12 Non-diagonalizable matrices: What can matter beyond eigenvalues and eigenvectors

12.1 Preliminaries

In the previous two Lectures and also in Lecture 5, we highlighted the important roles played by the eigenvalues and eigenvectors of a matrix in various applications. In this Lecture we will have a brief encounter with matrices for which eigenvalues and eigenvectors do *not* carry all the relevant information about the underlying problem. These are ***non-diagonalizable*** matrices, for which the number of linearly independent eigenvectors is less than the dimension of the matrix. In such a case, the eigenvectors alone cannot form a basis in the corresponding vector space and hence cannot fully describe the behavior of the solution. One requires so-called ***generalized eigenvectors*** to complete the set of the eigenvectors to a basis. In certain situations, it is the behavior of those generalized eigenvectors rather than that of "regular" eigenvectors that determines the behavior of the solution.

Let us now discuss how common (or uncommon) non-diagonalizable matrices are. More precisely, we do *not* mean to discuss here specific applications where such matrices occur.[13] Rather, we will consider the following issue. First, let us note that a matrix

$$\mathcal{N} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

is not diagonalizable. (Indeed, both its eigenvalues equal 1, yet it is not the identity matrix.) But matrices

$$\mathcal{N}_{\epsilon,1} = \begin{pmatrix} 1 & 1 \\ \epsilon & 1 \end{pmatrix} \qquad \text{and} \qquad \mathcal{N}_{\epsilon,2} = \begin{pmatrix} 1 & 1 \\ 0 & 1+\epsilon \end{pmatrix},$$

which are just some slight perturbations of $\mathcal{N}$, are diagonalizable. Then it may seem that studying properties of non-diagonalizable matrices is not worth one's while because by just slightly perturbing those matrices, one would obtain diagonalizable matrices. However, many properties of diagonalizable matrices obtained as slight perturbations of non-diagonalizable matrices are still **close to the properties of the primordial non-diagonalizable matrices**. You will explore this in a homework problem. Thus, understanding peculiar properties of non-diagonalizable matrices should help one to understand the behavior of matrices (diagonalizable or not) that are, in some sense, close to non-diagonalizable ones.

As a piece of nomenclature, let us note that *non-diagonalizable* matrices give an ultimate example of so-called ***non-normal*** matrices. The definition of a ***normal*** (and real-valued) matrix, $\mathcal{M}$, is that it commutes with its transpose:

$$\mathcal{M} \text{ is normal} \qquad \Leftrightarrow \qquad \mathcal{M}\mathcal{M}^T = \mathcal{M}^T\mathcal{M}.$$

Clearly, any real symmetric matrix is normal. Any normal matrix is diagonalizable. Moreover, eigenvalues and eigenvectors of a normal matrix $\mathcal{M}$ provide complete information for the large-n behavior of a product $\mathcal{M}^n \underline{x}$.

On the other hand, *not* all non-normal matrices are non-diagonalizable, but, vice versa, all non-diagonalizable matrices *are* non-normal. Also, any matrix that is in some sense close

[13] One example, concerning matrix (12.16), is given below. Many more examples and references can be found, e.g., in the book "Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators" by Lloyd N. Trefethen and Mark Embree (Princeton University Press, 2005) or in its accompanying website http://web.comlab.ox.ac.uk/pseudospectra/.

to a non-diagonalizable matrix is also non-normal. For example, both matrices $\mathcal{N}_{\epsilon,1}$ and $\mathcal{N}_{\epsilon,2}$ shown above are non-normal. In this lecture we will use both names, 'non-normal' and 'non-diagonalizable'. Many examples and applications of non-normal matrices are found in a definitive book by Trefethen and Embree cited in the footnote on the previous page. In this short lecture, we will only show that eigenvalues of a non-normal matrix $\mathcal{N}$ may bear no relation to the large-n behavior of $\mathcal{N}^n \underline{x}$.

