Numerical Analysis Lecture Notes

Peter J. Olver

6. Eigenvalues and Singular Values

In this section, we collect together the basic facts about eigenvalues and eigenvectors. From a geometrical viewpoint, the eigenvectors indicate the directions of pure stretch and the eigenvalues the extent of stretching. Most matrices are complete, meaning that their (complex) eigenvectors form a basis of the underlying vector space. A particularly important class are the symmetric matrices, whose eigenvectors form an orthogonal basis of $\mathbb{R}^n$. A non-square matrix A does not have eigenvalues. In their place, one uses the square roots of the eigenvalues of the associated square Gram matrix $K = A^T A$, which are called singular values of the original matrix. The numerical computation of eigenvalues and eigenvectors is a challenging issue, and must be be deferred until later.

6.1. Eigenvalues and Eigenvectors.

We inaugurate our discussion of eigenvalues and eigenvectors with the basic definition.

Definition 6.1. Let A be an $n \times n$ matrix. A scalar λ is called an *eigenvalue* of A if there is a *non-zero* vector $\mathbf{v} \neq \mathbf{0}$, called an *eigenvector*, such that

$$A\mathbf{v} = \lambda \mathbf{v}. \tag{6.1}$$

In other words, the matrix A stretches the eigenvector $\mathbf{v}$ by an amount specified by the eigenvalue λ.

Remark: The odd-looking terms "eigenvalue" and "eigenvector" are hybrid German–English words. In the original German, they are *Eigenwert* and *Eigenvektor*, which can be fully translated as "proper value" and "proper vector". For some reason, the half-translated terms have acquired a certain charm, and are now standard. The alternative English terms *characteristic value* and *characteristic vector* can be found in some (mostly older) texts. Oddly, the term *characteristic equation*, to be defined below, is still used.

The requirement that the eigenvector $\mathbf{v}$ be nonzero is important, since $\mathbf{v} = \mathbf{0}$ is a trivial solution to the eigenvalue equation (6.1) for *any* scalar λ. Moreover, as far as solving linear ordinary differential equations goes, the zero vector $\mathbf{v} = \mathbf{0}$ gives $\mathbf{u}(t) \equiv \mathbf{0}$, which is certainly a solution, but one that we already knew.

The eigenvalue equation (6.1) is a system of linear equations for the entries of the eigenvector $\mathbf{v}$ — provided that the eigenvalue λ is specified in advance — but is "mildly"

nonlinear as a combined system for λ and $\mathbf{v}$. Gaussian Elimination per se will not solve the problem, and we are in need of a new idea. Let us begin by rewriting the equation in the form

$$(A - \lambda\,\mathrm{I})\mathbf{v} = \mathbf{0}, \tag{6.2}$$

where I is the identity matrix of the correct size[†]. Now, for given λ, equation (6.2) is a homogeneous linear system for $\mathbf{v}$, and always has the trivial zero solution $\mathbf{v} = \mathbf{0}$. But we are specifically seeking a nonzero solution! A homogeneous linear system has a nonzero solution $\mathbf{v} \neq \mathbf{0}$ if and only if its coefficient matrix, which in this case is $A - \lambda\,\mathrm{I}$, is singular. This observation is the key to resolving the eigenvector equation.

Theorem 6.2. *A scalar λ is an eigenvalue of the $n \times n$ matrix A if and only if the matrix $A - \lambda\,\mathrm{I}$ is singular, i.e., of rank $< n$. The corresponding eigenvectors are the nonzero solutions to the eigenvalue equation $(A - \lambda\,\mathrm{I})\mathbf{v} = \mathbf{0}$.*

Proposition 6.3. *A scalar λ is an eigenvalue of the matrix A if and only if λ is a solution to the* characteristic equation

$$\det(A - \lambda\,\mathrm{I}) = 0. \tag{6.3}$$

In practice, when finding eigenvalues and eigenvectors by hand, one first solves the characteristic equation (6.3). Then, for each eigenvalue λ one uses standard linear algebra methods, i.e., Gaussian Elimination, to solve the corresponding linear system (6.2) for the eigenvector $\mathbf{v}$.

Example 6.4. Consider the 2×2 matrix

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

We compute the determinant in the characteristic equation using formula (3.8):

$$\det(A - \lambda\,\mathrm{I}) = \det \begin{pmatrix} 3 - \lambda & 1 \\ 1 & 3 - \lambda \end{pmatrix} = (3 - \lambda)^2 - 1 = \lambda^2 - 6\,\lambda + 8.$$

Thus, the characteristic equation is a quadratic polynomial equation, and can be solved by factorization:

$$\lambda^2 - 6\,\lambda + 8 = (\lambda - 4)\,(\lambda - 2) = 0.$$

We conclude that A has two eigenvalues: $\lambda_1 = 4$ and $\lambda_2 = 2$.

For each eigenvalue, the corresponding eigenvectors are found by solving the associated homogeneous linear system (6.2). For the first eigenvalue, the eigenvector equation is

$$(A - 4\,\mathrm{I})\,\mathbf{v} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad \text{or} \qquad \begin{aligned} -x + y &= 0, \\ x - y &= 0. \end{aligned}$$

[†] Note that it is not legal to write (6.2) in the form $(A - \lambda)\mathbf{v} = \mathbf{0}$ since we do not know how to subtract a scalar λ from a matrix A. Worse, if you type $A - \lambda$ in MATLAB or MATHEMATICA, the result will be to subtract λ from *all* the entries of A, which is *not* what we are after!

The general solution is

$$x = y = a, \qquad \text{so} \qquad \mathbf{v} = \begin{pmatrix} a \\ a \end{pmatrix} = a \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

where a is an arbitrary scalar. Only the nonzero solutions[†] count as eigenvectors, and so the eigenvectors for the eigenvalue $\lambda_1 = 4$ must have $a \neq 0$, i.e., they are all nonzero scalar multiples of the basic eigenvector $\mathbf{v}_1 = (1, 1)^T$.

