LECTURE 20

Quadratic forms

20.1 Bilinear forms

Underlying our discussion on adjoint maps is the concept of bilinear form. We have seen how adjoint maps are defined relative to inner products, which in turn are symmetric and positive-definite bilinear forms (see Definition (5.1) and the terminology following it). Let us provide this definition for the case where the two arguments are from the space vector space V.

Definition 20.1 Let V denote a vector space over a field $\mathbb{F}$. A **bilinear form** is a function of two arguments $\mathcal{B}(\cdot,\cdot) : V \times V \to \mathbb{F}$, which is linear in each argument: for all $a, b \in \mathbb{F}$ and $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$,

- $\mathcal{B}(a\mathbf{u} + b\mathbf{v}, \mathbf{w}) = a\mathcal{B}(\mathbf{u}, \mathbf{w}) + b\mathcal{B}(\mathbf{v}, \mathbf{w})$,
- $\mathcal{B}(\mathbf{u}, a\mathbf{v} + b\mathbf{w}) = a\mathcal{B}(\mathbf{u}, \mathbf{v}) + b\mathcal{B}(\mathbf{u}, \mathbf{w})$.

We say $\mathcal{B}(\cdot,\cdot)$ is

- **symmetric** if $\mathcal{B}(\mathbf{v}, \mathbf{w}) = \mathcal{B}(\mathbf{w}, \mathbf{v})$;
- **anti-symmetric** (or **skew-symmetric**) if $\mathcal{B}(\mathbf{v}, \mathbf{w}) = -\mathcal{B}(\mathbf{w}, \mathbf{v})$.

The following observation is evident.

Proposition 20.1 Let $\mathcal{B}(\cdot,\cdot)$ be a bilinear form on V. Then its restriction to $U \le V$ is also a bilinear form.

In the following, we will restrict our attention to the case where $\mathbb{F} = \mathbb{R}$.
Any general (non-symmetric) bilinear form can be written uniquely as the sum of two bilinear forms, one symmetric and one skew-symmetric.

Proposition 20.2 Let $\mathcal{B}(\cdot,\cdot) : V \times V \to \mathbb{R}$ be a bilinear form on a real vector space V. Then we can uniquely write for all $\mathbf{v}, \mathbf{w} \in V$

$$\mathcal{B}(\mathbf{v}, \mathbf{w}) = \mathcal{B}_{\text{sym}}(\mathbf{v}, \mathbf{w}) + \mathcal{B}_{\text{skew}}(\mathbf{v}, \mathbf{w}),$$

where

$$\mathcal{B}_{\text{sym}}(\mathbf{v}, \mathbf{w}) = \frac{1}{2}\left(\mathcal{B}(\mathbf{v}, \mathbf{w}) + \mathcal{B}(\mathbf{w}, \mathbf{v})\right), \qquad \mathcal{B}_{\text{skew}}(\mathbf{v}, \mathbf{w}) = \frac{1}{2}\left(\mathcal{B}(\mathbf{v}, \mathbf{w}) - \mathcal{B}(\mathbf{w}, \mathbf{v})\right).$$

Proof. The 'decomposition' into two bilinear forms can be verified readily. The uniqueness claim follows by contradiction. ■

Just as was the case with linear maps, a bilinear form has a matrix representation corresponding to a given choice of basis.

Definition 20.2 — Gram matrix. Let $\mathcal{B}(\cdot,\cdot) : V \times V \to \mathbb{R}$ be a bilinear form on V, where V is a real vector space with basis $B = \{\mathbf{v}_1, \ldots, v_n\}$. The matrix representation of $\mathcal{B}$ relative to B is known as the **Gram matrix** and is given by

$$G_{ij} = \mathcal{B}(\mathbf{v}_i, \mathbf{v}_j), \qquad i, j = 1, \ldots, n.$$

With the above definition in place, we note that any evaluation of $\mathcal{B}$ can be identified as a vector-matrix-vector product. To see this, let

$$\mathbf{v} = \sum_{i=1}^{n} x_i \mathbf{v}_i, \quad \mathbf{w} = \sum_{j=1}^{n} y_j \mathbf{v}_j.$$

Then, using the bilinear properties of $\mathcal{B}$, we get

$$\mathcal{B}(\mathbf{v}, \mathbf{w}) = \mathcal{B}\left(\sum_{i=1}^{n} x_i \mathbf{v}_i, \sum_{j=1}^{n} y_j \mathbf{v}_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i \mathcal{B}(\mathbf{v}_i, \mathbf{v}_j) y_j = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i G_{ij} y_j = \mathbf{x}^T G \mathbf{y}.$$

