Numerical Analysis homework 1

Jerry Xu 3230101329[1]

[1] *Mathematics and Applied Mathematics (Strengthening Basic Science Program),class 2301, Zhejiang University*

[1] *3230101329@zju.edu.cn*

Due time: September 25, 2025

Abstract

The abstract is not necessary for the theoretical homework, but for the programming project, you are encouraged to write one.

I. The content of the homework

question 1.8.1 I - VIII, Theoretical Questions, on page 7

I.

Consider the bisection method starting with the initial interval $[a, b] = [1.5, 3.5]$. In the following questions "the interval" refers to the bisection interval whose width changes across different loops.

- What is the width of the interval at the nth step?
- What is the supremum of the distance between the root r and the midpoint of the interval?

Solution.

Let the initial interval be $[a, b] = [1.5, 3.5]$, so $b - a = 2$.

Each bisection halves the interval width. If we measure the step count so that after n bisections the interval has been halved n times, the width at the n-th step is

$$w_n = \frac{b-a}{2^n}.$$

With $b - a = 2$ this becomes

$$\boxed{w_n = \frac{2}{2^n} = 2^{1-n}.}$$

Let m_n be the midpoint of the interval at the n-th step. For any root r that is in the current interval, the distance $|r - m_n|$ is at most half the interval width (because the midpoint splits the interval). Thus

$$|r - m_n| \le \frac{w_n}{2} = \frac{b-a}{2^{n+1}}.$$

Therefore the supremum (in all possible locations of r within the interval) is

$$\boxed{\sup |r - m_n| = \frac{b-a}{2^{n+1}}.}$$

With $b - a = 2$ this simplifies to

$$\boxed{\sup |r - m_n| = \frac{2}{2^{n+1}} = 2^{-n}.}$$

II.

In using the bisection algorithm with its initial interval as $[a_0, b_0]$ with $a_0 > 0$, we want to determine the root with its relative error no greater than ϵ. Prove that this goal of convergence is achieved with the following choice of the number of steps,

$$n \geq \frac{\log(b_0 - a_0) - \log\epsilon - \log a_0}{\log 2} - 1.$$

Solution.

We use the same notation as before. Let r be the root contained in the initial interval $[a_0, b_0]$ with $a_0 > 0$, and let m_n denote the midpoint after n bisection steps. From the bisection property we have the midpoint error bound

$$|r - m_n| \leq \frac{b_0 - a_0}{2^{n+1}}.$$

We require the relative error to satisfy

$$\frac{|r - m_n|}{|r|} \leq \epsilon.$$

Since $r \geq a_0 > 0$, a sufficient condition is

$$\frac{|r - m_n|}{a_0} \leq \epsilon,$$

and using precvious inequation of midpoint error bound this becomes

$$\frac{b_0 - a_0}{2^{n+1}a_0} \leq \epsilon.$$

then we have:

$$2^{n+1} \geq \frac{b_0 - a_0}{a_0\epsilon}.$$

Taking logarithms (base 2) gives

$$n + 1 \geq \log_2\left(\frac{b_0 - a_0}{a_0\epsilon}\right) = \frac{\log(b_0 - a_0) - \log a_0 - \log\epsilon}{\log 2}.$$

Hence

$$n \geq \frac{\log(b_0 - a_0) - \log a_0 - \log\epsilon}{\log 2} - 1,$$

which is the claimed bound.

III.

Perform four iterations of Newton's method for the polynomial equation $p(x) = 4x^3 - 2x^2 + 3 = 0$ with the starting point $x_0 = -1$. Use a hand calculator and organize results of the iterations in a table.

Solution.

