LINEAR ALGEBRA AND MATRICES
M3P9

December 11, 2000

PART ONE. LINEAR TRANSFORMATIONS

Let k be a field. Usually, k will be the field of complex numbers $\mathbf{C}$, but it can also be any other field, e.g., the field of real numbers $\mathbf{R}$.

Let V be an n-dimensional vector space over k, and $f : V \to V$ be a linear transformation.

Given a basis $e_1, \ldots, e_n$ of V we can write the matrix of f in this basis:

$$A = (a_{ij}), \quad \text{where} \quad f(e_j) = \sum_{i=1}^{n} a_{ij} e_i$$

This means that j-th column of A is the column vector $f(e_j)$. In another basis f will be given by a different matrix. How is it related to A?

Lemma. Let $p_1, \ldots, p_n$ be another basis of V, and let X be the matrix of f in this basis. Then

$$P^{-1}AP = X$$

where $P = (p_{ij})$ is a $n \times n$-matrix whose columns are the column vectors $p_1, \ldots, p_n$ in order; in other words, $p_j = \sum_{i=1}^{n} p_{ij} e_i$.

Proof. By definition, $A(p_j) = \sum_{i=1}^{n} x_{ij} p_i$. The l.h.s. of this expression is

$$A(\sum_{i=1}^{n} p_{ij} e_i) = \sum_{i=1}^{n} p_{ij} A e_i = \sum_{i=1}^{n} p_{ij} \sum_{k=1}^{n} a_{ki} e_k = \sum_{k=1}^{n} (\sum_{i=1}^{n} a_{ki} p_{ij}) e_k.$$

The r.h.s. is

$$\sum_{i=1}^{n} x_{ij} \sum_{k=1}^{n} p_{ki} e_k = \sum_{k=1}^{n} (\sum_{i=1}^{n} p_{ki} x_{ij}) e_k.$$

On comparing these expressions we conculde that

$$\sum_{i=1}^{n} a_{ki} p_{ij} = \sum_{i=1}^{n} p_{ki} x_{ij}$$

for any k and j. Thus we have $AP = PX$. The matrix P is invertible because $p_1, \ldots, p_n$ is a basis, hence we finally get what we claimed. QED

Defintion. Two square $n \times n$-matrices A and B are *similar* if for some invertible $n \times n$-matrix P we have $P^{-1}AP = B$.

This is an equivalence relation. The previous lemma shows that the matrices in a similarity class are precisely the matrices of one and the same linear transformation written with respect to all possible bases of V. The first part of the course is devoted to the study of canonical forms of matrices: the classification of matrices up to similarity, or, what is the same, to finding a basis in which a given linear transformation has the matrix of the simplest possible form. We want this canonical form to be unique.

The simplest possible matrices are diagonal matrices. One can ask a natural question: is any matrix similar to a diagonal one? Unfortunately, the answer is negative, as we shall see shortly.

We shall use the following result from the first year linear algebra course:
The following conditions are equivalent:
(i) $Ker(A) \neq 0$,
(ii) $det(A) = 0$,
(iii) A *is not invertible.*

Definition. The *characteristic polynomial* of A is defined as

$$f_A(t) := det(t.Id - A),$$

where $Id = \delta_{ij}$ is the identity matrix. The roots of $f_A(t)$ are called the *eigenvalues* of A.

Similar matrices have the same characteristic polynomial, so it gives us a means to distinguish between different similarity classes. Equivalently, $\lambda \in k$ is an eigenvalue of A if and only if $\lambda.Id - A$ is *not* invertible.

Example. Find the characteristic polynomial and the eigenvalues of an upper triangular matrix.

Definiton. A non-zero vector $v \in V$ is called an *eigenvector* of A if $Av = \lambda v$ for some $\lambda \in k$.

Properties. If λ is an eigenvalue of A, then there is an eigenvector with this eigenvalue. If $k = \mathbf{C}$, then any matrix has an eigenvalue, and hence an eigenvector. If $k = \mathbf{R}$, this is no longer true, for example the matrix

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

has no real eigenvalue, and hence no real eigenvector.

Theorem 1. A matrix is diagonalizable (that is, is similar to a diagonal matrix) if and only if it has a basis consisting of eigenvectors. If this is the case, then the entries of the corresponding diagonal matrix are the eigenvalues of A.

Proof. We shall usually denote by $e_1, \ldots, e_n$ the standard basis of V, $e_i = (0, \ldots, 1_i, \ldots, 0)$.

Suppose that $X = P^{-1}AP$ is diagonal. Then $e_1, \ldots, e_n$ are eigenvectors of X. Then $p_1 := Pe_1, \ldots, p_n := Pe_n$, is a basis of V consisting of eigenvectors of A.

Conversely, let $p_1, \ldots, p_n$ be a basis of A consisting of eigenvectors with eigenvalues $\lambda_1, \ldots, \lambda_n$, respectively. By Lemma above $X = P^{-1}AP$ is the matrix of A in the basis $p_1, \ldots, p_n$, where $p_i = Pe_i$. Then $Xe_i = \lambda_i.e_i$. QED

Examples. Let us consider the matrix

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

If $v = (v_1, v_2)$, then $Sv = (v_2, 0)$. Thus all eigenvectors of S are proportional to e_1. Hence S is not diagonalizable.

