UNIT 3: EIGENDECOMPOSITIONS: THE QUADRATIC FORMS PERSPECTIVE

Dr.P.M.Bajracharya

January 2, 2022

Contents

1 Quadratic forms and matrices

We have been viewing matrices in terms of transformations, i.e., a matrix $A \in \mathbb{R}^{m\times n}$ as a representation of a linear function from $\mathbb{R}^n$ to $\mathbb{R}^m$, but we can also view matrices in terms of quadratic forms. Viewing matrices in terms of quadratic forms is a very powerful approach in machine learning and data science. It makes many advanced high-dimensional concepts more intuitive. It also provides a very nice geometric way to think about eigenvalues and eigenvectors that complements the more algebraic approach we adopted previously.

Definition. A quadratic form $Q : \mathbb{R}^n \to \mathbb{R}$ is a polynomial in the variables $x_1, ..., x_n$, all of whose terms are of degree 2, that is,

$$Q(x) = x^T Ax = \sum_{i,j=1}^{n} Ax_i x_j,$$

where A is an $n \times n$ matrix.

Example 1.

1. x_1^2
2. $x_1^2 + x_2^2$
3. $4x_1^2 + 2x_1x_2 - x_2^2$

Terms of the form $x_i x_j$, for $i \neq j$ are sometimes called **cross terms**, since they involve the product of two different variables, rather than the product of a variable with itself. Terms of the form x_i^2 , that involve the product of a variable with itself, are sometimes called **diagonal terms**. A quadratic form can have diagonal terms or cross terms or both.

Quadratic forms that do not have any cross terms are sometimes called **diagonal quadratic forms**.

An expression

$$4(x_1 - 3)^2 + x_2^2 = 4x_1^2 - 24x_1 + 36 - x_2^2$$

is not a quadratic form in x_1 and x_2. It is, however, "almost" a quadratic form. The reason is that if we put $x_1' = x_1 - 3$, i.e., if we do a variable transformation that is a simple translation along the x_1 axis, then

$$4(x_1 - 3)^2 + x_2^2 = 4x_1'^2 + x_2^2.$$

By definition, it is a quadratic in x_1' and x_2. Redefining variables, i.e., performing a variable transformation, will permit us to remove those lower-order terms, which will help simplify/clarify the discussion. This will help us to define eigenvectors and eigenvalues more easily.

The connection with matrices. Consider the function $f : \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = a + bx.$$

This function is the sum of an afine part (a) and a linear part (bx). We know that we can view this as a linear function (in the sense that it satisfies the definition of a linear function) by "removing" the afine part and considering the function

$$g(x) = f(x) - a = bx.$$

This simply amounts to "shifting" or "translating" the function along the y-axis (if we view the function as $y = f(x)$), so that it goes through the origin.

Similarly, we can view the quadratic function given by the following equation

$$f(x) = cx^2 + bx + a$$

as a quadratic form by "removing" the linear and affine parts. To do so, one can use the procedure "completing the square". We have

$$\begin{aligned} f(x) &= cx^2 + bx + a \\ &= c\left(x^2 + \frac{b}{c}x + \frac{a}{c}\right) \\ &= c\left(\left(x + \frac{b}{2c}\right)^2 - \frac{b^2 - 4ac}{4c^2}\right) \\ &= c\left(x + \frac{b}{2c}\right)^2 - \frac{b^2 - 4ac}{4c} \end{aligned}$$

and thus we have a quadratic with a vertex at

$$\left(-\frac{b}{2c}, -\frac{b^2 - 4ac}{4c}\right).$$

If we define

$$\begin{aligned} g(x) &= f(x) + \frac{b^2 - 4ac}{4c} \\ x' &= x + \frac{b}{2c}, \end{aligned}$$

then we have

$$g(x) = cx'^2,$$

which is a quadratic form in the variable x'.

Next, let's go to the two-variable case. Consider the function

$f(x) : \mathbb{R}^2 \to \mathbb{R}$ given as

$$\begin{aligned} & f(x) = a + b_1 x_1 + b_2 x_2 \\ \Rightarrow \qquad & f(x) = a + b^T x, \end{aligned} \tag{1}$$

where $b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$. In this expression, b can be viewed in one of two complementary ways: first, b is a vector, in which case $b^T x$ is a number that is the dot product of b and x; and second, b^T is a 1×2 matrix that maps $x \in \mathbb{R}^2$ to a number in $\mathbb{R}$. In either case, the function f is the sum of an affine part (a) and a linear part ($b^T x$). As before, we can convert this function by Equation (1) to a linear function by considering the translated function

$$g(x) = f(x) - a = b^T x.$$

Next, we consider a function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x) = a + b_1 x_1 + b_2 x_2 + c_1 x_1^2 + c_2 x_2^2 + c_3 x_1 x_2.$$

Then we rewrite this equation as follows:

$$\begin{aligned} f(x) &= a + (b_1 \ b_2) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + (x_1 \ x_2) \begin{pmatrix} c_1 & c_3/2 \\ c_3/2 & c_2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= a + b^T x + x^T C x, \end{aligned} \tag{2}$$

where $C = \begin{pmatrix} c_1 & c_3/2 \\ c_3/2 & c_2 \end{pmatrix}$. As before, we will be able to remove the linear and affine parts ($b^T x$ and a, respectively), again using the completing the square and shifting procedures. This too will lead to a quadratic form, but one with cross terms of the form $x_1 x_2$. Removing these cross terms in order to get a much simpler quadratic form without any cross terms will be closely related to computing *eigenvectors* and *eigenvalues*.

Generalization

To go to the three-variable case and beyond, observe that while Equation (1) and Equation (2) have been derived for a function $f : \mathbb{R}^2 \to \mathbb{R}$, exactly the same expressions could be derived for $f : \mathbb{R}^n \to \mathbb{R}$. The reason is that the equations just involve dot products and matrix-vector multiplications, and there is no explicit dependence on the dimension of the input to this function. This suggests that we can write a general quadratic function $f : \mathbb{R}^n \to \mathbb{R}$ in terms of an $n \times n$ symmetric matrix, a n-dimensional vector, and a number. This is true. As an example, for $x \in \mathbb{R}^3$, the generalization of Equation (2) is

$$\begin{aligned} f(x) = a + (b_1\ b_2\ b_3) & \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \\ & + (x_1\ x_2\ x_3) \begin{pmatrix} c_{11} & c_{12}/2 & c_{13}/2 \\ c_{12}/2 & c_{22} & c_{23}/2 \\ c_{13}/2 & c_{23}/2 & c_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}. \end{aligned} \tag{3}$$

The above discussion gives rise of two problems:

Problem I. Express a quadratic form in terms of matrices.

Problem II. Given a matrix, find the associated quadratic form.

Problem I.

Let's consider the case where the function $f : \mathbb{R}^3 \to \mathbb{R}$ has just quadratic terms, i.e., where it is a quadratic form. In this case, we are considering

$$f(x) = c_{11}x_1^2 + c_{22}x_2^2 + c_{33}x_3^2 + c_{12}x_1x_2 + c_{13}x_1x_3 + c_{23}x_2x_3. \tag{4}$$

We can write this as the product of three matrices as follows:

$$f(x) = (x_1 \ x_2 \ x_3) \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ 0 & c_{22} & c_{23} \\ 0 & 0 & c_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x^T A x.$$

It follows that

$$A = \begin{pmatrix} c_{11} & 0 & 0 \\ c_{12} & c_{22} & 0 \\ c_{13} & c_{23} & c_{33} \end{pmatrix}$$

is the matrix associated with the given quadratic form.

Thus, the point here is that we can start with an arbitrary quadratic form and construct in a very natural way a matrix associated with the given quadratic form.

