

SOME RESULTS CONCERNING  
THE FUNDAMENTAL NATURE OF  
WYNN'S VECTOR EPSILON ALGORITHM

A Dissertation Submitted to the College of  
Graduate Studies and Research  
in Partial Fulfilment of the Requirements  
for the Degree of Doctor of Philosophy  
in the Department of Mechanical Engineering  
University of Saskatchewan  
Saskatoon

By

John Arthur Steele

Spring 2002

© Copyright John Arthur Steele, 2001. All rights reserved.

## **PERMISSION TO USE**

In presenting this thesis in partial fulfilment of the requirements for a postgraduate degree from the University of Saskatchewan, I agree that the Libraries of the University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor who supervised my thesis work or, in his absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or in part should be addressed to:

Head of the Department of Mechanical Engineering  
57 Campus Drive  
University of Saskatchewan  
Saskatoon, Saskatchewan S7N 5A9

## ABSTRACT

In this thesis, Wynn's Vector Epsilon Algorithm (VEA) is examined. Although the usefulness of this sequence-to-sequence transformation for inducing and enhancing convergence in vector sequences has been amply demonstrated by others, it is still not well understood. After reviewing some known important theoretical results for the VEA and its kernel (the full set of vector sequences which the VEA transforms to give a constant vector sequence), the author provides a sufficient and necessary condition for membership of a vector sequence in the real part of the kernel of the 1<sup>st</sup> order VEA. This kernel is shown to be the set of all real vector sequences  $\{\mathbf{x}_n\}$  converging toward, orbiting, or diverging away from some vector  $\mathbf{x}$  where each term of the error sequence  $\{\mathbf{x}_n - \mathbf{x}\}$  is a scaled and/or rotated version of the previous term of the error sequence, called  $\lambda\mathbf{R}$  sequences. This result is contrasted with one by McLeod and Graves-Morris. It is then shown that  $\lambda\mathbf{R}$  sequences may also be described as those sequences  $\{\mathbf{x}_n\}$  whose terms satisfy  $\mathbf{x}_n = \mathbf{x} + z^n \mathbf{w} + \bar{z}^n \bar{\mathbf{w}}$  where  $z \neq 0$ ,  $z \neq 1$ ,  $\|\mathbf{w}\| > 0$ , and  $\langle \mathbf{w}, \bar{\mathbf{w}} \rangle = 0$ .

Numerical experiments by the author on vector sequences generated by the formula  $\mathbf{x}_n = \mathbf{A}\mathbf{x}_{n-1} + \mathbf{b}$  are reported. Circumstances are found under which the VEA order of such sequences is lower than the upper bound given by Brezinski. The reduction is triggered by the presence of certain orthogonal relationships between eigenvector and generalised eigenvector components whose corresponding Jordan blocks in the Jordan canonical form of  $\mathbf{A}$  have complex conjugate eigenvalues. This empirical result anticipates the complex kernel of the 1<sup>st</sup> order VEA which is shown to be every sequence  $\{\mathbf{x}_n\}$  whose terms satisfy  $\mathbf{x}_n = \mathbf{x} + z^n \mathbf{w}_1 + \bar{z}^n \mathbf{w}_2$  with  $z \neq 0$ ,  $z \neq 1$ ,  $\|\mathbf{w}_1\| + \|\mathbf{w}_2\| > 0$ , and

$\langle \mathbf{w}_1, \mathbf{w}_2 \rangle = 0$  and no others. Some remaining open questions are noted in the final chapter.

## ACKNOWLEDGEMENTS

I wish to express my profound gratitude to Prof. Allan Dolovich for many things in the course of my engineering and research education. Beginning when I was an undergraduate, he validated my scepticism as an important value in the pursuit of knowledge. He has tempered my natural enthusiasm and arrogance by emphasising the distinction between conjecture and theorem, but has also encouraged me to verify or falsify my speculations. During that process, he has set a high standard for proof and for clarity, requiring me to face every loose or unclear thread in the argument. Most of all, he has willingly immersed himself in every one of those conjectures and joined me in the struggle to make them into theorems. For all of these and for encouraging me in my desire to make a contribution to the engineering sciences through numerical analysis, I shall always be grateful.

I also wish to thank the development engineers who designed the Hewlett Packard 48GX handheld computer. The ease with which it has allowed me to test ideas and experiment with numerical phenomena has played a major role in much of my work. Its built-in functions for linear algebra have been especially useful, but the power and elegance of its overall design have made every exploration a pleasure and greatly facilitated the development of code for experiments in both engineering and numerical analysis as well as other areas of inquiry.

Finally, I wish to thank my wife Julia Steele for reviving dreams I had lost and for her support as I undertook to finish what I had started.

## DEDICATION

*To my late father,*

*the Reverend Gerry Steele,*

*B. Eng. (Toronto), L.Th. (Wycliffe), B.S.T. (Trinity),*

*1926 - 1997*

*who loved linear algebra, too,*

*but was lost to me before I knew.*

# TABLE OF CONTENTS

Permission to Use .....	i
Abstract.....	ii
Acknowledgements .....	iv
Dedication.....	v
Table of Contents .....	vi
List of Tables.....	viii
List of Figures.....	ix
List of Theorems.....	xi
1. Introduction .....	1
2. Introduction to the VEA .....	6
2.1 The Shanks Transform .....	6
2.2 The Scalar Epsilon Algorithm .....	11
2.3 The Vector Epsilon Algorithm .....	16
3. The Real Kernel of the 1 <sup>st</sup> Order VEA .....	22
3.1 The Result.....	23
3.2 Proof of Necessity .....	24
3.3 Proof of Sufficiency.....	30
3.4 Comments on the Result.....	33
4. An Alternative Form for the Real Kernel of the 1 <sup>st</sup> Order VEA .....	35
4.1 An Intuitive Definition of a Real Scaled Rotation Sequence .....	36
4.1.1 A Real Scaled Rotation Sequence in Real Terms .....	37
4.1.2 A Real Scaled Rotation Sequence in Complex Terms .....	44
4.2 An Eigenvector Expansion for a Real Scaled Rotation Sequence.....	48
4.3 The Kernel of the 1st Order VEA, Real Part, Complex Form .....	57
5. Experimentation Toward the Kernel of the VEA.....	59
5.1 Background Material .....	61
5.2 Published Theoretical Results .....	67
5.3 The Initial Experimental Results .....	75
5.3.1 When $\mathbf{V}$ is Unitary .....	76
5.3.2 When $\mathbf{V}$ is Not Unitary .....	79
6. Orthogonality in $\mathbf{V}$ and VEA Order Reduction.....	81
6.1 Design of the Experiments .....	83
6.2 Test Results .....	89
6.2.1 $N = 2$ .....	89
6.2.2 $N = 4$ .....	90
6.2.3 $N = 6$ .....	91
6.2.4 $N = 8$ .....	94
6.3 Summary and Generalising Conjecture.....	96

6.4	Consequences of the Result.....	97
6.4.1	Hypersonic Flow Calculations .....	97
6.4.2	Consequences for the 1 <sup>st</sup> Order VEA Kernel.....	99
7.	The Complex Kernel for the 1 <sup>st</sup> Order VEA.....	102
7.1	Three Distinct Vectors Define a $zw$ Sequence .....	105
7.2	The Complex Kernel of the 1 <sup>st</sup> Order VEA.....	110
7.2.1	Proof that Being $zw$ is Sufficient.....	111
7.2.2	Proof that Being $zw$ is Necessary .....	116
8.	Future Work.....	121
8.1	Orthogonality & Multiple Complex Conjugate Jordan Blocks.....	121
8.2	Overshadowing and Linear Independence .....	121
8.3	The Generality of an Empirical Description of the 2 <sup>nd</sup> order VEA.....	122
8.4	Applications with VEA Order-Reducing Orthogonalities .....	123
	References .....	124
	Appendix A – A Derivation of the Kernel of the Shanks Transform.....	127
	Appendix B – The Jordan Canonical Form.....	131
	Appendix C – The Minimal Polynomial .....	136
C.1	The Characteristic Polynomial .....	136
C.2	The Minimal Polynomial of a Matrix.....	143
C.3	The Minimal Polynomial of a Matrix for a Vector .....	144
	Appendix D – Gekeler’s Proof of Theorem 5.1 [12], p. 68.....	152



**LIST OF TABLES**

2.1 A comparison of Shanks' transform sequences..... 8  
2.2 The scalar  $\varepsilon$ - table ..... 13  
2.3 The scalar  $\varepsilon$ -table for the sequence given by equation (2.3) ..... 14  
3.1 First order  $\varepsilon$ -table for arbitrary real  $\mathbf{x}_0$ ,  $\mathbf{x}_1$ , and  $\mathbf{x}_2$ . ..... 25  
6.1 The order used for orthogonality relations in this chapter..... 86  
7.1 The  $\varepsilon$ -table used for the complex kernel proof..... 112  
C.1 The effect of non-zero  $c_{ij}$  on  $K_1$ ,  $K_2$ , and  $K$  of equation (C.53)..... 151

## LIST OF FIGURES

3.1	Two sets of three ordered pairs with their VEA limits.....	22
3.2	VEA-1 applied to 3 arbitrary vectors.....	26
3.3	The heads of a vector sequence such that VEA-1 gives the vector $\mathbf{x}$ for every set of three consecutive vectors. ....	28
3.4	The heads of the vectors in an arbitrary $\lambda\mathbf{R}$ sequence $\{\mathbf{w}_n\}$ with the head of the limit (antilimit) vector $\mathbf{w}$ .....	31
3.5	A possible difference between $\mathbf{w}$ (having endpoint $Q$ ) and $\varepsilon_2^{(0)}$ (having endpoint $Q^{(0)}$ ). ....	32
4.1	The orthonormal basis, $\mathbf{q}_1$ and $\mathbf{q}_2$ , for the plane of rotation from $\mathbf{v}_{bef}$ and $\mathbf{v}_{aft}$ . ....	38
4.2	The action of $\mathbf{R}$ on vectors $\mathbf{q}_1$ and $\mathbf{q}_2$ . ....	48
6.1	The o-diagram for $\mathbf{V}$ with 2 vectors in each column block, and no 2 vectors orthogonal. ....	85
6.2	An ordering of the relationships in an o-diagram for $\mathbf{V}$ consisting of 2 column blocks with 2 vectors each. ....	87
6.3	O-diagram for unitary $\mathbf{V}$ consisting of 2 column blocks with 2 vectors each. ....	87
6.4	O-diagram for $\mathbf{V}$ consisting of 2 column blocks with 2 vectors each, with orthogonal relationships 001,10,1.....	87
6.5	O-diagram for $\mathbf{V}$ consisting of 2 column blocks, 3 vectors each, with orthogonality 10100,1000,001,10,1.....	88
6.6	O-diagram for $\mathbf{V}$ consisting of 2 column blocks, 4 vectors each, with orthogonality 0000001,000001,00001,0001,001,01,1.....	88
6.7	O-diagram for $\mathbf{V}$ having 2 column blocks with 1 vector each needed for a reduction of 1 VEA order below Brezinski's upper bound. ....	89
6.8	The 64 o-diagrams for $\mathbf{V}$ having 2 column blocks with 2 vectors each, grouped by order, 4 <sup>th</sup> order not framed, 3 <sup>rd</sup> order framed by dotted lines, and 2 <sup>nd</sup> order framed by solid lines. ....	90
6.9	O-diagrams for 9 symmetrically grouped sets of orthogonal relationships for $\mathbf{V}$ having 2 column blocks with 3 vectors each. ....	92
6.10	a. The maximal 6 <sup>th</sup> order o-diagram, and b. the minimal 5 <sup>th</sup> order o-diagram for $\mathbf{V}$ having 2 column blocks with 3 vectors each. ....	92
6.11	a. The minimal 5 <sup>th</sup> order o-diagram, b. the maximal 5 <sup>th</sup> order o-diagram, and c. the minimal 4 <sup>th</sup> order o-diagram for $\mathbf{V}$ having 2 column blocks with 3 vectors each. ....	93
6.12	a.,b. Two maximal 4 <sup>th</sup> order o-diagrams and c. the minimal 3 <sup>rd</sup> order o-diagram for $\mathbf{V}$ having 2 column blocks with 3 vectors each. ....	94
6.13	a. O-diagram for an expected reduction of 3 VEA orders and b.-g. six symmetric orthogonality sets to add to it in search of a 4 <sup>th</sup> VEA order reduction for $\mathbf{V}$ having 2 column blocks with 4 vectors each. ....	95
6.14	a. Two maximal 5 <sup>th</sup> order o-diagrams minus possible same-block relationships and b. the minimal 4 <sup>th</sup> order o-diagram for $\mathbf{V}$ having 2 column blocks with 4 vectors each. ....	96

6.15	The vertical format minimal 4 <sup>th</sup> order o-diagram with central open circles for $V$ having 2 column blocks with 4 vectors each. ....	97
6.16	Some of the eigenvalues of $A$ for inviscid hypersonic flow calculations from Cheung et al [9].....	98
6.17	Some of the eigenvalues of $A$ for viscous hypersonic flow calculations from Cheung et al [9].....	98

## LIST OF THEOREMS

2.1	The implicit kernel of the Shanks transform .....	9
2.2	The explicit kernel of the Shanks transform.....	10
2.3	Translativity of the SEA (Shanks transform) .....	13
2.4	Homogeneity of the SEA (Shanks transform).....	15
2.5	Translativity, homogeneity and unitary transformativity of the VEA.....	18
2.6	The sufficient but not necessary condition for membership in the kernel of the VEA by McLeod and Graves-Morris .....	19
3.1	The real vector sequences of the kernel of the 1 <sup>st</sup> order VEA.....	23
5.1	Gekeler's theorem showing that sequences generated by $\mathbf{x}_n = \mathbf{A}\mathbf{x}_{n-1} + \mathbf{b}$ with $\mathbf{I} - \mathbf{A}$ nonsingular are in the kernel of the VEA.....	68
5.2	Brezinski's refinement of Gekeler's theorem giving an upper bound on the VEA order as generalised null eigencomponents vanish from the error sequence.....	72
5.3	Brezinski's theorem giving an upper bound for the VEA order of some sequences generated by $\mathbf{x}_n = \mathbf{A}\mathbf{x}_{n-1} + \mathbf{b}$ with $\mathbf{I} - \mathbf{A}$ singular .....	73
7.1	The complex kernel of the 1 <sup>st</sup> order VEA.....	110

## 1. INTRODUCTION

The heart of this thesis is a search for insights into the fundamental nature of Wynn's Vector Epsilon Algorithm (VEA). The algorithm is a non-linear sequence-to-sequence transformation whose usefulness lies in its ability to transform many slowly convergent sequences of vectors into sequences that converge much more quickly to the same limit. It is also able to transform many sequences that diverge into ones that actually converge to the antilimit vector of the original sequence, the unique vector away from which the original sequence diverges.

Different orders of VEA transform sequence may be computed, each associated with a positive integer. Each vector in the 1<sup>st</sup> order VEA transform sequence is computed from three consecutive vectors from the original sequence. The formula for computing the VEA is a recursive one; in the process of computing a  $k^{\text{th}}$  order VEA transform sequence, portions of the 1<sup>st</sup> through  $(k - 1)^{\text{th}}$  order VEA transform sequences must first be computed. Typically, when the algorithm works at all, each order of VEA produces a transform sequence with better convergence properties than that of the previous order.

The VEA has been profitably applied to a variety of iterative methods which generate sequences of approximations to the desired solution vector, but having a rate of convergence which may not be satisfactory. Wynn [32] has shown this algorithm to be very effective and economical in accelerating the convergence of a wide variety of vector sequences. The VEA has been successfully used for convergence acceleration when iteratively solving non-singular [12] and singular [3,25] systems of linear equa-

tions and systems of non-linear equations [2, 12] as well. Brezinski [5] has given an algorithm for efficiently computing several eigenvalues and eigenvectors simultaneously using the VEA and an extension of the Rayleigh quotient method.

Successful engineering applications include the following. After Hafez and Cheng [14] and Hafez *et al* [15] used the SEA to quite significantly reduce convergence times in iterative transonic flow calculations with Richardson-type iteration, Cheung *et al* [9] did the same thing with hypersonic flow calculations using the VEA, achieving a 50% reduction in computation time. (See §6.4.1 for comment on this work.) Tan [29] showed how to obtain exact values for the partial derivatives of the eigenvalues and eigenvectors of parameter dependent matrices by using a small number of steps of the VEA on a slowly convergent iterative method. Tan and Andrew [30] have used the VEA and related algorithms to do the same for an iterative method by Andrew. The eigenvalues and eigenvectors of parameter dependent matrices are crucial to the fields of design optimisation, finite element solution sensitivity analysis and others. More recently, Dolovich and Brodland [11] have demonstrated the value of the VEA in significantly accelerating the convergence of iterative finite element methods, and Lowe [17] has shown that the VEA is very effective in accelerating the convergence of a variety of algebraic reconstruction algorithms in computerised tomography.

Despite abundant demonstrations of its capabilities and uses during the past 39 years, the VEA is still not well understood. The central mystery concerning this algorithm has been its kernel. The kernel is the set of all sequences that the VEA transforms to give a constant vector sequence, i.e. a vector sequence where every vector of the sequence is identical. This can be thought of as a sequence which, more than just

converging to the limit, has actually *already converged* to the (anti)limit vector. Prior to 2000, only subsets of the kernel of the VEA had been published. With Prof. Allan Dolovich, the author [27] has given the real part of the kernel for the 1<sup>st</sup> order VEA. In this thesis, the entire complex kernel for the 1<sup>st</sup> order VEA is provided as well. It is the author's hope that once its nature is understood fully, the algorithm will find use in those engineering and mathematical applications to which is ideally suited.

The thesis progresses as follows. Chapter 2 gives an introduction to the Vector Epsilon Algorithm including a brief history with some of the theoretical results pertaining to the VEA relevant to the work presented here. This will include the translatability, homogeneity, and unitary transformativity of the VEA. Also reviewed is some of what is known regarding the kernel of the VEA. The 1<sup>st</sup> fundamental result regarding the kernel was essentially a sufficient but not necessary condition for membership in the kernel published in 1971 for the real case by McLeod [18] and in 1983 for the complex case by Graves-Morris [13]. Other more recent results are also touched on.

Chapter 3 gives the proof for the real part of the kernel of the 1<sup>st</sup> order VEA as already published by this author [27]. It is shown that it is the set of all real vector sequences  $\{\mathbf{x}_n\}$  converging toward, orbiting, or diverging away from some vector  $\mathbf{x}$  for which each term of the error sequence  $\{\mathbf{x}_n - \mathbf{x}\}$  is a scaled and/or rotated version of its predecessor in the error sequence. The comments at the end of the chapter distinguish this result from the 1<sup>st</sup> order case of the McLeod – Graves-Morris sufficient condition.

Chapter 4 reworks the real kernel of Chapter 3 using two different derivations that show that the set of  $\lambda\mathbf{R}$  sequences (and the real part of the kernel of the 1<sup>st</sup> order

VEA) is identical to the set of sequences  $\{\mathbf{x}_n\}$  for which  $\mathbf{x}_n = \mathbf{x} + z^n \mathbf{w} + \bar{z}^n \bar{\mathbf{w}}$ ,  $n = 0, 1, \dots$  where  $z$  is complex,  $z \neq 0$ ,  $z \neq 1$ ,  $\|\mathbf{w}\| > 0$ , and  $\langle \mathbf{w}, \bar{\mathbf{w}} \rangle = 0$ .

Chapter 5 explores vector sequences  $\{\mathbf{x}_n\}$  all of whose terms satisfy  $\mathbf{x}_n = \mathbf{A}\mathbf{x}_{n-1} + \mathbf{b}$  where  $\mathbf{A}$  is a square matrix and  $\mathbf{b}$  is a vector. First, theoretical results by Gekeler and Brezinski based on the McLeod – Graves-Morris condition are reviewed and discussed. These results culminate in an upper bound by Brezinski on the order of VEA needed to give the (anti)limit of  $\{\mathbf{x}_n\}$ . Following this, a set of numerical experiments by the author are reported. These experiments were done using the above generating formula in order to search for circumstances under which an order of VEA lower than Brezinski's upper bound yields  $\mathbf{x}$ . The author has found no description of such circumstances in the literature. A reduction in VEA order below Brezinski's upper bound was found to occur whenever mutually orthogonal eigenvector and generalised eigenvector components were present in  $\mathbf{x}_n - \mathbf{x}$  whose corresponding Jordan blocks in the corresponding Jordan canonical form of  $\mathbf{A}$  have complex conjugate eigenvalues. In some cases, the VEA order of the resulting sequence was reduced by as much as half.

Chapter 6 determines exactly which orthogonal relationships between these components are critical to this reduction in VEA order. It describes further experiments on iteration matrices  $\mathbf{A}$  having two same-sized complex-conjugate Jordan blocks and varying combinations of orthogonality relations between the associated eigenvectors and generalised eigenvectors. These experiments reveal that the only orthogonal relationships instrumental in a reduction in the VEA order below Brezinski's upper bound are between vectors of differing column blocks and not between any two vectors



in the same column block. The results are then summarised and a generalising conjecture is put forth which suggests a possible explanation for the degree of success enjoyed by Cheung *et al* [9] in their application of the VEA to hypersonic flow calculations. The conclusion also anticipates the result of the next chapter.

Chapter 7 defines a  $zw$  sequence which is a generalisation of the reformulation in Chapter 4 of the  $\lambda R$  sequences introduced in Chapter 3. It is shown that every  $zw$  sequence is defined by three distinct vectors: two initial vectors  $\mathbf{x}_0$  and  $\mathbf{x}_1$  and a(n) (anti)limit vector  $\mathbf{x}$ . A method for finding the formula for  $\mathbf{x}_n$ ,  $n = 0, 1, \dots$  for every  $zw$  sequence from  $\mathbf{x}_0$ ,  $\mathbf{x}_1$  and  $\mathbf{x}$  is also provided. Then it is shown that the kernel of the 1<sup>st</sup> order VEA is the set containing every  $zw$  sequence. Finally, Chapter 8 offers some open questions to be answered in subsequent investigations.

It has been the author's intention to make this work accessible to readers with a background in engineering and not necessarily in applied mathematics. To that end, material has been included in both the main body and the appendices to provide useful background in some of the areas into which this dissertation ventures. Terminology used but not defined in the main text is usually defined in an appendix referred to prior to the first use of the term. Appendix C includes some original interpretation of minimal polynomials.

## 2. INTRODUCTION TO THE VEA

The Vector Epsilon Algorithm (VEA) was developed by Wynn [32] in 1962. It originated through a modification to the Scalar Epsilon Algorithm (SEA) discovered by Wynn [31] in 1956. The SEA, in turn, was a more computationally efficient way to calculate the Shanks transform [24, 23] which does for many scalar sequences what the VEA does for many vector sequences. This chapter, therefore, begins with the Shanks transform. Section 2.2 presents a brief look at the SEA. In §2.3, the VEA is introduced.

### 2.1 The Shanks Transform

The Shanks transform is a useful sequence-to-sequence transformation for inducing or accelerating convergence in scalar sequences. It is named for Daniel Shanks who published it in 1955 [24] although the same algorithm was described by Schmidt [23] and O'Beirne [19] previously. (See Brezinski [8, p. 79] for details.) From a convergent scalar sequence, the transform produces other scalar sequences that converge to the same limit. Depending on the nature of the original sequence, these transform sequences often converge much more quickly than the original sequence. When the original sequence diverges, the transform sequences sometimes converge to the scalar from which the original sequence diverges (the antilimit of the original sequence). Failing this, the transform sequences diverge from the same antilimit as the original sequence, if such an antilimit exists, and this divergence is often slower.

From a given sequence, several different transform sequences may be found depending on the order of transform applied. The calculation of a single term of a  $k^{\text{th}}$  order Shanks transform sequence requires  $2k + 1$  terms of the original sequence and the evaluation of two determinants of order  $k + 1$ . The 1<sup>st</sup> order Shanks transform is equivalent to Aitken's method [1].

Given a scalar sequence  $\{S_n\}$ , the terms of the  $k^{\text{th}}$  order Shanks transform sequence  $\{e_k(S_n)\}$  are defined by

$$e_k(S_n) = \frac{\begin{vmatrix} S_n & S_{n+1} & S_{n+2} & \cdots & S_{n+k} \\ S_{n+1} - S_n & S_{n+2} - S_{n+1} & S_{n+3} - S_{n+2} & \cdots & S_{n+k+1} - S_{n+k} \\ S_{n+2} - S_{n+1} & S_{n+3} - S_{n+2} & S_{n+4} - S_{n+3} & \cdots & S_{n+k+2} - S_{n+k+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ S_{n+k} - S_{n+k-1} & S_{n+k+1} - S_{n+k} & S_{n+k+2} - S_{n+k+1} & \cdots & S_{n+2k} - S_{n+2k-1} \end{vmatrix}}{\begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ S_{n+1} - S_n & S_{n+2} - S_{n+1} & S_{n+3} - S_{n+2} & \cdots & S_{n+k+1} - S_{n+k} \\ S_{n+2} - S_{n+1} & S_{n+3} - S_{n+2} & S_{n+4} - S_{n+3} & \cdots & S_{n+k+2} - S_{n+k+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ S_{n+k} - S_{n+k-1} & S_{n+k+1} - S_{n+k} & S_{n+k+2} - S_{n+k+1} & \cdots & S_{n+2k} - S_{n+2k-1} \end{vmatrix}}. \quad (2.1)$$

Notice that the numerator and denominator determinants are identical except for their first rows. Therefore, if each is expanded by the first row, the cofactors from the first may be reused in calculating the second reducing the real burden to little more than that of a single determinant of order  $k + 1$ .

The transient, which is so common in physical and numerical processes, is among the sequences for which the Shanks transform will yield a sequence all of whose terms are the limit or antilimit of the original sequence. An  $m^{\text{th}}$  order mathematical transient is a sequence of the form

$$S_n = S + \sum_{i=1}^m c_i r_i^n, \quad (2.2)$$

with  $r_i \neq 1$ , for  $i = 1, \dots, m$ . As an example, consider the second order transient sequence  $\{S_n\}$  whose generating formula is

$$S_n = 3 + 8\left(\frac{1}{2}\right)^n + 27\left(\frac{1}{3}\right)^n, \quad (2.3)$$

and whose values along with those of the first two Shanks transform sequences appear in Table 2.1, below. (All values are exact unless truncation is indicated by an ellipsis.) Note that each term of each transform sequence appears to the right of the latest term of the original sequence used in its computation. This arrangement makes apparent the improvement in the estimate of the limit of the original sequence due to the Shanks transform.

Table 2.1 - A comparison of Shanks' transform sequences

$S_n$	$e_1(S_n)$	$e_2(S_n)$
$S_0 = 38$		
$S_1 = 16$		
$S_2 = 8$	$e_1(S_0) = 3.42857\dots$	
$S_3 = 5$	$e_1(S_1) = 3.2$	
$S_4 = 3.833\dots$	$e_1(S_2) = 3.0909\dots$	$e_2(S_0) = 3$
$S_5 = 3.3611\dots$	$e_1(S_3) = 3.04$	$e_2(S_1) = 3$

Though  $S_2$  is 5 away from the sequence limit of 3,  $e_1(S_0)$  reduces this distance to only 0.42857..., eliminating more than 91% of the error. Similarly,  $e_1(S_3)$  removes just under 89% of the error present in  $S_5$ . Notice, however, that  $e_2(S_n)$  is a constant sequence which removes all of the transient components in  $S_n$ . This is because when  $k$ , the order of Shanks' transform in equation (2.1) equals  $m$ , the order of the transient defined by

equation (2.2), then  $e_k(S_n) = S$  of equation (2.2) for all  $n$ . The transient sequence defined by equation (2.3) is said to be part of the kernel of the second order Shanks transform, the set of all sequences that the second order algorithm transforms to a constant sequence. The terms of this constant sequence will be the limit or antilimit of the original sequence.

One form of the kernel of the transform was provided by Wynn [34], who proved

Theorem 2.1: If and only if, for some sequence  $\{S_n\}$ , there exist scalars  $S$  and  $a_i$ ,  $i = 0, 1, \dots, k$  such that  $\sum a_i \neq 0$ ,  $a_k \neq 0$  and

$$\sum_{i=0}^k a_i (S_{n+i} - S) = 0, \quad n = 0, 1, \dots, \quad (2.4)$$

then  $e_k(S_n) = S$ .

An illuminating derivation of this expression found in [4] is given in Appendix A.

If equation (2.4) is written in expanded form to give

$$a_0(S_n - S) + a_1(S_{n+1} - S) + \dots + a_{k-1}(S_{n+k-1} - S) + a_k(S_{n+k} - S) = 0, \quad (2.5)$$

since  $a_k \neq 0$ , it follows that

$$(S_{n+k} - S) = b_0(S_n - S) + b_1(S_{n+1} - S) + \dots + b_{k-1}(S_{n+k-1} - S) \quad (2.6)$$

where

$$b_i = \frac{-a_i}{a_k}. \quad (2.7)$$

If the related sequence  $\{S_n - S\}$  is called the error sequence associated with  $\{S_n\}$ , then it may be said that the kernel of  $e_k(S_n)$  is every scalar sequence having the property that each term of its error sequence is a linear combination of the previous  $k$  terms of its error sequence with the  $k$  coefficients  $b_i$  not totalling unity.

This kernel is called the implicit kernel for the Shanks transform because  $S_n$  in equation (2.4) is not isolated on one side of the equals sign. In 1970, Brezinski and Crouzeix [7] offered the following explicit form for the kernel of the Shanks transform using the classic solution for the difference equation (2.4).

**Theorem 2.2:** The quantity  $e_k(S_n) = S$  if and only if each term of  $\{S_n\}$  has the generating formula

$$S_n = S + \sum_{i=1}^p A_i(n)r_i^n + \sum_{i=p+1}^q [B_i(n)\cos(b_i n) + C_i(n)\sin(b_i n)]e^{w_i n} + \sum_{i=0}^m c_i \delta_{in} \quad (2.8)$$

where  $r_1, r_2, \dots, r_p \neq 1$ ,  $A_i, B_i$ , and  $C_i$  are polynomials in  $n$ ,  $c_i$  is a scalar for  $i = 0, \dots, m$ , and  $\delta_{in}$  is a Kronecker delta. The relation governing  $k$ , the order of Shanks' transform necessary to give  $S$ , is

$$m + \sum_{i=1}^p d_i + 2 \sum_{i=p+1}^q d_i = k - 1 \quad (2.9)$$

where  $d_i$  is equal to one plus the degree of  $A_i$  for  $i = 1, \dots, p$ , and equal to one plus the maximum of the degrees of  $B_i$  and  $C_i$  for  $i = p+1, \dots, q$ . By convention,  $m = -1$  if no  $c_i$  exist.

This somewhat daunting formulation does yield some useful insights upon closer examination. In equation (2.8), each summation represents one type of component in  $\{S_n - S\}$ . Apart from the polynomial factors, the summation from 1 to  $p$  in-

cludes geometrically varying components (which would include those of the transient from the previous example) while the summation from  $p + 1$  to  $q$  includes sinusoidal components with an exponential factor. The Kronecker delta in the final summation allows for the presence of arbitrary error  $c_n$  in  $S_n$  for  $n = 0, \dots, m$ . Equation (2.9) indicates how much each type of component influences the order of the Shanks transform required to obtain  $S$ . In particular, it shows that sinusoidally varying components in  $\{S_n\}$  are twice as taxing on the Shanks transform as geometrically varying components.

## 2.2 The Scalar Epsilon Algorithm

The Scalar Epsilon Algorithm or SEA was developed by Wynn [31] in 1956. It yields  $e_k(S_n)$  without the burdensome evaluation of determinants. However, each order of SEA first requires the calculation of terms of an intermediate sequence having little known practical importance. Because of this, a new notation is used in which each term of each sequence is described by the symbol  $\varepsilon$  with a subscript that denotes the sequence to which the term belongs and a superscript that denotes the index of the term within the sequence. It is really only the even numbered sequences  $\varepsilon_{2k}^{(n)}$ ,  $k = 0, 1, \dots$ , that are of interest which, where they exist, are equal to  $e_k(S_n)$ . Therefore, the sequence  $\{\varepsilon_{2k}^{(n)}\}$  will be called the  $k^{\text{th}}$  order transform sequence. Thus, the 1<sup>st</sup> order transform sequence is  $\{\varepsilon_2^{(n)}\}$ .

The SEA is defined by the following expressions.

$$\varepsilon_{-1}^{(n)} = 0, \quad \varepsilon_0^{(n)} = S_n, \quad n = 0, 1, \dots \quad (2.10)$$

$$\varepsilon_{j+1}^{(n)} = \varepsilon_{j-1}^{(n+1)} + \frac{1}{\varepsilon_j^{(n+1)} - \varepsilon_j^{(n)}} \quad j, n = 0, 1, \dots \quad (2.11)$$

As seen in equation (2.11), the terms of each new sequence are calculated by finding the differences between consecutive terms of the previous sequence, inverting the difference, and adding it to a term from a sequence previous to that. As long as consecutive terms of a sequence remain distinct, the differences will be invertible and the algorithm will give the same results as the Shanks transform, excluding machine roundoff effects. If some such terms are not distinct, the algorithm encounters division by zero, and the term being calculated may not be defined by equation (2.11). In some cases, such singularities can be avoided and stability enhanced through modifications to the basic algorithm. Wynn [33] has developed rules for jumping over such singularities called particular rules and Cordellier [10] has generalised them for consecutive singularities in adjacent transform sequences.

The calculation scheme of equations (2.10) and (2.11) lends itself well to a tabular arrangement. It is usually represented as shown in Table 2.2.



Table 2.2 - The scalar  $\varepsilon$ -table

	$\varepsilon_0^{(0)} = S_0$				
$\varepsilon_{-1}^{(1)} = 0$		$\varepsilon_1^{(0)}$			
	$\varepsilon_0^{(1)} = S_1$		$\varepsilon_2^{(0)}$		
$\varepsilon_{-1}^{(2)} = 0$		$\varepsilon_1^{(1)}$		$\varepsilon_3^{(0)}$	
	$\varepsilon_0^{(2)} = S_2$		$\varepsilon_2^{(1)}$		$\varepsilon_4^{(0)}$
$\varepsilon_{-1}^{(3)} = 0$		$\varepsilon_1^{(2)}$		$\varepsilon_3^{(1)}$	...
	$\varepsilon_0^{(3)} = S_3$		$\varepsilon_2^{(2)}$		$\varepsilon_4^{(1)}$
$\varepsilon_{-1}^{(4)} = 0$		$\varepsilon_1^{(3)}$		$\varepsilon_3^{(2)}$	...
	$\varepsilon_0^{(4)} = S_4$		$\varepsilon_2^{(3)}$		
$\vdots$		$\vdots$		$\vdots$	

In this table, notice that the sequences involved are represented as columns. Beginning with  $\{\varepsilon_{-1}^{(n)}\}$  and  $\{\varepsilon_0^{(n)}\}$  of equations (2.10), subsequent columns are calculated from left to right. The utility of this arrangement of terms is seen with a sample calculation of equation (2.11). Notice that each new term in the  $\varepsilon$ -table is defined by the three terms which are closest to it on its left.

The second order transient already considered in §2.1 is now revisited. Table 2.3 gives the  $\varepsilon$ -table for that example. (Again, all values are exact unless truncation is indicated by ellipsis.) By inspection of Tables 2.1 and 2.3, it is apparent that  $\{\varepsilon_{2k}^{(n)}\}$  is the same as  $\{e_k(S_n)\}$  for both  $k = 1$  and  $k = 2$ .