Remark: In general, if $\mathbf{v}$ is an eigenvector of A for the eigenvalue λ, then so is any nonzero scalar multiple of $\mathbf{v}$. In practice, we only distinguish linearly independent eigenvectors. Thus, in this example, we shall say "$\mathbf{v}_1 = (1, 1)^T$ is *the* eigenvector corresponding to the eigenvalue $\lambda_1 = 4$", when we really mean that the eigenvectors for $\lambda_1 = 4$ consist of all nonzero scalar multiples of $\mathbf{v}_1$.

Similarly, for the second eigenvalue $\lambda_2 = 2$, the eigenvector equation is

$$(A - 2\,\mathrm{I})\,\mathbf{v} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

The solution $(-a, a)^T = a\,(-1, 1)^T$ is the set of scalar multiples of the eigenvector $\mathbf{v}_2 = (-1, 1)^T$. Therefore, the complete list of eigenvalues and eigenvectors (up to scalar multiple) for this particular matrix is

$$\lambda_1 = 4, \qquad \mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad\qquad \lambda_2 = 2, \qquad \mathbf{v}_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Example 6.5. Consider the 3×3 matrix

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

Using the formula for a 3×3 determinant, we compute the characteristic equation

$$\begin{aligned}
0 = \det(A - \lambda\,\mathrm{I}) = \det &\begin{pmatrix} -\lambda & -1 & -1 \\ 1 & 2-\lambda & 1 \\ 1 & 1 & 2-\lambda \end{pmatrix} \\
= &(-\lambda)(2 - \lambda)^2 + (-1) \cdot 1 \cdot 1 + (-1) \cdot 1 \cdot 1 - \\
&\quad - 1 \cdot (2 - \lambda)(-1) - 1 \cdot 1 \cdot (-\lambda) - (2 - \lambda) \cdot 1 \cdot (-1) \\
= &-\lambda^3 + 4\lambda^2 - 5\lambda + 2.
\end{aligned}$$

The resulting cubic polynomial can be factored:

$$-\lambda^3 + 4\lambda^2 - 5\lambda + 2 = -(\lambda - 1)^2\,(\lambda - 2) = 0.$$

[†] If, at this stage, you end up with a linear system with only the trivial zero solution, you've done something wrong! Either you don't have a correct eigenvalue — maybe you made a mistake setting up and/or solving the characteristic equation — or you've made an error solving the homogeneous eigenvector system.

Most 3×3 matrices have three different eigenvalues, but this particular one has only two: $\lambda_1 = 1$, which is called a *double eigenvalue* since it is a double root of the characteristic equation, along with a *simple eigenvalue* $\lambda_2 = 2$.

The eigenvector equation (6.2) for the double eigenvalue $\lambda_1 = 1$ is

$$(A - \mathrm{I})\mathbf{v} = \begin{pmatrix} -1 & -1 & -1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

The general solution to this homogeneous linear system

$$\mathbf{v} = \begin{pmatrix} -a - b \\ a \\ b \end{pmatrix} = a \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + b \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

depends upon two free variables: $y = a$ and $z = b$. Any nonzero solution forms a valid eigenvector for the eigenvalue $\lambda_1 = 1$, and so the general eigenvector is any non-zero linear combination of the two "basis eigenvectors" $\mathbf{v}_1 = (-1, 1, 0)^T$, $\widehat{\mathbf{v}}_1 = (-1, 0, 1)^T$.

On the other hand, the eigenvector equation for the simple eigenvalue $\lambda_2 = 2$ is

$$(A - 2\,\mathrm{I})\mathbf{v} = \begin{pmatrix} -2 & -1 & -1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

The general solution

$$\mathbf{v} = \begin{pmatrix} -a \\ a \\ a \end{pmatrix} = a \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$$

consists of all scalar multiples of the eigenvector $\mathbf{v}_2 = (-1, 1, 1)^T$.

In summary, the eigenvalues and (basis) eigenvectors for this matrix are

$$\lambda_1 = 1, \qquad \mathbf{v}_1 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \qquad \widehat{\mathbf{v}}_1 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix},$$
$$\lambda_2 = 2, \qquad \mathbf{v}_2 = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}. \tag{6.4}$$

In general, given a real eigenvalue λ, the corresponding *eigenspace* $V_\lambda \subset \mathbb{R}^n$ is the subspace spanned by all its eigenvectors. Equivalently, the eigenspace is the kernel

$$V_\lambda = \ker(A - \lambda\,\mathrm{I}). \tag{6.5}$$

In particular, $\lambda \in \mathbb{R}$ is an eigenvalue if and only if $V_\lambda \neq \{\mathbf{0}\}$ is a nontrivial subspace, and then every nonzero element of V_λ is a corresponding eigenvector. The most economical way to indicate each eigenspace is by writing out a basis, as in (6.4) with $\mathbf{v}_1, \widehat{\mathbf{v}}_1$ giving a basis for V_1, while $\mathbf{v}_2$ is a basis for V_2.

Example 6.6. The characteristic equation of the matrix $A = \begin{pmatrix} 1 & 2 & 1 \\ 1 & -1 & 1 \\ 2 & 0 & 1 \end{pmatrix}$ is

$$0 = \det(A - \lambda\, \mathrm{I}) = -\lambda^3 + \lambda^2 + 5\lambda + 3 = -(\lambda + 1)^2 (\lambda - 3).$$

Again, there is a double eigenvalue $\lambda_1 = -1$ and a simple eigenvalue $\lambda_2 = 3$. However, in this case the matrix

$$A - \lambda_1\, \mathrm{I} = A + \mathrm{I} = \begin{pmatrix} 2 & 2 & 1 \\ 1 & 0 & 1 \\ 2 & 0 & 2 \end{pmatrix}$$

has only a one-dimensional kernel, spanned by $\mathbf{v}_1 = (\,2, -1, -2\,)^T$. Thus, even though λ_1 is a double eigenvalue, it only admits a one-dimensional eigenspace. The list of eigenvalues and eigenvectors is, in a sense, incomplete:

$$\lambda_1 = -1, \qquad \mathbf{v}_1 = \begin{pmatrix} 2 \\ -1 \\ -2 \end{pmatrix}, \qquad\qquad \lambda_2 = 3, \qquad \mathbf{v}_2 = \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}.$$

Example 6.7. Finally, consider the matrix $A = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & -2 \\ 2 & 2 & -1 \end{pmatrix}$. The characteristic equation is

$$0 = \det(A - \lambda\, \mathrm{I}) = -\lambda^3 + \lambda^2 - 3\lambda - 5 = -(\lambda + 1)(\lambda^2 - 2\lambda + 5).$$

The linear factor yields the eigenvalue -1. The quadratic factor leads to two complex roots, $1 + 2\,\mathrm{i}$ and $1 - 2\,\mathrm{i}$, which can be obtained via the quadratic formula. Hence A has one real and two complex eigenvalues:

$$\lambda_1 = -1, \qquad\qquad \lambda_2 = 1 + 2\,\mathrm{i}, \qquad\qquad \lambda_3 = 1 - 2\,\mathrm{i}.$$

Solving the associated linear system, the real eigenvalue is found to have corresponding eigenvector $\mathbf{v}_1 = (\,-1, 1, 1\,)^T$.

