1. Answer the following questions about  $\sin(x)$ . Recall that  $\lim_{x\to 0} \frac{\sin(x)}{x} = 1$  and  $\lim_{x\to 2\pi} \frac{\sin(x)}{x} = 0$ 

a) Compute the condition number of sin(x).

b) For which values of x the function sin(x) is well-conditioned?

c) For which values of x the function  $\sin(x)$  is ill-conditioned?

a) 
$$f(x) = \sin(x)$$
  
 $f'(x) = \cos(x)$   
 $condition number = \left| \frac{xf(x)}{f(x)} \right|$   
 $= \left| \frac{x \cos(x)}{\sin(x)} \right|$   
 $= \left| \frac{x \cos(x)}{\sin(x)} \right|$ 

b) 
$$\lim_{x\to 0} \frac{\sin x}{x} = 1$$
 $\lim_{x\to 0} |x \cot(x)| = 1$ 
 $\lim_{x\to 0} |x \cot(x)| = 1$ 

when  $x = 0$ ,  $\sin(x)$  is well condition

c) i 
$$\lim_{\chi \to 2\pi} \frac{\sin x}{x} = 0$$
 $\lim_{\chi \to 3\pi} ||\chi \cot(x)|| = \infty$ 

i when  $\chi = 2\pi$ ,  $\sin(x)$  is ill condition.

2. List the numerical problems one could have by solving the quadratic formula  $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$  in a floating-point system.

Overflow, underflow or catastrophic cancellation

Overflow, underflow or catastrophic cancellation

if the coefficients are very large or very small

when be or 4ac may overflow or underflow

when be or 4ac may overflow or underflow

## 3. (Heath, 2018)

- a) Using a four-digit decimal arithmetic compute the formula of the area of the planet using the formula  $A=4\pi r^2$  for r=6370 km.
- b) Using the same formula and precision, compute the difference in surface area if the value for the radius is increased by  $1~\rm km$ .
- c) Since  $dA/dr = 8\pi r$ , the change in surface area is approximated by  $8\pi rh$ , where h is the change in radius. Use this formula, still with four-digit arithmetic, to compute the difference in surface area due to an increase of 1 km in radius. How does the value obtained using this approximate formula compare with that obtained from the "exact" formula in part b?
- d) Determine which of the previous two answers is more nearly correct by repeating both computations using heigher precision, say, six-digit decimal arithmetic.
- e) Explain the results you obtained in parts a-d.

a) 
$$A = 4\pi r^{2} \quad r_{s} = 6370$$
 $A_{s} = 4\pi (6370)^{2}$ 
 $= 162307600\pi$ 
 $\approx 5.0970 \times 10^{6}$ 
 $A_{s} = (5 + \frac{0}{10} + \frac{9}{10^{2}} + \frac{9}{10^{4}}) \times 10^{8}$ 
 $A_{s} = (6371)^{2}$ 
 $= 162358564\pi$ 
 $\approx 5.1006 \times 10^{6}$ 
 $= 1.6000 \times 10^{5}$ 
 $= 1.6000 \times 10^{5}$ 
 $= (1 + \frac{1}{10} + \frac{1}{10^{2}} + \frac{1}{10^{4}}) \times 10^{5}$ 
 $= (1 + \frac{1}{10} + \frac{1}{10^{2}} + \frac{1}{10^{4}}) \times 10^{5}$ 
 $= (1 + \frac{1}{10} + \frac{1}{10^{2}} + \frac{1}{10^{4}}) \times 10^{5}$ 

The value one get using the form that is a little bit larger than given b.

d) sturh = 8 % 6370 ·1 = 50960 W ≈ 1.600955 × 605  $= ((+\frac{6}{10} + \frac{0}{10^2} + \frac{0}{10^3} + \frac{9}{10^4} + \frac{5}{10^4} + \frac{5}{10^6}) \times 10^5$ 4かり2-4かり、2 426371)2-420(6370)2 = 50964K ~1.601081 x 105  $= (|t| \frac{b}{10} + \frac{0}{10^2} + \frac{1}{10^3} + \frac{0}{10^4} + \frac{8}{10^5} + \frac{1}{10^5}) \times |0|^5$ Company two value, the way we very from part c is more correctly. e) for the floating point

system, if the ne

teep more ligits, it will move

correctly, if using the approximated formula, it will more

, - -

4. The 2-norm requires one to square all values of a given vector before taking the square root of their sum.

$$||x||_2 = \left(\sum_{i=1}^n |x_i|^2\right)^{1/2}$$

A problem with this norm is that the squared value of an entry  $x_i$  can overflow. Derive a procedure that prevents overflow from happening. As a hint, think of how you can modify the values  $x_i$  while not changing the value of the norm.

anging the value of the norm.

$$||x||_{2} = \left(\sum_{i=1}^{n} |x_{i}|^{2}\right)^{\frac{1}{2}}$$

$$= \sqrt{\sum_{i=1}^{n} |x_{i}|^{2}}$$

$$= \sqrt{\sum_{i=1}^{n} |x_{i}|^{2}}$$

$$= \sqrt{\sum_{i=1}^{n} |x_{i}|^{2}}$$

$$= \sqrt{\sqrt{\sum_{i=1}^{n} |x_{i}|^{2}}}$$

5. (Heath, 2018) Write the LU factorization for the following matrix. Show both L and U explicitly.

$$U := \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2} U := \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{P_1 + P_2}$$