

Matrix Analysis



Graduate Course
Beijing Jiaotong University

Xiaoming Huang
School of Science, BJTU
xmhuang@bjtu.edu.cn



Chapter 0

Review and Miscellanea

Chapter 0 Review and Miscellanea

- o.0 Linear Equations
- o.1 Vector spaces
- o.2 Matrices
- o.3 Determinants
- o.4 Rank
- o.5 Nonsingularity
- o.6 The Euclidean inner product and norm
- o.7 Partitioned sets and matrices
- o.8 Determinants again
- o.9 Special types of matrices
- o.10 Change of basis
- o.11 Equivalence relations

Topics of today

0.0 Linear Equations

Example: A fundamental problem that surfaces in all mathematical sciences is that of analyzing and solving m algebraic equations in n unknowns. The study of a system of simultaneous linear equations is in a natural and indivisible alliance with the study of the rectangular array of numbers defined by the coefficients of the equations. This link seems to have been made at the outset. The earliest recorded analysis of simultaneous equations is found in the ancient Chinese book *Chiu-chang Suan-shu* (*Nine Chapters on Arithmetic*), estimated to have been written some time around 200 B.C. In the beginning of Chapter VIII, there appears a problem of the following form.

Three sheafs of a good crop, two sheafs of a mediocre crop, and one sheaf of a bad crop are sold for 39 dou. Two sheafs of good, three mediocre, and one bad are sold for 34 dou; and one good, two mediocre, and three bad are sold for 26 dou. What is the price received for each sheaf of a good crop, each sheaf of a mediocre crop, and each sheaf of a bad crop?

0.0 Linear Equations

Today, this problem would be formulated as three equations in three unknowns by writing

$$3x + 2y + z = 39,$$

$$2x + 3y + z = 34,$$

$$x + 2y + 3z = 26,$$

where x , y , and z represent the price for one sheaf of a good, mediocre, and bad crop, respectively.

How to solve this problem?

0.0 Linear Equations

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

Problem: to calculate, if possible, a common solution for a system of m linear algebraic equations in n unknowns.

Let E_k denote the k_{th} equation in the above system:

$$E_k : a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{kn}x_n = b_k \quad (k = 1, 2, \cdots, m)$$

Three Possibilities:

- **UNIQUE SOLUTION:** There is one and only one set of values for the x_i 's that satisfies all equations simultaneously.
- **NO SOLUTION:** There is no set of values for the x_i 's that satisfies all equations simultaneously—the solution set is empty.
- **INFINITELY MANY SOLUTIONS:** There are infinitely many different sets of values for the x_i 's that satisfy all equations simultaneously. It is not difficult to prove that if a system has more than one solution, then it has infinitely many solutions. For example, it is impossible for a system to have exactly two different solutions.

Example: To solve the following equation system:

$$\begin{cases} 2x + y + z = 1 \\ 6x + 2y + z = -1 \\ -2x + 2y + z = 7 \end{cases}$$

At each step of solution, the strategy is to focus on one position, called the ***pivot position***, and to eliminate all terms below this position using the three elementary operations. The coefficient in the pivot position is called a ***pivotal element*** (or simply a ***pivot***), while the equation in which the pivot lies is referred to as the ***pivotal equation***. Only nonzero numbers are allowed to be pivots. If a coefficient in a pivot position is ever 0, then the pivotal equation is interchanged with an equation *below* the pivotal equation to produce a nonzero pivot. (This is always possible for square systems possessing a unique solution.) Unless it is 0, the first coefficient of the first equation is taken as the first pivot. For example, the circled 2 in the system below is the pivot for the first step:

Step 1. Eliminate all terms below the first pivot.

- Subtract three times the first equation from the second so as to produce the equivalent system:

$$\begin{array}{rclcl} \textcircled{2}x & + & y & + & z & = & 1, \\ & & -y & - & 2z & = & -4 \quad (E_2 - 3E_1), \\ -2x & + & 2y & + & z & = & 7. \end{array}$$

- Add the first equation to the third equation to produce the equivalent system:

$$\begin{array}{rclcl} \textcircled{2} x & + & y & + & z & = & 1, \\ & & - & y & - & 2z & = & -4, \\ & & & 3y & + & 2z & = & 8 & (E_3 + E_1). \end{array}$$

Step 2. Select a new pivot.

- For the time being, select a new pivot by moving down and to the right.² If this coefficient is not 0, then it is the next pivot. Otherwise, interchange with an equation *below* this position so as to bring a nonzero number into this pivotal position. In our example, -1 is the second pivot as identified below:

$$\begin{array}{rclcl} 2x & + & y & + & z & = & 1, \\ & & \textcircled{-1}y & - & 2z & = & -4, \\ & & 3y & + & 2z & = & 8. \end{array}$$

Step 3. Eliminate all terms below the second pivot.

- Add three times the second equation to the third equation so as to produce the equivalent system:

$$\begin{array}{rclcl} 2x & + & y & + & z & = & 1, \\ & & \textcircled{-1}y & - & 2z & = & -4, \\ & & & & -4z & = & -4 \end{array} \quad \begin{array}{l} \\ \\ (E_3 + 3E_2). \end{array} \quad (1.2.5)$$

- In general, at each step you move down and to the right to select the next pivot, then eliminate all terms below the pivot until you can no longer proceed. In this example, the third pivot is -4 , but since there is nothing below the third pivot to eliminate, the process is complete.

At this point, we say that the system has been *triangularized*. A triangular system is easily solved by a simple method known as *back substitution* in which the last equation is solved for the value of the last unknown and then substituted back into the penultimate equation, which is in turn solved for the penultimate unknown, etc., until each unknown has been determined. For our example, solve the last equation in (1.2.5) to obtain

$$z = 1.$$

Substitute $z = 1$ back into the second equation in (1.2.5) and determine

$$y = 4 - 2z = 4 - 2(1) = 2.$$

Finally, substitute $z = 1$ and $y = 2$ back into the first equation in (1.2.5) to get

$$x = \frac{1}{2}(1 - y - z) = \frac{1}{2}(1 - 2 - 1) = -1,$$

which completes the solution.

Matrix expression form

$$\left(\begin{array}{ccc|c} 2 & 1 & 1 & 1 \\ 6 & 2 & 1 & -1 \\ -2 & 2 & 1 & 7 \end{array} \right). \quad (\text{The line emphasizes where } = \text{ appeared.})$$

The array of coefficients—the numbers on the left-hand side of the vertical line—is called the *coefficient matrix* for the system. The entire array—the coefficient matrix augmented by the numbers from the right-hand side of the system—is called the *augmented matrix* associated with the system. If the coefficient matrix is denoted by $\mathbf{A}$ and the right-hand side is denoted by $\mathbf{b}$, then the augmented matrix associated with the system is denoted by $[\mathbf{A}|\mathbf{b}]$.

Matrix expression form

To solve the system (1.2.4) by using elementary row operations, start with the associated augmented matrix $[\mathbf{A}|\mathbf{b}]$ and triangularize the coefficient matrix $\mathbf{A}$ by performing exactly the same sequence of row operations that corresponds to the elementary operations executed on the equations themselves:

$$\begin{aligned} \left(\begin{array}{ccc|c} \textcircled{2} & 1 & 1 & 1 \\ 6 & 2 & 1 & -1 \\ -2 & 2 & 1 & 7 \end{array} \right) \begin{array}{l} R_2 - 3R_1 \\ R_3 + R_1 \end{array} &\longrightarrow \left(\begin{array}{ccc|c} 2 & 1 & 1 & 1 \\ 0 & \textcircled{-1} & -2 & -4 \\ 0 & 3 & 2 & 8 \end{array} \right) \begin{array}{l} \\ R_3 + 3R_2 \end{array} \\ &\longrightarrow \left(\begin{array}{ccc|c} 2 & 1 & 1 & 1 \\ 0 & -1 & -2 & -4 \\ 0 & 0 & -4 & -4 \end{array} \right). \end{aligned}$$

The final array represents the triangular system

$$\begin{array}{rcrcrcrcrcl} 2x & + & y & + & z & = & 1, \\ & & -y & - & 2z & = & -4, \\ & & & & -4z & = & -4 \end{array}$$

Example Solve the following system using Gaussian elimination with back substitution:

$$\begin{aligned}v - w &= 3, \\-2u + 4v - w &= 1, \\-2u + 5v - 4w &= -2.\end{aligned}$$

Solution: The associated augmented matrix is

$$\left(\begin{array}{ccc|c} 0 & 1 & -1 & 3 \\ -2 & 4 & -1 & 1 \\ -2 & 5 & -4 & -2 \end{array} \right).$$

Since the first pivotal position contains 0, interchange rows one and two before eliminating below the first pivot:

$$\begin{aligned} \left(\begin{array}{ccc|c} \textcircled{0} & 1 & -1 & 3 \\ -2 & 4 & -1 & 1 \\ -2 & 5 & -4 & -2 \end{array} \right) & \xrightarrow{\text{Interchange } R_1 \text{ and } R_2} \left(\begin{array}{ccc|c} \textcircled{-2} & 4 & -1 & 1 \\ 0 & 1 & -1 & 3 \\ -2 & 5 & -4 & -2 \end{array} \right) R_3 - R_1 \\ & \longrightarrow \left(\begin{array}{ccc|c} -2 & 4 & -1 & 1 \\ 0 & \textcircled{1} & -1 & 3 \\ 0 & 1 & -3 & -3 \end{array} \right) R_3 - R_2 \longrightarrow \left(\begin{array}{ccc|c} -2 & 4 & -1 & 1 \\ 0 & 1 & -1 & 3 \\ 0 & 0 & -2 & -6 \end{array} \right). \end{aligned}$$

Back substitution yields

$$w = \frac{-6}{-2} = 3, \quad v = 3 + w = 3 + 3 = 6, \quad u = \frac{1}{-2} (1 - 4v + w) = \frac{1}{-2} (1 - 24 + 3) = 10.$$

GAUSS–JORDAN METHOD

The purpose of this section is to introduce a variation of Gaussian elimination that is known as the *Gauss–Jordan method*.⁴ The two features that distinguish the Gauss–Jordan method from standard Gaussian elimination are as follows.

