Chapter 1. Find roots of an equation

1

Numerical errors

- Propagated vs. computational error
  - x = exact value, y = approx. value
  - F = exact function, G = its approximation
  - G(y) – F(x) = [G(y) - F(y)] + [F(y) - F(x)]

Total error = Computational error: affected by algorithm + Propagated data error: not affected by algorithm

- Rounding vs. truncation error
  - Rounding error: introduced by finite precision calculations in the computer arithmetic
  - Truncation error: introduced by algorithm via problem simplification, e.g. series truncation, iterative process truncation etc.

Computational error = Truncation error + rounding error

2

Truncation errors

- Truncation errors are problem specific
- Often, every step involves an approximation, e.g. a finite Taylor series
- The truncation errors accumulate
- Often, truncation errors can be calculated

Example 1 $\sin(x) \approx x - \frac{x^3}{6}$

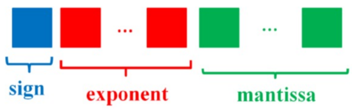
Example 2 Finite difference approximation for computing derivative

$$f'(x_i) = \lim_{h\to 0} \frac{f(x_{i+1}) - f(x_i)}{h} \Rightarrow f'(x_i) \approx \frac{f(x_{i+1}) - f(x_i)}{h}$$

3

Roundoff errors

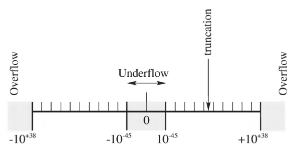
- Precision of representation of numbers is finite



	Sign (s)	Exponent (e)	Fraction (f)	Bias
Single Precision	1 [31]	8 [30-23]	23 [22-00]	127
Double Presicion	1 [63]	11 [62-52]	52 [51-00]	1023

Table: IEEE 754 standard for floating-point arithmetic.

The Limits of Single-precision Floating-Point Numbers



4

Example of roundoff errors

Two roots of the quadratic equation

$$ax^2 + bx + c = 0\ .$$

are

$$x_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \quad \text{and} \quad x_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a}\ .$$

When $b^2 \gg |ac|$, there is the danger of a subtractive cancellation in one of the expression.

We can rewrite the expression as

$$x_1 = \frac{q}{a} \quad \text{and} \quad x_2 = \frac{c}{q}\ .$$

with

$$q = -\frac{1}{2}\left[b + \mathrm{sgn}(b)\sqrt{b^2 - 4ac}\right]\ .$$

e.g., $0.001x^2+1000x+0.001=0$

5

数值计算应注意的问题

1.避免相近二数相减，避免大数和小数相加

两个相近数的前几位有效数字是相同的，相减后有效数字位会大大减少。例如，√1001≈31.64， √1000≈31.62，求（√1001-√1000)的值，直接相减结果为0.02（0.01580），只有一位有效数字，计算中损失了3位有效数字。

$$\sqrt{1001} - \sqrt{1000} = \frac{1}{\sqrt{1001}+\sqrt{1000}} \approx 0.01581$$

c: 计算机里的表示　　ε: 相对误差

$$a = b - c \Rightarrow a_c \simeq b_c - c_c \simeq b(1+\epsilon_b) - c(1+\epsilon_c)$$

$$\Rightarrow \frac{a_c}{a} \simeq 1 + \epsilon_b \frac{b}{a} - \frac{c}{a}\epsilon_c\ .$$

If $b \approx c$

$$\frac{a_c}{a} \overset{\text{def}}{=} 1 + \epsilon_a \simeq 1 + \frac{b}{a}(\epsilon_b - \epsilon_c) \simeq 1 + \frac{b}{a}\max(|\epsilon_b|, |\epsilon_c|)$$

6

2.避免小分母溢出　$\int dx \sin x/x$

3.避免大数溢出　$\frac{e^x}{e^x+1} = \frac{1}{1+e^{-x}}$

4.乘法快于除法： *0.5 v.s. /2.0

5. 最底层循环尽量优化：

```
do i=1,10000
 a=a+2.0*pi
end do
```

v.s.

```
b=2.0*pi
do i=1,10000
 a=a+b
end do
```

7

Roots of an equation

In physics, we often encounter situations in which we need to find the possible value of x that ensures the equation f(x)=0, where f(x) can either be an explicit or an implicit function of x. If such a value exists, we call it a root or zero of the equation.