Theorem 2. Let $v_1, \ldots, v_m$ be eigenvectors of A such that their eigenvalues λ_i are pairwise different, $\lambda_i \neq \lambda_j$ if $i \neq j$. Then $v_1, \ldots, v_m$ are linearly independent.

Proof. Define B_i as the product of $\lambda_j.Id - A$ for all $j \neq i$. (These matrices commute so the order is not important.) Then B_i sends v_j for $j \neq i$ to 0, and multiplies v_i by a non-zero number

$$(\lambda_1 - \lambda_i) \ldots (\lambda_{i-1} - \lambda_i)(\lambda_{i+1} - \lambda_i) \ldots (\lambda_n - \lambda_i),$$

because by assumption the λ_i are all different.

Suppose that

$$\mu_1 v_1 + \ldots + \mu_n v_n = 0.$$

Applying B_i to this we conclude that $\mu_i = 0$. Hence there is no non-trivial relation among the v_i. QED

A corollary of this theorem is a sufficient condition of diagonalizability.

Corollary. If $f_A(t)$ has n different roots, then A is diagonalizable.

A diagonal matrix is clearly upper triangular. Over an algebraically closed field any matrix is similar to an upper triangular matrix, as shows the following theorem.

Theorem 3. Let A be a matrix such that $f_A(t)$ is a product of linear factors over k, then A is similar to an upper triangular matrix (with eigenvalues on the main diagonal).

Proof. We proceed by induction in $dim(V)$. The assertion is trivially true for $dim(V) = 1$. Suppose it is true for vector spaces of dimension $n-1$

Let $\lambda_1 \in k$ be a root of $f_A(t)$, and $v_1 \in V$ be an eigenvector with eigenvalue λ_1. Let $v_2, \ldots, v_n$ be vectors in V such that $v_1, \ldots, v_n$ is a basis of V. Then the matrix of A in this basis is

$$X = \begin{pmatrix} \lambda_1 & * & \ldots & * \\ 0 & b_{11} & \ldots & b_{1,n-1} \\ 0 & b_{21} & \ldots & b_{2,n-1} \\ \ldots & & & \ldots \\ 0 & b_{n-1,1} & \ldots & b_{n-1,n-1} \end{pmatrix}$$

Then $f_X(t) = f_A(t) = (t-\lambda_1)f_B(t)$, where $B = (b_{ij})$, hence $f_B(t)$ is a product of linear factors. By the inductive assumption, $C := Q^{-1}BQ$ is upper triangular for some $(n-1) \times (n-1)$-matrix Q. Let T be the direct sum of Id of size 1×1 and Q. Then

$$T^{-1}XT = \begin{pmatrix} \lambda_1 & * & \ldots & * \\ 0 & c_{11} & \ldots & c_{1,n-1} \\ 0 & c_{21} & \ldots & c_{2,n-1} \\ \ldots & & & \ldots \\ 0 & c_{n-1,1} & \ldots & c_{n-1,n-1} \end{pmatrix}$$

This matrix is upper triangular. QED

This theorem shows that when k is an algebraically closed field, say $k = \mathbf{C}$, then any matrix is similar to an upper triangular matrix. This is often a big help in calculations, as shows the following proof.

Cayley–Hamilton theorem. $f_A(A) = 0$.

First proof. Before starting the proof we note that for any polynomial $g(t)$ we have $g(P^{-1}AP) = P^{-1}g(A)P$. Thus it is enough to check the statement on any matrix in the similarity class of A.

To check that $f_A(A) = 0$ we can go over to an extension of the ground field k over which $f_A(t)$ is a product of linear factors. Then by the previous theorem we can assume that

$$A = \begin{pmatrix} \lambda_1 & * & * & \ldots & * \\ 0 & \lambda_2 & * & \ldots & * \\ 0 & 0 & \lambda_3 & \ldots & * \\ \ldots & & & & \ldots \\ 0 & 0 & 0 & \ldots & \lambda_n \end{pmatrix}$$

One proves by induction that $(A-\lambda_1 Id)\dots(A-\lambda_m Id)$ kills e_1, ... ,e_m. (Note that $(A-\lambda_m Id)e_m$ is a linear combination of e_1, ..., e_{m-1}.) Thus $f_A(A)$ maps V to 0, that is, $f_A(A)$ is a zero matrix. QED

Second proof. Recall that the matrix $adj(B)$ (whose entries are principal minors of B with appropriate signs) is a matrix of the same size as B such that $det(B).Id = adj(B)B$. Let $adj(t.Id - A)$ be the adjacent matrix of $t.Id - A$. Then we have $f_A(t).Id = adj(t.Id - A).(t.Id - A)$. Both sides here are polynomials in t with matrix coefficients. Hence we can substitute t by A. Then the right hand side vanishes, and we get $f_A(A) = 0$. QED

There exists a unique monic polynomial $m_A(t)$ with coefficients in k such that $m_A(A) = 0$ and $deg(m_A(t))$ is minimal among the polynomials annihilating A. Indeed, by the Cayley–Hamilton theorem the set of non-trivial polynomials $p(t)$ such that $p(A) = 0$ is not empty, therefore it contains a monic polynomial of minimal degree. If there are two such polynomials, the degree of their difference is less than $deg(m_A(t))$, which is a contradiction. Hence such a polynomial is unique.

