Linear Algebra: Theory and Techniques

Xi Tan (tan19@purdue.edu)

December 24, 2016

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 Spaces

2.1 Vector Space

Definition 2.1.1 A **vector space** over a field $\mathcal{F}$ is a *nonempty* set V together with the operations of addition $V \times V \to V$ and scalar multiplication $\mathcal{F} \times V \to V$ 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 $a, b \in \mathcal{F}$, 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 $a, b \in \mathcal{F}$, we have

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

2.2 Subspaces

Definition 2.2.1 A subspace of $\mathcal{R}^n$ is any collection S of vectors in $\mathcal{R}^n$ 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 2.2.2 Let S, T be two subspaces of $\mathcal{R}^n$. 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 2.2.3 Let A be an $m \times n$ matrix.

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

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

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

Definition 2.2.4 Let A be an $m \times n$ matrix. The *null space* (or *kernel*) of A is the subspace of $\mathcal{R}^n$ consisting of solutions of the homogeneous linear system $A\mathbf{x} = \mathbf{0}$. It is denoted by *null(A)*.

Definition 2.2.5 A *basis* for a subspace S of $\mathcal{R}^n$ is a set of vectors in S that

(1) spans S and

(2) is linearly independent. [4]

Definition 2.2.6 If S is a subspace of $\mathcal{R}^n$, then the number of vectors in a basis for S is called the *dimension* of S, denoted *dim* S. [5]

Definition 2.2.7 The *rank* of a matrix A is the dimension of its row and column spaces and is denoted by *rank(A)*. [6]

Definition 2.2.8 The *nullity* of a matrix A is the dimension of its null space and is denoted by *nullity(A)*.

Theorem 2.2.1 The Rank Theorem. If A is an $m \times n$ matrix, then

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

.

Theorem 2.2.2 If A is invertible, then A is a product of elementary matrices.

Theorem 2.2.3 Let A be an $m \times n$ matrix. Then $rank(A^T A) = rank(A)$.

Definition 2.2.9 Let S be a subspace of $\mathcal{R}^n$ 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 2.2.10 A transformation $T : \mathcal{R}^n \to \mathcal{R}^m$ 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 $\mathcal{R}^n$ and scalars c_1, c_2.

2.3 Bases and Dimension

2.4 Coordinates

2.5 Linear Forms: One Vector as Argument

2.6 Bilinear and Quadratic Forms: Two Vectors as Argument

2.7 Jordan Canoical Forms

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

[5] The zero vector $\mathbf{0}$ is always a subspace of $\mathcal{R}^n$. 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 3

Eigenvalues and Eigenvectors

3.1 Definitions

Remark 3.1.1 eigenvectors are non-zero.

Definition 3.1.1 The set of all eigenvectors corresponding to the same eigenvalue, together with the zero vector, is called an *eigenspace*.

Definition 3.1.2 The characteristic polynomial of a matrix $\mathbf{A}$ of order n is

$$|\mathbf{A} - \lambda \mathbf{I}| = \prod_{i=1}^{n} (\lambda - \lambda_i) \tag{3.1}$$

Theorem 3.1.1 Every square matrix of order n has n eigenvalues, possibly complex and not necessarily all unique.

Definition 3.1.3 The algebraic multiplicity $\mu_A(\lambda_i)$ of a eigenvalue λ_i is the multiplicity as a root of the characteristic polynomial.

Definition 3.1.4 The eigenspace E_{λ_i} associated with λ_i is defined as

$$E_{\lambda_i} = \{\mathbf{v} : (\mathbf{A} - \lambda_i \mathbf{I})\mathbf{v} = 0\} \tag{3.2}$$

Definition 3.1.5 The dimension of the eigenspace $\mathbf{E}_{\lambda_i}$ is referred to as the geometric multiplicity $\gamma_A(\lambda_i)$ of λ_i.

Chapter 4

Vector Calculus

4.1 Inner Product (Dot 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.

4.2 Outer Product

4.3 Cross Product

Definition 4.3.1

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

It is also called the vector product.