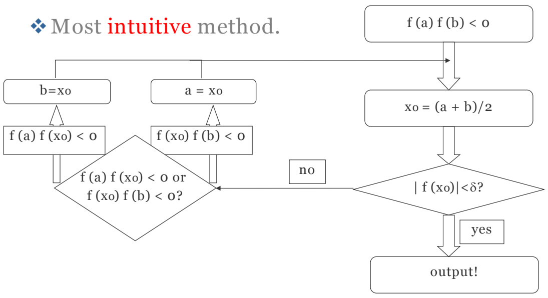
If we need to find a root for f(x)=a, then how?

define g(x)=f(x)-a, and find a root for g(x)=0.

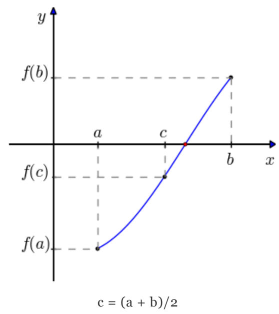
8

Bisection method

- If we know that there is a root x_r in the region [a,b] for f(x)=0, we can use the bisection method to find it within a required accuracy.
- Most intuitive method.

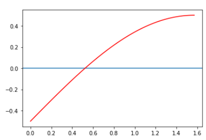


9



10

Code Example



- f(x)=sin(x)-0.5; x is within 0 to π/2.
- Analytically, we know the root is π/6.
- Numerically, the procedure is:

since [sin(0)-0.5]*[sin(π/2)-0.5]<0 and [sin(0)-0.5]*[sin(π/4)-0.5]<0, but [sin(π/2)-0.5]*[sin(π/4)-0.5]>0;

- then the root must be within (0,π/4).
- Then we calculate the value at π/8.
- ……
- Bisection.cpp

11

Bracket finding Strategies

- Graph the function
  - May require 1000's of points and thus function calls to determine visually where the function crosses the x-axis
- "Exhaustive searching" – a global bracket finder
  - Looks for changes in the sign of f(x)
  - Need small steps so that no roots are missed
  - Need still to take large enough steps that this doesn't take forever

12

Error Analysis and Convergence Criterion

- absolute value of the difference (ε_d)

$$\varepsilon_d = \left| x_{m,i+1} - x_{m,i} \right|$$

- relative percent error (ε_r)

$$\varepsilon_r = \left| \frac{x_{m,i+1} - x_{m,i}}{x_{m,i+1}} \right| \times 100$$

- true error (ε_t) in the ith iteration

$$\varepsilon_t = \left| \frac{x_t - x_{m,i}}{x_t} \right| \times 100$$

13

Convergence analysis of bisection method

At each iteration the interval $[a_i, b_i]$ is divided into halves

$$\varepsilon_d^i = \left| x_{m,\,i+1} - x_{m,\,i} \right| \le (b_i - a_i)$$

$$\varepsilon_d^{i+1} = \left| x_{m,\,i+2} - x_{m,\,i+1} \right| \le (b_{i+1} - a_{i+1}) = \frac{1}{2}(b_i - a_i)$$

$$\frac{\varepsilon_d^{i+1}}{\varepsilon_d^i} \approx \frac{1}{2}$$

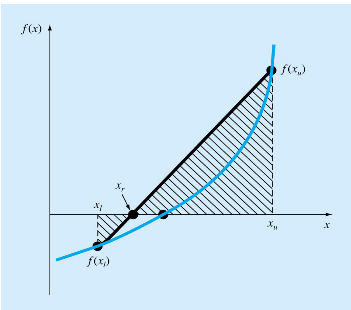
Rate of convergence is linear

14

Regula-Falsi Method (False-Position)

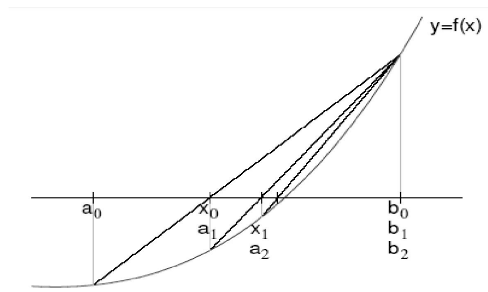
- We can approximate the solution by doing a *linear interpolation* between $f(x_u)$ and $f(x_l)$
- Find x_r such that $l(x_r)=0$, where $l(x)$ is the linear approximation of $f(x)$ between x_l and x_u
- Derive x_r using similar triangles

$$x_r = \frac{x_l f_u - x_u f_l}{f_u - f_l}$$



15

Iterations in Regula-Falsi Method



16

The Newton method

i.e., Newton-Raphson

Based on Taylor expansion

$$f(x_{i+1}) = f(x_i) + f'(x_i)h + \frac{f''(x_i)h^2}{2!} + \frac{f^{(3)}(x_i)h^3}{3!} + \ldots\ldots + \frac{f^{(n)}(x_i)h^n}{n!} + R_n$$

