MATH 325 Q1: LINEAR ALGEBRA III, PART 11

Bilinear forms

We will assume in this chapter that $F = \mathbb{R}$ and V is a vector space (of finite dimension) over $\mathbb{R}$. We are going to develop a theory of "inner products" on V without assuming axiom (P).

Definition. *A bilinear form on V is a map*

$$\langle -, - \rangle : V \times V \to F$$

satisfying

- **(L1)** $\langle \alpha_1 \cdot v_1 + \alpha_2 \cdot v_2, w \rangle = \alpha_1 \cdot \langle v_1, w \rangle + \alpha_2 \cdot \langle v_2, w \rangle$ *for all* $\alpha_1, \alpha_2 \in F$ *and all* $v_1, v_2, w \in V$*; and*
- **(L2)** $\langle w, \alpha_1 v_1 + \alpha_2 v_2 \rangle = \alpha_1 \cdot \langle w, v_1 \rangle + \alpha_2 \cdot \langle w, v_2 \rangle$ *for all* $\alpha_1, \alpha_2 \in F$ *and all* $v_1, v_2, w \in V$.

It is called symmetric if axiom **(H)** *holds, i.e.* $\langle v, w \rangle = \langle w, v \rangle$ *for all vectors* $v, w \in V$.

Example. Let $V = \mathbb{R}^n$. Then the Euclidean inner product

$$\langle x, y \rangle = \sum x_i y_i$$

where x_i, y_i are coordinates of the vectors x, y is a symmetric bilinear form.

Matrix of a bilinear form

Let $\langle -, - \rangle$ be a bilinear form on V. Fix a basis $\{v_1, \ldots, v_n\}$ of V. Set $b_{ij} = \langle v_i, v_j \rangle$ and let $B = (b_{ij})$.

Definition. *The matrix B is called the matrix of the bilinear form $\langle -, - \rangle$ in the basis $\{v_1, \ldots, v_n\}$.*

Remark. *Note that this matrix determines the bilinear form uniquely. Namely, if $x = \sum x_i v_i$ and $y = \sum y_i v_i$ then it is easy to check that*

$$\langle x, y \rangle = (x_1 \ldots x_n) \cdot B \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

Remark. *It is obvious that $\langle -, - \rangle$ is symmetric if and only if B is symmetric, i.e. $B^T = B$. In this chapter we will consider only symmetric bilinear forms so that $b_{ij} = b_{ji}$.*

Example. Let $\langle -, - \rangle$ be the Euclidean bilinear form on $\mathbb{R}^n$. Then its matrix B in the standard basis $\{e_1, \ldots, e_n$ is $B = \mathrm{Id}_n$.

Remark. *Our discussion shows that fixing a basis of V we have a natural one-to-one correspondence between the set of (symmetric) bilinear forms on n-dimensional vector space V and symmetric matrices of size $n \times n$.*

Base change

Let $\mathcal{B} = \{v_1, \ldots, v_n\}$ and $\mathcal{B}' = \{v_1', \ldots, v_n'\}$ be two bases of V. Then a bilinear form $\langle -, - \rangle$ has the matrix B in the first basis and the matrix B' in the second basis. We now want to address a natural question how B and B' are related to each other.

Recall that $b_{ij} = \langle v_i, v_j \rangle$ and similarly $b'_{ij} = \langle v_i', v_j' \rangle$. Let $S = (s_{ij})$ be the base change matrix, so that

$$v_i' = \sum_k s_{ki} v_k = S \cdot \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}.$$

It follows that

$$\begin{array}{rclcl} b'_{ij} & = & \langle v_i', v_j' \rangle & = & \langle \sum_k s_{ki} v_k, \sum_m s_{mj} v_m \rangle \\ & & & = & \sum_{k,m} s_{ki} s_{mj} \langle v_k, v_m \rangle \\ & & & = & \sum_{k,m} s_{ki} s_{mj} b_{km} \\ & & & = & (S^T B S)_{ij}. \end{array}$$

Thus, we proved

Lemma. *One has $B' = S^T B S$ where S is the base change matrix.*

Quadratic forms

Definition. *Let $\langle -, - \rangle$ be a (symmetric) bilinear form on V. The map $q : V \to F$ given by $q(x) = \langle x, x \rangle$ is called the quadratic form associated with $\langle -, - \rangle$. The corresponding matrix B is also called the matrix of q.*

Remark. Note that the quadratic form $q : V \to F$ allows us to restore the bilinear form $\langle -, - \rangle$ uniquely. Namely,

$$\langle v, w \rangle = \frac{1}{2}(q(v + w) - q(v) - q(w)).$$

Indeed, using properties of linearity and symmetry of our bilinear form we have

$$q(v+w) - q(v) - q(w) = \langle v+w, v+w\rangle - \langle v, v\rangle - \langle w, w\rangle = 2\langle v, w\rangle.$$

Algebraic definition of quadratic forms

We now pass to an algebraic description of quadratic forms. Let us fix a basis $\{v_1, \ldots, v_n\}$ of V. Then q is given by

$$q(x) = \langle x, x\rangle = (x_1 \ldots x_n) \cdot B \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \sum_{i,j} x_i b_{ij} x_j$$

where $x_1, \ldots, x_n$ are coordinates of x in the given basis.

