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

DS284: Numerical Linear Algebra

Final Exam – Aug 2022 Term

Faculty Instructor: Dr. Phani Motamarri

TAs: Ashish Rout, Dibya Nayak, Gourab Panigrahi, Nikhil Kodali

Duration: 9:00 hrs to 12:00 hrs | Max Points: 100

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) $\mathbf{I}_m$ denotes the identity matrix of order m. (v) ϵ_{mach} denotes machine epsilon. (vi) $\mathbf{0}_n$ denotes the n-dimensional column vector with each element being zero.
Start each problem on a new page.
There are **5 questions** and 4 pages.

Problem 1 [7x3=21 points]

Assert if the following statements are True or False. Give a detailed reasoning for your assertion. Marks will be awarded only for your reasoning.

(a) $\mathbf{A} \in \mathbb{R}^{m\times m}$ and $\mathbf{B} \in \mathbb{R}^{m\times m}$ are two symmetric matrices and are related by a similarity transformation. Let $\mathbf{Q} \in \mathbb{R}^{m\times m}$ be the eigenvector matrix of $\mathbf{A}$ and $\overline{\mathbf{Q}} \in \mathbb{R}^{m\times m}$ be the eigenvector matrix of $\mathbf{B}$. The row spaces of $\mathbf{Q}$ and $\overline{\mathbf{Q}}$ are the same.

(b) If $\mathbf{P}$ is a projector, then both (i) range($\mathbf{P}$) $\perp$ null($\mathbf{P}$), and (ii) null($\mathbf{P}$) $\cap$ range($\mathbf{P}$)=$\{0\}$ hold true.

(c) The relative error associated with the solution of a backward stable algorithm $\tilde{f}$ designed to solve problem $f : X \rightarrow Y$ is always $O(\epsilon_{mach})$

(d) For matrices $\mathbf{A} \in \mathbb{R}^{m\times m}$ and $\mathbf{B} \in \mathbb{R}^{m\times m}$, if $\mathbf{B}$ is positive-definite, then $\mathbf{AB}$ and $\mathbf{BA}$ are similar matrices.

(e) Let $\mathbf{M} \in \mathbb{R}^{n\times n}$ be a positive definite matrix and $\mathbf{R}$ is the Cholesky factor of $\mathbf{M}$. Then $\kappa_2(\mathbf{M}) = \kappa_2(\mathbf{R})$ where κ_2 is the matrix condition number defined using the 2-norm of matrix.

(f) If $\mathbf{A} \in \mathbb{R}^{m\times m}$ is a symmetric positive definite matrix and $\mathbf{X} \in \mathbb{R}^{m\times m}$ is any matrix, then $\mathbf{X}^T\mathbf{AX}$ is always symmetric positive definite.

(g) 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 the system of equations $\mathbf{V}_2\mathbf{x} = \mathbf{b}$ has a unique solution $\mathbf{x} \in \mathbb{R}^{n-m}$ for all vectors $\mathbf{b} \in \mathbb{R}^n$ satisfying $\mathbf{Ab} = \mathbf{0}$

Problem 2 [4+6=10 points]

Regression is one of the key aspects of machine learning. In most linear regression type problems, the aim is to find a linear map between input feature vector $\mathbf{x}$ and an output

target scalar y based on a given dataset. One intrinsic assumption in linear regression is that all the data samples are given equal weightage. However, the data is usually collected from multiple sources with varying levels of fidelity involved in the measurement of target scalar y. This necessitates attributing different weights to different data samples. In such a scenario, we define the loss function (or the objective function that one has to minimize to find the optimal parameters in our regression) corresponding to our linear regression problem as follows:

$$\mathbf{L} = \sum_{i=1}^{n} w_i(y_i - \theta_0 - \theta_1 x_{i1} - \theta_2 x_{i2} - \theta_3 x_{i3} - ... - \theta_m x_{im})^2 \tag{1}$$

