Linear Algebra: Theory and Techniques

Xi Tan (tan19@purdue.edu)

June 6, 2014

Contents

Preface

Some good books to consider:

1. Linear Algebra Its Applications, by Strang
2. Linear Algebra Right, by Axler
3. Linear Algebra, by Lang
4. Finite Dimensional Spaces Mathematics Studies, by Halmos
5. Linear Algebra Problem Book, by Halmos
6. Linear Algebra and Its Applications, by Lax
7. http://joshua.smcvt.edu/linearalgebra/book.pdf
8. http://www.math.brown.edu/ treil/papers/LADW/book.pdf

Part I

Theory

Chapter 1

Preliminaries

Definition 1.0.1 A **field** is a non-empty set F *closed* under two operations, usually called *addition* and *multiplication*[1], and denoted by $+$ and $\cdot$ respectively, such that the following *nine* axioms hold

(1-2). Associativity of addition and multiplication.

(3-4). Commutativity of addition and multiplication.

(5-6). Existence and uniqueness of additive and multiplicative identity elements.

(7-8). Existence and uniqueness of additive inverses and multiplicative inverses.

(9). Distributivity of multiplication over addition.

Definition 1.0.2 The characteristic of a ring R, $char(R)$, is the smallest positive integer n such that

$$\underbrace{1 + \cdots + 1}_{n \text{ summands}} = 0$$

Theorem 1.0.1 Any finite ring has nonzero characteristic.

[1]Subtraction and division are defined implicitly in terms of the inverse operations of addition and multiplication.

Chapter 2

Vector Calculus

2.1 Vector Algebra

2.1.1 Dot Product

2.1.2 Cross Product

2.1.3 Scalar Triple Product

2.1.4 Vector Triple Product

2.2 Line, Surface, and Volume Integrals

Chapter 3

Vector Spaces

3.1 Vector Space

Definition 3.1.1 A **vector space** over a field $\mathbb{F}$ is a *nonempty* set V together with the operations of addition $V \times V \to V$ and scalar multiplication $\mathbb{F} \times V \to V$ satisfying the following *eight* properties:

(-) Additive axioms. For every $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$, we have

(1) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$

(2) $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$

(3) $\mathbf{0} + \mathbf{u} = \mathbf{u} + \mathbf{0} = \mathbf{u}$, where $\mathbf{0} \in V$ is unique for all $\mathbf{u} \in V$

(4) $(-\mathbf{u}) + \mathbf{u} = \mathbf{u} + (-\mathbf{u}) = \mathbf{0}$, where $-\mathbf{u} \in V$ is unique for every $\mathbf{u} \in V$

(-) Multiplicative axioms. For every $\mathbf{u} \in V$ and scalars $a, b \in \mathbb{F}$, we have

(1) $1\mathbf{x} = \mathbf{x}$

(2) $(ab)\mathbf{x} = a(b\mathbf{x})$

(-) Distributive axioms. For every $\mathbf{u}, \mathbf{v} \in V$ and scalars $a, b \in \mathbb{F}$, we have

(1) a($\mathbf{u}$+$\mathbf{v}$) = a$\mathbf{u}$+ a$\mathbf{v}$

(2) (a+b)$\mathbf{u}$= a$\mathbf{u}$+ b$\mathbf{u}$

3.2 Subspaces

Definition 3.2.1 A subspace of $\mathbb{R}^n$ is any collection S of vectors in $\mathbb{R}^n$ such that

(1) The zero vector $\mathbf{0}$ is in S.

(2) If $\mathbf{u}$ and $\mathbf{v}$ are in S, then $\mathbf{u} + \mathbf{v}$ is in S. [1]

(3) If $\mathbf{u}$ is in S and c is a scalar, then $c\mathbf{u}$ is in S. [2]

Definition 3.2.2 Let S, T be two subspaces of $\mathbb{R}^n$. We say S is orthogonal to T if *every* vector in S is orthogonal to *every* vector in T. The subspace $\{\mathbf{0}\}$ is orthogonal to all subspaces. [3]

Definition 3.2.3 Let A be an $m \times n$ matrix.

(1) The *row space* of A is the subspace *row*(A) of $\mathbb{R}^n$ spanned by the rows of A.

(2) The *column space* (or *range*) of A is the subspace *col*(A) of $\mathbb{R}^m$ spanned by the columns of A.

