ELEN 162
A. I. Zecevic

Lecture Notes for Week 2

Matrix Norms

The notion of a norm can be extended to matrices as well. This can be done in several different ways, the most straightforward of which relies on the following definition.

Definition 1. Let $\|\cdot\|$ denote a vector norm in R^n. The matrix norm *induced* by $\|\cdot\|$ is defined as

$$\|A\| = \max_{\|x\|=1} \|Ax\| \tag{1}$$

The term "induced norm" is appropriate in this context because the matrix norm defined by (1) is uniquely determined by the choice of vector norm (note that Ax is a *vector*).

Since Definiton 1 is not very "user friendly", we need to establish what it means for different choices of vector norms. In order to do that, we will need the following definition.

Definition 2. Let A be a real matrix whose eigenvalues are $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$, and let

$$\rho(A) = \max_i |\lambda_i| \tag{2}$$

denote the eigenvalue with the *largest magnitude*. This eigenvalue is referred to as the *spectral radius* of A.

Combining this definition with Definition 1, it can be shown that the matrix norm induced by the Euclidean vector norm (which is denoted $\|A\|_2$) can be computed as

$$\|A\|_2 = \sqrt{\rho(A^T A)} \tag{3}$$

It can also be shown that the matrix norm induced by the weighted infinity norm (denoted $\|A\|_\infty^\omega$) can be computed as

$$\|A\|_\infty^\omega = \max_i \frac{1}{\omega_i} \sum_{j=1}^{n} |a_{ij}|\omega_j \tag{4}$$

In the special case when $\omega_1 = \ldots = \omega_n = 1$, this norm reduces to

$$\|A\|_\infty = \max_i \sum_{j=1}^{n} |a_{ij}| \tag{5}$$

Proofs that all three expressions conform to (1) are provided in the textbook.

It is not difficult to show that *any* induced matrix norm must satisfy the following four conditions:

1. $\|A\| > 0$ when $A \neq 0$, and $\|A\| = 0$ if and only if $A = 0$.
2. For any $c \in R$ we have that $\|cA\| = |c| \, \|A\|$.
3. Given a pair of matrices A and B, inequalities

$$\|A + B\| \leq \|A\| + \|B\| \tag{6}$$

and

$$\|AB\| \leq \|A\| \, \|B\| \tag{7}$$

must hold.

4. For any vector $x \in R^n$ the corresponding matrix norm must satisfy

$$\|Ax\| \leq \|A\| \, \|x\| \tag{8}$$

Our next example shows how different types of norms are computed, and allows us to make some comparisons between them.

Example 1. Consider matrix

$$A = \begin{bmatrix} 1 & 0 & -2 \\ 0 & 4 & -5 \\ -1 & 2 & 3 \end{bmatrix} \tag{9}$$

which is obviously *not* symmetric. In order to compute its Euclidean norm, we first need to form matrix

$$A^T A = \begin{bmatrix} 2 & -2 & -5 \\ -2 & 20 & -14 \\ -5 & -14 & 38 \end{bmatrix} \tag{10}$$

whose eigenvalues are

$$\lambda(A^T A) = \begin{cases} 0.30988 \\ 13.7757 \\ 45.91442 \end{cases} \tag{11}$$

From (11), it follows that the spectral radius of this matrix is

$$\rho(A^T A) = \max_i \left| \lambda_i(A^T A) \right| = 45.91442 \tag{12}$$

and we have that

$$\|A\|_2 = \sqrt{\rho(A^T A)} = 6.776 \tag{13}$$

To evaluate the infinity norm, we need to form matrix

$$|\, A \,| = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 4 & 5 \\ 1 & 2 & 3 \end{bmatrix} \tag{14}$$

and add up the elements in each row. If we do so, we will easily verify that

$$\|A\|_\infty = \max_i \sum_{j=1}^{3} |a_{ij}| = 9 \tag{15}$$

This value is obviously larger than the one obtained using the Euclidean norm, which is something that we should keep in mind.

When it comes to the *weighted infinity norm*, the result will depend on how we choose parameters ω_1, ω_2 and ω_3. If we set $\omega_1 = 1$, $\omega_2 = 0.5$ and $\omega_3 = 0.4$, for example, we obtain

$$\Sigma_1 = \frac{1}{\omega_1} \sum_{j=1}^{n} |a_{1j}| \omega_j = 1.8 \tag{16}$$

$$\Sigma_2 = \frac{1}{\omega_2} \sum_{j=1}^{n} |a_{2j}| \omega_j = 8 \tag{17}$$

and

$$\Sigma_3 = \frac{1}{\omega_3} \sum_{j=1}^{n} |a_{3j}| \omega_j = 8 \tag{18}$$

which implies that

$$\|A\|_\infty^\omega = \max_i \frac{1}{\omega_i} \sum_{j=1}^{n} |a_{ij}| \omega_j = 8 \tag{19}$$

The fact that this value is smaller than the one we obtained in (15) indicates that the value of $\|A\|_\infty^\omega$ can be reduced (to some extent, of course) by an appropriate choice of vector ω. This added degree of freedom can be very useful, particularly in the context of iterative methods.

Quadratic Forms

A *quadratic form* is defined as the scalar product

$$\langle x, Ax \rangle = x^T A x \tag{20}$$

where A is assumed to be a *symmetric* matrix. Expressions of this sort are commonly encountered in optimization problems, and are also associated with certain iterative techniques for solving systems of linear algebraic equations (such as the conjugate gradient method, for example).

Definition 3. A symmetric matrix A is said to be *positive definite* if all of its eigenvalues are positive.

