Exercise 1

Question a. *Given the data pairs:*

$$(x, y) \;=\; \{(-1, 0), (-0.5, -1), (0.5, 1), (1, 0)\}. \tag{1}$$

Give first the general form of the interpolation polynomial expressed in the Lagrange characteristic polynomials and next indicate how it is defined for an interpolation on the given data points.

The Lagrange characteristic polynomials are given by

$$\phi_k(x) = \prod_{\substack{j=0\\ j\neq k}}^{n} \frac{x - x_j}{x_k - x_j}$$

and the Lagrange form of the interpolant by

$$\Pi_n(x) = \sum_{k=0}^{n} y_k \phi_k(x)$$

Computing the Lagrange characteristic polynomials gives us,

$$\begin{aligned}
\phi_0(x) &= \left(\frac{x+0.5}{-1+0.5}\right)\left(\frac{x-0.5}{-1-0.5}\right) && = \frac{(x+0.5)(x-0.5)}{(-0.5)(-1.5)}\\
\phi_1(x) &= \left(\frac{x+1}{-0.5+1}\right)\left(\frac{x-0.5}{-0.5-0.5}\right) && = \frac{(x+1)(x-0.5)}{-0.5}\\
\phi_2(x) &= \left(\frac{x+1}{0.5+1}\right)\left(\frac{x+0.5}{0.5+0.5}\right) && = \frac{(x+1)(x+0.5)}{1.5}
\end{aligned}$$

so that Π_2 is given by, $\Pi_2(x) = 0\phi_0(x) - 1\phi_1(x) + 1\phi_2(x)$.

Question b. *The conditioning of interpolation is expressed by the inequality*

$$\max_{x\in I} |\Pi_n(x) - \tilde{\Pi}_n| \leq \Lambda \max_{k\in\{0,\ldots,n\}} |y_k - \tilde{y}_k|$$

wher $\Pi_n(x)$ is the interpolation polynomial based on the pairs (x_k, y_k) and $\tilde{\Pi}_n(x)$ on the pairs $(x_k, \tilde{y}_k)$, $k = 0, \ldots, n$. Show that Lebesque's constant Λ is given by $\Lambda = \sum_{k=0}^{n} \max_{x\in I} |\phi_k(x)|$, where $\phi_k(x), k = 0, \ldots n$, are the Lagrange characteristic polynomials. Give Λ for the interpolation on $[-1, 0.5]$ and data points give in part (a).

$$\begin{aligned}
\max_{x\in I} |\Pi_n(x) - \tilde{\Pi}_n| &= \max_{x\in I} |\sum_{k=0}^{n} y_k\phi_k(x) - \sum_{k=0}^{n} \tilde{y}_k\phi_k(x)|\\
&= \max_{x\in I} |\sum_{k=0}^{n} \phi_k(x)(y_k - \tilde{y}_k)|\\
&\leq \sum_{k=0}^{n} \max_{x\in I} |\phi_k(x)(y_k - \tilde{y}_k)|\\
&\leq \sum_{k=0}^{n} \max_{x\in I} |\phi_k(x)| \max_{k\in\{0,\ldots,n\}} |y_k - \tilde{y}_k|\\
&= \Lambda \max_{k\in\{0,\ldots,n\}} |y_k - \tilde{y}_k|
\end{aligned}$$

Question c. *Define both the midpoint rule and the composite midpoint rule for an integration of a function f over an interfval $[a, b]$.*

The midpoint rule is given by

$$\int_a^b f(x)\ dx \approx I_m(f) = (b-a)f\left(\frac{a+b}{2}\right)$$

The composite trapezoidal rule is given by

$$\int_a^b f(x)\ dx = \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} f(x)\ dx \approx \sum_{i=0}^{n-1} Hf\left(\frac{x_i + x_{i+1}}{2}\right)$$

where $H = (b-a)/n$ and $x_i = a + iH$, $i = 0, \ldots, 1$.

