

2016

Covariance expressions for eigenvalue and eigenvector problems

Andrew J. Liounis

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Liounis, Andrew J., "Covariance expressions for eigenvalue and eigenvector problems" (2016). *Graduate Theses, Dissertations, and Problem Reports*. 6095.

<https://researchrepository.wvu.edu/etd/6095>

This Thesis is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Thesis has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

COVARIANCE EXPRESSIONS FOR EIGENVALUE AND EIGENVECTOR PROBLEMS

Andrew J. Liounis

Thesis submitted to the
Benjamin M. Statler College of Engineering and Mineral Resources
at West Virginia University

in partial fulfillment of the requirements for the degree of
Master of Science
in
Aerospace Engineering

John A. Christian, Chair Ph.D.
Jason Gross, Ph.D.
Alfred Lynam, Ph.D.

Department of Mechanical and Aerospace Engineering

Morgantown, West Virginia
2016

Keywords: Eigenvalues, Eigenvectors, Eigenfunctions, Covariance Matrices,
Parameter Estimation

Copyright © 2016

ABSTRACT

Covariance Expressions for Eigenvalue and Eigenvector Problems

Andrew J. Liounis

There are a number of important scientific and engineering problems whose solutions take the form of an eigenvalue–eigenvector problem. Some notable examples include solutions to linear systems of ordinary differential equations, controllability of linear systems, finite element analysis, chemical kinetics, fitting ellipses to noisy data, and optimal estimation of attitude from unit vectors. In many of these problems, having knowledge of the eigenvalue and eigenvector Jacobians is either necessary or is nearly as important as having the solution itself. For instance, Jacobians are necessary to find the uncertainty in a computed eigenvalue or eigenvector estimate. This uncertainty, which is usually represented as a covariance matrix, has been well studied for problems similar to the eigenvalue and eigenvector problem, such as singular value decomposition. There has been substantially less research on the covariance of an optimal estimate originating from an eigenvalue-eigenvector problem.

In this thesis we develop two general expressions for the Jacobians of eigenvalues and eigenvectors with respect to the elements of their parent matrix. The expressions developed make use of only the parent matrix and the eigenvalue and eigenvector pair under consideration. In addition, they are applicable to any general matrix (including complex valued matrices, eigenvalues, and eigenvectors) as long as the eigenvalues are simple. Alongside this, we develop expressions that determine the uncertainty in a vector estimate obtained from an eigenvalue-eigenvector problem given the uncertainty of the terms of the matrix. The Jacobian expressions developed are numerically validated with forward finite, differencing and the covariance expressions are validated using Monte Carlo analysis. Finally, the results from this work are used to determine covariance expressions for a variety of estimation problem examples and are also applied to the design of a dynamical system.

I dedicate this work to my wife, Natalie.

Acknowledgments

First and foremost, I would like to thank my advisor, Dr. John Christian. Thank you for taking me on as an undergraduate with no experience and giving me an opportunity to research and eventually get a job in the field I have always wanted. Thank you for all of your support, guidance, and challenges that helped me to produce this document and grow dramatically in my skills as an engineer. The work you provided for me was both challenging and extremely interesting. Finally, thank you for your trust and confidence throughout my work as an undergraduate and graduate researcher. The timing of your arrival at West Virginia University could not have been better.

I would also like to thank the members of my committee, Dr. Jason Gross and Dr. Alfred Lynam. Thank you for reviewing this thesis and for providing feedback and suggestions throughout.

Thank you to Andrew Rhodes and Jacob Hikes for your help with all of the work in this document. Without you acting as a sounding board, and your help with the intricacies of this problem the results would not be nearly as complete as they are. Also, thank you to the rest of the members of the Applied Space Exploration Laboratory both past and present: Joshua Gerhard, Lylia Benhacine, Chris Gioia, Dr. Ashish Jagat, Eric Kim, Dr. John Woods, Drew Bittner, Jordan Sell, and Miranda Straub. You all helped to create an awesome environment and work-space, as well as provided valuable feedback and assistance with classwork.

Thank you to Shane Robinson for your assistance, feedback, and suggestions regarding much of this manuscript. Thank you also for sharing my passion and excitement for this work—it was certainly contagious. Without your feedback, we never would have created the improved version presented in this thesis.

Thank you to NASA's Johnson Space Center for the funding to do this work.

I would like to extend a thank you to my mentor Kenny Getzandanner at NASA's Goddard Space Flight Center and all of the other members of the Navigation and Mission Design Branch. Thank you for providing advice, feedback, and interesting work over my two summer internships and one semester Pathways Internship.

Thank you to the members of my family. To my parents, David Liounis and Barbara Liounis, thank you for your love, support, encouragement, and upbringing throughout my life and schooling. To my grandmother, Agnes Tobin, thank you for always being there for me, your encouragement, your wisdom, and your advice. To my siblings, Christopher and Alyssa, thanks for always being there to have fun and relax. To my parents-in-law, Joel and Debbie Carpini, and my brother-in-law, Joseph Carpini, thank you for your consistent encouragement and for accepting me into your family.

Thank you to my wife Natalie. Thank you for being there for support when things were stressful and for being there for fun and relaxation when they were not. Thank you for keeping me grounded in my faith. Thank you for your love, patience, and understanding throughout our time together. And most certainly, thank you for taking the time to review this thesis and all of my papers so that it looks like I actually understand how the English language works. Thank you so much, and I love you.

Finally, but certainly not least, thank you to God for granting me the strength, peace, and abilities to make this thesis into what it is.

Contents

Abstract	ii
Acknowledgments	iv
List of Figures	ix
List of Tables	x
1 Introduction	1
1.1 The History of Eigenvalues and Eigenvectors	1
1.2 Motivation	2
1.3 Contribution	4
1.4 Outline	5
2 Background and Literature Review	7
2.1 The Eigenvalue and Eigenvector Problem	7
2.1.1 Solutions to the Eigenvalue and Eigenvector Problem	8
2.1.2 Uniqueness of Eigenvectors	8
2.1.3 Repeated Eigenvalues	8
2.1.4 Properties of the Eigenvalue and Eigenvector Problem	9
2.2 Stability of Eigenvalues and Eigenvectors	9
2.3 Jacobians and the Rules of Matrix Vector Calculus	12
2.4 Analyticity and Differentiability	15
2.4.1 The Cauchy-Riemann Equations	16
2.4.2 The Implicit Function Theorem	16
2.5 Covariance Matrices and Uncertainty in a Problem	17
2.6 Eigenvalue and Eigenvector Derivatives	19

3	Covariance of Eigenvalues and Eigenvectors	22
4	Choosing a Normalization for the Eigenvectors	25
4.1	Determining if the Eigenvectors Are Analytic	26
4.1.1	Constraining the Eigenvectors to the Unit Hyper-Sphere	27
4.1.2	Normalizing Using the Standard Transpose	28
4.1.3	Constraining the Length of the Projection of the Eigenvector onto an Independent Vector	29
4.2	Choosing a Normalization	30
5	Development of Eigenvalue and Eigenvector Jacobians Via Exterior Algebra	33
5.1	Finding the Jacobian matrices for the $n = 2$ Case	33
5.2	Finding the Jacobian matrices for the $n = 3$ Case	38
5.3	Finding the Jacobian matrices for the General Case	39
5.4	Numeric Verification	41
6	Numerically Efficient Eigenvalue and Eigenvector Jacobians	44
6.1	The General Case	44
6.2	Simplified Cases	46
6.2.1	Real Symmetric Parent Matrix	46
6.2.2	Real Diagonal Parent Matrix	48
6.3	Numerical Validation	52
6.4	Comparison of Performance	53
6.5	Extension to the Generalized Eigenvalue Problem	54
6.5.1	A Solution for the Normalization $\mathbf{v}_0^H \mathbf{v} = \alpha$	56
6.5.2	A Solution for the Normalization $\mathbf{v}_0^H \mathbf{B} \mathbf{v} = \alpha$	57
6.5.3	Numeric Verification	58
7	Example Applications	61
7.1	A Simple 2×2 Case	61
7.2	Ellipse Fitting	62
7.3	Attitude Estimation	66
7.4	Horizon-based OPNAV	71
7.4.1	A Review of Horizon-based OPNAV	71
7.4.2	Covariance of the Ellipse Fit	72

7.4.3	Covariance of the CAD OPNAV Measurement	73
7.4.4	Jacobian Matrix of the Position Estimate	75
7.4.5	Numeric Verification of the Developed Covariance Expressions	78
7.5	The Design of Natural Frequencies of a Mass Spring System	79
8	Conclusion	83

List of Figures

2.1	Eigenvalue and eigenvector covariance as a function of eigenvalue/eigenvector condition.	12
4.1	The choice of an eigenvector normalization affects the condition of the system.	31
5.1	The eigenvector normalization constrains the eigenvector to a hyper-plane.	36
5.2	Histograms of percent difference for the exterior algebra Jacobians.	43
6.1	Histograms of percent difference for numerically efficient Jacobians.	53
6.2	Minimum computation time versus matrix size.	54
6.3	Histograms of percent difference for the generalized eigenvalue problem with a normalization of $\mathbf{v}_0^H \mathbf{v} = 1$	59
6.4	Histograms of percent difference for the generalized eigenvalue problem with a normalization of $\mathbf{v}_0^H \mathbf{B} \mathbf{v} = 1$	60
7.1	Ellipse fitting to noisy data.	63
7.2	Ellipse fitting covariance.	66
7.3	2D covariance projections for the Davenport q-method solution to the Wahba problem.	70
7.4	Lit horizons of celestial bodies are used to estimate the relative position of the spacecraft.	71
7.5	Monte Carlo results for horizon based OPNAV.	79
7.6	An example of a mass spring system.	79

List of Tables

2.1	The resulting sizes from differentiating the top by the side.	13
2.2	The complete resulting sizes from differentiating the top by the side.	14
2.3	Probability of realization of a random variable falling within a $k\sigma$ ellipsoid of the truth value.	19

Chapter 1

Introduction

1.1 The History of Eigenvalues and Eigenvectors

Much like the rest of mathematics, the algebraic eigenvalue problem developed in reverse of how it is taught in schools today. It first arose out of the study of discrete mechanical systems in the early 18th century. It was not until much later that the eigenvalue problem began to be considered a topic of linear algebra.¹

In the early part of the 18th century, Bernoulli was the first to discover that vibration problems would have specific modes and values corresponding with these modes that would govern the steady state solutions. Unfortunately, Bernoulli never recognized where these values/modes come from (which we now know to be the eigenvalues and eigenvectors of the coefficients of the linear differential equations governing the system). Euler furthered the work by Bernoulli for the vibrations of string instruments, but Euler still did not fully recognize that he was working with differential equations [1].

In the latter half of the 18th century, the eigenvalue problem began to appear during the solution of systems of linear ordinary differential equations (ODEs) with constant coefficients.² In an attempt to solve the differential equations governing a massless string swinging with three finite masses attached at various points, D'Alembert happened upon the standard eigenvalue problem and realized that the solution of the ordinary differential equation governing the motion of the system involved the exponential of the eigenvalues of the coefficient matrix (in modern linear algebra speak) [4-6].

Lagrange then began to expand upon D'Alembert's work and extend it to the solutions of other

¹This discussion closely follows that of [1], [2], and [3].

²It was not referred to as the eigenvalue problem until much later, and the idea of presenting the coefficients in a matrix form had not been thought of yet. Despite this the form and solution of the eigenvalue problem was certainly present and discovered around this time as we will discuss.

differential equations. Beginning to focus on how the solution to the system of linear ODEs changes as the eigenvalues change while mainly being concerned with the mechanical stability of the system,³ he discovered a better way to find the eigenvalues [7] when recognizing that the coefficient matrix was symmetric. Lagrange then turned his attention to celestial mechanics and applied his technique to determine solutions for the change in various orbital parameters over time [8].

Lagrange’s solution form for systems of linear equations was also extended by Laplace to orbital elements (although they considered different elements) [9,10]. Lagrange and Laplace both continued their work in this area off and on, with Lagrange being one of the first to encounter the generalized eigenvalue problem [8] and Laplace being one of the first to realize that properties of the eigenvalues could be deduced from the symmetry of the coefficient matrix [11].

The next big step in the eigenvalue and eigenvector problem was the study of eigenvectors. In [12], Euler determined that any rigid body would rotate about one of its principle axes. In an attempt to explain this further, Lagrange began to study this problem more and realized that the principle axes were the eigenvectors of the inertia matrix for the rigid body [13].

In 1826 [14], Cauchy was the first to realize that the eigenvalues of symmetric matrices are always real. He also was the first to recognize that the characteristic equation came from a determinant [3].⁴ Hermite later extended this idea to what are now known as Hermetian matrices [15].

Finally, in 1839, the modern terminology began to appear when Cauchy used the phrase “characteristic equation” to describe the polynomial governing the eigenvalues [16]. The terms eigenvalue and eigenvector were first used by Hilbert (“eigen” is the German word for “own”) in 1904 and have become the common terms used to describe the problem (in place of proper values and vectors) [17].

The rest of the history of eigenvalues and eigenvectors largely involves their application to different problems and work on how to calculate them more efficiently and with more accuracy. There has also been a move toward understanding the stability and derivatives of eigenvalues and eigenvectors in recent years that will be discussed in the Literature Review as well as within the primary work of this document.

1.2 Motivation

There are many problems that make use of the eigenvalue or eigenvector of a matrix in their solutions. It arises in techniques for the solution of linear ordinary differential equations [18], determining the

³Certain values for the eigenvalues can lead to infinitely oscillating or unbounded systems which are physically impossible.

⁴Interestingly, the concept of a determinant developed before a formal concept of a matrix and actually lead to matrices.

controllability of a linear system [19], approximation of vibration modes in finite element analysis [20], determining vibration modes in simpler vibration problems [21], determining buckling modes for beams under stress [22], modelling the reactions in chemical mixtures [23], finding the optimal solution to estimation problems [24–27], and many more [28]. Because of this, it is often beneficial to be able to calculate the Jacobians of eigenvalues and eigenvectors with respect to the elements of the matrix from which they were computed.

For example, eigenvalue and eigenvectors are used throughout finite-element analysis (FEA) solutions to vibration problems, where the goal is often to minimize a structure’s sensitivity to various parameters [20]. In these types of problems, it can be a great help to understand how the eigenvalues or eigenvectors change if we change a design parameter without having to resolve the system. In order to do this, we can use a Taylor series expansion along with the eigenvalue and eigenvector Jacobians to predict how our solution will change if we change part of the system.

There are also numerous examples of eigenvalue estimation problems from the field of computer vision, such as two-dimensional ellipse fitting [29], epipolar geometry [30], the eight-point algorithm [31], and pose estimation [32]. Examples from other fields are also numerous, including direction-of-arrival of plane waves [33], estimation of vibration modes [34], quaternion-based spacecraft attitude estimation [35, 36], and horizon-based spacecraft optical navigation [24]. In many estimation problems, an understanding of the uncertainty in the estimate is equally as important as the estimate itself—a quantity commonly described through a covariance matrix. Estimation of the solution’s covariance for the least squares problem is straightforward and well understood. The covariance of eigenvalues and eigenvectors, however, has received substantially less study to date. The present work aims to help address this discrepancy.

A variety of factors can give rise to estimation solutions that take the form of an eigenvalue problem. As a simple motivating example, consider the situation where we seek to find the optimal parameter vector, \mathbf{v} , that minimizes a quadratic cost function where there is also some constraint on the 2-norm (length) of \mathbf{v} . That is, we seek the optimal solution to a problem of the form

$$\text{Min } J(\mathbf{v}) = \mathbf{v}^T \mathbf{A} \mathbf{v} + \lambda (\alpha^2 - \mathbf{v}^T \mathbf{v}) \tag{1.1}$$

where the first term is the quadratic cost function and the second term is the constraint that has been adjoined with a Lagrange multiplier, λ . In this example, the length of \mathbf{v} is constrained to be α .

The optimal solution is found by applying the first differential condition [37]. If we assume that

\mathbf{A} is real valued and symmetric, then we have,

$$2\mathbf{v}^T \mathbf{A}^T - 2\lambda \mathbf{v}^T = 0 \tag{1.2}$$

or, equivalently,

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}, \tag{1.3}$$

which is simply the classic eigenvalue problem. The eigenvalues are given by λ and the eigenvectors are given by \mathbf{v} .

Here, we stress that simple problems like this motivating example are not the only source of eigenvalue problems in estimation applications. Some applications, such as the stability of nonconservative systems [38], produce matrices that are non-Hermitian. Others may also have matrices containing complex valued entries. Thus, we find it necessary to develop methods that are capable of addressing this more general problem.

1.3 Contribution

There is a large pool of existing literature on the subject of eigenvalue and eigenvector Jacobians. Unfortunately, this literature has several limitations that we seek to address in this thesis. In addition, while work has been done on the covariance of estimates making use of the singular value decomposition of a matrix, to the author’s knowledge, no work has been done on the covariance of an estimate resulting from an eigenvalue–eigenvector problem. To address these limitations, we offer contributions in four important areas.

First, we provide an in-depth derivation and discussion of the covariance for eigenvalues and eigenvectors that are produced from a noisy matrix \mathbf{A} with known statistics.

Second, we present two new methods for computing eigenvalue and eigenvector derivatives that are simple and straightforward to implement. In most of the prior work described in Chapter 2.6, the eigenvalue derivatives are obtained by either (1) using both the corresponding left eigenvectors and right eigenvectors, or (2) simultaneously solving for both the eigenvalue and eigenvector derivatives. Though there are instances of other methods that do not require the left eigenvectors or the simultaneous solution for the eigenvector derivatives [39, 40], they are difficult to implement; therefore, we propose new, simpler solutions along these lines for the eigenvalue and eigenvector derivatives.

Third, we present methods capable of calculating the derivatives of the eigenvalues with respect to every element of the matrix \mathbf{A} in a single expression. Our covariance approach considers the

possibility that the covariance matrix describing the uncertainty in the elements of \mathbf{A} may be fully correlated. This requires that the eigenvalue and eigenvector derivatives be computed with respect to every element in \mathbf{A} . In addition, having the full Jacobian matrices with respect to the elements of the parent matrix makes it easy to use the chain rule to calculate the Jacobian matrices with respect to multiple parameters that contribute to \mathbf{A} . This allows us to design dynamic systems with multiple parameters at once. Generally, the methods described in Chapter 2.6 only calculate the derivative of the eigenvalues and the eigenvectors with respect to a single scalar parameter. In fact, the full Jacobian expressions with respect to each term of a given matrix are only expressed once, in much brevity, and only for the symmetric case [41]. We therefore propose solutions for the full Jacobian matrices which, of course, can easily be simplified to find the derivatives with respect to a vector of parameters or a single scalar parameter through the use of the chain rule.

Fourth, and finally, we present methods capable of handling both real and complex valued eigenvectors. The majority of the work described in Chapter 2.6 utilizes the normalization of $\mathbf{v}^T \mathbf{v} = 1$, $\mathbf{u}^T \mathbf{v} = 1$, or some similar implementation in order to make the eigenvectors and eigenvector derivatives unique (where \mathbf{u} is the left eigenvector corresponding to λ). This normalization is excellent when the eigenvectors are real; however, as discussed in [41–44], this normalization is not sufficient to make imaginary eigenvectors unique. This leads to the eigenvectors not being analytic, and thus the derivatives developed are only able to handle real valued eigenvectors. Therefore, the choice of eigenvector normalization must be chosen with some care if the technique is to work for complex valued problems. We discuss our solution to this problem in some detail.

1.4 Outline

The rest of this document proceeds as follows. First, in Chapter 2 we provide background information on many of the techniques and methods that we will use in the rest of this work. Then, in Section 2.6 we provide a brief overview of the literature that already exists on this topic concluding the background portions of the work.

In Chapter 3, we continue with a discussion on the development of analytic expressions for the covariance in eigenvalues and eigenvectors due to uncertainty in the parent matrix. This chapter concludes with us still needing expressions for the Jacobian matrices of eigenvalues and eigenvectors with respect to the elements of their parent matrices. In Chapter 4, we discuss some prerequisites for ensuring that these derivatives exist, mainly considering the choice of a normalization for the eigenvectors.

Finally, in Chapters 5 and 6, having chosen a normalization for our eigenvectors, we turn our attention to the development of two new analytic expressions for the Jacobians of the eigenvalues and eigenvectors. Chapter 5 tackles the problem by first finding the Jacobian for the eigenvalue by considering the characteristic equation as expressed using exterior algebra. This result is then used to calculate the eigenvector Jacobian, completing the required expressions. Chapter 6 takes a different approach by first solving for the eigenvector derivative and then using this to calculate the eigenvalue derivative. This method represents a major improvement from the results in Chapter 5, as will be seen. This Chapter also introduces expressions for the derivatives of the eigenvalues and eigenvectors resulting from the general eigenvalue problem, and it contains a comparison of the numerical efficiency between the two methods.

Chapter 7 concludes the work by showing how each of the techniques may be applied to various estimation problems. Finally, the conclusion recounts the results of the work and discusses areas of future work that may need to be addressed.

Chapter 2

Background and Literature Review

2.1 The Eigenvalue and Eigenvector Problem

The eigenvalue problem is an attempt to find a scalar term λ such that the problem [28].

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v} \tag{2.1}$$

or equivalently

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = 0 \tag{2.2}$$

has a non-trivial solution for \mathbf{v} . Here \mathbf{A} is a $n \times n$ matrix, λ is a scaling term referred to as an eigenvalue, and \mathbf{v} is a vector in the direction of the null space of $\mathbf{A} - \lambda\mathbf{I}$ referred to as an eigenvector. The eigenvector problem is to find the non-trivial vector(s) that satisfy Eqs. (2.1) and (2.2) corresponding to a particular choice for the eigenvalue.

Overall, the eigenvalue/eigenvector problem seeks to find vectors that are only scaled and not rotated when left multiplied by the matrix \mathbf{A} .¹ It is an important problem from linear algebra that has application in numerous types of problems and analysis, such as the solution to systems of linear ordinary differential equations [18], controllability of linear systems [19], finite element analysis, vibration problems, buckling problems [20,22,45], chemical kinetics [23], optimal estimation problems [24-26], and many more [28]. Since it arises so frequently, the solution has been well studied.

¹Obviously, the physical meaning of rotation is somewhat lost when $n > 3$; however, the idea that the vector is changed in no other way than by scaling should still be relatively clear.

2.1.1 Solutions to the Eigenvalue and Eigenvector Problem

The solution to the eigenvalue/eigenvector usually proceeds in two steps (at least when done by hand). First we must solve for the eigenvalues. As mentioned, the eigenvectors are defined to be only the non-trivial solutions to Eq. (2.2). As such, the only way Eq. (2.2) can have a non-trivial solution is if the eigenvalue is chosen such that $\mathbf{A} - \lambda\mathbf{I}$ is rank deficient, or equivalently

$$|\mathbf{A} - \lambda\mathbf{I}| = 0 \tag{2.3}$$

where $|\bullet|$ means to take the determinant of the matrix. This relationship will result in a polynomial of degree n in λ where the n roots for λ are the eigenvalues for the matrix \mathbf{A} . The polynomial is referred to as the characteristic equation for the matrix \mathbf{A} . Solving for the n roots can be done either by hand (if n is small) or by using any number of numeric root finders.

Once the eigenvalues have been determined, the eigenvectors can be found by returning to Eq. (2.1) or (2.2) and solving the resulting linear system of equations. The resulting solution will only be determined up to an unknown scale, because the null space of a matrix applies to any vectors in the direction of that null space (arbitrarily scaled vectors). This can be further seen by replacing \mathbf{v} with $\gamma\mathbf{v}$ in Eq. (2.1) or (2.2) and noting that the result is the same.

2.1.2 Uniqueness of Eigenvectors

It should be clear that in general, eigenvectors are not unique. In fact, for just a single simple eigenvalue, there are an infinite number of eigenvectors that satisfy Eq. (2.1) (all in the direction of \mathbf{v}). Due to this non-uniqueness, a normalization constraining the length of the eigenvectors is usually enforced. For now, we will defer the discussion of eigenvector normalization, as this will be an important choice for some of the work discussed in Chapters 5 and 6; however, keep in mind this requirement for the uniqueness of eigenvectors.

2.1.3 Repeated Eigenvalues

Since the characteristic equation is a n^{th} order polynomial, it is possible for there to be roots with a multiplicity greater than one. For this reason, we give a specific identifier to eigenvalues to indicate how many times they repeat for a given matrix. An eigenvalue with a multiplicity of one is referred to as a simple eigenvalue. An eigenvalue with a multiplicity equal to k where $k > 1$ that also has

k linearly independent eigenvectors² is referred to as a semi-simple eigenvalue.³ The multiplicity of the eigenvalue will play an important role in the rest of this work.

2.1.4 Properties of the Eigenvalue and Eigenvector Problem

There are a few special properties of eigenvalues and eigenvectors that we will make use of throughout the rest of this work. We will represent them here without proof, as proofs can be found in most books on the topic.

The first property that we will make use of extensively is that for a real symmetric parent matrix, the eigenvalues and eigenvectors will always be real. As mentioned in the history section, this was first realized by Cauchy. The next property is the extension of the symmetric case to complex matrices. If a complex matrix is symmetric, then we do not have any information about the eigenvalues and eigenvectors. If a matrix is Hermetian,⁴ then its eigenvalues are real, although the eigenvectors can be either real or complex. Finally, if a matrix is skew symmetric (for real matrices) or skew Hermetian (for complex matrices), then the eigenvalues will be purely imaginary (or 0).⁵

Continuing on with the symmetry, for a symmetric or Hermetian matrix, the geometric multiplicity will always equal the algebraic multiplicity for the eigenvalues. Further, the eigenvectors of such a matrix will always form an orthogonal basis for \mathbb{R}^n or \mathbb{C}^n .

The final property that we will make use of is that for any matrices with distinct eigenvalues, the eigenvectors will always be linearly independent and form a basis for \mathbb{R}^n or \mathbb{C}^n .⁶

2.2 Stability of Eigenvalues and Eigenvectors⁷

In the upcoming chapters we will be developing expressions for the Jacobians and covariance of eigenvalues and eigenvectors with respect to the elements of their parent matrix. In general, whether we are computing the covariance or attempting to predict how the eigenvalues and eigenvectors will change if we change the matrix, we are using the derivatives to linearize a nonlinear problem.⁸ For such a linearization to be meaningful, the eigenvalues and eigenvectors must be well behaved in

²In other words, $\dim[N(\mathbf{A} - \lambda\mathbf{I})] = k$ where $N(\bullet)$ is the null space of \bullet and $\dim[\bullet]$ is the dimension of \bullet .

³The multiplicity of the eigenvalue is sometimes referred to as the algebraic multiplicity, while the dimension of the null space of $\mathbf{A} - \lambda\mathbf{I}$ is often referred to as the geometric multiplicity [46].

⁴A Hermetian matrix is one such that $\mathbf{A} = \mathbf{A}^H$ where a superscript of H indicates the Hermetian (complex conjugate) transpose $\mathbf{A}^H = \text{conj}(\mathbf{A}^T)$.

⁵We will not use this in the following work but have included it here for completeness.

⁶Although this is likely not an orthogonal basis if the matrix is not symmetric or Hermetian.

⁷A portion of this section appeared in the journal paper [47].

⁸As will be discussed later, if we chose certain normalizations for the eigenvectors, then the eigenvector problem actually becomes linear and thus the first derivative and the initial value for the eigenvector actually define the system.

the neighborhood of their nominal values. Thus, an understanding of eigenvalue and eigenvector stability is a prerequisite to the development of the Jacobian matrices.

As a simple motivating example, consider the case when matrix \mathbf{A} is the 3×3 identity matrix. The eigenvalues are obviously $\lambda = 1$ and the corresponding eigenvectors are the standard basis for \mathbb{R}^3 .⁹ Now, a small perturbation to \mathbf{A} can cause the eigenvectors to deviate wildly to some new orientation, breaking our linearization assumption.¹⁰ We therefore need some way to quantify how well behaved our eigenvalues and eigenvectors are.

Fortunately, the stability of the eigenvalue–eigenvector problem has been well studied in the area of perturbation theory [48–52]. In this case, we consider the analysis presented by Van Loan [48]. Van Loan states the stability of the eigenvalues and eigenvectors can be measured by two terms: one for an eigenvalue and one for an eigenvector. The condition of an eigenvalue λ is simply

$$c_\lambda = \frac{1}{\|\mathbf{u}^H \mathbf{v}\|} \quad (2.4)$$

where λ is the eigenvalue being considered, \mathbf{v} is the right eigenvector corresponding to λ , \mathbf{u} is the left eigenvector corresponding to λ , H indicates the Hermetian transpose, and $\|\bullet\|$ indicates the absolute value when applied to a scalar and the 2-norm when applied to a vector or matrix. In fact, it is argued in [48] that the perturbation on λ caused by a perturbation of order ϵ on \mathbf{A} will be approximated by $\|\hat{\lambda} - \lambda\| \approx c_\lambda \epsilon$, where $\hat{\lambda}$ is the perturbed eigenvalue and a perturbation of order ϵ is an additive perturbation matrix with a 2-norm equal to ϵ .

To estimate the condition of the eigenvector, we need to estimate the separation of the corresponding eigenvalue, which provides an estimate of the reciprocal of the distance between the trivial invariant subspaces spanned by the true and the perturbed eigenvectors. That is

$$\text{dist}(\text{span}\{\mathbf{v}\}, \text{span}\{\hat{\mathbf{v}}\}) \approx \frac{1}{\text{sep}(\lambda)} = c_{\mathbf{v}} \quad (2.5)$$

where $\text{dist}(S_1, S_2) = \|P_1 - P_2\|$ with S_i being subspaces and P_i being the orthogonal projection onto S_i , $\hat{\mathbf{v}}$ is the perturbed eigenvector, and $\text{span}\{\mathbf{v}\}$ is referring to the subspace spanned by \mathbf{v} . Calculating the separation of λ is a rather tedious process that involves determining a Householder matrix and calculating the singular value decomposition of a **QR** factorization; therefore, the reader

⁹In actuality, the eigenvectors can be any three linearly independent vectors in \mathbb{C}^3 , but in general we just present them as the standard basis. This is the main problem with semi-simple eigenvalues, as we will discuss in the next sentence.

¹⁰Again, here the eigenvectors are not really deviating wildly to some new orientation because in actuality they did not have an original orientation. Rather they just spanned \mathbb{C}^3 .

is referred to [48], which does an excellent job of presenting the algorithm in a clear manner.

We now attempt to give a basic understanding of the physical meaning of $\text{sep}(\lambda)$. Consider the unitary transformation \mathbf{Q} :

$$\mathbf{Q}^H \mathbf{A} \mathbf{Q} = \begin{bmatrix} \lambda & \mathbf{w}^H \\ \mathbf{0} & \mathbf{B} \end{bmatrix} \quad (2.6)$$

where \mathbf{B} and \mathbf{w} result from the transformation of \mathbf{Q} on \mathbf{A} . Now, it can be shown that we can find some matrix \mathbf{F} that satisfies $\|\mathbf{F}\| = \text{sep}(\lambda)$, such that we have a matrix

$$\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{Q} \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{F} \end{bmatrix} \mathbf{Q}^H \quad (2.7)$$

with λ as a repeated eigenvalue. Therefore, $\text{sep}(\lambda)$ is a measure of the size of the perturbation required to give \mathbf{A} a repeated eigenvalue of λ . Obviously, if λ is already a repeated eigenvalue, then $\text{sep}(\lambda) = 0$, indicating that without more information we cannot estimate how the eigenvectors corresponding to λ will change when \mathbf{A} is perturbed.