Where:

$$h = x_{i+1} - x_i$$

R_n is the remainder term to account for all terms from n+1 to infinity.

$$R_n = \frac{f^{(n+1)}(\xi)}{(n+1)!} h^{n+1}$$

And ξ is a value of x that lies somewhere between x_i and x_{i+1}

17

The Newton method

i.e., Newton-Raphson

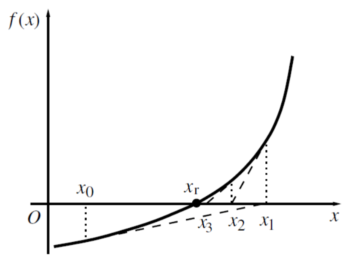
This method is based on linear approximation of a smooth function around its root. We can formally expand the function $f(x_r) = 0$ in the neighborhood of the root x_r through the Taylor expansion.

$$f(x_r) \approx f(x) + (x_r - x) f'(x) + \cdots = 0$$

where x can be viewed as a trial value for the root of x_i at the ith step and the approximate value of the next step x_{i+1} can be derived.

18

$$x_{i+1} = x_i + \Delta x_i = x_i - f_i / f_i'$$
(i = 0, 1, ...)



19

Code example

Example:

f(x)=sin(x)-0.5; f'(x)=cos(x)

i	x_i	f_i	f_i'
0	0	-0.5	1
1	0.5	……	……

❖ NewtonRoot.cpp

20

Newton Method – Convergence

$$0 = f(x^*) = f(x_k) + \frac{df(x_k)}{dx}(x^* - x_k) + \underbrace{\frac{1}{2}\frac{d^2 f(\tilde{x})}{dx^2}(x^* - x_k)^2}$$

Exact root

Mean Value theorem truncates Taylor series

But

$$0 = f(x_k) + \frac{df(x_k)}{dx}(x_{k+1} - x_k)$$

by Newton definition

21

Newton Method – Convergence

Subtracting
$$\frac{df(x_k)}{dx}(x_{k+1} - x^*) = \frac{1}{2}\frac{d^2 f(\tilde{x})}{dx^2}(x_k - x^*)^2$$

Dividing through
$$(x_{k+1} - x^*) = \frac{1}{2}\left[\frac{df(x_k)}{dx}\right]^{-1}\frac{d^2 f(\tilde{x})}{d^2 x}(x_k - x^*)^2$$

Let $\left|\frac{1}{2}\left[\frac{df(x_k)}{dx}\right]^{-1}\frac{d^2 f(\tilde{x})}{d^2 x}\right| = K_k$

then $\left|x_{k+1} - x^*\right| \le K_k \left|x_k - x^*\right|^2$

Convergence is quadratic

22

Newton Method – Convergence

Local Convergence Theorem

If

$$\left.\begin{array}{ll} a) \ \frac{df}{dx} & \text{bounded away from zero} \\ b) \ \frac{d^2 f}{dx^2} & \text{bounded} \end{array}\right\} K \ \text{is bounded}$$

Then Newton's method converges given a sufficiently close initial guess (and convergence is quadratic)

23

Newton Method – Convergence

Example 1

$$f(x) = x^2 - 1 = 0, \quad \text{find } x \ (x^* = 1)$$

$$\frac{df(x_k)}{dx} = 2x_k \qquad \frac{d^2 f(\tilde{x})}{dx^2} = 2$$

$$\frac{df(x_k)}{dx}(x_{k+1} - x^*) = \frac{1}{2}\frac{d^2 f(\tilde{x})}{dx^2}(x_k - x^*)^2$$

$$2x_k(x_{k+1} - x^*) = (x_k - x^*)^2$$

$$\text{or } (x_{k+1} - x^*) = \frac{1}{2x_k}(x_k - x^*)^2$$

Convergence is quadratic

24

Newton Method – Convergence

Example 2

$f(x) = x^2 = 0, \quad x^* = 0$

Note : $\left(\frac{df}{dx}\right)^{-1}$ not bounded away from zero

$$\frac{df(x_k)}{dx} = 2x_k \qquad \frac{d^2 f(\tilde{x})}{dx^2} = 2$$

$$\frac{df(x_k)}{dx}(x_{k+1} - x^*) = \frac{1}{2}\frac{d^2 f(\tilde{x})}{dx^2}(x_k - x^*)^2$$

$$\Rightarrow 2x_k(x_{k+1} - 0) = (x_k - 0)^2 \quad \text{for } x_k \neq x^* = 0$$

