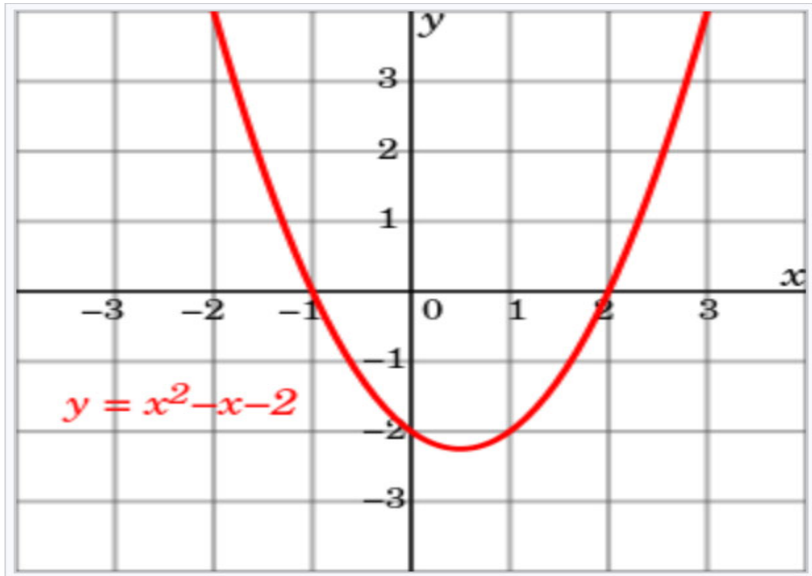
- Linear equation

$$ax + b = 0 \quad \Rightarrow \quad x = -\frac{b}{a}$$



- Quadratic equation

$$ax^2 + bx + c = 0 \quad \Rightarrow \quad x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

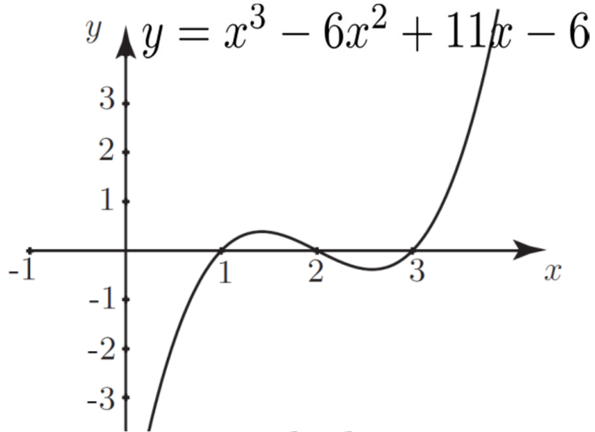


NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Still “simple” situation

- Cubic equation

$$ax^3 + bx^2 + cx + d = 0 \text{ where } a \neq 0$$



A basic idea to solving the cubic equation is first to find one root and then to convert the equation to a quadratic equation. A formula for one root was found by G.Cardano in 1545:

$$\begin{aligned} x &= \sqrt[3]{\left(\frac{-b^3}{27a^3} + \frac{bc}{6a^2} - \frac{d}{2a}\right) + \sqrt{\left(\frac{-b^3}{27a^3} + \frac{bc}{6a^2} - \frac{d}{2a}\right)^2 + \left(\frac{c}{3a} - \frac{b^2}{9a^2}\right)^3}} \\ &+ \sqrt[3]{\left(\frac{-b^3}{27a^3} + \frac{bc}{6a^2} - \frac{d}{2a}\right) - \sqrt{\left(\frac{-b^3}{27a^3} + \frac{bc}{6a^2} - \frac{d}{2a}\right)^2 + \left(\frac{c}{3a} - \frac{b^2}{9a^2}\right)^3}} \\ &- \frac{b}{3a} \end{aligned}$$

Complicated situations

- Quartic equation
  - There does exist an explicit formula (similar to the cubic formula), but it is very complicated.
- Polynomial of degree 5 or higher
  - No such a formula, proved by N.H.Abel in 1824
- Moreover, *f*(*x*) could be a general function, other than a polynomial.
  - E.g. $f(x) = \sqrt{e^x + x^3 - x^2 + x + 10} - 10\sin(2\pi x + 2)$

- **Conclusion:** *effective algorithms are needed!*

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Example

- Problem: find the value of the square root of 3.
- Method 1: construct function $f(x) = x^2-3$.

(1) Let $a = 1$ and $b = 2$. Obviously, $f(a) < 0$, $f(b) > 0$,

and $[a, b]$ contains the required root of $f(x)$.

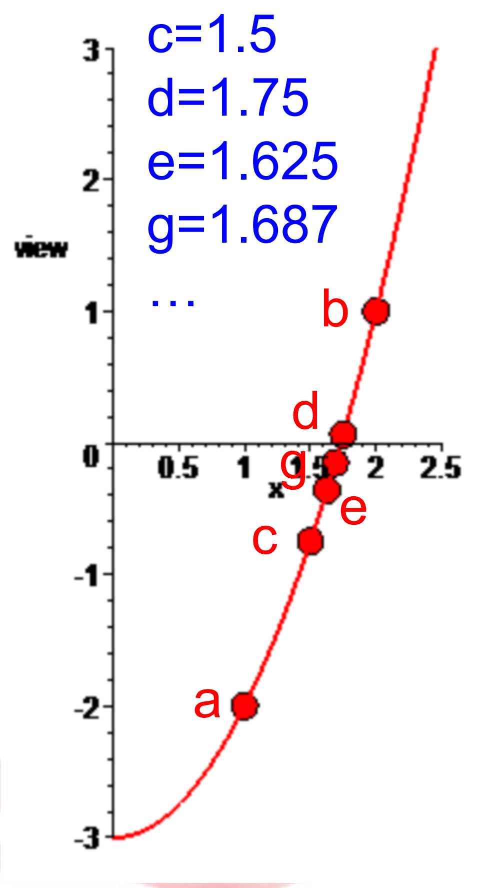
(2) Let $c = \dfrac{a+b}{2}$, and compute $f(c)$.