As a corollary to Proposition 20.2, we have the following result,

Corollary 20.3 Let $G \in \mathbb{R}^{n\times n}$ denote the Gram matrix for a bilinear form $\mathcal{B}$ on V relative to some basis B of V. Then the Gram matrices for $\mathcal{B}_{\text{sym}}$ and $\mathcal{B}_{skew}$ are

$$G_{\text{sym}} = \frac{1}{2}(G + G^T), \qquad G_{\text{skew}} = \frac{1}{2}(G - G^T),$$

so that

$$G = G_{\text{sym}} + G_{\text{skew}} = \frac{1}{2}(G + G^T) + \frac{1}{2}(G - G^T).$$

(R) The above result is just an algebraic result which can be verified immediately for any square matrix. However, it is reassuring to identify the parallel between bilinear forms and Gram matrices.

Recall that a bilinear form is called **positive-definite** if $\mathcal{B}(\mathbf{v}, \mathbf{v}) > 0$ for all $\mathbf{v} \neq \mathbf{0}$. We extend this terminology to matrices.

Definition 20.3 — Matrix definiteness. A matrix $G \in \mathbb{R}^{n\times n}$ is said to be definite if for all $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$

- $\mathbf{x}^T G \mathbf{x} > 0$ (positive-definite);
- $\mathbf{x}^T G \mathbf{x} < 0$ (negative-definite).

A matrix $G \in \mathbb{R}^{n\times n}$ is said to be semi-definite if for all $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$

- $\mathbf{x}^T G \mathbf{x} \geq 0$ (positive-semidefinite);
- $\mathbf{x}^T G \mathbf{x} \leq 0$ (negative-semidefinite).

The matrix is said to be indefinite if it is not definite.

(R) The above definition is extended to include general non-symmetric matrices. Note that some textbooks require the matrix G to be symmetric in order for the above terminology to apply.

Proposition 20.4 Symmetry and definiteness of a Gram matrix are preserved under a change of basis.

Proof. Exercise. ■

20.2 Quadratic forms

We have seen how, in addition to symmetry, positive-definiteness in a bilinear form induces an inner product, which in turn allows us to define a norm on V. Another quantity induced by a symmetric bilinear form is a so-called quadratic form.

Definition 20.4 — Quadratic form. Let $\mathcal{B} : V \times V \to \mathbb{R}$, be a symmetric bilinear form, where V is a real vector space. The quadratic form induced by $\mathcal{B}$ is the real-valued function (or form) $\mathcal{Q} : V \to \mathbb{R}$ defined via

$$\mathcal{Q}(\mathbf{v}) = \mathcal{B}(\mathbf{v}, \mathbf{v}).$$

There is in fact a one-to-one correspondence between a bilinear form and the quadratic form it induces: given a quadratic form q, we can recover the bilinear form via the relation

$$\mathcal{B}(\mathbf{v}, \mathbf{w}) = \frac{1}{2}\left(\mathcal{Q}(\mathbf{v}+\mathbf{w}) - \mathcal{Q}(\mathbf{v}) - \mathcal{Q}(\mathbf{w})\right).$$

The matrix representation of $\mathcal{Q}$ is given by

$$\mathcal{Q}(\mathbf{v}) := q(\mathbf{x}) := \mathbf{x}^T G \mathbf{x},$$

where G is the symmetric Gram matrix associated with $\mathcal{B}$ relative to a basis B of V and $\mathbf{x}$ is the vector of coordinates of $\mathbf{v}$ in the same basis.
Quadratic forms arise frequently in mathematical sciences. We provide below two examples (and note that the Linear Programming part of this module will provide a third!).

20.2.1 Eigenvalue representation

Let $A \in \mathbb{R}^{n\times n}$ (formerly G) denote a symmetric matrix. By the Real Spectral Theorem, A has real eigenpairs. This allows us to provide another representation of its eigenvalues:

$$A\mathbf{x} = \lambda\mathbf{x} \iff \mathbf{x}^T A\mathbf{x} = \lambda\mathbf{x}^T\mathbf{x} \iff \lambda = \frac{\mathbf{x}^T A\mathbf{x}}{\mathbf{x}^T\mathbf{x}}.$$

The expression on the right is known as the **Rayleigh quotient of A at $\mathbf{x}$**, denoted by $R_A(\mathbf{x})$. Note that since $\mathbf{x}^T\mathbf{x} = \|\mathbf{x}\|^2$, we can introduce the unit vector $\widehat{\mathbf{x}} := \mathbf{x}/\|\mathbf{x}\|$, so that the eigenvalue λ becomes the evaluation of a quadratic form:

$$\lambda = \frac{\mathbf{x}^T A\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \widehat{\mathbf{x}}^T A\widehat{\mathbf{x}}.$$

