Linear Algebra

Hechen Hu

December 4, 2017

Contents

1

Solving Linear Equations

1.1 Systems of Linear Equations

Definition. A *linear equation* in the variables $x_1, \cdots, x_n$ is an equation that can be written in the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

where b and coefficients a_i are real or complex numbers. A *linear system* is a collection of one or more linear equations involving the same variables. A *solution* of the system is a list of numbers that makes each equation a true statement when their values are substituted for $x_1, \cdots, x_n$ respectively. The set of all possible solutions is called the *solution set* of the linear system. Two linear systems are called *equivalent* if they have the same solution set.

Definition. A linear system is *consistent* if it has either one solution or infinitely many solutions. If it has no solution, it is called *inconsistent*.

Definition. The *coefficient matrix* is the matrix where the coefficients of each variable in a system aligned in columns. If additionally the coefficient of the right-hand side of equations are added to the coefficient matrix, a new matrix called *augmented matrix* is generated.

Definition. *Elementary row operations* on a matrix include:

- (*Replacement*) Replace one row by the sum of itself and a multiple of another row.
- (*Interchange*) Interchange two rows.
- (*Scaling*) Multiply all entries in a row by a nonzero constant.

Two matrices are *row equivalent* if there is a sequence of elementary operations that transforms one matrix into the other.

Theorem 1.1.1. *If the augmented matrices of two linear systems are row equivalent, then the two systems have the same solution set.*

1.2 Row Reduction and Echelon Forms

Definition. A rectangular matrix is in *echelon form* if it has the following properties:

- All nonzero rows are above any rows of all zeros;
- Each leading entry of a row is in a column to the right of the leading entry of the row above it;
- All entries in a column below a leading entry are zero.

If a matrix in echelon form satisfies the following additional conditions, then it is in *reduced echelon form*:

- The leading entry in each nonzero row is 1;
- Each leading 1 is the only nonzero entry in its column.

Theorem 1.2.1. *Each matrix is row equivalent to an unique reduced echelon matrix.*

If a matrix A is row equivalent to an (reduced)echelon matrix U, U is called an *(reduced) echelon form of A*. The abbreviation RREF and REF are used for reduced (row) echelon form and (row) echelon form respectively.

Definition. A *pivot position* in a matrix A is a location in A that corresponds to a leading entry in an echelon form of A. A *pivot column* is a column of A that contains a pivot position.

Theorem 1.2.2. *A linear system is consistent iff the rightmost column of the augmented matrix* ***is not*** *a pivot column, that is, iff an echelon form of the augmented matrix has* ***no*** *row of the form*

$$\begin{bmatrix} 0 & \cdots & 0 & b \end{bmatrix} \qquad \text{with } b \text{ nonzero}$$

If a linear system is consistent, then the solution set contains either

- *a unique solution, when there are no free variables.*
- *infinitely many solutions, when there is at least one free variable.*

1.3 Vector Equations

Definition. A matrix with only one column is called a *column vector*, or simply a *vector*.

Definition. Given vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p$ in $\mathbb{R}^n$ and given scalars $c_1, c_2, \cdots, c_p$, the vector $\mathbf{y}$ defined by

$$\mathbf{y} = c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p$$

is called a *linear combination of* $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p$ *using weights* $c_1, c_2, \cdots, c_p$.

Definition. If $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p$ are in $\mathbb{R}^n$, then the set of all linear combinations of them is denoted by $\text{Span}\{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$ and is called the *subset of* $\mathbb{R}^n$ *spanned (or generated) by* $\mathbf{v}_1, \cdots, \mathbf{v}_p$. That is, $\text{Span}\{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$ is the collection of all vectors that can be written in the form

$$c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p$$

with $c_1, c_2, \cdots, c_p$ scalars.

1.4 The Matrix Equation $\mathbf{Ax} = \mathbf{b}$

Definition. If $\mathbf{A}$ is an $m \times n$ matrix, with columns $a_1, \cdots, a_n$, and if $\mathbf{x}$ is in $\mathbb{R}^n$, then the *product of* $\mathbf{A}$ *and* $\mathbf{x}$, denoted by $\mathbf{Ax}$, is the *linear combination of the columns of* $\mathbf{A}$ *using the corresponding entries in* $\mathbf{x}$ *as weights*, that is,

$$\mathbf{Ax} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \begin{bmatrix} x_1 \\ \cdots \\ x_n \end{bmatrix} = x_1a_1 + x_2a_2 + \cdots + x_na_n$$

$\mathbf{Ax}$ is defined only if the number of columns of $\mathbf{A}$ equals the number of entries in $\mathbf{x}$.

