This is to certify that the

thesis entitled

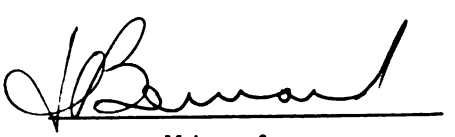
DESIGN SENSITIVITY IN DYNAMICAL SYSTEMS

presented by

JOSEPH EUGENE WHITESELL, JR.

has been accepted towards fulfillment
of the requirements for

PH.D. degree in MECHANICAL ENGINEERING



Major professor

Date 8-8-80

DESIGN SENSITIVITY IN DYNAMICAL SYSTEMS

By

Joseph Eugene Whitesell, Jr.

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Mechanical Engineering

1980

ABSTRACT

DESIGN SENSITIVITY IN DYNAMICAL SYSTEMS

By

Joseph Eugene Whitesell, Jr.

Modal analysis is a standard tool used in the design of dynamical systems. By numerically determining certain eigenvalues and eigenvectors associated with a linear structural model, its vibratory behavior can be investigated. The difficulty with this technique is that it may be costly in computer time if several redesigns (and re-evaluations of the eigenvalues and eigenvectors) are needed to find a suitable design.

The thesis presents a technique to represent selected eigenvalues and eigenvectors by a Taylor series in a design variable. Since, with the technique, the Taylor coefficients can be computed efficiently and to arbitrary order, the Taylor series can be used to accurately estimate the consequences of a design change. The technique, therefore, has the potential to substantially reduce the number of calculations necessary to re-evaluate the eigensystem after a design change.

Each Taylor coefficient is determined by evaluating a recursive formula which is derived through an application of generalized inverse theory. If the eigenvector is found through an inverse iteration, the technique is especially efficient. Most of computational effort spent to find the eigenvector can then be re-applied to find the Taylor coefficients. Two representative examples illustrate the power of the method for practical design problems.

DEDICATION

To my father and mother.

ACKNOWLEDGEMENTS

There are several people who deserve acknowledgement for helping me during my studies and through the preparation of this dissertation. The advice and friendship of Dr. James Bernard, my major professor, has been invaluable. His quick mathematical insight and his engineering intuition as well as his high personal values and his good humor have made working with him stimulating and enjoyable. It was Dr. Albert Andry who first stimulated my interest in vibration problems and in the topic of this thesis. His ability as a lecturer and his good advice and friendship were major influences during the past three years. Both Drs. Andry and Bernard were cited last year by Michigan State University for excellence-in-teaching. Dr. Robert Schlueter, who was my advisor during my Master's degree studies, and a member of my thesis committee, has helped me numerous times during my studies here and I am greatly indebted to him. The West Coast Branch of my thesis committee consists of Herb, Margie, Marta, Erik, Evan and Lorin Rauch who have been good friends and they have enriched my life in many ways.

I want to thank especially Dr. J.S. Frame for introducing me to linear algebra through his extraordinary teaching ability. Attending his course was a rare privilege.

For the difficult job of typing this thesis I thank Jan Swift of the Department of Mechanical Engineering. I appreciate the programming help of Mr. Mark Zykin of the Case Center for Computer-Aided Design. I have much enjoyed working with Mark. Finally, I would like to thank Dr. John Brighton, Chairman of the Department of Mechanical Engineering, for his consistant support and leadership and for his advice.

TABLE OF CONTENTS

Chapter 1 -		Introduction	1
	1.1	Some Historical Notes	2
	1.2	Thesis Overview	3
Chapter 2 -		Mathematical Preliminaries	6
	2.1	Concepts from Linear Algebra	7
	2.2	Concepts from Functional Analysis	19
Chapter 3 -		Eigensystem Differentiability	25
	3.1	Preliminary Differentiability Results	25
	3.2	The Reduction Process	28
	3.3	Eigenvector Differentiability	32
	3.4	Differentiation with Respect to Several Parameters	37
	3.5	Conclusions	42
Chapter 4 -		Differentiation of Eigensystems	43
	4.1	Differentiation of Fully Analytic Matrices	43
	4.2	Eigensystem Derivatives with Generalized Inverse	49
	4.3	Numerical Approach	59
Chapter 5 -		Design Sensitivity in Finite Element Models of Dynamical Structures	64
	5.1	The Finite Element Method	64
	5.2	Dynamical Problems	70
	5.3	Examples	78
Chapter 6 -		Discussion and Further Study	94
	6.1	Discussion	94
	6.2	Methods	95
	6.3	Suggestions for Further Study	96
Bibliography			97

LIST OF TABLES

Table		Page
5.3-6	The Effect of a Design Change on an Eigenvalue	85

LIST OF FIGURES

Figure		Page
5.1-11	The Triangulation of a Domain	68
5.3-2	Triangular Element	80
5.3-7	Eigenvalue Dependence on Design	86
5.3-8	Baseline Design	89
5.3-9	First Design Change	90
5.3-10	Second Design Change	91
5.3-11	Third Design Change	92
5.3-12	Fourth Design Change	93

Chapter 1

Introduction

The numerical determination of static and dynamic responses for linear structural models has become a routine procedure in engineering design [1,2]. Typically a sequence of design and redesign is followed until certain features such as weight, natural frequency, mode shape, and deflection and stress magnitudes attain suitable values. At each stage the designer changes the design and performs a numerical analysis to evaluate the consequences of the changes.

This approach relies heavily on human judgement to modify the design at each stage and since the appropriate design modifications may not be obvious, several redesigns and reanalyses are often necessary. Alternative methods to improve this procedure have been proposed. In one method, the full structural model is replaced by a Taylor series which approximates its behavior in a neighborhood of the current design [3]. Combined with interactive computer graphics hardware, this method allows a designer to preview the consequences of a proposed design while avoiding a costly reanalysis. Another approach is the optimal design method [3], which combines nonlinear programming methods with structural design techniques. The procedure identifies a cost functional with certain important design features and

attempts to minimize it by changing the design. When the method is successful, a sequence of numerically determined designs converge to a feasible design for which the selected cost functional attains a local minimum.

A major difficulty arises when these methods are applied to large problems since the determination of the system's state variables and especially their sensitivity values may require an excessive amount of computation. However, the potential benefits of these methods provide a strong incentive to develop improved numerical methods.

1.1 Some Historical Notes

When the design involves dynamical measures such as natural frequencies or mode shapes, efficient methods for determining the sensitivities of eigenvalues and eigenvectors to design changes become important. Although methods for computing the first derivative of an eigenvalue with respect to some system parameter have been known since the nineteenth century work of Jacobi [4], eigenvector derivatives have a much shorter history. The first work in this area was done by Fox and Kapoor [5], who presented two methods for determining an eigenvector derivative for the symmetric eigenvalue problem $(K - \lambda M)u = 0$. Unfortunately both methods are computationally inefficient. The first method, although it requires only knowledge of the specified eigenvalue and eigenvector, involves the multiplication of an $(n + 1)xn$ matrix by its transpose and the subsequent solution of a fully

populated symmetric system. The matrix multiplication is a lengthy computation and usually leads to a loss of any sparseness possessed by the factor matrices. The second method avoids these problems, but it expresses the derivative of an eigenvector in terms of a complete set of eigenvalues and eigenvectors. This formulation has only theoretical value when the eigensystem is large since it is difficult to extract a full set of eigenvectors. In refs. [6-8] the work of Fox and Kapoor was extended to non-symmetric systems but no improvement in computational efficiency was made.

These computational difficulties were first discussed by Nelson [9]. As a remedy, he proposed a technique by which the rank $(n - 1)$ matrix $L - \lambda I$ is modified by zeroing certain of its entries. The modified matrix, which describes a system of equations which must be solved to determine the eigenvector derivative, is well conditioned and retains the sparseness of the original system. This is a significant improvement since the problems of mathematical physics and engineering often involve sparse matrices. In this thesis further improvement in computing eigensystem sensitivities are presented. Two representative examples illustrate the power of the method for practical design problems.

1.2 Thesis Overview

Notation and preliminary mathematical concepts are presented in Chapter II. In Chapter III a survey of eigensystem differen-

tiability results are presented. The presentation is based on the literature of the perturbation theory for linear operators. After two example problems the chapter closes with a discussion of eigensystem dependence on several parameters.

Chapter IV is concerned with methods for differentiating eigenvalues and eigenvectors. The chapter states some basic results, followed by an application of generalized inverse theory which leads to further improvements in computing eigensystem derivatives. A recursive numerical method is then described for efficiently computing eigensystems derivatives of arbitrary order. Each eigenvalue and eigenvector derivative is determined by solving a sparse triangular system and involves no more than $O(n^2)$ multiplications.

Chapter V begins with a statement of the design sensitivity problem in finite element models. The methods of Chapter IV are then extended to the generalized eigenvalue problem $(K - \lambda M)u = 0$. The chapter ends with two examples which involve finite element formulations.

In the first Example (5.3-1), an eigenvalue problem associated with a single element plane elastic vibration problem is considered. After introducing a design variable to the problem, a Taylor series representing an eigenvalue as a function of the design variable is determined with the technique presented in this thesis. The results are compared with direct evaluations of the eigenvalue at various values of the design variable.

In the second example (5.3-7) an eigenvalue which depends on the boundary shape of a fixed-fixed vibrating plate is estimated using a first order approximation. The estimated eigenvalue is compared with a direct evaluation of the eigenvalue for several design modifications.

Chapter VI presents concluding remarks and suggestions for future work.

Chapter 2

Mathematical Preliminaries

This chapter contains a summary of definitions and theorems drawn from linear algebra and functional analysis which are used in the following chapters. The reader may find this material in many sources such as refs. [10-14].

In this thesis an n dimensional column vector is denoted by x. Its transpose to a row vector is denoted by x^T. The conjugation of x^T is denoted by x*. To avoid confusion between the ith vector in an indexed sequence of vectors and the ith component in a particular vector, $\underline{x}_i$ denotes an indexed vector and x_i denotes the ith component of the vector x. The jth component of the vector $\underline{x}_i$ is denoted by x_{ij}. An nxm matrix is denoted by A, B, C etc. The entry in the ith row and jth column of A is denoted by a_{ij}. The ith matrix in an indexed sequence of matrices is denoted by A_i. The entry in the jth row and kth column of the matrix A_i is denoted by a_{ijk}.

The real (complex) field is denoted by $\mathcal{R}(\mathcal{C})$. The real (complex) vector space of dimension n is denoted by $\mathcal{R}^n(\mathcal{C}^n)$. The space of real (complex) nxm matrices is denoted by $\mathcal{R}^{nxm}$ $(\mathcal{C}^{nxm})$.

2.1 Concepts from Linear Algebra

Definition (2.1-1): A linear space X is a set of elements called vectors which is closed under addition (if $x \in X$ and $y \in X$ then $x+y = z \in X$) and under multiplication by a real or complex scalar α (if $x \in X$ then $\alpha x = y \in X$).

Definition (2.1-2): Let X be a linear space over a field F and let V be a set of k vectors $\{\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_k\}$ of X. Then if for some set of k scalars $\{c_1, c_2, \ldots, c_k\}$ from F, not all zero, the linear combination

$$c_1\underline{x}_1 + c_2\underline{x}_2 + \ldots + c_k\underline{x}_k = 0,$$

the vectors $\{\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_k\}$ are said to be linearly dependent over F. If instead

$$c_1\underline{x}_1 + c_2\underline{x}_2 + \ldots + c_k\underline{x}_k = 0$$

only if each $c_i = 0$ then the vectors $\{\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_k\}$ are linearly independent.

Definition (2.1-3): A set of vectors $\{\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_k, \ldots\}$ is said to span X if every vector $x \in X$ can be expressed as a linear combination of the set $\{\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_k, \ldots\}$.

Definition (2.1-4): A set of vectors $\{\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_k\}$ of a linear space V is said to be a Hamel basis if and only if

i) the set spans V and

ii) the vectors $\{\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_k\}$ are linearly independent.

Definitions (2.1-1) through (2.1-4) are concerned with only the algebraic properties of a linear space (algebraic linear space). If these notions are combined with the topological notions of length, distance and convergence, a space with both algebraic and topological properties results (topological linear space).

Definition (2.1-5): A norm on a linear space X is a positive real valued function $||\cdot||$ with the following properties

i) $||x|| \neq 0$ if $x \neq 0$, $||x|| = 0$ otherwise

ii) $||\alpha x|| = |\alpha| \cdot ||x||$, $\alpha \in F$

iii) $||\underline{x}_1 + \underline{x}_2|| \leq ||\underline{x}_1|| + ||\underline{x}_2||$

Definition (2.1-6): An inner product is a scalar function, $(\cdot,\cdot)$, of two elements $x,y \in X$ such that

i) $(x_1,\underline{x}_2 + \underline{x}_3) = (\underline{x}_1,\underline{x}_2) + (\underline{x}_1,\underline{x}_3)$

ii) $(\underline{x}_1,\underline{x}_2) = (\underline{x}_2,\underline{x}_1)^*$ (* denotes conjugation)

iii) $(\underline{x}_1, \alpha\underline{x}_2) = \alpha(\underline{x}_1,\underline{x}_2)$, $\alpha \in F$

iv) $(x,x) > 0$ if $x \neq 0$, $(x,x) = 0$ otherwise.

Definition (2.1-7): Two vectors x and y are orthogonal with respect to an inner product $(\cdot,\cdot)$ if

$$(x,y) = 0.$$

Definition (2.1-8): A set of vectors $\underline{x}_i$ forms an orthonormal set with respect to an inner product $(\cdot,\cdot)$ if

$$(\underline{x}_i,\underline{x}_j) = \delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

Definition (2.1-9): Two sets of vectors $\underline{x}_i$ and $\underline{y}_i$ form a bi-orthonormal set with respect to an inner product $(\cdot,\cdot)$ if

$$(\underline{x}_i,\underline{y}_j) = \delta_{ij}.$$

Definition (2.1-10): The transpose of a matrix A is the matrix $A^T = B$ such that $a_{ij} = b_{ji}$. The Hermitian transpose of a matrix A is the matrix $A^* = B$ such that $a_{ij} = b^*_{ji}$.

Definition (2.1-11): A matrix A is called symmetric (Hermitian) if $A = A^T$ $(A = A^*)$ or skew-symmetric (skew-Hermitian) if $A = -A^T$ $(A = -A^*)$.

Definition (2.1-12): A matrix A is upper triangular if $a_{ij} = 0$ for $i > j$, lower triangular if $a_{ij} = 0$ for $i < j$ and diagonal if $a_{ij} = 0$ for $i \neq j$.

Definition (2.1-13): The matrix A whose elements are all zero with the exception that $a_{ij} = 1$ is given the special symbol, $A = e_{ij}$ and is called the ij matrix unit.

Definition (2.1-14): The product of an mxk matrix A and an kxn matrix B is the matrix C = AB such that

$$c_{ij} = \sum_{p=1}^{k} a_{ip}b_{pj}.$$

Definition (2.1-15): For positive integers, the powers of an nxn matrix $A=A^1$ are defined by

$$A^0 = I,\ A^2 = A^1 \cdot A^1, \ldots, A^n = A^{n-1} \cdot A^1.$$

Definition (2.1-16): A matrix A is idempotent if $A^2 = A$.

Definition (2.1-17): If the negative integral powers of the nxn matrix A exist, they are defined by

$$A^{-n} = (A^{-1})^n$$

where A^{-1} is called the inverse of A and

$$A^{1}A^{-1} = A^{0} = I.$$

If A^{-1} exists the matrix A is called non-singular otherwise it is called singular.

Definition (2.1-18): The rank r of an mxn matrix A is the maximum number of linearly independent columns of A.

Definition (2.1-19): A generalized inverse of an mxn matrix A of rank r is any nxm matrix A^{I} of rank r such that

$$AA^{I}A = A.$$

The mxm matrix AA^{I} and the nxn matrix $A^{I}A$ are each idempotent matrices [10].

Theorem (2.1-20): Any solution X of the linear system

$$AX = Y$$

where A is an mxn matrix of rank r

X is an nxk matrix

Y is an mxk matrix

may be expressed as

$$X = A^{I}Y + A_{0}Z$$

where A_0 satisfies $AA_0 = 0$ and Z is arbitrary [10].

Definition (2.1-21): The scalars λ and corresponding vectors u which satisfy the equation

$$Lu = \lambda u$$

are called the <u>eigenvalues</u> and corresponding <u>right eigenvectors</u> of the matrix L. The eigenvectors of L^T are called the <u>left eigenvectors</u> of L. The subspace consisting of the origin and all the right (left) eigenvectors corresponding to λ is called the right (left) eigenspace of L corresponding to λ.

