

1 · Eigenvalues

Eigenvalues are among the most successful tools of applied mathematics. Here are some of the fields where they are important, with a representative citation from each.

acoustics [563]	chemistry [722]
control theory [443]	earthquake engineering [151]
ecology [130]	economics [739]
fluid mechanics [669]	functional analysis [630]
helioseismology [331]	magnetohydrodynamics [135]
Markov chains [582]	matrix iterations [338]
partial differential equations [178]	physics of music [279]
quantum mechanics [666]	spectroscopy [349]
structural analysis [154]	vibration analysis [376]
numerical solution of differential equations [639]	

Figures 1.1 and 1.2 present images of eigenvalues in two quite different applications.

In the simplest context of matrices, the definitions are as follows. Let \mathbf{A} be an $N \times N$ matrix with real or complex coefficients; we write $\mathbf{A} \in \mathbb{C}^{N \times N}$. Let \mathbf{v} be a nonzero real or complex column vector of length N , and let λ be a real or complex scalar; we write $\mathbf{v} \in \mathbb{C}^N$ and $\lambda \in \mathbb{C}$. Then \mathbf{v} is an *eigenvector* of \mathbf{A} , and $\lambda \in \mathbb{C}$ is its corresponding *eigenvalue*, if

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}. \quad (1.1)$$

(Even if \mathbf{A} is real, its eigenvalues are in general complex unless \mathbf{A} is self-adjoint.) The set of all the eigenvalues of \mathbf{A} is the *spectrum* of \mathbf{A} , a nonempty subset of the complex plane \mathbb{C} that we denote by $\sigma(\mathbf{A})$. The spectrum can also be defined as the set of points $z \in \mathbb{C}$ where the *resolvent* matrix,

$$(z - \mathbf{A})^{-1},$$

does not exist. Throughout this book, $z - \mathbf{A}$ is shorthand for $z\mathbf{I} - \mathbf{A}$, where \mathbf{I} is the identity.

Unlike singular values [414, 776], eigenvalues conventionally make sense only for a matrix that is square. This reflects the fact that in applications, they are generally used where a matrix is to be compounded iteratively, for example, as a power \mathbf{A}^k or an exponential $e^{t\mathbf{A}} = \mathbf{I} + t\mathbf{A} + \frac{1}{2}(t\mathbf{A})^2 + \cdots$.

For most matrices \mathbf{A} , there exists a *complete set of eigenvectors*, a set of N linearly independent vectors $\mathbf{v}_1, \dots, \mathbf{v}_N$ with $\mathbf{A}\mathbf{v}_j = \lambda_j\mathbf{v}_j$. If \mathbf{A} has N distinct eigenvalues, then it is guaranteed to have a complete set of



Figure 1.1: Spectroscopic image of light from the sun. The black ‘Fraunhofer lines’ correspond to various differences of eigenvalues of the Schrödinger operator for atoms such as H, Fe, Ca, Na, and Mg that are present in the solar atmosphere. Light at these frequencies resonates with frequencies of the transitions between energy states in these atoms and is absorbed. Spectroscopic measurements such as these are a crucial tool in chemical analysis, not only of astronomical bodies, and by making possible the measurement of redshifts of distant galaxies, they led to the discovery of the expanding universe. Original image courtesy of the Observatories of the Carnegie Institution of Washington.

eigenvectors, and they are unique up to normalization by scalar factors. For any matrix \mathbf{A} with a complete set of eigenvectors $\{\mathbf{v}_j\}$, let \mathbf{V} be the $N \times N$ matrix whose j th column is \mathbf{v}_j , a *matrix of eigenvectors*. Then we can write all N eigenvalue conditions at once by the matrix equation

$$\mathbf{A}\mathbf{V} = \mathbf{V}\mathbf{\Lambda}, \quad (1.2)$$

where $\mathbf{\Lambda}$ is the diagonal $N \times N$ matrix whose j th diagonal entry is λ_j . Pictorially,

$$\mathbf{A} \left[\begin{array}{c|c|c|c|c} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_N & \end{array} \right] = \left[\begin{array}{c|c|c|c|c} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_N & \end{array} \right] \begin{pmatrix} \lambda_1 & & & & \\ & \lambda_2 & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & \lambda_N \end{pmatrix}.$$

Since the eigenvectors \mathbf{v}_j are linearly independent, \mathbf{V} is nonsingular, and thus we can multiply (1.2) on the right by \mathbf{V}^{-1} to obtain the factorization

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}, \quad (1.3)$$

known as an *eigenvalue decomposition* or a *diagonalization* of \mathbf{A} . In view of this formula, a matrix with a complete set of eigenvectors is said to be *diagonalizable*. An equivalent term is *nondefective*.