The following theorem provides necessary and sufficient conditions for positive definiteness. In this case I will actually provide a proof, since the arguments that I will use for this purpose will be helpful for Homework 1.

Theorem 1.4. Let A be a symmetric matrix with distinct eigenvalues. Then,

$$x^T A x > 0 \tag{21}$$

for any nonzero vector x in R^n if and only if A is positive definite.

Proof. If A is assumed to be positive definite, then all of its eigenvalues must be positive. Let us now set $y = T^{-1}x$, where T is a matrix whose columns are the normalized eigenvectors of A. This definiton obviously implies that $x = Ty$, which allows us to rewrite the quadratic form as

$$x^T Ax = y^T T^T ATy \tag{22}$$

Recalling that $T^{-1} = T^T$ when matrix A is symmetric, this becomes

$$x^T Ax = y^T T^T ATy = y^T T^{-1} ATy \tag{23}$$

From Lemma 1.7, we also know that

$$T^{-1}AT = \Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \tag{24}$$

where Λ represents the Jordan canonical form. Consequently, expression (23) becomes

$$x^T Ax = y^T T^{-1} ATy = y^T \Lambda y = \sum_{i=1}^{n} \lambda_i y_i^2 \geq 0 \tag{25}$$

To rule out the posibility that $y = 0$, we should observe that $x \neq 0$ by assumption. If the corresponding vector $y = T^{-1}x$ happened to be 0, we would have

$$T^{-1}x = y = 0 \tag{26}$$

which is a contradiction, since system (26) has a unique solution (which is $x = 0$, because $T^{-1}$is a nonsingular matrix).

In view of that, we can conclude that if $x \neq 0$, the corresponding vector $y = T^{-1}x$ will satisfy $y \neq 0$ as well. Recalling that $\lambda_i > 0$ $(i = 1, 2, \ldots, n)$ by assumption, it follows that

$$x^T Ax = \sum_{i=1}^{n} \lambda_i y_i^2 > 0 \tag{27}$$

Suppose now that

$$x^T Ax > 0 \tag{28}$$

for all $x \neq 0$, but that A is *not* positive definite. In that case, at least one of its eigenvalues will satisfy $\lambda_k \leq 0$. To see why this leads to a contradiction, let us denote the eigenvector that corresponds to λ_k by x_k. Substituting this vector into the quadratic form, we directly obtain

$$x_k^T Ax_k = \lambda_k x_k^T x_k = \lambda_k \|x_k\|_2^2 \leq 0 \tag{29}$$

which violates condition (21). **Q.E.D.**

Functions of Matrices

Given a function $f(x)$ whose Taylor series expansion has the form

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k \tag{30}$$

we can define the corresponding function of matrix A as

$$f(A) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} A^k \tag{31}$$

The individual terms in this sum have a straightforward interpretation, since they represent powers of A.

The problem with expression (31) is that it cannot be evaluated directly, since it involves an infinite sum. To see how this difficulty can be resolved, let $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ denote the eigenvalues of A, and let $\{x_1, x_2, \dots, x_n\}$ be the corresponding eigenvectors. For the sake of simplicity, in the following we will assume that the eigenvalues are *real* and *distinct*, and that function $f(x)$ and all of its derivatives are finite in points $x = \lambda_i$ $(i = 1, 2, \dots, n)$. When this is the case, we say that f is *defined on the spectrum of* A, and sequence (31) is guaranteed to converge.

We now proceed to show that $f(A)$ has the *same* eigenvectors as A, and that its eigenvalues are $\{f(\lambda_1),\ f(\lambda_2),\ \dots,\ f(\lambda_n)\}$. In order to do that, we should first observe that

$$A^k x_i = \lambda_i^k x_i \tag{32}$$

for any positive integer k. This is easily verified by induction, since

$$A^2 x_i = A\,[Ax_i] = A\,[\lambda_i x_i] = \lambda_i [Ax_i] = \lambda_i^2 x_i \tag{33}$$

$$A^3 x_i = A\left[A^2 x_i\right] = A\left[\lambda_i^2 x_i\right] = \lambda_i^2 [Ax_i] = \lambda_i^3 x_i \tag{34}$$

and so on. Recalling that

$$f(A)x_i = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} A^k x_i = \left[\sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} \lambda_i^k\right] x_i \tag{35}$$

and that

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} \lambda_i^k = f(\lambda_i) \tag{36}$$

the relationship

$$f(A)x_i = f(\lambda_i)x_i \tag{37}$$

follows directly.

How can we make use of this result? In order to see that, let us form a matrix

$$T = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \tag{38}$$

whose columns are the eigenvectors of A. When we multiply T on the left by $f(A)$, we obtain

$$f(A)T = [f(A)x_1 \, f(A)x_2 \, \ldots \, f(A)x_n] \tag{39}$$

which becomes

$$f(A)T = [f(\lambda_1)x_1 \, f(\lambda_2)x_2 \, \ldots \, f(\lambda_n)x_n] \tag{40}$$

by virtue of (37).

We should note at this point that the right hand side of expression (40) can be rewritten as

$$[f(\lambda_1)x_1 \, f(\lambda_2)x_2 \, \ldots \, f(\lambda_n)x_n] = Tf(\Lambda) \tag{41}$$

where

$$f(\Lambda) = \begin{bmatrix} f(\lambda_1) & 0 & \cdots & 0 \\ 0 & f(\lambda_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & f(\lambda_n) \end{bmatrix} \tag{42}$$

If we now combine (40) and (41), we obtain

$$f(A)T = Tf(\Lambda) \tag{43}$$

and therefore

$$f(A) = Tf(\Lambda)T^{-1} \tag{44}$$

This result is useful because it provides us with a straightforward way to compute $f(A)$ once we have determined the eigenvectors and eigenvalues of matrix A. The following example illustrates how this approach works in practice.

