University of Washington
Lecture Notes
Linear Algebra for Controls

Xu Chen
Bryan T. McMinn Endowed Research Professorship
Associate Professor
Department of Mechanical Engineering
University of Washington
chx [AT] uw [DOT] edu

Contents

1 Basic concepts of matrices and vectors 1
1.1 Matrix addition and multiplication . . . 2
1.2 Matrix transposition . . . 3

2 Linear systems of equations 4

3 Vector space, linear independence, basis, and span 7

4 Matrix properties 8
4.1 Rank . . . 8
4.2 Range and null spaces . . . 8
4.3 Determinants . . . 8

5 Matrix and linear equations 9

6 Eigenvector and eigenvalue 11
6.1 Matrix, mappings, and eigenvectors . . . 11
6.2 Computation of eigenvalue and eigenvectors . . . 12
6.3 Eigenbases and diagonalization . . . 15

7 Similarity transformation 17

8 Matrix inversion 18
8.1 Block matrix decomposition and inversion . . . 21
8.2 *LU and Cholesky decomposition . . . 22
8.3 Determinant and matrix inverse identity . . . 23
8.4 Matrix inversion lemma . . . 24
8.5 Special inverse equalities . . . 26

9 Spectral mapping theorem 27

10 Matrix exponentials 28

11 Inner product 29
11.1 Inner product spaces . . . 29
11.2 Trace (standard matrix inner product) . . . 30

12 Norms 31
12.1 Vector norm . . . 31
12.2 Induced matrix norm . . . 32
12.3 Frobenius norm and general matrix norms . . . 32
12.4 Norm inequalities . . . 34
12.5 Exercises . . . 35

13 Symmetric, skew-symmetric, and orthogonal matrices 35
13.1 Definitions and basic properties . . . 35
13.2 Symmetric eigenvalue decomposition (SED) . . . 37
13.3 Symmetric positive-definite matrices . . . 40
13.4 General positive-definite matrices . . . 42
13.5 *Positive-definite functions and non-constant matrices . . . 43

14 Singular value and singular value decomposition (SVD) 44
14.1 Motivation . . . 44
14.2 SVD . . . 44
14.3 Properties of singular values . . . 47
14.4 Exercises . . . 47

1 Basic concepts of matrices and vectors

A linear equation set

$$\begin{aligned} 3x_1 + 4x_2 + 10x_3 &= 6 \\ x_1 + 4x_2 - 10x_3 &= 5 \\ 4x_2 + 10x_3 &= -1 \end{aligned} \quad (1)$$

can be simply written as

$$\begin{bmatrix} 3 & 4 & 10 \\ 1 & 4 & -10 \\ 0 & 4 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 6 \\ 5 \\ -1 \end{bmatrix} \quad (2)$$

(2) wrote x_1, x_2, and x_3 just once rather than two or three times in (1). There are only three unknowns in the above linear equation set. The notational simplicity and many algebraic convenience that will arise, however, are significant when we have thousands of unknowns...

Formally, we write an $m \times n$ matrix A as

$$A = [a_{jk}] = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & \dots & \dots & a_{2n} \\ \vdots & \dots & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

- $m \times n$ (reads m by n) is the dimension/size of the matrix. It means that A has m rows and n columns.
- Each element a_{jk} is an entry of the matrix. For two matrices A and B to be equal, it must be that $a_{jk} = b_{jk}$ for any j and k.
- If $m = n$, A belongs to the class of square matrices. The entries a_{11}, a_{22}, $\dots$, a_{nn} are then called the diagonal entries of A.
  - Upper triangular matrices : square matrices with nonzero entries only on and above the main diagonal.
  - Lower triangular matrices : nonzero entries only on and below the main diagonal.
  - Diagonal matrices : nonzero entries only on the main diagonal.
  - Identity matrice : diagonal and all diagonal entries are 1.
- Vectors: special matrices whose row or column number is one.
  - A row vector: $a = [a_1, a_2, \dots, a_n]$; its dimension is $1 \times n$.
  - A $m \times 1$ column vector:

$$b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

Example (Matrix and quadratic forms). We can use matrices to express general quadratic functions of vectors. For instance

$$f(x) = x^T A x + 2bx + c$$

is equivalent to

$$f(x) = \begin{bmatrix} x \\ 1 \end{bmatrix}^T \begin{bmatrix} A & b \\ b^T & c \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix}$$

1.1 Matrix addition and multiplication

The **sum** of two matrices A *and* B *(of the same size)* is

$$A + B = [a_{jk} + b_{jk}]$$

The **product** between a $m \times n$ matrix A and a scalar c is

$$cA = [ca_{jk}]$$

i.e. each entry of A is multiplied by c to generate the corresponding entry of cA.

The **matrix product** $C = AB$ is meaningful only if the column number of A equals the row number of B. The computation is done as shown in the following example:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ \boxed{a_{21}} & \boxed{a_{22}} & \boxed{a_{23}} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{bmatrix} \begin{bmatrix} \boxed{b_{11}} & b_{12} \\ \boxed{b_{21}} & b_{22} \\ \boxed{b_{31}} & b_{32} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ \boxed{c_{21}} & c_{22} \\ c_{31} & c_{32} \\ c_{41} & c_{42} \end{bmatrix}$$

where

$$\begin{aligned} c_{21} &= a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} \\ &= [a_{21}, a_{22}, a_{23}] \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix} \\ &= \text{"second row of } A\text{"} \times \text{"first column of } B\text{"} \end{aligned}$$

More generally:

$$\begin{aligned} c_{jk} &= a_{j1}b_{1k} + a_{j2}b_{2k} + \cdots + a_{jn}b_{nk} \\ &= [a_{j1}, a_{j2}, \ldots, a_{jn}] \begin{bmatrix} b_{1k} \\ b_{2k} \\ \vdots \\ b_{nk} \end{bmatrix} \end{aligned} \tag{3}$$

namely, the jk entry of C is obtained by multiplying each entry in the jth row of A by the corresponding entry in the kth column of B and then adding these n products. This is called a multiplication of rows into columns.

Matrix multiplication is not commutative: It is a good habit to always check the matrix dimensions when doing matrix products:

$$\begin{array}{cccc} A & B & = & C \\ [m \times n] & [n \times p] & & [m \times p] \end{array}$$

This way it is clear that AB in general does not equal to BA, e.g.,

$$ABC = (AB)\,C = A\,(BC) \neq BCA$$

Matrices as combination of vectors: The matrix-vector product

$$Ax = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 \\ a_{41}x_1 + a_{42}x_2 + a_{43}x_3 \end{bmatrix}$$

is nothing but the weighted sum of the columns of A:

$$Ax = \left[\begin{array}{c|c|c} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{array}\right] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \\ a_{41} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \\ a_{42} \end{bmatrix} + x_3 \begin{bmatrix} a_{13} \\ a_{23} \\ a_{33} \\ a_{43} \end{bmatrix}$$

1.2 Matrix transposition

Definition 1 (Transpose). The transpose of an $m \times n$ matrix

$$A = [a_{jk}] = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & \dots & \dots & a_{2n} \\ \vdots & \dots & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

is the $n \times m$ matrix A^T (reads "A transpose") defined as

$$A^T = [a_{kj}] = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & \dots & \dots & a_{m2} \\ \vdots & \dots & \dots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{bmatrix}$$

Transposition has the following rules:

- $\left(A^T\right)^T = A$
- $(A+B)^T = A^T + B^T$
- $(cA)^T = cA^T$
- $(AB)^T = B^T A^T$

If $A = A^T$, then A is called symmetric. If $A = -A^T$ then A is called skew-symmetric.

2 Linear systems of equations

A linear system of m equations in n unknowns $x_1, \ldots, x_n$ is a set of equations of the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots a_{2n}x_n &= b_2 \\ &\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots a_{mn}x_n &= b_m \end{aligned} \tag{4}$$

- Linearity: each variable x_j appears in the first power only.
- If all the b_j are zero, then the linear equation is called a homogeneous system. Otherwise, it is a nonhomogeneous system.
- Homogeneous systems always have at least the trivial solution $x_1 = x_2 = \cdots = x_n = 0$.

The m equations (4) can be written as a single vector equation

$$Ax = b$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & \ldots & a_{2n} \\ \vdots & \vdots & \ldots & \ldots & \vdots \\ a_{m1} & a_{m2} & \ldots & \ldots & a_{mn} \end{bmatrix}, \; x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix}, \; b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

Gauss[1] elimination is a systematic method to solve linear equations. Consider

$$\underbrace{\begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 0 & 10 & 25 \\ 20 & 10 & 0 \end{bmatrix}}_{A} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 \\ 0 \\ 90 \\ 80 \end{bmatrix}}_{b}$$

1. Obtain the augmented matrix of the system

$$\left[\, A \mid b \,\right] = \left[\begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ -1 & 1 & -1 & 0 \\ 0 & 10 & 25 & 90 \\ 20 & 10 & 0 & 80 \end{array}\right]$$

[1] Johann Carl Friedrich Gauss, 1777-1855, German mathematician: contributed significantly to many fields, including number theory, algebra, statistics, analysis, differential geometry, geodesy, geophysics, electrostatics, astronomy, Matrix theory, and optics.

Gauss was an ardent perfectionist. He was never a prolific writer, refusing to publish work which he did not consider complete and above criticism. Mathematical historian Eric Temple Bell estimated that, had Gauss published all of his discoveries in a timely manner, he would have advanced mathematics by fifty years.

2. Perform elementary row operation on the augmented matrix, to obtain the Row Echelon Form. Adding the first row to the second row gives

$$\text{pivot role}: \left[\begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ \boxed{-1} & \boxed{1} & \boxed{-1} & \boxed{0} \\ 0 & 10 & 25 & 90 \\ 20 & 10 & 0 & 80 \end{array}\right] \xrightarrow[\text{add pivot role}]{\text{row 2}} \underline{\underline{\left[\begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ \boxed{0} & \boxed{0} & \boxed{0} & \boxed{0} \\ 0 & 10 & 25 & 90 \\ 20 & 10 & 0 & 80 \end{array}\right]}}$$

$$\xrightarrow[\text{add -20}\times\text{pivot role}]{\text{row 4}} \underline{\underline{\left[\begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 10 & 25 & 90 \\ 0 & 30 & -20 & 80 \end{array}\right]}}$$

What we have done is using the pivot row to eliminate x_1 in the other equations. At this stage, the linear equations look like

$$x_1 - x_2 + x_3 = 0 \tag{5}$$
$$0 = 0 \tag{6}$$
$$10x_2 + 25x_3 = 90 \tag{7}$$
$$30x_2 - 20x_3 = 80 \tag{8}$$

Re-arranging yields

$$x_1 - x_2 + x_3 = 0 \tag{9}$$
$$10x_2 + 25x_3 = 90 \tag{10}$$
$$30x_2 - 20x_3 = 80 \tag{11}$$
$$0 = 0 \tag{12}$$

Moving on, we can get ride of x_2 in the third equation, by adding to it -3 times the second equation. Correspondingly in the augmented matrix, we have

$$\left[\begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ 0 & 10 & 25 & 90 \\ 0 & 30 & -20 & 80 \\ 0 & 0 & 0 & 0 \end{array}\right] \rightarrow \left[\begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ 0 & 10 & 25 & 90 \\ 0 & 0 & -95 & -190 \\ 0 & 0 & 0 & 0 \end{array}\right] \xrightarrow[\text{normalizing}]{} \underbrace{\left[\begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ 0 & 1 & 5/2 & 9 \\ 0 & 0 & 1 & 38/19 \\ 0 & 0 & 0 & 0 \end{array}\right]}_{\text{the row echelon form}}$$

namely

$$x_3 = 38/19$$
$$x_2 + x_3 = 9$$
$$x_1 - x_2 + x_3 = 0$$

The unknowns can now be readily obtained by back substitution: $x_3 = 38/19$, $x_2 = 9 - x_3$, $x_1 = x_2 - x_3$.

Elementary Row Operations for Matrices What we have done can be summarized by the following elementary matrix row operations:

- Interchange of two rows
- Addition of a constant multiple of one row to another row
- Multiplication of a row by a nonzero constant c

Let the final row echelon form be denoted by

$$\left[\begin{array}{c|c} R & f \end{array}\right]$$

We have

1. The two systems $Ax = b$ and $Rx = f$ are equivalent.
2. At the end of the Gauss elimination (before the back substitution), the row echelon form of the augmented matrix will be

$$\left[\begin{array}{cccccc|c} r_{11} & r_{12} & \dots & \dots & \dots & r_{1n} & f_1 \\ & r_{22} & \dots & \dots & \dots & r_{2n} & f_2 \\ & & \ddots & \dots & \dots & \vdots & \vdots \\ & & & r_{rr} & \dots & r_{rn} & f_r \\ & & & & & & f_{r+1} \\ & & & & & & \vdots \\ & & & & & & f_m \end{array}\right]$$

   where all unfilled entries are zero.
3. The number of nonzero rows, r, in the row-reduced coefficient matrix R is called the rank of R and also the rank of A.
4. Solution concepts:
   (a) *No solution* / system is inconsistent: r is less than m and $f_{r+1}, f_{r+2}, \dots, f_m$ are not all zero.
   (b) *Unique solution*: if the system is consistent and $r = n$, there is exactly one solution, which can be found by back substitution.
   (c) *Infinitely many solutions*: if $f_{r+1} = f_{r+2} = \dots = f_m = 0$. To obtain any of these solutions, choose values of $x_{r+1}, \dots, x_n$ arbitrarily. Then solve the r-th equation for x_r (in terms of those arbitrary values), then the $(r-1)$-st equation for x_{r-1}, and so on up the line.