where n is the number of data samples in the given data set $(\mathbf{x}_i, y_i)_{i=1}^n$. Here $\mathbf{x}_i \in \mathbb{R}^m$, $y_i \in \mathbb{R}$, $w_i > 0$ are given and $n >> m$. The unknown quantities in linear regression are the parameter values $\theta_0, \theta_1, ..., \theta_m$, which need to be estimated through minimization of the loss function. Define a feature matrix $\mathbf{X} \in \mathbb{R}^{n\times(m+1)}$ such that the first entry of each row is 1, and the remaining m entries in each row correspond to the m entries of the feature vector $\mathbf{x}_i$. Based on the information above, answer the following questions:

(a) Show that if $\mathbf{X}$ is a full rank matrix, the optimal parameters obtained through minimization of $\mathbf{L}$ is unique. (*Hint:* Obtain the first derivative of $\mathbf{L}$ with respect to unknown parameters and set it to 0.)

(b) To improve the generalizability of the regression model, it is typical to define a modified loss function $\mathbf{J} = \mathbf{L} + \eta \sum_{j=0}^{m} \theta_i^2$, where $\mathbf{L}$ is defined as in equation (1) and $\eta > 0$ is a user-controllable parameter. The set of optimal parameters can be obtained by minimizing this modified loss function $\mathbf{J}$. Prove that the set of optimal parameters, i.e., $\theta_0, \theta_1, ..., \theta_m$ can always be made unique by varying the value of η irrespective of the rank of $\mathbf{X}$. Given a dataset $(\mathbf{x}_i, y_i)_{i=1}^n$, for what value of (or range of values of) η will the set of optimal parameters be unique?

Problem 3 [4+3+3+7=17 points]

Consider the eigenvalue problem corresponding to a large sparse and diagonally dominant symmetric matrix $\mathbf{A} \in \mathbb{R}^{m\times m}$. Let us say, we are interested in solving $\mathbf{A}\mathbf{x}_i = \lambda_i \mathbf{x}_i$ for $i = 1 \cdots n$ smallest eigenvalue and eigenvector pairs of $\mathbf{A}$ ($n << m$). Answering the following questions will make you deduce an iterative algorithm for solving this eigenvalue problem different from the methods discussed in the class.

(a) Let us begin the iteration $k = 0$ with a trial guess of orthogonal vectors spanning the n-dimensional subspace $\mathbb{V}_{(0)}^n = \{\widetilde{\mathbf{x}}_1^{(0)}, \widetilde{\mathbf{x}}_2^{(0)}, \ldots, \widetilde{\mathbf{x}}_n^{(0)}\}$. Assume each of these vectors act as an approximation to the corresponding exact eigenvectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ of the matrix $\mathbf{A}$. Let $\mathbf{t}_i = \mathbf{x}_i - \widetilde{\mathbf{x}}_i^{(0)}$ for $i = 1, 2, \ldots, n$ denote the correction vectors to the approximate eigenvectors. Show that $\mathbf{t}_i$ can be computed by solving the corrector equation $(\mathbf{A} - \epsilon_i \mathbf{I})\,\mathbf{t}_i = (\epsilon_i \mathbf{I} - \mathbf{A})\,\widetilde{\mathbf{x}}_i^{(0)}$, where ϵ_i is the exact eigenvalue of $\mathbf{A}$. Assuming that you know the exact eigenvalue ϵ_i, can this corrector equation be solved for $\mathbf{t}_i$?

