Chapter 4 Symmetric matrices and the second derivative test

In this chapter we are going to finish our description of the nature of nondegenerate critical points. But first we need to discuss some fascinating and important features of square matrices.

A. **Eigenvalues and eigenvectors**

Suppose that $A = (a_{ij})$ is a fixed $n \times n$ matrix. We are going to discuss linear equations of the form

$$\boxed{Ax \;=\; \lambda x,}$$

where $x \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$. (We sometimes will allow $x \in \mathbb{C}^n$ and $\lambda \in \mathbb{C}$.) Of course, $x = 0$ is always a solution of this equation, but not an interesting one. We say x is a *nontrivial* solution if it satisfies the equation and $x \neq 0$.

DEFINITION. If $Ax = \lambda x$ and $x \neq 0$, we say that λ is an *eigenvalue* of A and that the vector x is an *eigenvector* of A corresponding to λ.

EXAMPLE. Let $A = \begin{pmatrix} 0 & 3 \\ 1 & 2 \end{pmatrix}$. Then we notice that

$$A\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix} = 3\begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

so $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is an eigenvector corresponding to the eigenvalue 3. Also,

$$A\begin{pmatrix} 3 \\ -1 \end{pmatrix} = \begin{pmatrix} -3 \\ 1 \end{pmatrix} = -\begin{pmatrix} 3 \\ -1 \end{pmatrix},$$

so $\begin{pmatrix} 3 \\ -1 \end{pmatrix}$ is an eigenvector corresponding to the eigenvalue -1.

EXAMPLE. Let $A = \begin{pmatrix} 2 & 1 \\ 0 & 0 \end{pmatrix}$. Then $A\begin{pmatrix} 1 \\ 0 \end{pmatrix} = 2\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, so 2 is an eigenvalue, and $A\begin{pmatrix} 1 \\ -2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, so 0 is also an eigenvalue.

REMARK. The German word for eigenvalue is *eigenwert*. A literal translation into English would be "characteristic value," and this phrase appears in a few texts. The English word "eigenvalue" is clearly a sort of half translation, half transliteration, but this hybrid has stuck.

PROBLEM 4–1. Show that A is invertible $\iff$ 0 is not an eigenvalue of A.

The equation $Ax = \lambda x$ can be rewritten as $Ax = \lambda I x$, and then as $(A - \lambda I)x = 0$. In order that this equation have a *nonzero* x as a solution, Problem 3–52 shows that it is necessary and sufficient that

$$\det(A - \lambda I) = 0.$$

(Otherwise Cramer's rule yields $x = 0$.) This equation is quite interesting. The quantity

$$\det\begin{pmatrix} a_{11}-\lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22}-\lambda & \dots & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & \dots & a_{nn}-\lambda \end{pmatrix}$$

can in principle be written out in detail, and it is then seen that it is a polynomial in λ of degree n. This polynomial is called the *characteristic polynomial* of A; perhaps it would be more consistent to call it the *eigen*polynomial, but no one seems to do this.

The only term in the expansion of the determinant which contains n factors involving λ is the product

$$(a_{11} - \lambda)(a_{22} - \lambda) \dots (a_{nn} - \lambda).$$

Thus the coefficient of λ^n in the characteristic polynomial is $(-1)^n$. In fact, that product is also the only term which contains as many as $n-1$ factors involving λ, so the coefficient of λ^{n-1} is $(-1)^{n-1}\ (a_{11} + a_{22} + \cdots + a_{nn})$. This introduces us to an important number associated with the matrix A, called the *trace* of A:

$$\text{trace} A \ = \ a_{11} + a_{22} + \cdots + a_{nn}.$$

Notice also that the polynomial $\det(A - \lambda I)$ evaluated at $\lambda = 0$ is just $\det A$, so this is the constant term of the characteristic polynomial. In summary,

$$\det(A - \lambda I) = (-1)^n \lambda^n + (-1)^{n-1}(\text{trace} A)\lambda^{n-1} + \cdots + \det A.$$

PROBLEM 4–2. Prove that

$$\text{trace} AB \ = \ \text{trace} BA.$$

EXAMPLE. All of the above virtually provides an algorithm for finding eigenvalues and eigenvectors. For example, suppose

$$A = \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix}.$$

We first calculate the characteristic polynomial,

$$\begin{aligned} \det(A - \lambda I) &= \det \begin{pmatrix} 1-\lambda & 2 \\ 1 & 3-\lambda \end{pmatrix} \\ &= (1-\lambda)(3-\lambda) - 2 \\ &= \lambda^2 - 4\lambda + 1. \end{aligned}$$

Now we use the quadratic formula to find the zeros of this polynomial, and obtain $\lambda = 2 \pm \sqrt{3}$. These two numbers are the eigenvalues of A. We find corresponding eigenvectors x by considering $(A - \lambda I)x = 0$:

$$\begin{pmatrix} -1 \mp \sqrt{3} & 2 \\ 1 & 1 \mp \sqrt{3} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

We can for instance simply choose a solution of the lower equation, say $x_1 = 1 \mp \sqrt{3}$, $x_2 = -1$. The upper equation requires no verification, as it must be automatically satisfied! (Nevertheless, we calculate: $(-1 \mp \sqrt{3})(1 \mp \sqrt{3}) + 2(-1) = 2 - 2 = 0$.) Thus we have eigenvectors as follows:

$$\begin{aligned} A \begin{pmatrix} 1-\sqrt{3} \\ -1 \end{pmatrix} &= (2+\sqrt{3}) \begin{pmatrix} 1-\sqrt{3} \\ -1 \end{pmatrix}, \\ A \begin{pmatrix} 1+\sqrt{3} \\ -1 \end{pmatrix} &= (2-\sqrt{3}) \begin{pmatrix} 1+\sqrt{3} \\ -1 \end{pmatrix}. \end{aligned}$$

EXAMPLE. Let

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

The characteristic polynomial is $\lambda^2 + 1$, so the eigenvalues are *not* real: they are $\pm i$, where $i = \sqrt{-1}$. The eigenvectors also are not real:

$$\begin{aligned} A \begin{pmatrix} 1 \\ i \end{pmatrix} &= \begin{pmatrix} i \\ -1 \end{pmatrix} = i \begin{pmatrix} 1 \\ i \end{pmatrix}, \\ A \begin{pmatrix} 1 \\ -i \end{pmatrix} &= \begin{pmatrix} -i \\ -1 \end{pmatrix} = -i \begin{pmatrix} 1 \\ -i \end{pmatrix}. \end{aligned}$$

Of course, the moral of this example is that real matrices may have only nonreal eigenvalues and eigenvectors. (Notice that this matrix is not symmetric.)

EXAMPLE. Let

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

The characteristic polynomial is clearly $(2-\lambda)^3$, so $\lambda = 2$ is the only eigenvalue. To find an eigenvector, we need to solve $(A-2I)x = 0$. That is,

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

or equivalently,

$$\begin{cases} x_2 + x_3 & = 0, \\ x_3 & = 0. \end{cases}$$

Thus the only choice for x is $x = \begin{pmatrix} c \\ 0 \\ 0 \end{pmatrix}$. Thus there is only *one* linearly independent eigenvector.

PROBLEM 4–3. Modify the above example to produce a 3×3 real matrix B whose characteristic polynomial is also $(2-\lambda)^3$, but for which there are *two* linearly independent eigenvectors, but not three.

Moral: when λ is an eigenvalue which is *repeated*, in the sense that it is a multiple zero of the characteristic polynomial, there might not be as many linearly independent eigenvectors as the multiplicity of the zero.

PROBLEM 4–4. Let λ_0 be a fixed scalar and define the matrix B to be $B = A - \lambda_0 I$. Prove that λ is an eigenvalue of $A \iff \lambda - \lambda_0$ is an eigenvalue of B. What is the relation between the characteristic polynomials of A and B?

PROBLEM 4–5. If A is an $n \times n$ matrix whose characteristic polynomial is λ^n and for which there are n linearly independent eigenvectors, show that $A = 0$.

EXAMPLE. From Problem 3–29, take

$$A = \begin{pmatrix} 1 & -1 & 1 \\ -1 & 3 & 0 \\ 1 & 0 & 2 \end{pmatrix}.$$

The characteristic polynomial of A is

$$\begin{aligned} \det(A - \lambda I) &= \det \begin{pmatrix} 1-\lambda & -1 & 1 \\ -1 & 3-\lambda & 0 \\ 1 & 0 & 2-\lambda \end{pmatrix} \\ &= (1-\lambda)(3-\lambda)(2-\lambda) - (3-\lambda) - (2-\lambda) \\ &= -\lambda^3 + 6\lambda^2 - 9\lambda + 1. \end{aligned}$$

The eigenvalue equation is

$$\lambda^3 - 6\lambda^2 + 9\lambda - 1 = 0;$$

this cubic equation has three real roots, none of them easy to calculate. The moral here is that when $n > 2$, the eigenvalues of A may be difficult or impossible to calculate explicitly.