Due to the simplicity of the SEA, some useful and already known results are easily shown. For example,

Theorem 2.3: Given some sequence  $\{S_n\}$  to which the  $k^{\text{th}}$  order SEA is applied resulting in

Table 2.3 - The scalar  $\varepsilon$ -table for the sequence given by equation (2.3)

$\varepsilon_{-1}^{(1)} = 0$	$\varepsilon_0^{(0)} = 38$	$\varepsilon_1^{(0)} = -0.04545\dots$	$\varepsilon_2^{(0)} = 3.42857\dots$	$\varepsilon_3^{(0)} = -4.5$	$\varepsilon_4^{(0)} = 3$
$\varepsilon_{-1}^{(2)} = 0$	$\varepsilon_0^{(1)} = 16$	$\varepsilon_1^{(1)} = -0.125$	$\varepsilon_2^{(1)} = 3.2$	$\varepsilon_3^{(1)} = -9.5$	$\varepsilon_4^{(1)} = 3$
$\varepsilon_{-1}^{(3)} = 0$	$\varepsilon_0^{(2)} = 8$	$\varepsilon_1^{(2)} = -0.333\dots$	$\varepsilon_2^{(2)} = 3.0909\dots$	$\varepsilon_3^{(2)} = -20.5$	$\varepsilon_4^{(2)} = 3$
$\varepsilon_{-1}^{(4)} = 0$	$\varepsilon_0^{(3)} = 5$	$\varepsilon_1^{(3)} = -0.85714\dots$	$\varepsilon_2^{(3)} = 3.04$	$\vdots$	$\vdots$
$\varepsilon_{-1}^{(5)} = 0$	$\varepsilon_0^{(4)} = 3.833\dots$	$\varepsilon_1^{(4)} = -2.11764\dots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\varepsilon_0^{(5)} = 3.3611\dots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

---


$$\varepsilon_{2k-1}^{(n)} = p_n, \quad n = 0, 1, \dots \quad (2.12)$$

and

$$\varepsilon_{2k}^{(n)} = q_n, \quad n = 0, 1, \dots, \quad (2.13)$$

if a new sequence  $\{T_n\}$  is created where

$$T_n = S_n + c, \quad n = 0, 1, \dots \quad (2.14)$$

with  $c$  an arbitrary complex scalar and then the  $k^{\text{th}}$  order SEA is applied to  $\{T_n\}$ , then

$$\varepsilon_{2k-1}^{(n)} = p_n, \quad n = 0, 1, \dots, \quad (2.15)$$

and

$$\varepsilon_{2k}^{(n)} = q_n + c, \quad n = 0, 1, \dots \quad (2.16)$$

The SEA, and thus the Shanks transform, is therefore translative. In addition, there is

Theorem 2.4: Given some sequence  $\{S_n\}$  to which the  $k^{\text{th}}$  order SEA is applied resulting in

$$\varepsilon_{2k-1}^{(n)} = p_n, \quad n = 0, 1, \dots \quad (2.17)$$

and

$$\varepsilon_{2k}^{(n)} = q_n, \quad n = 0, 1, \dots, \quad (2.18)$$

if  $\{T_n\}$  is defined such that

$$T_n = cS_n, \quad n = 0, 1, \dots \quad (2.19)$$

with  $c$  an arbitrary complex scalar and then the  $k^{\text{th}}$  order SEA is applied to  $\{T_n\}$ , then

$$\varepsilon_{2k-1}^{(n)} = P_n / c, \quad n = 0, 1, \dots, \quad (2.20)$$

and

$$\varepsilon_{2k}^{(n)} = cq_n, \quad n = 0, 1, \dots \quad (2.21)$$

The SEA is thus homogeneous of degree one or, simply, homogeneous. For a given order, if  $\{S_n\}$  transforms to give  $\{q_n\}$ , then  $\{cS_n + d\}$  transforms to give  $\{cq_n + d\}$ . These properties are sometimes expressed by saying that the SEA (as well as the Shanks transform) is quasilinear.

Many other results are known for the SEA including a whole body of convergence results which do not directly bear on the work reported here. For an overview

and bibliography of these and other results, the interested reader is referred to Brezinski and Redivo-Zaglia [8].

### 2.3 The Vector Epsilon Algorithm

The SEA was the first important non-linear sequence-to-sequence transformation. The economy with which it obtained any order of Shanks' transform enabled much more vigorous empirical exploration of its applicability to accelerate the convergence of scalar sequences. It could also be applied to sequences of  $N$ -dimensional vectors by treating each vector sequence as though it were a set of  $N$  independent scalar sequences. However, whenever consecutive vectors had one or more identical coordinates, the SEA was prevented from giving results. (This was before Cordellier developed so called particular rules [10,33] which often enable the calculation of terms of the  $\varepsilon$ -table on all sides of a region of singularity.) In addition, what was happening in the transformation of one sequence of co-ordinates in no way informed what was happening in other co-ordinate sequence transformations. These limitations motivated the search for a new way to deal with vector sequences as a whole.

The idea pursued by Wynn was to modify the definition of the SEA as little as possible in order to enable it to handle vector sequences. The only real problem with equations (2.10) and (2.11) was the denominator in (2.11). How does one divide unity by a vector? This problem suggested the need to define a vector inverse. Wynn's proposal was what he called the Samelson inverse, named after the colleague who suggested it to him. This inverse is defined by

$$\mathbf{x}^{-1} = \frac{\bar{\mathbf{x}}}{\langle \mathbf{x}, \mathbf{x} \rangle} \quad (2.22)$$

where the overbar denotes the complex conjugate and the inner product in the denominator is

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^N x_i \bar{y}_i \quad (2.23)$$

and  $N$  is the dimension of  $\mathbf{x}$ . This inverse has the property expected of an inverse, which is that

$$\left(\mathbf{x}^{-1}\right)^{-1} = \frac{\left(\frac{\bar{\mathbf{x}}}{\langle \mathbf{x}, \mathbf{x} \rangle}\right)}{\left\langle \frac{\bar{\mathbf{x}}}{\langle \mathbf{x}, \mathbf{x} \rangle}, \frac{\bar{\mathbf{x}}}{\langle \mathbf{x}, \mathbf{x} \rangle} \right\rangle} = \frac{\frac{\mathbf{x}}{\langle \mathbf{x}, \mathbf{x} \rangle}}{\frac{\langle \bar{\mathbf{x}}, \bar{\mathbf{x}} \rangle}{\langle \mathbf{x}, \mathbf{x} \rangle^2}} = \frac{\mathbf{x} \langle \mathbf{x}, \mathbf{x} \rangle}{\langle \bar{\mathbf{x}}, \bar{\mathbf{x}} \rangle} = \mathbf{x}. \quad (2.24)$$

With this modification, Wynn created the Vector Epsilon Algorithm [32] which is defined by

$$\boldsymbol{\varepsilon}_{-1}^{(n)} = \mathbf{0}, \quad \boldsymbol{\varepsilon}_0^{(n)} = \mathbf{x}_n, \quad n = 0, 1, \dots \quad (2.25)$$

and

$$\boldsymbol{\varepsilon}_{j+1}^{(n)} = \boldsymbol{\varepsilon}_{j-1}^{(n+1)} + \left[ \boldsymbol{\varepsilon}_j^{(n+1)} - \boldsymbol{\varepsilon}_j^{(n)} \right]^{-1} \quad j, n = 0, 1, \dots \quad (2.26)$$

where the inverse is defined by equations (2.22) and (2.23).

Despite its origin with a seemingly very small modification of the SEA, this one change effectively divorced the VEA from much of the abundant theory known for its

predecessor, the SEA. However, several theoretical results for the VEA analogous to some of those reported above for the SEA are known.

Consider the Samelson inverse. If  $\mathbf{x}$  is the one dimensional vector  $[x]$ , then  $\mathbf{x}^{-1}$  is the one-dimensional vector  $[1/x]$ . Therefore, for sequences made up of one-dimensional vectors, the VEA gives the same results as the SEA gives for the corresponding scalar sequences. Therefore, the VEA must be quasilinear in the same sense that the SEA was quasilinear. However, it goes further than this.

Theorem 2.5: Given a vector sequence  $\{\mathbf{x}_n\}$  which, when the  $k^{\text{th}}$  order VEA is applied to it, results in

$$\boldsymbol{\varepsilon}_{2k-1}^{(n)} = \mathbf{p}_n, \quad n = 0, 1, \dots \quad (2.27)$$

and

$$\boldsymbol{\varepsilon}_{2k}^{(n)} = \mathbf{q}_n, \quad n = 0, 1, \dots, \quad (2.28)$$

if a new sequence  $\{\mathbf{t}_n\}$  is created such that

$$\mathbf{t}_n = c\mathbf{U}\mathbf{x}_n + \mathbf{y}, \quad n = 0, 1, \dots, \quad (2.29)$$

where  $c$  is an arbitrary complex scalar,  $\mathbf{U}$  is a unitary matrix and  $\mathbf{y}$  is an arbitrary complex vector, then it can be shown that the  $k^{\text{th}}$  order VEA applied to  $\{\mathbf{t}_n\}$  results in

$$\boldsymbol{\varepsilon}_{2k-1}^{(n)} = 1/c \overline{\mathbf{U}} \mathbf{p}_n, \quad n = 0, 1, \dots \quad (2.30)$$

and

$$\boldsymbol{\varepsilon}_{2k}^{(n)} = c\mathbf{U}\mathbf{q}_n + \mathbf{y}, \quad n = 0, 1, \dots \quad (2.31)$$

The reader may recall that a unitary matrix is a complex matrix whose inverse is its complex conjugate-transpose. It may be thought of as the generalisation of an orthogonal matrix to a complex vector space. Just as the columns and rows of an orthogonal matrix are of unit length and mutually orthogonal, so are the columns and rows of a unitary matrix. The straight-forward proof of the results in (2.30) and (2.31) exploits the fact that, as with premultiplication by an orthogonal matrix in real vector spaces, premultiplication by a unitary matrix can be thought of as causing some combination of rotations and reflections in a complex vector space. Such transformations preserve inner products. As a consequence of this result, in whatever way one may wish to scale uniformly, translate, rotate and reflect the vectors of a sequence, the transform sequence produced by any order of the VEA will be similarly scaled uniformly, translated, rotated and reflected.

In 1971, McLeod [18] gave a result almost analogous to the one given previously for the Shanks transform with equation (2.4). In a difficult proof, he showed

Theorem 2.6: If, for some vector sequence  $\{\mathbf{x}_n\}$ , there exist real scalars  $a_i, i = 0, 1, \dots, k$  such that  $\sum a_i \neq 0, a_k \neq 0$  and vector  $\mathbf{x}$  such that

$$\sum_{i=0}^k a_i (\mathbf{x}_{n+i} - \mathbf{x}) = 0, \quad n = 0, 1, \dots, \quad (2.32)$$

then  $\varepsilon_{2k}^{(n)} = \mathbf{x}, n = 0, 1, \dots$

In 1983, Graves-Morris [13] showed that the same was true for complex  $a_i$  using a completely different proof.

The results of McLeod and Graves-Morris with equation (2.32) were called *almost* analogous because they supply a sufficient but not necessary condition. This allows the possibility that  $\varepsilon_{2k}^{(n)} = \mathbf{x}$  may hold for  $n = 0, 1, \dots$  when the VEA is applied to some vector sequence  $\{\mathbf{x}_n\}$ , even though no scalars  $a_i, i = 0, 1, \dots, k$  exist that will satisfy equation (2.32). Therefore, the set of sequences satisfying equation (2.32) may well be a proper subset of the kernel of the  $k^{\text{th}}$ -order VEA. Also, some of the sequences satisfying the condition for a given value of  $k$  may, in fact, be members of the kernel for an order of VEA smaller than  $k$ . The reality of both of these phenomena will be demonstrated in Chapters 3, 5 and 6.

As mentioned previously, the origins of the VEA prevented many theoretical results for the SEA and the Shanks transform from providing counterparts for the VEA. One of these deficits has been the lack of a way to express each term of a VEA transform sequence as the ratio of two determinants similar to that given for the Shanks transform in equation (2.1). Apart from the VEA, practically all known recursively computed sequence-to-sequence transformations have such expressions.

In 1994, Salam [20] showed that each vector of each VEA transform sequence can be expressed as a ratio of two designants<sup>1</sup> whose elements are vectors from a Clifford algebra<sup>2</sup>. In 1998, Salam [21] modified the usual definition for a Hankel determinant<sup>3</sup> to one consisting of vectors in a Clifford algebra instead of scalars and

---

<sup>1</sup> Designants are related to determinants but do not rely on multiplication being commutative.

<sup>2</sup> Clifford algebras are generalisations of Hamilton's quaternions and contain scalars and vectors along with entities called bivectors and other multivectors. Noncommutative multiplication is defined between all entities in a Clifford algebra.

<sup>3</sup> The Shanks transform can also be defined as a ratio of two Hankel determinants.



showed that such a determinant would itself be a vector. In 2000 [22], he revised the definition of a Hankel determinant to one which no longer relied on Clifford algebras, and indicated at the end of the paper that the kernel of the VEA would be forthcoming.

Also in 2000, Brezinski published a review paper [6] indicating that a determinantal expression for each vector in a VEA transform sequence had been found. It was presented at a conference in Luminy last summer by Salam and Graves-Morris, and will appear very soon in Numerical Algorithms with other papers presented at the conference. The expression involves determinants of order  $2k + 1$ , rather than  $k + 1$  as with Shanks transform. This result will furnish the kernel of all orders of the VEA. It will mean that the unique character of the VEA can soon be fully exploited by those engineering and mathematical applications which, by their nature, generate vector sequences of the same fundamental character as those which the VEA extrapolates.

### 3. THE REAL KERNEL OF THE 1<sup>ST</sup> ORDER VEA

Early in the author's research into the VEA, in an effort to come to some intuitive understanding of the way in which the 1<sup>st</sup> order VEA behaves, he repeatedly plotted three ordered but random points on the Cartesian plane, applied the 1<sup>st</sup> order VEA to the correspondingly ordered 2-dimensional vectors terminating at those three points, and then plotted the endpoint of the result. It seemed from the lines drawn from each of the three points to the point given by the VEA that the points transformed lay at equally spaced angles around an equiangular spiral whose centre was the endpoint of the vector given by the VEA. Figure 3.1, below, gives two representative examples whose vector endpoints have integer co-ordinates. The VEA limit point for each example has a small circle around it. Freehand curves suggesting the relevant spirals have also been added.

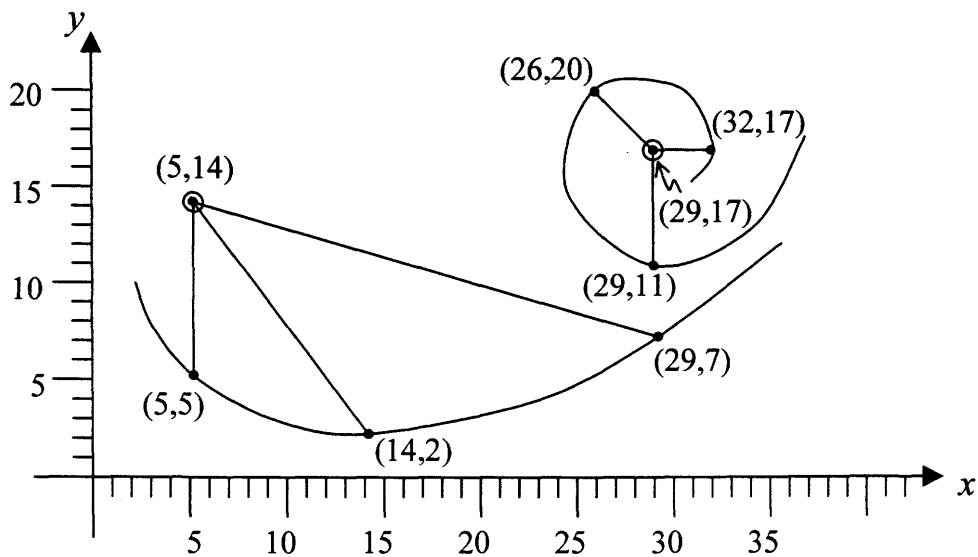


Figure 3.1 Two sets of three ordered pairs with their VEA limits.

The implication of this hypothesis was that each term  $\mathbf{x}_n - \mathbf{x}$  of the error vector sequence  $\{\mathbf{x}_n - \mathbf{x}\}$  is a scaled and rotated version of its predecessor. In this chapter, a geometric proof is given confirming that, for real vector sequences, this is indeed the case. This result first appeared in a paper by Steele and Dolovich [27] in June of 2000.

Section 3.1 gives, in symbolic terms, the condition for success of the 1<sup>st</sup> order VEA in transforming a real vector sequence into a constant sequence of its limit or antilimit. In §3.2 and §3.3, proofs that the condition is necessary and sufficient, respectively, are given. In §3.4 some comments are made on the result. Throughout this chapter, the term “VEA-1” will be used as an abbreviation for “the 1<sup>st</sup> order VEA”.

### 3.1 The Result

Theorem 3.1: Given a sequence of real  $N$ -dimensional vectors  $\{\mathbf{x}_n\}$ ,  $\varepsilon_2^{(n)} = \mathbf{x}$  for  $n = 0, 1, \dots$  if and only if

$$\mathbf{x}_n = \mathbf{x} + \lambda^n \mathbf{R}^n \mathbf{e}, \quad \text{for } n = 0, 1, \dots \quad (3.1)$$

where  $\lambda$  is a real scalar,  $\mathbf{R}$  is a real  $N \times N$  rotator,  $\mathbf{e}$  is a real non-zero vector existing solely within the plane of rotation of  $\mathbf{R}$ ,  $\lambda > 0$ , and  $\lambda \mathbf{R} \neq \mathbf{I}$ , the identity matrix. For  $N \geq 2$ , by rotator is meant a matrix which rotates vector components within a two-dimensional subspace but has no effect upon vector components within the remaining  $(N - 2)$ -dimensional subspace. Every Jordan canonical form of such a matrix will be diagonal, with two eigenvalues being complex conjugates of each other and of unit modulus and all other eigenvalues being unity. All eigenvectors will be mutually orthogonal. In addition, the eigenvector associated with the second complex eigenvalue will be the complex conjugate of an eigenvector associated with the first complex eigenvalue. (These attributes of a rotator are shown in §4.2 in the next chapter.) For  $N = 1$ , the con-

vention is adopted that the rotator is the  $1 \times 1$  identity matrix whose, rotation not being meaningful in a one-dimensional space.

Such sequences will be referred to as  $\lambda\mathbf{R}$  sequences. This result, therefore, may be restated as follows: with respect to real vector sequences, the kernel of VEA-1 is the set of all  $\lambda\mathbf{R}$  sequences.

### 3.2 Proof of Necessity

Recall the definition of the VEA from equations (2.22,23,25&26) in which

$$\boldsymbol{\varepsilon}_{-1}^{(n)} = \mathbf{0}, \quad \boldsymbol{\varepsilon}_0^{(n)} = \mathbf{x}_n, \quad n = 0, 1, \dots \quad (3.2)$$

and

$$\boldsymbol{\varepsilon}_{j+1}^{(n)} = \boldsymbol{\varepsilon}_{j-1}^{(n+1)} + \left[ \boldsymbol{\varepsilon}_j^{(n+1)} - \boldsymbol{\varepsilon}_j^{(n)} \right]^{-1} \quad j, n = 0, 1, \dots \quad (3.3)$$

where the inverse in (3.3) for real vectors is defined by

$$\mathbf{w}^{-1} = \frac{\mathbf{w}}{\langle \mathbf{w}, \mathbf{w} \rangle} \quad (3.4)$$

and the inner product for real vectors is

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^N x_i y_i. \quad (3.5)$$

The proof begins with a construction illustrating the geometry of VEA-1 when acting upon three real  $N$ -dimensional vectors. Consider three such vectors  $\mathbf{x}_0$ ,  $\mathbf{x}_1$ , and  $\mathbf{x}_2$  and

the vector  $\epsilon$ -table resulting from an application of VEA-1 to them which appears in Table 3.1, below.

Table 3.1 First order  $\epsilon$ -table for arbitrary real  $\mathbf{x}_0$ ,  $\mathbf{x}_1$ , and  $\mathbf{x}_2$ .

$$\begin{array}{rcc}
 & \epsilon_0^{(0)} = \mathbf{x}_0 & \\
 \epsilon_{-1}^{(1)} = \mathbf{0} & & \epsilon_1^{(0)} \\
 & \epsilon_0^{(1)} = \mathbf{x}_1 & \epsilon_2^{(0)} \\
 \epsilon_{-1}^{(2)} = \mathbf{0} & & \epsilon_1^{(1)} \\
 & \epsilon_0^{(2)} = \mathbf{x}_2 & \vdots \\
 \vdots & & \vdots \\
 & \vdots & 
 \end{array}$$

In seeking the necessary conditions for which  $\epsilon_2^{(n)} = \mathbf{x}$  for all non-negative integers  $n$ , sequences for which equation (3.3) will encounter division by zero must be eliminated. Therefore, a restriction is placed upon  $\mathbf{x}_1$  such that

$$\mathbf{x}_1 \neq \mathbf{x}_0 \tag{3.6}$$

$$\mathbf{x}_1 \neq \mathbf{x}_2 \tag{3.7}$$

and

$$\mathbf{x}_1 \neq \frac{(\mathbf{x}_0 + \mathbf{x}_2)}{2}. \tag{3.8}$$

In all other respects, vectors  $\mathbf{x}_0$ ,  $\mathbf{x}_1$ , and  $\mathbf{x}_2$  are arbitrary. Let their endpoints be called  $P_0$ ,  $P_1$ , and  $P_2$ , respectively, which define a plane depicted in Figure 3.2. Therefore, these vectors may also be called  $\overline{OP_0}$ ,  $\overline{OP_1}$ , and  $\overline{OP_2}$ , respectively,  $O$  being some

origin generally not in the plane of the page. (The case where  $P_0$ ,  $P_1$ , and  $P_2$  are collinear will be addressed below.)

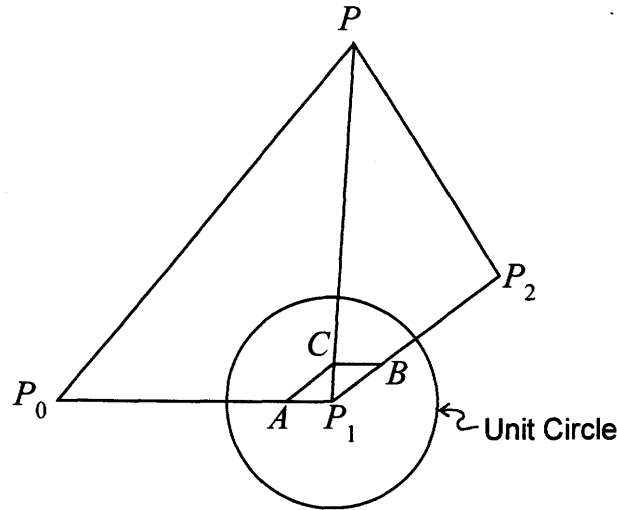


Figure 3.2. VEA-1 applied to 3 arbitrary vectors.

These three vectors correspond to  $\epsilon_0^{(0)}$ ,  $\epsilon_0^{(1)}$ , and  $\epsilon_0^{(2)}$ , respectively, in Table 3.1. By equations (3.6) and (3.7), the difference vectors  $\overline{P_0P_1}$  and  $\overline{P_1P_2}$  are invertible. When inverted (according to equation (3.4)) and added to the zero vectors ( $\epsilon_{-1}^{(1)}$  and  $\epsilon_{-1}^{(2)}$ ), vectors  $\overline{AP_1}$  and  $\overline{P_1B}$  ( $\epsilon_1^{(0)}$  and  $\epsilon_1^{(1)}$ ), respectively, are obtained. (The unit circle is shown as a visual reference for the inversion of vector lengths.) The difference between these vectors is  $\overline{P_1C}$  which, because of equation (3.8), inverts to give  $\overline{P_1P}$ . When added to the original second vector  $\overline{OP_1}$  ( $\epsilon_0^{(1)}$ ), the result is the vector  $\overline{OP}$  ( $\epsilon_2^{(0)}$ ). Finally, lines corresponding to vectors  $\overline{P_0P}$  and  $\overline{P_2P}$  are drawn, completing the construction.

The inverse relationships of  $\overline{P_0P_1}$  and  $\overline{P_1P}$  to  $\overline{AP_1}$  and  $\overline{P_1C}$ , respectively, cause  $\Delta P_0P_1P$  to be similar to  $\Delta CP_1A$ . For the same reasons,  $\Delta P_1P_2P$  is similar to  $\Delta P_1CB$ . It is also clear that  $\Delta CP_1A$  is congruent to  $\Delta P_1CB$ . Therefore,  $\Delta P_0P_1P$  must be similar to  $\Delta P_1P_2P$ . The arbitrariness of  $\mathbf{x}_0$ ,  $\mathbf{x}_1$ , and  $\mathbf{x}_2$  means that VEA-1 will always locate  $P$  in such a way that  $\Delta P_1P_2P$  will be a scaled and/or rotated version of  $\Delta P_0P_1P$ .

It can easily be shown that the scaling factor is given by

$$\lambda = \frac{\|\overline{P_1P_2}\|_2}{\|\overline{P_0P_1}\|_2}, \quad (3.9)$$

so that  $\lambda$  is real and  $\lambda \geq 0$ , and that the angle of rotation is given by

$$\theta = \pi - m\angle P_0P_1P_2, \quad (3.10)$$

where  $m\angle P_0P_1P_2$  is the measure of  $\angle P_0P_1P_2$ . For  $\theta = 0$  or  $\pi$ , endpoints  $P_0$ ,  $P_1$  and  $P_2$  are collinear and the triangles in Figure 3.2 are degenerate; each triangle is then a scaled but unrotated version of its neighbours. Despite this degeneracy, the steps in the construction above and the logic that follows them are still valid. This collinear case in two dimensions also illustrates the case for  $N = 1$ .

Now, if a vector sequence  $\{\mathbf{x}_n\}$  is such that VEA-1 gives the same vector  $\mathbf{x}$  when applied to any set of three consecutive vectors, then the geometric relationship between the head of  $\mathbf{x}$  and the heads of the sequence vectors is as shown in Figure 3.3. Because of the operation of VEA-1, every triangle is a scaled and/or rotated copy of its neighbours and thus every other triangle, and all lie on the same two-dimensional subspace.

Therefore, a unique matrix  $\mathbf{R}$  may be defined with rotation angle  $\theta$  given above such that

$$\mathbf{x}_{n+1} - \mathbf{x}_n = \lambda \mathbf{R}(\mathbf{x}_n - \mathbf{x}_{n-1}) \quad (3.11)$$

and

$$\mathbf{x}_{n+1} - \mathbf{x} = \lambda \mathbf{R}(\mathbf{x}_n - \mathbf{x}). \quad (3.12)$$

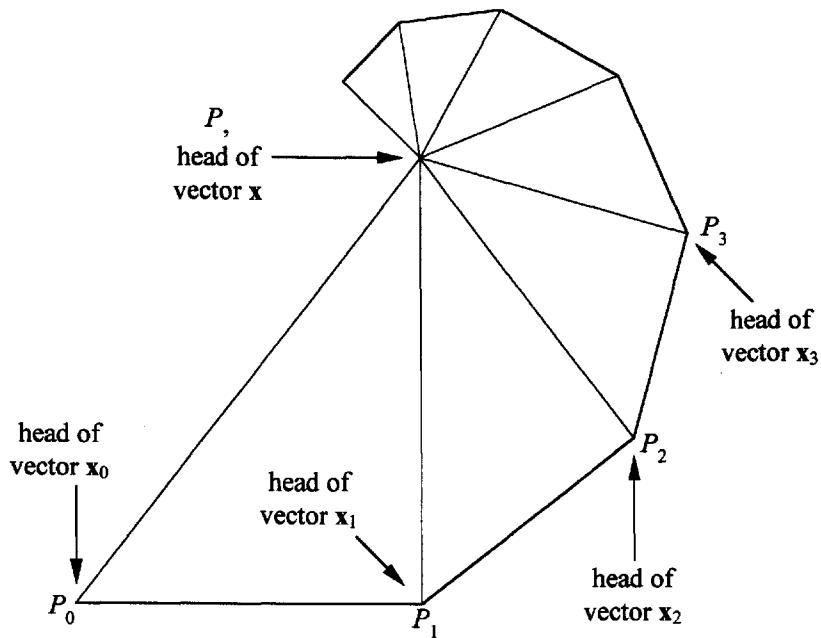


Figure 3.3. The heads of a vector sequence such that VEA-1 gives the vector  $\mathbf{x}$  for every set of three consecutive vectors.

If  $\mathbf{e} = \mathbf{x}_0 - \mathbf{x}$ , then

$$\mathbf{x}_0 = \mathbf{x} + \mathbf{e}, \quad (3.13)$$

and, by equation (3.12),



$$\mathbf{x}_1 = \mathbf{x} + \lambda \mathbf{R} \mathbf{e},$$

$$\mathbf{x}_2 = \mathbf{x} + \lambda^2 \mathbf{R}^2 \mathbf{e},$$

⋮

or

$$\mathbf{x}_n = \mathbf{x} + \lambda^n \mathbf{R}^n \mathbf{e}, \quad (3.14)$$

which is equation (3.1).

The restrictions placed on  $\mathbf{x}_1$  in equations (3.6), (3.7) and (3.8) must also be generalised to  $\mathbf{x}_n$ . In addition to equation (3.14), every  $\mathbf{x}_n$  must satisfy

$$\mathbf{x}_n \neq \mathbf{x}_{n+1} \quad (3.15)$$

and

$$\mathbf{x}_n \neq \frac{(\mathbf{x}_{n-1} + \mathbf{x}_{n+1})}{2}. \quad (3.16)$$

If we assume that  $\lambda = 0$ , equation (3.11) leads to a contradiction of equation (3.15), so  $\lambda > 0$ . If we assume that  $\lambda \mathbf{R} = \mathbf{I}$ , equation (3.11) leads to a contradiction of equation (3.16), so  $\lambda \mathbf{R} \neq \mathbf{I}$ . Finally, if we assume that  $\mathbf{e} = \mathbf{0}$ , then equation (3.12) leads to a contradiction of equation (3.15), so  $\mathbf{e}$  is non-zero. Thus, if VEA-1 is to give the same vector  $\mathbf{x}$  when applied to any three consecutive vectors in some sequence  $\{\mathbf{x}_n\}$ , it is necessary that  $\{\mathbf{x}_n\}$  be a  $\lambda \mathbf{R}$  sequence.

### 3.3 Proof of Sufficiency

The second part of the proof begins by considering the set of all  $\lambda\mathbf{R}$  sequences in some detail to enhance the clarity of what follows. By definition, the terms of any such sequence, say  $\{\mathbf{w}_n\}$ , are given by

$$\mathbf{w}_n = \mathbf{w} + \lambda^n \mathbf{R}^n \mathbf{e} \quad \text{for } n = 0, 1, 2, \dots \quad (3.17)$$

where  $\lambda > 0$  and  $\lambda\mathbf{R} \neq \mathbf{I}$  and  $\mathbf{e} \neq \mathbf{0}$ . Each  $\lambda\mathbf{R}$  sequence can be characterised by its parameters  $\mathbf{w}$ ,  $\lambda$ ,  $\mathbf{R}$ , and  $\mathbf{e}$ . If  $|\lambda| < 1$ , then  $\mathbf{w}$  is clearly the limit of the sequence because  $\lambda^n \mathbf{R}^n \mathbf{e}$  goes to  $\mathbf{0}$  as  $n$  approaches  $\infty$ . If  $|\lambda| > 1$ , then the sequence diverges from the unique vector  $\mathbf{w}$  which is the antilimit, about which more will be said below. In both of these cases, if  $\mathbf{R} = \mathbf{I}$ , then the heads of all vectors  $\mathbf{w}_n$  and of vector  $\mathbf{w}$  are collinear and the associated triangles are degenerate. If  $\lambda = 1$ , the prohibition  $\lambda\mathbf{R} \neq \mathbf{I}$  means that  $\mathbf{R} \neq \mathbf{I}$ , and that  $N \geq 2$ . In this case, each triangle is a rotated but unscaled version of its neighbours and, hence, isosceles. The heads of the vectors  $\mathbf{w}_n$  lie on the circumference of a circle centred on the head of  $\mathbf{w}$ .

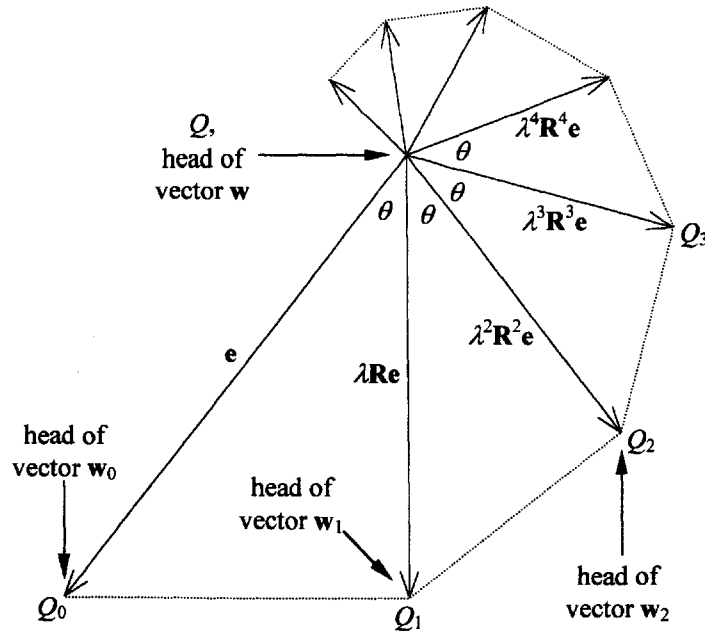


Figure 3.4. The heads of the vectors in an arbitrary  $\lambda R$  sequence  $\{w_n\}$  with the head of the limit (antilimit) vector  $w$ .

An arbitrary  $\lambda R$  sequence  $\{w_n\}$  is now considered. Let  $Q_n$  be the endpoint of vector  $w_n$  and let  $Q$  be the endpoint of vector  $w$ . Figure 3.4 shows the heads of these vectors with some of the accompanying geometric relationships for  $N \geq 2$ . It has already been shown that for each set of vectors  $w_n$ ,  $w_{n+1}$ , and  $w_{n+2}$ , VEA-1 will give a vector  $\varepsilon_2^{(n)}$  whose endpoint forms scaled and/or rotated similar triangles with  $Q_n$ ,  $Q_{n+1}$ , and  $Q_{n+2}$ . It remains only to show that such  $\varepsilon_2^{(n)} = w$  for an arbitrary  $n$ . The possibility that these two vectors are not equal is depicted in Figure 3.5.

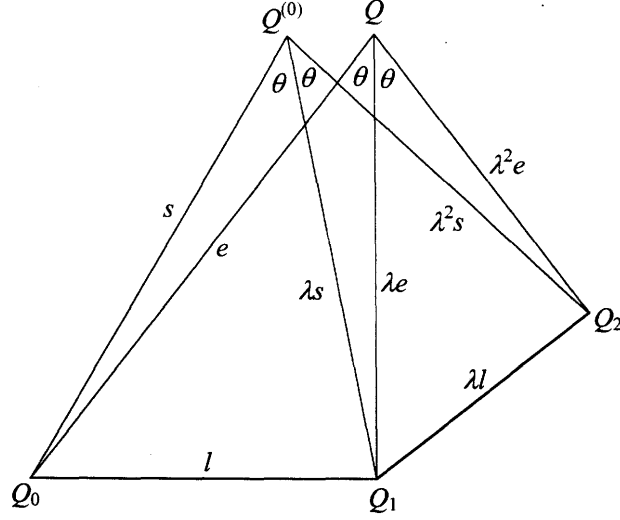


Figure 3.5. A possible difference between  $\mathbf{w}$  (having endpoint  $Q$ ) and  $\varepsilon_2^{(0)}$  (having endpoint  $Q^{(0)}$ ).

Consider segment  $\overline{Q_0Q_1}$ , the length of which will be called  $l$ . Let  $e$  be the length of  $\mathbf{e}$  which is  $\overline{QQ_0}$  and let  $s$  be the distance from  $Q_0$  to  $Q^{(0)}$ , the endpoint of  $\varepsilon_2^{(0)}$ . Both  $\theta$  and  $\lambda$  are known from the definition of the sequence. These will also determine the corresponding proportions for any triangles set up by VEA-1. The law of cosines gives expressions for  $l^2$  in terms of  $s$  and in terms of  $e$ :

$$l^2 = s^2 + \lambda^2 s^2 - 2s\lambda s \cos\theta = e^2 + \lambda^2 e^2 - 2e\lambda e \cos\theta \quad (3.18)$$

or

$$l^2 = s^2(1 + \lambda^2 - 2\lambda \cos\theta) = e^2(1 + \lambda^2 - 2\lambda \cos\theta) \quad (3.19)$$

so  $s^2 = e^2$ . Since both  $s$  and  $e$  are lengths,  $s = e$  and  $\Delta Q_0Q_1Q^{(0)}$  is congruent to  $\Delta Q_0Q_1Q$ . Therefore,  $Q^{(0)} = Q = Q^{(n)}$  for all  $n$ . If  $N = 1$ , the same logic holds and the

theorem is proven. In order for VEA-1 to give the limit of a vector sequence, it is both sufficient and necessary that the sequence be  $\lambda\mathbf{R}$ . ■

### 3.4 Comments on the Result

Note that a vector sequence may become  $\lambda\mathbf{R}$  without starting that way. In this case, it may be reindexed with 0 assigned to the first vector of the  $\lambda\mathbf{R}$  subsequence. When VEA-1 is applied to this  $\lambda\mathbf{R}$  subsequence, it will give the corresponding limit or antilimit.

The fact that VEA-1 locates the antilimit vector for a diverging  $\lambda\mathbf{R}$  sequence illustrates an often-useful property of the VEA. It is frequently able to obtain meaningful results from divergent computational processes. For example, its scalar counterpart, the SEA, obtains very accurate values for power series expansions outside the domain for which they are convergent.