3 Vector space, linear independence, basis, and span

Given a set of m vectors a_1, a_2, ..., a_m with the same size,

$$k_1 a_1 + k_2 a_2 + \cdots + k_m a_m$$

is called a linear combination of the vectors. If

$$a_1 = k_2 a_2 + k_3 a_3 + \cdots + k_m a_m$$

then a_1 is said to be *linearly dependent* on a_2, a_3, ..., a_m. The set

$$\{a_1, a_2, \ldots, a_m\} \tag{13}$$

is then a linearly dependent set. The same idea holds if a_2 or any vector in the set (13) is linearly dependent on others.

Generalizing, if

$$k_1 a_1 + k_2 a_2 + \cdots + k_m a_m = 0$$

holds if and only if

$$k_1 = k_2 = \cdots = k_m = 0$$

then the vectors in (13) are linearly dependent. This is saying that at least one of the vectors can be expressed as a linear combination of the other vectors.

Why is linear independence important? If a set of vectors is linearly dependent, then we can get rid of one or perhaps more of the vectors until we get a linearly independent set. This set is then the smallest "truly essential" set with which we can work.

Consider a set of n linearly independent vectors, a_1, a_2, ..., a_n, each with n components. All the possible linear combinations of a_1, a_2, ..., a_n form the vector space $\mathbb{R}^n$. This is the *span* of the n vectors.

Definition 2 (Basis). A *basis* of $\mathbf{V}$ is a set $\mathbf{B}$ of vectors in $\mathbf{V}$, such that any $v \in \mathbf{V}$ can be uniquely expressed as a finite linear combination of vectors in $\mathbf{B}$.

Example 3. In $\mathbb{R}^2$

$$v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \; v_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

is a linearly independent set and forms a basis.

$$v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \; v_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \; v_3 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

is not a linearly independent set.

4 Matrix properties

4.1 Rank

Definition 4 (Rank). The rank of a matrix A is the maximum number of linearly independent row or column vectors.

Theorem. *Row or column operations do not change the rank of a matrix.*

With the concept of linear dependence, many matrix-matrix operations can be understood from the view point of vector manipulations.

Example (Dyad). $A = uv^T$ is called a dyad, where u and v are vectors of proper dimensions. It is a rank 1 matrix, as can be seen that $A = uv^T$ is formed by linear combinations of the vector u, where the weights of the combinations are coefficients of v.

Fact. *For $A, B \in \mathbb{R}^{n\times n}$, if $rank\,(A) = n$ then $AB = 0$ implies $B = 0$. If $AB = 0$ but $A \neq 0$ and $B \neq 0$, then $rank\,(A) < n$ and $rank\,(B) < n$.*

4.2 Range and null spaces

Definition 5 (Range space). The range space of a matrix A, denoted as $\mathcal{R}\,(A)$, is the span of all the column vectors of A.

Definition 6 (Null space). The null space of a matrix $A \in \mathbb{R}^{n\times n}$, denoted as $\mathcal{N}\,(A)$, is the vector space

$$\{x \in \mathbb{R}^n : \; Ax = 0\}$$

The dimension of the null space is called *nullity* of the matrix.

Fact 7. *The following is true:*

$$\mathcal{N}\left(AA^T\right) = \mathcal{N}\left(A^T\right); \; \mathcal{R}\left(AA^T\right) = \mathcal{R}\,(A)$$

4.3 Determinants

Determinants were originally introduced for solving linear equations in the form of $Ax = y$, with a square A. They are cumbersome to compute for high-order matrices, but their definitions and concepts are partially very important.

We review only the computations of second- and third-order matrices

- 2×2 matrices:

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$$

- 3×3 matrices:

$$\det\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & k \end{bmatrix} = a\det\begin{bmatrix} e & f \\ h & k \end{bmatrix} - b\det\begin{bmatrix} d & f \\ g & k \end{bmatrix} + c\det\begin{bmatrix} d & e \\ g & h \end{bmatrix}$$
$$= aek + bfg + cdh - gec - bdk - ahf$$

where $\det\begin{bmatrix} e & f \\ h & k \end{bmatrix}$, $\det\begin{bmatrix} d & f \\ g & k \end{bmatrix}$, and $\det\begin{bmatrix} d & e \\ g & h \end{bmatrix}$ are called the minors of $\det\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & k \end{bmatrix}$.

Caution: $\det(cA) = c^n \det(A)$ (not $c\det(A)$!)

Theorem 8. *The determinant of A is nonzero if and only if A is full rank.*

You should be able to verify the theorem for 2×2 matrices. The proof will be immediate after introducing the concept of eigenvalues.

Definition 9. A linear transformation is called singular if the determinant of the corresponding transformation matrix is zero.

Fact 10. *Determinant facts:*

- *If A and B are square matrices, then*

$$\det(AB) = \det(BA) = \det A \det B$$
$$\det\left(A^T\right) = \det(A)$$
$$\det(A^*) = \det(A)$$

- *If X and Z are square, Y with compatible dimensions, then*

$$\det\left(\begin{bmatrix} X & Y \\ 0 & Z \end{bmatrix}\right) = \det X \det Z$$

5 Matrix and linear equations

Consider again, using now concepts in range and null spaces of matrices, the linear equations

$$Ax = y \tag{14}$$

- *Existence* of solutions requires that

$$y \in \mathcal{R}(A)$$

- The linear equation is called *overdetermined* if it has more equations than unknowns (i.e. A is a tall skinny matrix), *determined* if A is square, *undetermined* if it has fewer equations than unknowns (A is a wide matrix).

- *Solutions* of the above equation, provided that they exist, is constructed from

$$x = x_o + z : \; Az = 0 \tag{15}$$

  where x_0 is any (fixed) solution of (14) and z runs through all the homogeneous solutions of $Az = 0$, namely, z runs through all vectors in the null space of A.

- *Uniqueness* of a solution: if the null space of A is zero, the solution is unique.

You should be familiar with solving 2nd or 3rd-order linear equations by hand.

6 Eigenvector and eigenvalue

6.1 Matrix, mappings, and eigenvectors

Think of Ax this way: A defines a linear operator; Ax is a vector produced by feeding the vector x to this linear operator. In the two-dimensional case, we can look at Fig. 1. Certainly, Ax does not (at all) need to be in the same direction as x. An example is

$$A_0 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

which gives that

$$A_0 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ 0 \end{bmatrix}$$

namely, Ax is x projected on the first axis in the two-dimensional vector space, which will not be in the same direction as x as long as $x_2 \neq 0$.

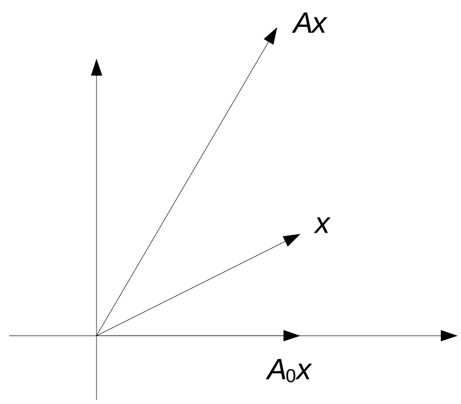


Figure 1: Example relationship between x and Ax

From here comes the concept of eigenvectors and eigenvalues. It says that there are certain "special directions/vectors" (denoted as v_1 and v_2 in our two-dimensional example) for A such that $Av_i = \lambda_i v_i$. Thus Av_i is on the same line as the original vector v_i, just scaled by the eigenvalue λ_i. It can be shown that if $\lambda_1 \neq \lambda_2$, then v_1 and v_2 are linearly independent (your homework). This is saying that any vector in $\mathbb{R}^2$ can be decomposed as

$$x = a_1 v_1 + a_2 v_2$$

Therefore

$$Ax = a_1 A v_1 + a_2 A v_2 = a_1 \lambda_1 v_1 + a_2 \lambda_2 v_2$$

Knowing λ_i and v_i thus can directly tell us how Ax looks like. More important, we have decomposed Ax into small modules that are from time to time more handy for analyzing the system properties. Figs. 2 and 3 demonstrate the above idea graphically.

Remark 11. The above geometric interpretations are for matrices with distinct real eigenvalues.

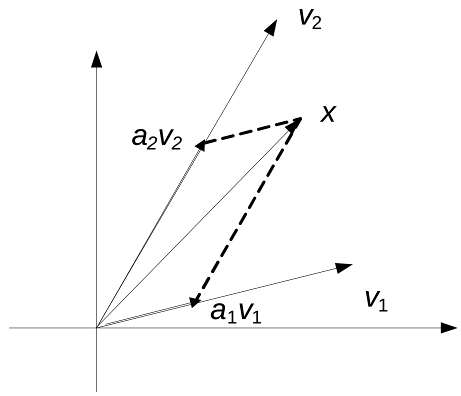


Figure 2: Decomposition of x

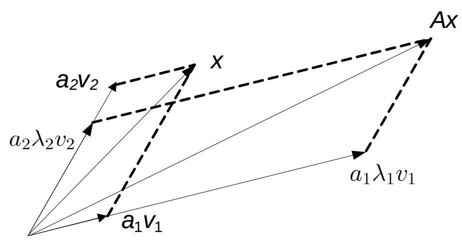


Figure 3: Construction of Ax

The geometric interpretation above makes eigenvalue a very important concept. *Eigenvalues* are also called *characteristic values* of a matrix. The set of all the eigenvalues of A is called the *spectrum* of A. The largest of the *absolute* values of the eigenvalues of A is called the *spectral radius* of A.

6.2 Computation of eigenvalue and eigenvectors

Formally, eigenvalue and eigenvector are defined as follows. For $A \in \mathbb{R}^{n\times n}$, an eigenvalue λ of A is one for which

$$Ax = \lambda x \tag{16}$$

has a nonzero solution $x \neq 0$. The corresponding solutions are called eigenvectors of A.

(16) is equivalent to

$$(A - \lambda I)\, x = 0 \tag{17}$$

As $x \neq 0$, the matrix $A - \lambda I$ must be singular, so

$$\det(A - \lambda I) = 0 \tag{18}$$

$\det(A - \lambda I)$ is a polynomial of λ, called the characteristic polynomial. Correspondingly, (18) is called the characteristic equation. So eigenvalues are roots of the characteristic equation. If an $n \times n$ matrix A has n eigenvalues $\lambda_1, \ldots, \lambda_n$, it must be that

$$\det(A - \lambda I) = (\lambda_1 - \lambda)\cdots(\lambda_n - \lambda)$$

After obtaining an eigenvalue λ, we can find the associated eigenvector by solving (17). This is nothing but solving a homogeneous system.

Example 12. Consider

$$A = \begin{bmatrix} -5 & 2 \\ 2 & -2 \end{bmatrix}$$

Then

$$\begin{aligned} \det(A - \lambda I) = 0 &\Rightarrow \det\left(\begin{bmatrix} -5-\lambda & 2 \\ 2 & -2-\lambda \end{bmatrix}\right) = 0 \\ &\Rightarrow (5+\lambda)(2+\lambda) - 4 = 0 \\ &\Rightarrow \lambda = -1 \text{ or } -6 \end{aligned}$$

So A has two eigenvalues: -1 and -6. The characteristic polynomial of A is $\lambda^2 + 7\lambda + 6$.

To obtain the eigenvector associated to $\lambda = -1$, we solve

$$(A - \lambda I)x = 0 \Leftrightarrow \left(\begin{bmatrix} -5 & 2 \\ 2 & -2 \end{bmatrix} + 1\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)x = \begin{bmatrix} -4 & 2 \\ 2 & -1 \end{bmatrix}x = 0$$

One solution is

$$x = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

As an exercise, show that an eigenvector associated to $\lambda = -6$ is $\begin{bmatrix} 2 & -1 \end{bmatrix}^T$.