Given any $n \times n$ matrix A with entries a_{ij} which are real numbers, or even *complex* numbers, the characteristic polynomial has at least one complex zero λ. This is an immediate consequence of the so-called "fundamental theorem of algebra." (This is proved in basic courses in *complex* analysis!) Thus A has at least one complex eigenvalue, and a corresponding eigenvector.

PROBLEM 4–6. Calculate the eigenvalues and eigenvectors of the matrix

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

PROBLEM 4–7. Learn how to use Matlab or Mathematica or some such program to find eigenvalues and eigenvectors of numerical matrices.

Now reconsider the characteristic polynomial of A. It is a polynomial $(-1)^n\lambda^n + \dots$ of degree n. The fundamental theorem of algebra guarantees this polynomial has a zero — let us call it λ_1. The polynomial is thus divisible by the first order polynomial $\lambda - \lambda_1$, the quotient being a polynomial of degree $n-1$. By induction we quickly conclude that the characteristic polynomial can be completely factored:

$$\det(A-\lambda I) = (-1)^n(\lambda-\lambda_1)\dots(\lambda-\lambda_n).$$

We think of $\lambda_1, \dots, \lambda_n$ as the eigenvalues of A, *though some may be repeated.* We can now read off two very interesting things. First, the constant term in the two sides of the above equation (which may be obtained by setting $\lambda = 0$) yields the marvelous fact that

$$\boxed{\det A = \lambda_1\ \lambda_2\ \dots\ \lambda_n.}$$

Second, look at the coefficient of λ^{n-1} in the two sides (see p. 4–2) to obtain

$$\boxed{\operatorname{trace} A = \lambda_1 + \lambda_2 + \dots + \lambda_n.}$$

These two wonderful equations reveal rather profound qualities of $\det A$ and $\operatorname{trace} A$. Although those numbers are explicitly computable in terms of algebraic operations on the entries of A, they are also intimately related to the more geometric ideas of eigenvalues and eigenvectors.

B. **Eigenvalues of symmetric matrices**

Now we come to the item we are most interested in. Remember, we are trying to understand Hessian matrices, and these are real symmetric matrices. For the record,

DEFINITION. An $n \times n$ matrix $A = (a_{ij})$ is *symmetric* if $a_{ij} = a_{ji}$ for all i, j. In other words, if $A^t = A$.

We have of course encountered these in the $n = 2$ case. The solution of Problem 3–18 shows that the eigenvalues of the 2×2 matrix

$$\begin{pmatrix} A & B \\ B & C \end{pmatrix}$$

are

$$\lambda = \frac{A + C \pm \sqrt{(A-C)^2 + 4B^2}}{2},$$

and these are both *real*. This latter fact is what we now generalize.

If A is an $n \times n$ matrix which is real and symmetric, then Problem 2–83 gives us

$$Ax \bullet y = x \bullet Ay \qquad \text{for all } \ x, y \in \mathbb{R}^n.$$

PROBLEM 4–8. Prove conversely that if $Ax \bullet y = x \bullet Ay$ for all x, $y \in \mathbb{R}^n$, then A is symmetric.

THEOREM. *If A is a real symmetric matrix, then its eigenvalues are all real.*

PROOF. Suppose λ is a possibly complex eigenvalue of A, with corresponding eigenvector $z \in \mathbb{C}^n$. Write λ and z in terms of their real and imaginary parts:

$$\begin{aligned} &\lambda = \alpha + i\beta, \quad \text{where} \quad \alpha, \beta \in \mathbb{R}, \\ &z = x + iy, \quad \text{where} \quad x, y \in \mathbb{R}^n \text{ and are not both } 0. \end{aligned}$$

Then the eigenvalue equation $Az = \lambda z$ becomes

$$A(x + iy) = (\alpha + i\beta)(x + iy).$$

That is,

$$Ax + iAy = \alpha x - \beta y + i(\alpha y + \beta x).$$

This complex equation is equivalent to the two real equations

$$\begin{cases} Ax &= \alpha x - \beta y, \\ Ay &= \alpha y + \beta x. \end{cases}$$

We now compute

$$\begin{cases} Ax \bullet y &= \alpha x \bullet y - \beta \|y\|^2, \\ Ay \bullet x &= \alpha x \bullet y + \beta \|x\|^2. \end{cases}$$

Since A is symmetric, the two left sides are equal. Therefore,

$$\alpha x \bullet y - \beta \|y\|^2 = \alpha x \bullet y + \beta \|x\|^2.$$

That is,

$$\beta(\|x\|^2 + \|y\|^2) = 0.$$

Since $\|x\|^2 + \|y\|^2 > 0$, we conclude $\beta = 0$. Thus $\lambda = \alpha$ is real.

QED

We conclude that a real symmetric matrix has at least one eigenvalue, and this eigenvalue is real. This result is a combination of the profound fundamental theorem of algebra and the above calculation we have just given in the proof of the theorem. It would seem strange to call upon complex analysis (the fund. thm. of alg.) to be guaranteed that a complex root exists, and then prove it must be real after all. That is indeed strange, so we now present an independent proof of the existence of an eigenvalue of a real symmetric matrix; this proof will not rely on complex analysis at all. This proof depends on rather elementary calculus.

Even so, it may seem strange to rely on calculus at all, since we are trying to prove a theorem about algebra — roots of polynomial equations. However, simple reasoning shows that something must be used beyond just algebra. For we are using the real numbers, a complete ordered field. The *completeness* is essential, as for example the polynomial $\lambda^2 - 2$ illustrates. Or even more challenging, imagine an equation such as $\lambda^{113} - \lambda + 5 = 0$; it definitely has a real solution. These two examples have only irrational solutions.

Let A be the $n \times n$ real symmetric matrix, and consider the quotient function $\mathbb{R}^n \xrightarrow{Q} \mathbb{R}$,

$$Q(x) = \frac{Ax \bullet x}{\|x\|^2} = \frac{Ax \bullet x}{x \bullet x}.$$

This is a rather natural function to consider. In a sense it measures something like the relative distortion of angles caused by A. "Relative," because the denominator $\|x\|^2$ is just right for $Q(x)$ to be scale invariant. Notice how *geometry* will be used in what follows to give our result in *algebra* — the existence of an eigenvalue. This function is known as the *Rayleigh quotient*.

This function is defined and of class C^∞ on $\mathbb{R}^n - \{0\}$, and we can compute its gradient quite easily. First, we have from Problem 2–84 a formula for the gradients of the numerator and the denominator:

$$\begin{aligned} \nabla Ax \bullet x &= 2Ax, \\ \nabla \|x\|^2 &= 2x. \end{aligned}$$

Thus the quotient rule yields

$$\begin{aligned} \nabla Q(x) &= \frac{\|x\|^2 2Ax - (Ax \bullet x)2x}{\|x\|^4} \\ &= \frac{2Ax}{\|x\|^2} - \frac{2Ax \bullet x}{\|x\|^4} x. \end{aligned}$$

The function Q is continuous on the unit sphere $\|x\| = 1$. Since this sphere $S(0,1)$ is closed and bounded, Q restricted to $S(0,1)$ attains its maximum value. Say at a point x_0, so that $\|x_0\| = 1$ and $Q(x) \le Q(x_0)$ for all $\|x\| = 1$. But the *homogeneity* of Q shows that $Q(x_0)$ is also the *global* maximum value of Q on $\mathbb{R}^n - \{0\}$. (This argument probably reminds you of Problem 3–18.) The details: if $x \ne 0$, then $x/\|x\|$ is on $S(0,1)$, so that

$$Q(x) = Q\left(\frac{x}{\|x\|}\right) \le Q(x_0).$$

Thus x_0 is a *critical point* of Q (p. 2–36). That is, $\nabla Q(x_0) = 0$. Let $\lambda = Ax_0 \bullet x_0$. Then the above expression for ∇Q gives

$$\begin{aligned} 0 = \nabla Q(x_0) &= \frac{2Ax_0}{\|x_0\|^2} - \frac{2Ax_0 \bullet x_0}{\|x_0\|^4} x_0 \\ &= 2Ax_0 - 2\lambda x_0. \end{aligned}$$

Therefore

$$Ax_0 = \lambda x_0, \quad \|x_0\| = 1.$$

We conclude that λ is an *eigenvalue* of A, and x_0 is a corresponding eigenvector! Moreover, this particular eigenvalue is given by

$$\lambda = \max\{Ax \bullet x \mid \|x\| = 1\},$$

and x_0 is a point where this maximum value is attained!