(3) if $f(c) = 0$, then c is the root. End.

else if $f(c) > 0$, let $b = c$;

else let $a = c$.

iterate steps (2) and (3) until $|a - b| \leq TOL$



NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

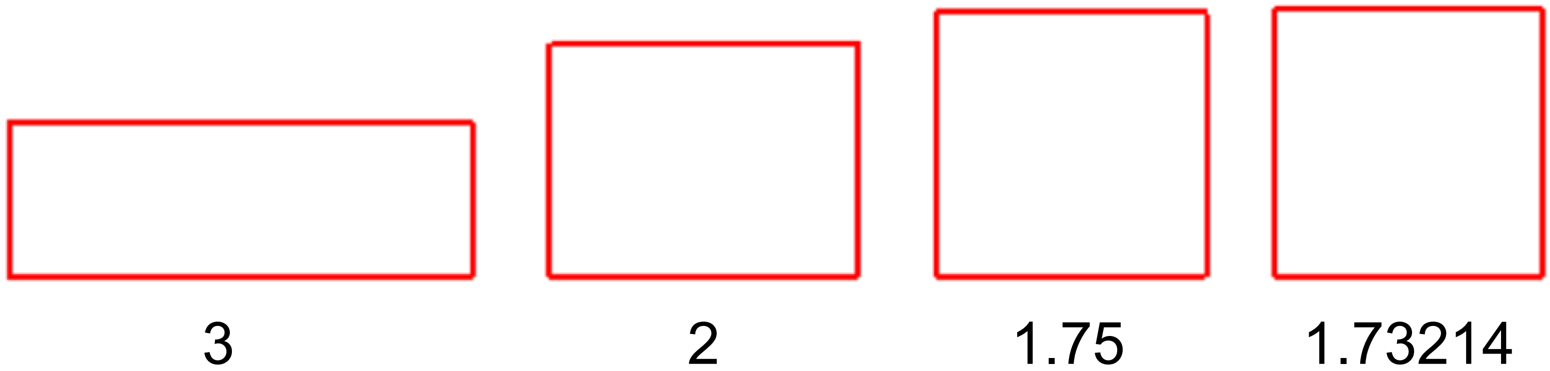
Example (cont)

- Method 2 (geometry approach)

(1) Start with a rectangle whose length $x_0 = 3$ and width $y_0 = 1$. The area of the rectangle is 3.

(2) Successively transform the rectangle into new rectangles with the same area that are closer and closer to a perfect square.

$$x_{i+1} = \frac{x_i + y_i}{2}, \quad y_i = \frac{3}{x_i}.$$

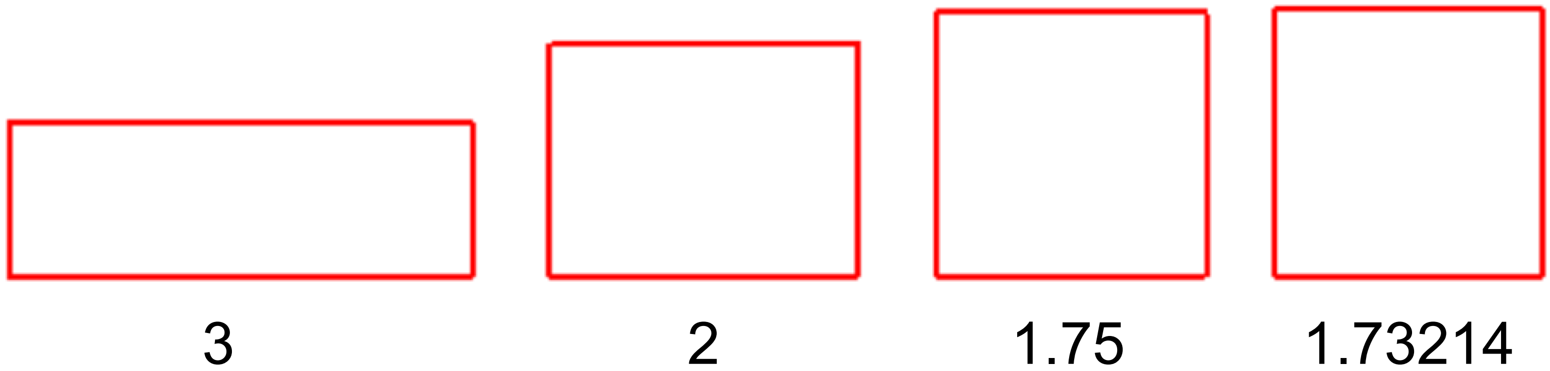


Example (cont)

(3) It is easy to verify that $\{x_i\}$ is decreasing and $\{y_i\}$ is increasing. They converge to the same number that is $\sqrt{3}$, i.e.,

$$\lim_{i \to +\infty} x_i = \lim_{i \to +\infty} y_i = \sqrt{3}.$$

The convergence is rather fast.



