Adv. Numerical Analysis Lecture 1: Matrix-Vector Multiplication; Orthogonal Vectors and Matrices

Panruo Wu

To accompany the *Numerical Linear Algebra* book, Lectures 1 and 2.

Matrix-vector Multiplication (1)

- New interpretation that is essential for numerical linear algebra: if b=Ax, then b is the linear combination of the columns of A.

Let a_j denote the jth column of A, an m-vector. Then (1.1) can be rewritten

$$b = Ax = \sum_{j=1}^{n} x_j a_j. \tag{1.2}$$

This equation can be displayed schematically as follows:

$$\left[\begin{array}{c} \\ b \\ \\ \end{array}\right] = \left[\begin{array}{c|c|c|c} & & & \\ a_1 & a_2 & \cdots & a_n \\ & & & \end{array}\right] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 \left[\begin{array}{c} \\ a_1 \\ \\ \end{array}\right] + x_2 \left[\begin{array}{c} \\ a_2 \\ \\ \end{array}\right] + \cdots + x_n \left[\begin{array}{c} \\ a_n \\ \\ \end{array}\right].$$

Matrix-vector Multiplication (2)

- Mathematician: Ax=b means matrix A (representing linear transformation) acts on vector x to produce b.
- Numerical analyst: Ax=b means vector x acts on A to produce b (as linear combination of columns of A).
- Function can be seen as <u>infinite dimensional vector</u>
  Matrix A's column can be function

Matrix Times a Matrix

- Matrix product B=A*C, each column of B is a linear combination of the columns of A.

$$b_j = Ac_j = \sum_{k=1}^{m} c_{kj} a_k.$$

$$\left[\begin{array}{c|c|c|c} & & & \\ b_1 & b_2 & \cdots & b_n \\ & & & \end{array}\right] = \left[\begin{array}{c|c|c|c} & & & \\ a_1 & a_2 & \cdots & a_m \\ & & & \end{array}\right] \left[\begin{array}{c|c|c|c} & & & \\ c_1 & c_2 & \cdots & c_n \\ & & & \end{array}\right]$$

Example 1: Outer Product

$$\left[\begin{array}{c} \\ u \\ \\ \end{array}\right]\left[\begin{array}{cccc} v_1 & v_2 & \cdots & v_n \end{array}\right] = \left[\begin{array}{c|c|c|c} & & & \\ v_1 u & v_2 u & \cdots & v_n u \\ & & & \end{array}\right] = \left[\begin{array}{ccc} v_1 u_1 & \cdots & v_n u_1 \\ \vdots & & \vdots \\ v_1 u_m & \cdots & v_n u_m \end{array}\right]$$

Example 2: Matrix Times Triangular Matrix

B = A*R

$$\begin{bmatrix} \Big| & & \Big| \\ b_1 & \cdots & b_n \\ \Big| & & \Big| \end{bmatrix} = \begin{bmatrix} \Big| & & \Big| \\ a_1 & \cdots & a_n \\ \Big| & & \Big| \end{bmatrix} \begin{bmatrix} 1 & \cdots & 1 \\ & \ddots & \vdots \\ & & 1 \end{bmatrix}.$$

$$b_j = Ar_j = \sum_{k=1}^{j} a_k.$$

The j-th column of B is the sum of the first j columns of A.

Range

- Range of matrix A (size m by n), denoted by range(A) or sometimes ran(A), means the set of vectors that can be expressed as Ax for some x. Namely

$$\text{Ran}(A) = \{Ax : \text{any } x\}$$

- Theorem 1.1 <u>Ran(A) is the space spanned by the columns of A.</u> Why?

$$Ax = a1*x1 + a2*x2 + \ldots + an*xn$$

- Because of the new interpretation that Ax is a linear combination of columns of A!
- Thus ran(A) is also sometimes called the <u>column space of A</u>

Nullspace

- The nullspace of matrix A, written as null(A), is the set of vectors x that satisfy Ax=0. i.e.
  null(A) = {x : Ax=0}

Rank

- Column rank is the dimension of the column space.
- Row rank is the dimension of the row space.
- Row rank = column rank (?!! Why?)
  So we are going to simply say rank(A) to refer to either column rank or row rank.