Definition. The unique polynomial $m_A(t)$ with coefficients in k such that $m_A(A) = 0$ and $deg(m_A(t))$ is minimal among the polynomials annihilating A is called the *minimal* polynomial of A.

It is clear that $m_A(t)$ divides $f_A(t)$. (We have $f_A(t) = g(t)m_A(t) + r(t)$ for some polynomials $g(t)$ and $r(t)$, $deg(r(t)) < deg(m_A(t))$; this implies that $r(A) = 0$ which is a contradiction.) Note that $m_A(t) = m_{P^{-1}AP}(t)$.

Example. If A is diagonal with paiwise different diagonal entries, then $m_A(t) = f_A(t)$. Note, however, that $m_{Id}(t) = t-1$, whereas $f_{Id}(t) = (t-1)^n$.

Proposition. Every eigenvalue of A is a root of $m_A(t)$.

Proof. If $Av = \lambda v$, $v \neq 0$, then $0 = m_A(A)v = m_A(\lambda)v$, hence $m_A(\lambda) = 0$. QED

Exercise. The minimal polynomial of the matrix

$$\begin{pmatrix} 0 & 0 & 0 & \dots & 0 & -c_0 \\ 1 & 0 & 0 & \dots & 0 & -c_1 \\ 0 & 1 & 0 & \dots & 0 & -c_2 \\ \dots & & & & & \dots \\ 0 & 0 & 0 & \dots & 1 & -c_{n-1} \end{pmatrix}$$

is $t^n + c_{n-1}t^{n-1} + \dots + c_1 t + c_0$.

From now on we suppose that $f_A(t)$ is a product of linear factors, in other words, that all eigenvalues of A are k.

We fix our notation. Let

$$f_A(t) = \prod_{i=1}^{m} (t - \lambda_i)^{d_i}, \quad \text{where} \quad \lambda_i \neq \lambda_j, \quad \text{if} \quad i \neq j, \quad \sum_{i=1}^{m} d_i = n;$$

$$f_j(t) = \prod_{i=1, i \neq j}^{m} (t - \lambda_i)^{d_i}.$$

Define $V_i := Ker((\lambda_i.Id - A)^{d_i})$. It is clear that $AV_i \subset V_i$, that is, the subspaces V_i are A-invariant.

Lemma. $V_i = Im(f_i(A))$.

Proof. Cayley–Hamilton theorem implies that $Im(f_i(A)) \subset V_i$, so it remains to show the opposite inclusion. The ideal in $k[t]$ generated by the polynomials $f_i(t)$ is the whole ring $k[t]$, hence $1 = \sum_{i=1}^{m} h_i(t) f_i(t)$ for some polynomials $h_i(t)$. (This follows from the fact that $k[t]$ has Euclid's algorithm.) Then we have for any $v \in V$

$$v = \sum_{i=1}^{m} h_i(A) f_i(A) v = \sum_{i=1}^{m} f_i(A) h_i(A) v.$$

Applying this to $v \in V_i$, and taking into account that $f_j(t)$ for $j \neq i$ is divisible by $(t - \lambda_i)^{d_i}$, we conclude that $v = h_i(A) f_i(A) v = f_i(A) h_i(A) v \in Im(f_i(A))$. QED

Theorem. $V = \oplus_{i=1}^{m} V_i$.

Proof. The last displayed formula shows that V is generated by the sum of the $V_i = Im(f_i(A))$. Let us show that the sum is direct. Suppose that $v_1 + \ldots + v_m = 0$, where $v_i \in V_i$. Since $f_i(A)$ kills all the v_j except possibly v_i, it must also kill v_i, that is, $f_i(A) v_i = 0$. Since $Id = \sum_{i=1}^{m} h_i(A) f_i(A)$ we have $v_i = h_i(A) f_i(A) v_i$. This implies that $v_i = 0$. QED

This reduces the classification of arbitrary linear transformations to that of nilpotent ones (a matrix M is *nilpotent* if $M^d = 0$ for some d).

Let us explain this. Recall that $AV_i \subset V_i$. If we choose some bases, say $p_1^{(i)}, \ldots, p_{k_i}^{(i)}$, in each of the V_i, we get a basis of V in which the matrix of A is the direct sum of blocks. Each block is the matrix of A restricted to V_i (in the basis $p_1^{(i)}, \ldots, p_{k_i}^{(i)}$). Let B_i be the restriction of $A - \lambda_i.Id$ to V_i. Then $B_i^{d_i} = 0$.

Remark. We have $f_{B_i}(t) = t^{dim(V_i)}$ because the only eigenvalue of any nilpotent matrix is 0. This implies that the characteristic polynomial of A restricted to V_i is $(t-\lambda_i)^{dim(V_i)}$. Since $f_A(t) = \prod_{i=1}^m (t-\lambda_i)^{d_i}$ is the product of these, we conclude that $dim(V_i) = d_i$.

It remains to construct a "good" basis of a nilpotent linear transformation.

Let B be a linear transformation of a vector space W such that $B^d = 0$ whereas $B^{d-1} \neq 0$. Define

$$W_i = Ker(B^i), \quad i = 0, \ldots, d.$$

Then $BW_i \subset W_{i-1}$. We have

$$W = W_d \supset W_{d-1} \supset \ldots \supset W_1 \supset W_0 = 0$$

Lemma. For any $j > i$ if $w \in W_j$ is not in W_{j-1}, then $B^{j-i}w$ is in $W_i \setminus W_{i-1}$ (obvious).

