Homework 1 Solutions

by WeiHao Li

Fudan University

(Numerical Algorithms with Case Studies I)

Problem 1.

Solution. Simply denote $E_{\text{rel}}(\hat{x})$ as E and $\tilde{E}_{\text{rel}}(\hat{x})$ as $\tilde{E}$, put

$$\frac{E}{\tilde{E}} = \frac{|\hat{x}|}{|x|} \leqslant \frac{|x| + |x - \hat{x}|}{|x|} = 1 + E,$$

then we have

$$E \leqslant \frac{\tilde{E}}{1 - \tilde{E}}.$$

By the same process, we have

$$\frac{\tilde{E}}{E} = \frac{|x|}{|\hat{x}|} \leqslant \frac{|\hat{x}| + |x - \hat{x}|}{|\hat{x}|} = 1 + \tilde{E},$$

then

$$E \geqslant \frac{\tilde{E}}{1 + \tilde{E}}.$$

Combine results above, we have

$$\frac{\tilde{E}_{\text{rel}}(\hat{x})}{1 + \tilde{E}_{\text{rel}}(\hat{x})} \leqslant E_{\text{rel}}(\hat{x}) \leqslant \frac{\tilde{E}_{\text{rel}}(\hat{x})}{1 - \tilde{E}_{\text{rel}}(\hat{x})}$$

Problem 2.

Solution. Put

$$\begin{aligned} f(x) &= \tan x - \sin x \\ &= \tan x (1 - \cos x) \\ &= 2 \tan x \sin \frac{x}{2}. \end{aligned}$$

Use $2 \tan x \sin \frac{x}{2}$ to evaluate $f(x)$ can avoid numerical cancellation.

Problem 3.

Solution. Let $A = \{a_{ij}\}$ and $E = \{e_{ij}\}$, and ε as the machine epsilon. For the ith element of vector Ax,

$$fl\left(\sum_{k=1}^{n} a_{ik} x_k \right) = (1+\varepsilon)^n (a_{i1} x_1 + a_{i2} x_2) + (1+\varepsilon)^{n-1} (a_{i3} x_3) + \cdots + (1+\varepsilon)^2 (a_{in} x_n),$$

then

$$\left| fl\left(\sum_{k=1}^{n} a_{ik} x_k \right) - \sum_{k=1}^{n} a_{ik} x_k \right| \leqslant [(1+\varepsilon)^n - 1] \sum_{k=1}^{n} a_{ik} x_k.$$

Let $\delta = (1+\varepsilon)^n - 1$.Then

$$|Ex| = |fl(Ax) - Ax| \leqslant \delta |Ax|$$

and

$$||Ex|| \leqslant \delta ||Ax||$$

for any $x \in \mathbb{R}^n$.

Because $\|E\|$ is the maximum of $||Ex||$ when $\|x\| = 1$ and $\|A\|$ is the maximum of $||Ax||$ when $\|x\| = 1$, so we have

$$\|E\| \leqslant \delta \|A\|$$

for any kind of matrix norm.

Problem 4.

Solution. See soltri.py for implementation.

Problem 5.

Solution. See solinear.py for implementation and visualization. The log-log plot visualize the execution time of program in terms of matrix size is shown below.