12.2 2×2 case

Consider a matrix

$$A = \begin{pmatrix} \lambda & a \\ 0 & \lambda \end{pmatrix}, \qquad a \neq 0, \tag{12.1}$$

whose doubly-repeated eigenvalue is λ and whose *only* eigenvector can be easily computed to be

$$\underline{v_1} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Since there is no second linearly independent eigenvector of this matrix, it is *not* diagonalizable, and hence is non-normal. Suppose we still want to use a basis in R^2 that contains eigenvector $\underline{v_1}$. Then an intuitively appealing second member of the basis is

$$\underline{u} = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

even though it is not an eigenvector of A.

Let us compute $A\underline{u}$:

$$A\underline{u} = \begin{pmatrix} a \\ \lambda \end{pmatrix} = a \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{12.2a}$$

or, equivalently,

$$A\underline{u} = a\underline{v_1} + \lambda \underline{u}. \tag{12.2b}$$

Note that the multiplication of $\underline{u}$ by A leads not only to the term $\lambda\underline{u}$, which would be typical of a true eigenvector, but also to an extra term $\text{const} \cdot \underline{v_1}$. A vector $\underline{u}$ satisfying Eq. (12.2b) is called a ***generalized eigenvector*** of A.

Next, to establish a formula for $A^n\underline{u}$, first compute $A^2\underline{u}$ and $A^3\underline{u}$:

$$\begin{aligned} A^2\underline{u} &= A(A\underline{u}) = A(a\underline{v_1} + \lambda\underline{u}) \\ &= a\lambda\underline{v_1} + \lambda(a\underline{v_1} + \lambda\underline{u}) \\ &= 2a\lambda\underline{v_1} + \lambda^2\underline{u}. \end{aligned} \tag{12.3a}$$

$$\begin{aligned} A^3\underline{u} &= A(A^2\underline{u}) = 2a\lambda \cdot A\underline{v_1} + \lambda^2 \cdot A\underline{u} \\ &= 2a\lambda^2\underline{v_1} + \lambda^2(a\underline{v_1} + \lambda\underline{u}) \\ &= 3a\lambda^2\underline{v_1} + \lambda^3\underline{u}. \end{aligned} \tag{12.3b}$$

From these calculations, the general pattern transpires:

$$A^n\underline{u} = na\lambda^{n-1}\underline{v_1} + \lambda^n\underline{u}. \tag{12.4}$$

Above, we have proved it for $n \geq 1$. In a homework problem you will use a similar method to prove Eq. (12.4) for $n < 0$.

From Eq. (12.4) we can make the following observation. As we noted earlier, the eigenvalues for the matrix A given by Eq. (12.1) do not fully characterize the long-time (i.e., the large-n, where n is defined in (12.4)) behavior of a system described by A, contrary to what was the case for the matrices in Lecture 5. Indeed, suppose $\lambda = 1$. For any of the matrices considered in Lecture 5, this would mean that as $n \to \infty$ and for the generic initial vector $\underline{v}$, $A^n\underline{v} \;\to\; 1^n \cdot \underline{v_1} = \underline{v_1}$. In particular, for the matrices of Lecture 5, $A^n\underline{v}$ would *remain bounded* (i.e., would not grow) for all n. In contrast, for the matrix A given by Eq. (12.1), $A^n\underline{v}$ for a generic $\underline{v}$ will have a component that grows with n: see the first term on the r.h.s. of (12.4). This growth, however, is only linear in n rather than exponential, as in λ^n.