The eigenvalue decomposition expresses a change of basis to ‘eigenvector coordinates’, i.e., coefficients in an expansion in eigenvectors. If $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$, for example, then we have

$$\mathbf{V}^{-1}(\mathbf{A}^k \mathbf{x}) = \mathbf{V}^{-1}(\mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1})^k \mathbf{x} = \mathbf{\Lambda}^k (\mathbf{V}^{-1} \mathbf{x}). \quad (1.4)$$

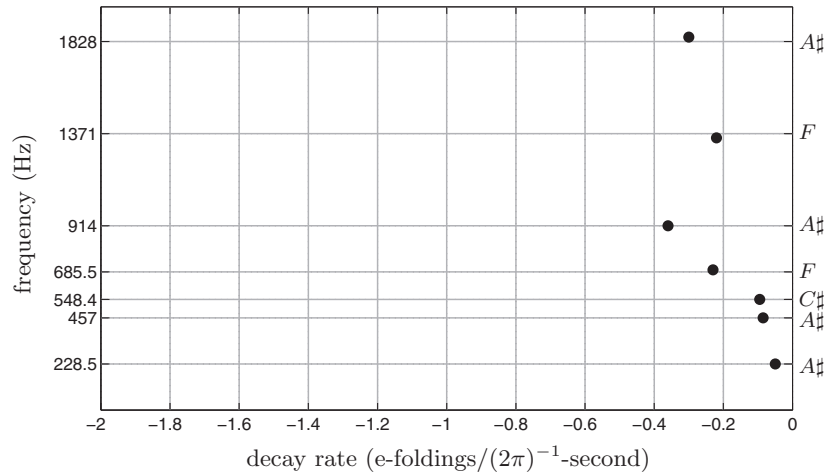


Figure 1.2: Measured eigenvalues in the complex plane of a minor third $A_4\sharp$ carillon bell (figure from [418] based on data from [696]). The grid lines show the positions of the frequencies corresponding to a minor third chord at 456.8 Hz, together with two octaves above the fundamental and one below. Immediately after the bell is struck, the ear hears all seven of the frequencies portrayed; a little later, the higher four have decayed and mostly the lowest three are heard; still later, the lowest mode, the ‘hum’, dominates. The simple rational relationships among these frequencies would not hold for arbitrarily shaped bells, but are the result of generations of evolution in bell shapes to achieve a pleasing effect.

Now the product $\mathbf{V}^{-1}(\mathbf{A}^k \mathbf{x})$ is equal to the vector \mathbf{c} of coefficients in an expansion $\mathbf{A}^k \mathbf{x} = \mathbf{V} \mathbf{c} = \sum c_j \mathbf{v}_j$ of $\mathbf{A}^k \mathbf{x}$ as a linear combination of the eigenvectors $\{\mathbf{v}_j\}$, and similarly, $\mathbf{V}^{-1} \mathbf{x}$ is the vector of coefficients in an expansion of \mathbf{x} . Thus, (1.4) asserts that to compute $\mathbf{A}^k \mathbf{x}$, we can expand \mathbf{x} in the basis of eigenvectors, apply the diagonal matrix $\mathbf{\Lambda}^k$, and interpret the result as the coefficients for another expansion in the basis of eigenvectors. In other words, the change of basis has rendered the problem diagonal and hence trivial. For $e^{t\mathbf{A}} \mathbf{x}$, similarly, we have

$$\mathbf{V}^{-1}(e^{t\mathbf{A}} \mathbf{x}) = \mathbf{V}^{-1}(\mathbf{V} e^{t\mathbf{\Lambda}} \mathbf{V}^{-1}) \mathbf{x} = e^{t\mathbf{\Lambda}} (\mathbf{V}^{-1} \mathbf{x}), \quad (1.5)$$

so diagonalization makes this problem trivial too, and likewise for other functions $f(\mathbf{A})$.

So far we have taken \mathbf{A} to be a matrix, but eigenvalues are also important when \mathbf{A} is a more general linear operator such as an infinite matrix, a differential operator, or an integral operator. Indeed, eigenvalue problems for matrices often come about through discretization of linear operators. The *spectrum* $\sigma(\mathbf{A})$ of a closed operator \mathbf{A} defined in a Banach

(or Hilbert) space is defined as the set of numbers $z \in \mathbb{C}$ for which the resolvent $(z - \mathbf{A})^{-1}$ does not exist as a bounded operator defined on the whole space (§4). It can be any closed set in the complex plane, including the empty set. Eigenvalues and eigenvectors (also called *eigenfunctions* or *eigenmodes*) are still defined by (1.1), but among the new features that arise in the operator case is the phenomenon that not every $z \in \sigma(\mathbf{A})$ is necessarily an eigenvalue. This book avoids fine points of spectral theory wherever possible, for the main issues to be investigated are orthogonal to the differences between matrices and operators. In particular, the distinction between spectra and pseudospectra has little to do with the distinction between point and continuous spectra. In certain contexts, of course, it will be necessary for us to be more precise.