Note however that, even if we restrict ourselves to quadratic functions that contain just terms of degree 2, this quadratic form does not uniquely define a matrix. For example, we could have put the c_{12}, c_{13} and c_{23} below the diagonal, with 0s above the diagonal as follows:

$$f(x) = (x_1 \ x_2 \ x_3) \begin{pmatrix} c_{11} & 0 & 0 \\ c_{12} & c_{22} & 0 \\ c_{13} & c_{23} & c_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x^T A^T x.$$

It follows that

$$A^T = \begin{pmatrix} c_{11} & 0 & 0 \\ c_{12} & c_{22} & 0 \\ c_{13} & c_{23} & c_{33} \end{pmatrix}$$

is the matrix associated with the given quadratic form.

Alternatively, we can write it in a more symmetric form as follows:

$$f(x) = (x_1\ x_2\ x_3) \begin{pmatrix} c_{11} & c_{12}/2 & c_{13}/2 \\ c_{12}/2 & c_{22} & c_{23}/2 \\ c_{13}/2 & c_{23}/2 & c_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

In this last expression, we have split the "off diagonal" terms, i.e., c_{ij} , for $i \neq j$, into half associated with the $x_i x_j$ term and half associated with the $x_j x_i$ term. Right now, we just want to note that this is possible, i.e., we didn't have to do this for the above expression to be true (as we saw with the previous two non-symmetric examples that also reproduce the same $f(x)$). It turns out, however, to be very convenient to do this. We'll get back to why this is the case soon.

Problem II.

Let's go the "other way," i.e., let's start with an arbitrary square matrix, starting with an arbitrary 3×3 matrix for simplicity,

$$A = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix} \in \mathbb{R}^{3\times 3},$$

where, in particular, we permit the possibility that $A_{12} = A_{21}$ as well as the possibility that $A_{12} \neq A_{21}$, and similarly for the other off-diagonal terms. In this case, we have

$$Ax = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} A_{11}x_1 + A_{12}x_2 + A_{13}x_3 \\ A_{21}x_1 + A_{22}x_2 + A_{23}x_3 \\ A_{31}x_1 + A_{32}x_2 + A_{33}x_3 \end{pmatrix} \in \mathbb{R}^3,$$

and we also have

$$\begin{aligned}
x^T A x &= (x_1\ x_2\ x_3) \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \\
&= A_{11}x_1^2 + (A_{12} + A_{21})x_1x_2 + (A_{13} + A_{31})x_1x_3 \\
&\quad + A_{22}x_2^2 + (A_{23} + A_{32})x_2x_3 + A_{33}x_3^2 \in \mathbb{R}^3.
\end{aligned}$$

The point of this is to show that we can start with an arbitrary square matrix and get a quadratic form. (i.e., we don't need to have a symmetric matrix to have a quadratic form.) Clearly, if we define the matrix A' to be

$$A' = \begin{pmatrix} A_{11} & \frac{A_{12}+A_{21}}{2} & \frac{A_{13}+A_{31}}{2} \\ \frac{A_{12}+A_{21}}{2} & A_{22} & \frac{A_{23}+A_{32}}{2} \\ \frac{A_{13}+A_{31}}{2} & \frac{A_{23}+A_{32}}{2} & A_{33} \end{pmatrix} \in \mathbb{R}^{3\times 3},$$

then $x^T A x = x^T A' x$, for all $x \in \mathbb{R}^3$, illustrating again the same non-uniqueness.

Here is a specific example of two matrices that correspond to the same quadratic form.

Example 2.

$$\text{If we let } A_1 = \begin{pmatrix} 1 & 2 & 4 \\ 0 & 6 & 8 \\ 0 & 0 & 9 \end{pmatrix} \text{ and } A_2 = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 6 & 4 \\ 2 & 4 & 9 \end{pmatrix},$$

then

$$
\begin{aligned}
x^T A_1 x &= (x_1\ x_2\ x_3) \begin{pmatrix} 1 & 2 & 4 \\ 0 & 6 & 8 \\ 0 & 0 & 9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \\
&= x_1^2 + 6x_2^2 + 9x_3^2 + 9x_1x_2 + 4x_1x_3 + 8x_2x_3 \\
&= (x_1\ x_2\ x_3) \begin{pmatrix} 1 & 1 & 2 \\ 1 & 6 & 4 \\ 2 & 4 & 9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \\
&= x^T A_2 x.
\end{aligned}
$$

This discussion illustrates that we can have many different square matrices that give rise to the same quadratic form. We might wonder whether there is some sort of standard or canonical form in which we can write a matrix that removes this non-uniqueness, since then we can write any quadratic form in terms of a unique matrix. The answer to this is yes. Basically, the way to do this is to do the above "average out" and write the quadratic form as a symmetric matrix (like we did above). That is, consider symmetric matrices. We could do it other ways, e.g., put everything above the diagonal, but we will do it this way since symmetric matrices have so many nice properties. Conversely, for any symmetric matrix, there is an associated quadratic form.

The point is the following. In general, if A is a square $n \times n$ matrix, then the function $f : \mathbb{R}^n \to \mathbb{R}$ given by

$$f(x) = x^T Ax$$

is a quadratic form. On the one hand, it is the product of three matrices, two of which are vectors, that for a given x and A yields a 1×1 matrix that is a number. On the other hand, it is the sum of a bunch of terms, each of which consists of elements $A_{ij} + A_{ji}$ of the matrix A multiplied by the product $x_i x_j$ of variables from the vector x. If we consider a matrix in which the off-diagonal position A_{ij} is replaced with $\frac{1}{2}(A_{ij} + A_{ji})$, then we get the same quadratic form, and so we will work with symmetric matrices.

Here is the statement of the basic result.

Theorem 1.1. *If $A \in \mathbb{R}^{3\times 3}$ with a quadratic form in 3 variables, then there is a symmetric matrix $B \in \mathbb{R}^{3\times 3}$) such that*

$$\forall\ x \in \mathbb{R}^3 \quad x^T A x = x^T B x.$$

Proof.

Let

$$A = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix}.$$

Then the quadratic form is given by

$$\begin{aligned} x^T A x &= (x_1\ x_2\ x_3) \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \\ &= A_{11}x_1^2 + (A_{12} + A_{21})x_1x_2 + (A_{13} + A_{31})x_1x_3 \\ &\quad + A_{22}x_2^2 + (A_{23} + A_{32})x_2x_3 + A_{33}x_3^2 \in \mathbb{R}^3. \end{aligned}$$

We put

$$B_{ij} = \frac{A_{ij} + A_{ji}}{2}.$$

Then the matrix $B = (B_{ij})_{3\times 3}$ is symmetric, since

$$B_{ij} = \frac{A_{ij} + A_{ji}}{2} = \frac{A_{ji} + A_{ij}}{2} = B_{ji}.$$

Moreover,

$$B_{ij} + B_{ji} = \frac{A_{ij} + A_{ji}}{2} + \frac{A_{ji} + A_{ij}}{2} = A_{ij} + A_{ji}.$$

Therefore,

$$\forall\, x \in \mathbb{R}^3 \;\; x^T A x = x^T B x.$$

In general we have the following theorem.

Theorem 1.2. *Given an arbitrary quadratic form in n variables (which, recall can be written as $x^T Ax$ for a square matrix $A \in \mathbb{R}^{n\times n}$), we can always find a symmetric matrix $B \in \mathbb{R}^{n\times n}$) such that*

$$\forall\, x \in \mathbb{R}^n \;\; x^T A x = x^T B x.$$

:

Due to this result, nothing is lost if we assume that A is symmetric. So, from now on, we do this.

By the way, we agree always to choose the symmetric matrix for two reasons. First, it gives us a unique (symmetric) matrix corresponding to a given quadratic form. Second, the choice of the symmetric matrix A allows us to apply the special theory available for symmetric matrices.

Note that from any matrix, we get a type of quadratic form using

$$x^T A y.$$

In this case, the variables x and y have different dimensions. For example, if

$$A = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{pmatrix},$$

then we obtain the following:

$$\begin{aligned} x^T Ax &= (x_1\ x_2) \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \\ &= A_{11}x_1y_1 + A_{12}x_1y_2 + A_{13}x_1y_3 \\ &\qquad + A_{21}x_2y_1 + A_{22}x_2y_2 + A_{23}x_2y_3 \in \mathbb{R}^3. \end{aligned}$$

Instead, we will focus on square matrices, in which case the vectors on the two sides of the matrix can be represented by the same variable.