Consider

$$p(x) = 4x^3 - 2x^2 + 3, \qquad p'(x) = 12x^2 - 4x.$$

Newton's iteration is

$$x_{n+1} = x_n - \frac{p(x_n)}{p'(x_n)}.$$

We perform four iterations starting from $x_0 = -1$. Below each step shows the values of $p(x_n)$, $p'(x_n)$ and the next iterate.

$$\begin{aligned}
x_0 &= -1, \\
p(x_0) &= 4(-1)^3 - 2(-1)^2 + 3 = -4 - 2 + 3 = -3, \\
p'(x_0) &= 12(-1)^2 - 4(-1) = 16, \\
x_1 &= x_0 - \frac{p(x_0)}{p'(x_0)} = -1 - \frac{-3}{16} = -1 + \frac{3}{16} = -\tfrac{13}{16} = -0.8125.
\end{aligned}$$

$$\begin{aligned}
x_1 &\approx -0.812500000000,\\
p(x_1) &\approx -0.465820312500,\\
p'(x_1) &\approx 11.171875000000,\\
x_2 &\approx x_1 - \frac{p(x_1)}{p'(x_1)} \approx -0.770804195804.
\end{aligned}$$

$$\begin{aligned}
x_2 &\approx -0.770804195804,\\
p(x_2) &\approx -0.020137886720,\\
p'(x_2) &\approx 10.212886082449,\\
x_3 &\approx x_2 - \frac{p(x_2)}{p'(x_2)} \approx -0.768832384256.
\end{aligned}$$

$$\begin{aligned}
x_3 &\approx -0.768832384256,\\
p(x_3) &\approx -0.000043708433,\\
p'(x_3) &\approx 10.168568357988,\\
x_4 &\approx x_3 - \frac{p(x_3)}{p'(x_3)} \approx -0.768828085870.
\end{aligned}$$

$$\begin{aligned}
x_4 &\approx -0.768828085870,\\
p(x_4) &\approx -2.0741 \times 10^{-10},\\
p'(x_4) &\approx 10.168471850942.
\end{aligned}$$

For clarity, the iteration data (rounded for a hand calculator) are tabulated below:

n	x_n	$p(x_n)$	$p'(x_n)$	x_{n+1}
0	-1.000000000000	-3.000000000000	16.000000000000	-0.812500000000
1	-0.812500000000	-0.465820312500	11.171875000000	-0.770804195804
2	-0.770804195804	-0.020137886720	10.212886082449	-0.768832384256
3	-0.768832384256	-0.000043708433	10.168568357988	-0.768828085870
4	-0.768828085870	-2.0741×10^{-10}	10.168471850942	—

After four Newton iterations we obtain

$$x_4 \approx -0.768828085870,$$

with residual $p(x_4) \approx -2.07 \times 10^{-10}$, indicating good convergence.

IV.

Consider a variation of Newton's method in which only the derivative at x_0 is used,

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_0)}.$$

Find C and s such that

$$e_{n+1} = Ce_n^s,$$

where e_n is the error of Newton's method at step n, s is a constant, and C may depend on x_n, the true solution α, and the derivative of the function f.

Solution.

Let α be the true root, and denote the error by $e_n = x_n - \alpha$. We use Taylor expansion of f about α:

$$f(x_n) = f(\alpha + e_n) = f'(\alpha)e_n + \tfrac{1}{2}f''(\alpha)e_n^2 + O(e_n^3).$$

Also write the denominator (the fixed derivative at x_0) as

$$f'(x_0) = f'(\alpha) + \big(f'(x_0) - f'(\alpha)\big).$$

For brevity set

$$\Delta := f'(x_0) - f'(\alpha).$$

The iteration is

$$e_{n+1} = x_{n+1} - \alpha = x_n - \frac{f(x_n)}{f'(x_0)} - \alpha = e_n - \frac{f(x_n)}{f'(x_0)}.$$

Substitute the Taylor series for $f(x_n)$:

$$e_{n+1} = e_n - \frac{f'(\alpha)e_n + \frac{1}{2}f''(\alpha)e_n^2 + O(e_n^3)}{f'(\alpha) + \Delta}.$$

Factor out e_n and simplify the leading terms:

$$\begin{aligned} e_{n+1} &= e_n\Big(1 - \frac{f'(\alpha)}{f'(\alpha)+\Delta}\Big) - \frac{\frac{1}{2}f''(\alpha)e_n^2}{f'(\alpha)+\Delta} + O(e_n^3) \\ &= e_n \cdot \frac{\Delta}{f'(\alpha)+\Delta} - \frac{1}{2}\frac{f''(\alpha)}{f'(\alpha)+\Delta}\, e_n^2 + O(e_n^3). \end{aligned}$$