This book is about the limitations of eigenvalues, and alternatives to them. In the remainder of this introductory section, let us accordingly consider the question, What are eigenvalues useful for? Why are eigenvalues and eigenfunctions—more generally, spectra and spectral theory—among the standard tools of applied mathematics? Various answers to these questions appear throughout this volume, but here, we shall make our best attempt to summarize them in a systematic way.

We begin with a one-paragraph history [96, 208, 721]. It is not too great an oversimplification to say that a major part of eigenvalue analysis originated early in the nineteenth century with Fourier's solution of the heat equation by series expansions. Fourier's ideas were extended by Poisson, and other highlights of the nineteenth century include Sturm and Liouville's treatment of more general second-order differential equations in the 1830s; Sylvester and Cayley's diagonalization of symmetric matrices in the 1850s (the origins of this idea go back to Cauchy, Jacobi, Lagrange, Euler, Fermat, and Descartes); Weber and Schwarz's treatment of a vibrating membrane in 1869 and 1885 (whose origins in vibrating strings go back to D. Bernoulli, Euler, d'Alembert, . . . , Pythagoras); Lord Rayleigh's treatise *The Theory of Sound* in 1877 [618]; and further developments by Poincaré around 1890. By 1900, eigenvalues and eigenfunction expansions were well-known, especially in the context of differential equations. The new century brought the mathematical theory of linear operators due to Fredholm, Hilbert, Schmidt, von Neumann, and others; the terms 'eigenvalue' and 'spectral theory' appear to have been coined by Hilbert. The influential book by Courant and Hilbert, first published in 1924, surveyed a large amount of material concerning eigenvalues of differential equations and vibration problems [168]. Just two years later came the explosive ideas of quantum mechanics, which in a short time, in the hands of Heisenberg, Jordan, Schrödinger, Dirac, and others, moved matrices and operators to center stage of the scientific world. Quantum 'matrix mechanics' revealed that energy states of atoms and molecules could be viewed as eigenfunctions of a Schrödinger operator, thereby explaining Figure 1.1, the periodic ta-

ble of the elements, and countless other scientific observations besides [384], and from that time on, every mathematical scientist has known the basics of matrices, operators, eigenvalues, and eigenfunctions.

What exactly do eigenvalues offer that makes them useful for so many problems? We believe there are three principal answers to this question, more than one of which may be important in a particular application.

1. *Diagonalization and separation of variables: use of the eigenfunctions as a basis.* One thing eigenvalues may accomplish is the decoupling, as in (1.3)–(1.5), of a problem involving vectors or functions into a collection of problems involving scalars, which may make subsequent computations easier. For example, in Fourier’s problem of heat conduction in a solid bar with zero temperature at both ends, the eigenmodes are sine waves that decay independently as a function of time. If an arbitrary initial temperature distribution is expanded as a sum of these sine waves, then the solution at a later time can be calculated by summing the components of the expansion.

2. *Resonance: heightened response to selected inputs.* Diagonalization is an algorithmic idea; the other uses of eigenvalues are more physical. One is the analysis of the phenomenon of resonance, perhaps most familiar in the context of vibrating strings, drums, and mechanical structures. Any visitor to science museums has seen demonstrations showing that certain systems respond preferentially to vibrations at special frequencies. These frequencies are the eigenvalues of the linear or linearized operator that governs the system in question, and the form of the response is associated with the corresponding eigenfunctions. Examples of resonance are familiar: One thinks of soldiers breaking step as they cross bridges; of the less fortunate Tacoma Narrows Bridge in the 1940s, whose collapse was initiated by a wind-induced flow oscillation too close to a structural eigenfrequency; of buildings and their response to the vibrations of earthquakes—an application where eigenvalues are written into legal codes; of that old cartoon standby, the soprano whose high E shatters windows. In other examples resonance is desired rather than feared: examples include AM radio, where the signal from a far-off station is selected from a sea of background noise by a finely tuned resonant circuit, and the cochlea of the human ear, whose basilar membrane resonates preferentially in different locations according to the frequency of the sound input and thus in a sense tunes in all stations at once. These last two examples illustrate the wide range of complexity in applications of eigenvalue ideas, for the radio problem is straightforward and almost perfectly linear, whereas the ear is a complicated nonlinear system, not yet fully understood, for which eigenmodes are only a crude first step.

