Commentary on Lang's Linear Algebra

Henry C. Pinkham

Henry C. Pinkham

Draft of May 31, 2013.

Contents

Preface

These notes were written to complement and supplement Lang's Linear Algebra [4] as a textbook in a Honors Linear Algebra class at Columbia University. The students in the class were gifted but had limited exposure to linear algebra. As Lang says in his introduction, his book is not meant as a substitute for an elementary text. The book is intended for students having had an elementary course in linear algebra. However, by spending a week on Gaussian elimination after covering the second chapter of [4], it was possible to make the book work in this class. I had spent a fair amount of time looking for an appropriate textbook, and I could find nothing that seemed more appropriate for budding mathematics majors than Lang's book. He restricts his ground field to the real and complex numbers, which is a reasonable compromise.

The book has many strength. No assumptions are made, everything is defined. The first chapter presents the rather tricky theory of dimension of vector spaces admirably. The next two chapters are remarkably clear and efficient in presently matrices and linear maps, so one has the two key ingredients of linear algebra: the allowable objects and the maps between them quite concisely, but with many examples of linear maps in the exercises. The presentation of determinants is good, and eigenvectors and eigenvalues is well handled. Hermitian forms and hermitian and unitary matrices are well covered, on a par with the corresponding concepts over the real numbers. Decomposition of a linear operator on a vector space is done using the fact that a polynomial ring over a field is a Euclidean domain, and therefore a principal ideal domain. These concepts are defined from scratch, and the proofs presented very concisely, again. The last chapter covers the elementary theory of convex sets: a beautiful topic if one has the time to cover it. Advanced students will enjoy reading Appendix II on the Iwasawa Decomposition. A motto for the book might be:

> A little thinking about the abstract principles underlying linear algebra
> can avoid a lot of computation.

Many of the sections in the book have excellent exercises. The most remarkable ones are in §II.3, III.3, III.4, VI.3, VII.1, VII.2, VII.3, VIII.3, VIII.4, XI.3, XI.6. A few of the sections have no exercises at all, and the remainder the standard set. There is a solutions manuel by Shakarni [5]. It has answers to all the exercises in the book, and good explanations for some of the more difficult ones. I have added some clarification to a few of the exercise solutions.

A detailed study of the book revealed flaws that go beyond the missing material on Gaussian elimination. The biggest problems are in Chapters IV and V. Certain of the key passages in Chapter IV are barely comprehensible, especially in §3. Admittedly this is the hardest material in the book, and is often omitted in elementary texts. In Chapter V there are problems of a different nature: because Lang does not use Gaussian elimination he is forced to resort to more complicated arguments to prove that row rank is equal to column rank, for example. While the coverage of duality is good, it is incomplete. The same is true for projections. The coverage of positive definite matrices over the reals misses the tests that are easiest to apply, again because they require Gaussian elimination. In his quest for efficiency in the definition of determinants, he produces a *key lemma* that has to do too much work.

While he covers some of the abstract algebra he needs, for my purposes he could have done more: cover elementary group theory and spend more time on permutations. One book that does more in this direction is Valenza's textbook [7]. Its point of view similar to Lang's and it covers roughly the same material. Another is Artin's Algebra text [1], which starts with a discussion of linear algebra. Both books work over an arbitrary ground field, which raises the bar a little higher for the students. Fortunately all the ground work for doing more algebra is laid in Lang's text: I was able to add it in my class without difficulty.

Lang covers few applications of linear algebra, with the exception of differential equations which come up in exercises in Chapter III, §3, and in the body of the text in Chapters VIII, §1 and §4, and XI, §4.

Finally, a minor problem of the book is that Lang does not refer to standard concepts and theorems by what is now their established name. The most egregious example is the Rank-Nullity Theorem, which for Lang is just Theorem 3.2 of Chapter III. He calls the Cayley-Hamilton Theorem the Hamilton-Cayley Theorem. He does not give names to the axioms he uses: associativity, commutativity, existence of inverses, etc.

Throughout the semester, I wrote notes to complement and clarify the exposition in [4] where needed. Since most of the book is clear and concise, I only covered a small number of sections.

- I start with an introduction to Gaussian elimination that I draw on later in these notes and simple extensions of some results.
- I continue this with a chapter on matrix inverses that fills one of the gaps in Lang's presentation.

- The next chapter develops some elementary group theory, especially in connection to permutation groups. I also introduce permutation matrices, which gives a rich source of examples of matrices.
- The next six chapters (4-9) address problems in Lang's presentation of key material in his Chapters IV, V and VI.
- In Chapter 10 I discuss the companion matrix, which is not mentioned at all in Lang, and can be skipped. It is interesting to compare the notion of cyclic vector it leads to with the different one developed in Lang's discussion of the Jordan normal form in [XI, §6].
- The last chapter greatly expands Lang's presentation of tests for positive (semi)definite matrices, which are only mentioned in the exercises, and which omit the ones that are most useful and easiest to apply.

To the extent possible I use the notation and the style of the author. For example I minimize the use of summations, and give many proofs by induction. One exception: I use A^t to denote the transpose of the matrix A. Those notes form the chapters of this book. There was no time in my class to cover Lang's Chapter XII on Convex Sets or Appendix II on the Iwasawa Decomposition, so they are not mentioned here. Otherwise we covered the entire book in class.

I hope that instructors and students reading Lang's book find these notes useful.

Comments, corrections, and other suggestions for improving these notes are welcome. Please email them to me at hcp3@columbia.edu.

HENRY C. PINKHAM
New York, NY
Draft of May 31, 2013

Chapter 1

Gaussian Elimination

The main topic that is missing in [4], and that needs to be covered in a first linear algebra class is Gaussian elimination. This can be easily fit into a course using [4]: after covering [4], Chapter II, spend two lectures covering Gaussian elimination, including elimination via left multiplication by elementary matrices. This chapter covers succinctly the material needed. Another source for this material is Lang's more elementary text [3].

1.1 Row Operations

Assume we have a system of m equations in n variables.

$$Ax = b,$$

where A is therefore a $m \times n$ matrix, x a n-vector and b a m-vector. Solving this system of equations in x is one of the fundamental problems of linear algebra.

Gaussian elimination is an effective technique for determining when this system has solutions in x, and what the solutions are.

The key remark is that the set of solutions do not change if we use the following three operations.

1. We multiply any equation in the system by a non-zero number;
2. We interchange the order of the equations;
3. We add to one equation a multiple of another equation.

1.1.1 Definition. Applied to the augmented matrix

$$A' = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{pmatrix} \tag{1.1.2}$$

these operations have the following effect:

1. Multiplying a row of A' by a non-zero number;
2. Interchanging the rows of A';
3. Adding to a row of A' a multiple of another row.

They are called *row operations*.

Using only these three matrix operations, we will put the coefficient matrix A into *row echelon form*. Recall that A_i denotes the i-th row of A.

1.1.3 Definition. The matrix A is in *row echelon form* if

1. All the rows of A that consist entirely of zeroes are below any row of A that has a non-zero entry;
2. If row A_i has its first non-zero entry in position j, then row A_{i+1} has its first non-zero entry in position $> j$.

1.1.4 Remark. This definition implies that the first non-zero entry of A_i in a position $j \geq i$. Thus if A is in row echelon form, $a_{ij} = 0$ for all $j < i$. If A is square, this means that A is upper triangular.

1.1.5 Example. The matrices

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 2 & 3 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}, \text{ and } \begin{pmatrix} -1 & 2 & 3 \\ 0 & -2 & 2 \\ 0 & 0 & 0 \end{pmatrix}$$

are in row echelon form, while

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 0 \\ 0 & 2 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \text{ and } \begin{pmatrix} 0 & 0 & 3 \\ 0 & 0 & 2 \\ 0 & 0 & 1 \end{pmatrix}$$

are not.

The central theorem is

1.1.6 Theorem. *Any matrix can be put in row echelon form by using only row operations.*

Proof. We prove this by induction on the number of rows of the matrix A. If A has one row, there is nothing to prove.

So assume the result for matrices with $m-1$ rows, and let A be a matrix with m rows. If the matrix is the 0 matrix, we are done. Otherwise consider the first column A^k of A that has a non-zero element. Then for some i, $a_{ik} \neq 0$, and for all $j < k$, the column A^j is the zero vector.

Since A^k has non-zero entry a_{ik}, interchange rows 1 and i of A. So the new matrix A has $a_{1k} \neq 0$. Then if another row A_i of A has k-th entry $a_{ik} \neq 0$, replace that row by

$$A_i - \frac{a_{ik}}{a_{1k}} A_1$$

which has j-th entry equal to 0. Repeat this on all rows that have a non-zero element in column k. The new matrix, which we still call A, has only zeroes in its first k columns except for a_{1k}.

Now consider the submatrix B of the new matrix A obtained by removing the first row of A and the first k columns of A. Then we may apply the induction hypothesis to B and put B in row echelon form by row operations. The key observation is that these same row operations may be applied to A, and do not affect the zeroes in the first k columns of A, since the only non-zero there is in row 1 and that row has been removed from B. □

We could of course make this proof constructive by repeating to B the row operations we did on A, and continuing in this way.

1.1.7 Exercise. Reduce the matrices in Example 1.1.5 that are not already in row echelon form to row echelon form.

At this point we defer the study of the inhomogeneous equation $Ax = b$. We will pick it up again in §7.2 when we have more tools.

1.2 Elementary Matrices

Following Lang [3], Chapter II §5, we introduce the three types of *elementary matrices*, and show they are invertible.

In order to do that we let I_{rs} be the square matrix with a $i_{rs} = 1$, and zeroes everywhere else. So for example in the 3×3 case

$$I_{23} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

1.2.1 Definition. Elementary matrices E are square matrices, say $m \times m$. We describe what E does to a matrix $m \times n$ A when we left-multiply E by A, and we write the formula for E. There are three types.

1. E multiplies a row of a matrix by a number $c \neq 0$; The elementary matrix
$$E_r(c) := I + (c-1)I_{rr}$$
multiplies the r-th row of A by c.
2. E interchange two rows of any matrix A; We denote by T_{rs}, $r \neq s$ the permutation matrix that interchanges row i of A with row j. This is a special case of a permutation matrix: see §3.3 for the 3×3 case. The permutation is a transposition.
3. E adds a constant times a row to another row; The matrix
$$E_{rs}(c) := I + cI_{rs}, r \neq s,$$
adds c times the s-th row of A to the r-th row of A.

1.2.2 Example. Here are some 3×3 examples, acting on the $3 \times n$ matrix
$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ a_{21} & \dots & a_{2n} \\ a_{31} & \dots & a_{3n} \end{pmatrix}.$$
First, since
$$E_1(c) = \begin{pmatrix} c & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix},$$
matrix multiplication gives
$$E_1(c)A = \begin{pmatrix} ca_{11} & \dots & ca_{1n} \\ a_{21} & \dots & a_{2n} \\ a_{31} & \dots & a_{3n} \end{pmatrix}.$$
Next since
$$T_{23} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$
we get
$$T_{23}A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ a_{31} & \dots & a_{3n} \\ a_{21} & \dots & a_{2n} \end{pmatrix}.$$

Finally, since

$$E_{13}(c) = \begin{pmatrix} 1 & 0 & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

we get

$$E_{13}(c)A = \begin{pmatrix} a_{11} + ca_{31} & \dots & a_{1n} + ca_{33} \\ a_{21} & \dots & a_{2n} \\ a_{31} & \dots & a_{3n} \end{pmatrix}.$$

1.2.3 Theorem. *All elementary matrices are invertible.*

Proof. For each type of elementary matrix E we write down an inverse, namely a matrix F such that $EF = I = FE$.

For $E_i(c)$ the inverse is $E_i(1/c)$. T_{ij} is its own inverse. Finally the inverse of $E_{ij}(c)$ is $E_{ij}(-c)$. □

1.3 Elimination via Matrices

The following result follows immediately from Theorem 1.1.6 once one understands left-multiplication by elementary matrices.

1.3.1 Theorem. *Given any $m \times n$ matrix A, one can left multiply it by elementary matrices of size m to bring into row echelon form.*

We are interested in the following

1.3.2 Corollary. *Assume A is a square matrix of size m. Then we can left-multiply A by elementary matrices E_i, $1 \le i \le k$ until we reach either the identity matrix, or a matrix whose bottom row is 0. In other words we can write*

$$A' = E_k \dots E_1 A. \tag{1.3.3}$$

where either $A' = I$ or the bottom row of A' is the zero vector.

Proof. Using Theorem 1.3.1, we can bring A into row echelon form. Call the new matrix AB. As we noted in Remark 1.1.4 this means that B is upper triangular. If any diagonal entry b_{ii} of B is 0, then the bottom row of B is zero, and we have reached one of the conclusions of the theorem.

So we may assume that all the diagonal entries of B are non-zero. We now do what is called *backsubstitution* to transform B into the identity matrix.

First we left multiply by the elementary matrices

$$E_m(1/b_{mm}), E_{m-1}(1/b_{m-1,m-1}), \dots, E_1(1/b_{11})$$

. This new matrix, that we still call B, now has 1s along the diagonal.

Next we left multiply by elementary matrices of type (3) in the order

$$\begin{aligned}
&E_{m,m-1}(c), E_{m,m-2}(c), \ldots, E_{m,1}(c),\\
&E_{m-1,m-2}(c), E_{m-1,m-3}(c), \ldots, E_{m-1,1}(c),\\
&\ldots, \ldots\\
&E_{2,1}(c).
\end{aligned}$$

where in each case the constant c is chosen so that a new zero in created in the matrix. This order is chosen so that no zeroes created by a multiplication is destroyed by a subsequent one. At the end of the process we get the identity matrix I so we are done.

□

1.3.4 Exercise. Reduce the matrices in Example 1.1.5 either to a matrix with bottom row zero or to the identiy matrix using left multiplcation by elementary matrices.

For example, the first matrix

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

backsubstitutes to

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \text{ then } \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \text{ then } \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

We will use these results on elementary matrices in §2.2 and then throughout Chapter 8.

Chapter 2

Matrix Inverses

This chapter fill in some gaps in the exposition in Lang, especially the break between his more elementary text [3] and the textbook [4] we use. There are some additional results: Proposition 2.2.5 and Theorem 2.2.8. They will be important later.

2.1 Background

On page 35 of our text [4], Lang gives the definition of an invertible matrix.

2.1.1 Definition. A $n \times n$ matrix A is *invertible* if there exists another $n \times n$ matrix B such that

$$AB = BA = I. \tag{2.1.2}$$

We say that B is both a left inverse and a right inverse for A. It is reasonable to require both since matrix multiplication is not commutative. Then Lang proves:

2.1.3 Theorem. *If A has an inverse, then its inverse B is unique.*

Proof. Indeed assume there is another matrix C satisfying (2.1.2) when C replaces B. Then

$$C = CI = C(AB) = (CA)B = IB = B \tag{2.1.4}$$

so we are done. □

Note that the proof only uses that C is a left inverse and B a right inverse.

We say, simply, that B is the *inverse* of A and A the *inverse* of B.

In [4], on p. 86, Chapter IV, §2, in the proof of Theorem 2.2, Lang only establishes that a certain matrix A has a left inverse, when in fact he needs to show it has an inverse. We recall this theorem and its proof in 4.1.4. We must prove that having a left inverse is enough to get an inverse. Indeed, we will establish this in Theorem 2.2.8.

2.2 More Elimination via Matrices

In 1.2.1 we introduced the three types of *elementary matrices* and showed they are invertible in Theorem 1.2.3.

2.2.1 Proposition. *Any product of elementary matrices is invertible.*

Proof. This follows easily from a more general result. Let A and B are two invertible $n \times n$ matrices, with inverses A^{-1} and B^{-1}. Then $B^{-1}A^{-1}$ in the inverse of AB

Indeed just compute

$$B^{-1}A^{-1}AB = B^{-1}IB = B^{-1}B = I$$

and

$$ABB^{-1}A^{-1} = A^{-1}IA = A^{-1}A = I.$$

□

Before stating the next result, we make a definition:

2.2.2 Definition. Two $m \times n$ matrices A and B are *row equivalent* if there is a product of elementary matrices E such that $B = EA$.

By Proposition 2.2.1 E is invertible, so if $B = EA$, then $A = E^{-1}B$.

2.2.3 Proposition. *If A is a square matrix, , and if B is row equivalent to A, then A has an inverse if and only if B has an inverse.*

Proof. Write $B = EA$. If A is invertible, then $A^{-1}E^{-1}$ is the inverse of B. We get the other implication by symmetry. □

2.2.4 Remark. This establishes an equivalence relation on $m \times n$ matrices, as discussed in §4.3.

We continue to assume that A is a square matrix. Corollary 1.3.2 tells us that A is either row equivalent to the identity matrix I, in which case it is obviously invertible by Proposition 2.2.3, or row equivalent to a matrix with bottom row the zero vector.

2.2.5 Proposition. *Let A be a square matrix with one row equal to the zero vector. Then A is not invertible.*

Proof. By [4] , Theorem 2.1 p. 30, the homogeneous system

$$AX = O \tag{2.2.6}$$

has a non-trivial solution X, since one equation is missing, so that the number of variables is greater than the number of equations. But if A were invertible, we could multiply (2.2.6) on the left by A^{-1} to get $A^{-1}AX = A^{-1}O$. This yields $X = O$, implying there cannot be a non-trivial solution, a contradiction to the assumption that A is invertible. □

Proposition 5.1.4 generalizes this result to matrices that are not square.

So we get the following useful theorem. The proof is immediate.

2.2.7 Theorem. *A square matrix A is invertible if and only if it is row equivalent to the identity I. If it is not invertible, it is row equivalent to a matrix whose last row is the zero vector. Furthermore any upper triangular matrix square A with non-zero diagonal elements is invertible.*

Finally, we get to the key result of this section.

2.2.8 Theorem. *Let A be a square matrix which has a left inverse B, so that $BA = I$. Then A is invertible and B is its inverse.*

Similarly, if A has a right inverse B, so that $AB = I$, the same conclusion holds.

Proof. Suppose first that $AB = I$, so B is a right inverse. Perform row reduction on A. By Theorem 2.2.7, there are elementary matrices $E_1, E_2, \dots, E_k$ so that the row reduced matrix $A' = E_k \dots E_1 A$ is the identity matrix or has bottom row zero. Then multiply by B on the right and use associativity:

$$A'B = (E_k \dots E_1 A)B = (E_k \dots E_1)(AB) = E_k \dots E_1.$$

This is invertible, because elementary matrices are invertible, so all rows of $A'B$ are non-zero by Proposition 2.2.5. Now if A' had a zero row, matrix multiplication tells us that $A'B$ would have a zero row, which is impossible. So $A' = I$. Finally Proposition 2.2.3 tells us that A is invertible, since it is row equivalent to I. In particular its left inverse $E_k \dots E_1$ and its right inverse B are equal.

To do the direction $BA = I$, just interchange the role of A and B.

□

2.2.9 Remark. The first five chapters of Artin's book [1] form a nice introduction to linear algebra at a slightly higher level than Lang's, with some group theory thrown in too. The main difference is that Artin allows his base field to be any field, including a finite field, while Lang only allows $\mathbb{R}$ and $\mathbb{C}$. The references to Artin's book are from the first edition.

2.3 Invertible Matrices

The group (for matrix multiplication) of invertible square matrices has a name: the general linear group, written $Gl(n)$. Its dimension is n^2. Most square matrices are invertible, in a sense one can make precise, for example by a dimension count. A single number, the determinant of A, tells us if A is invertible or not, as we will find out in [4], Chapter VI.

It is worth memorizing what happens in the 2×2 case. The matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is invertible if and only if $ad - bc \neq 0$, in which case the inverse of A is

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

as you should check by direct computation. See Exercise 4 of [I, §2].

Chapter 3

Groups and Permutation Matrices

In class we studied the motions of the equilateral triangle whose center of gravity is at the origin of the plane, so that the three vertices are equidistant from the origin. We set this up so that one side of the triangle is parallel to the x- axis. There are $6 = 3!$ motions of the triangle permuting the vertices, and they can be realized by linear transformations, namely multiplication by 2×2 matrices.

These notes describe these six matrices and show that they can be multiplied together without ever leaving them, because they form a *group*: they are closed under matrix multiplication. The multiplication is not commutative, however. This is the simplest example of a noncommutative group, and is worth remembering.

Finally, this leads us naturally to the $n \times n$ permutation matrices. We will see in §3.3 that they form a group. When $n = 3$ it is *isomorphic* to the group of motions of the equilateral triangle. From the point of view of linear algebra, permutation and permutation matrices are the most important concepts of these chapter. We will need them when we study determinants.

There are some exercises in this chapter. Lots of good practice in matrix multiplication.