(In particular, taking $j = d$ we see that all the inclusions in the dispayed formula above are strict, that is, $W_i \neq W_{i-1}$.) Based on this Lemma we construct a basis of W as follows.

Choose $w_1^{(1)}, \ldots, w_{s_1}^{(1)}$ to be linearly independent vectors in W such that

$$W =< w_1^{(1)}, \ldots, w_{s_1}^{(1)} > \oplus W_{d-1}.$$

By the previous Lemma we conclude that $Bw_1^{(1)}, \ldots, Bw_{s_1}^{(1)}$ are linearly independent vectors in W_{d-1}, and the intersecion of the space generated by them with W_{d-2} is zero. Let us choose linearly independent vectors $w_1^{(2)}, \ldots, w_{s_2}^{(2)}$ in W_{d-1} so that

$$W_{d-1} =< w_1^{(2)}, \ldots, w_{s_2}^{(2)} > \oplus < Bw_1^{(1)}, \ldots, Bw_{s_1}^{(1)} > \oplus W_{d-2}.$$

Then we continue like this, adding new vectors at each step to generate W_i by W_{i-1} and the images of vectors already chosen. In the end we get the set of vectors $\{B^k w_j^{(i)}\}$, where $i = 1, \ldots, d-1$, $j = 1, \ldots, s_i$, $k = 0, \ldots, d-i$. It is clear by construction that it generates the whole space W. We assert that the vectors $\{B^k w_j^{(i)}\}$, $i = 1, \ldots, d-1$, $j = 1, \ldots, s_i$, $k = 0, \ldots, d-i$, are linearly independent. Let

$$\sum a_{ijk} B^k w_j^{(i)} = 0$$

be a non-trivial linear combination. Apply B^{d-1} to it, the only possibly surviving terms are $\sum a_{1j0}B^{d-1}w_j^{(1)} = 0$, but then $\sum a_{1j0}w_j^{(1)} = 0$ (by the Lemma). However, these vectors are linearly independent, hence all $a_{1j0} = 0$. Next, we apply B^{d-2}, then the only possibly non-vanishing terms are $\sum a_{1j1}B^{d-1}w_j^{(1)} + \sum a_{2j0}B^{d-2}w_j^{(2)} = 0$. By the Lemma this implies that $\sum a_{1j1}Bw_j^{(1)} + \sum a_{2j0}w_j^{(2)} = 0$, and then again all coeficients must be zero since these vectors are linearly independent by constrution. Continuing in the same spirit we prove that all our vectors are linearly independent, and so form a basis of W.

Although the construction of this basis was somewhat involved, the matrix of B now takes a very simple form. Let us rearrange the basis vectors as follows:

$$B^{d-1}e_1^{(1)}, B^{d-2}e_1^{(1)}, \dots, e_1^{(1)}, B^{d-1}e_2^{(1)}, B^{d-2}e_2^{(1)}, \dots, e_2^{(1)}, \dots, B^{d-1}e_{s_1}^{(1)}, B^{d-2}e_{s_1}^{(1)}, \dots, e_{s_1}^{(1)},$$

$$B^{d-2}e_1^{(2)}, B^{d-3}e_1^{(2)}, \dots, e_1^{(2)}, B^{d-2}e_2^{(2)}, B^{d-3}e_2^{(2)}, \dots, e_2^{(2)}, \dots, B^{d-2}e_{s_2}^{(2)}, B^{d-3}e_{s_2}^{(2)}, \dots, e_{s_2}^{(2)},$$

$$\dots,$$

$$Be_1^{(d-1)}, e_1^{(d-1)}, Be_2^{(d-1)}, e_2^{(d-1)}, \dots, Be_{s_{d-1}}^{(d-1)}, e_{s_{d-1}}^{(d-1)}, e_1^{(d)}, \dots, e_{s_d}^{(d)}.$$

Then the matrix of B is the direct sum of blocks of the form

$$\begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ \dots & & & & & \dots \\ 0 & 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 \end{pmatrix}$$

Such a matrix is called a nilpotent Jordan block. There are s_1 such blocks of size d, s_2 such blocks of size $d-1$, and so on, finally, there are s_{d-1} blocks of the form

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

and s_d zero blocks of size 1.

Note that the form of this matrix implies that $rk(B^i) = (d-i)s_1 + (d-i-1)s_2 + \dots + s_{d-i}$. this implies that the numbers s_i are uniquely determined by the similarity class of B. We have proved that a nilpotent matrix B is similar to a direct sum of nilpotent Jordan blocks uniquely determined up to a permutation.

A matrix of the form

$$\begin{pmatrix} \lambda & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & \lambda & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & \lambda & 1 & \dots & 0 & 0 \\ \dots & & & & & & \dots \\ 0 & 0 & 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 & \lambda \end{pmatrix}$$

is called a Jordan block. Summarizing our preceding discussion we obtain the following statement.

Theorem. (The Jordan normal form.) Any matrix is similar to a direct sum of Jordan blocks. The Jordan normal form in each similarity class is unique up to a permutation of Jordan blocks.

