HA NOI UNIVERSITY OF SCIENCE AND TECHNOLOGY
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

CHAPTER 4: SOLVING NONLINEAR EQUATIONS

Vu Van Thieu, Dinh Viet Sang, Nguyen Khanh Phuong

SCIENTIFIC COMPUTING

Introduction

1. Problem
   - Existence and uniqueness of solutions
   - Sensitivity and conditions for solving nonlinear equations
   - Iterative procedure
2. Bisection method
3. Chord method
4. Newton's method
5. Secant method
6. Iterative method
7. Bairstow method
8. Summary

Solve nonlinear equations

Problem

Given the non-linear function $f(x)$, we need to find x satisfying

$$f(x) = 0.$$

The solution x is the solution of the equation and is also called the (zero point) solution of the function $f(x)$. The problem of finding x is called the root finding problem.

Solve nonlinear equations

Examples of problems finding solutions of nonlinear equations

1. $1 + 4x - 16x^2 + 3x^3 - 3x^4 = 0$
2. $\frac{x\sqrt{(2.1-0.5x)}}{(1-x)\sqrt{(1.1-0.5x)}} - 369 = 0$ with $(0 < x < 1)$
3. $tg(x) - tanh(x) = 0$ where $tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$

Solve nonlinear equations

Problem

If the equations $f(x)$ are nonlinear then

- it usually does not have an explicit formula solution
- numerical methods that allow us to find solutions based on *iterative procedure*

Solve nonlinear equations

Solution interval

For function $f : \mathbb{R} \mapsto \mathbb{R}$ the interval $[a, b]$ is called **solution interval** if function f has opposite signs at both ends a, b, i.e. $f(a)f(b) < 0$.

Existence of solutions

If f is a continuous function on the interval $[a, b]$ and $f(a)(f(b) < 0$ then there exists $x^* \in [a, b]$ such that $f(x^*) = 0$.

Solve nonlinear equations

Examples of solutions to nonlinear equations

1. $e^x + 1 = 0$ has no solution
2. $e^{-x} - x = 0$ has a solution
3. $x^2 - 4sin(x) = 0$ has two solutions
4. $x^3 - 6x^2 + 11x - 6 = 0$ has three solutions
5. $cos(x) = 0$ has infinitely many solutions

Solve nonlinear equations

Conditions for solving equations

- The absolute value of the number of conditions x^* of the function $f : \mathbb{R} \mapsto \mathbb{R}$ is $\frac{1}{|f'(x^*)|}$.
- A solution of x^* is said to be *ill-conditioned* (well) if the line tangent to it at the point of coordinates x^* is almost horizontal (vertical).

Solve nonlinear equations

Iterative procedure

Nonlinear equations often do not have an explicit solution. Therefore, to find our solution we often have to use numerical methods based on iterative procedures.

- **Stop condition**: $|f(x_k)| < \epsilon$ or $|x^* - x_k| < \epsilon$ where ϵ is the given *precision* and x_k is the approximate solution obtained at step k
- **Convergence rate**: We denote *error* iteration step k as: $e_k = x_k - x^*$. The sequence $\{e_k\}$ is said to converge to the degree r if

$$\lim_{k\to\infty} \frac{|e_{k+1}|}{|e_k|^r} = C$$

where C is a non-zero constant

Solve nonlinear equations

Iterative procedure

Nonlinear equations often do not have an explicit solution. Therefore, to find our solution we often have to use numerical methods based on iterative procedures.

- **Stop condition**: $|f(x_k)| < \epsilon$ or $|x^* - x_k| < \epsilon$ where ϵ is the given *precision* and x_k is the approximate solution obtained at step k
- **Convergence rate**: We denote *error* iteration step k as: $e_k = x_k - x^*$. The sequence $\{e_k\}$ is said to converge to the degree r if

$$\lim_{k \to \infty} \frac{|e_{k+1}|}{|e_k|^r} = C$$

where C is a non-zero constant

Name of the speed of convergence in some cases

- $r = 1$ - linear convergence rate
- $r > 1$ - linear convergence rate
- $r = 2$ - convergence rate squared