Definition. Equations having the form $\mathbf{Ax} = \mathbf{b}$ are called *matrix equations.*

Theorem 1.4.1. *If* $\mathbf{A}$ *is an* $m \times n$ *matrix, with columns* $a_1, \cdots, a_n$*, and* $\mathbf{b}$ *is in* $\mathbb{R}^m$*, the matrix equation*

$$\mathbf{Ax} = \mathbf{b}$$

has the same solution set as the vector equation

$$x_1a_1 + x_2a_1 + \cdots + x_na_n = \mathbf{b}$$

which has the same solution set as the system of linear equations whose augmented matrix is

$$\begin{bmatrix} a_1 & a_2 & \cdots & a_n & \mathbf{b} \end{bmatrix}$$

Definition. A set of vectors $\{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$ in $\mathbb{R}^m$ *spans (or generates)* $\mathbb{R}^m$ if $\text{Span}\{\mathbf{v}_1, \cdots, \mathbf{v}_p\} = \mathbb{R}^m$.

Theorem 1.4.2. *Let* $\mathbf{A}$ *be an* $m \times n$ *coefficient matrix. Then the following statements are logically equivalent, that is, for a particular* $\mathbf{A}$*, either they are all true statements or they are all false.*

- *For each* $\mathbf{b}$ *in* $\mathbb{R}^m$*, the equation* $\mathbf{Ax} = \mathbf{b}$ *has a solution.*
- *The columns of* $\mathbf{A}$ *spans* $\mathbb{R}^m$*.*
- $\mathbf{A}$ *has a pivot position in every row.*

Theorem 1.4.3. *If* $\mathbf{A}$ *is an* $m \times n$ *matrix,* $\mathbf{u}$ *and* $\mathbf{v}$ *are vectors in* $\mathbb{R}^n$*, and* c *is a scalar, then*

- $\mathbf{A}(\mathbf{u} + \mathbf{v}) = \mathbf{Au} + \mathbf{Av}$*.*
- $\mathbf{A}(c\mathbf{u}) = c(\mathbf{Au})$*.*

1.5 Solution Sets of Linear Systems

Definition. A system of Linear equations is said to be *homogeneous* if it can be written in the form $\mathbf{Ax} = \mathbf{0}$. Such a system always has at least one solution, namely, $\mathbf{x} = \mathbf{0}$, and this solution is usually called the *trivial solution.* A homogeneous equation has a nontrivial solution iff the equation has at least one free variable.

Definition. Vector addition can be considered as a *translation.* e.g. the vector $\mathbf{v}$ is *translated by* $\mathbf{p}$ to $\mathbf{v} + \mathbf{p}$.

Definition. A *parametric vector equation* can be written as

$$\mathbf{x} = s\mathbf{u} + t\mathbf{v} \qquad (s, t \in \mathbb{R})$$

which describes explicitly the spanned plane by $\mathbf{u}$ and $\mathbf{v}$. Whenever a solution set is described explicitly with vectors, we say that the solution is in *parametric vector form.*

Theorem 1.5.1. *Suppose the equation* $\mathbf{Ax} = \mathbf{b}$ *is consistent for some given* $\mathbf{b}$*, and let* $\mathbf{p}$ *be a nonzero solution. Then the solution set of it is the set of all vectors of the form* $\mathbf{w} = \mathbf{p} + \mathbf{v}_h$*, where* $\mathbf{v_h}$ *is any solution of the homogeneous equation* $\mathbf{Ax} = \mathbf{0}$*.*

1.6 Linear Independence

Definition. An indexed set of vectors $\{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$ in $\mathbb{R}^n$ is said to be *linearly independent* if the vector equation

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_p\mathbf{v}_p = \mathbf{0}$$

has only the trivial solution. The set $\{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$ is said to be *linearly dependent* if there exist weights $c_1, \cdots, c_p$, not all zero, such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p = \mathbf{0}$$

and this equation is called a *linear dependence relation* among $\mathbf{v}_1, \cdots, \mathbf{v}_p$.