[1] S is closed under addition.
[2] S is closed under scalar multiplication.
[3] A line can be orthogonal to another line, or it can be orthogonal to a plane, but a plane cannot be orthogonal to a plane.

3.2.1 Four Important Subspaces: the row, column, null, and left null space

Definition 3.2.4 Let A be an $m \times n$ matrix. The *null space* (or *kernel*) of A is the subspace of $\mathbb{R}^n$ consisting of solutions of the homogeneous linear system $A\mathbf{x} = \mathbf{0}$. It is denoted by *null(A)*.

Definition 3.2.5 A *basis* for a subspace S of $\mathbb{R}^n$ is a set of vectors in S that

(1) spans S and

(2) is linearly independent. [4]

Definition 3.2.6 If S is a subspace of $\mathbb{R}^n$, then the number of vectors in a basis for S is called the *dimension* of S, denoted *dim* S. [5]

Definition 3.2.7 The *rank* of a matrix A is the dimension of its row and column spaces and is denoted by *rank(A)*. [6]

Definition 3.2.8 The *nullity* of a matrix A is the dimension of its null space and is denoted by *nullity(A)*.

Theorem 3.2.1 The Rank Theorem. If A is an $m \times n$ matrix, then

$$rank(A) + nullity(A) = n$$

.

Theorem 3.2.2 If A is invertible, then A is a product of elementary matrices.

Theorem 3.2.3 Let A be an $m \times n$ matrix. Then $rank(A^T A) = rank(A)$.

Definition 3.2.9 Let S be a subspace of $\mathbb{R}^n$ and let $B = \{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ be a basis for S. Let $\mathbf{v}$ be a vector in S, and write $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$. Then $c_1, \cdots, c_k$ are called the coordinates of $\mathbf{v}$ with respect to B, and the column vector

$$[\mathbf{v}]_B = [c_1, \cdots, c_k]^T$$

is called the coordinate vector of $\mathbf{v}$ with respect to B. [7]

Definition 3.2.10 A transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is called a linear transformation if

$$T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2)$$

for all $\mathbf{v}_1, \mathbf{v}_2$ in $\mathbb{R}^n$ and scalars c_1, c_2.

3.3 Bases and Dimension

3.4 Coordinates

[4] It does not mean that they are orthogonal.

[5] The zero vector $\mathbf{0}$ is always a subspace of $\mathbb{R}^n$. Yet any set containing the zero vector is linearly dependent, so $\mathbf{0}$ cannot have a basis. We define *dim* $\mathbf{0}$ to be 0.

[6] The row and column spaces of a matrix A have the same dimension.

[7] This coordinate vector is unique.

Chapter 4

Vector and Matrix Calculus

4.1 Functions of Vectors

4.1.1 Inner Product

4.1.2 Outer Product

Definition 4.1.1

$$\mathbf{u} \otimes \mathbf{v} = \mathbf{u}\mathbf{v}^T$$

Remark 4.1.1 The inner product is the trace of the outer product.

4.1.3 Cross Product

Definition 4.1.2

$$\mathbf{a} \times \mathbf{b} = \|\mathbf{a}\|\|\mathbf{b}\|sin(\theta)\mathbf{n}$$

It is also called the vector product.

4.2 Functions of Matrices

4.2.1 Matrix Determinant

4.2.2 Matrix Exponential

4.3 Functions of Vectors and Matrices

4.3.1 Linear Forms: One Vector as Argument

4.3.2 Bilinear and Quadratic Forms: Two Vectors as Argument

4.4 Derivatives of Vectors and Matrices

4.4.1 Derivatives of a Vector or Matrix with Respect to a Scalar

Let $\mathbf{A}$ be a matrix, as a matrix-valued function

$$\mathbf{A}(x) : \mathbb{R} \to \mathbb{R}^{m \times n} \tag{4.1}$$

For vector- and matrix-valued functions there is a further manifestation of the linearity of the derivative: Suppose that f is a fixed linear function defined on $\mathbb{R}^n$ and that $\mathbf{A}$ is a differentiable vector- or matrix-valued function. Then

$$f(\mathbf{A})' = f(\mathbf{A}') \tag{4.2}$$

A useful example is the trace of $\mathbf{A}$, which is the sum of the diagonal elements of $\mathbf{A}$ (differentiable real-valued functions)

$$tr(\mathbf{A})' = tr(\mathbf{A}') \tag{4.3}$$

Another example is the inner product of two vectors, where we have [1]

$$(\mathbf{a}^T\mathbf{b})' = \mathbf{a}'^T\mathbf{b} + \mathbf{a}^T\mathbf{b}' \tag{4.4}$$

An important derivative of a matrix $\mathbf{A}$ is the derivative of its inverse.