Finally, it should be noticed that equation (3.12),

$$\mathbf{x}_{n+1} - \mathbf{x} = \lambda\mathbf{R}(\mathbf{x}_n - \mathbf{x}) \quad (3.20)$$

resembles equation (2.32) for  $k = 1$  which becomes

$$\mathbf{x}_{n+1} - \mathbf{x} = c(\mathbf{x}_n - \mathbf{x}) \quad (3.21)$$

where  $c = -a_0/a_1$ . Sequences satisfying (3.21) are such that each error vector is only a scaled version of the previous error vector. Such sequences cannot exhibit the rotation shown by typical  $\lambda\mathbf{R}$  sequences satisfying equation (3.20). Conversely, sequences

satisfying (3.21) are a subset of the set of  $\lambda\mathbf{R}$  sequences, namely, those for which  $\mathbf{R} = \mathbf{I}$ . It will be shown, however, that equation (2.32) for  $k = 2$  can be used to describe any  $\lambda\mathbf{R}$  sequence. Recall that for such a sequence, all error vectors lie on a subspace spanned by any two consecutive error vectors  $(\mathbf{x}_n - \mathbf{x})$  and  $(\mathbf{x}_{n+1} - \mathbf{x})$ . Clearly then, for a given  $n$ ,  $(\mathbf{x}_{n+2} - \mathbf{x})$  can be expressed as a linear combination of  $(\mathbf{x}_n - \mathbf{x})$  and  $(\mathbf{x}_{n+1} - \mathbf{x})$ . Therefore,

$$a_0(\mathbf{x}_n - \mathbf{x}) + a_1(\mathbf{x}_{n+1} - \mathbf{x}) + a_2(\mathbf{x}_{n+2} - \mathbf{x}) = \mathbf{0} \quad (3.22)$$

which is equation (2.32) with  $k = 2$ . If both sides of (3.22) are premultiplied by  $\lambda\mathbf{R}$ ,

$$a_0\lambda\mathbf{R}(\mathbf{x}_n - \mathbf{x}) + a_1\lambda\mathbf{R}(\mathbf{x}_{n+1} - \mathbf{x}) + a_2\lambda\mathbf{R}(\mathbf{x}_{n+2} - \mathbf{x}) = \mathbf{0} \quad (3.23)$$

is obtained. By equation (3.12), this becomes

$$a_0(\mathbf{x}_{n+1} - \mathbf{x}) + a_1(\mathbf{x}_{n+2} - \mathbf{x}) + a_2(\mathbf{x}_{n+3} - \mathbf{x}) = \mathbf{0}. \quad (3.24)$$

By induction, it is clear that for every  $\lambda\mathbf{R}$  sequence, there exist  $a_0$ ,  $a_1$ , and  $a_2$  such that (3.22) is satisfied for all  $n$ . According to equation (2.32) of Theorem 2.6, therefore, an application of the second order VEA is sufficient to obtain  $\mathbf{x}$  for any  $\lambda\mathbf{R}$  sequence. In contrast, the result of this chapter shows that the 1<sup>st</sup> order VEA is sufficient.

It will be seen in Chapters 5 and 6 that  $\lambda\mathbf{R}$  sequences are the simplest of several families of vector sequences requiring an order of VEA which is lower than that offered by the result of McLeod and Graves-Morris.

#### 4. AN ALTERNATIVE FORM FOR THE REAL KERNEL OF THE 1<sup>ST</sup> ORDER VEA

Looking back on the result of Chapter 3 and equation (3.1), it can be seen that there are more constraints in the definition of  $\mathbf{R}$  than the result actually requires. Recall that if  $\{\mathbf{x}_n - \mathbf{x}\}$  is the error sequence corresponding to a  $\lambda\mathbf{R}$  sequence, then by equation (3.1),

$$\mathbf{x}_n - \mathbf{x} = \lambda^n \mathbf{R}^n \mathbf{e} \quad (4.1)$$

where non-zero  $\mathbf{e}$  exists solely within the plane of rotation of  $\mathbf{R}$ . This is equivalent to saying that  $\mathbf{e}$  is a linear combination of the two eigenvectors of  $\mathbf{R}$  that span the plane of rotation. Because of this, the only relevant eigenvalues of  $\mathbf{R}$  are the two associated with these spanning eigenvectors. In this case, however, there are an infinite number of other matrices that could stand in for  $\mathbf{R}$  in equation (4.1) just as well. Instead of having  $N - 2$  eigenvalues equal to unity, those eigenvalues could equal anything without interfering with the definition. To succinctly capture the essence of a  $\lambda\mathbf{R}$  sequence and the real part of the kernel of the 1<sup>st</sup> order VEA, it is necessary to refine this definition until it relies upon a much leaner structure than a rotation matrix. The refinement process begins with further observations about rotation, as it is not covered in any depth in any of the several texts on linear algebra consulted by the author and the concept is central to the result of this chapter.

First, a new entity is defined. Recall from the previous chapter that the kernel of the 1<sup>st</sup> order VEA is the set of  $\lambda\mathbf{R}$  sequences. These sequences all have the property that each error term  $\mathbf{x}_n - \mathbf{x}$  is a scaled and rotated version of  $\mathbf{x}_{n-1} - \mathbf{x}$ . Therefore, a

definition of a *real scaled rotation sequence* or RSRS is given, stating it to be a sequence  $\{\mathbf{e}_n\}$  whose terms satisfy

$$\mathbf{e}_n = \lambda^n \mathbf{R}^n \mathbf{e}, \quad \text{for } n = 0, 1, 2, \dots \quad (4.2)$$

where  $\lambda$  is a real scalar,  $\mathbf{R}$  is a real  $N \times N$  rotation matrix,  $\mathbf{e}$  is a non-zero real vector existing solely within the plane of rotation of  $\mathbf{R}$ , with  $\lambda > 0$ , and  $\lambda \mathbf{R} \neq \mathbf{I}$ , the identity matrix. An RSRS may therefore be thought of as a  $\lambda \mathbf{R}$  sequence each of whose vectors has been translated so that its limit or antilimit is the origin.

Section 4.1 delves somewhat intuitively into the topic of scaled rotations in a real vector space, deriving another formula satisfied by the vectors of an RSRS. Section 4.2 does the same thing through the eigenvalue decomposition of the corresponding rotation matrix. Finally, §4.3 gives an alternative form for the condition proved in §§3.2 and 3.3.

Although the author believes that all of the elements of the analyses presented in this chapter must surely have appeared already in print previously for other purposes, the author knows of no such publications, allowing a claim at least of originality to be made if not of priority of publication.

## 4.1 An Intuitive Definition of a Real Scaled Rotation Sequence

The starting point for this section is the knowledge of three vectors:  $\mathbf{e}_0$ , the first vector in  $\{\mathbf{e}_n\}$  and  $\mathbf{v}_{bef}$  and  $\mathbf{v}_{aft}$  representing, respectively, an arbitrary vector in the plane of rotation of  $\mathbf{R}$  before and after undergoing the scaled rotation. Subsection 4.1.1



provides a derivation of an expression for the terms  $\mathbf{e}_n$ ,  $n = 0, 1, \dots$  of an RSRS using only real scalars and vectors in the definition. Then, in Subsection 4.1.2, a derivation is given for a different expression for each of the vectors of the same sequence  $\{\mathbf{e}_n\}$  using complex scalars along with vectors in the natural complex extension of the previous real vector space. In the process, the linear algebra of real scaled rotation sequences is explored.

#### 4.1.1 A Real Scaled Rotation Sequence in Real Terms

An RSRS  $\{\mathbf{e}_n\}$  may be defined quite intuitively using  $\mathbf{e}_0$ , the first vector in  $\{\mathbf{e}_n\}$ , and two real linearly independent vectors  $\mathbf{v}_{bef}$  and  $\mathbf{v}_{aft}$  respectively representing an arbitrary vector in the plane of rotation before and after undergoing the defined scaling and rotation. The vectors  $\mathbf{v}_{bef}$  and  $\mathbf{v}_{aft}$  simultaneously and conveniently define a plane of rotation, an angle of rotation, and a scaling factor. They may be used to define a unique rotation matrix  $\mathbf{R}$  and a scaling factor  $\lambda$  such that

$$\mathbf{v}_{aft} = \lambda \mathbf{R} \mathbf{v}_{bef}. \quad (4.3)$$

By equation (4.2) with  $\mathbf{R}^0 = \mathbf{I}$ ,

$$\mathbf{e}_0 = \mathbf{e} \quad (4.4)$$

which exists solely in the plane of rotation of  $\mathbf{R}$ . Therefore, an orthonormal basis for this plane can be obtained as follows. First,  $\mathbf{v}_{bef}$  is normalised to give  $\mathbf{q}_1$ , a unit vector in the plane of rotation. So, let

$$\mathbf{q}_1 = \frac{\mathbf{v}_{bef}}{\|\mathbf{v}_{bef}\|_2}. \quad (4.5)$$

Next,  $\mathbf{v}_\perp$  (“v-perp”), orthogonal to  $\mathbf{q}_1$  but also in the plane defined by  $\mathbf{v}_{bef}$  and  $\mathbf{v}_{aft}$ , is defined by removing from  $\mathbf{v}_{aft}$  its orthogonal projection onto  $\mathbf{q}_1$ . Thus,

$$\mathbf{v}_\perp = \mathbf{v}_{aft} - \langle \mathbf{v}_{aft}, \mathbf{q}_1 \rangle \mathbf{q}_1. \quad (4.6)$$

If  $\mathbf{v}_{aft} = p\mathbf{v}_{bef}$  where  $p$  is a real scalar, then  $\mathbf{v}_{bef}$  and  $\mathbf{v}_{aft}$  are not independent and equation (4.6) will yield  $\mathbf{v}_\perp = \mathbf{0}$ . If this is the case, the vector  $\mathbf{e}$  can be substituted for  $\mathbf{v}_{aft}$  in equation (4.6) to give  $\mathbf{v}_\perp$ . This will ensure that the plane defined contains  $\mathbf{e}$ . If  $\mathbf{e} = q\mathbf{v}_{bef}$  where  $q$  is also a real scalar, then any arbitrary vector may be used in place of  $\mathbf{v}_{aft}$  in equation (4.6) to give  $\mathbf{v}_\perp$ . Normalising  $\mathbf{v}_\perp$  to give

$$\mathbf{q}_2 = \frac{\mathbf{v}_\perp}{\|\mathbf{v}_\perp\|_2}, \quad (4.7)$$

the orthonormal basis for the plane of rotation is completed. These vectors are depicted in Figure 4.1, below.

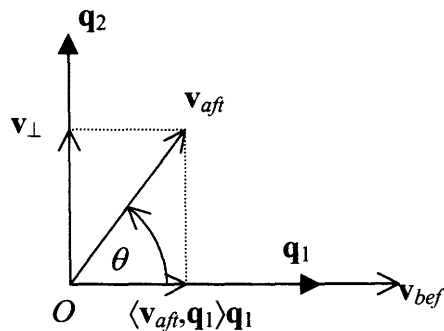


Figure 4.1 The orthonormal basis,  $\mathbf{q}_1$  and  $\mathbf{q}_2$ , for the plane of rotation from  $\mathbf{v}_{bef}$  and  $\mathbf{v}_{aft}$ .

Since  $\mathbf{q}_1$  and  $\mathbf{q}_2$  constitute an orthonormal basis for the plane of rotation, every vector lying in the plane of rotation may be expressed as a real linear combination of  $\mathbf{q}_1$  and  $\mathbf{q}_2$ , including the vectors of the RSRS  $\{\mathbf{e}_n\}$  which will satisfy

$$\mathbf{e}_n = r_{n,1}\mathbf{q}_1 + r_{n,2}\mathbf{q}_2, \quad \text{for } n = 0, 1, \dots \quad (4.8)$$

where the  $r_{n,i}$  are real. In particular,

$$\mathbf{e}_0 = r_{0,1}\mathbf{q}_1 + r_{0,2}\mathbf{q}_2 \quad (4.9)$$

or, in the terms of linear algebra,

$$\mathbf{e}_0 = \mathbf{Q}'\mathbf{r}_0' \quad (4.10)$$

where

$$\mathbf{Q}' = \begin{bmatrix} | & | \\ | & | \\ \mathbf{q}_1 & \mathbf{q}_2 \\ | & | \\ | & | \end{bmatrix} \quad (4.11)$$

and

$$\mathbf{r}_0' = \begin{bmatrix} r_{0,1} \\ r_{0,2} \end{bmatrix}. \quad (4.12)$$

The matrix  $\mathbf{Q}'$  may be extended by defining

$$\mathbf{Q} = \begin{bmatrix} | & | & | & \cdots & | \\ | & | & | & \cdots & | \\ \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{q}_3 & \cdots & \mathbf{q}_N \\ | & | & | & \cdots & | \\ | & | & | & \cdots & | \end{bmatrix}, \quad (4.13)$$

an orthogonal matrix. (Gram-Schmidt orthogonalisation will supply the additional orthogonal vectors  $\mathbf{q}_i$ ,  $3 \leq i \leq N$ , necessary to obtain a complete orthogonal basis for the space.) The vector  $\mathbf{r}'$  may also be extended by defining

$$\mathbf{r}_0 = \begin{bmatrix} r_{0,1} \\ r_{0,2} \\ r_{0,3} \\ \vdots \\ r_{0,N} \end{bmatrix} = \begin{bmatrix} r_{0,1} \\ r_{0,2} \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad (4.14)$$

so that

$$\mathbf{Q}\mathbf{r}_0 = \mathbf{e}_0 \quad (4.15)$$

also holds.

The coefficients  $r_{0,1}$  and  $r_{0,2}$  may be determined by premultiplying equation (4.15) by  $\mathbf{Q}^{-1}$  which is  $\mathbf{Q}^T$ ,  $\mathbf{Q}$  being orthogonal, to obtain

$$\mathbf{r}_0 = \mathbf{Q}^T \mathbf{e}_0 \quad (4.16)$$

or

$$\begin{bmatrix} r_{0,1} \\ r_{0,2} \\ r_{0,3} \\ \vdots \\ r_{0,N} \end{bmatrix} = \begin{bmatrix} - & - & \mathbf{q}_1^T & - & - \\ - & - & \mathbf{q}_2^T & - & - \\ - & - & \mathbf{q}_3^T & - & - \\ & & \vdots & & \\ - & - & \mathbf{q}_N^T & - & - \end{bmatrix} \begin{bmatrix} | \\ | \\ | \\ | \\ | \end{bmatrix} \mathbf{e}_0. \quad (4.17)$$

In fact, if it is known that  $\mathbf{e}_0$  is solely in the plane spanned by  $\mathbf{q}_1$  and  $\mathbf{q}_2$ , then it is necessary only to find  $\mathbf{r}_0'$  which, by equation (4.17) is

$$\begin{bmatrix} r_{0,1} \\ r_{0,2} \end{bmatrix} = \begin{bmatrix} - & - & \mathbf{q}_1^T & - & - \\ - & - & \mathbf{q}_2^T & - & - \end{bmatrix} \begin{bmatrix} | \\ | \\ | \\ | \\ | \end{bmatrix} \mathbf{e}_0 \quad (4.18)$$

or

$$\mathbf{r}_0' = \mathbf{Q}^T \mathbf{e}_0. \quad (4.19)$$

Suppose it were necessary to verify that  $\mathbf{e}_0$  is in the plane of rotation defined by  $\mathbf{v}_{bef}$  and  $\mathbf{v}_{aft}$ . Resort could be made to equation (4.17) with the knowledge that if, for some  $i > 2$ ,  $r_{0,i} \neq 0$ , then clearly  $\mathbf{e}_0$  contains components outside of the plane of rotation spanned by  $\mathbf{q}_1$  and  $\mathbf{q}_2$ . However, since multiplication by an orthogonal matrix is a norm-preserving operation, if

$$r_{0,1}^2 + r_{0,2}^2 < \|\mathbf{e}_0\|_2^2, \quad (4.20)$$

this also will indicate that  $\mathbf{e}_0$  contains components outside of the plane of rotation. Thus, the co-ordinates of  $\mathbf{e}_0$  may be determined and its membership in the plane of rotation verified using the shorter calculation of equation (4.19).

Once

$$e_0 = \|\mathbf{e}_0\|_2, \quad (4.21)$$

the Euclidean length of  $\mathbf{e}_0$ , is calculated,  $\mathbf{e}_0$  may be expressed in polar form as

$$\mathbf{e}_0 = e_0 (\cos \phi_0 \mathbf{q}_1 + \sin \phi_0 \mathbf{q}_2) \quad (4.22)$$

where  $\phi_0$  is the angle in the plane of rotation from  $\mathbf{q}_1$  to  $\mathbf{e}_0$  measured toward  $\mathbf{q}_2$ . Therefore, by comparison of equation (4.22) with equation (4.9)

$$r_{0,1} = e_0 \cos \phi_0 \quad (4.23)$$

and

$$r_{0,2} = e_0 \sin \phi_0. \quad (4.24)$$

If  $r_{0,2}$  is not negative, then

$$\phi_0 = \cos^{-1} \left( \frac{r_{0,1}}{e_0} \right). \quad (4.25)$$

Otherwise,

$$\phi_0 = 2\pi - \cos^{-1} \left( \frac{r_{0,1}}{e_0} \right). \quad (4.26)$$

Now, since  $\mathbf{v}_{bef}$  and  $\mathbf{v}_{aft}$  are real vectors, the angle of rotation from  $\mathbf{v}_{bef}$  to  $\mathbf{v}_{aft}$  may be found by normalising  $\mathbf{v}_{bef}$  and  $\mathbf{v}_{aft}$  to give  $\mathbf{q}_{bef} = \mathbf{q}_1$  from equation (4.5) and

$$\mathbf{q}_{aft} = \frac{\mathbf{v}_{aft}}{\|\mathbf{v}_{aft}\|_2}, \quad (4.27)$$

respectively, so that

$$\cos \theta = \langle \mathbf{q}_{bef}, \mathbf{q}_{aft} \rangle \quad (4.28)$$

and

$$\theta = \cos^{-1} \langle \mathbf{q}_{bef}, \mathbf{q}_{aft} \rangle \quad (4.29)$$

where  $\theta$  is the angle from  $\mathbf{q}_{bef}$  to  $\mathbf{q}_{aft}$  measured in the direction toward  $\mathbf{q}_2$ . Because of the way in which  $\mathbf{v}_\perp$  and then  $\mathbf{q}_2$  have been defined in equations (4.6) and (4.7),  $0 \leq \theta \leq \pi$ .

Finally,  $\lambda$ , the scaling factor, is defined by

$$\lambda = \frac{\|\mathbf{v}_{aft}\|_2}{\|\mathbf{v}_{bef}\|_2}. \quad (4.30)$$

Now,  $\mathbf{e}_1$ , the next vector of the RSRS, may be determined. This is the vector  $\mathbf{e}_0$  after it has undergone the rotation and scaling defined by  $\mathbf{v}_{bef}$  and  $\mathbf{v}_{aft}$ . Therefore, the angle from  $\mathbf{q}_1$  to  $\mathbf{e}_1$  measured toward  $\mathbf{q}_2$  will be the angle from  $\mathbf{q}_1$  to  $\mathbf{e}_0$  measured toward  $\mathbf{q}_2$  which is  $\phi_0$ , plus the angle from  $\mathbf{v}_{bef}$  to  $\mathbf{v}_{aft}$  measured toward  $\mathbf{q}_2$  which is  $\theta$ . Thus,

$$\mathbf{e}_1 = \lambda e_0 [\cos(\phi_0 + \theta) \mathbf{q}_1 + \sin(\phi_0 + \theta) \mathbf{q}_2]. \quad (4.31)$$

By induction,

$$\mathbf{e}_n = \lambda^n e_0 [\cos(\phi_0 + n\theta) \mathbf{q}_1 + \sin(\phi_0 + n\theta) \mathbf{q}_2]. \quad (4.32)$$

#### 4.1.2 A Real Scaled Rotation Sequence in Complex Terms

Now, for the purpose of defining the very same RSRS in a new way, the real vector space is extended to its natural complex counterpart and a complex unit vector

$$\mathbf{u} = \frac{\mathbf{q}_1 - i\mathbf{q}_2}{\sqrt{2}} \quad (4.33)$$

is defined so that its complex conjugate is

$$\bar{\mathbf{u}} = \frac{\mathbf{q}_1 + i\mathbf{q}_2}{\sqrt{2}}. \quad (4.34)$$

Therefore,

$$\mathbf{q}_1 = \frac{\mathbf{u} + \bar{\mathbf{u}}}{\sqrt{2}} \quad (4.35)$$

and

$$\mathbf{q}_2 = \frac{i(\mathbf{u} - \bar{\mathbf{u}})}{\sqrt{2}}. \quad (4.36)$$

Note an important property of  $\mathbf{u}$  revealed by the inner product  $\langle \mathbf{u}, \bar{\mathbf{u}} \rangle$ . By definition,



$$\langle \mathbf{u}, \bar{\mathbf{u}} \rangle = \left\langle \frac{\mathbf{q}_1 - i\mathbf{q}_2}{\sqrt{2}}, \frac{\mathbf{q}_1 + i\mathbf{q}_2}{\sqrt{2}} \right\rangle. \quad (4.37)$$

Because inner products are distributive,

$$\begin{aligned} \langle \mathbf{u}, \bar{\mathbf{u}} \rangle &= \left\langle \frac{\mathbf{q}_1}{\sqrt{2}}, \frac{\mathbf{q}_1}{\sqrt{2}} \right\rangle + \left\langle \frac{\mathbf{q}_1}{\sqrt{2}}, \frac{i\mathbf{q}_2}{\sqrt{2}} \right\rangle \\ &\quad + \left\langle \frac{-i\mathbf{q}_2}{\sqrt{2}}, \frac{\mathbf{q}_1}{\sqrt{2}} \right\rangle + \left\langle \frac{-i\mathbf{q}_2}{\sqrt{2}}, \frac{i\mathbf{q}_2}{\sqrt{2}} \right\rangle \end{aligned} \quad (4.38)$$

Recall, also, that inner products are conjugate linear, which is to say that for any complex scalar  $z$ ,

$$\langle z\mathbf{a}, \mathbf{b} \rangle = z\langle \mathbf{a}, \mathbf{b} \rangle \quad (4.39)$$

and

$$\langle \mathbf{a}, z\mathbf{b} \rangle = \bar{z}\langle \mathbf{a}, \mathbf{b} \rangle. \quad (4.40)$$

Therefore,

$$\begin{aligned} \langle \mathbf{u}, \bar{\mathbf{u}} \rangle &= \left\langle \frac{\mathbf{q}_1}{\sqrt{2}}, \frac{\mathbf{q}_1}{\sqrt{2}} \right\rangle - i\left\langle \frac{\mathbf{q}_1}{\sqrt{2}}, \frac{\mathbf{q}_2}{\sqrt{2}} \right\rangle \\ &\quad - i\left\langle \frac{\mathbf{q}_2}{\sqrt{2}}, \frac{\mathbf{q}_1}{\sqrt{2}} \right\rangle - \left\langle \frac{\mathbf{q}_2}{\sqrt{2}}, \frac{\mathbf{q}_2}{\sqrt{2}} \right\rangle \end{aligned} \quad (4.41)$$

Since  $\mathbf{q}_1$  and  $\mathbf{q}_2$  are mutually-orthogonal real unit vectors, this becomes

$$\langle \mathbf{u}, \bar{\mathbf{u}} \rangle = \frac{1}{2} - i0 - i0 - \frac{1}{2} = 0 \quad (4.42)$$

and it is seen that  $\mathbf{u}$  is orthogonal to its complex conjugate: In fact, for any complex scalars  $c_1$  and  $c_2$ , it is found that  $c_1\mathbf{u}$  is orthogonal to  $c_2\bar{\mathbf{u}}$  because

$$\langle c_1\mathbf{u}, c_2\bar{\mathbf{u}} \rangle = c_1\bar{c}_2\langle \mathbf{u}, \bar{\mathbf{u}} \rangle = 0. \quad (4.43)$$

Continuing on, if equations (4.35) and (4.36) are substituted into equation (4.32), then

$$\mathbf{e}_n = \lambda^n e_0 \left[ \cos(\phi_0 + n\theta) \frac{(\mathbf{u} + \bar{\mathbf{u}})}{\sqrt{2}} + \sin(\phi_0 + n\theta) \frac{i(\mathbf{u} - \bar{\mathbf{u}})}{\sqrt{2}} \right] \quad (4.44)$$

or

$$\mathbf{e}_n = \frac{\lambda^n e_0}{\sqrt{2}} \left\{ [\cos(\phi_0 + n\theta) + i \sin(\phi_0 + n\theta)] \mathbf{u} + [\cos(\phi_0 + n\theta) - i \sin(\phi_0 + n\theta)] \bar{\mathbf{u}} \right\} \quad (4.45)$$

is obtained. It follows from Euler's formula,

$$e^{i\alpha} = \cos \alpha + i \sin \alpha, \quad (4.46)$$

that

$$\mathbf{e}_n = \frac{\lambda^n e_0}{\sqrt{2}} \left\{ e^{i(\phi_0 + n\theta)} \mathbf{u} + e^{-i(\phi_0 + n\theta)} \bar{\mathbf{u}} \right\} \quad (4.47)$$

or

$$\mathbf{e}_n = \frac{\lambda^n e_0}{\sqrt{2}} e^{i\phi_0} e^{in\theta} \mathbf{u} + \frac{\lambda^n e_0}{\sqrt{2}} e^{-i\phi_0} e^{-in\theta} \bar{\mathbf{u}} \quad (4.48)$$

which is

$$\mathbf{e}_n = \frac{e_0}{\sqrt{2}} e^{i\phi_0} \lambda^n e^{in\theta} \mathbf{u} + \frac{e_0}{\sqrt{2}} e^{-i\phi_0} \lambda^n e^{-in\theta} \overline{\mathbf{u}} \quad (4.49)$$

or

$$\mathbf{e}_n = z^n \mathbf{w} + \overline{z}^n \overline{\mathbf{w}} \quad (4.50)$$

where

$$z = \lambda e^{i\theta} \quad (4.51)$$

and

$$\mathbf{w} = \frac{e_0}{\sqrt{2}} e^{i\phi_0} \mathbf{u}. \quad (4.52)$$

Finally, it is necessary to deal with the other restrictions in the definition of equation (4.2). If  $\lambda > 0$ , then by equation (4.51),  $z \neq 0$ . Also, if  $\lambda \mathbf{R} \neq \mathbf{I}$  then it must not be simultaneously true that  $\lambda = 1$  and  $\theta = 0$ . Thus, by equation (4.51),  $z \neq 1$ . If  $\mathbf{e} \neq \mathbf{0}$ , then by equations (4.4) and (4.21),  $e_0 > 0$ , so by equation (4.52),  $\mathbf{w} \neq \mathbf{0}$ . Recall also that by equations (4.43) and (4.52),  $\mathbf{w}$  is orthogonal to its complex conjugate. Therefore, an alternate definition for a real scaled rotation sequence to that of equation (4.2) is a sequence  $\{\mathbf{e}_n\}$  where, for  $n = 0, 1, \dots$ ,  $\mathbf{e}_n$  satisfies equation (4.50) with complex  $z \neq 0$ ,  $z \neq 1$ , and non-zero  $\mathbf{w}$  is orthogonal to  $\overline{\mathbf{w}}$ .

In the next section, equation (4.50) is obtained from the properties of a rotation matrix  $\mathbf{R}$  acting on two orthogonal unit vectors in the plane of rotation.

## 4.2 An Eigenvector Expansion for a Real Scaled Rotation Sequence

It is also possible to arrive at the expression given in equation (4.50) for an RSRS through an eigenvalue decomposition of the rotator  $\mathbf{R}$  in equation (4.2). Such a rotator may be defined by two real orthonormal vectors,  $\mathbf{q}_1$  and  $\mathbf{q}_2$ , which span the plane of rotation and by an angle  $\theta$  through which the vectors are rotated in the direction from  $\mathbf{q}_1$  toward  $\mathbf{q}_2$ . Thus,

$$\mathbf{R}\mathbf{q}_1 = \cos\theta\mathbf{q}_1 + \sin\theta\mathbf{q}_2 \quad (4.53)$$

and

$$\mathbf{R}\mathbf{q}_2 = -\sin\theta\mathbf{q}_1 + \cos\theta\mathbf{q}_2. \quad (4.54)$$

This is depicted in Figure 4.2, below.

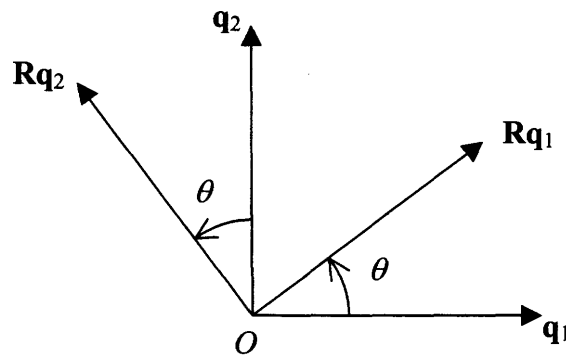


Figure 4.2 - The action of  $\mathbf{R}$  on vectors  $\mathbf{q}_1$  and  $\mathbf{q}_2$ .

Further, for every vector  $\mathbf{y}$  orthogonal to both  $\mathbf{q}_1$  and  $\mathbf{q}_2$ ,

$$\mathbf{R}\mathbf{y} = \mathbf{y} \quad (4.55)$$

An orthonormal set of vectors  $\mathbf{q}_i$ ,  $3 \leq i \leq N$ , are defined which are also orthogonal to both  $\mathbf{q}_1$  and  $\mathbf{q}_2$  so that they satisfy

$$\mathbf{R}\mathbf{q}_i = \mathbf{q}_i, \quad 3 \leq i \leq N. \quad (4.56)$$

In matrix form, equations (4.53), (4.54) and (4.56) are expressed as

$$\mathbf{R} \begin{bmatrix} | & | & | & & | \\ | & | & | & & | \\ \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{q}_3 & \cdots & \mathbf{q}_N \\ | & | & | & & | \\ | & | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & | & & | \\ | & | & | & & | \\ \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{q}_3 & \cdots & \mathbf{q}_N \\ | & | & | & & | \\ | & | & | & & | \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta & 0 & \cdots & 0 \\ \sin\theta & \cos\theta & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \quad (4.57)$$

or

$$\mathbf{R}\mathbf{Q} = \mathbf{Q}\mathbf{R}' \quad (4.58)$$

where

$$\mathbf{Q} = \begin{bmatrix} | & | & | & & | \\ | & | & | & & | \\ \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{q}_3 & \cdots & \mathbf{q}_N \\ | & | & | & & | \\ | & | & | & & | \end{bmatrix} \quad (4.59)$$

and

$$\mathbf{R}' = \begin{bmatrix} \cos\theta & -\sin\theta & 0 & \cdots & 0 \\ \sin\theta & \cos\theta & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}. \quad (4.60)$$

Postmultiplying both sides of (4.58) by  $\mathbf{Q}^{-1}$ ,

$$\mathbf{R} = \mathbf{Q}\mathbf{R}'\mathbf{Q}^{-1}. \quad (4.61)$$

This is equivalent to saying that  $\mathbf{R}$  is similar to  $\mathbf{R}'$ . Thus, every rotator is similar to  $\mathbf{R}'$  for some angle  $\theta$ .

Now, an eigenvalue decomposition of  $\mathbf{R}'$  is given. By inspection, it is plain that columns 3 through  $N$  of  $\mathbf{R}'$  are eigenvectors of  $\mathbf{R}'$  associated with unity eigenvalues. The remaining two eigenvalues and two eigenvectors can be found by examining the  $2 \times 2$  block in the top left-hand corner of  $\mathbf{R}'$  which will be called  $\mathbf{R}_2'$ . For each remaining eigenvalue  $\mu$  of  $\mathbf{R}'$ , it is known that  $\det(\mathbf{R}_2' - \mu\mathbf{I}) = 0$  or

$$\begin{vmatrix} \cos\theta - \mu & -\sin\theta \\ \sin\theta & \cos\theta - \mu \end{vmatrix} = \cos^2\theta - 2\mu\cos\theta + \mu^2 + \sin^2\theta = 0 \quad (4.62)$$

or

$$1 - 2\mu\cos\theta + \mu^2 = 0. \quad (4.63)$$

This quadratic polynomial in  $\mu$  is solved to give

$$\mu_1 = \cos\theta + i\sin\theta. \quad (4.64)$$

and

$$\mu_2 = \overline{\mu_1} = \cos\theta - i\sin\theta. \quad (4.65)$$

The constant  $\mu$  will be defined to be equal to  $\mu_1$  so that  $\mu$  and  $\bar{\mu}$  are the two eigenvalues of  $\mathbf{R}_2'$ . Therefore,

$$\mathbf{R}_2' - \mu \mathbf{I} = \begin{bmatrix} -i \sin \theta & -\sin \theta \\ \sin \theta & -i \sin \theta \end{bmatrix} = -\sin \theta \begin{bmatrix} i & 1 \\ -1 & i \end{bmatrix} \quad (4.66)$$

has a null space spanned by

$$\mathbf{u}' = \begin{bmatrix} 1 \\ -i \end{bmatrix} \quad (4.67)$$

while

$$\mathbf{R}_2' - \bar{\mu} \mathbf{I} = \begin{bmatrix} i \sin \theta & -\sin \theta \\ \sin \theta & i \sin \theta \end{bmatrix} = \sin \theta \begin{bmatrix} i & -1 \\ 1 & i \end{bmatrix} \quad (4.68)$$

has a null space spanned by

$$\bar{\mathbf{u}}' = \begin{bmatrix} 1 \\ i \end{bmatrix}. \quad (4.69)$$

Normalising  $\mathbf{u}'$  and  $\bar{\mathbf{u}}'$  and placing them into a column matrix, it is found that

$$\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -i/\sqrt{2} & i/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -i/\sqrt{2} & i/\sqrt{2} \end{bmatrix} \begin{bmatrix} \cos \theta + i \sin \theta & 0 \\ 0 & \cos \theta - i \sin \theta \end{bmatrix}. \quad (4.70)$$

By Euler's formula,

$$\begin{bmatrix} \cos \theta + i \sin \theta & 0 \\ 0 & \cos \theta - i \sin \theta \end{bmatrix} = \begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{bmatrix}. \quad (4.71)$$

so

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{bmatrix}. \quad (4.72)$$

or

$$\mathbf{R}_2' \mathbf{U}_2 = \mathbf{U}_2 \mathbf{M}_2 \quad (4.73)$$

where

$$\mathbf{M}_2 = \begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{bmatrix} \quad (4.74)$$

and

$$\mathbf{U}_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}. \quad (4.75)$$

Therefore, the full eigenvalue decomposition of  $\mathbf{R}'$  is

$$\mathbf{R}' \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & \cdots & 0 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & \cdots & 0 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} e^{i\theta} & 0 & 0 & \cdots & 0 \\ 0 & e^{-i\theta} & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \quad (4.76)$$

or

$$\mathbf{R}' \mathbf{U}' = \mathbf{U}' \mathbf{M} \quad (4.77)$$

where



$$\mathbf{U}' = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 & \cdots & 0 \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \quad (4.78)$$

is a unitary matrix, and

$$\mathbf{M} = \begin{bmatrix} e^{i\theta} & 0 & 0 & \cdots & 0 \\ 0 & e^{-i\theta} & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \quad (4.79)$$

Postmultiplying equation (4.77) by  $\mathbf{U}'^{-1}$  which is  $\mathbf{U}'^H$ , the complex conjugate transpose of  $\mathbf{U}'$ , gives

$$\mathbf{R}' = \mathbf{U}'\mathbf{M}\mathbf{U}'^H \quad (4.80)$$

If equation (4.80) is substituted into equation (4.61), then

$$\mathbf{R} = \mathbf{Q}\mathbf{U}'\mathbf{M}\mathbf{U}'^H\mathbf{Q}^H \quad (4.81)$$

( $\mathbf{Q}^H = \mathbf{Q}^T$  for real  $\mathbf{Q}$ .) or

$$\mathbf{R} = \mathbf{U}\mathbf{M}\mathbf{U}^H \quad (4.82)$$

where

$$\mathbf{U} = \mathbf{Q}\mathbf{U}', \quad (4.83)$$

a unitary matrix, being a product of unitary matrices. By this equation, the columns of  $\mathbf{U}$  which are the eigenvectors of  $\mathbf{R}$  are

$$\mathbf{U} = \begin{bmatrix} | & | & | & \cdots & | \\ | & | & | & \cdots & | \\ \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{q}_3 & \cdots & \mathbf{q}_N \\ | & | & | & \cdots & | \\ | & | & | & \cdots & | \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 & \cdots & 0 \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} = \begin{bmatrix} | & | & | & \cdots & | \\ | & | & | & \cdots & | \\ \frac{\mathbf{q}_1 - i\mathbf{q}_2}{\sqrt{2}} & \frac{\mathbf{q}_1 + i\mathbf{q}_2}{\sqrt{2}} & \mathbf{q}_3 & \cdots & \mathbf{q}_N \\ | & | & | & \cdots & | \\ | & | & | & \cdots & | \end{bmatrix} \quad (4.84)$$

Since all  $\mathbf{q}_i$  are real vectors, the first two columns of  $\mathbf{U}$  are clearly complex conjugates.

Therefore, the first two columns may be labelled  $\mathbf{u}$  and  $\bar{\mathbf{u}}$  so that

$$\mathbf{U} = \begin{bmatrix} | & | & | & \cdots & | \\ | & | & | & \cdots & | \\ \mathbf{u} & \bar{\mathbf{u}} & \mathbf{u}_3 & \cdots & \mathbf{u}_N \\ | & | & | & \cdots & | \\ | & | & | & \cdots & | \end{bmatrix} \quad (4.85)$$

where

$$\mathbf{u} = \frac{\mathbf{q}_1 - i\mathbf{q}_2}{\sqrt{2}} \quad (4.86)$$

and  $\mathbf{u}_i = \mathbf{q}_i$ ,  $3 \leq i \leq N$ .

