MATH 501 Problem Set 4

Pseudocode Programming

```
a = input("Enter x0 here: ");
b = input("Enter x1 here: ");
u = f(a);
v = f(b);

for k = 2:99
    if (abs(u) < abs(v))
        [a, b] = swap(a, b);
        [u, v] = swap(u, v);
        disp("Endpoints swapped");
    end
    s = (b - a)/(v - u);
    a = b;
    u = v;
    b = (b - v * s);
    v = (f(b));

    disp("Iteration " + double(k) + ": x = " +
         double(b) + " and f(x) = " +
         double(v));
    if (abs(v)<eps)
        disp("f(x) is small enough");
        break
    elseif (abs(b-a)<eps)
        disp("x is close enough to root,
             should it exist");
        break;
    end
end

function y = f(x)
y = x^3-sinh(x)+4*x^2+6*x+9;
end

function [b, a] = swap(a, b)
end
```

Output for $x_0 = 7$ and $x_1 = 8$:

```
Command Window
  Enter x0 here: 7
  Enter x1 here: 8
  Endpoints swapped
  Iteration 2: x = 7.0589 and f(x) = 20.7983
  Iteration 3: x = 7.1176 and f(x) = -1.8347
  Iteration 4: x = 7.1129 and f(x) = 0.071011
  Iteration 5: x = 7.1131 and f(x) = 0.00022912
  Iteration 6: x = 7.1131 and f(x) = -2.8751e-08
  Iteration 7: x = 7.1131 and f(x) = 2.0606e-13
  Iteration 8: x = 7.1131 and f(x) = -1.5632e-13
  Iteration 9: x = 7.1131 and f(x) = -1.5632e-13
  x is close enough to root, should it exist
fx >>
```

Output for problem 3.4.12 with $x_0 = 3$ and $x_1 = 10$:

```
Command Window
  Endpoints swapped
  Iteration 2: x = 3.0582 and f(x) = 82.7393
  Endpoints swapped
  Iteration 3: x = 1.3124 and f(x) = 24.3021
  Iteration 4: x = 0.57587 and f(x) = 13.3645
  Iteration 5: x = -0.32412 and f(x) = 7.7713
  Iteration 6: x = -1.5746 and f(x) = 7.8767
  Endpoints swapped
  Iteration 7: x = 91.8862 and f(x) = -4.023986394221349e+39
  Endpoints swapped
  Iteration 8: x = -0.32412 and f(x) = 7.7713
  x is close enough to root, should it exist
fx >>
```

Notice that we do not have a root at $x \approx -0.324$. The algorithm stopped because $x_8 \approx x_5$, whereas two consecutive swaps of endpoints after x_6 and x_7 computed resulted in the algorithm comparing x_8 directly with x_5, and it got tricked into believing that $x_n \approx x_{n+1}$.

The early iterations forced x_n's to move leftwards until x_6 got computed which resulted in a slightly larger value $f(x_6)$ than $f(x_5)$, despite $|x_6 - x_5|$ being relatively large. This then leads to a swap of endpoints and a large, positive x_7 which results in an enormous $f(x_7)$. Clearly picking $x_0 = 3$ and $x_1 = 10$ is a bad idea.

Textbook Problems

3.4.2 Prove that if $F : [a,b] \to \mathbb{R}$, if $F' \in C^0$, and if $|F'(x)| < 1$ on $[a,b]$, then F is a contraction. Does F necessarily have a fixed point?

> ***Proof.*** Since $[a,b]$ is compact and F' is continuous, the derivative attains its maximum $F'(\xi)$ for some $\xi \in [a,b]$. Then since $F'(\xi) < 1$, applying MVT gives
>
> $$|F(x) - F(y)| = F'(\xi^*)|x-y| \leqslant F'(\xi)|x-y|$$
>
> for some ξ^* between x and y, and this shows F is contractive.
>
> For fixed point, no. Consider $F(x) := 1 + x/2$ on $[0,1]$ — the only solution to $1 + x/2 = x$ is $x = 2$, outside the domain.
>
> *A stronger argument exists: even if $F : \mathbb{R} \to \mathbb{R}$, it still does not necessarily have a fixed point. Consider the Signoid function*
>
> $$S(x) = \frac{1}{1+e^{-x}}.$$
>
> *Taking the antiderivative of $1 - S(x)$ gives*
>
> $$F(x) = \int 1 - \frac{1}{1+e^{-x}}\,dx = x - \ln(e^x+1) + C.$$
>
> *Setting $C = 0$, this function has derivative $\in (0,1)$ everywhere and $F(x) < x$ for all x, as $\ln(e^x+1)$ is never 0.*
>
> *However, replacing $|F'(x)| < 1$ by $|F'(x)| \leqslant 1 - \epsilon < 1$ guarantees the existence of a fixed point, as one can then extract a Cauchy sequence with the iteration $x_{n+1} = F(x_n)$.* □