4.4 Scalar Triple Product

4.5 Vector Triple Product

4.6 Line, Surface, and Volume Integrals

4.7 Integration of Vectors and Matrices

Chapter 5

Matrix Calculus

5.1 Matrix Determinant

5.2 Kronecker Product and Vec

5.3 Hadamard Product and Diag

5.4 Matrix Exponential

Chapter 6

Vector and Matrix Derivatives

Suppose $\mathbf{Y}_{m\times n}$ and $\mathbf{X}_{p\times q}$ are both matrices (scalars, vectors are of course special cases). The derivative of $\mathbf{Y}$ with respect to $\mathbf{X}$ involves $mnpq$ partial derivatives, $\left[\frac{\partial Y_{ij}}{\partial X_{kl}}\right]$, for $i = 1, \cdots, m; j = 1, \cdots, n; k = 1, \cdots, p; l = 1, \cdots, q$. This immediately poses a question: What is a convenient (or logic) way of arraying these partial derivatives - as a row vector, as a column vector, or as a matrix (which is a natural choice), and if the latter of what shape/order?

Two competing notational conventions can be distinguished by whether the index of the derivative (matrix) is majored by the numerator or the denominator.

1. Numerator layout, i.e. according to $\mathbf{Y}$ and $\mathbf{X}^T$. This is sometimes known as the Jacobian layout.
2. Denominator layout, i.e. according to $\mathbf{Y}^T$ and $\mathbf{X}$. This is sometimes known as the gradient layout. It is named so because the gradient under this layout is a usual column vector.

The transpose of one layout is the same as the other. We use the **numerator-layout** notation throughout the paper.

	Scalar	Vector	Matrix
Scalar	$\frac{\partial y}{\partial x}$	$\frac{\partial \mathbf{y}}{\partial x} = [\frac{\partial y_i}{\partial x}]$	$\frac{d\mathbf{Y}}{\partial x} = [\frac{\partial y_{ij}}{\partial x}]$
Vector	$\frac{\partial y}{\partial \mathbf{x}} = [\frac{\partial y}{\partial x_j}]$	$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = [\frac{\partial y_i}{\partial x_j}]$	
Matrix	$\frac{\partial y}{d\mathbf{X}} = [\frac{\partial y}{\partial x_{ji}}]$		

The partials with respect to the numerator are laid out according to the shape Y while the partials with respect to the denominator are laid out according to the transpose of X. For example, $\partial y/\partial \mathbf{x}$ is a row vector[1] while $\partial \mathbf{y}/\partial x$ is a column vector.

Note:

1. derivative is a row vector; gradient is its transpose.
2. Hessian is the derivative of gradient.

6.1 Differentials

Example 6.1.1

$$d\mathbf{A} = \mathbf{A} - \mathbf{A} = \mathbf{0} \tag{6.1}$$

[1] We distinguish $\partial y/\partial \mathbf{x}$ and the gradient $\nabla_{\mathbf{x}} y$, which is the transpose of the former and hence a column vector.

Example 6.1.2

$$d(\alpha\mathbf{X}) = \alpha(X + d\mathbf{X}) - \alpha\mathbf{X} = \alpha d\mathbf{X} \tag{6.2}$$

Example 6.1.3

$$d(\mathbf{X}+\mathbf{Y}) = [(\mathbf{X}+\mathbf{Y}) + d(\mathbf{X}+\mathbf{Y})] - (\mathbf{X}+\mathbf{Y}) = d\mathbf{X} + d\mathbf{Y} \tag{6.3}$$

Example 6.1.4

$$d(\mathrm{tr}(\mathbf{X})) = \mathrm{tr}(\mathbf{X} + d\mathbf{X}) - \mathrm{tr}(\mathbf{X}) = \mathrm{tr}(\mathbf{X} + d\mathbf{X} - \mathbf{X}) = \mathrm{tr}(d\mathbf{X}) \tag{6.4}$$

Example 6.1.5

$$d(\mathbf{X}\mathbf{Y}) = (\mathbf{X} + d\mathbf{X})(\mathbf{Y} + d\mathbf{Y}) - \mathbf{X}\mathbf{Y} = [\mathbf{X}\mathbf{Y} + \mathbf{X}d\mathbf{Y} + (d\mathbf{X})\mathbf{Y} + d\mathbf{X}d\mathbf{Y}] - \mathbf{X}\mathbf{Y} = \mathbf{X}d\mathbf{Y} + (d\mathbf{X})\mathbf{Y} \tag{6.5}$$

Example 6.1.6

$$\mathbf{0} = \mathbf{dI} = d(\mathbf{X}\mathbf{X}^{-1}) = (d\mathbf{X})\mathbf{X}^{-1} + \mathbf{X}d\mathbf{X}^{-1} \tag{6.6}$$

$$d\mathbf{X}^{-1} = -\mathbf{X}^{-1}(d\mathbf{X})\mathbf{X}^{-1} \tag{6.7}$$

Another proof is:

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

Next, let's prove something not so trivial.

