Fixed point iteration and root finding

The sign function is defined as

\[ \text{sign}(x) = \begin{cases} 
1 & x > 0 \\
0 & x = 0 \\
-1 & x < 0 
\end{cases} \]

It can be evaluated via an iteration which is useful for some problems. One such iteration is given by

\[ x_{k+1} = \frac{3x_k - x_k^3}{2}. \quad (4) \]

a) Find all fixed points of this iteration. How many of the fixed points are stable?

b) Find the order of convergence of the iteration at the stable fixed point(s).

c) Show that for \( x \in [-1, 1] \), \( x \neq 0 \), the iteration (4) always converges.

SOLUTION:

a) Solving the equation \( x = (3x - x^3)/2 \equiv g(x) \) gives the fixed points \( x = \pm 1 \) and \( x = 0 \). The fixed points \( x = \pm 1 \) are stable since \( g'(\pm 1) = 0 \), and the fixed point \( x = 0 \) is unstable since \( g'(0) = 3/2 > 1 \).

b) Let \( \alpha \) denote the fixed point \( x = -1 \) or \( x = 1 \). Define the error \( \epsilon_n \equiv x_n - \alpha \). Taylor expanding \( x_{n+1} = g(x_n) \) around \( \alpha \) gives us

\[ x_{n+1} = g(x_n) = g(\alpha) + g'(\alpha)\epsilon_n + \frac{g''(\alpha)}{2}\epsilon_n^2 + O(\epsilon_n^3). \]

Since \( g(\alpha) = \alpha \), we have that \( \epsilon_{n+1} = g(x_n) - g(\alpha) \) and hence

\[ \epsilon_{n+1} = g'(\alpha)\epsilon_n + \frac{g''(\alpha)}{2}\epsilon_n^2 + O(\epsilon_n^3). \]

Since \( g'(\pm 1) = 0 \) and \( |g''(\alpha)| = 3 \) we have that the lowest non-zero term is \( O(\epsilon_n^2) \) and we have quadratic convergence.
c) We show that $g([0, 1]) \subset [0, 1]$ by first observing that $g'(x) = \frac{3}{2} - \frac{3x^2}{2} > 0$ for $x \in (0, 1)$. Therefore the function $g(x)$ is monotonically increasing from $g(0) = 0$ to the local maximum $g(1) = 1$ and hence $g([0, 1]) \subset [0, 1]$.

Since

$$
\frac{x_{k+1}}{x_k} = \frac{3x_k - x_k^3}{2} = \frac{3}{2} - \frac{x_k^2}{2},
$$

we have that $\frac{x_{k+1}}{x_k} > 1$ for $x \in (0, 1)$. Hence, if $x_0 \in (0, 1)$ then $\{x_k\}_{k=0}^\infty$ is a monotonically increasing sequence in the interval $[0, 1]$ where it must have a point of accumulation. Since $x = 0$ and $x = 1$ are the only fixed points in this interval, the monotonically increasing sequence must converge to $x = 1$.

(The argument when $x_0 \in [-1, 0)$ is analogous.)
Quadrature

Gaussian quadratures

$$\int_{-1}^{1} f(x) \, dx = \sum_{k=0}^{N} w_k f(x_k)$$

with nodes including the endpoints ($x_0 = -1$ and $x_N = 1$) are called \textit{Gauss-Legendre-Lobatto} quadratures.

a) Show that if the interior nodes $x_1, \ldots, x_{N-1}$ in the quadrature are given by the roots of $p_N'(x)$, where $p_N(x)$ denotes the $N$:th degree Legendre polynomial, then the quadrature is exact for polynomials up to degree $2N - 1$.

Hint: The $2N - 1$:th degree Legendre polynomial can be written as

$$p_{2N-1}(x) = s_{N-2}(x)(x^2 - 1)p_N'(x) + r_N(x)$$

where $s_{N-2}(x)$ and $r_N(x)$ are $N - 2$:th and $N$:th degree polynomials, respectively.

b) Find the 4-point ($N = 3$) Gauss-Legendre-Lobatto quadrature (nodes and weights) for the integrals $\int_{-1}^{1} f(x) \, dx$.