Having defined these condition values, we now return to the original example of the 3×3 identity matrix. Recall that for a symmetric matrix, the left and right eigenvectors will be the same; thus, the dot product will be 1. Therefore the eigenvalue condition for this matrix will be $c_\lambda = 1$, and we would expect the perturbation on λ caused by some small perturbation ϵ on \mathbf{A} to be approximately $\|\hat{\lambda} - \lambda\| \approx c_\lambda \epsilon = \epsilon$. This is exactly the case, as a perturbation on a diagonal element of the identity matrix will result precisely in $\hat{\lambda} = \lambda + \epsilon$, and a perturbation on the off diagonal will be less than or equal to this by the Gershgorin Circle Theorem [53, 54]. Now we consider the eigenvector condition number, which we discover is infinite because the separation of λ is exactly 0. This again makes intuitive sense because it shows that a small perturbation to \mathbf{A} can rotate the eigenvectors to dramatically different orientations.

In order to demonstrate the usefulness of the condition numbers for the eigenvalue and eigenvector, we have generated plots of the eigenvalue and eigenvector condition numbers versus the sample standard deviation of the eigenvalues and total covariance ($\sqrt{\text{Tr}[\mathbf{P}_{\mathbf{v}\mathbf{v}}]}$) of the eigenvectors. These plots were generated by considering 1,000 randomly generated matrices, randomly selecting an eigenpair, calculating the condition numbers for that eigenpair, and then running a Monte Carlo analysis of 50,000 runs, adding Gaussian noise to the elements of the matrix. This data is shown in Fig. 2.1. As can be seen, as the condition numbers grow, the covariance and standard deviations grow.

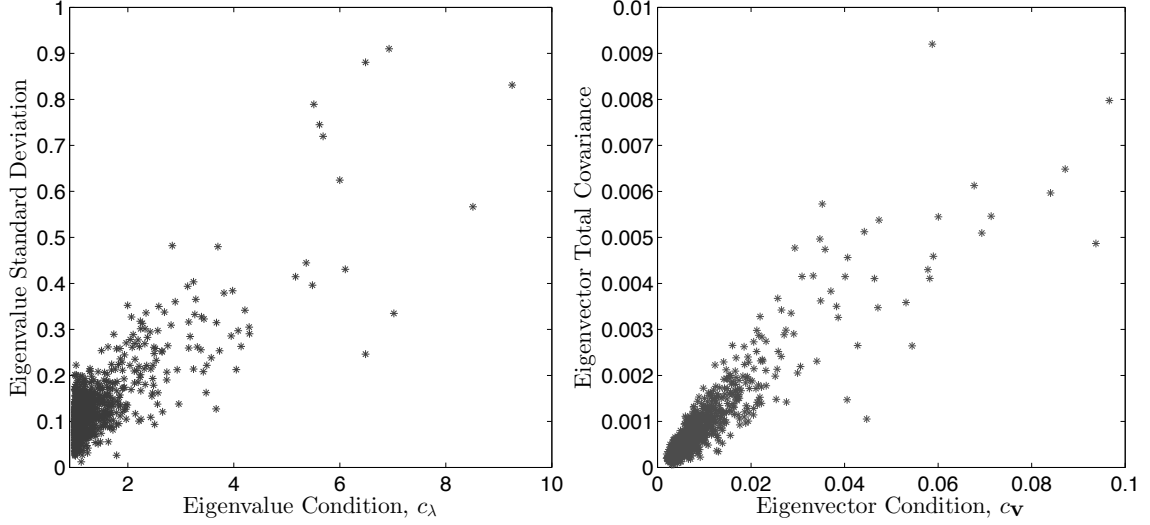


Figure 2.1: Sample standard deviation and covariance values as a function of the eigenvalue and eigenvector condition numbers.

2.3 Jacobians and the Rules of Matrix Vector Calculus

The Jacobian matrix is the extension of a single derivative to functions of multiple variables [55]. It collects all of the partial derivatives of a function with respect to all of the independent variables in that function. Mathematically, this is given by

$$\mathbf{J} = \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \frac{\partial f_1(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f_1(\mathbf{x})}{\partial x_n} \\ \frac{\partial f_2(\mathbf{x})}{\partial x_1} & \frac{\partial f_2(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f_2(\mathbf{x})}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m(\mathbf{x})}{\partial x_1} & \frac{\partial f_m(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f_m(\mathbf{x})}{\partial x_n} \end{bmatrix} \quad (2.8)$$

where $\mathbf{f} = \begin{bmatrix} f_1(\mathbf{x}) & f_2(\mathbf{x}) & \cdots & f_m(\mathbf{x}) \end{bmatrix}^T$ is a $m \times 1$ vector function and $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^T$ is a $n \times 1$ vector of independent variables. Thus, \mathbf{J} is a $m \times n$ matrix.

The Jacobian matrix (and its determinant when $m = n$) plays an important role in many multivariate problems [56]. For the purposes of this work, the most important role the Jacobian plays is that it is used in the formation of the best linear approximation of a function at a certain point through the Taylor series expansion of that function:

$$\mathbf{f}(\mathbf{x}) = \mathbf{x}_0 + \left. \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}_0} (\mathbf{x} - \mathbf{x}_0) = \mathbf{x}_0 + \mathbf{J}(\mathbf{x}_0) (\mathbf{x} - \mathbf{x}_0) \quad (2.9)$$

where \mathbf{x}_0 is the point we are linearizing about and $\left. \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}_0} = \mathbf{J}(\mathbf{x}_0)$ is the Jacobian of the

Table 2.1: The resulting sizes from differentiating the top by the side.

	1×1 scalar	$m \times 1$ vector	$m \times p$ matrix
1×1 scalar	1×1	$m \times 1$	$m \times p$
$n \times 1$ vector	$1 \times n$	$m \times n$?
$n \times l$ matrix	$n \times l$?	?

vector function \mathbf{f} with respect to \mathbf{x} evaluated at \mathbf{x}_0 . This relationship will be key in our development of covariance expressions for the eigenvalues and eigenvectors, as we will need it to be able to linearize the eigenvalues and eigenvectors.

The Jacobian also introduces an important concept from matrix vector calculus that we will make use of repeatedly throughout the rest of this work, namely that the derivative of a $m \times 1$ vector with respect to a $n \times 1$ is a $m \times n$ matrix. Using this rule for differentiation with respect to vectors allows us to create shortcuts for matrix vector calculus in the same ways that we have shortcuts for normal scalar calculus. Further, it allows us to maintain the definition of all the normal operators.

Unfortunately, these rules only work as shown in Table 2.1. In this work, we will need to be able to differentiate vectors and matrices with respect to vectors and matrices (the lower right side of the table); therefore, we need to choose how best to define these differentiations. There are numerous options for this, such as tensors or block matrices; however, we will proceed by “vectorizing” our derivatives and using the Kronecker product [57, 58].¹¹

By vectorizing a matrix, we mean stacking the columns of the matrix into a single column vector. Thus, if we have

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_n \end{bmatrix}$$

where \mathbf{a}_i is the i^{th} column of the matrix \mathbf{A} , then the vectorized version of \mathbf{A} is given by

$$\text{vec}[\mathbf{A}] = \mathbf{A}_{\text{vec}} = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_n \end{bmatrix}. \quad (2.10)$$

Now we can complete our table of matrix-vector derivatives as shown in Table 2.2.

The vectorization of matrices and matrix functions is made easier through the use of the Kro-

¹¹We choose this option because we feel that it leads to the cleanest derivatives, as all of the standard operators still work as in normal linear algebra and we only need to make use of one special operator, the Kronecker product.

Table 2.2: The complete resulting sizes from differentiating the top by the side.

	1×1 scalar	$m \times 1$ vector	$mp \times 1$ vectorized matrix
1×1 scalar	1×1	$m \times 1$	$mp \times 1$
$n \times 1$ vector	$1 \times n$	$m \times n$	$mp \times n$
$nl \times 1$ vectorized matrix	$nl \times 1$	$m \times nl$	$mp \times nl$

necker product and its identities. These are given by

$$\text{vec}[\mathbf{ABC}] = (\mathbf{C}^T \otimes \mathbf{A}) \mathbf{B}_{\text{vec}} = (\mathbf{I}_{n \times n} \otimes \mathbf{AB}) \mathbf{C}_{\text{vec}} = (\mathbf{C}^T \mathbf{B}^T \otimes \mathbf{I}_{k \times k}) \mathbf{A}_{\text{vec}} \quad (2.11)$$

$$\text{vec}[\mathbf{AB}] = (\mathbf{I}_{m \times m} \otimes \mathbf{A}) \mathbf{B}_{\text{vec}} = (\mathbf{B}^T \otimes \mathbf{I}_{k \times k}) \mathbf{A}_{\text{vec}} \quad (2.12)$$

where \mathbf{A} is a $k \times l$ matrix, \mathbf{B} is a $l \times m$ matrix, \mathbf{C} is a $m \times n$ matrix, and \otimes indicates the Kronecker product given by

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots & a_{1l}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \dots & a_{2l}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1}\mathbf{B} & a_{k2}\mathbf{B} & \dots & a_{kl}\mathbf{B} \end{bmatrix}. \quad (2.13)$$

There are a few other properties that will prove useful in simplifying our derivatives:

$$\mathbf{A} (\mathbf{x}^H \otimes \mathbf{I}_{l \times l}) = \mathbf{x}^H \otimes \mathbf{A} \quad (2.14)$$

$$(\mathbf{B} \otimes \mathbf{A}) + (\mathbf{B} \otimes \alpha \mathbf{A}) = \mathbf{B} \otimes (\mathbf{A} + \alpha \mathbf{A}) \quad (2.15)$$

$$\alpha (\mathbf{B} \otimes \mathbf{A}) = (\alpha \mathbf{B} \otimes \mathbf{A}) = (\mathbf{B} \otimes \alpha \mathbf{A}) \quad (2.16)$$

where \mathbf{x} is a $j \times 1$ column vector and α is a scalar.

Using these rules for matrix-vector calculus, the following shortcuts hold:

$$\frac{\partial}{\partial \mathbf{x}} \mathbf{A} \mathbf{x} = \mathbf{A} \quad (2.17)$$

$$\frac{\partial}{\partial \mathbf{x}} \mathbf{y}^H \mathbf{x} = \frac{\partial}{\partial \mathbf{x}} \mathbf{x}^H \mathbf{y} = \mathbf{y}^H \quad (2.18)$$

$$\frac{\partial}{\partial \mathbf{x}} \mathbf{x}^T \mathbf{x} = 2\mathbf{x}^T \quad (2.19)$$

$$\frac{\partial}{\partial \mathbf{A}} \text{Tr} [\mathbf{A}] = \mathbf{I} \quad (2.20)$$

$$\frac{\partial}{\partial \mathbf{A}} |\mathbf{A}| = |\mathbf{A}| \mathbf{A}^{-T} \quad (2.21)$$

$$\frac{\partial}{\partial \mathbf{A}_{\text{vec}}} \mathbf{A} \mathbf{x} = \mathbf{x}^H \otimes \mathbf{I} \quad (2.22)$$

$$\frac{\partial}{\partial \mathbf{A}} \text{Tr} [\mathbf{A}^k] = k (\mathbf{A}^{k-1})^T \quad (2.23)$$

$$\frac{\partial}{\partial \mathbf{A}_{\text{vec}}} \mathbf{y}^H \mathbf{A} \mathbf{x} = \mathbf{x}^T \otimes \mathbf{y}^H \quad (2.24)$$

In addition, the chain rule holds, which will be very important for our upcoming derivatives.

2.4 Analyticity and Differentiability

One of the keys to finding the Jacobians discussed in the rest of this work is to determine when the eigenvalues and eigenvectors are differentiable and analytic within some region about their nominal values. To prove this requires both the implicit function theorem and the Cauchy-Riemann equations, which we will discuss in this section. This also requires knowledge of what analyticity and differentiability imply.

A function is differentiable if at a single point the derivative can be taken. Therefore the function $f(x) = \frac{1}{x}$ is not differentiable at $x = 0$, but is differentiable at every other $x \in \mathbb{R}$. A function is analytic if it is differentiable in some region [28]. Therefore, the previous example is analytic in $x > 0$ and $x < 0$, but it is not analytic in $-1 \leq x \leq 1$.

In general, these definitions are interchangeable for real functions of real variables. If a real function is differentiable at a point, then it will be analytic within some region around that point (even if the region is very small). When considering complex functions of complex variables (as is done in the rest of this work), however, a function may be differentiable at a point while not being analytic in any region around that point. This is because of the Cauchy-Riemann equations, which govern complex differentiation.

2.4.1 The Cauchy-Riemann Equations

The Cauchy–Riemann Equations [28] provide a test for the differentiability and analyticity of complex functions. Given a complex function of a complex variable

$$f(z) = u(x, y) + iv(x, y) \tag{2.25}$$

where

$$z = x + iy, \tag{2.26}$$

$f(z)$ is differentiable at point $z = z_0 = x_0 + iy_0$ if and only if

$$\frac{\partial u}{\partial x}(z_0) = \frac{\partial v}{\partial y}(z_0), \text{ and } \frac{\partial u}{\partial y}(z_0) = -\frac{\partial v}{\partial x}(z_0). \tag{2.27}$$

Further, a complex function is analytic at z_0 if and only if the Cauchy-Riemann equations hold within some region around z_0 . As will be discussed later, the Cauchy-Riemann equations mean that we have to carefully consider the normalization that we use for our eigenvectors when dealing with complex values.

2.4.2 The Implicit Function Theorem

Suppose we have some relation

$$f(x, y) = 0 \tag{2.28}$$

on x and y where x and y are independent variables. Now suppose we would like to express y explicitly in terms of x (making y into a dependent variable). In some cases, this may be easy to find by rearranging Eq. (2.28) and solving for y , but in many cases, it may be extremely difficult or impossible to do this symbolically. In these cases, it may still be beneficial to be able to infer if it is even possible for such a representation to exist. The implicit function theorem can give us insight into this problem.

Given a vector function¹²

$$\mathbf{f}(\mathbf{x}, \mathbf{y}) = \mathbf{0} \tag{2.29}$$

that is analytic around some point $(\mathbf{x}_0, \mathbf{y}_0)$, if

$$|\mathbf{J}| \neq 0 \tag{2.30}$$

¹²We jump straight to the vector case here, as this is what will be used in the rest of this work.

where the Jacobian matrix is given by

$$\mathbf{J} = \left. \frac{\partial \mathbf{f}}{\partial \mathbf{y}} \right|_{\mathbf{x}=\mathbf{x}_0, \mathbf{y}=\mathbf{y}_0}, \quad (2.31)$$

then there exists a unique vector function

$$\mathbf{y}(\mathbf{x}) \quad (2.32)$$

defined and analytic in some neighborhood of $(\mathbf{x}_0, \mathbf{y}_0)$ such that

$$\mathbf{y}(\mathbf{x}_0) = \mathbf{y}_0. \quad (2.33)$$

Essentially we can use this theorem to prove that a derivative of \mathbf{y} with respect to \mathbf{x} exists in a region around $(\mathbf{x}_0, \mathbf{y}_0)$ [59].¹³ As with the Cauchy-Riemann equations, this theorem will be used later to help us choose our normalization condition.

2.5 Covariance Matrices and Uncertainty in a Problem

In real life, almost all data is corrupted by noise, which makes it impossible to actually determine the truth. Because of this, when we describe a measurement or the estimation of a state, we generally couple it with its covariance matrix (or variance value if it is a scalar measurement/estimate). A covariance matrix describes the uncertainty and correlation in a random vector, such as a measurement or state vector corrupted by random noise. The diagonal of the covariance matrix contains the variance for each element of the random vector, where the variance is the square of the standard deviation of the 1D probability density function (PDF) describing the uncertainty for the element.¹⁴ The off diagonal describes the correlation between the elements of the random vector, which is a measure of how dependent the elements are on each other.

The covariance matrix is defined as the second moment about the mean of the estimate/measurement vector. Mathematically, this is given as

$$\mathbf{P}_{\mathbf{xx}} = \mathbb{E} \left[(\mathbf{x} - \bar{\mathbf{x}}) (\mathbf{x} - \bar{\mathbf{x}})^T \right] \quad (2.34)$$

where $\mathbb{E}[\bullet]$ is the expected value operator and $\bar{\mathbf{x}}$ is the mean or expected value of \mathbf{x} . If we expand

¹³Note that if this theorem fails, it does not necessarily imply that such a function does not exist. Also note that this gives no insight into how big of a region this function is valid in.

¹⁴In general, the PDF of real random variables is almost always approximately Gaussian due to the Central Limit Theorem; therefore, the covariance matrix completely defines the PDF for all of \mathbf{x} since a Gaussian distribution is defined with just the standard deviation and mean [60].

this into matrix form for a 3 element vector, then we have

$$\mathbf{P}_{\mathbf{xx}} = \begin{bmatrix} \mathbb{E}[(x_1 - \bar{x}_1)^2] & \mathbb{E}[(x_1 - \bar{x}_1)(x_2 - \bar{x}_2)] & \mathbb{E}[(x_1 - \bar{x}_1)(x_3 - \bar{x}_3)] \\ \mathbb{E}[(x_1 - \bar{x}_1)(x_2 - \bar{x}_2)] & \mathbb{E}[(x_2 - \bar{x}_2)^2] & \mathbb{E}[(x_2 - \bar{x}_2)(x_3 - \bar{x}_3)] \\ \mathbb{E}[(x_1 - \bar{x}_1)(x_3 - \bar{x}_3)] & \mathbb{E}[(x_2 - \bar{x}_2)(x_3 - \bar{x}_3)] & \mathbb{E}[(x_3 - \bar{x}_3)^2] \end{bmatrix} \quad (2.35)$$

$$= \begin{bmatrix} \sigma_{x_1}^2 & \rho_{x_1x_2}\sigma_{x_1}\sigma_{x_2} & \rho_{x_1x_3}\sigma_{x_1}\sigma_{x_3} \\ \rho_{x_1x_2}\sigma_{x_1}\sigma_{x_2} & \sigma_{x_2}^2 & \rho_{x_2x_3}\sigma_{x_2}\sigma_{x_3} \\ \rho_{x_1x_3}\sigma_{x_1}\sigma_{x_3} & \rho_{x_2x_3}\sigma_{x_2}\sigma_{x_3} & \sigma_{x_3}^2 \end{bmatrix} \quad (2.36)$$

where σ_{x_i} is the standard deviation of the i^{th} element of \mathbf{x} and $\rho_{x_ix_j}$ is the correlation coefficient of the i^{th} and j^{th} terms of \mathbf{x} . The covariance matrix is always symmetric positive semi-definite and is usually restricted to be symmetric positive definite for filtering purposes.

Since the covariance matrix is always symmetric, it can always be diagonalized using eigenvalues and eigenvectors. When we do this, the eigenvalues and eigenvectors of the covariance matrix form the ‘‘principle frame’’ of the covariance, where the rotation matrix from the state’s frame to the principal frame is given by the unit eigenvector matrix. The principle frame is useful because it allows us to quickly see the overall uncertainty in the system (note we can also achieve this through the use of the trace since it is equal to the sum of the eigenvalues of a matrix). Further, it provides a frame in which covariance is entirely uncorrelated, and thus the error ellipses are unrotated, making them easier to visualize.

The standard deviation values from the covariance matrix can be used to create theoretical ellipsoids, which place probabilistic bounds on the potential realizations of a random variable. These bounds are very useful for both visualizing the uncertainty in a random variable as well as for making decisions about how to use the random variable. The ellipse bounds are referred to as $k\sigma$ ellipses where k is generally a positive integer. For each k there is a probabilistic expression detailing what percentage of realizations we should expect to see within an ellipsoid with a principal axis of $k\sigma$. This value is both dependent on the value of k as well as the number of degrees of freedom (for instance, if we are considering a three element random variable, then we have three degrees of freedom), but it is not dependent on the value of σ itself. Probability can be determined by considering the Mahalanobis distance, which is χ^2 distributed. Table 2.3 gives a few examples of the percentages contained within various $k\sigma$ ellipsoids for various degrees of freedom (DOF).

Table 2.3: Probability of realization of a random variable falling within a $k\sigma$ ellipsoid of the truth value.

DOF	k			
	1	2	3	4
1	68.27%	95.45%	99.73%	99.99%
2	39.35%	86.47%	98.89%	99.97%
3	19.87%	73.85%	97.07%	99.89%
4	9.02%	59.40%	93.89%	99.70%

2.6 Eigenvalue and Eigenvector Derivatives¹⁵

Despite the frequency with which eigenvalues and eigenvectors appear in various problems, there has been relatively little work focused specifically on quantifying the covariance of these quantities. There are, however, many nice studies on the covariance of closely related problems—such as the singular value decomposition [61].

While few have specifically addressed the issue of eigenvalue and eigenvector covariance, there is a vast body of relevant prior work that warrants a careful review. If the Jacobian of the eigenvector and eigenvalue can be computed, the covariance follows directly (as is discussed in §5). We will, therefore, now explore this topic more thoroughly.

Eigenvalue and eigenvector derivatives (i.e., Jacobians) have been well studied, particularly in the application of structural design. This has led to many different solutions for the derivative of eigenvalues and eigenvectors, although most of the work has been performed on the eigenvector derivatives. We can classify most of the existing solutions into three basic categories: modal expansion techniques, direct techniques, and techniques based on Nelson’s method [62]. Modal techniques express the eigenvector derivatives as a linear combination of all of the eigenvectors of the system, direct techniques rely on solving a linear system of equations for the eigenvalue and eigenvector derivatives, and techniques based on Nelson’s method calculate a homogeneous and particular solution in order to obtain the eigenvector derivatives.

We begin by considering modal techniques. The first known paper to fully detail both the eigenvalue and eigenvector derivatives was written by Fox and Kapoor [63], who considered symmetric matrices with distinct eigenvalues. Fox and Kapoor developed two methods for the eigenvector derivatives: a direct technique and a modal technique. Rogers [64] and Plaut and Huseyin [65] then simultaneously extended the modal technique of [63] to the general asymmetric case for both the eigenvalue and eigenvector derivative. A similar technique was also proposed by Kalaba, Spingarn,

¹⁵A portion of this section appeared in the journal [47].

and Tesfatsion much later [66]. Lim, Junkins, and Wang considered some issues with previous works and proposed a corrected modal technique [67]. Andrew and Tan then extended the modal technique to the case of repeated eigenvalues for a symmetric matrix [68,69]. This was then extended to the general asymmetric case by van der Aa, Morsche, and Mattheij [70].

Next we review the direct solutions. The second direct formulation was proposed in 1972 by Garg [71], who offered a solution for the general asymmetric case that involves simultaneously solving for the eigenvector and eigenvalue expressions by way of a linear system of equations. This formulation can be beneficial because it does not require knowledge of the left eigenvector or any of the other eigenvectors of the system. Rudisill then extended the system of equations method for just the eigenvector derivative to the general asymmetric case [72] and required knowledge of only the left and right eigenvectors corresponding to the eigenvalue being considered. Rudisill and Chu then derived an alternate system to simultaneously solve for both the eigenvalue and eigenvector derivatives, as well as higher order derivatives [73]. Juang, Lim, and Ghaemmaghami developed techniques to find the derivatives of repeated eigenvalues and their eigenvectors using the singular value decomposition [74–76]. Murthy and Haftka altered the techniques from [73] to use a better normalization condition [42]. Finally, Xu and Wu developed an independent technique for eigenvector derivatives for general asymmetric systems with complex eigenvectors and repeated eigenvalues in which they proposed an entirely new normalization [43].

We now turn to the body of work that relies on Nelson’s method. Nelson proposed a solution for symmetric matrices that involves solving for a particular solution for the eigenvector derivatives using just the left and right eigenvector being considered and then calculating the full solution by adding in a scalar multiple of the eigenvector [62]. Cardani and Mantegazza proposed a similar method to Nelson’s for the general asymmetric case that does not require the left eigenvectors [77] and also simultaneously solves for the eigenvalue derivative similar to [71] and [73]. Ojalvo [78], Dailey [79], Mills-Curran [80], Friswell [81], and Wu, Xu, and Li [82] all proposed various extensions of Nelson’s method to handle symmetric systems with repeated eigenvalues. Friswell and Adhikari also attempted to extend Nelson’s method to the general asymmetric case, though their choice of normalization may cause issues in some instances [83].

Finally we turn to four techniques that are difficult to classify into the above groupings. Magnus [41] proposed something similar to the direct techniques for general asymmetric complex matrices, but it uses a pseudo-inverse to circumvent the non-singular matrices that show up in the derivations. Magnus’s technique chooses a normalization that makes imaginary eigenvectors analytic (an important point that we will discuss more later). Meyer and Stewart proposed a new

technique that involves the use of the group inverse [84]. This technique is beneficial because it is easily adapted to any normalization chosen for the eigenvectors. Xie and Dai [85] proposed a technique that allows for the simultaneous determination of eigenpairs and their derivatives using the Davidson method [86], although it is only beneficial if the eigenpairs are not already known, and it only works for symmetric matrices. Finally, de Leeuw proposed a technique using the generalized inverse, which is similar to Magnus's approach [87].

Chapter 3

Covariance of Eigenvalues and Eigenvectors¹

In data fusion problems that take on the form of an eigenvalue problem as in Eq. (1.3), uncertainty typically enters the estimation process through the matrix \mathbf{A} . Thus, suppose that we have some $n \times n$ matrix $\hat{\mathbf{A}}$ that is generated using noisy data, such that the true \mathbf{A} is given by

$$\mathbf{A} = \hat{\mathbf{A}} + \Delta\mathbf{A} \quad (3.1)$$

where $\Delta\mathbf{A}$ is the error in $\hat{\mathbf{A}}$. In practice, both \mathbf{A} and $\Delta\mathbf{A}$ are unknown.

Now suppose that we want to estimate some meaningful quantity that depends on the eigenvalues (λ) or eigenvectors (\mathbf{v}) of the matrix \mathbf{A} . In this case we do not know the matrix \mathbf{A} because it has been obscured by the noisy data; therefore, the best we can do is find the eigenvalues and eigenvectors of $\hat{\mathbf{A}}$ (denoted as $\hat{\lambda}$ and $\hat{\mathbf{v}}$, respectively). Of equal importance is an understanding of the uncertainty in $\hat{\lambda}$ and $\hat{\mathbf{v}}$ based on some knowledge on the uncertainty in $\hat{\mathbf{A}}$. Specifically, we now seek to develop analytic expressions for the covariance of $\hat{\lambda}$ and $\hat{\mathbf{v}}$ as a function of the covariance of the terms in $\hat{\mathbf{A}}$.

Begin by taking a Taylor series expansion of the eigenvalue and eigenvector to first order:

$$\lambda \approx \hat{\lambda} + \left. \frac{\partial \lambda}{\partial \mathbf{A}_{\text{vec}}} \right|_{\hat{\mathbf{A}}_{\text{vec}}} (\mathbf{A}_{\text{vec}} - \hat{\mathbf{A}}_{\text{vec}}) \quad (3.2)$$

$$\mathbf{v} \approx \hat{\mathbf{v}} + \left. \frac{\partial \mathbf{v}}{\partial \mathbf{A}_{\text{vec}}} \right|_{\hat{\mathbf{A}}_{\text{vec}}} (\mathbf{A}_{\text{vec}} - \hat{\mathbf{A}}_{\text{vec}}) \quad (3.3)$$

¹A portion of this chapter appeared in the journal paper [47].

where \mathbf{A}_{vec} is the vectorized form of matrix \mathbf{A} . Now we can define $\Delta\mathbf{A}_{\text{vec}} = \mathbf{A}_{\text{vec}} - \hat{\mathbf{A}}_{\text{vec}}$, $\Delta\lambda = \lambda - \hat{\lambda}$, and $\Delta\mathbf{v} = \mathbf{v} - \hat{\mathbf{v}}$, resulting in

$$\Delta\lambda \approx \left. \frac{\partial\lambda}{\partial\mathbf{A}_{\text{vec}}} \right|_{\hat{\mathbf{A}}_{\text{vec}}} \Delta\mathbf{A}_{\text{vec}} \quad (3.4)$$

$$\Delta\mathbf{v} \approx \left. \frac{\partial\mathbf{v}}{\partial\mathbf{A}_{\text{vec}}} \right|_{\hat{\mathbf{A}}_{\text{vec}}} \Delta\mathbf{A}_{\text{vec}}. \quad (3.5)$$

To simplify notation, we also define the Jacobian matrices $\mathbf{H}_\lambda = \left. \frac{\partial\lambda}{\partial\mathbf{A}_{\text{vec}}} \right|_{\hat{\mathbf{A}}_{\text{vec}}}$ and $\mathbf{H}_\mathbf{v} = \left. \frac{\partial\mathbf{v}}{\partial\mathbf{A}_{\text{vec}}} \right|_{\hat{\mathbf{A}}_{\text{vec}}}$. From here we can expand the definition of the covariance matrix to obtain the mapping from the uncertainty in $\hat{\mathbf{A}}_{\text{vec}}$ to the uncertainty in $\hat{\lambda}$ and $\hat{\mathbf{v}}$:

$$\begin{aligned} \sigma_\lambda^2 &= \text{E} [\Delta\lambda^2] \\ &= \text{E} [(\mathbf{H}_\lambda \Delta\mathbf{A}_{\text{vec}}) (\mathbf{H}_\lambda \Delta\mathbf{A}_{\text{vec}})^T] \\ &= \mathbf{H}_\lambda \text{E} [\Delta\mathbf{A}_{\text{vec}} \Delta\mathbf{A}_{\text{vec}}^T] \mathbf{H}_\lambda^T \\ \mathbf{P}_{\mathbf{v}\mathbf{v}} &= \text{E} [\Delta\mathbf{v} \Delta\mathbf{v}^T] \\ &= \text{E} [(\mathbf{H}_\mathbf{v} \Delta\mathbf{A}_{\text{vec}}) (\mathbf{H}_\mathbf{v} \Delta\mathbf{A}_{\text{vec}})^T] \\ &= \mathbf{H}_\mathbf{v} \text{E} [\Delta\mathbf{A}_{\text{vec}} \Delta\mathbf{A}_{\text{vec}}^T] \mathbf{H}_\mathbf{v}^T \end{aligned}$$

where $\text{E}[\bullet]$ is the expected value operator. Now, noting that the covariance is defined by $\mathbf{P}_{\mathbf{A}_{\text{vec}}\mathbf{A}_{\text{vec}}} = \text{E} [\Delta\mathbf{A}_{\text{vec}} \Delta\mathbf{A}_{\text{vec}}^T]$, this can be simplified to

$$\sigma_\lambda^2 = \mathbf{H}_\lambda \mathbf{P}_{\mathbf{A}_{\text{vec}}\mathbf{A}_{\text{vec}}} \mathbf{H}_\lambda^T \quad (3.6)$$

$$\mathbf{P}_{\mathbf{v}\mathbf{v}} = \mathbf{H}_\mathbf{v} \mathbf{P}_{\mathbf{A}_{\text{vec}}\mathbf{A}_{\text{vec}}} \mathbf{H}_\mathbf{v}^T. \quad (3.7)$$

Therefore, as long as we have knowledge of the uncertainty of the elements of matrix $\hat{\mathbf{A}}$, we can determine the variance of the eigenvalues and covariance matrix for the eigenvectors.