- At each step, the pivot element is forced to be 1.
- At each step, all terms *above* the pivot as well as all terms below the pivot are eliminated.

In other words, if

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & b_n \end{array} \right)$$

is the augmented matrix associated with a linear system, then elementary row operations are used to reduce this matrix to

$$\left(\begin{array}{cccc|c} 1 & 0 & \cdots & 0 & s_1 \\ 0 & 1 & \cdots & 0 & s_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & s_n \end{array} \right).$$

The solution then appears in the last column (i.e., $x_i = s_i$) so that this procedure circumvents the need to perform back substitution.

Example Apply the Gauss–Jordan method to solve the following system:

$$\begin{aligned}2x_1 + 2x_2 + 6x_3 &= 4, \\2x_1 + x_2 + 7x_3 &= 6, \\-2x_1 - 6x_2 - 7x_3 &= -1.\end{aligned}$$

Solution: The sequence of operations is indicated in parentheses and the pivots are circled.

$$\begin{aligned}&\left(\begin{array}{ccc|c}\textcircled{2} & 2 & 6 & 4 \\2 & 1 & 7 & 6 \\-2 & -6 & -7 & -1\end{array}\right) R_1/2 \longrightarrow \left(\begin{array}{ccc|c}\textcircled{1} & 1 & 3 & 2 \\2 & 1 & 7 & 6 \\-2 & -6 & -7 & -1\end{array}\right) \begin{array}{l} R_2 - 2R_1 \\ R_3 + 2R_1 \end{array} \\&\longrightarrow \left(\begin{array}{ccc|c}\textcircled{1} & 1 & 3 & 2 \\0 & -1 & 1 & 2 \\0 & -4 & -1 & 3\end{array}\right) (-R_2) \longrightarrow \left(\begin{array}{ccc|c}1 & 1 & 3 & 2 \\0 & \textcircled{1} & -1 & -2 \\0 & -4 & -1 & 3\end{array}\right) \begin{array}{l} R_1 - R_2 \\ R_3 + 4R_2 \end{array} \\&\longrightarrow \left(\begin{array}{ccc|c}1 & 0 & 4 & 4 \\0 & \textcircled{1} & -1 & -2 \\0 & 0 & -5 & -5\end{array}\right) -R_3/5 \longrightarrow \left(\begin{array}{ccc|c}1 & 0 & 4 & 4 \\0 & 1 & -1 & -2 \\0 & 0 & \textcircled{1} & 1\end{array}\right) \begin{array}{l} R_1 - 4R_3 \\ R_2 + R_3 \end{array} \\&\longrightarrow \left(\begin{array}{ccc|c}1 & 0 & 0 & 0 \\0 & 1 & 0 & -1 \\0 & 0 & \textcircled{1} & 1\end{array}\right).\end{aligned}$$

Therefore, the solution is $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}$.

0.1 Vector spaces

0.1.1 Scalar field. Underlying a vector space is its *field*, or set of scalars. For our purposes, that underlying field is typically the real numbers $\mathbf{R}$ or the complex numbers $\mathbf{C}$ (see Appendix A), but it could be the rational numbers, the integers modulo a specified prime number, or some other field. When the field is unspecified, we denote it by the symbol $\mathbf{F}$. To qualify as a field, a set must be closed under two binary operations: “addition” and “multiplication.” Both operations must be associative and commutative, and each must have an identity element in the set; inverses must exist in the set for all elements under addition and for all elements except the additive identity under multiplication; multiplication must be distributive over addition.

0.1.2 Vector spaces. A *vector space* V over a field $\mathbf{F}$ is a set V of objects (called *vectors*) that is closed under a binary operation (“addition”) that is associative and commutative and has an identity (the zero vector, denoted by 0) and additive inverses in the set. The set is also closed under an operation of “scalar multiplication” of the vectors by elements of the scalar field $\mathbf{F}$, with the following properties for all $a, b \in \mathbf{F}$ and all $x, y \in V$: $a(x + y) = ax + ay$, $(a + b)x = ax + bx$, $a(bx) = (ab)x$, and $ex = x$ for the multiplicative identity $e \in \mathbf{F}$.

0.1 Vector spaces

Vector Space Definition

The set $\mathcal{V}$ is called a *vector space over* $\mathcal{F}$ when the vector addition and scalar multiplication operations satisfy the following properties.

- (A1) $\mathbf{x} + \mathbf{y} \in \mathcal{V}$ for all $\mathbf{x}, \mathbf{y} \in \mathcal{V}$. This is called the *closure property for vector addition*.
- (A2) $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$ for every $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{V}$.
- (A3) $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ for every $\mathbf{x}, \mathbf{y} \in \mathcal{V}$.
- (A4) There is an element $\mathbf{0} \in \mathcal{V}$ such that $\mathbf{x} + \mathbf{0} = \mathbf{x}$ for every $\mathbf{x} \in \mathcal{V}$.
- (A5) For each $\mathbf{x} \in \mathcal{V}$, there is an element $(-\mathbf{x}) \in \mathcal{V}$ such that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$.
- (M1) $\alpha \mathbf{x} \in \mathcal{V}$ for all $\alpha \in \mathcal{F}$ and $\mathbf{x} \in \mathcal{V}$. This is the *closure property for scalar multiplication*.
- (M2) $(\alpha\beta)\mathbf{x} = \alpha(\beta\mathbf{x})$ for all $\alpha, \beta \in \mathcal{F}$ and every $\mathbf{x} \in \mathcal{V}$.
- (M3) $\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$ for every $\alpha \in \mathcal{F}$ and all $\mathbf{x}, \mathbf{y} \in \mathcal{V}$.
- (M4) $(\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}$ for all $\alpha, \beta \in \mathcal{F}$ and every $\mathbf{x} \in \mathcal{V}$.
- (M5) $1\mathbf{x} = \mathbf{x}$ for every $\mathbf{x} \in \mathcal{V}$.

0.1 Vector spaces

Example 1.1

The *real coordinate spaces*

$$\mathbb{R}^{1 \times n} = \{ (x_1 \ x_2 \ \cdots \ x_n), x_i \in \mathbb{R} \} \quad \text{and} \quad \mathbb{R}^{n \times 1} = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, x_i \in \mathbb{R} \right\}$$

are special cases of the preceding example, and these will be the object of most of our attention. In the context of vector spaces, it usually makes no difference whether a coordinate vector is depicted as a row or as a column. When the row or column distinction is irrelevant, or when it is clear from the context, we will use the common symbol $\mathbb{R}^n$ to designate a coordinate space. In those cases where it is important to distinguish between rows and columns, we will explicitly write $\mathbb{R}^{1 \times n}$ or $\mathbb{R}^{n \times 1}$. Similar remarks hold for complex coordinate spaces.

0.1 Vector spaces

Example 1.2

Because (A1)–(A5) are generalized versions of the five additive properties of matrix addition, and (M1)–(M5) are generalizations of the five scalar multiplication properties given in §3.2, we can say that the following hold.

- The set $\mathfrak{R}^{m \times n}$ of $m \times n$ real matrices is a vector space over $\mathfrak{R}$.
- The set $\mathcal{C}^{m \times n}$ of $m \times n$ complex matrices is a vector space over $\mathcal{C}$.

0.1 Vector spaces

Example 1.3

With function addition and scalar multiplication defined by

$$(f + g)(x) = f(x) + g(x) \quad \text{and} \quad (\alpha f)(x) = \alpha f(x),$$

the following sets are vector spaces over $\mathbb{R}$:

- The set of functions mapping the interval $[0, 1]$ into $\mathbb{R}$.
- The set of all real-valued continuous functions defined on $[0, 1]$.
- The set of real-valued functions that are differentiable on $[0, 1]$.
- The set of all polynomials with real coefficients.

0.1 Vector spaces

Example 1.4

Consider the vector space $\mathbb{R}^2$, and let

$$\mathcal{L} = \{(x, y) \mid y = \alpha x\}$$

be a line through the origin. $\mathcal{L}$ is a subset of $\mathbb{R}^2$, but $\mathcal{L}$ is a special kind of subset because $\mathcal{L}$ also satisfies the properties **(A1)**–**(A5)** and **(M1)**–**(M5)** that define a vector space. This shows that it is possible for one vector space to properly contain other vector spaces.

0.1 Vector spaces

0.1.3 Subspaces, span, and linear combinations. A *subspace* of a vector space V over a field F is a subset of V that is, by itself, a vector space over F using the same operations of vector addition and scalar multiplication as in V . A subset of V is a subspace precisely when it is closed under these two operations. For example, $\{[a, b, 0]^T : a, b \in \mathbf{R}\}$ is a subspace of $\mathbf{R}^3$; see (0.2.5) for the transpose notation. An intersection of subspaces is always a subspace; a union of subspaces need not be a subspace. The subsets $\{0\}$ and V are always subspaces of V , so they are often called trivial subspaces; a subspace of V is said to be *nontrivial* if it is different from both $\{0\}$ and V . A subspace of V is said to be a proper subspace if it is not equal to V . We call $\{0\}$ the zero vector space. Since a vector space always contains the zero vector, a subspace cannot be empty.

If S is a subset of a vector space V over a field F , span S is the intersection of all subspaces of V that contain S . If S is nonempty, then $\text{span } S = \{a_1 v_1 + \cdots + a_k v_k : v_1, \dots, v_k \in S, a_1, \dots, a_k \in F, \text{ and } k = 1, 2, \dots\}$; if S is empty, it is contained in every subspace of V . The intersection of every subspace of V is the subspace $\{0\}$, so the definition ensures that $\text{span } S = \{0\}$. Notice that $\text{span } S$ is always a subspace even if S is not a subspace; S is said to *span* V if $\text{span } S = V$.