(b) *A priori* one does not know the value of ϵ_i. In this regard what is the best approximation that can be used for ϵ_i and why? (Hint:- Think how the space $\mathbb{V}_{(0)}^n$ is constructed). If ϵ_i is approximated to be $\widetilde{\epsilon}_i^{(0)}$, then the corrector equation becomes $(\mathbf{A} - \widetilde{\epsilon}_i^{(0)}\mathbf{I})\mathbf{t}_i = (\widetilde{\epsilon}_i^{(0)}\mathbf{I} - \mathbf{A})\widetilde{\mathbf{x}}_i^{(0)}$. To solve for $\mathbf{t}_i$ efficiently, one approximates $\mathbf{A}$ in L.H.S of the above corrector equation to the matrix $\mathbf{D} = diag(\mathbf{A})$. To this end, write an

expression for solution to the corrector equation (let $\widetilde{\mathbf{t}}_i$ denote this solution vector, an approximation to the correction vector $\mathbf{t}_i$).

(c) Now we construct a $2n$ dimensional space $\mathbb{V}_{(0)}^{2n} = \{\widetilde{\mathbf{x}}_1^{(0)}, \widetilde{\mathbf{t}}_1^{(0)}, \widetilde{\mathbf{x}}_2^{(0)}, \widetilde{\mathbf{t}}_2^{(0)}, \ldots, \widetilde{\mathbf{x}}_n^{(0)}, \widetilde{\mathbf{t}}_n^{(0)}\}$. Argue why is $\mathbb{V}_{(k)}^{2n}$ a better subspace than $\mathbb{V}_{(k)}^{n}$ to look for eigenvectors of $\mathbf{A}$ at any given iteration k.

(d) We now look for approximate eigenvector, eigenvalue pair of $\mathbf{A}$ in the space $\mathbb{V}_{(0)}^{2n}$. Let us denote $\widetilde{\mathbf{x}}_i^{(1)} \in \mathbb{V}_{(0)}^{2n}$ for $i = 1, 2, \ldots, n$ to be the eigenvector approximations we seek to find in $\mathbb{V}_{(0)}^{2n}$. Define the residual vector $\mathbf{r}_i = \mathbf{A}\widetilde{\mathbf{x}}_i^{(1)} - \widetilde{\epsilon}_i^{(1)}\widetilde{\mathbf{x}}_i^{(1)}$ where $\widetilde{\epsilon}_i^{(1)}$ is the best approximation to the eigenvalue corresponding to the eigenvector approximation $\widetilde{\mathbf{x}}_i^{(1)}$. This approximate eigenvector, eigenvalue pair $(\widetilde{\mathbf{x}}_i^{(1)}, \widetilde{\epsilon}_i^{(1)})$ is obtained by imposing the Galerkin condition that states that $\mathbf{r}_i$ is orthogonal to the space $\mathbb{V}_{(0)}^{2n}$. Mathematically deduce the consequences of the imposition of this Galerkin condition and subsequently elaborate how should one go about finding $(\widetilde{\mathbf{x}}_i^{(1)}, \widetilde{\epsilon}_i^{(1)})$ after imposing the Galerkin condition. Finally, $(\widetilde{\mathbf{x}}_i^{(1)}, \widetilde{\epsilon}_i^{(1)})$ forms the eigenvector, eigenvalue approximations for $k = 1$ iteration.

Note: Once $(\widetilde{\mathbf{x}}_i^{(1)}, \widetilde{\epsilon}_i^{(1)})$ is obtained from $\mathbb{V}^{2n}$, the trial subspace gets updated to $\{\widetilde{\mathbf{x}}_i^{(1)}, \widetilde{\mathbf{t}}_i^{(1)}\}$ to seek $\{\widetilde{\mathbf{x}}_i^{(2)}, \widetilde{\mathbf{t}}_i^{(2)}\}$ and iterations are continued till convergence is reached.

Problem 4 [6+5+4+5=20 points]

We are usually confronted with large sparse matrix eigenvalue problems arising from the discretization of a partial differential equation (eigenproblem), where the unknown eigenfunctions are approximated in a finite-dimensional subspace as a linear combination of localized basis functions spanning the subspace. Let $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m \times m}$ be two such large symmetric positive definite sparse matrices and the problem of finding eigenvector, eigenvalue pairs $(\lambda_i, \mathbf{u}_i)$ satisfying the equation

$$\mathbf{A}\mathbf{u}_i = \lambda_i \mathbf{B}\mathbf{u}_i \tag{2}$$

is called a *generalized* eigenvalue problem (Note: The standard eigenvalue problem you are familiar with is a special case having $\mathbf{B} = \mathbf{I}$).