Exercises - corollaries of this theorem. (a) Let A be a Jordan normal form, then $m_A(t) = \prod_{i=1}^{m}(t-\lambda_i)^{\delta_i}$, where δ_i is the size of a maximal Jordan block with eigenvalue λ_i. (Use the fact that the minimal polynomial of a direct sum of matrices is the l.c.m. of their minimal polynomials.)

(b) *A matrix is diagonalizable iff its minimal polynomial is a product of distinct linear factors.* (This is read directly from the Jordan normal form.) A corollary of this fact: if the direct sum of two matrices is diagonalizable, then these matrices are also diagonalizable.

Definition. Let A be a linear transformation on a vector space V. The set of vectors $v \in V$ such that $Av = \lambda v$ is called the eigenspace with eigenvalue λ, and is denoted by V_λ.

Reformulating an earlier result we have the following criterion of diagonalizability: A is diagonalizable iff V is the direct sum of eigenspaces of A. (=iff V has a basis consisting of eigenvectors of A)

Another corollary of the theorem. The number of Jordan blocks with eigenvalue λ equals $dim(V_\lambda)$. (Each block contributes one eigenvector.)

Proposition. Two commuting diagonalizable matrices are simultaneously diagonalizable. (In other words, there exists an invertible matrix P such that $P^{-1}AP$ and $P^{-1}BP$ are both diagonal.)

Proof. If V_λ is an eigenspace of A, then $AB = BA$ implies that $BV_\lambda \subset V_\lambda$. We have seen that the restriction of B to V_λ is diagonalizable (its minimal polynomial is a divisor of $m_B(t)$ which is a product of diffirent linear factors, hence it is also a product of diffirent linear factors). Since the restriction of A

to V_λ is scalar, we can find a basis of V_λ in which both A and B are diagonal. Putting such bases of V_λ together we get a desired basis of V. QED

Actually, the same is true for any number of pairwise commuting diagonalizable matrices (induction on the number of matrices).

The following result has many applications in the theory of Lie groups and Lie algebras.

Theorem. (Jordan decomposition) Any (square) matrix A can be represented as a sum $A = D_A + N_A$, where D_A is diagonalizable and N_A is nilpotent, and $D_A = g_1(A)$, $N_A = g_2(A)$, for some polynomials $g_1(t)$ and $g_2(t)$. This implies that $D_A N_A = N_A D_A$, and, more generally, D_A and N_A commute with any matrix B such that $AB = BA$.

Hint. In our previous notation take $D_A = \sum_{i=1}^m \lambda_i h_i(A) f_i(A)$, and use the fact that $h_i(A) f_i(A)$ is the projector from V to V_i (that is, it is identity on V_i and zero on V_j if $j \neq i$).

PART TWO. BILINEAR FORMS

Definition. Let V be a vector space over a field k. A function $<,>$ from $V \times V$ to k is called a bilinear form if it is linear in each argument.

Definition. If $p_1, \dots, p_n$ is a basis of V, then *the Gram matrix* of $<,>$ in this basis is the matrix $A = (a_{ij})$ such that $a_{ij} =< p_i, p_j >$.

Then if v and u are column vectors

$$v = \sum_{i=1} v_i p_i = (v_1, \dots, v_n)^t, \quad u = \sum_{i=1} u_i p_i = (u_1, \dots, u_n)^t$$

we have

$$< v, u >= v^t A u.$$

Lemma. If $f_1, \dots, f_n$ is another basis of V, and $F = (f_{ij})$ is a matrix such that $f_j = \sum_{i=1}^n f_{ij} p_i$, then the Gram matrix of $<,>$ in the new basis $f_1, \dots, f_n$ is $F^t A F$, where F^t is the transpose of F.

The proof is a direct check.

Matrices A and B (square, of the same size) such that $B = F^t A F$ for some invertible matrix F are called *congruent*. Note that then we have $rk(B) = rk(A)$. If $rk(A) = n$, the bilinear form $<,>$ is called *non-singular* (or *non-degenerate*).

Theorem. $<,>$ is non-singular iff for any non-zero $v \in V$ there exists $u \in V$ such that $< v, u >\neq 0$.

Proof. The set of $v \in V$ such that $< v, u >= 0$ for any $u \in V$ is just the kernel of A^t. QED

Definition. If $U \subset V$ is a subspace, then

$$U^\perp := \{v \in V \ \text{ such that } < v, u >= 0 \ \text{ for all } \ u \in U\}$$

Lemma. If U is a subspace of V such that the restriction of $<,>$ to U is non-singular, then $V = U \oplus U^\perp$.

Proof. We need to prove that $U \cap U^\perp = 0$ (an immediate consequence of the non-singularity assumption), and that V is generated by U and $U^\perp$. Take any $v \in V$, then by non-singularity we can find $w \in U$ such that $< u, v >=< u, w >$ for all $u \in U$. Such a w can be defined by the formula

$$Bw = (< p_1, v >, \dots, < p_m, v >)^t,$$

where $p_1, \dots p_m$ is a basis of U, and B is the Gram matrix of $<,>$ restricted to U. Then $v - w \in U^\perp$, and $v = w + (v - w)$ is the desired decomposition. QED

Symmetric bilinear forms

Definition. A bilinear form is called symmetric if $< v, u >=< u, v >$ for all $v, u \in V$.

Bilinear forms correspond to symmetric Gram matrices (such that $A^t = A$).

