Question 1. [3 MARKS]

Show that if Q is an orthogonal matrix then $Q^{-1} = Q^T$.

Solution:
List out the columns of Q as the vectors $q_1, q_2, \ldots, q_n$.
Then, the rows of Q^T are given by $q_1^T, q_2^T, \ldots, q_n^T$.
Then, compute $A = Q^T Q$. The entries of A_{ij} are given by $q_i^T q_j$.
Since the columns are all vectors that are orthogonal to each other and are unit vectors, $q_i^T q_j = 1$ iff $i = j$.
Hence, A is the identity matrix and $Q^{-1} = Q^T$.

Question 2. [13 MARKS]

Let h be a small number and let x be around 1. Consider the following function:

$$f(x) = \frac{\sqrt{x+h} - \sqrt{x-h}}{2h} \tag{1}$$

Part (a) [3 MARKS]

Explain, in plain language, what f computes.

Solution:
f computes a second-order ($\mathcal{O}(h^2)$) approximation to the first derivative of the function $y = \sqrt{x}$.

Part (b) [2 MARKS]

Clearly describe two issues when computing f in floating-point.

Solution:
The two issues from computing f in floating-point are:

- Cancellation error from computing $\sqrt{x+h} - \sqrt{x-h}$
- Magnification of error by dividing by a small number $2h$.

Part (c) [5 MARKS]

Show how to resolve both, by computing a different, but mathematically equivalent expression.

Solution:
We use the conjugation trick to eliminate the cancellation in the numerator.

$$\begin{aligned} f(x) &= \frac{\sqrt{x+h} - \sqrt{x-h}}{2h} \\ &= \left(\frac{\sqrt{x+h} - \sqrt{x-h}}{2h}\right)\left(\frac{\sqrt{x+h} + \sqrt{x-h}}{\sqrt{x+h} + \sqrt{x-h}}\right) \\ &= \frac{(x+h) - (x-h)}{2h(\sqrt{x+h} + \sqrt{x-h})} \\ &= \frac{2h}{2h(\sqrt{x+h} + \sqrt{x-h})} \\ &= \frac{1}{\sqrt{x+h} + \sqrt{x-h}} \end{aligned}$$

Part (d) [3 MARKS]

What is the limit of the second formula as $h \to 0$? How are the limit and f related?

Solution:
To compute the limit, sub 0 in for h:

$$\lim_{h\to 0} \frac{1}{\sqrt{x+h}+\sqrt{x-h}} = \frac{1}{2\sqrt{x}} \tag{2}$$

The limit is the first derivative of $\sqrt{x}$. So it is an exact expression for what f was intended to calculate.

Question 3. [11 MARKS]

Consider solving the equation

$$f(x) = \exp(-x) - x = 0 \tag{3}$$

A bracket on the unique root x^* would be $[1/2, 2]$.

Part (a) [3 MARKS]

How many iterations does it take for bisection method to achieve an absolute accuracy of 10^{-6} from the bracket?

Solution:
The size of the initial bracket is $3/2$.
Each iteration of bisection method reduces the interval to a half of the size on the previous iteration.
Hence, after k iterations, the size of the bracket is given by

$$\text{sz} = \frac{3}{2}\frac{1}{2^k}. \tag{4}$$

If we want the size of the bracket to be reduced to 10^{-6}, then we have

$$10^{-6} = \frac{3}{2^{k+1}} \tag{5}$$

or

$$k = \lceil \log_2(3 \times 10^6) - 1 \rceil = 21. \tag{6}$$

Part (b) [5 MARKS]

Consider the fixed-point iteration

$$g(x) = \exp(-x) \tag{7}$$

Show that $|g'(x)| < 1$ on $[1/2, 2]$. What is ρ?

Solution:
$|g'(x)| = |-\exp(-x)| = \exp(-x)$. Since this is a decreasing function, the maximum occurs at $g'(1/2) = 0.6065$, which is less than 1.

Part (c) [3 MARKS]

Write out Newton's method for $f(x)$ and express it as a fixed-point iteration g_2.

Solution:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} = x_k - \frac{\exp(-x_k) - x_k}{-\exp(-x_k) - 1} = x_k + \frac{\exp(-x_k) - x_k}{\exp(-x_k) + 1} \equiv g_2(x). \tag{8}$$

Question 4. [12 MARKS]

Consider the matrix A:

$$A = \begin{bmatrix} 10 & 7 \\ 7 & 5 \end{bmatrix} \tag{9}$$

You may use without proof the fact that

$$\lambda_1(A), \lambda_2(A) = \frac{15 \pm \sqrt{221}}{2} \tag{10}$$

and the fact that A can be orthogonally diagonalized:

$$A = Q^T D Q \tag{11}$$

where Q is an orthogonal matrix and

$$D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \tag{12}$$

Part (a) [4 MARKS]

Show that A is symmetric positive definite.

Solution:
Since A can be orthogonally diagonalized, we can write $x^T A x = x^T Q^T D Q x$.
For any nonzero x, $y = Qx$ is also nonzero, and hence $y^T D y = \lambda_1 y_1^2 + \lambda_2 y_2^2$ is a positive combination of positive numbers.

For zero x, $x^T A x$ is obviously zero. Therefore A is SPD.

Part (b) [2 MARKS]

What is $\|A\|$? What is $\|A^{-1}\|$?

Solution:
The maximum stretching factor occurs for a vector x satisfying $Qx = [1, 0]^T$, and the maximum stretching factor of the inverse occurs for the vector x satisfying $Qx = [0, 1]^T$. Hence, $\|A\| = \lambda_1$ and $\|A^{-1}\| = \lambda_2$

Part (c) [2 MARKS]

What is the condition number of A?

Solution:
Use the formula given in class and the previous calculations:

$$\kappa(A) = \|A\|\|A^{-1}\| = \frac{\lambda_1}{\lambda_2} = \frac{15 + \sqrt{221}}{15 - \sqrt{221}} \approx 2.23 \times 10^2 \tag{13}$$

Part (d) [4 MARKS]

Describe how you would use the Cholesky factorization to solve the linear system $Ax = b$ for an arbitrary b.

Solution:
Knowing the Cholesky decomposition allows us to write the linear system as $RR^T x = b$ for a lower triangular matrix R.
Then, we first solve the lower triangular system $Ry = b$ with forward substitution.
Next, we solve the upper triangular system $R^T x = y$ with backward substitution.

Question 5. [3 MARKS]

Consider the following system of overdetermined equations:

$$\begin{bmatrix} 10 & 2 & 14 \\ 7 & 6 & 19 \\ 5 & 8 & 21 \\ 1 & 3 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \tag{14}$$

Describe how you would solve this system of linear equations and justify why.

Solution:
Check the columns for linear dependence:

$$\begin{bmatrix} 10 \\ 7 \\ 5 \\ 1 \end{bmatrix} + 2 \begin{bmatrix} 2 \\ 6 \\ 8 \\ 3 \end{bmatrix} = \begin{bmatrix} 14 \\ 19 \\ 21 \\ 7 \end{bmatrix}. \tag{15}$$

Hence, the matrix is rank-deficient, and the most appropriate method is to use truncated SVD, where we calculate the SVD of the matrix and drop the singular values that are zero or close to zero.