5

Functions of Vectors and Matrices

We have already seen that the linear operator is a function on a vector space. The linear operator transforms a vector from one space into another space, $A:\Re^n \rightarrow \Re^m$, or possibly from a space into itself. Among this group of operators is a special class of operators known as *linear functionals*, which are simply linear operators that map vectors into the real numbers, $\Re$. Linear functionals serve some special purposes in vector spaces, and it is worth considering them in more detail here. In addition, three other classes of functions are discussed: the multilinear functional, the quadratic form, and matrix functions. Multilinear functionals, such as bilinear forms, also map vectors into the reals, while quadratic forms map vectors into reals, but are not linear. The more general matrix functions are functions, not necessarily linear, that map matrices (or matrix forms of operators) into other matrices. Matrix functions will be especially important in the solution of state equations as discussed in Chapter 6.

5.1 Linear Functionals

As mentioned above, the linear functional is simply a linear function of a vector:

> ***Linear Functional:*** A *linear functional* $f(\boldsymbol{x}) \in \Re$ maps a vector $\boldsymbol{x} \in \mathbf{X}$ into the set of real numbers such that:
>
> $$\begin{aligned} f(c\boldsymbol{x}) &= cf(\boldsymbol{x}) \\ f(\boldsymbol{x}+\boldsymbol{y}) &= f(\boldsymbol{x}) + f(\boldsymbol{y}) \end{aligned} \tag{5.1}$$
>
> where c is a scalar in the field of $\mathbf{X}$ and $\boldsymbol{y} \in \mathbf{X}$.

Consider that the n-dimensional vector space **X** is provided with a basis $\{\boldsymbol{e}_i\}$. Then

$$f(\boldsymbol{x}) = f(\sum_{i=1}^{n} x_i \boldsymbol{e}_i) = \sum_{i=1}^{n} x_i f(\boldsymbol{e}_i) \tag{5.2}$$

We can determine the effect of a linear functional on a vector by first determining its effect on the basis vectors. If we know the functional's effect on the basis vectors, we can use this result to operate on arbitrary vectors with little computation, which we have done in Section 3.1.2 with linear operators. We can think of the coefficients x_i as being the "components" of the functional in basis $\{\boldsymbol{e}_i\}$.

5.1.1 Changing the Basis of a Functional

In the event that the basis of a vector space changes and we wish to determine the effect of a functional f in the new basis $\{\hat{\boldsymbol{e}}_i\}$, we use the relationship between basis vectors as expressed in Section 2.2.4:

$$\boldsymbol{e}_j = \sum_{i=1}^{n} b_{ij} \hat{\boldsymbol{e}}_i \tag{5.3}$$

or in matrix form:

$$\left[\boldsymbol{e}_1 \;\vdots\; \boldsymbol{e}_2 \;\vdots\; \cdots \;\vdots\; \boldsymbol{e}_n \right] = \left[\hat{\boldsymbol{e}}_1 \;\vdots\; \hat{\boldsymbol{e}}_2 \;\vdots\; \cdots \;\vdots\; \hat{\boldsymbol{e}}_n \right] B \tag{5.4}$$

Then

$$\begin{aligned} f(\boldsymbol{x}) &= \sum_{j=1}^{n} x_j f(\sum_{i=1}^{n} b_{ij} \hat{\boldsymbol{e}}_i) \\ &= \sum_{j=1}^{n} x_j \sum_{i=1}^{n} b_{ij} f(\hat{\boldsymbol{e}}_i) \\ &= \sum_{i=1}^{n} \left(\sum_{j=1}^{n} b_{ij} x_j \right) f(\hat{\boldsymbol{e}}_i) \\ &\stackrel{\Delta}{=} \sum_{i=1}^{n} \hat{x}_i f(\hat{\boldsymbol{e}}_i) \end{aligned} \tag{5.5}$$

where the $\hat{x}_i$'s are considered the component of the functional in the new basis. By equating the third and fourth lines of (5.5), it can be seen that

$$\hat{x}_i = \sum_{j=1}^{n} b_{ij} x_j$$

or in vector-matrix form:

$$\hat{\boldsymbol{x}} = B\boldsymbol{x} \tag{5.6}$$

which is the same result as Equation 2.33. This gives us the formula for changing the basis of the components.

5.2 *Multilinear Functionals*

As might be guessed from the name, a multilinear functional is a functional that is linear in a number of different vectors, which would be expressed as $A(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k)$. The most common form of the multilinear functional is the so-called "bilinear form." (The term *form* is generally used in finite-dimensional spaces, whereas the term *functional* is sometimes restricted to infinite-dimensional spaces such as function spaces.) We will discuss only these bilinear forms here. Other types of multilinear functionals can be found in books on tensors [1]. The bilinear form is a functional that is linear in two different vectors, according to the following definition:

> ***Bilinear Form:*** A *bilinear form B* is a functional that acts on *two* vectors from space **X** in such a way that if $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{z}$ are vectors in **X**, and α is a scalar in the field of **X**, then

$$\begin{aligned} B(\boldsymbol{x}+\boldsymbol{y},\boldsymbol{z}) &= B(\boldsymbol{x},\boldsymbol{z}) + B(\boldsymbol{y},\boldsymbol{z}) \\ B(\boldsymbol{x},\boldsymbol{z}+\boldsymbol{y}) &= B(\boldsymbol{x},\boldsymbol{z}) + B(\boldsymbol{x},\boldsymbol{y}) \\ B(\boldsymbol{x},\alpha\boldsymbol{y}) &= \alpha B(\boldsymbol{x},\boldsymbol{y}) \\ B(\alpha\boldsymbol{x},\boldsymbol{y}) &= \alpha B(\boldsymbol{x},\boldsymbol{y}) \end{aligned} \tag{5.7}$$

True to its name, the bilinear form $B(\boldsymbol{x},\boldsymbol{y})$ is linear in $\boldsymbol{x}$ and in $\boldsymbol{y}$. If the last condition is replaced by the condition

$$B(\alpha\boldsymbol{x},\boldsymbol{y}) = \bar{\alpha} B(\boldsymbol{x},\boldsymbol{y}) \tag{5.8}$$

then the form B is known strictly as a *sesquilinear form*, but this distinction is often neglected.

In an n-dimensional linear vector space, bilinear forms can be specified with a set of n^2 components, because each vector $\boldsymbol{x}$ and $\boldsymbol{y}$ can be expanded into, at most, n independent components. These n^2 numbers may be written as an $n \times n$ matrix, where the $(i, j)^{th}$ element of this matrix is:

$$b_{ij} = B(\boldsymbol{e}_i, \boldsymbol{e}_j) \tag{5.9}$$

so the bilinear form itself may be written as:

$$B(\boldsymbol{x}, \boldsymbol{y}) = \bar{\boldsymbol{x}}^{\mathrm{T}} \left[b_{ij} \right] \boldsymbol{y} \overset{\Delta}{=} \bar{\boldsymbol{x}}^{\mathrm{T}} B \boldsymbol{y} \tag{5.10}$$

Such a bilinear form is said to be *symmetric* if $B(\boldsymbol{x}, \boldsymbol{y}) = B(\boldsymbol{y}, \boldsymbol{x})$, and hermitian if $B(\boldsymbol{x}, \boldsymbol{y}) = \overline{B(\boldsymbol{y}, \boldsymbol{x})}$, in which case it can be shown that the matrix representing the form in (5.10) is symmetric (hermitian) in the usual matrix sense.