0.1 Vector spaces - Subspaces

Subspaces

Let $\mathcal{S}$ be a nonempty subset of a vector space $\mathcal{V}$ over $\mathcal{F}$ (symbolically, $\mathcal{S} \subseteq \mathcal{V}$). If $\mathcal{S}$ is also a vector space over $\mathcal{F}$ using the same addition and scalar multiplication operations, then $\mathcal{S}$ is said to be a ***subspace*** of $\mathcal{V}$. It's not necessary to check all 10 of the defining conditions in order to determine if a subset is also a subspace—only the closure conditions (A1) and (M1) need to be considered. That is, a nonempty subset $\mathcal{S}$ of a vector space $\mathcal{V}$ is a subspace of $\mathcal{V}$ if and only if

$$(A1) \quad \mathbf{x}, \mathbf{y} \in \mathcal{S} \implies \mathbf{x} + \mathbf{y} \in \mathcal{S}$$

and

$$(M1) \quad \mathbf{x} \in \mathcal{S} \implies \alpha \mathbf{x} \in \mathcal{S} \text{ for all } \alpha \in \mathcal{F}.$$

0.1 Vector spaces

A linear combination of vectors in a vector space V over a field $\mathbf{F}$ is any expression of the form $a_1v_1 + \cdots + a_kv_k$ in which k is a positive integer, $a_1, \dots, a_k \in \mathbf{F}$, and $v_1, \dots, v_k \in V$. Thus, the span of a nonempty subset S of V consists of all linear combinations of finitely many vectors in S . A linear combination $a_1v_1 + \cdots + a_kv_k$ is *trivial* if $a_1 = \cdots = a_k = 0$; otherwise, it is *nontrivial*. A linear combination is by definition a sum of *finitely many* elements of a vector space.

Let S_1 and S_2 be subspaces of a vector space over a field $\mathbf{F}$. The *sum* of S_1 and S_2 is the subspace

$$S_1 + S_2 = \text{span} \{S_1 \cup S_2\} = \{x + y : x \in S_1, y \in S_2\}$$

If $S_1 \cap S_2 = \{0\}$, we say that the sum of S_1 and S_2 is a direct sum and write it as $S_1 \oplus S_2$; every $z \in S_1 \oplus S_2$ can be written as $z = x + y$ with $x \in S_1$ and $y \in S_2$ in one and only one way.

0.1 Vector spaces

0.1.4 Linear dependence and linear independence. We say that a finite list of vectors $v_1, \dots, v_k$ in a vector space V over a field F is linearly dependent if and only if there are scalars $a_1, \dots, a_k \in F$, not all zero, such that $a_1x_1 + \dots + a_kx_k = 0$. Thus, a list of vectors $v_1, \dots, v_k$ is linearly dependent if and only if some nontrivial linear combination of $v_1, \dots, v_k$ is the zero vector. It is often convenient to say that “ $v_1, \dots, v_k$ are linearly dependent” instead of the more formal statement “the list of vectors $v_1, \dots, v_k$ is linearly dependent.” A list of vectors $v_1, \dots, v_k$ is said to have *length* k . A list of two or more vectors is linearly dependent if one of the vectors is a linear combination of some of the others; in particular, a list of two or more vectors in which two of the vectors in the list are identical is linearly dependent. Two vectors are linearly dependent if and only if one of the vectors is a scalar multiple of the other. A list consisting only of the zero vector is linearly dependent since $a_10 = 0$ for $a_1 = 1$.

A finite list of vectors $v_1, \dots, v_k$ in a vector space V over a field F is linearly independent if it is not linearly dependent. Again, it can be convenient to say that “ $v_1, \dots, v_k$ are linearly independent” instead of “the list of vectors $v_1, \dots, v_k$ is linearly independent.”

0.1 Vector spaces

0.1.5 Basis. A linearly independent list of vectors in a vector space V whose span is V is a basis for V . Each element of V can be represented as a linear combination of vectors in a basis in one and only one way; this is no longer true if any element whatsoever is appended to or deleted from the basis. A linearly independent list of vectors in V is a basis of V if and only if no list of vectors that properly contains it is linearly independent. A list of vectors that spans V is a basis for V if and only if no proper sublist of it spans V . The empty list is a basis for the zero vector space.

0.1.6 Extension to a basis. Any linearly independent list of vectors in a vector space V may be extended, perhaps in more than one way, to a basis of V . A vector space can have a basis that is not finite; for example, the infinite list of monomials $1, t, t^2, t^3, \dots$ is a basis for the real vector space of all polynomials with real coefficients; each polynomial is a unique linear combination of (finitely many) elements in the basis.

0.1 Vector spaces

0.1.7 Dimension. If there is a positive integer n such that a basis of the vector space V contains exactly n vectors, then every basis of V consists of exactly n vectors; this common cardinality of bases is the dimension of the vector space V and is denoted by $\dim V$. In this event, V is finite-dimensional; otherwise V is infinite-dimensional. In the infinite-dimensional case, there is a one-to-one correspondence between the elements of any two bases. The real vector space $\mathbf{R}^n$ has dimension n . The vector space $\mathbf{C}^n$ has dimension n over the field $\mathbf{C}$ but dimension $2n$ over the field $\mathbf{R}$. The basis $e_1, \dots, e_n$ of $\mathbf{F}^n$ in which each n -vector e_i has a 1 as its i th entry and 0s elsewhere is called the standard basis.

Let V be a finite-dimensional vector space and let S_1 and S_2 be two given subspaces of V . The subspace intersection lemma is

$$\dim(S_1 \cap S_2) + \dim(S_1 + S_2) = \dim S_1 + \dim S_2 \quad (0.1.7.1)$$

Rewriting this identity as

$$\begin{aligned} \dim(S_1 \cap S_2) &= \dim S_1 + \dim S_2 - \dim(S_1 + S_2) \\ &\geq \dim S_1 + \dim S_2 - \dim V \end{aligned} \quad (0.1.7.2)$$

0.2 Matrices

0.2.1 Rectangular arrays. A *matrix* is an m -by- n array of scalars from a field F . If $m = n$, the matrix is said to be square. The set of all m -by- n matrices over F is denoted by $M_{m,n}(F)$, and $M_{n,n}(F)$ is often denoted by $M_n(F)$. The vector spaces $M_{n,1}(F)$ and F^n are identical. If $F = \mathbb{C}$, then $M_n(\mathbb{C})$ is further abbreviated to M_n , and $M_{m,n}(\mathbb{C})$ to $M_{m,n}$. Matrices are typically denoted by capital letters, and their scalar entries are typically denoted by doubly subscripted lowercase letters.

Suppose that $A = [a_{ij}] \in M_{n,m}(F)$. The main diagonal of A is the list of entries $a_{11}, a_{22}, \dots, a_{qq}$, in which $q = \min\{n, m\}$. It is sometimes convenient to express the main diagonal of A as a vector $\text{diag } A = [a_{ii}]_{i=1}^q \in F^q$. The p th superdiagonal of A is the list $a_{1,p+1}, a_{2,p+2}, \dots, a_{k,p+k}$, in which $k = \min\{n, m - p\}$, $p = 0, 1, 2, \dots, m - 1$; the p th subdiagonal of A is the list $a_{p+1,1}, a_{p+2,2}, \dots, a_{p+\ell,\ell}$, in which $\ell = \min\{n - p, m\}$, $p = 0, 1, 2, \dots, n - 1$.

0.2 Matrices

0.2.2 Linear transformations. Let U be an n -dimensional vector space and let V be an m -dimensional vector space, both over the same field $\mathbf{F}$; let $\mathcal{B}_U$ be a basis of U and let $\mathcal{B}_V$ be a basis of V . We may use the isomorphisms $x \rightarrow [x]_{\mathcal{B}_U}$ and $y \rightarrow [y]_{\mathcal{B}_V}$ to represent vectors in U and V as n -vectors and m -vectors over $\mathbf{F}$, respectively. A linear transformation is a function $T : U \rightarrow V$ such that $T(a_1x_1 + a_2x_2) = a_1T(x_1) + a_2T(x_2)$ for any scalars a_1, a_2 and vectors x_1, x_2 . A matrix $A \in M_{m,n}(\mathbf{F})$ corresponds to a linear transformation $T : U \rightarrow V$ in the following way: $y = T(x)$ if and only if $[y]_{\mathcal{B}_V} = A[x]_{\mathcal{B}_U}$. The matrix A is said to represent the linear transformation T (relative to the bases $\mathcal{B}_U$ and $\mathcal{B}_V$); the representing matrix A depends on the bases chosen. When we study a matrix A , we realize that we are studying a linear transformation relative to a particular choice of bases, but explicit appeal to the bases is usually not necessary.

0.2.3 Vector spaces associated with a matrix or linear transformation. Any n -dimensional vector space over $\mathbf{F}$ may be identified with $\mathbf{F}^n$; we may think of $A \in M_{m,n}(\mathbf{F})$ as a linear transformation $x \rightarrow Ax$ from $\mathbf{F}^n$ to $\mathbf{F}^m$ (and also as an array). The domain of this linear transformation is $\mathbf{F}^n$; its range is $\text{range } A = \{y \in \mathbf{F}^m : y = Ax\}$ for some $x \in \mathbf{F}^n$; its null space is $\text{nullspace } A = \{x \in \mathbf{F}^n : Ax = 0\}$. The range of A is a subspace of $\mathbf{F}^m$, and the null space of A is a subspace of $\mathbf{F}^n$. The dimension of $\text{nullspace } A$ is denoted by $\text{nullity } A$; the dimension of $\text{range } A$ is denoted by $\text{rank } A$.