Theorem 4.4.1

$$\left(\mathbf{A}^{-1}\right)' = -\mathbf{A}^{-1}\mathbf{A}'\mathbf{A}^{-1}$$

Proof Since

$$\frac{\mathbf{A}^{-1}(x+h) - \mathbf{A}^{-1}(x)}{h} = \frac{\mathbf{A}^{-1}(x+h)[\mathbf{A}(x+h) - \mathbf{A}(x)]\mathbf{A}^{-1}(x)}{h}$$

Another easy proof is:

$$\mathbf{0} = \mathbf{I}' = (\mathbf{A}^{-1}\mathbf{A})' = (\mathbf{A}^{-1})'\mathbf{A} + \mathbf{A}^{-1}\mathbf{A}'$$

Post-multiply $\mathbf{A}^{-1}$ and obtain the desired proof.

4.5 Integration of Vectors and Matrices

[1] Actually, it should work for all dot product (not necessarily the inner product, which is in the context of Euclidean spaces.)

Chapter 5

Some Intuitive Explanations

5.1 Eigenvalues and Singular Values

5.2 SVD, PCA, and Change of Basis

Part II

Techniques

Chapter 6

Schur Complement and LU Decomposition

6.1 Preliminaries

Usually, $|\mathbf{AB}| \neq |\mathbf{BA}|$. For example

Example 6.1.1

$$\left|\begin{bmatrix}1 & 1\end{bmatrix}\begin{bmatrix}0\\1\end{bmatrix}\right| = |1| = 1$$

$$\left|\begin{bmatrix}0\\1\end{bmatrix}\begin{bmatrix}1 & 1\end{bmatrix}\right| = \left|\begin{bmatrix}0 & 0\\1 & 1\end{bmatrix}\right| = 0$$

However, the **Sylvester's Determinant Theorem** says, as long as $\mathbf{AB}$ and $\mathbf{BA}$ are both square matrices,

$$|\mathbf{I} + \mathbf{AB}| = |\mathbf{I} + \mathbf{BA}| \tag{6.1}$$

It is also not true in general that

$$\left|\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix}\right| = |\mathbf{AD} - \mathbf{BC}|$$

unless $\mathbf{C}$ and $\mathbf{D}$ are commutable, i.e., $\mathbf{CD} = \mathbf{DC}$. The general formula for block determinant is

$$\left|\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix}\right| = |\mathbf{A}|\,|\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B}| \tag{6.2}$$

which is based on Schur complement.

6.2 Schur Complement and LU Decomposition

Suppose we have a homogeneous linear system

$$\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix}\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} = \begin{pmatrix}\mathbf{0}\\ \mathbf{0}\end{pmatrix} \tag{6.3}$$

To solve for $\mathbf{y}$, if $\mathbf{A}$ is nonsingular, we may multiply the first row by $-\mathbf{CA}^{-1}$ and add to the second, and obtain

$$\begin{bmatrix}\mathbf{I} & \mathbf{0}\\ -\mathbf{CA}^{-1} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix} = \begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{0} & \mathbf{D} - \mathbf{CA}^{-1}\mathbf{B}\end{bmatrix} \tag{6.4}$$

Definition 6.2.1 Suppose $\mathbf{M}$ is a square matrix

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

and $\mathbf{A}$ nonsingular. We denote [1]

$$\mathbf{M}/\mathbf{A} = \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B} \tag{6.5}$$

and call it *the Schur complement of* $\mathbf{A}$ *in* $\mathbf{M}$, or *the Schur complement of* $\mathbf{M}$ *relative to* $\mathbf{A}$.

