Recitation 6

Recitation Instructor: Shivam Verma
Email: shivamverma@nyu.edu
Ph: 718-362-7836
Office hours: WWH 605 (2.50 - 4.50 pm, Tuesdays)

Brief Overview

- Solving overdetermined linear systems
  - Least squares method
  - QR algorithm
- Eigenvalues and eigenvectors
  - Jacobi's method for finding eigenvalues

Systems of linear equations

$A \in R^{m \times n}, x \in R^{n \times 1}, b \in R^{m \times 1}$

Overdetermined: When m > n (skinny)

This is a system with more equations than unknowns. Can have multiple solutions.

Underdetermined: When m < n (fat)

This is a system with fewer equations than unknowns.

Example.

Can you find the solutions for these?

1. x + y + z = 1, x + y + z = 0
2. x + y + z = 1, x + y + 2z = 3

Two kinds of underdetermined solutions:

- No solution (constraints not satisfied)
- Infinite solutions

Least Squares

Taking the case where m >= n,

- To solve $Ax = b$, minimize the 'residual sum of squares' or 'mean square error' or 'squared euclidean norm'
- Optimization problem:
  - $min_x \;\; \|Ax - b\|_2^{\;2}$
  - Has a closed-form solution, known as the **normal equation**:
    - $A^T Ax = A^T b$
  - Multiple ways of solving

Solve Normal Equation using LU, Cholesky etc.

- If A has full rank, $A^T A$ is invertible. In general, $A^T A$ is a symmetric positive definite. How?
  - $x^T A^T Ax = (Ax)^T (Ax) = \|Ax\|_2^{\;2} \geq 0$
  - This property is very useful in general (see Cholesky decomposition).

- Can use the usual methods (LU, Cholesky etc.) to solve this linear system in $O(mn^2)$.
- Disadvantage:
  - Computing
  - May be ill-conditioned, as $k(A^T A)$ = $k(A)^2$

QR decomposition

$$A = [Q_1 \quad Q_2] \begin{bmatrix} R_{11} \\ 0 \end{bmatrix} = Q_1 R_{11},$$

$$\begin{aligned}\|Ax-b\|^2 &= \|Q^T(Ax-b)\|^2 \\ &= \left\| \begin{bmatrix} R_{11} \\ 0 \end{bmatrix} x - \begin{bmatrix} Q_1^T b \\ Q_2^T b \end{bmatrix} \right\|^2 \\ &= \|R_{11}x - Q_1^T b\|^2 + \|Q_2^T b\|^2.\end{aligned}$$

- Since second term is independent of x, the minimum can be achieved when:
  - $R_{11}x = Q_1{}^T b$
  - This is a triangular linear system. Can be solved in $O(n^2)$

- This decomposition exists for any matrix - rectangular, non-symmetric etc.

- How can we calculate a QR decomposition?

  Givens rotations

  Use sequence of rotations in 2D subspaces:
  For $m \approx n$: $\sim n^2/2$ square roots, and $4/3n^3$ multiplications
  For $m \gg n$: $\sim nm$ square roots, and $2mn^2$ multiplications

  Householder reflections

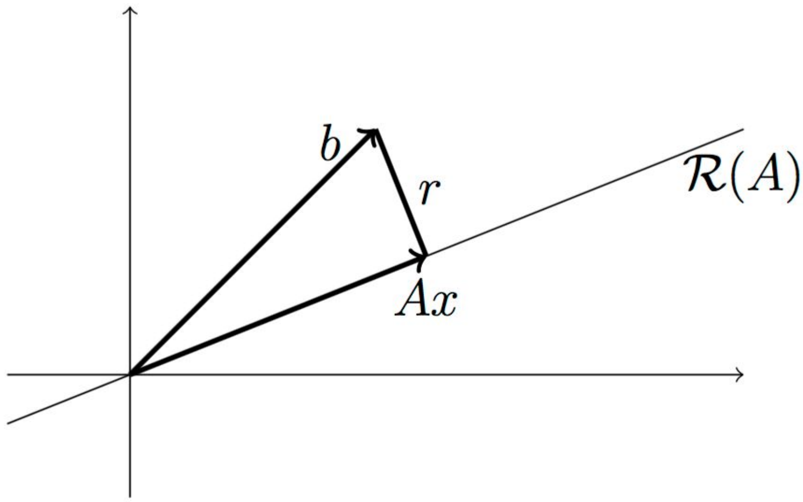
  Use sequence of reflections in 2D subspaces