0.2 Matrices

Rank-nullity-Theorem

$$\dim(\text{range } A) + \dim(\text{nullspace } A) = \text{rank } A + \text{nullity } A = n$$

for $A \in M_{m,n}(F)$. The null space of A is a set of vectors in F^n whose entries satisfy m homogeneous linear equations. That is,

$$Ax = 0$$

or

$$\begin{cases} a_{11} + a_{12} + \dots + a_{1n} = 0 \\ a_{21} + a_{22} + \dots + a_{2n} = 0 \\ \dots \\ a_{m1} + a_{m2} + \dots + a_{mn} = 0 \end{cases} \quad \text{where } A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

0.2 Matrices

0.2.4 Matrix operations

- Addition: $A + B$
- Multiplication: AB (defined only when $A \in M_{m,n}(\mathbf{F})$ and $B \in M_{n,q}(\mathbf{F})$.)

Note: It is associative, but not always commutative.

0.2 Matrices

0.2.5 The transpose, conjugate transpose, and trace

- *Transpose* A^T , of course, $(A^T)^T = A$.
- *Conjugate transpose* $A^* = (\bar{A})^T$
(adjoint / Hermitian adjoint)
- *Trace* $\text{tr}A = a_{11} + \cdots + a_{qq}$

$$\text{tr}AA^* = \text{tr}A^*A = \sum_{i,j} |a_{i,j}|^2$$

0.2 Matrices

Symmetric: $A^T = A$,

Skew symmetric: $A^T = -A$

Orthogonal: $A^T A = I$;

Hermitian: $A^* = A$,

Skew Hermitian: $A^* = -A$

0.2 Matrices

Essentially Hermitian:

if $\exp(i\theta)A$ is Hermitian for some $\theta \in \mathbf{R}$

Unitary: $A^* A = I$

Normal: $A^* A = A A^*$

0.2 Matrices

0.2.6 Metamechanics of matrix multiplication

1. If $A \in M_{m,n}(\mathbb{F})$, $x \in \mathbb{F}^n$, and $y \in \mathbb{F}^m$, then the (column) vector Ax is a linear combination of the columns of A ; the coefficients of the linear combination are the entries of x . The row vector $y^T A$ is a linear combination of the rows of A ; the coefficients of the linear combination are the entries of y .
2. If b_j is the j th column of B and a_i^T is the i th row of A , then the j th column of AB is Ab_j and the i th row of AB is $a_i^T B$.
3. If $m = n$, then $A^T B = [a_i^T b_j]$: the i, j entry of $A^T B$ is the scalar $a_i^T b_j$.
4. If $p = q$, then $AB^T = \sum_{k=1}^n a_k b_k^T$: each summand is an m -by- n matrix, the *outer product* of a_k and b_k .

0.3 Determinants

0.3.1 Laplace expansion by minors along a row or column. The determinant may be defined inductively for $A = [a_{ij}] \in M_n(\mathbb{F})$ in the following way. Assume that the determinant is defined over $M_{n-1}(\mathbb{F})$ and let $A_{ij} \in M_{n-1}(\mathbb{F})$ denote the submatrix of $A \in M_n(\mathbb{F})$ obtained by deleting row i and column j of A . Then, for any $i, j \in \{1, \dots, n\}$, we have

$$\det A = \sum_{k=1}^n (-1)^{i+k} a_{ik} \det A_{ik} = \sum_{k=1}^n (-1)^{k+j} a_{kj} \det A_{kj} \quad (0.3.1.1)$$

The first sum is the *Laplace expansion by minors along row i* ; the second sum is the *Laplace expansion by minors along column j* . This inductive presentation begins by defining the determinant of a 1-by-1 matrix to be the value of the single entry.

Notice that $\det A^T = \det A$, $\det A^* = \overline{\det A}$ if $A \in M_n(\mathbb{C})$, and $\det I = 1$.

0.3 Determinants

0.3.2 Alternating sums and permutations. A *permutation* of $\{1, \dots, n\}$ is a one-to-one function $\sigma : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$. The *identity permutation* satisfies $\sigma(i) = i$ for each $i = 1, \dots, n$. There are $n!$ distinct permutations of $\{1, \dots, n\}$, and the collection of all such permutations forms a group under composition of functions.

Consistent with the low-dimensional examples in (0.3.1), for $A = [a_{ij}] \in M_n(\mathbf{F})$ we have the alternative presentation

$$\det A = \sum_{\sigma} \left(\operatorname{sgn} \sigma \prod_{i=1}^n a_{i\sigma(i)} \right) \quad (0.3.2.1)$$

in which the sum is over all $n!$ permutations of $\{1, \dots, n\}$ and $\operatorname{sgn} \sigma$, the “sign” or “signum” of a permutation σ , is $+1$ or -1 according to whether the minimum number of transpositions (pairwise interchanges) necessary to achieve it starting from $\{1, \dots, n\}$ is even or odd. We say that a permutation σ is *even* if $\operatorname{sgn} \sigma = +1$; σ is *odd* if $\operatorname{sgn} \sigma = -1$.

0.3 Determinants

0.3.3 Elementary row and column operations.

Type 1: Interchange of two rows.

For $i \neq j$, interchange of rows i and j of A results from left multiplication of A by

$$\begin{bmatrix} 1 & & & & & & & \\ & \ddots & & & & & & \\ & & 0 & & & & 1 & \\ & & & 1 & & & & \\ & \vdots & & & \ddots & & & \vdots \\ & & 1 & & & & & 0 \\ & & & \dots & & & & \\ & & & & & 1 & & \\ & & & & & & 1 & \ddots \\ & & & & & & & & 1 \end{bmatrix}$$

The two off-diagonal 1s are in the i, j and j, i positions, the two diagonal 0s are in positions i, i and j, j , and all unspecified entries are 0.

0.3 Determinants

Type 2: Multiplication of a row by a nonzero scalar.

Multiplication of row i of A by a nonzero scalar c results from left multiplication of A by

$$\begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & c & & \\ & & & & 1 & \\ & & & & & \ddots \\ & & & & & & 1 \end{bmatrix}$$

The i, i entry is c , all other main diagonal entries are 1, and all unspecified entries are 0.

0.3 Determinants

Type 3: Addition of a scalar multiple of one row to another row.

For $i \neq j$, addition of c times row i of A to row j of A results from left multiplication of A by

$$\begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & 1 & & \\ & & c & & \ddots & \\ & & & & & 1 \end{bmatrix}$$

The j, i entry is c , all main diagonal entries are 1, and all unspecified entries are 0. The displayed matrix illustrates the case in which $j > i$.

0.3 Determinants

0.3.4 Reduced row echelon form. To each $A = [a_{ij}] \in M_{m,n}(\mathbb{F})$ there corresponds a (unique) canonical form in $M_{m,n}(\mathbb{F})$, the reduced row echelon form, also known as the Hermite normal form.

If a row of A is nonzero, its leading entry is its first nonzero entry. The defining specifications of the RREF are as follows:

RREF

- (a) Any zero rows occur at the bottom of the matrix.
- (b) The leading entry of any nonzero row is a 1.
- (c) All other entries in the column of a leading entry are zero.
- (d) The leading entries occur in a staircase pattern, left to right; that is, if row i is nonzero and a_{ik} is its leading entry, then either $i = m$, or row $i + 1$ is zero, or the leading entry in row $i + 1$ is $a_{i+1,\ell}$, in which $\ell > k$.

For example,

$$\begin{bmatrix} 0 & 1 & -1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 & 0 & \pi \\ 0 & 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

0.3 Determinants

Row Echelon Form for Matrix

Row Echelon Form

An $m \times n$ matrix $\mathbf{E}$ with rows $\mathbf{E}_{i*}$ and columns $\mathbf{E}_{*j}$ is said to be in *row echelon form* provided the following two conditions hold.

- If $\mathbf{E}_{i*}$ consists entirely of zeros, then all rows below $\mathbf{E}_{i*}$ are also entirely zero; i.e., all zero rows are at the bottom.
- If the first nonzero entry in $\mathbf{E}_{i*}$ lies in the j^{th} position, then all entries below the i^{th} position in columns $\mathbf{E}_{*1}, \mathbf{E}_{*2}, \dots, \mathbf{E}_{*j}$ are zero.

These two conditions say that the nonzero entries in an echelon form must lie on or above a stair-step line that emanates from the upper-left-hand corner and slopes down and to the right. The pivots are the first nonzero entries in each row. A typical structure for a matrix in row echelon form is illustrated below with the pivots circled.

$$\begin{pmatrix} (*) & * & * & * & * & * & * & * \\ 0 & 0 & (*) & * & * & * & * & * \\ 0 & 0 & 0 & (*) & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & (*) & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Find the Value of a Determinants

Special form of determinant

$$|A| = \begin{vmatrix} a_{11} & a_{11} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_{nn} \end{vmatrix} = ?$$

Find the Value of a Determinants

Special form of determinant

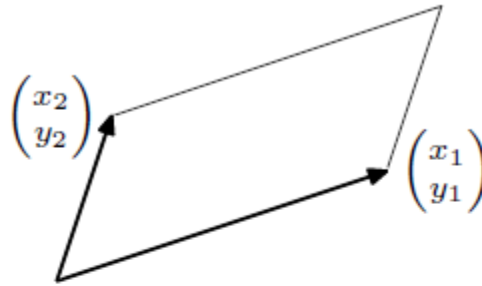
$$|A| = \begin{vmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_{nn} \end{vmatrix} = a_{11} \cdot a_{22} \cdot \dots \cdot a_{nn}$$

Geometry of Determinants

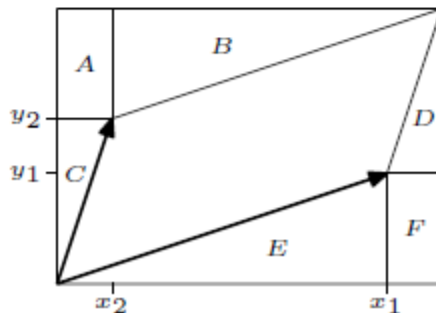
The prior section develops the determinant algebraically, by considering what formulas satisfy certain properties. This section complements that with a geometric approach. One advantage of this approach is that, while we have so far only considered whether or not a determinant is zero, here we shall give a meaning to the value of that determinant. (The prior section handles determinants as functions of the rows, but in this section columns are more convenient. The final result of the prior section says that we can make the switch.)