Example 2. Let

$$A = \begin{bmatrix} 2 & -2 \\ -2 & 4 \end{bmatrix} \tag{45}$$

and suppose that we want to compute function

$$f(A) = e^{-A}\cos(A) \tag{46}$$

In order to do that, we first need to determine the eigenvalues and eigenvectors of matrix A, which happen to be

$$\lambda_1(A) = 0.7639 \quad \longrightarrow \quad x_1 = \begin{bmatrix} -0.85065 \\ -0.52573 \end{bmatrix} \tag{47}$$

and

$$\lambda_2(A) = 5.2361 \quad \longrightarrow \quad x_2 = \begin{bmatrix} -0.52573 \\ 0.85065 \end{bmatrix} \tag{48}$$

in this case. Using this information, we can compute matrices T and Λ as

$$T = \begin{bmatrix} -0.85065 & -0.52573 \\ -0.52573 & 0.85065 \end{bmatrix} \tag{49}$$

and

$$\Lambda = \begin{bmatrix} 0.7639 & 0 \\ 0 & 5.2361 \end{bmatrix} \tag{50}$$

respectively.

Let us now define function $f(x)$ as

$$f(x) = e^{-x} \cos x \tag{51}$$

The corresponding matrix function $f(\Lambda)$ will then have the form

$$f(\Lambda) = \begin{bmatrix} e^{-\lambda_1} \cos \lambda_1 & 0 \\ 0 & e^{-\lambda_2} \cos \lambda_2 \end{bmatrix} = \begin{bmatrix} 0.33639 & 0 \\ 0 & 0.00266 \end{bmatrix} \tag{52}$$

Recalling that matrix A is symmetric, we know that $T^{-1} = T^T$, so matrix $f(A)$ can be computed as

$$f(A) = Tf(\Lambda)T^{-1} = Tf(\Lambda)T^T = \begin{bmatrix} 0.244147 & 0.149247 \\ 0.149247 & 0.094900 \end{bmatrix} \tag{53}$$

Remark 1. Note that matrix $f(A)$ is *symmetric*, which is to be expected (according to Lemma 1.9 in the textbook).

Functional Spaces

Functional spaces are linear vector spaces whose elements are *functions*. The one that we will be most interested in consists of complex-valued functions of real arguments which satisfy

$$\int_{-\infty}^{\infty} |f(x)|^2 dx < \infty \tag{54}$$

(such functions are said to be *square integrable* on interval $(-\infty\ \infty)$). In this space, it is assumed that the addition of two functions f and g produces a new function $h = f + g$ which is defined as

$$h(x) = f(x) + g(x) \tag{55}$$

Similarly, multiplying function f by some complex number c is assumed to produce a function $\xi = cf$ such that

$$\xi(x) = cf(x) \tag{56}$$

When addition and multiplication by a number are defined in this way, it is not difficult to show that all the properties of linear spaces are satisfied.

The scalar product in this space is defined as

$$\langle f, g\rangle = \int_{0}^{\infty} f^*(x)g(x)dx \tag{57}$$

where $f^*(x)$ denotes the *complex conjugate of* $f(x)$. Such a definition ensures that the following four identities hold for any choice of α, β, f and g.

$$\begin{aligned}
&1)\ \ \langle \alpha f, g\rangle = \alpha^*\langle f, g\rangle \\
&2)\ \ \langle f, \beta g\rangle = \beta\langle f, g\rangle \\
&3)\ \ \langle \sum_i \alpha_i f_i, g\rangle = \sum_i \alpha_i^*\langle f_i, g\rangle \\
&4)\ \ \langle f, \sum_i \beta_i g_i\rangle = \sum_i \beta_i\langle f, g_i\rangle
\end{aligned} \tag{58}$$

All of these relationships follow directly from expression (57) and the basic properties of integrals.

If we assume that functions $\{\phi_1(x), \phi_2(x), \dots\}$ constitute a *basis* in this space, then any other function $f(x)$ that belongs to it can be expressed as

$$f(x) = \sum_i \beta_i \phi_i(x) \tag{59}$$

where β_i $(i = 1, 2, \dots)$ are constant coefficients (which can be complex in general). Such an expression is very similar to the ones we have encountered before, except for the fact that set

$\{\phi_1(x), \phi_2(x), \dots\}$ now consists of *functions* rather than vectors or matrices. Such a basis is said to be *orthonormal* if

$$\langle \phi_i(x), \phi_j(x) \rangle = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \tag{60}$$

with respect to the scalar product defined in (57).

When discussing the properties of bases in functional spaces, it is important to recognize that they can contain *infinitely many* elements (unlike the bases that we considered previously). The following two examples illustrate what set $\{\phi_1(x), \phi_2(x), \dots\}$ might look like in such cases.

Example 3. A function $f(t)$ is said to be periodic with period T if it satisfies

$$f(t+T) = f(t) \tag{61}$$

for any choice of t. It is well known that any such function can be expressed in the form of a Fourier series

$$f(t) = a_0 + \sum_{n=1}^{\infty} a_n \cos(n\omega t) + \sum_{n=1}^{\infty} b_n \sin(n\omega t) \tag{62}$$

where $\omega = 2\pi/T$. The coefficients $\{a_0, a_1, a_2, \ldots\}$ and $\{b_1, b_2, \ldots\}$ in (62) are constant, and can be computed directly from function $f(t)$.