Hint: The three term recursion for Legendre polynomials is given by

$$P_0(x) = 1, \; P_1(x) = x \; \text{and} \; (k + 1)P_{k+1} - (2k + 1)xP_k + kP_{k-1} = 0.$$  

\textbf{SOLUTION:}

a) We want to show that with the given nodes, weights can be found such that

$$\int_{-1}^{1} p_{2N-1}(x) \, dx = \sum_{k=0}^{N} w_k p_{2N-1}(x_k).$$

From the hint we have that

$$\int_{-1}^{1} p_{2N-1}(x) \, dx = \int_{-1}^{1} s_{N-2}(x)(x^2 - 1)p_N'(x) \, dx + \int_{-1}^{1} r_N(x) \, dx.$$  

Integration by parts of the first integral to the right gives

$$\int_{-1}^{1} s_{N-2}(x)(x^2 - 1)p_N'(x) \, dx = -\int_{-1}^{1} \frac{d}{dx} \left( s_{N-2}(x)(x^2 - 1) \right) p_N(x) \, dx.$$
Since $\frac{d}{dx}(s_{N-2}(x)(x^2 - 1))$ is a polynomial of degree $N-1$, it can be expanded into $p_0(x), \ldots, p_{N-1}(x)$. Since these polynomials are orthogonal to $p_N(x)$, the integral $\int_{-1}^{1} s_{2N-1}(x^2 - 1)p_N'(x) \, dx = 0$, and hence
\[
\int_{-1}^{1} p_{2N-1}(x) \, dx = \int_{-1}^{1} r_N(x) \, dx. \quad (2)
\]

Since $x_0 = -1$, $x_N = 1$, and $\{x_k\}_{k=1}^{N-1}$ are the roots of $p_N'(x)$, it follows that
\[
\sum_{k=0}^{N} w_k p_{2N-1}(x_k) = \sum_{k=0}^{N} w_k s_{N-2}(x_k)(x_k^2 - 1)p_N'(x_k) \, dx + \sum_{k=0}^{N} w_k r_N(x_k) \, dx
\]
\[
= \sum_{k=0}^{N} w_k r_N(x_k) \, dx. \quad (3)
\]

To show (1) we see from (2) and (3) that we must require that $\int_{-1}^{1} r_N(x) \, dx = \sum_{k=0}^{N} w_k r_N(x_k)$ for $r_N(x) = x^j$, $j = 0, \ldots, N$. This gives a linear (Vandermonde) system for the weights $w_k$ which can be solved since the nodes are distinct.

b) Using the recursion formula for Legendre polynomials we find that $p_3(x) = \frac{5x^3 - 3x}{2}$, and hence $p_3'(x) = \frac{15x^2 - 3}{2}$ with the roots $\pm \frac{1}{\sqrt{5}}$. Hence, the nodes are $x_0 = -1$, $x_1 = -1/\sqrt{5}$, $x_2 = -\sqrt{3}$ and $x_3 = 1$.

We find the weights by solving the system
\[
\sum_{k=0}^{3} w_k x_k^j = \int_{-1}^{1} x^j \, dx, \quad j = 0, \ldots, 3 \quad \Leftrightarrow
\]
\[
\begin{align*}
w_0 + w_1 + w_2 + w_3 &= 2 \\
-w_0 - \frac{w_1}{\sqrt{5}} + \frac{w_2}{\sqrt{5}} + w_3 &= 0 \\
w_0 + \frac{w_1}{5} + \frac{w_2}{5} + w_3 &= \frac{2}{3} \\
-w_0 - \frac{w_1}{5\sqrt{5}} + \frac{w_2}{5\sqrt{5}} + w_3 &= 0
\end{align*}
\]
which by using symmetry ($w_0 = w_3$ and $w_1 = w_2$) is reduced to
\[
\begin{align*}
2w_0 + 2w_1 &= 2 \\
2w_0 + \frac{2w_1}{5} &= \frac{2}{3}
\end{align*}
\]
with the solution $w_0 = 1/6$ and $w_1 = 5/6$. 

2
Interpolation / Approximation:

3. a. Define what is meant by a cubic spline.

b. Show that giving just function values \( f(x_i), i = 1, \ldots, N \) is insufficient to determine uniquely a cubic spline through the points. Work out how many additional conditions that are required for this.

c. Two ways to obtain a unique cubic spline interpolant are known as 'natural spline' and 'not-a-knot' spline. Define these two spline types.

d. A cubic \( B \)-spline is the cubic spline that is non-zero over the narrowest interval. In the case of a unit-spaced grid, it will look as shown in the figure to the right; non-zero over four subintervals only (taking values at the nodes and at subinterval midpoints as shown, when normalized in the usual fashion).