  For $m \approx n$: $2/3n^3$ multiplications
  For $m \gg n$: $2mn^2$ multiplications

- See textbook or Deuflhard/Hohmann [2] for proof and discussion.

- Advantage: Better conditioned than least-squares, as $k(R_1) = k(A)$. How?
- $k(A^T A) = k(R_1{}^T Q_1{}^T Q_1 R_1) = k(R_1{}^T R_1)$

Geometric interpretation of least squares



- $A^T(Ax - b) = 0 \Rightarrow A^T r = 0$ where r is the residual
- This means residual vector is orthogonal to any vector in the range of A
- $||Ax||^2 + ||r||^2 = ||b||^2$
- Thus, least squares solves for the projection of 'b' on the range space of 'Ax', or, it solves $Ax = b_{projected}$, where $b_{projected} = b \cdot cos(\theta)$
- If $\theta \approx \pi/2$, then $b \cdot cos(\theta) \approx 0$, and corresponding solution will be bad (model doesn't fit data!)

- In general, it may be that columns of A are nearly linearly dependent, in which case problem becomes ill-conditioned, as $A^T A$ is not invertible.
    - One approach is called **regularization**. It involves adding a strictly positive constant to the diagonal elements to make eigenvalues non-zero.
    - $(A^T A + \lambda I)x = A^T b$
    - This is the solution of the minimization problem:

        - $min_x \ \ ||Ax - b||_2^{\ 2} + \lambda ||x||_2^{\ 2}$
        - This is known as L2-regularization, since the “regularization” term involves an L2-norm

    - **(Home Exercise)** Can you say whether we can use an L1-norm instead of the L2-norm for regularization? Is there a closed-form solution for this? why/why not?

- See [1] for an excellent discussion on this topic. Regularization is a very popular concept in applied math, statistics & machine learning, where the objective is also to solve a “system” of nonlinear equations.

Eigenvalue Problems

Theorem 5.1 *Suppose that* $A \in \mathbb{R}^{n\times n}_{\text{sym}}$; *then, the following statements are valid.*

(i) *There exist* n *linearly independent eigenvectors* $\boldsymbol{x}^{(i)} \in \mathbb{R}^n$ *and corresponding eigenvalues* $\lambda_i \in \mathbb{R}$ *such that* $A\boldsymbol{x}^{(i)} = \lambda_i \boldsymbol{x}^{(i)}$ *for all* $i = 1, 2, \dots, n$.

(ii) *The function*

$$\lambda \mapsto \det(A - \lambda I) \tag{5.2}$$

is a polynomial of degree n *with leading term* $(-1)^n\lambda^n$, *called the* **characteristic polynomial of** A. *The eigenvalues of* A *are the zeros of the characteristic polynomial.*

(iii) *If the eigenvalues* λ_i *and* λ_j *of* A *are distinct, then the corresponding eigenvectors* $\boldsymbol{x}^{(i)}$ *and* $\boldsymbol{x}^{(j)}$ *are orthogonal in* $\mathbb{R}^n$, *i.e.,*

$$\boldsymbol{x}^{(i)\mathrm{T}}\boldsymbol{x}^{(j)} = 0 \qquad if\ \lambda_i \neq \lambda_j\,, \qquad i, j \in \{1, 2, \dots, n\}\,.$$

(iv) *If* λ_i *is a root of multiplicity* m *of (5.2), then there is a linear subspace in* $\mathbb{R}^n$ *of dimension* m, *spanned by* m *mutually orthogonal eigenvectors associated with the eigenvalue* λ_i.