Now, recall from the definition of an RSRS in equation (4.2) that

$$\mathbf{e}_n = \lambda^n \mathbf{R}^n \mathbf{e} \quad (4.87)$$

where  $\mathbf{R}$  is a rotator, non-zero  $\mathbf{e}$  exists solely in the plane of rotation of  $\mathbf{R}$ ,  $\lambda$  is real and non-zero, and  $\lambda \mathbf{R} \neq \mathbf{I}$ . Therefore, by equation (4.82),

$$\mathbf{e}_n = \lambda^n (\mathbf{U}\mathbf{M}\mathbf{U}^H)^n \mathbf{e} \quad (4.88)$$

or

$$\mathbf{e}_n = \lambda^n \underbrace{\mathbf{U}\mathbf{M}\mathbf{U}^H \mathbf{U}\mathbf{M}\mathbf{U}^H \dots \mathbf{U}\mathbf{M}\mathbf{U}^H}_{n \text{ times}} \mathbf{e} \quad (4.89)$$

which, as  $\mathbf{U}^H\mathbf{U}$  is  $\mathbf{I}$ , gives

$$\mathbf{e}_n = \lambda^n \mathbf{U}\mathbf{M}^n \mathbf{U}^H \mathbf{e}. \quad (4.90)$$

Let

$$\mathbf{c} = \mathbf{U}^H \mathbf{e} \quad (4.91)$$

Then  $\mathbf{c}$  is the co-ordinate vector of  $\mathbf{e}$  in the basis of the columns of  $\mathbf{U}$  because premultiplying (4.91) by  $\mathbf{U}$  gives

$$\mathbf{U}\mathbf{c} = \mathbf{e} \quad (4.92)$$

or

$$\begin{bmatrix} | & | & | & & | \\ | & | & | & & | \\ | & | & | & & | \\ \mathbf{u} & \mathbf{u} & \mathbf{u}_3 & \dots & \mathbf{u}_N \\ | & | & | & & | \\ | & | & | & & | \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ \vdots \\ c_N \end{bmatrix} = \mathbf{e}. \quad (4.93)$$

Substituting equation (4.91) into equation (4.90) gives

$$\mathbf{e}_n = \lambda^n \mathbf{U}\mathbf{M}^n \mathbf{c}. \quad (4.94)$$

Since  $\mathbf{e}$  exists solely in the plane of rotation and is non-zero, it must be a linear combination of  $\mathbf{u}$  and  $\bar{\mathbf{u}}$  only. Therefore, the first two coefficients of  $\mathbf{c}$ ,  $c_1$  and  $c_2$ , are non-zero and  $c_3$  through  $c_N$  are zero. The fact that  $\mathbf{e}$  is real means that the coefficients  $c_1$  and  $c_2$  must, themselves, be complex conjugates. Therefore,  $c_1$  and  $c_2$  are renamed  $c$  and  $\bar{c}$  and

$$\mathbf{e}_n = \lambda^n \mathbf{U} \begin{bmatrix} e^{in\theta} & 0 & 0 & \dots & 0 \\ 0 & e^{-in\theta} & 0 & \dots & 0 \\ 0 & 0 & 1 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} c \\ \bar{c} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (4.95)$$

or

$$\mathbf{e}_n = \lambda^n \begin{bmatrix} | & | & | & & | \\ | & | & | & & | \\ \mathbf{u} & \bar{\mathbf{u}} & \mathbf{u}_3 & \dots & \mathbf{u}_N \\ | & | & | & & | \\ | & | & | & & | \end{bmatrix} \begin{bmatrix} ce^{in\theta} \\ \bar{c}e^{-in\theta} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (4.96)$$

which is

$$\mathbf{e}_n = \lambda^n ce^{in\theta} \mathbf{u} + \lambda^n \bar{c} e^{-in\theta} \bar{\mathbf{u}} \quad (4.97)$$

or

$$\mathbf{e}_n = z^n \mathbf{w} + \bar{z}^n \bar{\mathbf{w}} \quad (4.98)$$

where

$$z = \lambda e^{i\theta}, \quad (4.99)$$

and

$$\mathbf{w} = c\mathbf{u}. \quad (4.100)$$

As with equation (4.50), it is necessary to apply the restrictions that  $z \neq 0$  and  $z \neq 1$  and that  $\mathbf{w} \neq \mathbf{0}$ . Also, since the formula for  $\mathbf{w}$  in this derivation, beginning with the definitions of  $\mathbf{q}_1$  and  $\mathbf{q}_2$  and following on through equations (4.84) and (4.85), satisfies the assumptions leading to equation (4.43),  $\mathbf{w}$  is orthogonal to  $\overline{\mathbf{w}}$ .

### 4.3 The Kernel of the 1<sup>st</sup> Order VEA, Real Part, Complex Form

To recapitulate, it has been established, by two different routes, that the original definition of an RSRS given in equation (4.2),

$$\mathbf{e}_n = \lambda^n \mathbf{R}^n \mathbf{e}, \quad \text{for } n = 0, 1, 2, \dots \quad (4.101)$$

where  $\lambda$  is a real scalar,  $\mathbf{R}$  is a real  $N \times N$  rotation matrix,  $\mathbf{e}$  is a non-zero real vector existing solely within the plane of rotation of  $\mathbf{R}$ , with  $\lambda > 0$ , and  $\lambda \mathbf{R} \neq \mathbf{I}$ , the identity matrix, is equivalent to

$$\mathbf{e}_n = z^n \mathbf{w} + \overline{z}^n \overline{\mathbf{w}}, \quad \text{for } n = 0, 1, 2, \dots \quad (4.102)$$

where the complex scalar  $z \neq 0$ ,  $z \neq 1$ , and  $\mathbf{w}$  is non-zero and orthogonal to its complex conjugate. The equivalence of (4.101) to (4.102) means that an alternative form of the necessary and sufficient condition for the success of the 1<sup>st</sup> order VEA on real vector

sequences in equation (3.1), which is also the original definition of a  $\lambda\mathbf{R}$  sequence, can be given. By equations (4.101), (4.102) and (3.1), it has been shown that, given a vector sequence  $\{\mathbf{x}_n\}$  whose terms are real and  $N$ -dimensional, then  $\varepsilon_2^{(n)} = \mathbf{x}$  for  $n = 0, 1, \dots$  if and only if

$$\mathbf{x}_n = \mathbf{x} + z^n \mathbf{w} + \overline{z}^n \overline{\mathbf{w}}, \quad n = 0, 1, \dots \quad (4.103)$$

where  $z$  is a complex scalar,  $z \neq 0$ ,  $z \neq 1$ , and  $\mathbf{w}$  is a non-zero complex vector which is orthogonal to its complex conjugate. This new definition fulfils the expectations that emerged at the start of this chapter that there must be an alternative form for the kernel of the 1<sup>st</sup> order VEA that makes no reference to rotation matrices. This definition will strongly resemble the complex kernel for the 1<sup>st</sup> order VEA to be given in Chapter 7.

## 5. EXPERIMENTATION TOWARD THE KERNEL OF THE VEA

The focus of the work, up to this point, has been only the 1<sup>st</sup> order VEA. For this chapter and the next, the picture is broadened to include higher orders of the VEA. In addition, the methods employed are empirical: careful observation of what order of VEA is necessary to create a constant transform sequence from a variety of carefully constructed original sequences all generated in the same manner. The object of the experiments is to look for vector sequences where the minimum order of VEA transform necessary to give a constant sequence is lower than the sufficient order suggested by McLeod and Graves-Morris and reported in Theorem 2.6., *i.e.* vector sequences which do not satisfy the equation (2.32) when  $k$  is the order of VEA used. The intention for these experiments is that they would lead to the discovery of conditions that are both sufficient and necessary for membership of a vector sequence in the kernel of the VEA.

Throughout this chapter and the next, reference will be made to the VEA order of a sequence. By this is meant the lowest order of VEA which is capable of yielding a transform sequence all of whose terms are the same vector. If no order of the VEA is capable of making such a transformation, then the sequence will be said to have no VEA order.

The vector sequences used in these experiments are all generated by defining an  $N \times N$  matrix  $\mathbf{A}$ , and  $N$ -dimensional vectors  $\mathbf{b}$  and  $\mathbf{x}_0$ , and then generating the sequence  $\{\mathbf{x}_n\}$  to satisfy the formula

$$\mathbf{x}_n = \mathbf{A}\mathbf{x}_{n-1} + \mathbf{b}. \quad (5.1)$$

One reason for choosing this method of sequence generation is that, though none of them is practically implemented as such, all of the classic iterative techniques for the solution of large systems of linear equations generate vector sequences which satisfy equation (5.1). These include Jacobi iteration, Gauss-Seidel iteration, Jacobi underrelaxation, successive over-relaxation and Richardson iteration. Because of this, such sequences have already received much attention which has resulted in useful extensions of the McLeod – Graves-Morris condition. These extensions by Gekeler and by Brezinski have shown that, under certain mild conditions<sup>4</sup>, the sequence will have a VEA order and the vector of the constant transform sequence will be the limit or antilimit vector  $\mathbf{x}$  satisfying

$$\mathbf{x} = \mathbf{Ax} + \mathbf{b} . \tag{5.2}$$

One consequence of this is that the VEA is able to convert iterative solution methods that rely upon a matrix  $\mathbf{A}$  having a spectral radius less than unity into direct solution methods with no such spectral constraint.

In this chapter, some current theory pertaining to such vector sequences is introduced and explained somewhat in order that the experiments and their results may be understood in their proper context. In addition, the results of the an initial group of experiments are given which demonstrate that the VEA order of certain sequences is lower than the upper limit given in the current theory. A subsequent group of experiments designed to determine precisely when this is so is the subject of Chapter 6.

---

<sup>4</sup> By a mild condition is meant the disallowance of some parameter taking one of a finite number of values.



Section 5.1 introduces some important background pertaining to the iterative solution of systems of linear equations in general and these numerical experiments in particular. Following this, §5.2 presents and illustrates some theoretical results by Gekeler and Brezinski regarding an upper bound on the VEA order of a vector sequence generated by the iterative scheme of equation (5.1). Then, the results of the initial experiments are described in §5.3. They indicate that the VEA order of a vector sequence whose generating matrix  $\mathbf{A}$  has eigenvalues which are complex conjugates and corresponding eigenvectors and generalised eigenvectors that are mutually orthogonal is lower than the upper bounds given by Brezinski. However, it is not determined until the experiments of the next chapter which of the orthogonality relationships present between the eigenvectors and generalised eigenvectors of  $\mathbf{A}$  are critical to this reduction in the VEA order of the sequence below the upper bound.

## 5.1 Background Material

Under certain mild conditions, the sequence generated by equation (5.1) will either converge toward or diverge from some vector  $\mathbf{x}$  satisfying both equation (5.2) and

$$(\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{b}. \quad (5.3)$$

If  $\mathbf{I} - \mathbf{A}$  is non-singular, then equation (5.3) will have a unique solution and  $\{\mathbf{x}_n\}$  will have a limit or antilimit. If  $\mathbf{I} - \mathbf{A}$  is singular, then as long as  $\mathbf{b}$  is a linear combination of the columns of  $\mathbf{I} - \mathbf{A}$  (i.e. in the range or column space of  $\mathbf{I} - \mathbf{A}$ ), equation (5.3) will still have a solution, though there may or may not exist a limit or antilimit of  $\{\mathbf{x}_n\}$ .

If equation (5.2) is subtracted from equation (5.1), then

$$\mathbf{x}_n - \mathbf{x} = \mathbf{A}(\mathbf{x}_{n-1} - \mathbf{x}) \quad (5.4)$$

which is

$$\mathbf{e}_n = \mathbf{A}\mathbf{e}_{n-1} \quad (5.5)$$

where

$$\mathbf{e}_n = \mathbf{x}_n - \mathbf{x}. \quad (5.6)$$

This  $\mathbf{e}_n$  may be thought of as the error present in  $\mathbf{x}_n$ . By induction on equation (5.5),

$$\mathbf{e}_n = \mathbf{A}^n \mathbf{e}_0. \quad (5.7)$$

Plainly, the nature of the sequence  $\{\mathbf{e}_n\}$  depends completely upon  $\mathbf{A}$  and  $\mathbf{e}_0$ . By equation (5.6), however,  $\{\mathbf{e}_n\}$  is just a translation of the sequence  $\{\mathbf{x}_n\}$ . This has a clear implication for the experiments. Since, by Theorem 2.5, the VEA order of all translations of a given vector sequence are the same, the investigation may be confined to the sequences  $\{\mathbf{x}_n\}$  whose (anti)limit is  $\mathbf{0}$ , which is to say, sequences  $\{\mathbf{x}_n\}$  generated by equation (5.1) for which  $\mathbf{b} = \mathbf{0}$  or

$$\mathbf{x}_n = \mathbf{A}\mathbf{x}_{n-1} = \mathbf{A}^n \mathbf{x}_0. \quad (5.8)$$

These are convenient sequences to analyse because the error present in each term of such a sequence is simply the term, itself. To emphasise these aspects, these sequences will be denoted by  $\{\mathbf{e}_n\}$  instead of  $\{\mathbf{x}_n\}$ .

Now, consider what happens when the matrix  $\mathbf{A}^n$  premultiplies some vector  $\mathbf{e}_0$ . If  $\mathbf{J}$ , a Jordan canonical form of  $\mathbf{A}$ , and  $\mathbf{V}$ , the corresponding matrix of eigenvectors and generalised eigenvectors, are computed, then by definition

$$\mathbf{A} = \mathbf{V}\mathbf{J}\mathbf{V}^{-1}. \quad (5.9)$$

(See Appendix B for a short introduction to this form.) Therefore,

$$\mathbf{A}^n \mathbf{e}_0 = (\mathbf{V}\mathbf{J}\mathbf{V}^{-1})^n \mathbf{e}_0 \quad (5.10)$$

or

$$\mathbf{A}^n \mathbf{e}_0 = \underbrace{\mathbf{V}\mathbf{J}\mathbf{V}^{-1}\mathbf{V}\mathbf{J}\mathbf{V}^{-1}\dots\mathbf{V}\mathbf{J}\mathbf{V}^{-1}}_{n \text{ times}} \mathbf{e}_0. \quad (5.11)$$

Notice that every occurrence of  $\mathbf{V}^{-1}\mathbf{V}$  becomes the identity matrix  $\mathbf{I}$  and disappears from the expression to leave

$$\mathbf{A}^n \mathbf{e}_0 = \mathbf{V}\mathbf{J}^n \mathbf{V}^{-1} \mathbf{e}_0. \quad (5.12)$$

Let

$$\mathbf{c} = \mathbf{V}^{-1} \mathbf{e}_0. \quad (5.13)$$

Substituting equation (5.13) into equation (5.12) gives

$$\mathbf{A}^n \mathbf{e}_0 = \mathbf{V}\mathbf{J}^n \mathbf{c} \quad (5.14)$$

which, by equation (5.7), is just

$$\mathbf{e}_n = \mathbf{V}\mathbf{J}^n\mathbf{c}. \quad (5.15)$$

The meaning of equation (5.15) can be made clearer with the introduction of partitions of  $\mathbf{V}$ ,  $\mathbf{J}$  and  $\mathbf{c}$  according to the dimensions of Jordan blocks  $\mathbf{J}_i$ ,  $i = 1, 2, \dots, r$ . Thus,

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_1 & & & & \\ & \mathbf{J}_2 & & & \\ & & \mathbf{J}_3 & & \\ & & & \ddots & \\ & & & & \mathbf{J}_r \end{bmatrix}. \quad (5.16)$$

where, for each  $i$ ,

$$\mathbf{J}_i = \begin{bmatrix} \mu_i & 1 & 0 & \cdots & 0 & 0 \\ 0 & \mu_i & 1 & \ddots & 0 & 0 \\ 0 & 0 & \mu_i & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & \mu_i & 1 \\ 0 & 0 & 0 & \cdots & 0 & \mu_i \end{bmatrix}, \quad (5.17)$$

and each  $\mathbf{J}_i$  is an  $N_i \times N_i$  block. In like manner,  $\mathbf{V}$  is partitioned into column blocks  $\mathbf{V}_i$ ,  $i = 1, 2, \dots, r$  so that

$$\mathbf{V} = \begin{bmatrix} | & | & | & \cdots & | \\ | & | & | & & | \\ \mathbf{V}_1 & \mathbf{V}_2 & \mathbf{V}_3 & \cdots & \mathbf{V}_r \\ | & | & | & & | \\ | & | & | & & | \end{bmatrix} \quad (5.18)$$

and for each  $i$

$$\mathbf{V}_i = \begin{bmatrix} | & | & | & \cdots & | \\ \mathbf{v}_{i,1} & \mathbf{v}_{i,2} & \mathbf{v}_{i,3} & \cdots & \mathbf{v}_{i,N_i} \\ | & | & | & \cdots & | \end{bmatrix}. \quad (5.19)$$

Finally,  $\mathbf{c}$  is partitioned into segments  $\mathbf{c}_i$ ,  $i = 1, 2, \dots, r$  so that

$$\mathbf{c} = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \vdots \\ \mathbf{c}_r \end{bmatrix} \quad (5.20)$$

and the entries in each  $\mathbf{c}_i$  are designated

$$\mathbf{c}_i = \begin{bmatrix} c_{i,1} \\ c_{i,2} \\ \vdots \\ c_{i,N_i} \end{bmatrix}. \quad (5.21)$$

Now, premultiplying equation (5.13) by  $\mathbf{V}$ , the result is

$$\mathbf{Vc} = \mathbf{e}_0 \quad (5.22)$$

which is

$$\begin{bmatrix} | & & | & & | & & | & & | \\ | & & | & & | & & | & & | \\ | & & | & & | & & | & & | \\ \mathbf{v}_{1,1} & \cdots & \mathbf{v}_{1,N_1} & & \mathbf{v}_{2,1} & \cdots & \mathbf{v}_{2,N_2} & \cdots & \mathbf{v}_{r,1} & \cdots & \mathbf{v}_{r,N_r} \\ | & & | & & | & & | & & | \\ | & & | & & | & & | & & | \\ | & & | & & | & & | & & | \\ | & & | & & | & & | & & | \\ | & & | & & | & & | & & | \\ | & & | & & | & & | & & | \\ | & & | & & | & & | & & | \end{bmatrix} \begin{bmatrix} c_{1,1} \\ \vdots \\ c_{1,N_1} \\ \cdots \\ c_{2,1} \\ \vdots \\ c_{2,N_2} \\ \cdots \\ c_{r,1} \\ \vdots \\ c_{r,N_r} \end{bmatrix} = \mathbf{e}_0. \quad (5.23)$$

Thus,  $\mathbf{c}$  is the vector containing the coefficients in the linear combination of eigenvectors and generalised eigenvectors that make up  $\mathbf{e}_0$ . It represents a decomposition of  $\mathbf{e}_0$  into components in the columns of  $\mathbf{V}$ . This  $\mathbf{c}$  is sometimes called the co-ordinate vector of  $\mathbf{e}_0$  in the eigenbasis of  $\mathbf{A}$ .

If equation (5.15) is premultiplied by  $\mathbf{V}^{-1}$ , the result is a related sequence

$$\mathbf{V}^{-1}\mathbf{e}_n = \mathbf{J}^n\mathbf{c}. \quad (5.24)$$

If  $\mathbf{V}^{-1}$  is a unitary matrix (*i.e.* if  $\mathbf{V}$  is a unitary matrix), then this related sequence  $\{\mathbf{V}^{-1}\mathbf{e}_n\}$  is merely the sequence  $\{\mathbf{e}_n\}$  after it has undergone a unitary transformation. It will be of the class of sequences described by equation (2.29) in Theorem 2.5. Therefore, if the  $k^{\text{th}}$  order VEA applied to  $\{\mathbf{e}_n\}$  gives the zero sequence (the sequence all of whose terms are  $\mathbf{0}$ ), the zero vector being the (anti)limit of  $\{\mathbf{e}_n\}$ , then the  $k^{\text{th}}$  order VEA applied to  $\{\mathbf{V}^{-1}\mathbf{e}_n\}$  will give  $\{\mathbf{V}^{-1}\mathbf{0}\}$  which is also the zero sequence. This means that for each  $\mathbf{A}$  having mutually orthogonal unit-length eigenvectors and generalised eigenvectors, all sequences  $\{\mathbf{x}_n\}$  satisfying

$$\mathbf{x}_n = \mathbf{A}\mathbf{x}_{n-1} + \mathbf{b}, \quad n = 1, 2, \dots \quad (5.25)$$

can be fully investigated by examining only those sequences  $\{\mathbf{e}_n\}$  generated by

$$\mathbf{e}_n = \mathbf{J}^n \mathbf{c}, \quad n = 0, 1, \dots, \quad (5.26)$$

a much more transparent set of sequences with which to work. For the sake of understanding those matrices  $\mathbf{A}$  not having orthonormal eigenvectors and generalised eigenvectors, the experiments have included sequences satisfying equation (5.15) or

$$\mathbf{e}_n = \mathbf{V}\mathbf{J}^n \mathbf{c}, \quad n = 0, 1, \dots \quad (5.27)$$

where  $\mathbf{V}$  is not unitary, i.e. does not have mutually orthogonal columns of unit 2-norm. The contrast between the sets of results for these two cases provides the motivation for the experiments of the next chapter.

## 5.2 Published Theoretical Results

Theorem 2.6 gave the sufficient condition of McLeod [18] and Graves-Morris [13] for membership in the kernel of the  $k^{\text{th}}$  order VEA. They said that if there exist scalars  $a_i, i = 0, 1, \dots, k$ , and vector  $\mathbf{x}$  such that  $\sum a_i \neq 0, a_k \neq 0$ , and

$$\sum_{i=0}^k a_i (\mathbf{x}_{n+i} - \mathbf{x}) = \mathbf{0}, \quad n = 0, 1, \dots, \quad (5.28)$$

then  $\varepsilon_{2k}^{(n)} = \mathbf{x}$  for every  $n$ . Since these conditions are not necessary, there may exist sequences whose VEA order is  $k$  but do not satisfy them as well as sequences which do satisfy them but have lower VEA order than  $k$  of the conditions. (An example of this

was seen in the  $\lambda\mathbf{R}$  sequences of Chapter 3.) Hence, Theorem 2.6 gives an upper bound on the VEA order of any sequence satisfying equation (5.28).

In 1972, Gekeler [12] used this sufficient condition to prove

Theorem 5.1: If  $\{\mathbf{x}_n\}$  is a vector sequence generated by equation (5.1) with  $\mathbf{I} - \mathbf{A}$  non-singular and  $m$  is the degree of the minimal polynomial of  $\mathbf{A}$  for the vector  $\mathbf{x}_0 - \mathbf{x}$  where  $\mathbf{x}$  is the unique solution of the linear system of  $\mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{b}$ , then

$$\varepsilon_{2m}^{(n)} = \mathbf{x}, \quad n = 0, 1, \dots \quad (5.29)$$

By equation (5.6),  $\mathbf{x}_0 - \mathbf{x}$  is simply  $\mathbf{e}_0$ . Also, the minimal polynomial of a matrix  $\mathbf{A}$  for a vector  $\mathbf{w}$  is the monic<sup>5</sup> polynomial

$$p(\mu) = \mu^m + c_{m-1}\mu^{m-1} + c_{m-2}\mu^{m-2} + \dots + c_1\mu + c_0 \quad (5.30)$$

of lowest degree  $m$  such that

$$p(\mathbf{A})\mathbf{w} = (\mathbf{A}^m + c_{m-1}\mathbf{A}^{m-1} + c_{m-2}\mathbf{A}^{m-2} + \dots + c_1\mathbf{A} + c_0\mathbf{I})\mathbf{w} = \mathbf{0}. \quad (5.31)$$

It will be denoted  $p_{\min, \mathbf{A}, \mathbf{w}}(\mu)$ . (See Appendix C for a presentation of the minimal polynomial of a matrix for a vector.) This polynomial may be alternatively defined as follows. Let  $\nu_l$ ,  $l = 1, 2, \dots, q$  be the  $q$  distinct eigenvalues of  $\mathbf{A}$ , let  $c_{ij}$  be the coefficients defined in equation (5.21) and let  $K_l$  equal the highest index  $j$  for which  $c_{ij} \neq 0$  for every  $i$  whose corresponding Jordan block  $\mathbf{J}_i$  has eigenvalue  $\mu_i = \nu_l$ . Then, the minimal polynomial of  $\mathbf{A}$  for the vector  $\mathbf{w}$  may be defined as

---

<sup>5</sup> A monic polynomial is one where the coefficient for the term of highest degree is unity.



$$p_{\min, \mathbf{A}, \mathbf{w}}(\mu) = (\mu - \nu_1)^{K_1} (\mu - \nu_2)^{K_2} \dots (\mu - \nu_q)^{K_q} \quad (5.32)$$

with the convention that  $K_j = 0$  if  $c_{i,j} = 0$  for every  $j$  when  $\mu_i = \nu_j$ . In other words, for each distinct eigenvalue  $\nu_l$  of  $\mathbf{A}$ , of all the column vectors  $\mathbf{v}_{i,j}$  in all of the column blocks  $\mathbf{V}_i$  corresponding to  $\mu_i = \nu_l$ , the highest  $j$  for which some of vector  $\mathbf{v}_{i,j}$  is present as a component in  $\mathbf{w}$  is the power to which  $(\mu - \nu_k)$  is raised in  $p_{\min, \mathbf{A}, \mathbf{w}}(\mu)$ . If, for some  $\nu_l$  and every  $\mu_i = \nu_l$ , none of the  $\mathbf{v}_{i,j}, j = 1, 2, \dots, N_i$ , is present in  $\mathbf{w}$ , then  $(\mu - \nu_l)$  will not be a factor of  $p_{\min, \mathbf{A}, \mathbf{w}}(\mu)$ .

This theorem provided the first extension of the McLeod – Graves-Morris sufficient condition to vector sequences satisfying equation (5.1), providing an upper bound on the VEA order of such sequences. Appendix D gives the brief and enlightening proof of this theorem by Gekeler.

The appropriateness of the minimal polynomial of a matrix for a vector as a tool to indicate the complexity of a vector sequence is apparent when considering the following example. It is given using  $\mathbf{V}$ ,  $\mathbf{J}$  and  $\mathbf{c}$  in the form of equation (5.15) rather than using  $\mathbf{A}$  and  $\mathbf{e}_0$  in the form of equation (5.7).

$$\mathbf{e}_n = \left[ \begin{array}{cccc|cccc} | & | & | & | & | & | & | & | \\ | & | & | & | & | & | & | & | \\ | & | & | & | & | & | & | & | \\ \mathbf{v}_{1,1} & \mathbf{v}_{1,2} & \mathbf{v}_{1,3} & \mathbf{v}_{1,4} & \mathbf{v}_{2,1} & \mathbf{v}_{2,2} & \mathbf{v}_{2,3} & \\ | & | & | & | & | & | & | & | \\ | & | & | & | & | & | & | & | \end{array} \right] \left[ \begin{array}{cccc|ccc} \mu_1 & 1 & 0 & 0 & & & \\ 0 & \mu_1 & 1 & 0 & & & \\ 0 & 0 & \mu_1 & 1 & & & \\ 0 & 0 & 0 & \mu_1 & & & \\ \hline & & & & \mu_2 & 1 & 0 \\ & & & & 0 & \mu_2 & 1 \\ & & & & 0 & 0 & \mu_2 \end{array} \right] \left[ \begin{array}{c} 0 \\ c_{1,2} \\ c_{1,3} \\ 0 \\ \hline c_{2,1} \\ 0 \\ 0 \end{array} \right] \quad (5.33)$$

Here,  $\mathbf{J}$  is  $7 \times 7$  and consists of 2 Jordan blocks, but it multiplies a vector  $\mathbf{c}$  having some zero coefficients. Only nonzero coefficients in  $\mathbf{c}$  have been labelled. The matrix  $\mathbf{J}$  raised to the power  $n$  becomes

$$\mathbf{J}^n = \begin{bmatrix} \mu_1^n & \binom{n}{1}\mu_1^{n-1} & \binom{n}{2}\mu_1^{n-2} & \binom{n}{3}\mu_1^{n-3} & & & \\ 0 & \mu_1^n & \binom{n}{1}\mu_1^{n-1} & \binom{n}{2}\mu_1^{n-2} & & & \\ 0 & 0 & \mu_1^n & \binom{n}{1}\mu_1^{n-1} & & & \\ 0 & 0 & 0 & \mu_1^n & & & \\ \hline & & & & \mu_2^n & \binom{n}{1}\mu_2^{n-1} & \binom{n}{2}\mu_2^{n-2} \\ & & & & 0 & \mu_2^n & \binom{n}{1}\mu_2^{n-1} \\ & & & & 0 & 0 & \mu_2^n \end{bmatrix}, \quad (5.34)$$

where  $\binom{i}{j}$  is the binomial coefficient whose formula is

$$\binom{i}{j} = {}_i C_j = \frac{i!}{j!(i-j)!} \quad (5.35)$$

with the convention that  $\binom{i}{j} = 0$  if  $j > i$ . Therefore, substituting equation (5.34) into equation (5.33) and completing the multiplication  $\mathbf{J}^n \mathbf{c}$ ,  $\mathbf{e}_n$  can be written as

$$\mathbf{e}_n = \begin{bmatrix} | & | & | & | & | & | & | \\ | & | & | & | & | & | & | \\ | & | & | & | & | & | & | \\ \mathbf{v}_{1,1} & \mathbf{v}_{1,2} & \mathbf{v}_{1,3} & \mathbf{v}_{1,4} & \mathbf{v}_{2,1} & \mathbf{v}_{2,2} & \mathbf{v}_{2,3} \\ | & | & | & | & | & | & | \\ | & | & | & | & | & | & | \\ | & | & | & | & | & | & | \end{bmatrix} \begin{bmatrix} c_{1,2}\binom{n}{1}\mu_1^{n-1} + c_{1,3}\binom{n}{2}\mu_1^{n-2} \\ c_{1,2}\mu_1^n + c_{1,3}\binom{n}{1}\mu_1^{n-1} \\ c_{1,3}\mu_1^n \\ 0 \\ \hline c_{2,1}\mu_2^n \\ 0 \\ 0 \end{bmatrix} \quad (5.36)$$

or

$$\mathbf{e}_n = [c_{1,2} \binom{n}{1} \mu_1^{n-1} + c_{1,3} \binom{n}{2} \mu_1^{n-2}] \mathbf{v}_{1,1} + [c_{1,2} \mu_1^n + c_{1,3} \binom{n}{1} \mu_1^{n-1}] \mathbf{v}_{1,2} + [c_{1,3} \mu_1^n] \mathbf{v}_{1,3} + [c_{2,1} \mu_2^n] \mathbf{v}_{2,1} \quad (5.37)$$

which is also

$$\mathbf{e}_n = \mu_1^n [c_{1,2} \mathbf{v}_{1,2} + c_{1,3} \mathbf{v}_{1,3}] + \binom{n}{1} \mu_1^{n-1} [c_{1,2} \mathbf{v}_{1,1} + c_{1,3} \mathbf{v}_{1,2}] + \binom{n}{2} \mu_1^{n-2} c_{1,3} \mathbf{v}_{1,1} + \mu_2^n c_{2,1} \mathbf{v}_{2,1} \quad (5.38)$$

Equation (5.37) shows four distinct coefficients multiplying four different vectors. Equation (5.38) shows four distinct monomials multiplying four different combinations of vectors. Whichever way it is distributed, there are four fundamentally different component vector functions making up  $\mathbf{e}_n$ . In fact, this very same error sequence is generated by

$$\mathbf{e}_n = \begin{bmatrix} | & | & | & | \\ | & | & | & | \\ \mathbf{v}_{1,1} & \mathbf{v}_{1,2} & \mathbf{v}_{1,3} & \mathbf{v}_{2,1} \\ | & | & | & | \\ | & | & | & | \end{bmatrix} \begin{bmatrix} \mu_1 & 1 & 0 & \dots \\ 0 & \mu_1 & 1 & \dots \\ 0 & 0 & \mu_1 & \dots \\ \dots & \dots & \dots & \mu_2 \end{bmatrix}^n \begin{bmatrix} 0 \\ c_{1,2} \\ \frac{c_{1,3}}{\mu_1} \\ c_{2,1} \end{bmatrix}, \quad (5.39)$$

although the first matrix on the right-hand side of this expression is now  $7 \times 4$  and no longer square.

The fourth order character of the behaviour of this sequence is detected by the minimal polynomial of  $\mathbf{A}$  for the vector  $\mathbf{x}_0$  which (see Appendix C) is the same as the minimal polynomial of  $\mathbf{J}$  for the vector  $\mathbf{c}$ . With two distinct eigenvalues,  $\nu_1 = \mu_1$  and  $\nu_2 = \mu_2$ , each corresponds to only one Jordan block. The highest  $j$  for which  $c_{1j} \neq 0$  is

$j = 3$ , so  $K_1 = 3$ . The highest  $j$  for which  $c_{2,j} \neq 0$  is  $j = 1$ , so  $K_2 = 1$ . Thus, the minimal polynomial of  $\mathbf{J}$  for the vector  $\mathbf{c}$  in this example is

$$p_{\min, \mathbf{A}, \mathbf{e}_0}(\mu) = (\mu - \mu_1)^3 (\mu - \mu_2) \quad (5.40)$$

and the degree of  $p_{\min, \mathbf{A}, \mathbf{e}_0}(\mu)$  is 4.

Brezinski [3] refined Gekeler's result supplying

Theorem 5.2: Given a vector sequence  $\{\mathbf{x}_n\}$  generated by equation (5.1) with  $\mathbf{I} - \mathbf{A}$  non-singular, if  $m$  is the degree of the minimal polynomial of  $\mathbf{A}$  for the vector  $\mathbf{x}_0 - \mathbf{x}$  where  $\mathbf{x}$  is the unique solution of the linear system of equation (5.2) and  $k$  is the number of zeroes of the same polynomial that are equal to 0, then, for  $0 \leq p \leq k$ ,

$$\mathbf{e}_{2(m-p)}^{(n+p)} = \mathbf{x}, \quad n = 0, 1, \dots \quad (5.41)$$

The minimal polynomial of  $\mathbf{A}$  for the vector  $\mathbf{x}_0 - \mathbf{x}$  having  $k$  roots of zero means that present in  $\mathbf{x}_0 - \mathbf{x}$  is some part of the  $k^{\text{th}}$  vector of some column block associated with a Jordan block having zero as an eigenvalue. If  $\mathbf{v}_{i,k}$  is that generalised eigenvector, then (equation (B.12) with  $\mu_i = 0$ )

$$\mathbf{A} \mathbf{v}_{i,k} = \mathbf{v}_{i,k-1}. \quad (5.42)$$

Since, by equation (5.4),

$$\mathbf{x}_1 - \mathbf{x} = \mathbf{A}(\mathbf{x}_0 - \mathbf{x}), \quad (5.43)$$

that component of  $\mathbf{v}_{i,k}$  present in  $\mathbf{x}_0 - \mathbf{x}$  is, by equation (5.42), annihilated in  $\mathbf{x}_1 - \mathbf{x}$  and replaced by a corresponding portion of  $\mathbf{v}_{i,k-1}$ . This causes the degree  $m$  of the minimal polynomial of  $\mathbf{A}$  for the vector  $\mathbf{x}_1 - \mathbf{x}$  to be one less than that for  $\mathbf{x}_0 - \mathbf{x}$ . Therefore, by Theorem 5.1, one lower order of the VEA is needed to give the limit vector  $\mathbf{x}$  provided the sequence to be transformed begins with  $\mathbf{x}_1$  instead of  $\mathbf{x}_0$ . This explains the case for  $p = 1$ . This effect continues with each iteration until  $\mathbf{x}_k$  which no longer has any representative of any column block  $\mathbf{V}_i$  for which  $\mu_i = 0$ .

These results do not yield a sufficient condition if the matrix  $\mathbf{I} - \mathbf{A}$  is singular.

Brezinski [3] addressed this lack with

**Theorem 5.3:** Given a vector sequence  $\{\mathbf{x}_n\}$  generated by equation (5.1), if  $\mathbf{I} - \mathbf{A}$  is singular and  $\mathbf{b}$  is in the range of  $\mathbf{I} - \mathbf{A}$  (which means that  $\mathbf{x} = \mathbf{Ax} + \mathbf{b}$  has infinitely many solutions  $\mathbf{x}$ ), let  $m$  denote the degree of the minimal polynomial of  $\mathbf{A}$  for the vector  $\mathbf{x}_0 - \mathbf{x}$ , let  $k$  denote the number of zeros of the same polynomial that are equal to 0, and let  $q$  denote the number of zeros of the same polynomial that are equal to 1. Then, if  $q = 1$ , for  $0 \leq p \leq k$ ,

$$\mathbf{e}_{2(m-p)-2}^{(n+p)} = \mathbf{x}, \quad n = 0, 1, \dots, \quad (5.44)$$

and if  $q = 2$ ,

$$\mathbf{e}_{2(m-p)-3}^{(n+p)} = \mathbf{z}, \quad n = 0, 1, \dots \quad (5.45)$$

where  $\mathbf{z}$  is a constant vector.

If  $q = 1$ , there exists in  $\mathbf{x}_0 - \mathbf{x}$  one or more parts of vectors  $\mathbf{v}_{i,1}$  for which  $\mu_i = 1$ .

