Chapter 5

Matrices and multichannel time series

Familiarity with matrices is essential to computer modeling in both physical and social sciences. As this is a big subject covered by many excellent texts at all levels, our review will be a quick one. We focus on those properties required in the succeeding chapters. We avoid proofs, and although constructions given should be useful in most situations, there will be occasional matrices (which we will dismiss as pathological cases) in which our constructions will fail. In practice, the user should always check computed results. Unfortunately, the so-called pathological cases arise in practice far more often than might be expected. When matrix difficulties arise, the first tendency of the scientist is to use a higher-precision arithmetic. In the author's experience, physically meaningful calculations rarely require high precision. When higher precision seems to be needed, it is often because something is happening physically which shows that the problem being solved is a poorly posed problem. If a slight change in the problem should not make a drastic change in the answer, then it may happen that a different organization of the calculations will obviate the need for high precision. Anyway, our discussion here will focus on the nonpathological cases, but the reader is warned that pathological cases will certainly be encountered in practice and when they are they will be a stern test of the reader's mathematical knowledge and physical insight.

5.1 REVIEW OF MATRICES

A set of simultaneous equations may be written as

$$\mathbf{A}\mathbf{x} \quad = \quad \mathbf{b} \tag{5.1}$$

where A is a square matrix (nonsquare matrices are taken up in Chap. 6 on least squares) and x and b are column vectors. In a 2×2 case, (5.1) becomes

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad = \quad \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \quad = \quad x_1 \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix} \tag{5.2}$$

Equation (5.2) has a simple geometric interpretation. The two columns of the matrix and the column $\mathbf{b}$ are regarded as vectors in a plane. Equation (5.2) says x_1 times the first column vector plus x_2 times the second column vector equals the $\mathbf{b}$ column vector. Difficulty arises when the two column vectors of the matrix point in the same direction. Unless $\mathbf{b}$ just happens to be that direction, no solution x_1, x_2 is possible. The same thing may be said about the general case. A solution $\mathbf{x}$ to equation (5.1) exists if $\mathbf{b}$ lies in the space spanned by the columns of $\mathbf{A}$. In most practical situations the matrix $\mathbf{A}$ and the column $\mathbf{b}$ arise from independent considerations so that it is often reasonable to require the columns of $\mathbf{A}$ to span a space which will contain an arbitrary $\mathbf{b}$ vector. If $\mathbf{A}$ is an $n \times n$ matrix, then its columns are required to span an n-dimensional space. In particular, the n-dimensional parallelopiped with edges given by the columns of $\mathbf{A}$ should not have a zero volume. Such a volume is given by the determinant of $\mathbf{A}$.

Another set of simultaneous equations which arises frequently in practice is the so-called homogeneous equations

$$\mathbf{A}\mathbf{x} \quad = \quad 0 \tag{5.3}$$

This set always has the solution $\mathbf{x} = \mathbf{0}$ which is often called the trivial solution. For (5.3) to have nontrivial solution values for $\mathbf{x}$ the determinant of $\mathbf{A}$ should vanish, meaning that the columns of $\mathbf{A}$ do not span an n-dimensional space. We will return later to the subject of actually solving sets of simultaneous equations.

This set always has the solution $\mathbf{x} = \mathbf{0}$ which is often called the trivial solution. For (5.3) to have nontrivial solution values for $\mathbf{x}$ the determinant of $\mathbf{A}$ should vanish, meaning that the columns of $\mathbf{A}$ do not span an n-dimensional space. We will return later to the subject of actually solving sets of simultaneous equations.

A most useful feature of matrices is that their elements may be not only numbers but that they may be other matrices. Viewed differently, a big matrix may be partitioned into smaller submatrices. A surprising thing is that the product of two matrices is the same whether there are partitions or not. Study the identity

$$\left[\begin{array}{cc|c} a & b & c \\ d & e & f \end{array}\right] \left[\begin{array}{cc} g & h \\ i & j \\ \hline k & l \end{array}\right] \quad = \quad \begin{bmatrix} a & b \\ d & e \end{bmatrix} \begin{bmatrix} g & h \\ i & j \end{bmatrix} + \begin{bmatrix} c \\ f \end{bmatrix} \begin{bmatrix} k & l \end{bmatrix} \tag{5.4}$$

In terms of summation notation, the left-hand side of (5.4) means

$$C_{ik} \quad = \quad \sum_{j=1}^{3} A_{ij} B_{jk} \tag{5.5}$$

whereas the right-hand side means

$$C_{ik} \quad = \quad \sum_{j=1}^{2} A_{ij} B_{jk} + \sum_{j=3}^{3} A_{ij} B_{jk} \tag{5.6}$$

Equations (5.5) and (5.6) are obviously the same; this shows that this partitioning of a matrix product is merely rearranging the terms. Partitioning does not really do anything at all from a mathematical point of view, but it is extremely important from the point of view of computation or discussion.

We now utilize matrix partitioning to develop the bordering method of matrix inversion. The bordering method is not the fastest or the most accurate method but it is quite simple, even for nonsymmetric complex-valued matrices, and it also gives the determinant and works for homogeneous equations. The bordering method proceeds by recursion. Given the inverse to a $k \times k$ matrix, the method shows how to find the inverse of a $(k+1) \times (k+1)$ matrix, which is the same old $k \times k$ matrix with an additional row and column attached to its borders. Specifically, $\mathbf{A}, \mathbf{e}, \mathbf{f}, g$, and $\mathbf{A}^{-1}$ are taken to be known in (5.7). The task is to find $\mathbf{W}, \mathbf{x}, \mathbf{y}$, and z.

$$\begin{bmatrix} \mathbf{A} & \vdots & \mathbf{f} \\ \cdots & \vdots & \cdots \\ \mathbf{e} & \vdots & g \end{bmatrix} \begin{bmatrix} \mathbf{W} & \vdots & \mathbf{y} \\ \cdots & \vdots & \cdots \\ \mathbf{x} & \vdots & z \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \vdots & \mathbf{0} \\ \cdots & \vdots & \cdots \\ \mathbf{0} & \vdots & 1 \end{bmatrix} \tag{5.7}$$