In the special case when $f(t) = f(-t)$, we say that the function is *even*, and the corresponding Fourier series reduces to

$$f(t) = a_0 + \sum_{n=1}^{\infty} a_n \cos(n\omega t) \tag{63}$$

This expression indicates that all even periodic functions with period T belong to an infinite dimensional linear space whose basis are functions

$$\{1, \cos(\omega t), \cos(2\omega t), \cos(3\omega t) \ldots\} \tag{64}$$

Example 4. Any analytic function $f(x)$ can be expressed in terms of its Taylor series expansion

$$f(x) = \sum_{k=0}^{\infty} a_k x^k \tag{65}$$

In this expression, the coefficients a_k are given as

$$a_k = \frac{f^{(k)}(0)}{k!} \tag{66}$$

where $f^{(k)}(0)$ denotes the k-th derivative of $f(x)$ evaluated at $x = 0$. It is not difficult to see that all such functions belong to an infinite dimensional space whose basis are functions

$$\{1, x, x^2, x^3 \ldots\} \tag{67}$$

Linear Operators

Before applying these concepts to quantum mechanics, we also need to say a few words about *operators*, and how they act on the elements of functional spaces. In general, an operator can be thought of as a mapping that transforms a given function f into another function φ that belongs to the same (or possibly different) space. We can formally represent this transformation as

$$\varphi = \hat{A}f \tag{68}$$

where $\hat{A}$ denotes the operator.

An operator $\hat{A}$ is said to be *linear* if it satisfies

$$\hat{A}(\alpha f + \beta g) = \alpha \hat{A} f + \beta \hat{A} g \tag{69}$$

for any two functions f and g (the coefficients α and β in (69) are assumed to be complex numbers in general). Note that matrices satisfy this property as well. Indeed, given a vector $x \in R^n$, the operation

$$w = Ax \tag{70}$$

transforms x into some other vector $w \in R^n$, and it is easily verified that

$$A(\alpha x + \beta y) = \alpha A x + \beta A y \tag{71}$$

With any linear operator, we can associate a set of eigenvalues $\{\lambda_i\}$ and corresponding eigenfunctions $\{\phi_i\}$ which satisfy

$$\hat{A}\phi_i = \lambda_i \phi_i \quad (i = 1, 2, \ldots) \tag{72}$$

Such a definition closely resembles the one that we used in linear algebra, the main difference being that we are now dealing with *eigenfunctions* instead of eigenvectors.

Remark 2. Strictly speaking, this expression applies only to operators that have a *discrete spectrum*. We can use the above definition without any loss of generality, however, since this is the only type of operator that is of interest in quantum computing. You will see examples of operators with a *continuous spectrum* in Homework 2.

In cases when the basis consists of a finite number of elements, it is often convenient to specify how operator $\hat{A}$ acts on the basis functions. We will see that this actually gives us an alternative way to *define* operator $\hat{A}$, since it allows us to determine how $\hat{A}$ acts on *any* other function.

To see how that works, suppose that functions $\{\xi_1, \xi_2, \ldots, \xi_n\}$ constitute an orthonormal basis in space S, and that

$$\hat{A}\xi_i = \varphi_i \quad (i = 1, 2, \ldots, n) \tag{73}$$

(where functions $\{\varphi_i\}$ are known). Since any function f that belongs to this space can be represented as

$$f = \alpha_1 \xi_1 + \alpha_2 \xi_2 + \ldots + \alpha_n \xi_n \tag{74}$$

applying operator $\hat{A}$ to it will produce a new function

$$g = \hat{A}f = \alpha_1 \hat{A}\xi_1 + \alpha_2 \hat{A}\xi_2 + \ldots + \alpha_n \hat{A}\xi_n = \alpha_1 \varphi_1 + \alpha_2 \varphi_2 + \ldots + \alpha_n \varphi_n \tag{75}$$

This function is easy to evaluate if we are given f, because its components $\{\varphi_i\}$ are known, and coefficients $\{\alpha_i\}$ can be computed as

$$\alpha_i = \langle f, \xi_i \rangle \qquad (i = 1, 2, \ldots, n) \tag{76}$$

Remark 3. Expression (76) follows directly from the fact that functions $\{\xi_i\}$ are orthonormal.

The following two examples show how we can use this approach to compute eigenvalues of operators and the normalized eigenfunctions that correspond to them. In both cases, we will be dealing with a two dimensional functional space S in which functions ψ_0 and ψ_1 constitute an orthonormal basis. This scenario is of interest to us because it frequently arises in quantum computing.

Example 5. Suppose that we would like to determine the eigenvalues and eigenfunctions of operator $\hat{X}$, which transforms basis functions ψ_0 and ψ_1 as

$$\hat{X}\psi_0 = \psi_1 \tag{77}$$

and

$$\hat{X}\psi_1 = \psi_0 \tag{78}$$

respectively. If $\psi \in S$ is an eigenfunction of this operator, we know that it must satisfy

$$\hat{X}\psi = \lambda\psi \tag{79}$$

where λ is a number (which can be complex in general). Since ψ is an element of S, we also know that it can be expressed as

$$\psi = \alpha_0\psi_0 + \alpha_1\psi_1 \tag{80}$$

Using (77), (78) and (80), we can rewrite $\hat{X}\psi$ as

$$\hat{X}\psi = \alpha_0\hat{X}\psi_0 + \alpha_1\hat{X}\psi_1 = \alpha_0\psi_1 + \alpha_1\psi_0 \tag{81}$$

