Queen Mary
University of London

MTH6140 **Linear Algebra II**

Notes 6 **25th November 2010**

6 Quadratic forms

A lot of applications of mathematics involve dealing with quadratic forms: you meet them in statistics (analysis of variance) and mechanics (energy of rotating bodies), among other places. In this section we begin the study of quadratic forms.

6.1 Quadratic forms

For almost everything in this chapter, we assume that

> *the characteristic of the field K is not equal to* 2.

This means that $2 \neq 0$ in K, so that the element $1/2$ exists in K. Of our list of "standard" fields, this only excludes $\mathbb{F}_2$, the integers mod 2. (For example, in $\mathbb{F}_5$, we have $1/2 = 3$.)

A quadratic form is a function which, when written out in coordinates, is a polynomial in which every term has total degree 2 in the variables. For example,

$$Q(x,y,z) = x^2 + 4xy + 2xz - 3y^2 - 2yz - z^2$$

is a quadratic form in three variables.

We will meet a formal definition of a quadratic form later in the chapter, but for the moment we take the following.

Definition 6.1 A *quadratic form* in n variables $x_1, \ldots, x_n$ over a field K is a polynomial

$$\sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j$$

in the variables in which every term has degree two (that is, is a multiple of $x_i x_j$ for some i, j), and each A_{ij} belongs to K.

In the above representation of a quadratic form, we see that if $i \neq j$, then the term in $x_i x_j$ comes twice, so that the coefficient of $x_i x_j$ is $A_{ij} + A_{ji}$. We are free to choose any two values for A_{ij} and A_{ji} as long as they have the right sum; but we will always make the choice so that the two values are equal. That is, to obtain a term $cx_i x_j$, we take $A_{ij} = A_{ji} = c/2$. (This is why we require that the characteristic of the field is not 2.)

Any quadratic form is thus represented by a *symmetric* matrix A with (i, j) entry A_{ij} (that is, a matrix satisfying $A = A^\top$). *This is the third job of matrices in linear algebra:* **Symmetric matrices represent quadratic forms.**

We think of a quadratic form as defined above as being a function from the vector space K^n to the field K. It is clear from the definition that

$$Q(x_1, \ldots, x_n) = v^\top A v, \text{ where } v = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

Now if we change the basis for V, we obtain a different representation for the same function Q. The effect of a change of basis is a linear substitution $v = Pv'$ on the variables, where P is the transition matrix between the bases. Thus we have

$$v^\top A v = (Pv')^\top A (Pv') = (v')^\top (P^\top A P) v',$$

so we have the following:

Proposition 6.1 A basis change with transition matrix P replaces the symmetric matrix A representing a quadratic form by the matrix $P^\top AP$.

As for other situations where matrices represented objects on vector spaces, we make a definition:

Definition 6.2 Two symmetric matrices A, A' over a field K are *congruent* if $A' = P^\top AP$ for some invertible matrix P.

Proposition 6.2 Two symmetric matrices are congruent if and only if they represent the same quadratic form with respect to different bases.

Our next job, as you may expect, is to find a *canonical form* for symmetric matrices under congruence; that is, a choice of basis so that a quadratic form has a particularly simple shape. We will see that the answer to this question depends on the field over which we work. We will solve this problem for the fields of real and complex numbers.

6.2 Reduction of quadratic forms

Even if we cannot find a canonical form for quadratic forms, we can simplify them very greatly.

Theorem 6.3 Let Q be a quadratic form in n variables $x_1,\ldots,x_n$, over a field K whose characteristic is not 2. Then by a suitable linear substitution to new variables $y_1,\ldots,y_n$, we can obtain

$$Q = \alpha_1 y_1^2 + \alpha_2 y_2^2 + \cdots + \alpha_n y_n^2$$

for some $\alpha_1,\ldots,\alpha_n \in K$.