Proposition 6.1.1

$$d|\mathbf{X}| = |\mathbf{X}|\mathrm{tr}(\mathbf{X}^{-1}d\mathbf{X}) \tag{6.8}$$

Proof First, we see that

$$\mathrm{tr}(\mathbf{A}^T\mathbf{B}) = \sum_{i=1}^{n}\left(\sum_{j=1}^{n}(\mathbf{A}^T)_{ij}\mathbf{B}_{ji}\right) = \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{A}_{ji}\mathbf{B}_{ji} = \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{A}_{ij}\mathbf{B}_{ij} = \mathrm{vec}(\mathbf{A})^T\mathrm{vec}(\mathbf{B}) \tag{6.9}$$

which can be computed by first multiply **A** and **B** element-wise, and then sum all the elements in the resulting matrix (known as the *Frobenius inner product*)[2].

Next, applying the Laplace's formula

$$|\mathbf{X}| = \sum_{j} x_{ij} \cdot adj^T(\mathbf{X})_{ij} \tag{6.10}$$

we have,

$$d(|\mathbf{X}|) = \sum_{i}\sum_{j}\frac{\partial|\mathbf{X}|}{\partial x_{ij}}dx_{ij} \tag{6.11}$$

[2] The trace operator is a scalar function (of a matrix), that essentially turns matrices into vectors and computes a dot product between them.

$$= \sum_i \sum_j \frac{\partial \{\sum_k x_{ik} \cdot adj^T(\mathbf{X})_{ik}\}}{\partial x_{ij}} dx_{ij} \qquad \text{(expand by row } i\text{)} \tag{6.12}$$

$$= \sum_i \sum_j \left\{ \sum_k \frac{\partial x_{ik}}{\partial x_{ij}} \cdot adj^T(\mathbf{X})_{ik} + \sum_k x_{ik} \frac{\partial adj^T(\mathbf{X})_{ik}}{\partial x_{ij}} \right\} dx_{ij} \tag{6.13}$$

$$= \sum_i \sum_j adj^T(\mathbf{X})_{ij} dx_{ij} \qquad \left(\frac{\partial adj^T(\mathbf{X})_{ik}}{\partial x_{ij}} = 0, \forall k \neq j \right) \tag{6.14}$$

$$\tag{6.15}$$

Now, use $\sum_{i=1}^n \sum_{j=1}^n \mathbf{A}_{ij}\mathbf{B}_{ij} = \mathrm{tr}(\mathbf{A}^T\mathbf{B})$, we have

$$d(|\mathbf{X}|) = tr(adj(X) d\mathbf{X}) \tag{6.16}$$

Since $\mathbf{X}$ is invertible, and $adj(\mathbf{X}) = |\mathbf{X}|\mathbf{X}^{-1}$, finally,

$$d(|\mathbf{X}|) = |\mathbf{X}| tr(\mathbf{X}^{-1} d\mathbf{X}) \tag{6.17}$$

6.2 Vector-by-vector Derivatives

The first two important identities are

$$\frac{\partial A\mathbf{x}}{\partial \mathbf{x}} = A \tag{6.18}$$

$$\frac{\partial \mathbf{x}^T A}{\partial \mathbf{x}} = A^T \tag{6.19}$$

In the numerator-layout, the major index of the resulting matrix is based on the numerator, so when A is on the left hand side of $\mathbf{x}$, the derivative is the same size as A, on the other hand, if A is on the right hand side of $\mathbf{x}$, it needs to be transposed.

Example 6.2.1 Suppose $a = a(\mathbf{x})$ is a scalar function and $\mathbf{u} = \mathbf{u}(\mathbf{x})$ a vector function.

$$\frac{\partial a\mathbf{u}}{\partial \mathbf{x}} = \frac{\partial a\mathbf{u}}{\partial a}\frac{\partial a}{\partial \mathbf{x}} + \frac{\partial a\mathbf{u}}{\partial \mathbf{u}}\frac{\partial \mathbf{u}}{\partial \mathbf{x}} = \mathbf{u}\frac{\partial a}{\partial \mathbf{x}} + a\frac{\partial \mathbf{u}}{\partial \mathbf{x}} \tag{6.20}$$

Recall that $\frac{\partial a\mathbf{u}}{\partial a}$ is a row vector, and the chain rule is expanded from right to left, just as the composition of functions.