12.3 3×3 and $M \times M$ cases

First, consider a matrix

$$A = \begin{pmatrix} \lambda & a & 0 \\ 0 & \lambda & a \\ 0 & 0 & \lambda \end{pmatrix}, \qquad a \neq 0\,. \tag{12.5}$$

Again, it has only one eigenvector,

$$\underline{v_1} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix},$$

corresponding to the triply-repeated eigenvalue λ. As in Section 1, let us augment this vector to a basis in a straightforward way, taking the other basis vectors to be

$$\underline{u_1} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \underline{u_2} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \tag{12.6}$$

and consider $A\underline{u_1}$ and $A\underline{u_2}$:

$$A\underline{u_1} = \begin{pmatrix} a \\ \lambda \\ 0 \end{pmatrix} = a\underline{v_1} + \lambda\underline{u_1}\,, \tag{12.7a}$$

$$A\underline{u_2} = \begin{pmatrix} 0 \\ a \\ \lambda \end{pmatrix} = a\underline{u_1} + \lambda\underline{u_2}\,. \tag{12.7b}$$

Vectors $\underline{u_1}$ and $\underline{u_2}$ satisfying Eqs. (12.7), which are the counterparts of Eq. (12.2b), are called the first and second *generalized eigenvectors* of A, respectively.

Next, to establish the patterns for $A^n\underline{u_{1,2}}$, first perform explicit calculations for $n = 2$ and 3:

$$\begin{aligned} A^2\underline{u_1} = A(a\underline{v_1} + \lambda\underline{u_1}) &= a\lambda\underline{v_1} + \lambda(a\underline{v_1} + \lambda\underline{u}) \\ &= 2a\lambda\underline{v_1} + \lambda^2\underline{u_1}\,; \end{aligned} \tag{12.8a}$$

$$\begin{aligned} A^2\underline{u_2} = A(a\underline{u_1} + \lambda\underline{u_2}) &= a(a\underline{v_1} + \lambda\underline{u_1}) + \lambda(a\underline{u_1} + \lambda\underline{u_2}) \\ &= a^2\underline{v_1} + 2a\lambda\underline{u_1} + \lambda^2\underline{u_2}\,. \end{aligned} \tag{12.8b}$$

$$A^3\underline{u_1} = A(2a\lambda\underline{v_1} + \lambda^2\underline{u_1})\,\big|_{\text{similarly to (12.3b)}} = 3a\lambda^2\underline{v_1} + \lambda^3\underline{u_1}\,; \tag{12.9a}$$

$$\begin{aligned} A^3\underline{u_2} &= A(a^2\underline{v_1} + 2a\lambda\underline{u_1} + \lambda^2\underline{u_2}) \\ &= a^2\lambda\underline{v_1} + 2a\lambda(a\underline{v_1} + \lambda\underline{u_1}) + \lambda^2(a\underline{u_1} + \lambda\underline{u_2}) \\ &= 3a^2\lambda\underline{v_1} + 3a\lambda^2\underline{u_1} + \lambda^3\underline{u_2}\,. \end{aligned} \tag{12.9b}$$

The pattern for $A^n\underline{u_1}$ is the same as in Section 1. In particular, the coefficient of $\underline{v_1}$ is $n\,a\lambda^{n-1}$; see (12.4). However, in the expression for $A^n\underline{u_2}$, it is still unclear what is happening with the coefficient of $\underline{v_1}$. To find out, we extend our calculations to two more powers of A, i.e. $A^4\underline{u_2}$ and $A^5\underline{u_2}$:

$$\begin{aligned} A^4\underline{u_2} &= 3a^2\lambda\cdot A\underline{v_1} + 3a\lambda^2\cdot A\underline{u_1} + \lambda^3\cdot A\underline{u_2} \\ &= 3a^2\lambda^2\underline{v_1} + 3a\lambda^2(a\underline{v_1} + \lambda\underline{u_1}) + \lambda^3(a\underline{u_1} + \lambda\underline{u_2}) \\ &= 6a^2\lambda^2\underline{v_1} + 4a\lambda^3\underline{u_1} + \lambda^4\underline{u_2}\,; \end{aligned} \tag{12.10a}$$

$$\begin{aligned} A^5\underline{u_2} &= 6a^2\lambda^2\cdot A\underline{v_1} + 4a\lambda^3\cdot A\underline{u_1} + \lambda^4\cdot A\underline{u_2} \\ &= \underbrace{6a^2\lambda^3\underline{v_1}} + \underbrace{4a\lambda^3(a\underline{v_1}} + \lambda\underline{u_1}) + \lambda^4(a\underline{u_1} + \lambda\underline{u_2})\,. \end{aligned} \tag{12.10b}$$