(a) Rewrite the generalized eigenvalue problem in equation (2) as a standard eigenvalue problem $\mathbf{H}\mathbf{v}_i = \lambda_i \mathbf{v}_i$ and in doing so, ensure that $\mathbf{H}$ is a symmetric matrix. Express $\mathbf{u}_i$ in terms of $\mathbf{v}_i$. (Hint: Use the property of $\mathbf{B}$)

(b) Devise an iterative algorithm to find the eigenvalue λ_i closest to 2.0 (assume 2.0 is not an eigenvalue of equation (2)) and a corresponding eigenvector. Will your algorithm always converge, or is there any condition that needs to be satisfied for convergence?

(c) Identify the computationally dominant step in your algorithm and describe your method of choice for performing this step.

(d) How will you modify the algorithms proposed in parts (b) and (c) if you are told that the eigenvalue closest to 2.0 is, in fact, the lowest eigenvalue? What do you gain in doing so?

Problem 5 [2+3+5+4+3+4+6+3+2=32 points

Let $\mathbf{A} \in \mathbb{R}^{m\times m}$ not necessarily a symmetric matrix be a full rank matrix. Also let $\mathbf{A}_s = \frac{1}{2}(\mathbf{A} + \mathbf{A}^T)$ be a symmetric positive definite matrix. You are now interested in solving the linear system of equations $\mathbf{Ax} = \mathbf{b}$ for some non-zero $\mathbf{b} \in \mathbb{R}^m$ using an iterative solver $\mathcal{S}$. A salient feature of this solver $\mathcal{S}$ is to seek an iterate $\mathbf{x}_n$ lying in a Krylov subspace $\mathbf{K}_n = \{\mathbf{b}, \mathbf{Ab}, \mathbf{A^2b}, ..., \mathbf{A}^{n-1}\mathbf{b}\}$ minimizing the norm of the residual vector corresponding $\mathbf{Ax} = \mathbf{b}$ at every n^{th} iteration. You are now going to deduce the convergence behavior of the solver $\mathcal{S}$ by answering the following questions.

(a) Let $\mathbf{K}_n = \{\mathbf{b}, \mathbf{Ab}, \mathbf{A^2b}, ..., \mathbf{A}^{n-1}\mathbf{b}\}$ be a Krylov subspace of dimension n. Let $\mathbf{x}_n \in \mathbf{K}_n$ be an iterate at n^{th} iteration of the iterative solver $\mathcal{S}$ employed to solve $\mathbf{Ax} = \mathbf{b}$. Show that the residual vector $\mathbf{r}_n = \mathbf{b} - \mathbf{Ax}_n$ can be written as $\mathbf{r}_n = p_n(\mathbf{A})\mathbf{b}$ where $p_n \in P_n$ with $P_n = \{\text{Polynomials } p \text{ of degree} \leqslant n \text{ with } p(0) = 1\}$

(b) The solver $\mathcal{S}$ employed to solve $\mathbf{Ax} = \mathbf{b}$ seeks to find the iterate $\mathbf{x}_n \in \mathbf{K}_n$ such that $||\mathbf{r}_n||_2 = ||p_n(\mathbf{A})\mathbf{b}||_2$ is minimized. Deduce that for $n = 1$, $p_1(\mathbf{A}) = \mathbf{I} - c_0\mathbf{A}$ and show that minimal polynomial which minimizes $||p_1(\mathbf{A})\mathbf{b}||_2$ should have c_0 to be positive.

