Numerical Methods

ISSUES IN NUMERICAL ANALYSIS

- **WHAT IS NUMERICAL ANALYSIS?**
  - It is a way to do highly complicated mathematics problems on a computer.
  - It is also known as a technique widely used by scientists and engineers to solve their problems.
- **TWO ISSUES OF NUMERICAL ANALYSIS:**
  - How to compute? This corresponds to algorithmic aspects;
  - How accurate is it? That corresponds to error analysis aspects.

- **ADVANTAGES OF NUMERICAL ANALYSIS:**
  - It can obtain numerical answers of the problems that have no “analytic” solution.
  - It does NOT need special substitutions and integrations by parts. It needs only the basic mathematical operations: addition, subtraction, multiplication and division, plus making some comparisons.
- **IMPORTANT NOTES:**
  - Numerical analysis solution is always numerical.
  - Results from numerical analysis is an approximation.

- **NUMERICAL ERRORS**

When we get into the real world from an ideal world and finite to infinite, errors arise.

- SOURCES OF ERRORS:
  - Mathematical problems involving quantities of infinite precision.
  - Numerical methods bridge the precision gap by putting errors under firm control.
  - Computer can only handle quantities of finite precision.

– TYPES OF ERRORS:

- Truncation error (finite speed and time) - An example:

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = \left(1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!}\right) + \sum_{n=4}^{\infty} \frac{x^n}{n!}$$

$$= p_3(x) + \sum_{n=4}^{\infty} \frac{x^n}{n!}$$

- Round-off error (finite word length): All computing devices represent numbers with some imprecision, except for integers.

- Human errors: (a) Mathematical equation/model. (b) Computing tools/machines. (c) Error in original data. (d) Propagated error.

– MEASURE OF ERRORS:

Let a be a scalar to be computed and let $\bar{a}$ be its approximation.

Then, we define

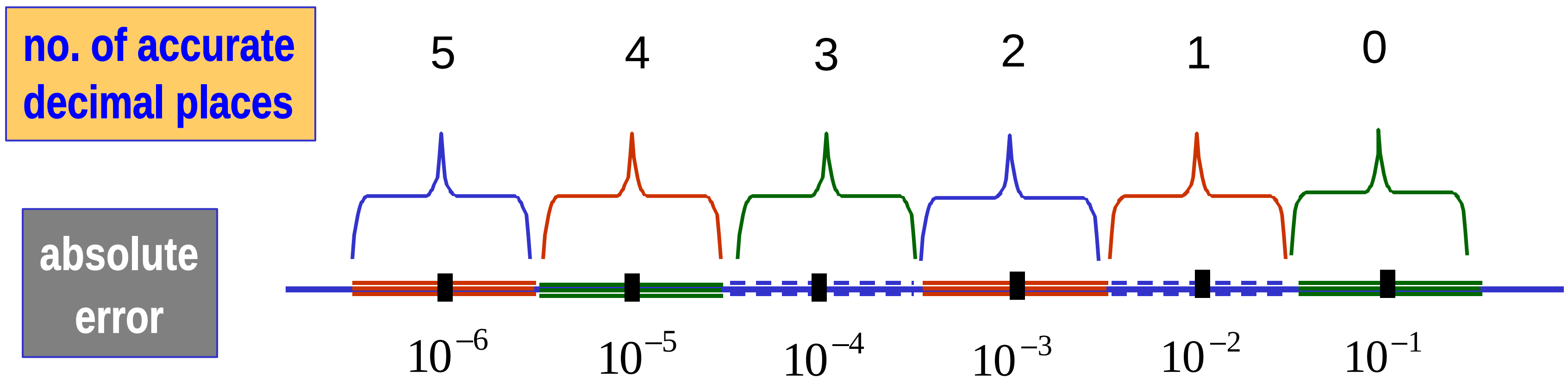
- Absolute error = | true value – approximated value |.