Example 13 (Multiple eigenvectors)**.** Obtain the eigenvalues and eigenvectors of

$$A = \begin{bmatrix} -2 & 2 & -3 \\ 2 & 1 & -6 \\ -1 & -2 & 0 \end{bmatrix}$$

Analogous procedures give that

$$\lambda_1 = 5,\ \lambda_2 = \lambda_3 = -3$$

So there are repeated eigenvalues. For $\lambda_2 = \lambda_3 = -3$, the characteristic matrix is

$$A + 3I = \begin{bmatrix} 1 & 2 & -3 \\ 2 & 4 & -6 \\ -1 & -2 & 3 \end{bmatrix}$$

The second row is the first row multiplied by 2. The third row is the negative of the first row. So the characteristic matrix has only rank 1. The characteristic equation

$$(A - \lambda_2 I)x = 0$$

has two linearly independent solutions

$$\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}$$

Theorem 14 (Eigenvalue and determinant). *Let* $A \in \mathbb{R}^{n\times n}$*. Then*

$$\det A = \prod_{i=1}^{n} \lambda_i$$

Proof. Letting $\lambda = 0$ in the characteristic polynomial

$$p(\lambda) = \det(A - \lambda I) = (\lambda_1 - \lambda)(\lambda_2 - \lambda)\dots$$

gives

$$\det(A) = p(0) = \prod_{i=1}^{n} \lambda_i$$

□

Example 15. For the two-dimensional case

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \Rightarrow p(\lambda) = \det(A - \lambda I) = (a_{11} - \lambda)(a_{22} - \lambda) - a_{12}a_{21}$$

On the other hand

$$p(\lambda) = (\lambda_1 - \lambda)(\lambda_2 - \lambda)$$

Matching the coefficients we get

$$\begin{aligned} \lambda_1 + \lambda_2 &= a_{11} + a_{22} \\ \lambda_1\lambda_2 &= a_{11}a_{22} - a_{12}a_{21} \end{aligned}$$

6.3 Eigenbases and diagonalization

Eigenvectors of an $n \times n$ matrix A may (or may not!) form a basis for $\mathbb{R}^n$. If we are interested in a transformation $y = Ax$, such an "eigenbasis" (basis of eigenvectors), if exists, is of great advantage because then we can represent any x in $\mathbb{R}^n$ uniquely as a linear combination of the eigenvectors $x_1, \ldots, x_n$, say, $x = c_1x_1 + c_2x_2 + \ldots + c_nx_n$. And, denoting the corresponding (not necessarily distinct) eigenvalues of the matrix A by $\lambda_1, \ldots, \lambda_n$, we have $Ax_j = \lambda_j x_j$, so that we simply obtain

$$\begin{aligned} y = Ax &= A\left(c_1x_1 + c_2x_2 + \ldots + c_nx_n\right) \\ &= c_1Ax_1 + c_2Ax_2 + \cdots + c_nAx_n \\ &= c_1\lambda_1x_1 + \cdots + c_n\lambda_nx_n \end{aligned}$$

This shows that we have decomposed the complicated action of A on an arbitrary vector x into a sum of simple actions (multiplication by scalars) on the eigenvectors of A.

Theorem 16 (Basis of Eigenvectors). *If an $n \times n$ matrix A has n distinct eigenvalues, then A has a basis of eigenvectors $x_1, \ldots, x_n$ for $\mathbb{R}^n$.*

Proof. We just need to prove that the n eigenvectors are linearly independent. If not, reorder the eigenvectors and suppose r of them, $\{x_1, x_2, \ldots, x_r\}$, are linearly independent and $x_{r+1}, \ldots, x_n$ are linearly dependent on $\{x_1, x_2, \ldots, x_r\}$. Consider x_{r+1}. There must exist $c_1, \ldots c_{n+1}$, not all zero, such that

$$c_1x_1 + \ldots c_{r+1}x_{r+1} = 0 \tag{19}$$

Multiplying A on both sides yields

$$c_1Ax_1 + \ldots c_{r+1}Ax_{r+1} = 0$$

Using $Ax_i = \lambda_i x_i$, we have

$$c_1\lambda_1x_1 + \cdots + c_{r+1}\lambda_{r+1}x_{r+1} = 0$$

But from (19) we know

$$c_1\lambda_{r+1}x_1 + \ldots c_{r+1}\lambda_{r+1}x_{r+1} = 0$$

Subtracting the last two equations gives

$$c_1\left(\lambda_1 - \lambda_{r+1}\right)x_1 + \cdots + c_r\left(\lambda_r - \lambda_{r+1}\right)x_r = 0$$

None of $\lambda_1 - \lambda_{r+1}, \ldots, \lambda_r - \lambda_{r+1}$ are zero, as the eigenvalues are distinct. Hence not all coefficients $c_1\left(\lambda_1 - \lambda_{r+1}\right), \ldots, c_r\left(\lambda_r - \lambda_{r+1}\right)$ are zero. Thus $\{x_1, x_2, \ldots, x_r\}$ is not linearly independent–a contradiction with the assumption at the beginning of the proof. □

Theorem 16 provides an important decomposition–called diagonalization–of matrices. To show that, we briefly review the concept of matrix inverses first.

Definition 17 (Matrix Inverse). The inverse A^{-1} of a square matrix A satisfies

$$AA^{-1} = A^{-1}A = I$$

If A^{-1} exists, A is called nonsingular; otherwise, A is singular.

Theorem 18 (Diagonalization of a Matrix). *Let an $n \times n$ matrix A have a basis of eigenvectors $\{x_1, x_2, \ldots, x_n\}$, associated to its n distinct eigenvectors $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$, respectively. Then*

$$A = XDX^{-1} = [x_1, x_2, \ldots, x_n] \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \lambda_n \end{bmatrix} [x_1, x_2, \ldots, x_n]^{-1} \tag{20}$$

Also,

$$A^m = XD^mX^{-1}, \quad (m = 2, 3, \ldots). \tag{21}$$

Remark 19. From (21), you can find some intuition about the benefit of (20): A^m can be tedious to compute while D^m is very simple!

Proof. From Theorem 16, the n linearly independent eigenvectors of A form a basis. Write

$$\begin{aligned} Ax_1 &= \lambda_1 x_1 \\ Ax_2 &= \lambda_2 x_2 \\ &\vdots \\ Ax_n &= \lambda_n x_n \end{aligned}$$

as

$$A[x_1, x_2, \ldots, x_n] = [x_1, x_2, \ldots, x_n] \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \lambda_n \end{bmatrix}$$

The matrix $[x_1, x_2, \ldots, x_n]$ is square. Linear independence of the eigenvectors implies that $[x_1, x_2, \ldots, x_n]$ is invertible. Multiplying $[x_1, x_2, \ldots, x_n]^{-1}$ on both sides gives (20).

(21) then immediately follows, as

$$A^m = \left(XDX^{-1}\right)^m = XDX^{-1}XDX \ldots XDX^{-1} = XD^mX^{-1}$$

□

Example 20. Let

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

The matrix has eigenvalues at 1 and -1, with associated eigenvectors

$$\begin{bmatrix} 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Then

$$X = \begin{bmatrix} 3 & 1 \\ 1 & 1 \end{bmatrix}, \ A = X \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} X^{-1}$$

Now if we are to compute A^{3000}. We just need to do

$$A^{3000} = X \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}^{3000} X^{-1} = I$$

7 Similarity transformation

Definition 21 (Similar Matrices. Similarity Transformation)**.** An $n \times n$ matrix $\hat{A}$ is called similar to an $n \times n$ matrix A if

$$\hat{A} = T^{-1}AT$$

for some **nonsingular** $n \times n$ matrix T. This transformation, which gives $\hat{A}$ from A, is called a similarity transformation.

Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be two vector spaces of the same dimension. Take the *same* point P. Let u be its coordinate in $\mathcal{S}_1$ and $\hat{u}$ be its coordinate in $\mathcal{S}_2$. These coordinates in the two vector spaces are related by some linear transformation T:

$$u = T\hat{u}, \ \hat{u} = T^{-1}u$$

Consider Fig. 4. Let the point P go through a linear transformation A in the vector space $\mathcal{S}_1$ to generate an output point P_o. P_o is physically the same point in both $\mathcal{S}_1$ and $\mathcal{S}_2$. However, the coordinates of P_o are different: if we see it from "standing inside" $\mathcal{S}_1$, then

$$y = Au$$

If we see it in $\mathcal{S}_2$, then the coordinate is some other value $\hat{y}$.

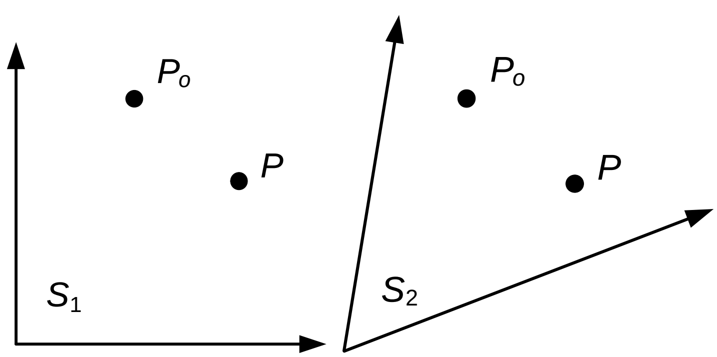


Figure 4: Same points in different vector spaces

How does the linear transformation A mathematically "look like" in $\mathcal{S}_2$?
Result:

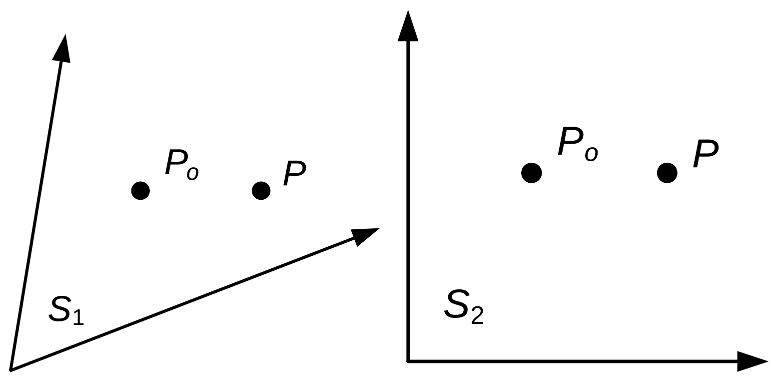
$$\hat{y} = T^{-1}y = T^{-1}Au = \left(T^{-1}AT\right)\hat{u}$$

namely, the linear transformation, viewed from $\mathcal{S}_2$, is

$$\hat{A} = T^{-1}AT$$

It is central to recognize that the physical operation is the same: P goes to another point P_o. Different is our perspective of viewing this transformation. $\hat{A}$ and A are in this sense called similar.

Purpose of doing similarity transformation: $\hat{A}$ can be simpler! Consider, for instance, the following example



In $\mathcal{S}_1$, the transformation changes both coordinates of P while in $\mathcal{S}_2$, only the first coordinate of P is changed.

Theorem 22 (Eigenvalues and Eigenvectors of Similar Matrices). *If $\hat{A}$ is similar to A, then $\hat{A}$ has the same eigenvalues as A. Furthermore, if x is an eigenvector of A, then $y = T^{-1}x$ is an eigenvector of $\hat{A}$ corresponding to the same eigenvalue.*

□

8 Matrix inversion

This section provides a more detailed description of matrix inversion. Recall that the inverse A^{-1} of a square nonsingular matrix A satisfies

$$AA^{-1} = A^{-1}A = I$$

Theorem 23 (Inverse is unique). *If A has an inverse, the inverse is unique.*

Concepts only. If both B and C are inverses of A, then $BA = AB = I$ and $CA = AC = I$ so that

$$B = IB = (CA)\,B = CAB = C\,(AB) = CI = C$$

Connection with previous topics: The set of all $n \times n$ matrices is not a field. Multiplicative inverse is unique. □

Definition 24 (Existence of a matrix inverse). The inverse A^{-1} of an $n \times n$ matrix A exists if and only if the rank of A is n. Hence A is nonsingular if rank$(A) = n$, and singular if rank$(A) < n$.

Proof. Let $A \in \mathbb{R}^{n\times n}$ and consider the linear equation

$$Ax = b$$

If A^{-1} exists, then

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

Hence $A^{-1}b$ is a solution to the linear equation. It is also unique. If not, then take another solution u; we should have $Au = b$ and $u = A^{-1}b$. Since A^{-1} is unique, it must be that $u = x$.

Conversely, if A has rank n. Then we can solve $Ax = b$ uniquely by Gauss elimination, to get

$$x = Bb$$

where B is the backward substitution linear transformation in Gauss elimination. Hence

$$Ax = A\left(Bb\right) = \left(AB\right)b = Ib$$

for any b. Hence

$$AB = I$$

Similarly, substituting $Ax = b$ into $x = Bb$ gives

$$x = B\left(Ax\right) = \left(BA\right)x = Ix$$

and hence

$$BA = I$$

Together $B = A^{-1}$ exists. □

There are several ways to compute the inverse of a matrix. One approach for low-order matrices is the method of using adjugate matrix (sometimes also called adjoint matrix):

$$A^{-1} = \frac{1}{\det\left(A\right)}\text{adj}\left(A\right)^T$$

We explain the computation by two examples. You can find additional details in your undergraduate linear algebra course.