5.2.1 Changing the Basis of a Bilinear Form

If the basis of the vector space is changed, we would expect that the matrix representing the bilinear form will change as well. To determine how it changes denote by B the matrix for the bilinear form in the original basis $\{\boldsymbol{e}_i\}$, and let $\hat{B}$ denote the form in a new basis $\{\hat{\boldsymbol{e}}_j\}$. Assume that the relationship between the old basis and the new basis is known, as it was in Chapter 2 [see Equation (2.28)]:

$$\boldsymbol{e}_i = \sum_{k=1}^{n} \beta_{ki} \hat{\boldsymbol{e}}_k \tag{5.11}$$

(Here we are using the symbol β_{ki} to avoid confusion with the elements of the bilinear form b.) Substituting this into the expression for the $(i, j)^{th}$ element of the matrix representing B, i.e., Equation (5.9), we obtain:

$$\begin{aligned} b_{ij} &= B(\boldsymbol{e}_i, \boldsymbol{e}_j) \\ &= B\left(\sum_{l=1}^{n} \beta_{li} \hat{\boldsymbol{e}}_l, \sum_{m=1}^{n} \beta_{mj} \hat{\boldsymbol{e}}_m \right) \end{aligned}$$

$$= \sum_{l=1}^{n} \beta_{li} \sum_{m=1}^{n} \beta_{mj} B(\hat{\boldsymbol{e}}_l, \hat{\boldsymbol{e}}_m)$$
$$\stackrel{\Delta}{=} \sum_{l=1}^{n} \beta_{li} \sum_{m=1}^{n} \beta_{mj} \hat{b}_{lm}$$

One can see that if the coefficients β_{ij} are arranged in a matrix β, then the summations above result in the relationship

$$B = \beta^{\mathrm{T}} \hat{B} \beta \tag{5.12}$$

which is somewhat different, yet similar to the similarity transformation that performs a change of basis on the matrix of a linear operator, as in Equation (3.24). Note, though, that if the change of basis matrix β is orthonormal, so that its inverse is equal to its transpose, the transformations are performed with the same operations.

5.2.2 Bilinear Forms as Inner Products

The bilinear form written as (5.10) for any *n*-dimensional space can also be written as $B(\boldsymbol{x}, \boldsymbol{y}) = \langle \boldsymbol{x}, B\boldsymbol{y} \rangle$. This suggests that the bilinear forms and inner products are equivalent. However, the bilinear form cannot be expressed as the matrix product $B(\boldsymbol{x}, \boldsymbol{y}) = \bar{\boldsymbol{x}}^{\mathrm{T}} B\boldsymbol{y}$ for function spaces, which are infinite dimensional. For example, consider the space of real-valued functions defined on the interval $t \in [0,1]$. A valid bilinear form for this space is

$$B(\boldsymbol{f}(t), \boldsymbol{g}(t)) = \int_0^1 \boldsymbol{f}(t)\boldsymbol{g}(t)w(t)\, dt \tag{5.13}$$

One can note that this bilinear form also appears to be a valid inner product on the space. This is generally true with one condition: inner products are bilinear forms defined on real vector spaces provided that $B(\boldsymbol{x}, \boldsymbol{x}) > 0$ holds for $\boldsymbol{x} \neq 0$ and $B(\boldsymbol{x}, \boldsymbol{x}) = 0$ holds for $\boldsymbol{x} = 0$. This condition is discussed in more detail in the next section.

If this were a space of complex-valued functions, then the inner product would be a sesquilinear form:

$$B(\boldsymbol{f}(t), \boldsymbol{g}(t)) = \int_0^1 \bar{\boldsymbol{f}}(t)\boldsymbol{g}(t)w(t)\, dt \tag{5.14}$$

In (5.13) and (5.10), the function $w(t)$ is known as a *weighting function*. We have used $w(t)=1$ in past examples of inner products, but it may be introduced into the form (inner product) for special purposes, such as to make an operator self-adjoint by changing the way the inner product is computed.

5.3 Quadratic Forms

We can define quadratic forms in the same way that we defined bilinear forms except that we shall use the same vector twice. For example, given a matrix A, a *quadratic* form for a finite-dimensional (complex-valued) vector may be expressed as

$$A(\boldsymbol{x},\boldsymbol{x}) = \bar{\boldsymbol{x}}^{\mathrm{T}} A\boldsymbol{x} \tag{5.15}$$

or

$$A(\boldsymbol{x},\boldsymbol{x}) = \langle \boldsymbol{x}, A\boldsymbol{x}\rangle \tag{5.16}$$

In most cases, it is convenient to speak only of symmetric or hermitian quadratic forms. This is because if $\bar{\boldsymbol{x}}^{\mathrm{T}} A\boldsymbol{x}$ is a scalar, then

$$\bar{\boldsymbol{x}}^{\mathrm{T}} A\boldsymbol{x} = \overline{\left(\bar{\boldsymbol{x}}^{\mathrm{T}} A\boldsymbol{x}\right)}^{\mathrm{T}} = \bar{\boldsymbol{x}}^{\mathrm{T}} \bar{A}^{\mathrm{T}} \boldsymbol{x}$$

Therefore,

$$\bar{\boldsymbol{x}}^{\mathrm{T}} A\boldsymbol{x} = \frac{1}{2}\left(\bar{\boldsymbol{x}}^{\mathrm{T}} A\boldsymbol{x} + \bar{\boldsymbol{x}}^{\mathrm{T}}\bar{A}^{\mathrm{T}}\boldsymbol{x}\right) = \bar{\boldsymbol{x}}^{\mathrm{T}}\left(\frac{A+\bar{A}^{\mathrm{T}}}{2}\right)\boldsymbol{x} \tag{5.17}$$

which is, by inspection, hermitian. Thus, when we speak of quadratic forms, we will always assume that we are working with a symmetric matrix.

We will be using quadratic forms as "weighted" squared norms; this approach is similar to the way we treated bilinear forms as inner products. However, there is a restriction on the use of $\langle \boldsymbol{x}, A\boldsymbol{x}\rangle^{1/2}$ as a norm because any norm we define must satisfy the conditions set forth in Chapter 2, including the requirement that they be nonnegative, and zero only when $\boldsymbol{x}=0$. For a quadratic form in general, this may not be the case. In fact, we may have one of five possibilities:

1. If $\bar{\boldsymbol{x}}^{\mathrm{T}} A\boldsymbol{x} > 0$ for all $\boldsymbol{x} \neq 0$, the (symmetric) quadratic form (5.15) is said to be *positive definite*.

2. If $\bar{\boldsymbol{x}}^{\mathrm{T}} A\boldsymbol{x} \geq 0$ for all $\boldsymbol{x} \neq 0$, the (symmetric) quadratic form (5.15) is said to be *positive semidefinite*.
3. If $\bar{\boldsymbol{x}}^{\mathrm{T}} A\boldsymbol{x} < 0$ for all $\boldsymbol{x} \neq 0$, the (symmetric) quadratic form (5.15) is said to be *negative definite*.
4. If $\bar{\boldsymbol{x}}^{\mathrm{T}} A\boldsymbol{x} \leq 0$ for all $\boldsymbol{x} \neq 0$, the (symmetric) quadratic form (5.15) is said to be *negative semidefinite*.
5. If $\bar{\boldsymbol{x}}^{\mathrm{T}} A\boldsymbol{x} > 0$ for some $\boldsymbol{x} \neq 0$ and $\bar{\boldsymbol{x}}^{\mathrm{T}} A\boldsymbol{x} < 0$ for other $\boldsymbol{x} \neq 0$, the (symmetric) quadratic form (5.15) is said to be *indefinite*.

These same terms are often applied to the matrix A itself, not just the quadratic form it defines.

5.3.1 Testing for Definiteness

There are a number of tests to determine whether a matrix it is positive definite, negative definite, etc. However, we will consider only the most commonly used test, which is given by the following theorem:

> THEOREM: A quadratic form (or equivalently, the symmetric matrix A defining it) is *positive (negative) definite* if *all* of the eigenvalues of A are positive (negative) or have positive (negative) real parts. The quadratic form is *positive (negative) semidefinite* if all of the eigenvalues of A are nonnegative (nonpositive) or have nonnegative (nonpositive) real parts.(5.18)

> PROOF: Given that we are working with a symmetric matrix, we can use the result of the previous chapter to realize that we can compute an orthonormal basis of the eigenvectors of A. This gives us the orthonormal modal matrix M with which we can change variables, $\boldsymbol{x} = M\hat{\boldsymbol{x}}$. Substituting this into (5.15),

$$
\begin{aligned}
\bar{\boldsymbol{x}}^{\mathrm{T}} A\boldsymbol{x} &= \bar{\hat{\boldsymbol{x}}}^{\mathrm{T}} M^{\mathrm{T}} AM\, \hat{\boldsymbol{x}} \\
&= \bar{\hat{\boldsymbol{x}}}^{\mathrm{T}} diag\{\lambda_1, \ldots, \lambda_n\} \hat{\boldsymbol{x}} \\
&= \sum_{i=1}^{n} |x_i|^2 \lambda_i
\end{aligned}
$$

> where $diag\{\ \}$ is a matrix with the elements of the set argument on the diagonal and zeros elsewhere. It is apparent from this expression that for $\bar{\boldsymbol{x}}^{\mathrm{T}} A\boldsymbol{x}$ to be positive (nonnegative) for *any* $\boldsymbol{x} \neq 0$, then $\lambda_i > 0$ $(\lambda_i \geq 0)$ for each i. Similarly, for this

expression to be negative (nonpositive) for *any* $\mathbf{x} \neq 0$, then $\lambda_i < 0$ $(\lambda_i \leq 0)$ for each *i*.