Theorem (2.1-22): The eigenvalues of a real symmetric matrix are real.

Theorem (2.1-23): Corresponding to any real symmetric matrix is a set of orthonormal eigenvectors.

Definition (2.1-24): A matrix A is <u>similar</u> to a matrix B if there exists a matrix S such that

$$B = S^{-1}AS.$$

Theorem (2.1-25): If A and B are similar then they have the same eigenvalues.

Definition (2.1-26): A matrix A is called <u>diagonable</u> if it is similar to a diagonal matrix such that $S^{-1}AS = \Lambda$, where the diagonal matrix Λ has the eigenvalues λ_i of A as its entries and where the ith column of S and the ith row of S^{-1} are respective right and left eigenvectors corresponding to λ_i.

Definition (2.1-27): A matrix A such that $AA^* = A^*A$ is called a <u>normal</u> matrix.

Definition (2.1-28): A matrix S is called <u>orthogonal</u> (unitary) if $S^TS = I$ ($S^*S = I$).

Theorem (2.1-29): Every real symmetric matrix A may be diagonalized by an orthogonal matrix S

$$S^TAS = \Lambda$$

where the diagonal matrix Λ has the eigenvalues, λ_i of A, as its entries.

Definition (2.1-30): If $f(\lambda)$ is a polynomial such that

$$f(\lambda) = f_0\lambda^m + f_1\lambda^{m-1} + \ldots + f_{m-1}\lambda + f_m$$

then

$$f(A) = f_0A^m + f_1A^{m-1} + \ldots + f_{m-1}A + f_mI$$

is the corresponding matrix polynomial.

Theorem (2.1-31): If $f(\lambda)$ is an analytic function defined on a simply connected domain $D \subseteq \mathbb{C}$ and if A is diagonable matrix such that

$$S^{-1}AS = \Lambda \text{ and } A = S\Lambda S^{-1}$$

then

$$f(A) = Sf(\Lambda)S^{-1}$$

where $f(\Lambda) = \sum_{i=1}^{n} f(\lambda_i)e_{ii}$.

Theorem (2.1-32): If $f(\lambda)$ and A are as defined in Theorem (2.1-31) then

$$f(A) = \sum_{i=1}^{n} f(\lambda_i)\, A_i$$

where $A_i = Se_{ii}S^{-1}$. The matrix $\underline{A}_i$ is called the ith rank one constituent idempotent matrix and $A_iA_j = \delta_{ij}A_i$ and $\sum_{i=1}^{n} A_i = I$. Furthermore $A_i = \underline{u}_i\, \underline{v}_i^T$ where u_i is the ith column of S (right eigenvector) and v_i^T is the ith row of S^{-1} (left eigenvector).

Alternatively, if A has s distinct eigenvalues λ_i with multiplicity m_i then

$$f(A) = \sum_{i=1}^{s} f(\lambda_i)\underline{A}_i$$

where

$$\underline{A}_i = \sum_{j}^{m_i} Se_{jj}S^{-1} = \sum_{j}^{m_i} A_j, \qquad j\varepsilon\{j \mid \lambda_i = \lambda_j\}.$$

The matrix $\underline{A}_i$ is called the ith rank m_i constituent idempotent matrix and $\underline{A}_i\underline{A}_j = \delta_{ij}\underline{A}_i$ and $\sum_{i=1}^{s} \underline{A}_i = I$.

Theorem (2.1-33): (Interpolation formula) The ith rank m_i consituent idempotent matrix A_i corresponding to a diagonable matrix A with s distinct eigenvalues is given by

$$A_i = \prod_{\substack{j=1 \\ j\neq i}}^{s} \frac{(A-\lambda_j I)}{(\lambda_i-\lambda_j)}\,.$$

Definition (2.1-34): The matrix $R_z = R(z) = (A-zI)^{-1}$ is called the resolvent matrix of A.

Theorem (2.1-35): (Residue Theorem) Let Γ be a closed curve in the complex plane enclosing eigenvalues λ_i of a matrix A, $i = 1,k$. Then

$$\sum_{i=1}^{k} \underline{A}_i = -\frac{1}{2\pi i}\int_{\Gamma} R(z)dz.$$

Theorem (2.1-36): (Dunford-Taylor Integral) Let f(z) be an analytic function defined on a simply connected domain $D \subseteq \mathscr{C}$ and let Γ be a closed curve in $\mathscr{C}$ enclosing all of the eigenvalues λ_i of a matrix A. Then

$$f(A) = -\frac{1}{2\pi i}\int_{\Gamma} f(z)R(z)dz \ .$$

Theorem (2.1-37): (First Resolvent Equation)

$$R(z_1) - R(z_2) = (z_1 - z_2)\, R(z_1)\, R(z_2)$$

if R is defined for z_1 and z_2.

Definition (2.1-38): The matrix $A_\lambda = A - \lambda I$ is called the characteristic matrix of A. The equation $\det(A_\lambda) = 0$ is called the characteristic equation of A. The notation $\det(\cdot)$ denotes the determinant of a matrix [10].

Theorem (2.1-39): (Cayley-Hamilton) Every matrix satisfies its characteristic equation.

Definition (2.1-40): A function $u(x)$: $\mathscr{C} \to \mathscr{C}^n$ is called a vector-valued function.

Definition (2.1-41): A function $A(x)$: $\mathscr{C} \to \mathscr{C}^{nxn}$ is called a matrix-valued function.

Definition (2.1-42): Suppose D is a simply connected domain in $\mathscr{C}$, $L(x)$ is an nxn matrix-valued function of $x \in D$ and $\{\lambda_i(x)\}_i^m$ is a set of eigenvalues of $L(x)$. Then if $\lambda_i(x) \to \lambda(x_1)$ as $x \to x_1$ where $\lambda(x_1)$ is an eigenvalue of $L(x_1)$ with multiplicity m and $x_1 \in D$, the set $\{\lambda_i(x)\}_i^m$ is called the <u>λ-group</u>.

Now let Γ be a closed curve in $\mathscr{C}$ enclosing a λ-group of eigenvalues. Then

$$P(x) = -\frac{1}{2\pi i}\int_{\Gamma} R(z_1x)dz \text{ where } R(z_1x) = (L(x) - zI)^{-1}$$

is called the <u>total projector</u>[1] for the λ-group [14].

---

[1]Note that $P(x_0)$ is identical to the rank m constituent idempotent matrix corresponding to $\lambda(x_0)$. Note also the connection to Theorem (2.1-35) which shows that the total projector is the sum of the individual consituent idempotent matrices of the λ-group. That P(x) is also idempotent follows from Theorems (2.1-32) and (2.1-35).

Theorem (2.1-43): (Product Rule)

i) If $A(x)$ and $B(x)$ are differentiable matrix-valued functions such that $C(x) = A(x)B(x)$ then

$$\frac{dC}{dx} = A \frac{dB}{dx} + \frac{dA}{dx} B.$$

ii) If $u(x)$ is a differentiable vector-valued function such that $v(x) = A(x)u(x)$ then

$$\frac{dv}{dx} = A \frac{du}{dx} + \frac{dA}{dx} u$$

where $\frac{dC}{dx}$ or $\frac{dv}{dx}$ indicates the differentiation of the individual entries c_{ij} or v_i with respect to x.

Theorem (2.1-44): If $A(x)$ is differentiable and n is a positive integer

$$\frac{dA^n}{dx} = \sum_{i=1}^{n} A^{i-1} \frac{dA}{dx} A^{n-i}$$

and if $A(x)$ is also non-singular

$$\frac{dA^{-n}}{dx} = - A^{-n} \frac{dA^n}{dx} A^{-n}.$$

Definition (2.1-45): An eigenvalue λ of a matrix A is called <u>simple</u> if it is not repeated. An eigenvalue λ of multiplicity m is called <u>semi-simple</u> if its eigenspace has dimension m. All the

eigenvalues of a diagonable matrix are semi-simple.

2.2 Concepts from Functional Analysis

Definition (2.2-1): Let $\{x_n\}$ be an infinite sequence of elements in a normed space X such that $\lim_{n\to\infty} ||x_n - x|| = 0$. Then $\{x_n\}$ is said to converge (strongly) to x.

Definition (2.2-2): A sequence $\{x_n\}$ in a normed space is called a Cauchy sequence if for any $\epsilon > 0$ there is an $N(\epsilon)$ such that

$$||x_m - x_n|| < \epsilon \text{ for every } m, n > N.$$

Definition (2.2-3): A normed space X is said to be complete if it contains the limit point of every Cauchy sequence in X. A complete normed linear space is called a Banach space.

Definition (2.2-4): An inner product space is a linear space X on which (x,y) is defined for each pair of elements x,y in X. A complete inner product space is called a Hilbert space. (Since every inner product generates a norm, any Hilbert space is also a Banach Space.)

Definition (2.2-5): A linear operator L is a mapping between linear spaces such that

i) The domain $\mathscr{D}(L)$ of L is a linear space and the range $\mathscr{R}(L)$ lies in a linear space over the same field F

ii) For all $x,y \in \mathscr{D}(L)$ and $\alpha \in F$

$$L(x + y) = Lx + Ly$$

$$L(\alpha x) = \alpha Lx.$$

Definition (2.2-6): The Null Space of L denoted $\mathscr{N}(L)$ is the set of all $x \in \mathscr{D}(L)$ such that $Lx = 0$.

The following standard mapping notation will be used. A function f which maps from a set A into a set B, is denoted by $f:A \to B$

1. A function $f:A \to B$ is surjective (onto) if and only if each $b \in B$ is an image of some element of A.
2. A function $f:A \to B$ is injective (one to one) if for each $b \in \mathscr{R}(f)$ there is exactly one $a \in A$ such that $b = f(a)$.
3. A function $f:A \to B$ is bijective (one to one and onto) if and only if it is subjective and injective (that is if and only if every $b \in B$ is the unique image of some $a \in A$.)

Definition (2.2-7): A function $f:A \to B$ is called a homeomorphism if it is bijective and if both f and $f^{-1}:B \to A$ are continuous.

Definition (2.2-8): Let X and Y be normed spaces and $L: \mathcal{D}(L) \to Y$ a linear operator, where $\mathcal{D}(L) \subseteq X$. The operator is said to be <u>bounded</u> if there is a positive real number c such that for all $x \in \mathcal{D}(L)$,

$$||Lx|| \leq c\ ||x||.$$

Theorem (2.2-9): If a normed space X is finite dimensional, then every linear operator on X is bounded.

Theorem (2.2-10): If L is a linear operator such that the mapping $L: \mathcal{D}(L) \to Y$ is injective then there exists the mapping $L^{-1}: \mathcal{R}(L) \to \mathcal{D}(L)$ which maps each $y \in \mathcal{R}(L)$ onto $x \in \mathcal{D}(L)$ for which $Lx = y$.

Theorem (2.2-11): Let X and Y be linear spaces and $L: \mathcal{D}(L) \to Y$. Then

a) $L^{-1}: \mathcal{R}(L) \to \mathcal{D}(L)$ exists if and only $\mathcal{N}(L)$ is empty.

b) L^{-1} is a linear operator.

Definition (2.2-12): Let X be a linear space, a function f defined for each $u \in X$ is a <u>linear functional</u> on X if

$$f[\alpha u + \beta v] = \alpha f[u] + \beta f[v]$$

for all u,v of X and $\alpha, \beta \in F$.

Definition (2.2-13): The linear space consisting of all the bounded linear functionals defined on X is called the dual space of X and is denoted by X'.

Theorem (2.2-14): Let X and Y be Hilbert spaces with inner products $(\cdot,\cdot)_X$ and $(\cdot,\cdot)_Y$ respectively and $L:X \to Y$ a linear operator. Then corresponding to L there exists a unique operator $L^*:Y \to X$ called the adjoint of L which satisfies

$$(L^*v,u)_X = (v,Lu)_Y$$

for every $u \in X$ and $v \in Y$. If $L^* = L$ then L is self adjoint.

Definition (2.2-15): A linear self-adjoint operator L is called positive definite $(u,Lu) > 0$ unless $u = 0$.

Theorem (2.2-16): Let λ be a non-zero scalar. Let X be a Hilbert space with inner product $(\cdot,\cdot)_X$ and $L:X \to X$ a bounded linear operator. Then if

$$Lx - \lambda x = y \tag{1}$$

$$Lx - \lambda x = 0 \tag{2}$$

$$L^*f - \lambda f = g \tag{3}$$

$$L^*f - \lambda f = 0 \tag{4}$$

i) (1) has a solution x for every $y \in X$ if and only if (2) has only the trivial solution $x = 0$.

ii) (3) has a solution f for every $g \in X$ if and only if (4) has only the trivial solution $f = 0$.

iii) (1) has a solution x (is normally solvable) if and only if $(f,y)_X = 0$ for all solutions of (4).

iv) (3) has a solution f (is normally solvable) if and only if $(g,x)_X = 0$ for all solutions x of (2).

Definition (2.2-17): Let X and Y be Banach spaces. Then a function $f:X \to Y$ is called Frechet differentiable at $x \in X$ if there exists a unique linear operator $f'(x):X \to Y$ such that

$$\lim_{||h||\to 0} \frac{||f(x + h) - f(x) - f'(x)h||}{||h||} = 0.$$

Definition (2.2-18): Let X be a Hilbert space with inner product $(\cdot,\cdot)_X$ and then if $f:X \to \mathscr{R}$ is a Frechet differentiable real valued function there exists a unique vector in X denoted by $\nabla f(x)$ called the gradient of f at x such that

$$f'(x)h = (h,\nabla f(x))_X.$$

Definition (2.2-19): Let X be a Hilbert space with inner product $(\cdot,\cdot)_X$ and $f:X \to \mathscr{R}$, then the directional derivative of f at x is defined by

$$Df(x,h) = \lim_{t \to 0^+} \frac{f(x + th) - f(x)}{t}$$

provided that this limit exists. If this limit exists for any direction h then f is called Gateaux differentiable at $x \in X$.

Remark: If f is Frechet differentiable then Df(x,h) exists for any direction h and

$$Df(x,h) = f'(x)h = (h, \nabla f(x))_X.$$

Theorem (2.2-20): (Projection Theorem [15]) Let X be a Hilbert space with inner product $(\cdot,\cdot)_X$, let $||\cdot||_X$ be the norm generated by $(\cdot,\cdot)_X$ and let M be a closed subspace of X. Then corresponding to any $x \in X$ there is a unique vector $\tilde{m} \in M$ such that

$$||v - m^*||_X \leq ||v - m||_X \text{ for all } m \in M. \qquad (2.2\text{-}21)$$

Furthermore $(x - m^*, m) = 0$ is a necessary and sufficient condition that $m^* \in M$ is the unique vector which satisfies (2.2-21).

Chapter 3

Eigensystem Differentiability

In this chapter various results related to the differentiability of eigenvalues and eigenvectors are presented.

3.1 Preliminary Differentiability Results

Suppose $L(x)$ is an nxn matrix-valued function (See 2.1-4) whose elements are analytic functions of $x \in D \subseteq \mathscr{C}$. It is natural to ask if and when $L(x)$ has eigenvalues and eigenvectors which are also analytic functions of x. This question, which has been studied extensively by Rellich, Kato and others [15-17], unfortunately does not have a succinct answer, since the problem inherits all of the complications associated with repeated eigenvalues and defective eigenspaces from the underlying eigenproblem. Nevertheless it is vital to recognize whether the eigenvalue problem behaves smoothly if eigenvalue and eigenvector derivatives are to be used in computations.

The following theorem gives an important sufficient condition for eigenvalue differentiability. Its proof may be found in cited literature.

Theorem (3.1-1): Let $L(x)$ be an nxn matrix-valued function of $x \in \mathscr{C}$, which is analytic in some neighborhood of $x = 0$. Then

a) If $\lambda(0)$ is a non-repeated eigenvalue of $L(0)$, $\lambda(x)$ is an analytic function in some neighborhood of $x = 0$ and there exist left and right eigenvectors $v(x)$ and $u(x)$ which are analytic vector-valued functions of x in some neighborhood of $x = 0$.

b) If $\lambda(0)$ has multiplicity m and $\{\lambda_i(x)\}_i^m$ is the λ-group, $\lambda_i(x)$ may be expanded in a *Puiseaux* series [18] as

$$\lambda_i(x) = \lambda(0) + \mu_1 x^{\frac{1}{p}} + \mu_2 x^{\frac{2}{p}} + \ldots \text{ , } p \text{ is an integer} \leq m$$

Proof: [14, 18 and 19].

