Zeros, Zeros, Zeros: How To Find Them Using Broyden’s Method

A Thesis

Presented in Partial Fulfillment of the Requirements for the Degree Master of Mathematical Sciences in the Graduate School of The Ohio State University

By

James Sutherland, B.S.

Graduate Program in Department of Mathematical Sciences

The Ohio State University

2017

Master’s Examination Committee:

Dr. Edward Overman, Advisor

Dr. Ching-Shan Chou
© Copyright by
James Sutherland
2017
Zero-finding methods are used in a wide variety of applications throughout computational mathematics. Newton’s method is arguably the most well-known iterative zero-finding method, but its need of the Jacobian matrix in higher dimensions can sometimes lead to serious problems. In 1965, C.G. Broyden developed two n-dimensional secant methods that use the Jacobian only at the initial iterate and perform a rank-one update at all other iterations. Over time these methods became known as Broyden’s “good” method and Broyden’s “bad” method. There have been many studies done with a few carefully selected initial points that conclude the naming of Broyden’s methods are accurate. However, rather than focusing on just a few isolated initial points, in this project convergence plots were utilized as a primary tool in the analysis of these methods to facilitate an investigation of several million initial points at once. Using a variety of examples involving nonlinear functions, the results show that each method has its benefits depending on the problem under investigation. The labels “good” and “bad” can be misleading and should not influence one method to be used more frequently than the other.
For Lucy, who is truly man’s best friend.
Acknowledgments

I would like to first thank my thesis advisor Dr. Edward Overman. He was an invaluable source of knowledge throughout both the course of this thesis and my graduate studies. Additionally, he constantly encouraged me to never stop having fun with math. I would also like to thank Dr. Ching-Shan Chou for serving on my defense committee. A very big thank you goes out to my anonymous donor, whose support this summer gave me the opportunity to devote my full attention to the completion of this thesis. I would like to also thank Willa Skeehan for her feedback and encouragement throughout this whole process. Finally, I would like to thank my fiancée Alli for her unending support and encouragement, and unfaltering faith in my ability to finish this thesis.
Vita

1991 ........................................... Born - Akron, OH, USA

2014 ......................................... B.S. Applied Mathematics,
                                  Kent State University

2015-2017 ................................. Graduate Teaching Associate,
                                  The Ohio State University.

Fields of Study

Major Field: Mathematical Sciences
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Dedication</td>
<td>iii</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>iv</td>
</tr>
<tr>
<td>Vita</td>
<td>v</td>
</tr>
<tr>
<td>List of Tables</td>
<td>ix</td>
</tr>
<tr>
<td>List of Figures</td>
<td>x</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Overview</td>
<td>1</td>
</tr>
<tr>
<td>1.2 One-Dimensional Methods</td>
<td>2</td>
</tr>
<tr>
<td>1.2.1 Newton’s Method</td>
<td>2</td>
</tr>
<tr>
<td>1.2.2 Secant Method</td>
<td>5</td>
</tr>
<tr>
<td>1.2.3 Fixed Chord Method</td>
<td>6</td>
</tr>
<tr>
<td>1.2.4 Convergence</td>
<td>6</td>
</tr>
<tr>
<td>1.2.5 A Global Modification: Line Search</td>
<td>12</td>
</tr>
<tr>
<td>1.3 N-Dimensional Methods</td>
<td>13</td>
</tr>
<tr>
<td>1.3.1 Linear Algebra Review: Matrix Norms</td>
<td>13</td>
</tr>
<tr>
<td>1.3.2 Newton’s Method</td>
<td>15</td>
</tr>
<tr>
<td>1.3.3 Fixed Chord Method</td>
<td>17</td>
</tr>
<tr>
<td>1.3.4 Broyden’s Method</td>
<td>17</td>
</tr>
<tr>
<td>1.3.5 Convergence</td>
<td>19</td>
</tr>
<tr>
<td>1.3.6 Global Modifications</td>
<td>23</td>
</tr>
<tr>
<td>1.4 Numerical Methods</td>
<td>23</td>
</tr>
<tr>
<td>1.4.1 Code # 1: Overall Convergence vs Nonconvergence</td>
<td>24</td>
</tr>
</tbody>
</table>
## Appendices 83

### A. Proof of Superlinear Convergence of Broyden’s Method . . . . . . . . . 83

#### A.1 The Method . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 83

#### A.2 Useful Lemmas . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 83

#### A.3 Theorems for q-Superlinear Convergence . . . . . . . . . . . . . . 86

### Bibliography . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 90
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>33</td>
</tr>
<tr>
<td>3.1</td>
<td>45</td>
</tr>
<tr>
<td>4.1</td>
<td>57</td>
</tr>
<tr>
<td>5.1</td>
<td>64</td>
</tr>
<tr>
<td>6.1</td>
<td>71</td>
</tr>
<tr>
<td>6.2</td>
<td>72</td>
</tr>
<tr>
<td>6.3</td>
<td>72</td>
</tr>
<tr>
<td>6.4</td>
<td>73</td>
</tr>
</tbody>
</table>
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Example of Newton iterates getting stuck in a loop.</td>
<td>4</td>
</tr>
<tr>
<td>2.1</td>
<td>Contour plot of Equation 2.1.</td>
<td>29</td>
</tr>
<tr>
<td>2.2</td>
<td>Newton’s local convergence regions of (2.1).</td>
<td>31</td>
</tr>
<tr>
<td>2.3</td>
<td>Self-similar behavior in Newton’s local convergence regions of (2.1).</td>
<td>32</td>
</tr>
<tr>
<td>2.4</td>
<td>“Good” Broyden’s local convergence regions of (2.1).</td>
<td>34</td>
</tr>
<tr>
<td>2.5</td>
<td>“Bad” Broyden’s local convergence regions of (2.1).</td>
<td>35</td>
</tr>
<tr>
<td>2.6</td>
<td>Detail of Broyden’s local convergence region near the zero of (2.1).</td>
<td>36</td>
</tr>
<tr>
<td>3.1</td>
<td>Contour plots of Equation 3.1.</td>
<td>40</td>
</tr>
<tr>
<td>3.2</td>
<td>Newton’s local convergence regions of all zeros to (3.1).</td>
<td>41</td>
</tr>
<tr>
<td>3.3</td>
<td>“Good” Broyden’s local convergence regions of all zeros of (3.1).</td>
<td>42</td>
</tr>
<tr>
<td>3.4</td>
<td>“Bad” Broyden’s local convergence regions of all zeros of (3.1).</td>
<td>43</td>
</tr>
<tr>
<td>3.5</td>
<td>Detail of Newton’s local convergence regions of (3.1) near its zeros.</td>
<td>45</td>
</tr>
<tr>
<td>3.6</td>
<td>Detail of Broyden’s local convergence regions of (3.1) near its zeros.</td>
<td>46</td>
</tr>
<tr>
<td>3.7</td>
<td>Newton’s global convergence regions of all zeros of (3.1).</td>
<td>47</td>
</tr>
<tr>
<td>4.1</td>
<td>Contour plot of Equation 4.1.</td>
<td>50</td>
</tr>
<tr>
<td>Section</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>4.2</td>
<td>Newton’s local convergence regions of all zeros of 4.1.</td>
<td>51</td>
</tr>
<tr>
<td>4.3</td>
<td>“Good” Broyden’s local convergence regions of all zeros of 4.1.</td>
<td>52</td>
</tr>
<tr>
<td>4.4</td>
<td>“Bad” Broyden’s local convergence regions of all zeros of 4.1.</td>
<td>53</td>
</tr>
<tr>
<td>4.5</td>
<td>Newton’s global convergence regions of all zeros of 4.1.</td>
<td>55</td>
</tr>
<tr>
<td>4.6</td>
<td>Newton’s local and global convergence regions over a larger area of all zeros of 4.1.</td>
<td>56</td>
</tr>
<tr>
<td>5.1</td>
<td>Contour plot of Equation 5.1.</td>
<td>59</td>
</tr>
<tr>
<td>5.2</td>
<td>Newton’s local convergence regions of all zeros of (5.1).</td>
<td>61</td>
</tr>
<tr>
<td>5.3</td>
<td>“Good” Broyden’s local convergence regions of all zeros of (5.1).</td>
<td>62</td>
</tr>
<tr>
<td>5.4</td>
<td>“Bad” Broyden’s local convergence regions of all zeros of (5.1).</td>
<td>63</td>
</tr>
<tr>
<td>5.5</td>
<td>Newton’s global convergence regions of (5.1) with varying step sizes.</td>
<td>65</td>
</tr>
<tr>
<td>6.1</td>
<td>Newton’s local individual convergence regions of (4.1) corresponding to each zero.</td>
<td>74</td>
</tr>
<tr>
<td>6.2</td>
<td>Self-similar behavior in Newton’s local convergence regions of (4.1).</td>
<td>75</td>
</tr>
<tr>
<td>6.3</td>
<td>Broyden’s local individual convergence regions of (4.1) corresponding to each zero.</td>
<td>77</td>
</tr>
<tr>
<td>6.4</td>
<td>Comparison of Newton and “bad” Broyden’s local individual convergence regions of (5.1) corresponding to each zero.</td>
<td>78</td>
</tr>
<tr>
<td>6.5</td>
<td>Function and error convergence plots for (4.1) using ( x^{(0)} = (2, -2) ) and ( x^{(0)} = (-1.6, 1.9) ).</td>
<td>80</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

1.1 Overview

The focus of this thesis will be the application of Newton’s method and Broyden’s secant method to a variety of nonlinear problems as a method of investigating the differences between the methods. Broyden’s secant method has two different formulations, which are more commonly known as the “good” and “bad” methods. A feature of our investigation will be the use of convergence plots to compare and contrast the effectiveness of each method.