While this theorem does not cover the indefinite case, it should be clear that an indefinite quadratic form is one whose matrix A has some positive eigenvalues and some negative eigenvalues.

There is another common test for definiteness that is often used for small matrices and is based on the *principal minors*. *Minors* of a matrix are determinants formed by striking out an equal number of rows and columns from a square matrix. *Principal minors* of a matrix are the minors whose diagonal elements are also diagonals of the original matrix. *Leading principal minors* are principal minors that contain *all* the diagonal elements of the original matrix, up to *n*, the size of the minor. For example, in Figure 5.1 below we indicate all the leading principal minors of an arbitrary matrix.

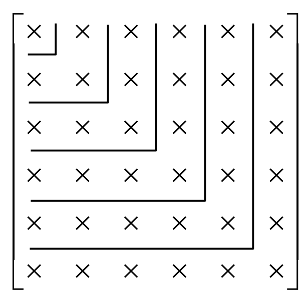

Figure 5.1 Leading principal minors are formed by making determinants of the leading square sub-matrices of the original matrix.

With this established, we can test a matrix for positive definiteness by establishing that all the leading principal minors are positive. The matrix will be positive semidefinite if all the leading principal minors are nonnegative. The matrix will be negative definite if the leading principal minors alternate in sign, beginning with a negative term as we start with the 1×1 case. (How would you state the rule for negative semidefiniteness?)

5.3.2 Changing the Basis of a Quadratic Form

Analogous to the basis change for bilinear forms and entirely consistent with the basis change performed in the example above, we can change the matrix for a quadratic form when the basis of the underlying space is changed. The derivation is the same as in Section 5.2.1 and provides the same result. If A is the matrix for a quadratic form on a space using basis $\{\mathbf{e}_i\}$ and $\hat{A}$ is the matrix for the same quadratic form on the same space, but using $\{\hat{\mathbf{e}}_i\}$ for the basis, then the two forms are related by

$$A = \beta^{\mathrm{T}} \hat{A} \beta$$

where β is the change of basis matrix denoted by B in Equation (2.33).

5.3.3 Geometry of Quadratic Forms

An interesting property of quadratic forms is that they may be efficiently used to represent the equations of geometric objects known as *quadrics*. An example already seen in Chapter 4 is the ellipsoid. If $\boldsymbol{x} \in \Re^n$ and the $n \times n$ matrix A is positive definite and symmetric, then

$$\boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x} = 1 \tag{5.19}$$

is the equation of an ellipsoid in $\Re^n$. If A is not positive definite, then other geometric possibilities arise, such as conics and hyperboloids, which are also quadrics.

To get a graphical image of the ellipsoid, we must draw the surface of the equation $\boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x} = 1$. In order to see the relative dimensions of the ellipsoid and not scale the entire ellipsoid, we normalize its size by considering only the values of $\boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x}$ such that $\boldsymbol{x}^{\mathrm{T}} \boldsymbol{x} = 1$. Now to get a picture of what the ellipsoid looks like, i.e., the directions of its principal axes and how far the surface of the ellipsoid extends in each of those directions, we will find the minimum and maximum distance of the surface of the ellipsoid from the origin. We will thus extremize $\boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x}$ subject to $\boldsymbol{x}^{\mathrm{T}} \boldsymbol{x} = 1$.

We solve this optimization problem in a manner similar to the underdetermined system problem of Chapter 3. Introducing a scalar LaGrange multiplier γ, we construct the hamiltonian

$$H = \boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x} + \gamma\left(1 - \boldsymbol{x}^{\mathrm{T}} \boldsymbol{x}\right) \tag{5.20}$$

Now taking the appropriate derivatives,

$$\frac{\partial^{\mathrm{T}} H}{\partial \boldsymbol{x}} = 2A\boldsymbol{x} - 2\gamma \boldsymbol{x} = 0 \tag{5.21}$$

and

$$\frac{\partial H}{\partial \gamma} = 1 - \boldsymbol{x}^{\mathrm{T}} \boldsymbol{x} = 0 \tag{5.22}$$

Simplifying (5.21), we arrive at

$$(A-\gamma I)\boldsymbol{x} = 0 \tag{5.23}$$

which implies that the LaGrange multiplier is equal to an eigenvalue and $\boldsymbol{x}$ is the corresponding eigenvector (recall that if A is positive definite, there will be n of these). Thus the extrema of the ellipsoid occur at the eigenvectors. [Equation (5.22) merely reinforces the normalization of the size of the ellipsoid.] If indeed $\boldsymbol{x}$ is a normalized eigenvector corresponding to eigenvalue γ, then at this point,

$$\begin{aligned} \boldsymbol{x}^{\mathrm{T}} A\boldsymbol{x} &= \boldsymbol{x}^{\mathrm{T}}\gamma\boldsymbol{x} \\ &= \gamma\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} \\ &= \gamma\|\boldsymbol{x}\|^2 \\ &= \gamma \end{aligned}$$

This means that the extremal value taken by the ellipsoid at this point is the eigenvalue γ. This will be true for the distance from the origin to the surface of the ellipsoid along any of the principal axes. Note that this result is true for a symmetric, positive definite matrix only; for a nonsymmetric matrix the principal axes are the left singular vectors and the axis lengths are the singular values.

The example below gives an application in which the computation of an ellipsoid is useful for analysis of physical systems. The subsequent diagram (Figure 5.2) then illustrates the geometry of the ellipse and the related eigenvalues and eigenvectors of matrix A.

Example 5.1: Statistical Error Analysis

Ellipsoids such as those depicted in Figure 4.5 in Example 4.13 are useful graphical representations for quantities that have different magnitudes in different directions. In that example, the ellipses represented the extent to which a robot is able to produce motion in specific directions of its workspace. A similar representation is commonly used for statistical error analysis wherein the error expected in a computed quantity may vary depending on the source of that error.

For example, consider that in a system with several sensors (such as the three-link robot example) wherein the joint angle sensors are noisy devices, each producing a measurement written as

$$\hat{\theta} = \theta + \delta\theta \tag{5.24}$$

where θ is the true value of the angle and $\delta\theta$ is an error term. If we assume that the error terms for the angle sensors are uncorrelated to each other, and if these errors are zero-mean, then we can define the covariance matrix of the vector $\Theta = \begin{bmatrix} \theta_1 & \theta_2 & \cdots & \theta_n \end{bmatrix}^{\mathrm{T}}$ of angle measurements as:

$$\begin{aligned} Q &\stackrel{\Delta}{=} \mathrm{E}\left[\left(\delta\Theta - \mathrm{E}[\delta\Theta] \right) \left(\delta\Theta - \mathrm{E}[\delta\Theta] \right)^{\mathrm{T}} \right] \\ &= \mathrm{E}\left[\delta\Theta \cdot \delta\Theta^{\mathrm{T}} \right] \\ &= diag\left(\nu_1^2, \ldots, \nu_n^2 \right) \end{aligned} \tag{5.25}$$

where the ν_i^2 terms are the variances of the *n* individual sensor noises, $\delta\Theta$ is a vector of angle errors as in Equation (5.24), and E is the expected value operator.

