b. Find a binary linear (5, 2)-code that can correct one error.

Exercise 8.8.14 Find the standard generator matrix G and the parity-check matrix H for each of the following systematic codes:

a. $\{00000,\ 11111\}$ over $\mathbb{Z}_2$.

b. Any systematic $(n,\ 1)$-code where $n \geq 2$.

c. The code in Exercise 8.8.10(a).

d. The code in Exercise 8.8.10(b).

Exercise 8.8.15 Let $\mathbf{c}$ be a word in F^n. Show that $B_t(\mathbf{c}) = \mathbf{c} + B_t(\mathbf{0})$, where we write

$$\mathbf{c} + B_t(\mathbf{0}) = \{\mathbf{c} + \mathbf{v} \mid \mathbf{v} \text{ in } B_t(\mathbf{0})\}$$

Exercise 8.8.16 If a $(n,\ k)$-code has two standard generator matrices G and G_1, show that $G = G_1$.

Exercise 8.8.17 Let C be a binary linear n-code (over $\mathbb{Z}_2$). Show that either each word in C has even weight, or half the words in C have even weight and half have odd weight. [*Hint*: The dimension theorem.]

8.9 An Application to Quadratic Forms

An expression like $x_1^2 + x_2^2 + x_3^2 - 2x_1x_3 + x_2x_3$ is called a quadratic form in the variables x_1, x_2, and x_3. In this section we show that new variables y_1, y_2, and y_3 can always be found so that the quadratic form, when expressed in terms of the new variables, has no cross terms y_1y_2, y_1y_3, or y_2y_3. Moreover, we do this for forms involving any finite number of variables using orthogonal diagonalization. This has far-reaching applications; quadratic forms arise in such diverse areas as statistics, physics, the theory of functions of several variables, number theory, and geometry.

Definition 8.21 Quadratic Form

A ***quadratic form*** q *in the* n *variables* $x_1,\ x_2,\ \dots,\ x_n$ *is a linear combination of terms* $x_1^2,\ x_2^2,\ \dots,\ x_n^2$, *and cross terms* $x_1x_2,\ x_1x_3,\ x_2x_3,\ \dots$.

If $n = 3$, q has the form

$$q = a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2 + a_{12}x_1x_2 + a_{21}x_2x_1 + a_{13}x_1x_3 + a_{31}x_3x_1 + a_{23}x_2x_3 + a_{32}x_3x_2$$

In general

$$q = a_{11}x_1^2 + a_{22}x_2^2 + \cdots + a_{nn}x_n^2 + a_{12}x_1x_2 + a_{13}x_1x_3 + \cdots$$

This sum can be written compactly as a matrix product

$$q = q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$$

where $\mathbf{x} = (x_1,\ x_2,\ \dots,\ x_n)$ is thought of as a column, and $A = \left[a_{ij}\right]$ is a real $n \times n$ matrix. Note that if $i \neq j$, two separate terms $a_{ij}x_ix_j$ and $a_{ji}x_jx_i$ are listed, each of which involves x_ix_j, and they can (rather cleverly) be replaced by

$$\tfrac{1}{2}(a_{ij} + a_{ji})x_ix_j \quad \text{and} \quad \tfrac{1}{2}(a_{ij} + a_{ji})x_jx_i$$

respectively, *without altering the quadratic form*. Hence there is no loss of generality in assuming that x_ix_j and x_jx_i have the same coefficient in the sum for q. In other words, **we may assume that A is symmetric**.

Example 8.9.1

Write $q = x_1^2 + 3x_3^2 + 2x_1x_2 - x_1x_3$ in the form $q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$, where A is a symmetric 3×3 matrix.

Solution. The cross terms are $2x_1x_2 = x_1x_2 + x_2x_1$ and $-x_1x_3 = -\frac{1}{2}x_1x_3 - \frac{1}{2}x_3x_1$. Of course, x_2x_3 and x_3x_2 both have coefficient zero, as does x_2^2. Hence

$$q(\mathbf{x}) = \left[\begin{array}{ccc} x_1 & x_2 & x_3 \end{array}\right] \left[\begin{array}{rrr} 1 & 1 & -\frac{1}{2} \\ 1 & 0 & 0 \\ -\frac{1}{2} & 0 & 3 \end{array}\right] \left[\begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array}\right]$$

is the required form (verify).