6.3 Derivatives of Vectors and Matrices

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

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

$$\mathbf{A}(x) : \mathcal{R} \to \mathcal{R}^{m \times n} \tag{6.21}$$

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 $\mathcal{R}^n$ and that $\mathbf{A}$ is a differentiable vector- or matrix-valued function. Then

$$f(\mathbf{A})' = f(\mathbf{A}') \tag{6.22}$$

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

$$tr(\mathbf{A})' = tr(\mathbf{A}') \tag{6.23}$$

Another example is the inner product of two vectors, where we have [3]

$$(\mathbf{a}^T\mathbf{b})' = \mathbf{a}'^T\mathbf{b} + \mathbf{a}^T\mathbf{b}' \tag{6.24}$$

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

Chapter 7

Vector and Matrix Integrals

Chapter 8

Some Intuitive Explanations

8.1 Eigenvalues and Singular Values

8.2 SVD, PCA, and Change of Basis

Part II

Techniques

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 Symmetric Matrices

9.4 Projection Matrices

Remark 9.4.1 $P = A(A^TA)^{-1}A^T$, $P = \frac{aa^T}{\|a\|}$

Remark 9.4.2 $P^2 = P$

Remark 9.4.3 Only two eigenvalues possible: 0 and 1. The corresponding eigenvectors form the kernel and range of A, respectively.

Remark 9.4.4 Projection is invertible.

9.5 Normal Matrix

Definition 9.5.1 A *normal matrix* is a square matrix which satisfies

$$\mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T \tag{9.1}$$

9.6 Orthogonal Matrices

Definition 9.6.1 An *orthogonal matrix* (*unitary* for a complex matrix) is a normal matrix which further satisfies

$$\mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T = \mathbf{I} \tag{9.2}$$

Or, alternatively,

gonal and unit
rs

Remark 9.6.1 An orthogonal matrix is a square matrix with orthonormal columns.

Remark 9.6.2 $Q^TQ = I$ even if Q is rectangular (but then left-inverse).

Remark 9.6.3 Any permutation matrix P is an orthogonal matrix.

Remark 9.6.4 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 9.6.5 Geometrically, an orthogonal Q is the product of a rotation and a reflection.

Remark 9.6.6 Orthogonal matrix is invariant to 2-norm, that is, suppose Q is an orthogonal matrix, and x a vector, then

$$\|Qx\| = \|x\| \tag{9.3}$$

Remark 9.6.7 Projection matrices are usually not orthogonal, since they are not invariant to 2-norm.

Remark 9.6.8 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 9.6.9 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.

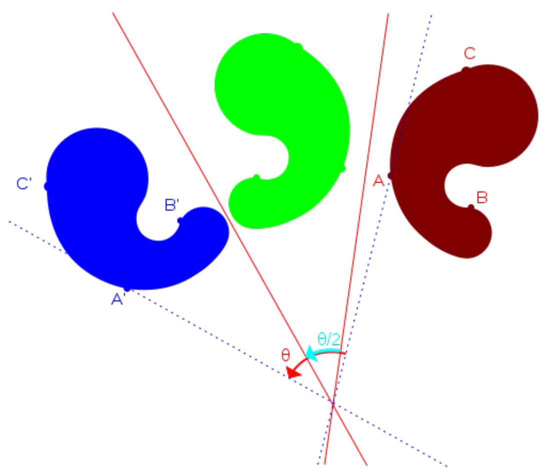

Figure 9.1: The product of two reflection matrices is a rotation matrix.

9.7 Positive Definite Matrices

Definition 9.7.1 Let $\mathbf{A}$ be an $n \times n$ square matrix. $\mathbf{A}$ is said to be positive definite if

$$\mathbf{x}^T\mathbf{A}\mathbf{x} > 0, \quad \forall \mathbf{x} \neq \mathbf{0} \tag{9.4}$$

Theorem 9.7.1 If $\mathbf{A}$ is positive definite, then

1. The diagonal elements of a positive definite 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 9.7.2 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 9.7.3 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 9.7.2 If $\mathbf{A}$ is positive definite 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 positive definite, 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 $\mathbf{x}_p^T\mathbf{A}_p\mathbf{x}_p = \mathbf{x}^T\mathbf{A}\mathbf{x} > 0$.

