Transformation Matrices

Between

Non-Linear and Linear

Differential Equations

Robert L. Sartain

Associate Professor

Department of Mathematics and Physics

Houston Baptist University

Houston, Texas

ABSTRACT

In the linearization of systems of non-linear differential equations, we consider those systems which can be exactly transformed into the second order linear differential equation $Y''-AY'-BY=0$ where Y, Y', and Y'' are n x 1 vectors and A and B are constant n x n matrices of real numbers. We use the 2n x 2n matrix $M=\begin{bmatrix} 0 & I \\ B & A \end{bmatrix}$ to transform the above matrix equation into the first order matrix equation $X' = MX$. We study specifically the matrix M and the conditions which will diagonalize or triangularize M. We indicate transformation matrices P and P^{-1} to accomplish this diagonalization or triangularization and how to use these to return to the solution of the second order matrix differential equation system from the first order system.

We conclude with a study of the relationship between the diagonalization of M to that of the submatrices A and B.

Introduction

(Szebehely, 1976), (Bond and Horn, 1978), and (Bond, 1982) have studied the linearization of systems of non-linear differenial equations. In each of these cases the transformation of one system to the other involves square matrices A and B of order n.

In the present paper we consider only those non-linear differential equation systems which can be exactly transformed into the second order linear differential equation system $Y''-AY'-BY=0$ where the matrices A and B are constant matrices of real numbers and where Y, Y', and Y'' are n x 1 vectors.

We form the matrix $M = \begin{bmatrix} 0 & I \\ B & A \end{bmatrix}$ which is a matrix of order 2n which arises in the standard transformation of the second-order linear equation to a first-order linear equation by the transformation

$$\left. \begin{array}{c} Z_1 = Y \\[2em] Z_2 = Y' \end{array} \right\}$$ where Y is the vector $(Y_1, Y_2, ..., Yn)^T$ and $Y' = (Y_1', Y_2', ..., Yn')^T$.

Observe that this transformation yields $Z_1' = Y' = Z_2$ and $Z_2' = Y'' = AZ_2 + BZ_1$.

Written in matrix form we have

$$\begin{bmatrix} Z_1' \\ Z_2' \end{bmatrix} = \begin{bmatrix} O & I \\ B & A \end{bmatrix} \cdot \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$$

Letting $X = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$ and $M = \begin{bmatrix} O & I \\ B & A \end{bmatrix}$ the equation $Y'' = AY' + BY$ can be reduced to the linear first-order matrix equation $X' = MX$.

This research is concerned with the following questions:

(1.) What are the conditions on the matrix M which will allow us to solve the matrix differential equation system $X' = MX$.

(2.) Under the assumption that this system can be solved, how does one actually recover the matrices to return to the original second order differential equation system $Y'' - AY' - BY = 0$.

This paper shall try to answer these questions as completely as possible.

Theory

The solution of the first-order linear differential equation system $X' = MX$ depends heavily on the matrix M given. We consider two cases. The most satisfactory solution occurs when the matrix M is similar to a diagonal matrix and much of our research has centered upon what must be known about the submatrices A and B of $M = \begin{bmatrix} 0 & I \\ B & A \end{bmatrix}$ in order to assure that M can be diagonalized. We give some elementary linear algebra theory to familiarize everyone with the ideas.

For a linear homogeneous transformation such as M above, there exist scalars λ_i and vectors X_i which satisfy the equation $MX_i = \lambda_i X_i$. The values of λ_i for which this equation is satisfied are called the eigenvalues of M and the vectors X_i which are fixed under the transformation M for each λ_i are called the eigenvectors. There are many other names associated with these values and vectors, some of which are characteristic values and vectors, latent values and vectors, and proper values and vectors.

Clearly, the zero vector will always satisfy the equation $MX = \lambda X$ for any λ chosen. However, we desire to find nontrivial solutions to the problem. Rewriting the equation as $(M-\lambda I) X = 0$ we see that any nontrivial vector X will satisfy the equation if and only if $\det (M-\lambda I) = 0$ where det stands for the determinant of the matrix $M-\lambda I$.