3. *Asymptotics and stability: dominant response to general inputs.* A related application of eigenvalues is to questions of the form, What will

happen as time elapses (or in the extreme, $t \rightarrow \infty$) to a system that has experienced some more or less random disturbance? Fourier's heat problem again affords an example: Whatever the shape of the initial temperature distribution, the higher sine waves decay faster than the lowest one, and therefore almost any initial distribution will eventually come to look like the half-wavelength sine with zeros just at the two ends of the interval. Similarly, what makes a church bell as in Figure 1.2 chime musically? As the clapper strikes, all frequencies are excited, but differential decay rates soon filter out all but a few dominant ones, and the result is a pleasing sound. Kettledrums operate on the same principle, as do Markov chains in probability theory. Sometimes the crucial issue is a question of stability: Are there modes that grow rather than decay with t ? For example, in fluid mechanics a standard technique to determine whether small perturbations to a laminar flow will be amplified into large ones—which may then trigger the onset of turbulence—is to calculate whether the eigenvalues of the system all lie in the left half of the complex plane. (We shall see in §20 that this technique is not always successful.) Similar questions arise in control theory and in numerical analysis, where time is discrete and stability depends on eigenvalues being less than 1 in modulus. Problems of convergence of matrix iterations in numerical analysis are also related, the convergence rate being determined by how close certain eigenvalues are to zero.

Principles 1, 2, and 3 account for most applications of eigenvalues. (Sometimes the latter two are hard to distinguish, as, for example, in the operation of bowed or blown musical instruments. The significance of eigenvalues in quantum mechanics also may have special features, not well captured by 1–3.) In view of the ubiquity of vibrations, oscillations, and linear or approximately linear processes in the physical world, they amply justify the great attention that has been given to eigenvalues over the years.

And we think there is a fourth reason, too, for the success of eigenvalues.

4. *They give a matrix a personality.* We humans like images; our brains are specially adapted to interpret them. Eigenvalues enable us to take the abstraction of a matrix or linear operator, for whose analysis we possess no hardwired talent, and portray it as a picture.

This book is about a class of problems for which eigenvalue methods may fail: problems involving matrices or operators for which the matrix \mathbf{V}^{-1} of (1.3)–(1.5), if it exists, contains very large entries:

$$\|\mathbf{V}^{-1}\| \gg 1. \tag{1.6}$$

(This often turns out to mean exponentially large with respect to a parameter.) This formulation of the matter assumes that the matrix \mathbf{V} itself is in

some sense reasonably scaled, with $\|\mathbf{V}\|$ roughly of order 1. If no assumptions are made about the scaling of $\|\mathbf{V}\|$, then (1.6) should be replaced by a statement about the *condition number* of \mathbf{V} in the norm $\|\cdot\|$,

$$\|\mathbf{V}\|\|\mathbf{V}^{-1}\| \gg 1, \quad (1.7)$$

and to be still more precise we should require that (1.7) hold not just for some eigenvector matrix \mathbf{V} , whose eigenvector columns might be badly scaled relative to one another, but for any eigenvector matrix \mathbf{V} . For operators as opposed to matrices, a suitable generalization of (1.7) can be applied in some cases, but not all.

The conditions (1.6) and (1.7) depend upon the choice of norm $\|\cdot\|$. Though sometimes it is essential to consider other possibilities (see, e.g., §56 and §57), most of our examples will be based on the use of the 2-norm, defined by $\|\mathbf{x}\|_2 = (\sum |x_j|^2)^{1/2}$ for a vector \mathbf{x} and then by

$$\|\mathbf{A}\|_2 = \max_{\mathbf{x}} \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \quad (1.8)$$

for a matrix \mathbf{A} . This choice of norm corresponds mathematically to formulation in a Hilbert space and physically to consideration of energy defined by a sum of squares, and in this important special case, (1.7) amounts to the condition that the eigenvectors of \mathbf{A} , if they exist, are far from orthogonal. At the other extreme is a *normal* matrix, one that has a complete set of orthogonal eigenvectors; real symmetric and Hermitian matrices fall in this category. In this case, if each \mathbf{v}_j is normalized by $\|\mathbf{v}_j\|_2 = 1$, then \mathbf{V} is a *unitary* matrix (in the real case we say *orthogonal*), with $\mathbf{V}^{-1} = \mathbf{V}^*$ (\mathbf{V}^* denotes the conjugate transpose) and $\|\mathbf{V}\|_2 = \|\mathbf{V}^{-1}\|_2 = 1$. Thus for $\|\cdot\| = \|\cdot\|_2$, (1.7) is a statement that \mathbf{A} is in some sense far from normal. In this norm, it is the nonnormal matrices for which eigenvalue analysis may fail, and in this book, starting with the subtitle on the cover, we often speak of problems that are ‘nonnormal’ or ‘far from normal’ when a more careful statement would refer to a more general condition, such as (1.7).

