Theorem 3.7.14 (Spectral theorem). *Let A be a symmetric $n \times n$ matrix with real entries. Then there exists an orthonormal basis $\vec{\mathbf{v}}_1, \dots, \vec{\mathbf{v}}_n$ of $\mathbb{R}^n$ and numbers $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ such that*

$$A\vec{\mathbf{v}}_i = \lambda_i \vec{\mathbf{v}}_i. \qquad 3.7.57$$

We use λ to denote both eigenvalues and the Lagrange multipliers of theorem 3.7.5. When working on a problem involving constrained critical points, Lagrange used λ to denote what we now call Lagrange multipliers. At that time, linear algebra did not exist, but later (after Hilbert proved the spectral theorem in the harder infinite-dimensional setting!) people realized that Lagrange had proved the finite-dimensional version, theorem 3.7.14.

Generalizing the spectral theorem to infinitely many dimensions is one of the central topics of functional analysis.

Recall that an $n \times n$ matrix whose columns form an orthonormal basis of $\mathbb{R}^n$ is called an orthogonal matrix. Thus another way of stating theorem 3.7.14 is that if A is a symmetric real matrix, then there exists an orthogonal matrix Q such that $Q^{-1}AQ$ is diagonal; see proposition 2.7.3.

Exercise 3.2.11 explores other results concerning orthogonal and symmetric matrices.

Remember that the maximum of Q_A is an element of its domain; the function achieves a maximal value at a maximum.

By proposition 2.7.3, theorem 3.7.14 is equivalent to the statement that $B^{-1}AB = B^\top AB$ is diagonal, where B is the orthogonal matrix whose columns are $\vec{\mathbf{v}}_1, \dots, \vec{\mathbf{v}}_n$. (Recall that a matrix B is orthogonal if and only if $B^\top B = I$; see exercise 2.4.7.)

Remark. Recall from definition 2.7.2 that a vector $\vec{\mathbf{v}}_i$ satisfying equation 3.7.57 is an eigenvector, with eigenvalue λ_i. So the orthonormal basis of theorem 3.7.14 is an eigenbasis. In section 2.7 we showed that every square matrix has at least one eigenvector, but since our procedure for finding it used the fundamental theorem of algebra, it had to allow for the possibility that the corresponding eigenvalue might be complex. Square real matrices exist that have no real eigenvectors or eigenvalues: for instance, the matrix $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ rotates vectors clockwise by $\pi/2$, so clearly there is no $\lambda \in \mathbb{R}$ and no vector $\vec{\mathbf{v}} \in \mathbb{R}^2$ such that $A\vec{\mathbf{v}} = \lambda\vec{\mathbf{v}}$; its eigenvalues are $\pm i$. Symmetric real matrices are a very special class of square real matrices, whose eigenvectors are guaranteed not only to exist but also to form an orthonormal basis. $\triangle$

Proof. Our strategy will be to consider the function $Q_A : \mathbb{R}^n \to \mathbb{R}$ given by $Q_A(\vec{\mathbf{x}}) = \vec{\mathbf{x}} \cdot A\vec{\mathbf{x}}$, subject to various constraints. The first constraint ensures that the first basis vector we find has length 1. Adding a second constraint ensures that the second basis vector has length 1 and is orthogonal to the first. Adding a third constraint ensures that the third basis vector has length 1 and is orthogonal to the first two, and so on.

Throughout we will use theorem 1.6.9, which guarantees that a continuous function on a compact set has a maximum. That theorem is nonconstructive: it proves existence of a maximum without showing how to find it. Thus theorem 3.7.14 is also nonconstructive.

First we restrict Q_A to the $(n-1)$-dimensional sphere $S \subset \mathbb{R}^n$ of equation $F_1(\vec{\mathbf{x}}) = |\vec{\mathbf{x}}|^2 = 1$; this is the first constraint. Since S is a compact subset of $\mathbb{R}^n$, Q_A restricted to S has a maximum, and the constraint $F_1(\vec{\mathbf{x}}) = |\vec{\mathbf{x}}|^2 = 1$ ensures that this maximum has length 1.