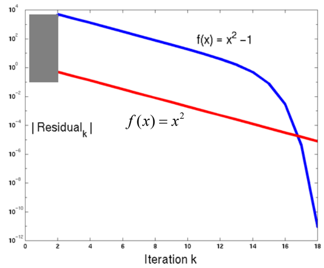
$$or \ (x_{k+1} - x^*) = \frac{1}{2}(x_k - x^*)$$

Convergence is linear

25

Newton Method – Convergence

Example 1,2



26

Possible failure

❖If the function is not monotonous

❖If f'_i=0 or very small at some points

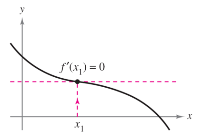
❖Works well when the function is monotonous, especially with moderate f'.

27

Nonconvergence Cases

❖ Case 1: If the initial estimate is selected such that the <u>derivative of the function equals zero.</u>

An example case of $f'(x_i) = 0$:



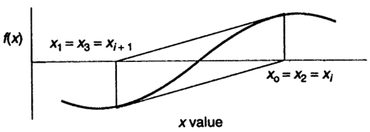
Newton's Method fails to converge if $f'(x_n) = 0$.

Way to solve this: choosing a different value for x_1

28

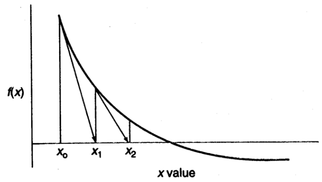
❖ Case 2: $\frac{f(x_i)}{f'(x_i)} = -\frac{f(x_{i+1})}{f'(x_{i+1})}$

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



29

❖ Case 3: A large number of iterations will be required if <u>$f'(x_i)$ is much larger than $f(x_i)$.</u> In such cases, $f(x_i)/f'(x_i)$ is small, which leads to a small adjustment at each iteration.



30

❖ Case 4: function f(x) is not differentiable at the root.

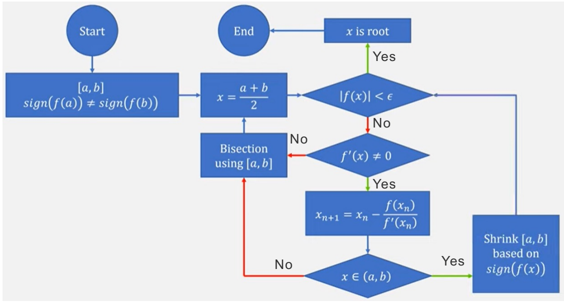
The function $f(x) = x^{1/3}$ is not differentiable at $x = 0$. Show that Newton's Method fails to converge using $x_1 = 0.1$.

Because $f'(x) = \frac{1}{3}x^{-2/3}$, the iterative formula is

$$\begin{aligned} x_{n+1} &= x_n - \frac{f(x_n)}{f'(x_n)} \\ &= x_n - \frac{x_n^{1/3}}{\frac{1}{3}x_n^{-2/3}} \\ &= x_n - 3x_n \\ &= -2x_n. \end{aligned}$$

31

Newton Bisection Hybrid (Newt-Safe)



Example: rtsafe.f from the famous Numerical Recipes

32

Secant method - discrete Newton method

In many cases, especially when f (x) has an implicit dependence on x, an analytic expression for the first-order derivative needed in the Newton method may not exist or may be very difficult to obtain.

We have to find an alternative scheme to achieve a similar algorithm. One way to do this is to replace the analytic f'(x) with the two-point formula for the first-order derivative, which gives

$$x_{i+1} = x_i - (x_i - x_{i-1}) f_i / (f_i - f_{i-1})$$

33

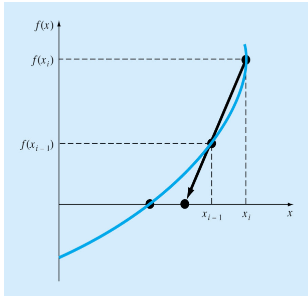
The Secant Method

$$x_{i+1} = x_i - f(x_i)\frac{x_i - x_{i-1}}{f(x_i) - f(x_{i-1})}$$

- Requires two initial estimates x_o, x_1.