Instead of completing the summation in (12.10b), let us focus on the two terms with underbraces, since it is their coefficient whose form we want to guess. Let $c_n\cdot a^2\lambda^{n-2}$ be this coefficient of $\underline{v_1}$ in the expression for $A^n\underline{u_2}$. Then from (12.10b) we can conclude that

$$c_{n+1} = c_n + n\,, \tag{12.11a}$$

where the '$+n$' comes from the fact that the coefficient of $\underline{v_1}$ in $A^n\underline{u_1}$ is $n\cdot a\lambda^{n-1}$. Note that the pattern (12.11a) is consistent with (12.10a) and (12.9b). In addition, from (12.7b) and (12.8b) we have that

$$c_1 = 0\,. \tag{12.11b}$$

From both of Eqs. (12.11), one has:

$$\begin{aligned} c_{n+1} &= c_n + n = (c_{n-1} + (n-1)) + n = \ldots \\ &= c_1 + 1 + 2 + \ldots + n = \frac{n(n+1)}{2}\,. \end{aligned} \tag{12.12}$$

Finally, from (12.9b), (12.10), and (12.12) we conclude that

$$A^n\underline{u_2} = \frac{(n-1)n}{2}a^2\lambda^{n-2}\underline{v_1} + na\lambda^{n-1}\underline{u_1} + \lambda^n\underline{u_2}\,. \tag{12.13}$$

Compare this with the formula (12.4). (For this comparison, it may be helpful to write, on one side of a page, Eqs. (12.7a) and (12.4), and on the other, Eqs. (12.7b) and (12.13).)

An important point to note about formula (12.13) is that $A^n\underline{u_2}$ has a term whose behavior for large n is:

$$\frac{n(n-1)}{2}\,a^2\lambda^{n-2}\underline{v_1} = O(n^2)\cdot a^2\lambda^{n-2}\cdot\underline{v_1}\,. \tag{12.14}$$

Now, as in the last paragraph of Section 1, suppose $\lambda = 1$. Then for the generic initial vector $\underline{v}$, $A^n\underline{v}$ will grow as $O(n^2)$. This, again, should be contrasted with the behavior of $A^n\underline{v}$ if A were a *normal* (e.g., symmetric) matrix with $\lambda = 1$.

Generalizing the above calculations for the 2×2 and 3×3 non-diagonalizable matrices (12.1) and (12.5), consider an $M \times M$ non-diagonalizable matrix

$$A = \begin{pmatrix} \lambda & a & 0 & 0 & \dots & 0 & 0 \\ 0 & \lambda & a & 0 & \dots & 0 & 0 \\ 0 & 0 & \lambda & a & \dots & 0 & 0 \\ \vdots & & & & & & \vdots \\ 0 & 0 & 0 & 0 & \dots & \lambda & a \\ 0 & 0 & 0 & 0 & \dots & 0 & \lambda \end{pmatrix}, \qquad a \neq 0\,. \tag{12.15}$$

An example of such a matrix is

$$A = \begin{pmatrix} 1 & -1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & -1 & \dots & 0 & 0 \\ \vdots & & & & & & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & -1 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix}, \tag{12.16}$$

which occurs when one writes the familiar formula

$$u'(t_j) \approx \frac{u(t_{j+1}) - u(t_j)}{h} \tag{12.17}$$

in matrix form:

$$\begin{pmatrix} u'(t_1) \\ u'(t_2) \\ \vdots \\ u'(t_{M-1}) \\ u'(t_M) \end{pmatrix} \approx -\frac{1}{h} \begin{pmatrix} 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 1 & -1 & \dots & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & \dots & 1 & -1 \\ 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_{M-1} \\ u_M \end{pmatrix}. \tag{12.18}$$

(In (12.18), we have arbitrarily set $u'(t_M) = -\frac{1}{h}u_M$ in order to have the matrix of the same form as in (12.16). While it is clear that such a choice does not make much sense, it is also intuitively clear that it only affects the last entry of the vector on the l.h.s. and hence does not affect the behavior of the remaining "bulk" of the vector of the derivatives.)