Theorem 1.6.1. *The columns of a matrix* $\mathbf{A}$ *are linearly independent iff the equation* $\mathbf{Ax} = \mathbf{0}$ *has **only** the trivial solution.*

Theorem 1.6.2. *A set of two vectors* $\{\mathbf{v}_1, \mathbf{v}_2\}$ *is linearly dependent iff one of the vectors is a multiple of the other.*

Theorem 1.6.3. *An indexed set* $S = \{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$ *of two or more vectors is linearly dependent iff at least one of the vectors in* S *is a linear combination of the others.*

Theorem 1.6.4. *Any set* $\{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$ *in* $\mathbb{R}^n$ *is linearly dependent if* $p > n$*(Same as the criterion for the existence of solutions in a system of equations).*

Theorem 1.6.5. *If a set* $S = \{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$ *in* $\mathbb{R}^n$ *contains the zero vector, then the set is linearly dependent.*

1.7 Linear Transformations

Definition. A *transformation*(or *function* or *mapping*) from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a rule that assigns to each vector $\mathbf{x} \in \mathbb{R}^n$ a vector $T(\mathbf{x}) \in \mathbb{R}^m$. $\mathbb{R}^n$ is called the *domain* of T, and $\mathbb{R}^m$ is called the *codomain* of T. For $\mathbf{x} \in \mathbb{R}^n$, the vector $T(\mathbf{x}) \in \mathbb{R}^m$ is called the *image* of $\mathbf{x}$ under T. The set of all images $T(\mathbf{x})$ is called the *range* of T.

***Example* 1.1.** *Given a scalar* r*, define* $T : \mathbb{R}^2 \to \mathbb{R}^2$ *by* $T(\mathbf{x}) = r\mathbf{x}$*.* T *is called a contraction when* $0 \leqslant r \leqslant 1$ *and a dilation when* $r > 1$*.*

Theorem 1.7.1. *Let* $T : \mathbb{R}^n \to \mathbb{R}^m$ *be a linear transformation. Then there exists a unique matrix* $\mathbf{A}$ *such that*

$$T(\mathbf{x}) = \mathbf{Ax} \quad \forall \mathbf{x} \in \mathbb{R}^n$$

In fact, $\mathbf{A}$ *is the* $m \times n$ *matrix whose* j*th column is the vector* $T(\mathbf{e}_j)$*, where* $\mathbf{e}_j$ *is the* j*th column of the identity matrix in* $\mathbb{R}^n$*.*

$$\mathbf{A} = \begin{bmatrix} T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n) \end{bmatrix}$$

The matrix $\mathbf{A}$ *is called the standard matrix for the linear transformation* T*.*

Theorem 1.7.2. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then T is injective iff the equation $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution.*

Theorem 1.7.3. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and let $\mathbf{A}$ be the standard matrix for T. Then*

- *T is surjective iff the columns of $\mathbf{A}$ span $\mathbb{R}^m$;*
- *T is injective iff the columns of $\mathbf{A}$ are linearly independent.*

Definition. If there is a matrix $\mathbf{A}$ such that

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k \quad \text{for } k = 0, 1, 2, \cdots$$

then the equation above is called a *linear difference equation* (or *recurrence relation*).

2

Matrices

2.1 Matrices and Arithmetic Operations on Them

Definition. A *diagonal matrix* is a square matrix whose nondiagonal entries are zero.

Definition. Two matrices are equal if they have the same size and each entries are equal.

Definition. The *sum* of two matrices is the sum of each corresponding entries in these two matrices. Thus the sum is defined only when they have the same size.

Definition. The *scalar multiple* of a matrix has entries of the product of the scalar and each corresponding original entries.

Theorem 2.1.1. *The set of matrices of the same size with respect to matrix addition and scalar multiplication over the field of real numbers is a vector space.*

Definition. If $\mathbf{A}$ is an $m \times n$ matrix, and if $\mathbf{B}$ is an $n \times p$ matrix with columns $\mathbf{b}_1, \cdots, \mathbf{b}_p$, then the *product* $\mathbf{AB}$ is the $m \times p$ matrix whose columns are $\mathbf{Ab}_1, \cdots, \mathbf{Ab}_p$. Multiplication of matrices corresponds to composition of linear transformations.

Theorem 2.1.2. *The multiplication has the following properties:*

- *Associativity of multiplication;*
- *Left distribution;*
- *Right distribution;*
- *Associativity over scalar multiplication;*

- *Identity for matrix multiplication; i.e. If* $\mathbf{A}$ *is a matrix of size* $m \times n$*, then*
$$\mathbf{I}_m\mathbf{A} = \mathbf{A} = \mathbf{A}\mathbf{I}_n$$
where $\mathbf{I}_n$ *is the* $n \times n$ *identity matrix.*

Definition. In general, matrix multiplication is not commutative and the cancellation law do not hold. When two matrices' multiplication is commutative, they are said to be *commute* with one another. Also, if a product $\mathbf{AB}$ is the zero matrix, in general it does not mean that either $\mathbf{A} = \mathbf{0}$ or $\mathbf{B} = \mathbf{0}$.

