EECS 275 Matrix Computation

Ming-Hsuan Yang

Electrical Engineering and Computer Science
University of California at Merced
Merced, CA 95344
http://faculty.ucmerced.edu/mhyang

Lecture 15

Overview

- Inverse iteration
- Rayleigh quotient iteration
- Condition
- Perturbation
- Stability

Reading

- Chapter 27 and Chapter 12 of *Numerical Linear Algebra* by Llyod Trefethen and David Bau
- Chapter 7 of *Matrix Computations* by Gene Golub and Charles Van Loan

Inverse iteration

- For any $\mu \in \mathbb{R}$ that is not an eigenvalue of A, the eigenvectors of $(A - \mu I)^{-1}$ are the same as the eigenvectors of A, and the the corresponding eigenvalues are $(\lambda_j - \mu)^{-1}$ where λ_j are the eigenvalues of A
- Suppose μ is close to an eigenvalue λ_J of A, then $(\lambda_J - \mu)^{-1}$ may be much larger than $(\lambda_j - \mu)^{-1}$ for all $j \neq J$
- If we apply power iteration to $(A - \mu I)^{-1}$, the process will converge rapidly to $\mathbf{q}_J$
- Algorithm:

    Initialize $\mathbf{v}^{(0)}$ randomly with $\|\mathbf{v}^{(0)}\| = 1$
    for $k = 1, 2, \dots$ **do**
        Solve $\mathbf{w} = (A - \mu I)^{-1}\mathbf{v}^{(k-1)}$ // apply $(A - \mu I)^{-1}$
        $\mathbf{v}^{(k)} = \frac{\mathbf{w}}{\|\mathbf{w}\|}$ // normalize
        $\lambda^{(k)} = (\mathbf{v}^{(k)})^\top A \mathbf{v}^{(k)}$ // Rayleigh quotient
    end for

Inverse iteration (cont'd)

- what if μ is or nearly is an eigenvalue of A, so that $A - \mu I$ is singular?
- can be handled numerically
- exhibit linear convergence
- unlike power iteration, we can choose the eigenvector that will be found by supplying an estimate μ of the corresponding eigenvalue
- if μ is much closer to one eigenvalue of A than to the others, then the largest eigenvalue o $(A - \mu I)^{-1}$ will be much larger than the rest

Theorem

Suppose λ_J is the closest eigenvalue of μ and λ_K is the second closest, that is $|\mu - \lambda_J| < |\mu - \lambda_K| < |\mu - \lambda_j|$ for each $j \neq J$. Furthermore, suppose $\mathbf{q}_J^\top \mathbf{v}^{(0)} \neq 0$, then after k iterations

$$\|\mathbf{v}^{(k)} - (\pm \mathbf{q}_J)\| = O\left(\left| \frac{\mu - \lambda_J}{\mu - \lambda_K} \right|^k \right), \quad |\lambda^{(k)} - \lambda_J| = O\left(\left| \frac{\mu - \lambda_J}{\mu - \lambda_K} \right|^{2k} \right)$$

as $k \to \infty$. The $\pm$ sign means that at each step k, one or the other choice of sign is to be taken, and then the indicated bound holds

Inverse iteration and Rayleigh quotient iteration

- Inverse iteration:
  - one of the most valuable tools of numerical linear algebra
  - a standard method of calculating one or more eigenvectors of a matrix if the eigenvalues are already known
- Obtaining an eigenvalue estimate from an eigenvector estimate (the Rayleigh quotient)
- Obtaining an eigenvector estimate from an eigenvalue estimate (inverse iteration)
- Rayleigh quotient algorithm:

Initialize $\mathbf{v}^{(0)}$ randomly with $\|\mathbf{v}^{(0)}\| = 1$
$\lambda^{(0)} = (\mathbf{v}^{(0)})^\top A\mathbf{v}^{(0)}$ = corresponding Rayleigh quotient
for $k = 1, 2, \ldots$ **do**
  Solve $\mathbf{w} = (A - \lambda^{(k-1)}I)^{-1}\mathbf{v}^{(k-1)}$ // apply $(A - \mu^{(k-1)}I)^{-1}$
  $\mathbf{v}^{(k)} = \frac{\mathbf{w}}{\|\mathbf{w}\|}$ // normalize
  $\lambda^{(k)} = (\mathbf{v}^{(k)})^\top A\mathbf{v}^{(k)}$ // Rayleigh quotient
end for

