Numerical Analysis MATH50003 (2023–24) Revision Sheet

Problem 1(a) State which real number is represented by an IEEE 16-bit floating point number (with $\sigma = 15, Q = 5$, and $S = 10$) with bits

$$\texttt{1 01000 0000000001}$$

Problem 1(b) How are the following real numbers rounded to the nearest F_{16}?

$$1/2, 1/2 + 2^{-12}, 3 + 2^{-9} + 2^{-10}, 3 + 2^{-10} + 2^{-11}.$$

Problem 2(a) Consider a Lower triangular matrix with floating point entries:

$$L = \begin{bmatrix} \ell_{11} \\ \ell_{21} & \ell_{22} \\ \vdots & \ddots & \ddots \\ \ell_{n1} & \cdots & \ell_{n,n-1} & \ell_{nn} \end{bmatrix} \in F_{\sigma,Q,S}^{n\times n}$$

and a vector $\boldsymbol{x} \in F_{\sigma,Q,S}^{n}$, where $F_{\sigma,Q,S}$ is a set of floating-point numbers. Denoting matrix-vector multiplication implemented using floating point arithmetic as

$$\boldsymbol{b} := \texttt{lowermul}(L, \boldsymbol{x})$$

express the entries $b_k := \mathbf{e}_k^\top \boldsymbol{b}$ in terms of ℓ_{kj} and $x_k := \mathbf{e}_k^\top \boldsymbol{x}$, using rounded floating-point operations $\oplus$ and $\otimes$.

Problem 2(b) Assuming all operations involve normal floating numbers, show that your approximation has the form

$$L\boldsymbol{x} = \texttt{lowermul}(L, \boldsymbol{x}) + \boldsymbol{\epsilon}$$

where, for $\epsilon_{\rm m}$ denoting machine epsilon and $E_{n,\epsilon} := \frac{n\epsilon}{1-n\epsilon}$ and assuming $n\epsilon_{\rm m} < 2$,

$$\|\boldsymbol{\epsilon}\|_1 \leq 2E_{n,\epsilon_{\rm m}/2}\|L\|_1\|\boldsymbol{x}\|_1.$$

Here we use the matrix norm $\|A\|_1 := \max_j \sum_{k=1}^n |a_{kj}|$ and the vector norm $\|\boldsymbol{x}\|_1 := \sum_{k=1}^n |x_k|$. You may use the fact that

$$x_1 \oplus \cdots \oplus x_n = x_1 + \cdots + x_n + \sigma_n$$

where

$$|\sigma_n| \leq \|\boldsymbol{x}\|_1 E_{n-1,\epsilon_{\rm m}/2}.$$

Problem 3 What is the dual extension of square-roots? I.e. what should $\sqrt{a + b\epsilon}$ equal assuming $a > 0$?

Problem 4 Use the Cholesky factorisation to determine whether the following matrix is symmetric positive definite:

$$\begin{bmatrix} 2 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 2 & 2 \end{bmatrix}$$

Problem 5 Use reflections to determine the entries of an orthogonal matrix Q such that

$$Q \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -3 \\ 0 \\ 0 \end{bmatrix}.$$

Problem 6 For the function $f(\theta) = \sin 3\theta$, state explicit formulae for its Fourier coefficients

$$\hat{f}_k := \frac{1}{2\pi} \int_0^{2\pi} f(\theta) \mathrm{e}^{-\mathrm{i}k\theta} \mathrm{d}\theta$$

and their discrete approximation:

$$\hat{f}_k^n := \frac{1}{n} \sum_{j=0}^{n-1} f(\theta_j) \mathrm{e}^{-\mathrm{i}k\theta_j}.$$

for *all* integers k, $n = 1, 2, \ldots$, where $\theta_j = 2\pi j/n$.

Problem 7 Consider orthogonal polynomials

$$H_n(x) = 2^n x^n + O(x^{n-1})$$

as $x \to \infty$ and $n = 0, 1, 2, \ldots$, orthogonal with respect to the inner product

$$\langle f, g \rangle = \int_{-\infty}^{\infty} f(x) g(x) w(x) \mathrm{d}x, \qquad w(x) = \exp(-x^2)$$

Construct $H_0(x)$, $H_1(x)$, $H_2(x)$ and hence show that $H_3(x) = 8x^3 - 12x$. You may use without proof the formulae

$$\int_{-\infty}^{\infty} w(x) \mathrm{d}x = \sqrt{\pi}, \int_{-\infty}^{\infty} x^2 w(x) \mathrm{d}x = \sqrt{\pi}/2, \int_{-\infty}^{\infty} x^4 w(x) \mathrm{d}x = 3\sqrt{\pi}/4.$$

Problem 8(a) Derive the 3-point Gauss quadrature formula

$$\int_{-\infty}^{\infty} f(x) \exp(-x^2) \mathrm{d}x \approx w_1 f(x_1) + w_2 f(x_2) + w_3 f(x_3)$$

with analytic expressions for x_j and w_j.

Problem 8(b) Compute the 2-point and 3-point Gaussian quadrature rules associated with $w(x) = 1$ on $[-1, 1]$.

Problem 9 Solve Problem 4(b) from PS8 using **Lemma 12 (discrete orthogonality)** with $w(x) = 1/\sqrt{1 - x^2}$ on $[-1, 1]$. That is, use the connection of $T_n(x)$ with $\cos n\theta$ to show that the Discrete Cosine Transform

$$C_n := \begin{bmatrix} \sqrt{1/n} & & & \\ & \sqrt{2/n} & & \\ & & \ddots & \\ & & & \sqrt{2/n} \end{bmatrix} \begin{bmatrix} 1 & \cdots & 1 \\ \cos \theta_1 & \cdots & \cos \theta_n \\ \vdots & \ddots & \vdots \\ \cos(n-1)\theta_1 & \cdots & \cos(n-1)\theta_n \end{bmatrix}$$

for $\theta_j = \pi(j - 1/2)/n$ is an orthogonal matrix.