If one knows the values of such a \( B \)-spline \( f(x) \) only at the node points, one can very easily tabulate it at increasingly denser sets of point locations by means of repeatedly using the relation

\[
f(x) = \frac{1}{3}(f(2x-2) + 4f(2x-1) + 6f(2x) + 4f(2x+1) + f(2x+2))
\]  

(2)

Show that (2) holds true for all values of \( x \).

Hint: One way to solve the problem, requiring only little algebra, is to let \( f(x) \) in the right hand side of (2) be the cubic \( B \)-spline, and then show that uniqueness of cubic splines implies that the whole expression must evaluate to the very same spline.

Solution:

a. A cubic spline \( s(x) \) is a piecewise cubic function which, at the nodes, feature continuity in \( s, s', s'' \) (but may jump in \( s''' \)).

b. With \( f(x_i), i = 1, \ldots, N \) given, we need to determine \( N-1 \) cubics, i.e. \( 4N-4 \) parameters. The conditions we have are

- Function values at the ends of each sub-interval \( \frac{2(N-1)\text{ }}{2} \)
- Continuity of \( s' \) at interior nodes \( N-2 \)
- Continuity of \( s'' \) at interior nodes \( N-2 \)

for a total of \( 4N-6 \). We fall short of 2 conditions.

c. Natural spline: Add two conditions by imposing also \( s''(x_1) = s''(x_N) = 0 \).

Not-a-knot spline: Remove two freedoms by not allowing the jump in \( s''' \) at \( x_2 \) and \( x_{N-1} \).

d. Substituting the \( B \)-spline \( f(x) \) into the right hand side of (2) gives a piecewise cubic result, which we call \( h(x) \). It will clearly satisfy:

- \( h(x) = 0 \) for \( x < -2 \) and \( x > +2 \).
- Continuity of \( h, h', h'' \) everywhere, and possible jumps in \( h''' \) at \( x = -2, -\frac{3}{2}, -1, -\frac{1}{2}, 0, \frac{1}{2}, 1, \frac{3}{2}, 2 \). In particular, \( h''(2) = h''(\frac{1}{2}) = 0 \).
- At the points \( x = -2, -\frac{3}{2}, -1, -\frac{1}{2}, 0, \frac{1}{2}, 1, \frac{3}{2}, 2 \), \( h(x) \) evaluates to the same values as are given in the figure for \( f(x) \).

A cubic spline which takes certain values at the nodes and also satisfies zero second derivative end conditions is the unique natural spline. Both \( h(x) \) and \( f(x) \) (the \( B \)-spline) satisfies all the properties just listed. So they have to be the same.
Linear Algebra Let $A$ be an $m \times n$ matrix, with $m \geq n$, and $b$ and $m \times 1$ vector. Consider the least-squares problem

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|^2,$$

which can be written as

$$Ax \simeq b. \quad (1)$$

a. Explain why any solution of the normal equations is a solution of this problem, regardless of the rank of $A$. Show that the solution is unique if the rank of $A$ is $n$.

b. Describe the solution of this problem in terms of the singular value decomposition of $A$.

Consider the regularized least-squares problem

$$\min_{x \in \mathbb{R}^n} \left(\|Ax - b\|^2 + \lambda^2 \|x\|^2\right).$$

c. Derive a form of the normal equations for this problem. Use it to argue that the regularized problem has a unique solution whenever $\lambda \neq 0$, regardless of the rank of $A$. (Hint: Write the regularized least-square problem in the form of equation (1) using an $(m+n) \times n$ linear system.)

d. Write the solution of the regularized problem in terms of the singular value decomposition of $A$.

e. Assume that $A$ has only one zero singular value and $b$ is not orthogonal to the corresponding left singular vector of $A$. Describe the effect on the solution, $x$, of the regularized problem as $\lambda \to 0$.