- For a matrix A of size m by n (m>=n), we say A is of <u>full rank</u> if A has the maximal rank (which is n). i.e. A has independent columns.
- Theorem1.2 A matrix A with size m>=n has full rank if and only if it maps no two distinct vectors to the same vector.

Inverse

- A full-rank square matrix is called non-singular. If not full rank the it's called singular.
- A non-singular matrix is invertible, and its inverse is denoted by A^{-1}, which satisfies $A \times A^{-1} = A^{-1} \times A = I$

Theorem 1.3. *For $A \in \mathbb{C}^{m \times m}$, the following conditions are equivalent:*

(a) *A has an inverse A^{-1},*
(b) $\operatorname{rank}(A) = m$,
(c) $\operatorname{range}(A) = \mathbb{C}^m$,
(d) $\operatorname{null}(A) = \{0\}$,
(e) *0 is not an eigenvalue of A,*
(f) *0 is not a singular value of A,*
(g) $\det(A) \neq 0$.

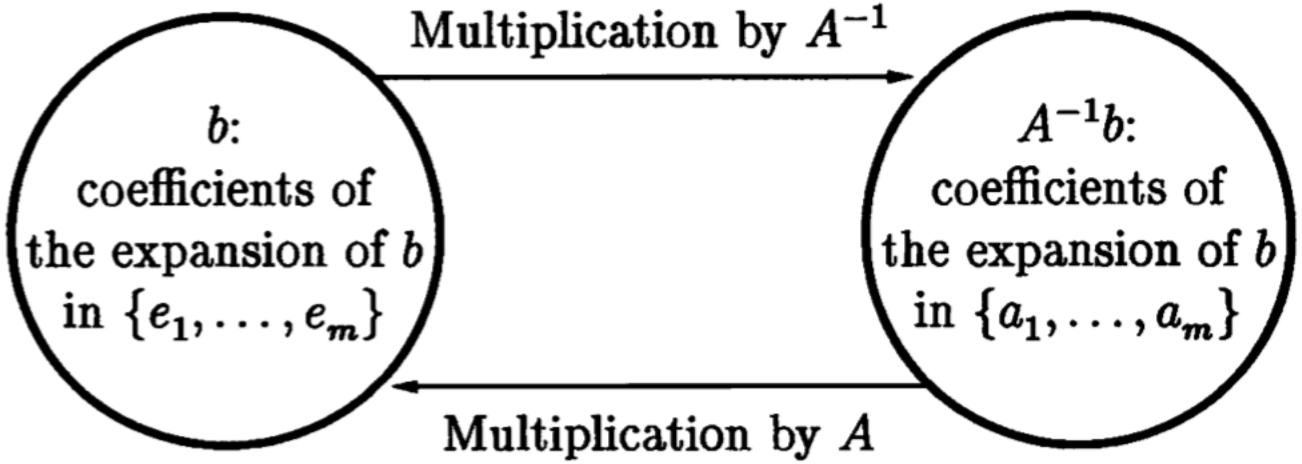
A Matrix Inverse Times a Vector

- What does

$$x = A^{-1}b$$

  mean?

- It means x is the unique vector that satisfies Ax=b.
- It also means <u>x is the vector of coefficients of the unique linear expansion of b in the basis of columns of A.</u>



Orthogonal Vectors and Matrices

- Orthogonality is the base for many of the best algorithms of numerical linear algebra.
- Two ingredients: orthogonal vectors and orthogonal (unitary) matrices.

Symmetric matrix

- We are going to focus on real matrix.
- A is symmetric iff $A = A^T$
- Symmetric matrices are a class of very nice and important matrices; And very useful in practice because a lot of applications will generate symmetric matrices!