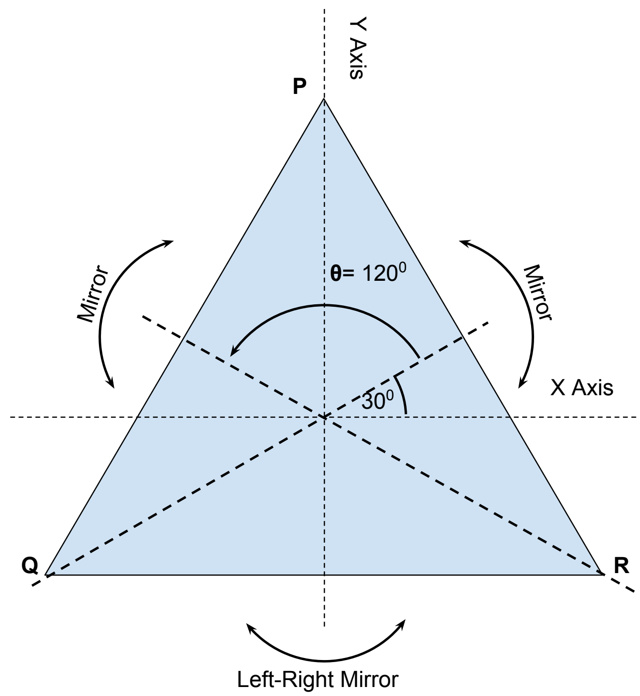
3.1 The Motions of the Equilateral Triangle

The goal of this section is to write down six 2×2 matrices such that multiplication by those matrices induce the motions of the equilateral triangle.

First the formal definition of a permutation:

3.1.1 Definition. A permutation σ of a finite set S is a *bijective* map from S to itself. See [4], p. 48.

We are interested in the permutations of the vertices of the triangle. Because we will be realizing these permutations by linear transformations and because the center of the triangle is fixed by all linear transformations, it turns out that each linear transformation fixes the entire triangle. Here is the graph.



To start, we have the identity motion given by the identity matrix

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Next the rotations. Since the angle of rotation is 120 degrees, or $\frac{2\pi}{3}$ radians, the first matrix of rotation is

$$R_1 = \begin{pmatrix} \cos\frac{2\pi}{3} & -\sin\frac{2\pi}{3} \\ \sin\frac{2\pi}{3} & \cos\frac{2\pi}{3} \end{pmatrix} = \begin{pmatrix} -1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{pmatrix}.$$

Lang [4] discusses this on page 85. The second one, R_2, corresponds rotation by 240 degrees. You should check directly what R_2 is, just like we did for R_1. You will get:

$$R_2 = \begin{pmatrix} -1/2 & \sqrt{3}/2 \\ -\sqrt{3}/2 & -1/2 \end{pmatrix}.$$

More importantly note that R_2 is obtained by repeating R_1, in matrix multiplication:

$$R_2 = R_1^2.$$

Also recall Lang's Exercise 23 of Chapter II, §3.

Next let's find the left-right mirror reflection of the triangle. What is its matrix? In the y coordinate nothing should change, and in the x coordinate, the sign should change. The matrix that does this is

$$M_1 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

This is an elementary matrix, by the way.

One property that all mirror reflection matrices share is that their square is the identity. True here.

Finally, the last two mirror reflections. Here they are. Once you understand why conceptually the product is correct, to do the product just change the sign of the terms in the first *column* of R_1 and R_2:

$$M_2 = R_1 M_1 = \begin{pmatrix} 1/2 & -\sqrt{3}/2 \\ -\sqrt{3}/2 & -1/2 \end{pmatrix} \text{ and } M_3 = R_2 M_1 = \begin{pmatrix} 1/2 & \sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{pmatrix}.$$

In the same way, to compute the product $M_1 R_1$ just change the sign of the terms in the first row of R_1.

3.1.2 Exercise. Check that the squares of M_2 and M_3 are the identity. Describe each geometrically: what is the invariant line of each one. In other words, if you input a vector $X \in \mathbb{R}^2$, what is $M_2 X$, etc.

3.2 Groups

3.2.1 Exercise. We now have six 2×2 matrices, all invertible. Why? The next exercise will give an answer.

3.2.2 Exercise. Construct a multiplication table for them. List the 6 elements in the order $I, R_1, R_2, M_1, M_2, M_3$, both horizontally and vertically, then have the elements in the table you construct represent the product, in the following order: if A is the row element, and B is the column element, then put AB in the table. See

name	I	R_1	R_2	M_1	M_2	M_3
I	I	R_1	R_2	M_1	M_2	M_3
R_1	R_1	R_2	I	M_2	M_3	M_1
R_2	R_2	I	R_1	M_3	M_1	M_2
M_1	M_1	M_3	M_2	I	R_2	R_1
M_2	M_2	M_1	M_3	R_1	I	R_2
M_3	M_3	M_2	M_1	R_2	R_1	I

Table 3.1: Multiplication Table.

Artin [1], chapter II, §1, page 40 for an example, and the last page of these notes for an unfinished example. The solution is given in Table 3.1.

Show that this multiplication is not commutative, and notice that each row and column contains all six elements of the group.

Table 3.1 is the solution to the exercise. The first row and the first column of the table just give the elements to be multiplied, and the entry in the table give the product. Notice that the 6×6 table divides into 4 3×3 subtables, each of which contains either M elements or R and I elements, but not both.

What is a group? The formal definition is given in Lang in Appendix II.

3.2.3 Definition. A group is a set G with a map $G \times G \to G$, written as multiplication, which to a pair (u, v) of elements in G associates uv in G, where this map, called the group law, satisfies three axioms:

GP 1. The group law is associative: Given elements u, v and w in G,

$$(uv)w = u(vw).$$

GR 2. Existence of a neutral element $1 \in G$ for the group law: for every element $u \in G$,

$$1u = u1 = u.$$

GR 3. Existence of an inverse for every element: For every $u \in G$ there is a $v \in G$ such that

$$uv = vu = 1.$$

The element v is written u^{-1}.

If you look at the first three axioms of a vector space, given in Lang p.3, you will see that they are identical. Only two changes: the group law is written like multiplication, while the law $+$ in vector spaces is addition. The neutral element for vector spaces is O, while here we write it as 1. These are just changes in notation, not in substance. Note that we lack axiom **VP 4**, which in our notation would be written:

GR 4. The group law is commutative: For all u and v in G

$$uv = vu.$$

So the group law is not required to be commutative, unlike addition in vector spaces. A group satisfying **GR 4** is called a commutative group. So vector spaces are commutative groups for their law of addition.

One important property of groups is that cancelation is possible:

3.2.4 Proposition. *Given elements u, v and w in G,*

1. *If $uw = vw$, then $u = v$.*
2. *If $wu = wv$, then $u = v$ again.*

Proof. For (1) multiply the equation on the right by the inverse of w to get

$$\begin{aligned} (uw)w^{-1} &= (vw)w^{-1} && \text{by existence of inverse,} \\ u(ww^{-1}) &= v(ww^{-1}) && \text{by associativity,} \\ u1 &= v1 && \text{by properties of inverses,} \\ u &= v && \text{by property of neutral element.} \end{aligned}$$

For (2), multiply by w^{-1} on the left. □

Because vector spaces are groups for addition, this gives a common framework for solving some of the problems in Lang: for example problems 4 and 5 of [I, §1].

3.2.5 Example. Here are some groups we will encounter.

1. Any vector space for its law of addition.
2. The integers $\mathbb{Z}$, with addition, so the neutral element is 0 and the inverse of n is $-n$.
3. Permutations on n elements, where the law is composition of permutations.
4. The group of all invertible $n \times n$ matrices, with matrix multiplication as its law. The neutral element is'

$$I = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

5. The group of all invertible $n \times n$ matrices with determinant 1.

Much of what we will do in this course is to prove results that will help us understand these groups. For example Theorem 8.3.3 clarifies the relationship between the last two example. See Remark 8.3.5.

3.3 Permutation Matrices

Permutations are discussed in Lang [6, §4] while studying determinants. The terminology of groups is not introduced until Appendix II, but in effect it is shown that permutations on the set of integers $(1, 2, \dots, n)$, which Lang calls J_n, is a group for composition of maps. It is called the *symmetric group*, and often written S_n. In §3.1 we studied S_3.

3.3.1 Exercise. Show that the number of elements in S_n is

$$n! = n(n-1)(n-2)\cdots(2)(1).$$

When $n = 3$ we get 6, as we noticed above.

Lang does not introduce permutation matrices, even though they form a natural linear algebra way of studying permutations, and are a rich store of examples. While we treat the general case of $n \times n$ permutation matrices here, you should first set $n = 3$ to get the case we studied in §3.1.

The objects to be permuted are the n unit column vectors in K^n:

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \dots, e_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Given a permutation σ on $\{1, 2, \dots, n\}$, we ask for a $n \times n$ matrix P^σ such that matrix multiplication $P^\sigma e_j$ yields $e_{\sigma(j)}$ for $j = 1, 2, \dots, n$. Because $P^\sigma I = P^\sigma$, where I is the identity matrix, the j-th column of P^σ must be the unit vector $e_{\sigma(j)}$. In particular P^σ must have exactly one 1 in each row and column with all other entries being 0.

Here is the formal definition:

3.3.2 Definition. Given a permutation σ on $\{1, 2, \dots, n\}$, we define the $n \times n$ *permutation matrix* $P^\sigma = (p^\sigma_{ij})$ of σ as follows:

$$p^\sigma_{ij} = \begin{cases} 1, & \text{if } i = \sigma(j); \\ 0, & \text{otherwise.} \end{cases}$$

Another way of saying this is that the j-th column of P^σ is $e_{\sigma(j)}$.

3.3.3 Example. Consider the permutation $\sigma(1) = 2$, $\sigma(2) = 3$, $\sigma(3) = 1$. Then its permutation matrix is

$$P = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

Here is another permutation matrix.

$$Q = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Notice that this is an elementary matrix, interchanging rows 1 and 2. Its square is I. We could call it M_3: M because it acts like a mirror in the previous sections, 3 because that is the variable that is not moved.

3.3.4 Exercise. What is the permutation σ associated to Q?

Solution: σ here is the permutation $\sigma(1) = 2$, $\sigma(2) = 1$, $\sigma(3) = 3$.

3.3.5 Exercise. Compute the matrix product PQ. Also compute the matrix product QP.

Solution:

$$PQ = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

while

$$QP = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

The answers are different, but each one is a permutation matrix.

3.3.6 Exercise. Here is another example.

$$R = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

What is the permutation associated to R? What is R^2?

3.3.7 Exercise. Write down all 3×3 permutation matrices. Show that they are invertible by finding their inverses. Make sure that your technique generalizes to $n \times n$ permutation matrices. (Hint: transpose)

3.3.8 Exercise. Construct a multiplication table for the 3×3 permutation matrices, just as in Exercise 3.2.2. Notice that it is the same table, given the proper identification of the elements.

Here is the solution to these exercises. One key concept is the notion of the *order* of an element, namely the smallest power of the element that makes it the identity. The order of I is obviously 1.

Table 3.1 shows that the order of all three elements M_i is 2. Note that this means that they are their own inverses. A little extra thought shows that the order of R_1 and R_2 is 3: Indeed: R_1 is the rotation by 120 degrees, and R_2 rotation by 240, so 3 times each one of them is a multiple of 360 degres.

For our permutation matrices, to have order 2 means to be symmetric, as you should check. The matrices Q and R are symmetric, but the matrix P is not. P^2 is the last non-symmetric permutation matrix:

$$P^2 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

It is now easy to put together a multiplication table that, in the language of group theory, shows that this group is *isomorphic* to the group of motions of the equilateral triangle of §3.2.

Chapter 4

Representation of a Linear Transformation by a Matrix

This chapter concerns [4], chapter IV. At the end of the chapter Lang gives the definition of similarity. Here we expand on this in §4.3. Then there are some solved exercises.

4.1 The First Two Sections of Chapter IV

In Chapter IV, §1, Lang considers an arbitrary $m \times n$ matrix A, and associates to it a linear map $L_A \colon K^n \to K^m$.

So he is building a map T from the set $M(m,n)$ of all $m \times n$ matrices to the set $\mathcal{L}(K^n, K^m)$ of linear maps from $K^n \to K^m$. From Exercise 2 p. 28, we learned that $M(m,n)$ is a vector space of dimension mn. $\mathcal{L}(K^n, K^m)$ is a vector space by Example 6, page 54.

So we have a map T between two vector spaces: we should check that it is a linear map. The two displayed equations on the top of p.86 establish that.

Lang's Theorem 1.1 shows that T is injective. Theorem 2.1 constructs an inverse to T. The construction of the inverse is important. It is the prototype of the difficult constructions in [IV, §3]. The following point should be emphasized.

4.1.1 Remark. As in Lang, let E^j, $1 \leq j \leq n$ be the unit coordinate vectors of K^n, and the e^i, $1 \leq i \leq m$ be those of K^m. By linearity the linear map L from K^n to K^m maps:

$$L(E^j) = a_{1j}e^1 + \cdots + a_{mj}e^m$$

for some $a_{ij} \in K$, which we have written here as in Lang. The choice of the order of the running indices i and j in a_{ij} is arbitrary: we could have written a_{ji} instead.

If we write the equations at the bottom of page 83 as a matrix multiplication, we get

$$\begin{pmatrix} L(E^1) \\ \vdots \\ L(E^n) \end{pmatrix} = \begin{pmatrix} a_{11} & \dots & a_{m1} \\ \vdots & \ddots & \vdots \\ a_{1n} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} L(e^1) \\ \vdots \\ L(e^m) \end{pmatrix}$$

The matrix (a_{ji}) is apparently written backwards. The point of choosing this order is that when one writes down the matrix showing how the coordinates transform (middle of page 84), the entries of the matrix are now written in the conventional order. This is the same computation as that given in the proof of Proposition 4.2.5 and Theorem 4.2.7 below.

4.1.2 Exercise. Show that Lang's Theorem 2.1 is a corollary of Theorem 4.2.7 below.

By Theorem 2.1, $\dim \mathcal{L}(K^n, K^m) = \dim M(m, n)$, namely mn.

Although we do not need it right away, you should study the following equation from p.86 carefully. If F is the linear map from K^n to K^m with matrix A, and G the linear map from K^m to K^s with matrix B, then we can find the matrix associated to the composite linear map $G \circ F$ without computation. Let X be an n-vector. Then

$$(G \circ F)(X) = G(F(X)) = B(AX) = (BA)X. \tag{4.1.3}$$

It is the matrix product BA. All we used is the associativity of matrix multiplication. We use this in the proof of Theorem 4.2.15.

It is worth amplifying Lang's proof of his Theorem 2.2 p.86.

4.1.4 Theorem. *Let A be an $n \times n$ matrix with columns A^1, ..., A^n. Then A is invertible if and only if A^1, ..., A^n are linearly independent.*

Proof. First assume A^1, ..., A^n are linearly independent. So $\{A^1, \dots, A^n\}$ is a basis of K^n. We apply Theorem 2.1 of Chapter III (p. 56), where the $\{A^1, \dots, A^n\}$ play the role of $\{v_1, \dots, v_n\}$ and the standard basis $\{E^1, \dots, E^n\}$ the role of $\{w_1, \dots, w_n\}$. Then that result tells us there is a unique linear mapping T sending A^j to E^j for all j:

$$T(A^j) = E^j \text{ for } j = 1, \dots, n.$$

From Theorem 2.1 of Chapter IV, §2, we know that T is multiplication by a matrix: it is a L_B. Thus we have

$$BA^j = E^j \text{ for } j = 1, \dots, n.$$

But now, just looking at the way matrix multiplication works (a column at a time for the right hand matrix), since E^j is the j-th column of the identity matrix, we get

$$BA = I$$

so B is a left inverse for A. Since Theorem 2.2.8 implies that A is invertible , we are done for this direction.

The converse is easy: we have already seen this argument before in the proof of Proposition 2.2.5. Assume A is invertible. Consider the linear map L_A:

$$L_A(X) = AX = x_1 A^1 + \cdots + x_n A^n,$$

where the A^i are the columns of A.

If $AX = O$, then multiplying by the inverse A^{-1} we get $A^{-1}AX = X = O$, so that the only solution of the equation $AX = O$ is $X = O$. So the only solution of

$$x_1 A^1 + \cdots + x_n A^n = O$$

is $X = O$, which says precisely that $A^1, \ldots, A^n$ are linearly independent.

□

4.2 Representation of $\mathcal{L}(V, W)$

We now get to the most problematic section of [4]: [IV, §3].

Suppose we have two vector spaces V and W, and a linear map F between them. Let $\mathcal{B} = \{v_1, \ldots, v_n\}$ be a basis for V and $\mathcal{E} = \{w_1, \ldots, w_m\}$ a basis for W, so that V has dimension n and W dimension m.

The goal is to find the matrix that Lang writes $M_{\mathcal{E}}^{\mathcal{B}}(F)$ and calls the matrix representing F in these bases. It is defined in Lang's Theorem 3.1.

Lang does not mention this up front (but see the set of equations at the bottom of page 83, and $(*)$ on page 89), but the linearity of F implies that there is a collection of numbers (c_{ij}) such that

$$\begin{aligned} F(v_1) &= c_{11}w_1 + \cdots + c_{1m}w_m \\ \vdots &= \vdots \\ F(v_n) &= c_{n1}w_1 + \cdots + c_{nm}w_m \end{aligned} \tag{4.2.1}$$

This gives us a $n \times m$ matrix C: this is **not** the matrix we want, but these numbers are typically what is known about the linear transformation. Although

Lang is unwilling to write this expression down at the beginning of the section, note that he does write it down in Example 3 p. 90.

Instead we use our basis $\mathcal{B}$ to construct an isomorphism from V to K^n we have studied before: see example 1 p.52. Lang now calls it $X_{\mathcal{B}}\colon V \to K^n$. To any element

$$v = x_1 v_1 + \cdots + x_n v_n \in V$$

it associates its vector of coordinates $(x_1, \ldots, x_n)$ in K^n. We do the same thing on the W side: we have an isomorphism $X_{\mathcal{E}}\colon W \to K^m$. To an arbitrary element

$$w = y_1 w_1 + \cdots + y_m w_m \in W$$

it associates its vector of coordinates $(y_1, \ldots, y_m)$ in K^m.

4.2.2 Remark. The inverse of $X_{\mathcal{E}}$ is the linear map

$$X_{\mathcal{E}}^{-1}\colon K^m \to W$$

which to $(y_1, \ldots, y_m) \in K^m$ associates $y_1 w_1 + \cdots + y_m w_m \in W$.

These isomorphisms (or their inverses) induce a unique linear map $K^n \to K^m$, in a sense that is made clear below, and it is the $m \times n$ matrix $M_{\mathcal{E}}^{\mathcal{B}}(F)$ representing this map, in the sense of §1 and §2, that we are interested in.

4.2.3 Notation. In the notation $M_{\mathcal{E}}^{\mathcal{B}}(F)$, the basis, here $\mathcal{B}$, of the domain V of the linear transformation F is written as the superscript, while that of W (often called the codomain), here $\mathcal{E}$, is the subscript.

We have what is called a commutative diagram:

$$\begin{array}{ccc}
V & \xrightarrow{F} & W \\
{\scriptstyle X_{\mathcal{B}}}\big\downarrow & & {\scriptstyle X_{\mathcal{E}}}\big\downarrow \\
K^n & \xrightarrow{M_{\mathcal{E}}^{\mathcal{B}}(F)} & K^m
\end{array}$$

4.2.4 Theorem. *The above diagram is commutative, meaning that for any $v \in V$,*

$$M_{\mathcal{E}}^{\mathcal{B}}(F) \circ X_{\mathcal{B}}(v) = X_{\mathcal{E}} \circ F(v)$$

This is exactly the content of the equation in Lang's Theorem 3.1 p. 88.

We want to compute $M_{\mathcal{E}}^{\mathcal{B}}(F)$, given information about F. Since $X_{\mathcal{B}}$ is an isomorphism, it has an inverse. Thus $M_{\mathcal{E}}^{\mathcal{B}}(F)$ can be written as the composition

$$M_{\mathcal{E}}^{\mathcal{B}}(F) = X_{\mathcal{E}} \circ F \circ X_{\mathcal{B}}^{-1}$$

so that if we know F and the bases, we can compute $M_{\mathcal{E}}^{\mathcal{B}}(F)$.

Following Lang, for simplicity of notation we write A for $M_{\mathcal{E}}^{\mathcal{B}}(F)$, and X for $X_{\mathcal{B}}(v)$, the coordinate vector of v with respect to the basis $\mathcal{B}$. Then we get the all important computation, where the scalar product is just the ordinary dot product in K^n.