3.4.3 Prove that if F is a continuous map from $[a,b] \to [a,b]$ then F must have a fixed point. Determine if this assertion holds for $F : \mathbb{R} \to \mathbb{R}$.

> ***Proof.*** Define $G : [a,b] \to \mathbb{R}$ by $G(x) := F(x) - x$. Since the minimal value F can attain is a, we see that $G(a) \geqslant 0$. Likewise, $G(b) \leqslant 0$. If one (or both) of these endpoints satisfies the equation we immediately have a fixed point. Otherwise, since G is the difference between two continuous functions, it is continuous, and by IVT there exists some $\xi \in (a,b)$ such that $G(\xi) = 0$, and such ξ is a fixed point.
>
> The claim is not true in general if we consider $F : \mathbb{R} \to \mathbb{R}$. The italic paragraphs above serve as an example. For a simpler one, consider $F(x) := x + 1$ from $\mathbb{R} \to \mathbb{R}$. □

3.4.5 Kepler's equation reads $x = y - \epsilon \sin y$, $\epsilon \in (0,1)$. Show that for each $x \in [0,\pi]$ there is a y satisfying the equation. Interpret this as a fixed-point problem.

> ***Proof.*** Define $F(y) := y - \epsilon \sin y$. Since $F(0) = 0$ and $F(\pi) = \pi$, this claim is trivial for $x = 0$ and $x = \pi$. We'll now assume $x \in (0,\pi)$. Let any such x_0 be given. Consider $G(y) := y - \epsilon \sin y - x_0$. Clearly since $x_0 > 0$ we have $G(0) < 0$, and since $x_0 < \pi$ we have $G(\pi) > 0$. Continuity of F implies that of G, so by MVT there exists some $\xi \in (0,\pi)$ such that $G(\xi) = 0$, i.e., $F(\xi) = x_0$, as desired. □

> ***Proof using fixed point.*** Let x be given. Define $h(y) := x + \epsilon \sin(y)$ on $[0,\pi]$, whose derivative, $\epsilon\cos(y)$, is always between $[0,\epsilon]$, hence a contractive mapping. Therefore by the contractive mapping theorem, h admits a fixed point $y' \in [0,\pi]$ such that $h(y') = y' = x + \epsilon\sin(y')$, i.e., $x = y' + \epsilon\sin(y')$. □

3.4.12 Let p be a positive number. Evaluate

$$x = \sqrt{p + \sqrt{p + \sqrt{p + \cdots}}}.$$

Solution

Let $x_1 = \sqrt{p}$ and $x_{n+1} = \sqrt{p + x_n}$. The above x is the same as $\lim_{n\to\infty} x_n$. We now show that the function $f(x) := \sqrt{p+x}$ is contractive for $p > 1$ (assuming $f : [-p, \infty) \to [0, \infty)$). Indeed, since f is continuous, for all $a < b$ (in the domain), $f(b) - f(a) = f'(\xi)(b-a)$ for some $\xi \in [a, b]$. However, notice that

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[\sqrt{p+x}\right] = \frac{1}{2\sqrt{p+x}} \in (0, 1/2]$$

with the assumption $p > 1$ and $x > -p$. This means $f'(\xi) \leqslant 1/2$ for all ξ in the domain of f, and thus f is contractive. Since $[-p, \infty)$ is closed in $\mathbb{R}$, by the contractive mapping theorem f admits a fixed point x where $\sqrt{p+x} = x \implies p + x = x^2 \implies x^2 - x - p = 0 \implies x = (1 + \sqrt{4p+1})/2$. (We discard the other solution because x is clearly positive in this context.)