Rayleigh quotient iteration

Theorem

Rayleigh quotient iteration converges to an eigenvalue/eigenvector pair for all except a set of measure zero of starting vectors $\mathbf{v}^{(0)}$*. When it converges, the convergence is ultimately cubic in the sense that if* λ_J *is an eigenvalue of* A *and* $\mathbf{v}^{(0)}$ *is sufficiently close to the eigenvector* $\mathbf{q}_J$*, then*

$$\|\mathbf{v}^{(k+1)} - (\pm\mathbf{q}_J)\| = O(\|\mathbf{v}^{(k)} - (\pm\mathbf{q}_J)\|^3)$$

and

$$|\lambda^{(k+1)} - \lambda_J| = O(|\lambda^{(k)} - \lambda_J|^3)$$

as $k \to \infty$*. The* $\pm$ *sign means that at each step* k*, one or the other choice of sign is to be taken, and then the indicated bound holds.*

Example

- Consider the symmetric matrix

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 4 \end{bmatrix}$$

  and let $\mathbf{v}^{(0)} = (1,1,1)^\top/\sqrt{3}$ the initial eigenvector estimate
- When Rayleigh quotient iteration is applied to A, the following values $\lambda^{(k)}$ are computed by the first 3 iterations:

$$\lambda^{(0)} = 5, \quad \lambda^{(1)} = 5.2131\ldots, \quad \lambda^{(2)} = 5.214319743184\ldots$$

  The actual value of the eigenvalue corresponding to the eigenvector closest to $\mathbf{v}^{(0)}$ is $\lambda = 5.214319743377$
- After three iterations, Rayleigh quotient iteration has produced a result accurate to 10 digits
- With three more iterations, the solution increases this figure to about 270 digits, if our machine precision were high enough

Operation counts

- Flops (floating point operations): addition, subtraction, multiplication, division, or square root counts as one flop
- Suppose $A \in \mathbb{R}^{m \times m}$ is a full matrix, each step of power iteration involves a matrix-vector multiplication, requiring $O(m^2)$ flops
- Each step of inverse iteration involves the solution of a linear system, which might seem to require $O(m^3)$ flops, but this can be reduced to $O(m^2)$ if the matrix is processed in advance by LU or QR factorization
- In Rayleigh quotient iteration, the matrix to be inverted changes at each step, and beating $O(m^3)$ flops per step
- These figures improve greatly if A is tridiagonal, and all three iterations requires just $O(m)$ flops per step
- For non-symmetric matrices, we must deal with Hessenberg instead of tridiagonal structure, and this figure increases to $O(m^3)$

Conditioning and condition numbers

- Conditioning: pertain to the perturbation behavior of a mathematical problem
- Stability: pertain to the perturbation behavior of an algorithm used to solve the problem on a computer
- We view a problem as a function $f : X \to Y$ from a normed vector space X of data to a normed vector space Y of solutions
- f is usually nonlinear (even in linear algebra), but most of the time it is at least continuous
- A well-conditioned problem: all small perturbations of $\mathbf{x}$ lead to only small changes in $f(\mathbf{x})$
- An ill-conditioned problem: some small perturbation of $\mathbf{x}$ leads to a large change in $f(\mathbf{x})$

Absolute condition number

- Let $\delta \mathbf{x}$ denote a small perturbation of $\mathbf{x}$ and write $\delta f = f(\mathbf{x} + \delta \mathbf{x}) - f(\mathbf{x})$
- Absolute condition number:
$$\hat{\kappa} = \hat{\kappa}(\mathbf{x}) = \lim_{\delta \to 0} \sup_{\|\delta \mathbf{x}\| \le \delta} \frac{\|\delta f\|}{\|\delta \mathbf{x}\|}$$
- For most problems, it can be interpreted as a supremum over all infinitesimal perturbations $\delta \mathbf{x}$, and can write simply as
$$\hat{\kappa} = \sup_{\delta \mathbf{x}} \frac{\|\delta f\|}{\|\delta \mathbf{x}\|}$$
- If f is differentiable, we can evaluate the condition number by means of the derivative of f
- Let $J(\mathbf{x})$ be the matrix where i, j entry is the partial derivative $\partial f_i / \partial \mathbf{x}_j$ evaluated at $\mathbf{x}$, known as the Jacobian of f at $\mathbf{x}$
- By the definition, $\delta f \approx J(\mathbf{x}) \delta \mathbf{x}$ with $\|\delta \mathbf{x}\| \to 0$, and the absolute condition number becomes
$$\hat{\kappa} = \|J(\mathbf{x})\|$$
where $\|J(\mathbf{x})\|$ represents the norm of $J(\mathbf{x})$

Relative condition number

- Relative condition number:

$$\kappa = \lim_{\delta \to 0} \sup_{\|\delta \mathbf{x}\| \le \delta} \left(\frac{\|\delta f\|}{\|f(\mathbf{x})\|} \Big/ \frac{\|\delta \mathbf{x}\|}{\|\mathbf{x}\|} \right)$$