Exercise 1.

1. Find the quadratic form corresponding to each of the following symmetric matrices.

   (a) $\begin{pmatrix} 4 & 1/\sqrt{2} \\ 1/2 & \sqrt{2} \end{pmatrix}$

   (b) $\begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & 3 \\ 0 & 3 & -1 \end{pmatrix}$

2. Find the corresponding symmetric matrix for each of the following quadratic forms.

   (a) $Q(x) = x_1^2 - 2x_1x_2 - 3x_2^2$

   (b) $Q(x) = 2x_1^2 + 3x_1x_2 - x_1x_3 + 4x_2^2 + x_3^2$

   (c) $Q(x) = x_1^2 + 2x_2^2 + 3x_3^2 + 4x_4^2$

2 Some simple examples

Example 3.

Let $Q(x) = 8x_1^2 - 4x_1x_2 + 5x_2^2$. Determine whether $Q(0,0)$ is the global minimum.

Solution. We can rewrite the given equation in quadratic form as follows:

$$Q(x) = x^T Ax, \text{ where } A = \begin{pmatrix} 8 & -2 \\ -2 & 5 \end{pmatrix}.$$

Then

$$\begin{aligned} & \begin{vmatrix} 8-\lambda & -2 \\ -2 & 5-\lambda \end{vmatrix} = 0 \\ \Rightarrow \quad & (8-\lambda)(5-\lambda) - 4 = 0 \\ \Rightarrow \quad & \lambda^2 - 13\lambda + 36 = 0 \\ \Rightarrow \quad & (\lambda - 9)(\lambda - 4) = 0. \end{aligned}$$

Clearly, the eigenvalues of A are $\lambda_1 = 9, \lambda_2 = 4$.

$\underline{\lambda_1 = 9}$

We have

$$\begin{pmatrix} -1 & -2 \\ -2 & -4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0.$$

So, the eigenvector associated with λ_1 is

$$v_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ -1 \end{pmatrix}$$

$\underline{\lambda_1 = 4}$

We have

$$\begin{pmatrix} 4 & -2 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0.$$

So, the eigenvector associated with λ_2 is

$$v_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Let $x = c_1 v_1 + c_2 v_2$, Now we have

$$Qx = x^T A x = x^T \lambda x = \lambda_1 c_1^2 + \lambda_2 c_2^2 = 9c_1^2 + 4c_2^2$$

Therefore, $Q(x) \geq 0$ and so $Q(0,0)$ is the global minimum. ◀

Example 4.

If $A = I_n$, then we get the following:

$$\begin{aligned} f(x) &= x_1^2 + x_2^2 + ... + x_n^2 \\ &= (x_1 \ ... \ x_n) \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \\ &= x^T I x \\ &= x^T x \\ &= \|x\|_2^2. \end{aligned}$$

This is the usual Euclidean norm of a vector $x \in \mathbb{R}^n$.

Example 5.

If $A = D$, a diagonal matrix, e.g.,

$$D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 9 \end{pmatrix},$$

then we get

$$
\begin{aligned}
f(x) &= x_1^2 + 4x_2^2 + 9x_3^2 \\
&= (x_1 \ x_2 \ x_3) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & \\ 0 & 0 & 9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \\
&= x^T D x \\
&= (Dx^T)^{1/2} D^{1/2} x \\
&= \|D^{1/2} x\|_2^2.
\end{aligned}
$$

Example 6.

If A is a matrix that can be written as $A = B^T B$, i.e., it is a correlation/covariance matrix, then we get

$$
\begin{aligned}
f(x) &= x^T B^T B x \\
&= (Bx)^T B x \\
&= \|Bx\|_2^2.
\end{aligned}
$$

Observe that a special case of this is the diagonal matrix with all positive entries that we saw in the previous example.

Problem 1. Consider the matrix

$$
D = \begin{pmatrix} \alpha_1 & 0 & 0 \\ 0 & \alpha_2 & 0 \\ 0 & 0 & \alpha_3 \end{pmatrix}.
$$

(a) Show that if $\alpha_i > 0$, for all$i$, then the quadratic form function $f(x) = x^T D x$ is a vector norm, in the sense that it satisfies the three conditions for a function to be a vector norm.

(b) Show that if $\alpha_i \geq 0$, for alli, but $\alpha_i = 0$ for at least one i, then determine which of the conditions for the quadratic form function to be a norm are satisfied and which are violated.

(c) Show that if $\alpha_i > 0$, for some i and $\alpha_i < 0$ for other i, then determine which of the conditions for the quadratic form function to be a norm are satisfied and which are violated.

3 Symmetric bi-linear functions

We kinow that any matrix can be viewed as a representation of a linear transformation with respect to a basis. We can also associate to any symmetric matrix something that is known as a **symmetric bilinear transformation**. Of course, symmetric matrices have more structure, and this will help us to say more about them. In this case, the "more" will be intuitive geometric things having to do with eigenvalues and eigenvectors that are of widespread importance in machine learning and data science.

Let's start with the definition.

Definition. Let V be a vector space, e.g., $\mathbb{R}^2$. Then a **symmetric bilinear function** on V is defined as a mapping $B : V \times V \to \mathbb{R}$ such that for any $u, v, w \in V$

(a) $B(u, v) = B(v, u)$

(b) $\forall\, a, b \in \mathbb{R} \quad B(au + bv, w) = aB(u, w) + bB(v, w)$.

Remark.

1. The first condition says that B is symmetric in its two arguments.
2. By combining these two conditions we also have that B is linear in its second argument.

Here are two special cases to compare this to the linear function perspective on matrices.

- Restricted to the case of 1×1 symmetric matrices, $A = (a)$, if we view this matrix as a linear function, then we are thinking of it as $y = Ax$, while if we think of this matrix as a symmetric bilinear function, then we are thinking of it as $y = Ax^2$.
- Restricted to the case of 2×2 symmetric matrices,
$$\begin{pmatrix} A_{11} & A_{12} \\ A_{12} & A_{22} \end{pmatrix},$$
if we view this as a linear function, then we are thinking of this as
$$y = \begin{pmatrix} A_{11} & A_{12} \\ A_{12} & A_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \in \mathbb{R}^2,$$
while if we are thinking of it as a symmetric bilinear function, then we are thinking of it as
$$y = (x_1 \; x_2) \begin{pmatrix} A_{11} & A_{12} \\ A_{12} & A_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}.$$

Viewing a symmetric matrix in terms of symmetric bilinear functions leads to the quadratic form perspective that has a cleaner geometric interpretation. It will also lead to a more geometric way to discuss eigenvalues/eigenvectors. This is the quadratic form analogue of the theorem that said that any specific matrix can be viewed as an abstract linear transformation, and vice versa.

Suppose that A is an $n \times n$ matrix. For $u, v \in \mathbb{R}^n$ we will define the function

$$f(u, v) = u^T A v \in \mathbb{R}.$$

Lets check then if this is a bilinear form.

We can see then that our defined function is bilinear. Looking at how this function is defined, especially the matrix A, it might give us a hint to a similarity between this bilinear form and the linear transformations

Theorem 3.1. *We have the following.*

(a) *If A is a symmetric $n \times n$ matrix, then $B_A(v, w) = v^T A w$ is a symmetric bilinear form.*

(b) *If B is a symmetric bilinear function on $\mathbb{R}^n$, then it is of the form $B = B_A(v, w) = v^T A w$, for some unique symmetric matrix A.*

Proof. (a) We know that $v^T A w \in \mathbb{R}$. So,

$$v^T A w = (v^T A w)^T.$$

This implies that

$$\begin{aligned}
B_A(v,w) &= (v^TAw)^T \\
&= (Aw)^Tv \\
&= w^TA^Tv \\
&= w^TAv \qquad \text{(because } A \text{ is symmetric)} \\
&= B_A(w,v).
\end{aligned}$$

Moreover, for any real number a

$$\begin{aligned}
B_A(av,w) &= (av)^TAw \\
&= a(v)^TAw \\
&= a(v^TAw) \\
&= aB_A(v,w).
\end{aligned}$$

Also,

$$\begin{aligned}
B_A(u+v,w) &= (u+v)^TAw \\
&= (u^T+v^T)Aw \\
&= u^TAw+v^TAw \\
&= B_A(u,w)+B_A(v,w).
\end{aligned}$$

Therefore, $B_A(v,w)=v^TAw$ is a symmetric bilinear form.