In order to properly facilitate a discussion on these topics, it is necessary to lay some groundwork on these two methods. The first section of this introduction will introduce the one-dimensional version of each method. Once this has been accomplished, it will be easier to generalize each method to its \( n \)-dimensional implementation. Additionally, the convergence behavior of each method will be presented along with the conditions required for optimal convergence. Although our investigation mainly features local methods, for completeness there is a brief discussion of global modifications. Finally, we describe the numerical methods used in the computation of the figures in this thesis.
In each of the following four chapters, we will explore the effectiveness of each method on a particular nonlinear system of equations. This exploration will be primarily through the use of convergence plots, although other methods will be utilized as necessary. In the last chapter, we summarize the debate on the merits of Broyden’s “good” method vs. Broyden’s “bad” method and offer our opinion based on our results.

**Remark.** While a lot of the material in this chapter is based on Dennis and Schnabel [8], we have adopted Overman and Baker’s [2] improved notation to make the material as easy to follow as possible.

### 1.2 One-Dimensional Methods

In order to more easily generalize the methods to their $n$-dimensional formulations, we will first introduce their one-dimensional formulations and discuss the advantages and disadvantages of each method.

#### 1.2.1 Newton’s Method

Although the method bears Sir Isaac Newton’s name, it did not appear in its current formulation until 1740, courtesy of the English mathematician Thomas Simpson [9]. In fact, methods that can be perceived as replacing $f'(x)$ by a finite difference approximation or the slope of the secant line actually predate the modern-day formulation of Newton’s method [14].

The method is derived by first considering the linear approximation (or more generally, the affine model) of some nonlinear function $f(x)$ at some initial point $x^{(0)}$ in the domain of $f$. This model is denoted by the linear polynomial $p^{(0)}(x)$. From
calculus,

\[ p^{(0)}(x) = f(x^{(0)}) + f'(x^{(0)})(x - x^{(0)}). \] (1.1)

Denote the zero of \( p^{(0)}(x) \) by \( x^{(1)} \), i.e.,

\[ 0 = f(x^{(0)}) + f'(x^{(0)})(x^{(1)} - x^{(0)}). \]

Solving for this zero yields

\[ x^{(1)} = x^{(0)} - \frac{f(x^{(0)})}{f'(x^{(0)})}. \] (1.2)

This process can be repeated iteratively until the desired accuracy is achieved, yielding the general iterative algorithm

\[ x^{(\nu+1)} = x^{(\nu)} - \frac{f(x^{(\nu)})}{f'(x^{(\nu)})} \] (1.3)

for any \( \nu \in \mathbb{N}[0, \infty) \). This will yield the sequence of iterates \( \{ x^{(\nu)} \} \) such that

\[ \lim_{\nu \to \infty} x^{(\nu)} = \bar{x} \]

if the initial guess \( x^{(0)} \) is “good enough.” Here \( \bar{x} \) denotes the exact zero of \( f(x) \). An explanation of what “good enough” constitutes will be outlined in Section 1.2.4.

**Potential Problems**

Unfortunately, there are many problems that can arise when Newton’s Method is being used. For example, the iterates can diverge or they could become stuck in a loop. An example of each problem is presented below.

- **Diverging Iterates:** Suppose Newton’s Method is applied to \( f(x) = x^{1/3} \).

Then by (1.3),

\[ x^{(\nu+1)} = x^{(\nu)} - \frac{(x^{(\nu)})^{1/3}}{\frac{1}{3}(x^{(\nu)})^{-2/3}} = x^{(\nu)} - 3x^{(\nu)} = -2x^{(\nu)}, \]
or in other words,

\[ x^{(\nu+1)} = (-2)^{\nu+1}x^{(0)}. \]

So for any \( x^{(0)} \neq \bar{x} \), the sequence of iterates \( \{x^{(\nu)}\} \) will diverge. This is also a poor function to apply Newton’s Method to since \( f'(x) \) is undefined at \( \bar{x} = 0 \).

- **Iterates Stuck in a Loop:** Consider the function \( f(x) = x^3 - 2x + 2 \). If \( x^{(0)} = 0 \), then \( x^{(1)} = 1 \) and \( x^{(2)} = 0 \), and the iterates will cycle forever.

![Figure 1.1: Example of Newton iterates getting stuck in a loop.](image)

In addition, \( f'(x^{(\nu)}) \) could be zero for any \( \nu \in \mathbb{N}[0, \infty) \), or \( f'(x) \) could be very expensive or difficult to evaluate. The former case is very unlikely for \( \nu \in \mathbb{N}[1, \infty) \) from a numerical standpoint. However, in the latter case, a nice example is in the solution of a two-point boundary value problem via shooting [11]. Solving the resulting system of ODEs using Newton’s Method requires the calculation of the derivative of the
function with respect to the slope at the boundary point. The secant method can help to overcome this difficulty.

### 1.2.2 Secant Method

The basic idea is to represent \( f'(x) \) by some finite difference approximation, e.g.,

\[
\frac{f(x + h) - f(x)}{h} \quad \text{or} \quad \frac{f(x + h) - f(x - h)}{2h}.
\]

A method that uses a finite approximation for \( f'(x) \) is called a quasi-Newton method. The first is a forward difference and requires an additional evaluation \( f(x + h) \) but does not require the evaluation of \( f'(x) \). However, the centered difference will require an additional function evaluation than Newton’s Method due to the two new evaluations of \( f(x + h) \) and \( f(x - h) \). If \( f(x) \) is subject to severe round-off errors, it might be impossible to reduce \( h \) enough such that the approximation is accurate enough without losing too many digits of accuracy after the evaluation.

An alternative scheme is to approximate \( f'(x^{(\nu)}) \) by

\[
\frac{f(x^{(\nu)}) - f(x^{(\nu-1)})}{x^{(\nu)} - x^{(\nu-1)}},
\]

i.e., to approximate the slope of the tangent line of \( f(x) \) at \( x^{(\nu)} \) with the slope of the secant line through that same point. Then, the general iterative method can be rewritten as

\[
x^{(\nu+1)} = x^{(\nu)} - \frac{x^{(\nu)} - x^{(\nu-1)}}{f(x^{(\nu)}) - f(x^{(\nu-1)})} f(x^{(\nu)}).
\]

Notice that this formulation has no additional function evaluations as \( f(x^{(\nu-1)}) \) is required for the previous iteration. The requirement of \( two \) initial iterates poses no problem. In order to better facilitate the jump to the \( n \)-dimensional formulation of
the secant method, consider an alternate representation of this method:

\[
\begin{align*}
    x^{(\nu+1)} &= x^{(\nu)} - a^{\nu-1}_\nu f(x^{(\nu)}) \\
    a_{\nu+1} &= \frac{f(x^{(\nu+1)}) - f(x^{(\nu)})}{x^{(\nu+1)} - x^{(\nu)}}.
\end{align*}
\]

Here \( a_\nu \) represents the approximate slope of \( f(x) \) at \( x^{(\nu)} \). In contrast, this representation requires only one initial point \( x^{(0)} \) but also an initial slope \( a_0 \). Mathematically this is equivalent to (1.5).

### 1.2.3 Fixed Chord Method

A method to reduce the number of function evaluations in Newton’s Method is the fixed chord method. At each iteration, \( f'(x^{(\nu)}) \) is replaced by \( f'(x^{(0)}) \). Then, the general iterative step is

\[
x^{(\nu+1)} = x^{(\nu)} - \frac{f(x^{(\nu)})}{f'(x^{(0)})}
\]

(1.7) for any \( \nu \in \mathbb{N}[0, \infty) \). Unfortunately, this method only has linear convergence near the zero. Most of the time, the number of function evaluations saved is not enough to offset the greater number of iterations required to converge with the same accuracy as Newton’s Method. This concept can be more useful in an \( n \)-dimensional context and will be revisited in a later section.

### 1.2.4 Convergence

Given a sequence of real numbers \( \{x^{(\nu)}\} \) with \( \nu \in \mathbb{N}[0, \infty) \), the sequence converges to \( \bar{x} \in \mathbb{R} \) if

\[
\lim_{\nu \to \infty} |x^{(\nu)} - \bar{x}| \iff \lim_{\nu \to \infty} |e^{(\nu)}| = 0,
\]

where \( e^{(\nu)} \) is the error at the \( \nu^{th} \) iterate. Furthermore, suppose there exists a real sequence \( \{C_\nu\} \) that converges to 0 as \( \nu \to \infty \). Then we say \( \{x^{(\nu)}\} \) converges to \( \bar{x} \).
$q$-superlinearly if

$$|x^{(\nu+1)} - \bar{x}| < C\nu|x^{(\nu)} - \bar{x}|.$$  \hspace{1cm} (1.8)

Equivalently,

$$\lim_{\nu \to \infty} \frac{|x^{(\nu+1)} - \bar{x}|}{|x^{(\nu)} - \bar{x}|} = \lim_{\nu \to \infty} \frac{|e^{(\nu+1)}|}{|e^{(\nu)}|} = 0.$$  \hspace{1cm} (1.9)

We say the sequence converges $q$-superlinearly with at least order $p$ if there exists $C > 0$ such that

$$|e^{(\nu+1)}| < C|e^{(\nu)}|^p,$$

for $\nu \gg 1$. Alternatively this is equivalent to

$$\lim_{\nu \to \infty} \frac{|e^{(\nu+1)}|}{|e^{(\nu)}|^p} = C$$  \hspace{1cm} (1.10)

for some nonzero positive $C$, which is called the asymptotic error constant. If $p = 1$, the method has first order or linear convergence. If $p = 2$, the method has second order or quadratic convergence. If $p \in (1, 2)$, we simply say the method has *superlinear convergence*. The prefix “$q$” stands for quotient and is used to differentiate from “$r$” (root) orders of convergence. The errors of a sequence with $r$-order $p$ are simply bounded above by another sequence of $q$-order $p$.

If the function $f(x)$ is “nice enough” and $x^{(0)}$ is “near enough” to $\bar{x}$, Newton’s Method will have quadratic convergence while the secant method will have superlinear convergence. The terms “nice enough” and “near enough” are more rigorously defined below and proofs are provided for the convergence rate of each method.