- 2×2 example:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc}\begin{bmatrix} \left(-1\right)^{1+1}d & \left(-1\right)^{1+2}b \\ \left(-1\right)^{2+1}c & \left(-1\right)^{2+2}a \end{bmatrix}$$

  where b in $\left(-1\right)^{1+2}b$ is obtained by:

  - noticing b is at row 1 column 2 of A;
  - looking at the element at row 2 column 1 of A (notice the transpose in adj $\left(A\right)^T$);

  - constructing a submatrix of A by removing row 2 and column 1 from it, i.e., $[b]$ in this 2×2 example;
  - computing the determinant of this submatrix.
  - adding $(-1)^{1+2}$ as a scalar

- 3×3 example:

$$A^{-1}=\begin{bmatrix} a & b & c\\ d & e & f\\ g & h & k\end{bmatrix}^{-1}=\frac{1}{\det A}\begin{bmatrix} \begin{vmatrix} e & f\\ h & k\end{vmatrix} & -\begin{vmatrix} b & c\\ h & k\end{vmatrix} & \begin{vmatrix} b & c\\ e & f\end{vmatrix}\\ -\begin{vmatrix} d & f\\ g & k\end{vmatrix} & \begin{vmatrix} a & c\\ g & k\end{vmatrix} & -\begin{vmatrix} a & c\\ d & f\end{vmatrix}\\ \begin{vmatrix} d & e\\ g & h\end{vmatrix} & -\begin{vmatrix} a & b\\ g & h\end{vmatrix} & \begin{vmatrix} a & b\\ d & e\end{vmatrix}\end{bmatrix}$$

where $|\cdot|$ denotes the determinant of a matrix. Similar as before, the row 1 column 2 element $-\begin{vmatrix} b & c\\ h & k\end{vmatrix}$ is obtained via

$$(-1)^{2+1}\det\left(A \text{ with } [d,e,f],\begin{bmatrix} a\\ d\\ g\end{bmatrix} \text{ removed}\right)$$

Example 25. Find the inverse matrices of

$$A=\begin{bmatrix} 3 & 1\\ 2 & 4\end{bmatrix},\ B=\begin{bmatrix} -1 & 1 & 2\\ 3 & -1 & 1\\ -1 & 3 & 4\end{bmatrix},\ C=\begin{bmatrix} -0.5 & 0 & 0\\ 0 & 4 & 0\\ 0 & 0 & 1\end{bmatrix}$$

The answers are:

$$A^{-1}=\begin{bmatrix} 0.4 & -0.1\\ -0.2 & 0.3\end{bmatrix},\ B^{-1}=\begin{bmatrix} -0.7 & 0.2 & 0.3\\ -1.3 & -0.2 & 0.7\\ -1 & 3 & 4\end{bmatrix},\ C^{-1}=\begin{bmatrix} -2 & 0 & 0\\ 0 & 0.25 & 0\\ 0 & 0 & 1\end{bmatrix}$$

The related MATLAB command for matrix inversion is *inv()*.

Theorem 26. *Inverse of products of matrices can be obtained from inverses of each factor:*

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

and more generally

$$(AB\dots YZ)^{-1}=Z^{-1}Y^{-1}\dots B^{-1}A^{-1} \tag{22}$$

Proof. By definition $(AB)(AB)^{-1} = I$. Multiplying A^{-1} on both sides from the left gives

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

Now multiplying the result by B^{-1} on both sides from the left, we get

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

The general case (22) follows by induction. □

Fact 27. **Inverse of upper (lower) triangular matrices are upper (lower) triangular*

Proof. (main idea) We can either use the adjoint matrix method or use the following decomposition of upper(lower) triangular matrices

$$A = D(I + N)$$

where D is diagonal and N is strictly upper (lower) triangular with zeros diagonal elements. Then using matrix Taylor expansion we have

$$\begin{aligned} A^{-1} &= (I + N)^{-1} D^{-1} \\ &= \left(I - N + N^2 - N^3 + N^4 - \dots\right) D^{-1} \end{aligned}$$

N is nilpotent: N^k are upper (lower) triangular and $N^n = 0$ for n larger than the row dimension of A. D^{-1} is diagonal. Hence A^{-1} is upper (lower) triangular. □

8.1 Block matrix decomposition and inversion

Consider

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

Recall the key step in performing row operations on matrices in Gauss elimination:

$$\begin{bmatrix} 3 & 4 \\ 1 & 2 \end{bmatrix} \rightarrow \begin{bmatrix} 3 & 4 \\ 0 & 2/3 \end{bmatrix}$$

where we had substracted one third of the first row in the second row. In matrix representations, the above looks like

$$\begin{bmatrix} 1 & 0 \\ -1/3 & 1 \end{bmatrix} \begin{bmatrix} 3 & 4 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 4 \\ 0 & 2/3 \end{bmatrix}$$

For more general two by two matrices, we have

$$\begin{bmatrix} 1 & 0 \\ -ca^{-1} & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ 0 & d - ca^{-1}b \end{bmatrix}$$

If we want to keep the second row unchanged and simplify the first row, we can do

$$\begin{bmatrix} 1 & -bd^{-1} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a - bd^{-1}c & 0 \\ c & d \end{bmatrix}$$

Generalizing the concept to blok matrices (with compatible dimensions), we have

$$\begin{bmatrix} I & 0 \\ -B^TA^{-1} & I \end{bmatrix} \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} = \begin{bmatrix} A & B \\ 0 & C - B^TAB \end{bmatrix}$$

and

$$\begin{bmatrix} A & B \\ 0 & C - B^TAB \end{bmatrix} \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & C - B^TAB \end{bmatrix}$$

Thus

$$\begin{bmatrix} I & 0 \\ -B^TA^{-1} & I \end{bmatrix} \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & C - B^TAB \end{bmatrix}$$

Inversion is now very easy:

$$\left\{ \begin{bmatrix} I & 0 \\ -B^TA^{-1} & I \end{bmatrix} \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} \right\}^{-1} = \begin{bmatrix} A & 0 \\ 0 & C - B^TAB \end{bmatrix}^{-1}$$

$$\Longrightarrow \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}^{-1} \begin{bmatrix} I & 0 \\ -B^TA^{-1} & I \end{bmatrix}^{-1} = \begin{bmatrix} A & 0 \\ 0 & C - B^TAB \end{bmatrix}^{-1}$$

and hence

$$\begin{aligned} \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}^{-1} &= \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & C - B^TAB \end{bmatrix}^{-1} \begin{bmatrix} I & 0 \\ -B^TA^{-1} & I \end{bmatrix} \\ &= \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} \begin{bmatrix} A^{-1} & 0 \\ 0 & \left(C - B^TAB\right)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -B^TA^{-1} & I \end{bmatrix} \end{aligned}$$

The above steps work for general partitioned 2 by 2 matrices as well. The result is as follows

$$\begin{aligned} \begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} I & -BA^{-1} \\ 0 & I \end{bmatrix} &= \begin{bmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{bmatrix} \\ \begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} &= \begin{bmatrix} I & -BA^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{bmatrix}^{-1} \begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix} \end{aligned}$$

or

$$\begin{aligned} \begin{bmatrix} I & -BD^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} I & 0 \\ -D^{-1}C & I \end{bmatrix} &= \begin{bmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{bmatrix} \\ \begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} &= \begin{bmatrix} I & -BD^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{bmatrix}^{-1} \begin{bmatrix} I & 0 \\ -D^{-1}C & I \end{bmatrix} \end{aligned}$$

8.2 *LU and Cholesky decomposition

Fact 28. *The following is true for upper and lower triangular matrices:*

$$\begin{bmatrix} I & 0 \\ M & I \end{bmatrix}^{-1} = \begin{bmatrix} I & 0 \\ -M & I \end{bmatrix}$$

$$\begin{bmatrix} I & M \\ 0 & I \end{bmatrix}^{-1} = \begin{bmatrix} I & -M \\ 0 & I \end{bmatrix}$$

From the last section

$$\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} I & -BA^{-1} \\ 0 & I \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{bmatrix}$$

Applying Fact 28 to the last equation gives the *block LU decomposition*:

$$\begin{aligned} \begin{bmatrix} A & B \\ C & D \end{bmatrix} &= \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix} \\ &= \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & B \\ 0 & D - CA^{-1}B \end{bmatrix} \end{aligned}$$

which shows *any square matrix can be decomposed into the product of a lower triangular matrix and an upper triangular matrix.*

There is also *block Cholesky decomposition*

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I \\ CA^{-1} \end{bmatrix} A \begin{bmatrix} I & A^{-1}B \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & D - CA^{-1}B \end{bmatrix}$$

or using half matrices

$$\begin{aligned} \begin{bmatrix} A & B \\ C & D \end{bmatrix} &= \begin{bmatrix} A^{\frac{1}{2}} \\ CA^{-\frac{*}{2}} \end{bmatrix} \begin{bmatrix} A^{\frac{*}{2}} & A^{-\frac{1}{2}}B \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & Q^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & Q^{\frac{*}{2}} \end{bmatrix} \\ Q &= D - CA^{-1}B \end{aligned}$$

where

$$A^{\frac{1}{2}} A^{\frac{*}{2}} = A, \; Q^{\frac{1}{2}} Q^{\frac{*}{2}} = Q$$

hence

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = LU$$

where

$$LU = \begin{bmatrix} A^{\frac{1}{2}} & 0 \\ CA^{-\frac{*}{2}} & 0 \end{bmatrix} \begin{bmatrix} A^{\frac{*}{2}} & A^{-\frac{1}{2}}B \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & Q^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & Q^{\frac{*}{2}} \end{bmatrix} = \begin{bmatrix} A^{\frac{1}{2}} & 0 \\ CA^{-\frac{*}{2}} & Q^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} A^{\frac{*}{2}} & A^{-\frac{1}{2}}B \\ 0 & Q^{\frac{*}{2}} \end{bmatrix}$$

8.3 Determinant and matrix inverse identity

Although $AB \neq BA$ in general, the determinants of products have the following property:

$$\det(AB) = \det(BA) = \det A \det B$$

where A and B should be square to start with.

Theorem 29 (Sylvester's determinant theorem)**.** *For $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times m}$,*

$$\det(I_m + AB) = \det(I_n + BA)$$

where I_m and I_n are the $m \times m$ and $n \times n$ identity matrices, respectively.

Proof. Construct

$$M = \begin{bmatrix} I_m & -A \\ B & I_n \end{bmatrix}$$

From the decomposition

$$M = \begin{bmatrix} I_m & 0 \\ B & I_n \end{bmatrix} \begin{bmatrix} I_m & -A \\ 0 & I_n + BA \end{bmatrix}$$

we have

$$\det M = \det (I_n + BA)$$

Alternatively

$$M = \begin{bmatrix} I_m + AB & -A \\ 0 & I_n \end{bmatrix} \begin{bmatrix} I_m & 0 \\ B & I_n \end{bmatrix}$$

hence

$$\det M = \det (I_m + AB)$$

Therefore

$$\det (I_m + AB) = \det M = \det (I_n + BA)$$

□

More generally, for any invertible $m \times m$ matrix X

$$\det (X + AB) = \det (X) \det \left(I_n + BX^{-1}A\right)$$

which comes from

$$\begin{aligned} X + AB &= X\left(I + X^{-1}AB\right) \\ \Rightarrow \det (X + AB) &= \det \left[X\left(I + X^{-1}AB\right)\right] = \det X \det \left(I + X^{-1}AB\right) \end{aligned}$$

8.4 Matrix inversion lemma

Fact 30 (Matrix inversion lemma). *Assume A is nonsingular and $(A + BC)^{-1}$ exists. The following is true*

$$(A + BC)^{-1} = A^{-1}\left(I - B\left(CA^{-1}B + I\right)^{-1} CA^{-1}\right) \tag{23}$$

Proof. Consider

$$(A + BC)\, x = y$$

We aim at getting $x = (*)\, y$, where $(*)$ will be our $(A + BC)^{-1}$. First, let

$$Cx = d$$

We have

$$\begin{aligned} Ax + Bd &= y \\ Cx - d &= 0 \end{aligned}$$

Solving the first equation yields

$$x = A^{-1}(y - Bd)$$

Then

$$CA^{-1}(y - Bd) = d$$

gives

$$d = \left(CA^{-1}B + I\right)^{-1} CA^{-1}y$$

Hence

$$\begin{aligned} x &= A^{-1}\left(y - B\left(CA^{-1}B + I\right)^{-1} CA^{-1}y\right) \\ &= A^{-1}\left(I - B\left(CA^{-1}B + I\right)^{-1} CA^{-1}\right) y \end{aligned}$$

and (23) follows. □

Exercise 31. The matrix inversion lemma is a powerful tool useful for many applications. One application in adaptive control and system identification uses

$$\left(A + \phi\phi^T\right)^{-1} = A^{-1}\left(I - \frac{\phi\phi^T A^{-1}}{\phi^T A^{-1}\phi + 1}\right)$$

Prove the above result. Prove also the general case (called rank one update):

$$\left(A + bc^T\right) = A^{-1} - \frac{1}{1 + c^T A^{-1} b}\left(A^{-1}b\right)\left(c^T A^{-1}\right)$$

Fact 32 (More extended matrix inversion lemma). *Assume A, C, and $A + BCB^T$ are nonsingular. The following is true*

$$\left(A + BCB^T\right)^{-1} = A^{-1}\left(I - B\left(CB^T A^{-1}B + I\right)^{-1} CB^T A^{-1}\right) \quad (24)$$

$$= A^{-1} - A^{-1}B\left(CB^T A^{-1}B + I\right)^{-1} CB^T A^{-1} \quad (25)$$

$$= A^{-1} - A^{-1}B\left(B^T A^{-1}B + C^{-1}\right)^{-1} B^T A^{-1} \quad (26)$$

Proof. Consider

$$\left(A + BCB^T\right) x = y$$

We aim at getting $x = (*)\, y$, where $(*)$ will be our $\left(A + BCB^T\right)^{-1}$. First, let

$$CB^T x = d$$

We have

$$\begin{aligned} Ax + Bd &= y \\ CB^T x - d &= 0 \end{aligned}$$

Solving the first equation yields

$$x = A^{-1}(y - Bd)$$

Then

$$CB^T A^{-1}(y - Bd) = d$$

gives

$$d = \left(CB^T A^{-1} B + I\right)^{-1} CB^T A^{-1} y$$

Hence

$$\begin{aligned} x &= A^{-1}\left(y - B\left(CB^T A^{-1} B + I\right)^{-1} CB^T A^{-1} y\right) \\ &= A^{-1}\left(I - B\left(CB^T A^{-1} B + I\right)^{-1} CB^T A^{-1}\right) y \end{aligned}$$

and (24) follows. □

The extended matrix inversion lemma is key in transforming the Kalman filter to the information filter when inverting the innovation of covariance matrices.