PROBLEM 4–9. Calculate the Hessian matrix of Q at a critical point x_0 with $\|x_0\| = 1$. Show that it is

$$H = 2A - 2\lambda I \qquad (\lambda = Q(x_0)).$$

The analysis we are going to do next will continue to use the quotient function and the formula we have obtained for its gradient, so we record here for later reference

$$\boxed{\nabla Q(x) = 2\,(Ax - Q(x)x) \quad \text{for} \quad \|x\| = 1.}$$

We are now doubly certain as to the existence of a real eigenvalue of the real symmetric matrix A. We proceed to a further examination of the eigenvector structure of A. First here is an incredibly important property with a ridiculously easy proof:

THEOREM. *Let x and y be eigenvectors of a real symmetric matrix, corresponding to different eigenvalues. Then x and y are orthogonal.*

PROOF. We know that $Ax = \lambda_1 x$ and $Ay = \lambda_2 y$ and $\lambda_1 \neq \lambda_2$. Therefore

$$\begin{aligned} \lambda_1(x \bullet y) &= (\lambda_1 x) \bullet y \\ &= Ax \bullet y \\ &= x \bullet Ay \qquad \text{(because } A \text{ is symmetric!)} \\ &= x \bullet (\lambda_2 y) \\ &= \lambda_2(x \bullet y). \end{aligned}$$

Subtract:

$$(\lambda_1 - \lambda_2)x \bullet y \;=\; 0.$$

Since $\lambda_1 - \lambda_2 \neq 0$, $x \bullet y = 0$.

QED

Next we give a very similar fact, based on the identical reasoning.

THEOREM. *Assume A is an $n \times n$ real symmetric matrix. Assume x is an eigenvector of A, and let M be the ($(n-1)$-dimensional) subspace of $\mathbb{R}^n$ consisting of all points orthogonal to x:*

$$M \;=\; \{y \in \mathbb{R}^n \mid x \bullet y \;=\; 0\}.$$

*Then M is **invariant** with respect to A. That is,*

$$y \in M \Longrightarrow Ay \in M.$$

PROOF. So simple: if $y \in M$,

$$Ay \bullet x \;=\; y \bullet Ax \;=\; y \bullet \lambda x \;=\; \lambda(y \bullet x) \;=\; 0.$$

Thus $Ay \in M$.

QED

Looking ahead to Section D, we now see a very nice situation. We have essentially split $\mathbb{R}^n$ into a one-dimensional space and an $(n-1)$-dimensional space, and on each of them the geometric action of multiplying by A is clear. For the one-dimensional space lies in the direction of an eigenvector of A, so that A times any vector there is just λ times the vector. On the $(n-1)$-dimensional space M we don't know what A does except that we know that multiplication of vectors in M by A produces vectors that are still in M. This situation effectively reduces our analysis of A by one dimension. Then we can proceed by induction until we have produced n linearly independent eigenvectors.

C. **Two-dimensional pause**

We are now quite amply prepared to finish our analysis of the structure of real symmetric matrices. However, I would like to spend a little time discussing a standard "analytic" geometry problem, but viewed with eigenvector eyes. Here is an example of this sort of

PROBLEM. Sketch the curve in the $x - y$ plane given as the level set

$$10x^2 - 12xy + 5y^2 \;=\; 1.$$

The associated symmetric matrix is

$$A = \begin{pmatrix} 10 & -6 \\ -6 & 5 \end{pmatrix},$$

and the curve is given in vector notation as

$$A \begin{pmatrix} x \\ y \end{pmatrix} \bullet \begin{pmatrix} x \\ y \end{pmatrix} = 1.$$

Now we find the eigenvalues of A:

$$\begin{aligned} \det \begin{pmatrix} 10-\lambda & -6 \\ -6 & 5-\lambda \end{pmatrix} &= \lambda^2 - 15\lambda + 14 \\ &= (\lambda - 1)(\lambda - 14), \end{aligned}$$

so the eigenvalues are 1 and 14. The eigenvector for $\lambda = 14$ is given by solving

$$\begin{aligned} (A - 14I) \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix}; \\ \begin{pmatrix} -4 & -6 \\ -6 & -9 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \end{aligned}$$

Thus we may use the vector

$$\begin{pmatrix} 3 \\ -2 \end{pmatrix}.$$

Normalize it and call it $\hat{\varphi}_1$:

$$\hat{\varphi}_1 = \frac{1}{\sqrt{13}} \begin{pmatrix} 3 \\ -2 \end{pmatrix}.$$

For the other eigenvalue 1 we can use a shortcut, as we know from Section B it must be orthogonal to $\hat{\varphi}_1$. Thus we let

$$\hat{\varphi}_2 = \frac{1}{\sqrt{13}} \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

and we are guaranteed this is an eigenvector! (Here's verification:

$$\begin{aligned} A\hat{\varphi}_2 &= \frac{1}{\sqrt{13}} \begin{pmatrix} 10 & -6 \\ -6 & 5 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} \\ &= \frac{1}{\sqrt{13}} \begin{pmatrix} 2 \\ 3 \end{pmatrix} \\ &= \hat{\varphi}_2.) \end{aligned}$$

Now we use the unit vectors $\hat{\varphi}_1$ and $\hat{\varphi}_2$ as new coordinate directions, and call the coordinates s and t, respectively:

$$\begin{pmatrix} x \\ y \end{pmatrix} = s\hat{\varphi}_1 + t\hat{\varphi}_2.$$

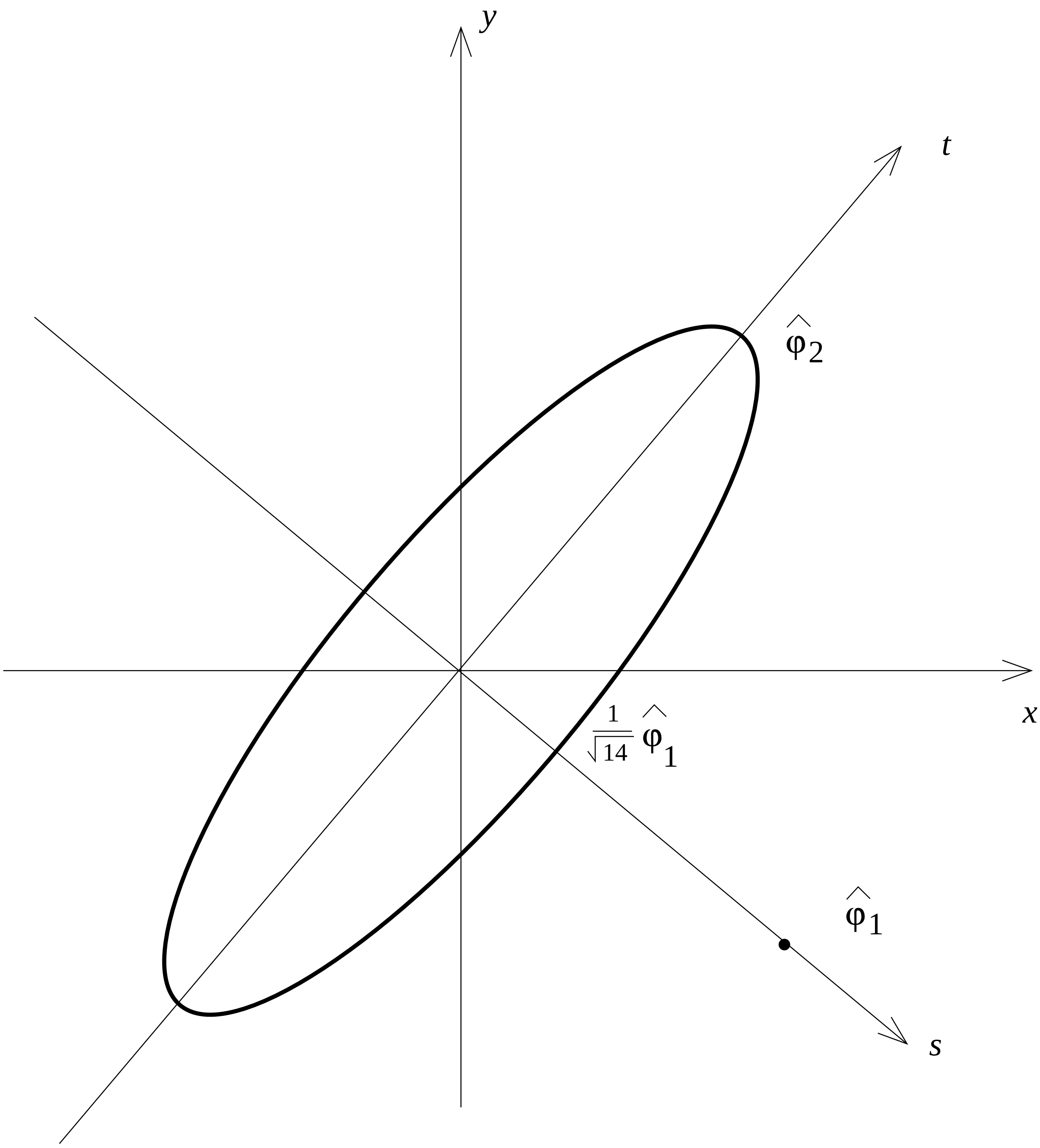
We calculate:

$$\begin{aligned} A\begin{pmatrix} x \\ y \end{pmatrix} \bullet \begin{pmatrix} x \\ y \end{pmatrix} &= (sA\hat{\varphi}_1 + tA\hat{\varphi}_2) \bullet (s\hat{\varphi}_1 + t\hat{\varphi}_2) \\ &= (14s\hat{\varphi}_1 + t\hat{\varphi}_2) \bullet (s\hat{\varphi}_1 + t\hat{\varphi}_2) \\ &= 14s^2 + t^2. \end{aligned}$$