The fact that  $\mathbf{v}_{i,1}$  is a true eigenvector associated with unity means that

$$\mathbf{A}\mathbf{v}_{i,1} = \mathbf{v}_{i,1}. \quad (5.46)$$

Therefore, any component of  $\mathbf{v}_{i,1}$  present in  $\mathbf{x}_0 - \mathbf{x}$  which is  $\mathbf{e}_0$  endures unchanged throughout all  $\mathbf{e}_n$ . These vector components represent merely a translation of the vector sequence  $\{\mathbf{x}_n\}$  along the null space of  $\mathbf{I} - \mathbf{A}$ . The presence of these vector components in  $\mathbf{x}_0 - \mathbf{x}$  places no added burden on the VEA because, by Theorem 2.5, a translation of the vectors in a vector sequence causes no change in the VEA order of the sequence. Therefore, the contribution of these vector components to the degree of the minimal polynomial is corrected for in the  $\epsilon$  subscript of equation (5.44). Thus, Brezinski has extended the upper bound on the VEA order of sequences generated by equation (5.1) to those for which  $\mathbf{I} - \mathbf{A}$  is singular.

When  $q = 2$ , the vector  $\mathbf{x}_0 - \mathbf{x}$  has at least some component of the second column vector  $\mathbf{v}_{i,2}$  from a column block  $\mathbf{V}_i$  for which  $\mu_i = 1$  in it, a generalised eigenvector component. The equation for this generalised eigenvector is (again, equation (B.12))

$$\mathbf{A}\mathbf{v}_{i,2} = \mathbf{v}_{i,2} + \mathbf{v}_{i,1}. \quad (5.47)$$

Therefore, with each premultiplication by  $\mathbf{A}$ , not only is the generalised eigenvector preserved, but there appears an added true eigenvector component. Therefore,  $\mathbf{e}_{12}$  will have had  $12\mathbf{v}_{i,1}$  added to it due to equation (5.7). Such a sequence can no more have a limit or antilimit than can the integer sequence  $\{n\}$ . Theorem 5.3 confirms this in equation (5.45) where the  $\epsilon$  subscript is odd. This represents a complete breakdown for the algorithm because if the vectors in an odd column of the  $\epsilon$  table are constant, then the previous transform sequence of vectors must not be constant. The terms of the next transform sequence are not defined, because their calculation requires the inversion of a zero difference vector. Clearly, the presence in  $\mathbf{x}_0 - \mathbf{x}$  of any part of a generalised

eigenvector of  $\mathbf{A}$  associated with a unity eigenvalue in the sequence vectors means that it can have no VEA order.

One commonly found description of the action of the VEA on vector sequences generated according to equation (5.1) is that the  $k^{\text{th}}$  order VEA produces a transform sequence which behaves as if  $\mathbf{A}$  had had its  $k$  largest magnitude eigenvalues set to zero. In effect, each order of VEA zeroes a dominant eigenvalue of  $\mathbf{A}$ . At the end of the next section, it shall be seen that under some circumstances, a single order of the VEA is effectively able to zero *two* of the largest magnitude eigenvalues of  $\mathbf{A}$ . The full description of these circumstances constitutes the subject of the next chapter.

### 5.3 The Initial Experimental Results

The experimental results are reported in two subsections. Subsection 5.3.1 gives the results of the experiments on sequences satisfying

$$\mathbf{e}_n = \mathbf{J}^n \mathbf{c}. \quad (5.48)$$

These sequences correspond to those generated according to

$$\mathbf{x}_n = \mathbf{A}\mathbf{x}_{n-1} + \mathbf{b}. \quad (5.49)$$

for which  $\mathbf{A}$  can be transformed into its Jordan canonical form  $\mathbf{J}$  by a unitary matrix  $\mathbf{V}$ .

Subsection 5.3.2 summarises the results of experiments on sequences satisfying

$$\mathbf{e}_n = \mathbf{V}\mathbf{J}^n \mathbf{c}. \quad (5.50)$$

where  $\mathbf{V}$  is not unitary. This set of sequences corresponds to those generated according to equation (5.49) for which  $\mathbf{A}$  can only be transformed into its Jordan canonical form  $\mathbf{J}$  by a matrix  $\mathbf{V}$  which is not unitary.

### 5.3.1 When $\mathbf{V}$ is Unitary

The first set of experiments done were on vector sequences  $\{\mathbf{e}_n\}$  satisfying equation (5.26) which is

$$\mathbf{e}_n = \mathbf{J}^n \mathbf{c}. \quad (5.51)$$

This set of experiments uncovered circumstances under which the VEA order of a sequence was lower than the upper bound given in Theorems 5.2 and 5.3. In some cases, the VEA order was half of the upper bound of the theorems reviewed. These circumstances are now given. In order to describe the circumstances, a phenomenon to which the author has given the name overshadowing is introduced.

For a matrix  $\mathbf{J}$  having both  $\nu_c$  and its complex conjugate  $\overline{\nu_c}$  for eigenvalues, let the powers to which  $(\mu - \nu_c)$  and  $(\mu - \overline{\nu_c})$  are raised in the factored form of  $p_{\min, \mathbf{J}, \mathbf{c}}$  be denoted  $K_{\nu_c}$  and  $K_{\overline{\nu_c}}$ , respectively. Then, if  $K_{\nu_c} > K_{\overline{\nu_c}}$ , it will be said that some Jordan block having  $\nu_c$  as an eigenvalue overshadows those Jordan blocks having  $\overline{\nu_c}$  for an eigenvalue. Similarly, if  $K_{\nu_c} < K_{\overline{\nu_c}}$ , then it will be said that some Jordan block having  $\overline{\nu_c}$  for an eigenvalue overshadows those Jordan blocks having  $\nu_c$  for an eigenvalue. Every instance of a complex number and its complex conjugate both being eigenvalues of  $\mathbf{J}$  will result in one or more overshadowed Jordan blocks, provided there



exist nonzero  $c_{ij}$  in  $\mathbf{c}$  corresponding both to Jordan blocks having  $\nu_c$  for an eigenvalue and to Jordan blocks having  $\overline{\nu_c}$  for an eigenvalue.

Now, it was found in this first set of experiments that the amount by which the VEA order of  $\{\mathbf{e}_n\}$  fell below the upper bound given by Theorems 5.2 and 5.3 was the sum of the  $K_{\nu_c}$  and/or  $K_{\overline{\nu_c}}$  corresponding to overshadowed Jordan blocks in  $\mathbf{J}$ . Another way to say it is that in order to arrive at the true VEA order using the formulae of Theorems 5.2 and 5.3, it is necessary to let  $m$  be the order of the minimal polynomial of the matrix  $\underline{\mathbf{J}}$  for the vector  $\mathbf{c}$  where  $\underline{\mathbf{J}}$  is the matrix  $\mathbf{J}$  after the eigenvalue of each overshadowed Jordan block has been changed to that of its complex conjugate.

To illustrate the concept of overshadowing, consider the sequence  $\{\mathbf{e}_n\}$  generated according to

$$\mathbf{e}_n = \begin{bmatrix} 3 & 1 & & & & & \\ 0 & 3 & & & & & \\ \hline & & 1+i & 1 & 0 & & \\ & & 0 & 1+i & 1 & & \\ & & 0 & 0 & 1+i & & \\ \hline & & & & & 1-i & 1 \\ & & & & & 0 & 1-i \end{bmatrix}^n \begin{bmatrix} c_{1,1} \\ 0 \\ 0 \\ c_{2,2} \\ c_{2,3} \\ c_{3,1} \\ c_{3,2} \end{bmatrix} \quad (5.52)$$

where  $i$  in this expression is the square root of negative one. In this equation, only non-zero scalars in  $\mathbf{c}$  have been given their labels. It can be seen that  $c_{1,1}$ ,  $c_{2,3}$ , and  $c_{3,2}$  are the non-zero coefficients with greatest second index in  $\mathbf{c}$ . Therefore, the minimal polynomial of  $\mathbf{J}$  for the vector  $\mathbf{c}$  will be

$$p_{\min, \mathbf{J}, \mathbf{c}}(\mu) = (\mu - 3)(\mu - 1 - i)^3(\mu - 1 + i)^2 \quad (5.53)$$

having degree 6. Since  $\mathbf{I} - \mathbf{J}$  is nonsingular and 0 is not a root of  $p_{\min, \mathbf{J}, \mathbf{c}}(\mu)$ , by Theorem 5.2, a sixth order application of the VEA to  $\{\mathbf{e}_n\}$  is sufficient to give a constant  $\mathbf{0}$  sequence. What is observed, however, is that a fourth order application of the VEA to  $\{\mathbf{e}_n\}$  yields the zero sequence. The second Jordan block overshadows the third Jordan block because the eigenvalues of the blocks are complex conjugates and the non-zero scalar of highest index in  $\mathbf{c}_2$  is three compared to that of  $\mathbf{c}_3$  which is only two.

If  $\mathbf{c}$  were changed and  $\{\mathbf{e}_n\}$  were, instead, generated by

$$\mathbf{e}_n = \begin{bmatrix} 3 & 1 & & & & & & & \\ 0 & 3 & & & & & & & \\ & & 1+i & 1 & 0 & & & & \\ & & 0 & 1+i & 1 & & & & \\ & & 0 & 0 & 1+i & & & & \\ & & & & & 1-i & 1 & & \\ & & & & & 0 & 1-i & & \end{bmatrix}^n \begin{bmatrix} c_{1,1} \\ 0 \\ c_{2,1} \\ 0 \\ 0 \\ c_{3,1} \\ c_{3,2} \end{bmatrix}, \quad (5.54)$$

then the third Jordan block would overshadow the second, even though the second block is bigger; it is the greatest index of non-zero coefficient in  $\mathbf{c}_2$  and not the size of  $\mathbf{J}_2$  which determines the power of the corresponding factor  $(\mu - 1 - i)$  in  $p_{\min, \mathbf{J}, \mathbf{c}}(\mu)$ . Therefore, only a third order application of the VEA is required instead of a fourth order application in order to give the zero sequence.

The most dramatic examples of this reduction are  $2k \times 2k$  matrices  $\mathbf{J}$  having two distinct eigenvalues in two  $k \times k$  Jordan blocks having non-zero non-unity complex conjugate eigenvalues with  $\mathbf{c}$  having nonzero coefficients  $c_{1,k}$  and  $c_{2,k}$ . In this case, the

degree of  $p_{\min, \mathbf{J}, \mathbf{c}}$  is  $2k$  as is Brezinski's upper bound for the VEA order of  $\{\mathbf{e}_n\}$  while the VEA order of  $\{\mathbf{e}_n\}$  is only  $k$ .

To summarise, what is seen is that two vector sequences  $\{\mathbf{e}'_n\}$  and  $\{\mathbf{e}''_n\}$  can be generated by matrices  $\mathbf{J}'$  and  $\mathbf{J}''$ , respectively, and vectors  $\mathbf{c}'$  and  $\mathbf{c}''$ , respectively, where Brezinski's upper bound for the VEA order in both cases is  $k$  and yet  $\{\mathbf{e}'_n\}$  will have a VEA order lower than  $k$  while  $\{\mathbf{e}''_n\}$  does not. This difference will be due to  $\mathbf{J}'$  having complex conjugate eigenvalues.

### 5.3.2 When $\mathbf{V}$ is Not Unitary

Experiments done on vector sequences  $\{\mathbf{e}_n\}$  satisfying equation (5.15)

$$\mathbf{e}_n = \mathbf{V}\mathbf{J}^n\mathbf{c}, \quad n = 0, 1, \dots \quad (5.55)$$

where  $\mathbf{V}$  is not unitary revealed that the phenomenon of overshadowing was necessary in order for the VEA order of  $\{\mathbf{e}_n\}$  to be lower than the upper bound of Theorems 5.2 and 5.3 but that overshadowing in itself was not sufficient. When the columns of  $\mathbf{V}$  were scaled so that they no longer had unit length but still remained mutually orthogonal, overshadowing continued to lower the VEA order as when  $\mathbf{V}$  was unitary. However, when some of the orthogonality relationships between the columns of  $\mathbf{V}$  were destroyed, the effect of overshadowing was sometimes unchanged and sometimes partially or completely eliminated. In other words, if the VEA order of the sequence  $\{\mathbf{J}^n\mathbf{c}\}$  is  $u$  orders lower than the upper bound of Theorems 5.2 and 5.3, the VEA order of  $\{\mathbf{V}\mathbf{J}^n\mathbf{c}\}$  will be  $v$  orders lower than the upper bound of Theorems 5.2 and 5.3 where

$0 \leq v \leq u$ , depending upon the presence or absence of certain not yet identified orthogonal relationships between the columns of  $\mathbf{V}$ .

In order to determine the precise correlation between combinations of orthogonal relationships among the columns of  $\mathbf{V}$  and the magnitude of the reduction of the VEA order of the resulting sequence below Brezinski's upper bounds, a second set of experiments was designed and executed. The design and results of these experiments is the subject of the next chapter.

## 6. ORTHOGONALITY IN V AND VEA ORDER REDUCTION

In the experiments of the previous chapter, it was observed (Subsection 5.3.1) that some vector sequences  $\{\mathbf{e}_n\}$  whose terms satisfy

$$\mathbf{e}_n = \mathbf{J}^n \mathbf{c} \quad (6.1)$$

do have a VEA order lower than the upper bound given by Brezinski in Theorems 5.2 and 5.3. In this expression,  $\mathbf{J}$  is in Jordan canonical form (i.e.  $N \times N$  and block diagonal), having the form

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_1 & & & & \\ & \mathbf{J}_2 & & & \\ & & \mathbf{J}_3 & & \\ & & & \ddots & \\ & & & & \mathbf{J}_r \end{bmatrix} \quad (6.2)$$

and each Jordan block  $\mathbf{J}_i$  is  $N_i \times N_i$  and has the form

$$\mathbf{J}_i = \begin{bmatrix} \mu_i & 1 & 0 & \cdots & 0 & 0 \\ 0 & \mu_i & 1 & \ddots & 0 & 0 \\ 0 & 0 & \mu_i & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & \mu_i & 1 \\ 0 & 0 & 0 & \cdots & 0 & \mu_i \end{bmatrix} \quad (6.3)$$

while  $\mathbf{c}$  is  $N \times 1$  and can be partitioned in a conformable fashion so that

$$\mathbf{c} = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \vdots \\ \mathbf{c}_r \end{bmatrix} \quad (6.4)$$

and each  $\mathbf{c}_i$  is  $N_i \times 1$  having elements labelled according to

$$\mathbf{c}_i = \begin{bmatrix} c_{i,1} \\ c_{i,2} \\ \vdots \\ c_{i,N_i} \end{bmatrix}. \quad (6.5)$$

Those sequences having a lower VEA order than Brezinski's upper bound are generated by a matrix  $\mathbf{J}$  having pairs of complex conjugate eigenvalues and a vector  $\mathbf{c}$  having some non-zero coefficients  $c_{i,j}$  corresponding to Jordan blocks  $\mathbf{J}_i$  of both such eigenvalues. How much lower the VEA order of the sequence is than Brezinski's upper bound is determined by how many degrees of  $p_{\min, \mathbf{J}, \mathbf{c}}(\mu)$  are due to overshadowed Jordan blocks in  $\mathbf{J}$ .

It was also observed that if a nonunitary matrix  $\mathbf{V}$  premultiplies this vector sequence to create  $\{\mathbf{V}\mathbf{e}_n\}$ , then the reduction in VEA order below Brezinski's upper bound due to overshadowed Jordan blocks may be either preserved, diminished, or eliminated. Which of these occurs depends not on the lengths of the column vectors of  $\mathbf{V}$ , which is to some extent arbitrary, but upon the presence or absence of certain orthogonal relationships among the column vectors in  $\mathbf{V}$ . This chapter describes the experiments carried out in order to determine the exact correlation between orthogonal relationships among the column vectors of  $\mathbf{V}$  and the degree of VEA order reduction preserved from

that observed when  $\mathbf{V}$  is unitary. It also seeks to extend these results with a conjecture for higher orders and notes some consequences of these results.

Section 6.1 describes the components of the experiments in some detail. In §6.2, the experimental details and specific results for  $N = 2, 4, 6,$  and  $8$  are given. Section 6.3 summarises these results and formulates a conjecture for all higher orders consistent with them. Finally, §6.4 describes consequences of these results, including a possible insight into one successful application of the VEA and a broader sufficient condition for membership of a vector sequence in the kernel of the 1<sup>st</sup> order VEA.

## 6.1 Design of the Experiments

The sequences used in these experiments all satisfy the expression

$$\mathbf{e}_n = \begin{bmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \end{bmatrix} \begin{bmatrix} \mathbf{J}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_2 \end{bmatrix} \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{bmatrix}, \quad n = 0, 1, \dots \quad (6.6)$$

where  $\mathbf{V}_1$  and  $\mathbf{V}_2$  are column blocks of equal size on whose column vectors some set of orthogonality relations has been imposed and  $\mathbf{J}_1$  and  $\mathbf{J}_2$  are Jordan blocks of equal size having unequal complex conjugate eigenvalues. Therefore, the size of  $\mathbf{V}$  and of  $\mathbf{J}$  is an even number. The eigenvalues of  $\mathbf{J}_1$  and  $\mathbf{J}_2$  are not equal to zero or one<sup>6</sup> so, by Theorem 5.2, the upper bound on the VEA order for all resulting sequences is simply the

---

<sup>6</sup> Actual eigenvalues used in tests were  $.8 \pm i.6$  and  $1 \pm i.5$ .

degree of  $p_{\min, \mathbf{J}, \mathbf{c}}(\mu)$ . The vector  $\mathbf{c}$  used in each case is random with the constraint that

$$c_{1, N_1} \neq 0 \neq c_{2, N_2}. \quad (6.7)$$

This ensures that the degree of  $p_{\min, \mathbf{J}, \mathbf{c}}(\mu)$  for every case is  $N$ . By turn, a list of combinations of orthogonality relationships is defined. For each combination in the list, a matrix  $\mathbf{V}$  is generated possessing those and only those orthogonalities called for in that combination. Then, the terms of the sequence  $\{\mathbf{e}_n\}$  are generated according to equation (6.6), the corresponding  $\varepsilon$ -table constructed, and the VEA-order of  $\{\mathbf{e}_n\}$  determined. In this way, it can be ascertained which vectors in  $\mathbf{V}_1$  and  $\mathbf{V}_2$  must be and which need not be orthogonal in order to obtain various reductions in the VEA order below  $N$ .

A tool of the author's invention called an o-diagram which will prove useful is now introduced. It will illustrate any set of orthogonality relationships present between the columns of  $\mathbf{V}$ . The diagram begins with a set of  $N$  dots, each dot representing a column vector of  $\mathbf{V}$  in a one-to-one correspondence, located at the vertices of a regular  $N$ -sided polygon. The dots are positioned so that there are two top-most dots beside each other, and since  $N$  is even, the same will be true of two bottom-most dots. The dots of the left half of the diagram correspond to vectors in  $\mathbf{V}_1$  and those of right half to vectors in  $\mathbf{V}_2$ . The top-most dots correspond to  $\mathbf{v}_{1,1}$  and  $\mathbf{v}_{2,1}$ . Consecutive dots continuing down each side of the diagram correspond to subsequent vectors within each column block of  $\mathbf{V}$ , the left and right bottom-most dots corresponding to  $\mathbf{v}_{1, N_1}$  and  $\mathbf{v}_{2, N_2}$ , respectively. The diagram will also have zero or more line segments connecting pairs of dots to form a simple graph. Each segment indicates an orthogonal relationship



between the vectors corresponding to the dots joined by the segment. This arrangement of dots permits all such segments to be distinguished visually because no three dots are collinear.

For the matrix

$$\mathbf{V} = \begin{bmatrix} | & | & | & | \\ | & | & | & | \\ \mathbf{v}_{1,1} & \mathbf{v}_{1,2} & \mathbf{v}_{2,1} & \mathbf{v}_{2,2} \\ | & | & | & | \\ | & | & | & | \end{bmatrix} \quad (6.8)$$

having no orthogonal relationship between any two column vectors, the corresponding o-diagram is given in Figure 6.1, below. In this figure, each dot is labelled with the vector it symbolises. Labels will generally be omitted from subsequent o-diagrams.

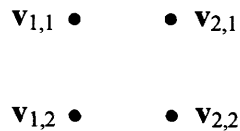


Figure 6.1 - The o-diagram for  $\mathbf{V}$  with 2 vectors in each column block, and no 2 vectors orthogonal.

Between the four column vectors of  $\mathbf{V}$  in equation (6.8), there exist six possible orthogonal relationships. As each relationship has only two possible states, orthogonal or not orthogonal, it is natural to use the binary number system to represent combinations of orthogonal relationships. However, this will require an ordering of the relationships. A natural ordering emerges when considering the left-to-right order of the vectors in  $\mathbf{V}$ , ignoring the column block separator. This order is shown in Table 6.1.

Notice that each row of the table records a vector relating to each of the vectors to its left in  $\mathbf{V}$ , each of which appears in sequence from left to right in  $\mathbf{V}$ .

Table 6.1 - The order used for orthogonality relations in this chapter.

$$\begin{array}{ccccccc}
 & & & & & & \mathbf{v}_{1,2} \perp \mathbf{v}_{1,1} ? \\
 & & & & & & \mathbf{v}_{1,3} \perp \mathbf{v}_{1,1} ? \quad \mathbf{v}_{1,3} \perp \mathbf{v}_{1,2} ? \\
 & & & & & & \mathbf{v}_{1,4} \perp \mathbf{v}_{1,1} ? \quad \mathbf{v}_{1,4} \perp \mathbf{v}_{1,2} ? \quad \mathbf{v}_{1,4} \perp \mathbf{v}_{1,3} ? \\
 & & & & & & \vdots \\
 & & & & & & \mathbf{v}_{1,N_1} \perp \mathbf{v}_{1,1} ? \quad \mathbf{v}_{1,N_1} \perp \mathbf{v}_{1,2} ? \quad \dots \quad \mathbf{v}_{1,N_1} \perp \mathbf{v}_{1,N_1-1} ? \\
 & & & & & & \mathbf{v}_{2,1} \perp \mathbf{v}_{1,1} ? \quad \mathbf{v}_{2,1} \perp \mathbf{v}_{1,2} ? \quad \dots \quad \mathbf{v}_{2,1} \perp \mathbf{v}_{1,N_1-1} ? \quad \mathbf{v}_{2,1} \perp \mathbf{v}_{1,N_1} ? \\
 & & & & & & \vdots \\
 & & & & & & \mathbf{v}_{2,N_2} \perp \mathbf{v}_{1,1} ? \quad \mathbf{v}_{2,N_2} \perp \mathbf{v}_{1,2} ? \quad \dots \quad \mathbf{v}_{2,N_2} \perp \mathbf{v}_{1,N_1} ? \quad \mathbf{v}_{2,N_2} \perp \mathbf{v}_{2,1} ? \quad \dots \quad \mathbf{v}_{2,N_2} \perp \mathbf{v}_{2,N_2-1} ?
 \end{array}$$

Clearly, the number of possible relationships between  $N$  vectors is  $1 + 2 + \dots + (N - 2) + (N - 1) = \frac{1}{2}(N - 1)N$ , and this is also the number of binary digits needed to represent any combination of orthogonalities in a  $\mathbf{V}$  matrix of  $N$  columns. The right-most binary digit has been made to correspond to the relationship at the top of Table 6.1 and the left-most binary digit to the bottom right-hand relation in Table 6.1. A 1 will indicate orthogonality and a 0 nonorthogonality. For the 4-column  $\mathbf{V}$  of equation (6.8), the  $2^6$  possible combinations of orthogonalities are represented by the binary numbers 000000 through 111111 or 0 through 63. The order of each of these relations is further illustrated by Figure 6.2.

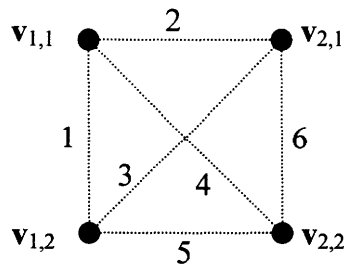


Figure 6.2 - An ordering of the relationships in an o-diagram for  $V$  consisting of 2 column blocks with 2 vectors each.

To ease the job of reading binary numbers, commas will be inserted between groups of digits corresponding to relationships on the same row of Table 6.1. Thus, Figure 6.1 shows the diagram corresponding to the number 0 or 000,00,0 in binary notation, Figure 6.3 gives the same for 63 or 111,11,1 in binary, and Figure 6.4 gives the diagram for 13 or 001,10,1 in binary.

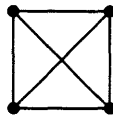


Figure 6.3 - O-diagram for unitary  $V$  consisting of 2 column blocks with 2 vectors each.

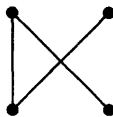


Figure 6.4 - O-diagram for  $V$  consisting of 2 column blocks with 2 vectors each, with orthogonal relationships 001,10,1.

Figures 6.5 and 6.6 give examples of o-diagrams for  $V$  matrices consisting of 2 column blocks having 3 vectors in each block and 4 vectors in each block, respectively.

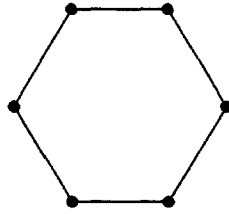


Figure 6.5 - O-diagram for  $V$  consisting of 2 column blocks, 3 vectors each, with orthogonality 10100,1000,001,10,1.

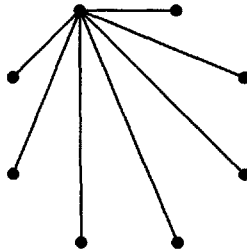


Figure 6.6 - O-diagram for  $V$  consisting of 2 column blocks, 4 vectors each, with orthogonality 0000001,000001,00001,0001,001,01,1.

For each combination of orthogonality relations, a matrix  $V$  having the prescribed relations must be generated. This is done by beginning with the  $N^{\text{th}}$ -order identity matrix and adjusting coefficients above the diagonal one at a time until the desired orthogonality relationships are realised. The order in which coefficients are adjusted is shown in alphabetical order by the letters  $a$  through  $f$  in the matrix

$$V = \begin{bmatrix} 1 & a & b & d \\ 0 & 1 & c & e \\ 0 & 0 & 1 & f \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (6.9)$$

Thus, the order of relationships in Table 6.1 is mirrored in the sequence of coefficients set in order to generate a suitable  $\mathbf{V}$ . For the number 13 or binary 001,10,1, the corresponding  $\mathbf{V}$  matrix generated for use in the experiments was

$$\mathbf{V} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (6.10)$$

and corresponds to the o-diagram given in Figure 6.4.

## 6.2 Test Results

### 6.2.1 $N=2$

Between the two column vectors of  $\mathbf{V}$  there is only one relationship to consider and therefore  $2^1 = 2$  matrices  $\mathbf{V}$  with 2 o-diagrams to generate and two sequences for which to determine the VEA order. It was found that if  $\mathbf{v}_{1,1} \perp \mathbf{v}_{2,1}$  then the corresponding sequence  $\{\mathbf{e}_n\}$  was 1<sup>st</sup> order and otherwise it was 2<sup>nd</sup> order. Therefore, the o-diagram illustrating the necessary orthogonality relationships for a reduction in VEA order of 1 is given in Figure 6.7. See §6.4.2 for further comments on this case.

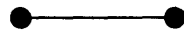


Figure 6.7 - O-diagram for  $\mathbf{V}$  having 2 column blocks with 1 vector each giving a reduction of 1 VEA order below Brezinski's upper bound.

### 6.2.2 $N = 4$

As already mentioned, for  $N = 4$  the 6 relations define 64 different combinations with their o-diagrams for each of which the corresponding VEA order must be determined. The results of all 64 o-diagrams is shown in Figure 6.8 with 4<sup>th</sup> order o-diagrams to the left and unframed, 2<sup>nd</sup> order to the right and solidly framed, and 3<sup>rd</sup> order in between with dotted frames. For each order, the o-diagram with the fewest orthogonal relationships is shaded.

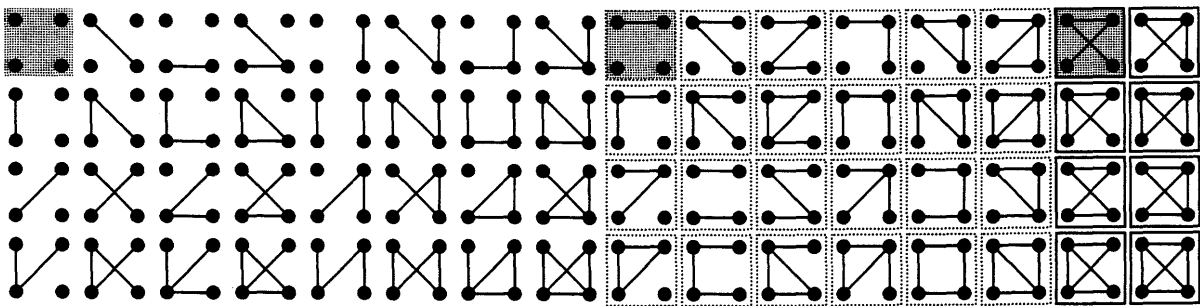


Figure 6.8 - The 64 o-diagrams for  $V$  having 2 column blocks with 2 vectors each, grouped by order, 4<sup>th</sup> order not framed, 3<sup>rd</sup> order framed by dotted lines, and 2<sup>nd</sup> order framed by solid lines.

Of the unframed 4<sup>th</sup> order o-diagrams, perhaps the most telling one is in the bottom right-hand corner of that group. It contains every orthogonal relationship possible except the one between the two eigenvectors  $v_{1,1}$  and  $v_{2,1}$ . Notice also that the 32 cases which do show a reduction in VEA order, members of the 3<sup>rd</sup> and 2<sup>nd</sup> order groups, all correspond to  $V$  matrices having orthogonal eigenvectors. Clearly, without that relationship, there is no reduction whatsoever in VEA order below the upper bound. In this respect, the results for  $N = 4$  are similar to those for  $N = 2$ .

Of the solidly framed 2<sup>nd</sup> order o-diagrams, the shaded one defines three orthogonal relationships which are present in every other 2<sup>nd</sup> order diagram. Every 3<sup>rd</sup> order o-diagram lacks either the orthogonality between  $v_{1,1}$  and  $v_{2,2}$  or the one between  $v_{1,2}$  and  $v_{2,1}$  or both. Therefore, it appears that the orthogonality between eigenvectors provides the first reduction in VEA order, and these two “cross orthogonalities” provide the second reduction in VEA order.

### 6.2.3 $N = 6$

With the increase in the number of vectors to six, the number of relationships increases to 15, and the number of possible combinations of these relationships to  $2^{15}$  or 32,768. It should be noted that all results thus far were obtained on an HP 48GX programmable scientific calculator. On it, the 64 cases could be run in just under an hour. The 32,768 cases for  $N = 6$  would take just over 45 days to complete, longer than the lifespan of one set of batteries. Clearly, something else had to be done.

It was noted that all of the  $N = 4$  results showed no advantage to the presence of an orthogonal relationship without its mirror image relationship (mirrored with respect to a central vertical axis on the o-diagram) also being orthogonal. Therefore, it was decided to determine how many orthogonal relationships could be grouped with their mirror images and either included or omitted together. This approach reduced the number of degrees of freedom to 9. The corresponding o-diagrams for these symmetrically grouped orthogonalities are shown in Figure 6.9.

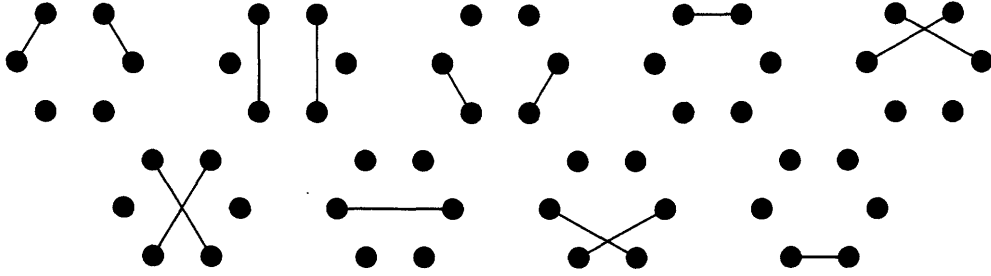


Figure 6.9 - O-diagrams for 9 symmetrically grouped sets of orthogonal relationships for  $V$  having 2 column blocks with 3 vectors each.

With this change, the number of cases to run became  $2^9$  or 512, a job of only 17 hours. Subsequent procurement of an HP 48GX simulator to run on a PC ultimately reduced the job to 1.5 hours. The results are summarised one order at a time beginning with the 1<sup>st</sup> reduction of one VEA order.



Figure 6.10 - a. The maximal 6<sup>th</sup> order o-diagram, and b. the minimal 5<sup>th</sup> order o-diagram for  $V$  having 2 column blocks with 3 vectors each.

It was found that every 6<sup>th</sup> order o-diagram was a subset of the 6<sup>th</sup> order o-diagram shown in Figure 6.10 a. (Such o-diagrams will be referred to as maximal.) Thus, every 6<sup>th</sup> order o-diagram, like those of the 4<sup>th</sup> order for  $N = 4$  and 2<sup>nd</sup> order for  $N = 2$ , lacked the top segment between  $v_{1,1}$  and  $v_{2,1}$ . It was also found the every o-diagram of 5<sup>th</sup> order was a superset of the one shown in Figure 6.10 b. (Such o-diagrams will be referred to as minimal.) All lower order o-diagrams were also supersets of Figure 6.10 b. Once again, the first reduction of one VEA order below Brezinski's upper bound relies solely on the presence of an orthogonality between the two eigenvectors of  $V$ .



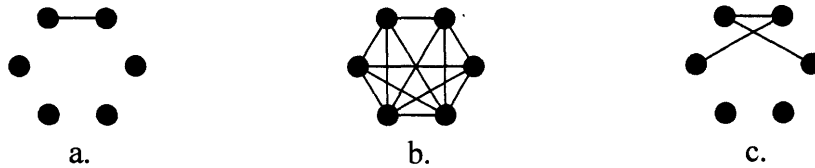


Figure 6.11 - a. The minimal 5<sup>th</sup> order o-diagram, b. the maximal 5<sup>th</sup> order o-diagram, and c. the minimal 4<sup>th</sup> order o-diagram for  $\mathbf{V}$  having 2 column blocks with 3 vectors each.

The minimal and maximal 5<sup>th</sup> order o-diagrams of Figure 6.11 a. and b. show that every 5<sup>th</sup> order o-diagram both included the orthogonality between eigenvectors and excluded the orthogonalities between  $\mathbf{v}_{1,1}$  and  $\mathbf{v}_{2,2}$  and between  $\mathbf{v}_{2,1}$  and  $\mathbf{v}_{1,2}$ . It was also observed that the three orthogonalities present in the minimal 4<sup>th</sup> order o-diagram of Figure 6.11 c. were present in every 4<sup>th</sup> order and every 3<sup>rd</sup> order diagram. Subsequent testing of the o-diagram in Figure 6.11 c. with either of these two orthogonalities removed also resulted in 5<sup>th</sup> order sequences. This parallels the situation found with  $N=4$ . The second reduction of one VEA order, given a first such reduction, hinges completely on the presence of both the orthogonality between  $\mathbf{v}_{1,2}$  and  $\mathbf{v}_{2,1}$  and that between  $\mathbf{v}_{1,1}$  and  $\mathbf{v}_{2,2}$ .

The 3<sup>rd</sup> VEA order reduction is a little trickier to analyse. The o-diagrams of Figure 6.12 a. and b. are two different maximal 4<sup>th</sup> order diagrams. (Every 4<sup>th</sup> order o-diagram was a subset of one of these two figures.) Figure 6.12 a. lacks only the orthogonality between  $\mathbf{v}_{1,2}$  and  $\mathbf{v}_{2,2}$ . The only orthogonalities lacking in Figure 6.12 b. are those between  $\mathbf{v}_{1,1}$  and  $\mathbf{v}_{2,3}$  and between  $\mathbf{v}_{2,1}$  and  $\mathbf{v}_{1,3}$ . The minimal 3<sup>rd</sup> order o-diagram of Figure 6.12 c. contains all three of these orthogonalities as well as those of the minimal 4<sup>th</sup> order o-diagram. It may therefore be concluded that the third reduction of

one VEA order requires all three of these orthogonalities in addition to those needed for the previous two reductions.

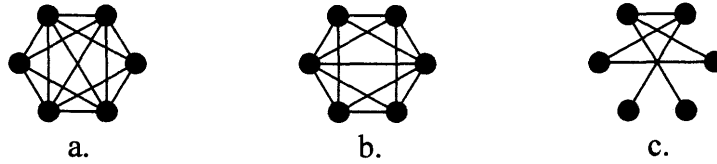


Figure 6.12 - a.,b. Two maximal 4<sup>th</sup> order o-diagrams and c. the minimal 3<sup>rd</sup> order o-diagram for  $V$  having 2 column blocks with 3 vectors each.

#### 6.2.4 $N = 8$

Between pairs chosen from eight vectors, 28 different relationships are possible defining 268,435,456 different  $8 \times 8$   $V$  matrices to create, sequences to generate, and 8<sup>th</sup> order  $\varepsilon$ -tables to calculate. Even a reduction to the 16 symmetric orthogonal relationship components leaves 65,536 cases to run, too much for the author's chosen research tools at the time.

Up to this point, none of the orthogonalities essential to any VEA order reduction have been between vectors of the same column block. If the 6 corresponding symmetric relationship sets are removed from the test, there remain still 10 symmetric degrees of freedom. If the test is further restricted to a search for a 4<sup>th</sup> VEA order of reduction given 3 VEA orders of reduction below the upper bound, then the 4 symmetric sets of orthogonalities necessary to obtain 3 reductions of VEA order can be included in all cases tested leaving only 6 symmetric degrees of freedom and 64 cases to run. The basic configuration and the 6 symmetric components to be added in combinations are shown in Figure 6.13.