Question

Bisection method

Iterative procedure

Assume that the solution interval $[a, c]$ has only one solution

1. Gradual reduction of the solution interval through division
2. The division to be performed is halving $b = \frac{(a+c)}{2}$
   If $f(b) = 0$ then b is the correct solution to be found, otherwise if $f(b) \neq 0$ we have
   - $f(a)f(b) < 0$ then the new solution interval is $[a, b]$
   - Otherwise, the new interval is $[b, c]$

Steps 1-2 are repeated until the given $[a, c] < \epsilon$

Iterative procedure

Assume that the solution interval $[a, c]$ has only one solution

1. Gradual reduction of the solution interval through division
2. The division to be performed is halving $b = \frac{(a+c)}{2}$
   If $f(b) = 0$ then b is the correct solution to be found, otherwise if $f(b) \neq 0$ we have
   - $f(a)f(b) < 0$ then the new solution interval is $[a, b]$
   - Otherwise, the new interval is $[b, c]$

Steps 1-2 are repeated until the given $[a, c] < \epsilon$

If precision ϵ is given, the number of iterations is an integer n satisfying

$$n \geq \log_2 \frac{ca}{\epsilon}$$

because of

$$\frac{ca}{2^n} < \epsilon$$

Bisection method

Example 1:

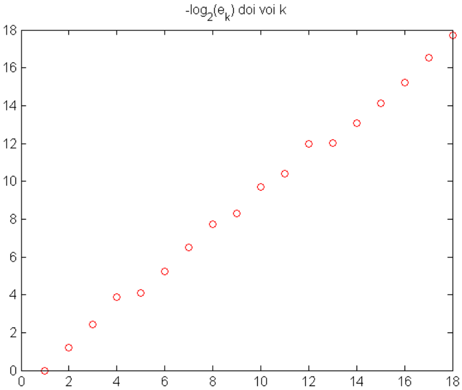
Find the solution of the equation $e^x - 2 = 0$ having the range of solutions [0.2] with precision $\epsilon = 0.01$

Loops	a	b	c	f(a)	f(b)	f(c)	error
1	0.0000	1.0000	20000	-1.0000	0.7183	5.0389	20000
2	0.0000	0.5000	1,00000	-1.0000	-0.3513	0.7183	1,0000
3	0.5000	0.7500	1,0000	-0.3513	0.1170	0.7183	0.5000
4	0.5000	0.6250	0.7500	-0.3513	-0.1318	0.1170	0.2500
5	0.6250	0.6875	0.7500	-0.1318	-0.0113	0.1170	0.1250
6	0.6875	0.7188	0.7500	-0.0113	0.0519	0.1170	0.0625
7	0.6875	0.7031	0.7188	-0.0113	0.0201	0.0519	0.0313
8	0.6875	0.6953	0.7031	-0.0113	0.0043	0.0201	0.0156
9	0.6875	0.6914	0.6953	-0.0113	-0.00349	0.0043	0.0078

Bisection method

Example 2

Consider the solution of the equation $f(x) = 1/(xe^{-x})$ having a solution interval $[0, 1]$ with precision $\epsilon = 0.00001$



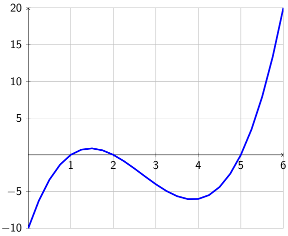
Error:
$e_k = \max\{x^* - a_k, c_k - x^*\}$
Horizontal axis: number of iterations k
Vertical axis: $-\log_2(e_k)$
Apparently $e_k \approx 2^{-k}$

Bisection method

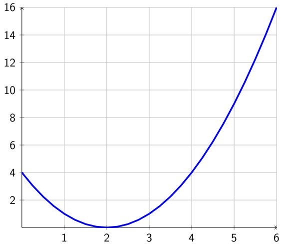
Comment on the split method

- Strengths: Works even with non-analytic functions.
- Weaknesses:
  - Need to determine the range of solutions and find only one solution.
  - Could not find a double solution.
  - When the function f has singularities, the bisection method can treat them as solutions.