**Newton: Quadratic Convergence [2]**

**Definition 1.2.1.** Let $f : S \to \mathbb{R}$ where $S \subset \mathbb{R}$. We say that $f$ is Lipschitz continuous at a point $x \in S$ if there exists a constant $\lambda > 0$ and a $\delta > 0$ such that for all
We say that \( f \) is Lipschitz continuous on a set \( T \subset S \) if there exists a constant \( \lambda > 0 \) such that for all \( x, y \in T \)

\[
|f(x) - f(y)| \leq \lambda |x - y|.
\]

(1.12)

In both cases \( \lambda \) is called a Lipschitz constant for \( f \). Furthermore if \( f \in C^1(S) \), then

\[
\lambda = \max_{\xi \in S} |g'(\xi)|.
\]

**Lemma 1.2.2.** Consider the function \( f : S \to \mathbb{R} \), where \( S \subset \mathbb{R} \). Assume that \( f'(x) \) is Lipschitz continuous on \( S \). For all \( x, y \in S \),

\[
|f(y) - f(x) - f'(x)(y - x)| \leq \frac{1}{2} \lambda |y - x|^2.
\]

(1.13)

**Proof.** Notice that

\[
f(y) - f(x) - f'(x)(y - x) = \int_x^y \left( f'(z) - f'(x) \right) dz.
\]

Then,

\[
|f(y) - f(x) - f'(x)(y - x)| \leq \int_x^y \left| f'(z) - f'(x) \right| dz
\]

\[
\leq \int_x^y \lambda |z - x| dz
\]

\[
= \lambda \int_0^{y-x} |t| dt
\]

\[
= \frac{1}{2} \lambda |y - x|^2,
\]

where the substitution \( t = z - x \) was made in the integral. \( \square \)

**Theorem 1.2.3** (Convergence of Newton’s Method). Let \( f \in C^1(S) \) where \( S \subset \mathbb{R} \) is an open interval with \( f(\bar{x}) = 0 \) for some \( \bar{x} \in S \). Assume that \( f \) is Lipschitz continuous
for all \( x, y \in \mathcal{N}_r(\bar{x}) \subset S \) and for some \( r > 0 \). Also assume that \( |1/f'(x)| \leq \beta \) for all \( x \in \mathcal{N}_r(\bar{x}) \).

Then, there exists an \( \tilde{r} \) that satisfies \( 0 < \tilde{r} < r \) such that if \( x^{(0)} \in \mathcal{N}_{\tilde{r}}(\bar{x}) \) then the sequence \( \{x^{(\nu)}\} \) generated by

\[
x^{(\nu + 1)} = x^{(\nu)} - \frac{f(x^{(\nu)})}{f'(x^{(\nu)})}
\]

exists, lies in \( \mathcal{N}_{\tilde{r}}(\bar{x}) \) and converges to \( \bar{x} \). Furthermore,

\[
|e^{(\nu + 1)}| \leq \frac{1}{2} \lambda \beta |e^{(\nu)}|^2 .
\] (1.14)

Proof. First, we have

\[
x^{(\nu + 1)} - \bar{x} = x^{(\nu)} - \bar{x} - \frac{f(x^{(\nu)})}{f'(x^{(\nu)})}
= x^{(\nu)} - \bar{x} - \frac{f(x^{(\nu)}) - f(\bar{x})}{f'(x^{(\nu)})}
= \frac{f(\bar{x}) - f(x^{(\nu)}) - f'(x^{(\nu)})(\bar{x} - x^{(\nu)})}{f'(x^{(\nu)})}
\] (1.15)

Therefore from Lemma 1.2.2, we immediately have

\[
|x^{(\nu + 1)} - \bar{x}| \leq \frac{1}{2} \frac{\lambda}{|f'(x^{(\nu)})|} |x^{(\nu)} - \bar{x}|^2
\leq \frac{1}{2} \lambda \beta |x^{(\nu)} - \bar{x}|^2 .
\]

\[\square\]

Secant: Superlinear Convergence [2]

Unfortunately, proving the secant method has superlinear convergence is not as simple. For simplicity, let \( f \in C^2(S) \) and \( f'(\bar{x}) \neq 0 \), so that the zero is simple. Furthermore, denote

\[
F[a, b] = \frac{f(b) - f(a)}{b - a} .
\]
With this notation, the secant method (1.5) can be rewritten as

$$x^{(\nu+1)} = x^{(\nu)} - \frac{f(x^{(\nu)})}{F[x^{(\nu-1)}, x^{(\nu)}]}.$$

(1.16)

Thus,

$$e^{(\nu+1)} = x^{(\nu+1)} - \bar{x} = x^{(\nu)} - \frac{f(x^{(\nu)})}{F[x^{(\nu-1)}, x^{(\nu)}]} - \bar{x} = e^{(\nu)} - \frac{f(x^{(\nu)})}{F[x^{(\nu-1)}, x^{(\nu)}]} \cdot$$

(1.17)

However, we need an expression that only contains the errors $e^{(\nu+1)}$ and $e^{(\nu)}$. From (1.17),

$$e^{(\nu+1)} = \frac{e^{(\nu)}}{F[x^{(\nu-1)}, x^{(\nu)}]} \left( F[x^{(\nu-1)}, x^{(\nu)}] - \frac{f(x^{(\nu)})}{e^{(\nu)}} \right) = \frac{e^{(\nu)}}{F[x^{(\nu-1)}, x^{(\nu)}]} G_{\nu}(x^{(\nu-1)}),$$

(1.18)

where

$$G_{\nu}(x) = \frac{f(x^{(\nu)}) - f(x)}{x^{(\nu)} - x} - \frac{f(x^{(\nu)})}{x^{(\nu)} - \bar{x}}.$$  

Notice that

$$G_{\nu}(\bar{x}) = \frac{f(x^{(\nu)}) - f(\bar{x}) - f(x^{(\nu)})}{x^{(\nu)} - \bar{x}} = \frac{-f(\bar{x})}{x^{(\nu)} - \bar{x}} = 0.$$  

Then by the MVT, there exists a $\xi_{\nu}$ between $x^{(\nu-1)}$ and $\bar{x}$ such that

$$G'_{\nu}(\xi_{\nu}) = \frac{G_{\nu}(x^{(\nu-1)}) - G_{\nu}(\bar{x})}{x^{(\nu-1)} - \bar{x}}.$$  

Combining this with the fact that $G_{\nu}(\bar{x}) = 0$, this is equivalent to

$$G_{\nu}(x^{(\nu-1)}) = G'_{\nu}(\xi_{\nu})(x^{(\nu-1)} - \bar{x}) = G'_{\nu}(\xi_{\nu}) e^{(\nu-1)}.$$  

(1.19)

Similarly, there also exists a $\zeta_{\nu}$ between $x^{(\nu-1)}$ and $x^{(\nu)}$ such that

$$f'_{\nu}(\zeta_{\nu}) = \frac{f(x^{(\nu)}) - f(x^{(\nu-1)})}{x^{(\nu)} - x^{(\nu-1)}} = F[x^{(\nu-1)}, x^{(\nu)}].$$  

(1.20)
Substituting (1.19) and (1.20) into (1.18) gives

\[ e^{(\nu+1)} = e^{(\nu)} e^{(\nu-1)} \frac{G'_\nu(\xi_\nu)}{f'(\xi_\nu)}. \] (1.21)

In order to determine the asymptotic error constant, we need \( \nu \gg 1 \), or

\[ \frac{G'_\nu(\xi_\nu)}{f'(\xi_\nu)} \approx \frac{G'_\nu(\bar{x})}{f'(\bar{x})} = R. \]

Since \( f \in C^2(S) \), for \( x^{(\nu)}, \bar{x} \in S \),

\[ f(x^{(\nu)}) \approx f(\bar{x}) + f'(\bar{x})(x^{(\nu)} - \bar{x}) + \frac{1}{2} f''(\bar{x})(x^{(\nu)} - \bar{x})^2. \]

We have \( f(\bar{x}) = 0 \) and \( f'(\bar{x}) \neq 0 \) so

\[ \frac{1}{2} f''(\bar{x}) \approx \frac{f(x^{(\nu)}) - f'(\bar{x})(x^{(\nu)} - \bar{x})}{(x^{(\nu)} - \bar{x})^2} = G'_\nu(\bar{x}). \]

Therefore,

\[ R = \frac{1}{2} \left( \frac{f''(\bar{x})}{f'(\bar{x})} \right). \] (1.22)

Returning to (1.21),

\[ e^{(\nu+1)} \approx R e^{(\nu)} e^{(\nu-1)} \]

\[ \implies R e^{(\nu+1)} \approx (R e^{(\nu)}) (R e^{(\nu-1)}). \] (1.23)

If we let \( y_\nu = \log |R e^{(\nu)}| = \log R |e^{(\nu)}| \), then (1.23) becomes the Fibonacci recursive formula

\[ y_{\nu+1} \approx y_\nu + y_{\nu-1}. \] (1.24)

The asymptotic formula can be shown to be \( y_{\nu+1} \approx r y_\nu \), where \( r \) is the golden ratio \( \phi \approx 1.618 \). Then,

\[ R |e^{(\nu+1)}| \approx (R |e^{(\nu)}|)^r \]

\[ \implies |e^{(\nu+1)}| \approx R^{r-1} |e^{(\nu)}|^r. \] (1.25)

Since \( r \approx 1.618 \in (1, 2) \), the secant method converges superlinearly.
1.2.5 A Global Modification: Line Search

Both Newton’s Method and the secant method as presented are local methods. This means that given a “nice enough” function \( f(x) \) and an initial guess that is “near enough” to \( \bar{x} \), the iterates are guaranteed to converge. Unfortunately, “near enough” is usually problem dependent, so it can be hard to determine in advance if a randomly selected initial point satisfies this criteria.

A global method will not guarantee that any initial point \( x^{(0)} \) converges to \( \bar{x} \), but just that the number will be greater than for a local method. In one dimension, global methods are easy to handle because there are only two directions for the iterate to move: either to the right or the left. In only one direction will \(|f(x)|\) decrease. All that is left to be determined is how big of a step the iteration takes.