Question d. *The error of the midpoint rule is given by $E^t = -(b-a)^3 f''(\xi)/24$, for some $\xi \in [a, b]$. What is the degree of exactness of this method? Why? Show that the error of the composite midpoint rule is given by $E^c = -(b-a)H^2 f''(\zeta)/24$, for some $\zeta \in [a, b]$, where H is the length of the subintervals in $[a, b]$.*
Hint: *You may use that for any continous function g it holds that there exist a ζ in $[a, b]$ such taht $ng(\zeta) = \sum_{i=0}^{n} g(x_i)$ for an arbitrary set of points x_i, $i = 1, \ldots, n$ in $[a, b]$.*

The degree of exactness of the midpoint rule is 1 because all linear functions are integrated exactly (the error is 0 since the second derivative f'' is zero).
The error using a composite midpoint rule is given by the sum of the errors made in each subinterval,

$$\begin{aligned} E^c = \sum_{i=0}^{n} E_i^t &= \sum_{i=0}^{n} -(x_{i+1} - x_i)^3 f''(\xi_i)/24 \\ &= -\frac{H^3}{24} \sum_{i=0}^{n} f''(\xi_i) \end{aligned}$$

Using the hint we get,

$$\begin{aligned} E^c = \sum_{i=0}^{n} E_i^t &= -\frac{H^3}{24} n f''(\zeta), \qquad n = (b-a)/H, \\ &= -(b-a)\frac{H^2}{24} f''(\zeta), \end{aligned}$$

for some $zeta \in [a, b]$.

Exercise 2

Question a. *Consider the linear system $Ax = b$ and suppose that the matrix A is an $m \times n$ matrix with $m > n$ of full rank (i.e. the columns form an independent set of vectors) leading to an overdetermined equation.*

Question a.i. *One way of solving this is minimizing $(Ax - b, Ax - b)$ over x. Show that this minimization leads to $A^T Ax = A^T b$, where $A^T A$ is a square matrix of order n.*

We want to minimize the dot product with respect to x, that is for each $i = 0, \dots n$ we want,

$$\begin{aligned} 0 &= \frac{\partial}{\partial x_i}(Ax - b, Ax - b) \\ &= (\frac{\partial}{\partial x_i}(Ax - b), Ax - b) + (Ax - b, \frac{\partial}{\partial x_i}(Ax - b)) \end{aligned}$$

Now using $\frac{\partial x}{\partial x_i} = e_i$ with e_i is the standard basis vector,

$$\begin{aligned} 0 &= (Ae_i, Ax - b) + (Ax - b, Ae_i)) \\ &= 2(Ae_i, Ax - b) \\ &= 2(e_i, A^T Ax - A^T b) \end{aligned}$$

Since this should hold for all $i = 0, \dots n$ it follows that we must have,

$$A^T Ax - A^T b = 0$$

This means that we need to solve the system $A^T Ax = A^T b$ where the matrix $A^T A$ is a $n \times n$ matrix.

Question a.ii. *What is the numerical problem with solving the equation in the previous part?*

By solving for $A^T Ax = A^T b$ we increase the condition number making the solution more sensitive to round off errors.

Question b. *Consider the iteration $x^{(k+1)} = Ax^{(k)}$ with $x^{(0)}$ given and suppose that one eigenvalue λ_1 of A is bigger in absolute value than all others. Moreover, A has a complete set of eigenvectors.*

Question b.i. *Show that $x^{(k)}$, will converge to the eigenvector associated to λ_1 if $x^{(0)}$ has a nonzero component in the direction of this eigenvector. Also indicate the convergence factor.*

First we write $x^{(0)} = \sum_{i=1}^{n} \alpha_i v_i$ where v_i is the eigenvector of A corresponding to the eigenvalue $lambda_i$. It follows that,

$$\begin{aligned} x^{(k)} = Ax^{(k)} = A^k x^{(0)} = A^k \sum_{i=1}^{n} \alpha_i v_i &= \sum_{i=1}^{n} \alpha_i \lambda_i^k v_i \\ = \lambda_1^k (\alpha_1 v_1 + & \underbrace{\sum_{i=2}^{n} \alpha_i (\frac{\lambda_i}{\lambda_1})^k v_i}_{\text{Goes to 0 as } k \to \infty, \text{ since } |\frac{\lambda_i}{\lambda_1}| < 1}) \end{aligned}$$

Thus $x^{(k)}$ converges to the direction of v_1 with a convergence factor of $\frac{\lambda_2}{\lambda_1}$.