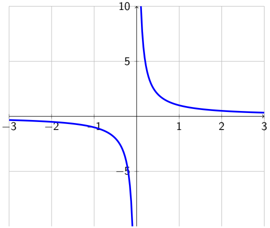
Bisection method



a) Many solutions



b) Double root



c) Singularity

Bisection method

Chord method

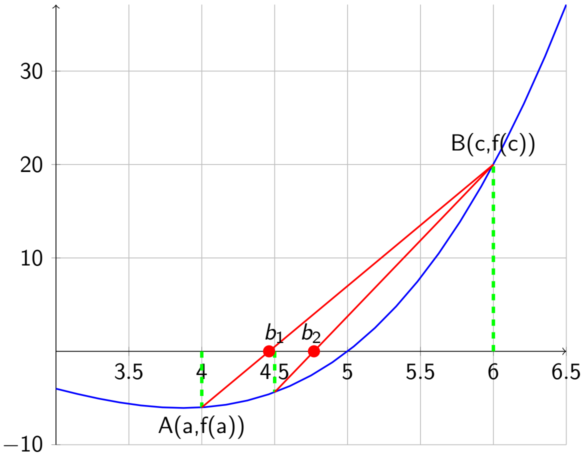
Iterative procedure

Assume that the solution interval $[a, c]$ has only one solution

1. Gradual reduction of the solution interval through division
2. The division to be performed is $b = a - \frac{ca}{f(c)-f(a)}f(a) = \frac{af(c)-cf(a)}{f(c)-f(a)}$ If $f(b) = 0$ then b is the solution to be found. Conversely, if $f(b) \neq 0$, we have:
   - If $f(a)f(b) < 0$ then the new integral is $[a, b]$
   - Otherwise, the new integral is $[b, c]$

Steps 1-2 are repeated until $[a, c] < \epsilon$ is given.
So b is the intersection of the horizontal axis with the line segment connecting A(a,f(a)) to B(c, f(c))

Chord method



Chord method

Comment

- Advantages: like bisection, we do not need the analytic form of the equation f
- Cons:
  - Need to know the solution interval
  - Single-sided convergence is slow, especially when the segment contains large solutions
  - Can be improved using the same halving method

Chord method

Newton's method

Description

The basic idea of the method is to replace the nonlinear equation $f(x) = 0$ with an approximate, linear equation for x. Built on top of Taylor.

Assuming $f(x)$ is continuously differentiable to the order $n+1$ then there exists $\xi \in (a, b)$

$$
\begin{aligned}
f(b) = f(a) + f(ba)f'(a) + \frac{(ba)^2}{2!} f''(a) + \cdots \\
+ \frac{(ba)^n}{n!} f^{(n)}(a) + \frac{(ba)^{(n+1)}}{(n+1)!} f^{(n+1)}(\xi)
\end{aligned}
$$

Newton's method

Description (continued)

Taylor expansion of $f(x)$ at the neighborhood of the original approximate solution x_0:

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + O(h^2)$$

where $h = x - x_0$.
Solve the approximate equation for x:

$$f(x_0) + f'(x_0)(x - x_0) = 0$$

Obtained: $x = x_0 - \frac{f(x_0)}{f'(x_0)}$
x is an incorrect solution, but this solution will be closer to the correct solution than the initial value x_0.