The determinant of this is a polynomial equation in λ. Since, in our case, M is of order 2n, the polynomial equation $f(\lambda) = 0$ will be of degree 2n, and hence there will be 2n eigenvalues associated with M. These values may or may not be distinct, and we shall discuss the consequences later in the paper.

For each distinct eigenvalue, λ, satisfying $f(\lambda) = \det(M-\lambda I) = 0$ there will exist at least one non-trivial eigenvector X satisfying $(M-\lambda I) X = 0$. (Stein, 1967).

The eigenvalues and eigenvectors play a central role in the solution of the differential equation $X' = MX$ and this is our purpose in discussing them here.

A matrix M is said to be similar to the matrix C if there exist a nonsingular matrix P such that $M = P C P^{-1}$.

The matrix M is diagonalizable if it is similar to a diagonal matrix

$$D = \begin{bmatrix} d_1 & & O \\ & d_2 & \\ & & \ddots \\ O & & d_n \end{bmatrix}$$

where the large zeros indicate that all elements off the main diagonal are zeros.

Suppose now that the matrix M of our system $X' = MX$ is known to be diagonalizable, then there exists a nonsingular matrix P such that $P^{-1}MP = D(\lambda_1, ..., \lambda_{2n})$ where the notation $D(\lambda_1, ..., \lambda_{2n})$ will always mean a diagonal matrix where the $\lambda_1, \lambda_2, ..., \lambda_{2n}$ are the eigenvalues of M and they lie on the main diagonal of D. See Theorems 6.7.1 and 6.8.1 of (Stein, 1967) for a proof of the above statement.

If we now set $X = PZ$ where P is the matrix which diagonalizes M, then $X' = PZ'$ and $Z = P^{-1}X$. So $PZ' = X' = MX$ or $Z' = P^{-1}MX = (P^{-1}MP) Z$ Thus $Z' = DZ$ where D is our diagonal matrix. Using elementary differential equation theory, it is known that a solution of $Z' = DZ$ is

$$Z = \begin{bmatrix} c_{11} e^{\lambda_1 t} \\ \vdots \\ c_{2n,2n} e^{\lambda_{2n} t} \end{bmatrix}$$

So X = PZ will then solve the original system simply by multiplying the matrix P by the vector Z. Thus it is of paramount importance first ᵗ determine whether or not the original matrix M can be diagonalized and if it can, then how does one obtain the matrix P which both diagonalizes M and gives the final solution for the X once we have Z as shown above.

Now consider the second case where the matrix M is not diagonalizable. Every square matrix M is similar to a triangular matrix T with the eigenvalues of M as the diagonal elements of T. A constructive proof of this fact is given in (Stein, 1967) as Theorem 6.8.5. The process again generates a nonsingular matrix P such that $P^{-1}MP = T$. Using the same construction of X = PZ given above, by elementary differential equation theory (Murdoch, 1957), one can solve the system for the original vector $X = (X_1, X_2, ..., X_{2n})^T$. If we assume the matrix T is upper triangular then after computing PZ, we will have $X_{2n} = C_{2n,2n} e^{\lambda_{2n} t}$ as in the preceding case, and in general depending on the multiplicity of the eigenvalues, if λ is an eigenvalue of multiplicity r, then the solution for this case will be of the form $e^{\lambda t} p(t)$ where $p(t)$ is a polynominal of degree r-1. Thus a solution is possible even in this case although not as easy to obtain or use. If the λ's are distinct then, for example, if $X_{2n} = C_{2n,2n} e^{\lambda_{2n} t}$, then the solution for

$$X_{2n-1} = C_{2n-1, 2n-1} e^{\lambda_{2n-1} t} + C_{2n-1, 2n} e^{\lambda_{2n} t}$$

where the c_{ij}'s are the nonzero elements in the upper triangular matrix which correspond to each row i and column j position. One can use these solutions then to substitute in and solve for the next X_i, and so on, until a complete solution is obtained.