(b) Let $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}, w = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \in \mathbb{R}^n$. Then we can write

$$v = \sum_{i=1}^{n} v_i e_i, \quad w = \sum_{j=1}^{n} w_j e_j,$$

where $\{e_1, ..., e_n\}$ is a basis for $\mathbb{R}^n$. Using the properties of bilinear functions, we have

$$\begin{aligned}
B(v,w) &= B\left(\sum_{i=1}^{n} v_i e_i, \sum_{j=1}^{n} w_j e_j\right) \\
&= \sum_{i=1}^{n} v_i \sum_{j=1}^{n} w_j B(e_i, e_j) \\
&= \sum_{i=1}^{n} v_i \sum_{j=1}^{n} B(e_i, e_j) w_j \\
&= \sum_{i=1}^{n} v_i \begin{pmatrix} B(e_1,e_1) & \cdots & B(e_1,e_n) \\ \vdots & \vdots & \vdots \\ B(e_n,e_1) & \cdots & B(e_n,e_n) \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \\
&= (v_1 \ \dots \ v_n) \begin{pmatrix} B(e_1,e_1) & \cdots & B(e_1,e_n) \\ \vdots & \vdots & \vdots \\ B(e_n,e_1) & \cdots & B(e_n,e_n) \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}
\end{aligned}$$

Setting $A = (B(e_i, e_j))_{n \times n}$, we have

$$B = v^T A w.$$

To prove the uniqueness of the matrix A, we assume that there is another matrix C such that

$$B = v^T C w.$$

Then

$$v^T A w - v^T C w = 0 \Rightarrow v^T (A - C) w = 0$$

Put $D = A - C$. Thus, for any $v, w \in \mathbb{R}^n$

$$v^T D w = 0.$$

In particular, if we choose $v, w \in \mathbb{R}^n$ such that $v_i = w_i = 1$ for all $i \in \{1, ..., n\}$, then we still have

$$v^T D w = 0.$$

This implies that

$$\begin{aligned} & \sum_{i=1}^{n} v_i \sum_{j=1}^{n} D_{ij} w_j = 0 \\ \Rightarrow \qquad & \sum_{i=1}^{n} \sum_{j=1}^{n} v_i D_{ij} w_j = 0. \end{aligned}$$

We now see that $v_i D_{ij} w_j = D_{ij}$ for all $i \in \{1, ..., n\}$. Therefore, we must have

$$D_{ij} = 0 \quad \text{for all } i \in \{1, ..., n\}.$$

That means,

$$\begin{aligned} & D = 0 \\ \Rightarrow \qquad & A - C = 0 \\ \Rightarrow \qquad & A = C. \end{aligned}$$

◀

4 Connections with conic sections

We saw earlier that subspaces, linear dependence, etc. generalize the intuitive geometric ideas that we have about lines through the origin, planes through the origin, etc. The quadratic forms too generalize intuitive geometric ideas that are related to conic sections and their generalizations.

- **Ellipse.** The equation in standard form is given by

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1.$$

  Note that the major and minor axes of an ellipse in standard form are just the canonical axes. WLOG, let's assume that $a \geq b \geq 0$, otherwise it is longer along the other axis.

  Remark. A circle is just an ellipse with $a = b$.

- **Hyperbola.** The equation in standard form is given by

$$\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = 1.$$

- **Degenerate cases.**
  - **Parabola.** This is the set of points in the plane that are equidistant from a fixed point and a fixed line.
  - **Line.** This arises when the quadratic and constant terms are zero.
  - **Point** This arises when the quadratic and linear and constant terms are zero.

Remark. Note that the a and b above basically correspond to stretching the units of the x_1 and x_2 axes.

Examples of more complex conic sections. Let's say that, instead of being given a conic section in standard form, we are given some quadratic function expression and we have to determine the type of conic section to which it corresponds.

Example 7.

Let's sketch the graph of the conic section

$$9x_1^2 - 4x_2^2 - 72x_1 + 8x_2 + 176 = 0.$$

To do so, let's try to write it in one of the standard conic forms. First write terms in x_1 and x_2 separately.

$$9(x_1^2 - 8x_1) - 4(x_2^2 - 2x_2) + 176 = 0.$$

Then, complete the square to get

$$9(x_1 - 4)^2 - 4(x^2 - 1)^2 + 176 - 144 + 4 = 0$$

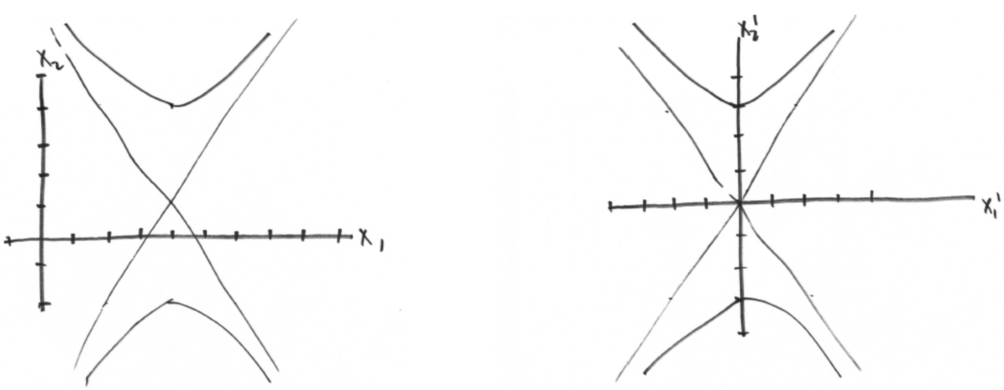
and then simplify this to get

$$9(x_1 - 4)^2 - 4(x^2 - 1)^2 + 36 = 0.$$

Then we get

$$-\frac{(x_1 - 4)^2}{4} + \frac{(x_2 - 1)^2}{9} = 1.$$

This is a hyperbola. See the figure given below.



(a) Original.

(b) Transformed.

There are two things to note about this example.

(a) Since the negative sign is on the x_1 and not the x_2, the hyperbola opens up-down and not left-right.

(b) Having those other linear and constant terms simply amounts to shifting the origin, i.e., defining a new set of variables. That is, if we define a new set of coordinates

$$x_1' = x_1 - 4$$
$$x_2' = x_2 - 1,$$

which simply amounts to shifting the origin, then we get the equation

$$-\frac{(x_1')^2}{4} + \frac{(x_2')^2}{9} = 1.$$

Example 8.

Let's sketch the graph of the conic section:

$$\frac{(x_1 - 4)^2}{4} + \frac{(x_2 - 1)^2}{9} = 1.$$

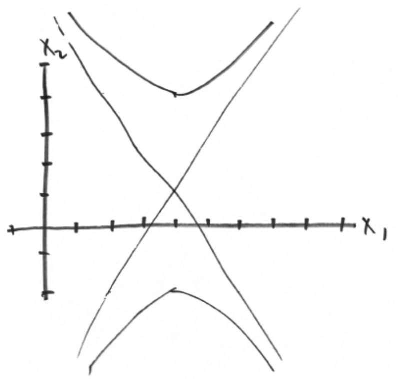
This is an ellipse. If we define a new set of coordinates

$$x_1' = x_1 - 4$$
$$x_2' = x_2 - 1,$$

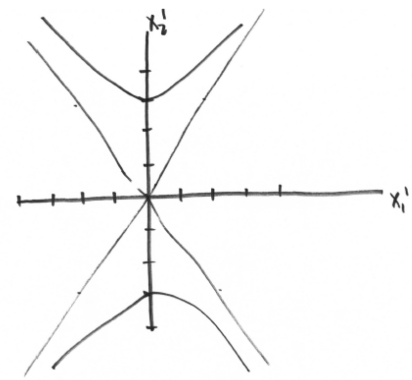
which simply amounts to shifting the origin, then we get the equation

$$\frac{(x_1')^2}{4} + \frac{(x_2')^2}{9} = 1.$$

See the figure (b) given below.