This formula leads us to another, more algebraic way, to say that a quadratic form is a homogeneous polynomial of degree 2, i.e. that q is a polynomial in n variables $x_1, x_2, \ldots, x_n$ having only terms of degree 2. That means that only terms ax_i^2 and cx_jx_k are allowed. Given such polynomial we can restore uniquely the matrix B (it is called the matrix associated to q) and hence we can restore the bilinear form.

Example. If $q(x) = x_1^2 + 3x_2^2 + 5x_3^2 + 4x_1x_2 - 7x_2x_3$ then

$$B = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 3 & -3.5 \\ 0 & -3.5 & 5 \end{pmatrix}.$$

Example. Let $V = \mathbb{R}^n$ and let $\langle -, -\rangle$ has the matrix

$$B = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}.$$

Then q is given by

$$q(x) = (x_1 x_2) \cdot \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1^2 + 4x_1x_2 + x_2^2.$$

Orthogonality

Definition. *Let $\langle -, -\rangle$ be a (symmetric) bilinear form on V. Vectors $v, w \in V$ are called orthogonal if*

$$\langle v, w\rangle = \langle w, v\rangle = 0.$$

Notation: $v \perp w$.

Definition. *Let $W \subset V$ be a vector subspace. The orthogonal complement $W^\perp$ consists of all vectors $v \in V$ such that v is orthogonal to all vectors $w \in W$.*

Lemma. *The set $W^\perp \subset V$ is a vector subspace in V.*

Proof. It is similar to the proof for inner products. □

Definition. *The vector subspace $V^\perp$ is called the radical of $(V, \langle -, - \rangle)$.*

Definition. *A bilinear form is called non-degenerate if $V^\perp = 0$.*

One can show that a bilinear form is non-degenerate if and only if the corresponding matrix B is non-degenerate. We will consider below only non-degenerate bilinear forms. For later use we now state and prove the following assertion.

Lemma. *Let $\langle -, - \rangle$ be a bilinear form on V. Then V has a basis consisting of orthogonal to each other vectors.*

Proof. We mimic the proof of a similar assertion for inner products. If $\langle -, - \rangle$ is zero, i.e. $\langle v, w \rangle = 0$ for all vectors $v, w \in V$ we can take any basis of V. Otherwise there is a vector $v_1 \in V$ such that $\langle v, v \rangle \neq 0$. Consider a linear map

$$\phi : V \longrightarrow F, \quad x \to \langle x, v \rangle.$$

Since $\phi(v) \neq 0$ this map is surjective. Hence $\dim(\operatorname{Ker} \phi) = n - 1$. Note that $\operatorname{Ker} \phi = v_1^\perp$, i.e. all vectors in $\operatorname{Ker} \phi$ are orthogonal to v_1. By induction, $\operatorname{Ker} \phi$ has an orthogonal basis, say $v_2, \ldots, v_n$. Then $\{v_1, v_2, \ldots, v_n\}$ is a basis of V and all vectors in this basis are orthogonal to each other. We are done. □

Diagonalization of quadratic forms

If we are given a general complicated quadratic form $q = \sum b_{ij} x_i x_j$ we want to simplify it as much as possible, for example to make it diagonal. The standard way of doing that is the change of variables.

Orthogonal diagonalization. Assume we are given a quadratic form $q(x) = \langle x, x \rangle$ on V. here $x \in V$. Take any basis $\{v_1, \ldots, v_n\}$ of V. In this basis q has a presentation $q(x) = \sum b_{ij} x_i x_j$ where $b_{ij} = \langle v_i, v_j \rangle$ and $x_1, \ldots, x_n$ are coordinates of $x \in V$ in the chosen basis. If we choose another basis $\{v'_1, \ldots, v'_n\}$ of V then q has presentation $q = \sum b'_{ij} y_i y_j$ where $b'_{ij} = \langle v'_i, v'_j \rangle$ and $y_1, \ldots, y_n$ are coordinates of x in the new basis, i.e. $x = \sum y_i v_i'$. If S is the base change matrix we know that $B' = S^T BS$ and the vector

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

can be written as $y = S^{-1}x$ or $x = Sy$. So, we want to find an invertible matrix S such that the matrix $S^T BS$ is diagonal. As we proved before there is an orthogonal basis of V. So if $\{v_1', \dots, v_n'\}$ is any such basis then the matrix B' is diagonal, say $B' = \mathrm{diag}(a_1, \dots, a_n)$, hence in this basis q is of the form $q = \sum a_i y_i^2$.