We shall assume from now on that all quadratic forms are given by

$$q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$$

where A is symmetric. Given such a form, the problem is to find new variables $y_1, y_2, \dots, y_n$, related to $x_1, x_2, \dots, x_n$, with the property that when q is expressed in terms of $y_1, y_2, \dots, y_n$, there are no cross terms. If we write

$$\mathbf{y} = (y_1, y_2, \dots, y_n)^T$$

this amounts to asking that $q = \mathbf{y}^T D\mathbf{y}$ where D is diagonal. It turns out that this can always be accomplished and, not surprisingly, that D is the matrix obtained when the symmetric matrix A is orthogonally diagonalized. In fact, as Theorem 8.2.2 shows, a matrix P can be found that is orthogonal (that is, $P^{-1} = P^T$) and diagonalizes A:

$$P^T AP = D = \left[\begin{array}{cccc} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{array}\right]$$

The diagonal entries $\lambda_1, \lambda_2, \dots, \lambda_n$ are the (not necessarily distinct) eigenvalues of A, repeated according to their multiplicities in $c_A(x)$, and the columns of P are corresponding (orthonormal) eigenvectors of A. As A is symmetric, the λ_i are real by Theorem 5.5.7.

Now define new variables $\mathbf{y}$ by the equations

$$\mathbf{x} = P\mathbf{y} \quad \text{equivalently} \quad \mathbf{y} = P^T\mathbf{x}$$

Then substitution in $q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ gives

$$q = (P\mathbf{y})^T A(P\mathbf{y}) = \mathbf{y}^T (P^T AP)\mathbf{y} = \mathbf{y}^T D\mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$$

Hence this change of variables produces the desired simplification in q.

Theorem 8.9.1: Diagonalization Theorem

Let $q = \mathbf{x}^T A\mathbf{x}$ be a quadratic form in the variables $x_1, x_2, \dots, x_n$, where $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ and A is a symmetric $n \times n$ matrix. Let P be an orthogonal matrix such that $P^T AP$ is diagonal, and *define new variables* $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$ *by*

$$\mathbf{x} = P\mathbf{y} \quad \text{equivalently} \quad \mathbf{y} = P^T\mathbf{x}$$

If q *is expressed in terms of these new variables* $y_1, y_2, \dots, y_n$*, the result is*

$$q = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$$

where $\lambda_1, \lambda_2, \dots, \lambda_n$ *are the eigenvalues of* A *repeated according to their multiplicities.*

Let $q = \mathbf{x}^T A\mathbf{x}$ be a quadratic form where A is a symmetric matrix and let $\lambda_1, \dots, \lambda_n$ be the (real) eigenvalues of A repeated according to their multiplicities. A corresponding set $\{\mathbf{f}_1, \dots, \mathbf{f}_n\}$ of orthonormal eigenvectors for A is called a set of **principal axes** for the quadratic form q. (The reason for the name will become clear later.) The orthogonal matrix P in Theorem 8.9.1 is given as $P = \begin{bmatrix} \mathbf{f}_1 & \cdots & \mathbf{f}_n \end{bmatrix}$, so the variables X and Y are related by

$$\mathbf{x} = P\mathbf{y} = \begin{bmatrix} \mathbf{f}_1 & \mathbf{f}_2 & \cdots & \mathbf{f}_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = y_1\mathbf{f}_1 + y_2\mathbf{f}_2 + \cdots + y_n\mathbf{f}_n$$

Thus the new variables y_i are the coefficients when $\mathbf{x}$ is expanded in terms of the orthonormal basis $\{\mathbf{f}_1, \dots, \mathbf{f}_n\}$ of $\mathbb{R}^n$. In particular, the coefficients y_i are given by $y_i = \mathbf{x} \cdot \mathbf{f}_i$ by the expansion theorem (Theorem 5.3.6). Hence q itself is easily computed from the eigenvalues λ_i and the principal axes $\mathbf{f}_i$:

$$q = q(\mathbf{x}) = \lambda_1(\mathbf{x} \cdot \mathbf{f}_1)^2 + \cdots + \lambda_n(\mathbf{x} \cdot \mathbf{f}_n)^2$$

Example 8.9.2

Find new variables y_1, y_2, y_3, and y_4 such that

$$q = 3(x_1^2 + x_2^2 + x_3^2 + x_4^2) + 2x_1x_2 - 10x_1x_3 + 10x_1x_4 + 10x_2x_3 - 10x_2x_4 + 2x_3x_4$$

has diagonal form, and find the corresponding principal axes.

