# Solutions
# Preliminary Examination in Numerical Analysis
# January, 2017

1. **Root Finding.**

   The roots are -1,0,1. **(a)** First consider $x_0 > 1$. Let $x_{n+1} = 1+\varepsilon$ and $x_n = 1+\delta$ with $\delta > 0$. The iteration gives $0 < \frac{\varepsilon}{\delta} < \frac{2}{3}$, which implies that $1 < x_{n+1} < x_n$. Newton's method will converge monotonically to 1. Next consider $1/\sqrt{3} < x_0 < 1$. As the signs of the numerator and denominator in the rational part of the iteration does not change on the interval under consideration we find that $x_1 > 1$. Finally, $x_0 = 1$ produces $x_1 = 1$. Note that an iteration starting at $1/\sqrt{3} < x_0 < 1$ is not monotonic since it first moves up past $x = 1$ then monotonically decreases back towards 1.

   To answer **(b)**, rewrite the iteration as $x_{n+1} = -\frac{2x_n^3}{1-3x_n^2}$, and note that for $0 \le x_0 < 1/\sqrt{3}$ the next iterate will be non-positive. Insisting that $-x_0 < x_1 \le 0$, so that $x_1$ will be closer to zero than $x_0$ gives the limiting case $x_1 = -x_0$, or $\alpha(1-3\alpha^2) = -2\alpha^3$, which has the solution $\alpha = 1/\sqrt{5}$. Furthermore, whenever $|x_n| < 1/\sqrt{5}$ one finds that $|x_{n+1}| < |x_n|$ so the sequence of absolute values decreases monotonically and must converge, the only possible limit being 0.

   Finally, for **(c)** we may consider the case $f''(x) > 0$ (otherwise consider $-f(x) = 0$). Assume first that $f'(x) > 0$ in the interval, $f(x_0) \ge 0$ by assumption. The situation is as the one pictured in Figure 1 and we thus conclude that $x_1 < x_0$ and that since the tangent lies to the right of the curve it is also true that $\alpha > x_1$. The case $f'(x) < 0$ is handled similarly and the results follows by induction.

2. **Numerical quadrature.**

   The error in the trapezoid rule over a single interval is

$$\int_a^b f(x)\mathrm{d}x = \frac{b-a}{2}(f(b)+f(a)) - \frac{(b-a)^3}{12}f''(\xi)$$

   for some unknown $\xi \in [a,b]$. In our example $f(x) = \ln(x)$ and each interval has unit length (from $k$ to $k+1$). The exact relationship between the integral and the composite trapezoid rule approximation in our case is therefore

$$\int_1^n \ln(x)\mathrm{d}x = \frac{1}{2}\sum_{k=1}^{n-1}(\ln(k)+\ln(k+1)) + \frac{1}{12}\sum_{k=1}^{n-1}\xi_k^{-2}$$

   where $\xi_k \in [k, k+1]$. Plugging this back in to the formula for $\ln(n!)$ we find

$$\ln(n!) = \int_1^n \ln(x)\mathrm{d}x - \frac{1}{12}\sum_{k=1}^{n-1}\xi_k^{-2} + \frac{1}{2}\ln(n).$$

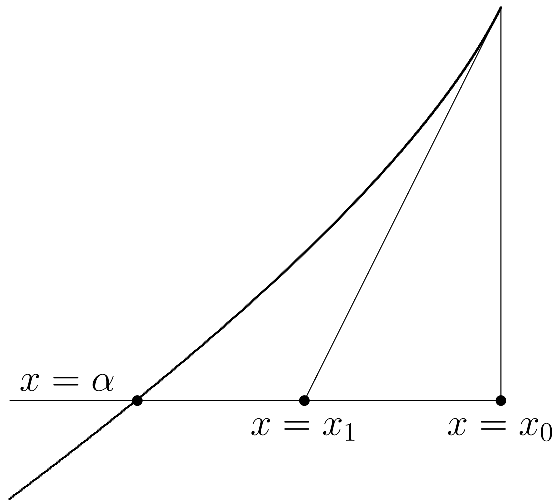


Figure 1: Monotonicity and convexity yields monotone convergence for Newton's method.