It is easy to show that the conditions of Theorem (3.1-1) (a) are not necessary by constructing a matrix which has repeated analytic eigenvalues and analytic eigenvectors.

Example (3.1-2): Let

$$L(x) = \sum_{i=1}^{n} \lambda_i(x)\underline{u}_i(x)\underline{v}_i^T(x) \qquad (3.1\text{-}3)$$

where the $\lambda_i(x)$ are analytic scalar functions not necessarily distinct for $x \in D \subseteq \mathscr{C}$ and the $\underline{u}_i(x)$ are analytic vector-valued functions of $x \in D$ such that the matrix $S(x)$, whose ith column is $\underline{u}_i(x)$, is invertible. Then $S^{-1}(x)$ is also

analytic [14], and we may take its ith row as $\underline{v}_i(x)$. Theorem (2.1-32) shows that the $\lambda_i(x)$ are analytic eigenvalues of L(x) and the $v_i(x)$ and $u_i(x)$ are analytic left and right eigenvectors of L(x), even if $\lambda_i(x_1) = \lambda_j(x_1)$, $i \neq j$ for some isolated value x_1 (exceptional point). So L(x) has the desired properties.

The matrix L(x) in example (3.1-2), although it has repeated eigenvalues, is diagonable due to the manner in which it was constructed [10]. That an analytic matrix is diagonable however, may not be taken as a necessary or sufficient condition for eigenvalue analyticity as the next example shows.

Example (3.1-4): The matrix

$$L(x) = \begin{bmatrix} 0 & x & 0 \\ 0 & 0 & x \\ x & 0 & 1 \end{bmatrix}$$

is diagonable for $x^3 \neq 4/27$ yet its eigenvalues cannot be written as a Taylor series in x at x = 0 [14] whereas the matrix

$$L(x) = \begin{bmatrix} x & 1 \\ 0 & x \end{bmatrix}$$

is not diagonable for any $x \in \mathbb{C}$ but its eigenvalue x is clearly analytic.

3.2 The Reduction Process

To gain further insight into the behavior of eigenvalue differentiability the reduction process of Kato [14] is useful. Suppose L(x) is an analytic matrix-valued function with the expansion

$$L(x) = \sum_{i=0}^{\infty} x^i L^{(i)}(0). \tag{3.2-1}$$

If $\lambda(x)$ is an analytic eigenvalue of L(x) with multiplicity m then a particular branch of $\lambda(x)$ may be expanded as

$$\lambda(x) = \lambda(0) + x\lambda^{(1)}(0) + x^2\lambda^{(2)}(0) + \ldots + \ldots \quad . \tag{3.2-2}$$

Kato has demonstrated that under some circumstances the sequence of coefficients $\lambda^{(i)}(0)$ may be identified with certain eigenvalues of a sequence of matrices, $\tilde{L}^{(i)}(x)$, which are described below. The following theorem is central to the process.

Theorem (3.2-3): Let $L(x)$ be an analytic matrix with expansion

$$L(x) = \sum_{i=0}^{\infty} x^i L^{(i)}(0) \tag{3.2-4}$$

and let $\lambda(0) = \lambda$ be a semi-simple eigenvalue of $L(0) = L$ of multiplicity m with total projector $P(0) = P$. If $\lambda_j^{(1)}$ is an eigenvalue of $\tilde{L}^{(1)} = PL^{(1)}P$ in the subspace $\mathscr{R}(P)$ with corresponding constituent idempotent matrix $P_j^{(i)}$ then $L(x)$ has exactly $m_j^{(1)} = \dim P_j^{(1)}$ repeated eigenvalues of the form $\lambda + x\lambda_j^{(1)} + O(x)$.

Proof: From the first resolvent equation Theorem (2.1-37) we find that

$$(L(x) - \lambda I)\, R(z,x) = I + (z - \lambda)\, R(z,x)\ . \tag{3.2-5}$$

Suppose that Γ is a closed curve in $\mathscr{C}$ which encloses the λ-group eigenvalues of $L(x)$ then

$$(L(x) - \lambda I)\, P(x) = -\frac{1}{2\pi i} \int_{\Gamma} (z - \lambda) R(z,x)\, dz \tag{3.2-6}$$

follows from applying theorem (2.1-35) to (3.2-5) and recalling Definition (2.1-42). Since $R(z,x)$ is analytic in x [14] (3.2-6) can be written as

$$(L(x) - \lambda I)\, P(x) = \sum_{i=1}^{\infty} x^i \tilde{L}^{(i)}(0) \tag{3.2-7}$$

where

$$\tilde{L}^{(i)}(0) = -\frac{1}{2\pi i} \int_{\Gamma} (z - \lambda)\, R^{(i)}(z,0)\, dz\ . \tag{3.2-8}$$

Since λ is a semi-simple eigenvalue both sides of (3.2-7) must vanish at $x = 0$ so $\tilde{L}^{(0)}(0) = 0$ and

$$\lim_{x\to 0} \frac{(L(x) - \lambda I)P(x)}{x} = \tilde{L}^{(1)}(0).$$

Now if $u_j(x)$ is an eigenvector of $\lambda_j(x)$ where $\lambda_j(x)$ is a member of the λ-group then

$$\lim_{x\to 0} \frac{(L(x) - \lambda I)}{x} P(x)u_j(x) = \lim_{x\to 0} \frac{(\lambda_j(x) - \lambda)}{x} u_j(x)$$

$$= \lambda_j^{(1)}(0)u_j(0)$$

which shows that $\lambda_j^{(1)}(0)$ is an eigenvalue of $\tilde{L}^{(1)}(0)$. To complete the proof we must evaluate (3.2-8) for $i = 1$.

From Theorem (2.1-44)

$$R^{(1)}(z,0) = -R(z,0)\, L^{(1)}(0)\, R(z,0)$$

we may expand $R(z,0)$ as

$$R(z,0) = (z - \lambda)^{-1}P(0) + S(z)$$

where $S(z)$ has no singularities so

$$R^{(1)}(z,0) = [(z - \lambda)^{-1}P(0) + S(z)]\, L^{(1)}(0)\, [(z - \lambda)^{-1}P(0) + S(z)].$$

Integrating $(z - \lambda)R^{(1)}(z,0)$ along Γ shows that

$$\tilde{L}^{(1)} = P(0)\, L^{(1)}(0)\, P(0).\square$$

This theorem can be re-applied to $\tilde{L}^{(1)}(x)$ if its eigenvalues are also semi-simple to derive a higher order expansion for $\lambda(x)$. In this case we have

$$\lambda(x) = \lambda + x\lambda_j^{(1)} + x^2\lambda_{jk}^{(2)} + O(x^2), \quad k = 1,m_j^{(1)} \tag{3.2-5}$$

where the $\lambda_{jk}^{(2)}$ are the repeated eigenvalues of

$$\tilde{L}^{(2)} = P_j^{(1)}L^{(2)}P_j^{(1)} - P_j^{(1)}L^{(1)}L_\lambda^{\dagger}L^{(1)}P_j^{(1)} \tag{3.2-6}$$

in the subspace $\mathcal{R}(P_j^{(1)})$. The matrix $L_\lambda^{\dagger}$ is defined by

$$L_\lambda^{\dagger} = -\sum_{i=1}^{n} (\lambda - \lambda_i I)^{-1}L_i \text{ if } L \text{ is diagonable.} \tag{3.2-7}$$

The reduction process then consists of repeated applications of Theorem (3.1-8) to the matrices $\tilde{L}^{(i)}$.

The reduction process may also be used to generate expansions for the constituent idempotent matrices, so if it can be shown that each $\tilde{L}^{(i)}$ has semi-simple eigenvalues then the process may be continued indefinitely to show that the eigenvalues of L(x) are analytic and to establish the existance of analytic eigenvectors (See Theorem (4.1-4)).

In the next section, the question of eigenvector differentiability is discussed through examples involving the interpolation formula Theorem (2.1-33) for constituent idempotent matrices.

3.3 Eigenvector Differentiability

Eigenvector differentiability is more complicated than eigenvalue differentiability. This is due, in part, to the possibility that the dimension of the eigenspace associated with a particular eigenvalue may change abruptly or disappear entirely. In this section we shall use the interpolation formula of Theorem (2.1-33) to study the behavior of eigenvectors corresponding to analytic semi-simple eigenvalues in two examples. In the first example the matrix L(x) has no exceptional point (See Example (3.2-2)) in its domain D.

Example (3.3-1): Suppose the eigenvalues of L(x) can be represented by the s analytic functions $\lambda_1(x)$, $\lambda_2(x),\ldots,\lambda_s(x)$ where $\lambda_i(x)$ is assumed to have multiplicity m_i and where L(x) is analytic and diagonable for $x \in D$.

Since L(x) is diagonable, we may express the rank m_i constituent idempotent matrices of L by the interpolation formula of Theorem 2.1-33.

For the ith constituent idempotent matrix, L_i, we have

$$\underline{L}_i(x) = \prod_{\substack{j=1 \\ j\neq i}}^{s} \frac{L(x)-\lambda_j(x)I}{(\lambda_i(x)-\lambda_j(x))}, \qquad x \in D \tag{3.3-2}$$

If there is no exceptional point in D, the $\underline{L}_i(x)$ are analytic on D. Any column of $\underline{L}_i(x)$ which is non-zero on D represents an analytic eigenvector $\underline{u}_i(x)$ corresponding to $\lambda_i(x)$ (See Theorem (2.132)). We now assume, without loss of generality, that x = 0 is in D. Then if $\underline{u}_i(0)$ is an eigenvector of L(0) corresponding to $\lambda_i(0)$, $L_i(x)\underline{u}_i(0)$ is an analytic eigenvector corresponding to $\lambda_i(x)$ in some neighborhood of x = 0. This is true since (1)

$$\underline{L}_i(0)\underline{u}_i(0) = \prod_{\substack{j=1 \\ j\neq i}}^{s} \frac{(\lambda_i(0)-\lambda_j(0))}{(\lambda_i(0)-\lambda_j(0))}\,\underline{u}_i(0) = \underline{u}_i(0) \neq 0$$

and the continuity of $\underline{L}_i(x)$ imply that $\underline{u}_i(x)$ is non-zero in a neighborhood of x = 0 and since (2) by the Cayley-Hamilton theorem, $L_i(x)\underline{u}_i(0)$ is in the null space of $L(x) - \lambda_i(x)I$.

Now suppose that $x = 0$ is an exceptional point in D such that $\lambda_k(0) = \lambda_i(0)$. Then $\underline{L}_i(x)$ is not continuous at $x = 0$. It may have a removable discontinuity at $x = 0$ however.[1] In the following example we study the behavior of $\underline{L}_i(x)$ in a deleted neighborhood of an exceptional point.

Example (3.3-3): Suppose $L(x)$ is an nxn matrix which is diagonable for $x \in D$ with n eigenvalues $\lambda_i(x)$ which are distinct analytic functions on D. If $\lambda_k(0) = \lambda_i(0)$ for $x = 0 \in D$ then $x = 0$ is an exceptional point in D since by hypothesis $\lambda_k(x)$ and $\lambda_i(x)$ are distinct analytic functions. We will now investigate the behavior of

$$L_i(x)\underline{u}(0) = \prod_{\substack{j=1 \\ j\neq i}}^{s} \frac{(L(x)-\lambda_j(x)I)}{(\lambda_i(x)-\lambda_j(x))}\ \underline{u}(0) \qquad (3.3\text{-}4)$$

in a deleted neighborhood of $x = 0$, where $\underline{u}(0)$ is any eigenvector taken from the two dimensional eigenspace corresponding to $\lambda_i(0) = \lambda_k(0)$.

Both the numerator and denominator of $L(x)u(0)$ become zero at $x = 0$ so

$$\lim_{x\to 0} L_i(x)\underline{u}(0) = \lim_{x\to 0} \frac{\frac{dL(x)}{dx} - \frac{d\lambda_k(x)}{dx}I}{\frac{d\lambda_i(x)}{dx} - \frac{d\lambda_k(x)}{dx}}\underline{u}(0) \qquad (3.3\text{-}5)$$

---

[1] For example if $L(x)$ is a normal matrix on $x \in D$ then $L_i^*(x) = L_i(x)$ [10] with $||L_i(x)|| = 1$ implying that $L_i(x)$ has no singularities on D [14].

after applying L'Hopital's rule. If by hypothesis, $\lambda_i(x)$, $\lambda_k(x)$ have different slopes at $x = 0$, the denominator of (3.3-4) is non-zero at 0 and the limit (3.3-5) exists.[1] However, (3.3-5) is an eigenvector only if its numerator is also non-zero.

To study the behavior of the numerator of (3.3-5) let $P = P(0)$ be the total projector corresponding to $\lambda_i(0) = \lambda_k(0)$. Then (3.3-15) may be written

$$\lim_{x\to 0} L_i(x)\underline{u}(0) = \lim_{x\to 0} \frac{\frac{dL(x)}{dx}P - \frac{d\lambda_k}{dx}(x)P}{\frac{d\lambda_i}{dx}(x) - \frac{d\lambda_k}{dx}(x)}\underline{u}(0) \qquad (3.3\text{-}6)$$

since $\underline{u}(0) = P\underline{u}(0)$.

Now suppose that $\underline{u}(0)$ is an eigenvector of

$$\tilde{L} = P\frac{dL(0)}{dx}P$$

corresponding to its eigenvalue $\frac{d\lambda_i}{dx}(0)$ (See Theorem (2.1-21)).[2] Then

$$P\frac{dL(0)}{dx}Pu(0) = \frac{d\lambda_i}{dx}(0)u(0) = \frac{d\lambda_i}{dx}(0)Pu(0) \qquad (3.3\text{-}7)$$

So the numerator of (3.3-6) is non-zero since otherwise

---

[1] Note also that the matrix $L_i(x)$ has only rank one in the limit as $x\to 0$ since $\lambda_i(x)$ is non-repeated in the deleted neighborhood of $x = 0$.

[2] An eigenvector of $\tilde{L}$ may be selected from the eigenspace of $\lambda_i(0)$ since the range of P is precisely that subspace.

$$\frac{dL(0)}{dx} P\underline{u}(0) = \frac{d\lambda_k(0)}{dx} P\underline{u}(0) \qquad (3.3\text{-}8)$$

and

$$P \frac{dL(0)}{dx} P\underline{u}(0) = \frac{d\lambda_k(0)}{dx} P\underline{u}(0) \qquad (3.3\text{-}9)$$

would contradict (3.3-7).

If $\underline{u}(0)$ is an eigenvector of $\tilde{L}$ corresponding to $\frac{d\lambda_k(0)}{dx}$ a similar analysis shows that the numerator of (3.1-16) does vanish. So if $\underline{u}(0)$ contains a component which is an eigenvector corresponding to $\frac{d\lambda_i(0)}{dx}$, (3.3-5) is an analytic eigenvector passing through the exceptional point $x = 0$ which corresponds to $\lambda_i(x)$. An analytic eigenvector corresponding to $\lambda_k(x)$ can be constructed similarly.

As the final topic of this chapter, we consider the question of differentiability of eigensystems which are functions of several parameters. As we shall see, the Frechet differentiability of such eigensystems may not be assumed for repeated eigenvalues even if they are analytic in a single variable.

3.4 Differentiation with Respect to Several Parameters

The preceding sections concerned differentiability of eigensystems with respect to a single variable. The situation is more complicated when several variables are involved.

Consider the following example [14]:

Let

$$L(x_1,x_2) = \begin{bmatrix} x_1 & x_2 \\ x_2 & -x_1 \end{bmatrix}, \quad (x_1,x_2) \in \mathscr{C}^2. \tag{3.4-1}$$

Even though $L(x_1,x_2)$ is symmetric and Frechet differentiable, its eigenvalues $\lambda_{1,2} = \pm (x_1^2 + x_2^2)^{1/2}$ are only Gateaux differentiable when $x_1 = x_2 = 0$ [20].

Using Definition (2.2-19) the directional derivative of λ_1 when $x_1 = x_2 = 0$ is determined as

$$D(\lambda_{1,2},h) = \lim_{t\to 0} \frac{((x_1 + th_1)^2 + (x_2 + th_2)^2)^{1/2} \mp (x_1^2 + x_2^2)^{1/2}}{t}$$

$$= \pm (h_1^2 + h_2^2)^{1/2}, \qquad \text{where } h = (h_1,h_2). \tag{3.4-2}$$

Since $D(\lambda_{1,2},h)$ exists for all h, $\lambda_{1,2}$ are Gateaux differentiable, but they are not Frechet differentiable since $D(\lambda_{1,2},h)$ is not linear in h.