The first thing to do is multiply the partitions in (5.7) together. For the first column of the product we obtain

$$\mathbf{AW} + \mathbf{fx} \quad = \quad \mathbf{I} \tag{5.8}$$

$$\mathbf{eW} + g\mathbf{x} \quad = \quad \mathbf{0} \tag{5.9}$$

A choice of $\mathbf{W}$ of

$$\mathbf{W} \quad = \quad \mathbf{A}^{-1}(\mathbf{I} - \mathbf{fx}) \tag{5.10}$$

leads to (5.8) being satisfied identically. This leaves $\mathbf{x}$ still unknown, but we may find it by substituting (5.10) into (5.9)

$$\mathbf{x} \quad = \quad \frac{\mathbf{eA}^{-1}}{\mathbf{eA}^{-1}\mathbf{f} - g} \tag{5.11}$$

Now, to get the column unknowns $\mathbf{y}$ and z, we compute the second column of the product (5.7)

$$\mathbf{Ay} + \mathbf{f}z \quad = \quad \mathbf{0} \tag{5.12}$$

$$\mathbf{ey} + gz \quad = \quad 1 \tag{5.13}$$

Multiply (5.12) by $\mathbf{A}^{-1}$

$$\mathbf{y} \quad = \quad -\mathbf{A}^{-1}\mathbf{f}z \tag{5.14}$$

This gives the column vector $\mathbf{y}$ within a scale factor z. To get the scale factor, we insert (5.14) into (5.13)

$$\begin{aligned} -\mathbf{e}\mathbf{A}^{-1}\mathbf{f}z + gz &= 1 \\ z &= \frac{1}{g - \mathbf{e}\mathbf{A}^{-1}\mathbf{f}} \end{aligned} \tag{5.15}$$

It may, in fact, be shown that the determinant of the matrix being inverted is given by the product over all the bordering steps of the denominator of (5.15). Thus, if at any time during the recursion the denominator of (5.15) goes to zero, the matrix is singular and the calculation cannot proceed.

Let us summarize the recursion: One begins with the upper left-hand corner of a matrix. The corner is a scalar and its inverse is trivial. Then it is considered to be bordered by a row and a column as shown in (5.7). Next, we find the inverse of this 2×2 matrix. The process is continued as long as one likes. A typical step is first compute z by (5.15) and then compute $\mathbf{A}^{-1}$ of one larger size by

$$\mathbf{A}^{-1} \quad \leftarrow \quad \begin{bmatrix} \mathbf{A}^{-1} & \vdots \\ \cdots & \\ \text{zeros} & \end{bmatrix} + z \begin{bmatrix} \mathbf{A}^{-1}\mathbf{f} \\ \cdots\cdots \\ -1 \end{bmatrix} [\mathbf{e}\mathbf{A}^{-1} \vdots -1] \tag{5.16}$$

where (5.16) was made up from (5.10), (5.11), and (5.14). A Fortran computer program for matrix inversion based on the bordering method is shown below:

```
      SUBROUTINE CMAINE (N,B,A)
C A=MATRIX INVERSE OF B
      COMPLEX B,A,C,R,DEL
      DIMENSION  A(N,N),B(N,N),R(100),C(100)
      DO 10 I=1,N
      DO 10 J=1,N
10    A(I,J)=0
      DO 40 L=1,N
      DEL=B(L,L)
      DO 30 I=1,L
      C(I)=0.
      R(I)=0.
      DO 20 J=1,L
      C(I)=C(I)+A(I,J)*B(J,L)
20    R(I)=R(I)+B(L,J)*C(I)
30    DEL=DEL-B(L,I)*C(I)
      C(L)=-1.
      R(L)=-1.
      DO 40 I=1,L
      C(I)=C(I)/DEL
      DO 40 J=1,L
40    A(I,J)=A(I,J)+C(I)*R(J)
      RETURN
      END
```

It is instructive to see what becomes of $\mathbf{A}^{-1}$ if $\mathbf{A}$ is perturbed steadily in such a way that the determinant of $\mathbf{A}$ becomes singular. If the element g in the matrix of (5.7) is moved closer and closer to $\mathbf{e}\mathbf{A}^{-1}\mathbf{f}$, then we see from (5.15) that z tends to infinity. What is interesting is that the second term in (5.16) comes to dominate the first, and the inverse tends to infinity times the product of a column $\mathbf{e}$ with a row $\mathbf{r}$.

The usual expressions $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$ or $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ in the limit of small z^{-1} tend to

$$\mathbf{A}\mathbf{c}\mathbf{r} \quad = \quad z^{-1}\mathbf{I} \tag{5.17}$$

or

$$\mathbf{c}\mathbf{r}\mathbf{A} \quad = \quad z^{-1}\mathbf{I} \tag{5.18}$$

In the usual case (rank $\mathbf{A} = n - 1$, not rank $\mathbf{A} < n - 1$) where neither $\mathbf{c}$ nor $\mathbf{r}$ vanish identically, (5.17) and (5.18) in the limit $z^{-1} = 0$ become

$$\mathbf{A}\mathbf{c} \quad = \quad \mathbf{0} \tag{5.19}$$

$$\mathbf{r}\mathbf{A} \quad = \quad \mathbf{0} \tag{5.20}$$

In summary, then, to solve an ordinary set of simultaneous equations like (5.1), one may compute the matrix inverse of $\mathbf{A}$ by the bordering method and then multiply (5.1) by $\mathbf{A}^{-1}$ obtaining

$$\mathbf{x} \quad = \quad \mathbf{A}^{-1}\mathbf{b} \tag{5.21}$$

In the event $\mathbf{b}$ vanishes, we are seeking the solution to homogeneous equations and we expect that z will explode in the last step of the bordering process. (If it happens earlier, one should be able to rearrange things.) The solution is then given by the column $\mathbf{c}$ in (5.19).