Outline

§1. Introduction

§2. Bisection method

§3. Newton's method

§4. Root-finding without derivatives

§5. Homework

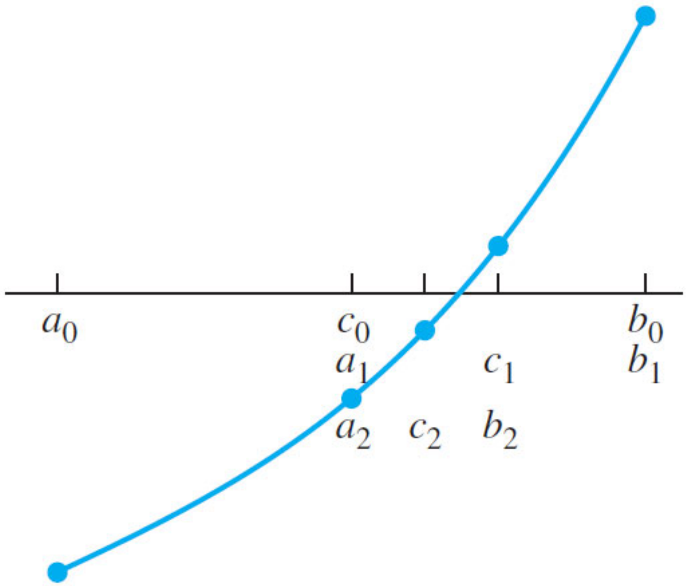
§6. Research problems

§7. Summary

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Bisection method

- Problem: given a continuous function $f(x)$, find r such that $f(r) = 0$.
- Basic idea is to **bracket** the root. That is to find interval $[a,b]$ such that $f(a)f(b) < 0$.
- **Algorithm**



Given initial interval $[a, b]$ such that $f(a)f(b) < 0$
while $(b - a)/2 > \text{TOL}$
  $c = (a + b)/2$
  if $f(c) = 0$, **stop**, **end**
  if $f(a)f(c) < 0$
    $b = c$
  else
    $a = c$
  end
end
The final interval $[a, b]$ contains a root.
The approximate root is $(a + b)/2$.

Example

- Question: find a root of $f(x) = x^3+x-1$ on [0,1].

  Let a = 0, b = 1. Apply the bisection method.

i	a_i	$f(a_i)$	c_i	$f(c_i)$	b_i	$f(b_i)$
0	0.0000	$-$	0.5000	$-$	1.0000	$+$
1	0.5000	$-$	0.7500	$+$	1.0000	$+$
2	0.5000	$-$	0.6250	$-$	0.7500	$+$
3	0.6250	$-$	0.6875	$+$	0.7500	$+$
4	0.6250	$-$	0.6562	$-$	0.6875	$+$
5	0.6562	$-$	0.6719	$-$	0.6875	$+$
6	0.6719	$-$	0.6797	$-$	0.6875	$+$
7	0.6797	$-$	0.6836	$+$	0.6875	$+$
8	0.6797	$-$	0.6816	$-$	0.6836	$+$
9	0.6816	$-$	0.6826	$+$	0.6836	$+$

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

So our best guess is $c_{10} = \dfrac{a_9 + c_9}{2} = 0.6821.$

Accuracy

- Process:
  - $[a,b] = [a_0, b_0] \to [a_1, b_1] \to [a_2, b_2] \to \cdots \to [a_n, b_n]$
  - choose the midpoint $r = \dfrac{a_n + b_n}{2}$

    as an estimate of the root.
- Computational cost: function evaluations = n+2
- Solution error: $|x_{root} - r| < \dfrac{b-a}{2^{n+1}}$
  – which also proves the convergence of the method.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Number of steps *n* = ?

- **Precision:** A solution is correct within *p* decimal places if the error is < 0.5×10^{-p}.
- E.g. If the error < 0.05, it is correct within 1 decimal place.

- To find the number of necessary steps to achieve the desired precision, we let $$\frac{b-a}{2^{n+1}} < \epsilon$$

  which gives $$n > \log_2 \frac{b-a}{2\epsilon}.$$

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Other considerations

- How to find initial [*a*,*b*]?
  - $f(a)f(b) < 0$
  - [*a*,*b*] as tight as possible
  - [*a*,*b*] contains the desired root
  - [*a*,*b*] contains only one root

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Advantages and disadvantages

- Pros
  - Very simple
  - Global convergence

- Cons
  - The error, in general, does not decrease monotonically.
  - Slow convergence rate. In fact, since the length of the intervals halves at each step, it can roughly be considered that the method converges **linearly** with rate ½.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Outline

§1. Introduction

§2. Bisection method

§3. Newton's method

§4. Root-finding without derivatives

§5. Homework

§6. Research problems

§7. Summary

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Newton's method

- Problem: given a C^1 continuous function $f(x)$ and an initial guess x_0, find r such that $f(r) = 0$.
- Basic mathematics

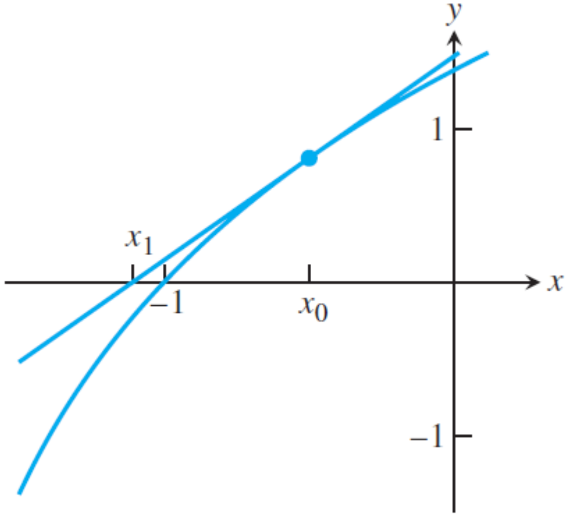
Using Taylor expansion of f about x_0, we have

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0).$$

Letting $f(x) = 0$ gives $x = x_0 - \frac{f(x_0)}{f'(x_0)}$.