Inner Product (1)

- The inner product of two column vectors x,y is the product of transpose of x by y:

$$x^T y = y^T x = \sum_{i=1}^{m} x_i y_i$$

- Sometimes also written as (x,y), or <x,y>
- The Euclidean length of x is the inner product of x by itself.

$$\|x\| = \sqrt{x^* x} = \left(\sum_{i=1}^{m} |x_i|^2 \right)^{1/2}$$

- The cosine of angle alpha between x and y can be expressed with inner product:

$$\cos \alpha = \frac{x^* y}{\|x\| \, \|y\|}.$$

Inner Product (2)

- Inner product is <u>bilinear</u>:

$$\begin{aligned}(x_1 + x_2)^* y &= x_1^* y + x_2^* y,\\ x^*(y_1 + y_2) &= x^* y_1 + x^* y_2,\\ (\alpha x)^*(\beta y) &= \overline{\alpha}\beta x^* y.\end{aligned}$$

- Some formulas for matrix products:

$$(AB)^* = B^* A^*.$$

$$(AB)^{-1} = B^{-1} A^{-1}.$$

Orthogonal Vectors

- Vectors x,y are said to be orthogonal, iff their inner product is 0; i.e. $x^T y = 0$
- Two *sets* of vectors X,Y are said to be orthogonal, if *every* x in X is orthogonal to *every* y in Y.
- A set of vectors S is orthogonal iff its elements are pairwise orthogonal, ie.e. for every x!=y in S, $x^T y = 0$
- A set of vector S is <u>orthonormal</u> if it's orthogonal and, in addition, every s in S has unit norm: $\|s\| = 1$
- Theorem 2.1. The vectors in an orthogonal set S are <u>linearly independent</u>.

Components of a Vector (1)

- The most important idea to draw from inner product and orthogonality is this:
  <u>Inner product can be used to decompose arbitrary vectors into orthogonal components</u>
- Suppose {q1,q2,…,qn} is an orthonormal set, and v is arbitrary vector. The quantity $q_j^T v$ is a scalar. Consider the vector:

$$r = v - (q_1^* v)q_1 - (q_2^* v)q_2 - \cdots - (q_n^* v)q_n$$

- I claim: r is orthogonal to, $\{q_1, q_2, \ldots, q_n\}$.
- Why? (hint: try compute $g_i^T v$).

Components of a Vector (2)

- We see that v can be decomposed into n+1 orthogonal components:
- If $\{q_i\}$ are a basis for $\mathbb{R}^{\mathrm{m}\times m}$, r must be 0.

$$v = r + \sum_{i=1}^{n} (q_i^T v) q_i = r + \sum_{i=1}^{n} (q_i q_i^T) v$$

$$v = \sum_{i=1}^{m} (q_i^T v) q_i = \sum_{i=1}^{n} (q_i q_i^T) v$$

$\downarrow$ projector.

why?