Complex eigenvalues are as important as real eigenvalues, and we need to be able to handle them too. To find the corresponding eigenvectors, which will also be complex, we need to solve the usual eigenvalue equation (6.2), which is now a complex homogeneous linear system. For example, the eigenvector(s) for $\lambda_2 = 1 + 2\,\mathrm{i}$ are found by solving

$$\left[\, A - (1 + 2\,\mathrm{i}\,)\, \mathrm{I} \,\right]\mathbf{v} = \begin{pmatrix} -2\,\mathrm{i} & 2 & 0 \\ 0 & -2\,\mathrm{i} & -2 \\ 2 & 2 & -2 - 2\,\mathrm{i} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

This linear system can be solved by Gaussian Elimination (with complex pivots). A simpler strategy is to work directly: the first equation $-2\,\mathrm{i}\,x + 2y = 0$ tells us that $y = \mathrm{i}\,x$, while the second equation $-2\,\mathrm{i}\,y - 2z = 0$ says $z = -\mathrm{i}\,y = x$. If we trust our calculations so far, we do not need to solve the final equation $2x + 2y + (-2 - 2\,\mathrm{i})z = 0$, since we know that the coefficient matrix is singular and hence this equation must be a consequence of

the first two. (However, it does serve as a useful check on our work.) So, the general solution $\mathbf{v} = (x, i x, x)^T$ is an arbitrary constant multiple of the complex eigenvector $\mathbf{v}_2 = (1, i, 1)^T$. The eigenvector equation for $\lambda_3 = 1 - 2i$ is similarly solved for the third eigenvector $\mathbf{v}_3 = (1, -i, 1)^T$.

Summarizing, the matrix under consideration has three complex eigenvalues and three corresponding eigenvectors, each unique up to (complex) scalar multiple:

$$\lambda_1 = -1, \qquad\qquad \lambda_2 = 1 + 2i, \qquad\qquad \lambda_3 = 1 - 2i,$$

$$\mathbf{v}_1 = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}, \qquad \mathbf{v}_2 = \begin{pmatrix} 1 \\ i \\ 1 \end{pmatrix}, \qquad \mathbf{v}_3 = \begin{pmatrix} 1 \\ -i \\ 1 \end{pmatrix}.$$

Note that the third complex eigenvalue is the complex conjugate of the second, and the eigenvectors are similarly related. This is indicative of a general fact for real matrices:

Proposition 6.8. *If A is a real matrix with a complex eigenvalue $\lambda = \mu + i\nu$ and corresponding complex eigenvector $\mathbf{v} = \mathbf{x} + i\mathbf{y}$, then the complex conjugate $\overline{\lambda} = \mu - i\nu$ is also an eigenvalue with complex conjugate eigenvector $\overline{\mathbf{v}} = \mathbf{x} - i\mathbf{y}$.*

Proof: First take complex conjugates of the eigenvalue equation (6.1):

$$\overline{A}\,\overline{\mathbf{v}} = \overline{A\mathbf{v}} = \overline{\lambda \mathbf{v}} = \overline{\lambda}\,\overline{\mathbf{v}}.$$

Using the fact that a real matrix is unaffected by conjugation, so $\overline{A} = A$, we conclude $A\,\overline{\mathbf{v}} = \overline{\lambda}\,\overline{\mathbf{v}}$, which is the equation for the eigenvalue $\overline{\lambda}$ and eigenvector $\overline{\mathbf{v}}$.   *Q.E.D.*

As a consequence, when dealing with real matrices, we only need to compute the eigenvectors for *one* of each complex conjugate pair of eigenvalues. This observation effectively halves the amount of work in the unfortunate event that we are confronted with complex eigenvalues.

The *eigenspace* associated with a complex eigenvalue λ is the subspace $V_\lambda \subset \mathbb{C}^n$ spanned by the associated eigenvectors. One might also consider complex eigenvectors associated with a real eigenvalue, but this doesn't add anything to the picture — they are merely complex linear combinations of the real eigenvalues. Thus, we only introduce complex eigenvectors when dealing with genuinely complex eigenvalues.

Remark: The reader may recall that we said one should never use determinants in practical computations. So why have we reverted to using determinants to find eigenvalues? The truthful answer is that the practical computation of eigenvalues and eigenvectors *never* resorts to the characteristic equation! The method is fraught with numerical traps and inefficiencies when (a) computing the determinant leading to the characteristic equation, then (b) solving the resulting polynomial equation, which is itself a nontrivial numerical problem[†], [**7**, **47**], and, finally, (c) solving each of the resulting linear eigenvector systems.

[†] In fact, one effective numerical strategy for finding the roots of a polynomial is to turn the procedure on its head, and calculate the eigenvalues of a matrix whose characteristic equation is the polynomial in question! See [**47**] for details.

Worse, if we only know an approximation $\widetilde{\lambda}$ to the true eigenvalue λ, the approximate eigenvector system $(A - \widetilde{\lambda})\mathbf{v} = \mathbf{0}$ will almost certainly have a nonsingular coefficient matrix, and hence only admits the trivial solution $\mathbf{v} = \mathbf{0}$ — which does not even qualify as an eigenvector!

Nevertheless, the characteristic equation does give us important theoretical insight into the structure of the eigenvalues of a matrix, and can be used when dealing with small matrices, e.g., 2×2 and 3×3, presuming exact arithmetic is employed. Numerical algorithms for computing eigenvalues and eigenvectors are based on completely different ideas.