From this expansion we see the following:

- **Generic case:** If $\Delta \neq 0$ (i.e. $f'(x_0) \neq f'(\alpha)$), the linear term is dominant. Thus

  $$e_{n+1} = C\, e_n^1 + (\text{higher order terms}),$$

  with

  $$\boxed{s = 1, \qquad C = \frac{\Delta}{f'(x_0)} = \frac{f'(x_0) - f'(\alpha)}{f'(x_0)}.}$$

  (Here we used $f'(x_0) = f'(\alpha) + \Delta$ to write C in this simple form.) So the method is generically *linearly* convergent with factor C.

- **Special case:** If $\Delta = 0$ (i.e. $f'(x_0) = f'(\alpha)$), then the linear coefficient vanishes and the next term is quadratic. In this special case

  $$e_{n+1} = -\frac{1}{2}\frac{f''(\alpha)}{f'(\alpha)}\, e_n^2 + O(e_n^3),$$

  so

  $$\boxed{s = 2, \qquad C = -\frac{1}{2}\frac{f''(\alpha)}{f'(\alpha)}.}$$

  That is, if the fixed derivative at x_0 happens to equal $f'(\alpha)$, then the method has quadratic leading behaviour (like usual Newton).

In short, generically $s = 1$ with $C = (f'(x_0) - f'(\alpha))/f'(x_0)$. Only when $f'(x_0) = f'(\alpha)$ the linear term disappears and we get $s = 2$ with $C = -\frac{1}{2}f''(\alpha)/f'(\alpha)$.

V.

Within $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, will the iteration $x_{n+1} = \tan^{-1} x_n$ converge?

Solution.

Consider the iteration $x_{n+1} = \arctan x_n$ with $x_0 \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$.

Case 1: $x_0 > 0$. Since $\tan y > y$ for $y \in (0, \frac{\pi}{2})$, we have

$$\arctan x < x \quad \text{for all } x > 0.$$

Hence $x_1 = \arctan x_0 < x_0$. By induction, if $x_n > 0$ then $x_{n+1} = \arctan x_n < x_n$, so the sequence (x_n) is positive and monotone decreasing. Also each $x_n \geq 0$ (clear from $\arctan x \geq 0$ when $x \geq 0$). A positive monotone decreasing sequence is bounded below and therefore converges.

Case 2: $x_0 < 0$. Note arctan is an odd function, so the same argument works: the sequence is negative and monotone increasing, hence it also converges.

Case 3: $x_0 = 0$. Then $x_n \equiv 0$ and it converges.

In all cases the limit must satisfy $L = \arctan L$, so $L = 0$. Thus the iteration converges for any initial $x_0 \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, and the limit is 0.

VI.

Let $p > 1$. What is the value of the following continued fraction?

$$x = \frac{1}{p + \frac{1}{p+\frac{1}{p+\cdots}}}$$

Prove that the sequence of values converges. (Hint: this can be interpreted as $x = \lim_{n\to\infty} x_n$, where $x_1 = \frac{1}{p}$, $x_2 = \frac{1}{p+\frac{1}{p}}$, $x_3 = \frac{1}{p+\frac{1}{p+\frac{1}{p}}}$, and so forth. Formulate x as a fixed point of some function.)

Solution.

Let $p > 1$ and define the finite-level values by

$$x_1 = \frac{1}{p}, \qquad x_{n+1} = \frac{1}{p + x_n} \quad (n \geq 1).$$

We show (x_n) converges by using monotone bounded subsequences.

1. *Basic bounds.*

Since $p > 1$ we have $x_1 = 1/p > 0$. For any $x \geq 0$, $\dfrac{1}{p+x} \leq \dfrac{1}{p}$. Thus every term satisfies

$$0 < x_n \leq \frac{1}{p} \qquad \text{for all } n.$$

So the sequence is positive and bounded.