Solution. The form can be written as $q = \mathbf{x}^T A\mathbf{x}$, where

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} 3 & 1 & -5 & 5 \\ 1 & 3 & 5 & -5 \\ -5 & 5 & 3 & 1 \\ 5 & -5 & 1 & 3 \end{bmatrix}$$

A routine calculation yields

$$c_A(x) = \det(xI - A) = (x-12)(x+8)(x-4)^2$$

so the eigenvalues are $\lambda_1 = 12$, $\lambda_2 = -8$, and $\lambda_3 = \lambda_4 = 4$. Corresponding orthonormal eigenvectors are the principal axes:

$$\mathbf{f}_1 = \frac{1}{2}\begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix} \quad \mathbf{f}_2 = \frac{1}{2}\begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix} \quad \mathbf{f}_3 = \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad \mathbf{f}_4 = \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}$$

The matrix

$$P = \begin{bmatrix} \mathbf{f}_1 & \mathbf{f}_2 & \mathbf{f}_3 & \mathbf{f}_4 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & -1 \end{bmatrix}$$

is thus orthogonal, and $P^{-1}AP = P^TAP$ is diagonal. Hence the new variables $\mathbf{y}$ and the old variables $\mathbf{x}$ are related by $\mathbf{y} = P^T\mathbf{x}$ and $\mathbf{x} = P\mathbf{y}$. Explicitly,

$$\begin{aligned} y_1 &= \tfrac{1}{2}(x_1 - x_2 - x_3 + x_4) & \qquad x_1 &= \tfrac{1}{2}(y_1 + y_2 + y_3 + y_4) \\ y_2 &= \tfrac{1}{2}(x_1 - x_2 + x_3 - x_4) & x_2 &= \tfrac{1}{2}(-y_1 - y_2 + y_3 + y_4) \\ y_3 &= \tfrac{1}{2}(x_1 + x_2 + x_3 + x_4) & x_3 &= \tfrac{1}{2}(-y_1 + y_2 + y_3 - y_4) \\ y_4 &= \tfrac{1}{2}(x_1 + x_2 - x_3 - x_4) & x_4 &= \tfrac{1}{2}(y_1 - y_2 + y_3 - y_4) \end{aligned}$$

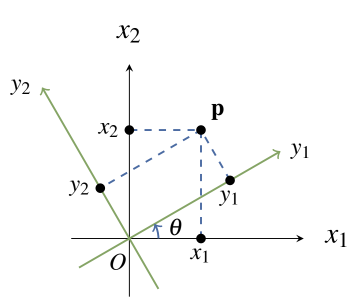
If these x_i are substituted in the original expression for q, the result is

$$q = 12y_1^2 - 8y_2^2 + 4y_3^2 + 4y_4^2$$

This is the required diagonal form.

It is instructive to look at the case of quadratic forms in two variables x_1 and x_2. Then the principal axes can always be found by rotating the x_1 and x_2 axes counterclockwise about the origin through an angle θ. This rotation is a linear transformation $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$, and it is shown in Theorem 2.6.4 that R_θ has matrix $P = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$. If $\{\mathbf{e}_1, \mathbf{e}_2\}$ denotes the standard basis of $\mathbb{R}^2$, the rotation produces a new basis $\{\mathbf{f}_1, \mathbf{f}_2\}$ given by

$$\mathbf{f}_1 = R_\theta(\mathbf{e}_1) = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \quad \text{and} \quad \mathbf{f}_2 = R_\theta(\mathbf{e}_2) = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix} \tag{8.7}$$



Given a point $\mathbf{p} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2$ in the original system, let y_1 and y_2 be the coordinates of $\mathbf{p}$ in the new system (see the diagram). That is,

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{p} = y_1\mathbf{f}_1 + y_2\mathbf{f}_2 = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \tag{8.8}$$

Writing $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$, this reads $\mathbf{x} = P\mathbf{y}$ so, since P is orthogonal, this is the change of variables formula for the rotation as in Theorem 8.9.1.

If $r \neq 0 \neq s$, the graph of the equation $rx_1^2 + sx_2^2 = 1$ is called an **ellipse** if $rs > 0$ and a **hyperbola** if $rs < 0$. More generally, given a quadratic form

$$q = ax_1^2 + bx_1x_2 + cx_2^2 \quad \text{where not all of } a,\ b, \text{ and } c \text{ are zero}$$

the graph of the equation $q = 1$ is called a **conic**. We can now completely describe this graph. There are two special cases which we leave to the reader.

1. *If exactly one of a and c is zero, then the graph of $q = 1$ is a* **parabola**.

So we assume that $a \neq 0$ and $c \neq 0$. In this case, the description depends on the quantity $b^2 - 4ac$, called the **discriminant** of the quadratic form q.