Determinants as Size Functions

This parallelogram picture



is familiar from the construction of the sum of the two vectors. One way to compute the area that it encloses is to draw this rectangle and subtract the area of each subregion.



area of parallelogram

$$\begin{aligned} &= \text{area of rectangle} - \text{area of } A - \text{area of } B \\ &\quad - \dots - \text{area of } F \\ &= (x_1 + x_2)(y_1 + y_2) - x_2y_1 - x_1y_1/2 \\ &\quad - x_2y_2/2 - x_2y_2/2 - x_1y_1/2 - x_2y_1 \\ &= x_1y_2 - x_2y_1 \end{aligned}$$

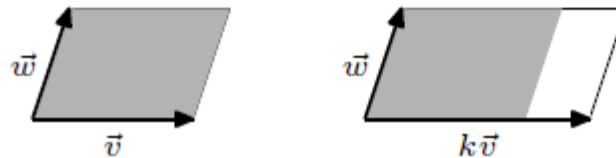
Geometry of Determinants

The fact that the area equals the value of the determinant

$$\begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} = x_1 y_2 - x_2 y_1$$

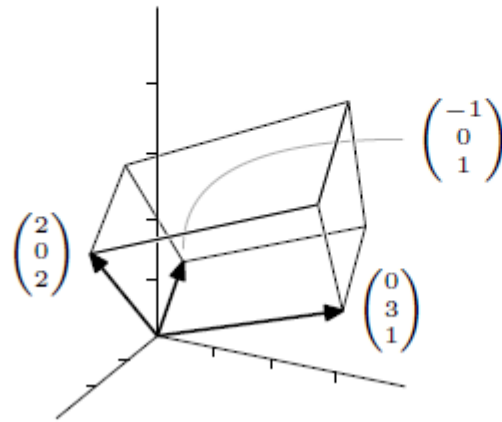
is no coincidence. The properties in the definition of determinants make reasonable postulates for a function that measures the size of the region enclosed by the vectors in the matrix.

For instance, this shows the effect of multiplying one of the box-defining vectors by a scalar (the scalar used is $k = 1.4$).



Geometry of Determinants

Example The volume of this parallelepiped, which can be found by the usual formula from high school geometry, is 12.



$$\begin{vmatrix} 2 & 0 & -1 \\ 0 & 3 & 0 \\ 2 & 1 & 1 \end{vmatrix} = 12$$

Volume, because it is an absolute value, does not depend on the order in which the vectors are given. The volume of the parallelepiped can also be computed as the absolute value of this determinant.

$$\begin{vmatrix} 0 & 2 & 0 \\ 3 & 0 & 3 \\ 1 & 2 & 1 \end{vmatrix} = -12$$

Geometry of Determinants

HomeWork

(a) Show that this gives the equation of a line in $\mathbb{R}^2$ thru (x_2, y_2) and (x_3, y_3) .

$$\begin{vmatrix} x & x_2 & x_3 \\ y & y_2 & y_3 \\ 1 & 1 & 1 \end{vmatrix} = 0$$

(b) Prove that the area of a triangle with vertices (x_1, y_1) , (x_2, y_2) , and (x_3, y_3) is

$$\frac{1}{2} \begin{vmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{vmatrix}.$$

(c) Prove that the area of a triangle with vertices at (x_1, y_1) , (x_2, y_2) , and (x_3, y_3) whose coordinates are integers has an area of N or $N/2$ for some positive integer N .

0.3 Determinants

0.3.5 Multiplicativity. A key property of the determinant function is that it is multiplicative: For $A, B \in M_n(\mathbf{F})$

$$\det AB = \det A \det B$$

This may be proved using elementary operations that row-reduce both A and B .

0.3.6 Functional characterization of the determinant. If we think of the determinant as a function of each row (or column) of a matrix separately with the others fixed, the Laplace expansion (0.3.1.1) reveals that the determinant is a linear function of the entries in any one given row (column). We summarize this property by saying that the function $A \rightarrow \det A$ is *multilinear* in the rows (columns) of A .

The determinant function $A \rightarrow \det A$ is the unique function $f : M_n(\mathbf{F}) \rightarrow \mathbf{F}$ that is

- (a) Multilinear in the rows of its argument
- (b) Alternating: any type 1 operation on A changes the sign of $f(A)$
- (c) Normalized: $f(I) = 1$

The permanent function is also multilinear (as are other generalized matrix functions), and it is normalized, but it is not alternating.

0.4 Rank

0.4.1 Definition. If $A \in M_{m,n}(\mathbf{F})$, $\text{rank } A = \dim \text{range } A$ is the length of a longest linearly independent list of columns of A . There can be more than one linearly independent list of columns whose length equals the rank. It is a remarkable fact that $\text{rank } A^T = \text{rank } A$. Therefore, an equivalent definition of rank is the length of a longest linearly independent list of rows of A : row rank = column rank.

0.4.2 Rank and linear systems. Let $A \in M_{m,n}(\mathbf{F})$ and $b \in \mathbf{F}^m$ be given. The linear system $Ax = b$ may have no solution, exactly one solution, or infinitely many solutions; these are the only possibilities. If there is at least one solution, the linear system is consistent; if there is no solution, the linear system is inconsistent. The linear system $Ax = b$ is consistent if and only if $\text{rank}[A \ b] = \text{rank } A$. The matrix $[A \ b] \in M_{m,n+1}(\mathbf{F})$ is the augmented matrix. To say that the augmented matrix and the coefficient matrix A of a linear system have the same rank is just to say that b is a linear combination of the columns of A . In this case, appending b to the columns of A does not increase the rank. A solution of the linear system $Ax = b$ is a vector x whose entries are the coefficients in a representation of b as a linear combination of the columns of A .

0.4 Rank

0.4.3 RREF and rank. Elementary operations do not change the rank of a matrix, and thus $\text{rank } A$ is the same as the rank of the RREF of A , which is just the number of nonzero rows in the RREF. As a practical matter, however, numerical calculation of the rank by calculation of the RREF is unwise. Round-off errors in intermediate numerical calculations can make zero rows of the RREF appear to be nonzero, thereby affecting perception of the rank.

0.4 Rank

0.4.4 Characterizations of rank. The following statements about a given matrix $A \in M_{m,n}(\mathbf{F})$ are equivalent; each can be useful in a different context. Note that in (b) and (c) the key issue is linear independence of *lists* of columns or rows of a matrix:

- (a) $\text{rank } A = k$.
- (b) k , and no more than k , rows of A are linearly independent.
- (c) k , and no more than k , columns of A are linearly independent.
- (d) Some k -by- k submatrix of A has nonzero determinant, and every $(k + 1)$ -by- $(k + 1)$ submatrix of A has zero determinant.
- (e) $\dim(\text{range } A) = k$.
- (f) There are k , but no more than k , linearly independent vectors $b_1, \dots, b_k$ such that the linear system $Ax = b_j$ is consistent for each $j = 1, \dots, k$.
- (g) $k = n - \dim(\text{nullspace } A)$ (the *rank-nullity theorem*).
- (h) $k = \min\{p : A = XY^T \text{ for some } X \in M_{m,p}(\mathbf{F}), Y \in M_{n,p}(\mathbf{F})\}$.
- (i) $k = \min\{p : A = x_1y_1^T + \dots + x_py_p^T\}$ for some $x_1, \dots, x_p \in \mathbf{F}^m, y_1, \dots, y_p \in \mathbf{F}^n$.

0.4 Rank

0.4.5 Rank inequalities. Some fundamental inequalities involving rank are:

- (a) If $A \in M_{m,n}(\mathbf{F})$, then $\text{rank } A \leq \min\{m, n\}$.
- (b) If one or more rows and/or columns are deleted from a matrix, the rank of the resulting submatrix is not greater than the rank of the original matrix.
- (c) Sylvester inequality: If $A \in M_{m,k}(\mathbf{F})$ and $B \in M_{k,n}(\mathbf{F})$, then

$$(\text{rank } A + \text{rank } B) - k \leq \text{rank } AB \leq \min\{\text{rank } A, \text{rank } B\}$$

- (d) The rank-sum inequality: If $A, B \in M_{m,n}(\mathbf{F})$, then

$$|\text{rank } A - \text{rank } B| \leq \text{rank}(A + B) \leq \text{rank } A + \text{rank } B \quad (0.4.5.1)$$

with equality in the second inequality if and only if $(\text{range } A) \cap (\text{range } B) = \{0\}$ and $(\text{range } A^T) \cap (\text{range } B^T) = \{0\}$. If $\text{rank } B = 1$ then

$$|\text{rank}(A + B) - \text{rank } A| \leq 1 \quad (0.4.5.2)$$

in particular, changing one entry of a matrix can change its rank by at most 1.

- (e) Frobenius inequality: If $A \in M_{m,k}(\mathbf{F})$, $B \in M_{k,p}(\mathbf{F})$, and $C \in M_{p,n}(\mathbf{F})$, then

$$\text{rank } AB + \text{rank } BC \leq \text{rank } B + \text{rank } ABC$$

with equality if and only if there are matrices X and Y such that $B = BCX + YAB$.