Remark 6.2.1 A very useful identity can be revealed from equation 6.4

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{C}\mathbf{A}^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{A}^{-1}\mathbf{B} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \tag{6.6}$$

which gives us the following identities

Theorem 6.2.1

$$det(\mathbf{M}) = det(\mathbf{M}/\mathbf{A}) \cdot det(\mathbf{A}) \tag{6.7}$$

$$rank(\mathbf{M}) = rank(\mathbf{M}/\mathbf{A}) + rank(\mathbf{A}) \tag{6.8}$$

Remark 6.2.2 For a non-homogeneous system of linear equations

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix}$$

We may use Schur complements to write the solution as

$$\mathbf{x} = (\mathbf{M}/\mathbf{D})^{-1}(\mathbf{u} - \mathbf{B}\mathbf{D}^{-1}\mathbf{v}) \tag{6.9}$$

$$\mathbf{y} = (\mathbf{M}/\mathbf{A})^{-1}(\mathbf{v} - \mathbf{C}\mathbf{A}^{-1}\mathbf{u}) \tag{6.10}$$

Theorem 6.2.2 If $\mathbf{M}$ is a positive-definite symmetric matrix, then so is the Schur complement of $\mathbf{D}$ in $\mathbf{M}$.

[1] It is easy to remember if you multiply the submatrices clockwise.

Chapter 7

The Ax = b Problem

7.1 Solving a Linear System of Equations

Theorem 7.1.1 If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then A is invertible if $ad - bc \neq 0$, in which case

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Remark 7.1.1 Three ways to solve a system of linear equations: by elimination, by determinants (Cramer's Rule???), or by matrix decomposition.

Remark 7.1.2 We prefer to use matrix decomposition to solve a linear system because

1. It takes $\mathcal{O}(n^3)$ to factorize, but once done it can be used to solve systems with different **b** (right hand side).

2. It is numerically more stable than computing $\mathbf{A}^{-1}\mathbf{b}$.

3. For a sparse matrix, the inverse may be dense and may hard to store in memory. Decomposition can overcome this problem.

Remark 7.1.3 The computation of elimination is $\mathcal{O}(n^3)$, but can be (non-trivially) reduced to $\mathcal{O}(n^{\log_2 7})$.

7.2 The Vector Spaces of a Matrix

Remark 7.2.1 Ax is a combination of the *columns* of A. $b^T A$ is a combination of the *rows* of A. Row picture can be seen as interchapter of (hyper-)planes. Column picture can be seen as combination of columns.

Remark 7.2.2 There are three different ways to look at matrix multiplication:

1. Each entry of AB is the product of a row (of A) and a column (of B)

2. Each *column* of AB is the product of a matrix (of A) and a column (of B)

3. Each *row* of AB is the product of a row (of A) and a matrix (of B)

Remark 7.2.3 Column space is perpendicular to the left null space. Row space is perpendicular to the null space.

7.3 Matrix Inverse: Binomial inverse theorem, Schur Complement, Blockwise Inversion

Remark 7.3.1 $\mathbf{A}^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{\det \mathbf{A}} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

Chapter 8

The $\mathbf{Ax} = \lambda\mathbf{x}$ Problem

Chapter 9

Special Square Matrices

9.1 Elementary Matrices

There are three types of elementary matrices: Row Switching, Row Multiplication, and Row Addition.

Remark 9.1.1 Left multiplication (pre-multiplication) by an elementary matrix represents elementary row operations, while right multiplication (post-multiplication) represents elementary column operations.

Remark 9.1.2 The inverse of elementary matrices has the same format as the orginal ones.

9.2 Permutation Matrices

Remark 9.2.1 When a permutation matrix P is multiplied with a matrix M from the left it will permute the rows of M, when P is multiplied with M from the right it will permute the columns of M.

Remark 9.2.2 The inverse of a permutation matrix is its transpose.

9.3 Projection Matrices

Remark 9.3.1 $P = A(A^TA)^{-1}A^T$, $P = \frac{aa^T}{\|a\|}$

Remark 9.3.2 $P^2 = P$

Remark 9.3.3 Only two eigenvalues possible: 0 and 1. The corresponding eigenvectors form the kernel and range of A, respectively.

Remark 9.3.4 Projection is invertible.

9.4 Orthogonal Matrices

Definition 9.4.1 An orthogonal matrix is a square matrix with orthonormal columns.

orthogonal and vectors

Remark 9.4.1 $Q^TQ = I$ even if Q is rectangular (but then left-inverse).

Remark 9.4.2 Any permutation matrix P is an orthogonal matrix.