4.2.5 Proposition.

$$F(v) = \langle X, A_1 \rangle w_1 + \dots \langle X, A_m \rangle w_m, \tag{4.2.6}$$

where A_i is the i-th row of A.

Proof. In the commutative diagram above, we first want to get from V to K^m via K^n. We get there from V by composing the maps $X_{\mathcal{B}}$ followed by $M_{\mathcal{E}}^{\mathcal{B}}(F)$. In our simplified notation, this is the matrix product AX, which is an m vector. Then to get back to W, we need to apply the inverse of $X_{\mathcal{E}}$. So we need to take each component of the m-vector AX and put it in front of the appropriate basis element in W, as we did in Remark 4.2.2. The components of AX are the products of the rows of A with X. That gives (4.2.6). □

Next we relate $M_{\mathcal{E}}^{\mathcal{B}}(F)$ to the matrix C of (4.2.1). Using (4.2.1) for the $F(v_i)$, we write out $F(v) = F(x_1 v_1 + \dots + x_n v_n)$ to get:

$$F(v) = x_1\big(c_{11}w_1 + \dots + c_{1m}w_m\big) + \dots + x_n\big(c_{n1}w_1 + \dots + c_{nm}w_m\big)$$

and collecting the terms in w_i we get:

$$F(v) = \big(x_1 c_{11} + \dots + x_n c_{n1}\big)w_1 + \dots + \big(x_1 c_{m1} + \dots + x_n c_{mn}\big)w_m$$

which we can write as

$$F(v) = \langle X, C^1 \rangle w_1 + \dots + \langle X, C^m \rangle w_m$$

the dot products with the columns of C. Compare this to (4.2.6): it says that A is the transpose of C.

4.2.7 Theorem. *The matrix $M_{\mathcal{E}}^{\mathcal{B}}(F)$ is the transpose of the matrix C of* (4.2.1). *In particular it is an $m \times n$ matrix.*

This key result is Lang's *assertion* of page 89.

4.2.8 Remark. A major source of confusion in Lang is the set of equations $(*)$ on p.89. Since a matrix A is already defined, the reader has every right to assume that the a_{ij} in these equations are the coefficients of the matrix A, especially since he already uses A_i to denote the rows of A. Nothing of the sort. The a_{ij} are being defined by these equations, just as here we defined the c_{ij} in the analogous (4.2.1). Because he has written the coefficients in $(*)$ backwards: a_{ji} for a_{ij}, he lands on his feet: they are the coefficients of A.

4.2.9 Example. Lang's Example 1 at the bottom of page 89 is illuminating. What we are calling C is the matrix

$$\begin{pmatrix} 3 & -1 & 17 \\ 1 & 1 & -1 \end{pmatrix}$$

while the matrix $A = M^{\mathcal{B}}_{\mathcal{B}'}$ (in Lang's notation) is its transpose, the matrix associated with F.

4.2.10 Example. We now clarify Lang's example 3 p.90, in changed notation to avoid confusion. Let V be an n dimensional vector space with a first basis $\mathcal{B} = \{v_1, \dots, v_n\}$, and a second basis $\mathcal{E} = \{w_1, \dots, w_n\}$. Thus there exists a change of basis matrix C such that:

$$\begin{aligned} w_1 &= c_{11}v_1 + \dots + c_{1n}v_n \\ \vdots &= \vdots \\ w_n &= c_{n1}v_1 + \dots + c_{nn}v_n \end{aligned} \tag{4.2.11}$$

If we interpret the left hand side as the identity linear transformation id applied to the basis elements w_i here, this is the same set of equations at (4.2.1), but with the role of the bases reversed: $\{w_1, \dots, w_n\}$ is now a basis in the domain of id, and $\{v_1, \dots, v_n\}$ a basis in the codomain

We compute $M^{\mathcal{E}}_{\mathcal{B}}(id)$. Note the change of position of the indices! We apply Proposition 4.2.5, untangling the meaning of all the terms, and using as before the abbreviation A for $M^{\mathcal{E}}_{\mathcal{B}}(Id)$. Then, as in (4.2.6):

$$w = \langle X, A_1 \rangle v_1 + \dots + \langle X, A_n \rangle v_n, \tag{4.2.12}$$

and by Theorem 4.2.7, A is the transpose of C. So we are done.

Next consider the unique linear map $F \colon V \to V$ such that

$$F(v_1) = w_1, \dots, F(v_n) = w_n.$$

The uniqueness follows from the important Theorem 2.1 of Chapter III of Lang.

We want to compute $M^{\mathcal{B}}_{\mathcal{B}}(F)$. The same basis on both sides. We write

$$\begin{aligned}F(v) &= F(x_1v_1+\dots+x_nv_n) = x_1F(v_1)+\dots x_nF(v_n)\\ &= x_1w_1+\dots+x_nw_n\\ &= x_1\big(c_{11}v_1+\dots+c_{1n}v_n\big)+\dots+x_n\big(c_{n1}v_1+\dots+c_{nn}v_n\big)\\ &= \big(x_1c_{11}+\dots x_nc_{n1}\big)v_1+\dots+\big(x_1c_{1n}+\dots x_nc_{nn}\big)v_1\\ &= \langle X,C^1\rangle v_1+\dots+\langle X,C^n\rangle v_n\\ &= AX,\end{aligned}$$

where we used (4.2.11) to express the w_i in terms of the v_i and we write A for the transpose of C. Thus $M^{\mathcal{B}}_{\mathcal{B}}(F)$ is the transpose of C.

The point of this exercise is that these two very different linear operators have the same matrix representing them: this is due to the different choice of bases.

Lang now states the following easy result.

4.2.13 Theorem. *Let V be a vector space of dimension n with basis $\mathcal{B}$ and W a vector space of dimension m with basis $\mathcal{E}$. Let f and g be linear maps from V to W. Then*

$$M^{\mathcal{B}}_{\mathcal{E}}(f+g) = M^{\mathcal{B}}_{\mathcal{E}}(f) + M^{\mathcal{B}}_{\mathcal{E}}(g)$$

and

$$M^{\mathcal{B}}_{\mathcal{E}}(cf) = cM^{\mathcal{B}}_{\mathcal{E}}(f).$$

Therefore $M^{\mathcal{B}}_{\mathcal{E}}$ is a linear map from $\mathcal{L}(V,W)$ to $M(m,n)$, the space of $n\times m$ matrices. Furthermore it is an isomorphism.

4.2.14 Remark. Some comments.

1. In the proof of the theorem, note the phrase "it is surjective since every linear map is represented by a matrix". This is a typo: it is surjective because every matrix represents a linear map. The phrase as written just shows the map is well defined. We could also finish the proof by showing that the dimension of $\mathcal{L}(V,W)$ is mn and concluding by the Rank-Nullity Theorem.
2. As noted in §4.1, Lang has already shown that $\mathcal{L}(K^n,K^m)$ is isomorphic to $M(m,n)$. But $\mathcal{L}(K^n,K^m)$ and $\mathcal{L}(V,W)$ are obviously isomorphic, by an isomorphism that depends on the choice of a basis for V and a basis for W, so we are done. The isomorphism comes from, for example, Theorem 2.1 of Chapter III, p.56.

On to the multiplicative properties of the associated matrix. Lang proves the next result beautifully: no computation is involved. It is also easy to see when the theorem applies: the middle basis index must be the same (in the statement below, $\mathcal{E}$); and how to apply it: compose the maps, and use the extreme bases.

4.2.15 Theorem. *Let V, W and U be vector spaces of dimensions n, m and p. Let $\mathcal{B}$, $\mathcal{E}$, $\mathcal{H}$ be bases for V, W, U respectively. Let*

$$F\colon V \to W \text{ and } G\colon W \to U$$

be linear maps. Then

$$M^{\mathcal{E}}_{\mathcal{H}}(G)M^{\mathcal{B}}_{\mathcal{E}}(F) = M^{\mathcal{B}}_{\mathcal{H}}(G \circ F),$$

where the left hand side is the matrix product of a matrix of size $p \times m$ by a matrix of size $m \times n$.

Proof. We can extend our previous commutative diagram:

$$\begin{tikzcd} V \arrow[r, "F"] \arrow[d, "X_{\mathcal{B}}"'] & W \arrow[r, "G"] \arrow[d, "X_{\mathcal{E}}"'] & U \arrow[d, "X_{\mathcal{H}}"'] \\ K^n \arrow[r, "M^{\mathcal{B}}_{\mathcal{E}}(F)"] & K^m \arrow[r, "M^{\mathcal{E}}_{\mathcal{H}}(G)"] & K^p \end{tikzcd}$$

and compare to the commutative diagram of the composite:

$$\begin{tikzcd} V \arrow[r, "G\circ F"] \arrow[d, "X_{\mathcal{B}}"'] & U \arrow[d, "X_{\mathcal{H}}"'] \\ K^n \arrow[r, "M^{\mathcal{B}}_{\mathcal{H}}(G\circ F)"] & K^p \end{tikzcd}$$

We need to show they are the same. This is nothing more than (4.1.3). □

Now assume the linear map F is a map from a vector space V to itself. Pick a basis $\mathcal{B}$ of V, and consider

$$M^{\mathcal{B}}_{\mathcal{B}}(F)$$

the matrix associated to F relative to $\mathcal{B}$. Then doing this for the identity map Id, we obviously get the identity matrix I:

$$M^{\mathcal{B}}_{\mathcal{B}}(Id) = I \tag{4.2.16}$$

The following elementary corollary will prove surprisingly useful.

4.2.17 Corollary. *Let V be a vector space and $\mathcal{B}$ and $\mathcal{E}$ two bases. Then*

$$M^{\mathcal{B}}_{\mathcal{E}}(Id)M^{\mathcal{E}}_{\mathcal{B}}(Id) = I = M^{\mathcal{E}}_{\mathcal{B}}(Id)M^{\mathcal{B}}_{\mathcal{E}}(Id),$$

so that $M^{\mathcal{E}}_{\mathcal{B}}(Id)$ is invertible.

Proof. In Theorem 4.2.15, let $V = U = W$, $F = G = Id$, and $\mathcal{H} = \mathcal{B}$. Use (4.2.16), and we are done. □

The next result is very important:

4.2.18 Theorem. *Let $F\colon V \to V$ be a linear map, and let $\mathcal{B}$ and $\mathcal{E}$ be bases of V. Then there is an invertible matrix N such that*

$$M_{\mathcal{E}}^{\mathcal{E}}(F) = N^{-1} M_{\mathcal{B}}^{\mathcal{B}}(F) N$$

Proof. Indeed take $N = M_{\mathcal{B}}^{\mathcal{E}}(id)$, which has inverse $M_{\mathcal{E}}^{\mathcal{B}}(id)$ by Corollary 4.2.17. Then the expression on the right becomes

$$M_{\mathcal{E}}^{\mathcal{B}}(id) M_{\mathcal{B}}^{\mathcal{B}}(F) M_{\mathcal{B}}^{\mathcal{E}}(id) \tag{4.2.19}$$

First apply Theorem 4.2.15 to the two terms on the right in (4.2.19):

$$M_{\mathcal{B}}^{\mathcal{B}}(F) M_{\mathcal{B}}^{\mathcal{E}}(id) = M_{\mathcal{B}}^{\mathcal{E}}(F \circ id) = M_{\mathcal{B}}^{\mathcal{E}}(F)$$

Now apply Theorem 4.2.15 again to the two remaining terms:

$$M_{\mathcal{E}}^{\mathcal{B}}(id) M_{\mathcal{B}}^{\mathcal{E}}(F) = M_{\mathcal{E}}^{\mathcal{E}}(id \circ F) = M_{\mathcal{E}}^{\mathcal{E}}(F)$$

so we are done. □

Two square matrices M and P are called **similar** if there is an invertible matrix N such that

$$P = N^{-1} M N$$

Theorem 4.2.18 shows is that two matrices that represent the same linear transformation $F\colon V \to V$ in different bases of V are similar. We have an easy converse that Lang does not mention:

4.2.20 Theorem. *Assume that two $n \times n$ matrices A and B are similar, so $B = N^{-1}AN$, for an invertible matrix N. Then they represent the same linear operator F.*

Proof. Choose an n dimensional vector space V, a basis $\mathcal{B} = \{v_1, \dots, v_n\}$ for V. Let F be the linear map represented by A in the $\mathcal{B}$ basis, so that if $(x_1, \dots, x_n)$ are the coordinates in the $\mathcal{B}$ basis, the action is by matrix multiplication AX, where X is the column vector of (x_i). Thus $A = M_{\mathcal{B}}^{\mathcal{B}}(F)$

We construct a second basis $\mathcal{E} = \{w_1, \dots, w_n\}$ of V, so that $N = M_{\mathcal{B}}^{\mathcal{E}}(id)$. Indeed, we already constructed it in the first part of Example 4.2.10: in (4.2.11) use for C the transpose of the matrix N, that gives us the basis $\mathcal{E} = \{w_1, \ldots, w_n\}$. By construction $N = M_{\mathcal{B}}^{\mathcal{E}}(id)$, so by Corollary 4.2.17

$$B = M_{\mathcal{E}}^{\mathcal{B}}(id) M_{\mathcal{B}}^{\mathcal{B}}(F) M_{\mathcal{B}}^{\mathcal{E}}(id) = N^{-1}AN$$

as required. □

We indicate that A is similar to B by $A \sim B$.

4.3 Equivalence Relations

We explore the notion of similarity of matrices by introducing equivalence relations.

If S is a set, a binary relation compares two elements of S. When we compare two elements s and t of S, the outcome can be either true or false. We express that the outcome is true by writing sRt.

We are only interested in a specific kind of binary relation, called an equivalence relation. It has three properties: it is reflexive, symmetric and transitive:

4.3.1 Definition. Let R be a binary relation on a set S. Then

1. R is *reflexive* when sRs for all $s \in S$.
2. R is *symmetric* when sRt implies tRs.
3. R is *transitive* when sRt and tRu imply sRu

Read "sRt" as "s is equivalent to t". The most familiar example of an equivalence relation is probably congruence:

4.3.2 Example. Congruence modulo a positive integer k is an equivalence relation on the set of integers $\mathbb{Z}$, defined as follows: Two integers a and b are congruent modulo k, if they have the same remainder under division by k. Each equivalence class contains all the integers whose remainder modulo division by k is a fixed integer. Thus there are k equivalence classes, often denoted $\tilde{0}$, $\tilde{1}$, …, $\widetilde{k-1}$. Thus the equivalence class $\bar{0}$ contains all the multiples of k:

$$\ldots, -2k, -k, 0, k, 2k, \ldots$$

A key fact about an equivalence relation on a set S is that it partitions S into non-overlapping *equivalence classes*.

4.3.3 Definition. A *partition* of a set S is a collection of non-overlapping subsets S_i, called *equivalence classes*, whose union is S. Thus for any two i and j in I, the intersection $S_i \cap S_j$ is empty, and the union $\cup_{i \in I} S_i = S$.

4.3.4 Proposition. *A partition $\{S_i, i \in I\}$ defines an equivalence relation P on $S \times S$, whose domain and range is all of S: sPt if s and t are in the same subset S_i. Conversely any equivalence relation R defines a partition of S, where each equivalence class S_s consists of the elements $t \in S$ that are equivalent to a given element s.*

Proof. It is easy to show that P satisfies the three properties of an equivalence relation. For the converse, just show that the sets S_s are either the same, or disjoint. Their union is obviously S. □

The key theorem is very easy to prove:

4.3.5 Theorem. *Similarity is an equivalence relation on $n \times n$ matrices.*

Proof. To prove that $\sim$ is an equivalence relation, we need to establish the following three points:

1. $A \sim A$: Use the identity matrix for C.
2. if $A \sim B$, then $B \sim A$:
   If $A \sim B$, there is an invertible C such that $B = C^{-1}AC$. Then, multiplying both sides of the equation on the right by C^{-1} and on the left by C, and letting $D = C^{-1}$, we see that $A = D^{-1}BD$, so B is similar to A.
3. if $A \sim B$ and $B \sim D$, then $A \sim D$:
   The hypotheses mean that there are invertible matrices C_1 and C_2 such that $B = C_1^{-1}AC_1$ and $D = C_2^{-1}BC_2$, so, substituting from the first equation into the second, we get

   $$D = C_2^{-1}C_1^{-1}AC_1C_2 = (C_1C_2)^{-1}A(C_1C_2),$$

   so A is similar to D using the matrix C_1C_2.

□

Since similarity is an equivalence relation on $n \times n$ matrices, it partitions these matrices into equivalence classes. View each $n \times n$ matrix as the matrix $M_{\mathcal{B}}^{\mathcal{B}}(T)$ of a linear operators T in the basis $\mathcal{B}$ of the n-dimensional vector space V. The importance of this relation is that the elements of a given equivalence class correspond to the same linear operator T, but expressed in different basis. One of the goals of the remainder of this course is to determine the common features of the matrices in a given similarity class. For example Lang shows in Chapter VIII, Theorem 2.1 that similar matrices have the same characteristic polynomial. However at the end of this course we will see that two matrices that have the same characteristic polynomial need not be similar: this follows from the construction of the

Jordan normal form in Chapter XI, §6. The simplest example is probably given by the matrices

$$\begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix} \; and \; \begin{pmatrix} \alpha & 1 \\ 0 & \alpha \end{pmatrix}$$

for any complex number α.

4.3.6 Remark. When you read Lang's Chapter XI, §6, be aware that Figure 1 on p. 264 is potentially misleading: the α_i are *not* assumed to be distinct.

4.3.7 Exercise. Show that row equivalence (see Definition 2.2.2) is an equivalence on $n \times m$ matrices.

We will study one further equivalence relation in Exercise 11.1.6.

4.4 Solved Exercises

Here are some easy exercises with solutions:

4.4.1 Exercise. Let F be the linear transformation from K^3 to K^2 given by

$$F \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 5x_1 - 7x_2 - x_3 \\ x_2 + x_3 \end{pmatrix}$$

What is the matrix A of F with respect to the standard bases of K^3 and K^2?

Solution: The x_i are coordinates with respect to the standard basis: indeed, a vector can be written $x_1 E^1 + x_2 E^2 + x_3 E^3$ where the E^i are our notation for the standard basis. Then the matrix is clearly

$$\begin{pmatrix} 5 & -7 & -1 \\ 0 & 1 & 1 \end{pmatrix}, \text{ so that when you multiply by } \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

on the right you get the original matrix back.

4.4.2 Exercise. Let V be a vector space of dimension 3, and let $\mathcal{B} = \{v_1, v_2, v_3\}$ be a basis for V. Let L be a linear operator from V to itself with matrix

$$M_{\mathcal{B}}^{\mathcal{B}}(L) = \begin{pmatrix} 2 & 0 & 1 \\ 1 & 1 & 5 \\ 2 & 1 & 2 \end{pmatrix}$$

relative to this basis. What is $L(2v_1 - v_2 + 5v_3)$? Is the operator L invertible?

Solution: Again, a very easy problem once you untangle the definitions. The coordinates of $2v_1-v_2+5v_3$ are $(2,-1,5)$, so the answer is just the matrix product:

$$\begin{pmatrix} 2 & 0 & 1 \\ 1 & 1 & 5 \\ 2 & 1 & 2 \end{pmatrix} \begin{pmatrix} 2 \\ -1 \\ 5 \end{pmatrix} = \begin{pmatrix} 9 \\ 26 \\ 13 \end{pmatrix}$$

The second part is unrelated. We answer in the usual way by doing row operations. First we interchange two rows to simplify the computations, and then we make the matrix upper triangular:

$$\begin{pmatrix} 1 & 1 & 5 \\ 2 & 0 & 1 \\ 2 & 1 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 5 \\ 0 & -2 & -9 \\ 0 & -1 & -8 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 5 \\ 0 & 0 & 7 \\ 0 & -1 & -8 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 5 \\ 0 & -1 & -8 \\ 0 & 0 & 7 \end{pmatrix}$$

so it is invertible by Theorem 2.2.7: all the diagonal elements are nonzero.

4.4.3 Exercise. Let V be a vector space of dimension 3, and let $\mathcal{B} = \{v_1, v_2, v_3\}$ be a basis for V. Let L be a linear operator from V to itself such that $L(v_1) = 2v_1 - v_3$, $L(v_2) = v_1 - v_2 + v_3$, $L(v_3) = v_2 + 2v_3$. Compute the matrix $M^{\mathcal{B}}_{\mathcal{B}}$ of the linear transformation.

Solution: Here we are given the transformation on the basis elements. On the basis elements the transformation is given by the matrix

$$\begin{pmatrix} 2 & 0 & -1 \\ 1 & -1 & 1 \\ 0 & 1 & 2 \end{pmatrix}$$

So by the big theorem of Lang, Chapter IV, §3, recalled in Theorem 4.2.7: the matrix representing the linear transformation is the transpose of the matrix on the basis elements. Thus the answer is

$$\begin{pmatrix} 2 & 1 & 0 \\ 0 & -1 & 1 \\ -1 & 1 & 2 \end{pmatrix}$$

And now two exercises from Lang [IV, 3].