The row homogeneous equations of (5.20) was introduced because such a set arises naturally for the solution to the row eigenvectors of a nonsymmetric matrix. In the next section, we will go into some detailed properties of eigenvectors. A column eigenvector $\mathbf{c}$ of a matrix $\mathbf{A}$ is defined by the solution to

$$\mathbf{A}\mathbf{c} \quad = \quad \lambda\mathbf{c} \tag{5.22}$$

where λ is the so-called eigenvalue. At the same time, one also considers a row eigenvector equation

$$\mathbf{r}\mathbf{A} \quad = \quad \lambda\mathbf{r} \tag{5.23}$$

To have a solution for (5.22) or (5.23), one must have $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$. After finding the roots λ_j of the polynomial $\det(\mathbf{A} - \lambda\mathbf{I})$, one may form a new matrix $\mathbf{A}'$ for each λ_j by

$$\mathbf{A}' \quad = \quad \mathbf{A} - \lambda_j\mathbf{I} \tag{5.24}$$

then the solution to

$$\mathbf{A}'\mathbf{x} \quad = \quad \mathbf{0} \tag{5.25}$$

arises from the column $\mathbf{c}$ at the last step of the bordering. It is the column eigenvector. Likewise, the row eigenvector is the row in the last step of the bordering algorithm.

EXERCISES:

1 Indicate the sizes of all the matrices in equations (5.8) to (5.15).

2 Show how (5.16) follows from (5.10), (5.11), (5.14), and (5.15).

5.2 SYLVESTER'S MATRIX THEOREM

Sylvester's theorem provides a rapid way to calculate functions of a matrix. Some simple functions of a matrix of frequent occurrence are $\mathbf{A}^{-1}$ and $\mathbf{A}^N$ (for N large). Two more matrix functions which are very important in wave propagation are $e^{\mathbf{A}}$ and $\mathbf{A}^{1/2}$. Before going into the somewhat abstract proof of Sylvester's theorem, we will take up a numerical example. Consider the matrix

$$\mathbf{A} \quad = \quad \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix} \tag{5.26}$$

It will be necessary to have the column eigenvectors and the eigenvalues of this matrix; they are given by

$$\begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad = \quad 1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{5.27}$$

$$\begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} \quad = \quad 2 \begin{bmatrix} 2 \\ 1 \end{bmatrix} \tag{5.28}$$

Since the matrix $\mathbf{A}$ is not symmetric, it has row eigenvectors which differ from the column vectors. These are

$$[-1 \quad 2] \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix} \quad = \quad 1\,[-1 \quad 2] \tag{5.29}$$

$$[1 \quad -1] \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix} \quad = \quad 2\,[1 \quad -1] \tag{5.30}$$

We may abbreviate equations (5.27) through (5.30) by

$$\begin{aligned} \mathbf{A} \ \mathbf{c}_1 &= \lambda_1 \ \mathbf{c}_1 \\ \mathbf{A} \ \mathbf{c}_2 &= \lambda_2 \ \mathbf{c}_2 \\ \mathbf{r}_1 \ \mathbf{A} &= \lambda_1 \ \mathbf{r}_1 \\ \mathbf{r}_2 \ \mathbf{A} &= \lambda_2 \ \mathbf{r}_2 \end{aligned} \tag{5.31}$$

The reader will observe that $\mathbf{r}$ or $\mathbf{c}$ could be multiplied by an arbitrary scale factor and (5.31) would still be valid. The eigenvectors are said to be normalized if scale factors have been chosen so that $\mathbf{r}_1 \cdot \mathbf{c}_1 = 1$ and $\mathbf{r}_2 \cdot \mathbf{c}_2 = 1$. It will be observed that $\mathbf{r}_2 \cdot \mathbf{c}_2 = 0$ and $\mathbf{r}_2 \cdot \mathbf{c}_1 = 0$, a general result to be established in the exercises.

Let us consider the behavior of the matrix $\mathbf{c}_1\mathbf{r}_1$.

$$\mathbf{c}_1\mathbf{r}_1 \quad = \quad \begin{bmatrix} 1 \\ 1 \end{bmatrix} [-1 \;\; 2] \quad = \quad \begin{bmatrix} -1 & 2 \\ -1 & 2 \end{bmatrix}$$

Any power of this matrix is the matrix itself, for example its square.

$$\begin{bmatrix} -1 & 2 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} -1 & 2 \\ -1 & 2 \end{bmatrix} \quad = \quad \begin{bmatrix} -1 & 2 \\ -1 & 2 \end{bmatrix}$$

This property is called idempotence (Latin for self-power). It arises because $(\mathbf{c}_1\mathbf{r}_1)(\mathbf{c}_1\mathbf{r}_1) = \mathbf{c}_1(\mathbf{r}_1 \cdot \mathbf{c}_1)\,\mathbf{r}_1 = \mathbf{c}_1\mathbf{r}_1$. The same thing is of course true of $\mathbf{c}_2\mathbf{r}_2$. Now notice that the matrix $\mathbf{c}_1\mathbf{r}_1$ is "perpendicular" to the matrix $\mathbf{c}_2\mathbf{r}_2$, that is

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

since $\mathbf{r}_2$ and $\mathbf{c}_2$ are perpendicular.

Sylvester's theorem says that any function f of the matrix $\mathbf{A}$ may be written

$$f(\mathbf{A}) \quad = \quad f(\lambda_1)\,\mathbf{c}_1\mathbf{r}_1 + f(\lambda_2)\,\mathbf{c}_2\mathbf{r}_2$$

The simplest example is $f(\mathbf{A}) = \mathbf{A}$

$$\begin{aligned} \mathbf{A} &= \lambda_1\,\mathbf{c}_1\mathbf{r}_1 + \lambda_2\,\mathbf{c}_2\mathbf{r}_2 \\ &= 1 \begin{bmatrix} -1 & 2 \\ -1 & 2 \end{bmatrix} + 2 \begin{bmatrix} 2 & -2 \\ 1 & -1 \end{bmatrix} \quad = \quad \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix} \end{aligned} \tag{5.32}$$