Definition. If $\mathbf{A}$ is an $m \times n$ matrix and k is a positive integer, then $\mathbf{A}^k$ denoted the product of k copies of $\mathbf{A}$, i.e. the kth power of $\mathbf{A}$. The 0th power of a matrix is the identity matrix.

Definition. If $\mathbf{A}$ is an $m \times n$ matrix, the *transpose* of $\mathbf{A}$ is the $n \times m$ matrix, denoted $\mathbf{A}^T$, whose columns are formed from the corresponding rows of $\mathbf{A}$.

Theorem 2.1.3. *The transpose operation has the following properties:*

- $(\mathbf{A}^T)^T = \mathbf{A}$*;*
- $(\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$*;*
- *Associativity with scalar multiplication;*
- $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$*, that is, the transpose of a product of arbitrary number of matrices equals the product of their transpose in the reverse order.*

2.2 The Inverse of a Matrix

Definition. If $\mathbf{A}$ is an $n \times n$ matrix, then if
$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_n$$
we say that $\mathbf{A}$ is *invertible* and $\mathbf{A}^{-1}$ an *inverse* of $\mathbf{A}$. The inverse of a matrix is unique. If a matrix is not invertible, it is called a *singular matrix.*

Definition. The *determinant* of the matrix $\mathbf{A}$
$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$
denoted $\det \mathbf{A}$ and equals $ad - bc$. Determinant is only defined for a square matrix, but the procedure above can be repeated on higher dimension matrices, for example
$$\det\begin{bmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \\ m & n & o & p \end{bmatrix} = a\det\begin{bmatrix} f & g & h \\ j & k & l \\ n & o & p \end{bmatrix} - b\det\begin{bmatrix} e & g & h \\ i & k & l \\ m & o & p \end{bmatrix} + c\det\begin{bmatrix} e & f & h \\ i & j & l \\ m & n & p \end{bmatrix} - d\det\begin{bmatrix} e & f & g \\ i & j & k \\ m & n & o \end{bmatrix}$$

By the Leibniz formula for the determinant of an $n \times n$ matrix $\mathbf{A}$ is

$$\det(\mathbf{A}) = \sum_{\sigma \in S_n} (\operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma_i})$$

Here the sum is computed over all permutations σ of the set $\{1, 2, \ldots, n\}$. A permutation is a function that reorders this set of integers. The value in the ith position after the reordering σ is denoted by σ_i. For example, for $n = 3$, the original sequence $1, 2, 3$ might be reordered to $\sigma = [2, 3, 1]$, with $\sigma_1 = 2$, $\sigma_2 = 3$, and $\sigma_3 = 1$. The set of all such permutations (also known as the symmetric group on n elements) is denoted by S_n.
For each permutation σ, $\operatorname{sgn}(\sigma)$ denotes the signature of σ, a value that is $+1$ whenever the reordering given by σ can be achieved by successively interchanging two entries an even number of times, and -1 whenever it can be achieved by an odd number of such interchanges.

Definition. If $\mathbf{A}$ is a square matrix, then the *minor* of the entry in the i-th row and j-th column (also called the (i, j) *minor*, or a *first minor*) is the determinant of the submatrix formed by deleting the i-th row and j-th column. This number is often denoted Mi, j. The (i, j) cofactor is obtained by multiplying the minor by $(-1)^{i+j}$.

In general, let A be an $m \times n$ matrix and k an integer with $0 < k \leqslant m$, and $k \leqslant n$. A $k \times k$ minor of $\mathbf{A}$, also called minor determinant of order k of $\mathbf{A}$ or, if $m = n$, $(n - k)$th minor determinant of $\mathbf{A}$, is the determinant of a $k \times k$ matrix obtained from $\mathbf{A}$ by deleting $m - k$ rows and $n - k$ columns.