This allows us to express (79) as

$$\hat{X}\psi = \alpha_0\psi_1 + \alpha_1\psi_0 = \lambda(\alpha_0\psi_0 + \alpha_1\psi_1) \tag{82}$$

which is convenient for our purposes. If we now group the terms next to ψ_0 and ψ_1, (29) becomes

$$(\alpha_1 - \lambda\alpha_0)\psi_0 + (\alpha_0 - \lambda\alpha_1)\psi_1 = 0 \tag{83}$$

If we now form the scalar products

$$\begin{aligned} &\langle \psi_0, (\alpha_1 - \lambda\alpha_0)\psi_0 + (\alpha_0 - \lambda\alpha_1)\psi_1 \rangle = (\alpha_1 - \lambda\alpha_0)\langle \psi_0, \psi_0 \rangle + \\ &+(\alpha_0 - \lambda\alpha_1)\langle \psi_0, \psi_1 \rangle = \alpha_1 - \lambda\alpha_0 = 0 \end{aligned} \tag{84}$$

and

$$\begin{aligned} &\langle \psi_1, (\alpha_1 - \lambda\alpha_0)\psi_0 + (\alpha_0 - \lambda\alpha_1)\psi_1 \rangle = (\alpha_1 - \lambda\alpha_0)\langle \psi 1, \psi_0 \rangle + \\ &+(\alpha_0 - \lambda\alpha_1)\langle \psi_1, \psi_1 \rangle = \alpha_0 - \lambda\alpha_1 = 0 \end{aligned} \tag{85}$$

we obtain two conditions that coefficients α_0 and α_1 must satisfy. In matrix form, these conditions can be expressed as

$$\begin{bmatrix} -\lambda & 1 \\ 1 & -\lambda \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{86}$$

which implies that coefficients α_0 and α_1 can have nonzero values only if and only if

$$\Delta(\lambda) = \lambda^2 - 1 = 0 \tag{87}$$

Based on this equation, we can conclude that the eigenvalues of operator $\hat{X}$ are $\lambda_1 = 1$ and $\lambda_2 = -1$.

In order to find the eigenfunction that corresponds to $\lambda_1 = 1$, we need to substitute this value into (86). If we do so, we obtain

$$\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{88}$$

which reduces to

$$\begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{89}$$

sfter one step of Gaussian elimination. Setting $\alpha_1 = t$ (where t is an unspecified number) and rewriting the first equation as

$$-\alpha_0 + \alpha_1 = 0 \tag{90}$$

we can easily compute α_0 as

$$\alpha_0 = t \tag{91}$$

This tells us that the eigenfunction that corresponds to $\lambda_1 = 1$ (which we will denote by ψ_A) has the general form

$$\psi_A = \alpha_0\psi_0 + \alpha_1\psi_1 = t\psi_0 + t\psi_1 \tag{92}$$

Since eigenfunctions need to be normalized in quantum mechanics, ψ_A must additionally satisfy

$$\langle \psi_A, \psi_A \rangle = \langle \alpha_0\psi_0 + \alpha_1\psi_1, \alpha_0\psi_0 + \alpha_1\psi_1 \rangle = 1 \tag{93}$$

If we expand this expression and recall that ψ_0 and ψ_1 are orthonormal, we obtain

$$\langle \psi_A, \psi_A \rangle = \alpha_0^*\alpha_0 \langle \psi_0, \psi_0 \rangle + \alpha_1^*\alpha_1 \langle \psi_1, \psi_1 \rangle = |\alpha_0|^2 + |\alpha_1|^2 = 1 \tag{94}$$

Observing that $\alpha_0 = \alpha_1 = t$ in this case, it follows that t should be chosen so that it satisfies

$$|\, t \,| = \frac{1}{\sqrt{2}} \tag{95}$$

We should note at this point that (95) *does not* specify parameter t uniquely, since it can take values such as $1/\sqrt{2}$, $-1/\sqrt{2}$, $i/\sqrt{2}$, etc. To keep things simple, we will set $t = 1/\sqrt{2}$, in which case ψ_A becomes

$$\psi_A = \frac{1}{\sqrt{2}}\psi_0 + \frac{1}{\sqrt{2}}\psi_1 \tag{96}$$

Proceeding in a similar manner, we can determine the eigenfunction that corresponds to $\lambda_2 = -1$. In this case, system (86) becomes

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{97}$$

and α_0 and α_1 have the form $\alpha_0 = -t$ and $\alpha_1 = t$. If we apply the normalization condition, we easily obtain

$$\psi_B = -\frac{1}{\sqrt{2}}\psi_0 + \frac{1}{\sqrt{2}}\psi_1 \tag{98}$$

Example 6. In this example we will consider the so-called *Hadamard operator*, which plays a key role in quantum computing. This operator (which is typically denoted by $\hat{H}$) transforms basis functions ψ_0 and ψ_1 as

$$\hat{H}\psi_0 = \frac{1}{\sqrt{2}}(\psi_0 + \psi_1) \tag{99}$$

and

$$\hat{H}\psi_1 = \frac{1}{\sqrt{2}}(\psi_0 - \psi_1) \tag{100}$$

Since the eigenfunctions of $\hat{H}$ must satisfy

$$\hat{H}\psi = \lambda\psi \tag{101}$$

for some number λ, we have that

$$\hat{H}\psi = \hat{H}(\alpha_0\psi_0 + \alpha_1\psi_1) = \alpha_0\hat{H}\psi_0 + \alpha_1\hat{H}\psi_1 = \lambda(\alpha_0\psi_0 + \alpha_1\psi_1) \tag{102}$$

Invoking (99) and (100), this becomes

$$\frac{\alpha_0}{\sqrt{2}}(\psi_0 + \psi_1) + \frac{\alpha_1}{\sqrt{2}}(\psi_0 - \psi_1) = \lambda(\alpha_0\psi_0 + \alpha_1\psi_1) \tag{103}$$