Example. Let $V = \mathbb{R}^2$ and $q(x) = 2x_1^2 + 2x_1x_2 + 2x_2^2$. Its matrix is

$$B = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

So the bilinear form in the standard basis is given by

$$\langle u, v\rangle = 2u_1v_1 + u_1v_2 + u_2v_1 + 2u_2v_2.$$

Here

$$u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \quad \text{and} \quad v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}.$$

One can easily checks that the following two vectors

$$e_1' = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad e_2' = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

are orthogonal to each other and that

$$\langle e_1', e_1'\rangle = 6, \quad \langle e_2', e_2'\rangle = 2.$$

Therefore q in the basis $\{e_1', e_2'\}$ is of the form

$$q = 6y_1^2 + 2y_2^2$$

where the variables y_1, y_2 are given by $S^{-1}x$ and S is the base change matrix whose columns are e_1', e_2'.

Diagonalization by completion of squares. Orthogonal diagonalization involves computing an orthogonal basis, so it may be time consuming to do it without computers for large n. We now discuss another method which is based on completion by squares.

Let us again consider the quadratic form of two variables,

$$q(x) = 2x_1^2 + 2x_1x_2 + 2x_2^2$$

(it is the same quadratic form as in the above example). Since

$$2(x_1 + \frac{1}{2}x_2)^2 = 2(x_1^2 + x_1x_2 + \frac{1}{4}x_2^2)$$

we get

$$2x_1^2 + 2x_1x_2 + 2x_2^2 = 2\left(x_1 + \frac{1}{2}x_2\right)^2 + \frac{3}{2}x_2^2 = 2y_1^2 + \frac{3}{2}y_2^2$$

where $y_1 = x_1 + \frac{1}{2}x_2$ and $y_2 = x_2$.

The same method can be applied to quadratic form of more than 2 variables. Let us consider, for example, a form $q(x)$ in $\mathbb{R}^3$,

$$q(x) = x_1^2 - 6x_1x_2 + 4x_1x_3 - 6x_2x_3 + 8x_2^2 - 3x_3^2.$$

Considering all terms involving the first variable x_1 (the first 3 terms in this case), let us pick a full square or a multiple of a full square which has exactly these terms (plus some other terms).

Since

$$(x_1 - 3x_2 + 2x_3)^2 = x_1^2 - 6x_1x_3 + 4x_1x_3 - 12x_2x_3 + 9x_2^2 + 4x_3^2$$

we can rewrite the quadratic form as

$$(x_1 - 3x_2 + 2x_3)^2 - x_2^2 + 6x_2x_3 - 7x_3^2.$$

Note, that the expression $-x_2^2 + 6x_2x_3 - 7x_3^2$ involves only variables x_2 and x_3. Since

$$-(x_2 - 3x_3)^2 = -(x_2^2 - 6x_2x_3 + 9x_3^2) = -x_2^2 + 6x_2x_3 - 9x_3^2$$

we have

$$-x_2^2 + 6x_2x_3 - 7x_3^2 = -(x_2 - 3x_3)^2 + 2x_3^2.$$

Thus we can write the quadratic form q as

$$q(x) = (x_1 - 3x_2 + 2x_3)^2 - (x_2 - 3x_3)^2 + 2x_3^2 = y_1^2 - y_2^2 + 2y_3^2$$

where

$$y_1 = x_1 - 3x_2 + 2x_3; \quad y_2 = x_2 - 3x_3; \quad y_3 = x_3.$$

Finally, let us address the question that an attentive reader is probably already asking: what to do if at some point we do have a product of two variables, but no corresponding squares? For example, how to diagonalize the form $q(x) = x_1x_2$? The answer follows immediately from the identity

$$4x_1x_2 = (x_1 + x_2)^2 - (x_1 - x_2)^2;$$

which gives us the representation

$$q = y_1^2 - y_2^2$$

where

$$y_1 = (x_1 + x_2)/2; \quad y_2 = (x_1 - x_2)/2.$$

Silvester's law of inertia

As we discussed above, there are many ways to diagonalize a quadratic form. Note, that a resulting diagonal matrix is not unique. For example, if we got a quadratic form with a diagonal matrix

$$A = \operatorname{diag}\{a_1, a_2, \ldots, a_n\}$$

we can take a diagonal matrix

$$S = \operatorname{diag}\{s_1, s_2, \ldots, s_s n\}$$

and transform A to

$$S^T A S = \operatorname{diag}\{a_1 s_1^2, a_2 s_2^2, \ldots, a_n s_n^2\}.$$

This transformation changes the diagonal entries of A. However, it does not change the signs of the diagonal entries. And this is always the case! Namely, the famous Silvester's Law of Inertia states that:

Theorem. *Let $q = \sum a_{ij} x_i x_j$ be a quadratic form with a matrix A. Then for any of its diagonalization $D = S^T A S$ the number of positive (negative, zero) diagonal entries of D depends only on q, but not on a particular choice of diagonalization.*

The idea of the proof of the Silvester's Law of Inertia is to express the number of positive (negative, zero) diagonal entries of a diagonalization $D = S^T A S$ in terms of A, not involving S or D. We will need the following definition.

Definition. *Let $q = \sum a_{ij} x_i x_j$ be a quadratic form on a vector space $V = \mathbb{R}^n$. The subspace $W \subset V$ is called positive (resp. negative, neutral) if $q(x) > 0$ (resp. $q(x) < 0, q(x) = 0$) for all nonzero vectors $x \in W$.*

Sometimes, to emphasize the role of A we will say A-positive (A negative, A-neutral). The Silvester's Law of inertia follows immediately from the following result.

Theorem. *Let A be a symmetric matrix, and let $D = S^T A S$ be its diagonalization by an invertible matrix S. Then the number of positive (resp. negative) diagonal entries of D coincides with the maximal dimension of an A-positive (resp. A-negative) subspace.*

The above theorem says that if r_+ is the number of positive diagonal entries of D, then there exists an A-positive subspace $W \subset \mathbb{R}^n$ of dimension r_+, but it is impossible to find a positive subspace E with $\dim E > r_+$. We will need the following lemma, which can be considered a particular case of the above theorem.

Lemma. *Let D be a diagonal matrix $D = \mathrm{diag}\{\lambda_1, \ldots, \lambda_n\}$. Then the number of positive (resp. negative) diagonal entries of D coincides with the maximal dimension of a D-positive (resp. D-negative) subspace.*

Proof. By rearranging the standard basis in $\mathbb{R}^n$ (changing the numeration) we can always assume without loss of generality that the positive diagonal entries of D are the first r_+ diagonal entries. Consider the subspace E_+ spanned by the first r_+ coordinate vectors $e_1, e_2, \ldots, e_{r_+}$. Clearly E_+ is a D-positive subspace, and $\dim E_+ = r_+$. It remains to show that for any other A-positive subspace E we have $\dim E \leq r_+$.

Consider the projection $p : \mathbb{R}^n \to E_+ \subset \mathbb{R}^n$ given by

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \longrightarrow \begin{pmatrix} x_1 \\ \vdots \\ x_{r_+} \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Let $p' : E \to E_+ \subset \mathbb{R}^n$ be the restriction $p|_E$ of p at E.

Claim: $\operatorname{Ker} p' = 0$. Indeed, let

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

be in the $\operatorname{Ker} p'$. Then $x_1 = x_2 = \ldots = x_{r_+} = 0$. Therefore,

$$q(x) = \sum_{i=r_+ +1}^{n} \lambda_i x_i^2 \leq 0.$$

But E is D-positive, hence $x = 0$, as claimed.

Thus,

$$\dim E = \dim \operatorname{Im} p' \leq \dim E_+ = r_+.$$

$\square$

We now are ready to prove the above theorem. Let $D = S^T A S$ be a diagonalization of q. Since for any vector $x \in \mathbb{R}^n$ one has

$$x^T D x = x^T (S^T A S) x = (Sx)^T A (SX)$$

it follows that for any D-positive subspace E, the subspace SE is an A-positive subspace. The same equlity implies that for any A-positive subspace F the subspace $S^{-1}F$ is D-positive. Since S and S^{-1} are invertible transformations $\dim E = \dim SE$ and $\mathrm{dim} F = \mathrm{dim} S^{-1}F$.

Therefore, for any D-positive subspace E we can find an A-positive subspace (namely SE) of the same dimension, and vice versa: for any A-positive subspace F we can find a D-positive subspace (namely $S^{-1}F$) of the same dimension. Therefore the maximal possible dimensions of a A-positive and a D-positive subspace coincide, and the theorem is proved.