Lemma. Assume $char(k) \neq 2$. A symmetric form is non-zero iff there exists $v \in V$ such that $< v, v > \neq 0$.

Proof. The form is non-zero iff there exist $v, u \in V$ such that $< v, u > \neq 0$. Then the formula

$$< u, v >= \frac{1}{2}(< v + u, v + u > - < v, v > - < u, u >)$$

implies that at least one of $< v + u, v + u >$, $< v, v >$, $< u, u >$ is non-zero. QED

The previous formula implies that provided the characteristic of the ground field is different from 2, a symmetric bilinear form $< u, v >$ can be recovered from its restriction to the diagonal, that is the function $V \to k$ given by $< v, v >$. This function is called *the associated quadratic form*, in coordinates it is simply the expression $\sum_{i,j=1}^{n} a_{ij} x_i x_j$. The well known process of reducing a quadratic polynomial to a sum of squares is what lies behind the idea of proof of the following theorem.

Theorem. Diagonalization of a symmetric bilinear form. Assume $char(k) \neq 2$. For any symmetric bilinear form there exist a basis of V in which its Gram matrix is diagonal.

Proof. If our form is zero there is nothing to prove. Suppose it is not, then by the previous lemma we can find $u \in V$ such that the restriction of $<,>$ to the subspace generated by u is non-singular. Call this subspace U. Then $V = U \oplus U^\perp$, and we repeat this argument for $<,>$ restricted to $U^\perp$, and so on. QED

Corollary 1. Assume $char(k) \neq 2$. Every symmetric matrix is congruent to a diagonal one.

Although in general the coefficients of this diagonal matrix are not uniquely defined, over some fields one can be more precise.

Corollary 2. If $k = \mathbf{C}$, then any symmetric matrix is congruent to the diagonal matrix $diag(1, \ldots, 1, 0, \ldots, 0)$. In particular, symmetric matrices A and B are congruent over $\mathbf{C}$ iff $rk(A) = rk(B)$. Hence there are $n + 1$ congruency classes.

Corollary 3. If $k = \mathbf{R}$, then any symmetric matrix is congruent to the diagonal matrix $diag(1, \ldots, 1, -1, \ldots, -1, 0, \ldots, 0)$.

Theorem. Sylvester's law of inertia. The number of 1's and -1's in a diagonal form of a real symmetric matrix are well defined (i.e., these numbers only depend on the congruency class of our matrix).

Proof. Let p be the number of 1's and let q be the number of -1's. The fact that $p + q$ is well defined is clear - this is just the rank of A. We need a definition. A symmetric bilinear form $<,>$ is *positive definite* if $< u, u >$ is positive for any non-zero u. (Likewise $<,>$ is *negative definite* if $< u, u >$ is negative for any non-zero u. If the form is neither positive definite nor negative definite it is called *indeterminate*.)

Consider the following number which clearly does not depend on the choice of a basis:

$$max_{U \subset V}\{dim(U)\},$$

where U is a subspace of V such that $<,>$ restricted to U is positive definite. Let us compute this number. Let $W \subset V$ be generated by the last $n - p$ basis vectors of the basis in which $<,>$ has the form of Corollary 3. Then if $dim(U) + dim(W) > n$, then $U \cap W$ contains a non-zero vector, say u. But then $< u, u > \leq 0$ by the construction of W. Hence our number cannot be greater than $n - (n - p) = p$. On the other hand, the restriction of $<,>$ to the subspace generated by the first p vectors of the above basis is clearly positive definite. Hence this number is p, which therefore is well defined. QED

Definition. The number $p - q$ is called the *signature* of the symmetric bilinear form $<,>$.

Alternating bilinear forms

Definition. A bilinear form $<,>$ is alternating if $< u, u >= 0$ for any $u \in V$. A bilinear form $<,>$ is skew-symmetric if $< u, v >= - < v, u >$ for any $u, v \in V$.

An alternating form is skew-symmetric, but the converse is only true when $char(k) \neq 2$ (take $u = v$ and divide by 2), and is wrong when $char(k) = 2$ (symmetric form over such a field is the same thing as skew-symmetric!).

A standard example of an alternating form is $< (v_1, v_2), (u_1, u_2) >= v_1u_2 - v_2u_1$. Its Gram matrix is

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

Theorem. Let k be a field of characteristic different from 2. Then for any alternating form there exists a basis in which its Gram matrix is the direct sum of copies of J and a zero matrix.

In other words, any skew-symmetric matrix (that is, $A^t = -A$) with zeros on the main diagonal is congruent to a matrix of the above form. Note that this theorem is valid over any field.

Proof. If $<,>$ is not zero, take a non-zero vector v such that there exists u with the property that $< v, u >\neq 0$. Then v and $u/ < v, u >$ are linearly independent. Let U be the space spanned by them. The Gram matrix of $<,>$ restricted to U is J. Since this restriction is non-singular we have $V = U \oplus U^{\perp}$. We repeat this process for $U^{\perp}$ until we get a zero form. QED

Corollary. The rank of an alternating form is even. In particular, non-singular alternating forms exist only on even-dimensional vector spaces. The determinant of a skew-symmetric matrix is a square in k. In particular, if $k = \mathbf{R}$, it is always non-negative.