- Geometric interpretation

Determining x by intersecting the tangent to the graph of f at $(x_0, f(x_0))$ with the x-axis.

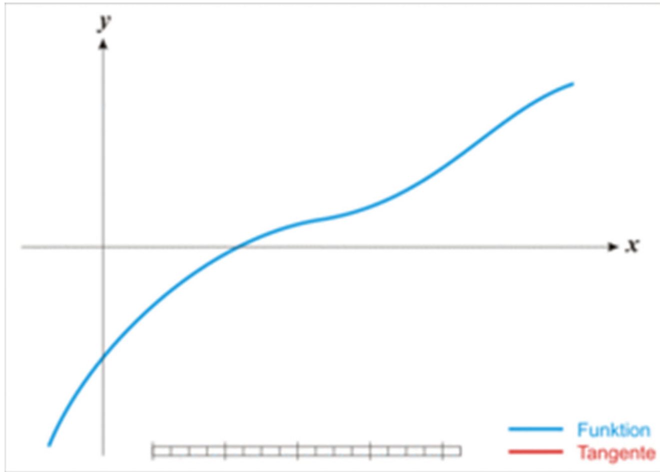


Newton's method

- Algorithm: iteratively generate a sequence $\{x_i\}$ according to the following steps:

$$x_0 = \text{initial guess}$$

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} \quad \textbf{for} \quad i = 0, 1, 2, \ldots.$$



until a stopping criterion is met.

- Stopping criteria: a decision for terminating the algorithm
  - Absolute error: $|x_{i+1}-x_i| <$ tolerance
  - Relative error: $|x_{i+1}-x_i| / |x_{i+1}| <$ tolerance

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Example sqrt(3) revisited

- Problem: find the value of the square root of 3.

(2) Successively transform the rectangle into new rectangles with the same area that are closer and closer to a perfect square.

$$x_{i+1} = \frac{x_i + y_i}{2}, \quad y_i = \frac{3}{x_i}.$$

Interpretation by Newton's method:

Construct function $f(x) = x^2 - 3$. Then $f'(x) = 2x$.

To keep the area, $y_i = \dfrac{3}{x_i}$. Newton's method gives

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} = x_i - \frac{x_i^2 - 3}{2x_i} = \frac{x_i + y_i}{2}$$

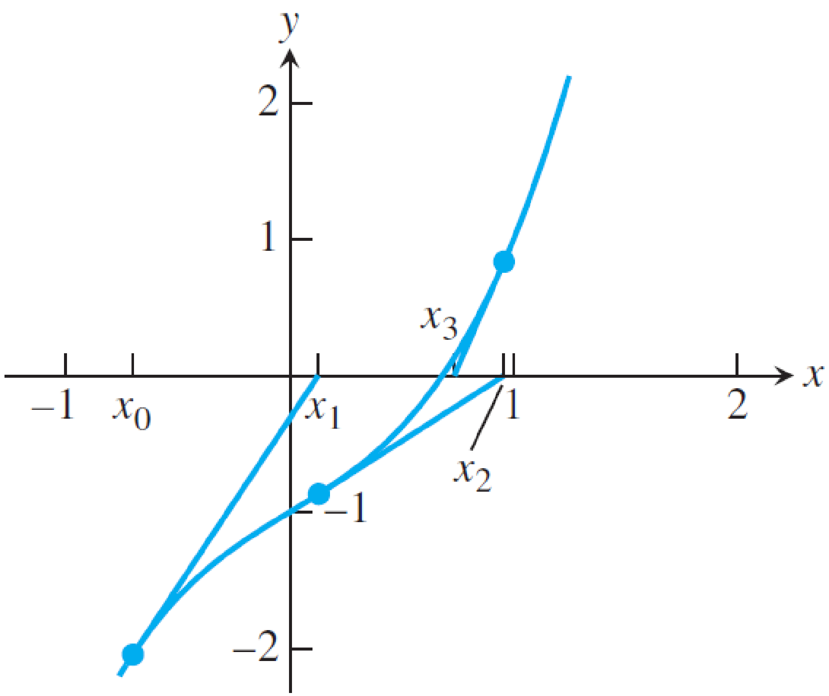
3 2 1.75 1.73214

Example

- Solve $x^3+x-1=0$ with initial guess x_0 = -0.7.