8.5 Special inverse equalities

Fact 33. *The following matrix equalities are true*

- $(I + GK)^{-1} G = G(I + KG)^{-1}$
  to prove the result, start with $G(I + KG) = (I + GK)G$
- $GK(I + GK)^{-1} = G(I + KG)^{-1} K = (I + GK)^{-1} GK$ (the proof uses the first equality twice)
- generalization 1: $(\sigma^2 I + GK)^{-1} G = G(\sigma^2 I + KG)^{-1}$
- generalization 2: if M is invertible then $(M + GK)^{-1} G = M^{-1} G(I + KM^{-1}G)^{-1}$

Exercise 34. Check validity of the following equality, assuming proper dimensions and invertibility of matrices:

- $Z(I + Z)^{-1} = I - (I + Z)^{-1}$
- $(I + XY)^{-1} = I - XY(I + XY)^{-1} = I - X(I + YX)^{-1} Y$
- extension

$$\begin{aligned} \left(I + XZ^{-1}Y\right)^{-1} &= I - XZ^{-1}Y\left(I + XZ^{-1}Y\right)^{-1} = I - XZ^{-1}\left(I + YXZ^{-1}\right)^{-1} Y \\ &= I - X(Z + YX)^{-1} Y \end{aligned}$$

9 Spectral mapping theorem

Theorem 35 (Spectral Mapping Theorem). *Take any $A \in \mathbb{C}^{n\times n}$ and a polynomial (in s) $f(s)$ (more generally, analytic functions). Then*

$$\text{eig}\,(f(A)) = f(\text{eig}\,(A))$$

Proof. Let

$$f(A) = x_0 I + x_1 A + x_2 A^2 + \dots$$

Let λ be an eigenvalue of A. We first observe that λ^n is an eigenvalue of A^n. This can be seen from $\det(\lambda^n I - A^n) = \det[(\lambda I - A)\, p(A)] = \det(\lambda I - A)\det(p(A))$ where $p(A)$ is a polynomial of A.

Now consider $f(\lambda) = x_0 + x_1\lambda + x_2\lambda^2 + \dots$.

$$\begin{aligned}\det(f(\lambda) I - f(A)) &= \det\left[x_1(\lambda I - A) + x_2\left(\lambda^2 I - A^2\right) + x_3\left(\lambda^3 I - A^3\right) + \dots\right]\\ &= \det[(\lambda I - A)\, q(A)]\\ &= \det(\lambda I - A)\det(q(A))\end{aligned}$$

Hence $f(\lambda)$ is an eigenvalue of $f(A)$.

Conversely, if γ is an eigenvalue of $f(A)$, we need to prove that γ is in the form of $f(\lambda)$. Factorize the polynomial

$$f(\lambda) - \gamma = a_0(\lambda - \alpha_1)(\lambda - \alpha_2)\dots(\lambda - \alpha_n)$$

On the other hand, we note that as a matrix polynomial with the same coefficients:

$$f(A) - \gamma I = a_0(A - \alpha_1 I)(A - \alpha_2 I)\dots(A - \alpha_n I)$$

But $\det(f(A) - \gamma I) = 0$, which means that there is at least one α_i such that

$$\det(A - \alpha_i I) = 0$$

which says that α_i is an eigenvalue of A. Hence

$$f(\lambda) - \gamma = a_0(\lambda - \alpha_i)\prod_{k\neq i}(\lambda - \alpha_k) = 0$$

i.e.

$$\gamma = f(\lambda)$$

where λ is an eigenvalue of A. □

Example 36. Compute the eigenvalues of

$$A = \begin{bmatrix} 99.8 & 2000 \\ -2000 & 99.8 \end{bmatrix}$$

Solution:

$$A = 99.8I + 2000\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

So

$$\text{eig}(A) = 99.8 + 2000\,\text{eig}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = 99.8 \pm 2000i$$

10 Matrix exponentials

Since the Taylor series

$$e^{st} = 1 + st + \frac{s^2t^2}{2!} + \frac{s^3t^3}{3!} + \dots$$

converges everywhere, we can define the exponential of a matrix $A \in \mathcal{C}^{n\times n}$ by

$$e^{At} = I + At + \frac{A^2t^2}{2!} + \frac{A^3t^3}{3!} + \dots$$

Fact 37. *Properties of matrix exponentials*

1. $e^{A0} = I$
2. $e^{A(t+s)} = e^{At}e^{As}$
3. *If* $AB = BA$ *then* $e^{(A+B)t} = e^{At}e^{Bt} = e^{Bt}e^{At}$
4. $\det\left(e^{At}\right) = e^{\textit{trace}(A)t}$
5. e^{At} *is nonsingular for all* $t \in \mathcal{R}$ *and* $\left(e^{At}\right)^{-1} = e^{-At}$
6. e^{At} *is the unique solution* X *of the linear system of ordinary differential equations*

$$\dot{X} = AX, \text{ subject to } X(0) = I$$

11 Inner product

11.1 Inner product spaces

Basics: Inner product, or dot product, brings a metric for vector lengths. It takes two vectors and generates a number. In $\mathbb{R}^n$, we have

$$\langle a, b\rangle \triangleq a^T b = [a_1, a_2, \dots, a_n] \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

Clearly, $\langle a, b\rangle \triangleq a^T b = \langle b, a\rangle$. Letting $b = a$ above, we get the square of the length of a:

$$||a|| = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2}$$

Formal definitions:

Definition 38. A real vector space $\mathbf{V}$ is called a real inner product space, if for any vectors a and b in $\mathbf{V}$ there is an associated real number $\langle a, b\rangle$, called the inner product of a and b, such that the following axioms hold:

- (linearity) For all scalars q_1 and q_2 and all vectors $a, b, c \in \mathbf{V}$
$$\langle q_1 a + q_2 b, c\rangle = q_1 \langle a, b\rangle + q_2 \langle b, c\rangle$$
- (symmetry) $\forall a, b \in \mathbf{V}$
$$\langle a, b\rangle = \langle b, a\rangle$$
- (positive definiteness) $\forall a \in \mathbf{V}$
$$\langle a, a\rangle \geq 0$$
where $\langle a, a\rangle = 0$ if and only if $a = 0$.

Definition 39 (2-norm of vectors)**.** The length of a vector in $\mathbf{V}$ is defined by

$$||a|| = \sqrt{\langle a, a\rangle} \geq 0$$

For $\mathbb{R}^n$,

$$||a|| = \sqrt{a^T a} = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2}$$

This is the Euclidean norm or 2-norm of the vector. $\mathbb{R}^n$ equiped with the inner product $\langle a, b\rangle = \sqrt{a^T b}$ is called the n-dimensional Euclidean space.

Example 40 (Inner product for functions, function spaces). The set of all real-valued continuous functions $f(x)$, $g(x)$, $\ldots$ $x \in [\alpha, \beta]$ is a real vector space under the usual addition of functions and multiplication by scalars. An inner product on this function space is

$$\langle f, g \rangle = \int_{\alpha}^{\beta} f(x)\, g(x)\, dx$$

and the norm of f is

$$||f(x)|| = \sqrt{\int_{\alpha}^{\beta} f(x)^2\, dx}$$

Inner products is also closely related to the geometric relationships between vectors. For the two-dimensional case, it is readily seen that

$$v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix},\ v_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

is a basis of the vector space. The two vectors are additionally orthogonal, by direct observation.

More generally, we have:

Definition 41 (Orthogonal vectors). Vectors whose inner product is zero are called orthogonal.

Definition 42 (Orthonormal vectors). Orthogonal vectors with unity norm is called orthonormal.

Definition 43. The angle between two vectors is defined by

$$\cos\angle(a, b) = \frac{\langle a, b \rangle}{||a|| \cdot ||b||} = \frac{\langle a, b \rangle}{\sqrt{\langle a, a \rangle} \cdot \sqrt{\langle b, b \rangle}}$$

11.2 Trace (standard matrix inner product)

The trace of an $n \times n$ matrix $A = [a_{jk}]$ is given by

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} a_{ii} \tag{27}$$

Trace defines the so-called **matrix inner product**:

$$\langle A, B \rangle = \mathrm{Tr}\left(A^T B\right) = \mathrm{Tr}\left(B^T A\right) = \langle B, A \rangle \tag{28}$$

which is closely related to vector inner products. Take an example in $\mathbb{R}^{3\times 3}$: write the matrices in the column-vector form $B = [\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3]$, $A = [\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3]$, then

$$A^T B = \begin{bmatrix} \mathbf{a}_1^T \mathbf{b}_1 & * & * \\ * & \mathbf{a}_2^T \mathbf{b}_2 & * \\ * & * & \mathbf{a}_3^T \mathbf{b}_3 \end{bmatrix} \tag{29}$$

So

$$\mathrm{Tr}\left(A^T B\right) = \mathbf{a}_1^T \mathbf{b}_1 + \mathbf{a}_2^T \mathbf{b}_2 + \mathbf{a}_3^T \mathbf{b}_3$$

which is nothing but the inner product of the two "stacked" vectors $\begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{bmatrix}$ and $\begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \end{bmatrix}$. Hence

$$\langle A, B\rangle = \mathrm{Tr}\left(A^T B\right) = \left\langle \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{bmatrix}, \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \end{bmatrix} \right\rangle$$

Exercise 44. If x is a vector, show that

$$\mathrm{Tr}(xx^T) = x^T x$$

12 Norms

Previously we have used $||\cdot||$ to denote the Euclidean length function. At different times, it is useful to have more general notions of size and distance in vector spaces. This section is devoted to such generalizations.

12.1 Vector norm

Definition 45. A *norm* is a function that assigns a real-valued length to each vector in a vector space $\mathbb{C}^m$. To develop a reasonable notion of length, a norm must satisfy the following properties: for any vectors a, b and scalars $\alpha \in \mathbb{C}$,

- the norm of a nonzero vector is positive: $||a|| \geq 0$, and $||a|| = 0$ if and only if $a = 0$
- scaling a vector scales its norm by the same amount: $||\alpha a|| = |\alpha| \, ||a||$
- triangle inequality: $||a + b|| \leq ||a|| + ||b||$

Let w_1 be a $n \times 1$ vector. The most important class of vector norms, the p norms, of w are defined by

$$||w||_p = \left(\sum_{i=1}^n |w_i|^p\right)^{1/p}, \ 1 \leq p < \infty$$

Specifically, we have

$\|w\|_1 = \sum_{i=1}^n |w_j|$ (absolute column sum)

$\|w\|_\infty = \max_i |w_i|$

$\|w\|_2 = \sqrt{w^H w}$ (Euclidean norm)

Remark 46. When unspecified, $||\cdot||$ refers to 2 norm in this set of notes.

Intuitions for the infinity norm By definition

$$||w||_\infty = \lim_{p\to\infty}\left(\sum_{i=1}^{n}|w_i|^p\right)^{1/p}$$

Intuitively, as p increases, $\max_i |w_i|$ takes more and more weighting in $\sum_{i=1}^{n}|w_i|^p$. More rigorously, we have

$$\lim_{p\to\infty}\left((\max|w_i|)^p\right)^{1/p} \le \lim_{p\to\infty}\left(\sum_{i=1}^{n}|w_i|^p\right)^{1/p} \le \lim_{p\to\infty}\left(\sum_{i=1}^{n}(\max|w_i|)^p\right)^{1/p}$$

Both $\lim_{p\to\infty}\left((\max|w_i|)^p\right)^{1/p}$ and $\lim_{p\to\infty}\left(\sum_{i=1}^{n}(\max|w_i|)^p\right)^{1/p}$ equals $\max_i|w_i|$. Hence $||w||_\infty = \max|w_i|$

12.2 Induced matrix norm

As matrices define linear transformations between vector spaces, it makes sense to have a measure of the "size" of the transformation. Induced matrix norms[2] are defined by

$$||M||_{p\leftarrow q} = \max_{x\neq 0}\frac{||Mx||_p}{||x||_q} \tag{30}$$

In other words, $||M||_{q\leftarrow q}$ is the maximum factor by which M can "stretch" a vector x.