2. *If $b^2 - 4ac = 0$, then either both $a \geq 0$ and $c \geq 0$, or both $a \leq 0$ and $c \leq 0$. Hence $q = (\sqrt{a}x_1 + \sqrt{c}x_2)^2$ or $q = (\sqrt{-a}x_1 + \sqrt{-c}x_2)^2$, so the graph of $q = 1$ is a* **pair of straight lines** *in either case.*

So we also assume that $b^2 - 4ac \neq 0$. But then the next theorem asserts that there exists a rotation of the plane about the origin which transforms the equation $ax_1^2 + bx_1x_2 + cx_2^2 = 1$ into either an ellipse or a hyperbola, and the theorem also provides a simple way to decide which conic it is.

Theorem 8.9.2

Consider the quadratic form $q = ax_1^2 + bx_1x_2 + cx_2^2$ where a, c, and $b^2 - 4ac$ are all nonzero.

1. *There is a counterclockwise rotation of the coordinate axes about the origin such that, in the new coordinate system, q has no cross term.*
2. *The graph of the equation*
$$ax_1^2 + bx_1x_2 + cx_2^2 = 1$$
is an ellipse if $b^2 - 4ac < 0$ and an hyperbola if $b^2 - 4ac > 0$.

Proof. If $b = 0$, q *already* has no cross term and (1) and (2) are clear. So assume $b \neq 0$. The matrix $A = \begin{bmatrix} a & \frac{1}{2}b \\ \frac{1}{2}b & c \end{bmatrix}$ of q has characteristic polynomial $c_A(x) = x^2 - (a+c)x - \frac{1}{4}(b^2 - 4ac)$. If we write $d = \sqrt{b^2 + (a-c)^2}$ for convenience; then the quadratic formula gives the eigenvalues

$$\lambda_1 = \frac{1}{2}[a + c - d] \quad \text{and} \quad \lambda_2 = \frac{1}{2}[a + c + d]$$

with corresponding principal axes

$$\mathbf{f}_1 = \frac{1}{\sqrt{b^2+(a-c-d)^2}} \begin{bmatrix} a - c - d \\ b \end{bmatrix} \quad \text{and}$$
$$\mathbf{f}_2 = \frac{1}{\sqrt{b^2+(a-c-d)^2}} \begin{bmatrix} -b \\ a - c - d \end{bmatrix}$$

as the reader can verify. These agree with equation (8.7) above if θ is an angle such that

$$\cos\theta = \frac{a-c-d}{\sqrt{b^2+(a-c-d)^2}} \quad \text{and} \quad \sin\theta = \frac{b}{\sqrt{b^2+(a-c-d)^2}}$$

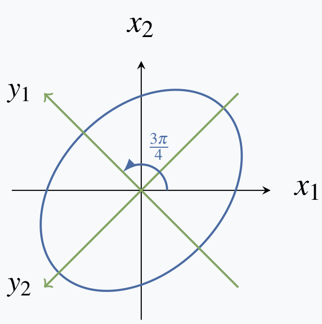
Then $P = \begin{bmatrix} \mathbf{f}_1 & \mathbf{f}_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ diagonalizes A and equation (8.8) becomes the formula $\mathbf{x} = P\mathbf{y}$ in Theorem 8.9.1. This proves (1).

Finally, A is similar to $\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$ so $\lambda_1\lambda_2 = \det A = \frac{1}{4}(4ac - b^2)$. Hence the graph of $\lambda_1 y_1^2 + \lambda_2 y_2^2 = 1$ is an ellipse if $b^2 < 4ac$ and an hyperbola if $b^2 > 4ac$. This proves (2). □

Example 8.9.3

Consider the equation $x^2 + xy + y^2 = 1$. Find a rotation so that the equation has no cross term.

Solution.



Here $a = b = c = 1$ in the notation of Theorem 8.9.2, so $\cos\theta = \frac{-1}{\sqrt{2}}$ and $\sin\theta = \frac{1}{\sqrt{2}}$. Hence $\theta = \frac{3\pi}{4}$ will do it. The new variables are $y_1 = \frac{1}{\sqrt{2}}(x_2 - x_1)$ and $y_2 = \frac{-1}{\sqrt{2}}(x_2 + x_1)$ by (8.8), and the equation becomes $y_1^2 + 3y_2^2 = 2$. The angle θ has been chosen such that the new y_1 and y_2 axes are the axes of symmetry of the ellipse (see the diagram). The eigenvectors $\mathbf{f}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and $\mathbf{f}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ -1 \end{bmatrix}$ point along these axes of symmetry, and this is the reason for the name *principal axes*.

The determinant of any orthogonal matrix P is either 1 or -1 (because $PP^T = I$). The orthogonal matrices $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ arising from rotations all have determinant 1. More generally, given any quadratic form $q = \mathbf{x}^T A\mathbf{x}$, the orthogonal matrix P such that $P^T AP$ is diagonal can always be chosen so that $\det P = 1$ by interchanging two eigenvalues (and hence the corresponding columns of P). It is shown in Theorem 10.4.4 that orthogonal 2×2 matrices with determinant 1 correspond to rotations. Similarly, it can be shown that orthogonal 3×3 matrices with determinant 1 correspond to rotations about a line through the origin. This extends Theorem 8.9.2: Every quadratic form in two or three variables can be diagonalized by a rotation of the coordinate system.