The Newton's formula becomes

$$x_{i+1} = x_i - \frac{x_i^3 + x_i - 1}{3x_i^2 + 1} = \frac{2x_i^3 + 1}{3x_i^2 + 1}.$$



i	x_i	$e_i = \lvert x_i - r \rvert$	e_i / e_{i-1}^2
0	−0.70000000	1.38232780	
1	0.12712551	0.55520230	0.2906
2	0.95767812	0.27535032	0.8933
3	0.73482779	0.05249999	0.6924
4	0.68459177	0.00226397	0.8214
5	0.68233217	0.00000437	0.8527
6	0.68232780	0.00000000	0.8541
7	0.68232780	0.00000000	

Quadratic convergence

- **Definition** Let e_i denote the error at step i of an iterative method. The iteration is **quadratically convergent** if

$$\lim_{i\to\infty} \frac{e_{i+1}}{e_i^2} = M < \infty,$$

- **Theorem** Let $f(x)$ be a C^2 continuous function and $f(r) = 0$. If $f'(r) \neq 0$, then Newton's method is locally and quadratically convergent to r. The error e_i at step i satisfies

$$\lim_{i\to\infty} \frac{e_{i+1}}{e_i^2} = M = \frac{f''(r)}{2f'(r)}.$$

Proof: Refer to Chapter 1.4.1 of the book.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Linear convergence

- **Definition** Let e_i denote the error at step i of an iterative method. If

$$\lim_{i\to\infty} \frac{e_{i+1}}{e_i} = S < 1,$$

the method is said to obey **linear convergence** with rate S.

- **Theorem** Assume $f(x)$ is a C^{m+1} continuous function and $0 = f(r) = f'(r) = \cdots = f^{(m-1)}(r)$, but $f^{(m)}(r) \neq 0$. Then Newton's method is locally convergent to r and the error e_i at step i satisfies

$$\lim_{i\to\infty} \frac{e_{i+1}}{e_i} = S = \frac{m-1}{m}.$$

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

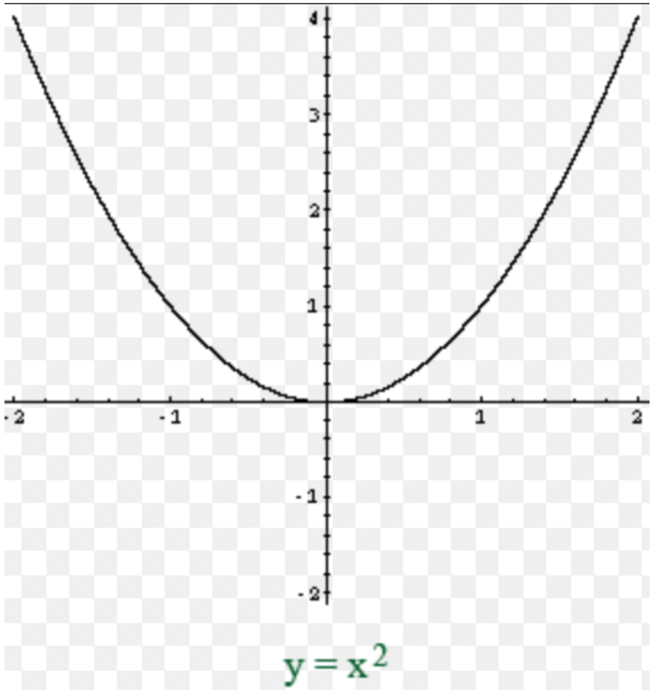
Example of linear convergence

- Apply Newton's method to $f(x) = x^2$ with an initial $x_0 = 1$.

The Newton's formula becomes

$$x_{i+1} = x_i - \frac{x_i^2}{2x_i} = \frac{x_i}{2}.$$

i	x_i	$e_i = \lvert x_i - r \rvert$	e_i / e_{i-1}
0	1.000	1.000	
1	0.500	0.500	0.500
2	0.250	0.250	0.500
3	0.125	0.125	0.500
$\vdots$	$\vdots$	$\vdots$	$\vdots$



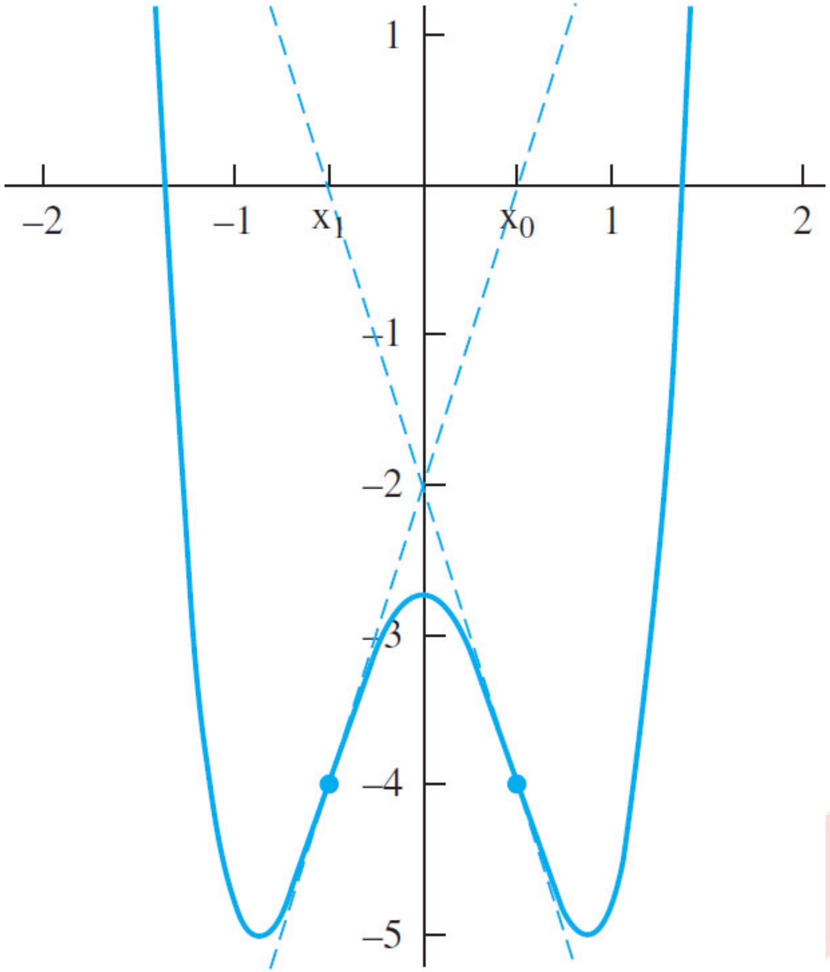
Failure example

- Apply Newton's method to $f(x) = 4x^4 - 6x^2 - 11/4$ with an initial $x_0 = ½$.