Using these noisy angle measurements to determine a different physical quantity, such as the tip of the third link of the robot, Equation (4.44) would give

$$\begin{aligned} \mathrm{X} &= \boldsymbol{f}(\hat{\Theta}) \\ &= \boldsymbol{f}(\Theta + \delta\Theta) \end{aligned} \tag{5.26}$$

Assuming the quantities $\delta\theta_i$ are small, we can expand (5.26) into its Taylor series equivalent to give

$$\mathrm{X} \approx \boldsymbol{f}(\Theta) + J(\Theta)\delta\Theta \tag{5.27}$$

where the jacobian $J(\Theta)$ is defined as $J(\Theta) \stackrel{\Delta}{=} \partial \boldsymbol{f}(\Theta) / \partial \Theta$. If we take the expected value of X to be $\mathrm{E}[\mathrm{X}] = \mathrm{E}[\boldsymbol{f}(\hat{\Theta})] = \boldsymbol{f}(\Theta)$, then the covariance of the computed endpoint location is (see [6], [8], and [10]):

$$\begin{aligned} R \stackrel{\Delta}{=} \mathrm{cov}[\mathrm{X}] &\approx \mathrm{E}\left[(\mathrm{X} - \mathrm{E}[\mathrm{X}])(\mathrm{X} - \mathrm{E}[\mathrm{X}])^{\mathrm{T}} \right] \\ &= \mathrm{E}\left[\left(\boldsymbol{f}(\hat{\Theta}) - \boldsymbol{f}(\Theta) \right) \left(\boldsymbol{f}(\hat{\Theta}) - \boldsymbol{f}(\Theta) \right)^{\mathrm{T}} \right] \\ &= \mathrm{E}\left[(J(\Theta)\delta\Theta)(J(\Theta)\delta\Theta)^{\mathrm{T}} \right] \\ &= \mathrm{E}\left[J(\Theta)\delta\Theta\delta\Theta^{\mathrm{T}} J(\Theta)^{\mathrm{T}} \right] \\ &= J(\Theta)\mathrm{E}\left[\delta\Theta\delta\Theta^{\mathrm{T}} \right] J^{\mathrm{T}}(\Theta) \\ &= J(\Theta) Q J^{\mathrm{T}}(\Theta) \end{aligned} \tag{5.28}$$

By definition, this covariance matrix R will be positive definite because it is assumed that there is at least some noise in each sensor.

REMARK: Similarly, an analogous analysis may be performed for the robot that indicates to what extent an error in the motor positions affects the endpoint position. This sort of sensitivity analysis is often critical in engineering design wherein it is desired that the inevitably noisy physical devices must produce minimal effect on the desired outcome.

Suppose

$$R = \begin{bmatrix} 3.25 & 1.3 \\ 1.3 & 1.75 \end{bmatrix}$$

The eigenvalues for this matrix are $\lambda_1 = 4$ and $\lambda_1 = 1$, so it is clearly positive definite. Corresponding to these eigenvalues are the normalized eigenvectors, arranged into the modal matrix:

$$M = \left[\boldsymbol{e}_1 \;\vdots\; \boldsymbol{e}_2 \right] = \begin{bmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{bmatrix}$$

Therefore, the ellipsoid (in two dimensions) given by the equation $\boldsymbol{x}^{\mathrm{T}} R \boldsymbol{x} = 1$ appears in Figure 5.2. We can interpret this diagram by saying that we expect that the error variance is about four times higher in the $\boldsymbol{e}_1$ direction than in the $\boldsymbol{e}_2$ direction, which is to say that the sensor noise affects measurements in one direction more so than in the other direction. If in fact the ellipsoid for the robot example is this eccentric (or more so), we could conclude that the robot is more precise in one direction than in another, a phenomenon that occurs frequently in realistic systems.

Example 5.2: Conic Sections

It can be easily shown that the equation for the general quadric shape:

$$ax^2 + dy^2 + fz^2 + 2bxy + 2cxz + 2eyz + 2jx + 2ky + 2lz + q = 0 \quad (5.29)$$

can be expressed in matrix-vector form assuming physical coordinates are arranged into the vector $X \stackrel{\Delta}{=} \begin{bmatrix} x & y & z \end{bmatrix}^{\mathrm{T}}$ as:

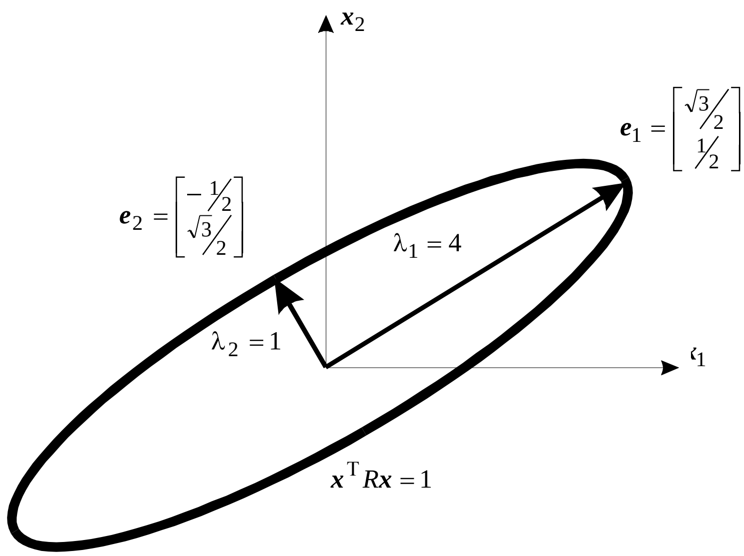

Figure 5.2 Graphical representation of the significance of the positive definite covariance matrix R in the robot example.

$$
\begin{aligned}
0 &= X^{T} \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix} X + 2\begin{bmatrix} j & k & l \end{bmatrix} X + q \\
&\triangleq X^{T} A X + 2 B X + q
\end{aligned}
\tag{5.30}
$$

or with $Y \triangleq \begin{bmatrix} x & y & z & 1 \end{bmatrix}^{T}$ as:

$$
\begin{aligned}
0 &= Y^{T} \begin{bmatrix} a & b & c & j \\ b & d & e & k \\ c & e & f & l \\ j & k & l & q \end{bmatrix} Y \\
&\triangleq Y^{T} \Lambda Y
\end{aligned}
\tag{5.31}
$$

Give conditions on the matrices A, B, and Λ, and the value of the scalar q such that Equations (5.30) and (5.31) represent various conic sections.

Solution:

It is readily apparent that if matrix A is diagonal, $B = \mathbf{0}$, and $q < 0$, then Equation (5.30) represents an ellipsoid whose axes are aligned with the coordinate axes x, y, and z. If in addition one of the values of a, d, or f is zero, the result is an elliptic cylinder. If $A \equiv 0$, Equation (5.30) represents either a plane or a line, depending on whether one or two of j, k, and l are zero.

However, many other conic sections can be represented with the matrix expressions in (5.30) and (5.31). Among these are:

1. If $rank(A) = 3$, $rank(\Lambda) = 4$, the eigenvalues of A all have the same sign, and $\det(\Lambda) > 0$, then the resulting shape is an ellipsoid that has been rotated so that its major axes are the eigenvectors of A and the values of B represent an offset from the origin. (Consider how to make the equations represent a sphere.)
2. If $rank(A) = 3$, $rank(\Lambda) = 4$, the eigenvalues of A have different signs, and $\det(\Lambda) > 0$, then the resulting shape is a hyperboloid of one sheet.
3. If $rank(A) = 3$, $rank(\Lambda) = 4$, the eigenvalues of A have different signs, and $\det(\Lambda) < 0$, then the resulting shape is a hyperboloid of two sheets.
4. If $rank(A) = 3$, $rank(\Lambda) = 3$, and the eigenvalues of A have different signs, the resulting shape is a quadric cone.
5. If $rank(A) = 2$, $rank(\Lambda) = 3$, and the eigenvalues of A have the same sign, the resulting shape is an elliptic cylinder (that is generally tilted in space).

There are numerous other possibilities, including "imaginary" shapes such as $x^2 + y^2 + z^2 = -1$.

