Indian Institute of Science, Bangalore
Department of Computational and Data Sciences (CDS)

DS284: Numerical Linear Algebra

Quiz 2 [Sept 23, 2023]

Faculty Instructor: Dr. Phani Motamarri

TAs: Kartick Ramakrishnan, Sayan Dutta, Sundaresan G, Raj Maddipati

Exam duration: 3:05 PM - 4:25 AM | Max Points: 30

Notations: (i) Vectors and matrices are denoted by bold faced lower case and upper case alphabets respectively. (ii) Set of all real numbers is denoted by $\mathbb{R}$ (iii) Set of all n dimensional vectors is denoted by $\mathbb{R}^n$ and set of all $m \times n$ matrices is denoted by $\mathbb{R}^{m\times n}$. (iv) $\mathbf{I}_n$ denotes the identity matrix of order n. (v) $\mathbf{0}_n$ denotes the null matrix of order $n \times n$.

1. Assert if the following statements are True or False. Give a short reasoning for your assertion. Marks will be awarded only for your reasoning. **6 × 3.5 = 21 points**

   (a) Condition number of a full rank matrix $\mathbf{A} \in \mathbb{R}^{m\times m}$, $\kappa(\mathbf{A})$ may be less than 1 for some vector induced norm.

   (b) Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ $(m < n)$ be a full rank matrix and $\mathbf{A}$ admits $\mathbf{U\Sigma V}^T$ as its full SVD. If $\mathbf{V} = [\mathbf{V}_1\mathbf{V}_2]$, where $\mathbf{V}_1 \in \mathbb{R}^{n\times m}$ and $\mathbf{V}_2 \in \mathbb{R}^{n\times(n-m)}$. Then for any $\boldsymbol{y} \in \mathbb{R}^n$,

   $$\boldsymbol{A}(\boldsymbol{I} - \boldsymbol{V}_1\boldsymbol{V}_1^T)\boldsymbol{y} = \mathbf{0}$$

   (c) Let $fl : \mathbb{R}^+ \to \mathbb{F}$ be a function that takes positive real numbers $x \in \mathbb{R}^+$ as input and returns the **nearest** floating point number approximation $x' \in \mathbb{F}$ (Assume that all inputs are representable as normalized floating point numbers and there is no risk of overflow/underflow in any of the operations). For $a, b, c \in \mathbb{R}^+$, assert individually if the following are true/false

   - $a + b = c \implies fl(fl(a) + fl(b)) = fl(c)$
   - $a \geq b \implies fl(a) \geq fl(b)$

   (d) The moon rover Vikram has collected a large number of high definition images of various facets of the moon surface. To send this to an observer on Earth, an image compression problem f is defined on these images. The problem $f : \mathbb{R}^{M\times N} \to \mathbb{R}^{m\times n}, M >> m, N > n$ has a relative condition number $1 \leq \kappa \leq 100$. Scientists at ISRO have come up with an innovative algorithm $\tilde{f}$ to solve the problem f. Assert if the below statements are true/false:

   - Since the problem f is well conditioned, we can conclude that the designed algorithm $\tilde{f}$ will always be stable.
   - If the designed algorithm $\tilde{f}$ is backward stable, then $||f(Image) - \tilde{f}(Image)||$ depends only on the precision used during the floating point arithmetic operations in executing $\tilde{f}$

   (e) Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be full rank and admits the SVD $\mathbf{A} = \mathbf{U\Sigma V^T}$ such that both $\mathbf{U}$ and $\mathbf{V}$ are upper-triangular matrices. Then $\mathbf{A}$ must be a diagonal matrix.

   (f) For input data $x \in \mathbb{R}$, the computation of $f(x) = x + 1$ on a computer is done by an algorithm $\tilde{f}(x) = fl(x) \oplus 1$ where $\oplus$ is floating point analogue of classical arithmetic addition operation. $\tilde{f}(x)$ is stable but not backward stable.

2. Suppose you are confronted with solving a linear system of equations $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A}$ is a very large full-rank square matrix. One would like to use an iterative solver to solve this system of equations, as the matrix size is large. An iterative solver starts with some initial guess $\mathbf{x}_0$ and tries to seek a sequence of approximations to $\mathbf{x}^*$ (the exact solution of $\mathbf{Ax} = \mathbf{b}$) in an iterative fashion so that the sequence of iterates $(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n, \ldots)$ converge to $\mathbf{x}^*$, within a given tolerance for a large value of n. One usually specifies the convergence criteria on the 2-norm of the relative residual $\varepsilon = \frac{||\mathbf{r}||_2}{||\mathbf{b}||_2} = \frac{||\mathbf{b}-\mathbf{Ax}||_2}{||\mathbf{b}||_2}$ and waits for $\mathbf{x}_i$ at i^{th} iteration to come close enough to $\mathbf{x}^*$ for achieving convergence.

   At a given iteration i, let the 2-norm of relative error in the solution be denoted as $\eta_i = \frac{||\mathbf{x}_i - \mathbf{x}^*||_2}{||\mathbf{x}^*||_2}$ and let the 2-norm of the relative residual be denoted as $\varepsilon_i = \frac{||\mathbf{r}||_2}{||\mathbf{b}||_2} = \frac{||\mathbf{b}-\mathbf{Ax}_i||_2}{||\mathbf{b}||_2}$
   Answer the following: **5+4 = 9 points**

   (a) Show that $\eta_i \leq \frac{||\mathbf{A}^{-1}||_2 ||\mathbf{b}||_2}{||\mathbf{A}^{-1}\mathbf{b}||_2}\varepsilon_i$.

   (b) Let $\mathbf{v}$ be the right singular vector of some full rank square matrix $\mathbf{M}$ corresponding to its largest singular value of 10. Also, $\mathbf{M}$ matrix has the smallest singular value to be 10^{-07}. Consider the linear system $\mathbf{Mx} = \mathbf{y}$, which has $\mathbf{v}$ to be the exact solution and assume that we have devised an iterative scheme to solve the above linear system such that the above inequality in (a) becomes equality at every iteration, i.e.

$$\eta_i = \frac{||\mathbf{M}^{-1}||_2 ||\mathbf{y}||_2}{||\mathbf{M}^{-1}\mathbf{y}||_2}\varepsilon_i \qquad \forall\ i \geq 1$$

   For such an iterative scheme devised above to solve $\mathbf{Mx} = \mathbf{y}$, let us assume we have reached the required tolerance in the 2-norm of the relative residual $\varepsilon_k = 10^{-12}$ for some iteration k. Can we ever achieve the 2-norm of the relative error in the solution to be $\eta_k = 10^{-6}$ for such an iteration k?