Exercise: The Pfaffian of a skew-symmetric matrix. The last corollary leaves open the question: what is "the square root" of $det(A)$, where $A = (a_{ij})$ is a skew-symmetric matrix of size $n = 2m$? If $m = 1$, then obviously $det(A) = a_{12}^2$, so here the answer is clear. Show by direct computation that for $m = 2$ we have

$$det(A) = (a_{12}a_{34} - a_{13}a_{24} + a_{14}a_{23})^2$$

This formula has an analogue for any m. In fact, there exists a polynomial in matrix entries of A, called the Pfaffian $Pf(A)$, such that $det(A) = Pf(A)^2$ (over any field).

Inner product spaces. Geometry of quadric surfaces

Let V be a vector space over $\mathbf{R}$. An *inner product* is a positively definite symmetric bilinear form $<,>$ on V.

An example of such a form is the standard *dot product* $u^t.v$. One defines orthonormal bases and proves the Cauchy–Schwarz inequality. In terms of the dot product one writes $< u, v >= u^t.A.v$, where A is the Gram matrix of $<,>$. One has $u^t.(A.v) = (A^t.u)^t.v$.

Definition. A square matrix P is an *orthogonal* matrix if $P^tP = Id$.

Such matrices preserve the dot product, and hence preserve the distances. Recall that the distance between v and v' is defined as $|v - v'| := \sqrt{(v - v')^t.(v - v')}$.

Proposition. All eigenvalues of a real symmetric matrix (considered as complex numbers) are real.

Proof. Let $z = (z_1, \ldots, z_n)$ be a complex eigenvector, then $\overline{z}^t.A.z = \lambda \sum_{i=1}^n |z_i|^2$. Since A is symmetric and real, this also equals $(\overline{A.z})^t.z = \overline{\lambda} \sum_{i=1}^n |z_i|^2$. Hence $\lambda \in \mathbf{R}$. QED

Theorem. Reduction to principal axes. For any real symmetric matrix there exists an orthonormal basis of V consisting of its eigenvectors.

Such a basis is not necessarily unique.

Proof. We proceed by induction on $n = dim(V)$. When $n = 1$ there is nothing to prove. Suppose the statement is true for $n - 1$. We can find an eigenvector u of length 1 (that is, $u^t.u = 1$) with some eigenvalue λ, and call U its linear span. Let U' be its orthogonal complement with respect to *the dot product*. Since $u^t.A.v = (A.u)^t.v = \lambda.u^t.v$, we see that if v is orthogonal to u with respect to the dot product, then Av is also. Hence $AU' \subset U'$. Then $V = U \oplus U'$ (since the restriction of the dot product to U is non-singular), and U and U' are A-invariant. Now take any othonormal basis in U', which exists by the inductive assumption. Together with u this gives an orthonormal basis of V. QED

The lines passing through the vectors of the basis constructed in this theorem are called *principal axes*.

Corollary. For any real symmetric matrix there exists an orthogonal matrix P such that $P^{-1}AP = P^tAP$ is diagonal (with eigenvalues on the diagonal).

Definition. A map $f : \mathbf{R}^n \to \mathbf{R}^n$ which preserves distances, that is, such that $|f(v) - f(v')| = |v - v'|$ for all $v, v' \in \mathbf{R}$, is called an *isometry*.

Proposition. Any isometry is a composition of an orthogonal transformation $v \mapsto Pv$ and a translation $v \mapsto v + w$.

Proof. Let $f : \mathbf{R}^n \to \mathbf{R}^n$ be an isometry. By composing f with the translation by $-f(0)$ we arrange that 0 goes to 0. So we can assume that $f(0) = 0$. Now taking $v' = 0$ in the definition of isometry, we see that $|f(u)| = |u|$ for any u. Now note that

$$u^t.v = (1/2)(|u|^2 + |v|^2 - |u - v|^2)$$

and all the terms in the right hand side are preserved by f. Therefore, we have $(f(u)^t.f(v)) = u^t.v$ for any u and v. (This is, in fact, the standard observation that the angles of a triangle are uniquely determined by its sides.) In other words, f preserves the dot product. In particular, the image of an orthonormal basis is again an orthonormal basis, say $p_i = f(e_i)$. Now any vector u is written as

$$u = \sum_{i=1}^{n} u_i e_i, \quad u_i = (u^t.e_i).$$

If we write $f(u) = \sum_{i=1}^{n} x_i p_i$ we must have $x_i = f(u)^t.p_i = f(u)^t.f(e_i) = u^t.e_i = u_i$. Therefore, we have $f : \sum_{i=1}^{n} u_i e_i \mapsto \sum_{i=1}^{n} u_i p_i$, hence f is linear, say $f(u) = Fu$. Then $(Fu)^t.Fv = u^t.v$ implies that F is an orthogonal matrix. QED

Proposition. Any equation of degree two

$$\sum_{i,j=1}^{n} a_{ij} x_i x_j + \sum_{i=1}^{n} b_i x_i + c = 0$$

can be reduced by an isometry to one of the following forms, where all the coefficients are non-zero:

$$\text{I}: \quad \sum_{i=1}^{m} \lambda_i x_i^2 - 1 = 0,$$

$$\text{II}: \quad \sum_{i=1}^{m} \lambda_i x_i^2 = 0,$$

$$\text{III}: \quad \sum_{i=1}^{m-1} \lambda_i x_i^2 + x_m = 0,$$

where $1 \leq m \leq n$, and in the last case $m - 1 \geq 1$.