Remark 9.4.3 Orthogonal matrices can be categorized into either the reflection matrix $Ref(\theta)$ which has determinant 1, or the rotation matrix $Rot(\theta)$, which has determinant -1.

Remark 9.4.4 Geometrically, an orthogonal Q is the product of a rotation and a reflection.

Remark 9.4.5 As a linear transformation, an orthogonal matrix preserves the dot product of vectors (therefore also norm and angle), and therefore acts as an isometry of Euclidean space, such as a rotation or reflection. In other words, it is a unitary transformation.

Remark 9.4.6 The product of two rotation matrices is a rotation matrix, and the product of two reflection matrices is also a rotation matrix. See figure 9.1.

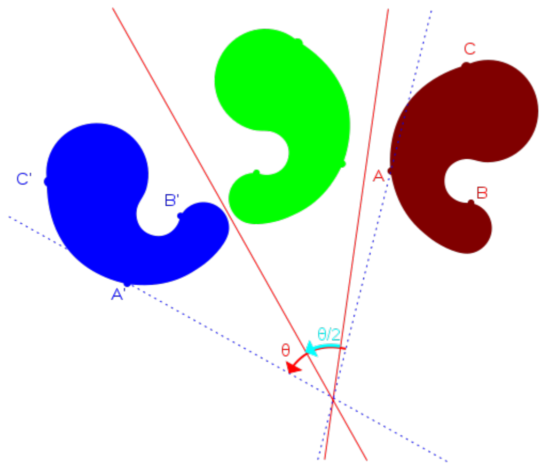

Figure 9.1: The product of two reflection matrices is a rotation matrix.

9.5 Positive Definite Matrices

Chapter 10

Matrix Decomposition

10.1 LU Decomposition

10.2 QR Decomposition

10.3 Cholesky Decomposition

10.4 Symmetric Positive Definite (s.p.d.) Matrices

10.4.1 Cholesky Decomposition

Definition 10.4.1 Let $\mathbf{A}$ be an $n \times n$ square matrix. $\mathbf{A}$ is said to be symmetric positive definite (s.p.d.) if

$$\mathbf{x}^T \mathbf{A} \mathbf{x} > 0, \quad \forall \mathbf{x} \neq \mathbf{0} \tag{10.1}$$

Theorem 10.4.1 If $\mathbf{A}$ is s.p.d., then

1. The diagonal elements of an s.p.d. matrix are positive.
2. All eigenvalues of $\mathbf{A}$ is positive.
3. Its determinant is positive.
4. It is nonsingular.

Proof The diagonal elements are positive because $a_{kk} = \mathbf{e}_k^T \mathbf{A} \mathbf{e}_k > 0$. The eigenvalues of an s.p.d. matrix are all positive is easy to prove by observing that

$$0 < \mathbf{x}^T \mathbf{A} \mathbf{x} = \mathbf{x}^T \lambda \mathbf{x} = \lambda \|\mathbf{x}\|_2^2$$

The positivity of determinant can be shown by looking at the LDU decomposition. Finally, it is nonsingular because the determinant is nonzero.

Definition 10.4.2 Let $\mathbf{A}$ be an $n \times n$ square matrix. A principal submatrix of $\mathbf{A}$ is obtained by selecting some rows and columns with the *same* index subset of $\{1, \cdots, n\}$.

Definition 10.4.3 Let $\mathbf{A}$ be an $n \times n$ square matrix. A *leading* principal submatrix of $\mathbf{A}$ is a principal submatrix of $\mathbf{A}$ with the index subset $\{1, \cdots, m\}$, for some $m \leq n$.

Theorem 10.4.2 If $\mathbf{A}$ is s.p.d. then every principle submatrix is s.p.d..

Proof Suppose $\mathbf{A}_p$ of size p is a principle submatrix of $\mathbf{A}$. Since $\mathbf{A}$ is s.p.d., for any nonzero vector $\mathbf{x}$ we have $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$. Remove the corresponding coordinates of $\mathbf{x}$, same as those removed when creating the principle submatrix, and call it $\mathbf{x}_p$. Then the resulting vector $\mathbf{x}_p^T \mathbf{A}_p \mathbf{x}_p = \mathbf{x}^T \mathbf{A} \mathbf{x} > 0$.