In the following two exercises, you are asked to find $M^{\mathcal{B}}_{\mathcal{B}'}(id)$, as defined in Lang p. 88. Denote by B the 3×3 matrix whose columns are the basis vectors of

$\mathcal{B}$, and by B' the 3×3 matrix whose columns are the basis vectors of $\mathcal{B}'$. As we will see, the answer is the columns of the matrix product

$$B'^{-1}B.$$

This is worth remembering. Thus once you unwind the definitions, the only difficulty is to compute the inverse of the matrix B' using Gaussian elimination.

Exercise 1 a)

The source basis $\mathcal{B}$ is

$$\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \tag{4.4.4}$$

and the target basis $\mathcal{B}'$ is

$$\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}.$$

We want to write each one of the source basis vectors as a combination of the target basis vectors. The key point is that we are solving the linear system

$$\begin{pmatrix} 2 & 0 & -1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix} \tag{4.4.5}$$

for the three choices of the right hand side corresponding to the three basis vectors of $\mathcal{B}$. So we compute the inverse C of B' by setting up, as usual:

$$\begin{pmatrix} 2 & 0 & -1 & | & 1 & 0 & 0 \\ 1 & 0 & 1 & | & 0 & 1 & 0 \\ 1 & 1 & 1 & | & 0 & 0 & 1 \end{pmatrix}$$

and doing row operations to make the left half of the matrix the identity matrix. The C will appear on the right hand side. First divide the first row by 2:

$$\begin{pmatrix} 1 & 0 & -1/2 & | & 1/2 & 0 & 0 \\ 1 & 0 & 1 & | & 0 & 1 & 0 \\ 1 & 1 & 1 & | & 0 & 0 & 1 \end{pmatrix}$$

Clear the first column:

$$\begin{pmatrix} 1 & 0 & -1/2 & | & 1/2 & 0 & 0 \\ 0 & 0 & 3/2 & | & -1/2 & 1 & 0 \\ 0 & 1 & 3/2 & | & -1/2 & 0 & 1 \end{pmatrix}$$

Interchange the second and third rows:

$$\begin{pmatrix} 1 & 0 & -1/2 & | & 1/2 & 0 & 0 \\ 0 & 1 & 3/2 & | & -1/2 & 0 & 1 \\ 0 & 0 & 3/2 & | & -1/2 & 1 & 0 \end{pmatrix}$$

Subtract the third row from the second, multiply the third row by $2/3$:

$$\begin{pmatrix} 1 & 0 & -1/2 & | & 1/2 & 0 & 0 \\ 0 & 1 & 0 & | & 0 & -1 & 1 \\ 0 & 0 & 1 & | & -1/3 & 2/3 & 0 \end{pmatrix}$$

Finally add to the first row $1/2$ times the third:

$$\begin{pmatrix} 1 & 0 & 0 & | & 1/3 & 1/3 & 0 \\ 0 & 1 & 0 & | & 0 & -1 & 1 \\ 0 & 0 & 1 & | & -1/3 & 2/3 & 0 \end{pmatrix}$$

So the inverse matrix C of B' is

$$\begin{pmatrix} 1/3 & 1/3 & 0 \\ 0 & -1 & 1 \\ -1/3 & 2/3 & 0 \end{pmatrix}$$

as you should check, and to get the answer just right multiply C by the matrix B whose columns are in (4.4.4)

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

a matrix whose columns give the same answer as in (4.4.6).

Here is a slightly different approach: start from (4.4.5), and solve for the x_i. A similar computation gives

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \frac{r_1+r_2}{3} \\ -r_2 + r_3 \\ \frac{-r_1+2r_2}{3} \end{pmatrix}$$

Now just plug in the values for the r_i. When they are $(1, 1, 0)$, $(-1, 1, 1)$, $(0, 1, 2)$ we get in turn for the $\{x_i\}$ the three column vectors:

$$\begin{pmatrix} 2/3 \\ -1 \\ 1/3 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1/3 \\ 1 \\ 2/3 \end{pmatrix}, \tag{4.4.6}$$

and they are the columns of the matrix we are looking for, since they are the coordinates, in the $\mathcal{B}'$ basis, of the unit coordinate vectors in the $\mathcal{B}$ basis.

Exercise 1 b) In this exercise the source basis $\mathcal{B}$ is:

$$\begin{pmatrix}3\\2\\1\end{pmatrix}, \begin{pmatrix}0\\-2\\5\end{pmatrix}, \begin{pmatrix}1\\1\\2\end{pmatrix}, \tag{4.4.7}$$

Using the same technique, since the target basis $\mathcal{B}'$ are the columns of the matrix below, we get the system

$$\begin{pmatrix}1 & -1 & 2\\1 & 2 & -1\\0 & 4 & 1\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} = \begin{pmatrix}r_1\\r_2\\r_3\end{pmatrix}$$

As before, we do Gaussian elimination. Here are the steps. First subtract the first row from the second row:

$$\begin{pmatrix}1 & -1 & 2\\0 & 3 & -3\\0 & 4 & 1\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} = \begin{pmatrix}r_1\\r_2-r_1\\r_3\end{pmatrix}$$

Next subtract $4/3$ times the second row from the third:

$$\begin{pmatrix}1 & -1 & 2\\0 & 3 & -3\\0 & 0 & 5\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} = \begin{pmatrix}r_1\\r_2-r_1\\r_3-(4/3)(r_2-r_1)\end{pmatrix}$$

Thus $x_3 = (r_3 - (4/3)(r_2 - r_1))/5 = (1/15)(4r_1 - 4r_2 + 3r_3)$, and now, back-substituting

$$3x_2 = r_2 - r_1 + 3x_2 = r_2 - r_1 + (1/5)(4r_1 - 4r_2 + 3r_3) = (1/5)(-r_1 + r_2 + 3r_3)$$

so

$$x_2 = (1/15)(-r_1 + r_2 + 3r_3).$$

Finally

$$\begin{aligned}x_1 &= r_1 + x_2 - 2x_3 = r_1 + (1/15)(-r_1 + r_2 + 3r_3) - (2/15)(4r_1 - 4r_2 + 3r_3)\\ &= (1/5)(2r_1 + 3r_2 - r_3).\end{aligned}$$

So

$$\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} = \begin{pmatrix}(1/5)(2r_1+3r_2-r_3)\\(1/15)(-r_1+r_2+3r_3)\\(1/15)(4r_1-4r_2+3r_3)\end{pmatrix} = \frac{1}{15}\begin{pmatrix}6r_1+9r_2-3r_3)\\-r_1+r_2+3r_3\\4r_1-4r_2+3r_3\end{pmatrix}.$$

For the first coordinate vector in $\mathcal{B}$ from (4.4.7) we have $r_1 = 3$, $r_2 = 2$, $r_3 = 1$, so we get

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 11/5 \\ 2/15 \\ 7/15 \end{pmatrix}$$

For the second coordinate vector $r_1 = 0$, $r_2 = -2$, $r_3 = 5$ in $\mathcal{B}$ from (4.4.7) we get

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -11/5 \\ -13/15 \\ 23/15 \end{pmatrix}$$

The last one is obtained in the same way.

Chapter 5

Row Rank is Column Rank

One of the most important theorems of linear algebra is the theorem that says that the row rank of a matrix is equal to its column rank. This result is sometimes called the fundamental theorem of linear algebra. The result, and its proof, is buried in Lang's Chapter V. It is stated as Theorem 3.2 p.114, but the proof requires a result from §6: Theorem 6.4. The goal here is to give the result the prominence it deserves, and to give a different proof. The one here uses Gaussian elimination.

5.1 The Rank of a Matrix

We start with a definition.

5.1.1 Definition. If A is a $m \times n$ matrix, then the *column rank* of A is the dimension of the subspace of K^m generated by the columns $\{A^1, \dots, A^n\}$ of A. The *row rank* of A is the dimension of the subspace of K^n generated by the rows $\{A_1, \dots, A_m\}$ of A.

First, as Lang points out on page 113, the space of solutions of a $m \times n$ homogeneous system of linear equations $AX = O$ can be interpreted in two ways:

1. either as those vectors X giving linear relations

$$x_1 A^1 + \cdots + x_n A^n = O$$

between the columns A^j of A,

2. or as those vectors X orthogonal to the row vectors A_i of A:

$$\langle X, A_1 \rangle = 0, \dots, \langle X, A_m \rangle = 0.$$

These are just two different ways of saying that X is in the kernel of the linear map L_A we investigated in Chapter IV. The second characterization uses the standard scalar product on K^n and K^m. Continuing in this direction:

5.1.2 Definition. For any subspace W of K^n, the orthogonal complement $W^\perp$ of W in K^n is defined by

$$W^\perp = \{y \in K^n \mid \langle w, y\rangle = 0, \text{for all } w \in W\}.$$

5.1.3 Proposition. *$W^\perp$ is a subspace of K^n, and its dimension is n - $\dim W$.*

Proof. If $W = \{O\}$, then $W^\perp$ is V, so we are done. We may therefore assume that $\dim W \geq 1$. Pick a basis $\{w_1, \dots, w_r\}$ of W. Write

$$w_i = (c_{i1}, c_{i2}, \dots, c_{in}) \in K^n.$$

Also write $y = (y_1, y_2, \dots, y_n) \in K^n$. Letting C be the $r \times n$ matrix (c_{ij}), it follows that y is in $W^\perp$ if and only if y is in the kernel of C. Since the w_i form a basis of W, there is no relation of linear dependence between the rows of C. Thus the row rank of C is r.

Now perform row reduction on C to put C in row echelon form, as defined in 1.1.3. Call the new matrix B. By definition, the row rank of C and the row rank of B are the same. We now prove an easy generalization of Proposition 2.2.5:

5.1.4 Proposition. *Let B be a $r \times n$ matrix, $r \leq n$, in row echelon form. Then B has row rank r if and only if it does not have a row of zeroes.*

Proof. Notice that Proposition 2.2.5 is the case $r = n$. To have row rank r, B must have linearly independent rows. If B has a row B_i of zeroes, then it satisfies the row equation $B_i = O$, a contradiction. Conversely, if B does not have a row of zeroes, then for each row i of B there is an index $\mu(i)$ so that the entry $b_{i,\mu(i)}$ is the first non-zero coordinate of row B_i. Note that $\mu(i)$ is a strictly increasing function of i, so that there are exactly $n - r$ indices $j \in [1, n]$ that are not of the form $\mu(i)$. The variables y_j for these $n - r$ indices are called the *free variables*. Assume the B_i satisfy an equation of linear dependence

$$\lambda_1 B_i + \lambda_2 B_2 + \dots + \lambda_r B_r = O.$$

Look at the equation involving the $\mu(1)$ coordinate. The only row with a non-zero entry there is B_1. Thus $\lambda_1 = 0$. Continuing in this way, we see that all the λ_i are 0, so this is not an equation of linear dependence. □

Back to the proof of Proposition 5.1.3. We can give arbitrary values to each one of the free variables and then do backsubstitution as in §1.3 to solve uniquely for the remaining variables. This implies that the space of solutions have dimension $n - r$, so we are done. □

Now we can prove one of the most important theorems of linear algebra.

5.1.5 Theorem. *The row rank and the column rank of any matrix are the same.*

Proof. Let L_A be the usual map: $K^n \to K^m$. By the Rank-Nullity Theorem, we have $\dim Im(L_A) + \dim Ker(L_A) = n$, which we can now write

$$\text{column rank } L_A + \dim Ker(L_A) = n.$$

Now the second interpretation of the space of solutions tells us that $Ker(L_A)$ is the subspace orthogonal to the space of row vectors of A. By Proposition 5.1.3, it has dimension n - row rank. So

$$\text{row rank } L_A + \dim Ker(L_A) = n.$$

Comparing the two equations, we are done.

□

Chapter 6

Duality

This chapter is a clarification and an expansion of Lang [4], section [V, §6] entitled The Dual Space and Scalar Products. The first three sections clarify the material there, the remaining sections contain new results.

If V is a vector space over the field K, V^* denotes the vector space of linear maps $f\colon V \to K$. These linear maps are called functionals. V^* is called the dual space of V, and the passage from V to V^* is called duality.

6.1 Non-degenerate Scalar Products

A scalar product on V is a map to K, written $\langle v, w\rangle$, $v \in V$, $w \in V$ that satisfies Lang's three axioms:

SP 1. $\langle v, w\rangle = \langle w, v\rangle$ for all v abd w in V;

SP 2. For all elements u, v and w of V, $\langle u, v + w\rangle = \langle u, v\rangle + \langle u, w\rangle$;

SP 3. If $x \in K$, then $\langle xv, w\rangle = x\langle v, w\rangle$.

In his definition (see p.94) Lang adds to **SP 3** the additional condition

$$\langle v, xw\rangle = x\langle v, w\rangle.$$

This is clearly unnecessary, since

$$\begin{aligned}
\langle v, xw\rangle &= \langle xw, v\rangle && \text{by } \textbf{SP 1}\\
&= x\langle w, v\rangle && \text{by } \textbf{SP 3}\\
&= x\langle v, w\rangle && \text{by } \textbf{SP 1}.
\end{aligned}$$

The scalar product is *non-degenerate* if the only $v \in V$ such that

$$\langle v, w \rangle = 0 \text{ for all } w \in V$$

is the zero vector O.

An important theorem Lang proves in Chapter V, §5, p.124 is

6.1.1 Theorem. *Any scalar product on a finite dimensional vector space V admits an orthogonal basis.*

Recall that an *orthogonal basis* $\{v_1, v_2, \dots, v_n\}$ is a basis where

$$\langle v_i, v_j \rangle = 0 \text{ whenever } i \neq j.$$

This is defined p.103. Lang proves this theorem by a variant of the Gram-Schmidt process.

As Lang notes in Theorem 4.1 of Chapter V, if X and Y are column vectors in K^n, then a scalar product on K^n is equivalent to specifying a unique symmetric $n \times n$ matrix A so that

$$\langle X, Y \rangle = X^t AY. \tag{6.1.2}$$

Returning to V, if we choose an orthogonal basis $\mathcal{B}$ for V and its scalar product, and let X and Y be the component vectors in K^n of vectors v and w in V, then the symmetric matrix A of (6.1.2) describing the scalar product is diagonal.

We get the easy corollary, not mentioned by Lang:

6.1.3 Corollary. *If the scalar product is non-degenerate, and an orthogonal basis $\mathcal{B}$ of V has been chosen, then the diagonal elements of the associated diagonal matrix A are non-zero.*

The importance of having a non-degenerate scalar product is revealed by Theorem 6.2 p.128: then there is an isomorphism (to which Lang does not give a name):

$$D \colon V \to V^*$$

given by $v \mapsto L_v$, where L_v is the functional taking the values

$$L_v(u) = \langle v, u \rangle, \text{ for all } u \in V.$$

So $D(v) = L_v$.

6.2 The Dual Basis

Lang defines the dual basis on p. 127. Here is the definition:

6.2.1 Definition. Let $\mathcal{B} = \{v_1, \dots, v_n\}$ be a basis for V. Then the dual basis $\mathcal{B}^* = \{\varphi_1, \dots, \varphi_n\}$ of V^* is defined as follows. For each φ_j, set

$$\varphi_j(v_k) = \begin{cases} 1, & \text{if } k = j; \\ 0, & \text{otherwise.} \end{cases}$$

By Lang's Theorem 2.1 of Chapter III, this specifies φ_j uniquely.

He uses the dual basis to prove his Theorem 6.3 of Chapter V.

One drawback of the isomorphism D defined above is that it depends not only on V, but also on the inner product on V.

6.2.2 Example. Let V be two-dimensional, with a basis $\mathcal{B} = \{v_1, v_2\}$. Let $\mathcal{B}^* = \{\varphi_1, \varphi_1\}$ be the dual basis on V^*.

First assume that the scalar product of V has as matrix A the identity matrix. Then for this scalar product the isomorphism D satisfies $D(v_i) = \varphi_1$ and $D(v_2) = \varphi_2$ as you should check.

Now instead assume the scalar product has in the same basis the matrix

$$A = \begin{pmatrix} a & b \\ b & d \end{pmatrix} \text{ with } ad - b^2 \neq 0.$$

The last condition is needed to insure that the scalar product is non-degenerate. Let's call the map from V to V^* given by this scalar product E instead of D, since we will get a different linear map. The definition shows that $E(v_1)$ is the functional which in the $\mathcal{B}$ for V takes the value

$$\begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ b & d \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = ax_1 + bx_2$$

on the vector $x_1v_1 + x_2v_2$. So $E(v_1)$ is the functional $a\varphi_1 + b\varphi_2$ in the dual basis on V^*. In the same way, $E(v_2)$ is the functional $b\varphi_1 + d\varphi_2$, so we get a different map when we choose a different scalar product..

This example shows that choosing a non-degenerate scalar product on V is equivalent to choosing a basis of V and the dual basis on V^*. See Remark 6.7.2 for details.

6.3 The Orthogonal Complement

In Lang's statement of Theorem 6.3, p. 129, the symbol $W^{\perp}$ is used to denote an object different from one that it was used for up to then. This is bad practice so here we will use instead:

$$W^{*\perp} = \{\varphi \in V^* \text{ such that } \varphi(W) = 0\}. \tag{6.3.1}$$

In this notation Theorem 6.3 says that

$$\dim W + \dim W^{*\perp} = \dim V.$$

Note that this theorem holds for all finite dimension vectors spaces V, without reference to a scalar product Its proof is very similar to that of Theorem 2.3 p. 106.

Now assume that V is an n-dimensional vector space with a non-degenerate scalar product. Then the easy Theorem 6.2 p.128 mentioned above and the comments in Lang starting at the middle of page 130 establish the following theorem in this context.

6.3.2 Theorem. *For any subspace W of V, D restricts to an isomorphism*

$$W^{\perp} \simeq W^{*\perp}.$$

This is the key point, so we should prove it, even though it is easy.

Proof. $W^{\perp}$, which Lang redefines as $perp_V(W)$ to avoid confusion with $W^{*\perp}$, is just

$$W^{\perp} = \{v \in V \text{ such that } \langle v, w\rangle = 0\}.$$

On the other hand $W^{*\perp}$, defined by Lang as $perp_{V^*}(W)$ is given in (6.3.1) above.

Take any $u \in W^{\perp}$, so that $\langle w, u\rangle = 0$ for all $w \in W$. Its image under the isomorphism of Theorem 6.2 is the map L_u, which is in $W^{*\perp}$ by definition. Conversely, if we start with a map L_u in $W^{*\perp}$, it comes from a u in $W^{\perp}$. □

Notice the logical progression: we want to establish Lang's Theorem 6.4 p. 131, in the case of a vector space V with a non-degenerate scalar product. Thus we want to prove:

$$\dim W + \dim W^{\perp} = \dim V.$$

To prove this, we first prove Lang's Theorem 6.3 which asserts that in any vector space V,

$$\dim W + \dim W^{*\perp} = \dim V.$$

(Alas, confusingly $W^{*\perp}$ is written $W^{\perp}$ in Lang.) Then we use Theorem 6.3.2 above to establish the isomorphism of $W^{\perp}$ with $W^{*\perp}$, when V has a non-degenerate scalar product. So $\dim W^{\perp} = \dim W^{*\perp}$ and we are done.

Why do we need Theorem 6.4? In order to prove that row rank equals column rank. In the approach of Chapter 5, we used the (non-degenerate) dot product on K^n to establish Proposition 5.1.3, our analog of Theorem 6.4. That meant that we did not have to pass through Theorem 6.3.2 above.

6.4 The Double Dual

Given a n-dimensional vector space V, we have constructed its dual space V^*, which has the same dimension by Lang, Theorem 6.1. The isomorphism between V and V^* depends on the choice of non-degenerate scalar product of V used to construct it, as we showed in Example 6.2.2.

Next we can take the dual of V^*, the *double dual* of V, written V^{**}. Obviously it has dimension n, but it also has a natural isomorphism with V, something V^* does not have. The material below is only mentioned in Lang's Exercise 6, p. 131.

6.4.1 Definition. Pick a $v \in V$. For any $\varphi \in V^*$, let $e_v(\varphi) = \varphi(v)$. The map $e_v \colon V^* \to K$ is easily seen to be linear. It is called *evaluation at* v.