- Again, assume $\delta \mathbf{x}$ and δf are infinitesimal,

$$\kappa = \sup_{\delta \mathbf{x}} \left(\frac{\|\delta f\|}{\|f(\mathbf{x})\|} \Big/ \frac{\|\delta \mathbf{x}\|}{\|\mathbf{x}\|} \right)$$

- If f is differentiable, we can express this in terms of Jacobian

$$\kappa = \frac{\|J(\mathbf{x})\|}{\|f(\mathbf{x})\| / \|\mathbf{x}\|}$$

- Relative condition number is more important in numerical analysis
- A problem is well-conditioned if κ is small (e.g., 1, 10, 10^2), and ill-conditioned if κ is large (e.g., 10^6, 10^{16})

Well-conditioned and ill-conditioned problems

- Consider the problem of obtaining the scalar $x/2$ from $x \in \mathbb{C}$. The Jacobian of the function $f : x \to x/2$ is the derivative $J = f' = 1/2$

$$\kappa = \frac{\|J\|}{\|f(x)\|/\|x\|} = \frac{1/2}{(x/2)/x} = 1.$$

  This problem is well-conditioned
- Consider $x^2 - 2x + 1 = (x-1)^2$, with a double root $x = 1$. A small perturbation in the coefficient may lead to large change in the roots

$$x^2 - 2x + 0.9999 = (x - 0.99)(x - 1.01).$$

  In fact, the roots can change in proportion to the square root of the change in the coefficients, so in this case the Jacobian is infinite (the problem is not differentiable), and $\kappa = \infty$

Computing eigenvalues of a non-symmetric matrix

- Computing the eigenvalues of a non-symmetric matrix is often ill-conditioned
- Consider the two matrices

$$A = \begin{bmatrix} 1 & 1000 \\ 0 & 1 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 1000 \\ 0.001 & 1 \end{bmatrix}$$

  and the eigenvalues are $\{1, 1\}$ and $\{0, 2\}$ respectively
- On the other hand, if a matrix A is symmetric (more generally, if it is normal), then its eigenvalues are well-conditioned
- It can be shown that if λ and $\lambda + \delta\lambda$ are corresponding eigenvalues of A and $A + \delta A$, then $|\delta\lambda| \leq \|\delta A\|_2$ with equality if δA is a multiple of the identity
- In that case, $\hat{\kappa} = 1$ if perturbations are measured in the 2-norm and the relative condition number is $\kappa = \|A\|_2 / |\lambda|$

Condition of matrix-vector multiplication

- Consider $A \in \mathbb{C}^{m\times n}$, determine a condition number corresponding to perturbations of $\mathbf{x}$

$$\kappa = \sum_{\delta\mathbf{x}} \left(\frac{\|A(\mathbf{x}+\delta x) - A\mathbf{x}\|}{\|A\mathbf{x}\|} \bigg/ \frac{\|\delta\mathbf{x}\|}{\|\mathbf{x}\|} \right) = \sup_{\delta\mathbf{x}} \frac{\|A\delta\mathbf{x}\|}{\|\delta\mathbf{x}\|} \bigg/ \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|},$$

  that is

$$\kappa = \|A\| \frac{\|\mathbf{x}\|}{\|A\mathbf{x}\|} \qquad (\kappa = \frac{\|J(\mathbf{x})\|}{\|f(\mathbf{x})\|/\|\mathbf{x}\|})$$

  which is an exact formula for κ, dependent on both A and $\mathbf{x}$

- If A happens to be square and non-singular, we can use the fact that $\|\mathbf{x}\|/\|A\mathbf{x}\| \leq \|A^{-1}\|$ to loosen the bound

$$\kappa \leq \|A\|\|A^{-1}\|,$$

  or

$$\kappa = \alpha\|A\|\|A^{-1}\|, \quad \alpha = \frac{\|\mathbf{x}\|}{\|A\mathbf{x}\|} \bigg/ \|A^{-1}\|$$

Condition of matrix-vector multiplication (cont'd)

- For certain choices of **x**, we have $\alpha = 1$, and consequently $\kappa = \|A\|\|A^{-1}\|$
- If we use 2-norm, this will occur when **x** is a multiple of a minimal right singular vector of A
- If $A \in \mathbb{C}^{m\times n}$ with $m \geq n$ has full rank, the above equations hold with A^{-1} replaced by the pseudo-inverse $A^{\dagger} = (A^H A)^{-1} A^H \in \mathbb{C}^{n\times m}$
- Given A, compute $A^{-1}\mathbf{b}$ from input $\mathbf{b}$?
- Mathematically, identical to the problem just considered except that A is repalced by A^{-1}

Condition of matrix-vector multiplication (cont'd)