Proof. By the principal axes theorem we can assume that the terms of order 2 have diagonal form. Use translations to complete the squares. If there are no linear terms left, after possibly multiplying by an overall constant, we find ourselves in case I or II. If there are some linear terms, change the coordinates by an orthogonal transformation in $x_{m+1}, \ldots, x_n$, which is identity on $x_1, \ldots, x_m$, so that there is only one linear term. (This uses the fact that any vector of unit length can be completed to an orthonormal basis.) After a translation (to eliminate the constant term) and a multiplication by a constant we arrive at case III. QED

If $m < n$, we say that the graph of the corresponding equation is a *cylinder*, such a graph can be obtained from a graph in a space of dimension $\leq n-1$. In case II the graph is a *cone*. Recall that if $n = m = 2$, then the quadric curve (a *conic*) is an ellipse if both coefficients are positive, a hyperbola if just one is positive, and an imaginary ellipse if both are negative. In case III we get a parabola.

Let us now consider the case $n = m = 3$. In case I if all the coefficients are positive, the quadric is called an *ellipsoid*, if just one coefficient is negative we get a *hyperboloid of one sheet*, if exactly two coefficients are negative, we have a *hyperboloid of two sheets*, and, finally, if all the coefficients are negative, we have an *imaginary ellipsoid*. In case III, if the coefficients have the same sign, we have an *elliptic paraboloid*, and in the opposite case an *hyperbolic paraboloid*.

It is clear that cylinders and cones are "made of lines" (there is a line contained in our surface passing through any point of it). One can prove that an ellipsoid, an elliptic paraboloid and a hypeboloid of 2 sheets do not contain lines. (In fact, they do not contain a subset of a straight line if it has at least 3 points!) On the contrary, there are lines on a hyperboloid of 1 sheet and on a hyperbolic paraboloid. Actually, we have the following stronger statement.

Theorem. Through every point of a hyperbolic paraboloid pass two different lines entirely contained in it. (And the same is true for the hyperboloid of 1 sheet.)

Proof. We have an equation $ax^2 - by^2 = z$, where $a, b > 0$. Let (x_0, y_0, z_0) be any solution of this equation. Consider the plane given by the equation $2ax_0x - 2by_0y - (ax_0^2 - by_0^2) = z$ (this is a tangent plane). Substituting this expression for z into the first equation we obtain $a(x - x_0)^2 = b(y - y_0)^2$, which is equivalent to the union of the line $\sqrt{a}(x - x_0) = \sqrt{b}(y - y_0)$ and the line $\sqrt{a}(x - x_0) = -\sqrt{b}(y - y_0)$ (in the plane given by the above equation). QED

Thus we have a really curved surface entirely made of absolutely straight lines! This has important applications in architecture: one can build a curved surface by using only straight metal rails. Hyperboloids of 1 sheet can be used to build towers...

Note that the lines we have found are never parallel to the plane given by $y = 0$. Thus such a line intersects this plane, and the resulting point is $(r, 0, ar^2)$ for some $r \in \mathbf{R}$. The two lines on our paraboloid passing through this point (which we just have found) are $L_r^{\pm}$ given by the equations $y = \pm\sqrt{a/b}(x - r)$, $z = ar(2x - r)$. We have found two one-parameter families of

lines entirely contained in our surface such that through every point of the surface passes exactly one line of each family.

Lemma. For any r and s the lines L_r^+ and L_s^- meet in exactly one point. The lines L_r^+ and L_s^+ never meet unless $r = s$. (easy)

One can rephrase this by saying that points of L_r^+ (for any r) parametrize lines of the family $\{L_s^-\}$, and vice versa.

This gives a bijective parametrization of the hyperbolic paraboloid by polynomials in two variables r and s:

$$x = \frac{1}{2}(r+s), \quad y = \frac{1}{2}\sqrt{\frac{a}{b}}(s-r), \quad z = ars.$$

With this parametrization the point (r, s) is mapped to the intersection point of L_r^+ and L_s^-.

Perhaps a more conceptual way of constructing the lines of a hyperbolic paraboloid is this: the equation $ax^2 - by^2 = z$ is satisfied by the points of the following two lines

$$\sqrt{a}x - \sqrt{b}y = rz, \sqrt{a}x + \sqrt{b}y = 1/r,$$

$$\sqrt{a}x + \sqrt{b}y = sz, \sqrt{a}x - \sqrt{b}y = 1/s,$$

for real parameters r and s. This can also be used on a hyperboloid of 1 sheet. Indeed, it is given by the equation $ax^2 - cz^2 = 1 - by^2$, $a, b, c > 0$, and this equation holds for any point of the following two lines:

$$\sqrt{a}x - \sqrt{c}z = r(1 - \sqrt{b}y), \sqrt{a}x + \sqrt{c}z = 1/r(1 + \sqrt{b}y),$$

$$\sqrt{a}x + \sqrt{c}z = s(1 - \sqrt{b}y), \sqrt{a}x - \sqrt{c}z = 1/s(1 + \sqrt{b}y).$$

These families can be used to prove the analogue of the previous theorem for hyperboloids of 1 sheet.

Note also that parabolic mirrors (elliptic paraboloids of round shape) are used from cars' lights to radiotelescopes, whenever one needs to focus a parallel beam of light (or vice versa).