Congruence

We return to the study of quadratic forms in general.

Theorem 8.9.3

If $q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is a quadratic form given by a symmetric matrix A, then A is uniquely determined by q.

Proof. Let $q(\mathbf{x}) = \mathbf{x}^T B\mathbf{x}$ for all $\mathbf{x}$ where $B^T = B$. If $C = A - B$, then $C^T = C$ and $\mathbf{x}^T C\mathbf{x} = 0$ for all $\mathbf{x}$. We must show that $C = 0$. Given $\mathbf{y}$ in $\mathbb{R}^n$,

$$\begin{aligned} 0 = (\mathbf{x}+\mathbf{y})^T C(\mathbf{x}+\mathbf{y}) &= \mathbf{x}^T C\mathbf{x} + \mathbf{x}^T C\mathbf{y} + \mathbf{y}^T C\mathbf{x} + \mathbf{y}^T C\mathbf{y} \\ &= \mathbf{x}^T C\mathbf{y} + \mathbf{y}^T C\mathbf{x} \end{aligned}$$

But $\mathbf{y}^T C\mathbf{x} = (\mathbf{x}^T C\mathbf{y})^T = \mathbf{x}^T C\mathbf{y}$ (it is 1×1). Hence $\mathbf{x}^T C\mathbf{y} = 0$ for all $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$. If $\mathbf{e}_j$ is column j of I_n, then the (i, j)-entry of C is $\mathbf{e}_i^T C\mathbf{e}_j = 0$. Thus $C = 0$. □

Hence we can speak of *the* symmetric matrix of a quadratic form.

On the other hand, a quadratic form q in variables x_i can be written in several ways as a linear combination of squares of new variables, even if the new variables are required to be linear combinations of the x_i. For example, if $q = 2x_1^2 - 4x_1x_2 + x_2^2$ then

$$q = 2(x_1 - x_2)^2 - x_2^2 \quad \text{and} \quad q = -2x_1^2 + (2x_1 - x_2)^2$$

The question arises: How are these changes of variables related, and what properties do they share? To investigate this, we need a new concept.

Let a quadratic form $q = q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ be given in terms of variables $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$. If the new variables $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$ are to be linear combinations of the x_i, then $\mathbf{y} = A\mathbf{x}$ for some $n \times n$ matrix A. Moreover, since we want to be able to solve for the x_i in terms of the y_i, we ask that the matrix A be invertible. Hence suppose U is an invertible matrix and that the new variables $\mathbf{y}$ are given by

$$\mathbf{y} = U^{-1}\mathbf{x}, \qquad \text{equivalently } \mathbf{x} = U\mathbf{y}$$

In terms of these new variables, q takes the form

$$q = q(\mathbf{x}) = (U\mathbf{y})^T A(U\mathbf{y}) = \mathbf{y}^T(U^T AU)\mathbf{y}$$

That is, q has matrix $U^T AU$ with respect to the new variables $\mathbf{y}$. Hence, to study changes of variables in quadratic forms, we study the following relationship on matrices: Two $n \times n$ matrices A and B are called **congruent**, written $A \overset{c}{\sim} B$, if $B = U^T AU$ for some invertible matrix U. Here are some properties of congruence:

1. *$A \overset{c}{\sim} A$ for all A.*
2. *If $A \overset{c}{\sim} B$, then $B \overset{c}{\sim} A$.*

3. *If* $A \overset{c}{\sim} B$ *and* $B \overset{c}{\sim} C$, *then* $A \overset{c}{\sim} C$.

4. *If* $A \overset{c}{\sim} B$, *then* A *is symmetric if and only if* B *is symmetric*.

5. *If* $A \overset{c}{\sim} B$, *then* $\text{rank } A = \text{rank } B$.

The converse to (5) can fail even for symmetric matrices.

Example 8.9.4

The symmetric matrices $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ have the same rank but are not congruent. Indeed, if $A \overset{c}{\sim} B$, an invertible matrix U exists such that $B = U^T AU = U^T U$. But then $-1 = \det B = (\det U)^2$, a contradiction.

The key distinction between A and B in Example 8.9.4 is that A has two positive eigenvalues (counting multiplicities) whereas B has only one.