Another example is

$$A^2 \quad = \quad 1^2 \begin{bmatrix} -1 & 2 \\ -1 & 2 \end{bmatrix} + 2^2 \begin{bmatrix} 2 & -2 \\ 1 & -1 \end{bmatrix} \quad = \quad \begin{bmatrix} 7 & -6 \\ 3 & -2 \end{bmatrix}$$

The inverse is

$$A^{-1} \quad = \quad 1^{-1} \begin{bmatrix} -1 & 2 \\ -1 & 2 \end{bmatrix} + 2^{-1} \begin{bmatrix} 2 & -2 \\ 1 & -1 \end{bmatrix} \quad = \quad \frac{1}{2} \begin{bmatrix} 0 & 2 \\ -1 & 3 \end{bmatrix}$$

The identity matrix may be expanded in terms of the eigenvectors of the matrix $\mathbf{A}$.

$$A^0 \quad = \quad \mathbf{I} \quad = \quad 1^0 \begin{bmatrix} -1 & 2 \\ -1 & 2 \end{bmatrix} + 2^0 \begin{bmatrix} 2 & -2 \\ 1 & -1 \end{bmatrix} \quad = \quad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Before illustrating some more complicated functions let us see what it takes to prove Sylvester's theorem. We will need one basic result which is in all the books on matrix theory, namely, that most matrices (see exercises) can be diagonalized. In terms of our 2×2 example this takes the form

$$\begin{bmatrix} \mathbf{r}_1 \\ \hline \mathbf{r}_2 \end{bmatrix} \mathbf{A}\,[\mathbf{c}_1 \mid \mathbf{c}_2] \quad = \quad \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \tag{5.33}$$

where

$$\begin{bmatrix} \mathbf{r}_1 \\ \hline \mathbf{r}_2 \end{bmatrix} [\mathbf{c}_1 \mid \mathbf{c}_2] \quad = \quad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{5.34}$$

Since a matrix commutes with its inverse, (5.34) implies

$$[\mathbf{c}_1 \mid \mathbf{c}_2] \begin{bmatrix} \mathbf{r}_1 \\ \hline \mathbf{r}_2 \end{bmatrix} \quad = \quad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{5.35}$$

Postmultiply (5.33) by the row matrix and premultiply by the column matrix. Using (5.35), we get

$$\mathbf{A} \quad = \quad [\mathbf{c}_1 \mid \mathbf{c}_2] \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \mathbf{r}_1 \\ \hline \mathbf{r}_2 \end{bmatrix} \tag{5.36}$$

Equation (5.36) is (5.32) in disguise, as we can see by writing (5.36) as

$$\begin{aligned} \mathbf{A} &= [\mathbf{c}_1 \mid \mathbf{c}_2] \left\{ \begin{bmatrix} \lambda_1 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \lambda_2 \end{bmatrix} \right\} \begin{bmatrix} \mathbf{r}_1 \\ \hline \mathbf{r}_2 \end{bmatrix} \\ &= [\mathbf{c}_1 \mid \mathbf{c}_2] \begin{bmatrix} \lambda_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{r}_1 \\ \hline \mathbf{r}_2 \end{bmatrix} + [\mathbf{c}_1 \mid \mathbf{c}_2] \begin{bmatrix} 0 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \mathbf{r}_1 \\ \hline \mathbf{r}_2 \end{bmatrix} \\ &= \lambda_1 \mathbf{c}_1 \mathbf{r}_1 + \lambda_2 \mathbf{c}_2 \mathbf{r}_2 \end{aligned}$$

Now to get $\mathbf{A}^2$ we have

$$\mathbf{A}^2 \quad = \quad (\lambda_1 \mathbf{c}_1 \mathbf{r}_1 + \lambda_2 \mathbf{c}_2 \mathbf{r}_2)(\lambda_1 \mathbf{c}_1 \mathbf{r}_1 + \lambda_2 \mathbf{c}_2 \mathbf{r}_2)$$

Using the orthonormality of $\mathbf{c}_1 \mathbf{r}_1$ and $\mathbf{c}_2 \mathbf{r}_2$ this reduces to

$$\mathbf{A}^2 \quad = \quad \lambda_1^2 \mathbf{c}_1 \mathbf{r}_1 + \lambda_2^2 \mathbf{c}_2 \mathbf{r}_2$$

It is clear how (5.36) can be used to prove Sylvester's theorem for any polynomial function of $\mathbf{A}$. Clearly, there is nothing peculiar about 2×2 matrices either. This works for $n \times n$. Likewise, one may consider infinite series functions in $\mathbf{A}$. Since almost any function can be made up of infinite series, we can consider also transcendental functions like sine, cosine, exponential.

Exponentials arise naturally as the solutions to differential equations. Consider the matrix differential equation

$$\frac{d}{dx} \mathbf{E} \quad = \quad \mathbf{A} \mathbf{E} \tag{5.37}$$

One may readily verify the power series solution

$$\mathbf{E} \quad = \quad \mathbf{I} + \mathbf{A} x + \frac{\mathbf{A}^2 x^2}{2!} + \cdots$$

This is the power series definition of an exponential function. If the matrix $\mathbf{A}$ is one of that vast majority which can be diagonalized, then the exponential can be more simply expressed by Sylvester's theorem. For the numerical example we have been considering, we have

$$\mathbf{E} \quad = \quad e^x \begin{bmatrix} -1 & 2 \\ -1 & 2 \end{bmatrix} e^{2x} \begin{bmatrix} 2 & -2 \\ 1 & -1 \end{bmatrix}$$

The exponential matrix is a solution to the differential equation (5.37) without regard to boundaries. It frequently happens that physics gives one a differential equation

$$\frac{d}{dx} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \quad = \quad \mathbf{A} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \tag{5.38}$$

Subject to two boundary conditions on either of y_1 or y_2 or a combination. One may verify that

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \quad = \quad e^{\mathbf{A}x} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix}$$

is the solution to (5.38) for arbitrary constants k_1 and k_2. Boundary conditions are then used to determine the numerical values of k_1 and k_2. Note that k_1 and k_2 are just $y_2(x = 0)$ and $y_2(x = 0)$.