Before we continue, recall that eigenvectors are not unique unless a normalization condition is enforced. Regardless of what normalization condition is chosen, the length of the eigenvector is always constrained to some specified value. However, because the length is constrained in the same way for each eigenvector, we lose a degree of freedom, and thus the eigenvector covariance matrix will always be rank deficient (it will have rank of $n - 1$). In the 2×2 case, the covariance

matrix will essentially represent a standard deviation value in the direction perpendicular to the eigenvector. In the 3×3 case, the covariance matrix will represent a 2 dimensional ellipse rather than a 3 dimensional ellipsoid. Further, the reduced covariance matrix can be found by rotating into a frame where the eigenvector lies along the primary axis, which will leave the reduced covariance matrix in a $(n - 1) \times (n - 1)$ sub-matrix.

Chapter 4

Choosing a Normalization for the Eigenvectors

As was mentioned before, in the standard eigenvalue problem, eigenvectors are only determined up to an unknown scale. For this reason, it is usually necessary to enforce a normalization on the eigenvectors to make them unique. When we are attempting to find the Jacobians of the eigenvectors, this is definitely the case. In this section, we will assume that the eigenvectors are complex.

One popular normalization for the eigenvectors is to constrain them to be unit length. Mathematically, this is expressed as

$$\|\mathbf{v}\|^2 = \mathbf{v}^H \mathbf{v} = 1. \quad (4.1)$$

This normalization is popular because it constrains the eigenvectors to lie on the unit hyper-sphere. Further, when the parent matrix is Hermetian, this ensures that the matrix of eigenvectors is an orthonormal matrix. That is

$$\mathbf{V}^H \mathbf{V} = \mathbf{I} \quad (4.2)$$

where

$$\mathbf{V} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_n \end{bmatrix}. \quad (4.3)$$

Another normalization used with less frequency is

$$\mathbf{v}^T \mathbf{v} = 1, \quad (4.4)$$

which has no physical meaning unless the eigenvectors are real (which we are assuming they are

not). This normalization also does not lead to an orthonormal matrix of eigenvectors when the eigenvectors are complex.

Another normalization not used commonly is

$$\mathbf{v}_0^H \mathbf{v} = 1 \tag{4.5}$$

where \mathbf{v}_0 is any independent vector that is not orthogonal to \mathbf{v} . Physically, this normalization constrains the length projection of \mathbf{v} onto \mathbf{v}_0 to be equal to 1.

4.1 Determining if the Eigenvectors Are Analytic

As has been mentioned before, the eigenvectors are not guaranteed to be analytic.¹ Therefore, we must be careful to choose a normalization that makes the eigenvectors analytic, or our Jacobian matrices will not represent the true derivative of the eigenvectors (as a true derivative does not exist). Now, the eigenvectors are only implicitly defined according to Eq. (2.2), and we have

$$\mathbf{f}(\mathbf{A}, \mathbf{y}) = 0 \tag{4.6}$$

where $\mathbf{y} = \begin{bmatrix} \mathbf{v}^T & \lambda \end{bmatrix}^T$. In addition, we can express our normalization implicitly as

$$\mathbf{x}\mathbf{v} - 1 = 0 \tag{4.7}$$

where \mathbf{x} is a row vector formed by choosing a column normalization vector and transposition method.² This leaves us with an expression for \mathbf{f} that includes our choice of normalization, given as

$$\mathbf{f}(\mathbf{A}, \mathbf{y}) = \begin{bmatrix} (\mathbf{A} - \lambda\mathbf{I}) \mathbf{v} \\ \mathbf{x}\mathbf{v} - 1 \end{bmatrix} = \mathbf{0}_{n+1 \times 1}. \tag{4.8}$$

We would like to determine if it is possible to express both λ and \mathbf{v} as explicit functions of \mathbf{A} to be sure that we can take the derivatives of \mathbf{v} and λ with respect to the elements of \mathbf{A} . Therefore, we need to use the Implicit Function theorem in order to test for analyticity.

As was mentioned before, the implicit function theorem states that as long as the determinant

¹The eigenvalues are, however, guaranteed to be analytic with respect to \mathbf{A} by the implicit function theorem as long as they are simple. We will not prove this, as this would require determining the partial derivative of the characteristic equation with respect to λ , which is actually a step in the generation of one of our expression for the eigenvalue derivatives. Therefore, a proof of this could easily be developed using the analysis and equations provided in this work.

²For instance, using the normalization in Eq. (4.5), \mathbf{x} would equal \mathbf{v}_0^H .

of the Jacobian of \mathbf{f} with respect to \mathbf{y} can be found and is nonzero, then \mathbf{y} can be expressed as a function of \mathbf{x} . Therefore, consider the Jacobian of \mathbf{f} as defined in Eq. (4.8):

$$\mathbf{J} = \frac{\partial \mathbf{f}}{\partial \mathbf{y}} = \begin{bmatrix} \mathbf{A} - \lambda \mathbf{I} & \mathbf{v} \\ \frac{\partial}{\partial \mathbf{v}}(\mathbf{x}\mathbf{v}) & 0 \end{bmatrix} \quad (4.9)$$

where $\frac{\partial}{\partial \mathbf{v}}(\mathbf{x}\mathbf{v})$ will be a row vector from the rules of matrix vector calculus. Recognizing the block structure of the Jacobian, it can be shown that the determinant is given by

$$\begin{aligned} |\mathbf{J}| &= -1 |\mathbf{A} - \lambda \mathbf{I}| + \left| \mathbf{A} - \lambda \mathbf{I} - \mathbf{v} \left[\frac{\partial}{\partial \mathbf{v}}(\mathbf{x}\mathbf{v}) \right] \right| \\ &= \left| \mathbf{A} - \lambda \mathbf{I} - \mathbf{v} \left[\frac{\partial}{\partial \mathbf{v}}(\mathbf{x}\mathbf{v}) \right] \right|. \end{aligned} \quad (4.10)$$

Therefore, in order to evaluate the determinant of the Jacobian, we need to consider the derivative of the normalization chosen with respect to the eigenvector itself. This is done for each of the normalizations discussed above in the following subsections.

4.1.1 Constraining the Eigenvectors to the Unit Hyper-Sphere

We begin with the most commonly used normalization for the standard eigenvalue–eigenvector problem. Expressing this normalization in terms of the elements of \mathbf{v} , we have

$$-1 + \sum_{k=1}^n \bar{v}_k v_k = -1 + \sum_{k=1}^n (x_k - iy_k)(x_k + iy_k) = -1 + \sum_{k=1}^n x_k^2 + y_k^2 = 0 \quad (4.11)$$

where $v_k = (x_k + iy_k)$ is the k^{th} element of the eigenvector \mathbf{v} and $\bar{\bullet}$ is the complex conjugate of \bullet . Now, let us attempt to determine if we can find the partial derivative of this normalization with respect to the eigenvector in a domain around the eigenvector. In order to do this, we need to use the Cauchy-Riemann Equations to check to see if it is analytic in a region around the eigenvector. Consequently, we need to check the partial derivatives of the real and the imaginary parts of the normalization with respect to the components of the j^{th} element of \mathbf{v} . Since the partial derivative and summation operators are linear, these are given by

$$\begin{aligned} \frac{\partial}{\partial x_j} \text{Re} \left[\sum_{k=1}^n x_k^2 + y_k^2 \right] &= 2x_j & \frac{\partial}{\partial x_j} \text{Im} \left[\sum_{k=1}^n x_k^2 + y_k^2 \right] &= 0 \\ \frac{\partial}{\partial y_j} \text{Re} \left[\sum_{k=1}^n x_k^2 + y_k^2 \right] &= 2y_j & \frac{\partial}{\partial y_j} \text{Im} \left[\sum_{k=1}^n x_k^2 + y_k^2 \right] &= 0 \end{aligned}$$

where $\text{Re}[\bullet]$ takes the real component of \bullet and $\text{Im}[\bullet]$ takes the imaginary component of \bullet . Now, using the Cauchy-Riemann equations, it is easy to see that this function is differentiable only when x_j and y_j are 0 and is analytic nowhere. Thus, this normalization is only differentiable when $\mathbf{v} = \mathbf{0}$, which by the definition of an eigenvector is impossible. Therefore, we cannot use this normalization when we have complex eigenvectors and want to find the eigenvector derivatives since the normalization is not analytic.

4.1.2 Normalizing Using the Standard Transpose

Let us now consider another possible normalization where we normalize using the standard transpose and see if this will allow us to find the derivatives. Expressing this normalization in the terms of the elements of \mathbf{v} , we have

$$-1 + \sum_{k=1}^n v_k v_k = -1 + \sum_{k=1}^n (x_k + iy_k)(x_k + iy_k) = -1 + \sum_{k=1}^n x_k^2 - y_k^2 + i2x_k y_k. \quad (4.12)$$

Again, let us check if this function is analytic with respect to \mathbf{v} using the Cauchy-Riemann equations. The partial derivatives with respect to the components of the j^{th} element of the eigenvector are

$$\begin{aligned} \frac{\partial}{\partial x_j} \text{Re} \left[\sum_{k=1}^n x_k^2 - y_k^2 + i2x_k y_k \right] &= 2x_j & \frac{\partial}{\partial x_j} \text{Im} \left[\sum_{k=1}^n x_k^2 - y_k^2 + i2x_k y_k \right] &= 2y_j \\ \frac{\partial}{\partial y_j} \text{Re} \left[\sum_{k=1}^n x_k^2 - y_k^2 + i2x_k y_k \right] &= -2y_j & \frac{\partial}{\partial y_j} \text{Im} \left[\sum_{k=1}^n x_k^2 - y_k^2 + i2x_k y_k \right] &= 2x_j, \end{aligned}$$

which meet the Cauchy-Riemann criteria. Since this satisfies the Cauchy-Riemann equations, the overall partial derivative can be expressed as

$$\frac{\partial}{\partial \mathbf{v}} (\mathbf{v}^T \mathbf{v} - 1) = 2\mathbf{v}^T \quad (4.13)$$

using the rules of matrix-vector calculus.

Returning to the implicit function theorem, we can check to see if we are guaranteed that \mathbf{v} and λ are analytic functions of \mathbf{A} on some domain in the vicinity of \mathbf{v} and λ . Substituting Eq. (4.13) into Eq. (4.10), we are left with

$$|\mathbf{J}| = |\mathbf{A} - \lambda \mathbf{I} - 2\mathbf{v}\mathbf{v}^T|. \quad (4.14)$$

The question becomes: is this determinant nonzero? The answer is “sometimes”—the determinant is non-zero only if λ is a simple eigenvalue. To see this, consider the term $\mathbf{A} - \lambda \mathbf{I}$, which by definition

has a null space in the direction of \mathbf{v} (and thus \mathbf{v} is not contained in the column space). Now consider the outer product term $2\mathbf{v}\mathbf{v}^T$, which is a matrix with a column space that only spans \mathbf{v} . Therefore, by adding (subtracting) this outer product to $\mathbf{A} - \lambda\mathbf{I}$ we have effectively added \mathbf{v} back into the column space (actually $-\mathbf{v}$) and made the term $\mathbf{A} - \lambda\mathbf{I} - 2\mathbf{v}\mathbf{v}^T$ full rank (which implies a non-zero determinant). Thus, using the Cauchy-Riemann equations and the implicit function theorem, we have proven that if we use the normalization given in Eq. (4.4), then the eigenvalues and eigenvectors can explicitly be expressed as a function of \mathbf{A} and are analytic in some domain around their nominal values.

4.1.3 Constraining the Length of the Projection of the Eigenvector onto an Independent Vector

Finally, let us consider the normalization which constrains the length of the projection of \mathbf{v} onto \mathbf{v}_0 . Again, expressing this normalization in the terms of the elements of \mathbf{v} and \mathbf{v}_0 , we have

$$-1 + \sum_{k=1}^n v_{0k}v_k = -1 \sum_{k=1}^n (a_k - ib_k)(x_k + iy_k) = -1 + \sum_{k=1}^n a_kx_k + b_ky_k + i(a_ky_k - b_kx_k) \quad (4.15)$$

where $v_{0k} = a_k - ib_k$ is the k^{th} element of \mathbf{v}_0 . The partial derivatives with respect to the components of the j^{th} elements of the eigenvector are

$$\begin{aligned} \frac{\partial}{\partial x_j} \text{Re} \left[\sum_{k=1}^n a_kx_k + b_ky_k + i(a_ky_k - b_kx_k) \right] &= a_j & \frac{\partial}{\partial x_j} \text{Im} \left[\sum_{k=1}^n a_kx_k + b_ky_k + i(a_ky_k - b_kx_k) \right] &= -b_j \\ \frac{\partial}{\partial y_j} \text{Re} \left[\sum_{k=1}^n a_kx_k + b_ky_k + i(a_ky_k - b_kx_k) \right] &= b_j & \frac{\partial}{\partial y_j} \text{Im} \left[\sum_{k=1}^n a_kx_k + b_ky_k + i(a_ky_k - b_kx_k) \right] &= a_j \end{aligned}$$

which again satisfy the Cauchy-Riemann equations for any choice of a_j and b_j , indicating that the normalization is analytic in all of \mathbb{C}^n . Since the Cauchy-Riemann equations are satisfied, the vector form of the partial derivative with respect to the eigenvector is given as

$$\frac{\partial}{\partial \mathbf{v}} (\mathbf{v}_0^H \mathbf{v} - 1) = \mathbf{v}_0^H \quad (4.16)$$

using the rules of matrix-vector calculus.

Given this result we can use the implicit function theorem to check to see if the eigenvalues and eigenvectors are guaranteed to be analytic functions of \mathbf{A} on some domain centered at their nominal

values. Substituting Eq. (4.16) into Eq. (4.10) gives

$$|\mathbf{J}| = |\mathbf{A} - \lambda\mathbf{I} - \mathbf{v}\mathbf{v}_0^H| \quad (4.17)$$

which, using the same logic as before, is guaranteed to be full rank and have a non-zero determinant. Therefore, using the implicit function theorem with this normalization, we are guaranteed that the eigenvectors and eigenvalues are analytic functions of \mathbf{A} in some domain around their nominal values.

4.2 Choosing a Normalization

After this analysis, we are left with two potential normalizations that we can use to develop our analytic derivatives. Unfortunately, the most common normalization of constraining the eigenvector to unit length is not one of the options.³ Therefore, we must choose between the two normalizations that are valid in order to proceed.⁴

Ideally, we would like this normalization to approximate the usual normalization of constraining the eigenvector to the unit hyper-sphere. The question then becomes which normalization is better for this. At first glance, it would appear that the normalization given in Eq. (4.4) would be the best approximation. In fact, if the eigenvector is real, then this normalization exactly constrains the eigenvectors to the unit sphere. However, in the case of complex eigenvectors (as is our focus in this work), this is not the case, and to the author’s knowledge this normalization has no “physical” meaning.⁵

Now let us consider the other potential normalization. At first glance, this would appear to be a poor approximation for constraining the length of the eigenvector to the unit hyper-sphere. However, note that this normalization still maintains a physical meaning, as it constrains the length of the projection of \mathbf{v} onto \mathbf{v}_0 to be 1. Also recall that the choice of \mathbf{v}_0 is arbitrary so long as $\mathbf{v}_0^H \mathbf{v} \neq 0$. While this is true mathematically, in finite precision computing, the choice of \mathbf{v}_0 can play a large role in the numerical stability of the eigenvector derivative. In particular, the closer the normalization vector gets to being orthogonal to the eigenvector, the more numerically unstable the

³If the eigenvectors can be guaranteed to be real, then the standard normalization can be used if desired; however, as we will argue, we feel that the normalization we choose provides a better result and more physical meaning than this normalization.

⁴Just as the normalization was crucial to being able to prove that derivatives exist for \mathbf{v} and λ , it will also be crucial in the development of the actual derivatives.

⁵Obviously, in general, complex numbers have lost their physical meaning, but we can still generally extend physical ideas to them (for instance constraining the “length” of a complex vector to a unit sphere). In this case, however, we cannot extend this physical meaning to the complex case. This normalization does constrain the length of the eigenvector, but it is not constraining it to the unit sphere. Further, in general, it actually constrains the length of the eigenvector to a complex number, further diluting the extended physical meaning.

derivative is. To demonstrate this fact, consider Fig. 4.1, which shows the condition number of the term $\mathbf{A} - \lambda\mathbf{I} - \mathbf{v}\mathbf{v}_0^H\mathbf{A}/\alpha$ as a function of the angle between \mathbf{v}_0 and \mathbf{v} for 5,000 randomly generated choices of normalization vector for a 10×10 matrix.⁶ Based on this information, it is clear that

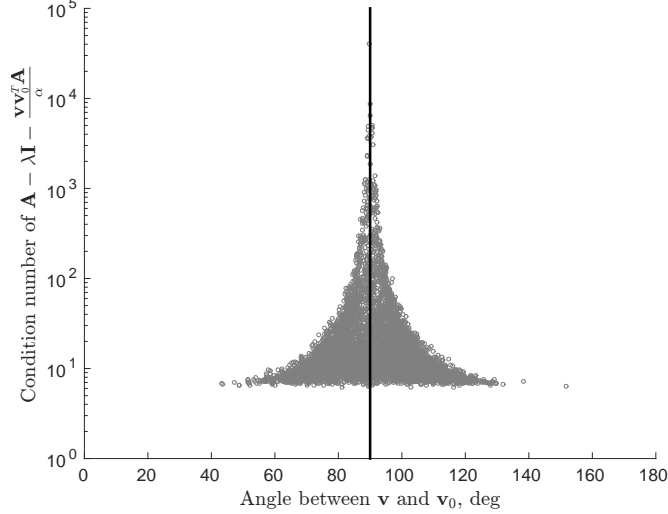


Figure 4.1: Condition number of $\mathbf{A} - \lambda\mathbf{I} - \mathbf{v}\mathbf{v}_0^H\mathbf{A}/\alpha$ as a function of the angle between \mathbf{v}_0 and \mathbf{v} for a 10×10 matrix and 5,000 randomly generated normalization vectors. A higher condition number indicates a more poorly conditioned matrix prone to issues in finite precision computing.

choosing the normalization vector such that $\mathbf{v}_0 = \mathbf{v}$ will ensure the best conditioning for the system if we are using this normalization.

So, since it appears that the normalization given in Eq. (4.5) performs best when $\mathbf{v}_0 = \mathbf{v}$, this is starting to look like it will be a good approximation of the normalization we truly desire. In fact, in general, this behaves exactly like the unit hyper-sphere constraint that we desire if we let $\mathbf{v}_0 = \hat{\mathbf{v}}$ where $\hat{\mathbf{v}}$ is the unit vector in the direction of \mathbf{v} . Further, as long as the eigenvectors are well conditioned, it still serves as a good approximation for that constraint so long as \mathbf{v}_0 was chosen to be \mathbf{v} in cases where this normalization does not behave as the usual constraint to the hyper-sphere. This is because the usual normalization constrains the eigenvectors to fall on the unit hyper-sphere, while this normalization constrains the perturbed eigenvectors to fall on the hyper-plane tangent to the hyper-sphere at \mathbf{v} . Therefore, as long as we remain close to the nominal value for the eigenvector, the proposed normalization approximates the normalization constraining the eigenvectors to the hyper-sphere to first order. Further, when \mathbf{v}_0 is chosen to be $\hat{\mathbf{v}}$, the difference between the usual normalization and Eq. (4.5) rarely needs to be considered in practice, except in very rare situations such as when attempting to numerically verify the analytic expressions as is done in Sect. 6.3.

⁶As will be shown later, the condition of this term will play a vital role in the numerical stability of our derivatives for the eigenvector.

In addition, since this normalization constrains the eigenvector to lie on the hyper-plane tangent to \mathbf{v}_0 , we have actually enforced that the eigenvector is a linear function of the elements of \mathbf{A} . Therefore, the explicit function for the eigenvector in terms of the elements of \mathbf{A} is given exactly by the first order Taylor series expansion of the eigenvector about its nominal value with respect to the change in the elements of \mathbf{A} . This is excellent for both estimating how the eigenvectors are changed when \mathbf{A} is changed and also for our covariance expressions. In addition, this helps to relax some of the requirements on the condition of the eigenvectors since we are no longer linearizing a non-linear problem (although we still need simple eigenvalues).

Based on this information, it appears the normalization proposed by Eq. (4.5) better approximates the unit hyper-sphere constraint for the length of the eigenvectors. This normalization also is beneficial because it maintains some of the properties of the eigenvectors when the matrix is Hermitian (namely that the matrix of eigenvectors is orthonormal). Finally, this normalization actually enforces that the eigenvector is a linear function of the elements of \mathbf{A} , which makes the eigenvector condition less important since we no longer need to linearize. Because of this, we choose to develop our derivatives using this normalization.⁷

⁷As will be seen, the derivatives that we develop are actually valid for all of these normalizations in the real case, and they can easily be manipulated to match the normalization given in Eq. (4.4) in the complex case if so desired. Even in the real case, however, we feel as though the chosen normalization actually describes what is physically happening better, whereas the other potential normalizations only approximate what is happening.

Chapter 5

Development of Eigenvalue and Eigenvector Jacobians Via Exterior Algebra¹

We now consider the derivation of a new technique for computing eigenvalue and eigenvector Jacobians. The technique developed only considers simple eigenvalue/eigenvector pairs—which is not especially surprising given the previous discussion of eigenvalue stability. We begin by discussing the simple 2×2 case, then proceed to the 3×3 case, and conclude with a solution to the general $n \times n$ case.

5.1 Finding the Jacobian matrices for the $n = 2$ Case

To present our technique, we begin with the simplest example—the case where $n = 2$. For this case, it is easy to verify that the characteristic equation of matrix \mathbf{A} can be written as

$$\lambda^2 - \text{Tr}[\mathbf{A}] \lambda + |\mathbf{A}| = 0. \quad (5.1)$$

where $\text{Tr}[\bullet]$ is the trace operator and $|\bullet|$ is the determinant operator. We now seek to determine the Jacobian of λ with respect to \mathbf{A}_{vec} . Begin by distributing the partial derivative and applying

¹Portions of this chapter appeared in the journal paper [47].

the chain rule,

$$2\lambda \frac{\partial \lambda}{\partial \mathbf{A}_{\text{vec}}} - \lambda \frac{\partial \text{Tr}[\mathbf{A}]}{\partial \mathbf{A}_{\text{vec}}} - \text{Tr}[\mathbf{A}] \frac{\partial \lambda}{\partial \mathbf{A}_{\text{vec}}} + \frac{\partial |\mathbf{A}|}{\partial \mathbf{A}_{\text{vec}}} = 0. \quad (5.2)$$

From here, the equation can be rearranged to find

$$(2\lambda - \text{Tr}[\mathbf{A}]) \frac{\partial \lambda}{\partial \mathbf{A}_{\text{vec}}} = \lambda \frac{\partial \text{Tr}[\mathbf{A}]}{\partial \mathbf{A}_{\text{vec}}} - \frac{\partial |\mathbf{A}|}{\partial \mathbf{A}_{\text{vec}}}. \quad (5.3)$$

Now, dividing by the scalar term on the right leads to the equation for the Jacobian:

$$\mathbf{H}_\lambda = \frac{\partial \lambda}{\partial \mathbf{A}_{\text{vec}}} = (2\lambda - \text{Tr}[\mathbf{A}])^{-1} \left(\lambda \frac{\partial \text{Tr}[\mathbf{A}]}{\partial \mathbf{A}_{\text{vec}}} - \frac{\partial |\mathbf{A}|}{\partial \mathbf{A}_{\text{vec}}} \right), \quad (5.4)$$

which is a function of the eigenvalue being considered and the matrix \mathbf{A} . All that is undefined in Eq. (5.4) is the partial derivatives of the trace and the determinant with respect to \mathbf{A}_{vec} . Utilizing the rules of matrix-vector calculus, these can be shown to be

$$\frac{\partial |\mathbf{A}|}{\partial \mathbf{A}_{\text{vec}}} = |\mathbf{A}| \text{vec}(\mathbf{A}^{-T})^T \quad (5.5)$$

$$\frac{\partial \text{Tr}[\mathbf{A}]}{\partial \mathbf{A}_{\text{vec}}} = \mathbf{I}_{\text{vec}}^T, \quad (5.6)$$

completely defining the change in the eigenvalue with respect to a change in the matrix \mathbf{A} .

Returning to Eq. (5.4), we can now see more clearly why we require that eigenvalues be unique. If we express the scalar term as

$$(2\lambda_1 - \text{Tr}[\mathbf{A}]) = \left(2\lambda_1 - \sum_{i=1}^2 \lambda_i \right)$$

where we have switched to using a subscript integer to indicate the eigenvalue being considered and have inserted the definition of the trace being the sum of the eigenvalues, then it is easy to see that if $\lambda_1 = \lambda_2$, this term will be zero. Thus, for repeated eigenvalues, we would find ourselves attempting to divide by zero. This applies to higher order cases as well, as any repeated eigenvalue will cause the scalar term to be zero. It is important to mention, however, that this singularity only exists when the specific eigenvalue being considered is repeated. There are no issues so long as the eigenvalue being considered is unique—this is true even if some of the other eigenvalues of \mathbf{A} are repeated. As a final note, we stress that this will never become an issue if we are only considering stable eigenvalues as discussed in Section 2.2. We now drop the subscript notation again for compactness and continue with our discussion.

It is also necessary to determine the Jacobian of the eigenvector with respect to a change in \mathbf{A} . Begin with the standard eigenvalue problem

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}. \quad (5.7)$$

Once again apply the partial derivative and the chain rule to obtain

$$\frac{\partial}{\partial \mathbf{A}_{\text{vec}}} \{\mathbf{A}\mathbf{v}\} = \mathbf{v} \frac{\partial \lambda}{\partial \mathbf{A}_{\text{vec}}} + \lambda \frac{\partial \mathbf{v}}{\partial \mathbf{A}_{\text{vec}}}. \quad (5.8)$$

Now consider the left-hand side a little more closely. The term $\mathbf{A}\mathbf{v}$ can be expressed as $\mathbf{A}\mathbf{v} = v_1\mathbf{a}_1 + v_2\mathbf{a}_2$, where \mathbf{a}_i is the i^{th} column of matrix \mathbf{A} and v_i is the i^{th} element of vector \mathbf{v} . Applying this expansion allows the usual rules of matrix-vector calculus to be utilized along with the chain rule. This results in

$$\mathbf{A} \frac{\partial \mathbf{v}}{\partial \mathbf{A}_{\text{vec}}} + v_1 \frac{\partial \mathbf{a}_1}{\partial \mathbf{A}_{\text{vec}}} + v_2 \frac{\partial \mathbf{a}_2}{\partial \mathbf{A}_{\text{vec}}} = \mathbf{v} \frac{\partial \lambda}{\partial \mathbf{A}_{\text{vec}}} + \lambda \frac{\partial \mathbf{v}}{\partial \mathbf{A}_{\text{vec}}}. \quad (5.9)$$

Now rearrange to solve for $\partial \mathbf{v} / \partial \mathbf{A}_{\text{vec}}$, resulting in a final expression of

$$(\mathbf{A} - \lambda \mathbf{I}) \frac{\partial \mathbf{v}}{\partial \mathbf{A}_{\text{vec}}} = \mathbf{v} \frac{\partial \lambda}{\partial \mathbf{A}_{\text{vec}}} - v_1 \frac{\partial \mathbf{a}_1}{\partial \mathbf{A}_{\text{vec}}} - v_2 \frac{\partial \mathbf{a}_2}{\partial \mathbf{A}_{\text{vec}}}, \quad (5.10)$$

which is simply a function of the eigenvector and corresponding eigenvalue, the matrix \mathbf{A} , and the eigenvalue Jacobian derived above.

In order to simplify notation later, we can also consider another expansion of the term $\mathbf{A}\mathbf{v}$ using the Kronecker product and vectorization. This expression is given as $\mathbf{A}\mathbf{v} = (\mathbf{v}^H \otimes \mathbf{I}_{2 \times 2})\mathbf{A}_{\text{vec}}$, where \otimes is the Kronecker product. If we use this expression instead, then we have

$$\mathbf{A} \frac{\partial \mathbf{v}}{\partial \mathbf{A}_{\text{vec}}} + \mathbf{v}^H \otimes \mathbf{I}_{2 \times 2} = \mathbf{v} \frac{\partial \lambda}{\partial \mathbf{A}_{\text{vec}}} + \lambda \frac{\partial \mathbf{v}}{\partial \mathbf{A}_{\text{vec}}}, \quad (5.11)$$

which can again be rearrange to solve for $\partial \mathbf{v} / \partial \mathbf{A}_{\text{vec}}$, resulting in a new final expression of

$$(\mathbf{A} - \lambda \mathbf{I}) \frac{\partial \mathbf{v}}{\partial \mathbf{A}_{\text{vec}}} = \mathbf{v} \frac{\partial \lambda}{\partial \mathbf{A}_{\text{vec}}} - \mathbf{v}^H \otimes \mathbf{I}_{2 \times 2}. \quad (5.12)$$

It is easy to verify here that Eqs. (5.10) and (5.12) are equivalent. From this point forward, we will use the vector notation to simplify the notation.

In Eq. (5.12) the term $\mathbf{A} - \lambda \mathbf{I}$ must be inverted to find the partial derivative of the eigenvector.

But this term is not invertible because substituting the eigenvalue into this expression will result in a rank deficient matrix (from the definition of an eigenvalue) with a null space in the direction of \mathbf{v} .

In order to overcome this invertibility issue, we need to make use of our normalization for the eigenvectors. Recall from Chapter 4 that we have chosen to use the normalization constraining the eigenvector to the hyper-plane tangent to the unit hyper-sphere at the normalization vector.² If we differentiate this normalization with respect to \mathbf{A}_{vec} , we get

$$\mathbf{v}_0^H \frac{\partial \mathbf{v}}{\partial \mathbf{A}_{\text{vec}}} = \mathbf{0} \quad (5.13)$$

which shows that the eigenvector derivative must also be constrained to lie along the hyper-plane perpendicular to the normalization vector if the eigenvector is constrained to this hyper-plane. These constraints can easily be demonstrated in \mathbb{R}^2 as shown in Fig. 5.1.

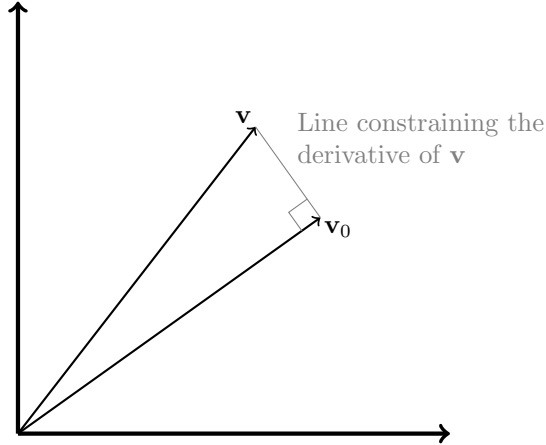


Figure 5.1: By constraining the projection of \mathbf{v} onto \mathbf{v}_0 , the eigenvector derivative is constrained to lie on a line perpendicular to \mathbf{v}_0 .