Newton's method

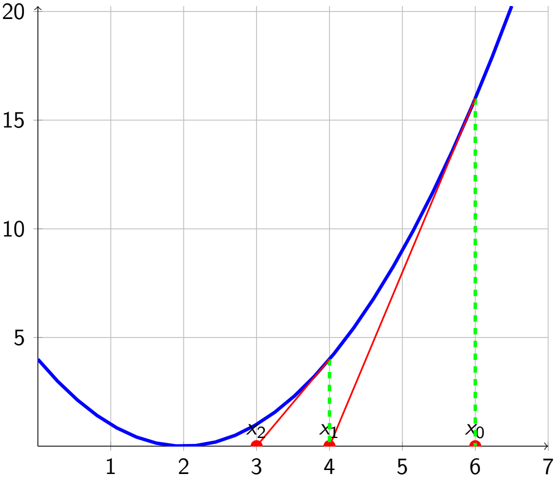
Iterative procedure

1. Initialize with x_0
2. Calculates with $k > 0$

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

3. Repeat step 2 until $|f(x_k)| < \epsilon$ where ϵ is the given precision

Newton's method



Newton's method

Comment

- Advantages:
  - For a smooth enough function and we start from the point near the solution, the convergence rate of the method is squared or $r = 2$
  - No need to know the solution dissociation, just the initial point x_0
- Cons:
  - Need to calculate the first derivative $f'(x_k)$, we can also approximate it with the formula $f'(x_k) = \frac{f(x_k+h)-f(x_k-h)}{2h}$ where h is a very small value eg $h = 0.001$
  - Not always the iterative procedure converges

Newton's method

Example 1

Use Newton's method to find the solution of the equation

$$f(x) = x^2 - 4\sin(x) = 0$$

First derivative of $f(x)$

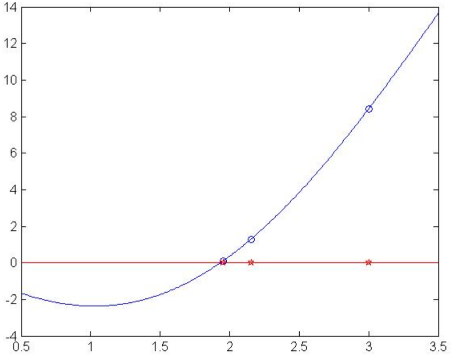
$$f'(x) = 2x - 4\cos(x)$$

The iterative formula of Newton's method is

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} = x_k - \frac{x_k^2 - 4\sin(x_k)}{2x_k - 4\cos(x_k)}$$

Approximate starting point $x_0 = 3$

Newton's method



k	x_k	$f(x_k)$
0	3.000000	8.435520
1	2.153058	1.294773
2	1.954039	0.108439

Newton's method

Example 2

Solve the equation $f(x) = x^2 - 2 = 0$ because $f'(x) = 2x$ so the iterative formula will be $x_{k+1} = x_k - \frac{x_k^2-2}{2x_k}$ error $e_k = x_k - x^* = x_k - \sqrt{2}$

k	x_k	e_k
0	4,000000000	2.5857864376
1	2.250000000	0.8357864376
2	1.569444444	0.1552308821
3	1.421890364	0.0076768014
4	1.414234286	0.0000207236
5	1.414213563	0.0000000002

Newton's method

Example 3

Solve the equation

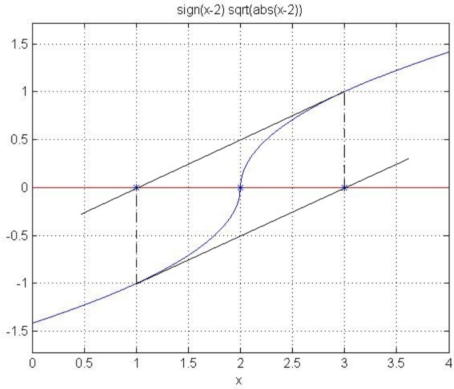
$$f(x) = \mathrm{sign}(xa)\sqrt{|xa|}$$

This equation satisfies:

$$xa - \frac{f(x)}{f'(x)} = -(xa)$$

The zero point of the function is $x^* = a$.
If we draw a tangent to the graph at any point, it always intersects the horizontal axis at the point of symmetry with the line $x = a$.
Newton's method infinitely iterative, neither convergent nor divergent.

Newton's method



Newton's method

Secant method

Iterative procedure

An improvement of Newton's method, instead of using the tangent $f'(x)$, we use an approximate difference based on two successive iterations.