An interesting situation arises with the square root of a matrix. A 2×2 matrix like $\mathbf{A}$ will have four square roots because there are four possible combinations for choice of plus or minus signs on $\sqrt{\lambda_1}$ and $\sqrt{\lambda_2}$. In general, an $n \times n$ matrix has 2^n square roots. An important application arises in a later chapter, where we will deal with the differential operator $(k^2 + \delta^2/\delta x^2)^{1/2}$. The square root of an operator is explained in very few books and few people even know what it means. The best way to visualize the square root of this differential operator is to relate it to the square root of the matrix $\mathbf{M}$ where

$$\mathbf{M} \quad = \quad k^2 \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 1 \end{bmatrix} + \frac{1}{(\Delta x)^2} \begin{bmatrix} ? & ? & & & \\ 1 & -2 & 1 & & \\ & 1 & -2 & 1 & \\ & & 1 & -2 & 1 \\ & & & ? & ? \end{bmatrix}$$

The right-hand matrix is a second difference approximation to a second partial derivative. Let us define

$$\mathbf{M} \quad = \quad k^2\mathbf{I} + \mathbf{T}$$

Clearly we wish to consider $\mathbf{M}$ generalized to a very large size so that the end effects may be minimized. In concept, we can make $\mathbf{M}$ as large as we like and for any size we can get $2^{\mathbf{M}}$ square roots. In practice there will be only two square roots of interest, one with the plus roots of all the eigenvalues and the other with all the minus roots. How can we find these "principal value" square roots? An important case of interest is where we can use the binomial theorem so that

$$\begin{aligned} (k^2\mathbf{I} + \mathbf{T})^{1/2} \quad &= \quad \pm k \left(\mathbf{I} + \frac{\mathbf{T}}{k^2} \right)^{1/2} \\ &= \quad \pm k \left(\mathbf{I} + \frac{\mathbf{T}}{2k^2} - \frac{\mathbf{T}^2}{8k^4} + \cdots \right) \end{aligned}$$

The result is justified by merely squaring the assumed square root. Alternatively, it may be justified by means of Sylvester's theorem. It should be noted that on squaring the assumed square root one utilizes the fact that $\mathbf{I}$ and $\mathbf{T}$ commute. We are led to the idea that the square root of the differential operator may be interpreted as

$$\left(k^2 + \frac{\partial^2}{\partial x^2}\right)^{1/2} \quad = \quad k + \frac{1}{2k}\frac{\partial^2}{\partial x^2} + \cdots$$

provided that k is not a function of x. If k is a function of x, the square root of the differential operator still has meaning but is not so simply computed with the binomial theorem.

EXERCISES:

1 Premultiply (5.31)b by $\mathbf{r}_1$ and postmultiply (5.31)c by $\mathbf{c}_2$, then subtract. Is $\lambda_1 \neq \lambda_2$ a necessary condition for $\mathbf{r}_1$ and $\mathbf{c}_2$ to be perpendicular? Is it a sufficient condition?

2 Show the Cayley-Hamilton theorem, that is, if

$$0 \quad = \quad f(\lambda) \quad = \quad \det(\mathbf{A} - \lambda \mathbf{I}) \quad = \quad p_0 + p_1\lambda + p_2\lambda^2 + \cdots p_n\lambda^n$$

then

$$f(\mathbf{A}) \quad = \quad p_0 + p_1\mathbf{A} + p_2\mathbf{A}^2 + \cdots + p_n\mathbf{A}^n \quad = \quad 0$$

3 Verify that, for a general 2×2 matrix $\mathbf{A}$, for which

$$\begin{aligned} \lambda_1 &\neq \lambda_2 \\ \mathbf{c}_1\mathbf{r}_1 &= (\lambda_2\mathbf{I} - \mathbf{A})/(\lambda_2 - \lambda_1) \end{aligned}$$

where λ_1 and λ_2 are eigenvalues of $\mathbf{A}$. What is the general form for $\mathbf{c}_2\mathbf{r}_2$?

4 For a symmetric matrix it can be shown that there is always a complete set of eigenvectors. A problem sometimes arises with nonsymmetric matrices. Study the matrix

$$\begin{bmatrix} 1 & 1-\epsilon^2 \\ -1 & 3 \end{bmatrix}$$

as $\epsilon \to 0$ to see why one eigenvector is lost. This is called a defective matrix. (This example is from T. R. Madden.)

5 A wide variety of wave-propagation problems in a stratified medium reduce to the equation

$$\frac{d}{dx}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \quad = \quad \begin{bmatrix} 0 & b \\ a & 0 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

What is the x dependence of the solution when ab is positive? When ab is negative? Assume a and b are independent of x. Use Sylvester's theorem. What would it take to get a defective matrix? What are the solutions in the case of a defective matrix?

6 Consider a matrix of the form $\mathbf{I} + \mathbf{v}\mathbf{v}^T$ where $\mathbf{v}$ is a column vector and $\mathbf{v}^T$ is its transpose. Find $(\mathbf{I}+\mathbf{v}\mathbf{v}^T)^{-1}$ in terms of a power series in $\mathbf{v}\mathbf{v}^T$. [Note that $(\mathbf{v}\mathbf{v}^T)^n$ collapses to $\mathbf{v}\mathbf{v}^T$ times a scaling factor, so the power series reduces considerably.]

7 The following "cross-product" matrix often arises in electrodynamics. Let $\mathbf{B} = (B_x, B_y, B_z)$

$$U \quad = \quad \frac{1}{\sqrt{\mathbf{B}\cdot\mathbf{B}}} \begin{bmatrix} 0 & -B_z & B_y \\ B_z & 0 & -B_x \\ -B_y & B_x & 0 \end{bmatrix}$$

(a) Write out elements of $\mathbf{I} + \mathbf{U}^2$.

(b) Show that $\mathbf{U}(\mathbf{I} + \mathbf{U}^2) = 0$ or $\mathbf{U}^3 = -\mathbf{U}$.

(c) Let $\mathbf{v}$ be an arbitrary vector. In what geometrical directions do $\mathbf{U}\mathbf{v}$, $\mathbf{U}^2\mathbf{v}$, and $(\mathbf{I} + \mathbf{U}^2)\mathbf{v}$ point?