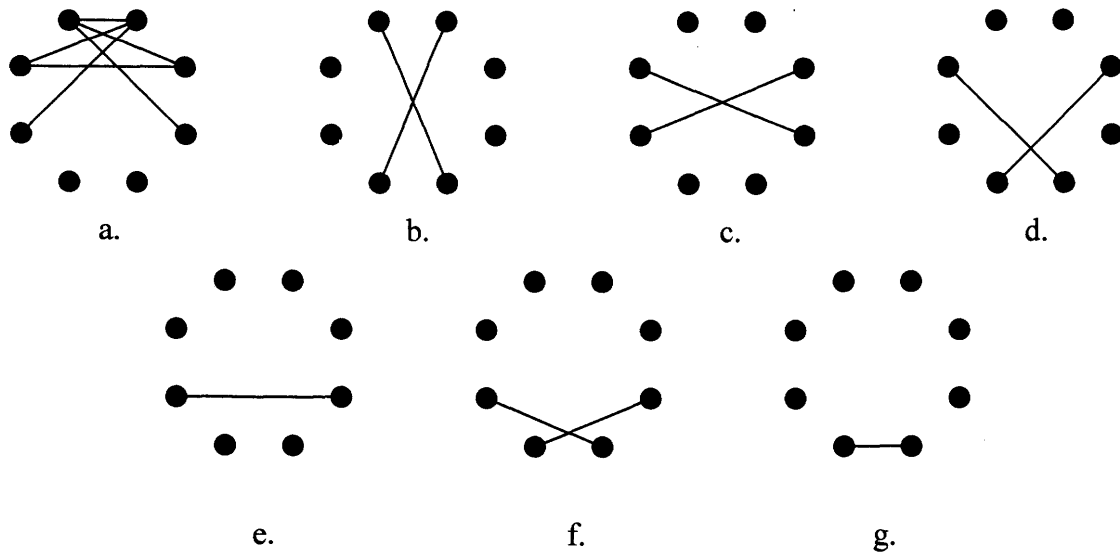


Figure 6.13 - a. O-diagram for an expected reduction of 3 VEA orders and b.-g. six symmetric orthogonality sets to add to it in search of a 4<sup>th</sup> VEA order reduction for  $V$  having 2 column blocks with 4 vectors each.

Because all 64 combinations include what are expected to be the necessary orthogonalities for 3 VEA orders of reduction below the upper bound for  $N = 8$ , it is expected that no case will produce a sequence of higher than 5<sup>th</sup> order. This turned out to be the case. Figure 6.14 shows the two maximal 5<sup>th</sup> order diagrams and the minimal 4<sup>th</sup> order o-diagram. Of all the orthogonalities tested, the ones missing from Figure 6.14 a. are those between  $v_{1,2}$  and  $v_{2,3}$  and between  $v_{1,3}$  and  $v_{2,2}$ . Those missing from Figure 6.14 b. are between  $v_{1,1}$  and  $v_{2,4}$  and between  $v_{1,4}$  and  $v_{2,1}$ . These four orthogonalities are the ones that appear along with those orthogonalities necessary for a reduction of 3 VEA orders in Figure 6.14 c. which together resulted in 4 VEA orders of reduction. Subsequent testing showed that the removal of any of these four added orthogonal relationships in Figure 6.14 c. disabled the 4<sup>th</sup> VEA order reduction and produced a 5<sup>th</sup> order sequence. Thus, the fourth reduction of one VEA order requires the presence of

these four orthogonalities in addition to those required for the first 3 VEA orders of reduction.

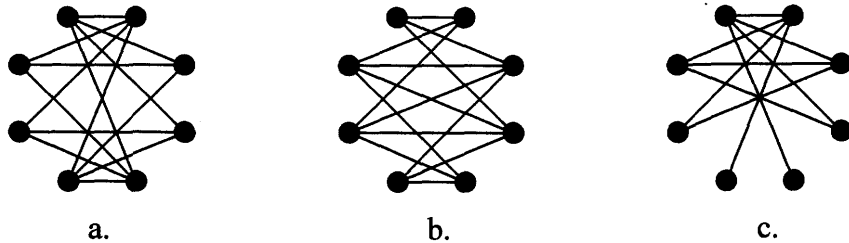


Figure 6.14 - a., b. Two maximal 5<sup>th</sup> order o-diagrams minus possible same-block relationships and c. the minimal 4<sup>th</sup> order o-diagram for  $V$  having 2 column blocks with 4 vectors each.

### 6.3 Summary and Generalising Conjecture

In light of the surprising finding (to the author) that none of the orthogonalities advantageous to reductions in VEA order are between column vectors in the same column block, an alternative form for such diagrams suggests itself in which the vectors within a column block appear evenly spaced along an undrawn vertical line with the same top-to-bottom ordering as previously. The two column blocks give rise to two such vertical columns. It is instructive to consider the orthogonalities shown in Figure 6.14 c. on such a revised o-diagram. It is shown in Figure 6.15 with 4 open circles strategically placed.

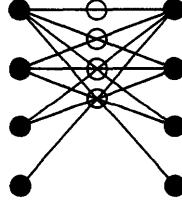


Figure 6.15 - The vertical format minimal 4<sup>th</sup> order o-diagram with central open circles for  $V$  having 2 column blocks with 4 vectors each.

Notice that the  $i^{\text{th}}$  open circle from the top has  $i$  segments passing through it and that the orthogonalities corresponding to these segments are those required for the  $i^{\text{th}}$  reduction in VEA order given a pre-existent reduction of  $i - 1$  VEA orders. This observation suggests a conjecture for all subsequent reductions in VEA order below Brezinski's upper bound. It is that a sufficient condition for a reduction of  $k$  VEA orders below Brezinski's upper bound is that every relation in the set of relations

$$\bigcup_{i=1}^k \bigcup_{j=1}^i \mathbf{v}_{1,j} \perp \mathbf{v}_{2,i+1-j} \quad (6.11)$$

hold. The author is convinced that it is true, though the experiments reported here only demonstrate its sufficiency up to  $N = 8$ . The author has also confirmed its sufficiency for  $N = 10$ . It may be that the condition is also necessary, though these experiments do not address this question.

## 6.4 Consequences of the Result

### 6.4.1 Hypersonic Flow Calculations

In light of this phenomenon whereby each order of the VEA negates the effect of two complex conjugate eigenvalues when associated with mutually orthogonal eigenvectors, it is interesting to note results published by Cheung *et al* [9]. Their work

applied Richardson relaxation to the problem of hypersonic flow calculations which make use of Steger-Warming flux-vector-splitting in the formation of their iteration matrix. (Recall that Richardson relaxation is one of the iterative schemes which generates a sequence according to equation (5.1).) They then accelerated the convergence of the resulting sequence using the VEA, consistently achieving a 50% reduction in computing time. Figure 6.16 shows a typical subset of the eigenvalues of the Richardson relaxation matrix  $A$  used in their solution procedures applied to inviscid flows. Figure 6.17 shows the same for viscous flow calculations.

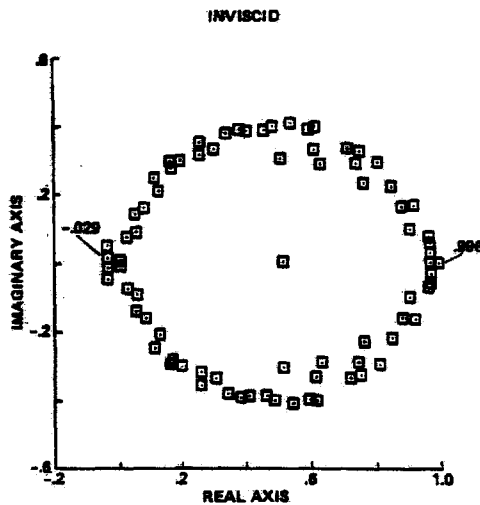


Fig. 1 Eigenvalues spectrum, inviscid calculation:  $M_\infty = 20$ ;  $\alpha = 0$ ; blunt cone grid  $63 \times 60$  with  $s = 1 \times 10^{-2}$ .

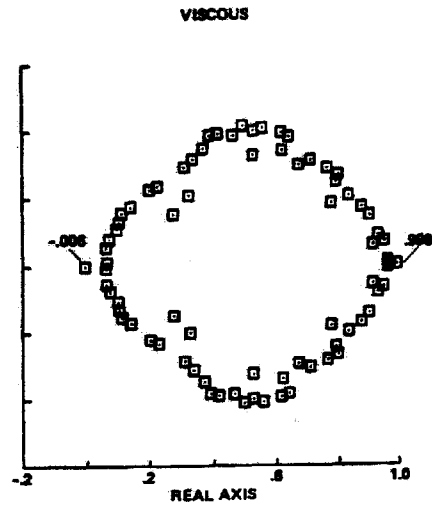


Fig. 2 Eigenvalues spectrum, viscous calculation:  $M_\infty = 20$ ;  $Re_\tau = 4 \times 10^6$ ;  $\alpha = 0$ ; blunt cone grid  $63 \times 60$  with  $s = 1 \times 10^{-2}$ .

Figure 6.16. Some of the eigenvalues of  $A$  for inviscid hypersonic flow calculations from Cheung *et al* [9].

Figure 6.17. Some of the eigenvalues of  $A$  for viscous hypersonic flow calculations from Cheung *et al* [9].

Applying the vertical axis for Figure 6.16 to both figures, notice that in both cases the eigenvalues come in complex conjugate pairs. If, in the presence of mutually orthogonal eigenvectors, each order of the VEA effectively zeroes two complex conjugate eigenvalues of  $A$ , it seems somewhat likely that this property of the VEA may have

been unwittingly exploited by Cheung *et al.* It would be very interesting to learn whether or not the complex conjugate eigenvalues of their  $\mathbf{A}$  matrices were in fact accompanied by mutually orthogonal eigenvectors.

This observation suggests to the author that it would be worth finding out which engineering applications routinely give rise to iteration matrices having complex conjugate eigenvalues whose corresponding eigenvectors are mutually orthogonal. Such applications are ripe for acceleration by the VEA.

#### 6.4.2 Consequences for the 1<sup>st</sup> Order VEA Kernel

The results of this chapter for  $N = 2$  are now explored a little more fully. They support the sufficient condition that every sequence  $\{\mathbf{e}_n\}$  whose terms  $\mathbf{e}_n$  satisfy

$$\mathbf{e}_n = \begin{bmatrix} | & | & | & | \\ \mathbf{v}_1 & \mathbf{v}_2 & & \\ | & | & | & | \end{bmatrix} \begin{bmatrix} \mu & \\ & \mu \end{bmatrix}^n \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}, \quad n = 0, 1, \dots \quad (6.12)$$

with nonzero 2-dimensional vectors  $\mathbf{v}_1 \perp \mathbf{v}_2$ , with  $\mu \neq 0$  and  $\mu \neq 1$  has VEA order of 1. Inequality (6.7) ensured that Brezinski's upper bound was 2. If it is permitted that either  $c_1$  or  $c_2$  of equation (6.12) is zero, then this lowers the upper bound to 1, giving the other way to obtain a VEA order of 1. Therefore, the condition

$$|c_1| + |c_2| > 0 \quad (6.13)$$

is added to the constraints on equation (6.12). Otherwise, allowing  $|c_1| + |c_2| = 0$  would mean that  $\mathbf{e}_n = \mathbf{0}$ ,  $n = 0, 1, \dots$  resulting in a constant vector sequence which may be thought of as having a VEA order of 0.

If each of  $\mathbf{v}_1$  and  $\mathbf{v}_2$  were extended to some dimension  $k > 2$  by zero padding (filling the new coefficient positions in each vector with zeroes), it is obvious that the resulting sequence will have the same order. By Theorem 2.5, each term of this sequence may, in turn, be premultiplied by any unitary matrix  $\mathbf{U}$  to produce a new sequence having the same VEA order and zero (anti)limit vector. For this reason, we may remove the constraint on the vectors  $\mathbf{v}_1$  and  $\mathbf{v}_2$  that they be 2-dimensional. Any two orthogonal nonzero vectors may be used to define  $\{\mathbf{e}_n\}$  and still give a sequence with VEA order of 1.

Calculating the product of equation (6.12) yields

$$\mathbf{e}_n = c_1 \mu^n \mathbf{v}_1 + c_2 \bar{\mu}^n \mathbf{v}_2, \quad n = 0, 1, \dots \quad (6.14)$$

Since the scaling of any eigenvector of any matrix is arbitrary, let  $\mathbf{w}_1 = c_1 \mathbf{v}_1$  and  $\mathbf{w}_2 = c_2 \mathbf{v}_2$ , giving the simpler result

$$\mathbf{e}_n = \mu^n \mathbf{w}_1 + \bar{\mu}^n \mathbf{w}_2, \quad n = 0, 1, \dots \quad (6.15)$$

where  $\mu \neq 0$ ,  $\mu \neq 1$ ,  $\|\mathbf{w}_1\| + \|\mathbf{w}_2\| > 0$ , and  $\mathbf{w}_1 \perp \mathbf{w}_2$ .

Notice that if  $\mu$  is real, the expression simplifies to

$$\mathbf{e}_n = \mu^n \mathbf{w}, \quad n = 0, 1, \dots \quad (6.16)$$

where  $\mathbf{w} = \mathbf{w}_1 + \mathbf{w}_2$ ,  $\mu \neq 0$ ,  $\mu \neq 1$  and  $\|\mathbf{w}\| > 0$ . If  $\mathbf{w}_2 = 0$ , then equation (6.16) still holds with  $\mathbf{w} = \mathbf{w}_1$ . In both cases, equation (6.16) is just a geometric progression which satisfies McLeod's sufficient condition for a 1<sup>st</sup> order VEA sequence.



Now, recall that throughout this thesis,  $\mathbf{e}_n$  has been defined by

$$\mathbf{e}_n = \mathbf{x}_n - \mathbf{x}, \quad n = 0, 1, \dots \quad (6.17)$$

Substituting this into equation (6.15) yields

$$\mathbf{x}_n = \mathbf{x} + \mu^n \mathbf{w}_1 + \bar{\mu}^n \mathbf{w}_2, \quad n = 0, 1, \dots \quad (6.18)$$

where  $\mu \neq 0$ ,  $\mu \neq 1$ ,  $\|\mathbf{w}_1\| + \|\mathbf{w}_2\| > 0$ , and  $\mathbf{w}_1 \perp \mathbf{w}_2$ . Thus, the experiments of this chapter indicate that equation (6.18) with subsequent qualifications is a sufficient condition for membership of a vector sequence in the kernel of the 1<sup>st</sup> order VEA.

The resemblance between equation (6.18) and equation (4.103) which is

$$\mathbf{x}_n = \mathbf{x} + z^n \mathbf{w} + \bar{z}^n \bar{\mathbf{w}}, \quad \text{for } n = 0, 1, 2, \dots \quad (6.19)$$

is striking. The only difference apart from nomenclature is that two orthogonal vectors which are complex conjugates are now replaced by two orthogonal vectors which are not complex conjugates and do not even necessarily have the same 2-norm. One of them may even be the zero vector. Recall that the condition that the terms of a *real* sequence  $\{\mathbf{x}_n\}$  satisfy equation (4.103) was both sufficient and necessary for  $\varepsilon_2^{(n)} = \mathbf{x}$ . Now that complex sequences are under consideration, this condition is only sufficient as only real vector sequences can satisfy it. By inspection, however, it is clear that equation (6.18) includes equation (6.19) as a special case.

It will be shown in the next chapter that the sufficient condition of equation (6.18) is also a necessary one.

## 7. THE COMPLEX KERNEL FOR THE 1<sup>ST</sup> ORDER VEA

In Chapters 3 and 4, the real part of the kernel of the 1<sup>st</sup> order VEA was shown. The result at the end of Chapter 4, given in equation (4.103), was that the kernel of the 1<sup>st</sup> order VEA is every vector sequence  $\{\mathbf{x}_n\}$  whose terms satisfy

$$\mathbf{x}_n = \mathbf{x} + z^n \mathbf{w} + \bar{z}^n \bar{\mathbf{w}}, \quad n = 0, 1, \dots \quad (7.1)$$

with complex  $z \neq 0$ ,  $z \neq 1$ ,  $\langle \mathbf{w}, \bar{\mathbf{w}} \rangle = 0$ , but  $\|\mathbf{w}\| > 0$ . Then, Chapters 5 and 6 explored complex vector sequences experimentally. It was determined (equation (6.18) with a small change in nomenclature) that every sequence  $\{\mathbf{x}_n\}$  whose terms  $\mathbf{x}_n$  satisfy

$$\mathbf{x}_n = \mathbf{x} + z^n \mathbf{w}_1 + \bar{z}^n \mathbf{w}_2, \quad n = 0, 1, \dots \quad (7.2)$$

with  $z \neq 0$ ,  $z \neq 1$ ,  $\|\mathbf{w}_1\| + \|\mathbf{w}_2\| > 0$ , and  $\mathbf{w}_1 \perp \mathbf{w}_2$  is a member of the kernel of the 1<sup>st</sup> order VEA. This condition, however, is only a sufficient one. There is yet no reason to believe that every possible type of complex 1<sup>st</sup> order VEA sequence has been examined and that every term of such a sequence must satisfy equation (7.2).

One step toward a necessary and sufficient condition for membership in the kernel of the 1<sup>st</sup> order VEA is to transform equation (7.1) according to equation (2.29) and apply Theorem 2.5. Therefore, let

$$\mathbf{t}_n = c\mathbf{U}\mathbf{x}_n + \mathbf{y}, \quad n = 0, 1, \dots \quad (7.3)$$

where  $c$  is a complex scalar,  $U$  is a unitary matrix,  $\mathbf{x}_n$  is the  $n^{\text{th}}$  term of  $\{\mathbf{x}_n\}$  which is a member of the real kernel of the 1<sup>st</sup> order VEA, and  $\mathbf{y}$  is an arbitrary complex vector. Substituting equation (7.1) into equation (7.3),

$$\mathbf{t}_n = cU(\mathbf{x} + z^n \mathbf{w} + \bar{z}^n \bar{\mathbf{w}}) + \mathbf{y} \quad (7.4)$$

or

$$\mathbf{t}_n = \mathbf{t} + z^n \mathbf{w}_1 + \bar{z}^n \mathbf{w}_2 \quad (7.5)$$

where

$$\mathbf{t} = cU\mathbf{x} + \mathbf{y}, \quad (7.6)$$

$$\mathbf{w}_1 = cU\mathbf{w}, \quad (7.7)$$

and

$$\mathbf{w}_2 = cU\bar{\mathbf{w}}. \quad (7.8)$$

If  $\mathbf{w}$  and its complex conjugate vector are orthogonal before undergoing a unitary transformation, then that will not change after the transformation. They will also continue to have the same lengths. Therefore,  $\mathbf{w}_1$  and  $\mathbf{w}_2$  are also mutually orthogonal and of equal length. By Theorem 2.5, sequences  $\{\mathbf{t}_n\}$  whose terms satisfy equation (7.5) with  $z \neq 0$ ,  $z \neq 1$ , and  $\mathbf{w}_1$  and  $\mathbf{w}_2$  orthogonal, of equal 2-norm and having 2-norms not totalling zero are also members of the kernel of the 1<sup>st</sup> order VEA.

Now, recall the origins of the constraint that the lengths of  $\mathbf{w}_1$  and  $\mathbf{w}_2$  be equal. In §4.1, it was the fact that  $\mathbf{e}_n$  was real that led to  $r_{n,1}$  and  $r_{n,2}$  in equation (4.8) being real. These, in turn, gave complex conjugate coefficients  $e^{i\phi_0}$  and  $e^{-i\phi_0}$  to the complex conjugate vectors  $\mathbf{u}$  and  $\overline{\mathbf{u}}$  creating  $\mathbf{w}$  and  $\overline{\mathbf{w}}$  in equation (4.50). In §4.2,  $\mathbf{e}_0$  being real led to  $c_1$  and  $c_2$  of equation (4.93) being complex conjugates and giving equation (4.95). This extension of the real part of the kernel to vector sequences whose terms lie in complex  $N$ -space leads naturally to the removal of this constraint.

Notice also that equation (7.5) is of the same form and satisfies all of the constraints of equation (7.2). Therefore, this expression will be used in the following definition. Let a  $\mathbf{zw}$  sequence be defined as a vector sequence  $\{\mathbf{x}_n\}$  whose terms are in complex  $N$ -space and satisfy

$$\mathbf{x}_n = \mathbf{x} + z^n \mathbf{w}_1 + \overline{z}^n \mathbf{w}_2 \quad (7.9)$$

with  $z \neq 0$ ,  $z \neq 1$ ,

$$\|\mathbf{w}_1\| + \|\mathbf{w}_2\| > 0 \quad (7.10)$$

and

$$\langle \mathbf{w}_1, \mathbf{w}_2 \rangle = 0. \quad (7.11)$$

If  $\mathbf{w}_2 = \overline{\mathbf{w}_1}$ , then  $\{\mathbf{x}_n\}$  is simply a  $\lambda\mathbf{R}$  sequence. Thus, the set of  $\mathbf{zw}$  sequences includes the set of  $\lambda\mathbf{R}$  sequences.

In §7.1, it will be demonstrated that any three distinct  $N$ -dimensional complex vectors  $\mathbf{x}_0$ ,  $\mathbf{x}_1$  and  $\mathbf{x}$  define a  $zw$  sequence. It will be shown how to solve for  $z$ ,  $\mathbf{w}_1$  and  $\mathbf{w}_2$  satisfying relations (7.9) through (7.11) for  $n = 0$  and  $n = 1$ . It will be seen that as long as  $\mathbf{x}_0$ ,  $\mathbf{x}_1$  and  $\mathbf{x}$  are distinct, some  $zw$  sequence can be defined which begins with  $\mathbf{x}_0$  and  $\mathbf{x}_1$  and has  $\mathbf{x}$  for its (anti)limit. In §7.2, it is proven that being  $zw$  is a necessary and sufficient condition for membership of a vector sequence in the kernel of the 1<sup>st</sup> order VEA.

### 7.1 Three Distinct Vectors Define a $zw$ Sequence

In order for  $\mathbf{x}_0$  and  $\mathbf{x}_1$  to be the first two vectors of a  $zw$  sequence having  $\mathbf{x}$  as a limit or antilimit, they must satisfy, respectively,

$$\mathbf{x}_0 = \mathbf{x} + \mathbf{w}_1 + \mathbf{w}_2 \quad (7.12)$$

and

$$\mathbf{x}_1 = \mathbf{x} + z\mathbf{w}_1 + \bar{z}\mathbf{w}_2 \quad (7.13)$$

where  $z \neq 0$ ,  $z \neq 1$ ,

$$\langle \mathbf{w}_1, \mathbf{w}_2 \rangle = 0 \quad (7.14)$$

and

$$\langle \mathbf{w}_1, \mathbf{w}_1 \rangle + \langle \mathbf{w}_2, \mathbf{w}_2 \rangle \neq 0. \quad (7.15)$$

If  $\mathbf{x}_0 = \mathbf{x}$ , then by equation (7.12),  $\mathbf{w}_1 = -\mathbf{w}_2$ . The only way, then, to satisfy equation (7.14) is for them to violate inequality (7.15). If  $\mathbf{x}_1 = \mathbf{x}$ , then by equation (7.13),  $z$  must violate  $z \neq 0$ . If  $\mathbf{x}_0 = \mathbf{x}_1$ , then  $z \neq 1$  is violated. Therefore,  $\mathbf{x}_0$ ,  $\mathbf{x}_1$  and  $\mathbf{x}$  satisfying relations (7.12-15) must be distinct.

These expressions are now solved for  $z$ ,  $\mathbf{w}_1$  and  $\mathbf{w}_2$  by defining

$$\mathbf{e}_0 = \mathbf{x}_0 - \mathbf{x} \quad (7.16)$$

and

$$\mathbf{e}_1 = \mathbf{x}_1 - \mathbf{x}. \quad (7.17)$$

Therefore,  $z$ ,  $\mathbf{w}_1$  and  $\mathbf{w}_2$  must be found such that  $\mathbf{e}_0$  and  $\mathbf{e}_1$  satisfy

$$\mathbf{e}_0 = \mathbf{w}_1 + \mathbf{w}_2. \quad (7.18)$$

and

$$\mathbf{e}_1 = z\mathbf{w}_1 + \bar{z}\mathbf{w}_2. \quad (7.19)$$

The three inner products involving  $\mathbf{e}_0$  and  $\mathbf{e}_1$  are now scrutinised in turn. First, consider the inner product

$$\langle \mathbf{e}_0, \mathbf{e}_0 \rangle = \langle \mathbf{w}_1 + \mathbf{w}_2, \mathbf{w}_1 + \mathbf{w}_2 \rangle. \quad (7.20)$$

This distributes to give

$$\langle \mathbf{e}_0, \mathbf{e}_0 \rangle = \langle \mathbf{w}_1, \mathbf{w}_1 \rangle + \langle \mathbf{w}_1, \mathbf{w}_2 \rangle + \langle \mathbf{w}_2, \mathbf{w}_1 \rangle + \langle \mathbf{w}_2, \mathbf{w}_2 \rangle. \quad (7.21)$$

By equation (7.14), this becomes

$$\langle \mathbf{e}_0, \mathbf{e}_0 \rangle = \langle \mathbf{w}_1, \mathbf{w}_1 \rangle + \langle \mathbf{w}_2, \mathbf{w}_2 \rangle. \quad (7.22)$$

Similarly, the inner product

$$\langle \mathbf{e}_1, \mathbf{e}_1 \rangle = \langle z\mathbf{w}_1 + \bar{z}\mathbf{w}_2, z\mathbf{w}_1 + \bar{z}\mathbf{w}_2 \rangle \quad (7.23)$$

becomes

$$\langle \mathbf{e}_1, \mathbf{e}_1 \rangle = z\bar{z}(\langle \mathbf{w}_1, \mathbf{w}_1 \rangle + \langle \mathbf{w}_2, \mathbf{w}_2 \rangle). \quad (7.24)$$

The scalar  $z$  may be expressed as

$$z = z_r + iz_i \quad (7.25)$$

where  $z_r$  and  $z_i$  are real scalars representing the real and imaginary parts of  $z$ , respectively. Substituting this into equation (7.24), the result is

$$\langle \mathbf{e}_1, \mathbf{e}_1 \rangle = (z_r + iz_i)(z_r - iz_i)(\langle \mathbf{w}_1, \mathbf{w}_1 \rangle + \langle \mathbf{w}_2, \mathbf{w}_2 \rangle) \quad (7.26)$$

which is

$$\langle \mathbf{e}_1, \mathbf{e}_1 \rangle = (z_r^2 + z_i^2)(\langle \mathbf{w}_1, \mathbf{w}_1 \rangle + \langle \mathbf{w}_2, \mathbf{w}_2 \rangle). \quad (7.27)$$

Therefore, by equations (7.27) and (7.22),

$$z_r^2 + z_i^2 = \frac{\langle \mathbf{e}_1, \mathbf{e}_1 \rangle}{\langle \mathbf{e}_0, \mathbf{e}_0 \rangle} \quad (7.28)$$

holds.

Now, consider

$$\langle \mathbf{e}_0, \mathbf{e}_1 \rangle = \langle \mathbf{w}_1 + \mathbf{w}_2, z\mathbf{w}_1 + \bar{z}\mathbf{w}_2 \rangle \quad (7.29)$$

which becomes

$$\langle \mathbf{e}_0, \mathbf{e}_1 \rangle = \bar{z}\langle \mathbf{w}_1, \mathbf{w}_1 \rangle + z\langle \mathbf{w}_2, \mathbf{w}_2 \rangle. \quad (7.30)$$

Substituting equation (7.25) into equation (7.30) gives

$$\langle \mathbf{e}_0, \mathbf{e}_1 \rangle = (z_r - iz_i)\langle \mathbf{w}_1, \mathbf{w}_1 \rangle + (z_r + iz_i)\langle \mathbf{w}_2, \mathbf{w}_2 \rangle \quad (7.31)$$

or

$$\langle \mathbf{e}_0, \mathbf{e}_1 \rangle = z_r(\langle \mathbf{w}_1, \mathbf{w}_1 \rangle + \langle \mathbf{w}_2, \mathbf{w}_2 \rangle) + iz_i(\langle \mathbf{w}_2, \mathbf{w}_2 \rangle - \langle \mathbf{w}_1, \mathbf{w}_1 \rangle). \quad (7.32)$$

Therefore, by equation (7.22),

$$z_r = \frac{\operatorname{Re}(\langle \mathbf{e}_1, \mathbf{e}_0 \rangle)}{\langle \mathbf{e}_0, \mathbf{e}_0 \rangle}, \quad (7.33)$$

and by equation (7.28),

$$z_i^2 = \frac{\langle \mathbf{e}_1, \mathbf{e}_1 \rangle}{\langle \mathbf{e}_0, \mathbf{e}_0 \rangle} - \left( \frac{\operatorname{Re}(\langle \mathbf{e}_1, \mathbf{e}_0 \rangle)}{\langle \mathbf{e}_0, \mathbf{e}_0 \rangle} \right)^2. \quad (7.34)$$

It must now be decided whether  $z_i$  will be the positive or the negative root. For now, the positive root will be chosen in what will be called case A. Case B where  $z_i$  is the negative root will be addressed at the end of the section. Therefore, for case A,



$$z_i = \sqrt{\frac{\langle \mathbf{e}_1, \mathbf{e}_1 \rangle}{\langle \mathbf{e}_0, \mathbf{e}_0 \rangle} - \left( \frac{\text{Re}(\langle \mathbf{e}_1, \mathbf{e}_0 \rangle)}{\langle \mathbf{e}_0, \mathbf{e}_0 \rangle} \right)^2} \quad (7.35)$$

and  $z$  is found. At this point, equations (7.18) and (7.19) can be solved for  $\mathbf{w}_1$  and  $\mathbf{w}_2$ . If  $z$  is real, then  $z = \bar{z}$  and equations (7.18) and (7.19) are not independent. In this case, the values

$$\mathbf{w}_1 = \mathbf{e}_0 \quad (7.36)$$

and

$$\mathbf{w}_2 = \mathbf{0} \quad (7.37)$$

satisfy both equations and all the constraints on  $\mathbf{w}_1$  and  $\mathbf{w}_2$ . If  $z$  is not real, then equations (7.18) and (7.19) are independent and equation (7.18) may be multiplied by  $\bar{z}$  and subtracted from equation (7.19) to give

$$\mathbf{e}_1 - \bar{z}\mathbf{e}_0 = (z - \bar{z})\mathbf{w}_1 \quad (7.38)$$

or

$$\mathbf{w}_1 = \frac{\mathbf{e}_1 - \bar{z}\mathbf{e}_0}{(z - \bar{z})}. \quad (7.39)$$

Similarly,  $\mathbf{w}_1$  may be eliminated to obtain

$$\mathbf{w}_2 = \frac{\mathbf{e}_1 - z\mathbf{e}_0}{(\bar{z} - z)}. \quad (7.40)$$

Now, consider case B where  $z_i$  is the negative of the root in equation (7.35). In this case, the resulting value for  $z$  would be the complex conjugate of  $z$  for case A. From equations (7.39) and (7.40),  $\mathbf{w}_1$  for case B would equal  $\mathbf{w}_2$  for case A and  $\mathbf{w}_2$  for case B would equal  $\mathbf{w}_1$  for case A. Therefore, the only consequence of the choice is which vector receives which label.

## 7.2 The Complex Kernel of the 1<sup>st</sup> Order VEA

The section begins with

Theorem 7.1: Given a vector sequence  $\{\mathbf{x}_n\}$  whose terms are complex and  $N$ -dimensional,  $\mathbf{x}_2^{(n)} = \mathbf{x}$  for  $n = 0, 1, 2, \dots$  if and only if

$$\mathbf{x}_n = \mathbf{x} + z^n \mathbf{w}_1 + \bar{z}^n \mathbf{w}_2, \quad \text{for } n = 0, 1, \dots \quad (7.41)$$

where  $z$  is a complex scalar,  $z \neq 0$ ,  $z \neq 1$ ,

$$\langle \mathbf{w}_1, \mathbf{w}_1 \rangle + \langle \mathbf{w}_2, \mathbf{w}_2 \rangle > 0, \quad (7.42)$$

and

$$\langle \mathbf{w}_1, \mathbf{w}_2 \rangle = 0. \quad (7.43)$$

The conditions given in Theorem 7.1 are simply the defining attributes of a  $zw$  sequence. Therefore, the theorem may be restated as follows: the kernel of the 1<sup>st</sup> order VEA is the set of all  $zw$  sequences.

Notice that any three consecutive vectors of a  $zw$  sequence

$$\mathbf{x}_n = \mathbf{x} + z^n \mathbf{w}_1 + \bar{z}^n \mathbf{w}_2, \quad n = k, k+1, k+2 \quad (7.44)$$

may be chosen to define the first three vectors

$$\mathbf{x}'_n = \mathbf{x} + z^n \mathbf{w}'_1 + \bar{z}^n \mathbf{w}'_2, \quad n = 0, 1, 2, \quad (7.45)$$

of the related  $zw$  sequence  $\{\mathbf{x}'_n\}$  for which

$$\mathbf{w}'_1 = z^k \mathbf{w}_1 \quad (7.46)$$

and

$$\mathbf{w}'_2 = \bar{z}^k \mathbf{w}_2. \quad (7.47)$$

Therefore, a proof for  $\mathbf{x}_0$ ,  $\mathbf{x}_1$  and  $\mathbf{x}_2$  will apply for any  $n$  as long as every  $\mathbf{x}_n$  satisfies the constraints given in the theorem and the definition of a  $zw$  sequence.

The proof for this theorem appears in the following two subsections. Subsection 7.2.2 gives the proof that a sequence being  $zw$  is sufficient to make it a member of the kernel of the 1<sup>st</sup> order VEA, while Subsection 7.2. gives the proof that a sequence being  $zw$  is necessary for it to be a member of the kernel of the 1<sup>st</sup> order VEA.

### 7.2.1 Proof that Being $zw$ is Sufficient

In order to make the differences between vectors of the  $\varepsilon$ -table more plain than is possible using superscripts and subscripts, the symbols  $\mathbf{x}_0$ ,  $\mathbf{x}_1$ ,  $\mathbf{x}_2$ ,  $\mathbf{a}$ ,  $\mathbf{b}$  and  $\mathbf{c}$  will be used in place of  $\varepsilon_0^{(0)}$ ,  $\varepsilon_0^{(1)}$ ,  $\varepsilon_0^{(2)}$ ,  $\varepsilon_1^{(0)}$ ,  $\varepsilon_1^{(1)}$ , and  $\varepsilon_2^{(0)}$ , respectively, as shown in table 7.1.

Table 7.1 The  $\varepsilon$ -table used for the complex kernel proof.

$$\begin{array}{ccc}
 & \varepsilon_0^{(0)} = \mathbf{x}_0 & \\
 \varepsilon_{-1}^{(1)} = \mathbf{0} & & \varepsilon_1^{(0)} = \mathbf{a} \\
 & \varepsilon_0^{(1)} = \mathbf{x}_1 & \varepsilon_2^{(0)} = \mathbf{c} \\
 \varepsilon_{-1}^{(2)} = \mathbf{0} & & \varepsilon_1^{(1)} = \mathbf{b} \\
 & \varepsilon_0^{(2)} = \mathbf{x}_2 & \vdots \\
 \vdots & & \vdots \\
 & \vdots & 
 \end{array}$$

To simplify the repeated calculations in this proof, an identity is presented.