Exercise 3.7.9 asks you to justify that the derivative of Q_A is

$$[\mathbf{D}Q_A(\vec{\mathbf{a}})]\vec{\mathbf{h}} = \vec{\mathbf{a}} \cdot (A\vec{\mathbf{h}}) + \vec{\mathbf{h}} \cdot (A\vec{\mathbf{a}}) = \vec{\mathbf{a}}^\top A\vec{\mathbf{h}} + \vec{\mathbf{h}}^\top A\vec{\mathbf{a}} = 2\vec{\mathbf{a}}^\top A\vec{\mathbf{h}}. \qquad 3.7.58$$

The derivative of the constraint function is

$$[\mathbf{D}F_1(\vec{\mathbf{a}})]\vec{\mathbf{h}} = 2\vec{\mathbf{a}}^\top \vec{\mathbf{h}}. \qquad 3.7.59$$

Theorem 3.7.5 tells us that if $\mathbf{v}_1$ is a maximum of Q_A restricted to S, then there exists λ_1 such that

In equation 3.7.60, remember that

$$(AB)^\top = B^\top A^\top.$$

$$2\vec{\mathbf{v}}_1^\top A = \lambda_1 2\vec{\mathbf{v}}_1^\top, \text{ so } (\vec{\mathbf{v}}_1^\top A)^\top = (\lambda_1 \vec{\mathbf{v}}_1^\top)^\top, \text{ i.e., } A^\top \vec{\mathbf{v}}_1 = \lambda_1 \vec{\mathbf{v}}_1. \qquad 3.7.60$$

Since A is symmetric, we can rewrite equation 3.7.60 as

$$A\vec{\mathbf{v}}_1 = \lambda_1 \vec{\mathbf{v}}_1. \qquad 3.7.61$$

Thus $\vec{\mathbf{v}}_1$, the maximum of Q_A restricted to S, satisfies equation 3.7.57 and has length 1; it is the first vector in our orthonormal basis of eigenvectors.

Remark. We can also prove the existence of the first eigenvector by a more geometric argument. By theorem 3.7.1, at a critical point $\vec{\mathbf{v}}_1$, the derivative $[\mathbf{D}Q_A(\vec{\mathbf{v}}_1)]$ vanishes on $T_{\vec{\mathbf{v}}_1}S$: $[\mathbf{D}Q_A(\vec{\mathbf{v}}_1)]\vec{\mathbf{w}} = 0$ for all $\vec{\mathbf{w}} \in T_{\vec{\mathbf{v}}_1}S$. Since $T_{\vec{\mathbf{v}}_1}S$ consists of vectors tangent to the sphere at $\vec{\mathbf{v}}_1$, hence orthogonal to $\vec{\mathbf{v}}_1$, this means that for all $\vec{\mathbf{w}}$ such that $\vec{\mathbf{w}} \cdot \vec{\mathbf{v}}_1 = 0$,

$$[\mathbf{D}Q_A(\vec{\mathbf{v}}_1)]\vec{\mathbf{w}} = 2(\vec{\mathbf{v}}_1^\top A)\vec{\mathbf{w}} = 0. \qquad 3.7.62$$

Taking transposes of both sides gives $\vec{\mathbf{w}}^\top A\vec{\mathbf{v}}_1 = \vec{\mathbf{w}} \cdot A\vec{\mathbf{v}}_1 = 0$. Since $\vec{\mathbf{w}} \perp \vec{\mathbf{v}}_1$, this says that $A\vec{\mathbf{v}}_1$ is orthogonal to any vector orthogonal to $\vec{\mathbf{v}}_1$. Thus it points in the same direction as $\vec{\mathbf{v}}_1$: $A\vec{\mathbf{v}}_1 = \lambda_1 \vec{\mathbf{v}}_1$ for some number λ_1. △

The space $(\vec{\mathbf{v}}_1)^\perp$ is called the *orthogonal complement* to the subspace spanned by $\vec{\mathbf{v}}_1$. More generally, the set of vectors $\vec{\mathbf{x}}$ satisfying

$$\vec{\mathbf{x}} \cdot \vec{\mathbf{v}}_1 = 0, \quad \vec{\mathbf{x}} \cdot \vec{\mathbf{v}}_2 = 0$$

(see equation 3.7.69) is called the *orthogonal complement* of the subspace spanned by $\vec{\mathbf{v}}_1$ and $\vec{\mathbf{v}}_2$.

For the second eigenvalue, denote by $\vec{\mathbf{v}}_2$ the maximum of Q_A when Q_A is subject to the two constraints

$$F_1(\vec{\mathbf{x}}) = |\vec{\mathbf{x}}|^2 = 1 \quad \text{and} \quad F_2(\vec{\mathbf{x}}) = \vec{\mathbf{x}} \cdot \vec{\mathbf{v}}_1 = 0. \qquad 3.7.63$$

In other words, we are considering the maximum of Q_A restricted to the space $S \cap (\vec{\mathbf{v}}_1)^\perp$, where $(\vec{\mathbf{v}}_1)^\perp$ is the space of vectors orthogonal to $\vec{\mathbf{v}}_1$.