6.4.2 Theorem. *The map $D_2 \colon v \mapsto e_v$ is an isomorphism of V with V^{**}.*

Proof. First we need to show D_2 is a linear map. The main point is that for two elements v and w of V,

$$e_{v+w} = e_v + e_w.$$

To show this we evaluate e_{v+w} on any $\varphi \in V^*$:

$$e_{v+w}(\varphi) = \varphi(v+w) = \varphi(v) + \varphi(w) = e_v(\varphi) + e_w(\varphi)$$

just using the linearity of φ. Thus $D_2(v+w) = D_2(v) + D_2(w)$. The remaining point $D_2(cv) = cD_2(v)$ is left to you.

Next, to show D_2 is an isomorphism, since the spaces are of the same dimension, by the Rank-Nullity theorem (Lang Theorem 3.2 of Chapter III) all we have to do is show D_2 is injective. Suppose not: that means that there is a v such that e_v evaluates to 0 on all $\varphi \in V^*$, so $\varphi(v) = 0$. But that is absurd: all functionals cannot vanish at a point. For example extend v to a basis of V and let ϕ be the element in the dual basis dual to v, so $\phi(v) = 1$. □

Thus we can treat V and V^{**} as the same space.

6.5 The Dual Map

As in Lang, Chapter IV, §3, suppose we have a vector space V of dimension n, a vector space W of dimension m, and a linear map F between them:

$$F\colon V \to W.$$

To each $\psi \in W^*$, which is of course a linear map $W \to K$, we can associate the composite map:

$$\psi \circ F\colon V \to W \to K.$$

In this way we get a linear map, that we call F^*, from W^* to V^*.

$$F^*\colon \psi \in W^* \mapsto \varphi = \psi \circ F \in V^*.$$

Our goal is to understand the relationship between the $m \times n$ matrix of F and the $n \times m$ matrix of F^*, in suitable bases, namely the dual bases.

For the vector space V^* we use the dual basis $\mathcal{B}^* = \{\varphi_1, \dots, \varphi_n\}$ of the basis $\mathcal{B}$ and for the vector space W^* we use the dual basis $\mathcal{E}^* = \{\psi_1, \dots, \psi_m\}$ of the basis $\mathcal{E}$.

The $m \times n$ matrix for F associated to $\mathcal{B}$ and $\mathcal{E}$ is denoted $M^{\mathcal{B}}_{\mathcal{E}}(F)$, as per Lang p. 88, while the matrix for F^* associated to $\mathcal{E}^*$ and $\mathcal{B}^*$ is denoted $M^{\mathcal{E}^*}_{\mathcal{B}^*}(F^*)$.

What is the relationship between these two matrices? Here is the answer.

6.5.1 Theorem. *One matrix is the transpose of the other:*

$$M^{\mathcal{E}^*}_{\mathcal{B}^*}(F^*) = M^{\mathcal{B}}_{\mathcal{E}}(F)^t.$$

Proof. Any functional $\varphi \in V^*$ can be written in terms of the dual basis as:

$$\varphi = \varphi(v_1)\varphi_1 + \cdots + \varphi(v_n)\varphi_n. \tag{6.5.2}$$

Just test the equality by applying the functionals to any basis vector v_j to see that both sides agree, since all the terms on the right hand side except the j-th one vanish by definition of the dual basis. That means they are the same.

Writing A for $M^{\mathcal{B}}_{\mathcal{E}}(F)$, then for any j:

$$F(v_j) = a_{1j}w_1 + \cdots + a_{mj}w_m. \tag{6.5.3}$$

This is what we learned in Chapter IV, §3. Note the indexing going in the wrong order: in other words, we are taking the dot product of w with the j-th column of A.

Now we compute. For any dual basis vector ψ_j we get:

$$\begin{aligned}
F^*(\psi_j) &= \psi_j \circ F && \text{in } V^*. \\
&= [\psi_j \circ F](v_1)\varphi_1 + \cdots + [\psi_j \circ F](v_n)\varphi_n && \text{by (6.5.2)} \\
&= \psi_j(F(v_1))\varphi_1 + \cdots + \psi_j(F(v_n))\varphi_n \\
&= \psi_j(a_{11}w_1 + \cdots + a_{m1}w_m)\varphi_1 + \\
&\cdots + \psi_j(a_{1n}w_1 + \cdots + a_{mn}w_m)\varphi_n && \text{by (6.5.3)} \\
&= a_{j1}\varphi_1 + \cdots + a_{jn}\varphi_n && \text{since } \psi_j \text{ is dual to } w_i,
\end{aligned}$$

in each sum we only keep the j-th term, and use $\psi_j(w_j) = 1$. Then from Chapter IV, §3, we know that the matrix $M_{\mathcal{B}^*}^{\mathcal{E}^*}(F^*)$ is the transpose of the matrix implied by this formula: therefore it is A^t. $\square$

6.6 The Four Subspaces

We now bring everything together. We start with a linear map $F\colon V \to W$, vector spaces of dimension n and m. We pick bases $\mathcal{B}$ and $\mathcal{E}$ of V and W, as in the previous section. Call A the matrix $M_{\mathcal{E}}^{\mathcal{B}}(F)$ associated to F.

As we saw in §6.2, we get a linear map F^* from W^* to V^*, and the matrix $M_{\mathcal{B}^*}^{\mathcal{E}^*}(F^*)$ of F^* in the dual bases is just the transpose of A.

We also have the linear maps $D_{\mathcal{B}}\colon V \to V^*$ and $D_{\mathcal{E}}\colon W \to W^*$ defined in Theorem 6.3.2, which says that they are isomorphisms if the scalar product on each vector space is non-degenerate. Let us assume that. Then we get a diagram of linear maps between vector spaces, where the vertical maps are isomorphisms.

$$\begin{array}{ccc}
V & \xrightarrow{F} & W \\
{\scriptstyle D_{\mathcal{B}}}\downarrow & & {\scriptstyle D_{\mathcal{E}}}\downarrow \\
V^* & \xleftarrow[F^*]{} & W^*
\end{array}$$

Now identify V and V^* each to K^n using the basis $\mathcal{B}$ and the dual basis $\mathcal{B}^*$; also identify W and W^* each to K^m using the basis $\mathcal{E}$ and the dual basis $\mathcal{E}^*$.

Let $B = M_{\mathcal{B}^*}^{\mathcal{B}}(D_{\mathcal{B}})$ be the matrix representing $D_{\mathcal{B}}$ is these two bases, and $E = M_{\mathcal{E}^*}^{\mathcal{E}}(D_{\mathcal{E}})$ the matrix representing $D_{\mathcal{E}}$. Then we get the corresponding diagram of matrix multiplication:

$$\begin{array}{ccc}
K^n & \xrightarrow{A} & K^m \\
{\scriptstyle B}\downarrow & & {\scriptstyle E}\downarrow \\
K^n & \xleftarrow[A^t]{} & K^m
\end{array}$$

This is most interesting if the scalar product on V is the ordinary dot product in the basis $\mathcal{B}$, and the same holds on W in the basis $\mathcal{E}$. Then it is an easy exercise to show that the matrices B and E are the identity matrices, so the diagram becomes:

$$\begin{tikzcd} K^n \arrow[r, "A"] \arrow[d, "I"'] & K^m \arrow[d, "I"'] \\ K^n & K^m \arrow[l, "A^t"] \end{tikzcd}$$

Then we get the following rephrasing of Theorem 5.1.5:

6.6.1 Theorem. *Let A be a $m \times n$ matrix, so that L_A is a linear map from K^n to K^m, with both given the ordinary dot product. Consider also the linear map $L_{A^t}\colon K^m \to K^n$ associated to the transpose of A. Then the kernel of L_A is the orthogonal complement of the image of L_{A^t} in K^n, and the kernel of L_{A^t} is the orthogonal complement of the image of L_A in K^m. In particular this implies*

$$Ker(L_A) \oplus Image(L_{A^t}) = K^n,$$
$$Image(L_A) \oplus Ker(L_{A^t}) = K^m.$$

Thus the four subspaces associated to the matrix A are the kernels and images of A and A^t. The text by Strang [6] builds systematically on this theorem.

6.7 A Different Approach

We give a completely elementary approach, without scalar products, to the above material.

Start with a vector space V over K with a basis $\mathcal{B} = \{v_1, \dots, v_n\}$. To every $v \in V$ we construct maps to K as follows. The element v can be written uniquely as

$$v = x_1 v_1 + x_2 v_2 + \cdots + x_n v_n$$

so let φ_i be the map from V to K given by $\varphi_i(v) = x_i$.

These are functionals, called the coordinate functions. See Lang Example 1 [V, §6]. We must show they are linear. Indeed, if

$$w = y_1 v_1 + y_2 v_2 + \cdots + y_n v_n$$

then

$$\varphi_i(v + w) = x_i + y_i = \varphi_i(v) + \varphi_i(w)$$

and

$$\varphi_i(cv) = cx_i = c\varphi_i(v).$$

Note that $\varphi_i(v_j) = 0$ if $i \neq j$, and $\varphi_i(v_1) = 1$.

Then to every element w we can associate a functional called φ_w given by

$$\varphi_w = y_1\varphi_1 + y_2\varphi_2 + \cdots + y_n\varphi_n.$$

.

Thus we have constructed a map from V to V^* that associates to any w in V the functional φ_w. We need to show this is a linear map: this follows from the previous computations.

6.7.1 Proposition. *The φ_i, $1 \leq i \leq n$ form a basis of V^*.*

Proof. First we show the φ_i are linearly independent. Assume not. Then there is an equation of linear dependence:

$$a_1\varphi_1 + a_2\varphi_i + \cdots + a_n\varphi_n = O.$$

Apply the equation to v_j. Then we get $a_j = 0$, so all the coefficients are zero, and this is not an equation of dependence. Now take an arbitrary functional ψ on V. Let $a_j = \psi(v_j)$. Now consider the functional

$$\psi - a_1\varphi_1 - a_2\varphi_i - \cdots - a_n\varphi_n$$

Applied to any basis vector v_i this functional gives 0. So it is the zero functional and we are done. □

So we have proved that the dimension of V^* is n. The basis $\mathcal{B}^* = \{\varphi_1, \ldots, \varphi_n\}$ is called the dual basis of the basis $\mathcal{B}$.

6.7.2 Remark. Now, just to connect to Lang's approach, let's define a scalar product on V using the bases $\mathcal{B}$ and $\mathcal{B}^*$. Let

$$\langle v, w\rangle = \varphi_v(w).$$

The linearity in both variables is trivial. To conclude, we must show we have symmetry, so

$$\langle v, w\rangle = \langle v, w\rangle = \varphi_w(v).$$

With the same notation as before for the components of v and w, we get

$$\begin{aligned}
\varphi_v(w) &= x_1\varphi_1(w) + \cdots + x_n\varphi_n(w) \\
&= x_1y_1 + \cdots + x_ny_n \\
&= y_1\varphi_1(v) + \cdots + y_n\varphi_n(v) \\
&= \varphi_w(v)
\end{aligned}$$

so we have symmetry. This product is clearly non-degenerate: to establish that, we must check that for every non-zero $v \in V$ there is a $w \in V$ such that $\langle v, w \rangle \neq 0$. If v is non zero, it has at least one non zero coefficient x_i in the $\mathcal{B}$ basis. Then take $\langle v, v_i \rangle = x_i \neq 0$, so we are done.

We now use these ideas to show that the row rank and the column rank of a matrix are the same.

Let A be a $m \times n$ matrix. Let N be the kernel of the usual map $K^n \to K^m$ given by matrix multiplication:

$$x \mapsto Ax.$$

View the rows of A as elements of V^*. Indeed, they are the elements of V^* such that when applied to N, give 0. Pick any collection of linearly independent rows of A, then extend this collection to a basis of V^*. Pick the dual basis for $V^{**} = V$. By definition N is the orthogonal complement of the collection of independent rows of A. So

$$\dim N + \text{row rank} A = n.$$

Combining with the Rank-Nullity Theorem, we get the desired result.

Chapter 7

Orthogonal Projection

Some additional material about orthogonality, to complement what is in Lang, Chapter V. As usual A^t denotes the transpose of the matrix A. A vector x is always a column vector, so x^t is a row vector.

7.1 Orthogonal Projection

Start with a real n-dimensional vector space V of dimension n with a positive definite inner product. Assume you have a subspace $U \subset V$ of dimension $m > 0$. By Theorem 2.3 of Chapter V, we know its orthogonal complement $W = U^{\perp}$ has dimension $r = n - m$. Thus we can write $V = U \oplus W$.

Pick an orthogonal basis $\{u_1, \dots, u_m\}$ of U, an orthogonal basis $\{w_1, \dots, w_r\}$ of W. The two together form an orthogonal basis of V. Write any $v \in V$ in terms of this basis:

$$v = c_1 u_1 + \dots + c_m u_m + d_1 w_1 + \dots + d_r w_r \tag{7.1.1}$$

for suitable real numbers c_i and d_j.

7.1.2 Proposition. *c_i is the* component *of v along u_i.*

We use the definition of component in Lang, page 99. Dotting (7.1.1) with u_i gives $(v - c_i u_i) \cdot u_i = 0$, which is exactly what is required. In the same way, the d_i is the component of v along w_i.

Now we apply Lang's Theorem 1.3 of Chapter V. First to v and $u_i, \dots, u_m$. The theorem tells us that the point $p = c_1 u_1 + \dots + c_m u_m$ is the point in U closest to v. Similarly, the point $q = d_1 w_1 + \dots + d_r w_r$ is the point in W closest to v.

7.1.3 Definition. The *orthogonal projection* of V to U is the linear transformation P from V to V that sends

$$v \mapsto p \in U \text{ such that } \langle v - p, u\rangle = 0 \text{ for all } u \in U.$$

Here we map
$$v \mapsto p = c_1u_1 + \cdots + c_mu_m.$$
as per (7.1.1), and we have shown in Proposition 7.1.2 that this linear map P is the orthogonal projection to U.

The kernel of P is W. The image of P is of course U. Obviously $P^2 = P$. So Exercise 10 of Lang, Chapter III, §4 applies. Notice that we have, similarly, a projection to W that we call Q. It sends

$$v \mapsto q = d_1w_1 + \cdots + d_rw_r.$$

We have the set up of Exercises 11-12 of Lang, Chapter III, §4, p. 71. Finally note that the matrix of P in our basis can be written in block form as

$$A_m = \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix}$$

where I_m is the $m \times m$ identity matrix, and the other matrices are all zero matrices. In particular it is a symmetric matrix.

Conversely we can establish:

7.1.4 Theorem. *Any square matrix P that*

- *is symmetric ($P^t = P$),*
- *and satisfies $P^2 = P$;*

is the matrix of the orthogonal projection to the image of P.

Proof. We establish Definition 7.1.3 just using the two properties. For all v and w in V:

$$\langle v - Pv, Pw\rangle = (v - Pv)^tPw = v^tPw - v^tP^tPw = v^tPw - v^tP^2w = 0.$$

In the next-to-the-last step we replaced P^tP by P^2 because P is symmetric, and then in the last step we used $P^2 = P$.

□

7.2 Solving the Inhomogeneous System

We work in K^n and K^m, equipped with the ordinary dot product. Our goal is to solve the inhomogeneous equation

$$Ax = b$$

where A be a $m \times n$ matrix, and b a m-column vector. We started studying this equation in §1.1.

7.2.1 Theorem. *The inhomogeneous equation $Ax = b$ can be solved if and only if*

$$\langle b, y \rangle = 0, \textit{ for every } m\textit{-vector } y \textit{ such that } y^t A = 0.$$

Proof. To say that the equation $Ax = b$ can be solved is simply to say that b is in the image of the linear map with matrix A. By the four subspaces theorem 6.6.1 the image of A in K^m is the orthogonal complement of the kernel of A^t. Thus for any b orthogonal to the kernel of A^t, the equation $Ax = b$ can be solved, and only for such b. This is precisely our assertion. $\square$

7.2.2 Example. Now a 3×3 example. We want to solve the system:

$$\begin{aligned} x_1 - x_2 &= b_1 \\ x_2 - x_3 &= b_2 \\ x_3 - x_1 &= b_3 \end{aligned}$$

So

$$A = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{pmatrix}.$$

Now A has rank 2, so up to a scalar, there is only one non-zero vector y such that $y^t A = 0$. To find y add the three equations. We get

$$0 = b_1 + b_2 + b_3.$$

This says that the scalar product of $(1, 1, 1)$ with b is 0. So by the theorem the system has a solution for all b such that $b_1 + b_2 + b_3 = 0$.

Let's work it out. Write $b_3 = -b_1 - b_2$. Then the third equation is a linear combination of the first two, so can be omitted. It is sufficient to solve the system:

$$\begin{aligned} x_1 - x_2 &= b_1 \\ x_2 - x_3 &= b_2 \end{aligned}$$

x_3 can be arbitrary, and then $x_2 = x_3 + b_2$ and

$$x_1 = x_2 + b_1 = x_3 + b_1 + b_2$$

so the system can be solved for any choice of x_3.

7.2.3 Remark. It is worth thinking about what happens when one does Gaussian elimination on the inhomogeneous system, using the augmented matrix 1.1.2. As in the proof of Proposition 5.1.3, reduce the matrix A to row echelon form C, getting an equivalent system

$$Cx = d$$

The matrix C may have rows of zeroes - necessarily the last rows. Assume it has p rows of zeroes. Then for the new system to have a solution, the last p components of d must be 0. The rank of C is $m - p$: just imitate the proof of Proposition 5.1.4. This can be at most n, since the row rank is the column rank. Then by the Rank-Nullity Theorem, any vector d whose last p components are 0 is in the image of C, and in that case the system has a solution. The matrix C has $m - p$ columns of index $\mu(i)$, $1 \le i \le m - p$, where $\mu(i)$ is a strictly increasing function of i, such that the entry $c_{i,\mu(i)}$ is the first non-zero coordinate of row C_i of C. The remaining columns correspond to the free variables x_i. Thus there are $n - (m - p)$ of them. For any choice of the free variables the system admits a unique solution in the remaining $(m - p)$ variables.

7.2.4 Example. Now we redo Example 7.2.2 via Gaussian elimination to illustrate the remark above. Here $n = m = 3$. The augmented matrix is

$$\begin{pmatrix} 1 & -1 & 0 & b_1 \\ 0 & 1 & -1 & b_2 \\ -1 & 0 & 1 & b_3 \end{pmatrix}.$$

We reduce A to row echelon form C:

$$\begin{pmatrix} 1 & -1 & 0 & b_1 \\ 0 & 1 & -1 & b_2 \\ 0 & 0 & 0 & b_1 + b_2 + b_3 \end{pmatrix}.$$

so $p = 1$. $\mu(1) = 1$, $\mu(2) = 2$, so x_3 is the only free variable. The only condition on b is that $b_1 + b_2 + b_3 = 0$.

7.3 Solving the Inconsistent Inhomogeneous System

We continue with the inhomogeneous system:

$$Ax = b$$

where A is a $m \times n$ matrix. Here, we assume that m is larger, perhaps much larger than n, and that the rank of A is n - the biggest it could be under the circumstances, meaning that its columns are linearly independent. In Theorem 7.3.3 below we show that this implies that the $n \times n$ matrix $A^t A$ is invertible. We also assume that we are working over $\mathbb{R}$, since we need results from §7.1.

Multiplication by the matrix A gives us an injective linear map from $\mathbb{R}^n$ to $\mathbb{R}^m = V$, and the image of A is a subspace U of dimension n of V by the Rank-Nullity theorem.

So typically, a right hand vector $b \in \mathbb{R}^m$ will not lie in the subspace U, so the equation $Ax = b$ will not be solvable, because b is not in the image of A. If this is the case, we say the system is *inconsistent*. Still, one would like to find the best possible approximate solution of the system. We know what to do: project b into U as in §7.1: why? because this is the best approximation by Lang, Theorem 1.3 p. 102.

The columns $A^1, \dots, A^n$ form a basis (not orthogonal) of the image U, so the projection point p of b can be written uniquely as a linear combination

$$p = x_1 A^1 + \cdots + x_n A^n$$

for real variables x_i. This is the matrix product $p = Ax$: check the boxed formula of Lang, page 113. To use the method of §7.1, we need the orthogonal complement of U, namely the subspace of V orthogonal to the columns of A. That is the kernel of A^t, since that is precisely what is orthogonal to all the columns.