Question b.ii. *How can we obtain an estimate of λ_1 during the iteration?*

Since $x^{(k+1)} = Ax^{(k)} \approx \lambda_1 x^{(k)}$, $x^k \approx \gamma v_1$. We can approximate λ_1 by,

$$\lambda_1^{(k)} = \frac{(x^{(k+1)}, x^k)}{(x^{(k)}, x^k)} \qquad \text{or} \qquad \lambda_1^{(k)} = \frac{x_i^{(k+1)}}{x_i^{(k)}}$$

wher the index i is chosen to be corresponding to the largest element of $x^{(k)}$ in absolute sense.

Question b.iii. *Assume $|\lambda_1| \neq 1$. Depending on whether it is bigger or less than one, what will eventually happen if we perform the iteration on a computer? And what is done to prevent this situation if we are only interested in finding $lambda_1$ and the associated eigenvector?*

When $|\lambda| < 1$ then $x^{(k)} \to 0$ and at some point it will be rounded to 0 due to floating point arithmetic.
When $|\lambda| > 1$ then $x^{(k)} \to \infty$ and it will become too large to fit in the floating point representation used by Matlab.
We may prevent this by scaling x^k in each iteration.

$$y^{(k+1)} = Ax^{(k)}, \qquad x^{(k+1)} = \frac{y^{(k+1)}}{||y^{(k+1)}||}$$

Exercise 3

Consider the nonlinear system $\mathbf{f}(\mathbf{x}) = 0$, where $\mathbf{f}$ is a mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$.

Question a. *Derive Newton's method for the above system and indicate which linear system has to be solved in each step.*

We may use the taylor series of $\mathbf{f}$ to derive Newton's method,

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{x}_0) + D\mathbf{f}(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + \text{h.o.t.}$$

where $D\mathbf{f}$ is the Jacobian matrix of f. If $\mathbf{f}(+x) = \mathbf{0}$ then, (ignoring higher order terms) we get,

$$0 = \mathbf{f}(\mathbf{x}_0) + D\mathbf{f}(\mathbf{x}_0)\underbrace{(\mathbf{x} - \mathbf{x}_0)}_{\Delta\mathbf{x}}$$

Here $\Delta\mathbf{x}$ is unknown and needs to found by solving $D\mathbf{f}(\mathbf{x}_0)\Delta\mathbf{x} = -\mathbf{f}(\mathbf{x}_0)$. Newton's method for systems is given by the following process,

$$\begin{aligned} \text{Solve for } \Delta\mathbf{x} \qquad & D\mathbf{f}(\mathbf{x}_k)\Delta\mathbf{x} = -\mathbf{f}(\mathbf{x}_k) \\ \text{Update} \qquad & \mathbf{x}_{k+1} = \mathbf{x}_k + \Delta\mathbf{x} \end{aligned}$$

Question b. *Suppose* $\mathbf{f}_1 = sin(\mathbf{x}_1 + 2\mathbf{x}_2 - 1)$, $\mathbf{f}_2 = arctan(\mathbf{x}_2 - \mathbf{x}_1)$. *Give the Jacobian matrix of* $\mathbf{f}$.

$$J = \begin{bmatrix} \frac{\partial \mathbf{f}_1}{\partial \mathbf{x}_1} & \frac{\partial \mathbf{f}_1}{\partial \mathbf{x}_2} \\ \frac{\partial \mathbf{f}_2}{\partial \mathbf{x}_1} & \frac{\partial \mathbf{f}_2}{\partial \mathbf{x}_2} \end{bmatrix} = \begin{bmatrix} \cos(\mathbf{x}_1 + 2\mathbf{x}_2 - 1) & 2\cos(\mathbf{x}_1 + 2\mathbf{x}_2 - 1) \\ \frac{-1}{1+(\mathbf{x}_2-\mathbf{x}_1)^2} & \frac{1}{1+(\mathbf{x}_2-\mathbf{x}_1)^2} \end{bmatrix}$$

Question c. *Zeros of functions can be found by a fixed point method* $x^{(k+1)} = \phi(x^{(k)})$. *Show that this fixed point method will converge if* $|\phi'(\alpha) < 1$ *and* $x^{(0)}$ *close enough to the fixed point* α.