0.4 Rank

0.4.6 Rank equalities. Some fundamental equalities involving rank are:

- (a) If $A \in M_{m,n}(\mathbb{C})$, then $\text{rank } A^* = \text{rank } A^T = \text{rank } \bar{A} = \text{rank } A$.
- (b) If $A \in M_m(\mathbb{F})$ and $C \in M_n(\mathbb{F})$ are nonsingular and $B \in M_{m,n}(\mathbb{F})$, then $\text{rank } AB = \text{rank } B = \text{rank } BC = \text{rank } ABC$; that is, left or right multiplication by a nonsingular matrix leaves rank unchanged.
- (c) If $A, B \in M_{m,n}(\mathbb{F})$, then $\text{rank } A = \text{rank } B$ if and only if there exist a nonsingular $X \in M_m(\mathbb{F})$ and a nonsingular $Y \in M_n(\mathbb{F})$ such that $B = XAY$.
- (d) If $A \in M_{m,n}(\mathbb{C})$, then $\text{rank } A^* A = \text{rank } A$.
- (e) Full-rank factorization: If $A \in M_{m,n}(\mathbb{F})$, then $\text{rank } A = k$ if and only if $A = XY^T$ for some $X \in M_{m,k}(\mathbb{F})$ and $Y \in M_{n,k}(\mathbb{F})$ that each have independent columns. The equivalent factorization $A = XBY^T$ for some nonsingular $B \in M_k(\mathbb{F})$ can also be useful. In particular, $\text{rank } A = 1$ if and only if $A = xy^T$ for some nonzero vectors $x \in \mathbb{F}^m$ and $y \in \mathbb{F}^n$.
- (f) If $A \in M_{m,n}(\mathbb{F})$, then $\text{rank } A = k$ if and only if there exist nonsingular matrices $S \in M_m(\mathbb{F})$ and $T \in M_n(\mathbb{F})$ such that $A = S \begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix} T$.

0.5 Nonsingularity

A linear transformation or matrix is said to be nonsingular if it produces the output 0 only for the input 0. Otherwise, it is *singular*. If $A \in M_{m,n}(\mathbf{F})$ and $m < n$, then A is necessarily singular. An $A \in M_n(\mathbf{F})$ is invertible if there is a matrix $A^{-1} \in M_n(\mathbf{F})$ (the *inverse* of A) such that $A^{-1}A = I$. If $A \in M_n$ and $A^{-1}A = I$, then $AA^{-1} = I$; that is, A^{-1} is a left inverse if and only if it is a right inverse; A^{-1} is unique whenever it exists.

It is useful to be able to call on a variety of criteria for a square matrix to be nonsingular. The following are equivalent for a given $A \in M_n(\mathbf{F})$:

Equivalent formula for a given $A \in \text{Mn}(F)$:

- (a) A is nonsingular.
- (b) A^{-1} exists.
- (c) $\text{rank } A = n$.
- (d) The rows of A are linearly independent.
- (e) The columns of A are linearly independent.
- (f) $\det A \neq 0$.
- (g) The dimension of the range of A is n .
- (h) The dimension of the null space of A is 0.
- (i) $Ax = b$ is consistent for each $b \in F^n$.
- (j) If $Ax = b$ is consistent, then the solution is unique.
- (k) $Ax = b$ has a unique solution for each $b \in F^n$.
- (l) The only solution to $Ax = 0$ is $x = 0$.
- (m) 0 is not an eigenvalue of A (see Chapter 1).

Chapter 0 Review and Miscellanea

- o.0 Introduction
- o.1 Vector spaces
- o.2 Matrices
- o.3 Determinants
- o.4 Rank
- o.5 Nonsingularity
- o.6 The Euclidean inner product and norm**
- o.7 Partitioned sets and matrices**
- o.8 Determinants again**
- o.9 Special types of matrices**
- o.10 Change of basis**
- o.11 Equivalence relations**



Topics of today

0.6 The Euclidean inner product and norm

0.6.1 Definitions. The scalar $\langle x, y \rangle = y^*x$ is the Euclidean inner product (standard inner product, usual inner product, scalar product, dot product) of $x, y \in \mathbb{C}^n$. The Euclidean norm (usual norm, Euclidean length) function on $\mathbb{C}^n$ is the real-valued function $\|x\|_2 = \langle x, x \rangle^{1/2} = (x^*x)^{1/2}$; two important properties of this function are that $\|x\|_2 > 0$ for all nonzero $x \in \mathbb{C}^n$ and $\|\alpha x\|_2 = |\alpha| \|x\|_2$ for all $x \in \mathbb{C}^n$ and all $\alpha \in \mathbb{C}$.

The function $\langle \cdot, \cdot \rangle : \mathbb{C}^n \times \mathbb{C}^n \rightarrow \mathbb{C}$ is linear in the first argument and conjugate linear in the second; that is, $\langle \alpha x_1 + \beta x_2, y \rangle = \alpha \langle x_1, y \rangle + \beta \langle x_2, y \rangle$ and $\langle x, \alpha y_1 + \beta y_2 \rangle = \bar{\alpha} \langle x, y_1 \rangle + \bar{\beta} \langle x, y_2 \rangle$ for all $\alpha, \beta \in \mathbb{C}$ and $y_1, y_2 \in \mathbb{C}^n$. If V is a real or complex vector space and $f : V \times V \rightarrow \mathbb{C}$ is a function that is linear in its first argument and conjugate linear in its second argument, we say that f is sesquilinear on V ; f is a semi-inner product on V if it is sesquilinear on V and $f(x, x) \geq 0$ for every $x \in V$; f is an inner product on V if it is sesquilinear on V and $f(x, x) > 0$ for every nonzero $x \in V$. An inner product space is a pair (V, f) in which V is a real or complex vector space and f is an inner product on V .

0.6 The Euclidean inner product and norm

0.6.2 Orthogonality and orthonormality. Two vectors $x, y \in \mathbb{C}^n$ are orthogonal if $\langle x, y \rangle = 0$. In $\mathbb{R}^2$ and $\mathbb{R}^3$, “orthogonal” has the conventional geometric interpretation of “perpendicular.” A list of vectors $x_1, \dots, x_m \in \mathbb{C}^n$ is said to be orthogonal if $\langle x_i, x_j \rangle = 0$ for all distinct $i, j \in \{1, \dots, m\}$. An orthogonal list of nonzero vectors is linearly independent. A vector whose Euclidean norm is 1 is said to be normalized (a unit vector). For any nonzero $x \in \mathbb{C}^n$, $x / \|x\|_2$ is a unit vector. An orthogonal list of vectors is an orthonormal list if each of its elements is a unit vector. An orthonormal list of vectors is linearly independent. Each of these concepts has a straightforward generalization to the context of an inner product space.

0.6.3 The Cauchy–Schwarz inequality. The *Cauchy–Schwarz inequality* states that

$$|\langle x, y \rangle| \leq \|x\|_2 \|y\|_2$$

for all $x, y \in \mathbb{C}^n$, with equality if and only if one of the vectors is a scalar multiple of the other. The angle θ between two nonzero real vectors $x, y \in \mathbb{R}^n$ is defined by

$$\cos \theta = \frac{\langle x, y \rangle}{\|x\|_2 \|y\|_2}, \quad 0 \leq \theta \leq \pi \quad (0.6.3.1)$$

Properties of norm:

(1) $\|\alpha\| \geq 0$ if and only if $\|\alpha\| = 0$ when $\alpha = 0$

(2) $\|k\alpha\| = |k|\|\alpha\|, \quad k \in \mathbb{C}$

Triangular
Inequality

(3) $\|\alpha + \beta\| \leq \|\alpha\| + \|\beta\|$

(4) $|(\alpha, \beta)| \leq \|\alpha\| \|\beta\|$

Proof: $\|\alpha + \beta\|^2 = (\alpha + \beta, \alpha + \beta)$
 $= (\alpha, \alpha) + (\alpha, \beta) + (\beta, \alpha) + (\beta, \beta)$
 $= \|\alpha\|^2 + 2\operatorname{Re}(\alpha, \beta) + \|\beta\|^2$
 $\leq \|\alpha\|^2 + 2|(\alpha, \beta)| + \|\beta\|^2$
 $\leq \|\alpha\|^2 + 2\|\alpha\|\|\beta\| + \|\beta\|^2$
 $\leq (\|\alpha\| + \|\beta\|)^2$

Proof: $0 \leq \|\alpha - k\beta\|^2 = (\alpha - k\beta, \alpha - k\beta)$
 $= (\alpha, \alpha) - \bar{k}(\alpha, \beta) - k(\beta, \alpha) + k\bar{k}(\beta, \beta)$
Take $k = \frac{(\alpha, \beta)}{(\beta, \beta)}$, then...

Cauchy-Schwarz 不等式

Example 1. Let $M_{n \times n}(C)$ be the space of all the matrices with order $m \times n$. Prove that for $\forall A, B \in M_{n \times n}(C)$

$$\left| \text{Tr}(AB^H) \right| \leq \sqrt{\text{Tr}(AA^H)} \sqrt{\text{Tr}(BB^H)}$$

Example 2. Let $\tilde{C}[a, b]$ be the space of all the functions on the interval $[a, b]$. Prove that for $\forall f(x), g(x) \in \tilde{C}[a, b]$

$$\left| \int_a^b f(x) \overline{g(x)} dx \right| \leq \sqrt{\int_a^b |f(x)|^2 dx} \sqrt{\int_a^b |g(x)|^2 dx}$$

0.6 The Euclidean inner product and norm

0.6.4 Gram–Schmidt orthonormalization. Any finite independent list of vectors in an inner product space may be replaced by an orthonormal list with the same span. This replacement may be carried out in many ways, but there is a systematic way to do so that has a useful special property. The Gram–Schmidt process starts with a list of vectors $v_1, \dots, v_n$ and (if the given list is linearly independent) produces an orthonormal list of vectors $z_1, \dots, z_n$ such that $\text{span}\{z_1, \dots, z_k\} = \text{span}\{x_1, \dots, x_k\}$ for each $k = 1, \dots, n$. The vectors z_i may be calculated in turn as follows: Let $y_1 = x_1$ and normalize it: $z_1 = y_1 / \|y_1\|_2$. Let $y_2 = x_2 - \langle x_2, z_1 \rangle z_1$ (y_2 is orthogonal to z_1) and normalize it: $z_2 = y_2 / \|y_2\|_2$. Once $z_1, \dots, z_{k-1}$ have been determined, the vector

$$y_k = x_k - \langle x_k, z_{k-1} \rangle z_{k-1} - \langle x_k, z_{k-2} \rangle z_{k-2} - \dots - \langle x_k, z_1 \rangle z_1$$

is orthogonal to $z_1, \dots, z_{k-1}$; normalize it: $z_k = y_k / \|y_k\|_2$. Continue until $k = n$. If we denote $Z = [z_1 \ \dots \ z_n]$ and $X = [x_1 \ \dots \ x_n]$, the Gram–Schmidt process gives a factorization $X = ZR$, in which the square matrix $R = [r_{ij}]$ is nonsingular and upper triangular; that is, $r_{ij} = 0$ whenever $i > j$.

