Homework I

Due on Jan 23, 2024

Use single precision (24 bits) unless specified otherwise. Whenever you are asked to explain some results, you should explain the observation in a quantitative manner. For example, if a curve has a bend at say $n = 1000$, you need to explain why you would expect the bend to happen roughly around this value of n.

1. **(10 marks)** Consider the function $f : \Re^2 \to \Re$ defined by $f(x, y) = x \cdot y$. Measure the size of the input (x, y) by $|x| + |y|$. What is the condition number of this function? When is the condition number very high ? Can you intuitively explain why it is high ?

2. In class, we defined the notion of a backward stable algorithm as follows: an algorithm for computing a function $f(x)$ computes $f(\widehat{x})$. Then the algorithm is backward stable if $\frac{|\widehat{x}-x|}{|x|}$ is $O(\varepsilon_{\text{mach}})$. Sometimes, it is more convenient to work with a slightly more general defintion. We say that the algorithm for computing $f(x)$ is stable if it computes a quantity y which is close to $f(\widehat{x})$ for some vector $\widehat{x}$ in the following sense:

$$\frac{|y - f(\widehat{x})|}{|f(\widehat{x})|} = O(\varepsilon_{\text{mach}}), \text{ where } \quad \frac{|\widehat{x} - x|}{|x|} = O(\varepsilon_{\text{mach}}).$$

   (i) **(5 marks)** Show that if an algorithm for computing f is stable, then the relative error in the output is roughly $O(k_f(x) \cdot \varepsilon_{\text{mach}})$, where $k_f(x)$ is the condition number of f at x. Hence, we can use either the notion of stability or backward stability.

   (ii) **(5 marks)** Consider the function $f(x) = 1 + x$. Show that if we compute this function using this expression, then the algorithm may not be backward stable, but is always stable.

   (iii) **(5 marks)** Suppose we compute $f(x) = \sin(x)$ for $x \in [0, 2\pi]$ using an algorithm that outputs a value in the range $\sin(x)(1 \pm \varepsilon_{\text{mach}})$. Show that such an algorithm may not be backward stable, but is always stable.

3. Consider the function $f(x) = \frac{e^x - 1}{x}$.

   (i) **(2 marks)** Show that the function is well-conditioned for x close to 0.

   (ii) **(5 marks)** Suppose we compute $f(x)$ as in the expression above. Assume that the computation of e^x has relative error at most $\varepsilon_{\text{mach}}$. Prove that this computation is not backward stable or stable (i.e., you need to show that the relative error can be large and since the problem is well-conditioned, it must be the case that the algorithm is unstable).

   (iii) **(3 marks)** Suppose we compute $f(x)$ as $\frac{e^x - 1}{\ln(e^x)}$. Assume that the computation of the exponential and the logarithm function have relative error at most $\varepsilon_{\text{mach}}$. Prove that this algorithm is backward stable or stable.

4. Consider the function
$$f(x) = \frac{1}{1-x} - \frac{1}{1+x},$$
assuming $x \neq \pm 1$.

   (a) **(2 marks)** When is the above function ill-conditioned?

   (b) **(3 marks)** Suppose we compute this function using the expression above. Show that the algorithm is unstable for values close to 0.

   (c) **(5 marks)** Suppose we compute the function as $f(x) = \frac{2x}{1-x^2}$. Show that this computation is stable or backward stable for all x.

5. The *fibonacci* numbers f_k are defined by $f_0 = 1, f_1 = 1$, and
$$f_{k+1} = f_k + f_{k-1} \tag{1}$$
for any integer $k > 1$. A small perturbation of them, the *pib* numbers, p_k, are defined by $p_0 = 1, p_1 = 1$ and
$$p_{k+1} = c \cdot p_k + p_{k-1} \tag{2}$$
for any integer $k > 1$ where $c = 1 + \frac{\sqrt{3}}{100}$.

   (a) **(5 marks)** Plot the numbers f_n and p_n together in one log scale plot. On the plot, mark $1/\epsilon_{mach}$ for single and double precision arithmetic.

   (b) **(4 marks)** Rewrite (1) to express f_{k-1} in terms of f_k and f_{k+1}. Use the computed f_n and f_{n-1} to recompute f_k for $k = n-2, n-3, \ldots, 0$. Make a plot of the difference between the original $f_0 = 1$ and the recomputed f_0 as a function of n. What n values result in low accuracy for the recomputed f_0 ? How do the results in single and double precision differ ?

   (c) **(6 marks)** Repeat part (b) for the pib numbers. Comment on the striking difference in the way precision is lost in the two cases. Explain the results.

6. Write a program to generate the first n terms in the sequence given by the difference equation:
$$x_{k+1} = 111 - (1130 - 3000/x_{k-1})/x_k,$$
with starting values $x_1 = 11/2, x_2 = 61/11$ **(2 marks)**. Use $n = 10$ for single precision.

   (i) **(2 marks)** The exact solution is monotonically converging to 6, but what do you observe?

   (ii) In order to explain the results, consider the function $F : \Re^2 \to \Re^2$ as
$$F\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 111 - (1130 - 3000/y)/x \\ x \end{bmatrix}.$$
The recurrence above can be expressed as $\begin{bmatrix} x_{k+1} \\ x_k \end{bmatrix} = F\left(\begin{bmatrix} x_k \\ x_{k-1} \end{bmatrix}\right)$. A vector z is a fixed point of F if $F(z) = z$. What are the possible fixed points of this function? **(3 marks)**

(iii) **(5 marks)** To understand the behaviour of the above recurrence, give 2D plots of how the vector $\begin{bmatrix} x_k \\ x_{k-1} \end{bmatrix}$ evolves as we start from points close to $\begin{bmatrix} 5 \\ 5 \end{bmatrix}$. Use high precision here so that rounding errors do not affect your conclusion.