**Solution:**

a. Any solution of the least-squares problem satisfies the orthogonality condition

$$< Ax - b, Az >= 0 \quad \forall z \in \mathbb{R}^n$$

and, thus, $x$ satisfies the normal equations, which are given by

$$A^tAx = A^tb.$$

If $A$ has rank $n$, then $A^tA$ is nonsingular and

$$x = (A^tA)^{-1}A^tb.$$
b. Let the singular value decomposition be given as $A = U\Lambda V^*$ where $U$ is an $m \times m$ unitary matrix, $V$ is an $n \times n$ unitary matrix and $\Lambda$ is an $m \times n$ matrix such that

$$\Lambda = \text{diag}\{\sigma_1, \sigma_2, \ldots, \sigma_n\}, \quad 0 \leq \sigma_1 \leq \sigma_2 \leq \cdots \leq \sigma_n$$

Substituting in the equation above and simplifying, we have

$$x = V\Lambda^tU^*b + z$$

where

$$\Lambda^t = \text{diag}\{\sigma^*_1, \sigma^*_2, \ldots, \sigma^*_n\}, \quad \sigma^*_j = \begin{cases} 0 & \sigma_j = 0 \\ \frac{1}{\sigma_j} & \sigma_j \neq 0 \end{cases}$$

and $z$ is any null vector of $A$. Writing it out in terms of the columns of $U$ and $V$ yields

$$x = \sum_{\sigma_j \neq 0} \frac{1}{\sigma_j}v_ju_j^tb + z$$

c. Write the system as

$$\begin{bmatrix} A \\ \lambda I \end{bmatrix} x \approx \begin{bmatrix} b \\ 0 \end{bmatrix}.$$ 

The normal equations associated with the regularized problem are

$$\begin{bmatrix} A^t \lambda I \\ \lambda I \end{bmatrix} \begin{bmatrix} A \\ \lambda I \end{bmatrix} x = \begin{bmatrix} A^t \lambda I \end{bmatrix} \begin{bmatrix} b \\ 0 \end{bmatrix}$$

which become

$$(A^tA + \lambda^2I)x = A^tb.$$  \hspace{1cm} (2)

Since, $A^tA$ is positive semidefinite and, if $\lambda \neq 0$, $\lambda^2I$ is positive definite, then $A^tA + \lambda^2I$ is nonsingular. This implies that equation(2) has a unique solution.

d. Plugging the singular value decomposition in to these new normal equations, we have

$$x = VDU^*b$$

where

$$D = \text{diag}\left\{\frac{\sigma_j}{\sigma_j^2 + \lambda^2}, \ldots\right\}.$$
Writing it out in terms of the columns of $U$ and $V$ yields

$$x = \sum_{j=1}^{n} \frac{\sigma_j}{\sigma_j^2 + \lambda^2} v_j u_j^t b = \sum_{\sigma_j \neq 0} \frac{\sigma_j}{\sigma_j^2 + \lambda^2} v_j u_j^t b$$

e. Consider the formula above. If $\sigma_j = 0$, it play no role in the solution. Thus,

$$\lim_{\lambda \to 0} x = \sum_{\sigma_j \neq 0} \frac{1}{\sigma_j} v_j u_j^t b$$

which is the minimal length least squares solution.
5. A general linear multistep formula for solving \( y' = f(t, y) \) takes the form
\[
y_{n+1} = \sum_{j=0}^{p} a_j y_{n-j} + k \sum_{j=-1}^{q} b_j f(t_{n-j}, y_{n-j}).
\]

a. Derive the coefficients \( a_j \) and \( b_j \) for the second order Adams-Bashforth and Adams-Moulton methods.

b. The figure to the right shows the stability domain for the second order Adams-Bashforth method (interior of the closed curve). Can this method be used to solve \( y' = y, \ y' = -y \), and/or \( y' = i y \)? Explain!

c. Determine the stability domain for the second order Adams-Moulton method.

Solution:

a. The AB2-formula for solving \( y' = f(t, y) \) takes the form \( y_{n+1} = a_0 y_n + k [b_0 f_n + b_1 f_{n-1}] \). It should be exact for example when \( k = 1 \) and \( \{ y = 1, f = 0 \}, \{ y = t, f = 1 \}, \{ y = t^2, f = 2t \} \). This gives three equations:
\[
\begin{cases}
1 = a_0 + 0b_0 + 0b_1 \\
1 = 0a_0 + 1b_0 + 1b_1 \\
1 = 0a_0 + 0b_0 - 2b_1 
\end{cases}
\]
with the solution \( a_0 = 1, b_0 = \frac{3}{2}, b_1 = -\frac{1}{2} \).