Theorem 8.9.4: Sylvester's Law of Inertia

If $A \overset{c}{\sim} B$, *then* A *and* B *have the same number of positive eigenvalues, counting multiplicities.*

The proof is given at the end of this section.

The **index** of a symmetric matrix A is the number of positive eigenvalues of A. If $q = q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is a quadratic form, the **index** and **rank** of q are defined to be, respectively, the index and rank of the matrix A. As we saw before, if the variables expressing a quadratic form q are changed, the new matrix is congruent to the old one. Hence the index and rank depend only on q and not on the way it is expressed.

Now let $q = q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ be any quadratic form in n variables, of index k and rank r, where A is symmetric. We claim that new variables $\mathbf{z}$ can be found so that q is **completely diagonalized**—that is,

$$q(\mathbf{z}) = z_1^2 + \cdots + z_k^2 - z_{k+1}^2 - \cdots - z_r^2$$

If $k \leq r \leq n$, let $D_n(k, r)$ denote the $n \times n$ diagonal matrix whose main diagonal consists of k ones, followed by $r - k$ minus ones, followed by $n - r$ zeros. Then we seek new variables $\mathbf{z}$ such that

$$q(\mathbf{z}) = \mathbf{z}^T D_n(k, r)\mathbf{z}$$

To determine $\mathbf{z}$, first diagonalize A as follows: Find an orthogonal matrix P_0 such that

$$P_0^T AP_0 = D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_r, 0, \dots, 0)$$

is diagonal with the nonzero eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_r$ of A on the main diagonal (followed by $n - r$ zeros). By reordering the columns of P_0, if necessary, we may assume that $\lambda_1, \dots, \lambda_k$ are positive and $\lambda_{k+1}, \dots, \lambda_r$ are negative. This being the case, let D_0 be the $n \times n$ diagonal matrix

$$D_0 = \text{diag}\left(\frac{1}{\sqrt{\lambda_1}}, \dots, \frac{1}{\sqrt{\lambda_k}}, \frac{1}{\sqrt{-\lambda_{k+1}}}, \dots, \frac{1}{\sqrt{-\lambda_r}}, 1, \dots, 1\right)$$

Then $D_0^T D D_0 = D_n(k,\ r)$, so if new variables $\mathbf{z}$ are given by $\mathbf{x} = (P_0 D_0)\mathbf{z}$, we obtain

$$q(\mathbf{z}) = \mathbf{z}^T D_n(k,\ r)\mathbf{z} = z_1^2 + \cdots + z_k^2 - z_{k+1}^2 - \cdots - z_r^2$$

as required. Note that the change-of-variables matrix $P_0 D_0$ from $\mathbf{z}$ to $\mathbf{x}$ has orthogonal columns (in fact, scalar multiples of the columns of P_0).

Example 8.9.5

Completely diagonalize the quadratic form q in Example 8.9.2 and find the index and rank.

Solution. In the notation of Example 8.9.2, the eigenvalues of the matrix A of q are 12, −8, 4, 4; so the index is 3 and the rank is 4. Moreover, the corresponding orthogonal eigenvectors are $\mathbf{f}_1$, $\mathbf{f}_2$, $\mathbf{f}_3$ (see Example 8.9.2), and $\mathbf{f}_4$. Hence $P_0 = \left[\begin{array}{cccc} \mathbf{f}_1 & \mathbf{f}_3 & \mathbf{f}_4 & \mathbf{f}_2 \end{array}\right]$ is orthogonal and

$$P_0^T A P_0 = \text{diag}\,(12,\ 4,\ 4,\ -8)$$

As before, take $D_0 = \text{diag}\,(\frac{1}{\sqrt{12}},\ \frac{1}{2},\ \frac{1}{2},\ \frac{1}{\sqrt{8}})$ and define the new variables $\mathbf{z}$ by $\mathbf{x} = (P_0 D_0)\mathbf{z}$. Hence the new variables are given by $\mathbf{z} = D_0^{-1} P_0^T \mathbf{x}$. The result is

$$\begin{aligned}
z_1 &= \sqrt{3}(x_1 - x_2 - x_3 + x_4) \\
z_2 &= x_1 + x_2 + x_3 + x_4 \\
z_3 &= x_1 + x_2 - x_3 - x_4 \\
z_4 &= \sqrt{2}(x_1 - x_2 + x_3 - x_4)
\end{aligned}$$

This discussion gives the following information about symmetric matrices.

Theorem 8.9.5

Let A and B be symmetric $n \times n$ matrices, and let $0 \leq k \leq r \leq n$.

1. *A has index k and rank r if and only if $A \overset{c}{\sim} D_n(k,\ r)$.*
2. *$A \overset{c}{\sim} B$ if and only if they have the same rank and index.*

Proof.