So we conclude that $b - p$ must be in the kernel of A^t. Writing this out, we get the key condition:

$$A^t(b - Ax) = 0, \text{ or } A^t Ax = A^t b. \tag{7.3.1}$$

Because $A^t A$ is invertible - see Theorem 7.3.3 below, we can solve for the unknowns x:

$$x = (A^t A)^{-1} A^t b.$$

So finally we can find the projection point:

$$p = Ax = A(A^t A)^{-1} A^t b.$$

So we get from any b to its projection point p by the linear transformation with matrix

$$P = A(A^t A)^{-1} A^t$$

Since A is a $m \times n$ matrix , it is easy to see that P is $m \times m$. Notice that $P^2 = P$:

$$P^2 = A(A^t A)^{-1} A^t A(A^t A)^{-1} A^t = A(A^t A)^{-1} A^t = P$$

by cancellation of one of the $(A^tA)^{-1}$ by A^tA in the middle. Also notice that P is symmetric by computing its transpose:

$$P^t = (A^t)^t((A^tA)^{-1})^tA^t = A((A^tA)^t)^{-1}A^t = A(A^tA)^{-1}A^t = P.$$

We used $(A^t)^{-1} = (A^{-1})^t$ (see Lang's exercise 32 p. 41), and of course we used $(A^t)^t = A$. So we have shown (no surprise, since it is a projection matrix):

7.3.2 Theorem. *The matrix P above satisfies Theorem 7.1.4: it is symmetric and $P^2 = P$.*

We now prove a result we used above. We reprove a special case of this result in Theorem 11.2.1. Positive definite matrices are considered at great length in Chapter 11.

7.3.3 Theorem. *Let A be a an $m \times n$ matrix, with $m \geq n$. If A has maximal rank n, then the $n \times n$ matrix A^tA is positive definite, and therefore is invertible. It is also symmetric.*

Proof. Because A has maximal rank, its kernel is trivial, meaning that the only n-vector x such that $Ax = 0$ is the zero vector. So assume $Ax \neq 0$. Then

$$x^t(A^tA)x = (x^tA^t)Ax = (Ax)^tAx = \|Ax\|^2 \geq 0$$

Now $\|Ax\|^2 = 0$ implies that Ax is the zero vector, and this cannot be the case. Thus $x^t(A^tA)x > 0$ whenever $x \neq 0$. This is precisely the definition that A^tA is positive definite. It is symmetric because

$$(A^tA)^t = A^t(A^t)^t = A^tA.$$

□

7.3.4 Exercise. Compute A^tA for the rank 2 matrix

$$A = \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix}$$

and show that it is positive definite.

$$A^tA = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}$$

This is obviously positive definite. In this case it is easy to work out the projection matrix $A(A^tA)^{-1}A^t$:

$$P = \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1/3 & 0 \\ 0 & 1/2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 5/6 & 2/6 & -1/6 \\ 2/6 & 2/6 & 2/6 \\ -1/6 & 2/6 & 5/6 \end{pmatrix}$$

which is of course symmetric, and $P^2 = P$ as you should check.

In conclusion, given an inconsistent system $Ax = b$, the technique explained above shows how to replace it by the system $Ax = p$ that matches it most closely.

Orthogonal projections will be used when (if) we study the method of least squares.

7.3.5 Remark. In Lang's Theorem 1.3 of Chapter V mentioned above, a result minimizing distance is proved just using Pythagoras. In this section, we might want to do the same thing when faced with the expression

$$\|Ax - b\|$$

It is not so simple because the columns of A are not mutually perpendicular. So instead we can use the standard minimization technique from multivariable calculus. First, to have an easier function to deal with, we take the square, which we write as a matrix product:

$$f(x) = (x^tA^t - b^t)(Ax - b) = x^tA^tAx - x^tA^tb - b^tAx + b^tb$$

Notice that each term is a number: check the size of the matrices and the vectors involved. Calculus tells us $f(x)$, which is a quadratic polynomial, has an extremum (minimum or maximum or inflection point) only when all the partial derivatives with respect to the x_i vanish. It is an exercise to see that the gradient ∇f of f in x is $A^tAx - A^tb$, so setting this to 0 gives the key condition (7.3.1) back. No surprise.

Of course Theorem 7.1.4 shows that it is possible to bypass distance minimization entirely, and use perpendicularity instead.

Chapter 8

Uniqueness of Determinants

Using the material from Lang [3] on elementary matrices, summarized in Chapter 1, we give an alternative quick proof of the uniqueness of the determinant function. This is the approach of Artin [1], Chapter I, §3, and is different from the one using permutations that Lang uses in our textbook [4]. You may wonder how Lang covers this material in [3]: the answer is that he omits the proofs.

8.1 The Properties of Expansion by Minors

We start with a $n \times n$ matrix A. In [4], [VI, §2-3], Lang shows that any expansion $F(A)$ by minors along rows (columns) satisfies certain properties. Recall that the expansion along the i-th row can be written inductively as

$$F(A) = (-1)^{i+1}a_{i1}F(A_{i1}) + \cdots + (-1)^{i+j}a_{ij}F(A_{ij}) + \cdots + (-1)^{i+n}a_{in}F(A_{in})$$

where the F on the right hand side refers to the function for $(n-1) \times (n-1)$ matrices, and A_{ij} is the ij-th *minor* of A. Note that Lang does not use this word, which is universally used. We will have more to say about minors in Chapter 11.

The expansion along the j-th column can be written inductively as

$$F(A) = (-1)^{j+1}a_{1j}F(A_{1j}) + \cdots + (-1)^{i+j}a_{ij}F(A_{ij}) + \cdots + (-1)^{j+n}a_{nj}F(A_{nj}).$$

These expansions are widely known as the Laplace expansions.

This is a place where summation notation is useful. The two expressions above can be written

$$\sum_{i=1}^{n}(-1)^{i+j}a_{ij}F(A_{ij}) \text{ and } \sum_{j=1}^{n}(-1)^{i+j}a_{ij}F(A_{ij}).$$

Only the index changes!

Now we recall the three defining properties of the determinant, then the additional properties we deduce from them. They are numbered here as in Lang:

1. Multilinearity in the columns (rows);
2. Alternation: if two adjacent columns (rows) are the same, the function F is 0;
3. Identity: $F(I) = 1$;
4. Enhanced alternation: If any two columns (rows) are interchanged, the function F is multiplied by -1.
5. More alternation: If any two columns (rows) are equal, the function F is 0;
6. If a scalar multiple of a column (row) is added to a different column (row), the function F does not change;
7. If a matrix A has a column (row) of zeroes, the function $F(A) = 0$.

Here is how you read this list: if you expand along rows (as is done in [4], VI, §2-3) read column everywhere, while if you expand along columns read (row) everywhere.

(7) is not proved in [4]. Here is the easy proof, which we only do for row expansion, so for a column of zeroes. Suppose the j-th column of A is 0. Multiply that column by the non zero constant c. This does not change the matrix. By linearity (property (1)) we then have $F(A) = cF(A)$. The only way this can be true is if $F(A) = 0$.

8.2 The Determinant of an Elementary Matrix

In this section we will deal with expansion along columns, so that we can make an easy connection with the row operations of Gaussian elimination. First we compute the determinant of the three different types of elementary matrices 1.2.1.

Multilinearity (property (1)) says, in particular, that if E_c is an elementary matrix of type (1), for the constant c, then

$$F(E_c A) = cF(A). \tag{8.2.1}$$

since $E_c A$ just multiplies the corresponding row of A by c. Let $A = I$, the identity matrix. Property (3) says that $F(I) = 1$, so substituting I for A in (8.2.1) we get $F(E_c) = c$.

Next let E be an elementary matrix of type (2): interchanging rows. Then by Property (4),

$$F(EA) = -F(A). \tag{8.2.2}$$

Again take $A = I$ to get $F(E) = -1$.

Finally let E be an elementary matrix of type (3). Property (6) says

$$F(EA) = F(A). \tag{8.2.3}$$

Once again take $A = I$ to get, this time, $F(E) = 1$.

8.2.4 Remark. It is important to notice that the computation of $F(E)$ does not depend on whether F is a row or column expansion. We only used Property (3): $F(I) = 1$ that holds in both directions, as does the rest of the computation.

Next we get the easy but critical:

8.2.5 Theorem. *For any elementary matrix E and any square matrix A,*

$$F(EA) = F(E)F(A).$$

Furthermore if $E_1, \dots, E_k$ are elementary matrices, then

$$F(E_k \dots E_1 A) = F(E_k) \dots F(E_1)F(A).$$

Proof. Now that we have computed the value of $F(E)$ for any elementary matrix, we just plug it into the equations (8.2.1), (8.2.2) and (8.2.3) to get the desired result. For the last statement, just peel off one of the elementary matrices at a time, by induction, starting with the one on the left. $\square$

8.3 Uniqueness

In this section we follow Artin [1], p. 23-24 closely.

By Theorem 8.2.5, if (1.3.3) is satisfied, then

$$F(A') = F(E_k) \dots F(E_1)F(A).$$

Now A' is either I, in which case $F(A') = 1$, or the bottom row of A' is 0, in which case by Property (7), $F(A') = 0$.

In either case we can compute $F(A)$: if $F(A') = 1$, we get the known value

$$F(A) = \frac{1}{F(E_k) \dots F(E_1)}.$$

If $F(A') = 0$, we get 0. By remark 8.2.4 we get the same result whether we do row or column expansion. So we get a uniquely defined value for the function F, only using the three defining properties, valid for multilinearity and alternation with respect to either rows or columns. So we have:

8.3.1 Theorem. *This function F, which we now call the determinant, is the only function of the rows (columns) of $n \times n$ matrices to K satisfying Properties (1), (2) and (3).*

8.3.2 Corollary. *The determinant of A is $\neq 0$ if and only if A is invertible.*

This follows from Gaussian elimination by matrices: A is invertible if and only if in (1.3.3) the matrix $A' = I$, which is the only case in which the determinant of A is non-zero.

We now get one of the most important results concerning determinants:

8.3.3 Theorem. *For any two $n \times n$ matrices A and B,*

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

Proof. First we assume A is invertible. By (1.3.3) we know A is the product of elementary matrices:

$$A = E_k \dots E_1.$$

Note that these are not the same $E_1, \dots, E_k$ as before. By using Theorem 8.2.5 we get

$$\det(A) = \det(E_k) \dots \det(E_1) \tag{8.3.4}$$

and

$$\begin{aligned} \det(AB) &= \det(E_k \dots E_1 B) && \text{by definition of } A \\ &= \det(E_k) \dots \det(E_1)\det(B) && \text{by Theorem 8.2.5} \\ &= \det(A)\det(B). && \text{using (8.3.4)} \end{aligned}$$

Next assume that A is not invertible, so $\det(A) = 0$. We must show $\det(AB) = 0$. Now apply (1.3.3) to A' with bottom row equal to 0. Matrix multiplication shows that the bottom row of $A'B$ is 0, so by Property (7), $\det(A'B) = 0$. So using Theorem 8.2.5 again

$$\begin{aligned} 0 &= \det(A'B) && \text{as noted} \\ &= \det(E_k \dots E_1 AB) && \text{by definition of } A' \\ &= \det(E_k) \dots \det(E_1)\det(AB) && \text{Th. 8.2.5 applied to } AB \end{aligned}$$

Since the $\det(E_i)$ are non-zero, this forces $\det(AB) = 0$, and we are done. □

As a trivial corollary we get, when A is invertible, so $A^{-1}A = I$:

$$\det(A^{-1}) = \frac{1}{\det(A)}.$$

These last two results are proved in Lang [4] p.172: Theorems 7.3 and Corollary 7.4 of chapter VI.

8.3.5 Remark. Theorem 8.3.3 says that the map which associates to any $n \times n$ matrix its determinant has a special property. For those who know group theory this can be expressed as follows. Look at the set of all invertible $n \times n$ matrices, usually denoted $Gl(n)$ as we mentioned in §2.3. Now $Gl(n)$ is a group for matrix multiplication, with I as in neutral element. As we already know the determinant of an invertible matrix is non-zero, so when restricted to $Gl(n)$ the determinant function maps to K^* which denotes the non zero elements of the field K. Now K^* is a group for multiplication, as you should check. Theorem 8.3.3 says that the determinant function preserves multiplication, i.e.,

> The determinant of the product is the product of the determinants.

Maps that do this are called *group homomorphisms* and have wonderful properties that you can learn about in an abstract algebra book such as [1]. In particular it implies that the kernel of the determinant function, meaning the matrices that have determinant equal to the neutral element of K^*, namely 1, form a subgroup of $Gl(n)$, called $Sl(n)$. Lang studies $Sl(n)$ in Appendix II.

8.4 The Determinant of the Transpose

It is now trivial to prove

8.4.1 Theorem. *For any $n \times n$ matrix A,*

$$\det(A^t) = \det(A)$$

Proof. Indeed computing the expansion by minors of A^t along columns is the same as computing that of A along rows, so the computation will give us $\det(A)$ both ways. The key point is again Remark 8.2.4.

□

This is Lang's Theorem 7.5 p.172. He proves it using the expansion formula for the determinant in terms of permutations: see (9.1.3).

Chapter 9

The Key Lemma on Determinants

The point of this chapter is to clarify Lang's key lemma 7.1 and the material that precedes it in §7 of [4]. A good reference for this material is Hoffman and Kunze [2], §5.3.

These notes can be read before reading about permutations in Lang's Chapter VI, §6. We do assume a knowledge of Lang [4], §2-3 of Chapter VI: the notation from there will be used.

9.1 The Computation

As in Chapter 8, we use $F(A)$ to denote a function from $n \times n$ matrices to the base field K that satisfies Properties (1) and (2) with respect to the columns: thus, as a function of the columns it is multilinear and alternating.

Let $e^1, \dots, e^n$ denote the columns (in order) of the identity matrix. Then we can, somewhat foolishly, perhaps, write:

$$A^j = a_{1j}e^1 + a_{2j}e^2 + \cdots + a_{nj}e^n. \tag{9.1.1}$$

A^j denotes, as usual, the j-th column of A.

So using multilinearity of F in A^1, we get

$$F(A^1, A^2, \dots, A^n) = \sum_{i=1}^{n} a_{i1} F(e^i, A^2, \dots, A^n).$$

Now repeat with respect to A^2. We get a second sum:

$$F(A^1, A^2, \dots, A^n) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{i1} a_{j2} F(e^i, e^j, A^3, \dots, A^n).$$

Now a third time with respect to A^3:

$$F(A) = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} a_{i1}a_{j2}a_{k3}F(e^i, e^j, e^k, A^4, \dots, A^n).$$

Now we have to solve an issue of notation: we call the indices k_1, k_2, ..., k_n, and we write only one summation sign. Once we have expanded all the columns, we get:

$$F(A) = \sum_{k_1,\dots,k_n=1}^{n} a_{k_1 1}a_{k_2 2}\dots a_{k_n n}F(e^{k_1}, e^{k_2}, \dots, e^{k_n}). \tag{9.1.2}$$

What does the summation mean? We need to take all possible choices of $k_1, k_2, \dots, k_n$ between 1 and n. It looks like we have n^n terms.

Now we use the hypothesis that F is alternating. Then

$$F(e^{k_1}, e^{k_2}, e^{k_3}, \dots, e^{k_n}) = 0$$

unless all of the indices are distinct: otherwise the matrix has two equal columns. This is how **permutations** enter the picture. When $k_1, k_2, \dots, k_n$ are distinct, then the mapping

$$1 \to k_1, 2 \to k_2, \dots, n \to k_n$$

is a permutation of $J_n = \{1, \dots, n\}$. Our sum is over all permutations.

Now at last we assume Property (3), so that $F(I) = 1$. Then

$$F(e^{k_1}, e^{k_2}, e^{k_3}, \dots, e^{k_n}) = \pm 1,$$

because we have the columns of the identity matrix, in a different order. At each interchange of two columns, the sign of F changes. So we get 1 if we need an even number of interchanges, and -1 if we need an odd number of interchanges.

The number $F(e^{k_1}, e^{k_2}, e^{k_3}, \dots, e^{k_n})$ is what is called the **sign** of the permutation. Lang writes it ϵ.

Thus if we write the permutations σ, we have proved

$$\boxed{F(A) = \sum_{\sigma} \epsilon(\sigma)a_{\sigma(1),1}a_{\sigma(2),2}\dots a_{\sigma(n),n}} \tag{9.1.3}$$

where the sum is over all permutations in J_n. This is a special case of the formula in the key Lemma 7.1, and gives us Theorem 7.2 directly.

9.2 Matrix Multiplication

The next step is to see how matrix multiplication allows us to produce the generalization of (9.1.1) Lang uses to get his Theorem 7.3.

First work out the 2×2 case. If we assume $C = AB$, it is easy to see the following relationship between the columns of C and those of A, with coefficients the entries of B.

$$\begin{aligned} \begin{pmatrix} C^1 & C^2 \end{pmatrix} &= \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{pmatrix} \\ &= \begin{pmatrix} b_{11}A^1 + b_{21}A^2 & b_{12}A^1 + b_{22}A^2 \end{pmatrix} \end{aligned}$$

This generalizes to all n: if $C = AB$ for $n \times n$ matrices, then for every i,

$$C^j = b_{1j}A^1 + b_{2j}A^2 + \cdots + b_{nj}A^n.$$

Now we compute $F(C)$ by expanding the columns using these equations and multilinearity, exactly as we did in the previous section. So C plays the role of A, the b_{ij} replace the a_{ij}, and the A^i replace the e^i.

Now we assume F is also alternating. We will not make use of Property (3) until we mention it. By the same argument as before we get the analog of (9.1.2):

$$F(C) = \sum_{k_1,\dots,k_n=1}^{n} b_{k_1 1} b_{k_2 2} \dots b_{k_n n} F(A^{k_1}, A^{k_2}, \dots, A^{k_n}).$$

Finally, using permutations as before, we can rewrite this as

$$\boxed{F(C) = \sum_{\sigma} \epsilon(\sigma) b_{\sigma(1),1} b_{\sigma(2),2} \dots b_{\sigma(n),n} F(A)}$$

which is the precise analog of the expression in Lang's Lemma 7.1. Using (9.1.3) for B, we get:

$$F(AB) = F(A)\det(B).$$

Finally if we assume that F satisfies Property (3), by the result of the previous section we know that F is the determinant, so we get the major theorem:

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

Note in passing that this solves Lang's Exercise 2 of §7.

Chapter 10

The Companion Matrix

This important matrix is not discussed in [4], but it provides a good introduction to the notion of a cyclic vector, which becomes important in [4], XI, §4.

10.1 Introduction

We start with a monic polynomial of degree n:

$$f(t) = t^n - a_{n-1}t^{n-1} - a_{n-2}t^{n-2} - \cdots - a_0 \tag{10.1.1}$$

with coefficients in an field K. Monic just means that the coefficient of the leading term t^n is 1.

We associates to this matrix an $n \times n$ matrix known as its *companion matrix*:

$$A = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 & a_0 \\ 1 & 0 & 0 & \dots & 0 & 0 & a_1 \\ 0 & 1 & 0 & \dots & 0 & 0 & a_2 \\ 0 & 0 & 1 & \dots & 0 & 0 & a_3 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 & a_{n-2} \\ 0 & 0 & 0 & \dots & 0 & 1 & a_{n-1} \end{pmatrix} \tag{10.1.2}$$

Thus the last column of A has the coefficients of f in increasing order, omitting the coefficient 1 of the leading term, while the *subdiagonal* of A, namely the terms $a_{i+1,i}$ are all equal to 1. All other terms are 0.

10.1.3 Theorem. *The characteristic polynomial of the companion matrix of a polynomial $f(t)$ is $f(t)$.*

Proof. We compute the characteristic polynomial of A, i.e. the determinant of

$$\begin{pmatrix} t & 0 & 0 & \dots & 0 & -a_0 \\ -1 & t & \dots\dots\dots & & 0 & -a_1 \\ 0 & -1 & t & \dots & 0 & -a_2 \\ 0 & 0 & -1 & \dots & 0 & -a_3 \\ \vdots & \vdots & \ddots & \ddots & t & \vdots \\ 0 & 0 & 0 & \dots & -1 & t-a_{n-1} \end{pmatrix}$$

by expansion along the first row, and induction on n. First we do the case $n = 2$. The determinant we need is

$$\begin{vmatrix} t & -a_0 \\ -1 & t-a_1 \end{vmatrix} = t(t-a_1) - a_0 = t^2 - a_1 t - a_0,$$

as required.

Now we do the case n. By Laplace expansion of the determinant along the first row we get two terms:

$$t \begin{vmatrix} t & \dots\dots\dots & 0 & -a_1 \\ -1 & t & \dots & 0 & -a_2 \\ 0 & -1 & \dots & 0 & -a_3 \\ \vdots & \ddots & \ddots & t & \vdots \\ 0 & 0 & \dots & -1 & t-a_{n-1} \end{vmatrix} + a_0(-1)^n \begin{vmatrix} -1 & t & \dots\dots\dots & & 0 \\ 0 & -1 & t & \dots & 0 \\ 0 & 0 & -1 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & t \\ 0 & 0 & 0 & \dots & -1 \end{vmatrix}$$

By induction we see that the first term gives

$$t(t^{n-1} - a_{n-1}t^{n-2} - -a_{n-2}t^{n-3} - \dots - a_2 t - a_1)$$

while the second term gives $a_0(-1)^n(-1)^{n-1} = -a_0$, since the matrix is triangular with -1 along the diagonal.