Proposition 6.9. *A matrix A is singular if and only if 0 is an eigenvalue.*

Proof: By definition, 0 is an eigenvalue of A if and only if there is a nonzero solution to the eigenvector equation $A\mathbf{v} = 0\mathbf{v} = \mathbf{0}$. Thus, 0 is an eigenvector of A if and only if it has a non-zero vector in its kernel, $\ker A \neq \{\mathbf{0}\}$, and hence A is necessarily singular.   *Q.E.D.*

Basic Properties of Eigenvalues

If A is an $n \times n$ matrix, then its *characteristic polynomial* is

$$p_A(\lambda) = \det(A - \lambda\,\mathrm{I}) = c_n\,\lambda^n + c_{n-1}\,\lambda^{n-1} + \cdots + c_1\,\lambda + c_0. \tag{6.6}$$

The fact that $p_A(\lambda)$ is a polynomial of degree n is a consequence of the general determinantal formula. Indeed, every term is prescribed by a permutation π of the rows of the matrix, and equals plus or minus a product of n distinct matrix entries including one from each row and one from each column. The term corresponding to the identity permutation is obtained by multiplying the diagonal entries together, which, in this case, is

$$(a_{11}-\lambda)\,(a_{22}-\lambda) \cdots (a_{nn}-\lambda) = (-1)^n\lambda^n + (-1)^{n-1}\big(a_{11} + a_{22} + \cdots + a_{nn}\big)\lambda^{n-1} + \cdots. \tag{6.7}$$

All of the other terms have at most $n-2$ diagonal factors $a_{ii} - \lambda$, and so are polynomials of degree $\leq n-2$ in λ. Thus, (6.7) is the only summand containing the monomials λ^n and λ^{n-1}, and so their respective coefficients are

$$c_n = (-1)^n, \qquad c_{n-1} = (-1)^{n-1}(a_{11} + a_{22} + \cdots + a_{nn}) = (-1)^{n-1}\,\mathrm{tr}\,A, \tag{6.8}$$

where $\mathrm{tr}\,A$, the sum of its diagonal entries, is called the *trace* of the matrix A. The other coefficients $c_{n-2}, \ldots, c_1, c_0$ in (6.6) are more complicated combinations of the entries of A. However, setting $\lambda = 0$ implies

$$p_A(0) = \det A = c_0,$$

and hence the constant term in the characteristic polynomial equals the determinant of the matrix. In particular, if $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is a 2×2 matrix, its characteristic polynomial has the explicit form

$$p_A(\lambda) = \det(A - \lambda\,\mathrm{I}) = \det \begin{pmatrix} a - \lambda & b \\ c & d - \lambda \end{pmatrix}$$
$$= \lambda^2 - (a + d)\lambda + (a\,d - b\,c) = \lambda^2 - (\mathrm{tr}\,A)\lambda + (\det A). \tag{6.9}$$

As a result of these considerations, the characteristic equation of an $n \times n$ matrix A is a polynomial equation of degree n. According to the Fundamental Theorem of Algebra, [**17**], every (complex) polynomial of degree $n \geq 1$ can be completely factored, and so we can write the characteristic polynomial in factored form:

$$p_A(\lambda) = (-1)^n(\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n). \tag{6.10}$$

The complex numbers $\lambda_1, \ldots, \lambda_n$, some of which may be repeated, are the *roots* of the characteristic equation $p_A(\lambda) = 0$, and hence the eigenvalues of the matrix A. Therefore, we immediately conclude:

Theorem 6.10. *An $n \times n$ matrix A has at least one and at most n distinct complex eigenvalues.*

Most $n \times n$ matrices — meaning those for which the characteristic polynomial factors into n *distinct* factors — have *exactly* n complex eigenvalues. More generally, an eigenvalue λ_j is said to have *multiplicity* m if the factor $(\lambda - \lambda_j)$ appears exactly m times in the factorization (6.10) of the characteristic polynomial. An eigenvalue is *simple* if it has multiplicity 1. In particular, A has n distinct eigenvalues if and only if all its eigenvalues are simple. In all cases, when the repeated eigenvalues are counted in accordance with their multiplicity, every $n \times n$ matrix has a total of n, possibly repeated, eigenvalues.

An example of a matrix with just one eigenvalue, of multiplicity n, is the $n \times n$ identity matrix I, whose only eigenvalue is $\lambda = 1$. In this case, *every* nonzero vector in $\mathbb{R}^n$ is an eigenvector of the identity matrix, and so the eigenspace is all of $\mathbb{R}^n$. At the other extreme, the "bidiagonal" *Jordan block matrix*[†]

$$J_a = \begin{pmatrix} a & 1 & & & & \\ & a & 1 & & & \\ & & a & 1 & & \\ & & & \ddots & \ddots & \\ & & & & a & 1 \\ & & & & & a \end{pmatrix} \tag{6.11}$$

also has only one eigenvalue, $\lambda = a$, again of multiplicity n. But in this case, J_a has only one eigenvector (up to scalar multiple), which is the first standard basis vector $\mathbf{e}_1$, and so its eigenspace is one-dimensional.

Remark: If λ is a complex eigenvalue of multiplicity k for the real matrix A, then its complex conjugate $\overline{\lambda}$ also has multiplicity k. This is because complex conjugate roots of a real polynomial necessarily appear with identical multiplicities.

If we explicitly multiply out the factored product (6.10) and equate the result to the characteristic polynomial (6.6), we find that its coefficients $c_0, c_1, \ldots c_{n-1}$ can be written

---

[†] All non-displayed entries are zero.

as certain polynomials of the roots, known as the *elementary symmetric polynomials*. The first and last are of particular importance:

$$c_0 = \lambda_1 \lambda_2 \, \cdots \, \lambda_n, \qquad\qquad c_{n-1} = (-1)^{n-1} (\lambda_1 + \lambda_2 + \, \cdots \, + \lambda_n). \qquad (6.12)$$

Comparison with our previous formulae for the coefficients c_0 and c_{n-1} leads to the following useful result.