Proof Our proof is by induction on n. We call a quadratic form which is written as in the conclusion of the theorem *diagonal*. A form in one variable is certainly diagonal, so the induction starts. Now assume that the theorem is true for forms in $n-1$ variables. Take

$$Q(x_1,\ldots,x_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}x_ix_j,$$

where $A_{ij} = A_{ji}$ for $i \neq j$.

Case 1: Assume that $A_{ii} \neq 0$ for some i. By a permutation of the variables (which is certainly a linear substitution), we can assume that $A_{11} \neq 0$. Let

$$y_1 = x_1 + \sum_{i=2}^{n}(A_{1i}/A_{11})x_i.$$

Then we have

$$A_{11}y_1^2 = A_{11}x_1^2 + 2\sum_{i=2}^{n} A_{1i}x_1x_i + Q'(x_2,\ldots,x_n),$$

where Q' is a quadratic form in $x_2,\ldots,x_n$. That is, all the terms involving x_1 in Q have been incorporated into $A_{11}y_1^2$. So we have

$$Q(x_1,\ldots,x_n) = A_{11}y_1^2 + Q''(x_2,\ldots,x_n),$$

where Q'' is the part of Q not containing x_1 minus Q'.

By induction, there is a change of variable so that

$$Q''(x_2,\ldots,x_n) = \sum_{i=2}^{n} \alpha_i y_i^2,$$

and so we are done (taking $\alpha_1 = A_{11}$).

Case 2: All A_{ii} are zero, but $A_{ij} \neq 0$ for some $i \neq j$. Now

$$x_i x_j = \tfrac{1}{4}\left((x_i + x_j)^2 - (x_i - x_j)^2\right),$$

so taking $x'_i = \frac{1}{2}(x_i + x_j)$ and $x'_j = \frac{1}{2}(x_i - x_j)$, we obtain a new form for Q which does contain a non-zero diagonal term. Now we apply the method of Case 1.

Case 3: All A_{ij} are zero. Now Q is the zero form, and there is nothing to prove: take $\alpha_1 = \cdots = \alpha_n = 0$.

Example 6.1 Consider the quadratic form $Q(x,y,z) = x^2 + 2xy + 4xz + y^2 + 4z^2$. We have

$$(x+y+2z)^2 = x^2 + 2xy + 4xz + y^2 + 4z^2 + 4yz,$$

and so

$$\begin{aligned} Q &= (x+y+2z)^2 - 4yz \\ &= (x+y+2z)^2 - (y+z)^2 + (y-z)^2 \\ &= X^2 + Y^2 - Z^2, \end{aligned}$$

where $X = x+y+2z$, $Y = y-z$, $Z = y+z$. Otherwise said, the matrix representing the quadratic form, namely

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 1 & 0 \\ 2 & 0 & 4 \end{bmatrix}$$

is congruent to the diagonal matrix

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

How do we find an invertible matrix P such that $P^\top AP = A'$? Here is how:

If v is the vector consisting of the 'original' variables, so $v = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$, and v' is the vector consisting of 'new' variables, so $v' = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$, then P is defined by $v = Pv'$ (see the argument on page 2 of this chapter).

Now in the current example we have

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & -1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix},$$

so P is the *inverse* of the above matrix, in other words

$$P = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & -1 \\ 0 & 1 & 1 \end{bmatrix}^{-1}.$$

Thus any quadratic form can be reduced to the diagonal shape

$$\alpha_1 x_1^2 + \cdots + \alpha_n x_n^2$$

by a linear substitution. But this is still not a "canonical form for congruence". For example, if $y_1 = x_1/c$, then $\alpha_1 x_1^2 = (\alpha_1 c^2) y_1^2$. In other words, we can multiply any α_i by any factor which is a perfect square in K.