$$e = \left| a - \bar{a} \right|$$

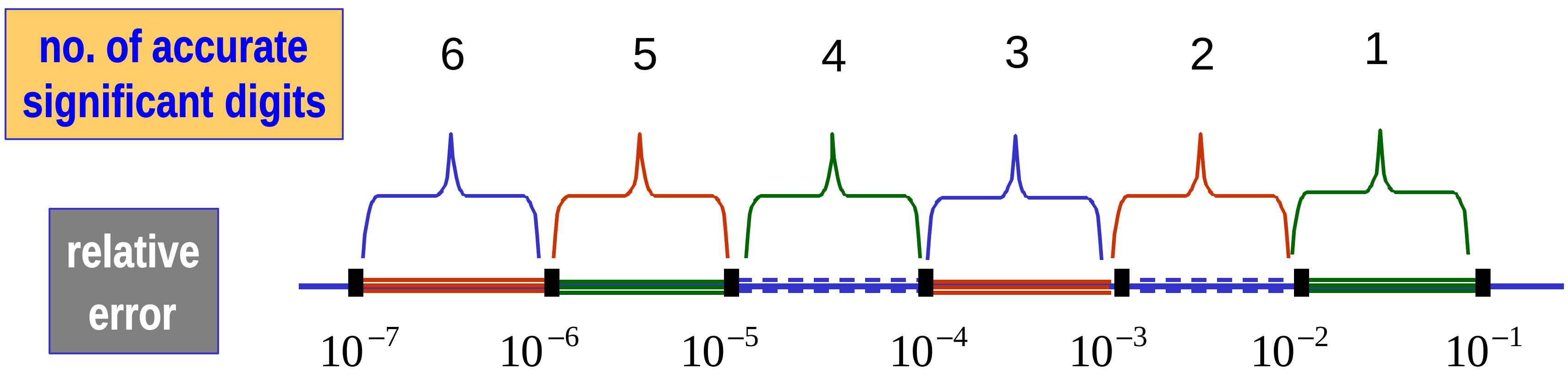
- Relative error = $\left| \dfrac{\text{true value} - \text{approximated value}}{\text{true value}} \right|$

$$e_r = \left| \frac{a - \bar{a}}{a} \right|$$

Absolute Error and Accuracy in Decimal Places



Relative Error and Accuracy in Significant Digits



n

Example: Let the true value of π be 3.1415926535898 and its approximation be 3.14 as usual. Compute the absolute error and relative error of such an approximation.

The absolute error:

$$e = \left| p - \overline{p} \right| = \left| 3.1415926535898 - 3.14 \right| = 0.0015926535898$$

which implies that the approximation is accurate up to 2 decimal places.

The relative error:

$$e_r = \left| \frac{p - \overline{p}}{p} \right| = \frac{0.0015926535898}{3.1415926535898} = 0.000506957382897$$

which implies that the approximation has a accuracy of 3 significant figures.

- **STABILITY AND CONVERGENCE**
  - STABILITY in numerical analysis refers to the trend of error change iterative scheme. It is related to the concept of convergence.

    It is stable if initial errors or small errors at any time remain small when iteration progresses. It is unstable if initial errors or small errors at any time get larger and larger, or eventually get unbounded.

  - CONVERGENCE: There are two different meanings of convergence in numerical analysis:

a. If the discretized interval is getting finer and finer after dicretizing the continuous problems, the solution is convergent to the true solution.

b. For an iterative scheme, convergence means the iteration will get closer to the true solution when it progresses.

Solutions to Nonlinear Equations (Computing Zeros)

- **Problem:** Given a function $f(x)$, which normally is nonlinear, the problem of "computing zeros" means to find all possible points, say

$$\tilde{x}_0, \tilde{x}_1, \cdots, \tilde{x}_n$$

such that

$$f(\tilde{x}_0) = f(\tilde{x}_1) = \ldots = f(\tilde{x}_n) = 0$$

However, it is often that we are required to find a single point $\tilde{x}_0$ in certain interval, say [a,b] such that

$$f(\tilde{x}_0) = 0$$

General strategy is to design an iterative process of the form

$$x_{n+1} = g(x_n)$$

with some starting point x_0. So that the numerical solution as

$$x_n \to \tilde{x}_0, \qquad \text{as } n \to \infty$$

Thus, instead of finding the exact solution, we find an approximation.