Since $[\mathbf{D}F_2(\vec{\mathbf{v}}_2)] = \vec{\mathbf{v}}_1^\top$, equations 3.7.58 and 3.7.59 and theorem 3.7.5 tell us that there exist numbers λ_2 and $\mu_{2,1}$ such that

Equation 3.7.64: The subscript 2, 1 for μ indicates that we are finding our second basis vector; the 1 indicates that we are choosing it orthogonal to the first basis vector.

$$\underbrace{2\vec{\mathbf{v}}_2^\top A}_{[\mathbf{D}Q_A(\vec{\mathbf{v}}_2)]} = \underbrace{\lambda_2 2\vec{\mathbf{v}}_2^\top}_{\lambda_1[\mathbf{D}F_1(\vec{\mathbf{v}}_2)]} + \underbrace{\mu_{2,1}\vec{\mathbf{v}}_1^\top}_{\lambda_2[\mathbf{D}F_2(\vec{\mathbf{v}}_2)]}. \qquad 3.7.64$$

(Note that here λ_2 corresponds to λ_1 in equation 3.7.13, and $\mu_{2,1}$ corresponds to λ_2 in that equation.) Take transposes of both sides (remembering that $A = A^\top$) to get

$$A\vec{\mathbf{v}}_2 = \lambda_2 \vec{\mathbf{v}}_2 + \frac{\mu_{2,1}}{2}\vec{\mathbf{v}}_1. \qquad 3.7.65$$

The first equality of equation 3.7.67 uses the symmetry of A: if A is symmetric,

$$\begin{aligned} \vec{\mathbf{v}} \cdot (A\vec{\mathbf{w}}) &= \vec{\mathbf{v}}^\top (A\vec{\mathbf{w}}) = (\vec{\mathbf{v}}^\top A)\vec{\mathbf{w}} \\ &= (\vec{\mathbf{v}}^\top A^\top)\vec{\mathbf{w}} = (A\vec{\mathbf{v}})^\top \vec{\mathbf{w}} \\ &= (A\vec{\mathbf{v}}) \cdot \vec{\mathbf{w}}. \end{aligned}$$

The second uses equation 3.7.61, and the third the fact that

$$\vec{\mathbf{v}}_2 \in S \cap (\vec{\mathbf{v}}_1)^\perp.$$

Now take the dot product of each side with $\vec{\mathbf{v}}_1$, to find

$$(A\vec{\mathbf{v}}_2) \cdot \vec{\mathbf{v}}_1 = \lambda_2 \underbrace{\vec{\mathbf{v}}_2 \cdot \vec{\mathbf{v}}_1}_{0} + \frac{\mu_{2,1}}{2} \underbrace{\vec{\mathbf{v}}_1 \cdot \vec{\mathbf{v}}_1}_{1} = \frac{\mu_{2,1}}{2}. \qquad 3.7.66$$

(We have $\vec{\mathbf{v}}_1 \cdot \vec{\mathbf{v}}_1 = 1$ because $\vec{\mathbf{v}}_1$ has length 1; we have $\vec{\mathbf{v}}_2 \cdot \vec{\mathbf{v}}_1 = 0$ because the two vectors are orthogonal to each other.) Using

$$(A\vec{\mathbf{v}}_2) \cdot \vec{\mathbf{v}}_1 = \vec{\mathbf{v}}_2 \cdot (A\vec{\mathbf{v}}_1) \underbrace{=}_{\text{eq. 3.7.61}} \vec{\mathbf{v}}_2 \cdot (\lambda_1 \vec{\mathbf{v}}_1) = 0, \qquad 3.7.67$$

equation 3.7.66 becomes $0 = \mu_{2,1}$, so equation 3.7.65 becomes

$$A\vec{\mathbf{v}}_2 = \lambda_2 \vec{\mathbf{v}}_2. \qquad 3.7.68$$

Thus we have found a second eigenvector: it is the maximum of Q_A constrained to the sets given by $F_1 = 0$ and $F_2 = 0$.