3.4.23 Find the order of convergence of these sequences:

$$(a)\ x_n = \sqrt{1/n} \qquad (b)\ x_n = \sqrt[n]{n} \qquad (c)\ x_n = \sqrt{1 + 1/n} \qquad (d)\ x_{n+1} = \tan^{-1} x_n$$

Solution

(a) We see that $\lim_{n\to\infty} \sqrt{1/n} = 0$, therefore $e_n = x_n$. Notice that $\lim_{n\to\infty} (\sqrt{1/(n+1)})/(\sqrt{1/n}) = 1$ whereas if $p > 1$ (power of denominator), using L'Hôpital's rule gives a divergent limit. Hence the order of convergence is 1.

(b) Notice that $x_n \to 1$ as $n \to \infty$. This sequence converges with order 1 because

$$\lim_{n\to\infty} \frac{1 - (n+1)^{1/(n+1)}}{1 - n^{1/n}} = \lim_{n\to\infty} \frac{(n+1)^{1/(n+1)-2}(\log(n+1) - 1)}{n^{1/n-2}(\log(n) - 1)} = 1,$$

while the limit does not exist if $1 - n^{1/n}$ is raised to some power > 1.

(c) Once again, $x_n \to 1$. The convergence is of order 1 since

$$\lim_{n\to\infty} \frac{1 - (1 + 1/(n+1))^{1/2}}{(1 - (1 + 1/n)^{1/2}} = \lim_{n\to\infty} \frac{2n^2\sqrt{1 + 1/n}}{2(n+1)^2\sqrt{1 + 1/(n+1)}} = 1$$

while if $p > 1$ then the limit does not exist.

(d) First observe that $x_n \to 0$: $f(x) = \tan^{-1}(x)$ is contractive and the only solution to $x = arc\tan x$, i.e., fixed point, is $x = 0$. Then since

$$\frac{\mathrm{d}}{\mathrm{d}x} \tan^{-1}(0) = \frac{1}{0^2 + 1} = 1$$

we see that this sequence converges with order 1 (since arctan's first derivative does not vanish at 0).

3.4.25 Prove that $F(x) := 4x(1-x)$ that maps $[0,1]$ into itself is not a contraction but has a fixed point. Why does this not contradict the contractive mapping theorem?

Solution

Indeed, $|F(1) - F(0.5)| = |0 - 1| = 1 > |1 - 0.5|$, so F is not contractive. On the other hand, $F(0) = 0$ so that is a fixed point. (In fact, $x = 0.75$ is also a fixed point.) This does not contradict the contractive mapping theorem because the latter says nothing about functions that are not contractive; it only says contractive mappings on a closed set has a fixed point.

4.1.6 A **monomial** matrix is a square matrix in which each row and column contains exactly one nonzero entry. Prove that a monomial matrix is nonsingular.

Proof. If a $n \times n$ matrix is monomial, then its columns form a standard basis of $\mathbb{R}^n$: there exists exactly one column in which the first entry (entry from first row) is 1 and all others are 0, corresponding to $[1\ 0 \dots\ 0]^T$. Then there exists exactly one column, different from the one before, that is of form $[0\ 1\ 0 \dots\ 0]^T$. Eventually we have n linearly independent column vectors, and their linear independence implies the matrix's nonsingularity. □

4.1.7 Let A have the block form

$$A = \begin{bmatrix} B & C \\ 0 & I \end{bmatrix}$$

in which the blocks are $n \times n$. Prove that if $B - I$ is nonsingular then for $k \geqslant 1$,

$$A^k = \begin{bmatrix} B^k & (B^k - I)(B - I)^{-1}C \\ 0 & I \end{bmatrix}.$$

Proof. This is highly analogous to the following equation (of numbers):

$$x^n - 1 = (x - 1)(x^{n-1} + x^{n-2} + \cdots + 1).$$

Similarly, for matrix B and identity I,

$$B^k - I = (B - I)(B^{k-1} + B^{k-2} + \cdots + I).$$

One can easily check the above equation by expanding all terms and seeing all B^n cancel out except for B^{k+1} (and I).