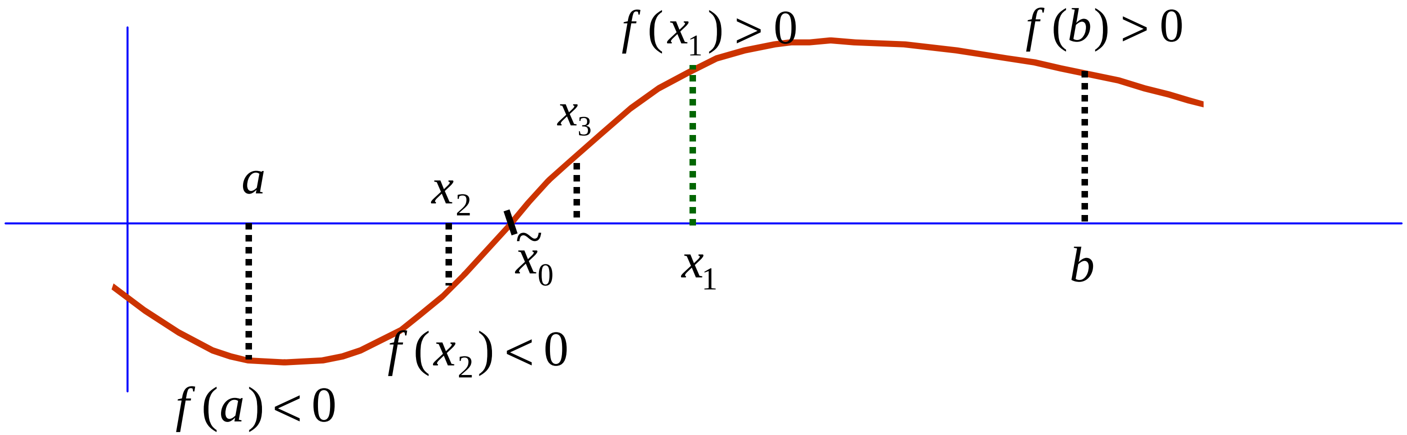
We focus on the following methods for this subject:

Bisection Method + False Position Method + Newton Method +

Secant Method + Fixed Point Method + Your Own Method

BISECTION METHOD

Given a function $f(x)$ in $[a, b]$ satisfying $f(a)\,f(b)<0$, find a zero of $f(x)$ in $[a, b]$.



Step 0: Let $x_a := a; \quad x_b := b; \quad n = 1.$

Step 1: Cut the interval in the middle, i.e., find $x_n := \dfrac{x_a + x_b}{2}$

Step 2: Define $\begin{cases} x_a := x_a; \quad x_b := x_n; & \text{if } f(x_a)\,f(x_n) < 0 \\ x_a := x_n; \quad x_b := x_b; & \text{if } f(x_b)\,f(x_n) < 0 \end{cases}$

Step 3: If x_n is close enough to $\tilde{x}_0$, stop. Otherwise, $n:=n+1$ & go to Step 1.

Advantages:

1. It is guaranteed to work if $f(x)$ is continuous in $[a, b]$ and a zero actually exists.

2. A specific accuracy of iterations is known in advance. Few other root-finding methods share this advantage.

Disadvantages:

a. It requires the values of a and b.

b. The convergence of interval halving is very slow.

c. Multiple zeros between a and b can cause problem.

Example: Let $f(x) = x^2 - 1$. Find its zero in $[0, 1.5]$

Of course, we know $f(x)$ has a root at $\tilde{x}_0 = 1$. Let us find it using the Bisection Method:

Step 0: $x_a = 0, \quad x_b = 1.5, \quad n = 1$

Step 1: $x_1 = \dfrac{0 + 1.5}{2} = 0.75$

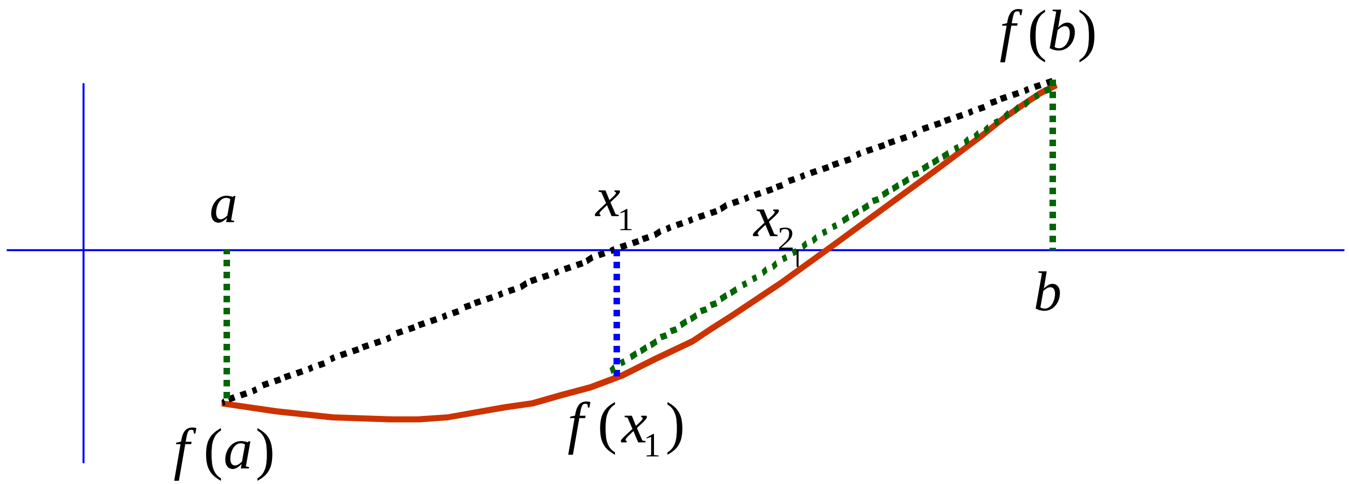
Step 2: $f(x_b)f(x_1) = (1.5^2 - 1)(0.75^2 - 1) = -0.546875 < 0$

$$\Rightarrow \quad x_2 = \frac{0.75 + 1.5}{2} = 1.125$$

$$\Rightarrow \quad x_3 = \frac{0.75 + 1.125}{2} = 0.9375 \quad \Rightarrow \quad x_4 = 1.03125 \quad \cdots$$

FALSE POSITION METHOD

The graph used in this method is shown in the following figure.



The key idea is to approximate the curve by a straight line within the interval and identify a "false" position x_1, which of course may not be a true solution. We can keep repeating this procedure to get approximations of the solution, x_2, x_3, …. Mathematically,

$$x_{n+1} = x_n - \frac{b - x_n}{f(b) - f(x_n)} f(x_n), \quad x_0 = a$$

Advantage:

Convergence is faster than bisection method.

Disadvantages:

1. It requires a and b.

2. The convergence is generally slow.

3. It is only applicable to $f(x)$ of certain fixed curvature in $[a, b]$.

4. It cannot handle multiple zeros.

Example: $f(x) = x^2 - 1$. Find its root in [0,1.5]