$(q_i^T v) q_i$

$= q_i (q_i^T v)$

$= (q_i q_i^T) v$

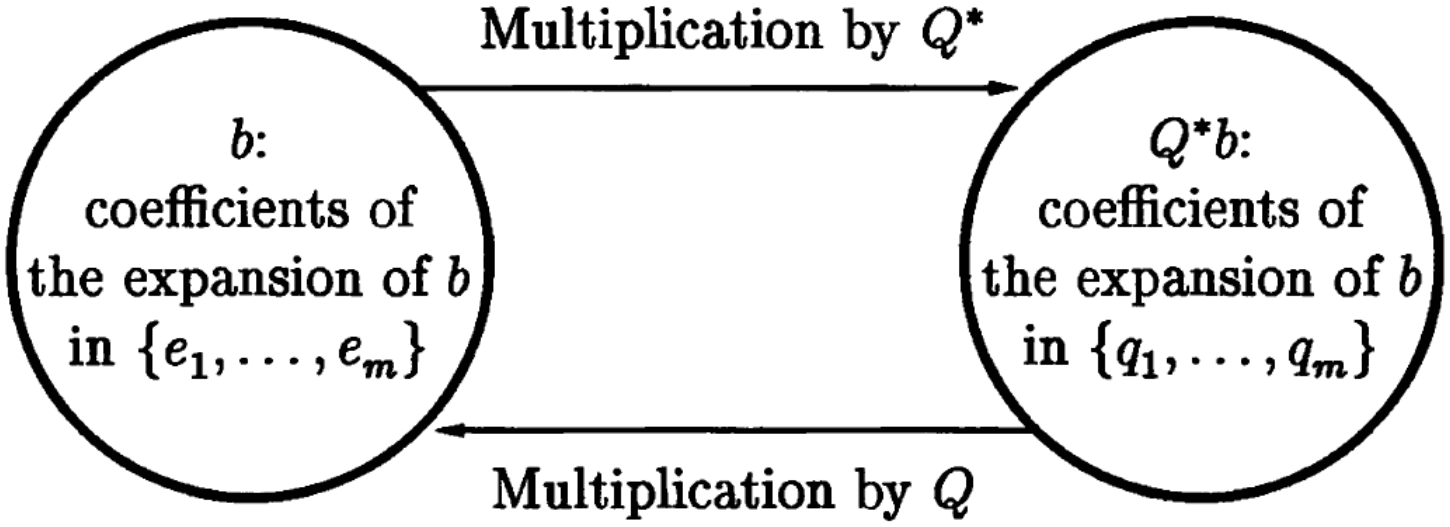
Orthogonal(Unitary) Matrices

- A square matrix $Q \in \mathbb{C}^{m\times m}$ is <u>unitary</u> (in the real case, we say <u>orthogonal</u>) iff $Q^* = Q^{-1}$, i.e. $Q^*Q = I$

$$\begin{bmatrix} \quad q_1^* \quad \\ \hline q_2^* \\ \hline \vdots \\ \hline q_m^* \end{bmatrix} \left[\begin{array}{c|c|c|c} q_1 & q_2 & \cdots & q_m \end{array}\right] = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}.$$

Multiplication by a Unitary Matrix (1)

- Last time, we discussed the interpretation of matrix-vector products Ax and $A^{-1}b$. If A is a unitary matrix Q these products become Qx and Q^*b, and the same interpretation is still valid.



Multiplication by a Unitary Matrix (2)

- The multiplication by a unitary matrix preserve geometric structure in Euclidean sense, i.e.

$$(Qx)^*(Qy) = x^*y$$
$$\|Qx\| = \|x\|$$

- The inner product is <u>invariant</u> under unitary transformation.
- Subsequently, unitary transformation preserves Euclidean norm (length).
- In the real case, multiplication by an orthogonal matrix Q corresponds to a rigid rotation (if $\det(Q) = 1$) or reflection (if $det(Q) = -1$) of the vector space.

Norms

Corresponds to Lecture 3 in NLA book.

Vector Norm

- Why norm?
  The notion to capture size and distance in a vector space.
- How is norm useful?
  To measure approximations and convergence.
- What is norm?
  A norm is a function $\|v\|: \mathbb{C}^m \rightarrow \mathbb{R}$ that assigns a real-value length to each vector.
  Norm must satisfy the following three conditions:

$$(1) \quad \|x\| \geq 0, \text{ and } \|x\| = 0 \text{ only if } x = 0,$$

$$(2) \quad \|x + y\| \leq \|x\| + \|y\|,$$

$$(3) \quad \|\alpha x\| = |\alpha| \, \|x\|.$$

The p-norm of Vectors

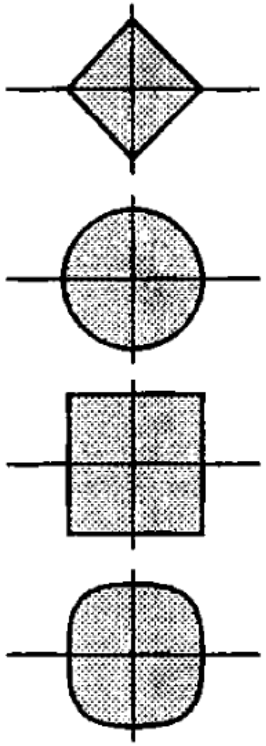
- The most important class of vector norms, the p-norm and their closed unit ball $\{x \in \mathbb{C}^m : \|x\| \leq 1\}$ for m=2:

$$\|x\|_1 = \sum_{i=1}^{m} |x_i|,$$

$$\|x\|_2 = \left(\sum_{i=1}^{m} |x_i|^2 \right)^{1/2} = \sqrt{x^* x},$$

$$\|x\|_\infty = \max_{1 \leq i \leq m} |x_i|,$$

$$\|x\|_p = \left(\sum_{i=1}^{m} |x_i|^p \right)^{1/p} \quad (1 \leq p < \infty).$$



Matrix Norms Induced by Vector Norms

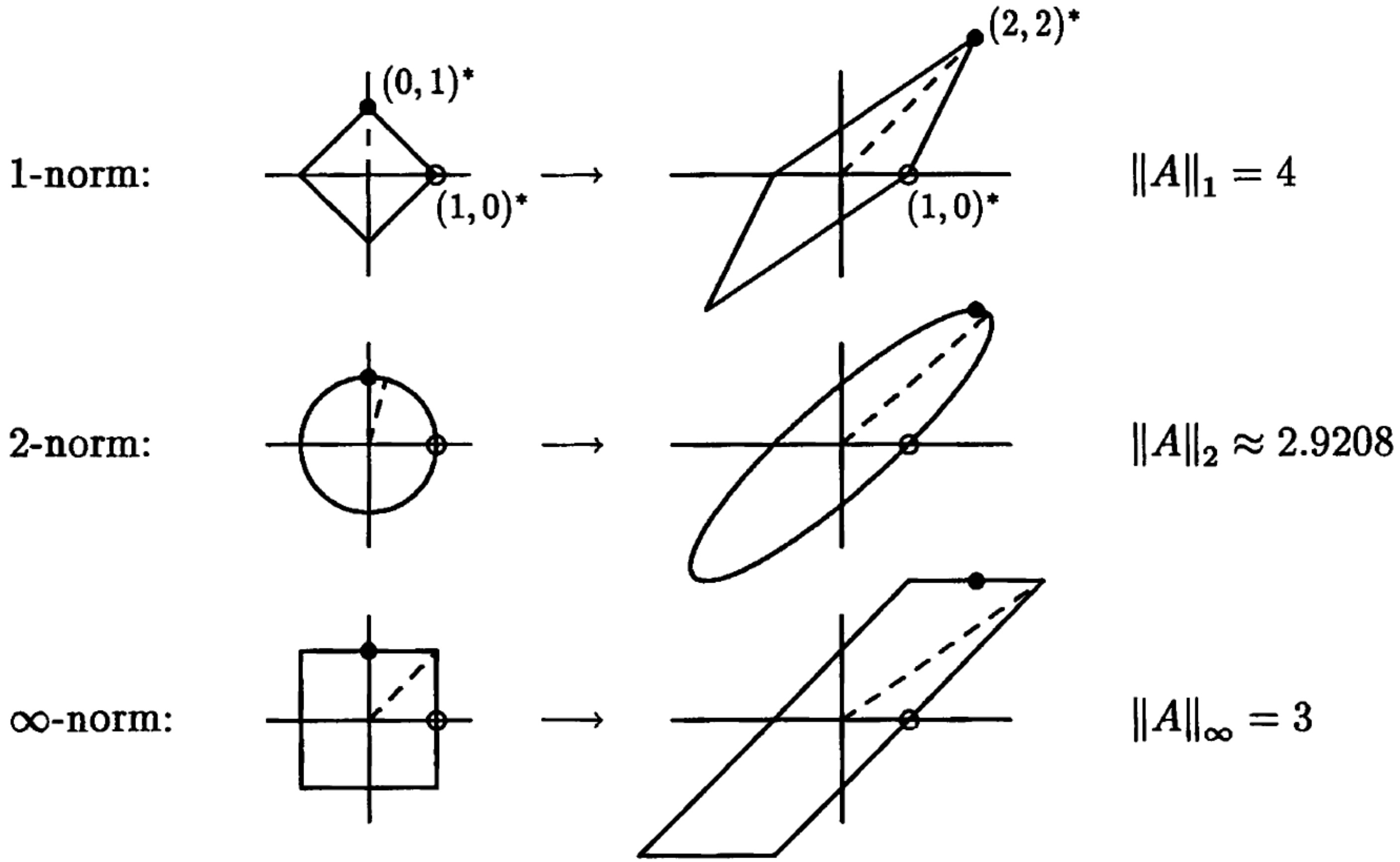
- We know the norm of vectors, how about norms of matrices?
- Induced matrix norm: $A \in \mathbb{C}^{m\times n}$,