Grouping the terms next to ψ_0 and ψ_1, we obtain

$$\psi_0\left[\frac{\alpha_0 + \alpha_1}{\sqrt{2}} - \lambda\alpha_0\right] + \psi_1\left[\frac{\alpha_0 - \alpha_1}{\sqrt{2}} - \lambda\alpha_1\right] = 0 \tag{104}$$

Since ψ_0 and ψ_1 are linearly independent, condition (104) will be satisfied if and only if

$$\frac{\alpha_0 + \alpha_1}{\sqrt{2}} - \lambda\alpha_0 = 0 \tag{105}$$

and

$$\frac{\alpha_0 - \alpha_1}{\sqrt{2}} - \lambda\alpha_1 = 0 \tag{106}$$

If we express these two equations as

$$\begin{bmatrix} (1/\sqrt{2} - \lambda) & 1/\sqrt{2} \\ 1/\sqrt{2} & -(1/\sqrt{2} + \lambda) \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{107}$$

it become apparent that α_0 and α_1 can take nonzero values only if λ satisfies

$$\Delta(\lambda) = 2\lambda^2 - 2 = 0 \tag{108}$$

From this, we can conclude that the eigenvalues of $\hat{H}$ are $\lambda_1 = 1$ and $\lambda_2 = -1$.

The eigenfunction that corresponds to $\lambda_1 = 1$ can be found by substituting this value into (107) and solving it. If we do that, we obtain

$$\frac{1}{\sqrt{2}} \begin{bmatrix} (1-\sqrt{2}) & 1 \\ 1 & -(1+\sqrt{2}) \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{109}$$

which reduces to

$$\frac{1}{\sqrt{2}} \begin{bmatrix} -0.4142 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{110}$$

after one step of Gaussian elimination. Setting $\alpha_1 = t$, we can now easily compute α_0 as

$$\alpha_0 = 2.4142t \tag{111}$$

from the first equation.

Because eigenfunction

$$\psi_A = \alpha_0 \psi_0 + \alpha_1 \psi_1 \tag{112}$$

needs to be normalized, coefficients α_0 and α_1 must satisfy

$$|\alpha_0|^2 + |\alpha_1|^2 = 1 \tag{113}$$

Given the expressions that we obtained for α_0 and α_1, this condition is equivalent to

$$(2.4142)^2 |\, t\, |^2 + |\, t\, |^2 = 1 \tag{114}$$

Since $t = 0.38268$ is a solution of equation (114), we can conclude that

$$\psi_A = 0.92388\psi_0 + 0.38268\psi_1 \tag{115}$$

is an eigenfunction of $\hat{H}$ that corresponds to $\lambda_1 = 1$.

Proceeding in a similar manner, it is easily verified that the normalized eigenfunction that corresponds to $\lambda_2 = -1$ has the form

$$\psi_B = -0.38268\psi_0 + 0.92388\psi_1 \tag{116}$$

We will not derive this result explicitly, but it could be a useful exercise for those who want some additional practice.

Self-Adjoint and Unitary Operators

In order to distinguish between different types of linear operators, it will be useful to introduce the following two definitions.

Definition 4. Given a linear operator $\hat{A}$, its *adjoint* operator (denoted $\hat{A}^\dagger$) is an operator that satisfies

$$\langle \hat{A}f, g\rangle = \langle f, \hat{A}^\dagger g\rangle \tag{117}$$

for any pair of functions f and g that belong to the functional space.

Definition 5. Operator $\hat{A}$ is said to be *self-adjoint* (or *Hermitian*) if $\hat{A} = \hat{A}^\dagger$.

Self-adjoint operators have many similarities with symmetric matrices. The following theorem describes one of them (this particular property is very important in quantum mechanics).

Theorem 1. Let $\hat{A}$ be a self-adjoint operator with eigenvalues $\{\lambda_1, \lambda_2, \ldots\}$. Then, each λ_i must be a *real* number, and the corresponding eigenfunctions $\{\phi_1, \phi_2, \ldots\}$ must be *orthogonal*.

Another important class of operators are so-called *unitary operators*, which are defined in the following way.

Definition 6. Operator $\hat{U}$ is said to be unitary if $\hat{U}\hat{U}^\dagger = \hat{U}^\dagger\hat{U} = \hat{I}$, where $\hat{I}$ denotes the identity operator (this operator satisfies $\hat{I}f = f$, which means that it leaves function f unchanged).

This definition has two corollaries that are very useful in quantum computing. Since they are quite straightforward, we will prove both of them.

Corollary 1. If $\hat{U}$ is *both* unitary *and* self-adjoint, it satisfies $\hat{U} = \hat{U}^{-1}$.

Proof. The proof follows directly from the definition of a unitary operator, since

$$\hat{U}^\dagger\hat{U} = \hat{I} \tag{118}$$

implies that $\hat{U}^\dagger = \hat{U}^{-1}$. Given that $\hat{U}$ is assumed to be self-adjoint, we also have that $\hat{U}^\dagger = \hat{U}$, so we can conclude that $\hat{U} = \hat{U}^{-1}$. **Q.E.D.**

Corollary 2. All the eigenvalues of a unitary operator $\hat{U}$ must have the form $\lambda_k = e^{i\theta_k}$.