Evaluating the integral

$$\ln(n!) = \left(n + \frac{1}{2}\right)\ln(n) - n + 1 - \frac{1}{12}\sum_{k=1}^{n-1} \xi_k^{-2}.$$

Exponentiating:

$$n! = \sqrt{n}(n/e)^n e^{1 - \frac{1}{12}\sum_{k=1}^{n-1} \xi_k^{-2}}.$$

The coefficient in Stirling's formula is clearly

$$C_n = \exp\left\{1 - \frac{1}{12}\sum_{k=1}^{n-1} \xi_k^{-2}\right\}.$$

The sum can be bounded as follows

$$1 \leq \sum_{k=1}^{n-1} \xi_k^{-2} \leq \sum_{k=1}^{n-1} k^{-2} \leq \sum_{k=1}^{\infty} k^{-2} = \frac{\pi^2}{6}$$

which means that the coefficient is bounded by

$$\exp\left\{1 - \frac{\pi^2}{72}\right\} \leq C_n \leq \exp\left\{1 - \frac{1}{12}\right\}.$$

There are, in fact, sharper estimates on $C_n$.

3. **Interpolation/Approximation.**
**(a)** Let $W = V^{-1}$, and the the elements of $W$ be $w_{ij}$. Note that $WV = I$, i.e. that row $i$ satisfies

$$\sum_{j=1}^{n} w_{ij} x_k^j = \delta_{ik}, \quad k = 1, \ldots, n.$$

This interpolation problem is solved by $l_i(x)$, that is:

$$l_i(x_k) = \prod_{\substack{i=1 \\ i \neq j}}^{n} \frac{(x_k - x_i)}{(x_j - x_i)} = \sum_{j=1}^{n} w_{ij} x_k^j = \delta_{ik}, \quad k = 1, \ldots, n,$$

which shows that $V$ is non-singular if and only if $x_i - x_j \neq 0$ when $i \neq j$.
**(b)** Finding the elements $w_{ij}$ is equivalent to finding the coefficients of $l_i(x)$, $i = 1, \ldots, n$. Noting that $l_i(x) = q_i(x)/q_i(x_i)$ we must thus find all the coefficients of each $q_i(x)$ in $\mathcal{O}(n)$ operations. We must also evaluate $q_i(x_i)$. Horner's rule can be used to carry out both tasks. Recall that for the synthetic division of a polynomial $P(x) = \sum_{l=0}^{m} \alpha_l x^l$ by $(x - x_i)$ we must find the polynomial $Q(x) = \sum_{l=1}^{m} \beta_l x^{l-1}$ that satisfies $P(x) = (x - x_i)Q(x) + \beta_0$, (with $\beta_0 = P(x_i)$). A direct computation, matching the coefficients on the sides of the equality sign, shows that the coefficients $\beta_k$ can be computed by the Horner recursion:

$$\beta_k = \alpha_k + \beta_{k+1} x_i, \; k = m-1, m-2, \ldots, 1, 0,$$

and $\beta_m = \alpha_m$.
Applying this to $\Phi_n(x)/(x - x_i)$ we thus may find the $n$ coefficients of each $q_i(x)$ at cost $\mathcal{O}(n)$. Once the coefficients are known $n$ additional applications of Horner's rule yields the $n$ scalars $q_i(x_i)$ at a cost of $\mathcal{O}(n)$ each.

4. **Linear Algebra**

We present a solution with $C = \mathbf{u}\mathbf{v}^T$ based on the ideas presented in the classic 1977 SIAM Review paper *Eigenproblems for Matrices Associated with Periodic Boundary Conditions* by Bjorck and Golub but note that since the rank 2 matrix is not unique other solutions are possible. For example the choice $C = \mathbf{e}_1\mathbf{e}_N^T + \mathbf{e}_N\mathbf{e}_1^T$ will leave the tridiagonal part of the matrix intact allowing for the possibility to exploit the diagonal dominance of the resulting $A'$.