The Newton's formula becomes

$$x_{i+1} = x_i - \frac{4x_i^4 - 6x_i^2 - \frac{11}{4}}{16x_i^3 - 12x_i}.$$

The iteration alternates between $\frac{1}{2}$ and $-\frac{1}{2}$.



Advantages and disadvantages

- Pros
  - Quadratic convergence in general

- Cons
  - Initial guess is important.
  - Function behavior affects the method.
  - May not converge
  - May be just linearly convergent

Outline

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Root-finding without derivatives

- Background: **how about the situation where the derivative may not be available?**

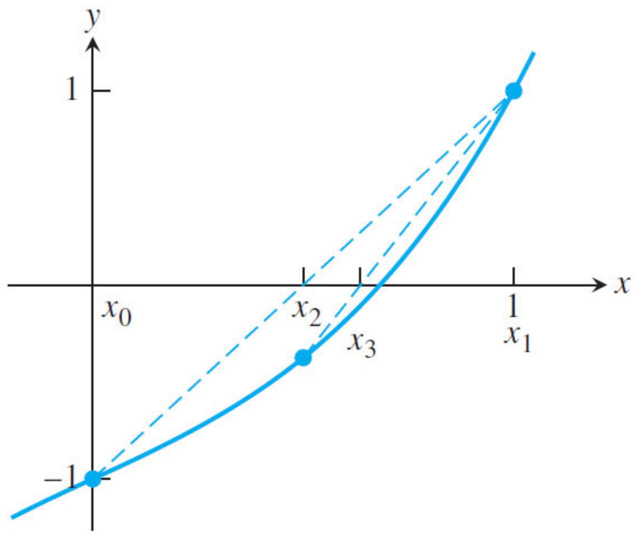
- Substitute for Newton's method
  - Secant method
  - Method of false position (Regula Falsi)
  - Muller's method
  - Inverse Quadratic Interpolation (IQI)
  - Brent's method

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Secant method

- Secant method: replace the derivative by a difference quotient

$$f'(x) \approx \frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}}$$



- Algorithm

$$x_0, x_1 = \text{initial guesses}$$

$$x_{i+1} = x_i - \frac{f(x_i)(x_i - x_{i-1})}{f(x_i) - f(x_{i-1})} \quad \textbf{for } i = 1, 2, 3, \ldots.$$

- Superlinear convergence

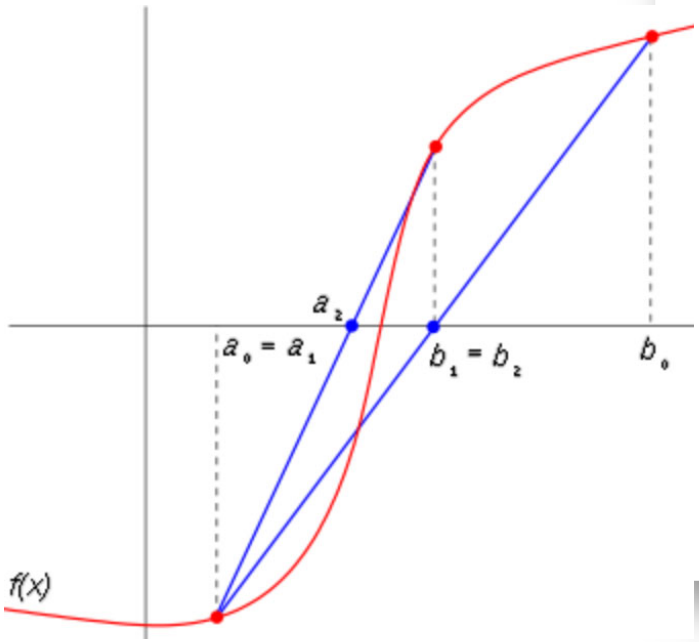
$$e_{i+1} \approx \left| \frac{f''(r)}{2f'(r)} \right|^{\alpha - 1} e_i^{\alpha}$$

where $\alpha = (1 + \sqrt{5})/2 \approx 1.62.$

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Method of False Position

- Basic idea: combine the Bisection method and the Secant method.
  - Use the point obtained from the Secant as the midpoint in the Bisection method
- Algorithm