Proof. Suppose that $\hat{U}$ is a unitary operator whose eigenfunctions are $\{\xi_1, \xi_2, \ldots\}$. In that case

$$\langle \hat{U}\xi_k, \hat{U}\xi_k \rangle = \langle \lambda_k\xi_k, \lambda_k\xi_k \rangle = \lambda_k^*\lambda_k\langle \xi_k, \xi_k \rangle = |\lambda_k|^2\langle \xi_k, \xi_k \rangle \tag{119}$$

Since $\hat{U}$ is unitary, this scalar product can also be expressed as

$$\langle \hat{U}\xi_k, \hat{U}\xi_k \rangle = \langle \xi_k, \hat{U}^\dagger\hat{U}\xi_k \rangle = \langle \xi_k, \xi_k \rangle \tag{120}$$

Subtracting (120) from (119), we obtain

$$(|\lambda_k|^2 - 1)\langle \xi_k, \xi_k \rangle = 0 \tag{121}$$

which implies that $|\lambda_k| = 1$ (since $\langle \xi_k, \xi_k \rangle \neq 0$). Given that λ_k is a complex number in general, it follows that

$$\lambda_k = |\lambda_k|e^{i\theta_k} = e^{i\theta_k} \tag{122}$$

Q.E.D.

Matrix Representations of Operators

When operators have a finite number of eigenfunctions, it is often convenient to represent them in matrix form. To see how this works, suppose once again that S is a two dimensional functional space with an orthonormal basis $\{\psi_0, \psi_1\}$, and that $\hat{A}$ is a linear operator in this space. If $\hat{A}$ is applied to function

$$\psi = \alpha_0\psi_0 + \alpha_1\psi_1 \tag{123}$$

we obtain a different function

$$\varphi = \hat{A}\psi \tag{124}$$

whose representation in basis $\{\psi_0, \psi_1\}$ has the general form

$$\varphi = \rho_0\psi_0 + \rho_1\psi_1 \tag{125}$$

It is not difficult to see that operator $\hat{A}$ will be fully defined if we are able to compute coefficients ρ_0 and ρ_1 for any choice of α_0 and α_1. We can say this because ρ_0 and ρ_1 specify function φ uniquely, and α_0 and α_1 are known once we are given function ψ.

To see how this can be done, we should first observe that both $\hat{A}\psi_0$ and $\hat{A}\psi_1$ belong to space S. As a result, we can represent these functions as

$$\hat{A}\psi_0 = a_{11}\psi_0 + a_{12}\psi_1 \tag{126}$$

and

$$\hat{A}\psi_1 = a_{21}\psi_0 + a_{22}\psi_1 \tag{127}$$

This allows us to rewrite expression (124) as

$$\varphi = \hat{A}\psi = \alpha_0\hat{A}\psi_0 + \alpha_1\hat{A}\psi_1 = \alpha_0(a_{11}\psi_0 + a_{12}\psi_1) + \alpha_1(a_{21}\psi_0 + a_{22}\psi_1) \tag{128}$$

After grouping the terms on the right hand side, function φ can be rewritten as

$$\varphi = (\alpha_0 a_{11} + \alpha_1 a_{21})\psi_0 + (\alpha_0 a_{12} + \alpha_1 a_{22})\psi_1 \tag{129}$$

Subtracting (125) from (129), we now obtain

$$(\alpha_0 a_{11} + \alpha_1 a_{21} - \rho_0)\psi_0 + (\alpha_0 a_{12} + \alpha_1 a_{22} - \rho_1)\psi_1 = 0 \tag{130}$$

Recalling that functions ψ_0 and ψ_1 are orthonormal, this implies that

$$\rho_0 = \alpha_0 a_{11} + \alpha_1 a_{21} \tag{131}$$

and

$$\rho_1 = \alpha_0 a_{12} + \alpha_1 a_{22} \tag{132}$$

We can express these two relationships in matrix form as

$$\begin{bmatrix} \rho_0 \\ \rho_1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} \tag{133}$$

Since the coefficients of matrix

$$A = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix} \tag{134}$$

correspond to $\hat{A}\psi_0$ and $\hat{A}\psi_1$, it follows that matrix A is uniquely defined by operator $\hat{A}$ and basis $\{\psi_0, \psi_1\}$. We therefore refer to it as the *matrix representation of operator* $\hat{A}$ *in basis* $\{\psi_0, \psi_1\}$.

Once we know matrix A, we can easily compute coefficients ρ_0 and ρ_1, and therefore

$$\varphi = \hat{A}\psi = \rho_0\psi_0 + \rho_1\psi_1 \tag{135}$$

for *any* choice of function $\psi \in S$. The only information that we need for this purpose are the coefficients α_0 and α_1 that correspond to ψ in this basis. The following two examples illustrate how this idea can be applied to operators $\hat{X}$ and $\hat{H}$ (which we already considered in Examples 5 and 6).

Example 7. Suppose that operator $\hat{X}$ transforms function $\psi = \alpha_0\psi_0 + \alpha_1\psi_1$ into function

$$\varphi = \hat{X}\psi = \rho_0\psi_0 + \rho_1\psi_1 \tag{136}$$

In order to relate coefficients ρ_0 and ρ_1 to α_0 and α_1, we need to express $\hat{X}\psi_0$ and $\hat{X}\psi_1$ as

$$\hat{X}\psi_0 = a_{11}\psi_0 + a_{12}\psi_1 \tag{137}$$

and

$$\hat{X}\psi_1 = a_{21}\psi_0 + a_{22}\psi_1 \tag{138}$$

and compute coefficients a_{11}, a_{12}, a_{21} and a_{22}.