Proposition 6.11. *The sum of the eigenvalues of a matrix equals its trace:*

$$\lambda_1 + \lambda_2 + \, \cdots \, + \lambda_n = \operatorname{tr} A = a_{11} + a_{22} + \, \cdots \, + a_{nn}. \qquad (6.13)$$

The product of the eigenvalues equals its determinant:

$$\lambda_1 \lambda_2 \, \cdots \, \lambda_n = \det A. \qquad (6.14)$$

Remark: For repeated eigenvalues, one must add or multiply them in the formulae (6.13–14) according to their multiplicity.

Example 6.12. The matrix $A = \begin{pmatrix} 1 & 2 & 1 \\ 1 & -1 & 1 \\ 2 & 0 & 1 \end{pmatrix}$ considered in Example 6.6 has trace and determinant

$$\operatorname{tr} A = 1, \qquad\qquad \det A = 3,$$

which fix, respectively, the coefficient of λ^2 and the constant term in its characteristic equation. This matrix has two distinct eigenvalues: -1, which is a double eigenvalue, and 3, which is simple. For this particular matrix, formulae (6.13–14) become

$$1 = \operatorname{tr} A = (-1) + (-1) + 3, \qquad\qquad 3 = \det A = (-1)(-1) \, 3.$$

Note that the double eigenvalue contributes twice to the sum and to the product.

6.2. Completeness.

Most of the vector space bases that play a distinguished role in applications are assembled from the eigenvectors of a particular matrix. In this section, we show that the eigenvectors of any "complete" matrix automatically form a basis for $\mathbb{R}^n$ or, in the complex case, $\mathbb{C}^n$. In the following subsection, we use the eigenvector basis to rewrite the linear transformation determined by the matrix in a simple diagonal form. The most important cases — symmetric and positive definite matrices — will be treated in the following section.

The first task is to show that eigenvectors corresponding to distinct eigenvalues are automatically linearly independent.

Lemma 6.13. *If $\lambda_1, \ldots, \lambda_k$ are distinct eigenvalues of the same matrix A, then the corresponding eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are linearly independent.*

Proof: The result is proved by induction on the number of eigenvalues. The case $k = 1$ is immediate since an eigenvector cannot be zero. Assume that we know the result is valid for $k - 1$ eigenvalues. Suppose we have a vanishing linear combination:

$$c_1 \mathbf{v}_1 + \cdots + c_{k-1} \mathbf{v}_{k-1} + c_k \mathbf{v}_k = \mathbf{0}. \tag{6.15}$$

Let us multiply this equation by the matrix A:

$$A\bigl(c_1 \mathbf{v}_1 + \cdots + c_{k-1} \mathbf{v}_{k-1} + c_k \mathbf{v}_k\bigr) = c_1 A\mathbf{v}_1 + \cdots + c_{k-1} A\mathbf{v}_{k-1} + c_k A\mathbf{v}_k$$
$$= c_1 \lambda_1 \mathbf{v}_1 + \cdots + c_{k-1} \lambda_{k-1} \mathbf{v}_{k-1} + c_k \lambda_k \mathbf{v}_k = \mathbf{0}.$$

On the other hand, if we multiply the original equation (6.15) by λ_k, we also have

$$c_1 \lambda_k \mathbf{v}_1 + \cdots + c_{k-1} \lambda_k \mathbf{v}_{k-1} + c_k \lambda_k \mathbf{v}_k = \mathbf{0}.$$

Subtracting this from the previous equation, the final terms cancel and we are left with the equation

$$c_1(\lambda_1 - \lambda_k)\mathbf{v}_1 + \cdots + c_{k-1}(\lambda_{k-1} - \lambda_k)\mathbf{v}_{k-1} = \mathbf{0}.$$

This is a vanishing linear combination of the first $k - 1$ eigenvectors, and so, by our induction hypothesis, can only happen if all the coefficients are zero:

$$c_1(\lambda_1 - \lambda_k) = 0, \qquad \cdots \qquad c_{k-1}(\lambda_{k-1} - \lambda_k) = 0.$$

The eigenvalues were assumed to be distinct, so $\lambda_j \neq \lambda_k$ when $j \neq k$. Consequently, $c_1 = \cdots = c_{k-1} = 0$. Substituting these values back into (6.15), we find $c_k \mathbf{v}_k = \mathbf{0}$, and so $c_k = 0$ also, since the eigenvector $\mathbf{v}_k \neq \mathbf{0}$. Thus we have proved that (6.15) holds if and only if $c_1 = \cdots = c_k = 0$, which implies the linear independence of the eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$. This completes the induction step. *Q.E.D.*

The most important consequence of this result is when a matrix has the maximum allotment of eigenvalues.

Theorem 6.14. *If the $n \times n$ real matrix A has n distinct real eigenvalues $\lambda_1, \ldots, \lambda_n$, then the corresponding real eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ form a basis of $\mathbb{R}^n$. If A (which may now be either a real or a complex matrix) has n distinct complex eigenvalues, then the corresponding eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ form a basis of $\mathbb{C}^n$.*

For instance, the 2×2 matrix in Example 6.4 has two distinct real eigenvalues, and its two independent eigenvectors form a basis of $\mathbb{R}^2$. The 3×3 matrix in Example 6.7 has three distinct complex eigenvalues, and its eigenvectors form a basis for $\mathbb{C}^3$. If a matrix has multiple eigenvalues, then there may or may not be an eigenvector basis of $\mathbb{R}^n$ (or $\mathbb{C}^n$). The matrix in Example 6.5 admits an eigenvector basis, whereas the matrix in Example 6.6 does not. In general, it can be proved that the dimension of the eigenspace is less than or equal to the eigenvalue's multiplicity. In particular, every simple eigenvalue has a one-dimensional eigenspace, and hence, up to scalar multiple, only one associated eigenvector.

Definition 6.15. An eigenvalue λ of a matrix A is called *complete* if the corresponding eigenspace $V_\lambda = \ker(A - \lambda\,\mathrm{I})$ has the same dimension as its multiplicity. The matrix A is *complete* if all its eigenvalues are.

Note that a simple eigenvalue is automatically complete, since its eigenspace is the one-dimensional subspace spanned by the corresponding eigenvector. Thus, only multiple eigenvalues can cause a matrix to be incomplete.