  However, it is not a "bracketing" method.
- *The Secant Method* has the same properties as *Newton*'s method.

  Convergence is not guaranteed for all x_o, f(x).



34

34

Code example

Example:

g(x)=sin(x)-0.5

i	x_i	f_i
0	0	-0.5
1	π/2	0.5
2	π/4	……

$$x_{i+1} = x_i - (x_i - x_{i-1})\frac{f(x_i)}{f(x_i) - f(x_{i-1})}$$

$$x_2 = \frac{\pi}{2} - (\frac{\pi}{2} - 0)\cdot 0.5/(0.5+0.5) = \frac{\pi}{4}$$

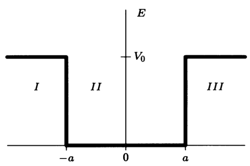
Secant.cpp

35

Physics problem: Finite Square-Well Potential

❖The finite square-well potential is:

$$V(x) = \begin{cases} V_0 & x \le -a \quad \text{Region I} \\ 0 & -a < x < a \quad \text{Region II} \\ V_0 & x \ge a \quad \text{Region III} \end{cases}$$



Schrödinger Equation $-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + V(x)\psi(x) = E\psi(x)$

The solution outside the finite well in regions I and III, where $E < V_0$, is:

$$\psi_I = Ce^{\beta x} \qquad \beta = \sqrt{2m(V_0 - E)}/\hbar$$

$$\psi_{II}(x) = A\sin\alpha x + B\cos\alpha x \qquad \alpha = \sqrt{2mE}/\hbar$$

$$\psi_{III} = Fe^{-\beta x}$$

the wave function must be zero at $x = \pm\infty$.

36

$$\psi_{\mathrm{I}} = Ce^{\beta x} \qquad \psi_{\mathrm{II}}(x) = A\sin\alpha x + B\cos\alpha x \qquad \psi_{\mathrm{III}} = Fe^{-\beta x}$$

Now, the **boundary conditions** require that:

$\psi_{\mathrm{I}} = \psi_{\mathrm{II}}$ at $x = -a$ $\qquad -A\sin\alpha a + B\cos\alpha a = Ce^{-\beta a}$

$\psi'_{I}(\mathrm{x}) = \psi'_{II}(\mathrm{x})$ at $x = -a$ $\qquad \alpha A\cos\alpha a + \alpha B\sin\alpha a = C\beta e^{-\beta a}$

$\psi_{\mathrm{II}} = \psi_{\mathrm{III}}$ at $x = a$ $\qquad A\sin\alpha a + B\cos\alpha a = Fe^{-\beta a}$

$\psi'_{II}(\mathrm{x}) = \psi'_{III}(\mathrm{x})$ at $x = a$ $\qquad \alpha A\cos\alpha a - \alpha B\sin\alpha a = -F\beta e^{-\beta a}$

Two kinds of solutions (different parity, see group theory):

Even states: $\quad A = 0, B \neq 0, C = F \qquad \alpha\sin\alpha a = \beta\cos\alpha a$

Odd states: $\quad A \neq 0, B = 0, C = -F \qquad \alpha\cos\alpha a = -\beta\sin\alpha a$

37

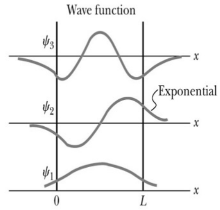
Our target: determine A,B,C,F and energy E

e.g., for even states, solve

$$f(\mathrm{E}) = \alpha\sin\alpha a - \beta\cos\alpha a = 0$$

$$\beta = \sqrt{2m(V_0 - E)}/\hbar$$

$$\alpha = \sqrt{2mE}/\hbar$$



Use hybrid method to find E

38

❖Roots of an equation

❖**Extremes of a function**

39

Extremes of a function

❖ An associated problem to find the root of an equation is finding the maxima and/or minima of a function.

❖ Examples of such situations in physics occur when considering the equilibrium position of an object, the potential surface of a field, and the optimized structures of molecules and small clusters.

40

❖ We know that an extreme of g(x) occurs at the point with

$$f(x) = \frac{dg(x)}{dx} = 0$$

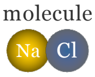
which is a minimum (maximum) if f'(x) = g''(x) is greater (less) than zero. So all the root-search schemes discussed so far can be generalized here to search for the extremes of a single-variable function.