For AM2, we get similarly \( y_{n+1} = a_0 y_n + k [b_{-1} f_{n+1} + b_0 f_n] \), leading to
\[
\begin{cases}
1 = a_0 + 0b_{-1} + 0b_0 \\
1 = 0a_0 + 1b_{-1} + 1b_0 \\
1 = 0a_0 + 2b_{-1} + 0b_0 
\end{cases}
\]
with the solution \( a_0 = 1, b_{-1} = \frac{1}{2}, b_0 = \frac{1}{2} \).

b. The stability domain tells for which values of the time step \( k \) solutions to the ODE \( y' = \lambda y \) can feature growing solutions. That is not the information we need to decide convergence. This instead requires
- consistency (here OK; already first order accuracy would be slightly more than what is needed)
- stability (root condition OK; the equation to test is \( r = 1 \), which obviously has as only root \( r = 1 \))

So the method can be used for all the cases mentioned.

c. We find the stability domain by applying the method to \( y' = \xi y \), giving \( y_{n+1} + y_n + \frac{k\xi}{2} [y_{n+1} + y_n] \). Writing \( k\lambda = \xi \), this becomes \( y_{n+1} = \frac{1 + \xi/2}{1 - \xi/2} y_n \). The relation \( \frac{1 + \xi/2}{1 - \xi/2} < 1 \) holds precisely when \( \text{Re} \xi < 0 \). Hence, the stability domain is the left half plane.
6. Finite Difference Equations for PDEs

Consider Richardson’s difference scheme for the heat equation, \( u_t = u_{xx} \):

\[
\frac{1}{2k} (u(x, t + k) - u(x, t - k)) = \frac{1}{h^2} (u(x - h, t) - 2u(x, t) + u(x + h, t))
\]

a. Show that this scheme has second-order truncation error.

b. Use either ODE principles or von Neumann analysis to show that this scheme is unconditionally unstable.

c. Demonstrate a minor modification of the left-side of Richardson’s scheme that yields a familiar unconditionally stable scheme and prove it.

Solution

a. Taylor expansion of left-hand side yields

\[
\frac{1}{2k} (u(x, t + k) - u(x, t - k)) = u_t(x, t) + \frac{1}{6} u_{ttt}(x, t) k^2 + O(k^4)
\]

Likewise, a Taylor expansion of the right-hand side yields

\[
\frac{1}{h^2} (u(x-h, t) - 2u(x, t) + u(x+h, t)) = u_{xx}(x, t) + \frac{1}{12} u_{xxxx}(x, t) h^2 + O(h^4)
\]

b. Using von Neumann analysis, we set \( u(x, t) = a^t e^{i\omega x} \) and solve for \( a^k \).

This yields the equation

\[
a^k - \frac{1}{a^k} = \frac{2k}{h^2} (e^{-i\omega h} - 2 + e^{i\omega h}) = \frac{4k}{h^2} (\cos(\omega h) - 1) = \frac{8k}{h^2} \sin^2(\omega h/2)
\]

Multiplying by \( a^k \) yields

\[
(a^k)^2 + \frac{8k}{h^2} \sin^2(\omega h/2) a^k - 1 = 0
\]

Plugging in the values \( a^k = 0 \) and \( a^k = -1 \) demonstrates that the equation has a real root \( a^k < -1 \) for any \( k, h > 0 \), and \( \omega \).
c. The backward Euler scheme is
\[
\frac{1}{k}(u(x, t) - u(x, t - k)) = \frac{1}{h^2}(u(x - h, t) - 2u(x, t) + u(x + h, t))
\]

A von Neumann analysis yields
\[
a^k - 1 = a^k \frac{2k}{h^2}(e^{-i\omega h} - 2 + e^{i\omega h}) = a^k \frac{4k}{h^2} \cos(\omega h) - 1 = a^k \frac{8k}{h^2} \sin^2(\omega h / 2)
\]

Solving for \(a^k\) yileds
\[
a^k = \frac{1}{1 + \frac{8k}{h^2} \sin^2(\omega h / 2)} < 1.0
\]

for any \(k, h > 0\), and \(\omega\).