Given interval $[a, b]$ such that $f(a)f(b) < 0$

for $i = 1, 2, 3, \ldots$

$$c = \frac{bf(a) - af(b)}{f(a) - f(b)} \quad \Leftarrow \quad c = a - \frac{f(a)(a-b)}{f(a) - f(b)}$$

  if $f(c) = 0$, **stop, end**

  if $f(a)f(c) < 0$

    $b = c$

  else

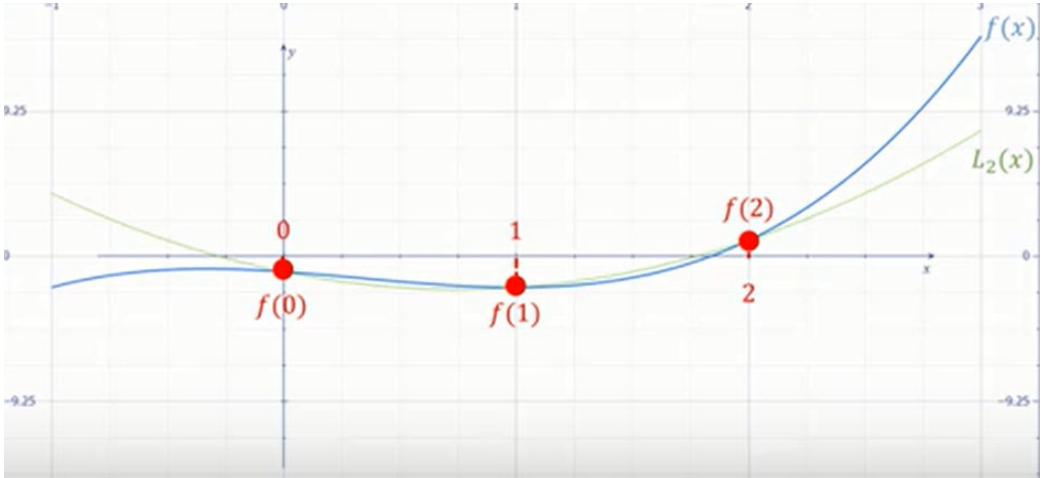
    $a = c$

  end

end

Muller's method

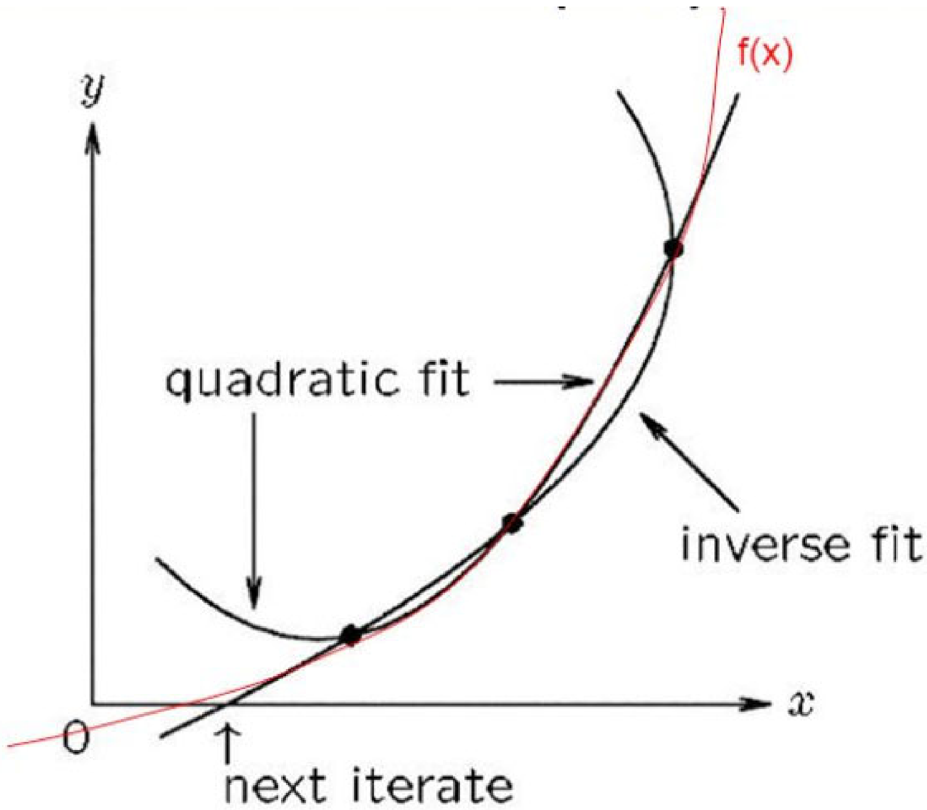
- Basic idea: extend line segments to parabola
  - Instead of intersecting the line through 2 previous points with the *x*-axis, we use 3 previous points x_0, x_1 and x_2 to construct a parabola $y=p(x)$ through them.
  - Intersect the parabola with the *x*-axis
  - If there are two intersection points, choose the one nearest to the last point x_2.



Ref: Muller, David E., "A Method for Solving Algebraic Equations Using an Automatic Computer," *Mathematical Tables and Other Aids to Computation*, 10 (1956), 208-215

Inverse Quadratic Interpolation (IQI)

- Basic idea: extend line segments to parabola of form $x = p(y)$, which avoids the two intersection points with the x-axis in Muller's method
  - use 3 previous points x_0, x_1 and x_2 to construct a parabola $x=p(y)$ through them.
  - the next guess is $x_3 = p(0)$.



Brent's method

- Basic idea: a **hybrid** method that combines the property of guaranteed from the bisection method with the property of fast convergence from other methods.
- The method begins with two points *a* and *b* such that $f(a)f(b) < 0$, and a third point *c*.
  - IQI is attempted and the result is used to replace one of *a*, *b* and *c* if (1) the error improves and (2)the bracketing interval is cut at least in half.
  - If IQI fails, the Secant Method is attempted.
  - If the Secant method fails too, the Bisection method is taken.