Given a vector  $\mathbf{y}$  satisfying

$$\mathbf{y} = \frac{\alpha \mathbf{w}_1 + \bar{\alpha} \mathbf{w}_2}{\beta} \tag{7.48}$$

where  $\alpha$  is a complex scalar and  $\beta$  is a real scalar, with  $\mathbf{w}_1$  and  $\mathbf{w}_2$  being mutually orthogonal complex vectors, the inverse of  $\mathbf{y}$  according to equation (2.22) which is

$$\mathbf{x}^{-1} = \frac{\bar{\mathbf{x}}}{\langle \mathbf{x}, \mathbf{x} \rangle} \tag{7.49}$$

will be

$$\mathbf{y}^{-1} = \frac{\overline{\left( \frac{\alpha \mathbf{w}_1 + \bar{\alpha} \mathbf{w}_2}{\beta} \right)}}{\left\langle \frac{\alpha \mathbf{w}_1 + \bar{\alpha} \mathbf{w}_2}{\beta}, \frac{\alpha \mathbf{w}_1 + \bar{\alpha} \mathbf{w}_2}{\beta} \right\rangle} \tag{7.50}$$

which is

$$y^{-1} = \frac{\frac{\overline{\alpha w_1 + \alpha w_2}}{\beta}}{\frac{\alpha \overline{\langle w_1, w_1 \rangle + \langle w_2, w_2 \rangle}}{\beta \beta}} \quad (7.51)$$

or

$$y^{-1} = \frac{\beta(\overline{\alpha w_1 + \alpha w_2})}{\alpha \overline{\langle w_1, w_1 \rangle + \langle w_2, w_2 \rangle}}. \quad (7.52)$$

To begin, by equation (7.41),

$$x_0 = x + w_1 + w_2, \quad (7.53)$$

and

$$x_1 = x + z w_1 + \bar{z} w_2. \quad (7.54)$$

By equation (2.26),

$$a = (x_1 - x_0)^{-1}, \quad (7.55)$$

so equation (7.53) is subtracted from equation (7.54) to obtain

$$x_1 - x_0 = (z-1)w_1 + (\bar{z}-1)w_2. \quad (7.56)$$

By equation (7.55) and using the identity of equation (7.52) with the definitions in equation (7.48),

$$a = \frac{(\bar{z}-1)\overline{w_1} + (z-1)\overline{w_2}}{(z-1)(\bar{z}-1)(\langle w_1, w_1 \rangle + \langle w_2, w_2 \rangle)} \quad (7.57)$$

is obtained. Also by equation (7.41),

$$\mathbf{x}_2 = \mathbf{x} + z^2 \mathbf{w}_1 + \bar{z}^{-2} \mathbf{w}_2. \quad (7.58)$$

By equation (2.26),

$$\mathbf{b} = (\mathbf{x}_2 - \mathbf{x}_1)^{-1}, \quad (7.59)$$

so equation (7.54) is subtracted from equation (7.58) and

$$\mathbf{x}_2 - \mathbf{x}_1 = (z^2 - z) \mathbf{w}_1 + \left( \bar{z}^{-2} - \bar{z} \right) \mathbf{w}_2 \quad (7.60)$$

or

$$\mathbf{x}_2 - \mathbf{x}_1 = z(z-1) \mathbf{w}_1 + \bar{z}(\bar{z}-1) \mathbf{w}_2 \quad (7.61)$$

is obtained. By equation (7.59) and using equations (7.48) and (7.52),

$$\mathbf{b} = \frac{\bar{z}(\bar{z}-1) \overline{\mathbf{w}_1} + z(z-1) \overline{\mathbf{w}_2}}{z \bar{z} (z-1) (\bar{z}-1) (\langle \mathbf{w}_1, \mathbf{w}_1 \rangle + \langle \mathbf{w}_2, \mathbf{w}_2 \rangle)}. \quad (7.62)$$

Also by equation (2.26),

$$\mathbf{c} = \mathbf{x}_1 + (\mathbf{b} - \mathbf{a})^{-1}. \quad (7.63)$$

Therefore, equation (7.57) is given a common denominator with (7.62) and

$$\mathbf{a} = \frac{z \bar{z} (\bar{z}-1) \overline{\mathbf{w}_1} + \bar{z} z (z-1) \overline{\mathbf{w}_2}}{z \bar{z} (z-1) (\bar{z}-1) (\langle \mathbf{w}_1, \mathbf{w}_1 \rangle + \langle \mathbf{w}_2, \mathbf{w}_2 \rangle)} \quad (7.64)$$

is obtained. Subtracting this from equation (7.62), the resulting expression is

$$\mathbf{b} - \mathbf{a} = \frac{(1-z)\bar{z}(\bar{z}-1)\overline{\mathbf{w}_1} + (1-\bar{z})z(z-1)\overline{\mathbf{w}_2}}{z\bar{z}(z-1)(\bar{z}-1)(\langle \mathbf{w}_1, \mathbf{w}_1 \rangle + \langle \mathbf{w}_2, \mathbf{w}_2 \rangle)} \quad (7.65)$$

which simplifies to become

$$\mathbf{b} - \mathbf{a} = \frac{-z\overline{\mathbf{w}_1} - \bar{z}\overline{\mathbf{w}_2}}{z\bar{z}(\langle \mathbf{w}_1, \mathbf{w}_1 \rangle + \langle \mathbf{w}_2, \mathbf{w}_2 \rangle)}. \quad (7.66)$$

Inverting this according to equations (7.48) and (7.52) gives

$$(\mathbf{b} - \mathbf{a})^{-1} = \frac{z\bar{z}(\langle \mathbf{w}_1, \mathbf{w}_1 \rangle + \langle \mathbf{w}_2, \mathbf{w}_2 \rangle)(-z\overline{\mathbf{w}_1} - \bar{z}\overline{\mathbf{w}_2})}{z\bar{z}(\langle \mathbf{w}_1, \mathbf{w}_1 \rangle + \langle \mathbf{w}_2, \mathbf{w}_2 \rangle)} \quad (7.67)$$

or

$$(\mathbf{b} - \mathbf{a})^{-1} = -z\overline{\mathbf{w}_1} - \bar{z}\overline{\mathbf{w}_2}. \quad (7.68)$$

When this result with equation (7.54) is substituted into equation (7.63), the result is

$$\mathbf{c} = \mathbf{x} + z\overline{\mathbf{w}_1} + \bar{z}\overline{\mathbf{w}_2} - z\overline{\mathbf{w}_1} - \bar{z}\overline{\mathbf{w}_2} \quad (7.69)$$

or

$$\mathbf{c} = \mathbf{x}. \quad (7.70)$$

Notice the presence of  $\langle \mathbf{w}_1, \mathbf{w}_1 \rangle + \langle \mathbf{w}_2, \mathbf{w}_2 \rangle$  as well as  $z$  and  $z - 1$  with their complex conjugates in the denominator in equation (7.62). This leads to the constraints on  $z$ ,  $\mathbf{w}_1$  and  $\mathbf{w}_2$  given in the theorem as well as the definition of a  $z\overline{\mathbf{w}}$  sequence. Thus, a

sequence being  $\mathbf{zw}$  is a sufficient condition for its membership in the kernel of the 1<sup>st</sup> order VEA.

### 7.2.2 Proof that being $\mathbf{zw}$ is necessary

In order to prove that being  $\mathbf{zw}$  is a necessary condition for membership of a sequence in the kernel of the 1<sup>st</sup> order VEA, it must be shown that every sequence that the 1<sup>st</sup> order VEA transforms to its limit must be a  $\mathbf{zw}$  sequence. Therefore, it will be assumed that the 1<sup>st</sup> order VEA applied to  $\mathbf{x}_0$ ,  $\mathbf{x}_1$  and  $\mathbf{x}_2$  results in  $\boldsymbol{\varepsilon}_2^{(0)} = \mathbf{x}$ . Then, it will be determined what conditions these assumptions impose upon  $\mathbf{x}_2$ .

By definition, if  $\{\mathbf{x}_n\}$  is in the kernel of the 1<sup>st</sup> order VEA, then by equations (2.25) and (2.26),

$$\boldsymbol{\varepsilon}_1^{(n)} = (\mathbf{x}_{n+1} - \mathbf{x}_n)^{-1}. \quad (7.71)$$

Also by equation (2.26),

$$\boldsymbol{\varepsilon}_2^{(n)} = \mathbf{x}_{n+1} + \left[ (\mathbf{x}_{n+2} - \mathbf{x}_{n+1})^{-1} - (\mathbf{x}_{n+1} - \mathbf{x}_n)^{-1} \right]. \quad (7.72)$$

For  $n = 0$ , given  $\boldsymbol{\varepsilon}_2^{(0)} = \mathbf{x}$ ,

$$\mathbf{x} = \mathbf{x}_1 + \left[ (\mathbf{x}_2 - \mathbf{x}_1)^{-1} - (\mathbf{x}_1 - \mathbf{x}_0)^{-1} \right]^{-1}. \quad (7.73)$$

This expression can be solved for  $\mathbf{x}_2$ . Subtracting  $\mathbf{x}_1$  and inverting yields

$$(\mathbf{x} - \mathbf{x}_1)^{-1} = (\mathbf{x}_2 - \mathbf{x}_1)^{-1} - (\mathbf{x}_1 - \mathbf{x}_0)^{-1}. \quad (7.74)$$



Adding  $(\mathbf{x}_1 - \mathbf{x}_0)^{-1}$  and inverting gives

$$\mathbf{x}_2 - \mathbf{x}_1 = \left[ (\mathbf{x} - \mathbf{x}_1)^{-1} + (\mathbf{x}_1 - \mathbf{x}_0)^{-1} \right]^{-1} \quad (7.75)$$

or

$$\mathbf{x}_2 = \mathbf{x}_1 + \left[ (\mathbf{x} - \mathbf{x}_1)^{-1} + (\mathbf{x}_1 - \mathbf{x}_0)^{-1} \right]^{-1}. \quad (7.76)$$

If a breakdown in the computation of  $\mathbf{x}_2$  is to be avoided, then  $\mathbf{x}_1 \neq \mathbf{x}$ ,  $\mathbf{x}_0 \neq \mathbf{x}_1$ , and  $\mathbf{x} \neq \mathbf{x}_0$ .<sup>7</sup> Thus, if  $\{\mathbf{x}_n\}$  is in the kernel of the 1<sup>st</sup> order VEA, then  $\mathbf{x}_0$ ,  $\mathbf{x}_1$ , and  $\mathbf{x}$  are distinct.

Recall from §7.1 that any three distinct points  $\mathbf{x}_0$ ,  $\mathbf{x}_1$  and  $\mathbf{x}$  may be used to define a  $z\mathbf{w}$  sequence. Is  $\mathbf{x}_2$  as defined by equation (7.76) also a member of the  $z\mathbf{w}$  sequence defined by  $\mathbf{x}_0$ ,  $\mathbf{x}_1$  and  $\mathbf{x}$ ? If so, then being  $z\mathbf{w}$  is necessary for membership in the kernel of the 1<sup>st</sup> order VEA.

Since  $z$ ,  $\mathbf{w}_1$  and  $\mathbf{w}_2$  can always be found from distinct  $\mathbf{x}_0$ ,  $\mathbf{x}_1$  and  $\mathbf{x}$ , such that

$$\mathbf{x}_0 = \mathbf{x} + \mathbf{w}_1 + \mathbf{w}_2, \quad (7.77)$$

and

$$\mathbf{x}_1 = \mathbf{x} + z\mathbf{w}_1 + \bar{z}\mathbf{w}_2 \quad (7.78)$$

where  $z$  is a complex scalar,  $z \neq 0$ ,  $z \neq 1$ ,

$$\langle \mathbf{w}_1, \mathbf{w}_1 \rangle + \langle \mathbf{w}_2, \mathbf{w}_2 \rangle \neq 0, \quad (7.79)$$

---

<sup>7</sup> Note that two of these breakdowns are easily overcome by replacing  $\mathbf{x}_2$  with the limit of  $\mathbf{x}_2$  as  $\mathbf{x}_1$  approaches  $\mathbf{x}$  or  $\mathbf{x}_0$ .

and

$$\langle \mathbf{w}_1, \mathbf{w}_2 \rangle = 0, \quad (7.80)$$

equations (7.77) and (7.78) may be substituted into equation (7.76) to determine a formula for  $\mathbf{x}_2$  such that  $\{\mathbf{x}_n\}$  is in the 1<sup>st</sup> order VEA kernel.

By equation (7.78),

$$\mathbf{x} - \mathbf{x}_1 = -z\mathbf{w}_1 - \bar{z}\mathbf{w}_2. \quad (7.81)$$

By the inverting identity of equation (7.52),

$$(\mathbf{x} - \mathbf{x}_1)^{-1} = \frac{-\bar{z}\mathbf{w}_1 - z\mathbf{w}_2}{z\bar{z}(\langle \mathbf{w}_1, \mathbf{w}_1 \rangle + \langle \mathbf{w}_2, \mathbf{w}_2 \rangle)}. \quad (7.82)$$

Subtracting equation (7.77) from (7.78),

$$\mathbf{x}_1 - \mathbf{x}_0 = (z-1)\mathbf{w}_1 + (\bar{z}-1)\mathbf{w}_2 \quad (7.83)$$

Inverting this by equation (7.52),

$$(\mathbf{x}_1 - \mathbf{x}_0)^{-1} = \frac{(\bar{z}-1)\bar{\mathbf{w}}_1 + (z-1)\bar{\mathbf{w}}_2}{(z-1)(\bar{z}-1)(\langle \mathbf{w}_1, \mathbf{w}_1 \rangle + \langle \mathbf{w}_2, \mathbf{w}_2 \rangle)}. \quad (7.84)$$

Giving equations (7.82) and (7.84) common denominators,

$$(\mathbf{x} - \mathbf{x}_1)^{-1} = \frac{-\bar{z}(z-1)(\bar{z}-1)\bar{\mathbf{w}}_1 - z(z-1)(\bar{z}-1)\bar{\mathbf{w}}_2}{z\bar{z}(z-1)(\bar{z}-1)(\langle \mathbf{w}_1, \mathbf{w}_1 \rangle + \langle \mathbf{w}_2, \mathbf{w}_2 \rangle)}, \quad (7.85)$$

and

$$(\mathbf{x}_1 - \mathbf{x}_0)^{-1} = \frac{z\bar{z}(\bar{z}-1)\bar{\mathbf{w}}_1 + z\bar{z}(z-1)\bar{\mathbf{w}}_2}{z\bar{z}(z-1)(\bar{z}-1)(\langle \mathbf{w}_1, \mathbf{w}_1 \rangle + \langle \mathbf{w}_2, \mathbf{w}_2 \rangle)}. \quad (7.86)$$

Adding equations (7.85) and (7.86) and simplifying gives

$$(\mathbf{x} - \mathbf{x}_1)^{-1} + (\mathbf{x}_1 - \mathbf{x}_0)^{-1} = \frac{\bar{z}(\bar{z}-1)\bar{\mathbf{w}}_1 + z(z-1)\bar{\mathbf{w}}_2}{z\bar{z}(z-1)(\bar{z}-1)(\langle \mathbf{w}_1, \mathbf{w}_1 \rangle + \langle \mathbf{w}_2, \mathbf{w}_2 \rangle)}. \quad (7.87)$$

Inverting this by equation (7.52) and simplifying yields

$$\left[ (\mathbf{x} - \mathbf{x}_1)^{-1} + (\mathbf{x}_1 - \mathbf{x}_0)^{-1} \right]^{-1} = z(z-1)\mathbf{w}_1 + \bar{z}(\bar{z}-1)\mathbf{w}_2. \quad (7.88)$$

Substituting this and equation (7.78) into equation (7.76) gives

$$\mathbf{x}_2 = \mathbf{x} + z\mathbf{w}_1 + \bar{z}\mathbf{w}_2 + z(z-1)\mathbf{w}_1 + \bar{z}(\bar{z}-1)\mathbf{w}_2, \quad (7.89)$$

which simplifies to become

$$\mathbf{x}_2 = \mathbf{x} + z^2\mathbf{w}_1 + \bar{z}^2\mathbf{w}_2. \quad (7.90)$$

This procedure may be repeated with  $\mathbf{x}_0'$ ,  $\mathbf{x}_1'$  and  $\mathbf{x}$  where

$$\mathbf{x}'_n = \mathbf{x} + z^n(z\mathbf{w}_1) + \bar{z}^n(\bar{z}\mathbf{w}_2) = \mathbf{x}_{n+1} \quad (7.91)$$

to show by induction that

$$\mathbf{x}_n = \mathbf{x} + z^n\mathbf{w}_1 + \bar{z}^n\mathbf{w}_2, \quad n = 0, 1, \dots \quad (7.92)$$

Thus, in order for the 1<sup>st</sup> order VEA to yield a constant sequence, the original sequence must be a  $zw$  sequence. ■

## 8. FUTURE WORK

There are a number of unanswered questions which arise fairly directly from the work reported in this thesis, and seem to the author to be not out of reach. The answers to each of them will contribute in some significant way to a practical understanding of the VEA, its kernel, and its most apt uses. These questions include the following.

### 8.1 Orthogonality & Multiple Complex Conjugate Jordan Blocks

The experiments in Chapter 6 laid some useful ground work but fall short of the broadest possible sufficient condition for vector sequences  $\{\mathbf{x}_n\}$  whose terms  $\mathbf{x}_n$ ,  $n = 0, 1, \dots$  satisfy equation (5.1) to be members of the lowest possible order kernel of the VEA. What are the orthogonality relations required between members of each of the column blocks for each reduction in VEA order when the Jordan form of  $\mathbf{A}$  in equation (5.1) has multiple Jordan blocks for each of two complex conjugate eigenvalues? Do the critical relationships continue to be between the  $\mathbf{v}_{ij}$  or are they actually between the linear combinations of the  $\mathbf{v}_{ij}$  associated with each of the two eigenvalues present in  $\mathbf{e}_0$ ?

Answers to this question should emerge readily from slightly more sophisticated sets of experiments (run on something a little bigger and faster than an HP 48GX).

### 8.2 Overshadowing and Linear Independence

Due to the procedure illustrated in equation (6.9), all the matrices  $\mathbf{V}$  generated for the experiments of Chapter 6 were nonsingular. This was done somewhat arbitrarily for consistency with the theory presented earlier where  $\mathbf{A} = \mathbf{V}\mathbf{J}\mathbf{V}^{-1}$ . If  $\mathbf{V}$  were singular, there could be no  $\mathbf{V}^{-1}$  and the resulting sequence would no longer be computable by

equation (5.1). That said, are there any constraints on the linear independence of the column vectors  $\mathbf{v}_{ij}$  that pertain to the phenomenon of overshadowing for vector sequences  $\{\mathbf{x}_n\}$  where  $\mathbf{x}_n = \mathbf{V}\mathbf{J}^n\mathbf{c}$  with  $\mathbf{J}$  being in Jordan canonical form having two complex conjugate Jordan blocks? Can the same column vector appear in more than one place in a given  $\mathbf{V}_i$  without affecting the VEA order of the resulting sequence?

### 8.3 The Generality of an Empirical Description of the 2<sup>nd</sup> order VEA

It is hoped that, once the experiments described in the previous sections have been completed, an experimentally determined formula will be suggested for the kernel of the 2<sup>nd</sup> order VEA. Can the parameters present in such a formula be adjusted in such a way that it fits any sequence of 5 vectors  $\mathbf{x}_n$ ,  $n = 0, 1, 2, 3$ , and 4 for which  $\epsilon_4^{(0)}$  is defined? Can these parameters be solved for uniquely? Alternatively, can the parameters present in such a formula be adjusted in such a way that it fits any sequence of 4 vectors  $\mathbf{x}_n$ ,  $n = 0, 1, 2$ , and 3, and an arbitrary limit vector  $\epsilon_4^{(0)}$  which will allow  $\mathbf{x}_4 = \epsilon_0^{(4)}$  to be calculated? If so, then there must be a proof analogous to that given in Chapter 7 for the kernel of the 2<sup>nd</sup> order VEA. If there do exist sequences having a 2<sup>nd</sup> order VEA limit which cannot be obtained by adjusting the parameters for such a formula, what is the nature of those sequences  $\{\mathbf{x}_n\}$  not fitting the formula and yet for which there exists (anti)limit  $\mathbf{x} = \epsilon_4^{(n)}$ ,  $n = 0, 1, \dots$ ? Such sequences, if they exist will probably suggest another set of experiments that may encompass all sequences beginning with any  $\mathbf{x}_0$ ,  $\mathbf{x}_1$ ,  $\mathbf{x}_2$ , and  $\mathbf{x}_3$ , having a defined  $\epsilon_4^{(0)}$ , leading to a proof for the second order kernel of the VEA analogous to that of Chapter 7.

## 8.4 Applications with VEA Order-Reducing Orthogonalities

In Subsection 6.4.1, an application of the VEA involving an iteration matrix  $A$  having many pairs of complex conjugate eigenvalues was noted. The results and conjecture of §6.3 indicate that if the eigenvectors associated with complex conjugate eigenvalues in the iteration matrices of such applications are mutually orthogonal, then the VEA can be expected to accelerate the convergence of these calculations especially well. Do hypersonic flow calculations which use Steger-Warming flux-vector-splitting indeed employ iteration matrices whose eigenvectors associated with complex conjugate eigenvalues are mutually orthogonal? If the associated Jordan blocks are larger than  $1 \times 1$ , do they possess any of the orthogonal relationships shown to result in further reductions in VEA order? What other iterative solution procedures currently used in engineering problems generate such an iteration matrix? Can iterative procedures which do not create such an iteration matrix be modified to do so in order to take advantage of this particular strength of the VEA?

## REFERENCES

- [1] A. C. Aitken. On Bernoulli's numerical solution of algebraic equations, *Proceedings of the Royal Society of Edinburgh*, 46:289-305, 1926.
- [2] C. Brezinski. Application de l' $\epsilon$ -algorithme à la résolution des systèmes non linéaires, *Comptes Rendus Hebdomadaire des Séances de l'Académie des Sciences Paris, Serie A*, 271 A:1174-1177, 1970.
- [3] C. Brezinski. Some results in the theory of the vector  $\epsilon$ -algorithm, *Linear Algebra and its Applications*, 8:77-86, 1974
- [4] C. Brezinski. *Accélération de la convergence en analyse numérique*, Lecture Notes in Mathematics v. 584. Springer, New York, NY, 1977.
- [5] C. Brezinski. *Algorithmes d'Accélération de la Convergence, Etude Numérique*. Editions Technip, Paris, 1978.
- [6] C. Brezinski. Convergence acceleration during the 20<sup>th</sup> century, *Journal of Computational and Applied Mathematics*, 122:1-21, 2000.
- [7] C. Brezinski and M. Crouzeix. Remarques sur le procédé  $\Delta^2$  d'Aitken, *Comptes Rendus Hebdomadaire des Séances de l'Académie des Sciences Paris, Serie A*, 270 A:896-898, 1970.
- [8] C. Brezinski and M. Redivo-Zaglia. *Extrapolation methods: theory and practice*, North-Holland, Amsterdam, 1991.
- [9] S. Cheung, A. Cheer, M. Hafez, and J. Flores. Convergence acceleration of viscous and inviscid hypersonic flow calculations. *AIAA Journal*, 29:1214-1222, 1991.
- [10] F. Cordellier. Démonstration algébrique de l'extension de l'identité de Wynn aux tables de Padé non normales. In L. Wuytack, ed., *Padé Approximation and its Applications*, v. 765 of *LNM*, pp. 36-60, Berlin, 1979. Springer-Verlag.
- [11] A. Dolovich and G.W. Brodland. Convergence acceleration for iterative finite-element methods. *Journal of Engineering Mechanics*, 121:1-6, 1995.
- [12] E. Gekeler. On the solution of systems of equations by the epsilon algorithm of Wynn. *Mathematics of Computation*, 26(118):427-436, April 1972.
- [13] P.R. Graves-Morris. Vector valued rational interpolants I. *Numerische Mathematik*, 42:331-348, 1983.
- [14] M. M. Hafez and H. K. Cheng. An acceleration technique related to Wynn's algorithm with application to transonic flow calculations. *Proceedings of the 5th International Conference on Numerical Methods in Fluid Dynamics*, 1976.
- [15] M.M. Hafez, S. Palaniswamy, G. Kuruvila, and M.D. Salas. Applications of Wynn's  $\epsilon$ -algorithm to transonic flow calculation. *AIAA Paper 87-1143*, 1987.



- [16] S. Lipschutz. *Theory and Problems of Linear Algebra, SI (metric) edition*. Schaum's outline series, McGraw-Hill, New York, 1974.
- [17] D.S. Lowe. *Enhanced Convergence of Algebraic Reconstruction Algorithms Used in Computerised Tomography*. M.Sc. thesis, University of Saskatchewan, 1996.
- [18] J.B. McLeod. A note on the  $\varepsilon$ -algorithm. *Computing*, 7:17-24, 1971.
- [19] T.H. O'Beirne. On linear iterative processes and on methods of improving the convergence of certain types of iterated sequences. Technical report, Torpedo Experimental Establishment, Greenock, May 1947.
- [20] A. Salam. Non-commutative extrapolation algorithms. *Numerical Algorithms*, 7:225-251, 1994.
- [21] A. Salam. What is a vector Hankel determinant. *Linear Algebra and its Applications*, 278:147-161, 1998.
- [22] A. Salam. On vector Hankel determinants. *Linear Algebra and its Applications*, 313:127-139, 2000.
- [23] R.J. Schmidt. On the numerical solution of linear simultaneous equations by an iterative method. *Phil. Mag.*, 32(7):369-383, 1941.
- [24] Daniel Shanks. Non-linear transformations of divergent and slowly convergent sequences. *Journal of Mathematics and Physics*, 34:1-42, 1955.
- [25] Avram Sidi. Application of vector extrapolation methods to consistent singular linear systems. *Applied Numerical Mathematics*, pages 487-500, 1989/90.
- [26] Smith, David A. and William F. Ford. Acceleration of linear and logarithmic convergence. *SIAM Journal of Numerical Analysis*, 16(2):223-240, 1979.
- [27] J. Steele and A. Dolovich. Toward the kernel of the vector epsilon algorithm. *International Journal for Numerical Methods in Engineering*, 48(5):721-730, 2000.
- [28] G. Strang. *Linear Algebra and its Applications*, 3<sup>rd</sup> ed. Harcourt Brace Jovanovich College Publishers, Fort Worth, 1988.
- [29] R.C.E. Tan. Computing derivatives of eigensystems by the vector  $\varepsilon$ -algorithm. *IMA Journal of Numerical Analysis*, 7:485-494, 1987.
- [30] R.C.E. Tan and A.L. Andrew. Computing derivatives of eigenvalues and eigenvectors by simultaneous iteration. *IMA Journal of Numerical Analysis*, 9:111-122, 1989.
- [31] P. Wynn. On a device for computing the  $e_m(s_n)$  transformation. *Mathematical Tables and Other Aids to Computation*, 10:91-96, 1956.
- [32] P. Wynn. Acceleration techniques for iterated vector and matrix problems. *Mathematics of Computation*, 16:301-322, 1962.
- [33] P. Wynn. Singular rules for certain nonlinear algorithms. *BIT*, 3:175-195, 1963.

- [34] P. Wynn. Continued fractions whose coefficients obey a non-commutative law of multiplication. *Archives for Rational Mechanics and Analysis*, 12:273-312, 1963.

## APPENDIX A – A DERIVATION OF THE KERNEL OF THE SHANKS TRANSFORM

In this appendix, the conditions under which  $e_k(S_n) = S$ , for some  $k$  and  $n = 0, 1, \dots$  will be derived.

Each term of the  $k^{\text{th}}$  order Shanks transform sequence  $\{e_k(S_n)\}$  is calculated according to the formula

$$e_k(S_n) = \frac{\begin{vmatrix} S_n & S_{n+1} & S_{n+2} & \cdots & S_{n+k} \\ S_{n+1} - S_n & S_{n+2} - S_{n+1} & S_{n+3} - S_{n+2} & \cdots & S_{n+k+1} - S_{n+k} \\ S_{n+2} - S_{n+1} & S_{n+3} - S_{n+2} & S_{n+4} - S_{n+3} & \cdots & S_{n+k+2} - S_{n+k+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ S_{n+k} - S_{n+k-1} & S_{n+k+1} - S_{n+k} & S_{n+k+2} - S_{n+k+1} & \cdots & S_{n+2k} - S_{n+2k-1} \end{vmatrix}}{\begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ S_{n+1} - S_n & S_{n+2} - S_{n+1} & S_{n+3} - S_{n+2} & \cdots & S_{n+k+1} - S_{n+k} \\ S_{n+2} - S_{n+1} & S_{n+3} - S_{n+2} & S_{n+4} - S_{n+3} & \cdots & S_{n+k+2} - S_{n+k+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ S_{n+k} - S_{n+k-1} & S_{n+k+1} - S_{n+k} & S_{n+k+2} - S_{n+k+1} & \cdots & S_{n+2k} - S_{n+2k-1} \end{vmatrix}} \quad (\text{A.1})$$

where the  $S_i$  in the determinants above are terms of the original sequence  $\{S_n\}$ . Now, suppose that  $e_k(S_n) = S$  for some  $k$  and every  $n = 0, 1, \dots$ . Then, if both sides of equation (A.1) are multiplied by the denominator of the right-hand-side and the  $S$  now on the left-hand-side is multiplied through the first row of the determinant on the new left-hand-side, the resulting equation is

$$\begin{aligned}
& \begin{vmatrix} S & S & \cdots & S \\ S_{n+1} - S_n & S_{n+2} - S_{n+1} & \cdots & S_{n+k+1} - S_{n+k} \\ \vdots & \vdots & \ddots & \vdots \\ S_{n+k} - S_{n+k-1} & S_{n+k+1} - S_{n+k} & \cdots & S_{n+2k} - S_{n+2k-1} \end{vmatrix} \\
&= \begin{vmatrix} S_n & S_{n+1} & \cdots & S_{n+k} \\ S_{n+1} - S_n & S_{n+2} - S_{n+1} & \cdots & S_{n+k+1} - S_{n+k} \\ \vdots & \vdots & \ddots & \vdots \\ S_{n+k} - S_{n+k-1} & S_{n+k+1} - S_{n+k} & \cdots & S_{n+2k} - S_{n+2k-1} \end{vmatrix}, n = 0, 1, \dots \quad (\text{A.2})
\end{aligned}$$

If both determinants are expanded by their top rows, they will have the same cofactors and the computation will look like

$$Sc_n + Sc_{n+1} + \dots + Sc_{n+k} = S_n c_n + S_{n+1} c_{n+1} + \dots + S_{n+k} c_{n+k}, \quad n = 0, 1, \dots \quad (\text{A.3})$$

with  $c_i$  being the cofactor corresponding to  $S_i$  in the top row of the right-hand-side of equation (A.2). Subtracting the left-hand-side from the right-hand-side of equation (A.3) and grouping like terms gives

$$0 = (S_n - S)c_n + (S_{n+1} - S)c_{n+1} + \dots + (S_{n+k} - S)c_{n+k}, \quad n = 0, 1, \dots \quad (\text{A.4})$$

This is equivalent to saying

$$\begin{vmatrix} S_n - S & S_{n+1} - S & \cdots & S_{n+k} - S \\ S_{n+1} - S_n & S_{n+2} - S_{n+1} & \cdots & S_{n+k+1} - S_{n+k} \\ \vdots & \vdots & \ddots & \vdots \\ S_{n+k} - S_{n+k-1} & S_{n+k+1} - S_{n+k} & \cdots & S_{n+2k} - S_{n+2k-1} \end{vmatrix} = 0, \quad n = 0, 1, \dots \quad (\text{A.5})$$

If the first row of this determinant is added to the second row, this new second row is added to the third row, and so on to the bottom row, the result

$$\begin{vmatrix} S_n - S & S_{n+1} - S & \cdots & S_{n+k} - S \\ S_{n+1} - S & S_{n+2} - S & \cdots & S_{n+k+1} - S \\ \vdots & \vdots & \ddots & \vdots \\ S_{n+k} - S & S_{n+k+1} - S & \cdots & S_{n+2k} - S \end{vmatrix} = 0, \quad n = 0, 1, \dots \quad (\text{A.6})$$

still holds. From this, it follows directly that the set of homogeneous linear equations

$$\begin{bmatrix} S_n - S & S_{n+1} - S & \cdots & S_{n+k} - S \\ S_{n+1} - S & S_{n+2} - S & \cdots & S_{n+k+1} - S \\ \vdots & \vdots & \ddots & \vdots \\ S_{n+k} - S & S_{n+k+1} - S & \cdots & S_{n+2k} - S \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad n = 0, 1, \dots \quad (\text{A.7})$$

is singular. Thus, if  $e_k(S_n) = S$ , then there must exist  $a_i, i = 0, 1, \dots, k$  such that equation (A.7) is satisfied. Since all of the above operations are reversible, if there exist  $a_i, i = 0, 1, \dots, k$  such that equation (A.7) is satisfied then  $e_k(S_n) = S$ , assuming that the denominator of equation (A.1) is not zero. It is easy to show that if the  $a_i$  sum to zero and equation (A.7) is satisfied by some value of  $S$ , then it will also be satisfied by any other value of  $S$ . It can also be shown that if, using  $a_i$  totalling zero, a sequence is generated that satisfies equation (A.7), then each of the cofactors  $c_i$  in equation (A.3) is zero and both the numerator and the denominator of equation (A.1) are zero leaving  $e_k(S_n)$  undefined. Hence, the  $a_i$  must not add to give zero.

Notice that if equation (A.7) is true, then the dimension of any solution vector to equation (A.7) for which  $k = m$  may be extended with  $p$  zero elements and the matrix in (A.7) extended to include  $p$  additional rows and columns to give the system

$$\begin{bmatrix} S_n - S & \cdots & S_{n+m} - S & S_{n+m+1} - S & \cdots & S_{n+m+p} - S \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ S_{n+m} - S & \cdots & S_{n+2m} - S & S_{n+2m+1} - S & \cdots & S_{n+2m+p} - S \\ S_{n+m+1} - S & \cdots & S_{n+2m+1} - S & S_{n+2m+2} - S & \cdots & S_{n+2m+p+1} - S \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ S_{n+m+p} - S & \cdots & S_{n+2m+p} - S & S_{n+2m+p+1} - S & \cdots & S_{n+2m+2p} - S \end{bmatrix} \begin{bmatrix} a_0 \\ \vdots \\ a_m \\ 0 \\ \vdots \\ 0 \end{bmatrix} = 0 \quad (\text{A.8})$$

The result is that equation (A.7) is satisfied with  $k = m + p$ . Similarly, if the coefficient  $a_k$  in equation (A.7) equals zero, then the last column of the matrix is not involved in the linear combination of columns that sum to zero. Therefore, the matrix of equation (A.7) may be trimmed of its last row and column and the solution vector made to stop at  $a_{k-1}$  and the system of equations will still be singular. Thus, the requirement that  $a_k$  not equal zero ensures that  $k$  is minimal.

Finally, it is noted that equation (A.7) is equivalent to

$$\sum_{i=0}^k a_i (S_{n+i} - S) = 0, \quad n = 0, 1, \dots \quad (\text{A.9})$$

## APPENDIX B – THE JORDAN CANONICAL FORM

The Jordan canonical form of a square matrix and the corresponding decomposition is most instructive. It not only gives the eigenvalues and eigenvectors but also the generalised eigenvectors, which occur when the matrix cannot be diagonalised. It clearly reveals the effect that repeated premultiplication by such a matrix will have on any vector or matrix which may be viewed as an ordered set of column vectors. It also plays an important role in the work reported in this thesis. The author first encountered this theoretical tool in the illuminating introductory text on linear algebra by Gilbert Strang [28]. This exposition elaborates on an introduction to the Jordan canonical form given in an article by Avram Sidi [25].

To begin, every square matrix  $\mathbf{A}$  has a Jordan canonical form  $\mathbf{J}$ , i.e. a diagonal or nearly diagonal square matrix to which it is similar. This may be expressed symbolically by saying that for every  $N \times N$  matrix  $\mathbf{A}$ , there exists some non-singular matrix  $\mathbf{V}$  such that

$$\mathbf{A} = \mathbf{V}\mathbf{J}\mathbf{V}^{-1} \quad (\text{B.1})$$

and  $\mathbf{J}$  is block-diagonal, having the form

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_1 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_2 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{J}_3 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{J}_r \end{bmatrix} \quad (\text{B.2})$$

where each  $\mathbf{J}_i$  (called a Jordan block) is  $N_i \times N_i$ , and has the form

$$\mathbf{J}_i = \begin{bmatrix} \mu_i & 1 & 0 & \cdots & 0 & 0 \\ 0 & \mu_i & 1 & \ddots & 0 & 0 \\ 0 & 0 & \mu_i & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & \mu_i & 1 \\ 0 & 0 & 0 & \cdots & 0 & \mu_i \end{bmatrix}. \quad (\text{B.3})$$

(Some texts show the 1's appearing below the diagonal.) When  $\mathbf{J}$  consists of 2 or more distinct Jordan blocks, then  $\mathbf{J}$  is not unique. However, the differences will lie only in the permutation of the Jordan blocks. The order in which they appear is determined by the matrix  $\mathbf{V}$  used for the decomposition.

It is useful to partition  $\mathbf{V}$  into blocks of column vectors (column blocks) in such a way as to make apparent which column block each Jordan block multiplies. Thus,

$$\mathbf{V} = \begin{bmatrix} | & | & | & \cdots & | \\ | & | & | & \cdots & | \\ \mathbf{V}_1 & \mathbf{V}_2 & \mathbf{V}_3 & \cdots & \mathbf{V}_r \\ | & | & | & \cdots & | \\ | & | & | & \cdots & | \end{bmatrix} \quad (\text{B.4})$$

and each  $\mathbf{V}_i$  is  $N \times N_i$  in size and has column vectors labelled

$$\mathbf{V}_i = \begin{bmatrix} | & | & | & \cdots & | \\ | & | & | & \cdots & | \\ \mathbf{v}_{i,1} & \mathbf{v}_{i,2} & \mathbf{v}_{i,3} & \cdots & \mathbf{v}_{i,N_i} \\ | & | & | & \cdots & | \\ | & | & | & \cdots & | \end{bmatrix}. \quad (\text{B.5})$$

If both sides of equation (B.1) are post-multiplied by  $\mathbf{V}$  to get



$$\mathbf{AV} = \mathbf{VJ} \quad (\text{B.6})$$

and then the result is viewed in blocks, the result is

$$\mathbf{A} \begin{bmatrix} | & | & | & \cdots & | \\ | & | & | & \cdots & | \\ | & | & | & \cdots & | \\ \mathbf{V}_1 & \mathbf{V}_2 & \mathbf{V}_3 & \cdots & \mathbf{V}_r \\ | & | & | & \cdots & | \\ | & | & | & \cdots & | \end{bmatrix} = \begin{bmatrix} | & | & | & \cdots & | \\ | & | & | & \cdots & | \\ | & | & | & \cdots & | \\ \mathbf{V}_1 & \mathbf{V}_2 & \mathbf{V}_3 & \cdots & \mathbf{V}_r \\ | & | & | & \cdots & | \\ | & | & | & \cdots & | \end{bmatrix} \begin{bmatrix} \mathbf{J}_1 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_2 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{J}_3 & \ddots & \mathbf{0} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{J}_r \end{bmatrix} \quad (\text{B.7})$$

or

$$\mathbf{A} \begin{bmatrix} | & | & | & \cdots & | \\ | & | & | & \cdots & | \\ | & | & | & \cdots & | \\ \mathbf{V}_1 & \mathbf{V}_2 & \mathbf{V}_3 & \cdots & \mathbf{V}_r \\ | & | & | & \cdots & | \\ | & | & | & \cdots & | \end{bmatrix} = \begin{bmatrix} | & | & | & \cdots & | \\ | & | & | & \cdots & | \\ | & | & | & \cdots & | \\ \mathbf{V}_1 \mathbf{J}_1 & \mathbf{V}_2 \mathbf{J}_2 & \mathbf{V}_3 \mathbf{J}_3 & \cdots & \mathbf{V}_r \mathbf{J}_r \\ | & | & | & \cdots & | \\ | & | & | & \cdots & | \end{bmatrix} \quad (\text{B.8})$$

which resembles the classic eigenvalue equation except that the expression uses Jordan blocks and column blocks instead of eigenvalues and eigenvectors. If  $\mathbf{A}$  is diagonalisable, then  $N_i = 1, i = 1, 2, \dots, r$ , which means that every Jordan block  $\mathbf{J}_i$  in equation (B.8) is merely an eigenvalue and every column block  $\mathbf{V}_i$  is but an eigenvector.