Theorem

Let $A \in \mathbb{C}^{m\times n}$ be nonsingular and consider the equation $A\mathbf{x} = \mathbf{b}$. The problem of computing $\mathbf{b}$ given $\mathbf{x}$, has condition number

$$\kappa = \|A\|\frac{\|\mathbf{x}\|}{\|\mathbf{b}\|} \leq \|A\|\|A^{-1}\| \tag{1}$$

with respect to perturbation of $\mathbf{x}$. The problem of computing $\mathbf{x}$ given $\mathbf{b}$, has condition number

$$\kappa = \|A^{-1}\|\frac{\|\mathbf{b}\|}{\|\mathbf{x}\|} \leq \|A\|\|A^{-1}\| \tag{2}$$

with respect to perturbation of $\mathbf{b}$. If we use 2-norm, then the equality holds in (1) if $\mathbf{x}$ is a multiple of a right singular vector of A corresponding to the minimal singular value σ_m. Likewise, the equality hold in (2) if $\mathbf{b}$ is a multiple of a left singular vector of A corresponding to the maximal singular value σ_1.

Condition number of a matrix

- The product $\|A\|\|A^{-1}\|$ is the condition number of A (relative to the norm $\|\cdot\|$), denoted by $\kappa(A)$

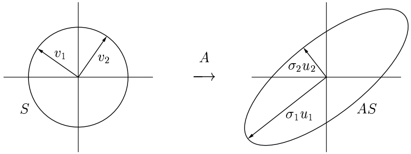
$$\kappa(A) = \|A\|\|A^{-1}\|$$

- The condition number is attached to a matrix, not a problem
- If $\kappa(A)$ is small, A is said to be well-conditioned; if $\kappa(A)$ is large, A is ill-conditioned.
- If A is singular, it is customary to write $\kappa(A) = \infty$
- If $\|\cdot\| = \|\cdot\|_2$, then $\|A\| = \sigma_1$, and $\|A^{-1}\| = 1/\sigma_m$, and thus

$$\kappa(A) = \frac{\sigma_1}{\sigma_m}$$

  in the 2-norm, which is the formula for computing 2-norm condition numbers of matrices.
- The ratio σ_1/σ_m can be interpreted as the eccentricity of the hyperellipse that is the image of the unit sphere of $\mathbb{C}^m$ under A

Condition number of a matrix (cont'd)



- For a rectangular matrix $A \in C^{m \times n}$ of full rank, $m \geq n$, the condition number is defined in terms of pseudo-inverse

$$\kappa(A) = \|A\| \|A^\dagger\|$$

since $A^\dagger$ is motivated by least squares problem, this definition is most useful with 2-norm, where we have

$$\kappa(A) = \frac{\sigma_1}{\sigma_n}$$

Condition number of a system of equations

- Let $\mathbf{b}$ be fixed, and consider the behavior of the problem $A \mapsto \mathbf{x} = A^{-1}\mathbf{b}$ when A is perturbed by infinitesimal δA, then

$$(A + \delta A)(\mathbf{x} + \delta \mathbf{x}) = \mathbf{b}$$

- Using $A\mathbf{x} = \mathbf{b}$ and dropping the term $(\delta A)(\delta \mathbf{x})$, we obtain $(\delta A)\mathbf{x} + A(\delta \mathbf{x}) = 0$, that is, $\delta \mathbf{x} = -A^{-1}(\delta A)\mathbf{x}$
- It implies $\|\delta \mathbf{x}\| \leq \|A^{-1}\| \|\delta A\| \|\mathbf{x}\|$, or equivalently

$$\frac{\|\delta \mathbf{x}\|}{\|\mathbf{x}\|} \Big/ \frac{\|\delta A\|}{\|A\|} \leq \|A^{-1}\| \|A\| = \kappa(A)$$

- The equality in this bound will hold when δA is such that

$$\|A^{-1}(\delta A)\mathbf{x}\| = \|A^{-1}\| \|\delta A\| \|\mathbf{x}\|$$

  and it can be shown that for any A and $\mathbf{b}$ and norm $\|\cdot\|$, such perturbations of δA exists

Condition number of a system of equations (cont'd)

Theorem

Let $\mathbf{b}$ *be fixed and consider the problem of computing* $\mathbf{x} = A^{-1}\mathbf{b}$ *where* A *is square and nonsingular. The condition number of this problem with respect to perturbation in* A *is*

$$\kappa = \|A\|\|A^{-1}\| = \kappa(A)$$

- Both theorems regarding matrix-vector multiplication and system of equations are of fundamental importance in numerical linear algebra
- They determine how accurately one can solve systems of equations
- If a problem $A\mathbf{x} = \mathbf{b}$ contains an ill-conditioned matrix A, one must always expect to "lose $\log_{10} \kappa(A)$ digits" in computing the solution (except under very special cases)