It is clear that the top-left, bottom-left, and bottom-right entries of A^k are B^k, 0, and I. It remains to show that the top-right entry is $(B^k - I)(B - I)^{-1}C$. Let $\varphi(n)$ be the statement that B^n indeed has top-right entry of this form. Immediately we see $\varphi(1)$ holds:

$$(B^1 - I)(B - I)^{-1}C = IC = C.$$

For the inductive step, assume $\varphi(k)$ holds. Then, the top-right entry of B^{k+1} is given by

$$\begin{aligned}\begin{bmatrix} B^k & (B^k - I)(B-I)^{-1}C \end{bmatrix}\begin{bmatrix} C \\ I \end{bmatrix} &= B^kC + (B^k - I)(B-I)^{-1}C \\ &= B^kC + (B^{k-1} + B^{k-2} + \cdots + I)C \\ &= (B^k + B^{k-1} + \cdots + I)C \\ &= (B^{k+1} - I)(B-I)^{-1}C,\end{aligned}$$

as long as $B - I$ is invertible. The claim then follows from the induction. □

4.1.10 Prove that the set of upper triangular $n \times n$ matrices is a subalgebra of the algebra of all $n \times n$ matrices.

Proof. Closures of addition and of scalar multiplication between upper triangular matrices are trivial. Now let A, B be two upper triangular $n \times n$ matrices and let $C := AB$.

$$AB = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1,n-1} & a_{1,n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2,n-1} & a_{2,n} \\ 0 & 0 & a_{33} & \cdots & a_{3,n-1} & a_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & a_{n,n} \end{bmatrix}\begin{bmatrix} b_{11} & b_{12} & b_{13} & \cdots & b_{1,n-1} & b_{n,n} \\ 0 & b_{22} & b_{23} & \cdots & b_{2,n-1} & b_{2,n} \\ 0 & 0 & b_{33} & \cdots & b_{3,n-1} & b_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & b_{n,n} \end{bmatrix}.$$

Now we consider entries of C. If $i > j$, i.e., C_{ij} is below its diagonal, then it is 0 because

$$C_{ij} = a_{i,1}b_{1,j} + a_{i,2}b_{2,j} + \cdots + a_{i,i-1}b_{i-1,j} + a_{i,i}b_{i,j} + \cdots + a_{i,n}b_{n,j}.$$

The terms highlighted in red are 0 because they are below A's diagonal, while those highlighted in violet are 0 because they are below B's diagonal.

We could verify that $i \leqslant j$ implies $C_{i.j}$ is possibly nonzero (for example the cyan row times the violet column), but such check is unnecessary because showing any entry below the diagonal is zero suffices to prove C is upper-triangular. □

4.1.13 Let A be an invertible $n \times n$ matrix and let $u, v \in \mathbb{R}^n$. Find the necessary and sufficient conditions on u and v in order that the matrix

$$\begin{bmatrix} A_{n\times n} & \mathbf{u}_{n\times 1} \\ \mathbf{v}^T_{1\times n} & 0_{1\times 1} \end{bmatrix}$$

be invertible, and give a formula for the inverse when it exists.

†Bold letter such as $\mathbf{v}$ refer to vectors, in this context $1 \times n$ or $n \times 1$ matrices.

Solution

For convenience we call this $(n+1) \times (n+1)$ matrix M. If M is invertible then it must be full rank, and so block Gaussian elimination characterizes the invertibility of M: it is invertible if and only if

$$\begin{bmatrix} A_{n\times n} & \mathbf{u}_{n\times 1} \\ \mathbf{v}^T_{1\times n} & 0_{1\times 1} \end{bmatrix} \sim \begin{bmatrix} A_{n\times n} & \mathbf{u}_{n\times 1} \\ \mathbf{0}_{1\times n} & -v^TA^{-1}u \end{bmatrix} \text{ is invertible } \iff \boxed{\text{the Schur complement } -\mathbf{v}^TA^{-1}\mathbf{u} \neq 0.}$$

To find its inverse, it suffices to find its right inverse since M is a square matrix. Suppose

$$\begin{bmatrix} A_{n\times n} & \mathbf{u}_{n\times 1} \\ \mathbf{v}^T_{1\times n} & 0 \end{bmatrix}\begin{bmatrix} B_{n\times n} & \mathbf{x}_{n\times 1} \\ \mathbf{y}^T_{1\times n} & c \end{bmatrix} = I_{(n+1)\times(n+1)}.$$