Chapter 10

Numerical Linear Algebra Algorithms

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

Remark 10.1.1 $\mathbf{A}^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{\det \mathbf{A}} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

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

Example 10.1.1

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

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

However, the **Sylvester's Determinant Theorem** says, as long as **AB** and **BA** are both square matrices,

$$\left|\mathbf{I} + \mathbf{AB}\right| = \left|\mathbf{I} + \mathbf{BA}\right| \tag{10.1}$$

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

$$\left| \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \right| = |\mathbf{A}| \left|\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B}\right| \tag{10.2}$$

which is based on Schur complement.

Chapter 11

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

11.1 Solving a Linear System of Equations

Theorem 11.1.1 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 11.1.1 Three ways to solve a system of linear equations: by elimination, by determinants (Cramer's Rule???), or by matrix decomposition.

Remark 11.1.2 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 $\mathbf{b}$ (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 11.1.3 Cofactors and Minors. Laplaces Theorem.

Remark 11.1.4 The computation of elimination is $\mathcal{O}(n^3)$, but can be (non-trivially) reduced to $\mathcal{O}(n^{\log_2 7})$.

11.2 The Vector Spaces of a Matrix

Remark 11.2.1 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 11.2.2 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 11.2.3 Column space is perpendicular to the left null space. Row space is perpendicular to the null space.

Chapter 12

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

Chapter 13

Matrix Decomposition

13.1 Decomposition related to solving $\mathbf{Ax} = \mathbf{b}$

13.1.1 LU Decomposition: Schur Complement

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

Example 13.1.1

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

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

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

$$\left| \mathbf{I} + \mathbf{AB} \right| = \left| \mathbf{I} + \mathbf{BA} \right| \tag{13.1}$$

It is also not true in general that

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

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

$$\left| \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \right| = |\mathbf{A}| \left| \mathbf{D} - \mathbf{CA}^{-1}\mathbf{B} \right| \tag{13.2}$$

which is based on Schur complement.

Now suppose we have a homogeneous linear system

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix} \tag{13.3}$$

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

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{CA}^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{D} - \mathbf{CA}^{-1}\mathbf{B} \end{bmatrix} \tag{13.4}$$

Definition 13.1.1 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]

$$\mathbf{M}/\mathbf{A} = \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B} \tag{13.5}$$

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

Remark 13.1.1 A very useful identity can be revealed from equation 13.4

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{C}\mathbf{A}^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{A}^{-1}\mathbf{B} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \tag{13.6}$$

which gives us the following identities

Theorem 13.1.1

$$det(\mathbf{M}) = det(\mathbf{M}/\mathbf{A}) \cdot det(\mathbf{A}) \tag{13.7}$$

$$rank(\mathbf{M}) = rank(\mathbf{M}/\mathbf{A}) + rank(\mathbf{A}) \tag{13.8}$$

Remark 13.1.2 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

$$\mathbf{x} = (\mathbf{M}/\mathbf{D})^{-1}(\mathbf{u} - \mathbf{B}\mathbf{D}^{-1}\mathbf{v}) \tag{13.9}$$

$$\mathbf{y} = (\mathbf{M}/\mathbf{A})^{-1}(\mathbf{v} - \mathbf{C}\mathbf{A}^{-1}\mathbf{u}) \tag{13.10}$$

Theorem 13.1.2 If $\mathbf{M}$ is a positive-definite symmetric matrix, then so is the Schur complement of $\mathbf{D}$ in $\mathbf{M}$.

13.1.2 LDU Decomposition

The LDU decomposition can be viewed as the matrix form of Gaussian elimination. It is used to find the inverse of a matrix, **or computing the determinant of a matrix**.

Remark 13.1.3 The triangular factorization can be written $A = LDU$, where L and U have $1's$ on the diagonal and D is the diagonal matrix of pivots.