- the basis of MATLAB's fzero routine

Outline

§1. Introduction

§2. Bisection method

§3. Newton's method

§4. Root-finding without derivatives

§5. Homework

§6. Research problems

§7. Summary

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Homework

Sheila is a student and she drives a typical student car: it is old, slow, rusty, and falling apart. Recently, the needle on the speedometer fell off. She glued it back on, but she might have placed it at the wrong angle. Thus, when the speedometer reads s, her true speed is $s + c$, where c is an unknown constant (possibly negative). Sheila made a careful record of a recent journey and wants to use this to compute c. The journey consisted of n segments. In the i-th segment she travelled a distance of d_i and the speedometer read s_i for the entire segment. This whole journey took time t. Help Sheila by computing c. Note that while Sheilas speedometer might have negative readings, her true speed was greater than zero for each segment of the journey.

Write a program that takes as input:

- the number n of Sheilas journey segments;
- the total time t (in hours) of her journey;
- the distances d_i (in miles) and speedometer readings s_i (in miles per hour) of the i-th segment, $i = 1, \cdots, n$.

You can assume all input variables to be integers. Compute and print out the constant c (in miles per hour).

Homework (cont)

Hint: use Newton's method.
Test the program for the following input:

- $n = 4$;
- $t = 10$;
- $d_1 = 5, s_1 = 3; d_2 = 2, s_2 = 2; d_3 = 3, s_3 = 6; d_4 = 3, s_4 = 1.$

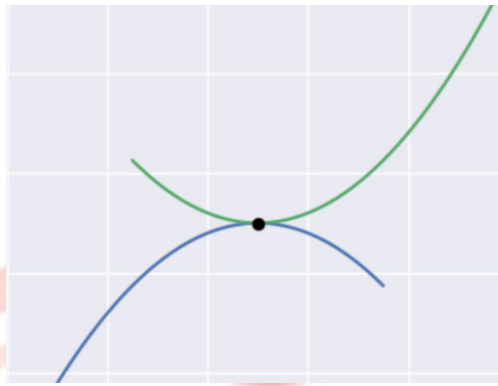
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Outline

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

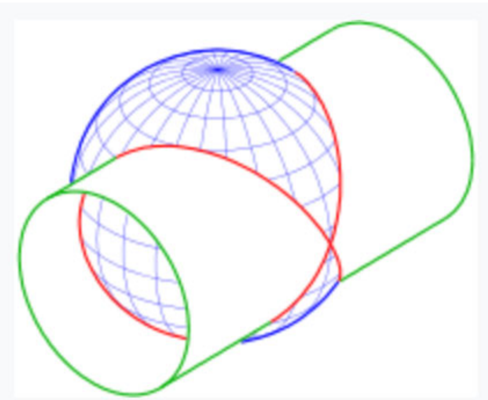
Curve/curve intersection

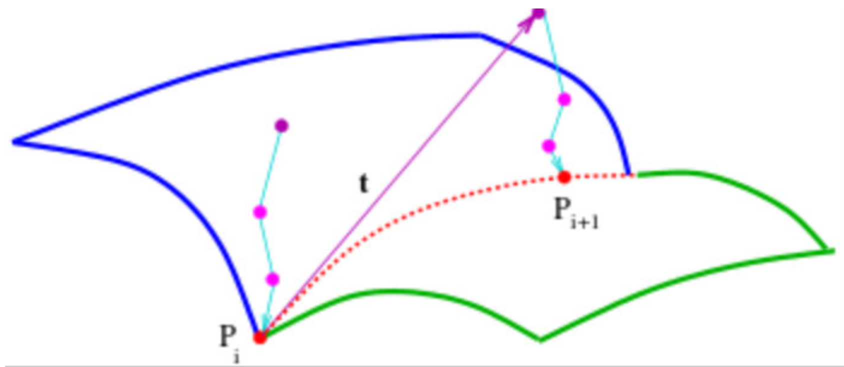
- **Problem:** given two parametric curves $\boldsymbol{r}_1(t) = (x(t), y(t), z(t))$, $t \in [0,1]$, and $\boldsymbol{r}_2(s) = (x(s), y(s), z(s))$, $s \in [0,1]$, find the intersection points of the two curves.
  http://web.mit.edu/hyperbook/Patrikalakis-Maekawa-Cho/node82.html
- Ideas
  - Use "geometric version" of Newton's method: choose two points on the two curves, compute their tangent lines and find the intersections of the lines; project the points on the curves and iterate the process. How to find the initial points?
  - How to use the idea of the bisection method to guarantee the convergence?



Surface/surface intersection

- **Problem:** given two parametric surface $\boldsymbol{r}_1(s,t) = (x(s,t), y(s,t), z(s,t))$, $s,t \in [0,1]$, and $\boldsymbol{r}_2(u,v) = (x(u,v), y(u,v), z(u,v))$, $u,v \in [0,1]$, find the intersection curves of the two surfaces.
  http://web.mit.edu/hyperbook/Patrikalakis-Maekawa-Cho/node99.html
- Ideas
  - Use "geometric version" of Newton's method: choose two points on the two surfaces, compute their tangent planes and find the intersection of the planes with a 3rd plane; project the point on the surfaces and iterate the process.
  - How to use the idea of the bisection method to guarantee the convergence?





NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Outline

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Summary

- Bisection method
  - Choice of initial bracketing interval [a,b]
- Newton's method
  - Initial guess x_0
  - Influence of the function's behavior
- Other methods
  - Secant, regula falsi, Muller, IQI
- Some concepts
  - Precision
  - Convergence rate

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

End