Results

We next describe the process of actually obtaining P which will diagonalize or triangularize the matrix to give us the solutions indicated above. We consider each of the cases separately.

Suppose first that M is diagonalizable, then there exists a set of 2n linearly independent eigenvectors, say U_1, U_2, ..., U_{2n} and this is true regardless of whether or not any eigenvalue has multiplicity greater than one or not. See (Edelen, 1976) for a discussion of this property. If we suppose that U_1 corresponds to λ_1, U_2 to λ_2, ..., U_{2n} to λ_{2n} even if some of the λ 's are equal, then the matrix $P = (U_1, U_2, ..., U_{2n})$ where each column of P is a vector of length 2n will be nonsingular since these 2n vectors are linearly independent. We observe that if $\lambda_i = \lambda_j$ for instance, there will still be distinct eigenvectors U_i and U_j which are linearly independent when M is diagonalizable. P nonsingular implies that it has an inverse and the product $P^{-1}MP$ will actually be a diagonal matrix with the eigenvalues down the main diagonal.

We will address the process of determining the eigenvectors for a given λ later in the paper.

We now consider the problem of finding the nonsingular matrix P when the matrix M is <u>not</u> diagonalizable.

The process is given as follows. <u>Process:</u>

    (1.) Find an arbitrary eigenvector, X_1 for the first eigenvalue λ_1 of your matrix. (We shall assume λ_1, λ_2, ..., λ_{2n} are successively on the diagonal from top left to lower right.)

(2.) Form $P_1 = (X_1, e_2, ..., e_{2n})$ provided that P so formed is nonsingular.

(3.) Compute P_1^{-1} and calculate the product $P_1^{-1} M P_1$ where M is our original matrix.

(4.) Using $P_1^{-1} M P_1$ we now have λ_1 in upper left corner. Call the submatrix formed by crossing out the row and column which contains λ_1, M_1. If M_1 is a 2x2 matrix skip to step 10. Otherwise compute an eigenvector for λ_2 using M_1 (not M, but it will be the same eigenvalue, λ, as for M).

(5.) Form $P_2 = (X_2, e_3, ..., e_{2n})$ where X_2 is the 2n-1 eigenvector found in step 4.

(6.) Compute P_2^{-1} and calculate the product $P_2^{-1} M_1 P_2$. This matrix is of order 2n-1.

(7.) $P_2^{-1} M_1 P_2$ will now have λ_2 in the upper left hand corner and zeros in the column below it.

(8.) Form the submatrix M_2 by crossing out the row and column containing λ_2.

(9.) If M_2 is 2x2 goto step 10. If not continue the process as illustrated in steps 5-8 until a submatrix which is 2x2 is finally obtained.

(10.) When M_j is reached which is 2x2 the process will be changed as follows:

Process (Continued)

(a) Compute an eigenvector using M_j for λ_{2n-1}.

(b) Compute an eigenvector using M_j for λ_{2n}.

(c) Let $P_{j+1} = (X_{2n-1}, X_{2n})$. Note it does not contain standard basis vectors as in previous cases.

(11.) Finally form the matrix P as follows:

$$P = P_1 \cdot \begin{bmatrix} I_1 & O \\ O & P_2 \end{bmatrix} \cdot \begin{bmatrix} I_2 & O \\ O & P_3 \end{bmatrix} \cdots \begin{bmatrix} I_{n-2} & O \\ O & P_{2n-1} \end{bmatrix} \quad where \quad I_1 = 1, \quad I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, etc.$$

(12.) Compute P^{-1}. Then the final product $P^{-1}MP$ will triangularize M and leave the eigenvalues on the diagonal of $P^{-1}MP$. Note that this form will create an upper triangular matrix.

Note: Each successive eigenvector is found by reducing $M_j - \lambda I$ to canonical form for j = 0, ...,2n-1. Then multiplying the reduced matrix by the appropriately sized vector made up of the X_j 's to be in the eigenvector. Solve this system of linear equations by choosing appropriate values for the arbitrary X's. You can choose X_i = 1 for arbitrary X's if so desired. This then will make up the necessary eigenvector for that λ.