Thus we do get the polynomial (10.1.1) as characteristic polynomial of (10.1.2). □

10.2 Use of the Companion Matrix

We denote by $e^1, e^2, \dots, e^n$ the standard unit basis of K^n. The form of (10.1.2) shows that the linear transformation L associated to the matrix A acts as follows:

$$L(e^1) = e^2, L(e^2) = e^3, \dots, L(e^n) = a_0 e^1 + a_1 e^2 + \dots + a_{n-1} e^n,$$

since e^2 is the first column of A, e^3 the second column of A, e^n the $n-1$-th column of A, and $a_0e^1 + a_1e^2 + \cdots + a_{n-1}e^n$ the last column of A.

This means that e^1 is a *cyclic vector* for L: as we apply powers of the operator L to e_1, we generate a basis of the vector space K^n: in other words the vectors e^1, $L(e^1)$, $L^2(e^1)$, $L^{n-1}(e^1)$ are linearly independent.

Conversely,

10.2.1 Theorem. *Assume that the operator $L\colon V \to V$ has no invariant subspaces, meaning that there is no proper subspace $W \subset V$ such that $L(W) \subset W$. Pick any non-zero vector $v \in V$. Then the vectors v, $L(v)$, $L^2(v)$, …, $L^{n-1}(v)$ are linearly independent, and therefore form a basis of V.*

Proof. To establish linear independence we show that there is no equation of linear dependence between v, $L(v)$, $L^2(v)$, …, $L^{n-1}(v)$. By contradiction, let k be the smallest positive integer such that there is an equation of dependence

$$b_0v + b_1L(v) + b_2L^2(v) + \cdots + b_kL^k(v) = 0$$

The minimality of k means that $b_k \neq 0$, so that we can solve for $L^k(v)$ in terms of the previous basis elements. But this gives us an invariant subspace for L: the one generated by v, $L(v)$, $L^2(v)$, …, $L^k(v)$, a contradiction. □

The matrix of L in the basis $\{v, L(v), L^2(v), \dots, L^{n-1}(v)\}$ is the companion matrix of the characteristic polynomial of L.

10.3 When the Base Field is the Complex Numbers

If $K = \mathbb{C}$, then one can easily find an even simpler $n \times n$ matrix D over $\mathbb{C}$ whose characteristic polynomial is $f(t)$. In that case $f(t)$ factors linearly with n roots α_1, …, α_n, not necessarily distinct. Then take the diagonal matrix D with diagonal entries the α_i. The characteristic polynomial of D is clearly $f(t)$.

The construction we will make for the Jordan normal form ([4], Chapter XI, §6) is a variant of the construction if the companion polynomial. Here we only treat the case where the characteristic polynomial of L is $(t-\alpha)^n$. So it has only one root with high multiplicity. Assume furthermore that there is a vector v that is cyclic, not for L, but for $L - \alpha I$, meaning that there is a k such that $(L-\alpha I)^k v = 0$. The smallest k for which this is true is called the *period* of v. The case that parallels what we did in the previous section is the case where $k = n$, the dimension of V.

An argument similar to the one above (given in [4], Lemma 6.1 p. 262) shows that the elements

$$w_0 = v, w_1 = (L-\alpha)v, w_2 = (L-\alpha)^2v, \dots, w_{n-1} = (L-\alpha)^{n-1}v$$

are linearly independent. If we use these as a basis, in this order, we see that

$$\begin{aligned}
Lw_0 &= Lv = (L-\alpha)v + \alpha v = w_1 + \alpha w_0 \\
Lw_1 &= L(L-\alpha)v = (L-\alpha)^2 v + \alpha(L-\alpha)v = w_2 + \alpha w_1 \\
Lw_k &= L(L-\alpha)^k v = (L-\alpha)^{k+1} v + \alpha(L-\alpha)^k v = w_{k+1} + \alpha w_k \\
Lw_{n-1} &= L(L-\alpha)^{n-1} v = (L-\alpha)^n v + \alpha(L-\alpha)^{n-1} v = \alpha w_{n-1}
\end{aligned}$$

Thus the matrix of L in this basis is

$$A = \begin{pmatrix} \alpha & 0 & 0 & \dots & 0 & 0 \\ 1 & \alpha & 0 & \dots & 0 & 0 \\ 0 & 1 & \alpha & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \alpha & 0 \\ 0 & 0 & 0 & \dots & 1 & \alpha \end{pmatrix} \tag{10.3.1}$$

Thus it is lower triangular, with α along the diagonal, and 1 on the subdiagonal. Because Lang takes the reverse order for the elements of the basis, he gets the transpose of this matrix: see p. 263. This confirms that the characteristic polynomial of this matrix is $(t-\alpha)^n$.

Finally we ask for the eigenvectors of (10.3.1). It is an exercise ([4], XI, §6, exercise 2) to show that there is only one eigenvector: w_{n-1}.

10.4 When the Base Field is the Real Numbers

When $K = \mathbb{R}$ the situation is slightly more complicated, since the irreducible factorization of $f(t)$ over $\mathbb{R}$ contains polynomials of degree 1 and degree 2. Let us look at the 2×2 case: the real matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{10.4.1}$$

with characteristic polynomial

$$g(t) = t^2 - (a+d)t + (ad - bc).$$

The trace of this matrix is $a+d$ and its determinant $ad - bc$. Notice how they appear in the characteristic polynomial. For this to be irreducible over $\mathbb{R}$, by the quadratic formula we must have

$$(a+d)^2 - 4(ad - bc) = (a-d)^2 + 4bc < 0$$

or $(a-d)^2 < 4bc$.

The companion matrix of this polynomial is

$$\begin{pmatrix} 0 & -(ad-bc) \\ 1 & a+d \end{pmatrix}$$

The full value of the companion matrix only reveals itself when one takes smaller subfields of $\mathbb{C}$, for example the field of rational numbers $\mathbb{Q}$. Over such a field there are irreducible polynomials of arbitrary high degree: for example the *cyclotomic polynomial*

$$\Phi_p(t) = t^{p-1} + t^{p-2} + \cdots + t + 1$$

for p a prime number. Since $(t-1)\Phi_p(t) = t^p - 1$, the roots of $\Phi(t)$ are complex numbers on the circle of radius 1, thus certainly not rational. It is a bit harder to show that $\Phi_p(t)$ is irreducible, but it only requires the elementary theory of polynomials in one variable. A good reference is Steven H. Weintraub's paper Several Proofs of the Irreducibility of the Cyclotomic Polynomial.

Chapter 11

Positive Definite Matrices

Real symmetric matrices come up in all areas of mathematics and applied mathematics. It is useful to develop easy tests to determine when a real symmetric matrix is positive definite or positive semidefinite.

The material in Lang [4] on positive (semi)definite matrices is scattered in exercises in Chapter VII, §1 and in Chapter VIII, §4. The goal of this chapter is to bring it all together, and to establish a few other important results in this direction.

11.1 Introduction

Throughout this chapter V with denote a real vector space of dimension n with a positive definite scalar product written $\langle v, w\rangle$ as usual. By the Gram-Schmidt orthonormalization process (Lang V.2) this guarantees that there exists an orthonormal basis $\mathcal{B} = \{v_1, v_2, \dots, v_n\}$ for V.

Recall that a *symmetric operator* T on V satisfies

$$\langle T(v), w\rangle = \langle Tv, w\rangle$$

for all v and w in V. A symmetric operator T is *positive semidefinite*

$$\langle Tv, v\rangle \geq 0 \text{ for all } v \in V. \tag{11.1.1}$$

This is the definition in Exercise 4 of VIII §4. Similarly a symmetric operator T is *positive definite* if (11.1.1) is replaced by

$$\langle Tv, v\rangle > 0 \text{ for all } v \neq O. \tag{11.1.2}$$

In particular, positive definite operators are a subset of positive semidefinite operators.

11.1.3 Remark. Lang uses the term *semipositive* instead of positive semidefinite: Exercise 10 of VII §1and Exercise 4 of VIII §4. Here we will use positive semidefinite, since that is the term commonly used.

Lang also defines a positive definite matrix A on $\mathbb{R}^n$ with the standard dot product. It is a real $n \times n$ symmetric matrix A such that for all coordinate vectors $X \neq O$, $X^t AX > 0$. See Exercise 9 of VII §1. In the same way we can define a positive semidefinite matrix.

Here is an important fact that is only stated explicitly by Lang in Exercise 15 b) of VIII §4:

11.1.4 Proposition. *If T is a positive definite operator on V, then for an orthonormal basis $\mathcal{B}$ of V the matrix $A = M_{\mathcal{B}}^{\mathcal{B}}(T)$ representing it in the basis $\mathcal{B}$ is positive definite, in particular it is symmetric. Thus if $v = x_1 v_1 + \cdots + x_n v_n$ and $X = (x_1, \dots, x_n)$ is the coordinate vector of v in $\mathbf{R}^n$ of v,*

$$X^t AX > 0 \text{ for all } X \neq O.$$

Similarly for positive semidefinite operators.

This follows immediately from the results of Chapter IV, §3 and the definitions.

Next it is useful to see how the matrix A changes as one varies the basis of V. Following the notation of Lang V.4, we write $g(v, w) = \langle Tv, w\rangle$. This is a symmetric bilinear form: see Lang, V.4, Exercises 1 and 2. We have just seen that for an orthonormal basis $\mathcal{B}$ of V,

$$g(v, w) = X^t AY$$

where X and Y are the coordinates of v and w in that basis. Now consider another basis $\mathcal{B}'$. By Lang's Corollary 3.2 of Chapter IV, if U is the invertible change of basis matrix $M_{\mathcal{B}'}^{\mathcal{B}}(id)$ and X' and Y' are the coordinate vectors of x and y in the $\mathcal{B}'$ basis, then

$$X' = UX \text{ and } Y' = UY.$$

Thus if A' denotes the matrix for $g(v, w)$ in the $\mathcal{B}'$ basis, we must have

$$g(v, w) = X^t AY = X'^t A'Y' = X^t U^t AUY.$$

Since this must be true for all v and w, we get

$$\boxed{A' = U^t AU} \tag{11.1.5}$$

Therefore the matrix A' is symmetric: just compute its transpose.

11.1.6 Exercise. Two $n \times n$ symmetric matrices A and A' are *congruent* if there is an invertible matrix U such that $A' = U^t A U$. Show that congruence is an equivalence relation on symmetric matrices: see §4.3.

11.1.7 Exercise. Show that two congruent matrices represent the same symmetric bilinear form on V in different bases. Thus they have the same index of positivity and same index of nullity. In particular, if one is positive definite or positive semidefinite, the other is also. See Lang V.8.

11.2 The First Three Criteria

From Lang's account in Chapter VIII, §4, one can piece together the following theorem.

11.2.1 Theorem. *The following conditions on the symmetric operator T are equivalent.*

1. *T is positive (semi)definite.*
2. *The eigenvalues of T are all positive (non negative).*
3. *There exists a operator S on V such that $T = S^t S$, which is invertible if and only if T is positive definite.*
4. *The index of negativity of T is 0. If T is positive definite, the index of nullity is also 0*

These equivalences are contained in the exercises of Chapter 8, §4, in particular Exercises 3, 6, 8, 15 and 25. The proof relies on the Spectral Theorem, which allows us to find an orthonormal basis of eigenvectors for any symmetric operator. We only consider the positive definite case, and leave the semidefinite case as an exercise.

Proof. For $(1) \Rightarrow (2)$, if $Tv = \lambda v$, so that v is an eigenvector with eigenvalue λ, then

$$\langle v, Tv \rangle = \langle v, \lambda v \rangle = \lambda \langle v, v \rangle.$$

This must be positive. Since $\langle v, v \rangle > 0$, we must have $\lambda > 0$. Thus all the eigenvalues are positive. Conversely, if all the eigenvalues are positive, write an arbitrary element $v \in V$ as a linear combination of the orthonormal eigenvectors v_i, which form a basis of V:

$$v = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n$$

for real numbers c_i. Then

$$\begin{aligned}\langle v, Tv\rangle &= c_1^2\langle v_1, v_1\rangle + c_2^2\langle v_2, v_2\rangle + \cdots + c_n^2\langle v_1, v_1\rangle \\ &= \lambda_1 c_1^2 + \lambda_2 c_2^2 + \cdots + \lambda_n c_n^2\end{aligned}$$

which is positive unless all the coefficients c_i are 0.

For $(1) \Rightarrow (3)$, again let $\{v_1, v_2, \dots, v_n\}$ be an orthonormal basis of eigenvectors for T. By the first equivalence T is positive definite if and only if the λ_i are all positive. In which case take for S the operator that maps $v_i \mapsto \sqrt{\lambda_i} v_i$. S is invertible if and only if all the λ_i are positive. It is easy to see that S is symmetric.

If $v = c_1 v_1 + \cdots + c_n v_n$ and $w = d_1 v_1 + \cdots + d_n v_n$, then

$$\begin{aligned}\langle Sv, w\rangle &= \sqrt{\lambda_1} c_1 d_1 \langle v_1, v_1\rangle + \cdots + \sqrt{\lambda_n} c_n d_n \langle v_n, v_n\rangle \\ &= \langle v, Sw\rangle\end{aligned}$$

S is called the square root of T. Note that a positive semidefinite operator also has a square root, but it is invertible only if T is positive definite.

$(3) \Rightarrow (1)$ is a special case of Theorem 7.3.3. We reprove it in this context in the language of operators. We take any operator S on V. $S^t S$ is symmetric because

$$\langle S^t S v, w\rangle = \langle Sv, Sw\rangle = \langle v, S^t S w\rangle.$$

It is positive semidefinite because, by the same computation,

$$\langle S^t S v, v\rangle = \langle Sv, Sv\rangle \geq 0$$

for all $v \in V$. If S is invertible, then $Sv \neq O$ when $v \neq O$. Then $\langle Sv, Sv\rangle > 0$ because the scalar product itself is positive definite.

$(1) \Leftrightarrow (4)$ follows easily from the definition of the index of positivity: see Lang, Chapter V, §8, p. 138. More generally, here is the definition of the more commonly used *inertia* of A, which is not defined in Lang: it is the triple of non-negative integers (n_+, n_-, n_0), where n_+ is the index of positivity, n_0 the index of nullity (defined p.137 of Lang) and $n_- = n - n_1 - n_3$ what could be called by analogy the index of negativity: thus if $\{v_1, \dots, v_n\}$ is any orthogonal basis of V, then

1. n_+ is the number of basis elements v_i such that $v_i^t A v_i > 0$,
2. n_- is the number of basis elements v_i such that $v_i^t A v_i < 0$,
3. n_0 is the number of basis elements v_i such that $v_i^t A v_i = 0$,

The fact that these numbers do not depend on the choice of orthogonal basis is the content of Sylvester's Theorem 8.1 of Chapter V. Thus, essentially by definition:

- A is positive definite if and only if its inertia is $(n, 0, 0)$. In Lang's terminology, A is positive definite if and only its index of positivity is n.

- A is positive semidefinite if and only if its inertia is $(n_+, 0, n_0)$.

This concludes the proof. □

As a corollary of this theorem, we get

11.2.2 Corollary. *If T is positive definite, then it is invertible.*

Proof. Indeed if v is a non-zero vector in the kernel of T, then $Tv = O$, so $\langle T(v), v\rangle = 0$, a contradiction. □

There are two other useful criteria. The purpose of the rest of this chapter is to explain them. They are not mentioned in Lang, but following easily from earlier results.

11.3 The Terms of the Characteristic Polynomial

In this section we establish a result for all square matrices that is interesting in its own right. We will use it in the next tests for positive (semi)definiteness.

Write the characteristic polynomial of any $n \times n$ matrix A over a field K as

$$P(t) = t^n - p_1 t^{n-1} + p_2 t^{n-2} - \dots + (-1)^n p_n, \tag{11.3.1}$$

where $p_i \in K$ for all i, $1 \le i \le n$.

Before stating the main result of this section, we need some definitions.

11.3.2 Definition. Let $J = \{i_1, \dots, i_k\}$ be a set of k distinct integers in the interval $[1, n]$, listed in increasing order. If A is a square matrix, let A_J or $A(i_1, \dots, i_k)$ be the square submatrix of A formed by the rows and the columns of A with indices J. Let D_J be its determinant. Then A_J is called a *principal submatrix* of A, and its determinant D_J or $D(i_1, \dots, i_k)$ a *principal minor* of A.

For each k between 1 and n, the matrices $A(1, 2, \dots, k)$ are called the *leading* principal submatrices, and their determinants the *leading* principal minors of A.

11.3.3 Remark. Lang looks at certain minors in VI.2, while studying the Laplace expansion along a row or column. See for example the discussion on p.148. He does not use the word minor, but his notation A_{ij} denotes the submatrix of size $(n-1) \times (n-1)$ where the i-th row i and the j-th column have been removed. Then the $\det A_{ij}$ are minors, and the $\det A_{ii}$ are principal minors.

11.3.4 Example. So if A is the matrix

$$\begin{pmatrix} 2 & 0 & 0 & 1 \\ 0 & 4 & 3 & 0 \\ 0 & 3 & 4 & 0 \\ 1 & 0 & 0 & 2 \end{pmatrix}$$

then

$$A(1,2) = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}, A(2,3) = \begin{pmatrix} 4 & 3 \\ 3 & 4 \end{pmatrix}, \text{and } A(1,2,3) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 3 \\ 0 & 3 & 4 \end{pmatrix}$$

Obviously, if A is symmetric as in this example, A_J is symmetric. For each k, there is only only leading principal matrix, but $\binom{n}{k}$ principal minors of order k.

For the matrix A above, we have already computed $A(1,2)$ and $A(1,2,3)$; $A(1)$ is the 1×1 matrix (2) and $A(1,2,3,4)$ is A. So the determinant $D(1) = 2$, $D(1,2) = 8$, and $D(1,2,3) = 2(16-9) = 14$. So far, so good. We compute the determinant of A itself by expansion along the first row. We get $2(16-9)2-7 > 0$. As we will see in the next test, these simple computations are enough to show that$A$ is positive definite. In Lang's notation our $A(1,2,3)$ would be A_{44}.

Then the main theorem is:

11.3.5 Theorem. *For each index j, $1 \leq j \leq n$ we have*

$$p_j = \sum_J \det A_J = \sum_J D_J$$

where the sum is over all choices of j elements $J = \{i_1, \dots, i_j\}$ from the set of the first n integers, and A_J is the corresponding principal submatrix of A. Thus the sum has $\binom{n}{j}$ terms.

11.3.6 Example. We know two cases of the theorem already.

- If $j = n$, then there is only one choice for J: all the integers between 1 and n. The theorem then says that $p_n = \det A$, which is clear: just evaluate at $t = 0$.

- If $j = 1$, we know that p_1 is the trace of the matrix. The sets J have just one element, so the theorem says

  $$p_1 = a_{11} + a_{22} + \dots a_{nn}$$

  which is indeed the trace.

Proof. To do the general case, we use the expansion of the determinant in terms of permutations in Theorem 9.1.3. We consider the entries a_{ij} of the matrix A as variables.

To pass from the determinant to the characteristic equation, we make the following substitutions: we replace each off-diagonal term a_{ij} by $-a_{ij}$ and each diagonal terms a_{ii} by $t - a_{ii}$, where t is a new variable. How do we write this substitution in the term

$$\epsilon(\sigma)a_{\sigma(1),1}a_{\sigma(2),2}\dots a_{\sigma(n),n}$$

of the determinant expansion? Let f be the number of integers i fixed by σ, meaning that $\sigma(i) = i$, and let the fixed integers be $\{j_1, \dots, j_f\}$ Designate the remaining $n - f$ integers in $[1, n]$ by $\{i_1, \dots, i_{n-f}\}$. Then the term associated to σ in the characteristic polynomial can be written

$$\epsilon(\sigma)\prod_{l=1}^{f}(t - a_{j_l,j_l})\prod_{l=1}^{n-f}(-a_{\sigma(i_l),i_l}). \tag{11.3.7}$$

We fix an integer k in order to study the coefficient p_k of t^{n-k} in the characteristic polynomial. How can a term of degree exactly $n - k$ in t be produced by the term (11.3.7) corresponding to σ?