It should be clear how to continue, but let us take one further step. Suppose $\vec{\mathbf{v}}_3$ is a maximum of Q_A restricted to $S \cap \vec{\mathbf{v}}_1^\perp \cap \vec{\mathbf{v}}_2^\perp$, that is, maximize Q_A subject to the three constraints

$$F_1(\vec{\mathbf{x}}) = 1, \qquad F_2(\vec{\mathbf{x}}) = \vec{\mathbf{x}} \cdot \vec{\mathbf{v}}_1 = 0, \qquad F_3(\vec{\mathbf{x}}) = \vec{\mathbf{x}} \cdot \vec{\mathbf{v}}_2 = 0. \qquad 3.7.69$$

The argument above says that there exist numbers λ_3, $\mu_{3,1}$, $\mu_{3,2}$ such that

$$A\vec{\mathbf{v}}_3 = \mu_{3,1}\vec{\mathbf{v}}_1 + \mu_{3,2}\vec{\mathbf{v}}_2 + \lambda_3\vec{\mathbf{v}}_3. \qquad 3.7.70$$

Dot this entire equation with $\vec{\mathbf{v}}_1$ (respectively, with $\vec{\mathbf{v}}_2$); you will find $\mu_{3,1} = \mu_{3,2} = 0$. Thus $A\vec{\mathbf{v}}_3 = \lambda_3\vec{\mathbf{v}}_3$. □

You should be be impressed by how easily the existence of eigenvectors dropped out of Lagrange multipliers. Of course, we could not have done this without the nonconstructive theorem 1.6.9 guaranteeing that the function Q_A has a maximum and minimum. In addition, we've only proved existence: there is no obvious way to find these constrained maxima of Q_A.

The spectral theorem gives another approach to quadratic forms, geometrically more appealing than completing squares.

Theorem 3.7.15. *Let A be a symmetric matrix. A quadratic form Q_A has signature (k, l) if and only if A has k linearly independent eigenvectors with positive eigenvalues and l linearly independent eigenvectors with negative eigenvalues.*

Proof. By equation 3.5.34 and proposition 3.5.17, $Q_A(\vec{\mathbf{x}}) = \vec{\mathbf{x}}^\top A\vec{\mathbf{x}}$, where A is a symmetric matrix. The spectral theorem says that there exists an orthonormal basis $\vec{\mathbf{v}}_1, \dots, \vec{\mathbf{v}}_n$ such that $A\vec{\mathbf{v}}_i = \lambda_i\vec{\mathbf{v}}_i$. Any $\vec{\mathbf{x}}$ can be written

$$\vec{\mathbf{x}} = \alpha_1(\vec{\mathbf{x}})\vec{\mathbf{v}}_1 + \dots + \alpha_n(\vec{\mathbf{x}})\vec{\mathbf{v}}_n, \qquad 3.7.71$$

where $\alpha_i(\vec{\mathbf{x}}) = \vec{\mathbf{v}}_i^\top\vec{\mathbf{x}}$; i.e., α_i is the line matrix $\vec{\mathbf{v}}_i^\top$. (These α_i are, of course, linearly independent functions, since they are the transposes of the basis vectors.) Then

$$\begin{aligned} Q_A(\vec{\mathbf{x}}) = \vec{\mathbf{x}}^\top A\vec{\mathbf{x}} &= \Big(\alpha_1(\vec{\mathbf{x}})\vec{\mathbf{v}}_1 + \dots + \alpha_n(\vec{\mathbf{x}})\vec{\mathbf{v}}_n\Big)^\top A\Big(\alpha_1(\vec{\mathbf{x}})\vec{\mathbf{v}}_1 + \dots + \alpha_n(\vec{\mathbf{x}})\vec{\mathbf{v}}_n\Big) \\ &= \Big(\alpha_1(\vec{\mathbf{x}})\vec{\mathbf{v}}_1 + \dots + \alpha_n(\vec{\mathbf{x}})\vec{\mathbf{v}}_n\Big)^\top \Big(\alpha_1(\vec{\mathbf{x}})\lambda_1\vec{\mathbf{v}}_1 + \dots + \alpha_n(\vec{\mathbf{x}})\lambda_n\vec{\mathbf{v}}_n\Big) \\ &= \lambda_1\big(\alpha_1(\vec{\mathbf{x}})\big)^2 + \dots + \lambda_n\big(\alpha_n(\vec{\mathbf{x}})\big)^2. \end{aligned} \qquad 3.7.72$$

Equation 3.7.72: To go from line 2 to line 3 we use the fact the $\vec{\mathbf{v}}_1, \dots, \vec{\mathbf{v}}_n$ form an orthonormal basis, so $\vec{\mathbf{v}}_i \cdot \vec{\mathbf{v}}_j$ is 0 if $i \neq j$ and 1 if $i = j$.

If $\lambda_1 > 0, \dots, \lambda_k > 0$ and $\lambda_{k+1} < 0, \dots, \lambda_{k+l} < 0$, with all other $\lambda_j = 0$, then Q_A has signature (k, l). Conversely, since the signature is well defined, if Q_A has signature (k, l), then k of the λ_i are > 0 and l are < 0, and the remainder are 0. □

Exercises for Section 3.7

3.7.1 What is the maximum volume of a box of surface area 10, for which one side is exactly twice as long as another?