Over the complex numbers $\mathbb{C}$, every element has a square root. Suppose that $\alpha_1, \ldots, \alpha_r \neq 0$, and $\alpha_{r+1} = \cdots = \alpha_n = 0$. Putting

$$y_i = \begin{cases} (\sqrt{\alpha_i}) x_i & \text{for } 1 \leq i \leq r, \\ x_i & \text{for } r+1 \leq i \leq n, \end{cases}$$

we have

$$Q = y_1^2 + \cdots + y_r^2.$$

We will see later that r is an "invariant" of Q: however we do the reduction, we arrive at the same value of r.

Over the real numbers $\mathbb{R}$, things are not much worse. Since any positive real number has a square root, we may suppose that $\alpha_1, \ldots, \alpha_s > 0$, $\alpha_{s+1}, \ldots, \alpha_{s+t} < 0$, and $\alpha_{s+t+1}, \ldots, \alpha_n = 0$. Now putting

$$y_i = \begin{cases} (\sqrt{\alpha_i}) x_i & \text{for } 1 \leq i \leq s, \\ (\sqrt{-\alpha_i}) x_i & \text{for } s+1 \leq i \leq s+t, \\ x_i & \text{for } s+t+1 \leq i \leq n, \end{cases}$$

we get

$$Q = x_1^2 + \cdots + x_s^2 - x_{s+1}^2 - \cdots - x_{s+t}^2.$$

Again, we will see later that s and t don't depend on how we do the reduction. [This is the theorem known as *Sylvester's Law of Inertia*.]

6.3 Linear forms and dual space

Now we begin dealing with quadratic forms in a more abstract way. We begin with linear forms, that is, functions of degree 1. The definition is simple:

Definition 6.3 Let V be a vector space over K. A *linear form* on V is a linear map from V to K, where K is regarded as a 1-dimensional vector space over K: that is, it is a function from V to K satisfying

$$f(v_1+v_2)=f(v_1)+f(v_2), \qquad f(cv)=cf(v)$$

for all $v_1,v_2,v\in V$ and $c\in K$.

If $\dim(V)=n$, then a linear form is represented by a $1\times n$ matrix over K, that is, a *row vector* of length n over K. If $f=[a_1 \quad a_2 \quad \dots \quad a_n]$ represents a linear form, then for $v=[x_1 \quad x_2 \quad \dots \quad x_n]^\top$ we have

$$f(v)=[a_1 \quad a_2 \quad \dots \quad a_n]\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}=a_1x_1+a_2x_2+\cdots+a_nx_n.$$

Conversely, any row vector of length n represents a linear form on K^n.

Definition 6.4 Linear forms can be added and multiplied by scalars in the obvious way:

$$(f_1+f_2)(v)=f_1(v)+f_2(v), \qquad (cf)(v)=cf(v).$$

So they form a vector space, which is called the *dual space* of V and is denoted by V^*.

Not surprisingly, we have:

Proposition 6.4 If V is finite-dimensional, then so is V^*, and $\dim(V^*)=\dim(V)$.

Proof We begin by observing that, if $(v_1,\dots,v_n)$ is a basis for V, and $a_1,\dots,a_n$ are any scalars whatsoever, then there is a unique linear map f with the property that $f(v_i)=a_i$ for $i=1,\dots,n$. It is given by

$$f(c_1v_1+\cdots+c_nv_n)=a_1c_1+\cdots+a_nc_n,$$

in other words, it is represented by the row vector $[a_1 \quad a_2 \quad \dots \quad a_n]$, and its action on K^n is by matrix multiplication as we saw earlier.

Now let f_i be the linear map defined by the rule that

$$f_i(v_j) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Then $(f_1, \ldots, f_n)$ form a basis for V^*; indeed, the linear form f defined in the preceding paragraph is $a_1 f_1 + \cdots + a_n f_n$. This basis is called the *dual basis* of V^* corresponding to the given basis for V. Since it has n elements, we see that $\dim(V^*) = n = \dim(V)$.

We can describe the basis in the preceding proof as follows.