Remark: The multiplicity of an eigenvalue λ_i is sometimes referred to as its *algebraic multiplicity*. The dimension of the eigenspace V_λ is its *geometric multiplicity*, and so completeness requires that the two multiplicities are equal. The word "complete" is not completely standard; other common terms for such matrices are *perfect*, *semi-simple* and, as discussed shortly, *diagonalizable*.

Theorem 6.16. *An $n \times n$ real or complex matrix A is complete if and only if its eigenvectors span $\mathbb{C}^n$. In particular, any $n \times n$ matrix that has n distinct eigenvalues is complete.*

Or, stated another way, a matrix is complete if and only if its eigenvectors can be used to form a basis of $\mathbb{C}^n$. Most matrices are complete. Incomplete $n \times n$ matrices, which have fewer than n linearly independent complex eigenvectors, are considerably less pleasant to deal with.

6.3. Eigenvalues of Symmetric Matrices.

Fortunately, the matrices that arise in most applications are complete and, in fact, possess some additional structure that ameliorates the calculation of their eigenvalues and eigenvectors. The most important class are the symmetric, including positive definite, matrices. In fact, not only are the eigenvalues of a symmetric matrix necessarily real, the eigenvectors always form an *orthogonal basis* of the underlying Euclidean space. In fact, this is by far the most common way for orthogonal bases to appear — as the eigenvector bases of symmetric matrices. Let us state this important result, but defer its proof until the end of the section.

Theorem 6.17. *Let $A = A^T$ be a real symmetric $n \times n$ matrix. Then*
(a) *All the eigenvalues of A are real.*
(b) *Eigenvectors corresponding to distinct eigenvalues are orthogonal.*
(c) *There is an orthonormal basis of $\mathbb{R}^n$ consisting of n eigenvectors of A.*
In particular, all symmetric matrices are complete.

Example 6.18. The 2×2 matrix $A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$ considered in Example 6.4 is symmetric, and so has real eigenvalues $\lambda_1 = 4$ and $\lambda_2 = 2$. You can easily check that the corresponding eigenvectors $\mathbf{v}_1 = (1, 1)^T$ and $\mathbf{v}_2 = (-1, 1)^T$ are orthogonal: $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$, and hence form an orthogonal basis of $\mathbb{R}^2$. The orthonormal eigenvector basis promised by Theorem 6.17 is obtained by dividing each eigenvector by its Euclidean norm:

$$\mathbf{u}_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \qquad \mathbf{u}_2 = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}.$$

Example 6.19. Consider the symmetric matrix $A = \begin{pmatrix} 5 & -4 & 2 \\ -4 & 5 & 2 \\ 2 & 2 & -1 \end{pmatrix}$. A straight-forward computation produces its eigenvalues and eigenvectors:

$$\lambda_1 = 9, \qquad\qquad \lambda_2 = 3, \qquad\qquad \lambda_3 = -3,$$

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \qquad \mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \qquad \mathbf{v}_3 = \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}.$$

As the reader can check, the eigenvectors form an orthogonal basis of $\mathbb{R}^3$. An orthonormal basis is provided by the unit eigenvectors

$$\mathbf{u}_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}, \qquad \mathbf{u}_2 = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}, \qquad \mathbf{u}_3 = \begin{pmatrix} \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{6}} \end{pmatrix}.$$

In particular, the eigenvalues of a symmetric matrix can be used to test its positive definiteness.

Theorem 6.20. *A symmetric matrix $K = K^T$ is positive definite if and only if all of its eigenvalues are strictly positive.*

Example 6.21. Consider the symmetric matrix $K = \begin{pmatrix} 8 & 0 & 1 \\ 0 & 8 & 1 \\ 1 & 1 & 7 \end{pmatrix}$. Its characteristic equation is

$$\det(K - \lambda\,I) = -\lambda^3 + 23\,\lambda^2 - 174\,\lambda + 432 = -(\lambda - 9)(\lambda - 8)(\lambda - 6),$$

and so its eigenvalues are $9, 8$, and 6. Since they are all positive, K is a positive definite matrix. The associated eigenvectors are

$$\lambda_1 = 9, \quad \mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \lambda_2 = 8, \quad \mathbf{v}_2 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \quad \lambda_3 = 6, \quad \mathbf{v}_3 = \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}.$$

Note that the eigenvectors form an orthogonal basis of $\mathbb{R}^3$, as guaranteed by Theorem 6.17. As usual, we can construct an corresponding orthonormal eigenvector basis

$$\mathbf{u}_1 = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}, \qquad \mathbf{u}_2 = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}, \qquad \mathbf{u}_3 = \begin{pmatrix} -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{pmatrix},$$

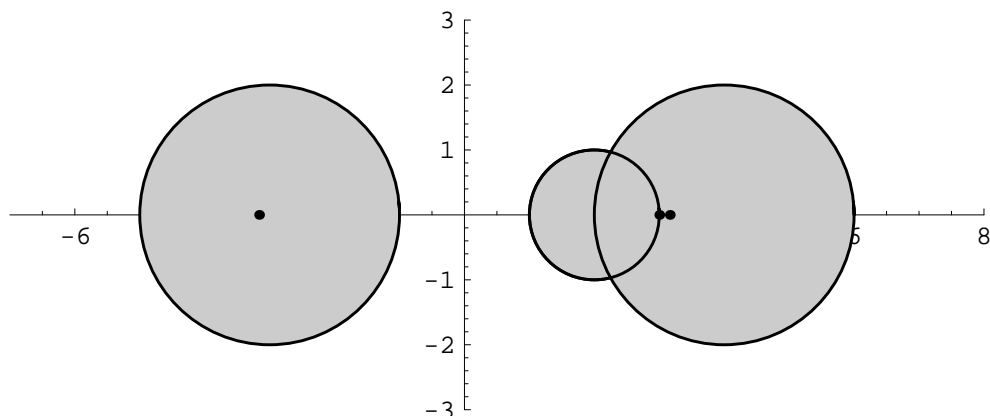by dividing each eigenvector by its norm.

Figure 6.1. Gerschgorin Disks and Eigenvalues.