1. If A has index k and rank r, take $U = P_0 D_0$ where P_0 and D_0 are as described prior to Example 8.9.5. Then $U^T A U = D_n(k,\ r)$. The converse is true because $D_n(k,\ r)$ has index k and rank r (using Theorem 8.9.4).
2. If A and B both have index k and rank r, then $A \overset{c}{\sim} D_n(k,\ r) \overset{c}{\sim} B$ by (1). The converse was given earlier.

□

Proof of Theorem 8.9.4.

By Theorem 8.9.1, $A \overset{c}{\sim} D_1$ and $B \overset{c}{\sim} D_2$ where D_1 and D_2 are diagonal and have the same eigenvalues as A and B, respectively. We have $D_1 \overset{c}{\sim} D_2$, (because $A \overset{c}{\sim} B$), so we may assume that A and B are both diagonal. Consider the quadratic form $q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$. If A has k positive eigenvalues, q has the form

$$q(\mathbf{x}) = a_1x_1^2 + \cdots + a_kx_k^2 - a_{k+1}x_{k+1}^2 - \cdots - a_rx_r^2, \quad a_i > 0$$

where $r = \text{rank } A = \text{rank } B$. The subspace $W_1 = \{\mathbf{x} \mid x_{k+1} = \cdots = x_r = 0\}$ of $\mathbb{R}^n$ has dimension $n - r + k$ and satisfies $q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$ in W_1.

On the other hand, if $B = U^T AU$, define new variables $\mathbf{y}$ by $\mathbf{x} = U\mathbf{y}$. If B has k' positive eigenvalues, q has the form

$$q(\mathbf{x}) = b_1y_1^2 + \cdots + b_{k'}y_{k'}^2 - b_{k'+1}y_{k'+1}^2 - \cdots - b_ry_r^2, \quad b_i > 0$$

Let $\mathbf{f}_1, \dots, \mathbf{f}_n$ denote the columns of U. They are a basis of $\mathbb{R}^n$ and

$$\mathbf{x} = U\mathbf{y} = \left[\begin{array}{ccc} \mathbf{f}_1 & \cdots & \mathbf{f}_n \end{array}\right] \left[\begin{array}{c} y_1 \\ \vdots \\ y_n \end{array}\right] = y_1\mathbf{f}_1 + \cdots + y_n\mathbf{f}_n$$

Hence the subspace $W_2 = \text{span}\{\mathbf{f}_{k'+1}, \dots, \mathbf{f}_r\}$ satisfies $q(\mathbf{x}) < 0$ for all $\mathbf{x} \neq \mathbf{0}$ in W_2. Note $\dim W_2 = r - k'$. It follows that W_1 and W_2 have only the zero vector in common. Hence, if B_1 and B_2 are bases of W_1 and W_2, respectively, then (Exercise 6.3.33) $B_1 \cup B_2$ is an independent set of $(n - r + k) + (r - k') = n + k - k'$ vectors in $\mathbb{R}^n$. This implies that $k \leq k'$, and a similar argument shows $k' \leq k$. □

Exercises for 8.9

Exercise 8.9.1 In each case, find a symmetric matrix A such that $q = \mathbf{x}^T B\mathbf{x}$ takes the form $q = \mathbf{x}^T A\mathbf{x}$.

a. $\left[\begin{array}{cc} 1 & 1 \\ 0 & 1 \end{array}\right]$

b. $\left[\begin{array}{cc} 1 & 1 \\ -1 & 2 \end{array}\right]$

c. $\left[\begin{array}{ccc} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{array}\right]$

d. $\left[\begin{array}{ccc} 1 & 2 & -1 \\ 4 & 1 & 0 \\ 5 & -2 & 3 \end{array}\right]$

Exercise 8.9.2 In each case, find a change of variables that will diagonalize the quadratic form q. Determine the index and rank of q.

a. $q = x_1^2 + 2x_1x_2 + x_2^2$

b. $q = x_1^2 + 4x_1x_2 + x_2^2$

c. $q = x_1^2 + x_2^2 + x_3^2 - 4(x_1x_2 + x_1x_3 + x_2x_3)$

d. $q = 7x_1^2 + x_2^2 + x_3^2 + 8x_1x_2 + 8x_1x_3 - 16x_2x_3$

e. $q = 2(x_1^2 + x_2^2 + x_3^2 - x_1x_2 + x_1x_3 - x_2x_3)$

f. $q = 5x_1^2 + 8x_2^2 + 5x_3^2 - 4(x_1x_2 + 2x_1x_3 + x_2x_3)$

g. $q = x_1^2 - x_3^2 - 4x_1x_2 + 4x_2x_3$

h. $q = x_1^2 + x_3^2 - 2x_1x_2 + 2x_2x_3$

Exercise 8.9.3 For each of the following, write the equation in terms of new variables so that it is in standard position, and identify the curve.