$$x_1 = x_0 - \frac{b-a}{f(b)-f(a)} f(a)$$

$$= 0 - \frac{1.5-0}{1.25-(-1)}(-1) = 0.6667$$

$$x_2 = x_1 - \frac{b-x_1}{f(b)-f(x_1)} f(x_1)$$

$$= 0.6667 - \frac{1.5-0.6667}{1.25-(-0.5556)}(-0.5556)$$

$$= 0.9231$$

$$x_3 = x_2 - \frac{b-x_2}{f(b)-f(x_2)} f(x_2)$$

$$= 0.9231 - \frac{1.5-0.9231}{1.25-(-0.1479)}(-0.1479)$$

$$= 0.9841$$

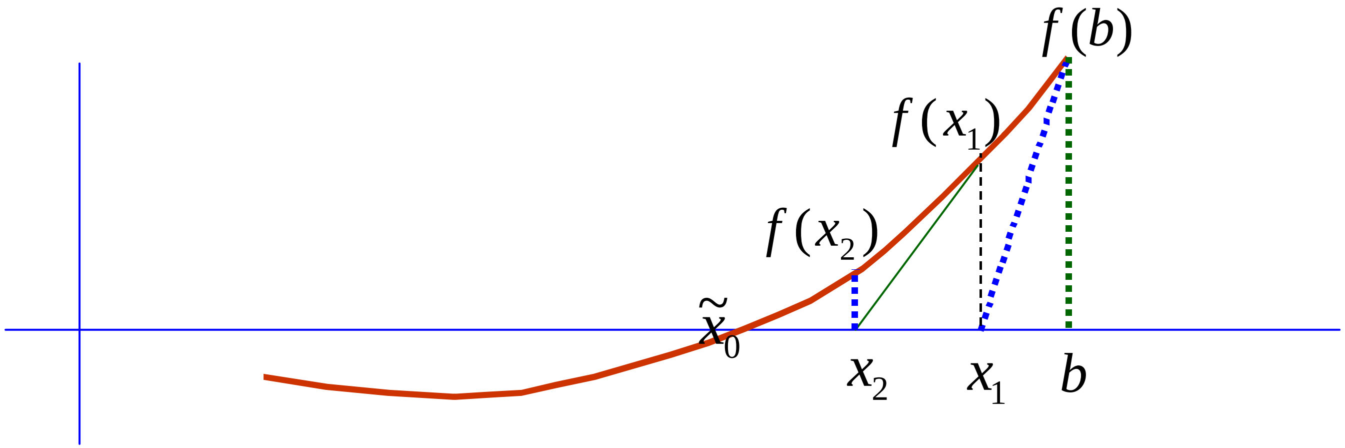
...

$x_n \to 1$ as $n \to \infty$

n

Newton Method

Assume that $f'(x)$ exists and nonzero at x_n for all n.



Zero-finding: The linear approximation based on one point (x_1, y_1) only is given by

$$y = y_1 + (x - x_1) f'(x)$$

We look for a point x for which $y = 0$. As such we have the following iteration:

$$y_{n+1} = y_n + (x_{n+1} - x_n) f'(x_n) = 0 \quad \Rightarrow \quad x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

Advantages:

1. Starting point x_1 can be arbitrary.

2. The convergence is faster than the previous two methods.

Disadvantages:

1. It needs $f'(x)$.

2. The divergence may occur.

Example: Find zero of $f(x) = x^2 - 1$ in [0,1.5] using Newton's Method

$$f'(x) = 2x \Rightarrow x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{x_n^2 - 1}{2x_n} = \frac{1}{2x_n}(x_n^2 + 1)$$

Starting with any initial point, say $x_0 = 0.1$, we have

$$x_1 = \frac{1}{2 \times 0.1}(0.1^2 + 1) = 5.05$$

$$x_2 = \frac{1}{2 \times 5.05}(5.05^2 + 1) = 2.624$$

$$x_3 = \frac{1}{2 \times 2.624}(2.624^2 + 1) = 1.5026$$

$$x_4 = \frac{1}{2 \times 1.5026}(1.5026^2 + 1) = 1.084$$

$$x_5 = \frac{1}{2 \times 1.084}(1.084^2 + 1) = 1.003$$

...

Again, $x_n \to 1$ as $n \to \infty$

Secant Method

Secant Method is a modified version of Newton's method in which $f'(x_n)$ is approximated by

$$f'(x_n) = \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}$$

Substituting this into the iteration scheme of Newton's method, we obtain

$$x_{n+1} = x_{n-1} - f(x_{n-1}) \frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})}$$

- Advantage: 1) Convergence is fast and 2) it does dot need derivative.
- Disadvantage: The method may fail.

Fixed Point Method

Start from $f(x) = 0$ and derive a relation

$$x = g(x)$$

Example: Compute zero for $f(x)$ with $f(x) = e^x - 4 - 2x$ or find x such that

$$e^x - 4 - 2x = 0$$

$$\Rightarrow \quad x = \frac{1}{2}(e^x - 4) = g(x) \qquad (1)$$

$$\Rightarrow \quad e^x = 4 + 2x \quad \Rightarrow \quad x = \ln(4 + 2x) = g(x) \qquad (2)$$

The fixed-point method is simply given by

$$\boxed{x_{n+1} = g(x_n)}$$

Q: Does it work? Does it converge? A: Maybe yes and maybe not.