First, denote the Newton step

\[
\Delta x^{(\nu)} = -\frac{f(x^{(\nu)})}{f'(x^{(\nu)})},
\] (1.26)

Notice that this quantity is the distance (and direction) along the \( x \)-axis from \( x^{(\nu)} \) to \( x^{(\nu+1)} \). Next define

\[
x^{(\nu+1)[\lambda]} \equiv x^{(\nu)} + \lambda \Delta x^{(\nu)},
\] (1.27)

a function of \( \lambda \) where \( \lambda \in (0, 1] \). Newton’s Method is simply \( x^{(\nu+1)} = x^{(\nu+1)[1]} \). Because we would like the iterates to decrease, we have only one condition:

\[
|f(x^{(\nu+1)[\lambda]})| < |f(x^{(\nu)})|.
\] (1.28)

If the new iterate is better, it is used. However, if \(|f(x^{(\nu+1)[\lambda]})| > |f(x^{(\nu)})|\) we halve the step-length by setting \( \lambda = \frac{1}{2} \lambda \) and check again. This process can be repeated until (1.28) is satisfied. It is advisable to set a minimum \( \lambda \) or the method can converge too slowly if \( \lambda \ll 1 \).
One final thing to consider when using a global method is the case when \( x^{(\nu)} \) is close to a local minimum. The Newton step \( \Delta x^{(\nu)} \) can be very large in this case since \( f'(x^{(\nu)}) \) will be close to 0. It is a good idea to have some limit on how large \( |\Delta x^{(\nu)}| \) is allowed to be.

Unfortunately, global secant methods can fail since an approximation to \( f'(x) \) is being used and we are no longer guaranteed that \( -f(x^{(\nu)})/a_{\nu} \) is the descent direction. For example, consider

\[
 f(x) = x^2 - 4x + 3.51, \tag{1.29}
\]

which has zeros at \( x = 1.3 \) and \( x = 2.7 \), and a minimum at \( x = 2 \). If we choose \( x^{(-1)} = 0.5 \) and \( x^{(0)} = 3 \), then \( f(x^{(-1)}) = 1.76 \), \( f(x^{(0)}) = 0.51 \) and \( x^{(1)}[\lambda] = 3 + 1.02\lambda \). However, \( f(x) \) is increasing for \( x > 2 \), so there is no \( \lambda \) for which \( |f(x^{(\nu+1)}[\lambda])| < |f(x^{(\nu)})| \) and the line-search method will fail. The problem is that the average slope of \( f(x) \) between \( x^{(-1)} \) and \( x^{(0)} \) is \(-0.5\), while \( f'(x^{(0)}) = 8 \), so the line-search method is searching in the wrong direction.

### 1.3 N-Dimensional Methods

With the foundation of the one-dimensional methods in place, the jump can be made to the \( n \)-dimensional methods. First we will define matrix norms, and their role in the \( n \)-dimensional methods being discussed.

#### 1.3.1 Linear Algebra Review: Matrix Norms

The analogs of the absolute value function \( |\cdot| \) in one dimension are vector and matrix norms in \( n \) dimensions. Even in vector and matrix spaces, there needs to be a way to measure the “size” of elements from these spaces. Different norms may be
acceptable or appropriate depending on the setting, so general properties of all norms will be considered as opposed to specific ones.

**Definition 1.3.1** (Matrix Norm). A function $||·|| : \mathbb{R}^{n\times n} \rightarrow \mathbb{R}$ is a matrix norm if for all $A, B \in \mathbb{R}^{n\times n}$ it satisfies the following four axioms:

1. $||A|| \geq 0$

1a) $||A|| = 0 \iff A = 0$

2. $||cA|| = |c| \cdot ||A||$

3. $||A + B|| \leq ||A|| + ||B||$

Additionally, some norms also satisfy the *submultiplicative property*, namely

$$||AB|| \leq ||A|| \cdot ||B||.$$  \hspace{1cm} (1.30)

For example, if the matrix norm defined $||A||_{\triangle} = \max |a_{ij}|$ is applied to

$$A = B = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

by direct calculation we find $||AB||_{\triangle} > ||A||_{\triangle}||B||_{\triangle}$ [6], which doesn’t satisfy the submultiplicative property. In this thesis, we will only use norms that satisfy this property.

**Remark.** There are some instances in the literature where a matrix norm is denoted $|||·|||$ while a vector norm simply $||·||$. For simplicity, we use the latter to denote both. The reader should be able to easily distinguish between the two from the context.
\(\ell_2\) Matrix Norm

The \(\ell_2\) matrix norm is an induced norm, meaning it relies on the definition of the \(\|\cdot\|_2\) vector norm. It is defined as

\[
\|A\|_2 = \max_{\|x\|_2 = 1} \|Ax\|_2.
\]

(1.31)

Frobenius Matrix Norm

The Frobenius matrix norm is an example of an “entry-wise” norm. It is equivalent to the \(\|\cdot\|_2\) vector norm of the vectorized matrix \(A\). There are a couple of different definitions, including

\[
\|A\|_F = \left( \sum_j \sum_i |a_{ij}|^2 \right)^{1/2} = \sqrt{\text{trace}(A^TA)}.
\]

(1.32)

Even though this is not an induced norm, it still satisfies the submultiplicative property.

1.3.2 Newton’s Method

Now consider \(f : \mathbb{R}^n \rightarrow \mathbb{R}^n\). Using a similar notation as Section 1.2 we still have \(f(\bar{x}) = 0\). By Taylor’s expansion, for \(i = 1, \ldots, n\),

\[
f_i(x + \delta x) = f_i(x) + \sum_{j=1}^n \frac{\partial f_i(x)}{\partial x_j} \delta x_j + \mathcal{O}(\delta x^2)
\]

\[
\approx f_i(x) + \sum_{j=1}^n \frac{\partial f_i(x)}{\partial x_j} \delta x_j,
\]

(1.33)

with the approximation holding if \(f(x)\) is nonlinear. Denote the affine approximation to \(f(x)\) at \(x^{(\nu)}\) as

\[
p^{(\nu)}(x) = f(x^{(\nu)}) + J(x^{(\nu)})(x - x^{(\nu)}),
\]
where $\mathcal{J}(\mathbf{x})$ denotes the Jacobian of the nonlinear system $\mathbf{f}(\mathbf{x})$, namely

$$
\mathcal{J}(\mathbf{x}) = \begin{pmatrix}
\frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\
\vdots & \ddots & \vdots \\
\frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n}
\end{pmatrix}.
$$

Then, (1.33) can be written in vector equation form as

$$
\mathbf{p}(\nu)(\mathbf{x}(\nu) + \delta \mathbf{x}(\nu)) = \mathbf{f}(\mathbf{x}(\nu)) + \mathcal{J}(\mathbf{x}(\nu))\delta \mathbf{x}(\nu).
$$

(1.34)

Similar to the one-dimensional method, we denote $\mathbf{x}(\nu+1) = \mathbf{x}(\nu) + \delta \mathbf{x}(\nu)$, such that $\mathbf{p}(\nu)(\mathbf{x}(\nu+1)) = \mathbf{0}$. Solving for $\delta \mathbf{x}(\nu)$, we have

$$
\delta \mathbf{x}(\nu) = -\mathcal{J}(\mathbf{x}(\nu))^{-1}\mathbf{f}(\mathbf{x}(\nu)),
$$

(1.35)

leading to the general iteration for the $n$-dimensional Newton’s Method

$$
\mathbf{x}(\nu+1) = \mathbf{x}(\nu) - \mathcal{J}(\mathbf{x}(\nu))^{-1}\mathbf{f}(\mathbf{x}(\nu)),
$$

(1.36)

for $\nu \in \mathbb{N}[0, \infty)$. If $\mathbf{x}(0)$ is “near enough” to $\bar{\mathbf{x}}$ and $\mathcal{J}(\bar{\mathbf{x}})$ is nonsingular, the sequence of iterates $\{\mathbf{x}(\nu)\}$ will also converge quadratically to $\bar{\mathbf{x}}$. This is also a local method and will be no better at converging from a random initial point as the one-dimensional method was. For an affine system $\mathbf{f}(\mathbf{x})$, the Newton’s $n$-dimensional method will converge in one iteration.

There are several problems with this method. First, $\mathcal{J}(\mathbf{x}(\nu))$ is required at every iteration. Due to the number of function evaluations, this can be very expensive. In addition, $\mathcal{J}(\mathbf{x})$ could be very difficult to calculate analytically, similar to the calculation of $f'(x)$ in the one-dimensional method. This problem can be eased by using a finite-difference approximation of $\mathcal{J}(\mathbf{x})$. Secondly, at each step, the method relies on the solution of the linear system of equations $\mathcal{J}(\mathbf{x}(\nu))\delta \mathbf{x}(\nu) = \mathbf{f}(\mathbf{x}(\nu))$. If $\mathcal{J}(\mathbf{x}(\nu))$ is singular or ill-conditioned, the solution can be wildly inaccurate.
1.3.3 Fixed Chord Method

Newton’s Method requires the solution of the matrix system \( \mathcal{J}(x^{(v)}) \delta x^{(v)} = -f(x^{(v)}) \) at each iteration. For a full matrix, the LU factorization will cost \( \sim \frac{2}{3}n^3 \) flops, which can be expensive for large \( n \), in addition to the evaluation of \( \mathcal{J}(x^{(v)}) \).

If the fixed chord method is used, at each step \( \mathcal{J}(x^{(v)}) \) becomes \( \mathcal{J}(x^{(0)}) \). After the initial LU decomposition, each iteration will only require \( \sim 2n^2 \) flops to calculate the Newton step \( \delta x^{(v)} \).

Similar to the method in one dimension, the convergence near the zero will only be linear. However far from the zero, most methods usually have linear convergence. If the Jacobian has a ”nice” structure, the cost of LU factorization can be reduced and the fixed chord method may not be required. Often, the choice on whether to use this method is problem dependent and we do not discuss it further here.