(Notice: no term with st!) Thus we recognize our curve in this new coordinate system as the *ellipse*

$$14s^2 + t^2 = 1.$$

Now the sketch is easily finished: we simply locate $\hat{\varphi}_1$ and $\hat{\varphi}_2$, and the rest is easy. Here is the result:



What has happened here is clear. This ellipse is not well situated in the $x-y$ coordinate system. In other words, the directions $\hat{e}_1$ and $\hat{e}_2$ are not of much geometrical interest for it. But the directions $\hat{\varphi}_1$ and $\hat{\varphi}_2$ are extremely significant for this ellipse! In fact, $\hat{\varphi}_1$ is the direction of its minor axis, $\hat{\varphi}_2$ of its major axis. We say that $\hat{\varphi}_1$ and $\hat{\varphi}_2$ are the "principal axes" for the ellipse and for the matrix A. Notice of course that $\hat{\varphi}_1$ and $\hat{\varphi}_2$ are orthogonal.

(Another way of expressing this achievement is to think of the bilinear form $10x^2 - 12xy + 5y^2$ as the square of a certain *norm* of the vector $\begin{pmatrix} x \\ y \end{pmatrix}$. This is definitely not the Euclidean norm, of course. But it has essentially all the same properties, and in fact in the new coordinates $s' = \sqrt{14}s$ and $t' = t$ we have

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{\sqrt{14}} s' \hat{\varphi}_1 + t' \hat{\varphi}_2$$

and

$$10x^2 - 12xy + 5y^2 = (s')^2 + (t')^2,$$

so that the ellipse looks like the unit circle in the new coordinates.)

In the next section we are going to extend all of that to the n-dimensional case, and the result will be called the *principal axis theorem*.

Here are some exercises for you to try.

PROBLEM 4–10. Carry out the same procedure and thus accurately sketch the curve in the $x - y$ plane given by the level set $16x^2 + 4xy + 19y^2 = 300$.

PROBLEM 4–11. Repeat the preceding problem for the curve $23x^2 - 72xy + 2y^2 = 50$.

PROBLEM 4–12. A further wrinkle in problems of the sort just presented is the presence of first order terms in the equation. Here is the n-dimensional case. Let A be an $n \times n$ real symmetric matrix and $c \in \mathbb{R}^n$ and $a \in \mathbb{R}$ and consider the set described by

$$Ax \bullet x + c \bullet x \;=\; a.$$

Suppose $\det A \neq 0$. Then reduce this situation to one of the form

$$Ay \bullet y \;=\; b$$

by a judicious choice of x_0 in the translation $x = x_0 + y$. This is called "completing the square." The point is that in the x coordinates the center of the figure is x_0, but in the y coordinates it is 0.

PROBLEM 4–13. Accurately sketch the curve in the $x - y$ plane given as the level set $(x - 2y)^2 + 5y = 0$. Show that it is a parabola, and calculate its vertex.

D. **The principal axis theorem**

Now we come to the result we have been eagerly anticipating. This result is of major importance in a wide variety of applications in mathematics, physics, engineering, etc. In our case it is definitive in understanding the Hessian matrix at a nondegenerate critical point. It has a variety of names, including "The Spectral Theorem" and "Diagonalization of Symmetric Matrices." There is an important term used in the statement which we now define.

DEFINITION. If $\varphi_1, \varphi_2, \dots, \varphi_k$ are vectors in $\mathbb{R}^n$ which are mutually orthogonal and which have norms equal to 1, they are said to be *orthonormal*. In terms of the Kronecker symbol,

$$\varphi_i \bullet \varphi_j = \delta_{ij}.$$

Since the vectors have unit norm, we distinguish them with our usual symbol for unit vectors, $\hat{\varphi}_i$.

PROBLEM 4–14. Prove that the vectors in an orthonormal set are linearly independent.
(HINT: if $\sum_{i=1}^{k} c_i \hat{\varphi}_i = 0$, compute the inner product of both sides of the equation with $\hat{\varphi}_j$.)

Therefore it follows that if we have n orthonormal vectors in $\mathbb{R}^n$ (same n), they must form a basis for $\mathbb{R}^n$. See p. 3–37. We then say that they form an *orthonormal basis*. The coordinate vectors $\hat{e}_1, \hat{e}_2, \dots, \hat{e}_n$ are a standard example.

PROBLEM 4–15. Here is an orthonormal basis for $\mathbb{R}^2$:

$$\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}, \quad \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}.$$

Similarly, find an orthonormal basis for $\mathbb{R}^4$ for which each vector has the form

$$\frac{1}{2}\begin{pmatrix}\pm1\\\pm1\\\pm1\\\pm1\end{pmatrix}.$$

Find an analogous orthonormal basis for $\mathbb{R}^8$.

We found an orthonormal basis for $\mathbb{R}^2$ in our ellipse problem at the end of Section C, namely

$$\hat{\varphi}_1 = \frac{1}{\sqrt{13}}\begin{pmatrix}3\\-2\end{pmatrix}, \quad \hat{\varphi}_2 = \frac{1}{\sqrt{13}}\begin{pmatrix}2\\3\end{pmatrix}.$$

PROBLEM 4–16. Suppose $\hat{\varphi}_1, \hat{\varphi}_2, \dots, \hat{\varphi}_n$ are an orthonormal basis for $\mathbb{R}^n$. Prove that every x in $\mathbb{R}^n$ has the representation

$$x = \sum_{i=1}^{n} x \bullet \hat{\varphi}_i \hat{\varphi}_i.$$

Notice how very much the formula for x in the problem resembles the formula

$$x = \begin{pmatrix}x_1\\\vdots\\x_n\end{pmatrix} = \sum_{i=1}^{n} x_i \hat{e}_i.$$

In fact, it is the generalization to an arbitrary orthonormal basis instead of the basis of coordinate vectors.

PROBLEM 4–17. Orthonormal bases often provide nice information about determinants. Suppose $\hat{\varphi}_1, \hat{\varphi}_2, \dots, \hat{\varphi}_n$ are an orthonormal basis for $\mathbb{R}^n$, written as column vectors. Define the $n \times n$ matrix having them as columns:

$$\Phi = (\hat{\varphi}_1 \; \hat{\varphi}_2 \; \dots \; \hat{\varphi}_n).$$

a. Prove that $\Phi^t \Phi = I$.

b. Prove that $\det \Phi = \pm 1$.

c. Suppose A is a matrix such that the $\hat{\varphi}_i$'s are eigenvectors:

$$A\hat{\varphi}_i = \lambda_i \hat{\varphi}_i.$$

Prove that

$$A\Phi = (\lambda_1 \hat{\varphi}_1 \; \dots \; \lambda_n \hat{\varphi}_n).$$

d. Prove that

$$\det A = \lambda_1 \lambda_2 \dots \lambda_n.$$

PRINCIPAL AXIS THEOREM. *Let A be an $n \times n$ real symmetric matrix. Then there exist eigenvectors $\hat{\varphi}_1, \hat{\varphi}_2, \dots, \hat{\varphi}_n$ for A which form an orthonormal basis:*

$$A\hat{\varphi}_i = \lambda_i \hat{\varphi}_i, \qquad 1 \le i \le n.$$

The eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ are real numbers and are the zeros of the characteristic polynomial of A, repeated according to multiplicity.

PROOF. We are confident about using the quotient function $Q(x) = Ax \bullet x/\|x\|^2$. We have already proved in Section B that an eigenvector $\hat{\varphi}_1$ exists, and we are going to carry out a proof by induction on k, presuming we know an orthonormal sequence $\hat{\varphi}_1, \dots, \hat{\varphi}_k$ of eigenvectors. We assume $1 \le k \le n-1$. We define

$$M = \{y \in \mathbb{R}^n \mid y \bullet \hat{\varphi}_1 = \dots = y \bullet \hat{\varphi}_k = 0\}.$$

(This is a subspace of $\mathbb{R}^n$ of dimension $n-k$.) We restrict the continuous function Q to the closed bounded set $M \cap S(0,1)$. It attains a maximum value there, say at a point $\hat{x}_0$. Thus $\|\hat{x}_0\| = 1$ and $\hat{x}_0 \bullet \hat{\varphi}_1 = \dots = \hat{x}_0 \bullet \hat{\varphi}_k = 0$. Because Q is homogeneous of degree 0, we know in fact that $Q(x) \le Q(x_0)$ for all $x \in M$; this is the same argument we used on p. 4–8.