Using these constraints, we can now define a matrix

$$\mathbf{N} = \sigma \mathbf{v}_0 \mathbf{v}_0^T \quad (5.14)$$

as the outer-product of \mathbf{v}_0 , which we term the Null Space Matrix. Here, σ is a scaling term included for reasons we will discuss shortly. The Null Space Matrix is clearly of rank 1. Additionally, the Null Space Matrix only contains information in the direction of \mathbf{v}_0 and subsequently projects any non-zero vector onto \mathbf{v}_0 . Using the relationship in Eq. (5.13), we also note that $\mathbf{N}(\partial \mathbf{v} / \partial \mathbf{A}_{\text{vec}}) = \mathbf{0}$, and thus we can add this term to the left-hand side of Eq. (5.10) without changing the solution.

²Also recall that, in general, it is best to choose the normalization vector to be the unit vector in the direction of the eigenvector.

Therefore, we have

$$(\mathbf{A} - \lambda\mathbf{I}) \frac{\partial \mathbf{v}}{\partial \mathbf{A}_{\text{vec}}} + \mathbf{N} \frac{\partial \mathbf{v}}{\partial \mathbf{A}_{\text{vec}}} = \mathbf{v} \frac{\partial \lambda}{\partial \mathbf{A}_{\text{vec}}} - \mathbf{v}^H \otimes \mathbf{I}_{2 \times 2}, \quad (5.15)$$

which we can simply rearrange to get

$$(\mathbf{A} - \lambda\mathbf{I} + \mathbf{N}) \frac{\partial \mathbf{v}}{\partial \mathbf{A}_{\text{vec}}} = \mathbf{v} \frac{\partial \lambda}{\partial \mathbf{A}_{\text{vec}}} - \mathbf{v}^H \otimes \mathbf{I}_{2 \times 2}. \quad (5.16)$$

Now, since the Null Space Matrix is rank 1 with information only in the direction of \mathbf{v}_0 , and $\mathbf{A} - \lambda\mathbf{I}$ is of rank $n - 1$ and is missing information in the direction of \mathbf{v}_0 , we have created a full rank matrix, which we can invert to solve for the eigenvector derivative. In most cases, the Null Space Matrix will not be of the same order as $\mathbf{A} - \lambda\mathbf{I}$ due to normalization conditions placed on \mathbf{v}_0 . This can cause numerical condition issues, so we allow σ to scale \mathbf{N} to be approximately the same order as $\mathbf{A} - \lambda\mathbf{I}$ in order to avoid these issues.

Finally, we can write the full eigenvector derivative as

$$\mathbf{H}_{\mathbf{v}} = \frac{\partial \mathbf{v}}{\partial \mathbf{A}_{\text{vec}}} = (\mathbf{A} - \lambda\mathbf{I} + \mathbf{N})^{-1} \left[\mathbf{v} \frac{\partial \lambda}{\partial \mathbf{A}_{\text{vec}}} - \mathbf{v}^H \otimes \mathbf{I}_{2 \times 2} \right]. \quad (5.17)$$

We note here that this technique is similar to that proposed in [41] which makes use of the pseudo-inverse instead of using a Null Space Matrix, but we find this explanation to be more intuitive. In addition, both the Null Space Matrix and the pseudo-inverse are in the worst case $\mathcal{O}(n^3)$ and can be optimized to be less than this in many instances [88]. Finally, the solution in [41] requires both the left and right eigenvectors in the general complex case, whereas our method only requires the right eigenvector.

In the case when the eigenvalues and eigenvectors are real, the Null Space Matrix technique can be used as an approximation for the normalization of $\mathbf{v}^T \mathbf{v} = 1$. In order to use this technique to approximate the derivatives for eigenvectors that are normalized to be of unit length, we only need to be certain that the eigenvectors are well behaved (i.e., they will only move within a small area when the perturbation to \mathbf{A} is small). By enforcing that a perturbation of \mathbf{A} only causes a small perturbation in the eigenvectors, we ensure that the projection of the perturbed eigenvector onto the unperturbed eigenvector is still nearly unitary, which is the requirement for this technique (that is, we need the perturbation of the eigenvector to stay within a range of the true eigenvector so the tangent to the unit circle is a good approximation to the unit circle).

5.2 Finding the Jacobian matrices for the $n = 3$ Case

We now consider the slightly more difficult case when $n = 3$. Once again we begin by expanding the characteristic equation of matrix \mathbf{A} :

$$|\mathbf{A} - \lambda\mathbf{I}| = \lambda^3 - \text{Tr}[\mathbf{A}]\lambda^2 + \frac{1}{2}\left(\text{Tr}[\mathbf{A}]^2 - \text{Tr}[\mathbf{A}^2]\right) - |\mathbf{A}| = 0. \quad (5.18)$$

We are still looking for the Jacobian of λ with respect to \mathbf{A}_{vec} . Applying the chain rule leads to

$$\begin{aligned} \frac{\partial}{\partial \mathbf{A}_{\text{vec}}} \{|\mathbf{A} - \lambda\mathbf{I}|\} &= \frac{\lambda}{2} \left(2\text{Tr}[\mathbf{A}] \frac{\partial \text{Tr}[\mathbf{A}]}{\partial \mathbf{A}_{\text{vec}}} - \frac{\partial \text{Tr}[\mathbf{A}^2]}{\partial \mathbf{A}_{\text{vec}}} \right) \\ &\quad + 3\lambda^2 \frac{\partial \lambda}{\partial \mathbf{A}_{\text{vec}}} - \text{Tr}[\mathbf{A}] 2\lambda \frac{\partial \lambda}{\partial \mathbf{A}_{\text{vec}}} - \lambda^2 \frac{\partial \text{Tr}[\mathbf{A}]}{\partial \mathbf{A}_{\text{vec}}} \\ &\quad + \frac{1}{2} \left(\text{Tr}[\mathbf{A}]^2 - \text{Tr}[\mathbf{A}^2] \right) \frac{\partial \lambda}{\partial \mathbf{A}_{\text{vec}}} - \frac{\partial |\mathbf{A}|}{\partial \mathbf{A}_{\text{vec}}} = 0. \end{aligned} \quad (5.19)$$

From here, we can rearrange to solve for $\partial\lambda/\partial\mathbf{A}_{\text{vec}}$, resulting in

$$a = \left(3\lambda^2 - 2\text{Tr}[\mathbf{A}]\lambda + \frac{1}{2} \left(\text{Tr}[\mathbf{A}]^2 - \text{Tr}[\mathbf{A}^2] \right) \right)^{-1} \quad (5.20)$$

$$\frac{\partial \lambda}{\partial \mathbf{A}_{\text{vec}}} = a \left[\lambda^2 \frac{\partial \text{Tr}[\mathbf{A}]}{\partial \mathbf{A}_{\text{vec}}} + \frac{\partial |\mathbf{A}|}{\partial \mathbf{A}_{\text{vec}}} - \frac{\lambda}{2} \left(2\text{Tr}[\mathbf{A}] \frac{\partial \text{Tr}[\mathbf{A}]}{\partial \mathbf{A}_{\text{vec}}} - \frac{\partial \text{Tr}[\mathbf{A}^2]}{\partial \mathbf{A}_{\text{vec}}} \right) \right], \quad (5.21)$$

which again is simply a function of the eigenvalue being considered and the matrix \mathbf{A} . Remember here that if the eigenvalue is repeated, then a will be singular as mentioned in the 2×2 case. After recalling Eqs. (5.5) and (5.6), only one partial derivative remains undefined, and it can be shown to be

$$\frac{\partial \text{Tr}[\mathbf{A}^2]}{\partial \mathbf{A}_{\text{vec}}} = 2\text{vec}(\mathbf{A}^T)^T, \quad (5.22)$$

which completely defines the Jacobian of λ .

Now we need the Jacobian of the eigenvector. Using the same techniques discussed in the 2×2 case, this can be shown to be given by

$$(\mathbf{A} - \lambda\mathbf{I}) \frac{\partial \mathbf{v}}{\partial \mathbf{A}_{\text{vec}}} = \mathbf{v} \frac{\partial \lambda}{\partial \mathbf{A}_{\text{vec}}} - \mathbf{v}^H \otimes \mathbf{I}_{3 \times 3}. \quad (5.23)$$

As before, we must add the Null Space Matrix to the left-hand side in order to solve this system for $\mathbf{H}_{\mathbf{v}}$:

$$\mathbf{H}_{\mathbf{v}} = (\mathbf{A} - \lambda\mathbf{I} + \mathbf{N})^{-1} \left[\mathbf{v} \frac{\partial \lambda}{\partial \mathbf{A}_{\text{vec}}} - \mathbf{v}^H \otimes \mathbf{I}_{3 \times 3} \right]. \quad (5.24)$$

5.3 Finding the Jacobian matrices for the General Case

The above cases will cover many of the real world applications of this technique. In some instances, however, it may be necessary to determine the uncertainty in an eigenvalue or eigenvector for systems with dimension greater than three. Therefore, in this section, we apply the above techniques to the case of a matrix of size $n \times n$.

Making use of exterior algebra [89], the characteristic equation of matrix \mathbf{A} can be expressed as

$$|\lambda \mathbf{I}_{n \times n} - \mathbf{A}| = \sum_{k=0}^n \lambda^{n-k} (-1)^k \text{Tr} [\Lambda^k \mathbf{A}] = 0 \quad (5.25)$$

where $\Lambda^k \mathbf{A}$ is the k^{th} exterior power of matrix \mathbf{A} , as shown in [90]. Utilizing the following substitution,

$$\text{Tr} [\Lambda^k \mathbf{A}] = \frac{|\mathbf{Q}_k|}{k!} \quad (5.26)$$

where \mathbf{Q}_k is given by

$$\mathbf{Q}_k = \begin{bmatrix} \text{Tr} [\mathbf{A}] & k-1 & 0 & 0 & \dots & 0 \\ \text{Tr} [\mathbf{A}^2] & \text{Tr} [\mathbf{A}] & k-2 & 0 & \dots & 0 \\ \text{Tr} [\mathbf{A}^3] & \text{Tr} [\mathbf{A}^2] & \text{Tr} [\mathbf{A}] & k-3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \text{Tr} [\mathbf{A}^{k-1}] & \text{Tr} [\mathbf{A}^{k-2}] & \text{Tr} [\mathbf{A}^{k-3}] & \dots & \text{Tr} [\mathbf{A}] & 1 \\ \text{Tr} [\mathbf{A}^k] & \text{Tr} [\mathbf{A}^{k-1}] & \text{Tr} [\mathbf{A}^{k-2}] & \text{Tr} [\mathbf{A}^{k-3}] & \dots & \text{Tr} [\mathbf{A}] \end{bmatrix}, \quad (5.27)$$

the characteristic equation can be written as

$$|\lambda \mathbf{I}_{n \times n} - \mathbf{A}| = \sum_{k=0}^n \frac{\lambda^{n-k} (-1)^k |\mathbf{Q}_k|}{k!} = \sum_{k=0}^n c_k \lambda^{n-k} = 0 \quad (5.42)$$

where $c_k = (-1)^k |\mathbf{Q}_k| / (k!)$ are the coefficients of the characteristic equation. We note here that there are other techniques for finding the coefficients c_k , such as solving the system that arises from the Cayley-Hamilton theorem, which states that every square matrix satisfies its own characteristic equation [28, 91–93]. For now we have chosen to utilize this solution due the compact notation, but note that other solutions may be more efficient for calculating the derivatives.

From here the partial derivative with respect to \mathbf{A}_{vec} can be applied to obtain

$$\frac{\partial}{\partial \mathbf{A}_{\text{vec}}} \{|\lambda \mathbf{I}_{n \times n} - \mathbf{A}|\} = \frac{\partial}{\partial \mathbf{A}_{\text{vec}}} \left\{ \sum_{k=0}^n c_k \lambda^{n-k} \right\} = 0. \quad (5.43)$$

The goal is to determine the partial derivative of λ ; therefore, it is necessary to isolate terms that will include this partial derivative from terms that will not. This can be performed by splitting the summation in Eq. (5.43) as follows:

$$\frac{\partial}{\partial \mathbf{A}_{\text{vec}}} \left\{ \sum_{k=0}^n c_k \lambda^{n-k} \right\} = \frac{\partial}{\partial \mathbf{A}_{\text{vec}}} \left\{ \sum_{k=0}^{n-1} c_k \lambda^{n-k} \right\} + \frac{\partial c_n}{\partial \mathbf{A}_{\text{vec}}} = 0.$$

Since the partial derivative and the summation are both linear operators, the partial derivative operator can be distributed inside the sum. Performing this step and taking into account the chain rule lead to the following:

$$\sum_{k=0}^{n-1} \left(c_k \lambda^{n-k-1} \frac{\partial \lambda}{\partial \mathbf{A}_{\text{vec}}} + \lambda^{n-k} \frac{\partial c_k}{\partial \mathbf{A}_{\text{vec}}} \right) + \frac{\partial c_n}{\partial \mathbf{A}_{\text{vec}}} = 0. \quad (5.44)$$

Here, the partial derivative of the coefficients of the characteristic equation are given by

$$\frac{\partial c_k}{\partial \mathbf{A}_{\text{vec}}} = \frac{(-1)^k}{k!} \frac{\partial |\mathbf{Q}_k|}{\partial \mathbf{A}_{\text{vec}}}. \quad (5.45)$$

From here, we can rearrange to solve for an equation for the Jacobian of the eigenvalue, resulting in

$$\mathbf{H}_\lambda = \frac{\partial \lambda}{\partial \mathbf{A}_{\text{vec}}} = \frac{- \left[\sum_{k=0}^{n-1} \left(\lambda^{n-k} \frac{\partial c_k}{\partial \mathbf{A}_{\text{vec}}} \right) + \frac{\partial c_n}{\partial \mathbf{A}_{\text{vec}}} \right]}{\sum_{k=0}^{n-1} c_k (n-k) \lambda^{n-k-1}}. \quad (5.46)$$

Once again, in the case of repeated eigenvalues, we will have a singularity with the denominator being equal to 0, thus why we only consider unique eigenvalues.

In Eqs. (5.45) and (5.46), every term is known except for $\partial |\mathbf{Q}_m| / \partial \mathbf{A}_{\text{vec}}$. This can be determined by first applying the chain rule as follows:

$$\frac{\partial |\mathbf{Q}_m|}{\partial \mathbf{A}_{\text{vec}}} = \frac{\partial |\mathbf{Q}_m|}{\partial (\mathbf{Q}_m)_{\text{vec}}} \frac{\partial (\mathbf{Q}_m)_{\text{vec}}}{\partial \mathbf{A}_{\text{vec}}}. \quad (5.47)$$

Now, using the rules of matrix-vector calculus, we show that

$$\frac{\partial |\mathbf{Q}_m|}{\partial (\mathbf{Q}_m)_{\text{vec}}} = |\mathbf{Q}_m| \text{vec} \left((\mathbf{Q}_m^{-T}) \right)^T. \quad (5.48)$$

Further, the term $\partial(\mathbf{Q}_m)_{\text{vec}}/\partial\mathbf{A}_{\text{vec}}$ can be expanded into the $m^2 \times n^2$ matrix given by

$$\frac{\partial(\mathbf{Q}_m)_{\text{vec}}}{\partial\mathbf{A}_{\text{vec}}} = \begin{bmatrix} (\text{vec}(\mathbf{I}_{n \times n}))^T \\ 2(\text{vec}(\mathbf{A}^T))^T \\ 3(\text{vec}((\mathbf{A}^2)^T))^T \\ \vdots \\ m(\text{vec}((\mathbf{A}^{m-1})^T))^T \\ \mathbf{0}_{1 \times n^2} \\ (\text{vec}(\mathbf{I}_{n \times n}))^T \\ 2(\text{vec}(\mathbf{A}^T))^T \\ \vdots \\ (m-1)(\text{vec}((\mathbf{A}^{m-2})^T))^T \\ \mathbf{0}_{2 \times n^2} \\ \vdots \end{bmatrix} \quad (5.49)$$

by considering the form of \mathbf{Q}_m . This completely defines the Jacobian of the eigenvalue.

As before, it is necessary to find the Jacobian of the eigenvectors, as well. Starting again by differentiating the standard eigenvalue problem with respect to \mathbf{A}_{vec} gives

$$(\mathbf{A} - \lambda\mathbf{I}) \frac{\partial\mathbf{v}}{\partial\mathbf{A}_{\text{vec}}} = \mathbf{v} \frac{\partial\lambda}{\partial\mathbf{A}_{\text{vec}}} - \mathbf{v}^H \otimes \mathbf{I}. \quad (5.50)$$

Now, utilizing the same techniques as in the individual cases, it can be shown that the general solution is given by

$$\mathbf{H}_{\mathbf{v}} = \frac{\partial\mathbf{v}}{\partial\mathbf{A}_{\text{vec}}} = (\mathbf{A} - \lambda\mathbf{I} + \mathbf{N})^{-1} \left[\mathbf{v} \frac{\partial\lambda}{\partial\mathbf{A}_{\text{vec}}} - \mathbf{v}^T \otimes \mathbf{I}_{n \times n} \right]. \quad (5.51)$$

5.4 Numeric Verification

In order to validate the derivatives presented in the above discussions, the analytic results were compared with numerical results obtained through forward finite differencing. This involved perturbing each element of \mathbf{A} and then calculating the eigenvalue and eigenvector of the perturbed matrix, such

that

$$\tilde{\mathbf{A}}_{i,j} = \mathbf{A} + \mathbf{A}_{\text{pert}}(i,j)$$

$$\mathbf{A}_{\text{pert}}(i,j) = \begin{cases} a_{kl} = 0 & k \neq i, l \neq j \\ a_{kl} = \text{pert} & k = i, l = j \end{cases} \quad (5.52)$$

$$\tilde{\lambda}_{i,j} = \text{eigval}(\tilde{\mathbf{A}}_{i,j}) \quad \tilde{\mathbf{v}}_{i,j} = \text{eigvec}(\tilde{\mathbf{A}}_{i,j}) \quad (5.53)$$

$$\left(\frac{\partial \lambda}{\partial \mathbf{A}_{\text{vec}}} \right)_{\text{num}} = \begin{bmatrix} \frac{\lambda - \tilde{\lambda}_{1,1}}{\text{pert}} & \frac{\lambda - \tilde{\lambda}_{2,1}}{\text{pert}} & \dots \\ \frac{\lambda - \tilde{\lambda}_{n,1}}{\text{pert}} & \dots & \frac{\lambda - \tilde{\lambda}_{n,n}}{\text{pert}} \end{bmatrix} \quad (5.54)$$

$$\left(\frac{\partial \mathbf{v}}{\partial \mathbf{A}_{\text{vec}}} \right)_{\text{num}} = \begin{bmatrix} \frac{\mathbf{v} - \tilde{\mathbf{v}}_{1,1}}{\text{pert}} & \frac{\mathbf{v} - \tilde{\mathbf{v}}_{2,1}}{\text{pert}} & \dots \\ \frac{\mathbf{v} - \tilde{\mathbf{v}}_{n,1}}{\text{pert}} & \dots & \frac{\mathbf{v} - \tilde{\mathbf{v}}_{n,n}}{\text{pert}} \end{bmatrix} \quad (5.55)$$

where *pert* is the size of the perturbation used for forward finite differencing.

The forward finite differencing described was applied to 5,000 random matrices of size 2×2 , 5,000 random matrices of size 3×3 , and 5,000 random matrices of size 10×10 . Then, the percent differences between the analytic and numeric derivatives were calculated. These data are shown as histograms in Fig. 5.2. As can be seen, the analytic differences matched numeric forward finite differencing well in all cases. We note that because of finite precision issues, we had to choose random matrices where the smallest of the elements in the eigenvector derivatives (in absolute value) was larger than the perturbation we used to perform the forward finite differencing. That is,

$$\min_{i,j} \left| \frac{\partial \mathbf{v}}{\partial \mathbf{A}_{\text{vec}}}(i,j) \right| > \text{pert}. \quad (5.56)$$

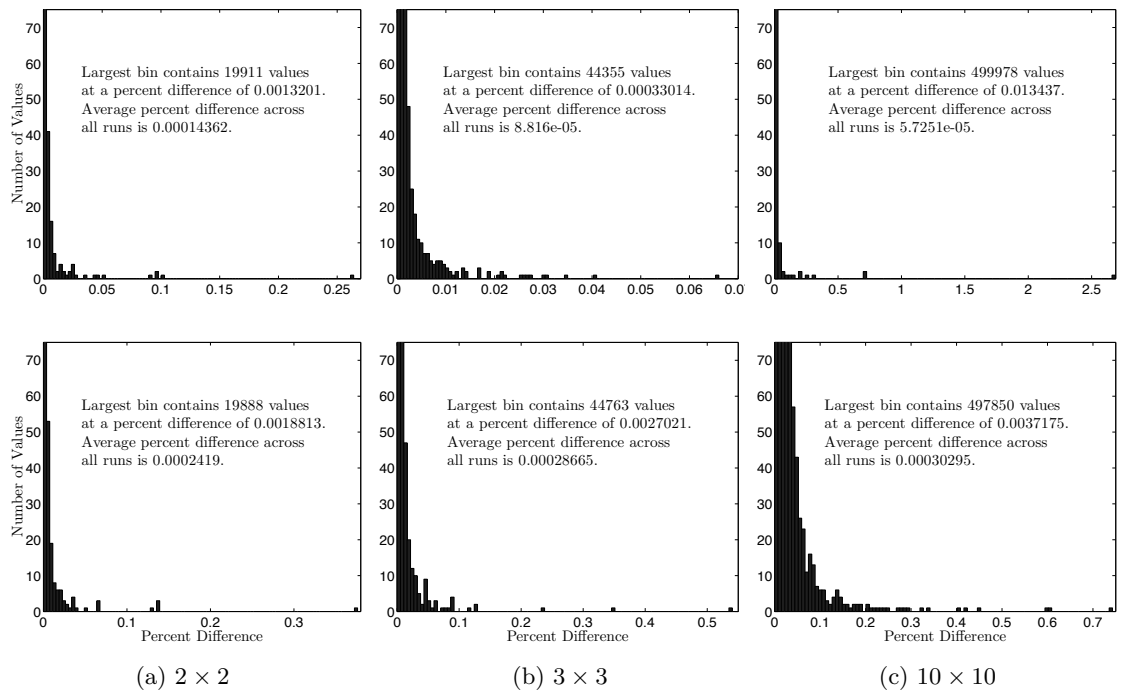


Figure 5.2: Histograms of percent difference between analytic derivatives (eigenvalue derivatives top and eigenvector derivatives bottom) and finite forward differencing for 5,000 randomly generated matrices.

Chapter 6

Numerically Efficient Eigenvalue and Eigenvector Jacobians

6.1 The General Case

Using the insights from Chapter 5, it is possible to arrive at a simpler and more numerically efficient solution. Beginning again with Eq. (2.1), we can left multiply by the Hermetian transpose of the normalization vector \mathbf{v}_0^H in order to form a new equation:

$$\mathbf{v}_0^H \mathbf{A} \mathbf{v} = \lambda \mathbf{v}_0^H \mathbf{v}. \quad (6.1)$$

Recall that while it is common to set $\mathbf{v}_0 = \mathbf{v}$, \mathbf{v}_0 is not a function of \mathbf{A} . Taking the derivative of Eq. (6.1) leads to

$$\mathbf{v}_0^H \mathbf{A} \frac{\partial \mathbf{v}}{\partial \mathbf{A}_{\text{vec}}} + \mathbf{v}^T \otimes \mathbf{v}_0^H = \mathbf{v}_0^H \mathbf{v} \frac{\partial \lambda}{\partial \mathbf{A}_{\text{vec}}} + \lambda \mathbf{v}_0^H \frac{\partial \mathbf{v}}{\partial \mathbf{A}_{\text{vec}}} \quad (6.2)$$

through simple application of the chain rule and identities pertaining to the vectorization of a matrix. Note again that a superscript of T indicates a standard transpose, while a superscript of H indicates the Hermetian (or conjugate) transpose. Now, recalling Eq. (5.13), the right-most term of Eq. (6.2) vanishes. Thus, after a simple rearrangement, we can find

$$\frac{\partial \lambda}{\partial \mathbf{A}_{\text{vec}}} = \mathbf{v}_0^H \mathbf{A} \frac{\partial \mathbf{v}}{\partial \mathbf{A}_{\text{vec}}} + \mathbf{v}^T \otimes \mathbf{v}_0^H, \quad (6.3)$$

which expresses the eigenvalue Jacobian as a function of \mathbf{A} , \mathbf{v} , \mathbf{v}_0 , $\partial \mathbf{v} / \partial \mathbf{A}_{\text{vec}}$, and α .

With an expression of the eigenvalue Jacobian in hand, we can now determine a compact expression for the eigenvector Jacobian. Observe that Eq. (5.50) and Eq. (6.3) create a system of two equations with two unknowns. Therefore, substituting Eq. (6.3) into Eq. (5.50) yields

$$(\mathbf{A} - \lambda \mathbf{I}) \frac{\partial \mathbf{v}}{\partial \mathbf{A}_{\text{vec}}} = \frac{\mathbf{v} \left[\mathbf{v}_0^H \mathbf{A} \frac{\partial \mathbf{v}}{\partial \mathbf{A}_{\text{vec}}} + \mathbf{v}^T \otimes \mathbf{v}_0^H \right]}{\alpha} - \mathbf{v}^T \otimes \mathbf{I}, \quad (6.4)$$

which can be arranged to give

$$\left[\mathbf{A} - \lambda \mathbf{I} - \frac{\mathbf{v} \mathbf{v}_0^H \mathbf{A}}{\alpha} \right] \frac{\partial \mathbf{v}}{\partial \mathbf{A}_{\text{vec}}} = \frac{\mathbf{v} (\mathbf{v}^T \otimes \mathbf{v}_0^H)}{\alpha} - \mathbf{v}^T \otimes \mathbf{I} \quad (6.5)$$

as an equation that isolates the eigenvector derivative. This expression can be simplified to

$$\left[\mathbf{A} - \lambda \mathbf{I} - \frac{\mathbf{v} \mathbf{v}_0^H \mathbf{A}}{\alpha} \right] \frac{\partial \mathbf{v}}{\partial \mathbf{A}_{\text{vec}}} = \mathbf{v}^T \otimes \left(\frac{\mathbf{v} \mathbf{v}_0^H}{\alpha} - \mathbf{I} \right) \quad (6.6)$$

by manipulating the Kronecker products.

The objective is now to solve Eq. (6.6) for $\partial \mathbf{v} / \partial \mathbf{A}_{\text{vec}}$. Unlike the result from Chapter 5, there is no need to incorporate a Null Space Matrix (or to use a pseudo-inverse) since the term $\mathbf{A} - \lambda \mathbf{I} - \mathbf{v} \mathbf{v}_0^H \mathbf{A} / \alpha$ is already full rank and invertible. This fact is straightforward to see by considering the column space of the term $\mathbf{A} - \lambda \mathbf{I}$, which will generally be rank $n - 1$ (assuming λ is simple and \mathbf{A} is full rank). Specifically, $\mathbf{A} - \lambda \mathbf{I}$ spans \mathbb{R}^{n-1} with a null space in the direction of \mathbf{v} . Now we can consider the column space of the term $\mathbf{v} \mathbf{v}_0^H \mathbf{A} / \alpha$, which is rank one and spans only \mathbf{v} . Therefore, by adding $\mathbf{v} \mathbf{v}_0^H \mathbf{A} / \alpha$ to $\mathbf{A} - \lambda \mathbf{I}$, the resulting column space spans all of \mathbb{R}^n , making the overall term full rank and invertible.

In light of this fact, the solution for the eigenvector Jacobian is given by

$$\frac{\partial \mathbf{v}}{\partial \mathbf{A}_{\text{vec}}} = \left[\mathbf{A} - \lambda \mathbf{I} - \frac{\mathbf{v} (\mathbf{v}_0^H \mathbf{A})}{\alpha} \right]^{-1} \left[\mathbf{v}^T \otimes \left(\frac{\mathbf{v} \mathbf{v}_0^H}{\alpha} - \mathbf{I} \right) \right], \quad (6.7)$$

which is a function of only \mathbf{A} , λ , \mathbf{v} , \mathbf{v}_0 , and α . Additionally, manipulating the Kronecker products allows for a final form of

$$\frac{\partial \mathbf{v}}{\partial \mathbf{A}_{\text{vec}}} = \mathbf{v}^T \otimes \left[\left(\mathbf{A} - \lambda \mathbf{I} - \frac{\mathbf{v} \mathbf{v}_0^H \mathbf{A}}{\alpha} \right)^{-1} \left(\frac{\mathbf{v} \mathbf{v}_0^H}{\alpha} - \mathbf{I} \right) \right]. \quad (6.8)$$

Once the eigenvector Jacobian has been calculated, we can calculate the eigenvalue Jacobian by simply plugging the result of Eq. (6.8) into Eq. (6.3). Making this substitution and simplifying

yields the final expression for the eigenvalue Jacobian:

$$\frac{\partial \lambda}{\partial \mathbf{A}_{\text{vec}}} = \frac{\mathbf{v}^T}{\alpha} \otimes \left[\mathbf{v}_0^H \mathbf{A} \left(\mathbf{A} - \lambda \mathbf{I} - \frac{\mathbf{v} \mathbf{v}_0^H \mathbf{A}}{\alpha} \right)^{-1} \left(\frac{\mathbf{v} \mathbf{v}_0^H}{\alpha} - \mathbf{I} \right) + \mathbf{v}_0^H \right]. \quad (6.9)$$

6.2 Simplified Cases

The overall goal of this work was to create a numerically efficient algorithm for the calculation of the Jacobians of an eigenvalue and eigenvector with respect to the elements of the parent matrix. Now that we have developed these expressions, it is beneficial to discuss how they simplify as assumptions are placed on the parent matrix. A variety of simplifications are possible by imposing structure on \mathbf{A} and subsequently Eqs. (6.8) and (6.9). Here we will focus on two important and particularly useful simplifications as a way to gain key insights and show connections with existing literature.

6.2.1 Real Symmetric Parent Matrix

The first simplified case we will consider, when the parent matrix of the eigenvalues and eigenvectors is real and symmetric, frequently appears in problems (for instance, the Davenport solution to Wahba's Problem [36]). In order to simplify and be able to match what is in the existing literature, it is also necessary to make a choice for \mathbf{v}_0 ; therefore, we can choose that $\mathbf{v}_0 = \mathbf{v}$. Note that as discussed in Sect. 4.2, this is the choice usually made in practice regardless, as it leads to the best condition for the calculation of the eigenvector derivative.