6.4. The Gerschgorin Circle Theorem.

In general, precisely computing the eigenvalues is not easy, and, in most cases, must be done through a numerical eigenvalue routine. In applications, though, we may not require their exact numerical values, but only approximate locations. The *Gerschgorin Circle Theorem*, due to the early twentieth century Russian mathematician Semen Gerschgorin, serves to restrict the eigenvalues to a certain well-defined region in the complex plane.

Definition 6.22. Let A be an $n \times n$ matrix, either real or complex. For each $1 \leq i \leq n$, define the *Gerschgorin disk*

$$D_i = \{\, |\, z - a_{ii}\,| \leq r_i \mid z \in \mathbb{C} \,\}, \qquad \text{where} \qquad r_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} |\, a_{ij}\,|. \qquad (6.16)$$

The *Gerschgorin domain* $D_A = \bigcup_{i=1}^{n} D_i \subset \mathbb{C}$ is the union of the Gerschgorin disks.

Thus, the ith Gerschgorin disk D_i is centered at the ith diagonal entry a_{ii}, and has radius r_i equal to the sum of the absolute values of the off-diagonal entries that are in the ith row of A. We can now state the Gerschgorin Circle Theorem.

Theorem 6.23. *All real and complex eigenvalues of the matrix A lie in its Gerschgorin domain D_A.*

Example 6.24. The matrix $A = \begin{pmatrix} 2 & -1 & 0 \\ 1 & 4 & -1 \\ -1 & -1 & -3 \end{pmatrix}$ has Gerschgorin disks

$$D_1 = \{\, |\, z - 2\,| \leq 1 \,\}, \qquad D_2 = \{\, |\, z - 4\,| \leq 2 \,\}, \qquad D_3 = \{\, |\, z + 3\,| \leq 2 \,\},$$

which are plotted in Figure 6.1. The eigenvalues of A are

$$\lambda_1 = 3, \qquad \lambda_2 = \sqrt{10} = 3.1623\ldots, \qquad \lambda_3 = -\sqrt{10} = -3.1623\ldots.$$

Observe that λ_1 belongs to both D_1 and D_2, while λ_2 lies in D_2, and λ_3 is in D_3. We thus confirm that all three eigenvalues are in the Gerschgorin domain $D_A = D_1 \cup D_2 \cup D_3$.

Proof of Theorem 6.23: Let $\mathbf{v}$ be an eigenvector of A with eigenvalue λ. Let $\mathbf{u} = \mathbf{v}/\|\mathbf{v}\|_\infty$ be the corresponding unit eigenvector with respect to the ∞ norm, so

$$\|\mathbf{u}\|_\infty = \max\{\,|u_1|,\ \ldots\ ,|u_n|\,\} = 1.$$

Let u_i be an entry of $\mathbf{u}$ that achieves the maximum: $|u_i| = 1$. Writing out the ith component of the eigenvalue equation $A\mathbf{u} = \lambda\mathbf{u}$, we find

$$\sum_{j=1}^n a_{ij}\,u_j = \lambda\,u_i, \qquad \text{which we rewrite as} \qquad \sum_{\substack{j=1\\j\neq i}}^n a_{ij}\,u_j = (\lambda - a_{ii})\,u_i.$$

Therefore, since all $|u_j| \le 1$ while $|u_i| = 1$,

$$|\lambda - a_{ii}| = |\lambda - a_{ii}|\,|u_i| = \left|\ \sum_{j\neq i} a_{ij}\,u_j\ \right| \le \sum_{j\neq i}|a_{ij}|\,|u_j| \le \sum_{j\neq i}|a_{ij}| = r_i.$$

This immediately implies that $\lambda \in D_i \subset D_A$ belongs to the ith Gerschgorin disk. *Q.E.D.*

One application is a simple direct test that guarantees invertibility of a matrix without requiring Gaussian Elimination or computing determinants. According to Proposition 6.9, a matrix A is nonsingular if and only if it does not admit zero as an eigenvalue. Thus, if its Gerschgorin domain does not contain 0, it cannot be an eigenvalue, and hence A is necessarily invertible. The condition $0 \notin D_A$ requires that the matrix have large diagonal entries, as quantified by the following definition.

Definition 6.25. A square matrix A is called *strictly diagonally dominant* if

$$|a_{ii}| > \sum_{\substack{j=1\\j\neq i}}^n |a_{ij}|, \qquad \text{for all} \qquad i = 1,\ldots,n. \tag{6.17}$$

In other words, strict diagonal dominance requires each diagonal entry to be larger, in absolute value, than the sum of the absolute values of *all* the other entries in its row. For example, the matrix $\begin{pmatrix} 3 & -1 & 1 \\ 1 & -4 & 2 \\ -2 & -1 & 5 \end{pmatrix}$ is strictly diagonally dominant since

$$|3| > |-1| + |1|, \qquad |-4| > |1| + |2|, \qquad |5| > |-2| + |-1|.$$

Diagonally dominant matrices appear frequently in numerical solution methods for both ordinary and partial differential equations. As we shall see, they are the most common class of matrices to which iterative solution methods can be successfully applied.

Theorem 6.26. *A strictly diagonally dominant matrix is nonsingular.*

Proof: The diagonal dominance inequalities (6.17) imply that the radius of the ith Gerschgorin disk is strictly less than the modulus of its center: $r_i < |a_{ii}|$. Thus, the disk cannot contain 0; indeed, if $z \in D_i$, then, by the triangle inequality,

$$r_i > |z - a_{ii}| \ge |a_{ii}| - |z| > r_i - |z|, \qquad \text{and hence} \qquad |z| > 0.$$

Thus, $0 \notin D_A$ does not lie in the Gerschgorin domain and so cannot be an eigenvalue.
$$\textit{Q.E.D.}$$

Warning: The converse to this result is obviously not true; there are plenty of non-singular matrices that are not diagonally dominant.