(a) Original.



(b) Transformed.

Question: What if there is an $x_1 x_2$ term?

Answer: We still complete the square, but using the other variables. In this case, we will see that A has a rotation, i.e., we don't have just a scaling and translation shift of the axes like in the previous example. Let's illustrate that.

Example 9.

Consider the expression

$$5x_1^2 - 4x_1x_2 + 5x_2^2 = 48$$

We do know that it is a symmetric bilinear equation, and so it satisfies the algebraic conditions of being a conic section, so it will be a conic section of some form. This expression can be written as follows:

$$48 = (x_1 \; x_2) \begin{pmatrix} 5 & -2 \\ -2 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x^T A x,$$

where the symmetric matrix associated with this quadratic form is

$$A = \begin{pmatrix} 5 & -2 \\ -2 & 5 \end{pmatrix}.$$

Recall that a rotation in the x_1x_2 plane takes the form

$$R_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

and the coefficients of x_1x_2 term in the quadratic form is related to the angle θ. Since this is a x_1x_2 matrix, this is a counterclockwise rotation by θ degrees. If we choose $\theta = 45°$, then we have

$$R_\theta = \begin{pmatrix} \cos(45°) & -\sin(45°) \\ \sin(45°) & \cos(45°) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Let's define a new coordinate system to be

$$x' = R_{\theta=45°} x$$

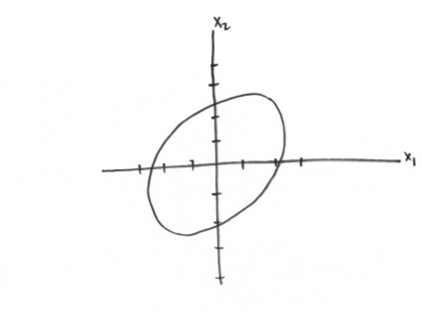
In more detail, we can write this new coordinate system as follows:

$$\begin{pmatrix} x_1' \\ x_2' \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} x_1 + x_2 \\ x_1 - x_2 \end{pmatrix}.$$

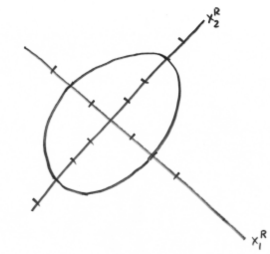
In the new (x_1', x_2') coordinate system, the associated conic section is

$$48 = 3(x_1')^2 + 7(x_2')^2 = 3\left(\frac{x_1 + x_2}{\sqrt{2}}\right)^2 + 7\left(\frac{x_1 - x_2}{\sqrt{2}}\right)^2.$$

Now, it is clear that the given quadratic form is an ellipse, with major and minor axes rotated by a 45° degree angle, relative to the canonical axes. See the figure given below.



(a) Conic section in original basis.



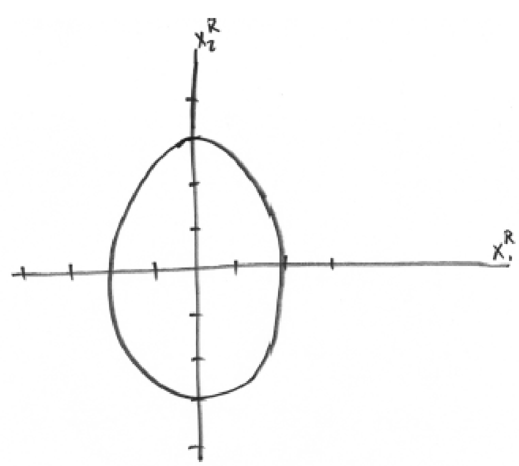
(b) Rotated basis for conic section.

In this new (x_1', x_2') coordinate system, the original expression can be written as follows:

$$48 = (x_1' \; x_2') \begin{pmatrix} 3 & 0 \\ 0 & 7 \end{pmatrix} \begin{pmatrix} x_1' \\ x_2' \end{pmatrix} = x'^T A' x'.$$

What this expression says is that, in the new coordinate system, the matrix A becomes a diagonal matrix A', defined as

$$A = \begin{pmatrix} 3 & 0 \\ 0 & 7 \end{pmatrix}.$$



Conic section in rotated basis.

What this example indicates that when we have cross terms such as x_1x_2, we have a conic section with major and minor axes that are rotated relative to the canonical basis, and that we can view the coordinate transformation illustrated in Figure 10.6 that makes the ellipse axis-aligned in one of two complementary ways:

- As de

  ning a new coordinate system as follows:

  $$x \to x' = RX,$$

  where $R = R_{\theta=45^\circ}$
- As defining a new matrix as follows:

  $$A \to A' = R^T AR.$$

Notation. Put

- $V = R^T$
- $D = A'$

Then $x' \in \mathbb{R}^2$ is given by $x' = V^T x$, and if we then look at the quadratic form $x^T Ax$ in the new coordinate system, then we have

$$x^T Ax = x^T(VDV^T)x = x^T VDV^T x = (V^T x)^T D(V^T x) = (x')^T Dx' :$$

Either way, we get

$$\begin{aligned} x^T A x &= 5x_1^2 - 4x_1x_2 + 5x_2^2 \\ &= 3\left(\frac{x_1 + x_2}{\sqrt{2}}\right)^2 + 7\left(\frac{x_1 - x_2}{\sqrt{2}}\right)^2 \\ &= 3(x')^2 + 7(x')^2 \\ &= x'^T A' x'. \end{aligned}$$

The bottom line is that the cross term amounts to doing some sort of orthogonal transformation, and this involves defining a new coordinate system where the matrix is diagonal.

5 Definiteness and indefiniteness of quadratic forms

The taxonomy of conic sections into ellipses, hyperbolas, and parabolas, as well as lines and points, is very useful for quadratic forms in 2 variables, but it quickly becomes awkward for quadratic forms in more than 2 variables. For quadratic functions in n variables, a related but slightly different classification is more convenient.

To understand the generalization, recall that the distinction into ellipses, hyperbolas, and parabolas depended on whether, when written in standard form, the coefficients of the variables were all the same sign (both positive, or both negative) or were different signs (one positive and one negative) or had one zero (one positive and one zero). The generalization we will consider will be a generalization of this condition.

- *Definite*: all positive or all negative.

- *Degenerate*: all non-negative including some zero, or all non-positive including some zero, i.e., all are either positive or zero, with not all positive, or all are either negative or zero, with not all negative.
- *Indefinite*: some positive or some negative, including potentially some that are zero.

Definition. Let A be an $n \times n$ symmetric matrix, and recall that $Q(x) = x^T Ax$ is the corresponding quadratic form. Then A (as well as Q) is

(a) **positive definite** if $x^T Ax > 0$, for all x

(b) **negative definite** if $x^T Ax < 0$, for all x

(c) **indefinite** if $x^T Ax > 0$ for some x's and $x^T Ax < 0$ for others.

(d) **positive semidefinite** if $x^T Ax \geq 0$, for all x

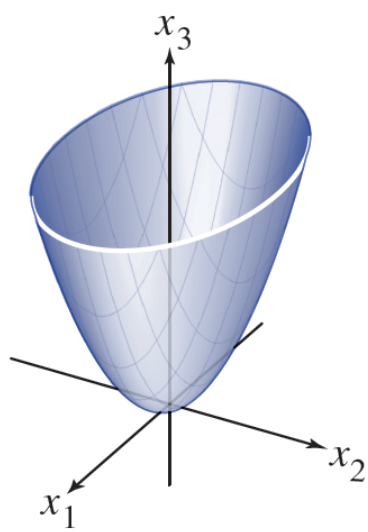
(e) **negative semidefinite** if $x^T Ax \leq 0$, for all x

Example 10.