In particular, the following matrix norms are common:

$\|M\|_{1\leftarrow 1} = \max_j \sum_{i=1}^{n}|M_{ij}|$ maximum absolute column sum

$\|M\|_{\infty\leftarrow\infty} = \max_i \sum_{j=1}^{m}|M_{ij}|$ maximum absolute row sum

$\|M\|_{2\leftarrow 2} = \sqrt{\lambda_{\max}(M^*M)}$ maximum singular value

The induced 2 norm can be understood as follows:

$$||M||_{2\leftarrow 2} = \max_{x\neq 0}\frac{||Mx||_2}{||x||_2} = \max_{x\neq 0}\sqrt{\frac{x^*M^*Mx}{\langle x,x\rangle^2}} = \sqrt{\lambda_{\max}(M^*M)}$$

Remark 47. When $p=q$ in (30), often the induced matrix norm is simply written as $||M||_p$.

12.3 Frobenius norm and general matrix norms

Matrix norms do not have to be induced by vector norms.

[2]It is 'induced' from other vector norms as shown in the definition.

Formal definition: Let $\mathcal{M}_n$ be the set of all $n \times n$ real- or complex-valued matrices. We call a function $||\cdot|| : \mathcal{M}_n \to \mathbb{R}$ a matrix norm if for all $A, B \in \mathcal{M}_n$ it satisfies the following axioms:

1. $||A|| \geq 0$
2. $\|A\| = 0$ if and only if $A = 0$
3. $||cA\| = |c|\|A\|$ for all complex scalars c
4. $\|A + B\| \leq \|A\| + \|B\|$
5. $\|AB\| \leq \|A\|\|B\|$

The formal definition of matrix norms is slightly amended from vector norms. This is because although $\mathcal{M}_n$ is itself a vector space of dimension n^2, it has a natural multiplication operation that is absent in regular vector spaces. A vector norm on matrices that satisfies the first four axioms and not necessarily axiom 5 is often called a generalized matrix norm.

Frobenius norm: The most important matrix norm which is not induced by a vector norm is the Frobenius norm, defined by

$$\|A\|_F \triangleq \sqrt{\operatorname{Tr}(A^*A)} = \sqrt{< A, A >} = \sqrt{\sum_{i,j} |a_{i,j}|^2}$$

Frobenius norm is just the Euclidean norm of the matrix, written out as a long column vector:

$$||A||_F = (\operatorname{Tr}(A^*A))^{\frac{1}{2}} = \left(\sum_{i=1}^{m}\sum_{j=1}^{m} |a_{i,j}|^2\right)^{\frac{1}{2}}$$

We also have bounds for Frobenius norms:

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

Transforming from one matrix norm to another:

Theorem 48. *If $||\cdot||$ is a matrix norm on $\mathcal{M}_n$ and if $S \in \mathcal{M}_n$ is nonsingular, then*

$$||A||_S = ||S^{-1}AS|| \ \forall A \in \mathcal{M}_n$$

is a matrix norm.

Exercise 49. Prove Theorem 48.

12.4 Norm inequalities

1. Cauchy-Schwartz Inequality:
$$|\langle x, y\rangle| \leq ||x||_2||y||_2$$
which is the special case of the Holder inequality
$$|\langle x, y\rangle| \leq ||x||_p||y||_q,\ \frac{1}{p}+\frac{1}{q}=1,\ 1\leq p,q\leq\infty \tag{31}$$
Both bounds are tight: for certain choices of x and y, the inequalities become equalities.

2. Bounding induced matrix norms:
$$||AB||_{l\leftarrow n} \leq ||A||_{l\leftarrow m}||B||_{m\leftarrow n} \tag{32}$$
which comes from
$$||ABx||_l \leq ||A||_{l\leftarrow m}||Bx||_m \leq ||A||_{l\leftarrow m}||B||_{m\leftarrow n}||x||_n$$
In general, the bound is not tight. For instance, $||A^n|| = ||A||^n$ does not hold for $n \geq 2$ unless A has special structures.

3. (31) and (32) are useful for computing bounds of difficult-to-compute norms. For instance, $||A||_2^2$ is expensive to compute but $||A||_1$ and $||A||_\infty$ are not. As a special case of (32), we have
$$||A||_2^2 \leq ||A||_1||A||_\infty$$
We can obtain an upper bound of $||A||_2^2$ by computing $||A||_1||A||_\infty$.

4. Any matrix induced norms of A are larger than the absolute eigenvalues of A:
$$|\lambda(A)| \leq ||A||_p$$
Hence as a special case, the spectral radius is upper bounded by any matrix norms:
$$\rho(A) \leq \|A\|$$

5. Let $A \in \mathcal{M}_n$ and $\epsilon > 0$ be given. There is a matrix norm such that
$$\rho(A) \leq \|A\| \leq \rho(A) + \epsilon$$
Hint: A can be decomposed as $A = U^*\Delta U$ where U is unitary and Δ is upper triangular [Schur triangulariztion theorem]. Let $D_t = \text{diag}(t, t^2, \ldots, t^n)$ and compute
$$D_t\Delta D_t^{-1} = \begin{bmatrix} \lambda_1 & t^{-1}d_{12} & \ldots & & \ldots & t^{-n+1}d_{1n} \\ 0 & \lambda_2 & t^{-1}d_{23} & & \ldots & t^{-n+2}d_{1n} \\ \vdots & \ddots & \lambda_3 & \ddots & & \vdots \\ & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & t^{-1}d_{n-1,n} \\ 0 & \ldots & & \ldots & 0 & \lambda_n \end{bmatrix}$$
For t large enough, the sum of the absolute values of the off-diagonal entries is less than ϵ and in particular
$$\|D_t\Delta D_t^{-1}\|_1 \leq \rho(A) + \epsilon$$

12.5 Exercises

1. Let x be an m vector and A be an $m \times n$ matrix. Verify each of the following inequalities, and give an example when the equality is achieved.

   (a) $||x||_\infty \leq ||x||_2$

   (b) $||x||_2 \leq \sqrt{m}||x||_\infty$

   (c) $||A||_\infty \leq \sqrt{n}||A||_2$

   (d) $||A||_2 \leq \sqrt{m}||A||_\infty$

2. Let x be a random vector with mean $\mathsf{E}[x] = 0$ and covariance $\mathsf{E}\left(xx^T\right) = I$, then

$$\|A\|_F^2 = \mathrm{E}\left[\|Ax\|_2^2\right]$$

   Hint: use Exercise 44.

13 Symmetric, skew-symmetric, and orthogonal matrices

13.1 Definitions and basic properties

A real square matrix A is called **symmetric** if $A = A^T$, **skew-symmetric** if $A = -A^T$.

Fact 50. *Any real square matrix A may be written as the sum of a symmetric matrix R and a skew-symmetric matrix S, where*

$$R = \frac{1}{2}\left(A + A^T\right), \; S = \frac{1}{2}\left(A - A^T\right)$$

If $A = [a_{jk}]$, then the **complex conjugate** of A is denoted as$\overline{A} = [\overline{a}_{jk}]$, i.e., each element $a_{jk} = \alpha + i\beta$ is replaced with its complex conjugate $\overline{a}_{jk} = \alpha - i\beta$.

A square matrix A is called **Hermitian** if $A^T = \overline{A}$; **skew-Hermitian** if $A^T = -\overline{A}$.

Example 51. Find the symmetric, skew-symmetric, Hermitian, and skew-Hermitian matrices in the set:

$$\left\{ \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2i \\ 2i & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2i \\ -2i & 1 \end{bmatrix}, \begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 2+2i \\ 2-2i & 0 \end{bmatrix} \right\}$$

We introduce one more class of important matrices: a real square matrix A is called **orthogonal**[3] if

$$A^T A = AA^T = I \tag{33}$$

Writing A in the column-vector notation

$$A = [a_1, a_2, \ldots, a_n]$$

[3]Some people also call use the notion of orthonormal matrix. But orthogonal matrix is more often used (we can say orthonormal basis).

we get the equivalent form of (33):

$$A^TA = \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_n^T \end{bmatrix} \begin{bmatrix} a_1, & a_2, & \dots, & a_n \end{bmatrix} = \begin{bmatrix} a_1^Ta_1 & a_1^Ta_2 & \dots & a_1^Ta_n \\ a_2^Ta_1 & a_2^Ta_2 & \dots & a_2^Ta_n \\ \vdots & \vdots & \vdots & \vdots \\ a_n^Ta_1 & a_n^Ta_2 & \dots & a_n^Ta_n \end{bmatrix} = I$$

Hence it must be that

$$a_j^Ta_j = 1$$
$$a_j^Ta_m = 0 \ \forall j \neq m$$

namely, a_j and a_m are orthonormal for any $j \neq m$.

The complex version of an orthogonal matrix is the **unitary matrix**. A square matrix A is called unitary if $A\overline{A}^T = \overline{A}^TA = I$, namely $A^{-1} = \overline{A}^T$.

Remark 52. Often the complex conjugate transpose $\overline{A}^T$ is written as A^*.

Theorem 53. *The eigenvalues of symmetric matrices are all real.*

Proof. $\forall : A \in \mathbb{R}^{n\times n}$ with $A^T = A$. $Au = \lambda u \Rightarrow \overline{u}^TAu = \lambda\overline{u}^Tu$, where $\overline{u}$ is the complex conjugate of u. $\overline{u}^TAu$ is a real number, as

$$\begin{aligned} \overline{\overline{u}^TAu} &= u^T\overline{A}\overline{u} \\ &= u^TA\overline{u} \quad \because A \in \mathbb{R}^{n\times n} \\ &= u^TA^T\overline{u} \quad \because A = A^T \\ &= \lambda u^T\overline{u} \quad \because (Au)^T = (\lambda u)^T \\ &= \lambda\overline{u}^Tu \quad \because u^T\overline{u} \in \mathbb{R} \\ &= \overline{u}^TAu \quad \because Au = \lambda u \end{aligned}$$

. By definition of complex conjugate numbers, $\overline{u}^Tu \in \mathbb{R}$. So $\lambda = \frac{\overline{u}^TAu}{\overline{u}^Tu}$ is also a real number. □

Theorem 54. *The eigenvalues of skew-symmetric matrices are all imaginary or zero.*

The proof is left as an exercise.

Fact 55. *An orthogonal transformation preserves the value of the inner product of vectors a and b in $\mathbb{R}^n$. That is, for any a and b in $\mathbb{R}^n$, orthogonal $n \times n$ matrix A, and $u = Aa$, $v = Ab$ we have $\langle u, v\rangle = \langle a, b\rangle$, as*

$$u^Tv = a^TA^TAb = a^Tb$$

Hence the transformation also preserves the length or 2-norm of any vector a in $\mathbb{R}^n$ given by $||a||_2 = \sqrt{\langle a, a\rangle}$.

Theorem 56. *The determinant of an orthogonal matrix is either 1 or -1.*

Proof. $UU^T = I \Rightarrow \det U \det U^T = (\det U)^2 = 1$ □

Theorem 57. *The eigenvalues of an orthogonal matrix A are real or complex conjugates in pairs and have absolute value 1.*

Proof. $Au = \lambda u \Rightarrow A^T A u = \lambda A^T u \Rightarrow u = \lambda A^T u \Rightarrow \overline{u}^T u = \lambda \overline{u}^T A^T u \Rightarrow \overline{u}^T u = \lambda \overline{u}^T \overline{A}^T u = \lambda\overline{\lambda}\overline{u}^T u \Rightarrow \left(|\lambda|^2 - 1\right)\overline{u}^T u = 0$ □

Properties of the special matrices

real matrix	complex matrix	properties
symmetric ($A = A^T$)	Hermitian ($A^* = A$)	eigenvalues are all real
orthogonal ($A^T A = AA^T = I$)	unitary ($A^* A = AA^* = I$)	eigenvalues have unity magnitude; Ax maintains the 2-norm of x
skew-symmetric ($A^T = -A$)	skew-Hermitian ($A^* = -A$)	eigenvalues are all imaginary or zero

Based on the eigenvalue characteristics:

- symmetric and Hermitian matrices are like the real line in the complex domain
- skew-symmetric/Hermitian matrices are like the imaginary line
- orthogonal/unitary matrices are like the unit circle

Exercise 58 (Representation of matrices using special matrices)**.** Many unitary matrices can be mapped as functions of skew-Hermitian matrices as follows

$$U = (I - S)^{-1}(I + S)$$

where $S \neq I$. Show that if S is skew-Hermitian, then U is unitary.