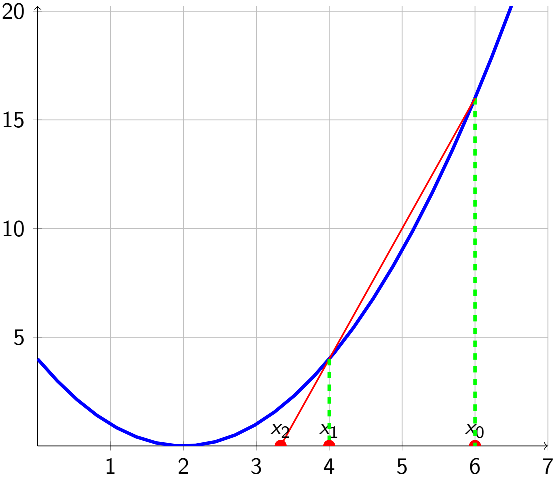
1. Starts with two starting points x_0 and x_1
2. With $k \geq 2$, we iterate by the formula

$$s_k = \frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}}$$

$$x_{k+1} = x_k - \frac{f(x_k)}{s_k}$$

3. Repeat step 2 until $|f(x_k)| < \epsilon$ given small positive.

Secant method



Secant method

Comment

- Advantages:
  - No need to know the solution dissociation, just two initial points x_0 and x_1
  - No need to calculate first derivative $f'(x_k)$
- Cons:
  - Two initialization points are required
  - Convergence rate of method on linear $1 < r < 2$, specifically golden ratio $r \approx \frac{1+\sqrt{5}}{2} = 1.618$

Secant method

Example 1

Solve the equation

$$f(x) = \mathrm{sign}(x-2)\sqrt{|x-2|} = 0$$

with two starting points $x_0 = 4$, $x_1 = 3$ and $\epsilon = 0.001$

$$s_k = \frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}}$$
$$= \frac{\mathrm{sign}(x_k-2)\sqrt{|x_k-2|} - \mathrm{sign}(x_{k-1}-2)\sqrt{|x_{k-1}-2|}}{x_k - x_{k-1}}$$

$$x_{k+1} = x_k - \frac{f(x_n)}{s_n} = x_k - \frac{\mathrm{sign}(x_k-2)\sqrt{|x_k-2|}}{s_k}$$

Secant method

k	x_k	e_k
0	4,000000000	2,000,000000000
1	3,000000000	1,000,000,000,000
2	0.585786438	1.4142135624
3	1.897220119	0.1027798813
$\vdots$	$\vdots$	$\vdots$
26	1.999989913	0.0000100868
27	1.999998528	0.0000014716
28	2,000003853	0.0000038528
29	2.000000562	0.0000005621

Secant method

Example 2

Solve the equation

$$f(x) = e^{-x/2}\sin(3x) = 0$$

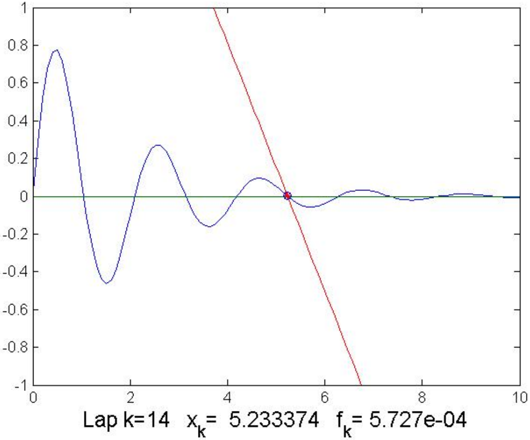
with two starting points x_0, x_1 and precision ϵ entered from the keyboard

$$s_k = \frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}}$$

$$= \frac{e^{-x_k/2}\sin(3x_k) - e^{-x_{k-1}/2}\sin(3x_{k-1})}{x_k - x_{k-first}}$$