To begin the simplifications for the symmetric case, consider Eq. (6.5), repeated here for convenience:

$$[\mathbf{A} - \lambda \mathbf{I} - \mathbf{v} \mathbf{v}^T \mathbf{A}] \frac{\partial \mathbf{v}}{\partial \mathbf{A}_{\text{vec}}} = -\mathbf{v}^T \otimes (\mathbf{I} - \mathbf{v} \mathbf{v}^T). \quad (6.10)$$

Note that we have replaced the Hermetian transposes with standard transposes, since the eigenvalues and eigenvectors are guaranteed to be real since \mathbf{A} is real symmetric. Also note that \mathbf{v}_0 has been replaced by \mathbf{v} . Now, as discussed before, the coefficient matrix of the eigenvector Jacobian matrix is full rank and invertible as long as the matrix \mathbf{A} is full rank and invertible. Making use of the fact that for an invertible matrix

$$\mathbf{A}^\dagger = \mathbf{A}^{-1}, \quad (6.11)$$

where \mathbf{A}^\dagger is the Moore-Penrose pseudo-inverse of \mathbf{A} [94], it is possible to write

$$\frac{\partial \mathbf{v}}{\partial \mathbf{A}_{\text{vec}}} = -\mathbf{v}^T \otimes \left[(\mathbf{A} - \lambda \mathbf{I} - \mathbf{v} \mathbf{v}^T \mathbf{A})^\dagger (\mathbf{I} - \mathbf{v} \mathbf{v}^T) \right]. \quad (6.12)$$

From here it is necessary to further consider the pseudo-inverse term. Recognize that the pseudo-inverse term in Eq. (6.12) can be expressed as the addition of a matrix and an outer product

$$(\mathbf{A} - \lambda\mathbf{I} - \mathbf{v}\mathbf{v}^T\mathbf{A})^\dagger = (\mathbf{B} + \mathbf{c}\mathbf{d}^T)^\dagger \quad (6.13)$$

where $\mathbf{B} = \mathbf{A} - \lambda\mathbf{I}$, $\mathbf{c} = -\mathbf{v}$, and $\mathbf{d} = \mathbf{A}\mathbf{v} = \lambda\mathbf{v}$. Using the case i identities presented in [95], when \mathbf{A} is real and symmetric, we can find that

$$(\mathbf{A} - \lambda\mathbf{I} - \mathbf{v}\mathbf{v}^T\mathbf{A})^\dagger = (\mathbf{A} - \lambda\mathbf{I})^\dagger - \lambda^{-1}\mathbf{v}\mathbf{v}^T. \quad (6.14)$$

Substituting this result into Eq. (6.12) yields

$$\frac{\partial\mathbf{v}}{\partial\mathbf{A}_{\text{vec}}} = -\mathbf{v}^T \otimes \left([(\mathbf{A} - \lambda\mathbf{I})^\dagger - \lambda^{-1}\mathbf{v}\mathbf{v}^T] [\mathbf{I} - \mathbf{v}\mathbf{v}^T] \right). \quad (6.15)$$

Now, expanding the matrix multiplication in the Kronecker product gives

$$\frac{\partial\mathbf{v}}{\partial\mathbf{A}_{\text{vec}}} = -\mathbf{v}^T \otimes \left[(\mathbf{A} - \lambda\mathbf{I})^\dagger - \lambda^{-1}\mathbf{v}\mathbf{v}^T - (\mathbf{A} - \lambda\mathbf{I})^\dagger \mathbf{v}\mathbf{v}^T + \lambda^{-1}\mathbf{v}\mathbf{v}^T \mathbf{v}\mathbf{v}^T \right], \quad (6.16)$$

which, when taking into account that for symmetric matrices the pseudo-inverse has the same null space as the matrix itself, simplifies to

$$\frac{\partial\mathbf{v}}{\partial\mathbf{A}_{\text{vec}}} = -\mathbf{v}^T \otimes (\mathbf{A} - \lambda\mathbf{I})^\dagger. \quad (6.17)$$

This is exactly the same result presented in [41].

With the simplified version of the eigenvector derivative in hand, the simplified eigenvalue Jacobian is trivial to find. To begin, substitute Eq. (6.17) into Eq. (6.3) to obtain

$$\frac{\partial\lambda}{\partial\mathbf{A}_{\text{vec}}} = \mathbf{v}^T \otimes \left[-\mathbf{v}^T \mathbf{A} (\mathbf{A} - \lambda\mathbf{I})^\dagger \right] + \mathbf{v}^T \otimes \mathbf{v}^T. \quad (6.18)$$

Now, making use of the fact that $\mathbf{v}^T \mathbf{A} = \lambda\mathbf{v}^T$ for symmetric matrices and the sharing of the null spaces, this reduces to

$$\frac{\partial\lambda}{\partial\mathbf{A}_{\text{vec}}} = \mathbf{v}^T \otimes \mathbf{v}^T \quad (6.19)$$

which, again, is exactly the same as that presented in [41]. Note that these expressions do not assume that the parent matrix is perturbed symmetrically.

Thus, the general eigenvalue and eigenvector Jacobians presented in Section 6.1 cleanly simplify to the results from [41] for the special case when \mathbf{A} is symmetric. This same result can be achieved by assuming the symmetry of \mathbf{A} and enforcing $\mathbf{v}_0 \equiv \mathbf{v}$ in Eqs. (5.50) and (6.3), as was done in [41].

6.2.2 Real Diagonal Parent Matrix

The second simplified case considered is a diagonal parent matrix with only real valued elements. In this case, the eigenvalues of the matrix are simply the diagonal elements of \mathbf{A} , and the eigenvectors are the standard basis. While this case is trivial, it leads to some powerful insights into the overall problem.

Simplified Jacobians for a Diagonal Matrix

To develop the simplified derivatives for a diagonal matrix, begin with the simplified derivatives for the symmetric case given in Eq. (6.17) and Eq. (6.19). Now, recognizing that the pseudo-inverse of a diagonal matrix is just the reciprocal of the non-zero diagonal elements, the eigenvector derivative simplifies to

$$\begin{aligned} \frac{\partial \mathbf{v}_i}{\partial \mathbf{A}_{\text{vec}}} &= -\mathbf{e}_i \otimes \text{diag} [(\lambda_1 - \lambda_i)^{-1}, \dots, (\lambda_{i-1} - \lambda_i)^{-1}, 0, (\lambda_{i+1} - \lambda_i)^{-1}, \dots, (\lambda_n - \lambda_i)^{-1}] \\ &= - \begin{bmatrix} \mathbf{0}_{n \times (i-1)n} & \text{diag} [\{(\lambda_k - \lambda_i)^\dagger\}] & \mathbf{0}_{n \times (n-i)n} \end{bmatrix} \end{aligned} \quad (6.20)$$

where \mathbf{e}_i is the i^{th} standard basis vector and the derivatives presented are for the i^{th} eigenvalue and eigenvector (at this point, it becomes necessary to distinguish the eigenvalue and eigenvector being considered). Recall that the pseudo-inverse of a non-zero scalar is the reciprocal, while the pseudo-inverse of a zero scalar is 0.

The eigenvalue derivative simplifies similarly:

$$\begin{aligned} \frac{\partial \lambda_i}{\partial \mathbf{A}_{\text{vec}}} &= \mathbf{e}_i \otimes \mathbf{e}_i \\ &= \begin{bmatrix} \mathbf{0}_{1 \times (i-1)n} & \mathbf{e}_i^T & \mathbf{0}_{1 \times (n-i)n} \end{bmatrix}. \end{aligned} \quad (6.21)$$

Perturbations to the Eigen-space of a Diagonal Matrix

With the simplified relationships in hand, we can make some interesting observations on the perturbation of the eigen-space. The first observation is that to perturb the i^{th} eigenvalue, we must perturb the i^{th} diagonal element of the diagonal parent matrix. Further, the perturbation to the

eigenvalue in this case is exactly the perturbation to the parent matrix. While this observation should be trivial (since the diagonal elements are the eigenvalues themselves), it leads to a more interesting observation for the general case, as will be discussed later. This observation can be expressed mathematically as

$$\Delta\lambda_i = \delta_i \tag{6.22}$$

when

$$\Delta\mathbf{A} = \delta_i \mathbf{e}_i \mathbf{e}_i^T = \delta_i \mathbf{v}_i \mathbf{v}_i^T. \tag{6.23}$$

The next observation is that the eigenvector is only perturbed when the i^{th} column of the parent matrix is perturbed. Further, there is an analytic relationship between the change to the eigenvector, the eigenvalues, and the perturbation itself. Mathematically, this is expressed as

$$\Delta\mathbf{v}_i = - \sum_{k=1, k \neq i}^n \frac{\delta_k}{\lambda_k - \lambda_i} \mathbf{e}_k = - \sum_{k=1, k \neq i}^n \frac{\delta_k}{\lambda_k - \lambda_i} \mathbf{v}_k \tag{6.24}$$

when

$$\Delta\mathbf{A} = \sum_{k=1, k \neq i}^n \delta_k \mathbf{e}_k \mathbf{e}_i^T = \sum_{k=1, k \neq i}^n \delta_k \mathbf{v}_k \mathbf{v}_i^T. \tag{6.25}$$

These relations show that we can express the eigenvector derivative for a diagonal matrix as a modal expansion of the other eigenvectors if the coefficients δ_k can be calculated (which is quite simple for a diagonal matrix since the eigenvectors are the standard basis).

Perturbations to the Eigen-space of a Diagonalizable Matrix

Now, reconsider the case when the parent matrix is real symmetric (as was done for the previous section). Since the parent matrix is symmetric, the eigenvectors will form an orthonormal basis for \mathbb{R}^n , and the matrix is diagonalizable as

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T \tag{6.26}$$

where $\mathbf{\Lambda}$ is a diagonal matrix of the eigenvalues and \mathbf{V} is an orthogonal matrix whose columns are the eigenvectors of \mathbf{A} . Substituting this into the standard eigenvalue problem gives

$$\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T \mathbf{v}_i = \lambda_i \mathbf{v}_i,$$

which can be rewritten as

$$\mathbf{\Lambda} \mathbf{V}^T \mathbf{v}_i = \lambda_i \mathbf{V}^T \mathbf{v}_i. \quad (6.27)$$

Since the columns of \mathbf{V} are made up of the orthogonal eigenvectors of \mathbf{A} ,

$$\mathbf{V}^T \mathbf{v}_i = \mathbf{e}_i, \quad (6.28)$$

which quickly reduces Eq. (6.27) down to the diagonalized eigen-system:

$$\mathbf{\Lambda} \mathbf{e}_i = \lambda_i \mathbf{e}_i. \quad (6.29)$$

Now suppose that the matrix \mathbf{A} is perturbed by adding a matrix $\Delta \mathbf{A}$. In the diagonalized space, this matrix perturbation can be expressed as

$$\Delta \mathbf{\Lambda} = \mathbf{V}^T \Delta \mathbf{A} \mathbf{V} \quad (6.30)$$

where $\Delta \mathbf{\Lambda}$ is an additive update to $\mathbf{\Lambda}$. The matrix $\Delta \mathbf{\Lambda}$ will not generally be diagonal.

Now that the problem has been diagonalized, the observations described above can be utilized. Since the eigenvectors in the diagonalized space form the standard basis, it is clear that the matrix $\Delta \mathbf{\Lambda}$ can be decomposed as

$$\Delta \mathbf{\Lambda} = \sum_{i=1}^n \sum_{k=1}^n \delta_{ki} \mathbf{e}_k \mathbf{e}_i^T \quad (6.31)$$

where

$$\Delta \mathbf{\Lambda} = \begin{bmatrix} \delta_{11} & \delta_{12} & \dots & \delta_{1n} \\ \delta_{21} & \delta_{22} & \dots & \delta_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{n1} & \delta_{n2} & \dots & \delta_{nn} \end{bmatrix} \quad (6.32)$$

and

$$\delta_{ki} = \mathbf{v}_k^T \Delta \mathbf{A} \mathbf{v}_i. \quad (6.33)$$

Now, analogous to relations in Eq. (6.22) and Eq. (6.24), the perturbations in the diagonalized space

are described by

$$\Delta\lambda_i = \delta_{ii} \quad (6.34)$$

$$\Delta\mathbf{e}_i = - \sum_{k=1, k \neq i}^n \frac{\delta_{ki}}{\lambda_k - \lambda_i} \mathbf{e}_k. \quad (6.35)$$

These perturbations must now be related to perturbations in the original eigenvectors, \mathbf{v}_i . For an additive update of the diagonalized eigenvector we have

$$\mathbf{v}_i + \Delta\mathbf{v}_i = \mathbf{V}(\mathbf{e}_i + \Delta\mathbf{e}_i). \quad (6.36)$$

Thus, it becomes apparent that the update to the original eigenvector is given by

$$\begin{aligned} \Delta\mathbf{v}_i &= \mathbf{V}\Delta\mathbf{e}_i \\ &= - \sum_{k=1, k \neq i}^n \frac{\delta_{ki}}{\lambda_k - \lambda_i} \mathbf{v}_k. \end{aligned} \quad (6.37)$$

Further, taking the inverse of Eq. (6.30) combined with Eq. (6.31) gives

$$\begin{aligned} \Delta\mathbf{A} &= \mathbf{V} \sum_{i=1}^n \sum_{k=1}^n \delta_{ki} \mathbf{e}_k \mathbf{e}_i^T \mathbf{V}^T \\ &= \sum_{i=1}^n \sum_{k=1}^n \delta_{ki} \mathbf{v}_k \mathbf{v}_i^T, \end{aligned} \quad (6.38)$$

which shows that any perturbation can be expressed as a linear combination of the outer products of the eigenvectors of any symmetric matrix where the coefficients are found using Eq. (6.33). While this approach is not efficient, it provides an interesting parallel to the modal expansion techniques discussed in [63–70], as well as stability theory for eigenvectors [96]. Interestingly, the result arrived at in Eq. (6.37) is exactly that from [96] for the symmetric case.

Further, if the update to \mathbf{A} is defined to be $\Delta\mathbf{A} = \partial A_{ij} \mathbf{e}_i \mathbf{e}_j^T$, then it becomes possible to find the derivatives for a perturbation to any element of \mathbf{A} as

$$\frac{\partial \mathbf{v}_i}{\partial A_{lm}} = - \sum_{k=1, k \neq i}^n \frac{v_{kl} v_{im}}{\lambda_k - \lambda_i} \mathbf{v}_k \quad (6.39)$$

where $\partial \mathbf{v}_i / \partial A_{lm}$ is the partial derivative of the i^{th} eigenvector with respect to the $(l, m)^{\text{th}}$ element of the matrix \mathbf{A} , and v_{ab} is the b^{th} element of the a^{th} eigenvector of \mathbf{A} , which is very similar to

what is done in [63–70]. A similar proof using left eigenvectors can be shown for the case of any diagonalizable matrix.

In summary, it once again becomes evident that the very general and very efficient expressions for eigenvalue and eigenvector Jacobians presented in Section 6.1 may be reduced to a variety of important special cases presented in the literature. In addition, these simplifications provide powerful insight into the structure and dynamics of the eigenvalue and eigenvector Jacobian problem.

6.3 Numerical Validation

Forward finite differencing was used to validate the formulation of the new eigenvalue and eigenvector Jacobians presented in Section 6.1. This provides a numerical approximation of the Jacobians, which may be compared with the analytic expressions developed in this chapter.

The forward finite differencing was performed by perturbing each element of the parent matrix individually in order to calculate each element of the Jacobians. The analytic derivatives from Eq. (6.7) and Eq. (6.9) were then compared with the finite differences, and the percent differences were calculated. This was performed for 5,000 randomly generated complex matrices of size 2×2 , 5,000 randomly generated complex matrices of size 3×3 , and 5,000 randomly generated complex matrices of size 10×10 . The results for both the eigenvalue and eigenvector derivatives are shown in the histograms in Fig. 6.1. Note that due to finite precision issues, matrices had to be ignored where the smallest component of the eigenvector derivatives was less than the perturbation size used in the finite differencing. As can be seen in the figure, the new method performed well in every instance, well below 0.1% difference for each and every element of the eigenvalue and eigenvector Jacobians. In addition, the output from the techniques derived in this chapter matched to within machine precision the outputs from Eqs. (5.46) and (5.51).

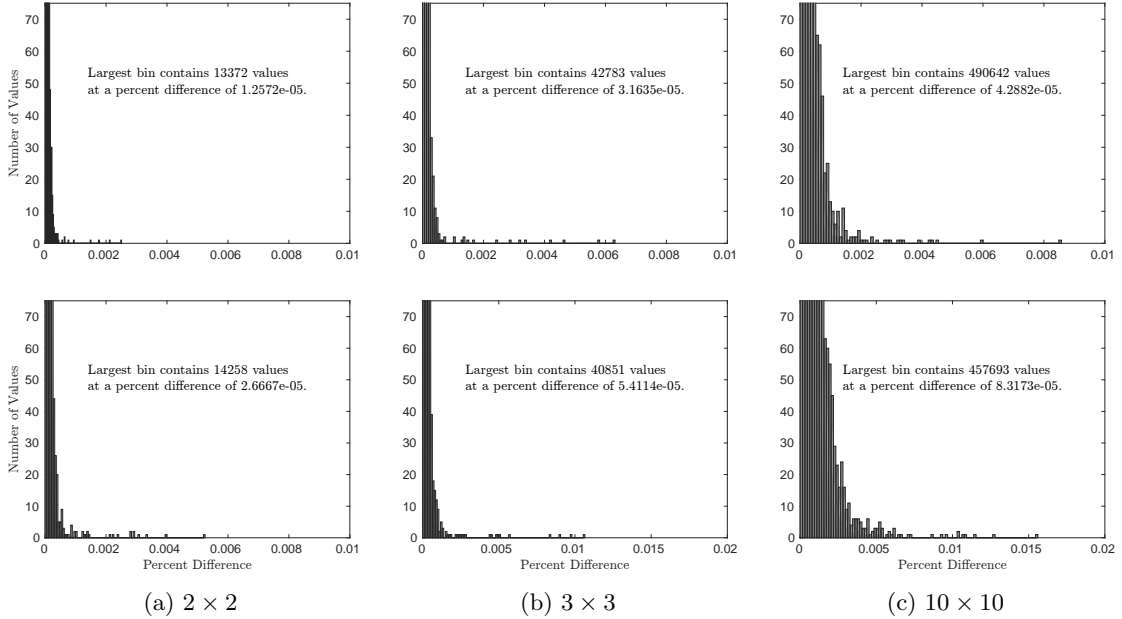


Figure 6.1: Histograms of percent difference between analytic derivatives computed using Eqs. (6.9) and (6.8) (eigenvalue derivatives top and eigenvector derivatives bottom) and finite forward differencing for 5,000 randomly generated matrices of each size. The histograms are of the percent difference for each element of the eigenvalue and eigenvector derivatives (for example, for each $n \times n$ matrix, there are n^2 eigenvalue derivative elements and $n \times n^2$ eigenvector derivative elements).

6.4 Comparison of Performance

The primary goal of the derivations presented in this chapter was to decrease the computational complexity of those presented in Chapter 5. An examination of the two formulations, however, indicates that both techniques are $\mathcal{O}(n^4)$ due to the $n \times n$ by $n \times n^2$ multiplication in Eq. (6.7) and the $\text{Tr}[\mathbf{A}^n]$ term in Eq. (5.27).¹ Despite the fact that both these formulations have the same upper limit on their computational complexity, it should be clear that the new formulation is much simpler, both in terms of operations performed² and (perhaps more importantly) in terms of memory use.

A simulation was run in an attempt to detail the increase in computational efficiency from the technique in [47] to the technique presented in this paper. The simulation was performed by applying each technique in turn to 50 randomly generated matrices (the same 50 matrices for each technique) at matrix sizes varying from 2 to 50. For each run, the computation time of each method was recorded. Finally, the minimum computation time for each matrix size was chosen for each method, and the results are shown in Fig. 6.2. As can be seen in the Figure, the new method is at minimum

¹This is assuming that the technique used to calculate the determinant is faster than $\mathcal{O}(n!)$, as this is the case in most modern linear algebra libraries.

²The formulation from Chapter 5 has two operations that are of order $\mathcal{O}(n^4)$ as opposed to one for the formulations proposed here.

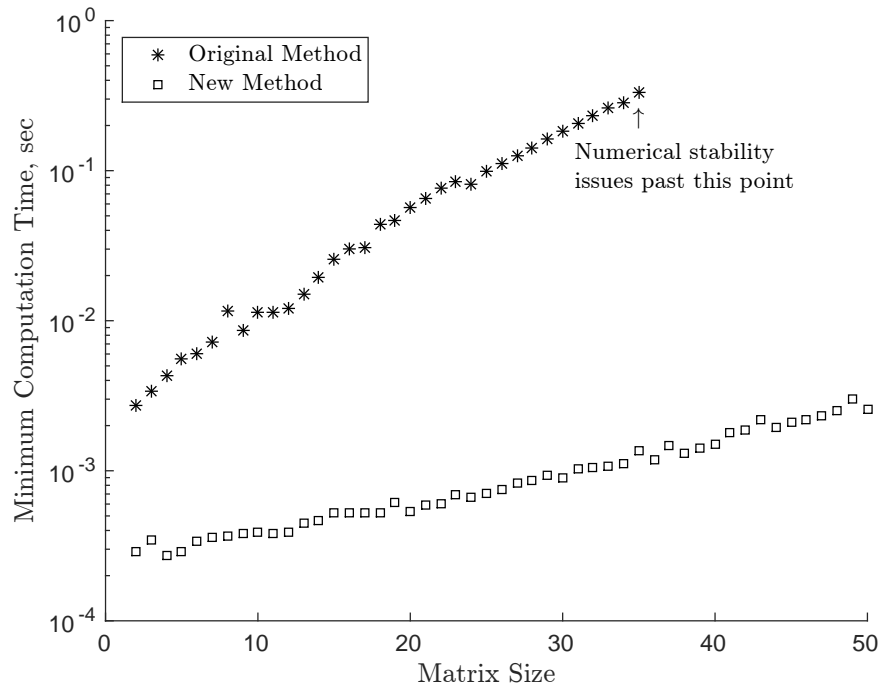


Figure 6.2: A plot of minimum computation time versus matrix size for the method from Chapter 5 (original method) and the method proposed in this Chapter (new method). Note that the method from Chapter 5 encounters numerical stability issues around a matrix size of 35 due to Eq. (5.48). This is why there is a cut-off in the data.

an order of magnitude faster, and the distance between the performance of the two methods increases as the matrix size increases. Also note that the new method is less susceptible to numerical precision issues, as is evidenced by the cut-off of the results for the method from Chapter 5.

These simulations were performed using an Intel Core i7-3770 processor at 3.4 GHz and the MATLAB programming language.

6.5 Extension to the Generalized Eigenvalue Problem

An additional benefit of the numerically efficient technique is that the ideas are easy to extend to the general eigenvalue and eigenvector problem [97]

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{B}\mathbf{v}. \tag{6.40}$$

To begin, differentiate with respect to a parameter vector \mathbf{x} to get

$$\mathbf{A} \frac{\partial \mathbf{v}}{\partial \mathbf{x}} + (\mathbf{v}^T \otimes \mathbf{I}) \frac{\partial \mathbf{A}_{\text{vec}}}{\partial \mathbf{x}} = \mathbf{B} \mathbf{v} \frac{\partial \lambda}{\partial \mathbf{x}} + \lambda (\mathbf{v} \otimes \mathbf{I}) \frac{\partial \mathbf{B}_{\text{vec}}}{\partial \mathbf{x}} + \lambda \mathbf{B} \frac{\partial \mathbf{v}}{\partial \mathbf{x}}, \quad (6.41)$$

which again provides a single equation with two unknowns. We can simplify this expression to result in

$$(\mathbf{A} - \lambda \mathbf{B}) \frac{\partial \mathbf{v}}{\partial \mathbf{x}} = \mathbf{B} \mathbf{v} \frac{\partial \lambda}{\partial \mathbf{x}} + \lambda (\mathbf{v}^T \otimes \mathbf{I}) \frac{\partial \mathbf{B}_{\text{vec}}}{\partial \mathbf{x}} - (\mathbf{v}^T \otimes \mathbf{I}) \frac{\partial \mathbf{A}_{\text{vec}}}{\partial \mathbf{x}}. \quad (6.42)$$

The parameter vector \mathbf{x} contains parameters that \mathbf{A} , \mathbf{B} , or both are functions of. In order to develop the full Jacobians for the eigenvalues of \mathbf{A} and \mathbf{B} , we can simply set the parameter vector to be the vectorized forms of \mathbf{A} and \mathbf{B} respectively, as we will do later. Regardless of what is chosen for the parameter vector, the resulting form is the same.

Now, we again need to make use of the normalization vector to provide the second equation and to make the eigenvectors unique and analytic, as discussed in Section 4.1. There are two different normalizations that make our eigenvectors analytic:

$$\mathbf{v}_0^H \mathbf{v} = \alpha \quad (6.43)$$

$$\mathbf{v}_0^H \mathbf{B} \mathbf{v} = \alpha. \quad (6.44)$$

The choice of the normalization used depends on the rest of the problem being solved; therefore, we will discuss the solution for each.

Regardless of the normalization chosen, the first step is to left multiply Eq. (6.41) by \mathbf{v}_0^H , which results in

$$\mathbf{v}_0^H (\mathbf{A} - \lambda \mathbf{B}) \frac{\partial \mathbf{v}}{\partial \mathbf{x}} = \mathbf{v}_0^H \mathbf{B} \mathbf{v} \frac{\partial \lambda}{\partial \mathbf{x}} + \mathbf{v}_0^H \lambda (\mathbf{v}^T \otimes \mathbf{I}) \frac{\partial \mathbf{B}_{\text{vec}}}{\partial \mathbf{x}} - \mathbf{v}_0^H (\mathbf{v}^T \otimes \mathbf{I}) \frac{\partial \mathbf{A}_{\text{vec}}}{\partial \mathbf{x}}. \quad (6.45)$$

This can be rearranged to get

$$\frac{\partial \lambda}{\partial \mathbf{x}} = \frac{1}{\mathbf{v}_0^H \mathbf{B} \mathbf{v}} \left[\mathbf{v}_0^H \mathbf{E} \frac{\partial \mathbf{v}}{\partial \mathbf{x}} - (\mathbf{v}^T \otimes \mathbf{v}_0^H) \left(\lambda \frac{\partial \mathbf{B}_{\text{vec}}}{\partial \mathbf{x}} - \frac{\partial \mathbf{A}_{\text{vec}}}{\partial \mathbf{x}} \right) \right] \quad (6.46)$$

where the system matrix \mathbf{E} is given by

$$\mathbf{E} = \mathbf{A} - \lambda \mathbf{B}. \quad (6.47)$$

Now we have a system of two equations with two unknowns that can easily be solved by substi-

tuting Eq. (6.46) into Eq. (6.42). Making this substitution yields

$$\begin{aligned} \mathbf{E} \frac{\partial \mathbf{v}}{\partial \mathbf{x}} = & \frac{\mathbf{B}\mathbf{v}}{\mathbf{v}_0^H \mathbf{B}\mathbf{v}} \left[\mathbf{v}_0^H \mathbf{E} \frac{\partial \mathbf{v}}{\partial \mathbf{x}} - (\mathbf{v}^T \otimes \mathbf{v}_0^H) \left(\lambda \frac{\partial \mathbf{B}_{\text{vec}}}{\partial \mathbf{x}} - \frac{\partial \mathbf{A}_{\text{vec}}}{\partial \mathbf{x}} \right) \right] \\ & + (\mathbf{v}^T \otimes \mathbf{I}) \left(\lambda \frac{\partial \mathbf{B}_{\text{vec}}}{\partial \mathbf{x}} - \frac{\partial \mathbf{A}_{\text{vec}}}{\partial \mathbf{x}} \right), \end{aligned} \quad (6.48)$$

which can be rearranged to get

$$\Psi \mathbf{E} \frac{\partial \mathbf{v}}{\partial \mathbf{x}} = (\mathbf{v}^T \otimes \Psi) \left(\lambda \frac{\partial \mathbf{B}_{\text{vec}}}{\partial \mathbf{x}} - \frac{\partial \mathbf{A}_{\text{vec}}}{\partial \mathbf{x}} \right) \quad (6.49)$$

where the projection matrix Ψ is given by

$$\Psi = \mathbf{I} - \frac{\mathbf{B}\mathbf{v}\mathbf{v}_0^H}{\mathbf{v}_0^H \mathbf{B}\mathbf{v}}. \quad (6.50)$$

The choice of normalization will determine how we solve this system, since the term $\Psi \mathbf{E}$ is rank deficient with a null space in the direction of \mathbf{v} . The following subsections address each of the two possible normalizations in turn.

6.5.1 A Solution for the Normalization $\mathbf{v}_0^H \mathbf{v} = \alpha$

First we consider the case where $\mathbf{v}_0^H \mathbf{v} = \alpha$. The derivative of this normalization is given by

$$\mathbf{v}_0^H \frac{\partial \mathbf{v}}{\partial \mathbf{x}} = \mathbf{0}_{n \times j} \quad (6.51)$$

where j is the length of \mathbf{x} . This tells us that the eigenvector Jacobian must be orthogonal to the normalization vector, as it did before. Again, this observation leads to a solution either using the Null Space Matrix or a pseudo-inverse. Therefore, the solution for the eigenvector Jacobian is given by

$$\frac{\partial \mathbf{v}}{\partial \mathbf{x}} = \left[\mathbf{v}^T \otimes (\Psi \mathbf{E} + \mathbf{N})^{-1} \Psi \right] \left(\lambda \frac{\partial \mathbf{B}_{\text{vec}}}{\partial \mathbf{x}} - \frac{\partial \mathbf{A}_{\text{vec}}}{\partial \mathbf{x}} \right) \quad (6.52)$$

where \mathbf{N} is the Null Space Matrix as given in Eq. (5.14). This approach is analogous to the technique used in section 5.1.

Substituting this result in Eq. (6.46) and simplifying gives the eigenvalue Jacobian as

$$\frac{\partial \lambda}{\partial \mathbf{x}} = \left(\frac{\mathbf{v}^T}{\mathbf{v}_0^H \mathbf{B}\mathbf{v}} \otimes \mathbf{v}_0^H \left[\mathbf{E} (\Psi \mathbf{E} + \mathbf{N})^{-1} \Psi - \mathbf{I} \right] \right) \left(\lambda \frac{\partial \mathbf{B}_{\text{vec}}}{\partial \mathbf{x}} - \frac{\partial \mathbf{A}_{\text{vec}}}{\partial \mathbf{x}} \right). \quad (6.53)$$

Using a pseudo-inverse instead of the Null Space Matrix results in the same forms with $(\Psi\mathbf{E} + \mathbf{N})^{-1}$ replaced by $(\Psi\mathbf{E})^\dagger$.