13.2 Symmetric eigenvalue decomposition (SED)

When $A \in \mathbb{R}^{n\times n}$ has n distinct eigenvalues, we have seen the useful result of matrix diagonalization:

$$A = U\Lambda U^{-1} = [u_1, \ldots, u_n]\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}[u_1, \ldots, u_n]^{-1} \tag{34}$$

where λ_i's are the distinct eigenvalues associated to the eigenvector u_i's.

The inverse matrix in (34) can be cumbersome to compute though.

The spectral theorem, aka symmetric eigenvalue decomposition theorem,[4] significantly simplifies the result when A is symmetric.

[4] Recall that the set of all the eigenvalues of A is called the spectrum of A. The largest of the absolute values of the eigenvalues of A is called the spectral radius of A.

Theorem 59. *$\forall : A \in \mathbb{R}^{n\times n},\ A^T = A$, there always exist λ_i and u_i, such that*

$$A = \sum_{i=1}^{n} \lambda_i u_i u_i^T = U\Lambda U^T \tag{35}$$

where:[5]

- *λ_i's: eigenvalues of A*
- *u_i: eigenvector associated to λ_i, normalized to have unity norms*
- *$U = [u_1, u_2, \cdots, u_n]^T$ is an orthogonal matrix, i.e., $U^TU = UU^T = I$*
- *$\{u_1, u_2, \cdots, u_n\}$ forms an orthonormal basis*
- $\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$

To understand the result, we show an important theorem first.

Theorem 60. *$\forall : A \in \mathbb{R}^{n\times n}$* with *$A^T = A$, then eigenvectors of A, associated with different eigenvalues, are orthogonal.*

Proof. Let $Au_i = \lambda_i u_i$ and $Au_j = \lambda_j u_j$. Then $u_i^T A u_j = u_i^T \lambda_j u_j = \lambda_j u_i^T u_j$. In the meantime, $u_i^T A u_j = u_i^T A^T u_j = (Au_i)^T u_j = \lambda_i u_i^T u_j$. So $\lambda_i u_i^T u_j = \lambda_j u_i^T u_j$. But $\lambda_i \neq \lambda_j$. It must be that $u_i^T u_j = 0$. □

Theorem 59 now follows. If A has distinct eigenvalues, then $U = [u_1, u_2, \cdots, u_n]^T$ is orthogonal if we normalize all the eigenvectors to unity norm. If A has $r(< n)$ distinct eigenvalues, we can *choose* multiple orthogonal eigenvectors for the eigenvalues with none-unity multiplicities.

Observations:

- If we "walk along" u_j, then

$$Au_j = \left(\sum_i \lambda_i u_i u_i^T\right) u_j = \lambda_j u_j u_j^T u_j = \lambda_j u_j \tag{36}$$

  where we used the orthonormal condition of $u_i^T u_j = 0$ if $i \neq j$. This confirms that u_j is an eigenvector.

[5] $u_i u_i^T \in \mathbb{R}^{n\times n}$ is a symmetric dyad, the so-called outerproduct of u_i and u_i. It has the following properties:

- $\forall\ v \in \mathbb{R}^{n\times 1}$, $(vv^T)_{ij} = v_i v_j$. (Proof: $(vv^T)_{ij} = e_i^T (vv^T) e_j = v_i v_j$, where e_i is the unit vector with all but the i_{th} elements being zero.)
- link with quadratic functions: $q(x) = x^T (vv^T) x = (v^T x)^2$

- $\{u_i\}_{i=1}^n$ is a orthonormal basis $\Rightarrow \forall x \in \mathbb{R}^n$, $\exists\ x = \sum_i \alpha_i u_i$, where $\alpha_i =< x, u_i >$. And we have

$$Ax = A\sum_i \alpha_i u_i = \sum_i \alpha_i A u_i = \sum_i \alpha_i \lambda_i u_i = \sum_i (\alpha_i \lambda_i) u_i \tag{37}$$

which gives the (intuitive) picture of the geometric meaning of Ax: decompose first x to the space spanned by the eigenvectors of A, scale each components by the corresponding eigenvalue, sum the results up, then we will get the vector Ax.

With the spectral theorem, next time we see a symmetric matrix A, we immediately know that

- λ_i is real for all i
- associated with λ_i, we can always find one or more real eigenvectors
- $\exists$ an orthonormal basis $\{u_i\}_{i=1}^n$, which consists of the eigenvectors
- if $A \in \mathbb{R}^{2\times 2}$, then if you compute first λ_1, λ_2 and u_1, you won't need to go through the regular math to get u_2, but can simply solve for a u_2 that is orthogonal to u_1 with $\|u_2\| = 1$.

Example 61. Consider the matrix $A = \begin{bmatrix} 5 & \sqrt{3} \\ \sqrt{3} & 7 \end{bmatrix}$. Computing the eigenvalues gives

$$\det \begin{bmatrix} 5-\lambda & \sqrt{3} \\ \sqrt{3} & 7-\lambda \end{bmatrix} = 35 - 12\lambda + \lambda^2 - 3 = (\lambda - 4)(\lambda - 8) = 0$$
$$\Rightarrow \lambda_1 = 4,\ \lambda_2 = 8$$

And we can know one of the eigenvectors from

$$(A - \lambda_1 I)\, t_1 = 0 \Rightarrow \begin{bmatrix} 1 & \sqrt{3} \\ \sqrt{3} & 3 \end{bmatrix} t_1 = 0 \Rightarrow t_1 = \begin{bmatrix} -\frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{bmatrix}$$

Note here we normalized t_1 such that $||t_1||_2 = 1$. With the above computation, we no more need to do $(A - \lambda_2 I)\, t_2 = 0$ for getting t_2. Keep in mind that A here is symmetric, so has eigenvectors orthogonal to each other. By direct observation, we can see that

$$x = \begin{bmatrix} \frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix}$$

is orthogonal to t_1 and $||x||_2 = 1$. So $t_2 = x$.

Theorem 62 (Eigenvalues of symmetric matrices). *If $A = A^T \in \mathbb{R}^{n\times n}$, then the eigenvalues of A satisfy*

$$\lambda_{\max} = \max_{x\in\mathbb{R}^n,\ x\neq 0} \frac{x^T A x}{\|x\|_2^2} \tag{38}$$

$$\lambda_{\min} = \min_{x\in\mathbb{R}^n,\ x\neq 0} \frac{x^T A x}{\|x\|_2^2} \tag{39}$$

Proof. Perform SED to get

$$A = \sum_{i=1}^{n} \lambda_i u_i^T u_i$$

where $\{u_i\}_{i=1}^n$ form a basis of $\mathbb{R}^n$. Then any vector $x \in \mathbb{R}^n$ can be decomposed as

$$x = \sum_{i=1}^{n} \alpha_i u_i$$

Thus

$$\max_{x\neq 0} \frac{x^T A x}{\|x\|_2^2} = \max_{\alpha_i} \frac{\left(\sum_i \alpha_i u_i\right)^T \sum_i \lambda_i \alpha_i u_i}{\sum_i \alpha_i^2} = \max_{\alpha_i} \frac{\sum_i \lambda_i \alpha_i^2}{\sum_i \alpha_i^2} = \lambda_{\max}$$

The proof for (39) is analogous and omitted. □

13.3 Symmetric positive-definite matrices

Definition 63. A symmetric matrix $P \in \mathbb{R}^{n\times n}$ is called **positive-definite**, written $P \succ 0$, if $x^T P x > 0$ for all $x\,(\neq 0) \in \mathbb{R}^n$. P is called **positive-semidefinite,** written $P \succeq 0$, if $x^T P x \geq 0$ for all $x \in \mathbb{R}^n$

Definition 64. A symmetric matrix $P \in \mathbb{R}^{n\times n}$ is called **negative-definite**, written $P \prec 0$, if $-P \succ 0$, i.e., $x^T P x < 0$ for all $x\,(\neq 0) \in \mathbb{R}^n$. P is called **negative-semidefinite,** written $P \preceq 0$, if $x^T P x \leq 0$ for all $x \in \mathbb{R}^n$

When A and B have compatible dimensions, $A \succ B$ means $A - B \succ 0$.
Positive-definite matrices can have negative entries, as shown in the next example.

Example 65. The following matrix is positive-definite

$$P = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

as $P = P^T$ and take any $v = [x, y]^T$, we have

$$v^T P v = \begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 2x^2 + 2y^2 - 2xy = x^2 + y^2 + (x+y)^2 \geq 0$$

and the equality sign holds only when $x = y = 0$.

Conversely, matrices whose entries are all positive are not necessarily positive-definite.

Example 66. The following matrix is not positive-definite

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

as

$$\begin{bmatrix} 1 \\ -1 \end{bmatrix}^T \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = -2 < 0$$

Theorem 67. *For a symmetric matrix P, $P \succ 0$ if and only if all the eigenvalues of P are positive.*

Proof. Since P is symmetric, we have

$$\lambda_{\max}(P) = \max_{x \in \mathbb{R}^n,\ x \neq 0} \frac{x^T A x}{\|x\|_2^2} \tag{40}$$

$$\lambda_{\min}(P) = \min_{x \in \mathbb{R}^n,\ x \neq 0} \frac{x^T A x}{\|x\|_2^2} \tag{41}$$

which gives

$$x^T A x \in \left[\lambda_{\min}\|x\|_2^2,\ \lambda_{\max}\|x\|_2^2\right]$$

For $x \neq 0$, $\|x\|_2^2$ is always positive. It can thus be seen that $x^T A x > 0,\ x \neq 0 \Leftrightarrow \lambda_{\min} > 0$. □

Lemma. *For a symmetric matrix P, $P \succeq 0$ if and only if all the eigenvalues of P are none-negative.*

Theorem. *If A is symmetric positive definite, X is full column rank, then $X^T A X$ is positive definite.*

Proof. Consider $y\left(X^T A X\right) y = x^T A x$, which is always positive unless $x = 0$. But X is full rank so $Xy = x = 0$ if and only if $y = 0$. This proves $X^T A X$ is positive definite. □

Fact. *All principle submatrices of A are positive definite.*

Proof. Use the last theorem. Take $X = e_1$, $X = [e_1, e_2]$, etc. Here e_i is a column vector whose ith-entry is 1 and all other entries are zero. □

Example 68. The following matrices are all not positive definite:

$$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} -1 & 1 \\ 1 & 2 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

Positive-definite matrices are like positive real numbers. We can have the concept of *square root* of positive-definite matrices.

Definition 69. Let $P \succeq 0$. We can perform symmetric eigenvalue decomposition to obtain $P = UDU^T$ where U is orthogonal with $UU^T = I$ and D is diagonal with all diagonal elements being none negative

$$D = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_n \end{bmatrix} \succeq 0$$

Then the square root of P, written $P^{\frac{1}{2}}$, is defined as

$$P^{\frac{1}{2}} = UD^{\frac{1}{2}}U^T$$

where

$$D^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\lambda_1} & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sqrt{\lambda_n} \end{bmatrix}$$

13.4 General positive-definite matrices

Definition 70. A general square matrix $Q \in \mathbb{R}^{n\times n}$ is called positive-definite, written as $Q \succ 0$, if $x^TQx > 0\ \forall x \neq 0$.

We have discussed the case when Q is symmetric. If not, recall that any real square matrix can be decomposed as the sum of a symmetric matrix and a skew symmetric matrix:

$$Q = \frac{Q+Q^T}{2} + \frac{Q-Q^T}{2}$$

where $\frac{Q+Q^T}{2}$ is symmetric.

Notice that $x^T\frac{Q-Q^T}{2}x = x^TQx - \left(x^TQx\right)^T = 0$. So for a general square real matrix:

$$Q \succ 0 \Leftrightarrow Q + Q^T \succ 0$$

Example 71. The following matrices are positive definite but not symmetric

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

For complex matrices with $Q = Q^* = Q_R + jQ_I$, we have

$$\begin{aligned} Q \succ 0 &\Leftrightarrow x^*Qx > 0,\ \forall x \neq 0 \\ &\Leftrightarrow \left(x_R^T - jx_I^T\right)(Q_R + jQ_I)(x_R + jx_I) > 0 \\ &\Leftrightarrow \begin{pmatrix} x_R \\ x_I \end{pmatrix}^T \begin{pmatrix} 1 \\ j \end{pmatrix} \begin{pmatrix} Q_R & Q_I \end{pmatrix} \begin{pmatrix} 1 \\ j \end{pmatrix} \begin{pmatrix} 1 \\ j \end{pmatrix}^T \begin{pmatrix} x_R \\ x_I \end{pmatrix} \\ &\Leftrightarrow \begin{pmatrix} x_R \\ x_I \end{pmatrix}^T \begin{pmatrix} Q_R & Q_I \\ -Q_I & Q_R \end{pmatrix} \begin{pmatrix} x_R \\ x_I \end{pmatrix} > 0 \\ &\Leftrightarrow x_R^TQ_Rx_R - x_I^TQ_Ix_R + x_R^TQ_Ix_I + x_I^TQ_Rx_I > 0 \end{aligned}$$

13.5 *Positive-definite functions and non-constant matrices

We can further extend the concept of positive definiteness to general and even time-varying functions, by placing upper and/or lower bounds that are "positive-definite like".