In determining eigenvectors for an M which is diagonalizable, there are two cases to consider. The first case is that for which all eigenvalues are distinct. In this case use Gaussian elimination to reduce the matrix $M - \lambda I$ to canonical form (we use this name to imply that the matrix has ones on the main diagonal and zeros off the main diagonal in so far as possible and always zeros below the main diagonal). Other authors use the phrase "row echelon" form to describe this.

If C is the row echelon form of $M - \lambda I$, then setting $CX = 0$, where X is the 2n x 1 vector $X = (X_1, X_2, ..., X_{2n})^T$, will allow us to obtain the elements X_1, X_2, ..., X_{2n} of the eigenvector for λ by solving the homogeneous system $CX = 0$. Any convenient choice of the arbitrary X's will give us an eigenvector for λ. Since all λ's are distinct in this case, then each of the eigenvectors will be linearly independent so the matrix P formed from the eigenvectors will be nonsingular.

The second case is that where some of the eigenvalues have multiplicity greater than one. In this case, if λ_i is an eigenvalue of multiplicity k_i then there will be k_i-linearly independent eigenvectors associated with each such λ_i. Recall that we are assuming the diagonalizability of M at this point, otherwise we would not know that this is possible. For each i, one can obtain distinct eigenvectors by choosing different values for the arbitrary variables in the equation $CX = 0$ when solved. Each of these eigenvectors will not only be linearly independent of each other but also linearly independent of eigenvectors obtained from distinct eigenvalues. See the proof of Theorem 6.8.4 of (Stein, 1967) for a proof of this fact.

Thus all 2n of the eigenvectors so obtained will be linearly independent and thus the matrix P formed from the eigenvectors as previously indicated will be nonsingular.

We have made the assumption in the above work that M is diagonalizable. Consider now the question of what conditions on M will guarantee that M is diagonalizable.

The most well-known theorem says that every real symmetric matrix is diagonalizable (Stein, 1967). See Theorem 6.8.6. However, our matrix $M = \begin{bmatrix} O & I \\ B & A \end{bmatrix}$ is in general not symmetric so we seek other criterion.

A second theorem (6.8.2) indicates that M is diagonalizable if and only if M has 2n linearly independent eigenvectors.

The difficulty with using this theorem is that we would like to determine whether or not M is diagonalizable before we compute the eigenvectors so we will know wh ther to use the diagonalization or triangularization process which we have described above in finding the matrix P and its inverse.

One further theorem (6.8.4.) tells us M is diagonalizable if and only if for each λ_i, the multiplicity of λ_i is equal to k=2n-r where r is the rank of $M-\lambda_i I$. (Stein, 1967).

Although this theorem is better than those above and works on all matrices, it still requires checking the multiplicity of every λ_i against the rank of $M-\lambda_i I$ which could be a rather formidable task.

Thus we look specifically at the form of our matrix $M = \begin{bmatrix} O & I \\ B & A \end{bmatrix}$ to determine, if possible, conditions on B and A which will help us decide the diagonalizability of M. To this end, consider the following little known theorem from (Hohn, 1964).

If P is m x n , Q is m x n, R is n x m and S is n x n and nonsingular, then

$$\det \begin{bmatrix} P & Q \\ R & S \end{bmatrix} = \det S \cdot \det (P - QS^{-1}R).$$

We now assume that P, Q, R, and S above are all square of size n x n and also that SQ = QS. We can then prove the following theorem.

$$\det \begin{bmatrix} P & Q \\ R & S \end{bmatrix} = \det (PS - QR)$$

__Proof:__ By the first theorem above $\det \begin{bmatrix} P & Q \\ R & S \end{bmatrix} = \det S \cdot \det (P - QS^{-1}R)$. Then using the well-known theorem that det AB = det A $\cdot$ det B when A and B are square n x n matrices we apply this to the right member above to have det S $\cdot$ det $(P - QS^{-1}R) = \det$ $(S(P - QS^{-1}R)) = \det (SP - SQS^{-1}R)$. Since we are assuming SQ = QS we have

$$\det (SP - SQS^{-1}R) = \det (SP - QSS^{-1}R) = \det (SP - QR) \text{ since } S \cdot S^{-1} = I.$$

We now apply this theorem to our matrix $M = \begin{bmatrix} O & I \\ B & A \end{bmatrix}$. Since I, the identity matrix, commutes with every matrix of the same size, we have IA = AI so det M = $|O \cdot A - B \cdot I|$ = det $(-B) = (-1)^n \cdot \det (B)$.