For a fixed point we have $\alpha = \phi(\alpha)$ and $x^{(k+1)} = \phi(x^{(k)})$. Using the mean value theorem we may find,

$$x^{(k+1)} - \alpha = \phi(x^{(k)}) - \phi(\alpha) = \phi'(\xi^{(k)})(x^{(k)} - \alpha)$$

where $\xi^{(k)} \in (x^k, \alpha)$. If $x^{(0)}$ is chosen such that $|\phi(\xi^{(k)})| < 1$ for all k, then it follows that,

$$\begin{aligned} |x^{(k+1)} - \alpha| &\leq |\phi'(\xi^{(k)})||(x^{(k)} - \alpha)| \\ &\leq |\phi'(\xi^{(k)})||\phi'(\xi^{(k-1)})||(x^{(k-1)} - \alpha)| \\ &\leq L^k|x^{(0)} - \alpha| \end{aligned}$$

where $L = \max_{i=0,\dots,k}\{|\phi'(\xi^{(k)})|\} < 1$. It follows that $x^{(k)}$ converges to α.

Question d. *Derive Aitken's extrapolation formula,*

$$\tilde{x}^{(k+1)} = \frac{x^{(k+1)}x^{(k-1)} - (x^{(k)})^2}{x^{(k+1)} - 2x^{(k)} + x^{(k-1)}}$$

where $\tilde{x}^{(k+1)}$ *is the extrapolated value based on* $x^{(k-1)}$, $x^{(k)} = \phi(x^{(k-1)})$, *and* $x^{(k+1)} = \phi(x^{(k)}) = \phi(\phi(x^{(k-1)}))$.

Recall $x_{k+1} = \phi(x_k)$ and

$$x^{(k+1)} - \alpha = \phi(x^{(k)}) - \phi(\alpha) = \phi'(\xi^{(k)})(x^{(k)} - \alpha)$$

where $\xi^{(k)} \in (x^k, \alpha)$. Rewriting gives,

$$\alpha\left(1 - \phi(x^{(k)})\right) = x_{k+1} - \phi(x^{(k)})x_k \Rightarrow \alpha = \frac{x_{k+1} - \phi(x^{(k)})x_k}{1 - \phi(x^{(k)})}$$

We now want to find a "nice" approximation of $\phi'(\xi^{(k)})$ using the forward finite difference method.

$$\phi'(\xi^k) \approx \frac{\phi(x_k) - \phi(x_{k-1})}{x_k - x_{k-1}} = \frac{x_{k+1} - x_k}{x_k - x_{k-1}} = \frac{\Delta x_{k+1}}{\Delta x_k}$$

we may now use this approximation to derive Aitken's extrapolation formula,

$$\begin{aligned}
\tilde{\alpha} = &= \frac{x_{k+1} - \frac{\Delta x_{k+1}}{\Delta x_k} x_k}{1 - \frac{\Delta x_{k+1}}{\Delta x_k}} \\
&= \frac{\Delta x_k x_{k+1} - \Delta x_{k+1} x_k}{\Delta x_k - \Delta x_{k+1}} \\
&= \frac{(x_k - x_{k-1})x_{k+1} - (x_{k+1} - x_k)x_k}{(x_k - x_{k-1}) - (x_{k+1} - x_k)} \\
&= \frac{x_{k-1}x_{k+1} - x_k^2}{x_{k+1} - 2x_k + x_{k-1}}
\end{aligned}$$

Exercise 4

Consider a system of ODEs,

$$\frac{d}{dt}y(t) = f(t, y(t)), \text{with } y(0) = y_0 \qquad (2)$$

Question a. *Consider the method* $u_{k+1} = u_{k-1} + 2\Delta t f(t_k, u_k)$

Question a.i. *State the root condition. Show that this method satisfies this condition. What does this mean for stability?*

If we denote with r_j *the roots of the characteristic polynomial,*

$$\pi(r) = r^{p+1} - \sum_{j=0}^{p} a_j r^{p-j}$$

then the numerical method satisfies the root condition if $|r_j| \leq 1$ *and if* $|r_j| = 1$ *then we must have* $\pi'(r_j) \neq 1$.