This complication has serious consequences if these derivatives are to be used in non-linear programming routines which use gradients to determine a descent direction. Since the definition of the gradient of a function depends crucially on the linearity of the Frechet differential. Recalling Definition (2.2-20), the Frechet differential of f may be represented in terms of the gradient of f as

$$D(f,h) = (\nabla f,h). \tag{3.4-3}$$

where $(\cdot,\cdot)$ is the inner product. It is this equivalence which provides the justification for using the negative gradient as the direction of steepest descent since if $||\nabla f|| = 1$ and $||h|| = 1$ $D(f,h) = (\nabla f,h)$ attains its minimum value for $h^* = -\nabla f$. When the Gateaux derivative is non-linear in the direction h, the descent direction is not so easily determined [21, 27].

We may also use Theorem (3.2-3) to compute the directional derivative for λ_1 or λ_2. Since $L(x_1,x_2)$ is Frechet differentiable we may represent $D(L,h)$ in terms of its gradient ∇L as

$$D(L,h) = (\nabla L,h) = \frac{\partial L}{\partial x_1} h_1 + \frac{\partial L}{\partial x_2} h_2 = \frac{dL}{dt}$$

where $$L(t) = \begin{bmatrix} x_1 + th_1 & x_2 + th_2 \\ x_2 + th_2 & -x_1 - th_1 \end{bmatrix} .$$

Then

$$D(L,h) = \begin{bmatrix} h_1 & h_2 \\ h_2 & -h_1 \end{bmatrix}$$

to which Theorem (3.2-3) may be applied since L(t) is an analytic symmetric function of t. The directional derivatives of the eigenvalues $\lambda_{1,2}$ at $x_1 = x_2 = 0$ are the eigenvalues of

$$\tilde{L} = \underline{P}\, D\,(L,h)\, P = P\, \frac{dL}{dt}\, P = D(L,h)$$

so

$$D(\lambda_{12},h) = \pm\, (h_1^2 + h_2^2)^{1/2} \quad \text{as before.}$$

Remark (3.4-4): If D(L,h) is a symmetric matrix and if B is any matrix representing an orthonormal basis for $\mathcal{R}(P)$ then the directional derivatives are also the eigenvalues of

$$B^T\tilde{L}B = B^TPD(L,h)PB = B^TD(L,h)B \qquad (3.4\text{-}5)$$

which is a formulation given by Haug and Rousselet [20] which may be easier to apply than Theorem (3.2-3).

The non-linear dependance of the directional derivatives of eigenvalues can perhaps best understood by writing $\tilde{L}$ as

$$\tilde{L} = PD(L,h)P = P \sum \frac{\partial L}{\partial x_i} h_i P = \sum P \frac{\partial L}{\partial x_i} P h_i$$

where although the matrix $\tilde{L}$ is linear in h_i its eigenvalues cannot expected to be, unless $\mathscr{R}(P)$ has dimension one (if the eigenvalue is simple).

The rank one constituent idempotent matrices and eigenvectors or $L(x_1,x_2)$ are even 'less' regular than the eigenvalues. To show this we construct an expansion of $L(x_1,x_2)$ in terms of its constituent idempotent matrices $L_i(x_1,x_2)$ (See Theorem (2.1-33).)

$$L(x_1,x_2) = \lambda_1(x_1,x_2)\, L_1(x_1,x_2) + \lambda_2(x_1,x_2)\, L_2(x_1,x_2) \qquad (3.4\text{-}6)$$

where

$$L_1 = \frac{L - \lambda_2}{(\lambda_1 - \lambda_2)} , \qquad L_2 = \frac{L - \lambda_1}{(\lambda_2 - \lambda_1)} . \qquad (3.4\text{-}7)$$

or

$$L_1(x_1,x_2) = \frac{1}{2(x_1^2 + x_2^2)^{1/2}} \begin{bmatrix} x_1 + (x_1^2+x_2^2)^{1/2} & x_2 \\ x_2 & -x_1 + (x_1^2+x_2^2)^{1/2} \end{bmatrix}$$

(3.4-8)

$$L_2(x_1,x_2) = -\frac{1}{2(x_1^2 + x_2^2)^{1/2}} \begin{bmatrix} x_1 - (x_1^2+x_2^2)^{1/2} & x_2 \\ x_2 & -x_1 - (x_1^2+x_2^2)^{1/2} \end{bmatrix}$$

(3.4-9)

However, the idempotents $L_i(x_1,x_2)$ possess no limit as $(x_1,x_2)\to(0,0)$ since

$$\lim_{x_2\to 0} L_1(0,x_2) = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \tag{3.4-10}$$

whereas

$$\lim_{x_1\to 0} L_1(x_1,0) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \tag{3.4-11}$$

and similarly

$$\lim_{x_2\to 0} L_2(0,x_2) = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} . \tag{3.4-12}$$

whereas

$$\lim_{x_1\to 0} L_2(x_1,0) = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} . \tag{3.4-13}$$

This illustrates that although in this example the rank one constituent idempotent matrices and therefore the eigenvectors can be continued smoothly through the origin along the x_1 or x_2 axis, their total limits do not exist at the origin. Consequently, the rank one constituent idempotent matrices and eigenvectors at not even continuous at $x_1 = x_2 = 0$. They do have directional derivatives since L(t) is symmetric and analytic in t (see Theorem (4.1-4)).

3.5 Conclusions

The theorems of this chapter provide sufficient conditions needed to justify the differentiation of eigenvalues and eigenvectors in a number of situations important to applications. In the next chapter formulations for the actual eigensystem derivatives are developed.

Chapter 4

Differentiation of Eigensystems

As the remarks of the previous chapter illustrate, the question of eigenvalue and eigenvector differentiability in its full generality is difficult. Some matrices however, have properties which lead to relatively simple eigensystem derivatives.

4.1 Differentiation of Fully Analytic Matrices

Definition (4.1-1): Let the diagonable matrix-valued function $L(x)$ have the decomposition

$$L(x) = \sum_{i=1}^{n} \lambda_i(x)\underline{u}_i(x)\underline{v}_i^T(x) = \sum_{i=1}^{n} \lambda_i(x)L_i(x) \qquad x \in D \subseteq \mathscr{C} \qquad (4.1\text{-}2)$$

where the $\underline{u}_i(x)$ and $\underline{v}_i(x)$ belong to an analytic biorthogonal set of eigenvectors such that $\underline{v}_i^T(x)u_j(x)=\delta_{ij}$, $\lambda_i(x)$ are corresponding analytic eigenvalues and the $L_i(x)$ are rank one constituent idempotent matrices. Then we say, in this thesis, that $L(x)$ is fully analytic on D.

Remark (4.1-3): Theorem (3.1-1) gives sufficient conditions for full analyticity of $L(x)$ on D if $L(x)$ has no repeated eigenvalues for any $x \in D$.

The following theorems form a basis for a study of eigensystem derivatives of fully analytic matrices and methods to compute them. The first of these (Theorem (4.1-4)) states a sufficient condition for full analyticity which is important in applications. The theorem was originally stated by Rellich [16] but the proof here follows Kato [14]. (See also [14] for extensions to normal matrices.)

Theorem (4.1-4) [14, 16]: If $L(x)$ is a Hermitian analytic matrix-valued function for $x \in D \subseteq \mathscr{R}$ then $L(x)$ is fully analytic on D.

Proof: If $L(x)$ is Hermitian on D then the $\tilde{L}^{(i)}$ in Theorem (3.2-2) are Hermitian and therefore diagonable. The reduction process can then be continued indefinitely.□

The next two theorems (Theorems (4.1-5 and 6)) present various algebraic properties of eigensystems derivatives. Part (e) is essentially the result given by Jacobi [4] although the derivation is different.

Theorem (4.1-5): Let $L(x)$ be an nxn fully analytic matrix-valued function for $x \in D$. Then

a) $$\frac{dL}{dx} = \sum_{i=1}^{n} \lambda_i \frac{dL_i}{dx} + \sum_{i=1}^{n} \frac{d\lambda_i}{dx} L_i$$

b) $\frac{dL}{dx}i = \frac{dL}{dx}i\ L_i + L_i \frac{dL}{dx}i = \underline{u}_i \frac{dv}{dx}^T i + \frac{du}{dx}i\ \underline{v}_i^T$

c) $\frac{dv}{dx}^T i\ \underline{u}_j = -\ \underline{v}_i^T \frac{du}{dx}\ j$

d) $\underline{v}_i^T \frac{dL}{dx}\ \underline{u}_j = (\lambda_j - \lambda_i)\underline{v}_i^T \frac{du}{dx}j = (\lambda_i - \lambda_j)\ \frac{dv}{dx}^T i\ \underline{u}_j, \qquad i \neq j$

e) $v_i^T \frac{dL}{dx}\ u_i = \frac{d\lambda}{dx}i$

where dependence on x is understood.

Proof: To prove part (a) apply the product rule of differentiation to equation (4.1-2). Part (b) follows after differentiating $L_i^2 = L_i$. For part (c), differentiate $v_i^T(x)u_j(x) = \delta_{ij}$. Part (d) and (e) are proved by forming the product

$$\underline{v}_i^T \frac{dL}{dx}\ \underline{u}_j = \sum \frac{d\lambda}{dx}k\ \underline{v}_i^T L_k \underline{u}_j + \sum \lambda_k\ \underline{v}_i^T \frac{dL}{dx}k\ \underline{u}_j$$

and applying (b) and (c).□

Theorem (4.1-6): Let L(x) be an nxn analytic real matrix-valued function for $x \in D \subseteq \mathscr{R}$ such that $L^T(x) = L(x)$. Then

a) $$\frac{dL}{dx} = \sum_{i=1}^{n} \lambda_i \frac{dL}{dx}i + \sum_{i=1}^{n} \frac{d\lambda}{dx}i\ L_i$$

b) $\frac{dL}{dx}j(x) = \frac{dL}{dx}jL_j + L_j \frac{dL}{dx}j = \underline{u}_j \frac{du}{dx}^T j + \frac{du}{dx}j\ \underline{u}_j^T$

c) $\frac{du_i}{dx}^T \underline{u}_j = - \underline{u}_i^T \frac{du_j}{dx}$

d) $\underline{u}_i^T \frac{dL}{dx} \underline{u}_j = (\lambda_j - \lambda_i) \underline{u}_i^T \frac{du_j}{dx} = (\lambda_i - \lambda_j) \frac{du_i}{dx}^T \underline{u}_j$

e) $\underline{u}_i^T \frac{dL}{dx} \underline{u}_i = \frac{d\lambda_i}{dx}$

Proof: Similar to (4.1-5) after noting that Theorem (4.1-4) guarantees that L(x) is fully analytic.□

Remark (4.1-7): In part (c) if i = j we have $\underline{u}_i^T \frac{du_i}{dx} = 0$, showing that $\underline{u}_i$ is orthogonal to $\frac{du_i}{dx}$. This results from the constant norm condition imposed by $\underline{u}_i^T(x)\underline{u}_j(x) = \delta_{ij}$. Note also that $L(x): \mathcal{R} \rightarrow \mathcal{R}^{nxn}$.

Further insight into the behavior of eigensystem derivatives may be obtained by studying the effect of changing an individual element of a matrix. In the following theorem we will assume that L(0) is a constant matrix such that $L(x) = L(0) + e_{ij}\, x$ is fully analytic for $x \in D$.

Theorem (4.1-8): Let $L(x) = L(0) + e_{ij}x$ be a fully analytic matrix-valued function for $x \in D$. Let $\lambda_i(x)$ be an analytic eigenvalue of L(x), let $\underline{v}_i(x)$ and $\underline{u}_i(x)$ belong to an analytic

set of left and right eigenvectors such that $\underline{v}_i^T(x)\underline{u}_j(x) = \delta_{ij}$ and let $L_i(x) = \underline{u}_i(x)\ \underline{v}_i^T(x)$ be the corresponding constituent idempotent matrix. Then

a) $\underline{v}_k^T \frac{dL}{dx} \underline{u}_k = \frac{d\lambda}{dx}k = \underline{v}_{ki}\ \underline{u}_{kj} = L_{kji}$

b) $\underline{v}_s^T \frac{dL}{dx} \underline{u}_t = (\lambda_t - \lambda_s)\ \underline{v}_s^T \frac{du}{dx}t = \underline{v}_{si}\ \underline{u}_{tj}$

c) $\sum_{k=1}^{n} \frac{d\lambda}{dx}k = \delta_{ij}$

d) $\sum_{k=1}^{n} \frac{d^2\lambda}{dx^2}k = 0$

Proof: To prove part a) apply Theorem (3.2-4) e) with $\frac{dL(x)}{dx} = e_{ij}$ and note that $\underline{v}_{ki}\ \underline{u}_{kj}$ is the jith entry in L_k. For part b) apply Theorem (3.2-4) d). Part c) follows from part a) and Theorem (2.1-32). Since the sum of eigenvalue derivatives is a constant function of x, the sum of second derivatives with respect to x is zero proving part d).□

Remark (4.1-9): Since the non-zero columns of the idempotent matrices L_i are eigenvectors corresponding to λ_i, part (a) above establishes a strong relationship between an eigenvalue's derivatives and its eigenvectors.

The following Theorem (4.1-10) states a formulation, given by Fox and Kapoor [5] (See also [12]), which expresses the derivative of an eigenvector of a symmetric matrix as a linear combination of eigenvectors.

Theorem (4.1-10) [5]: Let $L(x)$ be an analytic real matrix-valued function for $x \in D \subseteq \mathcal{R}$ such that $L^T(x) = L(x)$. Then

$$\frac{du}{dx}_i(x) = \sum_{j=1}^{n} c_j(x)\underline{u}_j(x) \qquad (4.1\text{-}11)$$

with

$$c_j(x) = \begin{cases} 0 & i = j \\ (\lambda_i(x) - \lambda_j(x))^{-1}\, u_j^T(x)\, \frac{dL}{dx}\, \underline{u}_i(x) & \\ = \underline{u}_j^T(x)\, \frac{du}{dx}_i(x) & i \neq j \end{cases}$$

Proof: Since $\frac{du}{dx}_i$ is a vector in an n dimensional space, we can express it as a linear combination of n orthonormal eigenvectors of $L(x)$

$$\frac{du}{dx}_i(x) = \sum_{k=1}^{n} c_k(x)\underline{u}_k(x)$$

using $\underline{u}_i^T(x)\underline{u}_j(x) = \delta_{ij}$ and $u_j^T \frac{du}{dx}_j = 0$ (See Remark (4.1-6), we have

$$\underline{u}_j^T \frac{du}{dx}i = c_j \underline{u}_j^T \underline{u}_j = c_j \text{ and } \underline{u}_i^T \frac{du}{dx}i = c_i = 0;$$

applying Theorem (4.1-5) (d) completes the proof.□

Although this formulation has theoretical value it is not useful when a complete set of eigenvectors is not available [9]. However, we may rewrite (4.1-11) as

$$\frac{du}{dx}i = -\sum_{j=1}^{n} (\lambda_j - \lambda_i)^{-1} \underline{u}_j \underline{u}_j^T \frac{dL}{dx} \underline{u}_i$$

$$= -\sum_{j=1}^{n} [(\lambda_j - \lambda_i)^{-1} L_j] \frac{dL}{dx} \underline{u}_i = -L_\lambda^\dagger \frac{dL}{dx} \underline{u}_i \qquad (4.1\text{-}12)$$

The matrix

$$L_\lambda^\dagger = -\sum_{j=1}^{n} [(\lambda_j - \lambda_i)^{-1} L_j]$$

as we shall see, in Section 4.3, can be easily computed and will lead to an efficient means to compute $\frac{du}{dx}i$.