(v) *Suppose that each of the eigenvectors* $\boldsymbol{x}^{(i)}$ *of* A *is* **normalised**, *in other words,* $\boldsymbol{x}^{(i)\mathrm{T}}\boldsymbol{x}^{(i)} = 1$ *for* $i = 1, 2, \dots, n$, *and let* X *denote the square matrix whose columns are the normalised (orthogonal) eigenvectors; then, the matrix* $\Lambda = X^{\mathrm{T}} A X$ *is diagonal, and the diagonal elements of* Λ *are the eigenvalues of* A.

(vi) *Let* $Q \in \mathbb{R}^{n\times n}$ *be an orthogonal matrix and define* $B \in \mathbb{R}^{n\times n}_{\text{sym}}$ *by* $B = Q^{\mathrm{T}} A Q$; *then,* $\det(B - \lambda I) = \det(A - \lambda I)$ *for each* $\lambda \in \mathbb{R}$. *The eigenvalues of* B *are the same as the eigenvalues of* A, *and the eigenvectors of* B *are the vectors* $Q^{\mathrm{T}}\boldsymbol{x}^{(i)}$, $i = 1, 2, \dots, n$.

(vii) *Any vector* $\boldsymbol{v} \in \mathbb{R}^n$ *can be expressed as a linear combination of the (ortho)normalised eigenvectors* $\boldsymbol{x}^{(i)}$, $i = 1, 2, \dots, n$, *of* A, *i.e.,*

$$\boldsymbol{v} = \sum_{i=1}^{n} \alpha_i \boldsymbol{x}^{(i)}, \qquad \alpha_i = \boldsymbol{x}^{(i)\mathrm{T}}\boldsymbol{v}\,.$$

(viii) *The trace of* A, $\mathrm{Trace}(A) = \sum_{i=1}^{n} a_{ii}$, *is equal to the sum of the eigenvalues of* A.

[Image source: Chap. 5, Introduction to Numerical Analysis, E. Suli & D. Mayers]

- Recap of some important properties (see image from book)

- If $A \in {R_{symm}}^{n \times n}$ (A is a real, symmetric matrix):
  - There exist 'n' linearly independent eigenvectors
  - $det(\lambda I - A) = 0$ gives the characteristic polynomial (in general)
  - If two eigenvalues are distinct, corresponding eigenvectors are orthogonal
  - Other properties in textbook.

- **Why are Eigenvalue problems important?** Ubiquitous in numerical linear algebra, especially solving ODEs, modelling symmetric physical systems or laws etc.