(d) What are the eigenvalues of $\mathbf{U}$? [Hint: Use part (b).]

(e) Why cannot $\mathbf{U}$ be canceled from $\mathbf{U}^3 = -\mathbf{U}$?

(f) Verify that the idempotent matrices of $\mathbf{U}$ are

$$\begin{aligned} \mathbf{c}_1\mathbf{r}_1 &= (\mathbf{I} + \mathbf{U}^2) \\ \mathbf{c}_2\mathbf{r}_2 &= \frac{1}{2}(i\mathbf{U} - \mathbf{U}^2) \\ \mathbf{c}_3\mathbf{r}_3 &= \frac{1}{2}(-i\mathbf{U} - \mathbf{U}^2) \end{aligned}$$

5.3 MATRIX FILTERS, SPECTRA, AND FACTORING

Two time series can be much more interesting than one because of the possibility of interactions between them. The general linear model for two series is depicted in Figure 5.1.

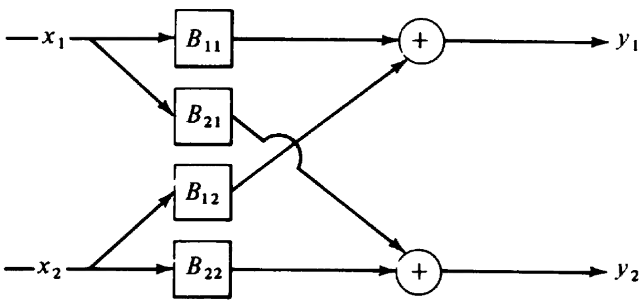

Figure 5.1: Two time series x_1 and x_2 input to a matrix of four filters illustrates the general linear model of multichannel filtering. $\boxed{\text{c5-5-2}}$ [NR]

The filtering operation in the figure can be expressed as a matrix times vector operation, where the elements of the matrix and vectors are Z transform polynomials.

That is,

$$\begin{bmatrix} Y_1(Z) \\ Y_2(Z) \end{bmatrix} = \begin{bmatrix} B_{11}(Z) & B_{12}(Z) \\ B_{21}(Z) & B_{22}(Z) \end{bmatrix} \begin{bmatrix} X_1(Z) \\ X_2(Z) \end{bmatrix} \tag{5.39}$$

One fact which is obvious but unfamiliar is that a matrix with polynomial elements is exactly the same thing as a polynomial with matrix coefficients. This is illustrated by the example:

$$\begin{bmatrix} 1+Z+2Z^2 & Z \\ 1 & 1+Z^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} Z + \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} Z^2$$

Now we can address ourselves to the inverse problem; given a filter $\mathbf{B}$ and the outputs $\mathbf{Y}$ how can we find the inputs $\mathbf{X}$? The solution is analogous to that of single time series. Let us regard $\mathbf{B}(Z)$ as a matrix of polynomials. One knowns, for example, that the inverse of any 2×2 matrix

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \text{is} \quad \frac{\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}}{ad - bc}$$

Consequently $\mathbf{Y} = \mathbf{BX}$ may be solved for $\mathbf{X}$ as $\mathbf{X} = \mathbf{B}^{-1}\mathbf{Y}$ where

$$\mathbf{B}^{-1} = \begin{bmatrix} B_{22}(Z) & -B_{12}(Z) \\ -B_{21}(Z) & B_{11}(Z) \end{bmatrix} \frac{1}{B_{11}(Z)B_{22}(Z) - B_{12}(Z)B_{21}(Z)}$$

The denominator is a scalar. We have treated scalar denominators before. If all the zeros lie outside the unit circle, we can use an ordinary power series for the inverse; otherwise, it is not minimum-phase and we use a Laurent series.

When one generalizes to many time series, the numerator matrix is the so-called adjoint matrix and the denominator is the determinant. The adjoint matrix can be formed without the use of any division operations. In other words, elements in the adjoint matrix are in the form of sums of products. For this reason, we may say that the criterion for a minimum-phase matrix wavelet is that the determinant of its Z transform has no zeros inside the unit circle.

Equation (5.39) is a useful description of Figure 5.1 in most applications. However in some applications (where the filter is an unknown to be determined), a transposed form of (5.39) is more useful. If b_{12} was interchanged with b_{21} in Figure 5.1, we could use the "row data" expression

$$[Y_1(Z) \quad Y_2(Z)] = [X_1(Z) \quad X_2(Z)] \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \tag{5.40}$$

Now that we have generalized the concept of filtering from scalar-valued time series to vector-valued series, it is natural to generalize the idea of spectrum. For vector-valued time functions, the spectrum is a matrix called the *spectral matrix* and it is given by

$$\begin{aligned}\mathbf{R}(\omega) &= \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} = \begin{bmatrix} \overline{Y}_1\left(\frac{1}{X}\right) \\ \overline{Y}_2\left(\frac{1}{X}\right) \end{bmatrix} [Y_1(Z) \quad Y_2(Z)] \\ &= \begin{bmatrix} \overline{Y}_1 Y_1 & \overline{Y}_1 Y_2 \\ \overline{Y}_2 Y_1 & \overline{Y}_2 Y_2 \end{bmatrix}\end{aligned} \tag{5.41}$$

It will be noticed that the vector times vector operation defining (5.41) is an "outer product" rather than the more usually occurring "inner product." The diagonals of the spectral matrix $\mathbf{R}$ contain the usual auto-spectrum of each channel. Off-diagonals contain the cross spectrum. Because (5.41) is an outer product, the matrix is singular. Now, instead of taking $[Y_1(Z) \quad Y_2(Z)]$ to have a time function with a finite amount of energy, let us suppose the filter inputs to (5.40), namely $(x_1(t), x_2(t))$ are made up of random numbers, independently drawn from some time series and their spectral matrix is defined like (5.41) but taking an expectation (average over the ensemble). We have

$$\mathbf{R}(\omega) = E \begin{bmatrix} \overline{Y}_1 \\ \overline{Y}_2 \end{bmatrix} [Y_1 \quad Y_2]$$

substituting from (5.40)

$$\mathbf{R}(\omega) = E \begin{bmatrix} \overline{B}_{11} & \overline{B}_{21} \\ \overline{B}_{12} & \overline{B}_{22} \end{bmatrix} \begin{bmatrix} \overline{X}_1 \\ \overline{X}_2 \end{bmatrix} [X_1 \quad X_2] \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