This implies that **for all h ∈ M** we have

$$Q(\hat{x}_0 + th) \;\le\; Q(\hat{x}_0), \qquad -\infty < t < \infty.$$

And this gives a maximum value at $t = 0$, so that

$$\frac{d}{dt} Q(\hat{x}_0 + th) \,\Big|_{t=0} \;=\; 0.$$

That is, the directional derivative $DQ(\hat{x}_0; h) = 0$. That is,

$$\nabla Q(\hat{x}_0) \bullet h \;=\; 0 \qquad \text{for all} \;\; h \in M.$$

Now the boxed formula on p. 4–9 asserts that

$$\nabla Q(\hat{x}_0) \;=\; 2\,(A\hat{x}_0 - Q(\hat{x}_0)\hat{x}_0)\,.$$

We know from the theorem on p. 4–10 that $A\hat{x}_0 \in M$, and thus $\nabla Q(\hat{x}_0) \in M$. But since $\nabla Q(\hat{x}_0)$ is orthogonal to all vectors in M, it is orthogonal to itself, and we conclude $\nabla Q(\hat{x}_0) = 0$. Thus $\hat{x}_0$ is a critical point for Q!

That does it, for $A\hat{x}_0 = Q(\hat{x}_0)\hat{x}_0$. We just name $\hat{x}_0 = \hat{\varphi}_{k+1}$ and $Q(\hat{x}_0) = \lambda_{k+1}$. We have thus produced an orthonormal sequence $\hat{\varphi}_1, \dots, \hat{\varphi}_{k+1}$ of eigenvectors of A. By induction, the proof is over, except for one small matter. That is the statement about the characteristic polynomial. But notice that

$$(A - \lambda I)\hat{\varphi}_i \;=\; (\lambda_i - \lambda)\hat{\varphi}_i,$$

and thus Problem 4–17 yields

$$\det(A - \lambda I) \;=\; (\lambda_1 - \lambda) \dots (\lambda_n - \lambda).$$

QED

REMARK. This is an unusual sort of induction argument. If you examine it carefully, you will notice that it really applies even in the case $k = 0$. There it is exactly the proof we gave in Section B. Thus we don't even actually need the proof of Section B, nor do we need a separate argument to "start" the induction. This is quite a happy situation: the starting point of the induction argument is not only easy, it is actually vacuous (there's nothing to check).

PROBLEM 4–18. This is sort of an easy "converse" of the principal axis theorem. Given any orthonormal sequence $\hat{\varphi}_1, \dots, \hat{\varphi}_n$ in $\mathbb{R}^n$ and any real numbers $\lambda_1, \dots, \lambda_n$, there exists one and only one $n \times n$ real matrix A such that

$$A\hat{\varphi}_i \;=\; \lambda_i \hat{\varphi}_i \qquad \text{for all} \;\; 1 \le i \le n.$$

Prove that A is symmetric.
(HINT: use Φ from Problem 4–17.)

PROBLEM 4–19. Find a 4×4 matrix A such that

$$A \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad A \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ -2 \\ -2 \end{pmatrix},$$

$$A \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad \text{and} \quad A \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 5 \\ -5 \end{pmatrix}.$$

The sort of matrix that was introduced in Problem 4–17 is exceedingly important in understanding both the algebra and the geometry of our Euclidean space $\mathbb{R}^n$. We need to understand all of this in great detail, so we pause to give the definition.

DEFINITION. A real $n \times n$ matrix Φ is an *orthogonal matrix* if its columns are an orthonormal basis for $\mathbb{R}^n$. That is,

$$\Phi = (\hat{\varphi}_1 \; \hat{\varphi}_2 \ldots \hat{\varphi}_n)$$

and $\hat{\varphi}_i \bullet \hat{\varphi}_j = \delta_{ij}$. The set of all orthogonal $n \times n$ matrices is denoted

$$O(n).$$

You noticed in Problem 4–17 that an equivalent way of asserting that Φ is orthogonal is the matrix formula $\Phi^t\Phi = I$. Thus, that Φ^t is a left inverse of Φ. But the theorem on p. 3–37 then asserts that Φ is invertible and has the inverse Φ^t. Thus, $\Phi\Phi^t = I$ as well. Here is a problem that summarizes this information, and more:

PROBLEM 4–20. Prove the following properties of $O(n)$:

a. $\Phi \in O(n) \iff$ the columns of Φ form an orthonormal basis for $\mathbb{R}^n$ (this is actually our definition).

b. $\Phi \in O(n) \iff \Phi^t$ is the inverse of Φ.

c. $\Phi \in O(n) \iff$ the rows of Φ form an orthonormal basis for $\mathbb{R}^n$.

d. $\Phi \in O(n) \implies \Phi^t \in O(n)$.

e. $\Phi \in O(n) \iff \Phi x \bullet \Phi y = x \bullet y$ for all x, $y \in \mathbb{R}^n$.

f. $\Phi \in O(n) \iff \|\Phi x\| = \|x\|$ for all $x \in \mathbb{R}^n$.

g. $\Phi \in O(n) \implies \Phi^{-1} \in O(n)$.

h. Φ, $\Phi' \in O(n) \implies \Phi\Phi' \in O(n)$.

(HINT for f: the hard part is $\Leftarrow$. Try showing that the condition in part e is satisfied, by verifying

$$2\Phi x \bullet \Phi y = \|\Phi(x+y)\|^2 - \|\Phi x\|^2 - \|\Phi y\|^2.)$$

DISCUSSION. Because of the last two properties in the problem, $O(n)$ is called the *orthogonal group*. The word "group" is a technical one which signifies the fact that products of group elements belong to the group, that there is an identity for the product (in this case it's the identity matrix I), and that each member of the group has a unique inverse (which also belongs to the group).

Notice how easy it is to compute the inverse of an orthogonal matrix!

DEFINITION. The set of all $n \times n$ invertible real matrices is called the *general linear group* and is denoted

$$\mathrm{GL}(n).$$

The set of all $n \times n$ real matrices with determinant 1 is called the *special linear group* and is denoted

$$\mathrm{SL}(n).$$

Every orthogonal matrix has determinant equal to ± 1 (Problem 4–17). The set of all orthogonal matrices with determinant 1 is called the *special orthogonal group* and is denoted

$$\mathrm{SO}(n).$$

Clearly,

$$\mathrm{SO}(n) \subset \mathrm{SL}(n) \subset \mathrm{GL}(n)$$

and

$$\mathrm{SO}(n) = O(n) \cap \mathrm{SL}(n).$$

PROBLEM 4–21. Prove that $\mathrm{GL}(n)$, $\mathrm{SL}(n)$, and $\mathrm{SO}(n)$ are all groups.

PROBLEM 4–22. Let $\hat{\varphi}_1, \dots, \hat{\varphi}_n$ be an orthonormal basis for $\mathbb{R}^n$, and A any $n \times n$ real or complex matrix. Prove that

$$\mathrm{trace} A = \sum_{i=1}^{n} A\hat{\varphi}_i \bullet \hat{\varphi}_i.$$

E. **Positive definite matrices**

In this section we lay the foundation for understanding the Hessian matrices we are so interested in.

Let A be an $n \times n$ real symmetric matrix. The principal axis theorem guarantees the existence of an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of A:

$$A\hat{\varphi}_i = \lambda_i \hat{\varphi}_i, \qquad 1 \le i \le n.$$

As we have discussed, the unit vectors $\hat{\varphi}_1, \dots, \hat{\varphi}_n$ are very natural as far as the matrix A is concerned. We now use them essentially as a new set of "coordinate axes" for $\mathbb{R}^n$. That is, every $x \in \mathbb{R}^n$ has a unique representation of the form

$$x = \sum_{i=1}^{n} s_i \hat{\varphi}_i.$$

The numbers $s_1, \dots, s_n$ are the "coordinates" of x in this new basis. They can be calculated directly by using the inner product:

$$s_i = x \bullet \hat{\varphi}_i.$$

Now we calculate the quadratic form we are interested in. In the Cartesian coordinates it is of course

$$Ax \bullet x = \sum_{i,j=1}^{n} a_{ij} x_i x_j.$$

In the more natural coordinates it is computed as follows:

$$\begin{aligned} Ax \bullet x &= \sum_{i=1}^{n} s_i A\hat{\varphi}_i \bullet \sum_{j=1}^{n} s_j \hat{\varphi}_j \\ &= \sum_{i=1}^{n} s_i \lambda_i \hat{\varphi}_i \bullet \sum_{j=1}^{n} s_j \hat{\varphi}_j \\ &= \sum_{i,j=1}^{n} \lambda_i s_i s_j \hat{\varphi}_i \bullet \hat{\varphi}_j \\ &= \sum_{i,j=1}^{n} \lambda_i s_i s_j \delta_{ij} \\ &= \sum_{i=1}^{n} \lambda_i s_i^2. \end{aligned}$$

Of course, the orthonormality was of crucial importance in that calculation. An example of this sort of calculation appears in Section 4C.