6.5.2 A Solution for the Normalization $\mathbf{v}_0^H \mathbf{B} \mathbf{v} = \alpha$

We now consider the case where $\mathbf{v}_0^H \mathbf{B} \mathbf{v} = \alpha$. The derivative of this normalization is given by

$$\mathbf{v}_0^H \mathbf{B} \frac{\partial \mathbf{v}}{\partial \mathbf{x}} + (\mathbf{v}^T \otimes \mathbf{v}_0^H) \frac{\partial \mathbf{B}_{\text{vec}}}{\partial \mathbf{x}} = \mathbf{0}_{n \times j}. \quad (6.54)$$

While this result is certainly more complicated than that in Eq. (6.51), it provides a means to avoid the Null Space Matrix and pseudo-inverse entirely. Consider the left-hand side of Eq. (6.49). If we expand the multiplication, we are left with

$$\mathbf{A} \frac{\partial \mathbf{v}}{\partial \mathbf{x}} - \lambda \mathbf{B} \frac{\partial \mathbf{v}}{\partial \mathbf{x}} - \frac{\mathbf{B} \mathbf{v} \mathbf{v}_0^H \mathbf{A}}{\alpha} \frac{\partial \mathbf{v}}{\partial \mathbf{x}} - \frac{\lambda}{\alpha} \mathbf{B} \mathbf{v} \mathbf{v}_0^H \mathbf{B} \frac{\partial \mathbf{v}}{\partial \mathbf{x}}. \quad (6.55)$$

Now we can make the substitution given in Eq. (6.54) on the right-most term, resulting in

$$\mathbf{A} \frac{\partial \mathbf{v}}{\partial \mathbf{x}} - \lambda \mathbf{B} \frac{\partial \mathbf{v}}{\partial \mathbf{x}} - \frac{\mathbf{B} \mathbf{v} \mathbf{v}_0^H \mathbf{A}}{\alpha} \frac{\partial \mathbf{v}}{\partial \mathbf{x}} + \frac{\lambda}{\alpha} \mathbf{B} \mathbf{v} (\mathbf{v}^T \otimes \mathbf{v}_0^H) \frac{\partial \mathbf{B}_{\text{vec}}}{\partial \mathbf{x}}. \quad (6.56)$$

Having made this substitution, we can return to Eq. (6.49), which, after algebraic simplification, becomes

$$(\Psi \mathbf{A} - \lambda \mathbf{B}) \frac{\partial \mathbf{v}}{\partial \mathbf{x}} = \lambda (\mathbf{v}^T \otimes \mathbf{I}) \frac{\partial \mathbf{B}_{\text{vec}}}{\partial \mathbf{x}} - (\mathbf{v}^T \otimes \Psi) \frac{\partial \mathbf{A}_{\text{vec}}}{\partial \mathbf{x}}. \quad (6.57)$$

All that remains is to solve for the eigenvector Jacobian. It turns out that as long as the eigenvalue being considered is not a zero eigenvalue and the matrix \mathbf{B} has a rank of at least $n - 1$, then the left-hand side of Eq. (6.57) will be full rank and invertible. These assumptions are generally met in practice; therefore, the solution to the eigenvector derivative is given by

$$\frac{\partial \mathbf{v}}{\partial \mathbf{x}} = (\Psi \mathbf{A} - \lambda \mathbf{B})^{-1} \left[\lambda (\mathbf{v}^T \otimes \mathbf{I}) \frac{\partial \mathbf{B}_{\text{vec}}}{\partial \mathbf{x}} - (\mathbf{v}^T \otimes \Psi) \frac{\partial \mathbf{A}_{\text{vec}}}{\partial \mathbf{x}} \right]. \quad (6.58)$$

With the eigenvector derivative in hand, it is easy to find the eigenvalue derivative. Substituting Eq. (6.58) into Eq. (6.46) and simplifying leaves us with

$$\begin{aligned} \frac{\partial \lambda}{\partial \mathbf{x}} = & \left(\frac{\lambda \mathbf{v}^T}{\alpha} \otimes \mathbf{v}_0^H \left[\mathbf{E} (\Psi \mathbf{A} - \lambda \mathbf{B})^{-1} - \mathbf{I} \right] \right) \frac{\partial \mathbf{B}_{\text{vec}}}{\partial \mathbf{x}} \\ & + \left(\frac{\mathbf{v}^T}{\alpha} \otimes \mathbf{v}_0^H \left[\mathbf{I} - \mathbf{E} (\Psi \mathbf{A} - \lambda \mathbf{B})^{-1} \Psi \right] \right) \frac{\partial \mathbf{A}_{\text{vec}}}{\partial \mathbf{x}} \end{aligned} \quad (6.59)$$

for the eigenvalue Jacobian.

6.5.3 Numerical Verification

Forward finite differencing was used to validate the formulation of the eigenvalue and eigenvector Jacobians for the generalized eigenvalue problem. This provides a numerical approximation of the Jacobians, which may be compared with the analytic expressions developed in this chapter.

The forward finite differencing was performed by perturbing each element of each of the parent matrices individually in order to calculate each element of the Jacobians. The analytic derivatives from Eqs. (6.53), (6.52), (6.59), and (6.58) were then compared with the finite differences, and the percent differences were calculated. This was performed for 5,000 randomly generated complex matrices of size 2×2 , 5,000 randomly generated complex matrices of size 3×3 , and 5,000 randomly generated complex matrices of size 10×10 for each normalization. The results for both the eigenvalue and eigenvector derivatives are shown in the histograms in Figs. 6.3 and 6.4. Note that due to finite precision issues, matrices had to be ignored where the smallest component of the eigenvector derivatives was less than the perturbation size used in the finite differencing. As can be seen in the figure, the methods performed well in every instance for each normalization, well below 0.1% difference for each and every element of the eigenvalue and eigenvector Jacobians.

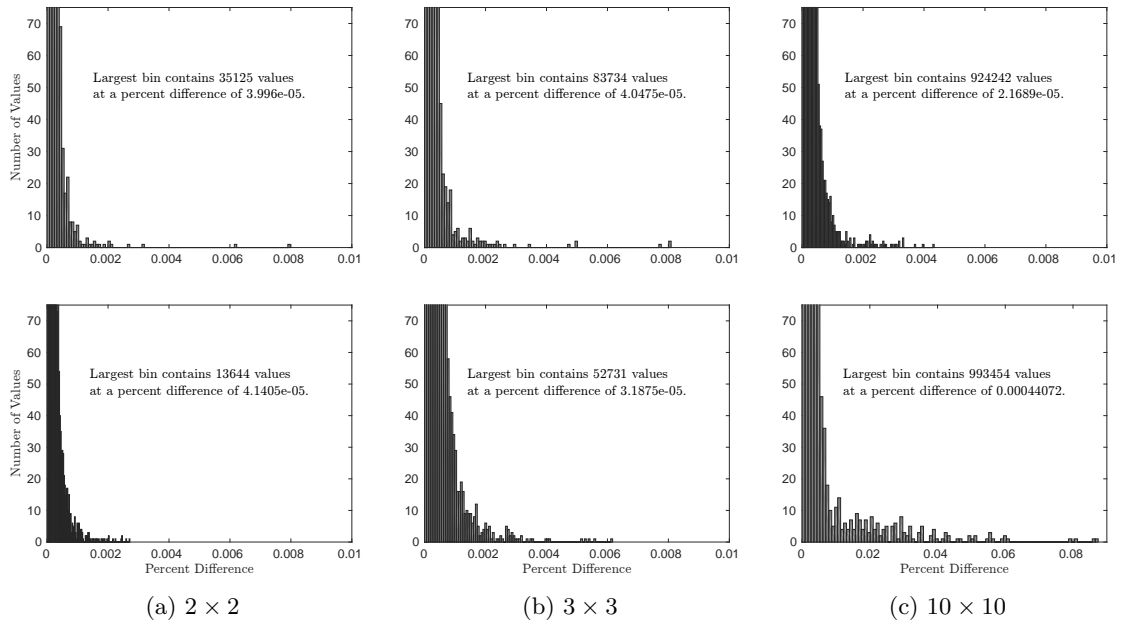


Figure 6.3: Histograms of percent difference between analytic derivatives computed using Eqs. (6.53) and (6.52) (eigenvalue derivatives top and eigenvector derivatives bottom) and finite forward differencing for 5,000 randomly generated matrices of each size with a normalization of $\mathbf{v}_0^H \mathbf{v} = 1$. The histograms are of the percent difference for each element of the eigenvalue and eigenvector Jacobians (for example, for each $n \times n$ matrix, there are $2n^2$ eigenvalue derivative elements and $2n^2$ eigenvector derivative elements).

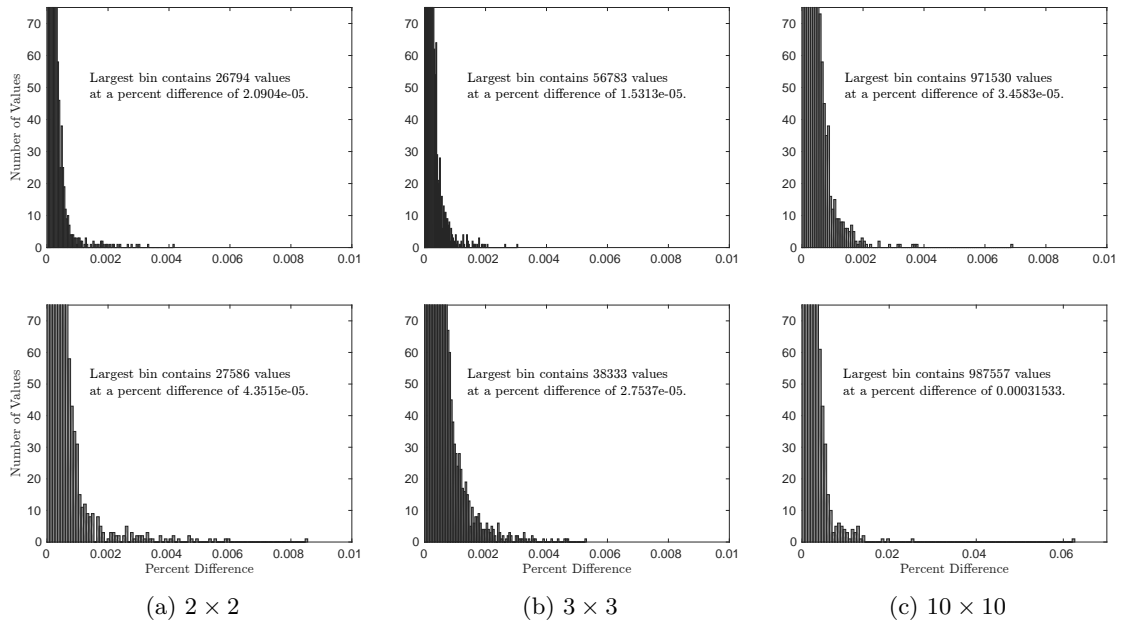


Figure 6.4: Histograms of percent difference between analytic derivatives computed using Eqs. (6.59) and (6.58) (eigenvalue derivatives top and eigenvector derivatives bottom) and finite forward differencing for 5,000 randomly generated matrices of each size with a normalization of $\mathbf{v}_0^H \mathbf{B} \mathbf{v} = 1$. The histograms are of the percent difference for each element of the eigenvalue and eigenvector derivatives (for example, for each $n \times n$ matrix, there are $2n^2$ eigenvalue derivative elements and $2n^2$ eigenvector derivative elements).

Chapter 7

Example Applications

Having developed the theoretical requirements for the Jacobians and covariance of the eigenvalues and eigenvectors, we now turn our attention to the application of these techniques to a variety of problems.

7.1 A Simple 2×2 Case

As with the theory development, we start with a simple case. Begin by supposing we have the following 2×2 matrix:

$$\mathbf{A} = \begin{bmatrix} 3 & 4 \\ -2 & -7 \end{bmatrix}. \quad (7.1)$$

Now let us assume that the vectorized version of this matrix has the following uncorrelated covariance matrix

$$\mathbf{P}_{\mathbf{A}_{\text{vec}}\mathbf{A}_{\text{vec}}} = \text{diag} \left[0.05 \quad 0.07 \quad 0.1 \quad 0.004 \right]. \quad (7.2)$$

This system has the following eigenvalues and eigenvectors:

$$\begin{aligned} \lambda_1 &= 2.123 & \mathbf{v}_1 &= \begin{bmatrix} 0.9768 \\ -0.2141 \end{bmatrix} \\ \lambda_2 &= -6.123 & \mathbf{v}_2 &= \begin{bmatrix} -0.4016 \\ 0.91584 \end{bmatrix}. \end{aligned}$$

We can apply the techniques presented in Section 5.1 or Section 6.1 directly for each eigenpair. This results in the following standard deviation value for the eigenvalue and covariance matrix for the eigenvector:

$$\begin{aligned} \sigma_{\lambda_1} &= 0.2891 & \mathbf{P}_{\mathbf{v}_1\mathbf{v}_1} &= \begin{bmatrix} 4.471 \times 10^{-4} & 2.039 \times 10^{-4} \\ 2.039 \times 10^{-4} & 9.303 \times 10^{-4} \end{bmatrix} \\ \sigma_{\lambda_2} &= 0.1668 & \mathbf{P}_{\mathbf{v}_2\mathbf{v}_2} &= \begin{bmatrix} 9.803 \times 10^{-4} & 4.298 \times 10^{-4} \\ 4.298 \times 10^{-4} & 1.885 \times 10^{-4} \end{bmatrix}. \end{aligned}$$

Now, to demonstrate that this technique correctly maps the covariance of the matrix into the variance and covariance of the eigenvalues and eigenvectors, we ran a Monte Carlo analysis of 1,000,000 runs by adding noise to the terms of the matrix according to the information encoded in the covariance matrix of the vector form of \mathbf{A} . For each iteration, we recorded the eigenvalues and eigenvectors and then used this data to generate a sample standard deviation for the eigenvalues and a sample covariance for the eigenvectors. Comparing these numerical results with the analytic resulted in the following percent differences (PD):

$$\begin{aligned} PD(\sigma_{\lambda_1}) &= 0.37\% & PD(\mathbf{P}_{\mathbf{v}_1\mathbf{v}_1}) &= 0.79\% \\ PD(\sigma_{\lambda_2}) &= 1.00\% & PD(\mathbf{P}_{\mathbf{v}_2\mathbf{v}_2}) &= 0.55\%. \end{aligned}$$

We make a note here that the percent differences for the covariance matrices indicate the percent difference for each term of the covariance matrix, since the covariance matrix is actually only of rank 1.

7.2 Ellipse Fitting

We now turn our attention to the application of these techniques to an actual estimation problem involving the use of eigenvectors in the form of ellipse fitting. Ellipse fitting to noisy data points, as demonstrated in Fig. 7.1, is a well studied problem [27, 29, 98–104] that has numerous applications, especially in computer vision and spacecraft optical navigation [24, 25, 105–109]. One technique that has received much attention for its speed and ease of implementation is the numerically stable Fitzgibbon’s method (NSFM) [27].

In the NSFM, noisy 2 dimensional data (in the form of $\{x, y\}$ pairs) are used to estimate ellipse

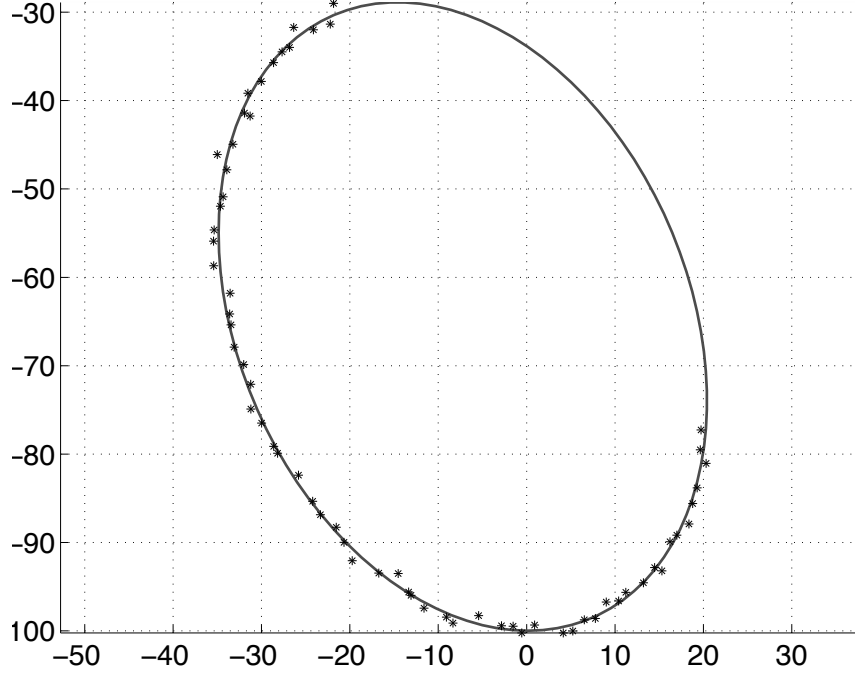


Figure 7.1: An example ellipse fit to noisy data. The noisy data points are shown as asterisks, and the fit ellipse is shown as a gray line.

parameters $\{ A \ B \ C \ D \ E \ F \}$ that form the coefficients in the implicit equation for a conic section

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0 \quad (7.3)$$

where the result is an ellipse if $4AC > B^2$. NSFMM computes the solution through the use of a 3×3 eigenvalue and eigenvector problem. We now briefly outline the steps in this technique. The reader is directed to [27] for the complete derivation.

The first step is to form the matrices \mathbf{D}_1 and \mathbf{D}_2 using the set of n noisy $\{x, y\}$ pairs as follows:

$$\mathbf{D}_1 = \begin{bmatrix} x_1^2 & x_1y_1 & y_1^2 \\ \vdots & \vdots & \vdots \\ x_i^2 & x_iy_i & y_i^2 \\ \vdots & \vdots & \vdots \\ x_n^2 & x_ny_n & y_n^2 \end{bmatrix} \quad \mathbf{D}_2 = \begin{bmatrix} x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots \\ x_i & y_i & 1 \\ \vdots & \vdots & \vdots \\ x_n & y_n & 1 \end{bmatrix} \quad (7.4)$$

where subscript i represents the i^{th} $\{x, y\}$ pair. Having formed the design matrices (\mathbf{D}_1 and \mathbf{D}_2),

we can calculate the scatter matrices as

$$\mathbf{S}_1 = \mathbf{D}_1^T \mathbf{D}_1 \quad \mathbf{S}_2 = \mathbf{D}_1^T \mathbf{D}_2 \quad \mathbf{S}_3 = \mathbf{D}_2^T \mathbf{D}_2 \quad (7.5)$$

and the constraint matrix as

$$\mathbf{C} = \begin{bmatrix} 0 & 0 & 2 \\ 0 & -1 & 0 \\ 2 & 0 & 0 \end{bmatrix}. \quad (7.6)$$

Finally, we can define the reduced scalar matrix as

$$\mathbf{M} = \mathbf{C}^{-1}(\mathbf{S}_1 - \mathbf{S}_2 \mathbf{S}_3^{-1} \mathbf{S}_2^T), \quad (7.7)$$

which is a 3×3 matrix.

The ellipse parameters can be found by solving the eigenvector–eigenvalue problem given by

$$\mathbf{M} \mathbf{v}_1 = \lambda \mathbf{v}_1 \quad (7.8)$$

where $\mathbf{v}_1 = [A \ B \ C]^T$. It can be shown that there is only one eigenvector for which $4AC - B^2 > 0$; this yields the optimal solution for the parameters A , B , and C . From here, the remaining parameters can be computed as

$$\mathbf{v}_2 = -\mathbf{S}_3^{-1} \mathbf{S}_2^T \mathbf{v}_1 \quad (7.9)$$

with $\mathbf{v}_2 = [D \ E \ F]^T$.

We now turn our attention to solving for the covariance of the best fit ellipse as a function of the covariance of the data points. Begin by assuming we have knowledge of the accuracy of the $\{x, y\}$ pairs in the form of the standard deviation of these coordinates, σ_x and σ_y respectively. Assuming noise in the x -direction is uncorrelated with noise in the y -direction, we can define a covariance matrix for each and every data point as

$$\mathbf{P}_{\mathbf{pp}} = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix} \quad (7.10)$$

where $\mathbf{p} = [x \ y]^T$ is a data point expressed as a column vector. To use the techniques outlined in this paper, we must find the values of the covariance matrix of \mathbf{M}_{vec} . The covariance matrix of

\mathbf{M}_{vec} is given by

$$\mathbf{P}_{\mathbf{M}_{\text{vec}}\mathbf{M}_{\text{vec}}} = \sum_{i=1}^n \frac{\partial \mathbf{M}_{\text{vec}}}{\partial \mathbf{p}_i} \mathbf{P}_{\text{pp}} \frac{\partial \mathbf{M}_{\text{vec}}}{\partial \mathbf{p}_i}^T \quad (7.11)$$

where \mathbf{p}_i is the i^{th} data point and

$$\frac{\partial \mathbf{M}_{\text{vec}}}{\partial \mathbf{p}_i} = \begin{bmatrix} \frac{\partial \mathbf{M}_{\text{vec}}}{\partial x_i} & \frac{\partial \mathbf{M}_{\text{vec}}}{\partial y_i} \end{bmatrix}. \quad (7.12)$$

Proceeding by direct computation, the derivations of $\partial \mathbf{M}_{\text{vec}} / \partial x_i$ and $\partial \mathbf{M}_{\text{vec}} / \partial y_i$ are straightforward. Note that both of these are column vectors. We omit the corresponding details for the sake of brevity.

Having calculated the covariance matrix for \mathbf{M}_{vec} , we can now directly compute \mathbf{M} and its covariance matrix using the results presented in Section 5.2 or 6.1, thus giving the expression for the covariance of the first three parameters of the optimal ellipse fit. Further, once the covariance of the first three parameters is obtained, it is easy to show that the covariance of the second three parameters is

$$\mathbf{P}_{\mathbf{v}_2\mathbf{v}_2} = \sum_{i=1}^n \frac{\partial \mathbf{v}_2}{\partial \mathbf{p}_i} \mathbf{P}_{\text{pp}} \frac{\partial \mathbf{v}_2}{\partial \mathbf{p}_i}^T \quad (7.13)$$

and the cross-covariance between \mathbf{v}_1 and \mathbf{v}_2 is given by

$$\mathbf{P}_{\mathbf{v}_1\mathbf{v}_2} = \sum_{i=1}^n \frac{\partial \mathbf{v}_1}{\partial \mathbf{p}_i} \mathbf{P}_{\text{pp}} \frac{\partial \mathbf{v}_2}{\partial \mathbf{p}_i}^T \quad (7.14)$$

where

$$\frac{\partial \mathbf{v}_1}{\partial \mathbf{p}_i} = \begin{bmatrix} \frac{\partial \mathbf{v}_1}{\partial x_i} & \frac{\partial \mathbf{v}_1}{\partial y_i} \end{bmatrix} \quad \frac{\partial \mathbf{v}_2}{\partial \mathbf{p}_i} = \begin{bmatrix} \frac{\partial \mathbf{v}_2}{\partial x_i} & \frac{\partial \mathbf{v}_2}{\partial y_i} \end{bmatrix}$$

are left to the reader to derive. This completes the 6×6 covariance matrix for all of the optimal ellipse parameters.

We implemented the above steps and calculated the analytic covariance for an ellipse with $\sigma_x = 0.01$ and $\sigma_y = 0.01$, and x and y were about 4 orders of magnitude larger. We then ran a Monte Carlo analysis of 20,000 runs, adding noise to the x and y data, and generated the plot shown in Fig. 7.2. This plot shows the results of the Monte Carlo analysis, as well as the analytic and numeric covariances for \mathbf{v}_1 . As can be seen, the numerical and analytic covariances match very well. In addition, this plot demonstrates how the covariance matrix is rank deficient, and thus the covariance is represented as a disk rather than an ellipsoid.

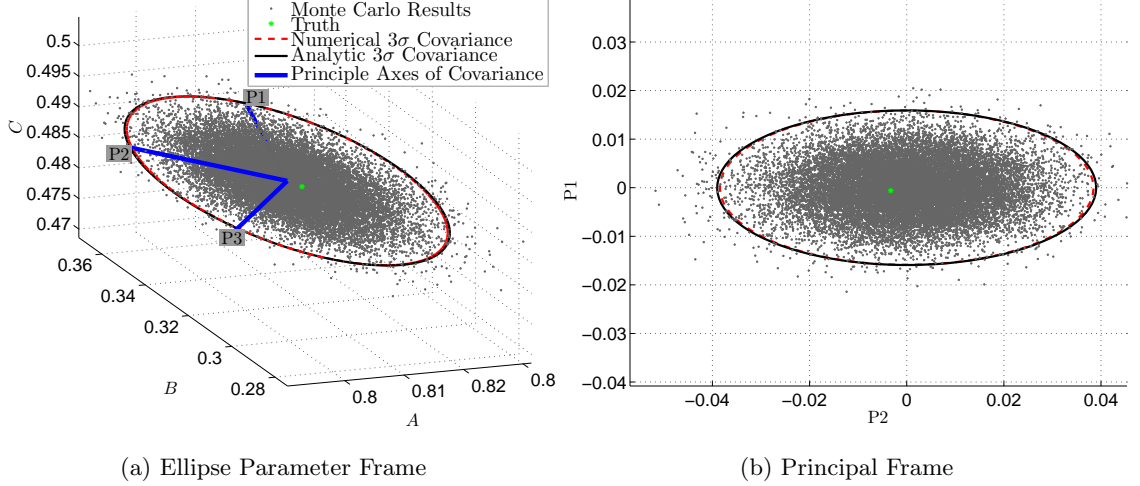


Figure 7.2: Covariance results for the first three parameters of an ellipse fit. The figure to the left shows the covariance in the ellipse parameter frame $\{A, B, C\}$, while the figure to the right shows the covariance and Monte Carlo results rotated in the principal axis frame of the covariance matrix. The legend is for both figures.

7.3 Attitude Estimation

Attitude estimation represents another important application of the eigenvector problem in the field of data fusion. Suppose that we collect a set of m noisy unit vector observations in a sensor frame, $\tilde{\mathbf{b}}_i$, along with their corresponding unit vectors in a reference frame, \mathbf{a}_i . In the case of attitude estimation, these are 3 dimensional unit vectors such that $\{\tilde{\mathbf{b}}_i, \mathbf{a}_i\} \in \mathbb{R}^3$. Finding the optimal rotation between the reference frame and the sensor frame, \mathbf{T} , is described by the well-known Wahba Problem [110]:

$$\text{Min } J(\mathbf{T}) = \frac{1}{2} \sum_{i=1}^m w_i \|\tilde{\mathbf{b}}_i - \mathbf{T}\mathbf{a}_i\|. \quad (7.15)$$

In [35], it was shown that the optimal solution to Eq. (7.15) is given by

$$\mathbf{K}\bar{\mathbf{q}} = \lambda\bar{\mathbf{q}} \quad (7.16)$$

where the attitude quaternion, $\bar{\mathbf{q}}$, is found by choosing the unit eigenvector associated with the largest eigenvalue. Although first published by Keat, this solution was developed by Paul Davenport and has since become the foundation for many popular numerical solutions for efficiently solving for only the eigenvector of interest [111, 112]. These are widely used in practice for spacecraft attitude estimation. The values in the symmetric 4×4 matrix \mathbf{K} are found using the pairs of m corresponding

unit vectors

$$\mathbf{K} = \begin{bmatrix} \mathbf{S} - \mu \mathbf{I}_{3 \times 3} & \mathbf{z} \\ \mathbf{z}^T & \mu \end{bmatrix} \quad (7.17)$$

$$\mathbf{B} = \sum_{i=1}^m \frac{1}{\sigma_i^2} \tilde{\mathbf{b}}_i \mathbf{a}_i^T \quad \mathbf{z} = \sum_{i=1}^m \frac{1}{\sigma_i^2} (\tilde{\mathbf{b}}_i \times \mathbf{a}_i) \quad (7.18)$$

$$\mu = \text{trace}[\mathbf{B}] \quad \mathbf{S} = \mathbf{B} + \mathbf{B}^T. \quad (7.19)$$

This form of the solution assumes that the quaternion has the vector part first and scalar part second, $\bar{\mathbf{q}}^T = [\mathbf{q}^T \ q_s]$. Thus, the attitude estimation problem has been reduced to finding the solution to a symmetric 4×4 eigenvector problem.

The application of the theory introduced in this paper now becomes clear. The noise in the measurements introduce perturbations in the matrix \mathbf{K} . The result is error in the estimated attitude quaternion. Because the attitude quaternion is an eigenvector of \mathbf{K} , the covariance of this eigenvector may be estimated with the techniques introduced in this paper.

There are two important notes to make before proceeding. First, both the “true” and “noisy” versions of \mathbf{K} are symmetric. Thus, the matrix \mathbf{K} only really has 10 unique values,

$$\mathbf{K} = \begin{bmatrix} k_{11} & k_{12} & k_{13} & k_{14} \\ k_{12} & k_{22} & k_{23} & k_{24} \\ k_{13} & k_{23} & k_{33} & k_{34} \\ k_{14} & k_{24} & k_{34} & k_{44} \end{bmatrix}. \quad (7.20)$$

If we define the unique entries of \mathbf{K} as the 10×1 vector \mathbf{k}

$$\mathbf{k}^T = [k_{11} \ k_{12} \ k_{13} \ k_{14} \ k_{22} \ k_{23} \ k_{24} \ k_{33} \ k_{34} \ k_{44}], \quad (7.21)$$

then we can write

$$\mathbf{K}_{\text{vec}} = \mathbf{G}\mathbf{k} \quad (7.22)$$

and eventually

$$\mathbf{P}_{\mathbf{K}_{\text{vec}} \mathbf{K}_{\text{vec}}} = \mathbf{G} \mathbf{P}_{\mathbf{k} \mathbf{k}} \mathbf{G}^T \quad (7.23)$$

where $\mathbf{P}_{\mathbf{k} \mathbf{k}}$ is the covariance of \mathbf{k} . Finding \mathbf{G} for this case is straightforward and is left to the reader.

This provides the necessary information to implement the general $n \times n$ solution for the specific 4×4

case.

The second note worth making has to do with the way attitude errors are expressed. The formulation developed in this paper assumes that the eigenvector errors (and, hence, the attitude quaternion errors in this example) are additive in nature. That is, $\bar{\mathbf{q}} = \hat{\mathbf{q}} + \Delta\bar{\mathbf{q}}$. We know, however, that attitude errors are actually multiplicative,

$$\bar{\mathbf{q}} = \delta\bar{\mathbf{q}} \odot \hat{\mathbf{q}} \quad (7.24)$$

where \odot is the quaternion multiplication operator. We use the non-Hamiltonian form [36], such that the above equation defines $\delta\bar{\mathbf{q}}$ to be the attitude error that transforms the estimated attitude quaternion, $\hat{\mathbf{q}}$, into the true attitude quaternion, $\bar{\mathbf{q}}$. Assuming that the attitude error is small, the two forms (additive error and multiplicative error) may be equated, and to first order we see that

$$\Delta\bar{\mathbf{q}} = \begin{bmatrix} \Delta\mathbf{q} \\ \Delta q_s \end{bmatrix} = \mathbf{M} \delta\mathbf{q} \quad (7.25)$$

where the 4×3 matrix \mathbf{M} is defined as

$$\mathbf{M} = \begin{bmatrix} \hat{q}_s \mathbf{I}_{3 \times 3} + [\hat{\mathbf{q}} \times] \\ -\hat{\mathbf{q}}^T \end{bmatrix}. \quad (7.26)$$

We now turn our attention to the application of the techniques outlined in this paper to this problem. The first step is to calculate the covariance of \mathbf{k} as a function of the measurement covariance. It can be shown that the measurement covariance for a unit vector observation is approximated by [26]

$$\mathbf{R}_i = \sigma_i^2 [\tilde{\mathbf{b}}_i \times] [\tilde{\mathbf{b}}_i \times]^T = \sigma_i^2 (\mathbf{I}_{3 \times 3} - \tilde{\mathbf{b}}_i \tilde{\mathbf{b}}_i^T) \quad (7.27)$$

where \mathbf{R}_i is the covariance matrix for the i^{th} measurement and σ_i is the standard deviation of the rotation error on the i^{th} measurement. With the measurement covariance in hand, the covariance of \mathbf{k} can be calculated as

$$\mathbf{P}_{\mathbf{k}\mathbf{k}} = \sum_{i=1}^m \frac{\partial \mathbf{k}}{\partial \tilde{\mathbf{b}}_i} \mathbf{R}_i \frac{\partial \mathbf{k}}{\partial \tilde{\mathbf{b}}_i}^T \quad (7.28)$$

where $\mathbf{P}_{\mathbf{k}\mathbf{k}}$ is the covariance matrix of vector \mathbf{k} and $\partial \mathbf{k} / \partial \tilde{\mathbf{b}}_i$ is the Jacobian of \mathbf{k} with respect to the i^{th} measurement vector.