5.4 Functions of Matrices

In this section we turn our attention to matrix functions that return matrix results. We have already seen simple examples, such as the addition of two operators (matrices), which is performed element-by-element, and the multiplication of two operators, which is *not* performed element-by-element. Multiplication of a matrix by itself leads to the further definitions:

$$\begin{aligned} AA &= A^2 \\ A\cdots A \ \ (n \text{ times}) &= A^n \\ \left(A^n\right)^m &= A^{nm} \\ A^n A^m &= A^{(n+m)} \\ \left(A^{-1}\right)^n &= A^{-n} \end{aligned} \tag{5.32}$$

(The last line in this list is not always accepted as proper notation.) By adding and subtracting scalar multiples of powers of *A*, we can generate polynomials in *A*. As a matter of notation, if we can define an n^{th}-order (monic) polynomial

$$\begin{aligned} p(\lambda) &= \lambda^n + \alpha_{n-1}\lambda^{n-1} + \cdots + \alpha_1\lambda^1 + \alpha_0 \\ &= (\lambda - p_1)(\lambda - p_2)\cdots(\lambda - p_n) \end{aligned} \tag{5.33}$$

then that same polynomial evaluated on the matrix *A* is[M]: `polyvalm(p,A)`

$$\begin{aligned} p(A) &= A^n + \alpha_{n-1}A^{n-1} + \cdots + \alpha_1 A^1 + \alpha_0 I \\ &= (A - p_1 I)(A - p_2 I)\cdots(A - p_n I) \end{aligned} \tag{5.34}$$

5.4.1 Cayley-Hamilton Theorem

Polynomials evaluated on matrices prove to have the following interesting property:

> THEOREM (Cayley-Hamilton theorem): If the characteristic polynomial of an arbitrary $n \times n$ matrix *A* is denoted by $\phi(\lambda)$, computed as $\phi(\lambda) = |A - \lambda I|$, then *A* satisfies its own characteristic equation, denoted by $\phi(A) = 0$. (5.35)

As a result of this theorem, we will never have reason to compute matrix powers higher than *n* because if

$$\phi(A) = A^n + \phi_{n-1}A^{n-1} + \cdots + \phi_1 A + \phi_0 I = 0$$

then

$$A^n = -\phi_{n-1}A^{n-1} - \cdots - \phi_1 A - \phi_0 I \tag{5.36}$$

Example 5.3: Application of the Cayley-Hamilton Theorem to Matrix Inverses

Let A be an $n \times n$ nonsingular matrix. Use the Cayley-Hamilton theorem to express the inverse A^{-1} without using powers of A higher than $n-1$.

Solution:

Let the characteristic equation for A be written as

$$\begin{aligned} \phi(\lambda) &= \lambda^n + \phi_{n-1}\lambda^{n-1} + \cdots + \phi_1\lambda + \phi_0 = 0 \\ &= \prod_{i=1}^{n} (\lambda - \lambda_i) \end{aligned} \tag{5.37}$$

Then because A is nonsingular, we know it cannot have any zero eigenvalues. Consideration of the polynomial expansion of (5.37) reveals that the coefficient ϕ_0 in the characteristic equation in (5.37) will be the product of all eigenvalues, which is therefore nonzero as well. Because matrix A satisfies its own characteristic equation,

$$A^n + \phi_{n-1}A^{n-1} + \cdots + \phi_1 A + \phi_0 I = 0 \tag{5.38}$$

can be solved by division by the nonzero ϕ_0:

$$I = -\frac{1}{\phi_0}\left(\phi_1 A + \phi_2 A^2 + \cdots + \phi_{n-1}A^{n-1} + A^n\right)$$

Multiplying both sides through by A^{-1} we get

$$A^{-1} = -\frac{1}{\phi_0}\left(\phi_1 I + \phi_2 A^1 + \cdots + \phi_{n-1}A^{n-2} + A^{n-1}\right)$$

We will want to produce functions of A of more generality, not limited to polynomials or inverses. Yet, we will want to retain the simplicity of matrix powers for computational purposes. For this we must establish the power series defined on a matrix. Recall that scalar functions are said to be *analytic* if and only if they can (at least locally) be represented by a convergent power series. The Taylor series is such a power series. The Taylor series for an analytic function $f(\lambda)$ is

$$f(\lambda)=\sum_{i=0}^{\infty}\frac{1}{i!}\left.\frac{d^i f}{d\lambda^i}\right|_{\lambda=0}\lambda^i \tag{5.39}$$

We can thus define the same function acting on a matrix *A* as

$$f(A)=\sum_{i=0}^{\infty}\frac{1}{i!}\left.\frac{d^i f}{d\lambda^i}\right|_{\lambda=0}A^i \tag{5.40}$$

Some functions whose Taylor series expansions are familiar can be computed with matrix arguments quite easily, e.g.,

$$\begin{aligned}\sin(A) &= A-\frac{A^3}{3!}+\frac{A^5}{5!}-\frac{A^7}{7!}+\cdots\\ \cos(A) &= I-\frac{A^2}{2!}+\frac{A^4}{4!}-\frac{A^6}{6!}+\cdots\\ e^{At} &= I+At+\frac{A^2t^2}{2!}+\frac{A^3t^3}{3!}+\cdots\end{aligned} \tag{5.41}$$

5.4.2 Using the Cayley-Hamilton Theorem to Compute Matrix Functions

The power series expansions in Equation (5.41), while convergent, are unsuitable for use in applications where closed-form functions are necessary. The best one could do with these expansions is to use an approximation obtained by truncating the series after some minimal number of terms. However, the Cayley-Hamilton theorem suggests a useful simplification. Because, as the theorem states, the matrix satisfies its own characteristic polynomial, $\phi(A)=0$, we can always express the n^{th} power of a matrix by a sum of its lower powers, 0 through $n-1$, as in Equation (5.38). This implies that if we express any analytic function of matrix *A* as its Taylor series, all powers of *A* higher than $n-1$ will be reducible to a sum of powers *up to* $n-1$. Then because the Taylor series converges, any analytic function $f(A)$ may be written

$$f(A)=\alpha_{n-1}A^{n-1}+\cdots+\alpha_1 A+\alpha_0 I \tag{5.42}$$

provided we can find the appropriate values of the *n* constant coefficients α_i, $i=1,\ldots,n$. To assist in the computation of these constants, the following theorem is helpful:

THEOREM: Suppose $f(\lambda)$ is an arbitrary function of scalar variable λ and $g(\lambda)$ is an $(n-1)^{st}$ order polynomial in λ. If $f(\lambda_i) = g(\lambda_i)$ for every eigenvalue λ_i of the $n \times n$ matrix *A*, then $f(A) = g(A)$. (5.43)

This theorem implies the useful property that if two functions, one polynomial and one perhaps not, agree on all the eigenvalues of a matrix, then they agree on the matrix itself. This can be used in conjunction with the Cayley-Hamilton theorem, as in the following example, to compute closed-form matrix functions.[M]

`funm(A,'fun')`

Example 5.4: Closed-Form Matrix Functions

Compute the values of $\sin A$, A^5, and e^{At} for the matrix

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

Solution:

Because matrix *A* is triangular, its eigenvalues are obvious: $\lambda_1 = -3$ and $\lambda_2 = -2$. Defining an arbitrary $(n-1)^{st}$ order polynomial as

$$g(\lambda) = \alpha_1 \lambda + \alpha_0$$

we will match values of this function with values of the three unknown functions to determine constants α_0 and α_1.

First, we solve the set of equations $f(\lambda_i) = g(\lambda_i)$ for $i = 1, \ldots, n$.

$$\begin{aligned} \sin(-3) &= \alpha_1(-3) + \alpha_0 \\ \sin(-2) &= \alpha_1(-2) + \alpha_0 \end{aligned} \tag{5.44}$$

to get $\alpha_1 = -0.768$ and $\alpha_0 = -2.45$. Thus,

$$g(\lambda) = -0.768\lambda - 2.45$$

Therefore,

$$f(A) = \sin A = g(A) = -0.768\,A - 2.45\,I = \begin{bmatrix} -0.141 & -0.768 \\ 0 & -0.909 \end{bmatrix}$$

Performing the same procedure on the function A^5,

$$\begin{aligned} (-3)^5 &= \alpha_1(-3) + \alpha_0 \\ (-2)^5 &= \alpha_1(-2) + \alpha_0 \end{aligned} \tag{5.45}$$

gives $\alpha_1 = 211$ and $\alpha_0 = 390$. This means

$$f(A) = A^5 = 211\,A + 390\,I = \begin{bmatrix} -243 & 211 \\ 0 & 32 \end{bmatrix}$$