The result is that $Ax \bullet x$ looks much nicer in the coordinates that come from the eigenvectors of A than in the original Cartesian coordinates. We reiterate,

$$Ax \bullet x = \sum_{i=1}^{n} \lambda_i s_i^2.$$

In this form we can deduce everything we need to know about the quadratic form $Ax \bullet x$. For instance, we know in case A is the Hessian matrix of a function at a critical point, then the critical point is a local minimum for the function if $Ax \bullet x > 0$ for all $x \neq 0$. We see instantly that this condition is equivalent to $\lambda_i > 0$ for all i:

THEOREM. *In the above situation the eigenvalues of A are all positive $\iff$ $Ax \bullet x > 0$ for all $x \in \mathbb{R}^n - \{0\}$.*

PROOF. For the direction $\Longleftarrow$, apply the given inequality to $x = \hat{\varphi}_i$. Then $0 < A\hat{\varphi}_i \bullet \hat{\varphi}_i = \lambda_i \hat{\varphi}_i \bullet \hat{\varphi}_i = \lambda_i$. Thus all the eigenvalues of A are positive. This much of the theorem did not require A to be symmetric. However, the converse direction $\Longrightarrow$ relies on the principal axis theorem. According to the calculation given above, $Ax \bullet x = \sum \lambda_i s_i^2 \geq 0$ since all $\lambda_i > 0$, and $Ax \bullet x = 0$ implies each $s_i = 0$ and thus $x = \sum_{i=1}^n s_i \hat{\varphi}_i = 0$.

QED

DEFINITION. The real symmetric matrix A is said to be *positive definite* in case the above equivalent conditions are satisfied. That is, all the eigenvalues of A are positive. Equivalently, $Ax \bullet x > 0$ for all $x \in \mathbb{R}^n - \{0\}$.

Of course, we say A is *negative definite* if all the eigenvalues of A are negative; equivalently, $Ax \bullet x < 0$ for all $x \neq 0$; equivalently, $-A$ is positive definite.

PROBLEM 4–23. Give an example of a real 2×2 matrix for which both eigenvalues are positive numbers but which does not satisfy $Ax \bullet x \geq 0$ for all $x \in \mathbb{R}^2$. (Of course, this matrix cannot be symmetric.)

It is quite an interesting phenomenon that positive definite matrices are analogous to positive numbers. The next result provides one of the similarities.

THEOREM. *A real symmetric matrix A is positive definite $\Longleftrightarrow$ there exists a real symmetric invertible matrix B such that $A = B^2$.*

PROOF. If $A = B^2$, then $Ax \bullet x = B^2 x \bullet x = Bx \bullet Bx = \|Bx\|^2 \geq 0$, and equality holds $\Longleftrightarrow Bx = 0 \Longleftrightarrow x = 0$ (since B is invertible). Conversely, use the eigenvectors $\hat{\varphi}_i$ of A to define the orthogonal matrix

$$\Phi = (\hat{\varphi}_1 \ \hat{\varphi}_2 \ \ldots \ \hat{\varphi}_n).$$

Then

$$\begin{aligned} \Phi^t A \Phi &= \begin{pmatrix} \hat{\varphi}_1^t \\ \vdots \\ \hat{\varphi}_n^t \end{pmatrix} (\lambda_1 \hat{\varphi}_1 \ldots \lambda_n \hat{\varphi}_n) \\ &= \begin{pmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & \ddots & \lambda_n \end{pmatrix}, \end{aligned}$$

so that

$$A = \Phi \begin{pmatrix} \lambda_1 & & 0 & \\ & \lambda_2 & & \\ & 0 & \ddots & \\ & & & \lambda_n \end{pmatrix} \Phi^t.$$

Now simply define

$$B = \Phi \begin{pmatrix} \sqrt{\lambda_1} & & & 0 \\ & \sqrt{\lambda_2} & & \\ & 0 & \ddots & \\ & & & \sqrt{\lambda_n} \end{pmatrix} \Phi^t.$$

Then B is even positive definite, and $B^2 = A$. (We say B is a *positive definite square root* of A.)

QED

What has happened in the above proof is tremendously interesting, probably more interesting than the theorem itself. Namely, starting with A we have used the principal axis theorem to represent it as simply as possible in coordinates tied closely to the geometry which A gives. In that coordinate system it is easy to find a square root of A, and then we "undo" the coordinate change to get the matrix B.

PROBLEM 4–24. Find a positive definite square root of

$$A = \begin{pmatrix} 16 & 2 \\ 2 & 19 \end{pmatrix}$$

(see Problem 4–10).

PROBLEM 4–25*. Prove that a positive definite matrix A has a *unique* positive definite square root. (For this reason, we can denote it $\sqrt{A}$.)
(HINT: suppose B is positive definite and $B^2 = A$. Show that if λ is an eigenvalue of A and $Ax = \lambda x$, then $Bx = \sqrt{\lambda}x$.)

PROBLEM 4–26. Show that $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ has no square root whatsoever. That is, there is no 2×2 matrix B even with complex entries such that B^2 equals the given matrix.

PROBLEM 4–27. Prove that if A is positive definite, then so is A^{-1}.

Now we are ready to focus our attention on the real issue we want to understand. Remember, we are trying to understand how to detect the nature of critical points of real-valued functions. Referring to Section 3H, "Recapitulation," we see that the crucial quantity is $Hy \bullet y$, where H is the Hessian matrix at the critical point of the function. We assume the critical point is nondegenerate, in other words that $\det H \neq 0$. Now think about whether we have a relative minimum. This translates to $Hy \bullet y > 0$ for all $y \neq 0$, as we shall prove in Section F. In other words, the condition for a relative minimum is going to be that H is positive definite.

Thus we are facing an algebra question: how can we tell whether a symmetric matrix is positive definite? The immediate but naive answer is just to respond: precisely when its eigenvalues are all positive.

However, we know that this is a potentially difficult matter for $n \times n$ matrices with $n > 2$, as calculating the eigenvalues may be difficult. In fact, usually only numerical approximations are available. The truly amazing thing is that there is an algorithm for detecting that all the eigenvalues of a symmetric matrix A are all positive, *without actually calculating the eigenvalues of A at all*. The fact is, we have observed this very feature in the $n = 2$ case. For we know (Problem 3–18) that A is positive definite $\iff$

$$a_{11} > 0, \quad a_{22} > 0, \quad \text{and} \quad a_{11}a_{22} - a_{12}^2 > 0.$$

In fact, we could drop the second inequality and simply write

$$a_{11} > 0 \quad \text{and} \quad \det A > 0.$$

Notice that *calculating* the eigenvalues in this case requires the square root of $(a_{11}-a_{22})^2+4a_{12}^2$, but our test requires no such thing.

The $n \times n$ case has a similar simplification:

THE DEFINITENESS CRITERION. *Let A be an $n \times n$ real symmetric matrix. For any $1 \leq k \leq n$, let $A(k)$ be the $k \times k$ "northwest" square submatrix of A:*

$$A(k) = \begin{pmatrix} a_{11} & \dots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \dots & a_{kk} \end{pmatrix}.$$

(Thus,

$$\begin{aligned} A(1) &= (a_{11}), \\ A(2) &= \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \\ &\vdots \\ A(n) &= A.) \end{aligned}$$

Then A *is positive definite* $\iff \det A(k) > 0$ *for all* $1 \le k \le n$.

(By the way, we could just as well have elected to employ the corresponding southeast submatrices instead. More about this after the theorem.)

PROOF. We first make a simple observation: if a matrix is positive definite, then its determinant is positive. The reason is that its determinant is equal to the product of its eigenvalues (p. 4–6), which are all positive.

It is rather evident that the direction $\Longrightarrow$ of the proof should be the easier one, so we attack it first. Suppose that A is positive definite. Then we prove directly that each $A(k)$ is positive definite; the above observation then completes this part of the proof. For a fixed $1 \le k \le n$, let $y \in \mathbb{R}^k$ be arbitrary, $y \ne 0$. Then define $x \in \mathbb{R}^n$ by

$$x = \begin{pmatrix} y_1 \\ \vdots \\ y_k \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Then it is true that

$$Ax \bullet x = A(k)y \bullet y.$$

Since A is positive definite, $Ax \bullet x > 0$. Thus $A(k)y \bullet y > 0$. Thus $A(k)$ is positive definite.