$$x_{k+1} = x_k - \frac{f(x_n)}{s_n} = x_k - \frac{e^{-x_k/2}\sin(3x_k)}{s_k}$$

Secant method



Two starting points

$x_0 = 4$

$x_1 = 5$

Accuracy

$\epsilon = 0.001$

Secant method

Iterative method

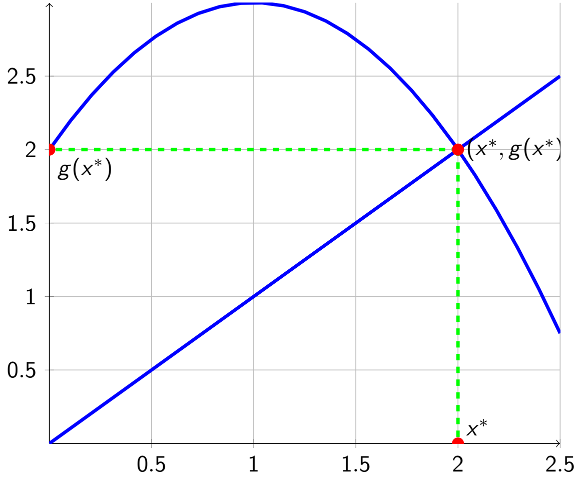
Fixed point

Instead of writing the equation as $f(x) = 0$, we rewrite it as a problem

$$\text{Find x satisfying } x = g(x)$$

The point x^* is called *fixed point* of the function $g(x)$ if $x^* = g(x^*)$, i.e. the point x^* is not transformed by g . mapping

Iterative method



Iterative method

Examples

- Newton's method, according to the formula $x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$, the function g that we need to find the fixed point x^* would be $g(x) = x - f(x)/f'(x)$
- Find the solution $f(x) = x - e^{-x} \Rightarrow g(x) = e^{-x}$
- Find the solution $f(x) = x^2 - x - 2 \Rightarrow g(x) = \sqrt{x+2}$ or $g(x) = x^2 - 2$
- Find the solution $f(x) = 2x^2 - x - 1 \Rightarrow g(x) = 2x^2 - 1$

Iterative method

Iterative procedure

Approach to problem solving

$$x_{k+1} = g(x_k) \text{ with } k = 1, 2, \cdots$$

The above iterative procedure is often called an iterative **find fixed point** with a given starting point x_1

Iterative method

Comment

- Advantages:
  - No need to know the solution interval
- Cons:
  - does not always converge

Iterative method

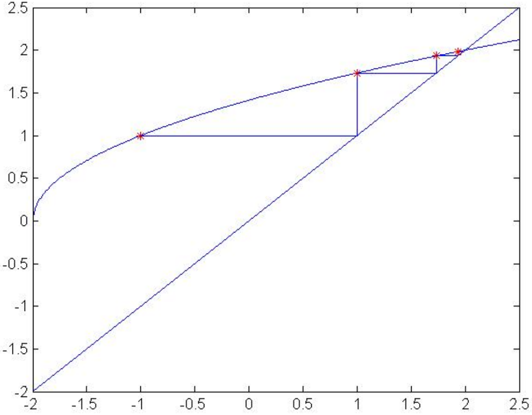
Example 1

Finding the solution of the equation $f(x) = x^2 - x - 2 = 0$ can lead to finding a fixed point

$$g(x) = \sqrt{x + 2}$$

Approximate starting point $x_1 = -1$

Iterative method



Starting point

$x_1 = -1$

Number of iterations

$n = 3$

Iterative method

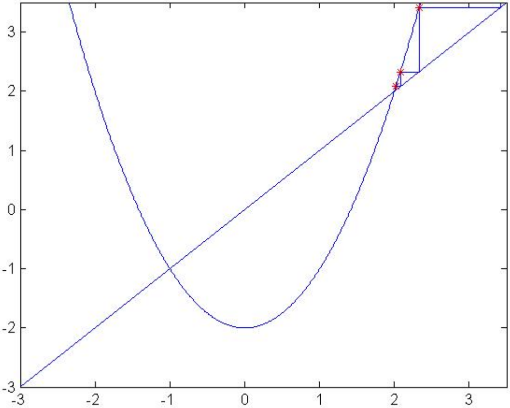
Example 2

Find the solution of the equation $f(x) = x^2 - x - 2$ by finding the point of disagreement of the function

$$g(x) = x^2 - 2$$

The starting point $x_1 = 2.02$ is very close to the solution

Iterative method



Starting point

$x_1 = 2.02$

Number of iterations

$n = 50$

Iterative method

Convergence theorem of iterative methods

Theorem 1: Assume the function $g(x)$ is continuous and the sequence repeats

$$x_{k+1} = g(x_k), k = 1, 2, \cdots$$

then if $x_k \to x^*$ when $k \to \infty$ then x^* is the fixed point of g.