In this case we have $\pi(r) = r^{1+1} - 1r^{1-1} = r^2 - 1 = 0 \Leftrightarrow r = \pm 1$ *and* $\pi'(r) = 2r$. *Hence this method satisfies the root condition which implies that the method is zero-stable.*

Question a.ii. *Show that the local truncation error is of second order in* Δt. *What is the conclusion for convergence, if you combine this with part (i).*

The local truncation error is given by,

$$\tau_{n+1}(\Delta t) = \frac{y_{n+1} - y_{n-1} - 2\Delta t f(t_n, y_n)}{\Delta t}$$

where y_n *is an exact solution to the ODE. We may approximate* y_{n+1} *and* $y - n - 1$ *using a Taylor series at* y_n,

$$y_{n+1} = y_n + \Delta t y'(t_n) + \frac{\Delta t^2}{2} y''(t_n) + \frac{\Delta t^3}{6} y'''(\xi)$$
$$y_{n-1} = y_n - \Delta t y'(t_n) + \frac{\Delta t^2}{2} y''(t_n) - \frac{\Delta t^3}{6} y'''(\eta)$$

subtracting both approximations gives,

$$y_{n+1} - y_{n-1} = 2\Delta t y'(t_n) + \frac{\Delta t^3}{6}(y'''(\xi) - y'''(\eta))$$

It follows that the local truncation error is given by,

$$\tau_{n+1}(\Delta t) = \frac{\Delta t^2}{6}(y'''(\xi) - y'''(\eta))$$

Which is of second order. Hence the method is consistent and as we've shown that it is also zero stable, it is convergent.

Question b. *Consider on* $[0, 1]$ *for* $u(x, t)$ *the diffusion equation* $\partial u/\partial t = \partial^2 u/\partial x^2 + x exp(-t)$ *with the initial condition* $u(x, 0) = sin(\pi x)$ *and boundary conditions* $u(0, t) = sin^2(t)$ *and* $u(1, t) = 0$. *Let the grid in x-direction be given by* $x_i = i\Delta x$ *where* $\Delta x = 1/m$. *Show that, by using* $\frac{\partial^2 u}{\partial x^2} \approx \frac{u(x_{i+1},t) - 2u(x_i,t) + u(x_{i+1},t)}{\Delta x^2}$ *in the PDE, one obtains a system of ordinary differential equations (ODEs) of the above form. Give the components of the vector function* f *and the initial vector.*

After discretization we have,

$$\frac{du_i}{dt} = \frac{u_{i+1} - 2u_i + u_{i-1}}{\Delta x^2} + x_i e^{-t}$$

for $i = 1, \ldots, m-1$ and for $i = 0$ and $i = m$ we have the boundary conditions,

$$u_0(t) = \sin^2(t),$$
$$u_m(t) = 0.$$

the initial condition for $i = 1, \ldots, m-1$ is given by,

$$u_i(0) = \sin(\pi x_i)$$

lastly the right hand side of (2) is given by,

$$f_1(t) = \frac{-2u_1 + u_2}{\Delta x^2} + x_1 e^{-t} + \frac{\sin^2(t)}{\Delta x^2}$$

$$f_i(t) = \frac{u_{i-1} - 2u_i + u_{i+1}}{\Delta x^2} + x_i e^{-t} \qquad\qquad i = 2, \ldots, m-2$$

$$f_{m-1}(t) = \frac{u_{m-2} - 2u_{m-1}}{\Delta x^2} + x_{m-1} e^{-t}$$