2. *Monotonicity of even and odd subsequences.*

The function $f(x) = \dfrac{1}{p+x}$ is strictly decreasing on $[0, \infty)$. Hence the composition $f \circ f$ is strictly increasing on $[0, \infty)$. But

$$x_{n+2} = f(f(x_n)) \quad \text{for all } n,$$

so the subsequence (x_{2n}) (even terms) is generated by iterating the increasing map $f \circ f$. In particular, one checks by induction that (x_{2n}) is monotone increasing, and similarly (x_{2n+1}) (odd terms) is monotone decreasing. More concretely:

$$x_2 = f(x_1) \leq x_1, \qquad x_4 = f(f(x_2)) \geq x_2, \ \ldots$$

so $x_2 \leq x_4 \leq x_6 \leq \cdots$ and $x_1 \geq x_3 \geq x_5 \geq \cdots$.

3. *Convergence of subsequences and same limit.*

Both subsequences are monotone and bounded (by step 1), so they converge: there exist limits L_{even} and L_{odd}. Passing to the limit in the relation $x_{n+1} = f(x_n)$ along even and odd indices shows these two limits must be fixed points of $f \circ f$. But any fixed point of $f \circ f$ that lies in $[0, 1/p]$ is also a fixed point of f itself (because if $y = f(f(y))$ then applying f to both sides gives $f(y) = f(f(f(y)))$, and by uniqueness one gets $y = f(y)$; alternately one can argue the two subsequence limits must be equal by continuity). Hence $L_{\text{even}} = L_{\text{odd}} =: a$. Therefore the whole sequence x_n converges to the same limit a.

4. *Value of the limit.*

Taking limit $n \to \infty$ in $x_{n+1} = \dfrac{1}{p + x_n}$ gives

$$a = \frac{1}{p+a}.$$

Multiply both sides by $p + a$ and rearrange:

$$a(p + a) = 1 \quad \Longrightarrow \quad a^2 + pa - 1 = 0.$$

Solve the quadratic:

$$a = \frac{-p \pm \sqrt{p^2+4}}{2}.$$

Since $a > 0$, we take the positive root. Thus

$$\boxed{a = \frac{-p + \sqrt{p^2+4}}{2}}$$

is the value of the continued fraction. This completes the proof that the sequence converges and gives the limit.

VII.

What happens in problem II if $a_0 < 0 < b_0$? Derive an inequality of the number of steps similar to that in II. In this case, is the relative error still an appropriate measure?

Solution.

We keep the same notation as in problem II. Now suppose the initial interval satisfies $a_0 < 0 < b_0$. Let r be the root in $[a_0, b_0]$, and let m_n be the midpoint after n bisections. As before the width of the interval after n steps is

$$w_n = \frac{b_0 - a_0}{2^n},$$

and the worst-case distance from the midpoint to the root is

$$|r - m_n| \leq \frac{w_n}{2} = \frac{b_0 - a_0}{2^{n+1}}.$$

1. *Absolute-error requirement.*

If we want the absolute error to be at most $\eta > 0$, i.e.

$$|r - m_n| \leq \eta,$$

then it is sufficient to require

$$\frac{b_0 - a_0}{2^{n+1}} \leq \eta.$$

Rearranging gives

$$2^{n+1} \geq \frac{b_0 - a_0}{\eta} \quad \Longrightarrow \quad n \geq \log_2\left(\frac{b_0 - a_0}{\eta}\right) - 1.$$

Equivalently, in natural logarithms,

$$\boxed{n \geq \frac{\log(b_0 - a_0) - \log \eta}{\log 2} - 1.}$$

This inequality is the analogue of the bound in problem II but written for an absolute-error tolerance η.

2. *Why relative error is problematic.*

In problem II we used a relative-error criterion $|r - m_n|/|r| \leq \epsilon$, and the derivation relied on a positive lower bound for $|r|$ (there we used $r \geq a_0 > 0$). When $a_0 < 0 < b_0$ such a positive lower bound need not exist: the true root r could be very close to 0. If r is near zero, the relative error $|r - m_n|/|r|$ can be arbitrarily large even when $|r - m_n|$ is small. Thus, *in general* a relative-error requirement is not appropriate when the interval crosses zero.