1.3.4 Broyden’s Method

The main alternative to Newton’s Method in \( n \) dimensions is Broyden’s secant method. We start with the \( n \)-dimensional secant equation

\[
\mathcal{J}_{\nu+1}(x^{(\nu+1)} - x^{(\nu)}) = f(x^{(\nu+1)}) - f(x^{(\nu)}).
\]  

For ease of notation, we will denote \( \Delta x^{(\nu)} = x^{(\nu+1)} - x^{(\nu)} \) and \( \Delta f^{(\nu)} = f(x^{(\nu+1)}) - f(x^{(\nu)}) \). Unfortunately, when \( n > 1 \), (1.37) is an underdetermined system and if \( \Delta x^{(\nu)} \neq 0 \), there is an \( n(n-1) \) affine subspace of matrices that satisfy the secant equation. In order to construct a successful secant approximation, we need to select a good way to choose from these possible matrices. Ideally we would like to enhance the Jacobian approximation properties of \( \mathcal{J}_{\nu+1} \) or facilitate its use in a quasi-Newton algorithm.
Strategy #1: Interpolation

An obvious choice is to require the affine model
\[
p^{(\nu+1)}(x) = f(x^{(\nu+1)}) + J_{\nu+1}(x - x^{(\nu+1)})
\]  
(1.38)
to interpolate \(f(x)\) at the past points \(x^{(0)}, x^{(1)}, \ldots, x^{(\nu)}\). This leads to the system of equations
\[
J_{\nu+1} \Delta x^{(k)} = \Delta f^{(k)}, \quad k \in \mathbb{N}[0, \nu],
\]
where \(\Delta x^{(k)} = x^{(\nu+1)} - x^{(k)}\) and \(\Delta f^{(k)} = f(x^{(\nu+1)}) - f^{(k)}\). Unfortunately, the directions \(\Delta x^{(\nu)}, \Delta x^{(\nu-1)}, \ldots, \Delta x^{(0)}\) tend to be linearly dependent or very close to it, so the computation of \(J_{\nu+1}\) is an ill-posed problem.

Strategy #2: Minimize Change in Affine Model

Aside from the secant equation, we have no new information about either the Jacobian or the model. In order to preserve as much as possible of what we already have, we choose \(J_{\nu+1}\) by trying to minimize the change in the affine model (1.38) subject to satisfying \(J_{\nu+1} \Delta x^{(\nu)} = \Delta f^{(\nu)}\).

The difference between the new and old affine models at any \(x \in \mathbb{R}^n\) is
\[
p^{(\nu+1)}(x) - p^{(\nu)}(x) = f(x^{(\nu+1)}) + J_{\nu+1}(x - x^{(\nu+1)}) - f(x^{(\nu)}) - J_{\nu}(x - x^{(\nu)})
\]
\[
= (J_{\nu+1} - J_{\nu})(x - x^{(\nu)})
\]
For any \(x \in \mathbb{R}^n\), define the unique vector \(x - x^{(\nu)} = \alpha \Delta x^{(\nu)} + t\), where \(t^T \Delta x^{(\nu)} = 0\). Thus, the goal is to minimize
\[
p^{(\nu+1)}(x) - p^{(\nu)}(x) = \alpha(J_{\nu+1} - J_{\nu}) \Delta x^{(\nu)} + (J_{\nu+1} - J_{\nu})t.
\]  
(1.39)
There is no control over the first term since \((J_{\nu+1} - J_{\nu}) \Delta x^{(\nu)} = \Delta f^{(\nu)} - J_{\nu} \Delta x^{(\nu)}\). However, the second term can be made zero for all \(x \in \mathbb{R}^n\) by choosing \(J_{\nu+1}\) such
that \((J_{\nu+1} - J_{\nu})t = 0\) for all \(t\) orthogonal to \(\Delta x^{(\nu)}\). Therefore, \(J_{\nu+1} - J_{\nu}\) will be a rank one matrix of the form \(u \Delta x^{(\nu)}^T\), for any \(u \in \mathbb{R}^n\). This leads to

\[
u = \frac{(\Delta f^{(\nu)} - J_{\nu} \Delta x^{(\nu)})}{||\Delta x^{(\nu)}||^2_2}
\]

which means that

\[
J_{\nu+1} = J_{\nu} + \frac{(\Delta f^{(\nu)} - J_{\nu} \Delta x^{(\nu)}) \Delta x^{(\nu)}^T}{||\Delta x^{(\nu)}||^2_2}
\]  

is the least change in the affine model consistent with \(J_{\nu+1} \Delta x^{(\nu)} = \Delta f^{(\nu)}\). In fact, Broyden’s update is the minimum change to \(J_{\nu}\) consistent with \(J_{\nu+1} \Delta x^{(\nu)} = \Delta f^{(\nu)}\), if the change \(J_{\nu+1} - J_{\nu}\) is measured with the Frobenius norm. The Frobenius norm is an appropriate choice since it measures the change in each component of the Jacobian approximation \(J_{\nu}\).

The matrix update given by (1.40) is commonly referred to as Broyden’s “good” update. Broyden derived another formulation that eliminates the need for a system of linear equations by updating \(J_{\nu}^{-1}\) directly.

**Lemma 1.3.2** (Sherman-Morrison formula). Let \(A \in \mathbb{R}^{n \times n}\) be an invertible matrix and let \(u, v \in \mathbb{R}^n\) such that \(v^T A u \neq 0\). Then

\[
(A + uv^T)^{-1} = A^{-1} + \frac{A^{-1}uv^TA^{-1}}{1 + v^TA^{-1}u}.
\]

Applying Lemma 1.3.2 directly to (1.40) gives Broyden’s “bad” update

\[
J_{\nu+1}^{-1} = J_{\nu}^{-1} + \frac{(\Delta x^{(\nu)} - J_{\nu}^{-1} \Delta f^{(\nu)}) \Delta f^{(\nu)}^T}{||\Delta f^{(\nu)}||^2_2}
\]  

for \(\nu \in \mathbb{N}[0, \infty)\).

**1.3.5 Convergence**

The definition of convergence that we use in the \(n\)-dimensional methods is the direct analog of the one-dimensional definition laid out in Section 1.2.4 using the
appropriate vectorized counterparts. The following lemma and theorem will be useful in the proof of the convergence of the $n$-dimensional Newton’s Method.

**Theorem 1.3.3.** Let $\| \cdot \|$ be any norm on $\mathbb{R}^{n \times n}$ that obeys the submultiplicative property and $\|I\| = 1$ and let $E \in \mathbb{R}^{n \times n}$. If $\|E\| < 1$, then $(I - E)^{-1}$ exists and

$$
\|(I - E)^{-1}\| \leq \frac{1}{1 - \|E\|}. \quad (1.42)
$$

If $A$ is nonsingular and $\|A^{-1}(B - A)\| < 1$, then $B$ is nonsingular and

$$
\|B^{-1}\| \leq \frac{\|A^{-1}\|}{1 - \|A^{-1}(B - A)\|}. \quad (1.43)
$$

**Proof.** First, we show

$$(I - E)^{-1} = \sum_{k=0}^{\infty} E^k. \quad (1.44)$$

Since

$$(I - E) \sum_{k=0}^{n} E^k = \sum_{k=0}^{n} E^k - \sum_{k=0}^{n} E^{k+1}$$

$$= I - E^{k+1}$$

and $\|E^{k+1}\| \leq \|E\|^{k+1} \to 0$ as $n \to \infty$, we must have $E^{k+1} \to 0$ as $n \to \infty$. So as $n \to \infty$,

$$(I - E) \sum_{k=0}^{\infty} E^k = I.$$

This proves (1.44). Therefore,

$$
\|(I - E)^{-1}\| = \left\| \sum_{k=0}^{\infty} E^k \right\| \leq \sum_{k=0}^{\infty} \|E^k\|$$

$$\leq \sum_{k=0}^{\infty} \|E\|^k = \frac{1}{1 - \|E\|}.$$

Equation (1.43) is a direct application of the first part of this theorem. \qed
Lemma 1.3.4. Let \( f : \mathbb{R}^n \to \mathbb{R}^m \) be continuously differentiable in the open convex set \( S \subset \mathbb{R}^n \), \( x \in S \), and let \( J \) be Lipschitz continuous at \( x \) in the neighborhood \( S \), using a vector norm and the induced matrix operator norm and the constant \( \gamma \). Then for any \( x + p \in S \),

\[
\| f(x + p) - f(x) - J(x)p \| \leq \gamma \frac{1}{2} \| p \|^2. \tag{1.45}
\]

Newton: Quadratic Convergence

Theorem 1.3.5. Let \( f : \mathbb{R}^n \to \mathbb{R}^n \) be continuously differentiable in an open convex set \( S \subset \mathbb{R}^n \). Assume there exists \( \bar{x} \in \mathbb{R}^n \) and constants \( r, \beta > 0 \) such that \( N_r(\bar{x}) \subset S \), \( f(\bar{x}) = 0 \), \( J(\bar{x})^{-1} \) exists with \( \| J(\bar{x})^{-1} \| \leq \beta \) and \( J(\bar{x}) \) Lipschitz continuous on \( N_r(\bar{x}) \) with Lipschitz constant \( \gamma \).

Then there exists \( \varepsilon > 0 \) such that for all \( x^{(0)} \in N_\varepsilon(\bar{x}) \), the sequence \( x^{(1)}, x^{(2)}, \ldots \) generated by

\[
x^{(\nu+1)} = x^{(\nu)} - J(x^{(\nu)})^{-1} f(x^{(\nu)}), \quad \nu \in \mathbb{N}[0, \infty)
\]

is well-defined and converges to \( \bar{x} \), and obeys

\[
\| x^{(\nu+1)} - \bar{x} \| \leq \beta \gamma \| x^{(\nu)} - \bar{x} \|^2 \tag{1.46}
\]

for \( \nu \in \mathbb{N}[0, \infty) \).