After computing the covariance matrix of \mathbf{k} , the covariance matrix of \mathbf{K}_{vec} can be computed using

Eq. (7.23). We now have all of the information needed to use the techniques outlined in this paper. Running the matrix \mathbf{K} , the largest eigenvalue and corresponding eigenvector, and the covariance matrix for \mathbf{K}_{vec} through the techniques for the $n \times n$ case will directly yield the covariance of the attitude quaternion estimate. We once again note that this covariance matrix is for additive errors in the quaternion. That is, application of the equations from Sections 5.3 or 6.1 yields $\mathbf{P}_{\Delta\bar{\mathbf{q}}\Delta\bar{\mathbf{q}}}$.

In order to validate this method, we performed a Monte Carlo analysis of 10,000 runs varying the multiplicative noise on the unit vector measurements. For the Monte Carlo analysis, we did not enforce the normalization conditions discussed in this paper, since quaternions must be unit length.

In addition, we compared the analytic results of our technique to another analytic expression for the attitude covariance, which is derived from the Fisher information matrix (FIM). This formulation is given in [26] (as later modified to this form in [113]):

$$\mathbf{F}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \text{Tr} \left[\mathbf{T}\mathbf{B}^T \right] \mathbf{I}_{3 \times 3} - \frac{1}{2}(\mathbf{T}\mathbf{B}^T + \mathbf{B}\mathbf{T}^T) \quad (7.29)$$

$$\mathbf{P}_{\delta\bar{\mathbf{q}}\delta\bar{\mathbf{q}}} = \begin{bmatrix} \frac{1}{4}\mathbf{F}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1} & \mathbf{0}_{3 \times 1} \\ \mathbf{0}_{1 \times 3} & 0 \end{bmatrix} \quad (7.30)$$

where $\mathbf{F}_{\boldsymbol{\theta}\boldsymbol{\theta}}$ is the FIM. This representation of the covariance is for the multiplicative errors. The additive error covariance can then be calculated using

$$\mathbf{P}_{\Delta\bar{\mathbf{q}}\Delta\bar{\mathbf{q}}} = \mathbf{M}\mathbf{F}^{-1}\mathbf{M}^T. \quad (7.31)$$

The results of the analytic covariance as discussed in this document, the analytic covariance using the FIM, and the numerical covariance of the Monte Carlo results (all for the additive noise) are shown in Fig. 7.3. As can be seen, all of the results match very well. In addition, we note that the two theoretical covariances (our new eigenvector covariance and the FIM covariance from [26]) are equivalent to machine precision.

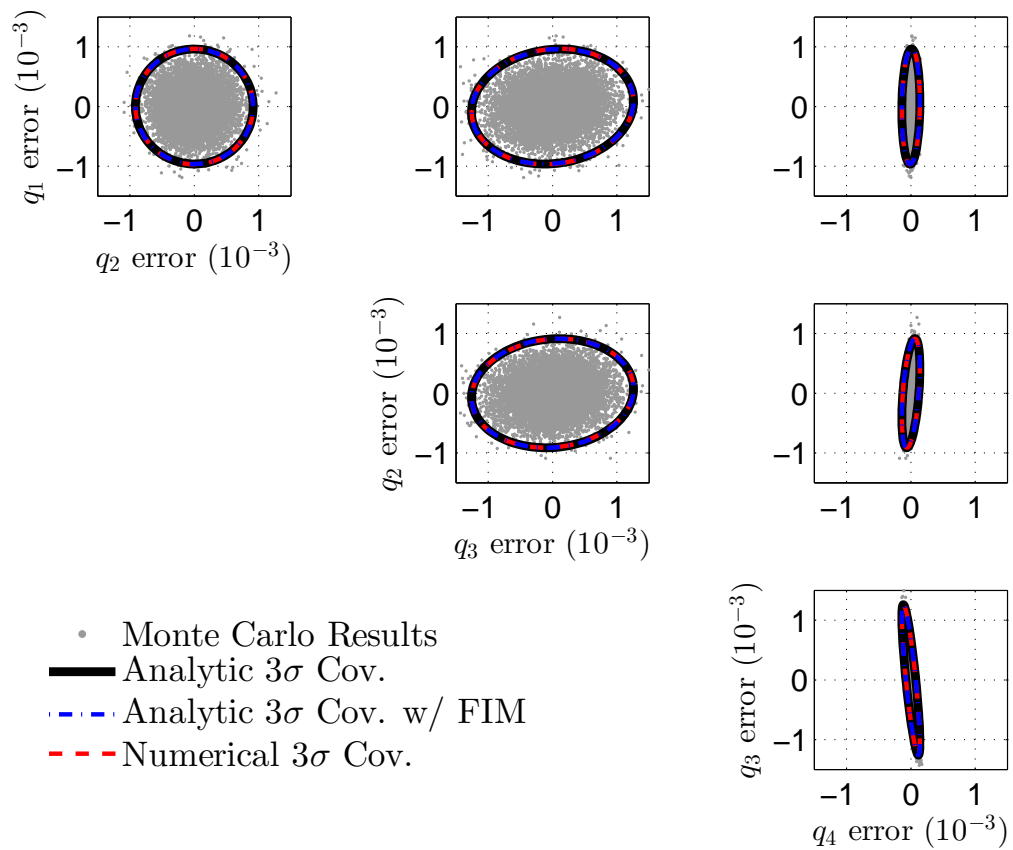


Figure 7.3: 2D covariance projections for the Davenport q-method solution to the Wahba problem.

7.4 Horizon-based OPNAV

7.4.1 A Review of Horizon-based OPNAV

There are a variety of optical navigation (OPNAV) problems that make use of a celestial body's horizon. One of the most popular techniques in horizon-based OPNAV is the central and apparent diameter (CAD) method. While a variety of solutions for CAD OPNAV exist, the present work will consider the method recently put forward by Christian [24].¹ A brief review of this technique is presented here, primarily as a way of introducing key notation. Readers interested in how this technique works (or CAD OPNAV in general) are directed to References [24] and [114] for a thorough discussion.

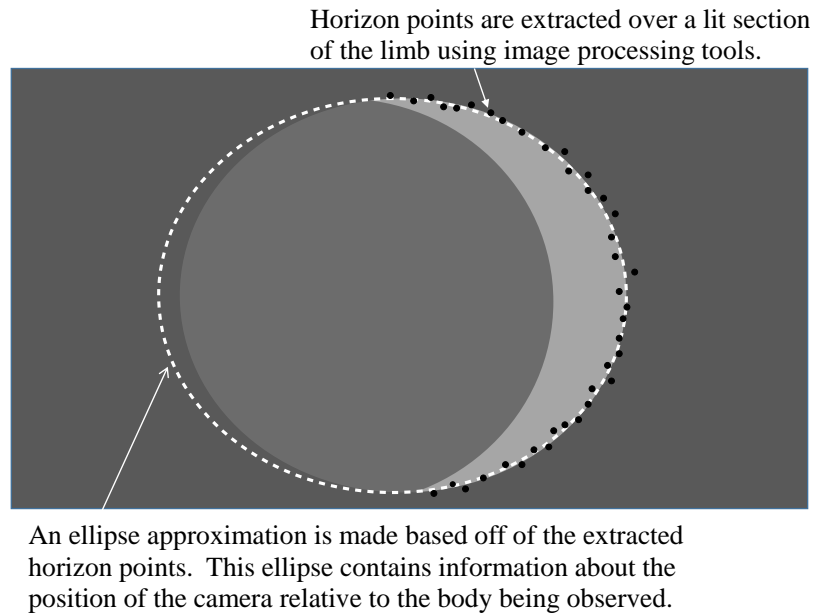


Figure 7.4: Lit horizons of celestial bodies are used to estimate the relative position of the spacecraft in CAD OPNAV. Diagram modeled after Reference [24].

To begin, suppose there is an ellipse in the camera image plane that has been fit to points on the planet's lit horizon (see Figure 7.4). This ellipse fit is described by the six coefficients of the implicit equation for a conic section given as

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + G = 0, \quad (7.32)$$

¹This earlier technique has since been superseded by newer methods, such as Iterative Horizon Reprojection [114] and a non-iterative technique using Cholesky factorization [115]. Despite this, the original CAD method is an excellent case study for eigenvector covariances from the generalized problem.

where x and y are coordinates of points on the edge of the ellipse and the result is an ellipse if $4AC > B^2$. Further, modeling the planet as a tri-axial ellipsoid and describing the shape through the 3×3 matrix \mathbf{A} , such that $\mathbf{p}_i^T \mathbf{A} \mathbf{p}_i = 1$ (where \mathbf{p}_i is a point on the ellipsoid's surface and everything is in the camera frame), the relationship of the camera's position with respect to the planet, \mathbf{r} , can be shown to be [24]

$$\lambda \mathbf{A} \mathbf{r} = \mathbf{C} \mathbf{r}. \quad (7.33)$$

Here, the matrices \mathbf{A} and \mathbf{C} are explicitly given as

$$\mathbf{C} = \frac{1}{2} \begin{bmatrix} 2Af^2 & Bf^2 & Df \\ Bf^2 & 2Cf^2 & Ff \\ Df & Ff & 2G \end{bmatrix} \quad \mathbf{A} = \mathbf{T}_C^P \begin{bmatrix} a^{-2} & 0 & 0 \\ 0 & b^{-2} & 0 \\ 0 & 0 & c^{-2} \end{bmatrix} \mathbf{T}_P^C,$$

where f is the camera focal length, $\{a, b, c\}$ are the parameters of the tri-axial ellipsoid model of the observed planet, and \mathbf{T}_C^P is the rotation matrix from the planet-fixed principal axis frame to the camera frame. Thus, it is known that \mathbf{r} is in the direction of one of the eigenvectors of \mathbf{A} and \mathbf{C} . As shown in [24], the eigenvector associated with the eigenvalue of unique sign will produce the correct solution, resulting in

$$\mathbf{r}_I = -\text{sign}(v_3) \rho \mathbf{T}_I^C \mathbf{v} \quad (7.34)$$

where

$$\rho = \left[\frac{\text{Tr}[\mathbf{C}] - \lambda \text{Tr}[\mathbf{A}]}{\lambda(\mathbf{v}^T \mathbf{A} \mathbf{A} \mathbf{v}) - \lambda(\mathbf{v}^T \mathbf{A} \mathbf{v}) \text{Tr}[\mathbf{A}]} \right]^{\frac{1}{2}} \quad (7.35)$$

and λ is the uniquely signed eigenvalue from Eq. (7.33), $\mathbf{v} = [v_1 \ v_2 \ v_3]^T$ is the unit-eigenvector corresponding to λ , and \mathbf{T}_I^C is the rotation matrix from the camera frame to the inertial frame.

7.4.2 Covariance of the Ellipse Fit

In order to develop the covariance expressions for the position measurement, we need to understand the covariance in the ellipse fit. Beginning with $\left\{ A \ B \ C \ D \ F \ G \right\}$ as ellipse parameters—and defining $\tilde{\mathbf{y}}_i = \begin{bmatrix} \tilde{x}_i & \tilde{y}_i \end{bmatrix}^T$ as a noisy data point that is thought to lie along the lit horizon and was extracted from an image—the algebraic distance can be defined as the residual from Eq. (7.32):

$$\tilde{d}_i = d(\tilde{\mathbf{y}}_i) = A\tilde{x}_i^2 + B\tilde{x}_i\tilde{y}_i + C\tilde{y}_i^2 + D\tilde{x}_i + F\tilde{y}_i + G. \quad (7.36)$$

Clearly, when $\tilde{\mathbf{y}}_i$ is a point directly on the ellipse, the algebraic distance will be 0, and when $\tilde{\mathbf{y}}_i$ is close to the ellipse, the algebraic distance should be close to 0.

From here, assuming that the ellipse parameters are normalized such that $A = 1$,² a new vector can be defined as

$$\boldsymbol{\alpha} = \begin{bmatrix} B & C & D & F & G \end{bmatrix}^T. \quad (7.37)$$

As shown in Reference [24], after making some assumptions about the body being observed and the extracted points being uncorrelated, the covariance of $\boldsymbol{\alpha}$ is

$$\mathbf{P}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} \approx \left(\hat{\mathbf{N}}_i \mathbf{R}_{\mathbf{y}_i \mathbf{y}_i} \hat{\mathbf{N}}_i^T \right) \left(\mathbf{H}^T \mathbf{H} \right)^{-1} \quad (7.38)$$

where

$$\hat{\mathbf{N}}_i = \frac{\partial \tilde{d}_i}{\partial \tilde{\mathbf{y}}_i} = \begin{bmatrix} 2A\tilde{x}_i + B\tilde{y}_i + D & B\tilde{x}_i + 2C\tilde{y}_i + F \end{bmatrix}$$

$$\mathbf{H} = \begin{bmatrix} \tilde{x}_1 \tilde{y}_1 & \tilde{y}_1^2 & \tilde{x}_1 & \tilde{y}_1 & 1 \\ \tilde{x}_2 \tilde{y}_2 & \tilde{y}_2^2 & \tilde{x}_2 & \tilde{y}_2 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \tilde{x}_n \tilde{y}_n & \tilde{y}_n^2 & \tilde{x}_n & \tilde{y}_n & 1 \end{bmatrix}$$

$$\mathbf{R}_{\mathbf{y}_i \mathbf{y}_i} \approx \sigma_{FPA}^2 \mathbf{I}_{2 \times 2}.$$

Here, σ_{FPA} is the accuracy of the horizon point extraction, thereby providing the required relationship between the accuracy of the edge detection method and the resulting covariance matrix of the best-fit ellipse parameters. It can be shown that this method results in approximately the same covariance matrix regardless of what point is chosen for nearly spherical planets.

7.4.3 Covariance of the CAD OPNAV Measurement

With the covariance of the ellipse fit in hand, the covariance of the CAD OPNAV measurement can be found by expanding Eq. (7.34) to first order using a Taylor series expansion about our best estimates. We need to do this with regard to some parameter vector \mathbf{x} , which we will define to include the horizon, $\boldsymbol{\alpha}$ (see Eq. (7.37)), the ellipsoidal model of the planet, \mathbf{b} , and the attitude error,

²In [24] and [25], a normalization of $G = 1$ was used. That normalization introduces a well known singularity when the edge of the ellipse passes through the origin of the plane in which it is defined [100]. Since any planetary body will project to an ellipse that is nearly a circle, the parameter A will never be near 0, and thus singularities are avoided by using this non-traditional normalization.

$\delta\theta$:

$$\mathbf{x} = \begin{bmatrix} \boldsymbol{\alpha}^T & \mathbf{b}^T & \delta\theta^T \end{bmatrix} \quad (7.39)$$

where $\mathbf{b} = \begin{bmatrix} a & b & c \end{bmatrix}^T$. This results in

$$\mathbf{r}_I \approx \hat{\mathbf{r}}_I + \left. \frac{\partial \mathbf{r}_I}{\partial \mathbf{x}} \right|_{\hat{\mathbf{x}}} (\mathbf{x} - \hat{\mathbf{x}}). \quad (7.40)$$

Defining $\Delta \mathbf{x} = (\mathbf{x} - \hat{\mathbf{x}})$, Eq. (7.40) may be rearranged to find

$$\Delta \mathbf{r}_I \approx \left. \frac{\partial \mathbf{r}_I}{\partial \mathbf{x}} \right|_{\hat{\mathbf{x}}} \Delta \mathbf{x}. \quad (7.41)$$

Now, letting $\mathbf{H}_{\mathbf{x}} = \left. \frac{\partial \mathbf{r}_I}{\partial \mathbf{x}} \right|_{\hat{\mathbf{x}}}$, we can directly find the covariance of the CAD OPNAV measurement as

$$\mathbf{P}_{\mathbf{r}_I \mathbf{r}_I} = \mathbb{E} [\Delta \mathbf{r}_I \Delta \mathbf{r}_I^T]. \quad (7.42)$$

Distributing the expected value operator, substituting in Eq. 7.41, and assuming the errors in the ellipse parameters are uncorrelated with errors in the model ellipsoid parameters (as well as the attitude error) leads to

$$\begin{aligned} \mathbf{P}_{\mathbf{r}_I \mathbf{r}_I} &= \mathbf{H}_{\mathbf{x}} \mathbb{E} [\Delta \mathbf{x} \Delta \mathbf{x}^T] \mathbf{H}_{\mathbf{x}}^T \\ &= \mathbf{H}_{\mathbf{x}} \mathbf{P}_{\mathbf{xx}} \mathbf{H}_{\mathbf{x}}^T, \end{aligned} \quad (7.43)$$

where

$$\mathbf{P}_{\mathbf{xx}} = \begin{bmatrix} \mathbf{P}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} & \mathbf{0}_{5 \times 3} & \mathbf{0}_{5 \times 3} \\ \mathbf{0}_{3 \times 5} & \mathbf{P}_{\mathbf{bb}} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 5} & \mathbf{0}_{3 \times 3} & \mathbf{P}_{\delta\theta\delta\theta} \end{bmatrix} \quad (7.44)$$

is the covariance matrix for the parameter vector.

Recall that $\mathbf{P}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}$ is known from Eq. (7.38), $\mathbf{P}_{\mathbf{bb}}$ is some assumed uncertainty in the planet shape model, and $\mathbf{P}_{\delta\theta\delta\theta}$ should be known from the attitude filter; therefore, all that is needed for Eq. (7.43) is to calculate the expression for $\mathbf{H}_{\mathbf{x}}$. This follows.

7.4.4 Jacobian Matrix of the Position Estimate

The matrix $\mathbf{H}_{\mathbf{x}}$ is simply the Jacobian matrix of the instantaneous CAD OPNAV position estimate with respect to \mathbf{x} (evaluated at the best estimate). Using matrix-vector calculus and the chain rule leads to

$$\mathbf{H}_{\mathbf{x}} = \frac{\partial \mathbf{r}}{\partial \mathbf{x}} = -\text{sign}(v_3) \left[\mathbf{T}_I^C \left(\mathbf{v} \frac{\partial \rho}{\partial \mathbf{x}} + \rho \frac{\partial \mathbf{v}}{\partial \mathbf{x}} \right) + (\rho \mathbf{v}^T \otimes \mathbf{I}_{3 \times 3}) \frac{\partial}{\partial \mathbf{x}} (\mathbf{T}_I^C)_{\text{vec}} \right]. \quad (7.45)$$

Additionally, the change in the range term (the partial derivative of ρ) can also be found by using the usual rules of matrix-vector calculus and the chain rule. The derivatives are given by

$$\frac{\partial \rho}{\partial \mathbf{x}} = \frac{\mathbf{t}}{m} \quad (7.46)$$

where

$$\begin{aligned} \mathbf{t} = d \left[\text{Tr}[\mathbf{C}]_{\mathbf{x}} - \lambda \text{Tr}[\mathbf{A}]_{\mathbf{x}} - \text{Tr}[\mathbf{A}] \frac{\partial \lambda}{\partial \mathbf{x}} \right] - n \left[\mathbf{v}^T \mathbf{Z} \left(2\lambda \frac{\partial \mathbf{v}}{\partial \mathbf{x}} + \mathbf{v} \frac{\partial \lambda}{\partial \mathbf{x}} \right) \right. \\ \left. + (\lambda \mathbf{v}^T \otimes \mathbf{v}^T) (\mathbf{A} \otimes \mathbf{I}_{3 \times 3} + \mathbf{I}_{3 \times 3} \otimes \mathbf{A} - \mathbf{A}_{\text{vec}} \mathbf{I}_{\text{vec}}^T - \text{Tr}[\mathbf{A}] \mathbf{I}_{9 \times 9}) \frac{\partial \mathbf{A}_{\text{vec}}}{\partial \mathbf{x}} \right], \end{aligned} \quad (7.47)$$

$m = 2\rho d^2$, $\mathbf{Z} = (\mathbf{A}\mathbf{A} - \text{Tr}[\mathbf{A}]\mathbf{A})$, $n = \text{Tr}[\mathbf{C}] - \lambda \text{Tr}[\mathbf{A}]$, and $d = \lambda \mathbf{v}^T \mathbf{Z} \mathbf{v}$ are defined to simplify the notation.

It is now apparent that the derivatives of eigenvalues and eigenvectors with respect to \mathbf{x} are required. Since this is a generalized eigenvalue problem, eigenvalue and eigenvector Jacobians can be directly found by applying Eqs. 6.53 and 6.52, respectively. The partial derivatives of the matrices with respect to \mathbf{x} are given by

$$\frac{\partial \mathbf{C}_{\text{vec}}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial \mathbf{C}_{\text{vec}}}{\partial \boldsymbol{\alpha}} & \mathbf{0}_{9 \times 6} \end{bmatrix} \quad (7.48)$$

$$\frac{\partial \mathbf{A}_{\text{vec}}}{\partial \mathbf{x}} = \begin{bmatrix} \mathbf{0}_{9 \times 5} & \frac{\partial \mathbf{A}_{\text{vec}}}{\partial \mathbf{b}} & \frac{\partial \mathbf{A}_{\text{vec}}}{\partial \boldsymbol{\theta}} \end{bmatrix}. \quad (7.49)$$

All that remains undefined are the derivatives of the trace operators. The derivatives of the traces can be shown to be

$$\text{Tr}[\mathbf{C}]_{\mathbf{x}} = \begin{bmatrix} \mathbf{I}_{\text{vec}}^T \frac{\partial \mathbf{C}_{\text{vec}}}{\partial \boldsymbol{\alpha}} & \mathbf{0}_{9 \times 6} \end{bmatrix} \quad (7.50)$$

$$\text{Tr}[\mathbf{A}]_{\mathbf{x}} = \begin{bmatrix} \mathbf{0}_{9 \times 5} & \mathbf{I}_{\text{vec}}^T \frac{\partial \mathbf{A}_{\text{vec}}}{\partial \mathbf{b}} & \mathbf{I}_{\text{vec}}^T \frac{\partial \mathbf{A}_{\text{vec}}}{\partial \boldsymbol{\theta}} \end{bmatrix} \quad (7.51)$$

by considering the vectorized form of the trace operator.

The derivative of \mathbf{C}_{vec} can easily be found by considering the structures of \mathbf{C} and $\boldsymbol{\alpha}$:

$$\frac{\partial \mathbf{C}_{\text{vec}}}{\partial \boldsymbol{\alpha}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ f^2/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & f/2 & 0 & 0 \\ f^2/2 & 0 & 0 & 0 & 0 \\ 0 & f^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & f/2 & 0 \\ 0 & 0 & f/2 & 0 & 0 \\ 0 & 0 & 0 & f/2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \quad (7.52)$$

The derivative of \mathbf{A}_{vec} with respect to \mathbf{b} is slightly more difficult to find. First, use vectorization to express

$$\mathbf{A}_{\text{vec}} = (\mathbf{T}_C^P \otimes \mathbf{T}_C^P) \mathbf{A}_{P\text{vec}} \quad (7.53)$$

where \mathbf{A}_P is the ellipsoidal model expressed in the principal frame of the planet being considered (i.e., $\mathbf{A}_P = \text{diag} \begin{bmatrix} a^{-2} & b^{-2} & c^{-2} \end{bmatrix}$ where $\text{diag}[\bullet]$ forms a square diagonal matrix when \bullet is a vector). Since $\mathbf{A}_{P\text{vec}}$ is now on the right, the partial derivative can be applied to obtain

$$\frac{\partial \mathbf{A}_{\text{vec}}}{\partial \mathbf{b}} = (\mathbf{T}_C^P \otimes \mathbf{T}_C^P) \frac{\partial \mathbf{A}_{P\text{vec}}}{\partial \mathbf{b}} \quad (7.54)$$

where the matrix derivative of \mathbf{A} in the principal frame is found by considering the structure of \mathbf{A}_P .

This results in

$$\frac{\partial \mathbf{A}_{P\text{vec}}}{\partial \mathbf{b}} = -2 \begin{bmatrix} a^{-3} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & b^{-3} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & c^{-3} \end{bmatrix}. \quad (7.55)$$

Similarly, to find the derivative of \mathbf{A}_{vec} , begin with the vectorization identities

$$\mathbf{A}_{\text{vec}} = \begin{cases} (\mathbf{T}_C^I \otimes \mathbf{T}_C^I) \mathbf{A}_{I\text{vec}} \\ (\mathbf{I} \otimes \mathbf{T}_C^I \mathbf{A}_I) (\mathbf{T}_I^C)_{\text{vec}} \\ (\mathbf{T}_C^I \mathbf{A}_I \otimes \mathbf{I}) (\mathbf{T}_C^I)_{\text{vec}} \end{cases} \quad (7.56)$$

where \mathbf{A}_I is the ellipsoidal model expressed in the inertial frame (i.e., $\mathbf{A}_I = \mathbf{T}_I^P \mathbf{A}_P \mathbf{T}_I^P$ where \mathbf{T}_I^P is the rotation matrix from the principal frame to the inertial frame). Now the application of the chain rule gives

$$\frac{\partial \mathbf{A}_{\text{vec}}}{\partial \boldsymbol{\delta \theta}} = (\mathbf{I} \otimes \mathbf{T}_C^I \mathbf{A}_I) \frac{\partial}{\partial \boldsymbol{\delta \theta}} \{(\mathbf{T}_I^C)_{\text{vec}}\} + (\mathbf{T}_C^I \mathbf{A}_I \otimes \mathbf{I}) \frac{\partial}{\partial \boldsymbol{\delta \theta}} \{(\mathbf{T}_C^I)_{\text{vec}}\} \quad (7.57)$$

since \mathbf{A}_I is not a function of $\boldsymbol{\delta \theta}$. Modeling the attitude error to first order as

$$\mathbf{T}_I^C = (\mathbf{I} - [\boldsymbol{\delta \theta} \times]) \widetilde{\mathbf{T}}_I^C, \quad (7.58)$$

then vectorization, the chain rule, and properties of a skew-symmetric matrix can be used to find

$$\frac{\partial}{\partial \boldsymbol{\delta \theta}} \{(\mathbf{T}_I^C)_{\text{vec}}\} = -(\mathbf{T}_C^I \otimes \mathbf{I}) \frac{\partial [\boldsymbol{\delta \theta} \times]_{\text{vec}}}{\partial \boldsymbol{\delta \theta}} \quad (7.59)$$

$$\frac{\partial}{\partial \boldsymbol{\delta \theta}} \{(\mathbf{T}_C^I)_{\text{vec}}\} = (\mathbf{I} \otimes \mathbf{T}_C^I) \frac{\partial [\boldsymbol{\delta \theta} \times]_{\text{vec}}}{\partial \boldsymbol{\delta \theta}} \quad (7.60)$$

where $\widetilde{\mathbf{T}}_I^C$ is the true rotation matrix from the camera to the inertial frame and $[\bullet \times]$ places \bullet into the skew-symmetric cross product matrix. Finally, by considering the structure of the skew-symmetric cross product matrix, it is easy to find that

$$\frac{\partial [\boldsymbol{\delta \theta} \times]_{\text{vec}}}{\partial \boldsymbol{\delta \theta}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad (7.61)$$

which fully defines the derivatives of the trace operators and the Jacobian matrix of the state vector with respect to the parameter vector.

7.4.5 Numeric Verification of the Developed Covariance Expressions

In order to verify the analytic expressions provided above, a Monte Carlo analysis of 1,000 runs was conducted to generate a numerical covariance to compare with the analytic covariance. The Monte Carlo was performed by generating synthetic 1024×1024 images of the Earth, using an edge extraction method to extract the edges, and then using the CAD OPNAV method on a portion of the extracted edges. The results are presented as 3D and 2D scatter plots with corresponding 3σ error ellipsoids in Figure 7.5. These results were run on a case where the Moon was located at $\begin{bmatrix} -750 & 350 & 100,000 \end{bmatrix}$ km in the camera frame, the Moon was nominally modeled as a sphere with radius 1737.4 km, the standard deviation for each ellipsoid parameter was 2.5 km [116], the attitude error was assumed to have a standard deviation of 0.02 degrees, the spacecraft was assumed to be on the equatorial plane, and the synthetic image was assumed to have a SNR of 13.125 with no blur (allowing the points to be extracted within 0.002 degrees or 0.07 pixels). As a point of comparison for the horizon extraction error, the Orion docking camera has a $40^\circ \times 30^\circ$ FOV with a 2400×1800 pixel FPA [117]. This results in approximately 0.017 deg/pixel. Thus, an error of 0.002 deg is equivalent to approximately 0.12 pixels on the docking camera FPA, which is more than achievable with modern edge extraction procedures.

It is clear from Figure 7.5 that the analytic and numerical covariance match very well. There is almost no noticeable difference between the analytic covariance (black covariance lines) and the numerically computed Monte Carlo covariance (red covariance lines). This verifies that the analytic expressions presented capture the important contributing factors to the CAD measurement error. Therefore, the above equations successfully trace errors originating on the camera focal plane array and in the planet model through the entire projection, image processing, and estimation processes, ending with an expression for the CAD OPNAV measurement covariance.

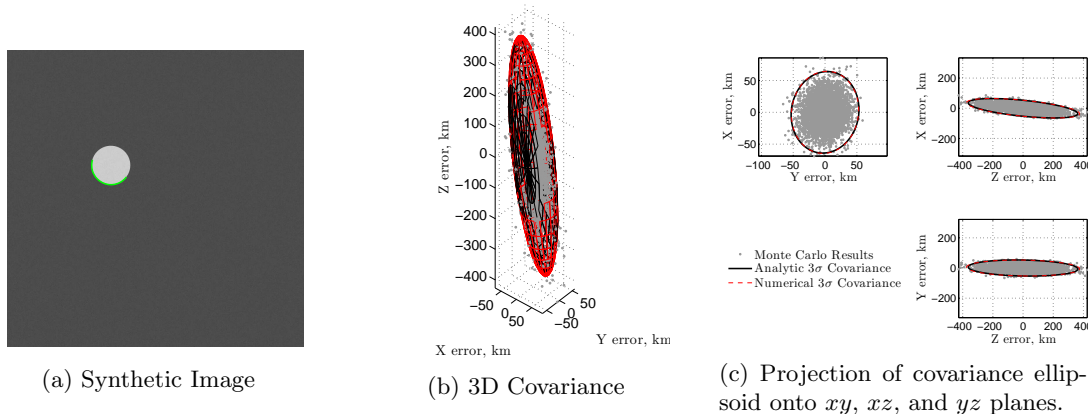


Figure 7.5: Synthetic image, error ellipses, and Monte Carlo results. CAD method performed only on the edge highlighted in green.