Since not all square matrices are diagonalisable, it is necessary to understand what goes on when a Jordan block multiplies a column block. This is shown below.

$$\mathbf{V}_i \mathbf{J}_i = \begin{bmatrix} | & | & | & \cdots & | \\ | & | & | & \cdots & | \\ | & | & | & \cdots & | \\ \mathbf{v}_{i,1} & \mathbf{v}_{i,2} & \mathbf{v}_{i,3} & \cdots & \mathbf{v}_{i,N_i} \\ | & | & | & \cdots & | \\ | & | & | & \cdots & | \end{bmatrix} \begin{bmatrix} \mu_i & 1 & 0 & \cdots & 0 \\ 0 & \mu_i & 1 & \cdots & 0 \\ 0 & 0 & \mu_i & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 1 \\ 0 & 0 & 0 & \cdots & \mu_i \end{bmatrix} \quad (\text{B.9})$$

or

$$\mathbf{V}_i \mathbf{J}_i = \begin{bmatrix} | & & | & & | & & | \\ | & & | & & | & & | \\ \mu_i \mathbf{v}_{i,1} & \mu_i \mathbf{v}_{i,2} + \mathbf{v}_{i,1} & \mu_i \mathbf{v}_{i,3} + \mathbf{v}_{i,2} & \cdots & \mu_i \mathbf{v}_{i,N_i} + \mathbf{v}_{i,N_i-1} & & \\ | & & | & & | & & | \\ | & & | & & | & & | \end{bmatrix} \quad (\text{B.10})$$

From equations (B.8) and (B.10), the first vector  $\mathbf{v}_{i,1}$  in each column block  $\mathbf{V}_i$  is an eigenvector which, with the accompanying eigenvalue  $\mu_i$  satisfies the equation

$$\mathbf{A} \mathbf{v}_{i,1} = \mu_i \mathbf{v}_{i,1} \quad (\text{B.11})$$

which is the classic eigenvalue formula. However, equations (B.8) and (B.10) also reveal that the equation governing the subsequent vectors  $\mathbf{v}_{i,j}$ ,  $2 \leq j \leq N_i$  of the column block  $\mathbf{V}_i$  is

$$\mathbf{A} \mathbf{v}_{i,j} = \mu_i \mathbf{v}_{i,j} + \mathbf{v}_{i,j-1}. \quad (\text{B.12})$$

These subsequent vectors  $\mathbf{v}_{i,j}$ ,  $2 \leq j \leq N_i$  are called generalised eigenvectors.

Notice the appearance of  $\mathbf{v}_{i,j-1}$  on the right-hand side of equation (B.12). When  $\mathbf{A}$  premultiplies a generalised eigenvector, the result pulls in a component of the previous vector in the column block. Thus, the vectors of a column block form a sort of chain beginning with a single eigenvector and followed by a sequence of generalised eigenvectors.

To obtain a Jordan form of a matrix  $\mathbf{A}$ , equation (B.11) is equivalent to

$$\mathbf{A}\mathbf{v}_{i,1} = \mu_i \mathbf{I}\mathbf{v}_{i,1} \quad (\text{B.13})$$

or

$$(\mathbf{A} - \mu_i \mathbf{I})\mathbf{v}_{i,1} = \mathbf{0}. \quad (\text{B.14})$$

For nontrivial  $\mathbf{v}_{i,1}$ , therefore, the matrix  $\mathbf{A} - \mu_i \mathbf{I}$  must be singular, which is to say that

$$\det(\mathbf{A} - \mu_i \mathbf{I}) = 0. \quad (\text{B.15})$$

Evaluating the left-hand side of this equation and equating it to zero will yield a polynomial of degree  $N$  which must be solved to give the eigenvalues  $\mu_i$ . Each eigenvalue is substituted into equation (B.14) which is solved for each  $\mathbf{v}_{i,1}$ , making sure to watch for multiple independent eigenvectors for each distinct eigenvalue. The integer  $r$  will be the total number of eigenvectors obtained for all distinct eigenvalues and, thus, the number of Jordan blocks in  $\mathbf{J}$ .

Now, equation (B.12) is equivalent to

$$\mathbf{A}\mathbf{v}_{i,j} = \mu_i \mathbf{I}\mathbf{v}_{i,j} + \mathbf{v}_{i,j-1} \quad (\text{B.16})$$

or

$$(\mathbf{A} - \mu_i \mathbf{I})\mathbf{v}_{i,j} = \mathbf{v}_{i,j-1}. \quad (\text{B.17})$$

Therefore, for  $i = 1$ , placing  $\mathbf{v}_{i,1}$  on the right-hand side, equation (B.17) may be solved recursively for  $\mathbf{v}_{i,j}$ ,  $j = 2, \dots$  until no further solution exists. The highest  $j$  is  $N_i$ . This process is then be repeated for  $i = 2, \dots, r$ .

## APPENDIX C – THE MINIMAL POLYNOMIAL

While it is not difficult to define the minimal polynomial of a square matrix, its significance is not obvious. The minimal polynomial of a square matrix *for a particular vector* is less commonly seen and also less obvious in meaning but turns out to be a very succinct tool in already known results concerning the capabilities of the Vector Epsilon Algorithm. This appendix defines and explains some of the significance of the characteristic polynomial of a matrix (§C.1), the minimal polynomial of a matrix (§C.2), and the minimal polynomial of a matrix for a particular vector (§C.3). An introduction to the minimal polynomial can also be found in Theory and Problems of Linear Algebra by Seymour Lipschutz [16] and other textbooks for a first or second course in linear algebra. Though a few of these books have been consulted, the use of the block diagonal Jordan form in the factored versions of the characteristic and minimal polynomials is original, as is all of §C.3.

### C.1 The Characteristic Polynomial

The characteristic polynomial of an  $N \times N$  matrix  $\mathbf{A}$ , denoted  $p_{ch,\mathbf{A}}$ , is defined by

$$p_{ch,\mathbf{A}}(\mu) = \det(\mathbf{A} - \mu\mathbf{I}) \quad (\text{C.1})$$

where  $\mathbf{I}$  is the identity matrix. When evaluated, the right-hand side of equation (C.1) becomes a polynomial of the form

$$p_{ch,\mathbf{A}}(\mu) = \mu^N + c_{N-1}\mu^{N-1} + c_{N-2}\mu^{N-2} + \dots + c_1\mu + c_0. \quad (\text{C.2})$$

Because the coefficient of the term of highest degree is one, this polynomial is called a monic polynomial. The characteristic equation of  $\mathbf{A}$  is then

$$p_{ch,\mathbf{A}}(\mu) = 0. \quad (\text{C.3})$$

The polynomial  $p_{ch,\mathbf{A}}(\mu)$  will have  $N$  zeros (*i.e.* equation (C.3) will have  $N$  roots) and these zeros will be the eigenvalues of  $\mathbf{A}$ . This is the same as saying that  $p_{ch,\mathbf{A}}(\mu)$  may be factored and expressed as

$$p_{ch,\mathbf{A}}(\mu) = (\mu - \mu'_1)(\mu - \mu'_2)\dots(\mu - \mu'_N) \quad (\text{C.4})$$

where the  $\mu'_i$  are the  $N$  roots of  $p_{ch,\mathbf{A}}(\mu)$  which are not necessarily distinct.

The Cayley-Hamilton theorem states that every square matrix  $\mathbf{A}$  satisfies its characteristic equation. Therefore, changing the notation to accommodate matrices, equation (C.3) becomes

$$p_{ch,\mathbf{A}}(\mathbf{A}) = \mathbf{0} \quad (\text{C.5})$$

or

$$\mathbf{A}^N + c_{N-1}\mathbf{A}^{N-1} + c_{N-2}\mathbf{A}^{N-2} + \dots + c_1\mathbf{A} + c_0\mathbf{I} = \mathbf{0}. \quad (\text{C.6})$$

where  $\mathbf{0}$  is now the  $N \times N$  zero matrix.

It is instructive to consider the Jordan canonical decomposition of  $\mathbf{A}$  (see Appendix B) which is

$$\mathbf{A} = \mathbf{V}\mathbf{J}\mathbf{V}^{-1}. \quad (\text{C.7})$$

where  $\mathbf{V}$  is a nonsingular matrix and  $\mathbf{J}$  is a Jordan canonical form of  $\mathbf{A}$  having the block diagonal form

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_1 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_2 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{J}_3 & \ddots & \mathbf{0} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{J}_r \end{bmatrix}. \quad (\text{C.8})$$

Each of the  $r$  Jordan blocks  $\mathbf{J}_i$  is  $N_i \times N_i$  and has the form

$$\mathbf{J}_i = \begin{bmatrix} \mu_i & 1 & 0 & \cdots & 0 & 0 \\ 0 & \mu_i & 1 & \ddots & 0 & 0 \\ 0 & 0 & \mu_i & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & \mu_i & 1 \\ 0 & 0 & 0 & \cdots & 0 & \mu_i \end{bmatrix}. \quad (\text{C.9})$$

From equation (C.7), for any positive integer  $n$ ,

$$\mathbf{A}^n = \underbrace{\mathbf{V}\mathbf{J}\mathbf{V}^{-1}\mathbf{V}\mathbf{J}\mathbf{V}^{-1}\cdots\mathbf{V}\mathbf{J}\mathbf{V}^{-1}}_{n \text{ times}} \quad (\text{C.10})$$

or

$$\mathbf{A}^n = \mathbf{V}\mathbf{J}^n\mathbf{V}^{-1}. \quad (\text{C.11})$$

If equation (C.11) is substituted into (C.6), the result is

$$\mathbf{V}\mathbf{J}^N\mathbf{V}^{-1} + c_{N-1}\mathbf{V}\mathbf{J}^{N-1}\mathbf{V}^{-1} + c_{N-2}\mathbf{V}\mathbf{J}^{N-2}\mathbf{V}^{-1} + \cdots + c_1\mathbf{V}\mathbf{J}\mathbf{V}^{-1} + c_0\mathbf{I} = \mathbf{0} \quad (\text{C.12})$$

which distributes to give

$$\mathbf{V}(\mathbf{J}^N + c_{N-1}\mathbf{J}^{N-1} + c_{N-2}\mathbf{J}^{N-2} + \dots + c_1\mathbf{J} + c_0\mathbf{I})\mathbf{V}^{-1} = \mathbf{0}. \quad (\text{C.13})$$

Since  $\mathbf{V}$  is nonsingular,

$$\mathbf{J}^N + c_{N-1}\mathbf{J}^{N-1} + c_{N-2}\mathbf{J}^{N-2} + \dots + c_1\mathbf{J} + c_0\mathbf{I} = \mathbf{0} \quad (\text{C.14})$$

which is to say that

$$p_{ch,A}(\mathbf{J}) = \mathbf{0}. \quad (\text{C.15})$$

Since equation (C.2) is equivalent to equation (C.4), equation (C.14) implies that

$$(\mathbf{J} - \mu'_1\mathbf{I})(\mathbf{J} - \mu'_2\mathbf{I})\dots(\mathbf{J} - \mu'_N\mathbf{I}) = \mathbf{0}. \quad (\text{C.16})$$

The truth of this can be seen in what follows.

Since all of the eigenvalues in a given Jordan block are identical, there can be no more than  $r$  distinct roots of  $p_{ch,A}(\mu)$  and possibly fewer if two or more Jordan blocks have the same eigenvalue. Therefore, the symbol  $\mu_i$  will be used for the eigenvalues of  $\mathbf{A}$  indexed from 1 through  $r$  to reflect the Jordan block in which they appear. Therefore, equation (C.16) may be rewritten as

$$(\mathbf{J} - \mu_1\mathbf{I})^{N_1}(\mathbf{J} - \mu_2\mathbf{I})^{N_2}\dots(\mathbf{J} - \mu_r\mathbf{I})^{N_r} = \mathbf{0} \quad (\text{C.17})$$

or

$$\prod_{i=1}^r (\mathbf{J} - \mu_i\mathbf{I})^{N_i} = \mathbf{0} \quad (\text{C.18})$$

which is

$$\begin{bmatrix} \mathbf{J}_{1,1} & & & & & \\ & \mathbf{J}_{2,1} & & & & \\ & & \mathbf{J}_{3,1} & & & \\ & & & \ddots & & \\ & & & & \mathbf{J}_{r,1} & \\ & & & & & \end{bmatrix}^{N_1} \begin{bmatrix} \mathbf{J}_{1,2} & & & & & \\ & \mathbf{J}_{2,2} & & & & \\ & & \mathbf{J}_{3,2} & & & \\ & & & \ddots & & \\ & & & & \mathbf{J}_{r,2} & \\ & & & & & \end{bmatrix}^{N_2} \dots \begin{bmatrix} \mathbf{J}_{1,r} & & & & & \\ & \mathbf{J}_{2,r} & & & & \\ & & \mathbf{J}_{3,r} & & & \\ & & & \ddots & & \\ & & & & \mathbf{J}_{r,r} & \\ & & & & & \end{bmatrix}^{N_r} = \mathbf{0} \quad (\text{C.19})$$

where  $\mathbf{J}_{i,j}$  is the  $N_i \times N_i$  block of the form

$$\mathbf{J}_{i,j} = \begin{bmatrix} \mu_i - \mu_j & 1 & 0 & \dots & 0 & 0 \\ 0 & \mu_i - \mu_j & 1 & \ddots & 0 & 0 \\ 0 & 0 & \mu_i - \mu_j & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & \mu_i - \mu_j & 1 \\ 0 & 0 & 0 & \dots & 0 & \mu_i - \mu_j \end{bmatrix}. \quad (\text{C.20})$$

Thus,

$$\mathbf{J}_{i,i} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \ddots & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & 0 & 1 \\ 0 & 0 & 0 & \dots & 0 & 0 \end{bmatrix}, \quad i = 1, 2, \dots, r. \quad (\text{C.21})$$

Now, consider the result of raising each  $\mathbf{J}_{i,j}$  to some power. It can easily be verified that if some  $N_{\mathbf{B}} \times N_{\mathbf{B}}$  matrix  $\mathbf{B}$  has the form



$$\mathbf{B} = \begin{bmatrix} \beta & 1 & 0 & \cdots & 0 & 0 \\ 0 & \beta & 1 & \ddots & 0 & 0 \\ 0 & 0 & \beta & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & \beta & 1 \\ 0 & 0 & 0 & \cdots & 0 & \beta \end{bmatrix}, \quad (\text{C.22})$$

as will every  $\mathbf{J}_{i,j}$ , then

$$\mathbf{B}^n = \begin{bmatrix} \beta^n & \binom{n}{1}\beta^{n-1} & \cdots & \binom{n}{N_{\mathbf{B}}-2}\beta^{n-N_{\mathbf{B}}+2} & \binom{n}{N_{\mathbf{B}}-1}\beta^{n-N_{\mathbf{B}}+1} \\ & \beta^n & \binom{n}{1}\beta^{n-1} & \cdots & \binom{n}{N_{\mathbf{B}}-2}\beta^{n-N_{\mathbf{B}}+2} \\ & & \beta^n & \ddots & \vdots \\ & \mathbf{0} & & \ddots & \binom{n}{1}\beta^{n-1} \\ & & & & \beta^n \end{bmatrix} \quad (\text{C.23})$$

where  $\binom{i}{j}$  is the binomial coefficient defined by

$$\binom{i}{j} = \frac{i!}{(i-j)!j!} \quad (\text{C.24})$$

with the convention that  $\binom{i}{j} = 0$  if  $j > i$ . For example, if  $\mathbf{B}$  is  $7 \times 7$ , then

$$\mathbf{B}^4 = \begin{bmatrix} \beta^4 & 4\beta^3 & 6\beta^2 & 4\beta & 1 & 0 & 0 \\ 0 & \beta^4 & 4\beta^3 & 6\beta^2 & 4\beta & 1 & 0 \\ 0 & 0 & \beta^4 & 4\beta^3 & 6\beta^2 & 4\beta & 1 \\ 0 & 0 & 0 & \beta^4 & 4\beta^3 & 6\beta^2 & 4\beta \\ 0 & 0 & 0 & 0 & \beta^4 & 4\beta^3 & 6\beta^2 \\ 0 & 0 & 0 & 0 & 0 & \beta^4 & 4\beta^3 \\ 0 & 0 & 0 & 0 & 0 & 0 & \beta^4 \end{bmatrix}, \quad (\text{C.25})$$



## C.2 The Minimal Polynomial of a Matrix

Now, notice the impact that two Jordan blocks having the same eigenvalue has on equation (C.19). Suppose  $\mu_1 = \mu_2$ . This means that both  $\mathbf{J}_{2,1}$  and  $\mathbf{J}_{1,2}$  (see equation (C.20)) are also nilpotent. If  $N_1 \geq N_2$ , then  $\mathbf{J}_{2,1}^{N_1}$  of the first factor matrix of equation (C.28) is also equal to  $\mathbf{0}$ . In this case, if the second factor matrix of equation (C.28) were removed from the left-hand side of the equation, the product would still equal the zero matrix because the second block of the diagonal would be annihilated by  $\mathbf{J}_{2,1}^{N_1}$ . If  $N_1 \leq N_2$ , then  $\mathbf{J}_{1,2}^{N_2}$  of the second factor matrix of equation (C.28) is also equal to  $\mathbf{0}$  and the first factor matrix of equation (C.28) can be removed from the equation with the product still equalling the zero matrix because the first block of the diagonal is annihilated by  $\mathbf{J}_{1,2}^{N_2}$ . This means that whenever two or more Jordan blocks in a Jordan canonical form of some matrix  $\mathbf{A}$  have the same eigenvalue, there exists a monic polynomial  $p(\mu)$  of lower degree than that of  $p_{ch,\mathbf{A}}(\mu)$  such that  $p(\mathbf{A}) = \mathbf{0}$ .

Briefly, the minimal polynomial of an  $N \times N$  square matrix  $\mathbf{A}$  is defined as the monic polynomial

$$p(\mu) = \mu^m + c_{m-1}\mu^{m-1} + c_{m-2}\mu^{m-2} + \dots + c_1\mu + c_0 \quad (\text{C.29})$$

of lowest degree  $m$  such that

$$p(\mathbf{A}) = \mathbf{A}^m + c_{m-1}\mathbf{A}^{m-1} + c_{m-2}\mathbf{A}^{m-2} + \dots + c_1\mathbf{A} + c_0\mathbf{I} = \mathbf{0}. \quad (\text{C.30})$$

This polynomial will be denoted by  $p_{\min, \mathbf{A}}(\mu)$ . It may be thought of as what is left of the characteristic polynomial after all of the factors  $(\mathbf{J} - \mu_i \mathbf{I})^{N_i}$  having redundant nilpotent blocks (arising from two or more Jordan blocks having the same eigenvalues) have been removed.

The factored form of  $p_{\min, \mathbf{A}}(\mu)$  may be constructed as follows. Let  $\nu_1, \nu_2, \dots, \nu_q$  be the  $q$  *distinct* eigenvalues of  $\mathbf{A}$ , and let  $M_k$  be the size of the largest Jordan block in  $\mathbf{J}$ , a Jordan canonical form of  $\mathbf{A}$ , having  $\nu_k$  as its eigenvalue. The minimal polynomial of  $\mathbf{A}$  will then be

$$p_{\min, \mathbf{A}}(\mu) = (\mu - \nu_1)^{M_1} (\mu - \nu_2)^{M_2} \dots (\mu - \nu_q)^{M_q}, \quad (\text{C.31})$$

and

$$p_{\min, \mathbf{A}}(\mathbf{J}) = (\mathbf{J} - \nu_1 \mathbf{I})^{M_1} (\mathbf{J} - \nu_2 \mathbf{I})^{M_2} \dots (\mathbf{J} - \nu_q \mathbf{I})^{M_q} = \mathbf{0} \quad (\text{C.32})$$

where  $m = \Sigma M_k$  is minimal. More concisely,

$$p_{\min, \mathbf{A}}(\mathbf{J}) = \prod_{k=1}^q (\mathbf{J} - \nu_k \mathbf{I})^{M_k} = \mathbf{0}. \quad (\text{C.33})$$

This time, there is one factor per distinct eigenvalue rather than one factor per Jordan block as in equation (C.18).

### C.3 The Minimal Polynomial of a Matrix for a Vector

The minimal polynomial of  $\mathbf{A}$  for the vector  $\mathbf{w}$  is the polynomial

$$p(\mu) = \mu^m + c_{m-1}\mu^{m-1} + c_{m-2}\mu^{m-2} + \dots + c_1\mu + c_0 \quad (\text{C.34})$$

of lowest degree  $m$  such that

$$p(\mathbf{A})\mathbf{w} = \mathbf{0}. \quad (\text{C.35})$$

The  $\mathbf{0}$  in this expression is the zero vector. The polynomial will be denoted by  $p_{\min, \mathbf{A}, \mathbf{w}}(\mu)$ . It may be of lower degree than that of  $p_{\min, \mathbf{A}}(\mu)$  because orthogonal relationships between the rows of  $p_{\min, \mathbf{A}, \mathbf{w}}(\mathbf{A})$  and the vector  $\mathbf{w}$  perform a part of the annihilation left undone by  $p_{\min, \mathbf{A}}(\mu)$ . The source of this reduction in  $m$  is illustrated below.

Substituting the Jordan decomposition of  $\mathbf{A}$  given in equation (C.7) into equation (C.35) yields

$$p_{\min, \mathbf{A}, \mathbf{w}}(\mathbf{V}\mathbf{J}\mathbf{V}^{-1})\mathbf{w} = \mathbf{0} \quad (\text{C.36})$$

which, through the process depicted in equations (C.10) through (C.14), becomes

$$\mathbf{V}p_{\min, \mathbf{A}, \mathbf{w}}(\mathbf{J})\mathbf{V}^{-1}\mathbf{w} = \mathbf{0}. \quad (\text{C.37})$$

If a vector  $\mathbf{c}$  is defined by

$$\mathbf{c} = \mathbf{V}^{-1}\mathbf{w}, \quad (\text{C.38})$$

then, premultiplying both sides of (C.38) by  $\mathbf{V}$ ,  $\mathbf{c}$  satisfies

$$\mathbf{V}\mathbf{c} = \mathbf{w}. \quad (\text{C.39})$$

Thus,  $\mathbf{c}$  is the co-ordinate vector of  $\mathbf{w}$  in the basis defined by the columns of  $\mathbf{V}$ . The scalar entries of  $\mathbf{c}$  are the coefficients of the linear combination of the columns of  $\mathbf{V}$  that make up  $\mathbf{w}$ .

Substituting equation (C.38) into equation (C.37) gives

$$\mathbf{V}P_{\min, A, \mathbf{w}}(\mathbf{J})\mathbf{c} = \mathbf{0}. \quad (\text{C.40})$$

Since  $\mathbf{V}$  is nonsingular,

$$P_{\min, A, \mathbf{w}}(\mathbf{J})\mathbf{c} = \mathbf{0}. \quad (\text{C.41})$$

Now,  $\mathbf{c}$  can likewise be partitioned in a fashion which is compatible with  $\mathbf{J}$  so that

$$\mathbf{c} = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \vdots \\ \mathbf{c}_r \end{bmatrix} \quad (\text{C.42})$$

and each  $\mathbf{c}_i$  is  $N_i \times 1$  with elements labelled

$$\mathbf{c}_i = \begin{bmatrix} c_{i,1} \\ c_{i,2} \\ \vdots \\ c_{i,N_i} \end{bmatrix}. \quad (\text{C.43})$$

Thus, equation (C.41) becomes

$$p_{\min, A, w} \begin{pmatrix} \mathbf{J}_1 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_2 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{J}_3 & \ddots & \mathbf{0} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{J}_r \end{pmatrix} \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \mathbf{c}_3 \\ \vdots \\ \mathbf{c}_r \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \quad (\text{C.44})$$

or

$$\begin{bmatrix} p_{\min, A, w}(\mathbf{J}_1) & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & p_{\min, A, w}(\mathbf{J}_2) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & p_{\min, A, w}(\mathbf{J}_3) & \ddots & \mathbf{0} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & p_{\min, A, w}(\mathbf{J}_r) \end{bmatrix} \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \mathbf{c}_3 \\ \vdots \\ \mathbf{c}_r \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \quad (\text{C.45})$$

Therefore,  $p_{\min, A, w}(\mu)$  must be such that

$$p_{\min, A, w}(\mathbf{J}_i) \mathbf{c}_i = \mathbf{0}, \quad i = 1, 2, \dots, r. \quad (\text{C.46})$$

Thus, the job of annihilating each of the  $r$  segments of the vector on the right-hand side of equation (C.45) is now shared between the block matrix  $p_{\min, A, w}(\mathbf{J}_i)$  and the vector segment  $\mathbf{c}_i$ . Some part of  $p_{\min, A, w}(\mathbf{J}_i)$  can be left nonzero as long as the nonzero part of  $p_{\min, A, w}(\mathbf{J}_i)$  only multiplies zero coefficients present in  $\mathbf{c}_i$ .

The polynomial  $p_{\min, A, w}$  has a factored form. Therefore, equation (C.46) may be written as

$$p_{\min, A, w}(\mathbf{J}_i) \mathbf{c}_i = (\mathbf{J}_i - \nu_1 \mathbf{I})^{K_1} (\mathbf{J}_i - \nu_2 \mathbf{I})^{K_2} \cdots (\mathbf{J}_i - \nu_q \mathbf{I})^{K_q} \mathbf{c}_i = \mathbf{0} \quad i = 1, 2, \dots, r, \quad (\text{C.47})$$

where  $K_k \leq M_k$ ,  $k = 1, \dots, q$ . This is just

$$p_{\min, A, w}(\mathbf{J}_i)\mathbf{c}_i = \prod_{k=1}^q (\mathbf{J}_i - \nu_k \mathbf{I})^{K_k} \mathbf{c}_i = \mathbf{0} \quad i = 1, 2, \dots, r. \quad (\text{C.48})$$

This is almost the same as equation (C.33).

How to determine the value of each  $K_k$  for this equation has not yet been shown. Notice that all of the blocks  $\mathbf{J}_i - \nu_k \mathbf{I}$  for which  $\nu_k \neq \mu_i$  are upper-triangular. (All coefficients below those on the diagonal are equal to zero.) In fact, for each  $i$ , exactly one of the blocks  $\mathbf{J}_i - \nu_k \mathbf{I}$  is *strictly* upper triangular: that for which  $\nu_k = \mu_i$ . (All entries below *and including* those on the diagonal are zero.) Notice what happens when a strictly upper triangular matrix is present as a factor surrounded by upper triangular matrices in a matrix product. If  $\mathbf{T}_{pre}$  and  $\mathbf{T}_{post}$  are upper-triangular  $N_i \times N_i$  matrices and the matrix  $\mathbf{J}_i - \nu_k \mathbf{I}$  is a strictly upper-triangular  $N_i \times N_i$  matrix identical to  $\mathbf{J}_{i,i}$  of equation (C.21), then the product  $\mathbf{T}_{pre}(\mathbf{J}_i - \nu_k \mathbf{I})^K \mathbf{T}_{post}$  where  $K < N_i$  will have the form

$$\mathbf{T}_{pre} \begin{bmatrix} 0 & 0 & \dots & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & 1 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 \end{bmatrix} \mathbf{T}_{post} = \begin{bmatrix} 0 & 0 & \dots & * & * & \dots & * & * \\ 0 & 0 & \dots & 0 & * & \dots & * & * \\ \vdots & \vdots & & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & * & * \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & * \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 \end{bmatrix} \quad (\text{C.49})$$

where an asterisk denotes each nonzero coefficient and the first non-zero coefficient in the top row of the product is at the top of the  $(K + 1)^{\text{th}}$  column. The first  $K$  columns of the product are zero, just as the first  $K$  columns of  $(\mathbf{J}_i - \nu_k \mathbf{I})^K$  are zero. Therefore, the presence of  $(\mathbf{J}_i - \nu_k \mathbf{I})^K$  as a factor surrounded by upper-triangular factor matrices has the



effect of squeezing all the nonzero coefficients in the product into the upper right-hand corner at and above the one non-zero super-diagonal of  $(\mathbf{J}_i - \nu_k \mathbf{I})^{K_k}$ , leaving the first  $K_k$  columns of the product all zeroes.

Now, if the resulting matrix is to premultiply a vector and the product is to be  $\mathbf{0}$ , then all of the nonzero coefficients of that vector must lie in the first  $K_k$  positions. Thus, the critical parameter  $K_k$  for each distinct eigenvalue  $\nu_k$  is no longer the size of the largest Jordan block  $\mathbf{J}_i$  for which  $\mu_i = \nu_k$  but rather the greatest index  $j$  for which  $c_{ij}$  (in  $\mathbf{c}_i$  where  $\mu_i = \nu_k$ ) is non-zero. Recall what was mentioned after equation (C.39), that each  $c_{ij}$  signifies how much of  $\mathbf{v}_{ij}$  is present in  $\mathbf{w}$ . Therefore, for each distinct eigenvalue  $\nu_k$  and all of the column vectors  $\mathbf{v}_{ij}$  associated with  $\mu_i = \nu_k$ ,  $K_k$  is the highest index  $j$  for which some  $\mathbf{v}_{ij}$  is present in  $\mathbf{w}$ . This defines the last of the parameters in equation (C.48).

To summarise, let  $K_k$  equal the highest index  $j$  for which  $c_{ij} \neq 0$  for every  $i$  satisfying  $\mu_i = \nu_k$ . Then, the minimal polynomial of  $\mathbf{A}$  for the vector  $\mathbf{w}$  may be defined as

$$p_{\min, \mathbf{A}, \mathbf{w}}(\mu) = (\mu - \nu_1)^{K_1} (\mu - \nu_2)^{K_2} \dots (\mu - \nu_q)^{K_q} \quad (\text{C.50})$$

with the convention that  $K_k = 0$  if  $c_{ij} = 0$  for every  $j$  when  $\mu_i = \nu_k$ . That is, for each distinct eigenvalue  $\nu_k$  of  $\mathbf{A}$ , of all the column vectors  $\mathbf{v}_{ij}$  in all of the column blocks  $\mathbf{V}_i$  associated with Jordan blocks whose eigenvalue  $\mu_i$  equals  $\nu_k$ , the highest  $j$  for which some of vector  $\mathbf{v}_{ij}$  is present as a component in  $\mathbf{w}$  is the power to which  $(\mu - \nu_k)$  is raised in  $p_{\min, \mathbf{A}, \mathbf{w}}(\mu)$ . If, for some  $\nu_k$  and every  $\mu_i = \nu_k$ , none of the  $\mathbf{v}_{ij}$ ,  $j = 1, 2, \dots, N_i$ , is present in  $\mathbf{w}$ , then  $(\mu - \nu_k)$  will not be a factor of  $p_{\min, \mathbf{A}, \mathbf{w}}(\mu)$ .

For example, suppose that  $\mathbf{A}$  and  $\mathbf{w}$  are such that for some nonsingular matrix  $\mathbf{V}$  satisfying  $\mathbf{A} = \mathbf{V}\mathbf{J}\mathbf{V}^{-1}$ ,

$$\mathbf{J} = \begin{bmatrix} a & 1 & 0 & & & \\ 0 & a & 1 & & & \\ 0 & 0 & a & & & \\ \hline & & & a & 1 & 0 \\ & & & 0 & a & 1 \\ & & & 0 & 0 & a \\ \hline & & & & & & b & 1 \\ & & & & & & 0 & b \end{bmatrix} \quad (\text{C.51})$$

and  $\mathbf{w} = \mathbf{V}\mathbf{c}$  with

$$\mathbf{c} = \begin{bmatrix} c_{1,1} \\ c_{1,2} \\ c_{1,3} \\ \hline c_{2,1} \\ c_{2,2} \\ c_{2,3} \\ \hline c_{3,1} \\ c_{3,2} \end{bmatrix}. \quad (\text{C.52})$$

Table C.1 shows the correlation between five different sets of zero and non-zero values for the  $c_{ij}$  and the exponents for each of the factors of

$$p_{\min, \mathbf{A}, \mathbf{w}}(\mu) = (\mu - a)^{K_1} (\mu - b)^{K_2} \quad (\text{C.53})$$

resulting in a polynomial of degree  $K$ . Again, asterisks denote non-zero entries in  $\mathbf{c}$ .

Table C.1 The effect of non-zero  $c_{ij}$  on  $K_1$ ,  $K_2$ , and  $K$  of equation (C.53).

<i>labels</i>	example1	example2	example3	example4	example5
$c_{1,1}$	*	0	0	*	0
$c_{1,2}$	*	0	*	0	0
$c_{1,3}$	*	*	0	0	0
$c_{2,1}$	*	0	*	0	0
$c_{2,2}$	*	0	0	0	*
$c_{2,3}$	*	0	0	0	0
$c_{3,1}$	*	0	*	0	0
$c_{3,2}$	*	*	0	*	0
$K_1$	3	3	2	1	2
$K_2$	2	2	1	2	0
$K$	5	5	3	3	2

## APPENDIX D – GEKELER’S PROOF OF THEOREM 5.1 [12], P. 68

Theorem 5.1: If  $\{\mathbf{x}_n\}$  is a vector sequence generated by  $\mathbf{x}_n = \mathbf{A}\mathbf{x}_{n-1} + \mathbf{b}$  with  $\mathbf{I} - \mathbf{A}$  non-singular and  $m$  is the degree of the minimal polynomial of  $\mathbf{A}$  for the vector  $\mathbf{x}_0 - \mathbf{x}$  where  $\mathbf{x}$  is the unique solution of the linear system of  $\mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{b}$ , then

$$\boldsymbol{\varepsilon}_{2m}^{(n)} = \mathbf{x}, n = 0, 1, \dots \quad (\text{D.1})$$

**Proof:** Given a sequence of vectors  $\mathbf{x}_n$  such that

$$\mathbf{x}_{n+1} = \mathbf{A}\mathbf{x}_n + \mathbf{b}, n = 0, 1, \dots \quad (\text{D.2})$$

where  $\mathbf{I} - \mathbf{A}$  is non-singular so that

$$\mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{b} \quad (\text{D.3})$$

has a unique solution  $\mathbf{x}$ , if equation (D.3) is subtracted from equation (D.2), the result is

$$(\mathbf{x}_{n+1} - \mathbf{x}) = \mathbf{A}(\mathbf{x}_n - \mathbf{x}) \quad (\text{D.4})$$

which implies that

$$(\mathbf{x}_n - \mathbf{x}) = \mathbf{A}^n(\mathbf{x}_0 - \mathbf{x}), n = 0, 1, \dots \quad (\text{D.5})$$

Let

$$p(t) = \sum_{i=0}^m a_i t^i \quad (\text{D.6})$$

be the minimum polynomial of  $\mathbf{A}$  for  $\mathbf{x}_0 - \mathbf{x}$ . That is

$$p(\mathbf{A})(\mathbf{x}_0 - \mathbf{x}) = \sum_{i=0}^m a_i \mathbf{A}^i (\mathbf{x}_0 - \mathbf{x}) = \mathbf{0}. \quad (\text{D.7})$$

Equation (D.7) can be pre-multiplied by  $\mathbf{A}^n$  to get

$$\mathbf{A}^n \sum_{i=0}^m a_i \mathbf{A}^i (\mathbf{x}_0 - \mathbf{x}) = \sum_{i=0}^m a_i \mathbf{A}^{n+i} (\mathbf{x}_0 - \mathbf{x}) = \mathbf{0} \quad (\text{D.8})$$

but by equation (D.5) this is the same as

$$\sum_{i=0}^m a_i (\mathbf{x}_{n+i} - \mathbf{x}) = \mathbf{0} \quad (\text{D.9})$$

which McLeod and Graves-Morris have shown gives

$$\boldsymbol{\varepsilon}_{2m}^{(n)} = \mathbf{x}, n = 0, 1, \dots \quad (\text{D.10})$$

if

$$\boldsymbol{\varepsilon}_{-1}^{(n)} = \mathbf{0}, \quad \boldsymbol{\varepsilon}_0^{(n)} = \mathbf{x}_n, n = 0, 1, \dots \quad (\text{D.11})$$

which was to be shown. ■