Q: When does it converge?

A: Convergence Theorem

Consider a function $f(x)$ and suppose it has a zero on the interval $[a, b]$. Also, consider the iteration scheme

$$x_{n+1} = g(x_n)$$

derived using fixed-point method. Then this scheme converges, i.e,

$$x_n \to \tilde{x}_0 \quad as \quad n \to \infty$$

if the following condition are satisfied:

(1) $|g'(x)| < 1$ for all $x \in [a, b]$. $g(x)$ is also said to be contraction in $[a, b]$.

(2) Start any initial point $x_0 \in [a, b]$.

Remark: If the above conditions are not satisfied, the iteration scheme might still converge as the above theorem only gives sufficient conditions.

Example: Compute zeros of $f(x) = e^x - 4 - 2x$

Scheme 1:

$$x = g(x) = \frac{1}{2}(e^x - 4)$$

$$x_{n+1} = \frac{1}{2}(e^{x_n} - 4)$$

$$g'(x) = \frac{1}{2}e^x$$

with $|g'(x)| < 1$ for $-\infty < x < 0.693$

Let us choose $x_0 = -2,$

$$\Rightarrow \quad x_1 = \frac{1}{2}(e^{-2} - 4) = -1.9323$$

$$x_2 = \frac{1}{2}(e^{-1.9323} - 4) = -1.9276$$

$x_3 = -1.9273,\ x_4 = -1.9272,$

$x_5 = -1.9272$

Scheme 2:

$$x = g(x) = \ln(4 + 2x)$$

$$x_{n+1} = \ln(4 + 2x_n), \qquad x > -2$$

$$g'(x) = \frac{2}{4 + 2x} \quad \Rightarrow$$

$$|g'(x)| < 1,\ x \in (-\infty, -3)$$

$$|g'(x)| < 1, \quad x \in (-1, \infty)$$

Let $x_0 = 0, \quad \Rightarrow$

$x_1 = 1.3863,\ x_2 = 1.9129,$

$x_3 = 2.0574,\ x_4 = 2.0937,$

$x_5 = 2.1026,\ x_6 = 2.1048,$

$x_7 = 2.1053,\ x_8 = 2.1054,$

$x_9 = 2.1054$

Applications: Compute $\sqrt{3}$.   (Actual value = 1.73205)

Solution: Let $x = \sqrt{3} \Rightarrow x^2 = 3 \quad \Rightarrow \quad f(x) = x^2 - 3 = 0$

The problem is transformed to a problem of finding zero (or root) for $f(x)$ in $[0, \infty)$.

We use Newton's Method with $x_0=1$.

$$f'(x) = 2x$$

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{x_n^2 - 3}{2x_n}$$

$$= \frac{1}{2x_n}(x_n^2 + 3)$$

$$x_1 = \frac{1}{2x_0}(x_0^2 + 3) = 2$$

$$x_2 = \frac{1}{2x_1}(x_1^2 + 3) = 1.75$$

$$x_3 = \frac{1}{2x_2}(x_2^2 + 3) = 1.73214$$

$$x_4 = \frac{1}{2x_3}(x_3^2 + 3) = 1.73205$$

(good enough)

n

SUMMARY

- Bisection Method

  Condition: Continuous function $f(x)$ in $[a, b]$ satisfying $f(a)f(b) < 0$.

  Convergence: Slow but sure. Linear.

- False position method

  Condition: Continuous function $f(x)$ in $[a, b]$ satisfying $f(a)f(b) < 0$.

  Convergence: Slow (linear).

- Newton Method

  Condition: Existence of nonzero $f'(x)$

  Convergence: Fast (quadratic).

- Secant Method

  Condition: Existence of nonzero $f(x_{n+1}) - f(x_n)$

  Convergence: Fast (quadratic).

- Fixed-point Method

  Condition: Contraction of $g(x)$.

  Convergence: Varying with the nature of $g(x)$.