6.5. Singular Values.

We have already indicated the central role played by the eigenvalues and eigenvectors of a square matrix in both theory and applications. Much more evidence to this effect will appear in the ensuing chapters. Alas, rectangular matrices do not have eigenvalues (why?), and so, at first glance, do not appear to possess any quantities of comparable significance. But you no doubt recall that our earlier treatment of least squares minimization problems, as well as the equilibrium equations for structures and circuits, made essential use of the symmetric, positive semi-definite *square* Gram matrix $K = A^T A$ — which can be naturally formed even when A is not square. Perhaps the eigenvalues of K might play a comparably important role for general matrices. Since they are not easily related to the eigenvalues of A — which, in the non-square case, don't even exist — we shall endow them with a new name.

Definition 6.27. The *singular values* $\sigma_1, \ldots, \sigma_r$ of an $m \times n$ matrix A are the positive square roots, $\sigma_i = \sqrt{\lambda_i} > 0$, of the nonzero eigenvalues of the associated Gram matrix $K = A^T A$. The corresponding eigenvectors of K are known as the *singular vectors* of A.

Since K is necessarily positive semi-definite, its eigenvalues are always non-negative, $\lambda_i \geq 0$, which justifies the positivity of the singular values of A — independently of whether A itself has positive, negative, or even complex eigenvalues — or is rectangular and has no eigenvalues at all. The standard convention is to label the singular values in decreasing order, so that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$. Thus, σ_1 will always denote the largest or *dominant* singular value. If $K = A^T A$ has repeated eigenvalues, the singular values of A are repeated with the same multiplicities. As we will see, the number r of singular values is always equal to the rank of the matrix.

Warning: Many texts include the zero eigenvalues of K as singular values of A. We find this to be somewhat less convenient, but you should be aware of the differences in the two conventions.

Example 6.28. Let $A = \begin{pmatrix} 3 & 5 \\ 4 & 0 \end{pmatrix}$. The associated Gram matrix $K = A^T A = \begin{pmatrix} 25 & 15 \\ 15 & 25 \end{pmatrix}$ has eigenvalues $\lambda_1 = 40$, $\lambda_2 = 10$, and corresponding eigenvectors $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $\mathbf{v}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Thus, the singular values of A are $\sigma_1 = \sqrt{40} \approx 6.3246$ and $\sigma_2 = \sqrt{10} \approx 3.1623$, with $\mathbf{v}_1, \mathbf{v}_2$ being the singular vectors. Note that the singular values are *not* the same as its eigenvalues, which are $\lambda_1 = \frac{1}{2}(3 + \sqrt{89}) \approx 6.2170$ and $\lambda_2 = \frac{1}{2}(3 - \sqrt{89}) \approx -3.2170$ — nor are the singular vectors eigenvectors of A.

Only in the special case of symmetric matrices is there a direct connection between the singular values and the eigenvalues.

Proposition 6.29. *If $A = A^T$ is a symmetric matrix, its singular values are the absolute values of its nonzero eigenvalues: $\sigma_i = |\lambda_i| > 0$; its singular vectors coincide with the associated non-null eigenvectors.*

Proof: When A is symmetric, $K = A^T A = A^2$. So, if $A\mathbf{v} = \lambda\mathbf{v}$, then $K\mathbf{v} = A^2\mathbf{v} = \lambda^2\mathbf{v}$. Thus, every eigenvector $\mathbf{v}$ of A is also an eigenvector of K with eigenvalue λ^2. Therefore, the eigenvector basis of A is also an eigenvector basis for K, and hence also forms a complete system of singular vectors for A. $\hspace{2cm}$ *Q.E.D.*

Condition Number, Rank, and Principal Component Analysis

The singular values not only provide a pretty geometric interpretation of the action of the matrix, they also play a key role in modern computational algorithms. The relative magnitudes of the singular values can be used to distinguish well-behaved linear systems from ill-conditioned systems which are much trickier to solve accurately. Since the number of singular values equals the matrix's rank, an $n \times n$ matrix with fewer than n singular values is singular. For the same reason, a square matrix with one or more very small singular values should be considered to be close to singular. The potential difficulty of accurately solving a linear algebraic system with coefficient matrix A is traditionally quantified as follows.

Definition 6.30. The *condition number* of a nonsingular $n \times n$ matrix is the ratio between its largest and smallest singular value: $\kappa(A) = \sigma_1/\sigma_n$.

If A is singular, it is said to have condition number ∞. A matrix with a very large condition number is said to be *ill-conditioned*; in practice, this occurs when the condition number is larger than the reciprocal of the machine's precision, e.g., 10^7 for typical single precision arithmetic. As the name implies, it is much harder to solve a linear system $A\mathbf{x} = \mathbf{b}$ when its coefficient matrix is ill-conditioned.

Determining the rank of a large matrix can be a numerical challenge. Small numerical errors in the entries can have an unpredictable effect. For example, the matrix $A = \begin{pmatrix} 1 & 1 & -1 \\ 2 & 2 & -2 \\ 3 & 3 & -3 \end{pmatrix}$ has rank $r = 1$, but a tiny change, e.g., $\widetilde{A} = \begin{pmatrix} 1.00001 & 1. & -1. \\ 2. & 2.00001 & -2. \\ 3. & 3. & -3.00001 \end{pmatrix}$, will produce a nonsingular matrix with rank $r = 3$. The latter matrix, however, is very close to singular, and this is highlighted by its singular values, which are $\sigma_1 \approx 6.48075$ while $\sigma_2 \approx \sigma_3 \approx .000001$. The fact that the second and third singular values are very small indicates that $\widetilde{A}$ is very close to a matrix of rank 1 and should be viewed as a numerical (or experimental) perturbation of such a matrix. Thus, an effective practical method for computing the rank of a matrix is to first assign a threshold, e.g., 10^{-5}, for singular values, and then treat any small singular value lying below the threshold as if it were zero.

This idea underlies the method of *Principal Component Analysis* that is assuming an increasingly visible role in modern statistics, data mining, imaging, speech recognition,

semantics, and a variety of other fields, [**30**]. The singular vectors associated with the larger singular values indicate the *principal components* of the matrix, while small singular values indicate relatively unimportant directions. In applications, the columns of the matrix A represent the data vectors, which are normalized to have mean **0**. The corresponding Gram matrix $K = A^T A$ can be identified as the associated *covariance matrix*, [**12**]. Its eigenvectors are the principal components that serve to indicate directions of correlation and clustering in the data.