Definition. The matrix formed by all of the cofactors of a square matrix A is called the *cofactor matrix.*

Definition. The *adjugate* is the transpose of the cofactor matrix of it, that is, if $\mathbf{A}$ is a matrix and $\mathbf{C}$ is its cofactor matrix, then

$$\operatorname{Adj}(\mathbf{A}) = \mathbf{C}^T$$

Theorem 2.2.1. *A matrix* $\mathbf{A}$ *is invertible only if* $\det(\mathbf{A}) \neq 0$, *and in this case*

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \operatorname{Adj}(\mathbf{A})$$

Theorem 2.2.2. *For a matrix* $\mathbf{A}$

$$\mathbf{A} \operatorname{Adj}(\mathbf{A}) = \det(\mathbf{A})\mathbf{I}$$

Theorem 2.2.3. *If* $\mathbf{A}$ *is an invertible* $n \times n$ *matrix, then for each* $\mathbf{b} \in \mathbb{R}^n$, *the equation* $\mathbf{Ax} = \mathbf{b}$ *has the unique solution* $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$.

Theorem 2.2.4.
- *The inverse of the inverse of a invertible matrix is the matrix itself.*
- *The inverse of the product of arbitrary number of invertible square matrices is the product of the inverse of themselves multiplied in the reverse order.*
- *The transpose of a invertible matrix is also invertible. Moreover, the inverse of a matrix's transpose is the transpose of the matrix's inverse.*

Definition. An *elementary matrix* is a matrix obtained by performing a single elementary row operation on a identity matrix.

Theorem 2.2.5. *If an elementary row operations is performed on an $m \times n$ matrix $\mathbf{A}$, the resulting matrix can be written as $\mathbf{EA}$, where the $m \times m$ matrix $\mathbf{E}$ is created by performing the same row operation on $\mathbf{I}_m$.*

Theorem 2.2.6. *Each elementary matrix $\mathbf{E}$ is invertible. The inverse of $\mathbf{E}$ is the elementary matrix of the same type that transforms $\mathbf{E}$ back into $\mathbf{I}$.*

Theorem 2.2.7. *An $n \times n$ matrix $\mathbf{A}$ is invertible iff $\mathbf{A}$ is a row equivalent to $\mathbf{I}_n$, and in this case, any sequence of elementary row operations that reduces $\mathbf{A}$ to $\mathbf{I}_n$ also transforms $\mathbf{I}_n$ into $\mathbf{A}^{-1}$.*

Theorem 2.2.8. *Let $\mathbf{A}$ be a square $n \times n$ matrix. Then the following statements are equivalent.*

- *$\mathbf{A}$ is an invertible matrix.*
- *$\mathbf{A}$ is row equivalent to the $n \times n$ identity matrix.*
- *$\mathbf{A}$ has n pivot positions.*

- *The equation $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution.*
- *The columns of $\mathbf{A}$ form a linearly independent set.*
- *The linear transformation $\mathbf{x} \mapsto \mathbf{Ax}$ is injective.*

- *The equation $\mathbf{Ax} = \mathbf{b}$ has at least one solution for each $\mathbf{b}$ in $\mathbb{R}^n$.*
- *The columns of $\mathbf{A}$ span $\mathbb{R}^n$.*
- *The linear transformation $\mathbf{x} \mapsto \mathbf{Ax}$ maps $\mathbb{R}^n$ onto $\mathbb{R}^n$.*

- *There is an* $n \times n$ *matrix* $\mathbf{C}$ *such that* $\mathbf{CA} = \mathbf{I}$.
- *There is an* $n \times n$ *matrix* $\mathbf{D}$ *such that* $\mathbf{AD} = \mathbf{I}$.
- $\mathbf{A}^T$ *is an invertible matrix.*

Proposition. Let $\mathbf{A}$ and $\mathbf{B}$ be square matrices. If $\mathbf{AB} = \mathbf{I}$, then $\mathbf{A}$ and $\mathbf{B}$ are both invertible, with $\mathbf{B} = \mathbf{A}^{-1}$ and $\mathbf{A} = \mathbf{B}^{-1}$

Definition. A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is *invertible* if there exists a function $S : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$\begin{array}{ll} (\forall \mathbf{x} \in \mathbb{R}^n) & S(T(\mathbf{x})) = \mathbf{x} \\ (\forall \mathbf{x} \in \mathbb{R}^n) & T(S(\mathbf{x})) = \mathbf{x} \end{array}$$

and S is called the *inverse* of T and denoted T^{-1}.

Theorem 2.2.9. *A linear transformation is invertible iff its standard matrix is invertible. In this case its inverse is unique.*

3

Vector Spaces and Subspaces

4

Orthogonality

5

Determinants

6

Eigenvalues and Eigenvectors

7

The Singular Value Decomposition(SVD)

8

Linear Transformations

9

Complex Vectors and Matrices