4.2 Eigensystem Derivatives with Generalized Inverse

If L(x) is a diagonable fully analytic nxn matrix-valued function in some neighborhood of x = 0, then we may write

$$[L(x) - \lambda(x)I]u(x) = \left(\sum_{i=0}^{\infty} x^i (L^{(i)} - \lambda^{(i)} I)\right)\left(\sum_{i=0}^{\infty} x^i u^{(i)}\right) = 0 \tag{4.2-1}$$

by expanding $L(x)$, $\lambda(x)$ and $u(x)$ in a Taylor series about $x = 0$. Collecting terms in x^i

$$\sum_{i=0}^{\infty} x^i \sum_{j=0}^{i} (L^{(i-j)} - \lambda^{(i-j)} I) u^{(j)} = 0 . \tag{4.2-2}$$

Each coefficient of x^i must vanish separately so

$$\sum_{j=0}^{i} (L^{(i-j)} - \lambda^{(i-j)} I) u^{(j)} = 0 \qquad i = 0,1,2,... \tag{4.2-3}$$

or

$$\sum_{j=0}^{i} L^{(i-j)} u^{(j)} = \sum_{j=0}^{i} \lambda^{(i-j)} u^{(j)} . \qquad i = 0,1,2,... \tag{4.2-4}$$

After subtracting the $i = j$ term from (4.2-3) the eigensystem derivatives are given recursively by

$$(L - \lambda I)u = 0, \qquad i = 0 \tag{4.2-5}$$

$$(L - \lambda I)u^{(i)} = - \sum_{j=0}^{i-1} (L^{(i-j)} - \lambda^{(i-j)} I) u^{(j)}, \quad i = 1,2,3,... \tag{4.2-6}$$

Since we have assumed that L(x) is fully analytic, these equations must be normally solvable (See Theorem (2.2-16)) for each i. Hence if v is any vector such that $v \in \mathcal{N}(L^T - \lambda I)$, then

$$v^T \sum_{j=0}^{i=1} (L^{(i-j)} - \lambda^{(i-j)}I)u^{(j)} = 0 \qquad (4.2\text{-}7)$$

follows from Theorem (2.2-16). Solving for $\lambda^{(i)}$ results in

$$\lambda^{(1)} = \frac{v^T L^{(1)} u}{v^T u} \qquad (4.2\text{-}8)$$

and

$$\lambda^{(i)} = \frac{v^T\left(\sum_{j=0}^{i-1} L^{(i-j)}u^{(j)} - \sum_{j=1}^{i-1} \lambda^{(i-j)}u^{(j)}\right)}{v^T u}, \quad i = 2,3,4,... \qquad (4.2\text{-}9)$$

where v is any vector such that $v^T u \neq 0$ and $v \in \mathcal{N}(L^T - \lambda I)$ and $u = u(0)$ lies on u(x), an analytic eigenvector.

We now turn our attention to solving equations (4.2-6) for $u^{(i)}$. Since the matrix $L_\lambda = L - \lambda I$ has rank n - m we use a generalized inverse L_λ^I and apply Theorem (2.1-20) to find that

$$u^{(i)} = - L_\lambda^I \sum_{j=0}^{i-1} (L^{(i-j)} - \lambda^{(i-j)}I)u^{(j)} + z, \qquad (4.2\text{-}10)$$

where $z \in \mathcal{N}(L_\lambda)$ and L_λ^I is any rank n - m matrix satisfying $L_\lambda L_\lambda^I L_\lambda = L_\lambda$.

We shall have immediate use for the following:

Lemma (4.2-11): Let A, A_R and A_L be nxn matrices satisfying $AA_R = 0$ and $A_L A = 0$. Then if A^I is any generalized inverse of A,

$$A^\# = (I - A_R)A^I (I - A_L) \tag{4.2-12}$$

is also a generalized inverse of A.

Proof. Multiplying (4.2-12) on both sides by A results in

$$AA^\#A = A(I - A_R)A^I(I - A_L)A = AA^IA = A.\square$$

Now suppose P(0) = P is the total projector associated with λ then

$$L_\lambda(0)P = PL_\lambda(0) = 0$$

since L is diagonable. From Lemma (4.2-11) if L_λ^I is any generalized inverse of L_λ then

$$L_\lambda^\dagger = (I - P)L_\lambda^I(I - P) \tag{4.2-13}$$

is also a generalized inverse of L_λ. Furthermore, since $Pu = u$ and $v^T P = v^T$, $L_\lambda^\dagger$ satisfies

$$L_\lambda^\dagger u = 0 \tag{4.2-14}$$

and

$$v^T L_\lambda^\dagger = 0 \, . \tag{4.2-15}$$

If

$$\tilde{u}^{(i)} = - L_\lambda^\dagger \sum_{j=0}^{i-1} (L^{(i-j)} - \lambda^{(i-j)} I) u^{(j)}, \qquad \tilde{u}^{(0)} = u$$

and

$$u^{(i)} = \tilde{u}^{(i)} + c_i u, \qquad i = 1,2,3,... \tag{4.2-16}$$

then Theorem (2.1-20) guarantees that $u^{(i)}$ is a solution of (4.2-6) for any constants c_i since $c_i u \in \mathcal{N}(L_\lambda)$.

We may use the properties (4.2-14 and 15) to simplify equations (4.2-8 and 9).

If $c_i = 0$ for all i, then equations (4.2-8 and 9) become

$$\lambda^{(i)} = \frac{v^T \sum_{j=0}^{i-1} L^{(i-j)} u^{(j)}}{v^T u} \tag{4.2-17}$$

where u is the value of an analytic eigenvector at x = 0 and v is any vector satisfying $v \in \mathcal{N}(L^T - \lambda I)$ and $v^T u \neq 0$.

If the eigenvector derivatives are to be used explicitly, then the constants c_i should be selected to satisfy some normalization conditions such as

$$v^T(x)\, u(x) = 1 \tag{4.2-18}$$

and

$$(u(x))^*\, u(x) = 1. \tag{4.2-19}$$

If we assume (4.2-18), we obtain

$$v^T u(i) = v^T \tilde{u}(i) + c_i v^T u \tag{4.2-20}$$

also

$$c_i = \frac{v^T u(i)}{v^T u}$$

follows from (4.2-16) since $v^T L_\lambda^\dagger = 0$.

Now suppose u is also normalized so that (4.2-19) holds. Then

$$(u + xu^{(1)} + x^2 u^{(2)} + \ldots)^*(u + xu^{(1)} + x^2 u^{(2)} + \ldots) = 1$$

must be true and the coefficients of x^i, $i \geq 1$, must be zero so

$$\sum_{j=0}^{i} (u^{(i-j)})*u^{(j)} = 0, \qquad i \geq 1$$

or

$$(u^{(0)})*u^{(i)} + (u^{(i)})*u^{(0)} = -\sum_{j=1}^{i-1} (u^{(i-j)})*u^{(j)}.$$

Equating real parts of both sides

$$\mathrm{Re}[u*u^{(i)}] = -\frac{1}{2}\,\mathrm{Re}[\sum_{j=1}^{i-1} (u^{(i-j)})*u^{(j)}]$$

and multiplying (4.2-16) by u* results in

$$u*u^{(i)} = u*\tilde{u}^{(i)} + c_i u*u$$

and

$$c_i = u*u^{(i)} - u*\tilde{u}^{(i)}. \qquad (4.2\text{-}21)$$

Then,

$$\mathrm{Re}[c_1] = -\,\mathrm{Re}[u*\tilde{u}^{(1)}]$$

and

$$\mathrm{Re}[c_i] = -\frac{1}{2}\mathrm{Re}\Big[\sum_{j=1}^{i-1}(u^{(i-j)})^{*}u^{(j)}\Big] - \mathrm{Re}[u^{*}\tilde{u}^{(i)}],\quad i \geq 2 \qquad (4.2\text{-}22)$$

must be satisfied for (4.2-18 and 19) to hold.

Remark (4.2-23): If $L(x)$ is a normal matrix then $u^{*}u^{(i)} = 0$ since $u^{*} = v^{T}$. If $L(x)$ is a real symmetric matrix then $u^{T} = v^{T}$,

$$c_1 = 0$$

and

$$c_i = -\frac{1}{2}\Big[\sum_{j=1}^{i-1}(u^{(i-j)})^{T}u^{(j)}\Big], \qquad i \geq 2.$$

To determine u in the event that $\lambda = \lambda(0)$ has multiplicity m in (4.2-1) we note that since v in (4.2-13) may be selected as any vector in the m-dimensional null space of $L^{T}(0) - \lambda(0)I$ and since $L(x)$ is fully analytic, $\mathcal{N}(L^{T}(0) - \lambda(0)I)$ is spanned by the values at $x = 0$ of m independent analytic vectors, which we may take as the rows of the $m \times n$ matrix $V^{T}(0) = V^{T}$. Then from (4.2-6)

$$V^{T}\sum_{j=0}^{i-1}(L^{(i-j)} - \lambda^{(i-j)}I)u^{(j)} = 0 \qquad (4.2\text{-}23)$$

or

$$V^{T}\sum_{j=0}^{i-1}L^{(i-j)}u^{(j)} = V^{T}\sum_{j=0}^{i-1}\lambda^{(i-j)}u^{(j)} \qquad (4.2\text{-}24)$$

Similarly we form an n x m matrix U(0) = U where the columns of U are derived from the analytic right eigenvectors of λ and

$$V^T \sum_{j=0}^{i-1} L^{(i-j)} U^{(j)} = V^T \sum_{j=0}^{i-1} U^{(j)} \Lambda^{(i-j)} = \Lambda^{(i)} \qquad (4.2\text{-}25)$$

In particular, if i = 1,

$$V^T L^{(1)} U = V^T P L^{(1)} P U = V^T U \Lambda^{(1)} = \Lambda^{(1)} \qquad (4.2\text{-}26)$$

where $\Lambda^{(1)}$ is an m x m diagonal matrix consisting of the m first derivatives of λ and the columns of V and U are normalized so that $V^T U = I$. Equation (4.2-26) shows that V and U must diagonalize the matrices $L^{(1)}$ and $\tilde{L}^{(1)} = P L^{(1)} P$. Furthermore, if $\Lambda^{(1)}$ has distinct values on its diagonal then V and U are unique. We now turn to an example.

Example (4.2-27): Let

$$L(x) = \begin{bmatrix} 1 & x \\ x & 1 \end{bmatrix} = L^{(0)} + x\, L^{(1)}(0) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + x \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

The eigenvalues of L(x) are $\lambda_{12} = 1 \pm x$ and $\lambda_{12}^{(1)} = \pm 1$ for all $x \in \mathscr{C}$.

Analytic eigenvectors of L(x) may be found from the non-zero columns of the rank one constituent idempotent matrices L_1 and L_2. Using Theorem (2.1-33) we arrive at

$$L_1(x) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \qquad L_2(x) = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Then

$$u_1(x) = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} \qquad u_2(x) = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix}$$

are the analytic unit eigenvectors of L(x) and they are the unique unit vectors that form the columns in the matrix S which diagonalizes $L^{(1)}(0)$:

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

Furthermore if v is any vector such that $v^T u_1 \neq 0$ or $v^T u_2 \neq 0$ and $v \in \mathcal{N}(L^T - \lambda I)$ then

$$\underbrace{v^T \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}}_{v^T u_1} u_1 = 1 \text{ or } \underbrace{v^T \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}}_{v^T u_2} u_2 = -1$$

Since if $v^T = (x,y)$ then

$$\frac{(x,y)\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}\frac{\sqrt{2}}{2}\\ \frac{\sqrt{2}}{2}\end{bmatrix}}{x\frac{\sqrt{2}}{2} + y\frac{\sqrt{2}}{2}} = \frac{x\frac{\sqrt{2}}{2} + y\frac{\sqrt{2}}{2}}{x\frac{\sqrt{2}}{2} + y\frac{\sqrt{2}}{2}} = 1$$

or

$$\frac{(x,y)\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}\frac{\sqrt{2}}{2}\\ -\frac{\sqrt{2}}{2}\end{bmatrix}}{x\frac{\sqrt{2}}{2} - y\frac{\sqrt{2}}{2}} = \frac{-x\frac{\sqrt{2}}{2} + y\frac{\sqrt{2}}{2}}{x\frac{\sqrt{2}}{2} - y\frac{\sqrt{2}}{2}} = -1$$

and we have agreement with (4.2-17) for i = 1. Note that it is essential that u (or v) lie on the trajectory of an analytic eigenvector for this formulation to work.

4.3 Numerical Approach

In numerical work the algorithm (inverse iteration) [22],

$$(L - \tilde{\lambda}I)u_{i+1} = v_i \qquad (4.3\text{-}1)$$

$$v_i = u_i/||u_i||_\infty \qquad (4.3\text{-}2)$$

is often used to determine an eigenvector u corresponding to the eigenvalue λ where $\tilde{\lambda} = \lambda + \varepsilon$. Solving (4.3-1) we have

$$u_{i+1} = (L - \tilde{\lambda}I)^{-1} v_i \tag{4.3-3}$$

In practice, the matrix $(L - \tilde{\lambda}I)$ is given an LU factorization and the iteration (4.3-1) is carried out as a series of back substitution stages. Since each back substitution stage only involves solving a triangular linear system the LU decomposition stage is the major effort in the computation [23].

These practical considerations provide a strong incentive to study the structure of $(L - \tilde{\lambda}I)^{-1}$ with the hope that it can be utilized in constructing $L_\lambda^\dagger$.

If $L = L(0)$ is a diagonable matrix with s distinct eigenvalues then

$$(L - \tilde{\lambda}_i I)^{-1} = \sum_{j=1}^{s} (\lambda_j - \tilde{\lambda}_i)^{-1} \underline{L}_j$$

$$(L - \tilde{\lambda}_i I)^{-1} = -\varepsilon^{-1} L_i + \sum_{\substack{j=1 \\ j \neq i}}^{s} (\lambda_j - \tilde{\lambda}_i)^{-1} L_j \tag{4.3-4}$$

follows from Theorem (2.1-32).

Now consider

$$\lim_{\varepsilon\to 0} (I - \underline{L}_i)(L - \lambda_i I)^{-1}(I - \underline{L}_i) = \lim_{\varepsilon\to 0} [\varepsilon^{-1}(I - \underline{L}_i)\underline{L}_i(I - \underline{L}_i)$$

$$+ \sum_{\substack{j=1 \\ j\neq i}}^{s} (\lambda_j - \tilde{\lambda}_i)^{-1}\underline{L}_j]. \qquad (4.3\text{-}5)$$

The first term of the right side of (4.3-5) vanishes since

$(I - \underline{L}_i)\ \underline{L}_i\ (I - \underline{L}_i) = 0$ is a constant and we have

$$L^I_{\lambda i} = (I - \underline{L}_i)\ (L - \lambda_i I)^{-1}\ (I - \underline{L}_i)$$

$$= - \sum_{\substack{j=1 \\ j\neq 1}}^{s} (\lambda_j - \lambda_i)^{-1}\underline{L}_j \qquad (4.3\text{-}6)$$

Furthermore, $L^I_{\lambda_i}$ is a generalized inverse of $L_{\lambda_i} = L - \lambda_i I$ since

$$(L - \lambda_i I)\ (I - \underline{L}_i)(L - \lambda_i(I)^{-1}(I - \underline{L}_i)(L - \lambda_i I) = L - \lambda_i I \qquad (4.3\text{-}7)$$

Finally since $P_i(0)$, the total projector of the λ_i - group, is equal to $\underline{L}_i$ we have (dropping the subscript i),

$$L^I_\lambda = (I - P)(L - \lambda I)^{-1}(I - P) \qquad (4.3\text{-}8)$$

From Lemma (4.2-11) however,

$$(I - P)L^I_\lambda\ (I - P) = L^\dagger_\lambda \qquad (4.3\text{-}9)$$

is also a generalized inverse of L_λ but since I - P is idempotent $L_\lambda^I = L_\lambda^\dagger$. Therefore we write

$$L_\lambda^\dagger = (I - P)(L - \lambda I)^{-1}(I - P) \tag{4.3-10}$$

and note that $L_\lambda^\dagger$ satisfies conditions (4.2-14) and (4.2-15), namely

$$L_\lambda^\dagger u = 0 \qquad \text{and} \qquad v^T L_\lambda^\dagger = 0.$$

This form of $L_\lambda^\dagger$ is especially convenient since $(L - \tilde{\lambda}I)^{-1}$ may be computed previously to determine the eigenvector u and since the idempotent matrix I - P is easily multiplied by a vector particularly if P has only rank one, i.e.,

$$(I - uv^T)b = b - (v^T b)u \tag{4.3-11}$$

or

$$b^T(I - uv^T) = b^T - (b^T u)v^T. \tag{4.3-12}$$

In practice, instead of directly inverting $(L - \tilde{\lambda}I)^{-1}$, an LU decomposition of $L - \tilde{\lambda}I$ is formed.

The following has proved to be an efficient procedure if λ is non-repeated.

1. Form an LU decomposition of $L - \tilde{\lambda}I$.
2. Use the LU decomposition in an inverse iteration scheme to find an eigenvector u.
3. Solve for z in $L_L z = (I - uv^T) \frac{dL}{dx} u$.
4. Solve for y in $L_u y = z$.
5. $\frac{du}{dx} = (I - uv^T)y = y - (v^T y)u + c_1 u$.

where L_u and L_L are upper and lower triangular matrices computed in the LU factorization of $L - \tilde{\lambda}I$.