Similarly to the calculations for the 2×2 and 3×3 cases, one can show that if

$$\underline{u_{M-1}} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}, \tag{12.19}$$

then:

$$\underline{\text{for } \; n \geq M-1:} \qquad A^n \underline{u_{M-1}} = O\left(\frac{n^{M-1}}{(M-1)!}\right) a^{M-1} \cdot \lambda^{n-(M-1)} \underline{v_1} + \left\{ \begin{array}{c} \text{other} \\ \text{terms} \end{array} \right\}. \tag{12.20}$$

Therefore, if $\lambda = 1$ (as, for example, for the matrix in Eq. (12.16)), then the behavior of $A^n\underline{v}$ for the generic initial vector $\underline{v}$ will be dominated by the $O(n^{M-1})$-term in (12.20), which for large M is a *very fast-growing* function of n. Thus, even though the eigenvalue ($\lambda = 1$) predicts that $A^n \underline{v}$ is to remain bounded as $n \to \infty$, the *actual* behavior of $A^n \underline{v}$ is *drastically different*.

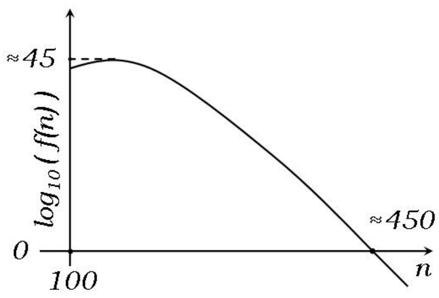
Moreover, suppose that $\lambda < 1$. If matrix A were normal (e.g., symmetric), then $A^n \underline{v} \sim \lambda^n \underline{v} \to 0$ as $n \to \infty$. *Theoretically*, this is also the case for the non-diagonalizable matrix A in (12.15), because the most troublesome term in (12.20) is $O(n^{M-1} \cdot \lambda^{n-(M-1)})$, and this can be shown, by a repeated application of L'Hôpital's Rule, to tend to zero:

$$\lim_{n\to\infty} (n^p \cdot \lambda^{n-p}) = 0 \quad \text{for } |\lambda| < 1 \text{ and for any fixed } p. \tag{12.21}$$

However, let us look at this from a *practical* perspective where we would have to compute such a term on a computer. For a specific example, let us take $\lambda = 1/2$ and $M = 100$. Then the graph of

$$f(n) = \frac{n^{99}}{99!} \cdot \left(\frac{1}{2}\right)^{n-99}, \qquad n \geq 100 \tag{12.22}$$

looks like this:



That is, for $n \sim 100$, this is a *HUGE* number of magnitude greater than $\sim 10^{40}$!!! Hence, for *any practical purpose*, the behavior of

$$O(n^{M-1}) \cdot \lambda^{n-(M-1)}, \qquad \textit{even for } |\lambda| < 1 \tag{12.23}$$

will be detected by the computer as a *very fast growth*, and even as a blow up. This is in stark contrast to the $\left(\frac{1}{2}\right)^n$-like behavior that would take place for a *normal* matrix A with $\lambda = 1/2$.

Let us summarize **the main point** of this Section. The large-n behavior of a *non-normal* matrix A of a large dimension can easily be very different from what could have been predicted based only on the eigenvalue(s) of A.