10.4.2 Cholesky Decomposition

Definition 10.4.4 The Cholesky decomposition of an s.p.d. matrix $\mathbf{A}$ is of the form

$$\mathbf{A} = \mathbf{L}\mathbf{L}^* \tag{10.2}$$

where $\mathbf{L}$ is a lower triangular matrix, with *real and positive diagonal elements.*

10.5 Singular Value Decomposition (SVD)

10.6 Eigendecomposition

10.7 Jordan Decomposition

10.8 Schur Decomposition

Part III

Topics

Chapter 11

Minors and Cofactors

11.1 Definition

Definition 11.1.1 General definition of a minor.

Let $\mathbf{A}$ be an $m \times n$ matrix and k an integer with $0 < k \leq \min m, n$. A $k \times k$ minor 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. For such a matrix there are a total of $\binom{m}{k} \cdot \binom{n}{k}$ minors of size $k \times k$.

Definition 11.1.2 First minors and cofactors.

If 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 M_{ij}. The (i, j) cofactor is obtained by multiplying the minor by $(-1)^{i+j}$.

Example 11.1.1 To illustrate these definitions, consider the following 3 by 3 matrix,

$$\begin{bmatrix} 1 & 4 & 7 \\ 3 & 0 & 5 \\ -1 & 9 & 11 \end{bmatrix} \tag{11.1}$$

To compute the minor M_{23} and the cofactor C_{23}, we find the determinant of the above matrix with row 2 and column 3 removed.

$$M_{2,3} = \det \begin{bmatrix} 1 & 4 & \square \\ \square & \square & \square \\ -1 & 9 & \square \end{bmatrix} = \det \begin{bmatrix} 1 & 4 \\ -1 & 9 \end{bmatrix} = (9 - (-4)) = 13$$

So the cofactor of the (2,3) entry is $C_{23} = (-1)^{2+3}(M_{23}) = -13$.

An important application of cofactors is the **Laplace's formula** for the expansion of determinants.

$$\det(\mathbf{A}) = \sum_{i=1}^{n} a_{ij} C_{ij} = \sum_{j=1}^{n} a_{ij} C_{ij} \tag{11.2}$$

If $k \neq i$, we see that

$$\sum_{j=1}^{n} a_{kj} C_{ij} = 0 \tag{11.3}$$

Similarly, if $k \neq j$

$$\sum_{i=1}^{n} a_{ik} C_{ij} = 0 \tag{11.4}$$

This is essentially the determinant of a matrix with the k-th row the same as the i-th row, or the k-th column the same as the j-th column, which is zero.

11.2 The Cramer's Rule and the Adjugate Matrix

$$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & b_2 \\ \vdots & \vdots & \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n & = & b_n \end{array} \tag{11.5}$$

If we multiply the above by the row vector of cofactors of the 1^{st} column, $[C_{11}, C_{21}, \cdots, C_{n1}]$, we obtain

$$[\det(\mathbf{A}), 0, \cdots, 0] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = [C_{11}, C_{21}, \cdots, C_{n1}]\mathbf{b} \tag{11.6}$$

The left hand side used Equation 11.4. The right hand side is nothing but the determinant of a matrix with the first column replaced by $\mathbf{b}$.

Similarly, we can multiply the linear system by the row vector of cofactors of the $2^{nd}, 3^{rd}, \cdots, n^{th}$, and we obtain

$$\det(\mathbf{A})\mathbf{x} = \begin{bmatrix} C_{11} & \cdots & C_{n1} \\ C_{12} & \cdots & C_{n2} \\ \vdots & & \vdots \\ C_{1n} & \cdots & C_{nn} \end{bmatrix} \mathbf{b} \tag{11.7}$$

which gives us

$$\det(\mathbf{A}) = \begin{bmatrix} C_{11} & \cdots & C_{1n} \\ C_{21} & \cdots & C_{2n} \\ \vdots & & \vdots \\ C_{n1} & \cdots & C_{nn} \end{bmatrix}^T \mathbf{A} \tag{11.8}$$

The matrix on the right

$$\mathrm{adj}(\mathbf{A}) = \mathbf{C}^T = \begin{bmatrix} C_{11} & \cdots & C_{1n} \\ C_{21} & \cdots & C_{2n} \\ \vdots & & \vdots \\ C_{n1} & \cdots & C_{nn} \end{bmatrix}^T \tag{11.9}$$

is called the adjugate matrix of $\mathbf{A}$, which is the transpose of the cofactor matrix $\mathbf{C}$.