7.5 The Design of Natural Frequencies of a Mass Spring System

We will now consider a simple problem in dynamics to see how the Jacobians can be used in the design of systems. Assume we have a simple undamped three degree of freedom mass spring system as shown in Fig. 7.6. Now, it is easy to see that this system is governed by the differential equation

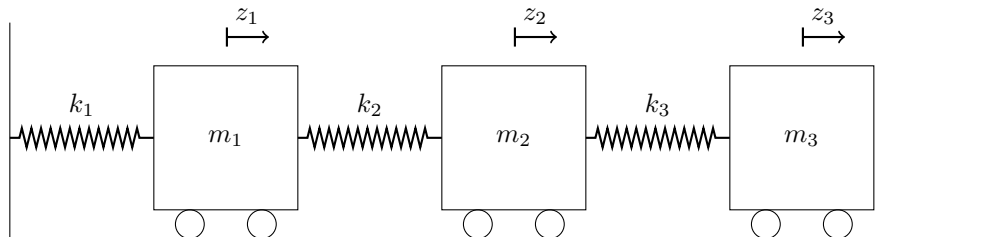


Figure 7.6: A simple undamped 3 degree of freedom mass spring system. m_i are the masses of the blocks and k_i are the spring constants of the springs. The wheels are allow the boxes to glide frictionlessly.

$$\mathbf{M}\ddot{\mathbf{z}} + \mathbf{K}\mathbf{z} = \mathbf{0} \tag{7.62}$$

where

$$\mathbf{M} = \text{diag} \begin{bmatrix} m_1 & m_2 & m_3 \end{bmatrix}, \quad (7.63)$$

$$\mathbf{K} = \begin{bmatrix} k_1 + k_2 & -k_2 & 0 \\ -k_2 & k_2 + k_3 & -k_3 \\ 0 & -k_3 & k_3 \end{bmatrix}, \quad (7.64)$$

and m_{1-3} and k_{1-3} are the mass and spring constants from Fig. 7.6, respectively.

It can be shown that an ODE of this type will have a solution of the form

$$\mathbf{z}(t) = \sum_{i=1}^3 c_i \mathbf{v}_i \cos(\sqrt{\lambda_i} t + \phi_i) \quad (7.65)$$

where c_i is a constant coefficient determined by the initial conditions, \mathbf{v}_i and λ_i are the i^{th} eigenvalue/eigenvector pair of

$$\mathbf{K}\mathbf{v} = \lambda\mathbf{M}\mathbf{v} \quad \mathbf{v}^T\mathbf{M}\mathbf{v} = 1, \quad (7.66)$$

and ϕ_i is a constant phase shift angle determined by the initial conditions of the system [21]. In these types of problems, the square roots of the eigenvalues are known as the “natural frequencies” of the system, and the eigenvectors are known as the normal modal shapes, with the reason coming from the way these values affect the solution of the problem.

When designing dynamical systems, it is generally a good idea to avoid applying a cyclic force to a system that has a similar frequency to one of the natural frequencies of the system, as this can quickly lead to instability and failure.³ In general, however, we cannot control the forcing conditions, only the design of the system itself. Therefore, it is important to be sure that any expected cyclic forces will not excite close to the natural frequency of the system, which requires “designing” the eigenvalues of the system. Having the Jacobians for the eigenvalues can be a great help in these cases.

For instance, return to our simple example problem and assume that we know this system will experience a cyclic excitation of $2Hz$. Also, assume we have a nominal design of $m_1 = 5kg$, $m_2 = 2kg$, $m_3 = 3kg$, $k_1 = 2N/m$, $k_2 = 5N/m$, and $k_3 = 1N/m$. The eigenvalues and eigenvectors of this

³For a great example of this, checkout some clips of the Takoma Narrows Bridge collapse [118].

nominal design are given by

$$\begin{aligned} \lambda_1 &= 0.1419 & \lambda_2 &= 0.5863 & \lambda_3 &= 4.0051 \\ \mathbf{v}_1 &= \begin{bmatrix} -0.1204 \\ -0.1515 \\ -0.2638 \end{bmatrix} & \mathbf{v}_2 &= \begin{bmatrix} 0.1716 \\ 0.1396 \\ -0.1840 \end{bmatrix} & \mathbf{v}_3 &= \begin{bmatrix} -0.1494 \\ -0.3893 \\ -0.0353 \end{bmatrix}. \end{aligned}$$

Therefore, it is easy to see that the nominal design has a natural frequency close to 2 from the third eigenvalue, which based on our expected excitation is not good. We can use the eigenvalue Jacobian for the third eigenvalue to see how we can adjust this value.⁴

The eigenvalue Jacobian can be calculated directly using Eq. 6.59 with $\mathbf{A} = \mathbf{K}$, $\mathbf{B} = \mathbf{M}$, and

$$\mathbf{x} = \begin{bmatrix} m_1 & m_2 & m_3 & k_1 & k_2 & k_3 \end{bmatrix}^T. \quad (7.67)$$

All we need for the Jacobian is the partial derivatives of \mathbf{K} and \mathbf{M} with respect to \mathbf{x} , which are given by

$$\frac{\partial \mathbf{K}_{\text{vec}}}{\partial \mathbf{x}} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 0 \\ \mathbf{0}_{9 \times 3} & 0 & 1 & 1 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix} \quad \frac{\partial \mathbf{M}_{\text{vec}}}{\partial \mathbf{x}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 & \mathbf{0}_{9 \times 3} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (7.68)$$

This allows us to calculate the eigenvalue Jacobian as

$$\frac{\partial \lambda_3}{\partial \mathbf{x}} = \begin{bmatrix} -0.2137 & -1.4503 & -0.0120 & 0.0534 & 0.6935 & 0.4309 \end{bmatrix}, \quad (7.69)$$

and thus we can predict the new eigenvalue by right multiplying $\partial \lambda_3 / \partial \mathbf{x}$ by our changes to \mathbf{x} . For instance, if we were to change our value of m_2 from $2kg$ to $2.25kg$,⁵ then we would expect the new

⁴Of course, this simple problem is symbolically solvable; thus, we do not need a nominal design, but in general, the systems are much more complex in real world applications.

⁵Here we chose to perturb m_2 since it will result in the largest change to the natural frequency according to the

value for λ_3 to change by -0.3626 , resulting in $\lambda_3 = 3.6425$ and a new natural frequency of 1.9085 . In actuality, we get a new eigenvalue of $\lambda_3 = 3.6539$, resulting in a new natural frequency of 1.9193 . This shows how we can use the Jacobians in the design of dynamic systems. A similar extension to FEA is also applicable.

eigenvalue Jacobian. This is another benefit of having symbolic expressions for the eigenvalue Jacobians.

Chapter 8

Conclusion

The covariance expressions for eigenvalues and eigenvectors were developed with respect to the covariance of the elements of the parent matrix. These expressions required the calculation of eigenvalue and eigenvector Jacobian matrices with respect to the elements of the parent matrix.

Two new formulations were derived for the complete Jacobians of eigenvalues and eigenvectors with respect to the elements of their parent matrix. The solutions rely on only the eigenvalue and eigenvector being considered and are valid for any simple complex eigenvalue–eigenvector pair. Further, the parent matrix may contain complex entries and need not be symmetric. As a result, the methods presented here are extremely general with applications to finite-element analysis (FEA) solutions to vibration problems, fitting of an ellipse to scattered data points, quaternion-based attitude estimation, and a host of other important scientific and engineering problems.

The second set of eigenvalue and eigenvector Jacobians developed in Chapter 6 was shown to collapse to well known results if the parent matrix is either (1) real and symmetric or (2) real and diagonal. This method may also be reinterpreted to gain a deeper understanding of perturbations of the eigenspace. In addition, this method was extended to the generalized eigenvalue problem for two different normalization conditions.

The eigenvalue and eigenvector Jacobians were validated by comparison with forward finite differencing. The computational performance speed of the technique in Chapter 6 was shown to be better by a factor of 10 (or greater for large matrices) when compared with the performance of the technique in Chapter 5.

Example applications were given detailing how to apply these Jacobians to real estimation and ODE problems. The results for each application were validated using Monte Carlo Analysis.

Future work on these topics includes extending the method presented in Chapter 6 to the case when the eigenvalues are semi-simple. The existing literature indicates that to do this will require the computation of the Hessian matrix (second derivative) of the eigenvalues with respect to the parent matrix. In addition, it would be useful to compare the numerical performance of this method with the techniques from the existing literature for the general case. Finally, further insight could be gained by examining more simplified cases.

Bibliography

- [1] Truesdell, C., “The Rational Mechanics of Elastic or Flexible Bodies, 1638-1788,” *Leonhardi Euleri opera omnia, Serie II*, Vol. 11, 1960.
- [2] Hawkins, T., “Cauchy and the spectral theory of matrices,” *Historia Mathematica*, Vol. 2, No. 1, 1975, pp. 1 – 29.
- [3] Kline, M., *Mathematical Thought From Ancient to Modern Times*, Oxford University Press, New York, 1972, pp. 799–812.
- [4] d’Alembert, J. L. R., *Traité de dynamique*, Paris, 1st ed., 1743.
- [5] DAlembert, J., “Suite des recherches sur le calcul intégral,” *Histoire de l’académie des sciences, Berlin*, Vol. 4, 1748, pp. 249–291.
- [6] d’Alembert, J. L. R., *Traité de dynamique*, Paris, 2nd ed., 1758.
- [7] Lagrange, J., “Solution de différents problèmes de calcul intégral, Miscellanea Tauriniensia, t. III (1762–1765),” *Reprinted in: Oeuvres de Lagrange*, Vol. 1, 1766, pp. 471–678.
- [8] Lagrange, J., “Recherches sur les équations séculaires des mouvements des noeuds et des inclinaisons des planètes,” *Mémoires de l’Académie des Sciences de Paris*, 1778.
- [9] de Laplace, P. S., “Mémoire Sur les Solutions particulières des Équations différentielles, et sur les inégalités séculaires des Planètes,” *Mémoires de l’Académie des Sciences de Paris*, 1772.
- [10] de Laplace, P. S., “Recherches sur le calcul intégral et sur le système du monde,” *Mémoires de l’Académie des Sciences de Paris*, 1772.
- [11] Laplace, P. S., *Traité de mécanique céleste*, Vol. 1, Paris, 1799.
- [12] Euler, L., “Du mouvement d’un corps solide quelconque lorsqu’il tourne autour d’un axe mobile,” *Mémoires de l’Académie des Sciences de Berlin*, Vol. 16, No. 1760, 1767, pp. 176–227.

- [13] Lagrange, J. L., *Mécanique analytique*, Vol. 1, Paris, 2nd ed., 1811.
- [14] Cauchy, A.-L., *Leçons sur les applications du calcul infinitesimal a la geometrie*, Vol. 1, De l'imprimerie royale, Paris, 1826.
- [15] Hermite, C., "Remarque Sur un Théorème de M. Cauchy," *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences*, Vol. 41, 1855, pp. 181–183.
- [16] Cauchy, B. A. L., *Mémoire sur l'intégration des équations linéaires*, 1839.
- [17] Hilbert, D., *Grundzuge einer allgemeinen theorie der linearen integralgleichungen*, , 1906.
- [18] Boyce, W. E. and DiPrima, R. C., *Elementary Differential Equations and Boundary Value Problems*, Wiley, Jefferson City, MO, 8th ed., 2005.
- [19] Ogata, K., *Modern control engineering*, Pearson Prentice Hall, Upper Saddle River, NJ, 5th ed., 2010.
- [20] Hughes, T. J., *The finite element method: linear static and dynamic finite element analysis*, Dover Publications, Inc., Mineola, NY, 2012.
- [21] Rao, S. S., *Mechanical Vibrations*, Prentice Hall, Columbus, 5th ed., 2011.
- [22] Bathe, K.-J. and Wilson, E. L., "Solution Methods for Eigenvalue Problems in Structural Mechanics," *International Journal for Numerical Methods in Engineering*, Vol. 6, No. 2, 1973, pp. 213–226.
- [23] Connors, K. A., *Chemical Kinetics: The Study of Reaction Rates in Solution*, John Wiley & Sons, 1990.
- [24] Christian, J. A., "Optical Navigation Using Planet's Centroid and Apparent Diameter in Image," *Journal of Guidance, Control, and Dynamics*, Vol. 38, No. 2, 2015, pp. 192–204.
- [25] Liounis, A. J., Gerhard, J., and Christian, J., "Performance Assessment of Horizon-Based Optical Navigation Techniques," *AAS Guidance, Navigation, and Control Conference*, Breckenridge, CO, January 2015.
- [26] Shuster, M., "Maximum Likelihood Estimation of Spacecraft Attitude," *Journal of the Astronautical Sciences*, Vol. 37, No. 1, 1989, pp. 79–88.

- [27] Halir, R. and Flusser, J., “Numerically Stable Direct Least Squares Fitting of Ellipses,” *Proc. 6th International Conference in Central Europe on Computer Graphics and Visualization. WSCG*, Vol. 98, Citeseer, 1998, pp. 125–132.
- [28] Greenberg, M. D., *Advanced Engineering Mathematics*, Prentice-Hall, Upper Saddle River, NJ, 1988.
- [29] Fitzgibbon, A., Pilu, M., and Fisher, R. B., “Direct Least Square Fitting of Ellipses,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 21, No. 5, 1999, pp. 476–480.
- [30] Ma, Y., Soatto, S., Košecák, J., and Sastry, S., *An Invitation to 3-D Vision: From Images to Geometric Models*, Springer, New York, NY, 2010.
- [31] Hartley, R., “In Defense of the Eight-Point Algorithm,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 6, 1997, pp. 580–593.
- [32] Horn, B., “Closed-form Solution of Absolute Orientation Using Unit Quaternions,” *Journal of the Optical Society of America*, Vol. 4, No. 4, 1987, pp. 629–642.
- [33] Roy, R. and Kailath, T., “ESPRIT — Estimation of Signal Parameters Via rotational Invariance Techniques,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 37, No. 7, 1989, pp. 984–995.
- [34] Takàcs, G. and Rohal’-Ilkiv, B., *Model Predictive Vibration Control*, chap. Basics of Vibration Dynamics, Springer, 2012, pp. 25–64.
- [35] Keat, J., “Analysis of Least-Squares Attitude Determination Routine DOAOP,” Tech. Rep. CSC/TM-77/6034, Computer Sciences Corp., Falls Church, VA, Feb 1977.
- [36] Markley, F. L. and Crassidis, J. L., *Fundamentals of Spacecraft Attitude Determination and Control*, Springer, New York, NY, 2014.
- [37] Hull, D. G., *Optimal Control Theory for Applications*, Springer, New York, NY, 2003.
- [38] Plaut, R., “Determining the Nature of Instability in Nonconservative Problems,” *AIAA Journal*, Vol. 10, No. 7, 1972, pp. 967–968.
- [39] Nail, J. B., Mitchell, J. R., and McDaniel, W. L., “Determination of Pole Sensitivities by Danilevskii’s Method,” *AIAA Journal*, Vol. 15, No. 10, 1977, pp. 1525–1527.

- [40] Mitchell, J. R., Nail, J. B., and McDaniel, W. L., “Computation of Cofactors of [sI-A] With Applications,” *IEEE Transactions on Automatic Control*, Vol. 24, No. 4, Aug 1979, pp. 637–638.
- [41] Magnus, J. R., “On Differentiating Eigenvalues and Eigenvectors,” *Econometric Theory*, Vol. 1, No. 02, 1985, pp. 179–191.
- [42] Murthy, D. V. and Haftka, R. T., “Derivatives of Eigenvalues and Eigenvectors of a General Complex Matrix,” *International Journal for Numerical Methods in Engineering*, Vol. 26, No. 2, 1988, pp. 293–311.
- [43] Xu, Z. and Wu, B., “Derivatives of Complex Eigenvectors With Distinct and Repeated Eigenvalues,” *International Journal for Numerical Methods in Engineering*, Vol. 75, No. 8, 2008, pp. 945–963.
- [44] Andrew, A. L., Chu, K.-W. E., and Lancaster, P., “Derivatives of Eigenvalues and Eigenvectors of Matrix Functions,” *SIAM Journal on Matrix Analysis and Applications*, Vol. 14, No. 4, 1993, pp. 903–926.
- [45] Meirovitch, L., *Principles and Techniques of Vibrations*, Vol. 1, Prentice Hall New Jersey, 1997.
- [46] Meyer, C. D., *Matrix Analysis and Applied Linear Algebra*, Vol. 2, Siam, 2000.
- [47] Liounis, A. J. and Christian, J. A., “Techniques for Generating Analytic Covariance Expressions for Eigenvalues and Eigenvectors,” *IEEE Transactions on Signal Processing*, Vol. 64, No. 7, April 2016, pp. 1808–1821.
- [48] Van Loan, C., “On Estimating the Condition of Eigenvalues and Eigenvectors,” *Linear Algebra and Its Applications*, Vol. 88, 1987, pp. 715–732.
- [49] Stewart, G. W., “Error and Perturbation Bounds for Subspaces Associated With Certain Eigenvalue Problems,” *SIAM Review*, Vol. 15, No. 4, 1973, pp. 727–764.
- [50] Wilkinson, J. H., *The Algebraic Eigenvalue Problem*, Vol. 87, Clarendon Press Oxford, 1965.
- [51] Golub, G. H. and Van Loan, C. F., *Matrix Computations*, Vol. 3, JHU Press, 1996.
- [52] Wilkinson, J., “Note on Matrices With a Very Ill-Conditioned Eigenproblem,” *Numerische Mathematik*, Vol. 19, No. 2, 1972, pp. 176–178.

- [53] Gershgorin, S. A., “Über Die Abgrenzung Der Eigenwerte Einer Matrix,” *Proceedings of the Russian Academy of Mathematical Sciences, Mathematical Series*, , No. 6, 1931, pp. 749–754.
- [54] Brualdi, R. A. and Mellendorf, S., “Regions in the complex plane containing the eigenvalues of a matrix,” *American Mathematical Monthly*, 1994, pp. 975–985.
- [55] Magnus, J. R. and Neudecker, H., *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley & Sons, New York, NY, 3rd ed., 2007.
- [56] Mathews, J. H. and Fink, K. D., *Numerical Methods Using MATLAB*, Prentice Hall, Upper Saddle River, NJ, 4th ed., 2004.
- [57] Zehfuss, G., “Über eine gewisse Determinante,” *Zeitschrift für Mathematik und Physik*, Vol. 3, 1858, pp. 298–301.
- [58] Henderson, H. V., Pukelsheim, F., and Searle, S. R., “On the history of the Kronecker product,” *Linear and Multilinear Algebra*, Vol. 14, No. 2, 1983, pp. 113–120.
- [59] Stewart, J., *Essential Calculus: Early Transcendentals*, Thomson Brooks/Cole, 2007.
- [60] Brown, R. G. and Hwang, P. Y., *Introduction to Random Signal Analysis and Kalman Filtering*, John Wiley & Sons, 4th ed., 2012.
- [61] Papadopoulos, T. and Lourakis, M., “Estimating the Jacobian of the Singular Value Decomposition: Theory and Applications,” *Lecture Notes in Computer Science*, Vol. 1842, 2000, pp. 554–570.
- [62] Nelson, R. B., “Simplified Calculation of Eigenvector Derivatives,” *AIAA Journal*, Vol. 14, No. 9, 1976, pp. 1201–1205.
- [63] Fox, R. and Kapoor, M., “Rates of Change of Eigenvalues and Eigenvectors.” *AIAA journal*, Vol. 6, No. 12, 1968, pp. 2426–2429.
- [64] Rogers, L. C., “Derivatives of Eigenvalues and Eigenvectors,” *AIAA journal*, Vol. 8, No. 5, 1970, pp. 943–944.
- [65] Plaut, R. and Huseyin, K., “Derivatives of Eigenvalues and Eigenvectors in Non-Self-Adjoint Systems.” *AIAA journal*, Vol. 11, No. 2, 1973, pp. 250–251.
- [66] Kalaba, R., Spingarn, K., and Tesfatsion, L., “Variational Equations for the Eigenvalues and Eigenvectors of Nonsymmetric Matrices,” *Journal of Optimization Theory and Applications*, Vol. 33, No. 1, 1981, pp. 1–8.

- [67] Lim, K., Junkins, J., and Wang, B., “Re-examination of eigenvector derivatives,” *Journal of Guidance, Control, and Dynamics*, Vol. 10, No. 6, 1987, pp. 581–587.
- [68] Andrew, A. L. and Tan, R. C., “Computation of derivatives of repeated eigenvalues and corresponding eigenvectors by simultaneous iteration,” *AIAA Journal*, Vol. 34, No. 10, 1996, pp. 2214–2216.
- [69] Andrew, A. L. and Tan, R. C., “Computation of Derivatives of Repeated Eigenvalues and the Corresponding Eigenvectors of Symmetric Matrix Pencils,” *SIAM Journal on Matrix Analysis and Applications*, Vol. 20, No. 1, 1998, pp. 78–100.
- [70] Van Der Aa, N., Ter Morsche, H., and Mattheij, R., “Computation of Eigenvalue and Eigenvector Derivatives for a General Complex-Valued Eigensystem,” *Electronic Journal of Linear Algebra*, Vol. 16, No. 1, 2007, pp. 300–314.
- [71] Garg, S., “Derivatives of Eigensolutions for a General Matrix.” *AIAA Journal*, Vol. 11, No. 8, 1973, pp. 1191–1194.
- [72] Rudisill, C. S., “Derivatives of Eigenvalues and Eigenvectors for a General Matrix,” *AIAA Journal*, Vol. 12, No. 5, 1974, pp. 721–722.
- [73] Rudisill, C. S. and Chu, Y.-Y., “Numerical Methods for Evaluating the Derivatives of Eigenvalues and Eigenvectors,” *AIAA Journal*, Vol. 13, No. 6, 1975, pp. 834–837.
- [74] Juang, J.-N. and Lim, K. B., “On the eigenvalue and eigenvector derivatives of a general matrix,” *NASA Technical Reports Server*, 1987.
- [75] Lim, K. B., Juang, J.-N., and Ghaemmaghani, P., “Eigenvector derivatives of repeated eigenvalues using singular value decomposition,” *Journal of Guidance, Control, and Dynamics*, Vol. 12, No. 2, 1989, pp. 282–283.
- [76] Juang, J.-N., Ghaemmaghani, P., and Lim, K. B., “Eigenvalue and eigenvector derivatives of a nondefective matrix,” *Journal of Guidance, Control, and Dynamics*, Vol. 12, No. 4, 1989, pp. 480–486.
- [77] Cardani, C. and Mantegazza, P., “Calculation of Eigenvalue and Eigenvector Derivatives for Algebraic Flutter and Divergence Eigenproblems,” *AIAA Journal*, Vol. 17, No. 4, 1979, pp. 408–412.

- [78] Ojalvo, I., “Gradients for Large Structural Models With Repeated Frequencies,” Tech. Rep. 861789, SAE Technical Paper, 1986.
- [79] Dailey, R. L., “Eigenvector Derivatives With Repeated Eigenvalues,” *AIAA Journal*, Vol. 27, No. 4, 1989, pp. 486–491.
- [80] Mills-Curran, W. C., “Calculation of Eigenvector Derivatives for Structures With Repeated Eigenvalues,” *AIAA Journal*, Vol. 26, No. 7, 1988, pp. 867–871.
- [81] Friswell, M., “The Derivatives of Repeated Eigenvalues and Their Associated Eigenvectors,” *Journal of Vibration and Acoustics*, Vol. 118, No. 3, 1996, pp. 390–397.
- [82] Wu, B., Xu, Z., and Li, Z., “Improved Nelson’s Method for Computing Eigenvector Derivatives with Distinct and Repeated Eigenvalues,” *AIAA Journal*, Vol. 45, No. 4, 2007, pp. 950–952.
- [83] Friswell, M. I. and Adhikari, S., “Derivatives of Complex Eigenvectors Using Nelson’s Method,” *AIAA Journal*, Vol. 38, No. 12, 2000, pp. 2355–2357.
- [84] Meyer, C. D. and Stewart, G. W., “Derivatives and Perturbations of Eigenvectors,” *SIAM Journal on Numerical Analysis*, Vol. 25, No. 3, 1988, pp. 679–691.
- [85] Xie, H. and Dai, H., “Davidson Method for Eigenpairs and Their Partial Derivatives of Generalized Eigenvalue Problems,” *Communications in Numerical Methods in Engineering*, Vol. 22, No. 2, 2006, pp. 155–165.
- [86] Crouzeix, M., Philippe, B., and Sadkane, M., “The Davidson Method,” *SIAM Journal on Scientific Computing*, Vol. 15, No. 1, 1994, pp. 62–76.
- [87] de Leeuw, J., “Derivatives of generalized eigensystems with applications,” *Department of Statistics, UCLA*, 2007.
- [88] Trefethen, L. N. and Bau III, D., *Numerical Linear Algebra*, Vol. 50, SIAM, Philadelphia, PA, 1997.
- [89] Winitzki, S., *Linear Algebra Via Exterior Products*, Sergei Winitzki, 2010.
- [90] Bourbaki, N., *Algebra I*, Hermann, Paris, France, 1971.
- [91] Hamilton, W. R., *Lectures on quaternions*, 1853.
- [92] Cayley, A., “A memoir on the theory of matrices,” *Philosophical Transactions of the Royal Society of London*, 1858, pp. 17–37.

- [93] Frobenius, G., “Ueber lineare Substitutionen und bilineare Formen.” *Journal für die reine und angewandte Mathematik*, Vol. 84, 1877, pp. 1–63.
- [94] Penrose, R., “A Generalized Inverse for Matrices,” *Proc. Cambridge Philos. Soc.*, Vol. 51, Cambridge University Press, 1955, pp. 406–413.
- [95] Meyer, C. D., “Generalized Inversion of Modified Matrices,” *SIAM Journal on Applied Mathematics*, Vol. 24, No. 3, 1973, pp. 315–323.
- [96] Atkinson, K. E., *An Introduction to Numerical Analysis*, John Wiley & Sons, 2008, pp. 591–601.
- [97] Moler, C. B. and Stewart, G. W., “An algorithm for generalized matrix eigenvalue problems,” *SIAM Journal on Numerical Analysis*, Vol. 10, No. 2, 1973, pp. 241–256.
- [98] Bookstein, F. L., “Fitting Conic Sections to Scattered Data,” *Computer Graphics and Image Processing*, Vol. 9, No. 1, 1979, pp. 56–71.
- [99] Sampson, P. D., “Fitting Conic Sections to “Very Scattered” Data: An Iterative Refinement of the Bookstein Algorithm,” *Computer graphics and image processing*, Vol. 18, No. 1, 1982, pp. 97–108.
- [100] Rosin, P. L., “A Note on the Least Squares Fitting of Ellipses,” *Pattern Recognition Letters*, Vol. 14, No. 10, 1993, pp. 799–808.
- [101] Zhang, Z., “Parameter Estimation Techniques: A Tutorial with Application to Conic Fitting,” *Image and Vision Computing*, Vol. 15, No. 1, 1997, pp. 59–76.
- [102] Chojnacki, W., Brooks, M. J., Van Den Hengel, A., and Gawley, D., “On the Fitting of Surfaces to Data with Covariances,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 11, 2000, pp. 1294–1303.
- [103] Kanatani, K. and Rangarajan, P., “Hyperaccurate Ellipse Fitting Without Iterations.” *VIS-APP (2)*, 2010, pp. 5–12.
- [104] Maini, E. S., “Robust Ellipse-Specific Fitting for Real-Time Machine Vision,” *Brain, Vision, and Artificial Intelligence*, Springer, 2005, pp. 318–327.
- [105] Cheng, Y., Johnson, A. E., Matthies, L. H., and Olson, C. F., “Optical Landmark Detection for Spacecraft Navigation,” *Advances in the Astronautical Sciences*, Vol. 114, No. 3, 2003, pp. 1785–1803.

- [106] Hanak, C., Crain, T., and Bishop, R., “Crater Identification Algorithm for the Lost in Low Lunar Orbit Scenario,” *AAS Guidance & Control Conference*, Breckenridge, CO, 2008.
- [107] Christian, J. A. and Lightsey, E. G., “Onboard Image-Processing Algorithm for a Spacecraft Optical Navigation Sensor System,” *Journal of Spacecraft and Rockets*, Vol. 49, No. 2, 2012, pp. 337–352.
- [108] Mortari, D., de Dilectis, F., and D’Souza, C., “Image Processing of Illuminated Ellipsoid,” *AAS/AIAA Astrodynamics Specialist Conference*, Paper 2013-853, American Astronautical Society, Hilton Head, SC, 2013.
- [109] Borissov, S. and Mortari, D., “Optimal Single-Point and Filtered Pose Estimation for Lunar Orbiters Using Visible Cameras,” *24th AAS/AIAA Space Flight Mechanics Meeting*, Paper AAS 14-247, Santa Fe, NM, January 2014.
- [110] Wahba, G., “A Least Square Estimate of Satellite Attitude,” *SIAM Review*, Vol. 7, No. 3, July 1965, pp. 409.
- [111] Shuster, M. and Oh, S., “Three-Axis Attitude Determination from Vector Observations,” *Journal of Guidance and Control*, Vol. 4, No. 1, 1981, pp. 70–77.
- [112] Mortari, D., “ESOQ: A Closed-Form Solution to the Wahba Problem,” *Journal of the Astronautical Sciences*, Vol. 45, No. 2, 1997, pp. 195–204.
- [113] Ainscough, T., Zanetti, R., Christian, J., and Spanos, P., “Q-Method Extended Kalman Filter,” *Journal of Guidance, Control, and Dynamics*, Vol. 38, No. 4, 2015, pp. 752–760.
- [114] Christian John A., “Optical Navigation Using Iterative Horizon Reprojection,” *Journal of Guidance, Control, and Dynamics*, 2016, pp. 112, doi: 10.2514/1.G001569.
- [115] Christian, J. and Robinson, S., “Observations of the Geometry of Horizon-Based Optical Navigation,” *39th Annual AAS Guidance, Navigation, and Control Conference*, No. AAS 16-151, AAS Rocky Mountain Section, Breckenridge, CO, February 2016.
- [116] Archinal, B. A., A’hearn, M., Bowell, E., Conrad, A., Consolmagno, G. J., Courtin, R., Fukushima, T., Hestroffer, D., Hilton, J. L., Krasinsky, G. A., et al., “Report of the IAU Working Group on Cartographic Coordinates and Rotational Elements: 2009,” *Celestial Mechanics and Dynamical Astronomy*, Vol. 109, No. 2, 2011, pp. 101–135.

- [117] Christian, J., Hinkel, H., D'Souza, C., Maguire, S., and Patangan, M., "The Sensor Test for Orion RelNav Risk Mitigation (STORRM) Development Test Objective," *AIAA Guidance, Navigation, and Control Conference*, Portland, OR, 8-11 August 2011.
- [118] Billah, K. Y. and Scanlan, R. H., "Resonance, Tacoma Narrows bridge failure, and undergraduate physics textbooks," *American Journal of Physics*, Vol. 59, No. 2, 1991, pp. 118–124.