Now we prove the converse direction $\Longleftarrow$. We do it by induction on n, the case $n = 1$ being obvious. Thus we assume the result is valid for the case $n - 1$, where $n \ge 2$, and we prove it for an $n \times n$ matrix A. Thus we are assuming that each $A(k)$ has positive determinant. By the induction hypothesis, $A(n-1)$ is positive definite.

We now use the principal axis theorem to produce orthonormal eigenvectors for A. Actually, for the present proof it is convenient to assume only that they are orthogonal (and nonzero), and that all of them with n^{th} coordinate nonzero have been rescaled to have n^{th} coordinate equal to 1:

$$\begin{aligned} &A\varphi_i = \lambda_i \varphi_i, \quad 1 \le i \le n; \\ &\varphi_1, \dots, \varphi_n \ \text{ are orthogonal and nonzero;} \\ &\text{the } n^{\text{th}} \text{ coordinate of each } \varphi_i \text{ is 0 or 1.} \end{aligned}$$

By Problem 4–17,

$$0 < \det A = \lambda_1 \lambda_2 \dots \lambda_n.$$

Each φ_i with n^{th} coordinate 0 can be written

$$\varphi_i = \begin{pmatrix} y_1 \\ \vdots \\ y_{n-1} \\ 0 \end{pmatrix},$$

where $y \in \mathbb{R}^{n-1}$ and $y \neq 0$, so that

$$\begin{aligned} \lambda_i \|\varphi_i\|^2 &= \lambda_i \varphi_i \bullet \varphi_i \\ &= A\varphi_i \bullet \varphi_i \\ &= A(n-1)y \bullet y \\ &> 0, \end{aligned}$$

since $A(n-1)$ is positive definite. Thus $\lambda_i > 0$.

Now suppose *two* of the eigenvectors φ_i and φ_j have n^{th} coordinate 1. Then $\varphi_i - \varphi_j$ has n^{th} coordinate 0 and is not itself 0, so as above we conclude that since $A(n-1)$ is positive definite,

$$\begin{aligned} 0 &< A(\varphi_i - \varphi_j) \bullet (\varphi_i - \varphi_j) \\ &= (\lambda_i \varphi_i - \lambda_j \varphi_j) \bullet (\varphi_i - \varphi_j) \\ &= \lambda_i \|\varphi_i\|^2 + \lambda_j \|\varphi_j\|^2 \quad \text{(by orthogonality).} \end{aligned}$$

Thus at least one of λ_i and λ_j is *positive*.

This leads to an interesting conclusion indeed! Among all the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, *at most one* of them is negative! Since their product is positive, they must all be positive. Therefore, A is positive definite.

QED

DISCUSSION. For any $n \times n$ real symmetric matrix A with $\det A \neq 0$, we now completely understand the criterion for A to be positive definite. Next, A is negative definite $\iff -A$ is positive definite $\iff \det(-A(k)) > 0$ for all $k \iff (-1)^k \det A(k) > 0$ for all k. Thus we obtain the negative definite result for free.

SUMMARY. When we examine the signs of the determinants $\det A(k)$ in order for $k = 1, 2, \ldots, n$, there are exactly three cases:

- $+, +, +, +, \ldots \iff A$ is positive definite.
- $-, +, -, +, \ldots \iff A$ is negative definite.

- any other pattern $\Longleftrightarrow$ A is neither positive nor negative definite.

PROBLEM 4–28. State and prove the corresponding criterion, using instead the $k \times k$ southeast square submatrices of A.

PROBLEM 4–29. Let A be an $n \times n$ real *diagonal* matrix with $\det A \neq 0$. Show that the definiteness criterion is virtually obvious for A. (Thus the useful content of the criterion is for matrices which are not diagonal.)

THE DEGENERATE CASE. The definiteness criterion of course deals only with the nondegenerate case in which $\det A \neq 0$. There is a companion result which is valid even if $\det A = 0$. Although this criterion appears to be of little interest in the classification of critical points, since we need them to be nondegenerate, we include the material in the rest of this section for the beautiful mathematics that is involved. We continue to work with an $n \times n$ real symmetric matrix. Such a matrix A is said to be *positive semidefinite* if

$$Ax \bullet x \ \geq \ 0 \quad \text{for all} \ \ x \in \mathbb{R}^n.$$

Equivalently, all the eigenvalues of A are nonnegative numbers.

What you might expect the definiteness criterion to assert is that the equivalent condition is $\det A(k) \geq 0$ for all $1 \leq k \leq n$. However, a simple 2×2 example belies this:

$$\begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}.$$

The key is to realize that our restriction to *northwest* square submatrices is rather artificial. Instead we should use all possible "symmetric square submatrices." These are matrices A' obtained from A by using only the entries a_{ij} where i, j are restricted to the same collection of indices. Put in negative terms, we have deleted some rows of A as well as the corresponding columns. Whereas there are n symmetric square submatrices of the form A_k, there are $2^n - 1$ symmetric square submatrices in all.

THE DEFINITENESS CRITERION BIS. *Let A be an $n \times n$ real symmetric matrix. Then A is positive semidefinite $\Longleftrightarrow$ every symmetric square submatrix A' satisfies*

$$\det A' \ \geq \ 0.$$

PROOF. The $\Longrightarrow$ direction of the proof is just as before. The $\Longleftarrow$ direction is again proved by induction on the dimension; the $n = 1$ case is trivial and we presume the $n - 1$ case is valid.

We again use a principal axis decomposition as before,

$$\begin{aligned} A\hat{\varphi}_i &= \lambda_i\hat{\varphi}_i \qquad \text{for } 1 \le i \le n, \\ \hat{\varphi}_1, \dots, \hat{\varphi}_n & \qquad \text{are orthonormal.} \end{aligned}$$

If $\det A \neq 0$, then the theorem is already known from the previous result, so there is nothing to prove. Thus we may assume $\det A = 0$, so that A has 0 as one of its eigenvalues. We may as well assume $\lambda_1 = 0$. Now consider any $2 \le i \le n$. There exists a scalar c such that $\hat{\varphi}_i - c\hat{\varphi}_1$ has at least one coordinate equal to 0 (it could happen that $c = 0$). Say its j^{th} coordinate is 0. Then we choose the particular symmetric square submatrix A' obtained by deleting the j^{th} row and the j^{th} column from A (thus A' equals the *minor* A_{jj} as defined on p. 3–29). Also let $y \in \mathbb{R}^{n-1}$ be obtained from $\varphi_i - c\varphi_1$ by simply deleting its j^{th} (zero) entry.

By the inductive hypothesis, A' is positive semidefinite. Therefore

$$\begin{aligned} 0 &\le A'y \bullet y \\ &= A(\hat{\varphi}_i - c\hat{\varphi}_1) \bullet (\hat{\varphi}_i - c\hat{\varphi}_1) \\ &= (A\hat{\varphi}_i - cA\hat{\varphi}_1) \bullet (\hat{\varphi}_i - c\hat{\varphi}_1) \\ &= (\lambda_i\hat{\varphi}_i - c\lambda_1\hat{\varphi}_1) \bullet (\hat{\varphi}_i - c\hat{\varphi}_1) \\ &= \lambda_i\hat{\varphi}_i \bullet (\hat{\varphi}_i - c\hat{\varphi}_1) \qquad (\lambda_1 = 0) \\ &= \lambda_i\hat{\varphi}_i \bullet \hat{\varphi}_i \qquad (\hat{\varphi}_i \text{ and } \hat{\varphi}_1 \text{ are orthogonal}) \\ &= \lambda_i. \end{aligned}$$

Thus $\lambda_i \ge 0$ for all $2 \le i \le n$ (and $\lambda_1 = 0$). Thus A is positive semidefinite.

QED

PROBLEM 4–30. Assume A is positive semidefinite and $a_{ii} = 0$. Prove that the i^{th} row and the i^{th} column of A consist of zeros. Prove that if A is positive definite, then $a_{ii} > 0$.

PROBLEM 4–31. Suppose A is an $n \times n$ positive definite matrix. Prove that

$$(\det A)^{\frac{1}{n}} \le \frac{\text{trace}A}{n}$$

and that equality holds $\iff A = cI$ for some $c > 0$.
(HINT: use the arithmetic-geometric mean inequality (Problem 5–31) for the eigenvalues of A.)

PROBLEM 4–32. Suppose A is an $n \times n$ positive definite matrix. Prove that

$$\det A \ \le \ a_{11}\, a_{22}\, \ldots\, a_{nn}.$$

Prove that equality holds $\iff$ A is a diagonal matrix.
(HINT: let $B =$ the diagonal matrix with entries $\sqrt{a_{ii}}$. Let $C = B^{-1}AB^{-1}$ and apply the preceding problem to C.)

PROBLEM 4–33. Suppose A is an $n \times n$ positive semidefinite matrix. Prove that

$$\det A \le a_{11}\, a_{22}\, \ldots\, a_{nn},$$

and that equality holds $\iff$ A is a diagonal matrix or some $a_{ii} = 0$.