0.6 The Euclidean inner product and norm

0.6.5 Orthonormal bases. An orthonormal basis of an inner product space is a basis whose elements constitute an orthonormal list. Since any finite ordered basis may be transformed with the Gram–Schmidt process to an orthonormal basis, any finite-dimensional inner product space has an orthonormal basis, and any orthonormal list may be extended to an orthonormal basis. Such a basis is pleasant to work with, since the cross terms in inner product calculations all vanish.

0.6 The Euclidean inner product and norm

0.6.6 Orthogonal complements. Given any subset $S \subset \mathbb{C}^n$, the orthogonal complement of S is the set $S^\perp = \{x \in \mathbb{C}^n : x^*y = 0 \text{ for all } y \in S\}$. Even if S is not a subspace, $S^\perp$ is always a subspace. We have $(S^\perp)^\perp = \text{span } S$, and $(S^\perp)^\perp = S$ if S is a subspace. It is always the case that $\dim S^\perp + \dim(S^\perp)^\perp = n$. If S_1 and S_2 are subspaces, then $(S_1 + S_2)^\perp = S_1^\perp \cap S_2^\perp$.

For a given $A \in M_{m,n}$, $\text{range } A$ is the orthogonal complement of nullspace A^* . Therefore, for a given $b \in \mathbb{C}^m$, the linear system $Ax = b$ has a solution (not necessarily unique) if and only if $b^*z = 0$ for every $z \in \mathbb{C}^m$ such that $A^*z = 0$. This equivalence is sometimes stated as the Fredholm alternative (theorem of the alternative) — exactly one of the following two statements is true: *Either* (1) $Ax = b$ has a solution *or* (2) $A^*y = 0$ has a solution such that $y^*b \neq 0$.

If $A \in M_{m,n}$ and $B \in M_{m,q}$, if $X \in M_{m,r}$ and $Y \in M_{m,s}$, and if $\text{range } X = \text{nullspace } A^*$ and $\text{range } Y = \text{nullspace } B^*$, then we have the following companion to (0.2.7.1) and (0.2.7.2):

$$\text{range } A \cap \text{range } B = \text{nullspace} \begin{bmatrix} X^* \\ Y^* \end{bmatrix} \quad (0.6.6.1)$$

0.7 Partitioned sets and matrices

0.7.1 Submatrices. Let $A \in M_{m,n}(\mathbf{F})$. For index sets $\alpha \subseteq \{1, \dots, m\}$ and $\beta \subseteq \{1, \dots, n\}$, we denote by $A[\alpha, \beta]$ the (sub)matrix of entries that lie in the rows of A indexed by α and the columns indexed by β . For example,

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} [[1, 3], [1, 2, 3]] = \begin{bmatrix} 1 & 2 & 3 \\ 7 & 8 & 9 \end{bmatrix}$$

If $\alpha = \beta$, the submatrix $A[\alpha] = A[\alpha, \alpha]$ is a principal submatrix of A . An n -by- n matrix has $\binom{n}{k}$ distinct principal submatrices of size k .

For $A \in M_n(\mathbf{F})$ and $k \in \{1, \dots, n\}$, $A[\{1, \dots, k\}]$ is a leading principal submatrix and $A[\{k, \dots, n\}]$ is a trailing principal submatrix.

The determinant of an r -by- r submatrix of A is called a *minor*; if we wish to indicate the size of the submatrix, we call its determinant a *minor of size r* . If the r -by- r submatrix is a *principal* submatrix, then its determinant is a principal minor (of size r); if the submatrix is a *leading principal* matrix, then its determinant is a *leading principal minor*; if the submatrix is a *trailing principal* submatrix, then its determinant is a *trailing principal minor*. By convention, the empty principal minor is 1; that is, $\det A[\emptyset] = 1$.

0.7*** Partitioned sets and matrices

0.7.2 Partitions, block matrices, and multiplication. If $\alpha_1, \dots, \alpha_t$ constitute a partition of $\{1, \dots, m\}$ and $\beta_1, \dots, \beta_s$ constitute a partition of $\{1, \dots, n\}$, then the matrices $A[\alpha_i, \beta_j]$ form a partition of the matrix $A \in M_{m,n}(\mathbb{F})$, $1 \leq i \leq t$, $1 \leq j \leq s$. If $A \in M_{m,n}(\mathbb{F})$ and $B \in M_{n,p}(\mathbb{F})$ are partitioned so that the two partitions of $\{1, \dots, n\}$ coincide, the two matrix partitions are said to be conformal. In this event,

$$(AB)[\alpha_i, \gamma_j] = \sum_{k=1}^s A[\alpha_i, \beta_k] B[\beta_k, \gamma_j] \quad (0.7.2.1)$$

$$(A + B)[\alpha_i, \beta_j] = A[\alpha_i, \beta_j] + B[\alpha_i, \beta_j]$$

If a matrix is partitioned by sequential partitions of its rows and columns, the resulting partitioned matrix is called a block matrix. For example, if the rows and columns of $A \in M_n(\mathbb{F})$ are partitioned by the same sequential partition $\alpha_1 = \{1, \dots, k\}$, $\alpha_2 = \{k + 1, \dots, n\}$, the resulting block matrix is

$$A = \begin{bmatrix} A[\alpha_1, \alpha_1] & A[\alpha_1, \alpha_2] \\ A[\alpha_2, \alpha_1] & A[\alpha_2, \alpha_2] \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

0.8*** Determinants again

(Skip)

0.9 Special types of matrices

0.9.1 Diagonal matrices. A matrix $D = [d_{ij}] \in M_{n,m}(\mathbf{F})$ is *diagonal* if $d_{ij} = 0$ whenever $j \neq i$. If all the diagonal entries of a diagonal matrix are positive (non-negative) real numbers, we refer to it as a *positive (nonnegative) diagonal matrix*. The term *positive diagonal matrix* means that the matrix is diagonal and has positive diagonal entries; it does not refer to a general matrix with positive diagonal entries. The identity matrix $I \in M_n$ is a positive diagonal matrix. A square diagonal matrix D is a *scalar matrix* if its diagonal entries are all equal, that is, $D = \alpha I$ for some $\alpha \in \mathbf{F}$. Left or right multiplication of a matrix by a scalar matrix has the same effect as multiplying it by the corresponding scalar.

0.9 Special types of matrices

0.9.2 Block diagonal matrices and direct sums. A matrix $A \in M_n(\mathbf{F})$ of the form

$$A = \begin{bmatrix} A_{11} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & A_{kk} \end{bmatrix}$$

in which $A_{ii} \in M_{n_i}(\mathbf{F})$, $i = 1, \dots, k$, $\sum_{i=1}^k n_i = n$, and all blocks above and below the block diagonal are zero blocks, is called *block diagonal*. It is convenient to write such a matrix as

$$A = A_{11} \oplus A_{22} \oplus \cdots \oplus A_{kk} = \bigoplus_{i=1}^k A_{ii}$$

This is the *direct sum* of the matrices $A_{11}, \dots, A_{kk}$. Many properties of block diagonal matrices generalize those of diagonal matrices.

0.9 Special types of matrices

0.9.3 Triangular matrices. A matrix $T = [t_{ij}] \in M_{n,m}(\mathbb{F})$ is *upper triangular* if $t_{ij} = 0$ whenever $i > j$. If $t_{ij} = 0$ whenever $i \geq j$, then T is said to be strictly upper triangular. Analogously, T is lower triangular (or strictly lower triangular) if its transpose is upper triangular (or strictly upper triangular). A *triangular matrix* is either lower or upper triangular; a strictly triangular matrix is either strictly upper triangular or strictly lower triangular. A unit triangular matrix is a triangular matrix (upper or lower) that has ones on its main diagonal. Sometimes the terms *right* (in place of *upper*) and *left* (in place of *lower*) are used to describe triangular matrices.

0.9.4 Block triangular matrices. A matrix $A \in M_n(\mathbb{F})$ of the form

$$A = \begin{bmatrix} A_{11} & \star & \star \\ & \ddots & \star \\ \mathbf{0} & & A_{kk} \end{bmatrix} \quad (0.9.4.1)$$

in which $A_{ii} \in M_{n_i}(\mathbb{F})$, $i = 1, \dots, k$, $\sum_{i=1}^k n_i = n$, and all blocks below the block diagonal are zero, is *block upper triangular*;

0.9 Special types of matrices

0.9.11 Vandermonde matrices and Lagrange interpolation. A *Vandermonde matrix* $A \in M_n(\mathbb{F})$ has the form

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{bmatrix} \quad (0.9.11.1)$$

in which $x_1, \dots, x_n \in \mathbb{F}$; that is, $A = [a_{ij}]$ with $a_{ij} = x_i^{j-1}$. It is a fact that

$$\det A = \prod_{\substack{i,j=1 \\ i>j}}^n (x_i - x_j) \quad (0.9.11.2)$$

so a Vandermonde matrix is nonsingular if and only if the parameters $x_1, \dots, x_n$ are distinct.