Recalling that operator $\hat{X}$ is defined by relations

$$\hat{X}\psi_0 = \psi_1 \tag{139}$$

and

$$\hat{X}\psi_1 = \psi_0 \tag{140}$$

it is easily verified that

$$\begin{aligned} a_{11} &= 0 \\ a_{12} &= 1 \\ a_{21} &= 1 \\ a_{22} &= 0 \end{aligned} \tag{141}$$

(these values are obtained by comparing (139) and (140) with (137) and (138)). From this, we can conclude that the matrix representation of $\hat{X}$ in basis $\{\psi_0, \psi_1\}$ is

$$X = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{142}$$

Example 8. Let us consider the Hadamard operator, which is defined by relations

$$\hat{H}\psi_0 = \frac{1}{\sqrt{2}}\psi_0 + \frac{1}{\sqrt{2}}\psi_1 \tag{143}$$

and

$$\hat{H}\psi_1 = \frac{1}{\sqrt{2}}\psi_0 - \frac{1}{\sqrt{2}}\psi_1 \tag{144}$$

As in the previous example, we will assume that this operator transforms the state $\psi = \alpha_0\psi_0 + \alpha_1\psi_1$ into

$$\varphi = \hat{H}\psi = \rho_0\psi_0 + \rho_1\psi_1 \tag{145}$$

To see how coefficients ρ_0 and ρ_1 are related to α_0 and α_1 in this case, let us express $\hat{H}\psi_0$ and $\hat{H}\psi_1$ as

$$\hat{H}\psi_0 = a_{11}\psi_0 + a_{12}\psi_1 \tag{146}$$

and

$$\hat{H}\psi_1 = a_{21}\psi_0 + a_{22}\psi_1 \tag{147}$$

Comparing (146) and (147) with (143) and (144) we obtain

$$\begin{aligned} a_{11} &= \frac{1}{\sqrt{2}} \\ a_{12} &= \frac{1}{\sqrt{2}} \\ a_{21} &= \frac{1}{\sqrt{2}} \\ a_{22} &= -\frac{1}{\sqrt{2}} \end{aligned} \tag{148}$$

which implies that the matrix representation of $\hat{H}$ in basis $\{\psi_0, \psi_1\}$ is

$$H = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{149}$$

If we compute the eigenvalues of matrices X and H, we will find that they match the eigenvalues of operators $\hat{X}$ and $\hat{H}$ (which we determined in Examples 5 and 6). This correlation points to a more general result, which can be stated as follows.

Theorem 2. Operator $\hat{A}$ and its matrix representation in basis $\{\psi_0, \psi_1\}$ have the *same* set of eigenvalues.

Coordinate Transformations

Suppose now that our functional space has *two* different orthonormal bases, $\{\psi_0, \psi_1\}$ and $\{\xi_0, \xi_1\}$. In that case, any function ψ that belongs to this space will have two different (but equivalent) representations

$$\psi = \alpha_0\psi_0 + \alpha_1\psi_1 \tag{150}$$

and

$$\psi = \beta_0 \xi_0 + \beta_1 \xi_1 \tag{151}$$

How can we relate these two representations?

Given that $\{\psi_0, \psi_1\}$ is a basis, we can represent ξ_0 and ξ_1 as

$$\xi_0 = c_{11}\psi_0 + c_{12}\psi_1 \tag{152}$$

and

$$\xi_1 = c_{21}\psi_0 + c_{22}\psi_1 \tag{153}$$

Substituting this into (151), we obtain

$$\psi = \beta_0\xi_0 + \beta_1\xi_1 = \beta_0(c_{11}\psi_0 + c_{12}\psi_1) + \beta_1(c_{21}\psi_0 + c_{22}\psi_1) \tag{154}$$

which becomes

$$\psi = (\beta_0 c_{11} + \beta_1 c_{21})\psi_0 + (\beta_0 c_{12} + \beta_1 c_{22})\psi_1 \tag{155}$$

when we rearrange the terms. If we now compare (150) and (155), it follows that coefficients α_0, α_1, β_0 and β_1 are related as

$$\alpha_0 = \beta_0 c_{11} + \beta_1 c_{21} \tag{156}$$

and

$$\alpha_1 = \beta_0 c_{12} + \beta_1 c_{22} \tag{157}$$

In matrix form, expressions (156) and (157) can be rewritten as

$$\begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} c_{11} & c_{21} \\ c_{12} & c_{22} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \tag{158}$$

and

$$T = \begin{bmatrix} c_{11} & c_{21} \\ c_{12} & c_{22} \end{bmatrix} \tag{159}$$

can be interpreted as the *transformation matrix* that relates representations of ψ in bases $\{\psi_0, \psi_1\}$ and $\{\xi_0, \xi_1\}$. Note that the elements of matrix T can be obtained directly from expressions (152) and (153) as

$$\begin{aligned} c_{11} &= \langle \xi_0, \psi_0 \rangle \\ c_{12} &= \langle \xi_0, \psi_1 \rangle \\ c_{21} &= \langle \xi_1, \psi_0 \rangle \\ c_{22} &= \langle \xi_1, \psi_1 \rangle \end{aligned} \tag{160}$$

Once we know this matrix, we can easily compute α_0 and α_1 given β_0 and β_1 (and vice versa).

Let us now consider a linear operator $\hat{A}$, whose matrix representation in basis $\{\psi_0, \psi_1\}$ is A. What would be the matrix representation of this operator in basis $\{\xi_0, \xi_1\}$? It is not difficult to show that this representation has the form

$$\tilde{A} = T^{-1}AT \tag{161}$$

where T is the matrix defined in (159). Since matrices $\tilde{A}$ and A are obviously similar, it follows that their eigenvalues are *exactly the same* (and also match the eigenvalues of operator $\hat{A}$, by virtue of Theorem 2).