VIII.

(*) Consider solving $f(x) = 0$ ($f \in C^{k+1}$) by Newton's method with the starting point x_0 being a root of multiplicity k. Note that α is a zero of multiplicity k of the function f.

- How can a multiple root be detected by examining the behavior of the points $(x_n, f(x_n))$?
- Prove that if r is a zero of multiplicity k of the function f, then quadratic convergence in Newton's iteration will be restored by making this modification:

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

Solution.

(1) Detecting a multiple root from the points $(x_n, f(x_n))$. As we learned in class, when α weren't a zero of multiplicity of the function f,we have the following equation to evaluate the convergence:

$$e_{n+1} = e_n^2 \frac{f''(\epsilon)}{2f'(x_n)}$$

Back to our situation,because α is a zero of multiplicity k of the function f and $f(x) = 0$ ($f \in C^{k+1}$),we can write $f(x)$ into $f(x) = (x - \alpha)^k g(x)$,$g \in C^1$

Similarily, we have:

$$e_{n+1} = e_n - \frac{f(x_n)}{f'(x_n)} = e_n - \frac{e_n g(x_n)}{kg(x_n) + e_n g'(x_n)} = e_n \frac{(k-1)g(x_n) - e_n g'(x_n)}{kg(x_n) + e_n g'(x_n)} = e_n \frac{(k-1) - e_n \frac{g'(x_n)}{g(x_n)}}{k + e_n \frac{g'(x_n)}{g(x_n)}}$$

Thus, as $n \to \infty$,

$$\frac{e_{n+1}}{e_n} \to \frac{k-1}{k} < 1,$$

so the convergence is only linear. Consequently, a root of multiplicity $k > 1$ can be detected numerically by observing that the ratio $\frac{|x_{n+1}-\alpha|}{|x_n-\alpha|}$ tends to the constant $\frac{k-1}{k}$, i.e. the error decreases slowly and at a constant rate.

(2) Proof that the modified iteration restores quadratic convergence. Write f near r in the form

$$f(x) = (x-r)^k g(x),$$

with $g \in C^1$ and $g(r) \neq 0$. Then

$$f'(x) = k(x-r)^{k-1} g(x) + (x-r)^k g'(x).$$

Let $e_n := x_n - r$. For the modified iteration

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

compute the new error $e_{n+1} = x_{n+1} - r$ using the above factorization. We have

$$\begin{aligned} k\frac{f(x_n)}{f'(x_n)} &= k\frac{e_n^k g(x_n)}{ke_n^{k-1} g(x_n) + e_n^k g'(x_n)} \\ &= e_n \cdot \frac{kg(x_n)}{kg(x_n) + e_n g'(x_n)}. \end{aligned}$$

Therefore

$$\begin{aligned} e_{n+1} = e_n - k\frac{f(x_n)}{f'(x_n)} &= e_n \left(1 - \frac{kg(x_n)}{kg(x_n) + e_n g'(x_n)}\right) \\ = e_n \cdot \frac{e_n g'(x_n)}{kg(x_n) + e_n g'(x_n)} &= e_n^2 \cdot \frac{g'(x_n)}{kg(x_n) + e_n g'(x_n)}. \end{aligned}$$

Since g is continuous and $g(r) \neq 0$, for x_n close to r the denominator $kg(x_n) + e_n g'(x_n)$ stays close to $kg(r) \neq 0$. Thus there exists a neighbourhood of r where the factor

$$C_n := \frac{g'(x_n)}{kg(x_n) + e_n g'(x_n)}$$

is bounded and tends to the limit $C := g'(r)/(kg(r))$ as $n \to \infty$. Hence for n large we get the error law

$$e_{n+1} = C_n \, e_n^2,$$

with $C_n \to C \neq 0$. This shows the iteration is *quadratically* convergent (error is proportional to square of previous error) once the iterates are sufficiently close to the root.

Thus the modification $x_{n+1} = x_n - k\, f(x_n)/f'(x_n)$ restores the usual quadratic convergence for a root of multiplicity k.