Considering the eigenmatrix M $- \lambda$I we have det (M $- \lambda$I) = det $\begin{bmatrix} -\lambda I & I \\ B & A - \lambda I \end{bmatrix}$ =

$$\det ((-\lambda I)(A - \lambda I) - B) = \det (-\lambda A + \lambda^2 I - B) = \det (\lambda^2 I - \lambda A - B).$$

Our purpose in looking at det (M $- \lambda$I) is to observe that det (M $- \lambda$I) = 0 is the characteristic equation of the matrix M, but since det (M $- \lambda$I) = det(λ^2I $- \lambda$A - B), then det (λ^2 I $- \lambda$A-B) = 0 is also the characteristic equation for M, and it is a polynomial equation in terms of the matrices A and B which are submatrices of the original matrix M. Assume first A = 0, then M = $\begin{bmatrix} O & I \\ B & O \end{bmatrix}$ and even though we cannot use the theorem just proved (S = 0 is not nonsingular) we can still see that det M = $(-1)^n \cdot$ det B

by Laplace's expansion. To compute the determinant of $M - \lambda I$ in this case we have det $(M - \lambda I) = \det \begin{bmatrix} -\lambda I & I \\ B & -\lambda I \end{bmatrix}$ and since $-\lambda I$ is nonsingular, the theorem does apply in this case to give us

det $(M - \lambda I) = \det(\lambda^2 I - B) = (-1)^n \det(B - \lambda^2 I)$. Since for a fixed n, $(-1)^n$ is a constant, it will not affect the eigenvalues in the polynomial characteristic equation. Thus if μ_i, $i = 1, ..., n$ are the eigenvalues of B, then $\lambda_i = \pm \sqrt{\mu_i}$ will be the eigenvalues of M. (Note: det $(B - \mu_i I) = 0$ implies det $(M - \lambda_i I) = 0$.)

Assume now that B is diagonalizable. Then for each μ_i, $i = 1, 2, ..., n$ there exists an eigenvector U_i, $i = 1, 2, ..., n$ such that the set $(U_1, U_2, ..., U_n)$ of eigenvectors is linearly independent. Thus let $T = (U_1, U_2, ..., U_n)$ be the matrix of eigenvectors as columns in the matrix. Then $T^{-1}BT = D(\mu_1, \mu_2, ..., \mu_n)$ will be the diagonal matrix of eigenvalues.

Let $S = \begin{bmatrix} \lambda_1 & & O \\ & \lambda_2 & \\ & & \ddots \\ O & & \lambda_n \end{bmatrix}$ be the matrix where the

λ_i's here are the <u>positive</u> square roots of the μ_i's and thus are half of the eigenvalues of M. Consider first the case where $\mu_i \neq 0$ for all $i = 1, 2, ..., n$. Then S^{-1} exists and

is given by $S^{-1} = \begin{bmatrix} \frac{1}{\lambda_1} & & O \\ & \frac{1}{\lambda_2} & \\ & & \ddots \\ O & & \frac{1}{\lambda_n} \end{bmatrix}$. In this case let $P = \begin{bmatrix} TS^{-1} & TS^{-1} \\ T & -T \end{bmatrix}$ and $P^{-1} = \frac{1}{2}\begin{bmatrix} ST^{-1} & T^{-1} \\ ST^{-1} & -T^{-1} \end{bmatrix}$

or $P^{-1}MP = \frac{1}{2}\begin{bmatrix} ST^{-1} & T^{-1} \\ ST^{-1} & -T^{-1} \end{bmatrix} \cdot \begin{bmatrix} O & I \\ B & O \end{bmatrix} \cdot \begin{bmatrix} TS^{-1} & TS^{-1} \\ T & -T \end{bmatrix} = \frac{1}{2}\begin{bmatrix} T^{-1}BTS^{-1}+S & T^{-1}BT^{-1}S^{-1}-S \\ -T^{-1}BTS^{-1}+S & -T^{-1}BTS^{-1}-S \end{bmatrix}$