Define first two special functions:

1. class-K function: $\psi \in C^0 : [0, a] \to [0, \infty)$ with $\psi(0) = 0$ and ψ strictly increasing

2. class-K_∞ function: if the domain $a = \infty$ and $\psi(r) \to \infty$ as $r \to \infty$

Note: ψ is continuous but does not need to be continuously differentiable, e.g.

$$\psi = \min\left\{x, x^2\right\}$$

is a class-K function.

Lemma 72. *Let $V : D \to \mathbb{R}$ be a continuous, positive definite function. Let $B_r \subset D$ for some $r > 0$. Then there exist class-K functions ψ and ϕ defined on $[0, r]$ s.t.*

$$\phi(\|x\|) \le V(x) \le \psi(\|x\|)$$

for all $x \in B_r$.

- if the domain $D = \mathbb{R}^n$ then $r = \infty$
- if $V(x)$ is radially unbounded, then ψ and ϕ can be class-K_∞

Definition 73. A time-dependent function $V(t, x)$ is positive-semidefinite if

$$V(t, x) \ge \phi(\|x\|)$$

where ϕ is class-K.

Definition 74. A time-varying matrix $P(t)$ is positive definite if there exists a lower-bounding positive definite matrix such that

$$P(t) \succeq c_3 I \succ 0, \ \forall t \ge 0$$

14 Singular value and singular value decomposition (SVD)

14.1 Motivation

Symmetric eigenvalue decomposition is great but many matrices are not symmetric. A general matrix A may actually not even be square. Singular value decomposition is an important matrix decomposition technique that works for arbitrary matrices.[6]

For a general none-square matrix $A \in \mathbb{C}^{m\times n}$, eigenvalues and eigenvectors are generalized to

$$Av_j = \sigma_j u_j \tag{42}$$

Be careful about the dimensions: if $m > n$, we have

$$\begin{bmatrix} & \vdots & \\ \cdots & A & \cdots \\ & \vdots & \end{bmatrix} \underbrace{\begin{bmatrix} v_1 \,|\, v_2 \,|\, \dots \,|\, v_n \end{bmatrix}}_{V} = \underbrace{\begin{bmatrix} u_1 \,|\, u_2 \,|\, \dots \,|\, u_n \end{bmatrix}}_{\hat{U}} \underbrace{\begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{bmatrix}}_{\hat{\Sigma}}$$

It turns out that, if A has full column rank n, then we can find a V that is unitary ($VV^* = V^*V = I$) and a $\hat{U}$ that has orthonormal columns. Hence

$$A = \hat{U}\hat{\Sigma}V^* \tag{43}$$

14.2 SVD

(43) forms the so-called reduced singular value decomposition (SVD). The idea of a "full" SVD is as follows. The columns of $\hat{U}$ are n orthonormal vectors in the m-dimensional space $\mathbb{C}^m$. They do not form a basis for $\mathbb{C}^m$ unless $m = n$. We can add additional $m - n$ orthonormal columns to $\hat{U}$ and augment it to a unitary matrix U. Now the matrix dimension has changed, $\hat{\Sigma}$ needs to be augmented to compatible dimensions as well. To maintain the equality (43), the newly added elements to $\hat{\Sigma}$ are set to zero.

Theorem 75. *Let $A \in \mathbb{C}^{m\times n}$ with rank r. Then we can find orthogonal matrices $U \in \mathbb{C}^{m\times m}$ and $V \in \mathbb{C}^{n\times n}$ such that*

$$A = U\Sigma V^*$$

[6]History of SVD: discovered between 1873 and 1889, independently by several pioneers; did not became widely known in applied mathematics until the late 1960s, when it was shown that SVD can be computed effectively and used as the basis for solving many problems.

where

$$\begin{aligned}\Sigma &\in \mathbb{R}^{m\times n} \text{ is diagonal}\\ U &\in \mathbb{C}^{m\times m} \text{ is unitary}\\ V &\in \mathbb{C}^{n\times n} \text{ is unitary}\end{aligned}$$

In addition, the diagonal entries σ_j of Σ are nonnegative and in nonincreasing order; that is, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$.

Proof. Notice that A^*A is positive semi-definite. Hence, A^*A has a full set of orthonormal eigenvectors; its eigenvalues are real and nonnegative. Order these eigenvalues as

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \lambda_{r+2} = \cdots = \lambda_n = 0$$

[7]Let $\{v_1, \ldots, v_n\}$ be an orthonormal choice of eigenvectors of A^*A corresponding to these eigenvalues:

$$A^*Av_i = \lambda_i v_i$$

Then,

$$||Av_i||^2 = v_i^*A^*Av_i = \lambda_i v_i^* v_i = \lambda_i$$

For $i > r$, it follows that $Av_i = 0$.

For $1 \leq i \leq r$, we have

$$A^*Av_i = \lambda_i v_i$$

Recall (42), we define $\sigma_i = \sqrt{\lambda_i}$ and get

$$\begin{aligned}Av_i &= \sigma_i u_i\\ A^*u_i &= \sigma_i v_i\end{aligned}$$

For $1 \leq i, j \leq r$, we have

$$\langle u_i, u_j\rangle = u_i^* u_j = \frac{1}{\sigma_i\sigma_j} v_i^* A^* A v_j = \frac{1}{\sigma_i\sigma_j}\lambda_j v_i^* v_j = \frac{\sigma_j}{\sigma_i} v_i^* v_j = \begin{cases}1 & i = j\\ 0 & i \neq j\end{cases}$$

Hence $\{u_1, \ldots, u_r\}$ is an orthonormal set of eigenvectors. Extending this set to form an orthonormal basis for $\mathbb{C}^m$ gives

$$U = \left[\begin{array}{ccc|ccc} u_1, & \ldots, & u_r & u_{r+1}, & \ldots, & u_m\end{array}\right]$$

For $i \leq r$, we already have

$$Av_i = \sigma_i u_i$$

[7]Fact: rank (A) = rank (A^*A). To see this, notice first, that rank $(A) \geq$ rank (A^*A) by definition of rank. Second, $A^*Ax = 0 \Rightarrow x^*A^*Ax = 0 \Rightarrow ||Ax|| = 0 \Rightarrow Ax = 0$, hence rank $(A) \leq$ rank (A^*A).

namely

$$A[v_1, \dots v_r] = [u_1, \dots, u_r] \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix}$$

$$= \left[\begin{array}{ccc|ccc} u_1, & \dots, & u_r & u_{r+1}, & \dots, & u_m \end{array}\right] \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \\ & & & 0 \\ & & & \vdots \\ & & & 0 \end{bmatrix}$$

For $v_{r+1}, \dots$, we have already seen that $Av_{r+1} = Av_{r+2} = \dots = 0$, hence

$$A \underbrace{[v_1, \dots v_r | v_{r+1}, \dots, v_n]}_{n\times n} = \underbrace{\left[\begin{array}{ccc|ccc} u_1, & \dots, & u_r & u_{r+1}, & \dots, & u_m \end{array}\right]}_{m\times m} \underbrace{\begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \\ & & & & & \vdots \\ & & & & & 0 \end{bmatrix}}_{m\times n}$$

$$\Rightarrow A = U\Sigma V^*$$

□

Theorem 76. *The range space of A is spanned by $\{u_1, \dots, u_r\}$. The null space of A is spanned by $\{v_{r+1}, \dots, v_n\}$.*

□

Theorem 77. *The nonzero singular values of A are the square roots of the nonzero eigenvalues of A^*A or AA^*.*

□

Theorem 78. $||A||_2 = \sigma_1$*, i.e., the induced two norm of A is the maximum singular value of A.*

The next important theorem can be easily proved via SVD.

Theorem (Fundermental theory of linear algebra). *Let $A \in \mathbb{R}^{m\times n}$. Then*

$$\mathcal{R}(A) + \mathcal{N}\left(A^T\right) = \mathbb{R}^m$$

and

$$\mathcal{R}(A) \perp \mathcal{N}\left(A^T\right)$$

Proof. By singular value decomposition

$$\begin{aligned} A &= U\Sigma V^T \\ A^T &= V\Sigma U^T \end{aligned}$$

Range of A is the first r columns of U, from the first equation; Null space of A^T is the last $m-r$ columns of U, from the second equation. □

New intuition of matrix vector operation With $A = U\Sigma V^*$, a new intuition for $Ax = U\Sigma V^* x$ is formed. Since V is unitary, it is norm-preserving, in the sense that V^*x does not change the 2-norm of the vector x. In other words, V^*x only rotates x in $\mathbb{C}^n$. The diagonal matrix Σ then functions to scale (by its diagonal values) the rotated vector. Finally, U is another rotation in $\mathbb{C}^m$.

14.3 Properties of singular values

Fact. *Let A and B be matrices with compatible dimensions. The following are true*
$\overline{\sigma}(A+B) \leq \overline{\sigma}(A) + \overline{\sigma}(B)$
$\overline{\sigma}(AB) \leq \overline{\sigma}(A)\overline{\sigma}(B)$

Proof. The first inequality comes from

$$\overline{\sigma}(A+B) = \max_{v\neq 0} \frac{||Av+Bv||_2}{||v||_2} \leq \max_{v\neq 0} \frac{||Av||_2 + ||Bv||_2}{||v||_2}$$

The second inequality uses

$$\overline{\sigma}(AB) = \max_{v\neq 0} \frac{||ABv||_2}{||v||_2} \leq \max_{v\neq 0} \frac{||A||_2 ||Bv||_2}{||v||_2}$$

□

14.4 Exercises

1. Compute the singular values of the following matrices

$$(a) \begin{bmatrix} 3 & \\ & -2 \end{bmatrix},\ (b) \begin{bmatrix} 2 & \\ & 3 \end{bmatrix},\ (c) \begin{bmatrix} 0 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},\ (d) \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix},\ (e) \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

2. Show that if A is real, then it has a real SVD (i.e., U and V are both real).

3. For any matrix $A \in \mathbb{R}^{n\times m}$, construct

$$M = \begin{bmatrix} \overbrace{0}^{n\times n} & \overbrace{A}^{n\times m} \\ \underbrace{A^T}_{m\times n} & \underbrace{0}_{m\times m} \end{bmatrix} \in \mathbb{R}^{(n+m)\times(n+m)}$$

which satisfies

$$M^T = M$$

M is Hermitian, and hence has real eigenvalues and eigenvectors:

$$\begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} u_j \\ v_j \end{bmatrix} = \sigma_j \begin{bmatrix} u_j \\ v_j \end{bmatrix} \tag{44}$$

(a) Show that

i. v_j is in the co-kernal (perpendicular to kernal/null space) of A and u_j is in the range of A.

ii. if σ_j and $\begin{bmatrix} u_j \\ v_j \end{bmatrix}$ form a eigen pair for M, then$-\sigma_j$ and $\left[u_j^T, -v_j^T\right]^T$ also form an eigen pair for M

iii. eigenvalues of M always appear in pairs that are symmetric to the imaginary axis.

(b) Use the results to show that, if

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 1 & 4 & 32 \end{bmatrix}$$

then M must have eigenvalues that are equal to 0.

4. Suppose $A \in \mathbb{C}^{m\times m}$ and has an SVD $A = U\Sigma V^*$. Find an eigenvalue decomposition of

$$\begin{bmatrix} 0 & A^* \\ A & 0 \end{bmatrix}$$

5. **Worst input direction** in matrix vector multiplications. Recall that any matrix defines a linear transformation:

$$Mw = z$$

What is the worst input direction for the vector w? Here *worst* means: if we fix the input norm, say $\|w\| = 1$, $\|z\|$ will reach a maximum value (the worst case) for a specific input direction in w.

(a) Show that the worst $||z||$ is $||M||$ when $||w|| = 1$.

(b) Provide procedures to obtain the w that gives the maximum $||z||$, for the cases of 1 norm, ∞ norm, and 2 norm.

Index

A
adjoint matrix, 19
adjugate matrix, 19

B
basis, 7

C
characteristic equation, 12
characteristic polynomial, 12
characteristic values of a matrix, 12

D
Determinants, 8
Diagonal matrices, 1
dyad, 8

E
eigenbasis, 15
eigenvalue, 11
eigenvectors, 11
Elementary Row Operations, 6

H
homogeneous system, 4

I
Identity matrix, 1
inverse, 15

L
linear combination, 7
linear equation set, 1
linearly independent, 7
lower triangular matrices, 1

M
Matrix inversion lemma, 24
matrix product, 2

N
nonhomogeneous system, 4
null space, 8
nullity, 8

O
overdetermined, 9

R
range space, 8
rank, 8
row echelon form, 6

S
singular, 9
skew-symmetric, 3
span, 7
spectral radius, 12
spectrum, 12
symmetric, 3

T
transpose, 3

U
undetermined, 9
Upper triangular matrices, 1

V
Vectors, 1