$$\|A\|_{(m,n)} = \max \frac{\|Ax\|_{(m)}}{\|x\|_{(n)}}$$

The maximum factor by which A can <u>stretch</u> a vector x.

Example

- The matrix $A = \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}$ maps $\mathbb{R}^2$ to $\mathbb{R}^2$; the following figure depicts its action on the unit ball defined by 1-, 2-, ∝-norms.



Example

- The p-norm of a Diagonal Matrix. Let

$$D = \begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_m \end{bmatrix}.$$

- In 2-norm, how does D map a unit ball?
  An ellipse whose semiaxis lengths are given by $|\mathrm{d}_i|$.
  The most amplified factor thus is $\|D\|_2 = \max_{1 \le \mathrm{i} \le m} \{|\mathrm{d}_i|\}$
- In fact, this generates to arbitrary p-norm (only for diagonal matrix!!!):

$$\|D\|_{\mathrm{p}} = \max_{1 \le \mathrm{i} \le m} \{|\mathrm{d}_i|\}$$

∞-norm and 1-norm of Matrix

- The 1-norm of a matrix A is the maximum column sum of A. Why?
  Hint: write

$$A = [a_1|a_2|\dots|a_n]$$

The ∞-norm of a matrix A is the maximum row sum of A. Why?

The 2-norm of an Outer-Product

- The rank 1 matrix of outer product of column vectors u and v:

$$\|uv^*\|_2 = \|u\|_2\|v\|_2$$

  why?

  Hint: try the definition. And Cauchy-Schwarz inequality

$$|x^*y| \leq \|x\|_2\|y\|_2$$

Bounding $\|AB\|$ in an Induced Matrix Norm

- The induced norm of matrix product can be bounded:

$$A \in \mathbb{R}^{l\times m}, B \in \mathbb{R}^{m\times n}$$
$$\|AB\|_{(l,n)} \leq \|A\|_{(l,m)} \times \|B\|_{(m,n)}$$

Frobenius Norm

- The most important matrix norm that is not induced by a vector norm is the Frobenius norm (F-norm)

$$\|A\|_F = (\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2)^{1/2}$$

- $\|A\|_F = \sqrt{\mathrm{tr}(A^*A)} = \sqrt{\mathrm{tr}(AA^*)}$, where tr(B) denotes the trace of B, the sum of diagonal entries.
- Like induced matrix norm, the F-norm can be used to bound products of matrices.

$$\|AB\|_F^2 \leq \|A\|_F\|B\|_F$$

Invariance under Unitary Multiplication

- Theorem 3.1 For any $A \in \mathbb{C}^{m\times n}$ and unitary $Q \in \mathbb{C}^{m\times m}$ we have
$$\|QA\|_2 = \|A\|_2$$
$$\|QA\|_F = \|A\|_F$$
why?
For 2-norm, think the definition of 2-norm.
For F-norm, use the trace equality $\|A\|_F = \sqrt{\mathrm{tr}(AA^*)}$

- The theorem remains valid if Q is a rectangular matrix with orthonormal columns, that is, $Q \in \mathbb{C}^{p\times m}$ with p>m.
- Analogous results hold for multiplication by unitary matrices on the right. The matrix can be generalized to rectangular matrix with orthonormal rows.