We can provide a more detailed characterisation of eigenvalues in terms of quadratic forms. In the following we use the notation

$$\lambda_{\min}(A) = \min \operatorname{sp} A, \quad \lambda_{\max}(A) = \max \operatorname{sp} A.$$

Proposition 20.5 Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Then

$$\lambda_{\min}(A) = \min_{\mathbf{x}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \frac{\mathbf{x}^T A\mathbf{x}}{\mathbf{x}^T\mathbf{x}}, \quad \lambda_{\max}(A) = \max_{\mathbf{x}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \frac{\mathbf{x}^T A\mathbf{x}}{\mathbf{x}^T\mathbf{x}}.$$

Proof. Let $\mathbf{q}_1, \dots, \mathbf{q}_n \in \mathbb{R}^n$ denote the orthonormal eigenvectors of A. Since they form a basis, we can write any $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ as

$$\mathbf{x} = c_1\mathbf{q}_1 + c_2\mathbf{q}_2 + \cdots + c_n\mathbf{q}_n.$$

Recall now that the spectral decomposition of A has the form

$$A = \lambda_1\mathbf{q}_1\mathbf{q}_1^T + \lambda_2\mathbf{q}_2\mathbf{q}_2^T + \cdots + \lambda_n\mathbf{q}_n\mathbf{q}_n^T,$$

We find

$$\mathbf{x}^T A\mathbf{x} = (c_1\mathbf{q}_1 + \cdots + c_n\mathbf{q}_n)^T(\lambda_1\mathbf{q}_1\mathbf{q}_1^T + \cdots + \lambda_n\mathbf{q}_n\mathbf{q}_n^T)(c_1\mathbf{q}_1 + \cdots + c_n\mathbf{q}_n) = \lambda_1 c_1^2 + \cdots + \lambda_n c_n^2$$

and

$$\mathbf{x}^T\mathbf{x} = (c_1\mathbf{q}_1 + \cdots + c_n\mathbf{q}_n)^T(c_1\mathbf{q}_1 + \cdots + c_n\mathbf{q}_n) = c_1^2 + \cdots + c_n^2.$$

Hence,

$$\max_{\mathbf{x}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \frac{\mathbf{x}^T A\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \max_{\mathbf{x}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \frac{\lambda_1 c_1^2 + \cdots + \lambda_n c_n^2}{c_1^2 + \cdots + c_n^2} \le \frac{\lambda_{\max}(A)(c_1^2 + \cdots + c_n^2)}{c_1^2 + \cdots + c_n^2} = \lambda_{\max}(A).$$

Since $\lambda_{\max}(A)$ is an eigenvalue of A, the maximum is attained when $\mathbf{x}$ is the eigenvector corresponding to $\lambda_{\max}(A)$. The expression for $\lambda_{\min}(A)$ is obtained similarly. ■

R Note that the above expressions for the extreme eigenvalues of A are quadratic forms evaluated at the corresponding unit eigenvectors:

$$\lambda_{\min}(A) = \min_{\mathbf{x}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \frac{\mathbf{x}^T A\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \min_{\|\mathbf{q}\|=1} \mathbf{q}^T A\mathbf{q}, \quad \lambda_{\max}(A) = \max_{\mathbf{x}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \frac{\mathbf{x}^T A\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \max_{\|\mathbf{q}\|=1} \mathbf{q}^T A\mathbf{q}.$$

R The results in the above proposition are selected from a more detailed version which provides expressions for all the eigenvalues of A (this is the Courant-Fischer-Weyl minimax theorem).

20.2.2 Conics and quadrics

Another application of quadratic form arises in geometry, in the classification of conic sections and quadric surfaces. The general formula for such objects is

$$\text{quadratic form} \quad + \quad \text{linear form} \quad = \quad \text{constant}.$$

This expression can be viewed as a constraint (in the form of a quadratic equation) on the points that describe a certain locus in 2D or 3D. To classify such objects, we use this generic form; in particular, the eigenvalues of the Gram matrix provide the required criteria. We illustrate this for three familiar conics in the examples below.