(a) The following BQF is *positive definite*:

$$f(x, y) = x^2 + y^2$$

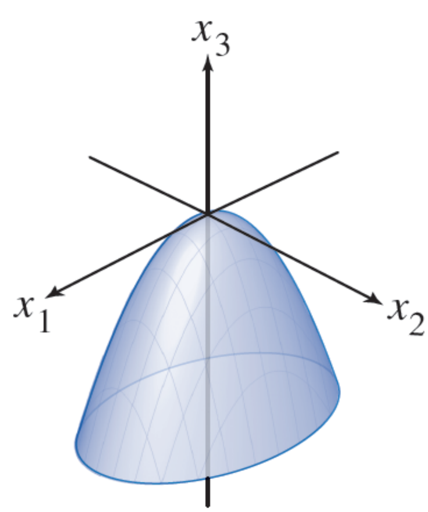
because $f(x, y) \geq 0$ for all $x, y \in \mathbb{R}$ and that $f(x, y) = 0$ if and only if $x = 0$ and $y = 0$.



(b) The following BQF is *negative definite*:

$$f(x, y) = -x^2 - y^2$$

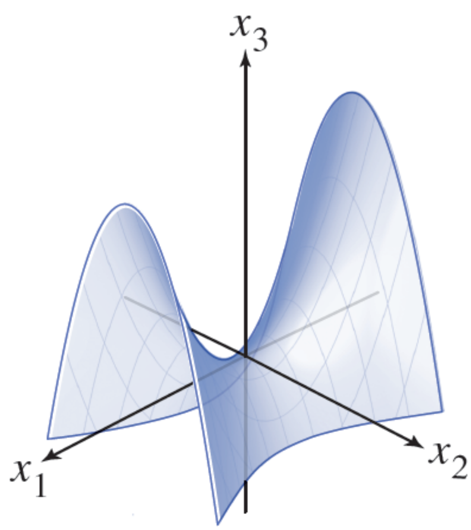
because $f(x, y) \leq 0$ for all $x, y \in \mathbb{R}$ and that $f(x, y) = 0$ if and only if $x = 0$ and $y = 0$.



(c) The following BQF is *indefinite*:

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

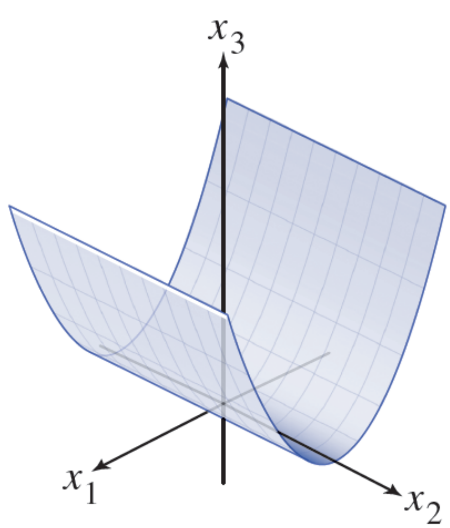
because $f(1, 1) = 1 > 0$, but $f(-1, 1) = -1 < 0$.



(d) The following BQF is *positive semidefinite1* if

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

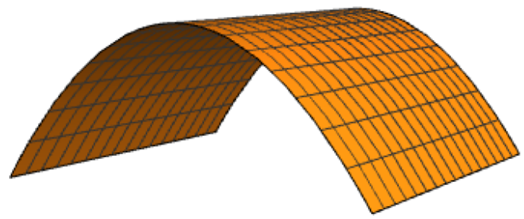
because $f(x, y) = (x + y)^2 \geq 0$ for all $x, y \in \mathbb{R}$. Observe that $f(x, y) = 0$ whenever $x = -y$.



(e) The following BQF is *negative semidefinite* it

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

because $f(x, y) = -(x+y)^2 \leq 0$ for all $x, y \in \mathbb{R}$. Observe that $f(x, y) = 0$ whenever $x = -y$.



Theorem 5.1 (Quadratic Forms and Eigenvalues)**.** *Let A be a 2×2 symmetric matrix. Then a quadratic form $x^T Ax$ is*

(a) positive definite if and only if the eigenvalues of A are both positive,

(b) negative definite if and only if the eigenvalues of A are both negative,

(c) indefinite if and only if A has both positive and negative eigenvalues,

(d) positive semidefinite if and only if the eigenvalues of A are nonnegative,

(e) negative semidefinite if and only if the eigenvalues of A are nonpositive.

Consequently, if $|A| = 0$, then the quadratic form is neither positive definite nor negative definite.

Note that this theorem can be generalized for an $n \times n$ matrix.

Example 11.

Determine the definiteness of the quadratic form $Q(x) = x_1^2 + 2x_1x_2 + x_2^2$.

Solution. This form can be written as

$$Q(x) = x^T \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} x.$$

Then we have

$$|A - \lambda I| = \begin{pmatrix} 1-\lambda & 1 \\ 1 & 1-\lambda \end{pmatrix} = (1-\lambda)^2 - 1 = \lambda(\lambda - 2).$$

Clearly, the eigenvalues of A are $\lambda_1 = 2$ and $\lambda_2 = 0$. Hence the quadratic form is positive semi-definite.

Example 12.

For which real numbers k is the quadratic form $Q(x) = kx_1^2 - 6x_1x_2 + kx_2^2$ positive semi-definite?

Solution. To determine the definiteness of this form we'll need to consider the matrix

$$\begin{pmatrix} k & -3 \\ -3 & k \end{pmatrix},$$

whose characterstic polynomial is

$$\begin{aligned} |A - \lambda I| &= \begin{pmatrix} k-\lambda & -3 \\ -3 & k-\lambda \end{pmatrix} \\ &= (k-\lambda)^2 - 9 \\ &= (k - \lambda + 3)(k - \lambda - 3). \end{aligned}$$

Clearly, the eigenvalues of A are $\lambda_1 = k+3$ and $\lambda_2 = k-3$. In order for Q to be positive definite, both of these eigenvalues must be positive. So $k > 3$ is a necessary and suffiient condition for Q to be a positive definite quadratic form.

Theorem 5.2 (Quadratic forms as a sum of squares)**.** *(a) For any quadratic form Q on $\mathbb{R}^n$, there exists $m = k+l$ linearly independent functions $\alpha_1, ..., \alpha_m$, such that*

$$Q(x) = (\alpha_1(x))^2 + ... + (\alpha_k(x))^2 - (\alpha_{k+1}(x))^2 - ... - (\alpha_{k+\ell}(x))^2.$$

(b) The number k of positive signs and the number ℓ of minus signs depends only on Q and not on the specific linear function chosen.

Definition. The **signature of a quadratic form** is the pair (k, ℓ), where k and ℓ are the numbers mentioned in the theorem.

Two things to note about the signature:

(1) It is unchanged if we use different linearly independent functions.

(2) It does not identify uniquely the quadratic form.

We now illustrate the theorem with the following examples.

Example 13.

Let's consider $f(x) = x_1^2 + x_1x_2$. By completing the square on x_1, we get

$$\begin{aligned} x_1^2 + x_1x_2 &= \left(x_1 + \frac{x_2}{2}\right)^2 - \left(\frac{x_2}{2}\right)^2 \\ &= (\alpha_1(x_1, x_2))^2 - (\alpha_2(x_1, x_2))^2. \end{aligned}$$

Note that what we have essentially done by completing the square here is that we have defined the two functions, $\alpha_1, \alpha_2 : \mathbb{R}^2 \to \mathbb{R}$ as follows:

$$\alpha_1 = \alpha_1\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) = x_1 + \frac{x_2}{2}$$
$$\alpha_2 = \alpha_2\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) = \frac{x_2}{2}.$$

Viewed as a matrix transformation, this can be written as a single function α as

$$\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

That is, $\alpha = Ax$.

Example 14.