For the function e^{At}, the equations

$$\begin{aligned} e^{-3t} &= \alpha_1(-3) + \alpha_0 \\ e^{-2t} &= \alpha_1(-2) + \alpha_0 \end{aligned} \tag{5.46}$$

are more difficult to solve in the sense that the constants α_0 and α_1 must now be functions of *t* as well. This can be done with symbolic math packages[M] such as Mathematica® or Maple®, or in simple cases such as this, they can be solved by hand:

`expm(A)`

$$\begin{bmatrix} e^{-3t} \\ e^{-2t} \end{bmatrix} = \begin{bmatrix} -3 & 1 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_0 \end{bmatrix}$$

so

$$\begin{aligned} \begin{bmatrix} \alpha_1 \\ \alpha_0 \end{bmatrix} &= \begin{bmatrix} -3 & 1 \\ -2 & 1 \end{bmatrix}^{-1} \begin{bmatrix} e^{-3t} \\ e^{-2t} \end{bmatrix} \\ &= \begin{bmatrix} -e^{-3t} + e^{-2t} \\ -2e^{-3t} + 3e^{-2t} \end{bmatrix} \end{aligned}$$

Then

$$\begin{aligned} e^{At} &= \left(-e^{-3t} + e^{-2t}\right)\begin{bmatrix} -3 & 1 \\ 0 & -2 \end{bmatrix} + \left(-2e^{-3t} + 3e^{-2t}\right)\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} e^{-3t} & -e^{-3t} + e^{-2t} \\ 0 & e^{-2t} \end{bmatrix} \end{aligned}$$

REMARK: In the above examples, the matrix A conveniently had distinct eigenvalues, so the set of equations $f(\lambda_i) = g(\lambda_i)$ as in (5.44), (5.45), and (5.46), constituted a sufficient set of n independent equations to solve. In the event that the eigenvalues are repeated, one can use the original equation $f(\lambda_i) = g(\lambda_i)$ and its derivative equation $df/d\lambda_i = dg/d\lambda_i$ as an independent set. Higher multiplicities will require higher derivatives.

Example 5.5: Matrix Functions with Repeated Eigenvalues

To illustrate the procedure for finding independent equations when an eigenvalue is repeated, we will find the value of e^{At} when

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

Solution:

Clearly, $\lambda_1 = \lambda_2 = -1$, so we will get only a single independent equation from

$$\begin{aligned} f(\lambda) = g(\lambda) &= \alpha_1\lambda + \alpha_0 \\ e^{-t} &= \alpha_1(-1) + \alpha_0 \end{aligned}$$

Using the derivative (with respect to λ, not t) of this equation, we get

$$\begin{aligned} f'(\lambda) = g'(\lambda) &= \alpha_1 \\ te^{-t} &= \alpha_1 \end{aligned}$$

Obviously, this differentiation simplifies the solution process, giving us $\alpha_1 = te^{-t}$ and $\alpha_0 = te^{-t} + e^{-t}$. Therefore,

$$
\begin{aligned}
e^{At} &= te^{-t}\begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix} + (te^{-t} + e^{-t})\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} -te^{-t} & te^{-t} \\ 0 & -te^{-t} \end{bmatrix} + \begin{bmatrix} te^{-t} + e^{-t} & 0 \\ 0 & te^{-t} + e^{-t} \end{bmatrix} \\
&= \begin{bmatrix} e^{-t} & te^{-t} \\ 0 & e^{-t} \end{bmatrix}
\end{aligned}
$$

Using Jordan Forms for Matrix Functions

The important point to note in the above examples is that the function of the matrix is not simply performed element-by-element. However, in the event that the matrix A is diagonal, then the function *will be* performed element-by-element. This property can be useful if we have a modal matrix with which to diagonalize an operator, compute a function of that diagonal matrix, and then transform the function back to the original basis. If the modal matrix M is used to diagonalize matrix A, then we have

$$A = M\hat{A}M^{-1} \tag{5.47}$$

Exploiting the Taylor series expansion of the function $f(A)$ of this matrix,

$$
\begin{aligned}
f(A) &= \sum_{i=0}^{\infty} \frac{1}{i!} \left.\frac{d^i f}{d\lambda^i}\right|_{\lambda=0} A^i \\
&= \sum_{i=0}^{\infty} \frac{1}{i!} \left.\frac{d^i f}{d\lambda^i}\right|_{\lambda=0} \left(M\hat{A}M^{-1}\right)^i
\end{aligned}
$$

This series contains powers of $M\hat{A}M^{-1}$ that can be reduced by realizing that

$$
\begin{aligned}
\left(M\hat{A}M^{-1}\right)^2 &= \left(M\hat{A}M^{-1}\right)\left(M\hat{A}M^{-1}\right) = M\hat{A}M^{-1}M\hat{A}M^{-1} = M\hat{A}^2M^{-1} \\
&\vdots \\
\left(M\hat{A}M^{-1}\right)^k &= M\hat{A}^kM^{-1}
\end{aligned}
\tag{5.48}
$$

Therefore,

$$\begin{aligned} f(A) &= \sum_{i=0}^{\infty} \frac{1}{i!} \frac{d^i f}{d\lambda^i}\bigg|_{\lambda=0} \left(M\hat{A}M^{-1}\right)^i \\ &= M\left(\sum_{i=0}^{\infty} \frac{1}{i!} \frac{d^i f}{d\lambda^i}\bigg|_{\lambda=0} \hat{A}^i\right)M^{-1} \\ &= Mf(\hat{A})M^{-1} \end{aligned} \tag{5.49}$$

Thus, if $\hat{A}$ is diagonal and therefore $f(\hat{A})$ is easily computed element-by-element, then the similarity transformation $f(A) = Mf(\hat{A})M^{-1}$ can be performed to find $f(A)$.

Note that this simplification holds not just for diagonalizable matrices, but for any similarity transformation on matrix *A*. If *A* is not diagonalizable, i.e., if its representation is a Jordan form, then $f(A)$ can still be found from $f(A) = Mf(\hat{A})M^{-1}$, and $f(\hat{A})$ can be computed by considering one Jordan block at a time. However, the function *f cannot* be computed element-by-element within a Jordan block. Nevertheless, we can find an easy way to compute the formulae for such a function.

Consider that $\hat{A}$ is a single $n \times n$ Jordan block. Then $\hat{A}$ corresponds to an eigenvalue λ_i that is repeated at least *n* times. Therefore, the part of the characteristic polynomial considering $\hat{A}$ alone is:

$$\phi(\lambda) = \left|\lambda - I\hat{A}\right| = \left(\lambda - \lambda_i\right)^n = 0 \tag{5.50}$$

This would imply that $(\lambda - \lambda_i)^k = 0$ for $k = 1, \ldots, n$. To find function $f(\hat{A})$, we would use an $(n-1)^{\text{st}}$-order polynomial $g(\lambda)$, which can be written without loss of generality as

$$g(\lambda) = \alpha_{n-1}(\lambda - \lambda_i)^{n-1} + \cdots + \alpha_2(\lambda - \lambda_i)^2 + \alpha_1(\lambda - \lambda_i)^1 + \alpha_0 \tag{5.51}$$

Because all the eigenvalues are the same within this Jordan block, we will need the equation $f(\lambda) = g(\lambda)$ and $n-1$ of its derivatives in order to find the constants α_j, $j = 0, \ldots, n-1$. Because of (5.51), though, these equations are simple to solve:

$$
\begin{array}{lcl}
g(\lambda_i) = f(\lambda_i) = \alpha_0 & \rightarrow & \alpha_0 = f(\lambda_i) \\
g'(\lambda_i) = f'(\lambda_i) = \alpha_1 & \rightarrow & \alpha_1 = f'(\lambda_i) \\
g''(\lambda_i) = f''(\lambda_i) = 2\alpha_2 & \rightarrow & \alpha_2 = \frac{1}{2} f''(\lambda_i) \\
g'''(\lambda_i) = f'''(\lambda_i) = (3)(2)\alpha_3 & \rightarrow & \alpha_3 = \frac{1}{3!} f'''(\lambda_i) \\
\vdots & \vdots & \vdots \\
g^{(n-1)}(\lambda_i) = f^{(n-1)}(\lambda_i) = (n-1)!\alpha_{n-1} & \rightarrow & \alpha_{n-1} = \frac{1}{(n-1)!} f^{(n-1)}(\lambda_i)
\end{array}
$$

where the notation $f^{(n-1)}(\lambda_i)$ stands for $\left.\frac{d^{n-1} f(\lambda)}{d\lambda^{n-1}}\right|_{\lambda=\lambda_i}$.