Multiplying the elimination matrix (obtained from block Gaussian elimination above) on the left:

$$\begin{bmatrix} I_{n\times n} & \mathbf{0}_{n\times 1} \\ -\mathbf{v}^T A^{-1}_{1\times n} & 1 \end{bmatrix} MM^{-1} = \begin{bmatrix} A_{n\times n} & \mathbf{u}_{n\times 1} \\ \mathbf{0}_{1\times n} & -\mathbf{v}^T A^{-1}\mathbf{u} \end{bmatrix}\begin{bmatrix} B_{n\times n} & \mathbf{x}_{n\times 1} \\ \mathbf{y}^T_{1\times n} & c \end{bmatrix} = \begin{bmatrix} I_{n\times n} & \mathbf{0}_{n\times 1} \\ -\mathbf{v}^T A^{-1}_{1\times n} & 1 \end{bmatrix}.$$

We can easily compute the bottom row entries $\mathbf{y}^T$ and c (since the bottom-left entry of M is 0):

$$(-\mathbf{v}^T A^{-1}\mathbf{u})\cdot c = 1 \implies c = -\frac{1}{\mathbf{v}^T A^{-1}\mathbf{u}}, \text{ and}$$

$$(-\mathbf{v}^T A^{-1}\mathbf{u})\cdot y^T = -\mathbf{v}^T A^{-1} \implies \mathbf{y}^T = \frac{\mathbf{v}^T A^{-1}}{\mathbf{v}^T A^{-1}\mathbf{u}}.$$

Also, since $A\mathbf{x} + c\mathbf{u} = 0$ (top-right of RHS),

$$\mathbf{x} = A^{-1}(-c\mathbf{u}) = \frac{A^{-1}\mathbf{u}}{\mathbf{v}^T A^{-1}\mathbf{u}},$$

and since $AB + \mathbf{u}\mathbf{y}^T = I$ (top-left of RHS),

$$B = A^{-1}AB = A^{-1}I - A^{-1}\mathbf{u}\mathbf{y}^T = A^{-1} - \frac{A^{-1}\mathbf{u}\mathbf{v}^T A^{-1}}{\mathbf{v}^T A^{-1}\mathbf{u}}.$$

To sum up, the inverse of M is

$$\begin{bmatrix} A^{-1} - \dfrac{A^{-1}\mathbf{u}\mathbf{v}^T A^{-1}}{\mathbf{v}^T A^{-1}\mathbf{u}} & \dfrac{A^{-1}\mathbf{u}}{\mathbf{v}^T A^{-1}\mathbf{u}} \\ \dfrac{\mathbf{v}^T A^{-1}}{\mathbf{v}^T A^{-1}\mathbf{u}} & -\dfrac{1}{\mathbf{v}^T A^{-1}\mathbf{u}} \end{bmatrix}.$$

4.1.14 Let D be a matrix in partitioned form $D = \begin{bmatrix} A & B \\ C & I \end{bmatrix}$. Prove that if $A - BC$ is nonsingular then so is D.

Proof. If $A - BC$ is nonsingular then $\det(A - BC) \neq 0$, and we want use this to prove $\det(D) \neq 0$. Similar to LU-decomposition (but with "reversed" order — eliminating backward), consider

$$\underbrace{\begin{bmatrix} I & -B \\ 0 & I \end{bmatrix}}_{U}\underbrace{\begin{bmatrix} A & B \\ C & I \end{bmatrix}}_{D} = \underbrace{\begin{bmatrix} A - BC & 0 \\ C & I \end{bmatrix}}_{L}.$$

Then $\det(U)\det(D) = \det(L) = \det(A - BC)$, where the last equality comes from the fact that Gaussian elimination on L has no effect on the rightmost column, and a pivot 1 has no effect on $\det(L)$, so $\det(L)$ is the same as the determinant of its top-left submaatrix, namely $\det(A - BC)$. Therefore if $\det(A - BC) \neq 0$, we know $\det(D) \neq 0$, i.e., D is nonsingular. □