- **How can we calculate Eigenvalues of large matrices?**

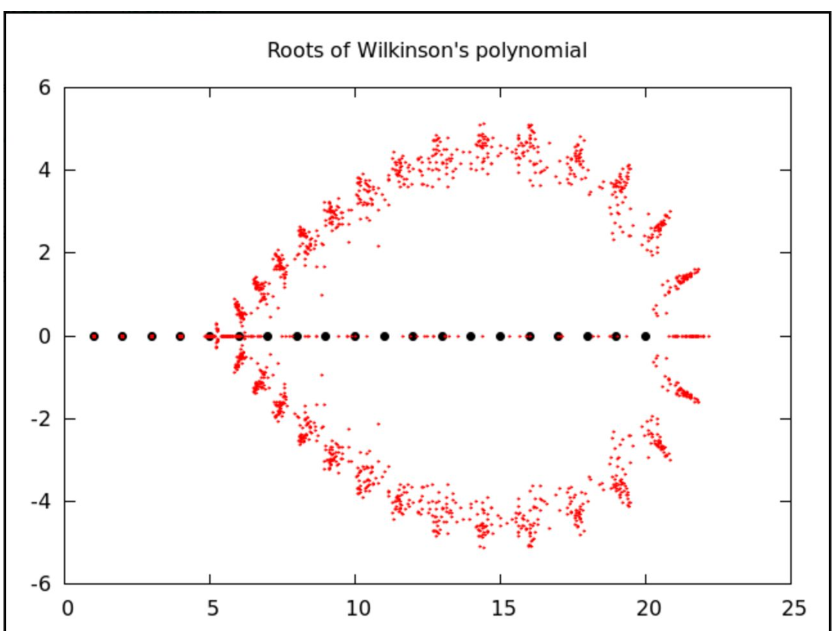
  - **Method 1**: Write down characteristic polynomial, and find its roots numerically. This is not very practical for three reasons:
    i. A 100x100 matrix will have 100 eigenvalues. Newton's method works well when starting very close to the optimal value, but may diverge otherwise.
    ii. May divide the characteristic polynomial once a root has been found, but polynomial division can be numerically dangerous/unstable.
    iii. To find eigenvectors, still have to solve 'n' linear equations, which will take $O(n^3)$ time!
    iv. Note: In general, **polynomial root-finding is an ill-conditioned problem**. See eg. 5.12 (p. 92, Numerical Linear Algebra, Trefethen & Bau) on Wilkinson's polynomial.

  - **Method 2**: Use an iterative method which may diagonalize the matrix, or lead to an eigenvector. Some popular methods are:
    i. Jacobi
    ii. QR
    iii. Sturm sequence
    iv. Power method
    v. Inverse Power or Inverse Iteration

- A bad example. Consider the matrix:

$$A = \begin{pmatrix} 0 & & & & & & \varepsilon \\ 1 & 0 & & & & & \\ & 1 & 0 & & & & \\ & & 1 & 0 & & & \\ & & & 1 & 0 & & \\ & & & & \ddots & \ddots & \\ & & & & & 1 & 0 \end{pmatrix}$$

  - Charac. polynomial: $\lambda^n - \varepsilon = 0$
    i. Case 1: $\varepsilon = 0, \lambda_i = 0$
    ii. Case 2: Let n = 40, and $\varepsilon = 10^{-40}$, relative error (to other elements) = $10^{-40}/1 = 10^{-40}$. One eigenvalue, $\lambda_k = 1/10 = 0.1$. Thus, adding an epsilon term changes one eigenvalue by $10^{39} \times \varepsilon$ times! This is an ill-conditioned problem, and numerically unstable.
  - See [1] for more info. on this problem.
- Another bad example: Wilkinson's polynomial.
  - $p(x) = (x-1)(x-2)(x-3)...(x-19)(x-20)$



Jacobi's method

Idea: Use orthogonal transformations (pre- and post- multiply) to convert matrix to diagonal form.

- Use a plane rotation matrix of the form:

$$R(\varphi) = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}$$

- Can check that this is an orthogonal matrix.
- Choose ϕ to make (p,q) and (q,p) element zero:

$$\varphi = \frac{1}{2}\tan^{-1}\frac{2a_{pq}}{a_{qq} - a_{pp}}$$

- See [4] for example. Also see sec. 5.2 in [3] for good discussion.

Helpful links

1. Comparison of Least Squares and QR
2. Chapter 3 of *Numerical Methods in Scientific Computing*, Deuflhard & Hohmann is excellent for least squares, QR.
3. See p. 1 & 2 on bad eigen value problems
4. See sec. 2 on Jacobi method
5. More about Jacobi's method
6. Found this nice MATLAB tutorial
7. Quick overview of linear algebra and relevant numerical algorithms
8. Jacobi convergence and eigenvalue problem examples
9. See topics 'markov chain 1 / 2' for applications of eigenvalue problem in probability
10. Underdetermined systems