Definition 6.5 The *Kronecker delta* δ_{ij} for $i, j \in \{1, \ldots, n\}$ is defined by the rule that

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Note that δ_{ij} is the (i, j) entry of the identity matrix. Now, if $(v_1, \ldots, v_n)$ is a basis for V, then the *dual basis* for the dual space V^* is the basis $(f_1, \ldots, f_n)$ satisfying

$$f_i(v_j) = \delta_{ij}.$$

There are some simple properties of the Kronecker delta with respect to summation. For example,

$$\sum_{i=1}^{n} \delta_{ij} a_i = a_j$$

for fixed $j \in \{1, \ldots, n\}$. This is because all terms of the sum except the term $i = j$ are zero.

6.4 Change of basis

Suppose that we change bases in V from $B = (v_1, \ldots, v_n)$ to $B' = (v'_1, \ldots, v'_n)$, with transition matrix $P = P_{B,B'}$. How do the dual bases change? In other words, if $B^* = (f_1, \ldots, f_n)$ is the dual basis of B, and $(B')^* = (f'_1, \ldots, f'_n)$ the dual basis of B', then what is the transition matrix $P_{B^*,(B')^*}$? The next result answers the question.

Proposition 6.5 Let B and B' be bases for V, and B^* and $(B')^*$ the dual bases of the dual space. Then

$$P_{B^*,(B')^*} = \left(P_{B,B'}^{\top}\right)^{-1}.$$

Proof Use the notation from just before the statement of this Proposition. If $P = P_{B,B'}$ has (i,j) entry p_{ij}, and $Q = P_{B^*,(B')^*}$ has (i,j) entry q_{ij}, we have

$$\begin{aligned} v'_i &= \sum_{k=1}^{n} p_{ki} v_k, \\ f'_j &= \sum_{l=1}^{n} q_{lj} f_l, \end{aligned}$$

and so

$$\begin{aligned} \delta_{ij} &= f'_j(v'_i) \\ &= \left(\sum_{l=1}^{n} q_{lj} f_l\right)\left(\sum_{k=1}^{n} p_{ki} v_i\right) \\ &= \sum_{l=1}^{n}\sum_{k=1}^{n} q_{lj}\delta_{ij} p_{ki} \\ &= \sum_{k=1}^{n} q_{kj} p_{ki}. \end{aligned}$$

Now q_{kj} is the (j,k) entry of $Q^\top$, and so we have

$$I = Q^\top P,$$

whence $Q^\top = P^{-1}$, so that $Q = \left(P^{-1}\right)^\top = \left(P^\top\right)^{-1}$, as required.

6.5 Quadratic and bilinear forms

The formal definition of a quadratic form looks a bit different from the version we gave earlier, though it amounts to the same thing. First we define a bilinear form.

Definition 6.6 (a) Let $b : V \times V \to K$ be a function of two variables from V with values in K. We say that b is a *bilinear form* if it is a linear function of each variable when the other is kept constant: that is,

$$b(v, w_1 + w_2) = b(v, w_1) + b(v, w_2), \qquad b(v, cw) = cb(v, w),$$

with two similar equations involving the first variable, namely

$$b(v_1 + v_2, w) = b(v_1, w) + b(v_2, w), \qquad b(cv, w) = cb(v, w).$$

A bilinear form b is *symmetric* if $b(v,w) = b(w,v)$ for all $v, w \in V$.

(b) Let $Q : V \to K$ be a function. We say that Q is a *quadratic form* if

(i) $Q(cv) = c^2Q(v)$ for all $c \in K$, $v \in V$, and

(ii) the function b defined by

$$b(v,w) = Q(v+w) - Q(v) - Q(w)$$

is a bilinear form on V. (Note that the bilinear form b is symmetric!)

If we think of the prototype of a quadratic form as being the function x^2, then the first equation says $(cx)^2 = c^2x^2$, while the second has the form

$$(x+y)^2 - x^2 - y^2 = 2xy,$$

and $2xy$ is the prototype of a bilinear form: it is a linear function of x when y is constant, and *vice versa*.