Proof. Let

\[
\varepsilon = \max \left\{ r, \frac{1}{2\beta \gamma} \right\}. \tag{1.47}
\]

We show by induction on \( \nu \) that at each step (1.46) holds, and also that

\[
\| x^{(\nu+1)} - \bar{x} \| \leq \frac{1}{2} \| x^{(\nu)} - \bar{x} \|,
\]

which implies that

\[
x^{(\nu+1)} \in N_\varepsilon(\bar{x}). \tag{1.48}
\]
First we show that \( J(x^{(0)}) \) is nonsingular. From \( \|x^{(0)} - \bar{x}\| \leq \varepsilon \), the Lipschitz continuity of \( J \) at \( \bar{x} \), and (1.47), it follows that
\[
\|J(\bar{x})^{-1}[J(x^{(0)}) - J(\bar{x})]\| \leq \|J(\bar{x})^{-1}\| \cdot \|J(x^{(0)}) - J(\bar{x})\| \\
\leq \beta \gamma \|x^{(0)} - \bar{x}\| \leq \beta \gamma \varepsilon \leq \frac{1}{2}.
\]
Thus by (1.43), \( J(x^{(0)}) \) is nonsingular and
\[
\|J(x^{(0)})^{-1}\| \leq \frac{\|J(\bar{x})^{-1}\|}{1 - \|J(\bar{x})^{-1}[J(x^{(0)}) - J(\bar{x})]\|} \\
\leq 2\|J(\bar{x})^{-1}\| \leq 2\beta.
\tag{1.49}
\]
Therefore \( x^{(1)} \) is well-defined and
\[
\begin{align*}
x^{(1)} - \bar{x} &= x^{(0)} - \bar{x} - J(x^{(0)})^{-1}f(x^{(0)}) \\
&= x^{(0)} - \bar{x} - J(x^{(0)})^{-1}[f(x^{(0)}) - f(\bar{x})] \\
&= J(x^{(0)})^{-1}[f(\bar{x}) - f(x^{(0)}) - J(x^{(0)})(\bar{x} - x^{(0)})].
\end{align*}
\]
By Lemma 1.3.4 and (1.49),
\[
\|x^{(1)} - \bar{x}\| \leq \|J(x^{(0)})^{-1}\| \cdot \|f(\bar{x}) - f(x^{(0)}) - J(x^{(0)})(\bar{x} - x^{(0)})\| \\
\leq 2\beta \cdot \frac{\gamma}{2} \|x^{(0)} - \bar{x}\|^2 = \beta \gamma \|x^{(0)} - \bar{x}\|^2,
\]
and (1.46) is proved. Since \( \|x^{(0)} - \bar{x}\| \leq \frac{1}{2\beta \gamma}, \|x^{(1)} - \bar{x}\| \leq \frac{1}{2} \|x^{(0)} - \bar{x}\| \), which shows (1.48) and completes the case for \( \nu = 0 \). The proof of the induction step follows identically.

**Secant: Superlinear Convergence**

The proof for the convergence of the \( n \)-dimensional secant method is much more lengthy than the proof for Newton’s Method. It is fully contained in Appendix A.
1.3.6 Global Modifications

Just as in the one-dimensional case, the easiest global modification to add to these three local $n$-dimensional methods is line-search. Now that there are an infinite number of directions for the iterates to move in, there can be many directions that the function decreases. However, we will continue to use the Newton step direction 
\[ \Delta x^{(\nu)} = -J(x^{(\nu)})^{-1}f(x^{(\nu)}) \] as the direction of descent. The rest of the process is exactly the same since we evaluate the function \[ \|f(x)\|_2 \] along the one-dimensional line \[ x^{(\nu)} + \lambda \Delta x^{(\nu)}. \]

As before, we only require that 
\[ \|f(x^{(\nu+1)}[\lambda])\| < \|f(x^{(\nu)})\|, \]
where \[ x^{(\nu+1)}[\lambda] \equiv x^{(\nu)} + \lambda \Delta x^{(\nu)}. \] We also impose a minimum value of \( \lambda \) so the iterates do not become stuck. Since \( J(x) \) is almost singular near a local minimum it is advisable to require a maximum step size \[ \|\Delta x^{(\nu)}\|_2 \] to keep the iterates from shooting very far away from the region of interest.

As discussed in Section 1.2.5, global secant methods can fail and the $n$-dimensional case is no different. Broyden’s method no longer guarantees that 
\[ -J^{-1}_\nu f(x^{(\nu)}) \] is a descent direction for the iterates and the global method can easily fail.

1.4 Numerical Methods

There are two fundamentally different types of convergence plots in this thesis. The majority of the plots show how fast an initial point converges and whether or not it will converge, without regard to which zero it converges to. However, there are a few plots featured in the concluding discussion in Chapter 6 that show not only how fast an initial point converges but also to which zero it converges.
The initial steps in each code are identical. The area of interest of the system is first discretized into a regular grid with the resolutions $\Delta x_1$ and $\Delta x_2$, corresponding to the points along $x_1$-axis and $x_2$-axis, respectively. The MATLAB function `meshgrid` is utilized to create two matrices containing all the $x_1$ and $x_2$ points. The specific details of each method are discussed in Section 1.4.1 and Section 1.4.2.

Each method runs for a specified number of iterations (max_iter) although Method #1 can exit a run if a stopping criterion is met. Success is achieved for a particular initial point if $\|f(x^{(\nu)})\|_2$ is less than the specified tolerance (tol) for some particular iterate $x^{(\nu)}$, with $\nu \in \mathbb{N}[0, \text{max}_\text{iter}]$. The details of how each convergence plot is graphically presented are also described in the following sections.

1.4.1 Code # 1: Overall Convergence vs Nonconvergence

Two additional matrices of the same dimensions as the discretized area of interest are initialized: `iter_matrix` and `cr_matrix`. The former stores the number of iterations that the corresponding initial point in the mesh needed to converge to a zero. The latter matrix stores a value of $-1$, $0$, or $1$. Depending on how a method is terminated, a value of $-1$, $0$, or $1$ is stored in the feedback variable `result`. The method can fail to converge for a particular initial point $x^{(0)}$ in the following two scenarios:

1. In order to maintain some level of accuracy for a given iterate, the condition number of the Jacobian $J(x^{(\nu)})$ or the Jacobian approximation $J_\nu$ is not allowed to exceed a prescribed value (max_cond). If such an event occurs, the method exits.
2. The method fails to converge within max_iter iterations but max_cond is never exceeded.

Since result is initialized to 0, after max_iter iterations only corresponding initial points that failed to converge by scenario two will have this value. A value of −1 indicates max_cond was exceeded and the method terminated while a value of 1 indicates success.

Each method was run to termination or completion on each point in the mesh and the results were saved in the two matrices described above. Before plotting, all entries in iter_matrix that had a value of −1 in the corresponding entry in iter_matrix were set to zero (since convergence was not achieved). The iteration matrix was then plotted using the MATLAB function imagesc, which scales the values of the matrix to an HSV colormap of 800 colors, as shown in the colorbars that accompany the plots. A black square indicates a failure to converge (via scenario #1) while a white square indicates the initial point failed to converge (via scenario #2).

1.4.2 Code # 2: Individual Convergence Regions

The code described in the previous section lacks efficiency since many of the discretized areas of interest contained several million initial points. In order to significantly decrease the run time and allow individual convergence regions corresponding to each zero to be plotted, another code was developed.

Instead of running the method over and over again for each initial point in the discretized domain, this code runs the method one iteration at a time over all the points in the domain. In order to facilitate this process, the matrices corresponding to the $x_1$ and $x_2$ elements of the initial points were vectorized. To calculate the current iterate,
each method requires the previous iterate, the value of the function at the previous iterate, and the Jacobian at the previous iterate (or previous Jacobian approximation). Since the Jacobian is of size $2 \times 2$, it can easily be stored as a $4 \times 1$ row vector. All the information is thus stored in a matrix of size $(\text{nr}_x\text{pts})(\text{nr}_y\text{pts}) \times 8$, where $\text{nr}_x\text{pts}$ and $\text{nr}_y\text{pts}$ are the number of discrete points along the $x_1$ and $x_2$ axes, respectively. Thus, each row of this matrix contains the $x_1$ and $x_2$ coordinates of the current and previous iterate, and the function and entries of the Jacobian evaluated at the previous iterate.

The methods as described in Section 1.3.2 and Section 1.3.4 are manually coded, including the calculation of $J^{-1}$ in Newton’s method and $J^{-1}_\nu$ in Broyden’s “good” method. For each system, the $x_1$ and $x_2$ coordinates of the zeros are defined in two separate vectors. At each iteration, and for each zero, the code checks which points have converged, i.e., which points are within the given tolerance ($\text{tol}$) of a particular zero. Those points are then plotted with a color corresponding to the zero and an intensity corresponding to the number of iterates that the initial point required to converge. These points are then deleted from the master matrix and the code advances to the next iteration.

This code does not put a limit on how large the condition number of the Jacobian (or approximation to the Jacobian) can be, so the convergence results may be slightly better than calculated using the previous code. However, we can gain more insight into the performance of the zero-finding methods by observing how the individual convergence regions are affected. Additionally, this code is significantly faster, since as points converge there are less to check and the speed of a run increases with each new iteration.
1.4.3 Global Newton’s Method

Newton’s global method is utilized for several plots. The local method is easily extended to a global method using the two additional requirements described in Section 1.3.6. For our plots, the maximum step size $\text{max\_step}$ was dependent on the system we were exploring. However, to prevent slow convergence, we set the line-search step length ($\lambda_{\text{min}}$) to $1/32$. If the condition $\|f(x^{(\nu+1)}[\lambda])\| < \|f(x^{(\nu)})\|$ is still not satisfied, the iterate is accepted anyway and the method proceeds to the next iteration.