Now, grouping the ensemble summation with the random variables, we get

$$\mathbf{R} = \begin{bmatrix} \overline{B}_{11} & \overline{B}_{21} \\ \overline{B}_{12} & \overline{B}_{22} \end{bmatrix} \left\{ E \begin{bmatrix} \overline{X}_1 \\ \overline{X}_2 \end{bmatrix} [X_1 \quad X_2] \right\} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \tag{5.42}$$

Next, we explicitly introduce the assumption that the random numbers $x_1(t)$ are drawn independently of $x_2(t)$, thus $E(\overline{X}_2(1/z)X_1(z)) = 0$ and the assumption that $x_i(t)$ is white $E[x_i(t)x_i(t+s)] = 0$ is $s \neq 0$ and of unit variance $E[x_i(t)^2] = 1$. Thus (5.42) becomes

$$\begin{aligned}\mathbf{R} &= \begin{bmatrix} \overline{B}_{11} & \overline{B}_{21} \\ \overline{B}_{12} & \overline{B}_{22} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \\ &= \begin{bmatrix} \overline{B}_{11} & \overline{B}_{21} \\ \overline{B}_{12} & \overline{B}_{22} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}\end{aligned} \tag{5.43}$$

Of course, in practice the spectral matrix must be estimated, say $\hat{\mathbf{R}}$, from finite samples of data. This means that ensemble summation must be simulated. If the ensembles sum in (5.42) is simulated by summation over one point (no summation), then (5.42) is a singular matrix like (5.41). As discussed earlier, the accuracy of the elements of the spectral matrix improves with the square root of the number of ensemble elements summed over.

Single-channel spectral factorization gives insight into numerous important problems in mathematical physics. We have seen that the concepts of filter and spectrum extend in quite a useful fashion to multichannel data. It was only natural that a great deal of effort should have gone to spectral factorization of multichannel data. This effort has been successful. However, in retrospect, from the point of view of computer modeling and interpretation of observed waves, it must be admitted that multichannel spectral factorization has not been especially useful. Nevertheless a brief summary of results will be given.

The root method. I extended the single-channel root method to the multichannel case.[1] The method is even more cumbersome in the multichannel case. A most surprising thing about the solution is that it includes a much broader result: that a polynomial with matrix coefficients may be factored. For example,

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + Z \begin{bmatrix} -3 & -1 \\ 14 & -11 \end{bmatrix} + Z^2 \begin{bmatrix} -4 & 4 \\ -58 & 28 \end{bmatrix}$$

factors 6 ways to

$$\begin{aligned}
&\left\{ I + Z \begin{bmatrix} 2 & -1 \\ 20 & -7 \end{bmatrix} \right\} \left\{ I + Z \begin{bmatrix} -5 & 0 \\ -6 & -4 \end{bmatrix} \right\} \\
&\left\{ I + Z \begin{bmatrix} -4 & 0 \\ -10 & -2 \end{bmatrix} \right\} \left\{ I + Z \begin{bmatrix} 1 & -1 \\ 24 & -9 \end{bmatrix} \right\} \\
&\left\{ I + Z \begin{bmatrix} 0 & -1 \\ 10 & -7 \end{bmatrix} \right\} \left\{ I + Z \begin{bmatrix} -3 & 0 \\ 4 & -4 \end{bmatrix} \right\} \\
&\left\{ I + Z \begin{bmatrix} -4 & 0 \\ -4 & -3 \end{bmatrix} \right\} \left\{ I + Z \begin{bmatrix} 1 & -1 \\ 18 & -8 \end{bmatrix} \right\} \\
&\left\{ I + Z \begin{bmatrix} -1 & -1 \\ 8 & -7 \end{bmatrix} \right\} \left\{ I + Z \begin{bmatrix} -2 & 0 \\ 6 & -4 \end{bmatrix} \right\} \\
&\left\{ I + Z \begin{bmatrix} -4 & 0 \\ 2 & -5 \end{bmatrix} \right\} \left\{ I + Z \begin{bmatrix} 1 & -1 \\ 12 & -6 \end{bmatrix} \right\}
\end{aligned}$$

The Toeplitz method. The only really practical method for finding an invertible matrix wavelet with a given spectrum is the multichannel Toeplitz method. The necessary algebra is developed in a later section on multichannel time series prediction.

The exp-log and Hilbert transform methods. A number of famous mathematicians including Norbert Wiener have worked on the problem from the point of view of extending the exp-log or the Hilbert transform method. The principal stumbling block is that $\exp(A+B)$ does not equal $\exp(A)\exp(B)$ unless A and B happen to commute, that is, $AB = BA$. This is usually not the case. Although many difficult papers have appeared on the subject (some stating that they solved the problem),

[1] Claerbout, J.F., Spectral Factorization of Multiple Time Series, Biometrica, Vol 53, Nos. 1 and 2, pp 264-6, 1966

the author is unaware of anyone who ever wrote a computer program which works at fast Fourier transform speeds as does the single-channel Hilbert transform method.

EXERCISES:

1 Think up a matrix filter where the two outputs $y_1(t)$ and $y_2(t)$ are the same but for a scale factor. Clearly $\mathbf{X}$ cannot be recovered from $\mathbf{Y}$. Show that the determinant of the filter vanishes. Find another example in which the determinant is zero at one frequency but nonzero elsewhere. Explain in the time domain in what sense the input cannot be recovered from the output.

2 Given a thermometer which measures temperature plus α times pressure and a pressure gage which measures pressure plus β times the time rate of change of the temperature, find the matrix filter which converts the observed series to temperature and pressure. [HINT: Use either the time derivative approximation $1 - Z$ or $2(1 - Z)/(1 + Z)$.]