Note that the formula

$$b(x,y) = Q(x+y) - Q(x) - Q(y)$$

(which is known as the *polarisation formula*) says that the bilinear form is determined by the quadratic form Q. Conversely, if we know the symmetric bilinear form b, then we have

$$2Q(v) = 4Q(v) - 2Q(v) = Q(v+v) - Q(v) - Q(v) = b(v,v),$$

so that $Q(v) = \frac{1}{2}b(v,v)$, and we see that the quadratic form is determined by the symmetric bilinear form. So these are equivalent objects.

6.6 Canonical forms for complex and real forms

Finally, in this section, we return to quadratic forms (or symmetric matrices) over the real and complex numbers, and find canonical forms under congruence. Recall that two symmetric matrices A and A' are congruent if $A' = P^\top AP$ for some invertible matrix P; as we have seen, this is the same as saying that they represent the same quadratic form relative to different bases.

Theorem 6.6 Any $n \times n$ complex symmetric matrix A is congruent to a matrix of the form

$$\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}$$

for some r. Moreover, $r = \text{rank}(A)$, and so if A is congruent to two matrices of this form then they both have the same value of r.

Proof We already saw that A is congruent to a matrix of this form. Moreover, if P is invertible, then so is $P^\top$, and so

$$r = \operatorname{rank}(P^\top AP) = \operatorname{rank}(A)$$

as claimed.

The next result is *Sylvester's Law of Inertia*.

Theorem 6.7 Any $n \times n$ real symmetric matrix A is congruent to a matrix of the form

$$\begin{bmatrix} I_s & O & O \\ O & -I_t & O \\ O & O & O \end{bmatrix}$$

for some s,t. Moreover, if A is congruent to two matrices of this form, then they have the same values of s and of t.

Proof Again we have seen that A is congruent to a matrix of this form. Arguing as in the complex case, we see that $s+t = \operatorname{rank}(A)$, and so any two matrices of this form congruent to A have the same values of $s+t$. Moreover, by restricting to a subspace on which A is invertible, we may assume without loss of generality that $s+t = n$.

Suppose that two different reductions give the values s,t and s',t' respectively, with $s+t = s'+t' = n$. Suppose (in order to obtain a contradiction) that $s < s'$. Now let Q be the quadratic form represented by A. Then we are told that there are linear functions $y_1,\ldots,y_n$ and $z_1,\ldots,z_n$ of the original variables $x_1,\ldots,x_n$ of Q such that

$$Q = y_1^2 + \cdots + y_s^2 - y_{s+1}^2 - \cdots - y_n^2 = z_1^2 + \cdots + z_{s'}^2 - z_{s'+1}^2 - \cdots - z_n^2.$$

Now consider the equations

$$y_1 = 0, \ldots, y_s = 0, z_{s'+1} = 0, \ldots z_n = 0$$

regarded as linear equations in the original variables $x_1,\ldots,x_n$. The number of equations is $s+(n-s') = n-(s'-s) < n$. According to a lemma from much earlier in the course (we used it in the proof of the Exchange Lemma!), the equations have a non-zero solution. That is, there are values of $x_1,\ldots,x_n$, not all zero, such that the variables $y_1,\ldots,y_s$ and $z_{s'+1},\ldots,z_n$ are all zero.

Since $y_1 = \cdots = y_s = 0$, we have for these values

$$Q = -y_{s+1}^2 - \cdots - y_n^2 < 0.$$

But since $z_{s'+1} = \cdots = z_n = 0$, we also have

$$Q = z_1^2 + \cdots + z_{s'}^2 > 0.$$

But this is a contradiction. So we cannot have $s < s'$. Similarly we cannot have $s' < s$ either. So we must have $s = s'$, as required to be proved.

We saw that $s+t$ is the rank of A.

Definition 6.7 The number $s-t$ is known as the *signature* of A.

Of course, both the rank and the signature are independent of how we reduce the matrix (or quadratic form); and if we know the rank and signature, we can easily recover s and t.