As with the one-dimensional Newton’s method, the method stops if $\|f(x^{(\nu)})\|_2$ is less than the specified tolerance $\text{tol}$, $\text{max\_cond}$ is exceeded, or the number of iterations reaches $\text{max\_iter}$.
Chapter 2: Freudenstein & Roth’s Function

2.1 The Function

2.1.1 Nonlinear System

The first function that Newton and Broyden’s methods will be tested on is a $2 \times 2$ polynomial system that is nonlinear in only one variable. The function was taken from [12] and the name references the original source.

$$f_1(x) = -13 + x_1 + ((5 - x_2)x_2 - 2)x_2$$
$$f_2(x) = -29 + x_1 + ((1 + x_2)x_2 - 14)x_2$$  \hspace{1cm} (2.1)$$

This system has a local minimum of 7 near $(1.141, -0.89)$ and a zero at $(5, 4)$. The Jacobian of the nonlinear system is

$$J(x) = \begin{pmatrix} 1 & -3x_2^2 + 10x_2 - 2 \\ 1 & 3x_2^2 + 2x_2 - 14 \end{pmatrix}.$$  \hspace{1cm} (2.2)$$

The Jacobian of a square matrix is singular if and only if its determinant is equal to zero. Setting $\det(J(x)) = 0$ gives the two singular lines

$$x_2 = \frac{1}{3}(2 \pm \sqrt{22}).$$  \hspace{1cm} (2.3)$$
2.1.2 Contour Plot

![Contour Plot](image)

Figure 2.1: Contour plot of Equation 2.1. The zero at (5, 4) is marked by the * and the local minimum of approximately 7 around (11.41, -0.897) is marked by the +. The dotted lines show where the Jacobian is singular. The contours immediately outside the zero and minimum are shown and the contours increase by 5. Notice the different scales of the $x_1$ and $x_2$ axes.

2.2 Local Convergence Regions

To investigate the effect of each method on the solution of the system, convergence plots were generated over the region of interest shown in Figure 2.1. The color bar denotes the number of iterations it took for each initial point to converge (within the maximum number of iterations). A black point indicates the method failed for the respective initial point due to the condition number of the Jacobian exceeding
the maximum limit set. Convergence was achieved when \( \|f(x)\|_2 \) was less than the specified tolerance.

When Newton’s method is applied to certain polynomials in the complex plane, the boundaries between convergence regions form a fractal. When we zoom in on a simple shape, such as a circle, it looks like a line. However, a fractal can be thought of as a set that remains complex on any arbitrarily small scale. For many fractals, zooming in on a small region will produce a set that is similar, but not necessarily identical, to the original fractal. No matter how far we zoom in, this property will always hold true, and is called \textit{self-similarity}.

For this particular function, the convergence plot from Newton’s method has self-similarity. Figure 2.3 shows successive magnifications of increasingly smaller areas of Figure 2.2, but that appear nearly identical. Convergence plots that are self-similar like this also display chaotic behavior. That is, two initial points that are very close can have very different results when Newton’s method is applied. For example, Figure 2.3c shows that \( x^{(0)} = (3.59, 0.3729) \) converges while \( x^{(0)} = (3.59, 0.3718) \) does not. This behavior will reoccur in future chapters.

As observed earlier, the system (2.1) is only nonlinear in \( x_2 \) and the convergence plot for Newton’s method clearly reflects this. The convergence plots for both Broyden methods in Figure 2.4 and Figure 2.5 show significantly different behavior. Despite the original Jacobian being linear in \( x_1 \), both Broyden convergence plots appear nonlinear. Both plots have similarly good convergence in the upper third of the area of interest. However, Broyden’s “bad” method performs much poorly in general over the remainder of the region. All three convergence plots show a thin band of good convergence at \( x_2 \approx -1 \). The cause of this behavior is unknown.
Figure 2.2: Local convergence regions of Newton's method with max-cond = \(10^5\), max-iter = 200, tol = \(10^{-10}\), and \(\Delta x_1 = \Delta x_2 = 0.01\). The zero at \((5, 4)\) is marked by the *.
Figure 2.3: Detail convergence regions corresponding to each zero of (2.1) using Newton’s local method with \( \text{max}_\text{iter} = 30 \) and the corresponding \( \Delta x_1 \) and \( \Delta x_2 \) values labeled on each subfigure. The region \( \{(x_1, x_2) : 0 \leq x_1 \leq 10, 0.35 \leq x_2 \leq 0.72\} \) in Figure 2.2 is enlarged and shown in Figure 2.3a and the region \( \{(x_1, x_2) : 0 \leq x_1 \leq 10, 0.367 \leq x_2 \leq 0.401\} \) in Figure 2.3a is enlarged and shown in Figure 2.3b.
Figure 2.3: (continued) The region \{(x_1, x_2) : 0 \leq x_1 \leq 10, 0.3682 \leq x_2 \leq 0.3735\} in Figure 2.3b is enlarged and shown in Figure 2.3c. The self-similarity of Figure 2.2 is evident from these examples.

Table 2.1: Comparison of the total number of initial points that converge for each of the local methods when applied to (2.1). Additionally, the convergence percentage of convergent initial points and total initial points within a certain number of iterations are compared. For the first 50 iterations, both Broyden methods show a similar convergence percentage of total points, but since more than 99% of the points that will converge already have for Broyden’s “bad” method, in the long run, Broyden’s “good” method comes out on top. Newton’s method significantly outperforms both Broyden methods in the total number of points that converge, either for a small number of iterations (50) or for a large number (400).
Figure 2.4: Local convergence regions of Broyden’s “good” method with $\max\text{-}\text{cond} = 10^5$, $\max\text{-}\text{iter} = 200$, $\text{tol} = 10^{-10}$, and $\Delta x_1 = \Delta x_2 = 0.01$. The zero at (5,4) is marked by the *.
Figure 2.5: Local convergence regions of Broyden's "bad" method with $\text{max-\text{cond}} = 10^5$, $\text{max-\text{iter}} = 200$, $\text{tol} = 10^{-10}$, and $\Delta x_1 = \Delta x_2 = 0.01$. The zero at (5,4) is marked by the *.
Figure 2.6: Detail of local convergence regions of both Broyden methods near the zero of (2.1) with $\max_{\text{cond}} = 10^5$, $\max_{\text{iter}} = 200$, $\text{tol} = 10^{-10}$, and $\Delta x_1 = \Delta x_2 = 5 \times 10^{-3}$. The location of the zero at $(5, 4)$ is marked by the *.
One would expect the initial points near (5, 4) to converge quickly since these are local methods. Figure 2.6 shows the behavior of both Broyden methods for such initial points. The range for the color map is smaller than the previous full page plots to allow for greater contrast between levels. The behavior of Broyden’s “bad” method near the zero more closely reflects the behavior of Newton’s method than Broyden’s “good” method. Additionally, the final Jacobian approximation $\mathbf{J}_\nu$ more closely approximates the Jacobian $\mathbf{J}(\bar{x})$. For example, if $\mathbf{x}^{(0)} = (6, 3.8)$, upon convergence we have

$$
\mathbf{J}(\bar{x}) = \begin{pmatrix} 1 & -10 \\ 1 & 42 \end{pmatrix}, \quad \mathbf{J}^{(\text{good})}_\nu = \begin{pmatrix} 1.31 & -7.52 \\ 0.41 & 37.3 \end{pmatrix}, \quad \mathbf{J}^{(\text{bad})}_\nu = \begin{pmatrix} 0.96 & -10.4 \\ 1.08 & 42.7 \end{pmatrix}.
$$

In this example, $\mathbf{J}^{(\text{good})}_\nu$ has a maximum relative error of 59% as an approximation to $\mathbf{J}(\bar{x})$ while $\mathbf{J}^{(\text{bad})}_\nu$ has a maximum relative error of 8% as an approximation to $\mathbf{J}(\bar{x})$. While neither of these approximations are very good for an initial point so close to $\bar{x}$, $\mathbf{J}^{(\text{good})}_\nu$ is significantly worse.

As mentioned previously, the Jacobian is singular for $x_2 \approx 2.23$ and $x_2 \approx -0.90$. In addition, the Jacobian only depends on $x_2$ and thus the convergence of a particular initial point $(x_1^{(0)}, x_2^{(0)})$ using Newton’s method is completely dependent on $x_2$. Figure 2.2 shows a clear separation at the line $x_2 = 2.23$ between two characteristically different regions of the convergence plot, directly corresponding to the first singular curve of the Jacobian.

If the condition number of $\mathbf{J}^{(\text{good})}_\nu$ or $\mathbf{J}^{(\text{bad})}_\nu$ is considered, both Broyden methods will fail to converge for initial points chosen on or near the curve $x_2 \approx 2.23$ as well. For both methods, we assume $\mathbf{J}_0 = \mathbf{J}(\mathbf{x}^{(0)})$ so this result is not surprising. In the upcoming chapters, we will find that the nonlinearity of the Jacobian singular curves
will have different and sometimes drastic impacts on the convergence of each method for initial points near or on these curves.
3.1 The Function

3.1.1 Nonlinear System

The next function we considered is a polynomial function like (2.1). However, Equation 3.1 is nonlinear in both variables and has two distinct zeros at (1, 1) and (4, 2). Both zeros lie in the same narrow valley, and are only separated by a ridge where $\|f(x)\|_2 \approx 0.06$. By these characteristics, we expect very sensitive convergence regions.

\[
\begin{align*}
  f_1(x) &= x_1 - x_2^2 \\
  f_2(x) &= (x_2 - 1)^2(x_2 - 2)^2 + (x_1 - x_2^2)^2
\end{align*}
\]  

(3.1)