0.10 Change of basis

Let V be an n -dimensional vector space over the field $\mathbf{F}$, and let the list $\mathcal{B}_1 = v_1, v_2, \dots, v_n$ be a basis for V . Any vector $x \in V$ can be represented as $x = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n$ because $\mathcal{B}_1$ spans V . If there were some other representation of $x = \beta_1 v_1 + \beta_2 v_2 + \dots + \beta_n v_n$ in the same basis, then

$$0 = x - x = (\alpha_1 - \beta_1)v_1 + (\alpha_2 - \beta_2)v_2 + \dots + (\alpha_n - \beta_n)v_n$$

from which it follows that all $\alpha_i - \beta_i = 0$ because the list $\mathcal{B}_1$ is independent. Given the basis $\mathcal{B}_1$, the linear mapping

$$x \rightarrow [x]_{\mathcal{B}_1} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}, \quad \text{in which } x = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n$$

from V to $\mathbf{F}^n$ is well-defined, one-to-one, and onto. The scalars α_i are the coordinates of x with respect to the basis $\mathcal{B}_1$, and the column vector $[x]_{\mathcal{B}_1}$ is the unique $\mathcal{B}_1$ -coordinate representation of x .

0.10 Change of basis

Let $T : V \rightarrow V$ be a given linear transformation. The action of T on any $x \in V$ is determined once one knows the n vectors $Tv_1, Tv_2, \dots, Tv_n$, because any $x \in V$ has a unique representation $x = \alpha_1 v_1 + \dots + \alpha_n v_n$ and $Tx = T(\alpha_1 v_1 + \dots + \alpha_n v_n) = T(\alpha_1 v_1) + \dots + T(\alpha_n v_n) = \alpha_1 Tv_1 + \dots + \alpha_n Tv_n$ by linearity. Thus, the value of Tx is determined once $[x]_{\mathcal{B}_1}$ is known.

Let $\mathcal{B}_2 = \{w_1, w_2, \dots, w_n\}$ also be a basis for V (either different from or the same as $\mathcal{B}_1$) and suppose that the $\mathcal{B}_2$ -coordinate representation of Tv_j is

$$[Tv_j]_{\mathcal{B}_2} = \begin{bmatrix} t_{1j} \\ \vdots \\ t_{nj} \end{bmatrix}, \quad j = 1, 2, \dots, n$$

Then, for any $x \in V$, we have

$$[Tx]_{\mathcal{B}_2} = \left[\sum_{j=1}^n \alpha_j Tv_j \right]_{\mathcal{B}_2} = \sum_{j=1}^n \alpha_j [Tv_j]_{\mathcal{B}_2} = \sum_{j=1}^n \alpha_j \begin{bmatrix} t_{1j} \\ \vdots \\ t_{nj} \end{bmatrix} = \begin{bmatrix} t_{11} & \cdots & t_{1n} \\ \vdots & \ddots & \vdots \\ t_{n1} & \cdots & t_{nn} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}$$

0.10 Change of basis

The n -by- n array $[t_{ij}]$ depends on T and on the choice of the bases $\mathcal{B}_1$ and $\mathcal{B}_2$, but it does not depend on x . We define the $\mathcal{B}_1$ - $\mathcal{B}_2$ basis representation of T to be

$${}_{\mathcal{B}_2}[T]_{\mathcal{B}_1} = \begin{bmatrix} t_{11} & \cdots & t_{1n} \\ \vdots & \ddots & \vdots \\ t_{n1} & \cdots & t_{nn} \end{bmatrix} = [[Tv_1]_{\mathcal{B}_2} \cdots [Tv_n]_{\mathcal{B}_2}]$$

We have just shown that $[Tx]_{\mathcal{B}_2} = {}_{\mathcal{B}_2}[T]_{\mathcal{B}_1}[x]_{\mathcal{B}_1}$ for any $x \in V$. In the important special case $\mathcal{B}_2 = \mathcal{B}_1$, we have ${}_{\mathcal{B}_1}[T]_{\mathcal{B}_1}$, which is called the $\mathcal{B}_1$ basis representation of T .

Consider the identity linear transformation $I : V \rightarrow V$ defined by $Ix = x$ for all x . Then

$$[x]_{\mathcal{B}_2} = [Ix]_{\mathcal{B}_2} = {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}[x]_{\mathcal{B}_1} = {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}[Ix]_{\mathcal{B}_1} = {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1} {}_{\mathcal{B}_1}[I]_{\mathcal{B}_2}[x]_{\mathcal{B}_2}$$

for all $x \in V$. By successively choosing $x = w_1, w_2, \dots, w_n$, this identity permits us to identify each column of ${}_{\mathcal{B}_2}[I]_{\mathcal{B}_1} {}_{\mathcal{B}_1}[I]_{\mathcal{B}_2}$ and shows that

$${}_{\mathcal{B}_2}[I]_{\mathcal{B}_1} {}_{\mathcal{B}_1}[I]_{\mathcal{B}_2} = I_n$$

0.10 Change of basis

If we do the same calculation starting with $[x]_{\mathcal{B}_1} = [Ix]_{\mathcal{B}_1} = \cdots$, we find that

$${}_{\mathcal{B}_1}[I]_{\mathcal{B}_2} {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1} = I_n$$

Thus, every matrix of the form ${}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}$ is invertible and ${}_{\mathcal{B}_1}[I]_{\mathcal{B}_2}$ is its inverse. Conversely, every invertible matrix $S = [s_1 \ s_2 \ \dots \ s_n] \in M_n(\mathbf{F})$ has the form ${}_{\mathcal{B}_1}[I]_{\mathcal{B}}$ for some basis $\mathcal{B}$. We may take $\mathcal{B}$ to be the vectors $\{\tilde{s}_1, \tilde{s}_2, \dots, \tilde{s}_n\}$ defined by $[\tilde{s}_i]_{\mathcal{B}_1} = s_i$, $i = 1, 2, \dots, n$. The list $\mathcal{B}$ is independent because S is invertible.

Notice that

$${}_{\mathcal{B}_2}[I]_{\mathcal{B}_1} = [[Iv_1]_{\mathcal{B}_2} \ \dots \ [Iv_n]_{\mathcal{B}_2}] = [[v_1]_{\mathcal{B}_2} \ \dots \ [v_n]_{\mathcal{B}_2}]$$

so ${}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}$ describes how the elements of the basis $\mathcal{B}_1$ are formed from elements of the basis $\mathcal{B}_2$. Now let $x \in V$ and compute

$$\begin{aligned} {}_{\mathcal{B}_1}[T]_{\mathcal{B}_2} [x]_{\mathcal{B}_2} &= [Tx]_{\mathcal{B}_1} = [I(Tx)]_{\mathcal{B}_1} = {}_{\mathcal{B}_1}[I]_{\mathcal{B}_2} [Tx]_{\mathcal{B}_2} \\ &= {}_{\mathcal{B}_1}[I]_{\mathcal{B}_2} {}_{\mathcal{B}_2}[T]_{\mathcal{B}_1} [x]_{\mathcal{B}_1} = {}_{\mathcal{B}_1}[I]_{\mathcal{B}_2} {}_{\mathcal{B}_2}[T]_{\mathcal{B}_1} [Ix]_{\mathcal{B}_1} \\ &= {}_{\mathcal{B}_1}[I]_{\mathcal{B}_2} {}_{\mathcal{B}_2}[T]_{\mathcal{B}_1} [I]_{\mathcal{B}_1} [x]_{\mathcal{B}_1} \end{aligned}$$

0.10 Change of basis

By choosing $x = w_1, w_2, \dots, w_n$ successively, we conclude that

$${}_{\mathcal{B}_2}[T]_{\mathcal{B}_2} = {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1} {}_{\mathcal{B}_1}[T]_{\mathcal{B}_1} [I]_{\mathcal{B}_2} \quad (0.10.1.1)$$

This identity shows how the $\mathcal{B}_1$ basis representation of T changes if the basis is changed to $\mathcal{B}_2$. For this reason, the matrix ${}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}$ is called the $\mathcal{B}_1 - \mathcal{B}_2$ change of basis matrix.

Any matrix $A \in M_n(\mathbf{F})$ is a basis representation of some linear transformation $T : V \rightarrow V$, for if $\mathcal{B}$ is any basis of V , we can determine Tx by $[Tx]_{\mathcal{B}} = A[x]_{\mathcal{B}}$. For this T , a computation reveals that ${}_{\mathcal{B}}[T]_{\mathcal{B}} = A$.

0.11 Equivalence relations

Let S be a given set and let Γ be a given subset of $S \times S = \{(a, b) : a \in S \text{ and } b \in S\}$. Then Γ defines a relation on S in the following way: We say that a is related to b , written $a \sim b$, if $(a, b) \in \Gamma$. A relation on S is said to be an equivalence relation if it is (a) reflexive ($a \sim a$ for every $a \in S$), (b) symmetric ($a \sim b$ whenever $b \sim a$), and (c) transitive ($a \sim c$ whenever $a \sim b$ and $b \sim c$). An equivalence relation on S gives a disjoint partition of S in a natural way: If we define the equivalence class of any $a \in S$ by $S_a = \{b \in S : b \sim a\}$, then $S = \cup_{a \in S} S_a$, and for each $a, b \in S$, either $S_a = S_b$ (if $a \sim b$) or $S_a \cap S_b = \emptyset$ (if $a \not\sim b$). Conversely, any disjoint partition of S can be used to define an equivalence relation on S .

0.11 Equivalence relations

The following table lists several equivalence relations that arise in matrix analysis. The factors D_1 , D_2 , S , T , L , and R are square and nonsingular; U and V are unitary; L is lower triangular; R is upper triangular; D_1 and D_2 are diagonal; and A and B need not be square for equivalence, unitary equivalence, triangular equivalence, or diagonal equivalence.

Equivalence Relation $\sim$	$A \sim B$
congruence	$A = SBS^T$
unitary congruence	$A = UBU^T$
*congruence	$A = SBS^*$
consimilarity	$A = SB\bar{S}^{-1}$
equivalence	$A = SBT$
unitary equivalence	$A = UBV$
diagonal equivalence	$A = D_1BD_2$
similarity	$A = SBS^{-1}$
unitary similarity	$A = UBU^*$
triangular equivalence	$A = LBR$



The End of Chapter O