Let's consider $f(x) = x_1^2 + x_1x_2 - x_2^2$. One might wonder whether we should start with x_1 or with x_2. Let's try both.

- First, try x_1 first. This gives us

$$\begin{aligned} x_1^2 + x_1x_2 - x_2^2 &= \left(x_1 + \frac{x_2}{2}\right)^2 - x_2^2 - \left(\frac{x_2}{2}\right)^2 \\ &= \left(x_1 + \frac{x_2}{2}\right)^2 - \left(\frac{\sqrt{3}}{2}x_2\right)^2, \end{aligned}$$

which implies that we are using the following transformation:

$$\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = Ax.$$

- Second, try x_2 first. This gives us

$$\begin{aligned} x_1^2 + x_1x_2 - x_2^2 &= -(x_2^2 - x_1x_2 - x_1^2) \\ &= -\left(\left(x_2 - \frac{x_1}{2}\right)^2 - x_1^2 - \left(\frac{x_1}{2}\right)^2\right) \\ &= \left(\frac{\sqrt{3}}{2}x_2\right)^2 - \left(x_2 - \frac{x_1}{2}\right)^2, \end{aligned}$$

  which implies that we are using the following transformation

$$\alpha' = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{3}}{2} & 0 \\ -\frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = A'x.$$

So, the point is that we don't get a unique answer. That's fine. We can use $\alpha = Ax$ or $\alpha' = A'x$ as our transformation. The theorem doesn't guarantee uniqueness. Moreover, if there is a subspace, then vectors aren't unique.

Example 15.

Consider $f(x) = f(x_1, x_2, x_3) = x_1^2 + 2x_1x_2 - 4x_1x_3 + 2x_2x_3 - 4x_3^2$. Then

$$\begin{aligned} f(x) &= x_1^2 + (2x_2 - 4x_3)x_1 + 2x_2x_3 - 4x_3^2 \\ &= (x_1 + x_2 - 2x_3)^2 + 2x_2x_3 - 4x_3^2 - (x_2 - 2x_3)^2 \\ &= (x_1 + x_2 - 2x_3)^2 - x_2^2 + 6x_2x_3 - 8x_3^2 \\ &= (x_1 + x_2 - 2x_3)^2 - (x_2^2 - 6x_3x_2) - 8x_3^2 \\ &= (x_1 + x_2 - 2x_3)^2 - (x_2 - 3x_3)^2 + x_3^2. \end{aligned}$$

This implies that we are using the following transformation

$$\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & -2 \\ 0 & 1 & -3 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = Ax.$$

If we had done the operations above in a different order, then we would have gotten a different matrix, $\alpha' = A'x$, but it would have had the same signature.

You can imagine that this is the sort of thing that quickly gets tedious and error-prone for a person to do, but it is very easy for a computer to do all day long.

Other linear algebra classes will spend a lot of time on this, e.g., the mechanics of how to do this (although often not making the connections with quadratic forms), typically viewing this as a procedure to compute a basis for solving systems of linear equations. We are basically doing the same thing here, but we are viewing it as completing the square repeatedly, since we are viewing the matrix as a quadratic form. The reason is that this perspective is a more intuitive geometric transformation that provides some intuition for high-dimensional data science problems more generally.

So, we just repeatedly complete the square.

- If we get a sum of squares, e.g., $(u(x_1, x_2))^2 + (v(x_1, x_2))^2 = 1$, then we get a generalized ellipse, in the subspace spanned by those two vectors.
- If we get a difference of squares, e.g., $(u(x_1, x_2))^2 - (v(x_1, x_2))^2 = 1$, then we get a generalized hyperbola, in the subspace spanned by those two vectors.

More generally, if we have more than two variables, then we might have more than two terms positive, more than two terms negative, as well as some terms zero. The number of terms with each sign, as well as the subspace they span, will be the same, even if the exact terms, i.e., basis functions, are different. This is what the theorem on *quadratic forms as a sum of squares* says.

Above we just followed a rule to complete the square.

Question: What if no single variable with quadratic term, e.g., $f(x_1, x_2) = x_1 x_2$, then what do we do?

Answer: Just juggle things around to make it work.

Example 16.

Let $Q(x) = x_1x_2 - x_1x_3 + x_2x_3$. In this case, introduce new variables, e.g., $x_1' = x_1 - x_2$, i.e., $x_1 = x_1' + x_2$. This amounts to a transformation

$$\alpha = \begin{pmatrix} x_1' \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

Then the given quadratic form can be written as

$$\begin{aligned} Q(x) &= (x_1' + x_2)x_2 - (x_1' + x_2)x_3 + x_2x_3 \\ &= x_1'x_2 + x_2^2 - x_1'x_3 - x_2x_3 + x_2x_3 \\ &= x_2^2 + x_1'x_2 - x_1'x_3 \end{aligned}$$

and from this we can just proceed with the previous rule we were following.

Importantly, this doesn't depend on the specific new set of variables, e.g., we could choose $x_1' = x_1 - x_2$, i.e., $x_1 = x_1' - x_2$. In this case, this amounts to a transformation

$$\alpha = \begin{pmatrix} x_1' \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

In this new set of variables, we have

$$\begin{aligned} Q(x) &= (x_1' - x_2)x_2 - (x_1' - x_2)x_3 + x_2x_3 \\ &= x_1'x_2 - x_2^2 - x_1'x_3 + x_2x_3 + x_2x_3 \\ &= -x_2^2 + x_1'x_2 - x_1'x_3 + 2x_2x_3 \end{aligned}$$

Again, now that we have a term of the form x_1^2 , we can just proceed with the previous rule we were following. (Or we could have chosen many other possibilities.)

Remark. The sum of squares may not be linearly independent. Consider the following example.

Example 17.

Consider the following:

$$\begin{aligned} f(x) &= 2x_1^2 + 2x_2^2 + 2x_1x_2 \\ &= \left(\sqrt{2}x_1 + \frac{x_2}{\sqrt{2}}\right)^2 + \left(\frac{3}{2}x_2\right)^2 \end{aligned}$$

In this example, the first line has 3 terms in 2 variables, so they are not independent, and they can be written as a sum of two terms. Alternatively, consider the following.

$$x_1^2 + x_2^2 + 2x_1x_2 = (x_1^2 + x_2)^2.$$

In this case, there is only 1 linearly-independent term (since, e.g., $B^2 - 4AC = 0$).

6 Two other topics

Connections with determinants and linear algebra. We provided several examples that illustrated that coefficients and cross terms in a symmetric matrix amount to defining new coordinate systems that are translated, rescaled, and/or rotated versions of the original coordinate system. Equivalently, they define a basis that is a translated, rescaled and/or rotated version of the original basis.

How general is this for an arbitrary quadratic form?

Relatedly, given an arbitrary quadratic form, is there a simple way to tell which type of conic section we have?

To answer this, recall that in general in $\mathbb{R}^2$, we can have the following:

$$Ax_1^2 + Bx_1x_2 + Cx_2^2 + Dx_1 + Ex_2 + F = 0, \tag{5}$$

where A, B, and C are numbers not all zero.

You may recall that we can classify conic sections by the discriminant: $B^2 - 4AC$. Assuming the conic section is non-degenerate, then we have the following.

- If $B^2 - 4AC < 0$, then the conic is an ellipse. (In particular, it is a circle if $A = C$ and $B = 0$.)
- If $B^2 - 4AC = 0$, then the conic is an parabola.

- If $B^2 - 4AC > 0$, then the conic is a hyperbola.

For a moment, let's ignore the D, E, and F terms, i.e., assume that we only have quadratic terms in the above equation, i.e., assume that we have taken care of the translation and rotation. In this case, we can write Equation (5) as follows:

$$(x_1\ x_2)\begin{pmatrix} A & B/2 \\ B/2 & C \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0.$$

Set

$$\Omega = \begin{pmatrix} A & B/2 \\ B/2 & C \end{pmatrix} m$$

Then the condition we want to check is

$$|\Omega| = B^2 - 4AC > 0 \text{ or } = 0 \text{ or } < 0$$

The condition $|\Omega = 0$ means that the matrix is rank-defficient, i.e., that

$$\begin{pmatrix} A \\ B/2 \end{pmatrix} = \alpha \begin{pmatrix} B/2 \\ C \end{pmatrix},$$

for some constant α. Recall that this is just the condition for linear dependence between two two-dimensional vectors. So, in particular, if $B^2 - 4AC = 0$, then we have linear dependence and a rank-defficient matrix. Alternatively, if $B^2 - 4AC \neq 0$, then we have linear independence.