If λ is a repeated eigenvalue with multiplicity m then it may be necessary to resort to a technique such as the method of diagonalizing $PL^{(1)}P$ given in Section 4.2 to find an analytic eigenvector u, before proceeding with,

3. Solve for z in $L_L z = (I - P) \frac{dL}{dx} u$.
4. Solve for y in $L_u y = z$.
5. $\frac{du}{dx} = (I - P)y + c_1 u$.

Chapter 5

Design Sensitivity in Finite Element Models of Dynamical Structures

The finite element method is used extensively for solving boundary value problems involving partial differential operators. Its success depends largely on its ability to deal effectively with complicated domains and boundary conditions. Its major disadvantage is that it often requires large amounts of computer time to find solutions. It makes sense then to develop efficient methods to update existing solutions when small design changes are made in order to avoid a complete re-analysis.

In this chapter, methods for computing a linear dynamical system's sensitivity to design changes are discussed and two illustrative examples are presented. The methods described are directly applicable to solutions generated by the finite element method [1, 24-26] or similar methods. To begin we shall outline how the finite element method can be derived for a static homogeneous boundary value problem in the plane.

5.1 The Finite Element Method

Consider the boundary value problem

$$\begin{aligned} Lu &= f \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \partial\Omega \end{aligned} \tag{5.1-1}$$

where $L: H^1(\Omega) \to L_2(\Omega)$ is the 2nd order partial differential operator:

$$Lu = -\frac{\partial}{\partial x}[p\,\frac{\partial u}{\partial x}] - \frac{\partial}{\partial y}[q\,\frac{\partial u}{\partial y}] + cu,$$

$H^1(\Omega) = \{u \in L_2(\Omega): \frac{\partial u}{\partial x} = u_x$ and $\frac{\partial u}{\partial y} = u_y \in L_2(\Omega)\}$, p,q,c and $f \in L_2(\Omega)$, and Ω is a connected, closed, and bounded domain in $\mathscr{R}^2$ with boundary $\partial\Omega$.

Now suppose that we define the following inner product on $H^1(\Omega)$ to be

$$(u,v)_{H^1} = \int_\Omega (u_x v_x + u_y v_y + uv) \tag{5.1-2}$$

for any $u, v \in H^1(\Omega)$, and consider the quantity

$$(Lu,v) \overset{\Delta}{=} (u,v)_L$$

with $v = 0$ on $\partial\Omega$.

If $(u,v)_L$ is integrated by parts over the domain Ω, an application of Green's identities yields

$$(u,v)_L = \int_\Omega (pu_x v_x + qu_y v_y + cuv) + \int_{\partial\Omega} (-pu_x v + qu_y v).$$

$$= \int_\Omega (pu_x v_x + qu_y v_y + cuv). \tag{5.1-4}$$

It can be shown that [25] if p,q and c are strictly positive and continuous then positive constants α and β exist such that

$$\alpha[(u,u)]^{1/2} \leq [(u,u)_L]^{1/2} \leq \beta[(u,u)]^{1/2} \qquad (5.1\text{-}5)$$

and so $(u,u)_L = 0$ if and only if $u = 0$.

Since $(u,v)_L = (v,u)_L$ (symmetric),

and

$$(\alpha u_1 + \beta u_1, v)_L = \alpha(u_1,v)_L + \beta(u_2,v) \qquad \text{(linear)}$$

we have established that $(u,v)_L$ is an inner product on $H_0^1(\Omega) = \{u \in H^1(\Omega): u = 0 \text{ on } \partial\Omega\}$.

Since $(u,v)_L$ is an inner product on $H_0^1(\Omega)$ we may now apply the Projection Theorem (2.2-20) to find a numerical solution for u.

To find an approximate solution to 5.1.1 let $M \subset H_0^1$ be spanned by the Hamel basis $B = \{\phi_1,\phi_2,\ldots,\phi_m\}$, then

$$\hat{u} = \sum_{i=1}^{m} z_i\phi_i \qquad (5.1\text{-}8)$$

is a vector in $M \subset H_0^1$. We seek $\tilde{u}$ such that

$$||L\tilde{u} - f|| \leq ||L\hat{u} - f|| \text{ , for all } \hat{u} \in M$$

By Theorem (2.2-20), $\tilde{u}$ must satisfy

$$(L\tilde{u} - f, \phi_i) = 0$$

$$(L\tilde{u}, \phi_j) = (f, \phi_j)$$

$$(\tilde{u}, \phi_j)_L = (f, \phi_j)$$

$$\left(\sum_{i=1}^{n} z_i \phi_i, \phi_j \right)_L = (f, \phi_j)$$

$$\sum_{i=1}^{n} (\phi_i, \phi_j)_L z_j = (f, \phi_j) \qquad \text{for all } \phi_j \in B \qquad (5.1\text{-}9)$$

which we write in matrix form as

$$Kz = f \qquad (5.1\text{-}10)$$

where $k_{ij} = (\phi_i, \phi_j)_L$.

The functions $\phi_i \in H_0^1(\Omega)$ may be constructed by triangulating the domain Ω [25] as shown in Figure (5.1-11). We then require $\phi_i(x,y)$ to satisfy

$$\phi_i(n_j) = \delta_{ij}$$

where $n_j = (x_j, y_j)$ is a vertex in the triangulation of Ω. This choice of the ϕ_i will tend to cause K to have a banded structure

since K is populated by inner products $(\phi_i, \phi_j)_L$ which are non-zero only if n_i and n_j are sufficiently close in the triangulation.

Since the matrix K was generated by using the Projection Theorem (2.2-20) with the inner product $(\cdot,\cdot)_L$ on the space $H_0^1(\Omega)$, we would, in general expect K to change if the domain Ω were to change. Thus, it is natural to consider the variable finite element problem

$$K(\Omega)z(\Omega) = f(\Omega) \tag{5.1-12}$$

where Ω is now considered to be a variable domain. Suppose

$$T(\Omega) = \Omega', \qquad T: \mathscr{R}^2 \to \mathscr{R}^2 \tag{5.1-13}$$

is a mapping operating on the domain Ω such that

$$\Omega' = \Omega(\alpha) = \Omega(0) + \alpha\Delta, \qquad \alpha \in D \tag{5.1-14}$$

where $\Delta \subset \mathscr{R}^2$ is a connected, closed and bounded set whose points are such that every point in Δ corresponds to some point in $\Omega(0) \subset \mathscr{R}^2$ where the addition operation in (5.1-13) is vector addition of the points in $\Omega(0)$ to the corresponding points in Δ and where T is a topological homeomorphism for $\alpha \in D$. Then if $\Omega(0)$ is given a triangulation, we may construct a matrix $K(\alpha)$ such that

$$K(\alpha) = K(\Omega(\alpha)) = K(\Omega(0) + \alpha\Delta)$$

and the variable problem $K(\alpha)x(\alpha) = f(\alpha)$ results. We now define the directional derivative of $K(\alpha)$ as

$$DK(\Omega(0),\Delta) = \lim_{\alpha \to 0} \frac{K(\Omega(0) + \alpha\Delta) - K(\Omega(0))}{\alpha} \qquad (5.1\text{-}15)$$

(If this limit exists for any Ω as defined above then $K(\alpha)$ is Gateaux differentiable and $DK(\Omega(0),\Delta)$ is its Gateaux derivative.)

5.2 Dynamical Problems

The distributed parameter forced vibration problem (without damping) can be written as

$$L_1 \frac{\partial^2 u}{\partial t^2}(x,t) + L_2 u(x,t) = f(x,t) \qquad (5.2\text{-}1)$$

$$B(u) = g(x,t) \text{ on } \partial\Omega x T \qquad (5.2\text{-}2)$$

$$C(u) = r(x,0) \text{ in } \Omega \qquad (5.2\text{-}3)$$

where L_1 and L_2 are self adjoint positive definite operators on Ω, u is taken from a function space $V(\Omega x T)$ defined on the product space $\Omega x T$ derived from the spacial domain Ω $(k = 2,3)$ and T is the temporal domain $[0,t]$. In addition, boundary conditions

(5.2-2) defined on $\partial\Omega x T$, where $\partial\Omega$ is the spatial boundary of Ω, may be given along with initial conditions (5.2-3).

When the finite element method is used to obtain an approximate solution to (5.2-1, 2 and 3), it is usual to arrive at a lumped parameter model of the form

$$M\ddot{z}(t) + Kz(t) = \bar{f}(t) \tag{5.2-4}$$

where M is an n x n matrix

K is an n x n matrix

z(t) is an n-vector consisting of n = km time varying coefficients of the spline basis functions $\phi_i \in B$ related to the triangulation of Ω

and $\bar{f}(t)$ is an n vector resulting from projecting f(x,t) onto the finite subspace spanned by B.

We are mainly concerned with the variable problem

$$M(\alpha)\ddot{z}(\alpha,t) + K(\alpha)z(\alpha,t) = \bar{f}(\alpha,t) \tag{5.2-5}$$

and the related generalized eigenvalue problem [28]

$$[K(\alpha) - \lambda(\alpha)\, M(\alpha)]u(\alpha) = 0 \tag{5.2-6}$$

where α accounts for a changing spacial domain

$$\Omega(\alpha) = \Omega(0) + \alpha\Delta$$

This problem can be related to the symmetric eigenvalue problem in standard form through the use of the following transformation [8],

$$u = M^{-1/2}v, \tag{5.2-7}$$

where $M^{-1/2}$ is the positive definite square root of M^{-1} and dependence on α is understood. Then

$$M^{-1/2}KM^{-1/2}v = \lambda v \tag{5.2-8}$$

Since $M^{-1/2}$ is positive definite and since K and M are analytic functions of $\alpha \in D$, $M^{-1/2}$ and therefore $M^{-1/2}KM^{-1/2}$ are analytic on D. Thus (5.2-6) is equivalent to the variable eigenvalue problem in standard form of the previous section and the eigenvalues $\lambda_i(\alpha)$ and particular corresponding eigenvectors $v_i(\alpha)$ $i = 1, n$, are analytic for $\alpha \in D$.

It is not convenient to determine the derivatives of $\lambda_i(\alpha)$ and $v_i(\alpha)$ from (5.2-8) since the transformation $M^{-1/2}(\alpha)$ may be difficult to compute and the transformed matrix in (5.2-8) is usually difficult to differentiate. Instead (5.2-6) is used. After differentiating and rearranging (5.2-6) we arrive at

$$[K - \lambda M]\frac{du}{d\alpha} = -\left[\frac{dK}{d\alpha} - \lambda\frac{dM}{d\alpha} - \frac{d\lambda}{d\alpha}M\right]u \tag{5.2-9}$$

Solving for $\frac{du}{d\alpha}$ with a generalized inverse D_λ^I, we obtain

$$\frac{du}{d\alpha} = - D_\lambda^I \left[\frac{dK}{d\alpha} - \lambda\frac{dM}{d\alpha} - \frac{d\lambda}{d\alpha} M\right]u + z \qquad (5.2\text{-}10)$$

where

$$D_\lambda = (K - \lambda M)$$

and

$$D_\lambda^I = (K - \lambda M)^I \qquad \text{satisfy}$$

$$D_\lambda \, D_\lambda^I \, D_\lambda = D_\lambda \qquad (5.2\text{-}11)$$

and $z \in \mathcal{N}(K - \lambda M)$. We may assume that analytic eigenvectors exist such that

$$u_i^T(\alpha)M(\alpha)u_j(\alpha) = \delta_{ij} \qquad (5.2\text{-}12)$$

since after substituting (5.2-7) into

$$v_i^T(\alpha)v_j(\alpha) = \delta_{ij} \qquad (5.2\text{-}13)$$

we have

$$u_i^T \, M^{1/2} \, M^{1/2} \, u_j = u_i^T \, M \, u_j = \delta_{ij}. \qquad (5.2\text{-}14)$$

Differentiating (5.2-12) gives

$$\frac{du_i^T}{d\alpha} Mu_j + u_i^T \frac{dM}{d\alpha} u_j + u_i^T M \frac{du_j}{d\alpha} = 0.$$

Now suppose u satisfies (5.2-12). Then from Lemma (4.2-11)

$$D_\lambda^\dagger = (I - uu^TM)D_\lambda^I(I - Muu^T) \tag{5.2-15}$$

is also a generalized inverse of D_λ since

$$D_\lambda\, uu^TM = Muu^T\, D_\lambda = 0. \tag{5.2-16}$$

Substituting $D_\lambda^\dagger$ in (5.2-10) with z = cu gives

$$\frac{du}{d\alpha} = - D_\lambda^\dagger[\frac{dK}{d\alpha} - \frac{dM}{d\alpha} - \frac{d\lambda}{d\alpha} M]u + z$$

$$= - D_\lambda^\dagger[\frac{dK}{d\alpha} - \lambda\frac{dM}{d\alpha}]u + cu \tag{5.2-17}$$

Multiplying (5.2-17) by u^TM and applying (5.2-12) shows that

$$c = - \frac{1}{2} u^T \frac{dM}{d\alpha} u \tag{5.2-18}$$

Again we must determine $D_\lambda^\dagger$ to complete the formulation. The approach is similar to the one used in the standard eigenvalue problem.

Suppose S is a matrix such that

$$S^T[K - \lambda M]S = \Lambda - \lambda I \tag{5.2-19}$$

Then we may expand $K - \lambda M$ as [29]

$$K - \lambda M = S(\Lambda - \lambda I)S^T = \sum_{i=1}^{n} (\lambda_i - \lambda)G_i \tag{5.2-20}$$

where

$$G_i = \frac{u_i u_i^T}{u_i^T M u_i}$$

and

$$G_i M G_j = \delta_{ij} G_i. \tag{5.2-22}$$

Similarly,

$$(K - \tilde{\lambda}M)^{-1} = \sum_{i=1}^{n} (\lambda_i - \tilde{\lambda})^{-1} G_i \tag{5.2-23}$$

If P_G is the sum of the m matrices G_k associated with eigenvalue λ of multiplicity m then

$$D_\lambda^\dagger = (I - P_G M)(K - \tilde{\lambda}M)^{-1}(I - MP_G) \tag{5.2-24}$$

is a generalized inverse of $K - \lambda M$ such that

$$D_\lambda^\dagger M u = 0 \tag{5.2-25}$$

$$u^T M D_\lambda^\dagger = 0. \tag{5.2-26}$$

If u is determined through the inverse iteration algorithm [26]

$$(K - \tilde{\lambda}M)u_{i+1} = Mv_i$$
$$v_i = u_i/||u_i||_\infty \tag{5.2-27}$$

then the algorithm given in (4.3) is essentially unchanged.

The eigenvalue derivative may be found by multiplying (5.2-9) by v^T, $v \in \mathcal{N}(K - \lambda M)$ and solving for $d\lambda/d\alpha$

$$\frac{d\lambda}{d\alpha} = \frac{v^T (\frac{dK}{d\alpha} - \lambda\frac{dM}{d\alpha})u}{v^T Mu}, \qquad \text{where } v^T Mu \neq 0. \tag{5.2-28}$$

The system (5.1-4) can be modified to include dissipative terms by adding the term $C\dot{z}$ to the left side [28]. We then have the damped system

$$M(\alpha)\ddot{z}(\alpha,t) + C\dot{z}(\alpha,t) + Kz(\alpha,t) = \bar{f}(\alpha,t) \tag{5.2-29}$$

and the related generalized eigenvalue problem

$$(\lambda^2(\alpha)M(\alpha) + \lambda(\alpha)C(\alpha) + K(\alpha))u(\alpha) = 0 \tag{5.2-30}$$

where the positive semi-definite matrix C accounts for viscous damping.

The 'damped' eigensystem of equation (5.2-30) is usually more difficult to solve than the corresponding system (5.2-6) without damping. However, under special circumstances the eigenvectors of the damped system are also eigenvectors of the corresponding undamped system. This is true, for example, if C is some linear combination of K and M, i.e.,

$$C = \beta K + \gamma M, \qquad \beta,\gamma \in \mathcal{R} \tag{5.2-31}$$

It has been shown that [30]

$$KM^{-1}C = CM^{-1}K \tag{5.2-32}$$

is a necessary and sufficient condition for the eigensystem of (5.2-30) to be uncoupled when transformed to the modal coordinates of the corresponding eigensystem of (5.2-6). If (5.2-32) holds for $\alpha \in D \subseteq \mathcal{R}$ then the eigenvectors of the damped system do not depend on $C(\alpha)$ and their derivatives may be computed directly from (5.2-17).