Computing $T^{-1}BTS^{-1}$ we obtain: $\begin{bmatrix} \mu_1 & & O \\ & \mu_2 & \\ & & \ddots \\ O & & \mu_n \end{bmatrix}\begin{bmatrix} \frac{1}{\lambda_1} & & O \\ & \frac{1}{\lambda_2} & \\ & & \ddots \\ O & & \frac{1}{\lambda_n} \end{bmatrix} = \begin{bmatrix} \lambda_1^2 & & O \\ & \lambda_2^2 & \\ & & \ddots \\ O & & \lambda_n^2 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\lambda_1} & & O \\ & \frac{1}{\lambda_2} & \\ & & \ddots \\ O & & \frac{1}{\lambda_n} \end{bmatrix}$

$$= \begin{bmatrix} \lambda_1 \lambda_2 & O \\ O & \ddots \lambda_n \end{bmatrix} = S. \quad \text{Since } T^{-1}BT \text{ diagonalizes B to its eigenvalue matrix.}$$

Thus $P^{-1}MP = \frac{1}{2} \begin{bmatrix} S+S & S-S \\ -S+S & -S-S \end{bmatrix} = \begin{bmatrix} S & O \\ O & -S \end{bmatrix}$.

Hence this choice for P will diagonalize M where the positive square roots of λ_i are down the diagonal of the upper-left n x n matrix and the negative square roots of λ_i are the diagonal elements of the lower right n x n matrix. Thus we have shown how to diagonalize M given any diagonalizable B provided the eigenvalues of B are nonzero. Thus M is diagonalizable when B is diagonalizable provided none of B's eigenvalues are zero, and provided that A = 0.

Now suppose A = β I, where β is any nonzero scalar. We then obtain

$$\det (M - \lambda I) = \det (\lambda^2 I - \lambda (\beta I) - B)$$
$$= \det((\lambda^2 - \lambda \beta) I - B)$$
$$= (-1)^n \cdot \det (B - (\lambda^2 - \lambda \beta)I).$$

Again let λ_i, i = 1, 2, ..., n be the eigenvalues of B. We then have $\lambda_i^2 - \lambda_i \beta = \mu_i$ or $\lambda_i^2 - \lambda_i \beta - \mu_i = 0$, for i = 1, 2, ..., n. By solving these quadratic equations we again obtain the eigenvalues for M from the eigenvalues for B.

If we assume $\mu_i \neq 0$ for all i and that all of the eigenvalues for B are distinct, then the eigenvalues for M will be distinct, and thus M is diagonalizable.

References

1. Bond, V. R., "Error Propagation in the Numerical Solutions of the Differential Equations of Orbital Mechanics", Celestial Mechanics, Vol. 27, No. 1 pp. 65-77, May 1982.

2. Bond, V. R. and Horn, M. K., "An Explicit Solution of the Linear Variational Equations of the Two-Body Problem", NASA 78-FM-47, JSC-14440, September 1978.

3. Edelen, D. G. B. and Kydoniefs, A. D. An Introduction to Linear Algebra for Science and Engineering, Elsevier, Lectures 12, 13 and 17, 1976.

4. Hohn, F. E., Elementary Matrix Algebra, Macmillan, Chapters 2, 4 and 8, 1964.

5. Murdoch, D. C, Linear Algebra for Undergraduates, John Wiley & Sons, Chapter 7, 1957.

6. Stein, F. M., Introduction to Matrices and Determinants, Wadsworth, Chapters 6 and 7, 1969.

7. Szebehely, V., "Three Theorems on Relative Motion", Celestial Mechanics, Vol. 14, pp. 113-119, 1976.