(c) Recall $\mathbf{A}_s = \frac{1}{2}(\mathbf{A} + \mathbf{A}^T)$ is positive definite. (i) Show that $\mathbf{x}^T\mathbf{Ax} = \mathbf{x}^T\mathbf{A}_s\mathbf{x}$, $\forall \mathbf{x} \in \mathbb{R}^m$. (ii) If λ_{min} is the lowest eigenvalue of the matrix $\mathbf{A}_s$, show that $\mathbf{x}^T\mathbf{A}_s\mathbf{x} \geqslant \lambda_{min}(\mathbf{A}_s) > 0$, $\forall \mathbf{x} \in \mathbb{R}^m$ such that $||\mathbf{x}||_2 = 1$. (iii) If λ_{max} is the highest eigenvalue of $\mathbf{A}^T\mathbf{A}$, show that $(\mathbf{Ax})^T(\mathbf{Ax}) \leqslant \lambda_{max}(\mathbf{A}^T\mathbf{A})$, $\forall \mathbf{x} \in \mathbb{R}^m$ such that $||\mathbf{x}||_2 = 1$.

(d) Show that $||p_1(\mathbf{A})||_2^2 \leqslant 1 - 2c_0\lambda_{min}(\mathbf{A}_s) + c_0^2\lambda_{max}(\mathbf{A}^T\mathbf{A})$ for any $c_0 > 0$. (Hint:- Use the fact that $p_1(\mathbf{A}) = \mathbf{I} - c_0\mathbf{A}$ and the results derived in part(c))

(e) For a general n^{th} iteration, let $\mathbf{x}_n^* \in \mathbf{K}_n$ minimizes $||\mathbf{r}_n||_2 = ||\mathbf{b} - \mathbf{Ax}_n||_2$ and let the corresponding residual vector $||\mathbf{b} - \mathbf{Ax}_n^*||$ be denoted as $\mathbf{r}_n^*$. Show that $\frac{||\mathbf{r}_n^*||_2}{||\mathbf{b}||_2} \leqslant \min_{p_n\in P_n} ||p_n(\mathbf{A})||_2$ where P_n is defined in (a).

(f) Let $p_n^* \in P_n$ minimize $||p_n(\mathbf{A})||_2$. Hence argue that $||p_n^*(\mathbf{A})||_2 \leqslant ||(p_1(\mathbf{A}))^n||_2$. Now show that for the matrix $\mathbf{A} \in \mathbb{R}^{m\times m}$, $||\mathbf{A}^k||_2 \leqslant ||\mathbf{A}||_2^k$ and hence argue that $||p_1(\mathbf{A})^n||_2 \leqslant ||p_1(\mathbf{A})||_2^n$.

(g) Show that $||p_1(\mathbf{A})||_2 \leqslant \sqrt{\left[1 - \frac{\lambda_{min}^2(\mathbf{A}_s)}{\lambda_{max}(\mathbf{A}^T\mathbf{A})}\right]}$. (Hint:- Use the result from (d) for a particular choice of c_0 which minimizes $1 - 2c_0\lambda_{min}(\mathbf{A}_s) + c_0^2\lambda_{max}(\mathbf{A}^T\mathbf{A})$)

(h) Use all the above results to show $\frac{||\mathbf{r}_n^*||_2}{||\mathbf{b}||_2} \leqslant \left[1 - \frac{\lambda_{min}^2(\mathbf{A}_s)}{\lambda_{max}(\mathbf{A}^T\mathbf{A})}\right]^{n/2}$.

(i) If $\mathbf{A}$ is symmetric and positive definite, show that $\frac{||\mathbf{r}_n^*||_2}{||\mathbf{b}||_2} \leqslant \left[\frac{(\kappa(\mathbf{A}))^2 - 1}{(\kappa(\mathbf{A}))^2}\right]^{n/2}$ where $\kappa(\mathbf{A})$ is the condition number of $\mathbf{A}$ in 2-norm. This result establishes the fact that the convergence behaviour of the solver $\mathcal{S}$ depends on condition number of matrix $\mathbf{A}$.