The eigenvalue derivatives may be determined by differentiating (5.2-30) and solving for $d\lambda/d\alpha$ to obtain

$$\frac{d\lambda}{d\alpha} = \frac{v^T(\lambda^2 \frac{dM}{d\alpha} + \lambda\frac{dC}{d\alpha} + \frac{dK}{d\alpha})u}{v^TCu + 2\lambda v^TMu} \tag{5.2-33}$$

Remark: Some caution must be taken when using this formula since the eigenvalues are not differentiable for particular values of the matrix C. For example, suppose u is an eigenvector of (5.2-30) then

$$\lambda^2 u^T M u + \lambda u^T C u + u^T K u = 0 \tag{5.2-34}$$

is satisfied by

$$\lambda = \frac{-u^T C u \pm [(u^T C u)^2 - 4u^T M u \; u^T K u]^{1/2}}{2u^T M u}. \tag{5.2-35}$$

If

$$(u^T C u)^2 = 4u^T M u \; u^T K u \tag{5.2-36}$$

then the mode corresponding to u is said to be critically damped [31] and the eigenvalue λ is not analytic. This can be verified by attempting to differentiate (5.2-35).

5.3 Examples

This section is devoted to two design sensitivity examples involving finite element vibration models.

In the first example we will use the methods presented in Chapters 4 and 5 to compute a Taylor series for the lowest natural frequency of a triangular plane elastic element. The design variation consists of a changing node (vertex) position. The

second example involves a fixed-fixed plate assembled from several plane elastic triangular elements which is undergoing boundary shape variations.

Example (5.3-1): Consider the plane elastic element shown in Figure (5.3-2). Following standard procedures and notation [1,2], the stiffness matrix is given as a function of x by

$$K(x) = \frac{t}{4}\begin{bmatrix} 4 & 0 & 2(x-1) & -2 & -2(x+1) & 2 \\ & 4 & 0 & 2(x+1) & 0 & -2(x+1) \\ & & x^2-2x+2 & -(x-1) & -x^2 & x-1 \\ & & & x^2-2x+2 & x+1 & -x^2 \\ & & & & x^2+2x+2 & -(x+1) \\ \text{symmetric} & & & & & x^2+2x+2 \end{bmatrix} \quad (5.3\text{-}3)$$

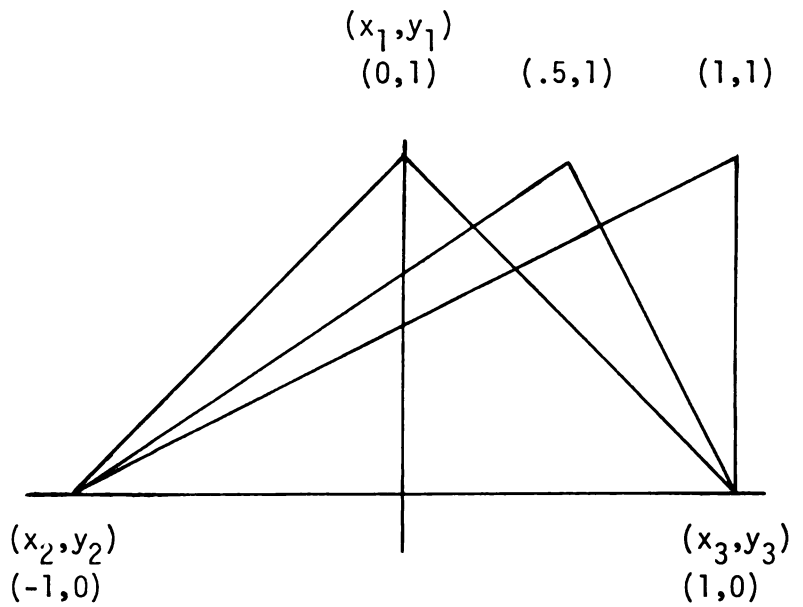


Figure 5.3-2 Triangular Element (shown in three design positions)

where t is the plate thickness,

$x = x_1$,

$$K(x) = \int_V B^T(x)\, C\, B(x)\, dV,$$

$$B(x) = \frac{1}{2A} \begin{bmatrix} b_1 & 0 & b_2 & 0 & b_3 & 0 \\ 0 & a_1 & 0 & a_2 & 0 & a_3 \\ a_1 & b_1 & u_2 & b_2 & a_2 & b_3 \end{bmatrix},$$

$$a_1 = x_3 - x_1 = 2 \qquad b_1 = y_2 - y_3 = 0$$
$$a_2 = x_1 - x_3 = x_1 - 1 \qquad b_2 = y_3 - y_1 = -1$$
$$a_3 = x_2 - x_1 = -x_1 - 1 \qquad b_3 = y_1 - y_2 = 1 \quad ,$$

$2A = b_1x_1 + b_2x_2 + b_3x_3 = 2$ and

$V = At.$

Then $K^{(1)} = \frac{dK}{dx} =$

$$\frac{t}{4} \begin{bmatrix} 0 & 0 & 2 & 0 & -2 & 0 \\ & 0 & 0 & 2 & 0 & -2 \\ & & 2x-1 & -1 & -2x & 1 \\ & & & 2x+2 & 1 & -2x \\ \text{symmetric} & & & & 2x+2 & -1 \\ & & & & & 2x+2 \end{bmatrix} \qquad (5.3\text{-}4)$$

$$K^{(2)} = \frac{1}{2}\frac{d^2K}{dx^2} = \frac{t}{8}\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 0 & 0 \\ & & 2 & 0 & -2 & 0 \\ & & & 2 & 0 & -2 \\ & & & & 2 & 0 \\ & \text{symmetric} & & & & 2 \end{bmatrix} \qquad (5.3\text{-}5)$$

$$K^{(i)} = 0 \qquad i \geq 3$$

For the mass matrix we use

$$M = \frac{At}{3}I = \frac{t}{3}I$$

which is a lumped mass formulation [2].

Since the area A is a constant

$$M^{(i)} = 0 \qquad i \geq 1$$

If the plate thickness t is set equal to 3, M = I and

$$D_\lambda(x) = K(x) - \lambda(x)I \ .$$

The smallest non-zero eigenvalue of K(0) is determined as

$$\lambda(0) = (\frac{3}{4})2 = \frac{3}{2}$$

and its corresponding unit eigenvector is

$$u(0) = \frac{\sqrt{2}}{2} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}$$

We may now use (4.2-16 and 17) to determine a Taylor series for $\lambda(x)$.

For $L_\lambda^\dagger = D_\lambda^\dagger$ we use (4.2-13), i.e.

$$D_\lambda^\dagger = ([I - uu^T])\ (K - \tilde{\lambda}I)^{-1}\ (I - uu^T)$$

to find that

$$D_\lambda^\dagger(0) = \frac{1}{18} \begin{bmatrix} -4 & 0 & -4 & -4 & -4 & 4 \\ 0 & 0 & 0 & -6 & 0 & -6 \\ -4 & 0 & -4 & 2 & -4 & -2 \\ -4 & -6 & 2 & -7 & 2 & 1 \\ -4 & 0 & -4 & 2 & -4 & -2 \\ 4 & -6 & -2 & 1 & -2 & 7 \end{bmatrix}$$

The first 10 Taylor series coefficients to single precision accuracy are

$$\lambda^{(1)} = 0.0$$
$$\lambda^{(2)} = -.500000$$
$$\lambda^{(3)} = 0.0$$
$$\lambda^{(4)} = .138890$$
$$\lambda^{(5)} = 0.0$$
$$\lambda^{(6)} = .001543$$
$$\lambda^{(7)} = 0.0$$
$$\lambda^{(8)} = -.052641$$
$$\lambda^{(9)} = 0.0$$
$$\lambda^{(10)} = .061548$$

With these coefficients, $\lambda(x)$ may be approximated as

$$\lambda(x) = \sum_{i=0}^{n} x^i \lambda^{(i)}.$$

In Table (5.3-6a) a Taylor series representing $\lambda(x)$ is evaluated for $x = 0$ to $x = 1.5$ and for $n = 2, 4, 6, 8, 10$ and 20. In Table (5.3-6b) a direct numerical solution for $\lambda(x)$ is presented for the same values of x. Figure (5.3-7) presents these data graphically.

A comparison between the values of $\lambda(x)$ given by the Taylor series and the values of $\lambda(x)$ obtained by direct evaluation show excellent agreement even when $n = 2$.

Table 5.3-6

x	n = 2	n = 4	n = 6	n = 8	n = 10	n = 20	x	λ(x)
0.00	1.500000	1.500000	1.500000	1.500000	1.500000	1.500000	0.00	1.500000
0.10	1.495000	1.495014	1.495014	1.495013	1.495013	1.495013	0.10	1.495013
0.20	1.480000	1.480222	1.480222	1.480222	1.480222	1.480222	0.20	1.480222
0.30	1.455000	1.456125	1.456126	1.456122	1.456122	1.456122	0.30	1.456123
0.40	1.420000	1.423556	1.423562	1.423527	1.423534	1.423532	0.40	1.423533
0.50	1.375000	1.383680	1.383704	1.383499	1.383559	1.383548	0.50	1.383548
0.60	1.320000	1.338000	1.338072	1.337187	1.337559	1.337471	0.60	1.337472
0.70	1.255000	1.288347	1.288528	1.285494	1.287232	1.286715	0.70	1.286711
0.80	1.180000	1.236889	1.237293	1.228461	1.235070	1.232753	0.80	1.232674
0.90	1.095000	1.186125	1.186944	1.164284	1.185745	1.177636	0.90	1.176675
1.00	1.000000	1.138889	1.140432	1.087791	1.149338	1.128685	1.00	1.119873
1.10	0.895000	1.098347	1.101080	0.988241	1.147878	1.127840	1.10	1.063237
1.20	0.780000	1.068000	1.072607	0.846263	1.227347	1.402026	1.20	1.007543
1.30	0.654999	1.051680	1.059128	0.629724	1.478205	3.021443	1.30	0.953379
1.40	0.519999	1.053555	1.065174	0.288312	2.068594	10.430609	1.40	0.901172
1.50	0.374999	1.078125	1.095701	-0.253417	3.295709	40.169945	1.50	0.851213

(a) Eigenvalue estimates with Taylor series of n terms.

(b) Direct evaluation of eigenvalue.

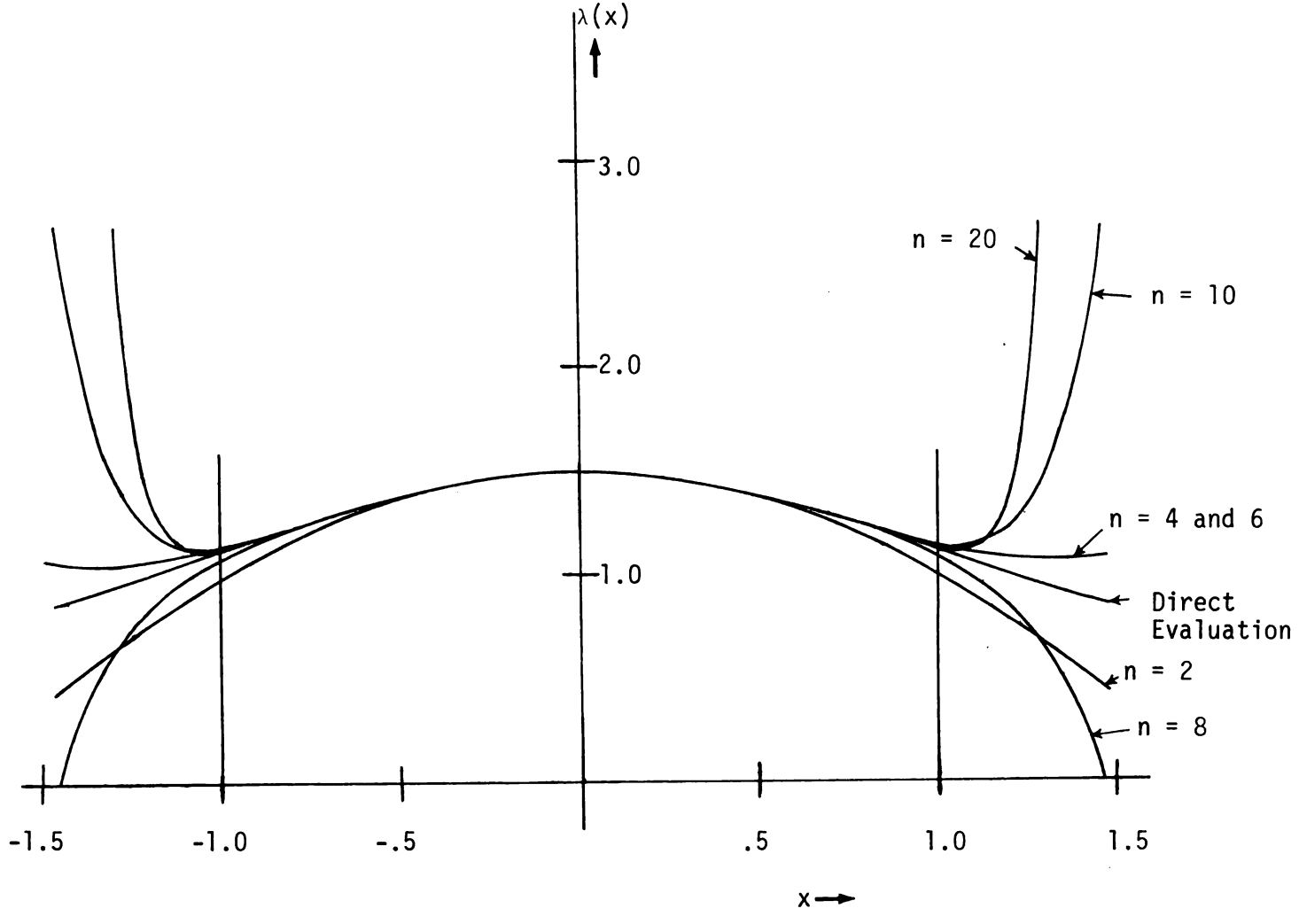


Figure 5.3-7 Eigenvalue Dependence on Design

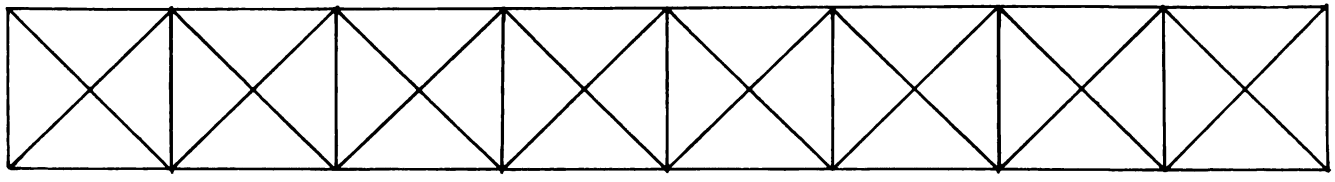
The efficiency of the technique may make design sensitivity calculations of eigenvalues and eigenvectors feasible in many practical situations. For example, if an inverse iteration method is used to determine the eigenvector at x = 0 then subsequent determinations of each of the Taylor coefficients $\lambda^{(i)}$ and $u^{(i)}$ involve little extra effort since the generalized inverse $D_\lambda^\dagger$ is easily determined from 4.2-13 (see Section 4.3). (The method is especially efficient if the matrix K(x) is only linear or quadratic in x since in these cases most of the matrices $K^{(i)} = L^{(i)}$ in Equations (4.2-16 and 17) are null.) Since the Taylor coefficients can be obtained so economically, there is a strong incentive to represent the design with a Taylor series. Then subsequent evaluations of the eigenvalue and eigenvector as the design is modified can be carried out with substantially less computational effort.

The next example involves a finite element model assembled from thirty two plane elastic elements similar to the element of example (5.3-1). The example illustrates the results of a design sensitivity calculation for an eigenvalue depending on a variable boundary shape.