13.1.3 Rank Decomposition

13.1.4 Cholesky Decomposition

Definition 13.1.2 The Cholesky decomposition of an s.p.d. matrix $\mathbf{A}$ is of the form

$$\mathbf{A} = \mathbf{L}\mathbf{L}^* \tag{13.11}$$

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

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

Definition 13.1.3 The Cholesky decomposition of a s.p.d. matrix $\mathbf{A}$ is of the form

$$\mathbf{A} = \mathbf{L}\mathbf{L}^T \tag{13.12}$$

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

Cholesky decomposition is unique. If $\mathbf{A}$ is symmetric semi-positive definite, it still has a decomposition of the form $\mathbf{A} = \mathbf{L}\mathbf{L}^*$, although may not be unique, if the diagonal entries of $\mathbf{L}$ are allowed to be zero. A closely related variant of the classical Cholesky decomposition is the LDL^T *decomposition*:

$$\mathbf{A} = \mathbf{LDL}^T = (\mathbf{LD}^{\frac{1}{2}})(\mathbf{D}^{\frac{1}{2}^T}\mathbf{L}^T) = (\mathbf{LD}^{\frac{1}{2}})(\mathbf{LD}^{\frac{1}{2}})^T \tag{13.13}$$

where the diagonal entries of $\mathbf{L}$ are all ones.

13.1.5 QR Decomposition: Givens Rotation, Householder Transformation

Any square matrix $\mathbf{A}$ may be decomposed as

$$\mathbf{A} = \mathbf{QR} \tag{13.14}$$

where $\mathbf{Q}$ is an orthogonal matrix, and $\mathbf{R}$ an upper triangular matrix. This is called the QR decomposition. It is essentially a change of basis process, and can be obtained by using the Gram-Schmidt process.

13.2 Decomposition related to solving $\mathbf{Ax} = \lambda\mathbf{x}$

13.2.1 Eigendecomposition

Suppose a square matrix $\mathbf{M}$ of order n is diagonalizable, i.e., it has n linearly independent eigenvectors, then since

$$\mathbf{MQ} = \mathbf{Q\Lambda} \tag{13.15}$$

where the columns of $\mathbf{Q}$ are eigenvectors of $\mathbf{M}$ (hence invertible), and $\mathbf{\Lambda}$ is a diagonal matrix with eigenvalues of $\mathbf{M}$ as entries. Then we have

$$\mathbf{M} = \mathbf{Q\Lambda Q}^{-1} \tag{13.16}$$

13.2.2 Jordan Decomposition

13.2.3 Schur Decomposition

13.2.4 Singular Value Decomposition (SVD)

13.2.5 QZ Decomposition

13.3 Other Decompositions

13.3.1 Polar Decomposition

Part III

Topics

Chapter 14

Minors and Cofactors

14.1 Definition

Definition 14.1.1 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 14.1.2 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 14.1.1 To illustrate these definitions, consider the following 3 by 3 matrix,

$$\begin{bmatrix} 1 & 4 & 7 \\ 3 & 0 & 5 \\ -1 & 9 & 11 \end{bmatrix} \tag{14.1}$$

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.

$$\det(\mathbf{A}) = \sum_{i=1}^{n} a_{ij} C_{ij} = \sum_{j=1}^{n} a_{ij} C_{ij} \tag{14.2}$$

If $k \neq i$, we see that

$$\sum_{j=1}^{n} a_{kj} C_{ij} = 0 \tag{14.3}$$

Similarly, if $k \neq j$

$$\sum_{i=1}^{n} a_{ik} C_{ij} = 0 \tag{14.4}$$

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.

14.2 The Cramer's Rule and the Adjugate Matrix

$$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & b_2 \\ \vdots & \vdots & \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n & = & b_n \end{array} \tag{14.5}$$

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

$$[\det(\mathbf{A}), 0, \cdots, 0] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = [C_{11}, C_{21}, \cdots, C_{n1}]\mathbf{b} \tag{14.6}$$

The left hand side used Equation 14.4. 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

$$\det(\mathbf{A})\mathbf{x} = \begin{bmatrix} C_{11} & \cdots & C_{n1} \\ C_{12} & \cdots & C_{n2} \\ \vdots & & \vdots \\ C_{1n} & \cdots & C_{nn} \end{bmatrix} \mathbf{b} \tag{14.7}$$

which gives us

$$\det(\mathbf{A}) = \begin{bmatrix} C_{11} & \cdots & C_{1n} \\ C_{21} & \cdots & C_{2n} \\ \vdots & & \vdots \\ C_{n1} & \cdots & C_{nn} \end{bmatrix}^T \mathbf{A} \tag{14.8}$$

The matrix on the right

$$\operatorname{adj}(\mathbf{A}) = \mathbf{C}^T = \begin{bmatrix} C_{11} & \cdots & C_{1n} \\ C_{21} & \cdots & C_{2n} \\ \vdots & & \vdots \\ C_{n1} & \cdots & C_{nn} \end{bmatrix}^T \tag{14.9}$$

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