12.4 The notion of the spectrum and pseudo-spectrum of a matrix

The spectrum of a matrix is simply the set of all its eigenvalues. For a normal matrix, its spectrum fully characterizes the large-n behavior of $A^n \underline{v}$ for the generic initial vector $\underline{v}$. On the other hand, we showed in the previous Section that for a non-normal matrix A, this is not so: its spectrum *may bear no relation* to the large-n behavior of $A^n \underline{v}$. In this Section, we will show that the spectrum of a large non-normal matrix A is also *very unstable*. That is, the spectra of A and $(A + \varepsilon B)$, where $\varepsilon \ll 1$ and B is some generic matrix, may differ by $O(1)$ (rather than by $O(\varepsilon)$).

To begin, let us note that the spectrum of a *normal* $M \times M$ matrix, which has M linearly independent eigenvectors, *is stable*. Instead of giving a general proof of this fact, we will illustrate it with a simple example. Namely, the spectrum of $(A + \varepsilon B)$ where

$$A = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_M)$$

is a diagonal (and hence normal) matrix and B is a matrix with a *single entry* anywhere off the main diagonal, is the same as the spectrum of A (i.e., it is the set $\{\lambda_1, \lambda_2, \ldots, \lambda_M\}$). Now, if the only entry of B is on the main diagonal, then the corresponding eigenvalue changes from

λ_j to $\lambda_j + \varepsilon b_{jj}$, i.e. by $O(\varepsilon) \ll 1$, while the other eigenvalues remain unchanged. In general, one can show that if A is symmetric (and hence diagonalizable), then the spectra of A and $A + \varepsilon B$, where B is now *any* matrix (of order one), differ only by an amount $O(\varepsilon)$. Such a stable behavior is what one would intuitively expect of a "well-behaved" matrix.

Now consider the non-diagonalizable matrix A given by Eq. (12.16) and perturb it by placing an entry ε in the bottom-left corner:

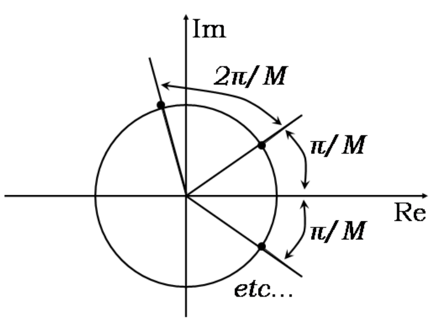
$$A + \varepsilon B = \begin{pmatrix} 1 & -1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & -1 & \dots & 0 & 0 \\ \vdots & & & & & & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & -1 \\ \boxed{\varepsilon} & 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix}. \tag{12.24}$$

The spectrum (i.e., the set of eigenvalues) of $(A + \varepsilon B)$ is found by computing

$$\begin{aligned} \det(A + \varepsilon B - \lambda I) &= (1-\lambda) \cdot \det \begin{pmatrix} (1-\lambda) & -1 & 0 & 0 & \dots & 0 & 0 \\ 0 & (1-\lambda) & -1 & 0 & \dots & 0 & 0 \\ 0 & 0 & (1-\lambda) & -1 & \dots & 0 & 0 \\ \vdots & & & & & & \vdots \\ 0 & 0 & 0 & 0 & \dots & (1-\lambda) & -1 \\ 0 & 0 & 0 & 0 & \dots & 0 & (1-\lambda) \end{pmatrix} \\ &\quad + \varepsilon \cdot (-1)^{M+1} \cdot \det \begin{pmatrix} -1 & 0 & 0 & 0 & \dots & 0 & 0 \\ (1-\lambda) & -1 & 0 & 0 & \dots & 0 & 0 \\ 0 & (1-\lambda) & -1 & 0 & \dots & 0 & 0 \\ \vdots & & & & & & \vdots \\ 0 & 0 & 0 & 0 & \dots & (1-\lambda) & -1 \end{pmatrix} \\ &= (1-\lambda) \cdot (1-\lambda)^{M-1} + \varepsilon (-1)^{M+1} \cdot (-1)^{M-1} \\ &= (1-\lambda)^M + \varepsilon \,. \end{aligned} \tag{12.25}$$