The majority of the familiar applications of eigenvalue analysis involve matrices or operators that are normal or close to normal, having eigenfunctions orthogonal or nearly so. Among the examples mentioned so far, all of the physical ones are in this category except certain problems of fluid mechanics. The familiar mechanical oscillations are governed by normal operators, for example, and so are the oscillations of quantum mechanics, at least in their standard formulation. As a consequence, our intuition about eigenvalues has been formed by the normal case. Two centuries of successes have generated confidence that the eigenvalue idea is both powerful in practice and fundamental in concept. It has not always been noted that as most of these successes involve problems governed by normal or near-normal operators, our grounds for confidence in the nonnormal case are less solid.

With this in mind, we shall now briefly indicate what can go wrong with 1, 2, and 3 in certain applications.

First, consider 2. If a linear operator is normal, then the degree of resonant amplification that may occur in response to an input at frequency ω is equal to the inverse of the distance in the complex plane between ω and the nearest eigenvalue. (This formula can be found in first-year physics textbooks, usually without the word ‘eigenvalue’.) For a nonnormal operator, however, the resonant amplification may be orders of magnitude greater. *The resonances of a nonnormal system are not determined by the eigenvalues alone.* This phenomenon is at the heart of the topic known as ‘receptivity’ in fluid mechanics (§23).

Next, consider 3. It is true that for a purely linear, constant-coefficient, homogeneous problem, eigenvalues govern the asymptotic behavior as $t \rightarrow \infty$. If the problem is normal, this statement is robust; the eigenvalues also have relevance to short-time or transient behavior, and moreover, their influence tends to persist if the problem is altered in small ways. If the problem is far from normal, however, conclusions based on eigenvalues are in general not robust. First, there may be a long transient that looks quite different from the asymptote and has no connection to the eigenvalues. Second, even the asymptote may change beyond recognition if the problem is modified slightly. *Eigenvalues do not always govern the transient behavior of a nonnormal system, nor the asymptotic behavior in the presence of nonlinear terms, variable coefficients, lower order terms, inhomogeneous forcing data, or other complications.* Few applied problems are free of all these effects. For those that are, it is rare that one is interested so purely in the limit $t \rightarrow \infty$ as one may at first imagine. These issues are at the heart of convergence and stability investigations in numerical analysis, and we discuss them, for example, in Parts VI and VII. For a high-level schema, see Figure 33.3.

This brings us to 1. Unlike 2 and 3, the algorithmic idea of diagonalization is not in general invalidated if $\|\mathbf{V}\| \|\mathbf{V}^{-1}\|$ is large (although in extreme cases there may be difficulties caused by rounding errors on a computer). On the other hand, there is a different difficulty that sometimes makes diagonalization less useful than one might expect, even for normal problems. In practice, for differential or other operators one works with truncated expansions; an infinite series is approximated by finite sum. The difficulty that arises sometimes is that the choice of the basis of eigenfunctions for such an expansion may necessitate taking an unacceptably large number of terms in the expansion to achieve the required accuracy. *Eigenfunction expansions may be exceedingly inefficient.* This fact was publicized by Orszag around 1970 in the context of spectral methods for the numerical solution of differential equations [588, 775]. Spectral methods, by contrast, are based on expansions in functions that have nothing to do with the eigenfunctions of the problem at hand, but which may converge ge-

ometrically, where an expansion in eigenfunctions converges only linearly. Thirty Chebyshev polynomials may resolve a problem as well as a thousand eigenfunctions. An example is considered in §59.

What about 4, a matrix or operator's personality? In the highly non-normal case, vivid though the image may be, the location of the eigenvalues may be as fragile an indicator of underlying character as the hair color of a Hollywood actor. We shall see that pseudospectra provide equally compelling images that may capture the spirit underneath more robustly.

In summary, eigenvalues and eigenfunctions have a distinguished history of application throughout the mathematical sciences; we could not get along without them. Their clearest successes, however, are associated with problems that involve well-behaved systems of eigenvectors, which in most contexts means matrices or operators that are normal or nearly so. This class of problems encompasses the majority of applications, but not all. For nonnormal problems, the record is less clear, and even the conceptual significance of eigenvalues is open to question.