For **(a)** notice that

$$\begin{pmatrix} 4 & 1 & & & 1 \\ 1 & 4 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 4 & 1 \\ 1 & & & 1 & 4 \end{pmatrix} = \begin{pmatrix} 0 & 1 & & & \\ 1 & 4 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 4 & 1 \\ & & & 1 & \frac{15}{4} \end{pmatrix} + \begin{pmatrix} 4 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & \cdots & 0 & \frac{1}{4} \end{pmatrix}.$$

So $A'$ is the tridiagonal matrix on the RHS, and $\mathbf{u} = [4, 0, \ldots, 0, 1]^T$, $\mathbf{v} = [1, 0 \ldots, 0, 1/4]$. For **(b)** let $B = A'$ and use the Sherman-Morrison-Woodbury formula

$$(B + \mathbf{u}\mathbf{v}^T)^{-1} = B^{-1} - \frac{B^{-1}\mathbf{u}\mathbf{v}^T B^{-1}}{1 + \mathbf{v}^T B^{-1}\mathbf{u}}.$$

Now $B\mathbf{y} = \mathbf{g}$ can be solved with $\sim CN$ cost ($C$ is a small integer like 7 or so). To solve $A\mathbf{w} = \mathbf{f}$, we perform the following.
1. Solve $B\mathbf{z} = \mathbf{f}$ at $\sim CN$ cost.
2. Solve $B\mathbf{y} = \mathbf{u}$ at $\sim CN$ cost.
3. Compute both $\alpha = \mathbf{v}^T\mathbf{z}$ and $\beta = \mathbf{v}^T\mathbf{y}$; each dot product costs $2N - 1$.
4. Form $\mathbf{z} - \alpha(1+\beta)^{-1}\mathbf{y}$ at $2N$ cost or so.

5. **ODEs**

(a) Explicit and Implicit Euler applied to the scalar problem $\dot{x} = \lambda x$ yield

$$x_{n+1} = (1+\mu)x_n, \qquad (1-\mu)x_{n+1} = x_n$$

where $\mu = h\lambda$ and $h$ is the time step size. The methods are stable when $|1+\mu| \leq 1$ and $|1-\mu|^{-1} \leq 1$, respectively, for $\mu \in \mathbb{C}$. Explicit Euler is stable for $\mu$ in a circle of unit radius centered at $-1$ in the complex plane; implicit Euler is stable for $\mu$ outside a circle of unit radius centered at 1 in the complex plane.

(b) Explicit Euler applied to this problem yields

$$u_{n+1} = u_n(1 - hu_n^2).$$

Take absolute values:

$$|u_{n+1}| = |u_n|\gamma_n, \qquad \gamma_n = |1 - hu_n^2|.$$

If $\gamma_n < 1$ for every $n$ then the sequence of absolute values is monotone decreasing and bounded below, so it must converge. $\gamma_n < 1$ when $u_n^2 < 2/h$. Clearly, if $u_0^2 < 2/h$ then $u_n^2 < 2/h$ for every $n$, so the sequence of absolute values converges. The limit must satisfy $|u_\infty| = |u_\infty|(1 - h|u_\infty|^2)$ so the only possible limit is $u_n \to 0$. Conversely, when $u_0^2 > 2/h$ all subsequent iterates also satisfy $u_n^2 > 2/h$ and $|u_{n+1}| > |u_n|$; the sequence $\{u_n\}$ alternates sign and can't converge.

(c) Implicit Euler applied to this problem yields

$$u_{n+1}(1 + hu_{n+1}^2) = u_n.$$

Since the function $g(u) = u + hu^3$ is a bijection for every $h \geq 0$, the nonlinear system $g(u_{n+1}) = u_n$ has a unique solution for every $h \geq 0$ and $u_n$. It's easy to see that $|u_{n+1}| < |u_n|$ for every $u_n$, so the sequence of absolute values is monotone decreasing and bounded below, and must converge. The limit must satisfy $|u_\infty|(1 + h|u_\infty^2|) = |u_\infty|$, so the limit is $\lim_{n\to\infty} u_n = 0$ for every $u_0$ and every $h \geq 0$.