If D, E, and F are non-zero, then we can use the procedures we described earlier to define a new set of coordinates in which they equal zero.

If the quadratic form involves 3 variables, then similar comments apply, except that one can obtain a zero determinant by more complex linear combinations of vectors. If the quadratic form involves 4 or more variables, then the connection with determinants tends to be less illuminating, as we discussed before.

Another way to convert degree 2 polynomials into quadratic forms. If it seems complicated to perform the procedures we described earlier to define a new set of coordinates in which the lower-order equal zero, then a "trick" that is sometimes employed is to add one extra coordinate and force it always to equal 1. In the case, of $x \in \mathbb{R}^2$, this means working with vectors $x \in \mathbb{R}^3$, where we add a third dimension to the vector x, but we always have the third component, call it x_3, equal to 1.

If we do this, then Equation (5) can be viewed in one of two ways.

- As a quadratic function in two variables $x \in \mathbb{R}^2$ with lower-order terms:
$$(x_1\ x_2) \begin{pmatrix} A & B/2 \\ B/2 & C \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + (D\ E) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + F = 0.$$
- As a quadratic form in three variables $x \in \mathbb{R}^3$ with no lower order terms but where the last component is always equal to one:
$$(x_1\ x_2\ 1) \begin{pmatrix} A & B/2 & D/2 \\ B/2 & C & E/2 \\ D/2 & E/2 & F \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix} = 0.$$

Problem. If it isn't clear that both of these expressions equal Equation (5), then simply multiply everything out.

Remark. If we let Ω be the entire 3×3 matrix, then the condition $|\Omega| = 0$ says that the three vectors are linearly-dependent.

For vectors in three dimensions, this isn't just that one is a scalar multiple of another, but that one of them can be written as a linear combination of the other two. In this case, we can have a point through the origin, or a line through the origin, etc.

Remark. BTW, just adding an extra dimension to all your vectors and matrices might seem strange, and it might seem like a big deal with you are working with vectors in $\mathbb{R}$ or $\mathbb{R}^2$ or $\mathbb{R}^3$. If your vectors are in R10;465, however, it seems like a relatively-minor change to work with vectors in $\mathbb{R}^{10,466}$, and doing this often makes things easier (which is the real justification).

7 Diagonalizatlion

We explain how to diagonalize a matrix if it is diagonalizable. As an example, we solve the following problem.

Diagonalize the matrix

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

by finding a nonsingular matrix S and a diagonal matrix D such that $S^{-1}AS = D$.

Diagonalization Problems and Examples: A Hermitian Matrix can be diagonalized by a unitary matrix More diagonalization problems

Here we explain how to diagonalize a matrix. We only describe the procedure of diagonalization, and no justification will be given. The process can be summarized as follows. A concrete example is provided below, and several exercise problems are presented at the end.

Diagonalization Procedure

Let A be the $n \times n$ matrix that you want to diagonalize (if possible).

Step 1. Find the characteristic polynomial $p(t)$ of A.

Step 2. Find eigenvalues λ of the matrix A and their algebraic multiplicities from the characteristic polynomial $p(t)$.

Step 3. For each eigenvalue λ of A, find a basis of the eigenspace E_λ.

Step 4. If there is an eigenvalue λ such that the geometric multiplicity of λ, $\dim(E_\lambda)$, is less than the algebraic multiplicity of λ, then the matrix A is not diagonalizable. If not, A is diagonalizable, and proceed to the next step.

Step 5. If we combine all basis vectors for all eigenspaces, we obtained n linearly independent eigenvectors $v_1, v_2, ..., v_n$.

Step 6. Define the nonsingular matrix

$$S = [v_1 \; v_2 \; ... \; v_n]$$

.

Step 7. Define the diagonal matrix D, whose (i, i)-entry is the eigenvalue λ such that the i-th column vector v_i is in the eigenspace E_λ.

Step 8. Then the matrix A is diagonalized as

$$S^{-1}AS = D.$$

Now let us examine these steps with an example.

Example 18.

Let us consider the following 3×3 matrix:

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

We want to diagonalize the matrix if possible.

Step 1. Find the characteristic polynomial.

$$\begin{aligned} p(t) = |A - \lambda I| &= \begin{vmatrix} 4-\lambda & -3 & -3 \\ 3 & -2-\lambda & -3 \\ -1 & 1 & 2-\lambda \end{vmatrix} \\ &= (4-\lambda)(-2-\lambda)(2-\lambda) - 9 - 9 \\ &\qquad - 3(-2-\lambda) + 9(2-\lambda) + 3(4-\lambda) \\ &= (4-\lambda)(-2-\lambda)(2-\lambda) + 18 - 9\lambda \\ &= (2-\lambda)((4-\lambda)(-2-\lambda) + 9) \\ &= (2-\lambda)(\lambda-1)^2. \end{aligned}$$

Hence the characteristic polynomial p(t) of A is

$$|A - \lambda I| = (2-\lambda)(\lambda-1)^2.$$

Step 2. Find eigenvalues.

From the characteristic polynomial obtained in Step 1, we see that eigenvalues are:

$$\lambda = 1 \text{ with algebraic multiplicity } 2$$

and

$$\lambda = 2 \text{ with algebraic multiplicity } 1.$$

We also see that the algebraic multiplicity of the eigenspace corresponding to the eigenvalue $\lambda = 1$ is 2 and that corresponding to the eigenvalue $\lambda = 2$ is 1.

Step 3. Find eigenvectors.

$\underline{\lambda = 1}$

We have

$$\begin{pmatrix} 3 & -3 & -3 \\ 3 & -3 & -3 \\ -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

It follows that

$$x_1 = x_2 + x_3.$$

The eigenvectors associated with $\lambda = 1$ are

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

$\underline{\lambda = 2}$

We have

$$A - \lambda I = \begin{pmatrix} 2 & -3 & -3 \\ 3 & -4 & -3 \\ -1 & 1 & 0 \end{pmatrix}$$

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

$$\sim \begin{pmatrix} 1 & -2 & -3 \\ 0 & 2 & 6 \\ 0 & -1 & -3 \end{pmatrix}$$

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

$$\sim \begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{pmatrix}.$$

This implies that

$$\begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

It follows that

$$x_1 = -3x_3, \quad x_2 = -3x_3.$$

Therefore, the eigenvector associated with $\lambda = 2$ is

$$v_3 = \begin{pmatrix} -3 \\ -3 \\ 1 \end{pmatrix}$$

From this we see that the set is a basis for the eigenspace E_2 and the geometric multiplicity is 1.

Since for both eigenvalues, the geometric multiplicity is equal to the algebraic multiplicity, the matrix A is not defective, and hence diagonalizable.

Step 4. Determine linearly independence of eigenvectors.

Let $S = (v_1\ v_2\ v_3)$. Then

$$|S| = \begin{vmatrix} 1 & 1 & -3 \\ 1 & 0 & -3 \\ 0 & 1 & 1 \end{vmatrix} = -3 - 1 + 3 = -1 \neq 0.$$

So, v_1, v_2, v_3 are linearly independent eigenvectors.

Step 5. If we combine all basis vectors for all eigenspaces, we obtained n linearly independent eigenvectors $v_1, v_2, ..., v_n$.

Step 6. Define the nonsingular matrix

$$S = [v_1\ v_2\ ...\ v_n]$$

.

Step 7. Define the diagonal matrix D.

Define the diagonal matrix

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

Step 8. Then the matrix A is diagonalized as

$$S^{-1}AS = D.$$