Therefore,

$$
f(\hat{A}) = \frac{1}{(n-1)!} f^{(n-1)}(\lambda_i)\left(\hat{A} - \lambda_i I\right)^{n-1} + \cdots + f'(\lambda_i)\left(\hat{A} - \lambda_i I\right) + f(\lambda_i) I \tag{5.52}
$$

To simplify this, we can realize that

$$
\hat{A} - \lambda_i I = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ & & 1 & & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ & & & & 1 \\ 0 & & \cdots & & 0 \end{bmatrix}
$$

$$
\left(\hat{A} - \lambda_i I\right)^2 = \begin{bmatrix} 0 & 0 & 1 & \cdots & 0 \\ & & 0 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 1 \\ & & & & 0 \\ 0 & & \cdots & & 0 \end{bmatrix}
$$

$$\vdots$$

$$
\left(\hat{A} - \lambda_i I\right)^{n-1} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ & & 0 & \ddots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ & & & & 0 \\ 0 & & \cdots & & 0 \end{bmatrix}
$$

and finally

$$\left(\hat{A}-\lambda_i I\right)^n=\begin{bmatrix}0 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & 0\end{bmatrix} \tag{5.53}$$

Therefore, (5.52) becomes

$$f(\hat{A})=\begin{bmatrix} f(\lambda_i) & f'(\lambda_i) & \frac{1}{2}f''(\lambda_i) & \cdots & \frac{1}{(n-1)!}f^{(n-1)}(\lambda_i)\\ 0 & f(\lambda_i) & f'(\lambda_i) & \ddots & \vdots\\ & 0 & f(\lambda_i) & \ddots & \frac{1}{2}f''(\lambda_i)\\ \vdots & & \ddots & \ddots & f'(\lambda_i)\\ 0 & & \cdots & 0 & f(\lambda_i)\end{bmatrix} \tag{5.54}$$

When this is computed, the similarity transformation $f(A)=Mf(\hat{A})M^{-1}$ may be applied as before.

Example 5.6: Matrix Exponential of Jordan Form

Find the value of e^{At} where $A=\begin{bmatrix}-3 & -4\\ 1 & 1\end{bmatrix}$.

Solution:

We will do this example using the Jordan form of A. We first find that the eigenvalues of A are $\lambda_1=\lambda_2=-1$ and that the geometric multiplicity of this eigenvalue is one, implying the existence of one regular eigenvector and one generalized eigenvector. Carrying out a top-down procedure for finding these vectors, we compute a modal matrix of

$$M=\begin{bmatrix}-4 & 0\\ 2 & 1\end{bmatrix}$$

so that

$$\hat{A}=M^{-1}AM=\begin{bmatrix}-1 & 1\\ 0 & -1\end{bmatrix}$$

Now because $f(\lambda)=e^{-t}$, we may use (5.54) to compute

$$f(\hat{A}) = \begin{bmatrix} f(-1) & f'(-1) \\ 0 & f(-1) \end{bmatrix} = \begin{bmatrix} e^{-t} & te^{-t} \\ 0 & e^{-t} \end{bmatrix}$$

which agrees with the result in the previous example. Finally,

$$\begin{aligned} f(A) = Mf(\hat{A})M^{-1} &= \begin{bmatrix} -4 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} e^{-t} & te^{-t} \\ 0 & e^{-t} \end{bmatrix} \begin{bmatrix} -\frac{1}{4} & 0 \\ \frac{1}{2} & 1 \end{bmatrix} \\ &= \begin{bmatrix} e^{-t} - 2te^{-t} & -4te^{-t} \\ te^{-t} & e^{-t} + 2te^{-t} \end{bmatrix} \end{aligned}$$

5.4.3 Minimal Polynomials

Although the Cayley-Hamilton theorem guarantees that there exists an equation, namely, the characteristic equation, that is satisfied by any matrix *A*, it is sometimes possible for other polynomial equations to be satisfied by *A*. If such a polynomial equation is of lower order, it may economize on some matrix calculations, as we have seen in the examples.

> ***Minimal Polynomial:*** The minimal polynomial of matrix *A* is the polynomial $\phi_m(\lambda)$ of lowest order such that $\phi_m(A) = 0$.
> (5.55)

By using the Jordan form, minimal polynomials are relatively easy to find. Recall that the characteristic polynomial of *A* can be written as

$$\phi(\lambda) = \prod_{i=1}^{n} (\lambda - \lambda_i) \tag{5.56}$$

for all *n* eigenvalues λ_i, $i = 1, \ldots, n$, some of which are possibly repeated. If all the eigenvalues are distinct, then the characteristic polynomial will be exactly the same as the minimal polynomial. If, however, there are some Jordan blocks in the Jordan form $\hat{A}$ of *A*, then the minimal polynomial will be of lower order

$$\phi_m(\lambda) = \prod_{i=1}^{k} (\lambda - \lambda_i)^{\eta_i}$$

or

$$\phi_m(\hat{A}) = \prod_{i=1}^{k} (\hat{A} - \lambda_i I)^{\eta_i} \tag{5.57}$$

where, as in Chapter 4, η_i is the index of each eigenvalue λ_i. Here, the summation is performed over the k *distinct* values for the eigenvalues. The discussion of the index revealed that η_i is the size of the largest Jordan block associated with λ_i. The sequence of computations ending with Equation (5.53) would indicate that if J_i is an $\eta_i \times \eta_i$ Jordan block belonging to eigenvalue λ_i, then

$$\begin{aligned} (\lambda_i I - J_i)^n &\neq 0 \quad \text{if } n < \eta_i \\ (\lambda_i I - J_i)^n &= 0 \quad \text{if } n \geq \eta_i \end{aligned}$$

So we must include η_i factors in (5.56) for each repeated eigenvalue λ_i, but additional factors would be redundant. Because $f(A) = Mf(\hat{A})M^{-1}$, it is a valid approach to determine the minimal polynomial based on the Jordan form; similar matrices will have the same characteristic polynomial and the same minimal polynomial.

Example 5.7: Minimal Polynomials for Jordan Forms

Determine the minimal polynomials for the following matrices, which are already in Jordan form:

$$A_1 = \begin{bmatrix} -5 & 1 & 0 \\ 0 & -5 & 1 \\ 0 & 0 & -5 \end{bmatrix} \qquad A_2 = \begin{bmatrix} -5 & 1 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & -5 \end{bmatrix} \qquad A_3 = \begin{bmatrix} -5 & 1 & 0 & 0 & 0 \\ 0 & -5 & 1 & 0 & 0 \\ 0 & 0 & -5 & 0 & 0 \\ 0 & 0 & 0 & -5 & 0 \\ 0 & 0 & 0 & 0 & -2 \end{bmatrix}$$

Solution:

$$\begin{aligned} \phi_{m1}(\lambda) &= (\lambda + 5)^3 \\ \phi_{m2}(\lambda) &= (\lambda + 5)^2 \\ \phi_{m3}(\lambda) &= (\lambda + 5)^3(\lambda + 2) \end{aligned}$$

5.5 Summary

Entire books are written and entire courses are taught on functional analysis. The introduction here is intended only to give some support to the concepts of bilinear and quadratic forms, which are important in the stability studies we will discuss in Chapter 7. The other matrix functions studied here, particularly the matrix exponential e^{At} and its computation through the Cayley-Hamilton theorem will be used in the next chapter on state space system solutions. Summarizing the main points in this chapter:

- A broad class of matrix (or vector) functions known as the functional produces scalar results from matrix (or vector) arguments. These may be linear, bilinear, sesquilinear, or nonlinear, of which the most important example is the quadratic form.
- Arbitrary analytic functions may be defined for matrices, which can be expressed with the help of power series expansions and the Cayley-Hamilton theorem. This theorem spares us the need to ever work with matrix powers higher than $n-1$, where n is the size of the matrix.
- The matrix exponential may be computed given the knowledge of a matrix's eigenvalues because we have a theorem that indicates that two functions that agree on the (scalar) eigenvalues of a matrix also agree on the matrix itself.
- Through an example, we have seen another application for singular values that also indicates the use of quadratic forms and the geometry they imply for their arguments.