3 Let

$$\mathbf{B}(Z) \quad = \quad \begin{bmatrix} 1 & 0 \\ 3 & 2 \end{bmatrix} + Z \begin{bmatrix} 2 & 1 \\ 0 & 0 \end{bmatrix}$$

Identify coefficients of powers of Z in $\mathbf{B}(Z)\mathbf{A}(Z) = \mathbf{I}$, to develop recursively the coefficients of $\mathbf{A}(Z) = [\mathbf{B}(Z)]^{-1}$.

4 Express the inverse of

$$\begin{bmatrix} 1 + 2Z & Z \\ 3 & 2 \end{bmatrix}$$

in a Taylor or Laurent series as is necessary.

5 The determinant of a polynomial with matrix coefficients may be independent of Z. Applied to matrix filters, this may mean that an inverse filter may have only a finite number of powers in Z instead of the infinite series one always has with scalar filters. What is the most nontrivial example you can find?

5.4 MARKOV PROCESSES

A Markov process is another mathematical model for a time series. Until now it has found little use in geophysics, but we will include it anyway because it might become useful and it is easily explained with the methods previously developed.

Suppose that x_t could take on only integer values. A given value is called a state. As time proceeds, transitions are made from the jth state to the ith state according to a probability matrix p_{ij}. The system has no memory. The next state is probabilistically dependent on the current state but independent of the previous states. The classic example is of a frog in a lily pond. As time goes by, the frog jumps from one lily pad to another. He may be more likely to jump to a near one than to a far one. The state of the system is the number of the pad the frog currently occupies. The transitions are his jumps.

To begin with, one defines a state probability $\pi_i(k)$, the probability that the system will occupy state i after k transitions if its state is known at $k = 0$. We also define the transition matrix P_{ij}. Then

$$\pi(k+1) \quad = \quad \mathbf{P}\pi(k) \tag{5.44}$$

The initial-state probability vector is $\pi(0)$. Since the initial state is known, then $\pi(0)$ is all zeros except for a one (1) in the position corresponding to the initial state. For example, see the state-transition diagram of Figure 5.2.

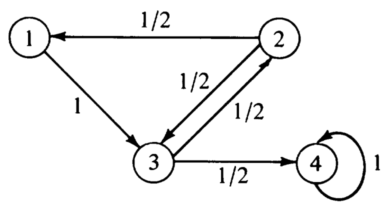

Figure 5.2: An example of a state-transition diagram. $\boxed{\text{c5-5-3}}$ [NR]

The diagram corresponds to the probability matrix

$$\begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \end{bmatrix}_{k+1} = \begin{bmatrix} 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 1 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \end{bmatrix}_k$$

Since at each time a transition must occur, we have that the sum of the elements in a column must be unity. In other words, the row vector $[1 \;\; 1 \;\; 1 \;\; 1]$ is an eigenvector of $\mathbf{P}$ with unit eigenvalue. Let us define the Z transform of the probability vector as

$$\Pi(Z) \quad = \quad \pi(0) + Z\pi(1) + Z^2\pi(2) + \cdots \tag{5.45}$$

In terms of Z transforms (5.44) becomes

$$\begin{aligned} [\pi(1) + Z\pi(2) + \cdots] &= \mathbf{P}[\pi(0) + Z\pi(1) + \cdots] \\ Z^{-1}[\Pi(Z) - \pi(0)] &= \mathbf{P}\Pi(Z) \\ (\mathbf{I} - Z\mathbf{P})\,\Pi(Z) &= \pi(0) \\ \Pi(Z) &= (\mathbf{I} - Z\mathbf{P})^{-1}\,\pi(0) \end{aligned} \tag{5.46}$$

Thus we have expressed the general solution to the problem as a function of the matrix $\mathbf{P}$ times an initial-state vector. There will be values of Z for which the inverse matrix to $(\mathbf{I} - Z\mathbf{P})$ does not exist. These values Z_j are given by $\det(\mathbf{I} - Z_j\mathbf{P}) = 0$ or $\det(\mathbf{P} - Z_j^{-1}\mathbf{I}) = 0$. Clearly the Z_j^{-1} are the eigenvalues of $\mathbf{P}$. Utilizing Sylvester's theorem, then, we have

$$(\mathbf{I} - Z\mathbf{P})^{-1} \quad = \quad \sum_j \frac{\mathbf{c}_j \mathbf{r}_j}{1 - \frac{Z}{Z_j}} \tag{5.47}$$

Some modification to (5.47) is required if there are repeated eigenvalues. Equation (5.47) is essentially a partial fraction expansion. A typical term has the form

$$\frac{1}{1-\frac{Z}{Z_j}} \quad = \quad 1 - \frac{Z}{Z_j} + \left(\frac{Z}{Z_j}\right)^2 + \cdots$$

Thus coefficients at successive powers of Z decline with time in the form $(Z_j^{-1})^t$. It is clear that, if probabilities are to be bounded, the roots $1/Z_j$ must be inside the unit circle (recall minimum phase). We have already shown that one the of the roots Z_1 is always unity. This leads to the "steady-state" solution $1^t = 1$. In our particular example, one can see by inspection that the steady-state probability vector is $[0 \;\; 0 \;\; 0 \;\; 1]^T$ so the general solution is of the form

$$\pi(t) \quad = \quad \left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} [1 \;\; 1 \;\; 1 \;\; 1] + \sum_{j=2}^{4} Z_j^{-t} \mathbf{c}_j \mathbf{r}_j \right\} \pi(0)$$

Finally, a word of caution must be added. Occasionally defective matrices arise (incomplete set of eigenvectors) and for these the Sylvester theorem does not apply. In such cases, the solutions turn out to contain not only terms like Z_j^{-t} but also terms like tZ^{-t} and t^2Z^{-t}. It is the same situation as that applying to ordinary differential equations with constant coefficients. Ordinarily, the solutions are of the form $(r_i)^t$ where r_i is the ith root of the *indicial* equation but the presence of repeated roots gives rise to solutions like tr_i^t.