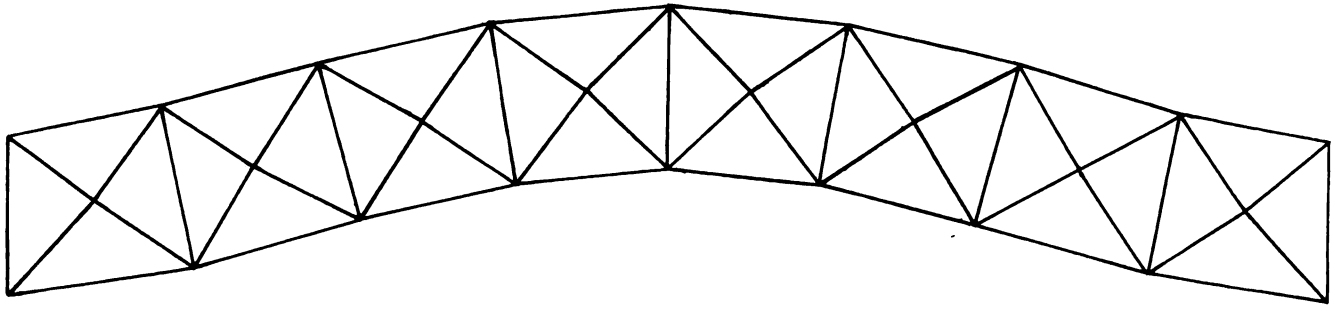
Example (5.3-7): Figures (5.3-8) through (5.3-12) illustrate the sensitivity on the lowest eigenvalue of a dynamic finite model to changes of the boundary shape. Figure (5.3-8a) shows the baseline design. The lowest eigenvalue for the baseline design is calculated as 14.44. (Figure (5.3-8b) illustrates

the mode shape corresponding to this eigenvalue.) In Figures (5.3-9) through (5.3-12) the shape of the lower edge of the model is changed. In each case an estimate of the new eigenvalue is compared with the result of a finite element run for the new configuration. The estimates are based on a single first order directional derivative of the eigenvalue with respect to some direction h (see Definition (2.2-10)) which is related to the variable boundary. The direction h is determined by a gradient projection method as a direction (in the design space) which causes the eigenvalue to increase without increasing the volume (mass) of the plate. (For a comprehensive discussion of this and similar techniques see [3].) Each subsequent variation in the design is determined by moving along the direction h an increased distance.

A comparison between a direct evaluation of λ and the first order estimate of λ shows good agreement for moderate changes in the design but for extreme design changes the estimate is substantially higher than the direct evaluation.



(a) Triangulation of Fixed-Fixed Plate



(b) Mode Shape Corresponding to Eigenvalue = 14.4436

Figure 5.3-8 Baseline Design

Estimated Eigenvalue = 20.0440

Computed Eigenvalue = 20.3973

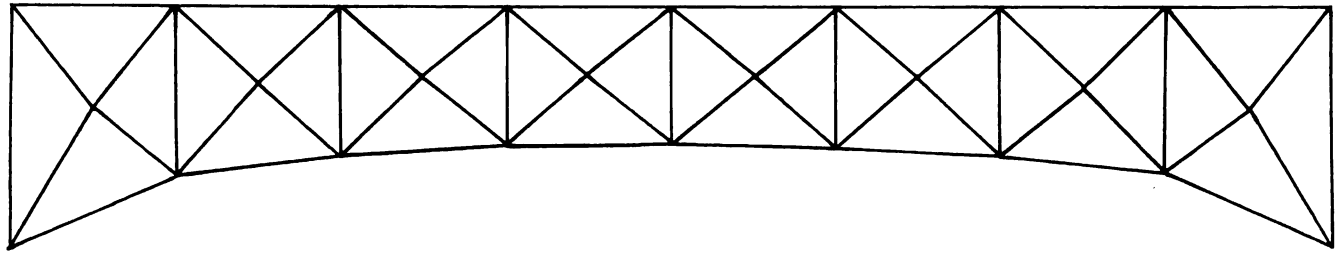


Figure 5.3-9 First Design Change

Estimated Eigenvalue = 25.6444

Computed Eigenvalue = 26.8654

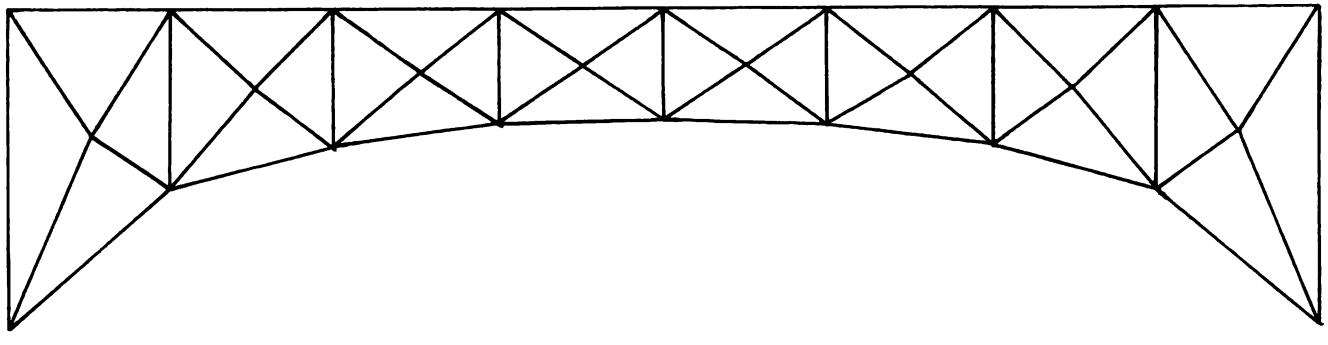


Figure 5.3-10 Second Design Change

Estimated Eigenvalue = 31.2447

Computed Eigenvalue = 34.3904

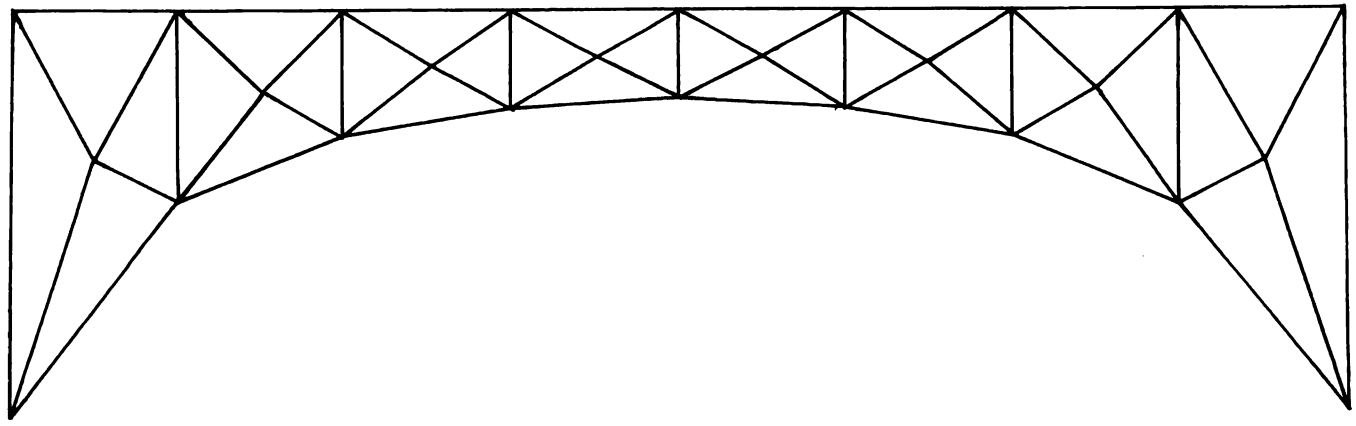


Figure 5.3-11 Third Design Change

Estimated Eigenvalue = 36.8451

Computed Eigenvalue = 44.1197

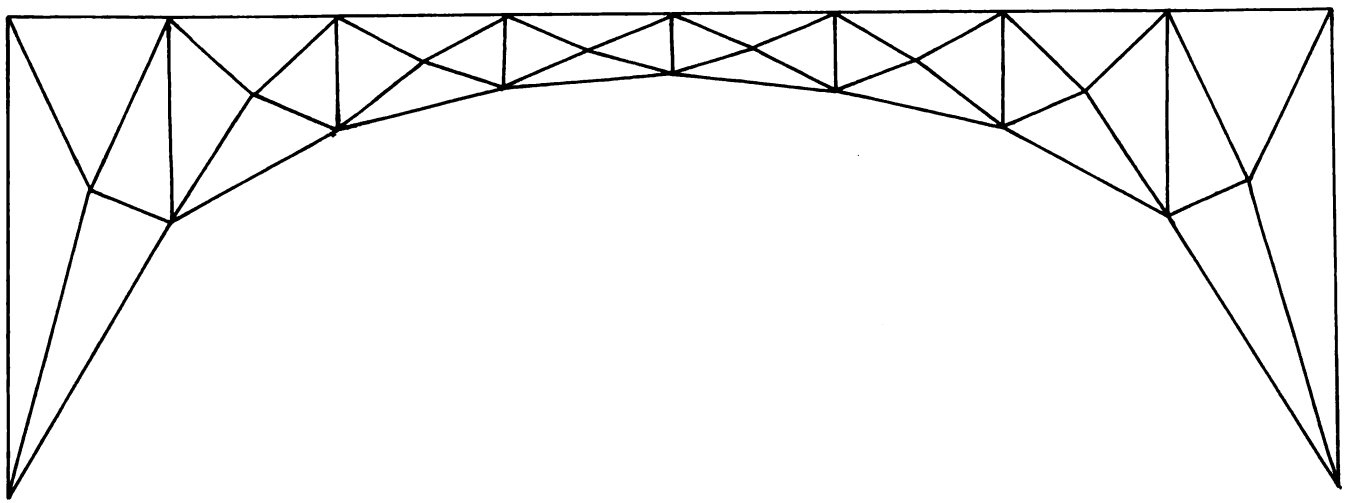


Figure 5.3-12 Fourth Design Change

Chapter 6

Discussion and Further Study

6.1 Discussion

Modal analysis has become a standard tool used in the design of dynamical systems. By numerically determining certain eigenvalues and eigenvectors associated with a linear structural model a designer can investigate how a particular structural design will perform dynamically. The eigenvalues are related to the natural frequencies of the structure and the eigenvectors indicate the associated vibratory modes.

If the designer is concerned with dynamical performance, several redesigns and re-evaluations of the eigenvalues and eigenvectors may be necessary before a suitable design is found. Since this procedure is costly in computer time, there has been an incentive to find improved methods.

This thesis presented a technique to represent an eigenvalue and eigenvector by a Taylor series in a design variable. Since the Taylor coefficients can be determined to arbitrary order, excellent estimates may be easily calculated for design changes which do not exceed the radius of convergence of the series. Since the coefficients of the Taylor series can be computed with substantially less effort than recomputing the eigenvalue or eigenvector, the procedure has the potential to reduce the number

of calculations necessary to carry out the design of dynamical structures. There is also the potential to use the method in the implementation of optimal design algorithms.

Two representative example problems were presented. In one of them, twenty Taylor coefficients of a specified eigenvalue were determined after a single finite element run. These coefficients were used to represent the eigenvalue as a power series in the design variable. Comparison between separate finite element runs and the Taylor series showed excellent agreement.

6.2 Methods

Each Taylor coefficient is determined by repeatedly evaluating a simple recursive formula which is derived through an application of generalized inverse theory. The procedure is simplified by using a generalized inverse matrix specially selected to annihilate certain terms in the formulation. The generalized inverse itself need not be formed explicitly if the eigenvector is determined through the inverse iteration method, with the result that each Taylor series coefficient is computed with no more than $O(n^2)$ multiplications.

The technique was extended to compute derivatives for the eigenvalues and eigenvectors of the generalized eigenvalue problem $(K - \lambda M)u = 0$ and there was some discussion of the damped eigenvalue problem $(M\lambda^2 + \lambda C + K)u = 0$.

6.3 Suggestions for Further Study

There are several questions involving eigensystem derivatives which deserve further study.

Necessary conditions for eigenvalue differentiability are in short supply. It may be possible to use Equation (4.2-25) to construct a sequence of necessary conditions for the full analyticity of diagonable matrices similar to the sufficient condition given by the reduction process. (Equation (4.2-26) amounts to such a necessary condition.)

A convenient computational method for determining the radius of convergence for Taylor series representing eigenvalues and eigenvectors would be useful in applications. Lower bounds have been given [14] which unfortunately often severely underestimate the actual convergence radius.

Most of the results of this thesis can be extended to eigenvalue problems involving linear operators on Hilbert spaces. Excellent references for continued study are Refs [10-14 and 19].

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Zienkiewicz, O.C., The Finite Element Method, McGraw-Hill, New York, 1978.

[2] Desai, C.S. and Abel, J.F., Introduction to the Finite Element Method, Van Nostrand Reinhold Co., New York, 1972.

[3] Haug, E.J., and Arora, J.S., Applied Optimal Design, John Wiley & Sons, New York, 1976.

[4] Jacobi, C.G.J., Crelle's Journal Für Die Reine Und Angewand Mathematik, Vol. 30, Berlin: De Gruyter, 1846, pp. 51-95.

[5] Fox, R.L. and Kapoor, M.P., "Rates of Change of Eigenvalues and Eigenvectors", AIAA Journal, Vol. 6, Dec. 1968, pp. 2426-2429.

[6] Plaut, R.H. and Huseyin, K., "Derivatives of Eigenvalues and Eigenvectors in Non-Self Adjoint Systems", AIAA Journal, Vol. II, Feb. 1973, pp. 250-251.

[7] Rogers, L.C., "Derivatives of Eigenvalues and Eigenvectors", AIAA Journal, Vol. 8, May 1970, pp. 943-944.

[8] Rudisill, C.S., "Derivatives of Eigenvalues and Eigenvectors for a General Matrix", AIAA Journal, Vol. 12, May 1974, pp. 721-722.

[9] Nelson, R.B., "Simplified Calculation of Eigenvector Derivatives", AIAA Journal, Vol. 14, Sept. 1976, pp. 1201-1205.

[10] Frame, J.S., "Matrix Functions and Applications, Part I", IEEE Spectrum, March, 1964.

[11] Frame, J.S., "Matrix Functions and Applications, Part II", IEEE Spectrum, April, 1964.

[12] Wilkinson, J.H., The Algebraic Eigenvalue Problem, Clarendon Press, London, 1965.

[13] Dunford, N. and Schwartz, Linear Operators, Part I, II and III, John Wiley & Sons, New York, 1971.

[14] Kato, T., Perturbation Theory for Linear Operators, Springer-Verlag, New York, 1976, 2nd Ed.

[15] Luenberger, D.G., Optimization by Vector Space Methods, John Wiley & Sons, Inc., New York, 1969.

[16] Rellich, F., Perturbation Theory of Eigenvalue Problems, Lecture Notes, New York University, 1953.

[17] Baumgärter, H., Englichdimensionale analytische Storungtheorie, Academie-Verlag, 1973.

[18] Knopp, K., Theory of Functions, Part II, Dover, New York, 1947 and 1975.

[19] Ben-Israel, A. and Greville, T.N.E., Generalized Inverses: Theory and Applications, John Wiley & Sons, New York, 1974.

[20] Haug, E.J. and Rousselet, B., "Design Sensitivity Analysis in Structural Mechanics: Eigenvalue Variations", Technical Report No. 52, NSF Project No. ENG 77-19967, Division of Material Engineering, College of Engineering, The University of Iowa, Iowa 52242.

[21] Choi, Kyung, Personal Conversation, May, 1980.

[22] Wilkinson, J.H. and Reinsch, C., Handbook for Automatic Computation, Vol. II, Linear Algebra, Springer-Verlag, New York, 1971.

[23] Forsythe, G.F. and Moler, C.B., Computer Solutions of Linear Algebraic Systems, Prentice-Hall, Inc.,

[24] Strang, G. and Fix, G.J., An Analysis of the Finite Element Method, Prentice Hall, Inc., Englewood Cliffs, N.J., 1973.

[25] Prenter, P.M., Splines and Variational Methods, John Wiley & Sons, New York, 1975.

[26] Bathe, K.J., Wilson, E.L., Numerical Methods in Finite Element Analysis, Prentice-Hall, Englewood Cliffs, N.J., 1976.

[27] Haug, E.J. and Rousselet, B., Personal Conversation, May, 1980.

[28] Meirovitch, L., Analytic Methods in Vibrations, MacMillan, New York, 1967.

[29] Lancaster, P. Lambda-Matrices and Vibrating Systems, Pergamon Press, New York, 1965.

[30] Caughey, T.K. and O'Kelley, M.E.S., "Classical Normal Modes in Damped Linear Dynamic Systems", Journal of Applied Mechanics, Vol. 32, 1965, pp. 583-588.

[31] Inman, D.J. and Andry, A.N., "Some Results on the Nature of Eigenvalues of Discrete Damped Linear Systems", Journal of Applied Mechanics, ASME, to appear.