First, the permutation σ must fix at least $n - k$ integers. So $f \geq n - k$. Then in (11.3.7) we pick the term t in $n - k$ factors, and the other term in the remaining $f - n + k$ factors involving t. So there are exactly $\binom{f}{n-k}$ ways of doing this. If we let I be a choice of $n - k$ indices fixed by σ, we get for the corresponding term of (11.3.7)

$$(-1)^k\epsilon(\sigma)t^{n-k}\prod_{l\notin I}a_{\sigma(l),l}$$

Remove from this expression the factors $(-1)^k$ and t^{n-k} which do not depend of σ. We call what is left

$$M(\sigma, I) = \epsilon(\sigma)\prod_{l\notin I}a_{\sigma(l),l}$$

if σ fixes I, and we set $M(\sigma, I) = 0$ otherwise. Now we fix the set of indices I. A permutation σ that fixes I gives a permutation τ on the complement $J = \{j_1, j_2, \dots, j_k\}$ of I in $[1, n]$. Notice that $\epsilon(\sigma) = \epsilon(\tau)$. Conversely any permutation τ of J lifts uniquely to a permutation σ of $[1, n]$ by requiring σ to fix all the integers not in J. This shows that

$$\sum_{\sigma}M(\sigma, I) = \sum_{\tau}\epsilon(\tau)a_{\tau(j_1),j_1}a_{\tau(j_2),j_2}\dots a_{\tau(j_k),j_k}$$

where the last sum is over all the permutations of J. This is just D_J. We have accounted for all the terms in the characteristic polynomial, so we are done. $\square$

11.4 The Principal Minors Test

The main theorem is

11.4.1 Theorem. *A symmetric matrix A is positive definite if and only if all its leading principal minors are positive. It is positive semidefinite if and only if all its principal minors are non-negative.*

Notice the subtle difference between the two cases: to establish that A is positive semidefinite, you need to check all the principal minors, not just the leading ones.

11.4.2 Example. Consider the matrix

$$\begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}$$

The leading principal minors of this matrix are both 0, and yet it obviously not positive semidefinite, since its eigenvalues (0 and -1) are not both non-negative.

To prove the theorem some more definitions are needed.

11.4.3 Definition. Given $J = \{i_1, \dots, i_k\}$ as in Definition 11.3.2, for any n-vector $X = (x_1, \dots, x_n)$ let X_J be the k-vector $(x_{i_1}, x_{i_2} \dots, x_{i_k})$. Let $\tilde{X}_J$ be the n-vector whose i-th entry $\tilde{x}_i$ is given by

$$\tilde{x}_i = \begin{cases} x_i, & \text{if } i \in J; \\ 0, & \text{otherwise.} \end{cases}$$

Then it is clear that

$$\boxed{\tilde{X}_J^t A \tilde{X}_J = X_J^t A_J X_J.} \tag{11.4.4}$$

11.4.5 Example. If $n = 4$, $k = 2$ and $J = \{2, 4\}$, then

$$X_J = \begin{pmatrix} x_2 \\ x_4 \end{pmatrix}, \text{ and } \tilde{X}_J = \begin{pmatrix} 0 \\ x_2 \\ 0 \\ x_4 \end{pmatrix}.$$

Since

$$A_J = \begin{pmatrix} a_{22} & a_{24} \\ a_{42} & a_{44} \end{pmatrix}$$

you can easily verify (11.4.4) in this case.

This allows us to prove:

11.4.6 Proposition. *If A is positive definite, the symmetric matrix A_J is positive definite. If A is positive samidefinite, then A_J is positive semidefinite.*

Proof. If A is positive definite, the left hand side of (11.4.4) is positive if $\tilde{X}_J \neq O$. So the right hand side is positive when $X_J \neq O$, since X_J is just $\tilde{X}_J$ with $n-k$ zero entries removed. That is the definition of positive definiteness for A_J, so we are done. The positive semidefinite case is even easier. $\square$

If A_J is positive definite, its determinant, which is the principal minor D_J of A, is positive: indeed the determinant is the product of the eigenvalues, which are all positive. This shows that all the principal minors are positive, and finishes the easy implication in the proof of the main theorem. A similar argument handles the positive semidefinite case.

Before proving the more difficult implication of the theorem, we look at some examples.

11.4.7 Example. When the set J has just one element, so $k = 1$, we are looking at 1×1 principal minors. So we get: $D(i) = a_{ii} > 0$. However there are symmetric matrices with positive diagonal entries that are not positive definite. The matrix

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

is not positive definite: test the vector $(1, 1)$:

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

When J has two elements, so $J = \{i, j\}$, we get:

11.4.8 Corollary. *Let A be a positive definite matrix. Then for $i \neq j$,*

$$|a_{ij}| \leq \sqrt{a_{ii}a_{jj}}.$$

In Example 11.4.7, $|a_{12}| = 2$, while $a_{11}a_{22} = 1$, so the matrix is not positive definite.

11.4.9 Example. The matrix

$$A = \begin{pmatrix} 2 & 0 & 0 & 2 \\ 0 & 4 & 3 & 0 \\ 0 & 3 & 4 & 0 \\ 2 & 0 & 0 & 2 \end{pmatrix}$$

is not positive definite, by applying the lemma to $i = 1$, $j = 4$. It is positive semidefinite, however.

A weaker result implied by this corollary is useful when just scanning the matrix.

11.4.10 Corollary. *If A is positive definite, the term of largest absolute value must be on the diagonal.*

Now we return to the proof of the main theorem. In the positive definite case it remains to show that if all the leading principal minors of the matrix A are positive, then A is positive definite. Here is the strategy. From Exercise 11.1.7, if U be any invertible $n \times n$ matrix, then the symmetric matrix A is positive definite if and only if $U^t AU$ is positive definite. Then we have the following obvious facts for diagonal matrices , which guide us.

11.4.11 Proposition. *If the matrix A is diagonal, then*

1. *It is positive definite if and only if all its diagonal entries are positive. This can be rewritten:*
2. *Iit is positive definite if and only if all its its leading principal minors are positive.*

Proof. Indeed, if the diagonal entries are $\{d_1, d_2, \dots, d_n\}$, the leading principal minors are $D_1 = d_1$, $D_2 = d_1 d_2$, …, $D_n = d_1 d_2 \cdots d_n$. So the positivity of the d_i is equivalent to that of the D_i. □

The key step in the proof is the following result:

11.4.12 Proposition. *If A is has positive leading minors, it can be diagonalized by an invertible lower triangular matrix U with $u_{ii} = 1$ for all i. In other words UAU^t is diagonal.*

Proof. Let D_k denote the k-th leading principal minor of A, so

$$D_k = \det A(1, \dots, k).$$

Thus $a_{11} = D_1 > 0$, so we can use it to clear all the entries of the first column of the matrix A below a_{11}. In other words we left-multiply A by the invertible matrix U_1 whose only non-zero elements are in the first column except for 1 along the diagonal:

$$U_1 = \begin{pmatrix} 1 & 0 & \dots & 0 \\ -\frac{a_{21}}{a_{11}} & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{a_{n1}}{a_{11}} & 0 & \dots & 1 \end{pmatrix}$$

This is simply the first step of Gaussian elimination on A, as described in §1.1. Now we right multiply by the transpose of U_1 to get $A^{(1)} = U_1 A U_1^t$. It is symmetric with zeroes in the first row and column, except at a_{11}, Because $A^{(1)}$ is obtained from A by using row and column operations, the passage from A to $A^{(1)}$ does not modify the determinants of the submatrices that contain the first row and column. So if we denote by $D_k^{(1)}$ the leading principal minors of $A^{(1)}$, we get

$$D_k^{(1)} = D_k > 0$$

From the partial diagonalization we have achieved, we see that

$$D_2^{(1)} = a_{11} a_{22}^{(1)}$$

Since $D_2^{(1)} > 0$ by hypothesis, its diagonal element $a_{22}^{(1)} > 0$. Now we can iterate the process. We use this diagonal element to clear the elements in the second column of $A^{(1)}$ below $a_{22}^{(1)}$ by using Gaussian elimination to left-multiply by the invertible U_2:

$$U_2 = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & -\frac{a_{32}^{(1)}}{a_{22}^{(1)}} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & -\frac{a_{n2}^{(1)}}{a_{22}^{(1)}} & 0 & \dots & 1 \end{pmatrix}$$

Define $A^{(2)} = U_2 A^{(1)} U_2^t$. It is symmetric with zeroes in the first two rows and columns, except at $a_{11}^{(2)} = a_{11}$ and $a_{22}^{(2)} = a_{22}^{(1)}$. Since$A^{(2)}$ is obtained from $A^{(1)}$ using row and column operations involving the second row and second column, we have

$$D_k^{(2)} = D_k^{(2)} > 0.$$

Now from partial diagonalization

$$D_3^{(2)} = a_{11} a_{22}^{(1)} a_{33}^{(2)}.$$

So $a_{33}^{(2)} > 0$. Continuing in this way, we get a diagonal matrix $A^{(n-1)}$ with positive diagonal elements, obtained from A by left-multiplying by

$$U = U_{n-1} U_{n-2} \cdots U_1$$

and right-multiplying by its conjugate U^t. Since U is the product of lower triangular matrices with 1's along the diagonal, it is also lower triangular, so we are done. □

More generally, the proof implies a result that is interesting in its own right.

11.4.13 Proposition. *Let A be any symmetric matrix that can be diagonalized by a product U of lower triangular matrices as in Proposition 11.4.12. Then the leading principal minors of $A' = UAU^t$ are equal to those of A.*

Proof. As before, this immediately follows from the fact that if you add to a row (or column) of a square matrix a multiple of another row (or another column), then the determinant of the matrix does not change. Just apply this to the leading principal minors. □

11.4.14 Exercise. State the result concerning negative definite matrices that is analogous to the main theorem, noting that Proposition 11.4.13 applies. Do the same for Theorem 11.2.1.

We now finish the proof of the main theorem in the positive definite case.

Assume that A is a symmetric matrix whose leading principal minors D_k are all positive. Proposition 11.4.12 tells us that A can be diagonalized to a matrix $A^{(n-1)} = UAU^t$ by a lower diagonal matrix U with 1's on the diagonal The diagonal matrix $A^{(n-1)}$ obtained has all its diagonal entries a_{11}, $a_{22}^{(1)}$, ..., $a_{nn}^{(n-1)}$ positive, so it is positive definite by the easy Proposition 11.4.11. By Exercise 11.1.7 A is positive definite, so we are done.

We now prove Theorem 11.4.1 in the positive semidefinite case. Proposition 11.4.6 establishes one of the implications. For the other implication we build on Lang's Exercise 13 of VII.1. Use the proof of this exercise in [5] to prove:

11.4.15 Proposition. *If A is positive semidefinite, then $A + \epsilon I$ is positive definite for any $\epsilon > 0$.*

We also need Theorem 11.3.5. Write the characteristic polynomial of A as in (11.3.1):

$$P(t) = t^n - p_1 t^{n-1} + p_2 t^{n-2} - \cdots + (-1)^n p_n.$$

Since all the principal minors of A are non-negative, Theorem 11.3.5 says that all the p_i are non-negative. We have the elementary proposition:

11.4.16 Proposition. *If the characteristic polynomial of A is written as in* (11.3.1)*, then if all the p_i are non-negative, A is positive semidefinite. If all the p_i are positive, then A is positive definite.*

Proof. We first note that all the roots of $P(t)$ are non negative, only using the non-negativity of the p_i. Assume we have a negative root λ. Then all the terms of $P(\lambda)$ have the same sign, meaning that if n is even, all the terms are non-negative, while if n is odd, all the terms are non-positive. Since the leading term λ^n is non-zero, this is a contradiction. Thus all the roots are non-negative, and A is therefore positive semidefinite by Theorem 11.2.1, (1). If the p_i are all positive (in fact if a single one of them is positive) then the polynomial cannot have 0 as a root, so by the same criterion A is positive definite. □

This concludes the proof of Theorem 11.4.1.

For some worked examples, see §11.5.2, which is Lang's exercise 16 of Chapter VIII, §4.

11.5 The Characteristic Polynomial Test

We write the characteristic polynomial of A as in (11.3.1). With this notation, we get a new test for positive definiteness.

11.5.1 Theorem. *A is positive definite if and only if all the p_i, $1 \le i \le n$, are positive. A is positive semidefinite if and only if all the p_i, $1 \le i \le n$, are nonnegative.*

Proof. One implication follows immediately from Theorem 11.4.16.

For the reverse implication, we must show that if A is positive definite, then all the constants p_i are positive. This follows immediately from Theorem 11.3.5 and Proposition 11.4.6: all the principal minors are positive (non-negative) and the p_i are sums of them.

□

11.5.2 Exercise. Consider Exercise 16 of [VIII, §4].

How to determine if a matrix is positive definite? The easiest test is often the leading principal minors test of Theorem 11.4.1 which is not presented in [4], or the characteristic polynomial test above.

Exercise 16 a) and c). Not positive definite because the determinant is negative. The determinant is the product of the eigenvalues, which are all positive by Exercise 3 a), so this cannot be so.

Exercise 16 b). Positive definite, because the upper right hand corner entry is positive, and the determinant is positive.

Exercise 16 d) and e). Not positive definite because a diagonal element is 0. If the k-th diagonal element is 0, and e_k is the k unit vector, the $e_k^t A e_k = 0$, contradicting positive definiteness.

Note that the answers in Shakarni [5] are wrong for b), c) and e).

Chapter 12

Homework Solutions

This chapter contains solutions of a few exercises in Lang, where the solution manual is too terse and where the method of solution is as important as the solution itself.

12.1 Chapter XI, §3

The purpose here is to fill in all the details in the answer of the following problem in [5], which seems (from your solutions) to be unclear.

Exercise 8. Assume $A\colon V \to V$ is diagonalizable over K. This implies that the characteristic polynomial $P_A(t)$ can be written

$$P_A(t) = (t-\alpha_1)^{m_1}(t-\alpha_1)^{m_2}\cdots(t-\alpha_r)^{m_r},$$

where the m_i are the multiplicities of the eigenvalues. The goal of the exercise is to show that the *minimal polynomial* of A is

$$\mu(t) = (t-\alpha_1)(t-\alpha_1)\cdots(t-\alpha_r).$$

Thus all the factors appear, but with multiplicity one.

There are two things to prove.

First that the $\mu(t)$ vanishes on A. Any element in A can be written as a linear combination of eigenvectors v_i. Since $(A-\alpha_i I)$ vanishes on all the eigenvectors with eigenvalue α_i, this is clear.

So the harder step is to show that there is no polynomial $g(t)$ of smaller degree that vanishes on A. Write $A = BDB^{-1}$ where D is diagonal. Assume by contradiction that there is a

$$g(t) = t^k + g_{k-1}t^{k-1} + \cdots + g_1 t + g_0$$

of degree k, where $k < r$ that vanishes on A. Note that

$$A^i = (BDB^{-1})^i = BDB^{-1}BDB^{-1} \cdots BDB^{-1} = BD^iB^{-1}$$

because all the $B^{-1}B$ in the middle simplify to I. This implies

$$\begin{aligned} g(A) &= A^k + g_{k-1}A^{k-1} + \cdots + g_1 A + g_0 I \\ &= BD^kB^{-1} + g_{k-1}BD^{k-1}B^{-1} + \cdots + g_1 BDB^{-1} + g_0 BIB^{-1} \\ &= B\Big(D^k + g_{k-1}D^{k-1} + \cdots + g_1 D + g_0 I\Big)B^{-1} \\ &= Bg(D)B^{-1}. \end{aligned}$$

By assumption $g(A) = O$, where O is the $n \times n$ zero matrix. So

$$Bg(D)B^{-1} = O.$$

Multiplying by B^{-1} on the left, and B on the right, we get

$$g(D) = O. \tag{12.1.1}$$

If the diagonal entries of D are denoted $d_1, d_2, \ldots, d_n$, then D^i is diagonal with diagonal entries $d_1^i, d_2^i, \ldots, d_n^i$. Then (12.1.1) is satisfied if and only if

$$g(d_i) = 0 \text{ for all } i. \tag{12.1.2}$$

But $g(t)$ has only k roots, and there are $r > k$ distinct numbers among the d_i. So (12.1.2) is impossible, and the minimal degree for g is r.

Finally you may ask, could there be a polynomial $g(t)$ of degree r that vanishes on A, other than $\mu(t)$? No: if there were, we could divide $\mu(t)$ by $g(t)$. If there is no remainder, the polynomials are the same. If there is a remainder, it has degree $< r$ and vanishes on A: this is impossible as we have just established.

12.2 Chapter XI, §4

Exercise 2. $A\colon V \to V$ is an operator, V is finite dimensional. Assume that $A^3 = A$. Show

$$V = V_0 \oplus V_1 \oplus V_{-1}$$

where V_i is the eigenspace of A with eigenvalue i.

Solution The polynomial $t^3 - t$ vanishes at A. Since $t^3 - t$ factors as $t(t-1)(t+1)$, and these three polynomials are relatively prime, we can apply Theorem 4.2 of Chapter XI, which says that $V = V_0 \oplus V_1 \oplus V_{-1}$, where V_i is the kernel of

$(A - iI)$. When $i = 0$, we have the kernel of A, and when $i = \pm 1$ we get the corresponding eigenspaces.

Note that since we are not told that $t^3 - t$ is the characteristic polynomial of A, it is possible that there is a polynomial of lower degree on which A vanishes. This polynomial would have to divide $t^3 - t$: for example it could be $t^2 - 1$. In that case the space V_0 would have dimension 0. Thus, more generally, we cannot assert that any of the spaces V_i are positive dimensional.

Appendix A

Symbols and Notational Conventions

A.1 Number Systems

$\mathbb{N}$ denotes the natural numbers, namely the positive integers.

$\mathbb{Z}$ denotes the integers.

$\mathbb{Q}$ denotes the field of rational numbers, $\mathbb{R}$ denotes the field of real numbers, and $\mathbb{C}$ the field of complex numbers.

When a neutral name for a field is required, we use K. In these notes, as in Lang, it usually stands for $\mathbb{R}$ or for $\mathbb{C}$.

$[a, b] = \{x \in \mathbb{R} \mid a \leq x \leq b\}$.

$(a, b) = \{x \in \mathbb{R} \mid a < x < b\}$.

A.2 Vector Spaces

The n-th cartesian product of the real numbers $\mathbb{R}$ is written $\mathbb{R}^n$; similarly we have $\mathbb{C}^n$. Lower-case letters such as x and a denote vectors in $\mathbb{R}^n$ or $\mathbb{C}^n$, each with coordinates represented by $(x_1, \ldots, x_n)$ and $(a_1, \ldots, a_n)$, respectively.

Vectors are also called points, depending on the context. When the direction is being emphasized, it is called a vector.

When unspecified, vectors are column matrices.

In the body of the text, an expression such as $[a_1, a_2, \ldots, a_n]$ denotes a column vector while $(a_1, a_2, \ldots, a_n)$ denotes a row vector.

The length of a vector v is written $\|v\|$, and the inner product of v and w is $\langle v, w\rangle$, or $v \cdot w$.

A.3 Matrices

Matrices are written with parentheses. Matrices are denoted by capital roman letters such as A, and have as entries the corresponding lower case letter. So $A = (a_{ij})$. A is an $m \times n$ matrix if it has m rows and n columns, so $1 \leq i \leq m$ and $1 \leq j \leq n$. We write the columns of A as A^j and the rows as A_i, following Lang.
A^t is the transpose of the matrix A. This differs from Lang.
$D(d_1, d_2, \dots, d_n)$ is the $n \times n$ diagonal matrix

$$\begin{pmatrix} d_1 & 0 & 0 & \dots & 0 \\ 0 & d_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & 0 \\ 0 & 0 & 0 & \dots & d_n \end{pmatrix}$$

I_n or just I stands for the $n \times n$ identity matrix $D(1, 1, \dots, 1)$.
If A is an $m \times n$ matrix, L_A is the linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ given by $L_A(x) = Ax$, the matrix product of A by the n-column vector x. The kernel of L_A is written $Ker(A)$, and its image $Image(A)$.

References

[1] Michael Artin, *Algebra*, First Edition, Prentice Hall, New York, 1991. ↑vi, 10, 14, 56, 58, 60

[2] Kenneth Hoffman and Ray Kunz, *Linear Algebra*, Second Edition, Prentice Hall, Englewood Cliffs, NJ, 1971. ↑61

[3] Serge Lang, *Introduction to Linear Algebra*, Second Edition, Springer, New York, 1997. ↑1, 3, 7, 56

[4] ———, *Linear Algebra*, Third Edition, Springer, New York, 1987. ↑v, vi, 1, 7, 9, 10, 11, 12, 19, 21, 39, 56, 57, 59, 61, 64, 66, 67, 69, 81

[5] Rami Shakarni, *Solutions Manuel for Lang's Linear Algebra*, Springer, New York, 1996. ↑vi, 80, 81, 82

[6] Gilbert Strang, *Linear Algebra and its Applications*, Third Edition, Harcourt Brace Jovanovich, San Diego, CA, 1988. ↑46

[7] Robert J. Valenza, *Linear Algebra*: *An Introduction to Abstract Mathematics*, Undergraduate Texts in Mathematics, Springer, New York, 1993. ↑vi