Theorem 2: Suppose $g \in C^1$ and $|g'(x)| < 1$ in some interval containing the fixed point x^*. If x_0 is also in this range, the iterative sequence $\{x_k\}$ converges to x^*.

Theorem 3: If g is a **co function** then it has only one fixed point and the iterative sequence $\{x_k\}$ converges to x^* for all points starting x_0.

Note: The function g is called a co function if there is a constant $L < 1$ such that for every x, y we have: $|g(x) - g(y)| < L(xy))$.

Iterative method

Bairstow method

Description

This is the method used to find the solution of a polynomial,:

$$y = a_0 + a_1 x + \cdots + a_N x_N$$

It can be rewritten as a quadratic factor plus a remainder

$$y = (x^2 + px + q)G(x) + R(x)$$

Inside,

- p and q are arbitrary values.
- $G(x)$ is a polynomial of degree $N - 2$
- $R(x)$ is the remainder, usually a first degree polynomial.

Bairstow method

Description (continued)

So the polynomial $G(x)$ and the remainder $R(x)$ have the form

$$G(x) = b_2 + b_3 x + b_4 x^2 + \cdots + b_N x^{N-2}$$
$$R(x) = b_0 + b_1 x$$

The value of b_0 and b_1 depends on the choice of p and q, the goal is to find $p = p^*$ and $q = q^*$ such that

- $b_0(p^*, q^*) = b_1(p^*, q^*) = 0 \Rightarrow R(x) = 0$
- $(x^2 + p^* x + q^*) \Rightarrow$ square factor of y

Bairstow method

Iterative procedure

1. Initializes p and q and calculates b_0 and b_1 (see ct in the book)
2. Calculates the values $(b_0)_p$,$(b_1)_p$ and $(b_0)_q$,$(b_1)_q$ (see ct in the book)
3. Find Δp and Δq when solving equation (9)
4. Obtained p^* and q^* by the formula $p^* = p + \Delta p$ and $q^* = q + \Delta q$

Bairstow method

Comment

- Advantages:
  - method that converges to the quadratic factor $(x^2 + px + q)$ regardless of the initialization value p, q
  - the coefficients of the polynomial $G(x)$ are also automatically obtained
- Cons:
  - the accuracy of the obtained test is not high
  - to improve, you can use Newton's method to recalculate each solution

Summary

Methods Dharma	Ranges Dissociation solutions	Requirements Disparity of Tao first order functions	Styles Oriental equations	Features Special
Split function	Yes	No	Any	Apply to the has no analytic fo
Chord	Yes	Yes	Any	Slow Convergence large separation
Newton	No	Yes	Any	Fast Convergence Need to calculate
Cat route	No	Yes	Any	nt
Repeat	No	Yes	Any	May not converge
Bairstow	No	Yes	Polynomials	Factorials of 2nd c Can find complex

Methods Dharma	Ranges Dissociation solutions	Requirements Disparity of Tao first order functions	Styles Oriental equations	Features Special
Split function	Yes	No	Any	Apply to the has no analytic fo
Chord	Yes	Yes	Any	Slow Convergence large separation
Newton	No	Yes	Any	Fast Convergence Need to calculate
Cat route	No	Yes	Any	nt
Repeat	No	Yes	Any	May not converge
Bairstow	No	Yes	Polynomials	Factorials of 2nd c Can find complex

The first two methods (division, chord) both require a solution dissociation interval. The Newton method and the linear sand need to have an initial guess. The iterative method has the problem that it does not always converge. The Bairstow method is limited to solving polynomial equations, which may also not converge.

Read more

Functions to find solutions in Matlab

- **X = roots(C)** find polynomial roots
- **X = fzero(FUN,X0)** find solutions to nonlinear equations

VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Thank you for your attentions!

solct.hust.edu.vn/ fb.com/groups/solct