Example 20.1 Consider the shifted conics below

$$\mathcal{C}_\pm := \left\{(x,y)\in\mathbb{R}^2 : \frac{(x-x_0)^2}{a^2} \pm \frac{(y-y_0)^2}{b^2} = 1\right\}.$$

Note that $\mathcal{C}_+$ is an ellipse, while $\mathcal{C}_-$ is a hyperbola. Both these conics can be written as indicated above:

$$q(\mathbf{x}) + \ell(\mathbf{x}) = 1 - q(\mathbf{x}_0),$$

where $q(\mathbf{x}) = \mathbf{x}^T G\mathbf{x}$, $\ell(x) = \mathbf{c}^T\mathbf{x}$ with

$$G = \begin{bmatrix} 1/a^2 & 0 \\ 0 & \pm 1/b^2 \end{bmatrix}, \quad \mathbf{c} = -2\begin{bmatrix} x_0/a^2 \\ \pm y_0/b^2 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}, \quad \mathbf{x}_0 = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}.$$

Let us include one more example.

Example 20.2 Consider the shifted parabola

$$\mathcal{C} = \left\{(x, y) \in \mathbb{R}^2 : \frac{(x - x_0)^2}{a^2} - (y - y_0) = 0\right\}.$$

This expression can also be given a representation involving a quadratic and a linear form:

$$q(\mathbf{x}) + \ell(\mathbf{x}) = y_0 - q(\mathbf{x}_0),$$

where $q(\mathbf{x}) = \mathbf{x}^T G\mathbf{x}$, $\ell(x) = \mathbf{c}^T\mathbf{x}$ with

$$G = \begin{bmatrix} 1/a^2 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{c} = -\begin{bmatrix} 2x_0/a^2 \\ 1 \end{bmatrix},$$

We distinguish the following classification based on the eigenvalues of G:

- $\lambda_1 > 0, \lambda_2 > 0$: ellipse
- $\lambda_1 > 0, \lambda_2 < 0$: hyperbola
- $\lambda_1 > 0, \lambda_2 = 0$: parabola

Let us consider now the case of a general quadratic equation, where the Gram matrix is not diagonal. Given that G is symmetric, we can use its eigenvalue decomposition to manipulate the definition of the conic:

$$q(\mathbf{x}) + \ell(\mathbf{x}) = \mathbf{x}^T G\mathbf{x} + \mathbf{c}^T\mathbf{x} = \mathbf{x}^T QDQ^T\mathbf{x} + \mathbf{c}^T\mathbf{x} = \tilde{\mathbf{x}}^T D\tilde{\mathbf{x}} + \tilde{\mathbf{c}}^T\tilde{\mathbf{x}}, \quad \tilde{\mathbf{x}} = Q^T\mathbf{x}, \ \tilde{\mathbf{c}} = Q^T\mathbf{c}.$$

Hence, our general conic can be identified as a standard conic in different coordinates $(\tilde{x}, \tilde{\mathbf{y}})$, related to the original coordinates through a multiplication by an orthogonal matrix. This allows us apply the above classification by verifying the diagonal entries of D (the eigenvalues of G). We illustrate this approach with an example.

Example 20.3 Identify the conic

$$\mathcal{C} := \left\{(x, y) \in \mathbb{R}^2 : x^2 - xy + y^2 - x - y = -\frac{1}{2}\right\}.$$

We can still identify a quadratic and a linear form in the above expression. In particular, note that we can write the expression for a general quadratic form as a sum of diagonal and off-diagonal terms:

$$q(\mathbf{x}) = \mathbf{x}^T G\mathbf{x} = \sum_{i,j=1}^{n} x_i G_{ij} x_j = \sum_{i=1}^{n} x_i^2 G_{ii} + 2\sum_{i<j} x_i G_{ij} x_j.$$

In our case ($n = 2$), we find $G_{11} = G_{22} = 1$, while $G_{12} = -1/2$. Hence, the conic can be described as

$$\mathcal{C} := \left\{(x, y) \in \mathbb{R}^2 : \frac{1}{2}\mathbf{x}^T \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \mathbf{x} - \mathbf{1}^T\mathbf{x} = -\frac{1}{2}\right\}.$$

So what is $\mathcal{C}$? The eigenvalues of G are both positive, so this is an ellipse. Further calculations can identify the coordinates $(\tilde{x}, \tilde{y})$, together with the corresponding semi-axes etc.

We end this lecture by noting the classification of quadrics. These are surfaces in 3D which have the same form (quadratic form + linear form), with the Gram matrix being a 3-by-3 symmetric matrix. The classification is as follows:

- $\lambda_1 > 0, \lambda_2 > 0, \lambda_3 > 0$: ellipsoid;
- $\lambda_1 > 0, \lambda_2\lambda_3 < 0$: hyperboloid;
- $\lambda_1\lambda_2 > 0, \lambda_3 = 0$: elliptic paraboloid;
- $\lambda_1\lambda_2 < 0, \lambda_3 = 0$: hyperbolic paraboloid.