PROBLEM 4–34. Suppose A is an $n \times n$ real matrix with columns $a_1, \ldots, a_n$ in $\mathbb{R}^n$:

$$A = (a_1\ a_2\ \ldots\ a_n).$$

Show that Problem 4–33 may be applied to the matrix $B = A^t A$, and results in what is known as **Hadamard's inequality**:

$$|\det A| \ \le \ \|a_1\|\, \|a_2\|\, \ldots\, \|a_n\|.$$

When can equality hold?

We shall see in Section 8A that Hadamard's inequality has a very appealing geometric interpretation: the volume of an n-dimensional parallelogram is no greater than the product of its edge lengths.

There is another approach to the analysis of positive semidefinite matrices that is quite elegant. This approach is completely algebraic in nature and thus entirely different from that we have seen thus far. It begins with a discussion of the determinant of a *sum* of two matrices. Suppose then that A and B are matrices represented in terms of column vectors in the usual way:

$$A = (a_1\ a_2\ \ldots\ a_n),$$
$$B = (b_1\ b_2\ \ldots\ b_n).$$

Thus a_j and $b_j \in \mathbb{R}^n$. Then the multilinearity of the determinant represents $\det(A+B)$ as a sum of 2^n determinants, where each summand is the determinant of a matrix C of the form

$$(c_1\ c_2\ \dots\ c_n),$$

where each c_j is either a_j or b_j.

(This is very similar to the binomial expansion of $(x+y)^n$ when the power is regarded as

$$(x+y)^n = (x+y)\ (x+y)\ \dots\ (x+y)$$

and all the multiplications are carried out, resulting in 2^n terms.)

Now specialize this formula to the case $B = \lambda I$. Then $b_j = \lambda \hat{e}_j$, and when $n-k$ of the columns of C come from the b_j's, the resulting determinant is

$$\det C = \lambda^{n-k} \det A',$$

where A' is the $k \times k$ square submatrix of A resulting from eliminating the particular $n-k$ rows and columns of A corresponding to this choice of C. Thus

$$\det(A+\lambda I) = \sum_{k=0}^{n} \lambda^{n-k} \sum_{A' \text{ is } k\times k} \det A', \tag{$*$}$$

where each A' in the inner sum is one of the $\binom{n}{k}$ square submatrices of A resulting from deleting k rows and the same k columns. For instance, the $n=2$ case is

$$\det(A+\lambda I) = \lambda^2 + \lambda(a_{11}+a_{22}) + \det A.$$

(Notice that when $k=0$ we are required to interpret the coefficient of λ^n as 1.)

Notice that replacing λ by $-\lambda$ in $(*)$ gives an explicit formula for the characteristic polynomial of A.

Here then is the punch line. Suppose we want to prove the hard direction $\Longleftarrow$ of the definiteness criterion for positive semidefiniteness. We thus assume A is symmetric and every A' satisfies $\det A' \geq 0$. Then for all $\lambda > 0$ we have $\det(A+\lambda I) > 0$. Therefore, if λ is an eigenvalue of A, $\det(A-\lambda I) = 0$ and we conclude that $-\lambda \leq 0$. Thus all the eigenvalues of A are nonnegative, proving that A is positive semidefinite.

PROBLEM 4–35. Prove that a positive semidefinite matrix has a unique positive semidefinite square root.

F. **The second derivative test**

We return at last to the calculus problem we were interested in, as summarized at the close of Chapter 3. We use that outline exactly as written, and we assume that the critical point x_0 for the function f is nondegenerate, so that the determinant of the Hessian matrix H is not zero. The test we are going to state is in terms of the *definiteness* of H, and we realize that the definiteness criterion of Section E may be useful in deciding this in a given case. However, we do not need to refer to the rule in the statement of the result.

THEOREM. *Assume the following:*

- $\mathbb{R}^n \xrightarrow{f} \mathbb{R}$ *is of class* C^2 *in a neighborhood of* x_0.
- x_0 *is a critical point of* x_0.
- x_0 *is nondegenerate.*
- H *is the Hessian matrix of* f *at* x_0.

Then the conclusion is definitive:

- f *has a strict local minimum at* $x_0 \Longleftrightarrow H$ *is positive definite.*
- f *has a strict local maximum at* $x_0 \Longleftrightarrow H$ *is negative definite.*
- f *has a saddle point at* $x_0 \Longleftrightarrow H$ *is neither positive nor negative definite.*

PROOF. We have the Taylor expansion from Section 3B,

$$f(x_0 + y) \;=\; f(x_0) + \frac{1}{2} Hy \bullet y + R,$$

where $|R|$ is smaller than quadratic as $y \to 0$.

• Assume H is positive definite. Then we use a principal axis representation for H as on p. 4–22, writing

$$y \;=\; \sum_{i=1}^{n} s_i \hat{\varphi}_i,$$

so that

$$Hy \bullet y \;=\; \sum_{i=1}^{n} \lambda_i s_i^2.$$

All $\lambda_i > 0$, so let $\lambda = \min(\lambda_1, \ldots, \lambda_n)$. Then $\lambda > 0$ and

$$Hy \bullet y \;\geq\; \lambda \sum_{i=1}^{n} s_i^2 \;=\; \lambda \|y\|^2.$$

Choose $\delta > 0$ such that for $\|y\| \leq \delta$ we have $|R| \leq \frac{\lambda}{4}\|y\|^2$. Then $0 < \|y\| \leq \delta \Longrightarrow$

$$\begin{aligned} f(x_0 + y) &\geq f(x_0) + \frac{1}{2}\lambda\|y\|^2 - |R| \\ &\geq f(x_0) + \frac{1}{2}\lambda\|y\|^2 - \frac{1}{4}\lambda\|y\|^2 \\ &= f(x_0) + \frac{1}{4}\lambda\|y\|^2 \\ &> f(x_0). \end{aligned}$$

Thus f has a strict local minimun at x_0.

• If H is negative definite, the same proof yields a strict local maximum at x_0 (or simply apply the previous result to $-f$).

• If H is neither positive nor negative definite, then since all its eigenvalues are nonzero, it must have a positive eigenvalue and a negative eigenvalue. Suppose for example that $\lambda_i < 0$. Then

$$\begin{aligned} f(x_0 + t\hat{\varphi}_i) &= f(x_0) + \frac{t^2}{2}H\hat{\varphi}_i \bullet \hat{\varphi}_i + R \\ &= f(x_0) + \frac{1}{2}\lambda_i t^2 + R \\ &\leq f(x_0) + \frac{1}{2}\lambda_i t^2 + |R|. \end{aligned}$$

Now choose $\delta > 0$ so that

$$\|y\| \leq \delta \Longrightarrow |R(y)| \leq -\frac{1}{4}\lambda_i\|y\|^2.$$

Then $0 < |t| \leq \delta \Longrightarrow$

$$\begin{aligned} f(x_0 + t\hat{\varphi}_i) &\leq f(x_0) + \frac{1}{2}\lambda_i t^2 - \frac{1}{4}\lambda_i t^2 \\ &= f(x_0) + \frac{1}{4}\lambda_i t^2 \\ &< f(x_0). \end{aligned}$$

Thus f does not have a local minimum at x_0. Likewise, using a positive eigenvalue shows that f does not have a local maximum at x_0. Thus x_0 is a saddle point.

Thus far we have covered the three implications $\Longleftarrow$. But since the three assertions on the left sides of the statements as well as on the right sides are mutually exclusive, the proof is finished.

QED

G. **A little matrix calculus**

We take the viewpoint of Section 3I, thinking of $n \times n$ real matrices as being the Euclidean space $\mathbb{R}^{n^2}$. Now we want to think of the calculus of the real-valued function det.

PROBLEM 4–36. Use the formula (∗) of Section 4E to write

$$\det(A + \lambda I) = \lambda^n + \lambda^{n-1}\text{trace}A + \lambda^{n-2}R + \dots,$$

where

$$R = \sum_{1 \le i \le j \le n} (a_{ii}a_{jj} - a_{ij}a_{ji}).$$

PROBLEM 4–37. In the preceding problem perform algebraic manipulations to rewrite

$$R = \frac{1}{2}\sum_{i,j}(a_{ii}a_{jj} - a_{ij}a_{ji})$$
$$= \frac{1}{2}[(\text{trace}A)^2 - \text{trace}(A^2)].$$

PROBLEM 4–38. Manipulate Problem 4–35 in such a way to achieve the polynomial equation

$$\det(I + tB) = I + t\text{trace}B + \text{ higher order terms in } t.$$

Conclude that the differential of det at I is the linear mapping trace. In terms of directional derivatives,

$$D\det(I; B) = \text{trace}B.$$

PROBLEM 4–39. Generalize the preceding result to obtain

$$D\det(A; B) = \operatorname{trace}(B\operatorname{adj}A).$$