The next part of the book begins with Chapter 6. Beginning in this chapter, we concentrate again on state space descriptions of physical systems governed by linear differential equations. All of the tools necessary for the analysis, design, and control of such systems are found in Chapters 1 through 5. Chapter 6 itself concerns the analytical solution to state variable equations, emphasizing SISO linear, time-invariant systems. We will also revisit the basis of eigenvectors in a discussion of the decomposition of system solutions along a natural basis set.

5.6 Problems

5.1 Let $\{\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3\}$ and $\{\boldsymbol{g}_1, \boldsymbol{g}_2, \boldsymbol{g}_3\}$ be two bases for $\Re^3$, such that

$$\begin{aligned} \boldsymbol{g}_1 &= 2\boldsymbol{f}_1 + \boldsymbol{f}_2 + \boldsymbol{f}_3 \\ \boldsymbol{g}_2 &= -1\boldsymbol{f}_1 + 3\boldsymbol{f}_2 + 2\boldsymbol{f}_3 \\ \boldsymbol{g}_3 &= 2\boldsymbol{f}_1 - \boldsymbol{f}_2 + \boldsymbol{f}_3 \end{aligned}$$

a) If the matrix of a linear operator in this space is given in basis $\{\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3\}$ by

$$T = \begin{bmatrix} 8 & -20 & 24 \\ -4 & 20 & -2 \\ -6 & -10 & 32 \end{bmatrix}$$

find the matrix in the basis $\{\boldsymbol{g}_1, \boldsymbol{g}_2, \boldsymbol{g}_3\}$.

b) If we are given a vector $\boldsymbol{x} = -4\boldsymbol{f}_1 + 4\boldsymbol{f}_2 + 4\boldsymbol{f}_3$, find the components of $\boldsymbol{x}$ in the basis $\{\boldsymbol{g}_1, \boldsymbol{g}_2, \boldsymbol{g}_3\}$.

c) Given a linear functional ϕ on the space that is specified by the values $\phi(\boldsymbol{f}_1) = -4, \phi(\boldsymbol{f}_2) = 4$, and $\phi(\boldsymbol{f}_3) = 4$, find the three numbers that specify the functional in the basis $\{\boldsymbol{g}_1, \boldsymbol{g}_2, \boldsymbol{g}_3\}$.

5.2 Consider two bases for space $\Re^3$:

$$\{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3\} = \left\{ \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 3 \\ -2 \\ -2 \end{bmatrix}, \begin{bmatrix} -4 \\ 2 \\ 1 \end{bmatrix} \right\} \qquad \{\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3\} = \left\{ \begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} -4 \\ -3 \\ -2 \end{bmatrix}, \begin{bmatrix} 5 \\ -2 \\ 0 \end{bmatrix} \right\}$$

If a linear functional on this space is given by $\gamma(\boldsymbol{a}_1) = 2$, $\gamma(\boldsymbol{a}_2) = -1$, , and $\gamma(\boldsymbol{a}_3) = -4$, find the representation for the functional in the basis $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3\}$.

5.3 Consider linear vector space V with standard basis $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$. Let bilinear form $B(\boldsymbol{x}, \boldsymbol{y})$ be given in this space by the coefficients

$$B(\boldsymbol{e}_i, \boldsymbol{e}_j) = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 4 \end{bmatrix}$$

Let a linear operator T be given in the same basis as

$$T = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 2 \\ 2 & 0 & 1 \end{bmatrix}$$

a) Show that $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = B(\boldsymbol{x}, \boldsymbol{y})$ is a valid inner product for this space.

b) Find the adjoint of T with respect to the inner product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = B(\boldsymbol{x}, \boldsymbol{y})$.

5.4 Let $p(t)$ be a polynomial, A a linear operator on vector space **V**, λ an eigenvalue of A, and $\boldsymbol{x}$ its corresponding eigenvector. Show that the same eigenvector $\boldsymbol{x}$ will also be an eigenvector of the operator $p(A)$ and that the eigenvalue it corresponds to will be $p(\lambda)$.

5.5 Let M be an $n \times n$ hermitian matrix. Let $\boldsymbol{e}_i$, $i = 1, \ldots, n$, be a set of linearly independent vectors. If it is known that $\boldsymbol{e}_i^* M \boldsymbol{e}_i > 0$ for all $i = 1, \ldots, n$, is it necessarily true that M is a positive-definite matrix?

5.6 Determine a closed-form expression for e^A, where $A = \begin{bmatrix} -1 & -2 \\ -3 & -6 \end{bmatrix}$.

5.7 If $A = \begin{bmatrix} -4 & 2 & 5 \\ 1 & -1 & -1 \\ -1 & 2 & 2 \end{bmatrix}$, find A^{10}.

5.8 Compute e^{At} for the following two matrices:

a) $\begin{bmatrix} -14 & -2 & -14 \\ -7 & -3 & -8 \\ 11 & 2 & 11 \end{bmatrix}$ b) $\begin{bmatrix} -60 & 30 & 210 \\ 40 & -20 & -140 \\ -20 & 10 & 70 \end{bmatrix}$

5.9 Find e^{At} for the given matrices:

a) $\begin{bmatrix} -4 & 0 & 0 \\ 0 & -7 & 1 \\ 0 & -9 & -1 \end{bmatrix}$ b) $\begin{bmatrix} \alpha & \omega \\ -\omega & \alpha \end{bmatrix}$ c) $\begin{bmatrix} -8 & 7 \\ -4 & 3 \end{bmatrix}$

5.10 Find a closed-form expression for A^k, $k \geq 1$,

where $A = \begin{bmatrix} 0 & 0 & 0 \\ 9 & 23 & 30 \\ -7 & -18 & -23.5 \end{bmatrix}$.

5.7 References and Further Reading

As in the previous chapter, the material given here on functions of matrices is available from any standard text on matrix computations, such as [4] and [5]. Multilinear functionals can be treated in increased generality as *tensors*, and this material is discussed in [1].

Because of its critical importance, the matrix exponential e^{At} is the subject of particular attention, being discussed in [7] and [9]. Its computation from the Jordan form is given in [2] and [3]. Note, though, that like eigenvalues, eigenvectors, and singular values, these "pencil-and-paper" computation methods are presented to reinforce an understanding of the nature and behavior of the matrix exponential. They are *not* the best ways to compute the exponential numerically. For details on numerical issues, see [7].

The robot joint-error analysis example can be pursued further through [6], [8], and [10].

[1] Bishop, Richard L., and Samuel I. Goldberg, *Tensor Analysis on Manifolds*, Dover, 1980.

[2] Brogan, William L., *Modern Control Theory*, 3rd edition, Prentice-Hall, 1991.

[3] Chen, Tsi-Chong, *Linear System Theory and Design*, Holt, Rinehart, and Winston, 1984.

[4] Gantmacher, Feliks R., *The Theory of Matrices, Vols I and II*, Chelsea Publishing Co., 1959.

[5] Golub, Gene H., and Charles F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, 1989.

[6] Menq, Chia-Hsiang, and Jin-Hwan Borm, “Statistical Measure and Characterization of Robot Errors,” *IEEE International Conference on Robotics and Automation*, 1988, pp. 926-931.

[7] Moler, Cleve B., and C. F. Van Loan, “Nineteen Dubious Ways to Compute the Exponential of a Matrix,” *SIAM Review*, vol. 20, 1978, pp. 810-836.

[8] Nakamura, Yoshihiko, and Yingti Xu, “Geometrical Fusion Method for Multi-Sensor Robotic Systems,” *IEEE International Conference on Robotics and Automation*, 1989, pp. 668-673.

[9] Vidyasagar, M., “A Characterization of e^{At} and a Constructive Proof of the Controllability Condition,” *IEEE Transactions on Automatic Control*, vol. AC-16, no. 4, 1971, pp. 370-371.

[10] Yufeng, Long, and Lu Jianqin, “On the Optimum Tolerances of Structural Parameters and Kinematic Parameters of Robot Manipulators,” *Mechanism and Machine Theory*, vol. 28, no. 6, 1993, pp. 819-824.