(d) Taylor series says

$$u_0 = u(h) + hu(h)^3 - \frac{3h^2}{2}u(\xi)^2$$

for some $\xi \in [0, h]$. The implicit Euler approximation is

$$u_0 = u_1 + hu_1^3.$$

Subtracting yields

$$\frac{3h^2}{2}u(\xi)^2 = (u(h) - u_1)(1 + h(u(h)^2 + u(h)u_1 + u_1^2)).$$

Note that $u(h)^2 + u(h)u_1 + u_1^2 \geq 0$ for every $u(h)$ and $u_1$, so

$$|u(h) - u_1| = \frac{3h^2 u(\xi)^2}{2(1 + h(u(h)^2 + u(h)u_1 + u_1^2))} \leq \frac{3h^2 u(\xi)^2}{2}.$$

If you wish you can further use the fact that $u(\xi)^2 \leq u(0)^2$. The method has second-order truncation error. As $h \to \infty$ this bound on the truncation error also goes to $\infty$. As $h \to \infty$ the implicit Euler approximation satisfies $u_1 \to 0$, and so does the true solution, so the error also goes to zero.

6. **PDEs**

Using the exact solution $u(x - at)$, we have to evaluate

$$\psi(h_t, h_x) = u(x_j - at_{n+1}) - c_{-1} u(x_{j-1} - at_n) - c_0 u(x_j - at_n) - c_1 u(x_{j+1} - at_n)$$

given that $x_{j-1} = x_j - h_x$, $x_{j+1} = x_j + h_x$ and $t_n = t_{n+1} - h_t$. Denoting $x_j - at_{n+1} = s$ for convenience and using the Taylor expansion, we have

$$u(s - h_x + ah_t) = u(s) + u'(s)(ah_t - h_x) + \frac{1}{2}u''(s)(ah_t - h_x)^2 + \dots$$

$$u(s + ah_t) = u(s) + u'(s)\, ah_t + \frac{1}{2}u''(s)(ah_t)^2 + \dots$$

and

$$u(s + h_x + ah_t) = u(s) + u'(s)(ah_t + h_x) + \frac{1}{2}u''(s)(ah_t + h_x)^2 + \dots$$

Thus, we obtain

$$\begin{aligned} \psi(h_t, h_x) &= u(s)(1 - c_{-1} - c_0 - c_1) - u'(s)\left[c_{-1}(ah_t - h_x) + c_0 a h_t + c_1(ah_t + h_x)\right] \\ &- \frac{1}{2}u''(s)\left[c_{-1}(ah_t - h_x)^2 + c_0(ah_t)^2 + c_1(ah_t + h_x)^2\right] + \dots \end{aligned}$$

and arrive at the linear system

$$\begin{cases} c_{-1} + c_0 + c_1 = 1 \\ c_{-1}(ah_t - h_x) + c_0 a h_t + c_1(ah_t + h_x) = 0 \\ c_{-1}(ah_t - h_x)^2 + c_0(ah_t)^2 + c_1(ah_t + h_x)^2 = 0. \end{cases}$$

Setting

$$\nu = a\frac{h_t}{h_x}$$

we obtain

$$c_{-1} = \frac{1}{2}\left(\nu^2 + \nu\right), \quad c_0 = 1 - \nu^2, \text{ and } c_1 = \frac{1}{2}\left(\nu^2 - \nu\right).$$

Assuming periodic boundary conditions, the matrix of this explicit scheme is circulant so that we know the eigenvectors and use them to compute the eigenvalues. For the $k$th eigenvector $e^{2\pi i k h_x j}$ we have

$$\begin{aligned} e^{2\pi i k h_x (j-1)} c_{-1} + e^{2\pi i k h_x j} c_0 + e^{2\pi i k h_x (j+1)} c_1 &= e^{2\pi i k h_x j}\left(e^{-2\pi i k h_x} c_{-1} + c_0 + e^{2\pi i k h_x} c_1\right) \\ &= e^{2\pi i k h_x j}\left(1 - \nu^2 + \nu^2 \cos(2\pi k h_x) + i\nu \sin(2\pi k h_x)\right) \end{aligned}$$

and compute the absolute value of the eigenvalues,

$$|\lambda_k(\nu)|^2 = \left(1 - \nu^2 \left(1 - \cos\left(2\pi k h_x\right)\right)\right)^2 + \nu^2 \sin^2\left(2\pi k h_x\right).$$

We require

$$|\lambda_k|^2 \leq 1$$

for all $k$.

It is optional to show that this inequality holds if $\nu \leq 1$ .