By setting $\det(A + \varepsilon B - \lambda I) = 0$, we obtain:

$$(1-\lambda)^M + \varepsilon = 0 \quad \Rightarrow \quad 1 - \lambda = (-\varepsilon)^{1/M} \quad \Rightarrow \quad \lambda = 1 - (-1)^{1/M} \cdot \sqrt[M]{\varepsilon}\,. \tag{12.26}$$



In a Complex Analysis course it is shown that there are M complex-valued M-th roots of (-1) (similarly to how $\pm i$ are the two square roots of -1). These roots are all located on the unit circle in the complex plane (see the figure on the left). Thus, there are now M different eigenvalues of the perturbed matrix $(A + \varepsilon B)$, even though A had a single M-times repeated eigenvalue $\lambda = 1$. In such cases one says that the perturbation has split the degenerate eigenvalue, or "removed the degeneracy".

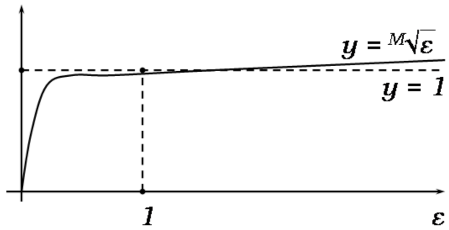
A more important observation concerns the *magnitude of the difference* between the spectra of $(A + \varepsilon B)$ and A. This difference is given by the last term in (12.26), i.e.:

$$|\,\text{spectrum}(A + \varepsilon B) - \text{spectrum}(A)\,| = O(\sqrt[M]{\varepsilon})\,. \tag{12.27}$$

Let $M = 100$ and $\varepsilon = 10^{-16}$, which is Matlab's round-off error, a really tiny number. But

$$\sqrt[100]{10^{-16}} = 10^{-0.16} \simeq 0.69 = O(1)\,, \tag{12.28}$$

i.e. a number of order one! Thus, even a tiny numerical error can drastically change the spectrum of a large non-diagonalizable (or, more generally, strongly non-normal) matrix.



Moreover, *for* $M \gg 1$,

$$\sqrt[M]{\varepsilon} \approx 1 \quad \text{for } 10^{-M} \ll \varepsilon \ll 10^{M}\,, \tag{12.29}$$

i.e. for all values of ε that may be encountered in practice. Therefore, for all "practical" values of ε, Eq. (12.26) is essentially independent of ε:

$$\lambda \approx 1 - (-1)^{1/M}\,. \tag{12.26$'$}$$

This is referred to as the ***pseudo-spectrum*** of the matrix A of Eq. (12.16). The theory of pseudospectra of (non-normal) matrices is covered in advanced graduate courses on Numerical Linear Algebra.

Let us summarize **the main points of this Lecture**:

1) The spectrum and the set of eigenvectors of a *normal* (e.g., symmetric) or a close-to-normal matrix A completely determines the evolution $A^n \underline{v}$ for any initial vector $\underline{v}$.

2) The set of eigenvectors of a *non-diagonalizable* matrix A does not form a basis in the corresponding vector space. Consequently, the spectrum of a non-diagonalizable matrix does *not* determine the behavior of $A^n \underline{v}$. Moreover, for a *large* non-diagonalizable matrix, such a behavior may be completely opposite of what the spectrum alone would predict (see Section 2). Similar statements hold for any "strongly" non-normal matrix, i.e. a matrix that may be diagonalizable but is, in some sense, close to some non-diagonalizable matrix.

3) While the spectrum of a normal matrix is *stable* with respect to small perturbations of its entries, the spectrum of a large non-normal matrix can be *very unstable*. Consequently, the concept that is of practical importance for large strongly non-normal matrices is *not* their spectra but pseudo-spectra (see Section 3).