41

Example

The (ionic) bond length of the diatomic molecule

$$V(r) = -\frac{e^2}{4\pi\varepsilon_0 r} + V_0\exp(-\frac{r}{r_0})$$



where e is the charge of a proton, ε_0 is the electric permittivity of vacuum, and V_0 and r_0 are parameters of this effective interaction.

The first term comes from the Coulomb interaction between the two ions, but the second term is the result of the electron distribution in the system.

42

The force:

$$f(r) = -\frac{dV(r)}{dr} = -\frac{e^2}{4\pi\varepsilon_0 r^2} + \frac{V_0}{r_0}\exp(-\frac{r}{r_0})$$

At equilibrium, the force between the two ions is zero. Therefore, we search for the root of f(x) = -dV(x)/dx = 0.

43

code example

- parameters for NaCl
- $e^2/4\pi\varepsilon_0$ = 14.4 AeV
- V_0 = 1.09 x 10^3 eV
- r_0 = 0.33 A
- r starts from 1 A
- 3.6.NaCl.cpp

44

Steepest-descent method

In principle, the search process should be forced to move along the direction of descending the function g(x) when looking for a minimum. In other words, for $x_{i+1}=x_i+\Delta x_i$, the increment Δx_i has the sign opposite to $g'(x_i)$.

Thus, an update scheme can be formulated:

$$\Delta x_i = -f_i / f_i' \Rightarrow \Delta x_i = -a \cdot g'_i = -a \cdot f_i$$

with 'a' being a positive, small, and adjustable parameter. For the minimum, f' (or g'') must be positive.

45

This scheme can be generalized to the multivariable case as

$$x_{i+1} = x_i + \Delta x_i = x_i - a \cdot \nabla g(x_i) / |\nabla g(x_i)|$$

where $x = (x_1, x_2, \ldots, x_l)$ and

$\nabla g(x) = (\partial g/\partial x_1, \partial g/\partial x_2, \ldots, \partial g/\partial x_l)$.

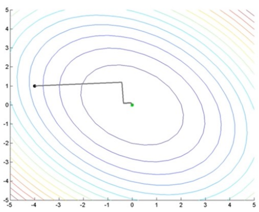
Note that step Δx_i here is scaled by $|\nabla g(x_i)|$ and is forced to move toward the direction of the steepest descent. This is why this method is known as the steepest-descent method.

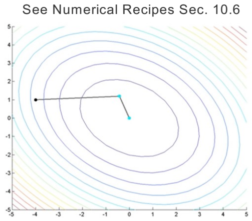
Widely used in machine learning!

46

Conjugate gradient method

See Numerical Recipes Sec. 10.6





Gradient-descent method: may need many steps to reach the minimum

Conjugate gradient method: fast convergence (maybe n step for n-dim)

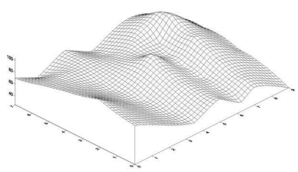
Search directions are orthogonal to each other $u^T v = 0$

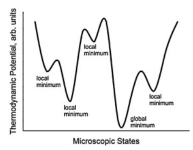
Search directions are conjugate to each other $\langle u, v\rangle = u^T Q v = 0$

$$f(x) = \frac{1}{2} x^T Q x - x^T b$$

47

Other methods for finding minima of a function





Local minima: quasi-Newton method *etc*

Global minima: Simulated annealing, Genetic algorithm, Particle swarm optimization, Differential evolution *etc*
(Derivative-free Optimization)

48

Homework

1. Sketch the function $x^3 - 5x + 3 = 0$.
(i) Determine the two positive roots to 4 decimal places using the bisection method. Note: You first need to bracket each of the roots.
(ii) Take the two roots that you found in the previous question (accurate to 4 decimal places) and "polish them up" to 14 decimal places using the Newton-Raphson method.
(iii) Determine the two positive roots to 14 decimal places using the hybrid method.

2. Search for the minimum of the function
g(x,y)=sin(x+y)+cos(x+2*y)
in the whole space.

49

3. Temperature dependence of magnetization

Determine M(T) the magnetization as a function of temperature T for simple magnetic materials. (see https://en.wikipedia.org/wiki/Curie%27s_law)

$$m(t) = \tanh\left(\frac{m(t)}{t}\right),$$

$$m(T) = \frac{M(T)}{N\mu}, \quad t = \frac{T}{T_c}, \quad T_c = \frac{N\mu^2\lambda}{k_B}$$

m: reduced magnetization
t: reduced temperature

For a given t, solve m, plot m as a function of t

50