a. $xy = 1$

b. $3x^2 - 4xy = 2$

c. $6x^2 + 6xy - 2y^2 = 5$

d. $2x^2 + 4xy + 5y^2 = 1$

Exercise 8.9.4 Consider the equation $ax^2 + bxy + cy^2 = d$, where $b \neq 0$. Introduce new variables x_1 and y_1 by rotating the axes counterclockwise through an angle θ. Show that the resulting equation has no x_1y_1-term if θ is given by

$$\cos 2\theta = \frac{a-c}{\sqrt{b^2+(a-c)^2}}$$
$$\sin 2\theta = \frac{b}{\sqrt{b^2+(a-c)^2}}$$

[*Hint*: Use equation (8.8) preceding Theorem 8.9.2 to get x and y in terms of x_1 and y_1, and substitute.]

Exercise 8.9.5 Prove properties (1)–(5) preceding Example 8.9.4.

Exercise 8.9.6 If $A \overset{c}{\sim} B$ show that A is invertible if and only if B is invertible.

Exercise 8.9.7 If $\mathbf{x} = (x_1, \dots, x_n)^T$ is a column of variables, $A = A^T$ is $n \times n$, B is $1 \times n$, and c is a constant, $\mathbf{x}^T A\mathbf{x} + B\mathbf{x} = c$ is called a **quadratic equation** in the variables x_i.

a. Show that new variables $y_1, \dots, y_n$ can be found such that the equation takes the form

$$\lambda_1 y_1^2 + \cdots + \lambda_r y_r^2 + k_1 y_1 + \cdots + k_n y_n = c$$

b. Put $x_1^2 + 3x_2^2 + 3x_3^2 + 4x_1x_2 - 4x_1x_3 + 5x_1 - 6x_3 = 7$ in this form and find variables y_1, y_2, y_3 as in (a).

Exercise 8.9.8 Given a symmetric matrix A, define $q_A(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$. Show that $B \overset{c}{\sim} A$ if and only if B is symmetric and there is an invertible matrix U such that $q_B(\mathbf{x}) = q_A(U\mathbf{x})$ for all $\mathbf{x}$. [*Hint*: Theorem 8.9.3.]

Exercise 8.9.9 Let $q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ be a quadratic form where $A = A^T$.

a. Show that $q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$, if and only if A is positive definite (all eigenvalues are positive). In this case, q is called **positive definite**.

b. Show that new variables $\mathbf{y}$ can be found such that $q = \|\mathbf{y}\|^2$ and $\mathbf{y} = U\mathbf{x}$ where U is upper triangular with positive diagonal entries. [*Hint*: Theorem 8.3.3.]

Exercise 8.9.10 A **bilinear form** β on $\mathbb{R}^n$ is a function that assigns to every pair $\mathbf{x}$, $\mathbf{y}$ of columns in $\mathbb{R}^n$ a number $\beta(\mathbf{x}, \mathbf{y})$ in such a way that

$$\beta(r\mathbf{x} + s\mathbf{y}, \mathbf{z}) = r\beta(\mathbf{x}, \mathbf{z}) + s\beta(\mathbf{y}, \mathbf{z})$$
$$\beta(\mathbf{x}, r\mathbf{y} + s\mathbf{z}) = r\beta(\mathbf{x}, \mathbf{z}) + s\beta(\mathbf{x}, \mathbf{z})$$

for all $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ in $\mathbb{R}^n$ and r, s in $\mathbb{R}$. If $\beta(\mathbf{x}, \mathbf{y}) = \beta(\mathbf{y}, \mathbf{x})$ for all $\mathbf{x}$, $\mathbf{y}$, β is called **symmetric**.

a. If β is a bilinear form, show that an $n \times n$ matrix A exists such that $\beta(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T A\mathbf{y}$ for all $\mathbf{x}$, $\mathbf{y}$.

b. Show that A is uniquely determined by β.

c. Show that β is symmetric if and only if $A = A^T$.

8.10 An Application to Constrained Optimization

It is a frequent occurrence in applications that a function $q = q(x_1, x_2, \dots, x_n)$ of n variables, called an **objective function**, is to be made as large or as small as possible among all vectors $\mathbf{x} = (x_1, x_2, \dots, x_n)$ lying in a certain region of $\mathbb{R}^n$ called the **feasible region**. A wide variety of objective functions q arise in practice; our primary concern here is to examine one important situation where q is a quadratic form. The next example gives some indication of how such problems arise.