3.1.2 Contour Plot

The area of interest that we considered for the convergence plots was the same as the one Spedicato and Huang considered in their original paper [13]. Akin to (2.1), $\|f(x)\|_2$ grows rapidly as $x_2$ moves farther from the zeros in either the positive or negative direction. Due to closeness of the contour lines, Figure 3.1a shows a restricted range along the $x_2$ axis. Figure 3.1b illustrates the contours immediately surrounding and near to the zeros of the function.
Figure 3.1: (3.1a) Contour plot of Equation 3.1 showing the center of the area of interest. The levels increase by 10 after the level curve for which $\|f(x)\|_2 = 5$. (3.1b) Detail contour plot of Equation 3.1 showing the area near the zeros (1, 1) and (4, 2) which are marked on both plots by the *’s. The level curves immediately surrounding the zeros are plotted and the levels increase by one after the level curve for which $\|f(x)\|_2 = 1$. The dotted lines denoted the $x_2$ values for which the Jacobian is singular.
Figure 3.2: Local convergence regions corresponding of (3.1) using Newton’s local method with $\text{max\_cond} = 5 \times 10^5$, $\text{max\_iter} = 400$, $\text{tol} = 10^{-10}$, and $\Delta x_1 = \Delta x_2 = 0.02$. The zeros at $(1,1)$ and $(4,2)$ are denoted by the *'s.
Figure 3.3: Local convergence “regions” corresponding to (3.1) using Broyden’s “good” local method with $\text{max\_cond} = 5 \times 10^5$, $\text{max\_iter} = 400$, $\text{tol} = 10^{-10}$, and $\Delta x_1 = \Delta x_2 = 0.02$. The zeros at (1,1) and (4,2) are denoted by the *’s.
Figure 3.4: Local convergence regions to (3.1) using Broyden's "bad" local method with $\max_{\text{cond}} = 5 \times 10^5$, $\max_{\text{iter}} = 400$, $\text{tol} = 10^{-10}$, and $\Delta x_1 = \Delta x_2 = 0.02$. The zeros at (1, 1) and (4, 2) are denoted by the *'s.
3.2 Local Convergence Regions

Despite the function’s nonlinearity, like (2.1) the singular curves of the Jacobian are all constant, specifically \( x_2 = 1 \), \( x_2 = 1.5 \), and \( x_2 = 2 \). As can be seen, two of the curves pass directly through both zeros of (3.1). As seen in the previous example, these curves will separate the convergence plot from Newton’s method into characteristically different regions.

For initial runs of the methods, the characteristic parameters `max_cond` and `tol` were kept constant with all other examples we looked at, but for this function the parameters we had decided on were an unfortunate combination and Newton’s method failed for all initial points due to the maximum condition number of \( \mathcal{J}(x) \) being exceeded. Thus for all runs on this function, `max_cond` was increased to \( 5 \times 10^5 \). The convergence regions of (3.1) using all three local methods are shown in Figures 3.2 - 3.4. Despite Figure 3.2 showing good convergence areas below \( x_2 = 1 \), due to the singular Jacobian curve through the zero at \((1, 1)\), the convergence rate is only linear throughout all the convergence regions, as \( C_\nu \to 0.5 \) as the iterates approach the zero.

For two of the three methods, (3.1) caused mediocre convergence results. With the exception of Broyden’s “bad” method, many initial points took many iterations to converge, and even after 400 iterations, most still failed to converge. Graphically, Broyden’s “bad” method is a lot better than Broyden’s “good” method and Table 3.1 reinforces this. Far and above, this function presented the most difficulties for all methods. There are almost no connected regions where initial points converge when using Broyden’s “good” method. Even for points near to the zeros, the methods are extremely sensitive and convergence is spotty at best. Figures 3.5 - 3.6b illustrate this local sensitivity.
<table>
<thead>
<tr>
<th>METHOD</th>
<th>% TOTAL</th>
<th>THRES</th>
<th>% CONV ≤ THRES</th>
<th>% TOTAL ≤ THRES</th>
</tr>
</thead>
<tbody>
<tr>
<td>NWT</td>
<td>54.319</td>
<td>50</td>
<td>100</td>
<td>54.319</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>65.600</td>
<td>35.631</td>
</tr>
<tr>
<td>BROYG</td>
<td>10.565</td>
<td>50</td>
<td>28.748</td>
<td>3.073</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>1.674</td>
<td>0.177</td>
</tr>
<tr>
<td>BROYB</td>
<td>61.613</td>
<td>50</td>
<td>88.034</td>
<td>54.240</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>1.322</td>
<td>0.814</td>
</tr>
</tbody>
</table>

Table 3.1: Comparison of the total number of initial points that converge for each of the local methods when applied to (3.1). Additionally, the convergence percentage of convergent initial points and total initial points within a certain number of iterations are compared. Within 50 iterations, the percentage of initial points that converge from Broyden’s “bad” is comparable to Newton’s method but the performance of Broyden’s “good” method is terrible on all counts. The results also show the difficulties that each Broyden method had when the number of iterations was small (25). All initial points that will converge have already done so within 50 iterations for Newton’s method and ultimately Broyden’s “bad” method has the best results.

Figure 3.5: Detail of Newton’s local convergence regions near the zeros of (3.1) with \( \text{max\_cond} = 5 \times 10^5 \), \( \text{max\_iter} = 400 \), \( \text{tol} = 10^{-16} \), and \( \Delta x_1 = \Delta x_2 = 0.01 \). The zeros are denoted by the *’s.
Figure 3.6: Detail of Broyden’s local convergence regions near the zeros of (3.1) with \( \max_{\text{cond}} = 5 \times 10^5 \), \( \max_{\text{iter}} = 400 \), \( \text{tol} = 10^{-10} \), and \( \Delta x_1 = \Delta x_2 = 0.01 \). The zeros are denoted by the *'s.
Both Broyden’s “good” method and Newton’s method show extremely poor convergence for initial points picked near the zero at (4, 2), although Broyden’s “bad” method does not perform much better. The singular curves of the Jacobian clearly cause problems for each of the methods for initial points on or near the curves.

### 3.3 Global Convergence Regions

![Convergence regions of (3.1) using Newton’s global method with $\lambda_{\text{min}} = 1/32$ and $\max \{\|\Delta x^{(\nu)}\|_2\} = 10$. The other parameters were kept the same as used in the local convergence computation.](image)

Figure 3.7: Convergence regions of (3.1) using Newton’s global method with $\lambda_{\text{min}} = 1/32$ and $\max \{\|\Delta x^{(\nu)}\|_2\} = 10$. The other parameters were kept the same as used in the local convergence computation.

The area of interest we have chosen for this problem doesn’t do a good job satisfying the “near enough” requirement that initial points of local methods must have to ensure convergence. Additionally, the convergence plots in Figures 3.2 - 3.4 show large areas of nonconvergence far from the zeros. In Section 1.3.6, we discussed
couple of easy modifications that could be applied to a local method to transform it to a global method. In the interest of examining how Newton’s global method affects the convergence regions, Figure 3.7 was generated. There was little improvement in the convergence regions save for the area between $x_2 = 1.5$ and $x_2 = 2$. The convergence was still $q$-linear from all initial points so there is little overall improvement when using Newton’s global method on this system.
4.1 The Function

4.1.1 Nonlinear System

\[ f_1(x) = x_1^2 - x_2 - 1 \]
\[ f_2(x) = -x_1 + x_2^2 - 1 \]  

Equation 4.1 is the second two-dimensional function whose convergence regions were calculated in the paper by Spedicato and Huang [13]. The authors tested the performance of all quasi-Newton methods with the exception of Broyden’s “bad” method and Greenstadt’s second method on this function. However, none of their graphical results for this function were displayed in the paper. The contour plot of (4.1) is shown in Figure 4.1.

4.1.2 Contour Plot

This function is the first one explored in this thesis to this point that has nonlinear Jacobian singular curves, one of which passes between several of the zeros. Similar to (3.1), it features three of its four zeros clustered tightly in a shallow basin. The fourth zero is isolated from this grouping by a saddle point and we want to see if the convergence to the zero at \(((1 + \sqrt{5})/2, (1 + \sqrt{5})/2)\) will be characteristically different
due to this separation. Unlike either of the previous functions, (4.1) is relatively gentler sloped away from the zeros, and the area of interest is smaller and more local in nature.

Figure 4.1: Contour plot of Equation 4.1. The zeros at \((0, -1), (-1, 0), ((1 + \sqrt{5})/2, (1 + \sqrt{5})/2)\) and \(((1 - \sqrt{5})/2, (1 - \sqrt{5})/2)\) are marked by the *’s. The contour lines immediately around the group of three zeros correspond to \(\|f(x)\|_2 = 0.06\). The remaining level curves differ by 0.5. The dotted lines indicate where the Jacobian is singular.

4.2 Local Convergence Regions

Away from the Jacobian singular curves, all three methods converged quickly. Almost all of the initial points converged in 20 iterations or less for Newton’s local method and 30 iterations or less for both Broyden local methods. Figure 4.2 shows some fractal-like curves spawning from the Jacobian singular curves.
Figure 4.2: Convergence regions of all zeros of (4.1) using Newton’s local method with $\max_{\text{cond}} = 10^5$, $\max_{\text{iter}} = 400$, $\text{tol} = 10^{-10}$ and $\Delta x_1 = \Delta x_2 = 5 \times 10^{-3}$. The location of the zeros are marked by the *’s.
Figure 4.3: Convergence regions of all zeros of (4.1) using Broyden’s “good” local method with $\text{max}_\text{cond} = 10^6$, $\text{max}_\text{iter} = 400$, $\text{tol} = 10^{-10}$ and $\Delta x_1 = \Delta x_2 = 5 \times 10^{-3}$. The location of the zeros are marked by the *'s.
Figure 4.4: Convergence regions of all zeros of (4.1) using Broyden’s “bad” local method with $\text{max\_cond} = 10^5$, $\text{max\_iter} = 400$, $\text{tol} = 10^{-10}$ and $\Delta x_1 = \Delta x_2 = 5 \times 10^{-3}$. The location of the zeros are marked by the *'s.
With Broyden’s “good” and “bad” methods, the regions near these curves display wider areas of non-convergence. Figure 4.4 shows that Broyden’s “bad” method has the most trouble with initial points lying on or near the two Jacobian singular curves.

4.3 Global Convergence Regions

As mentioned previously, there is no stable global Broyden method, but we again tested Newton’s global method first on the same area of interest (Figure 4.5) and then on a larger area (Figure 4.6b). While a maximum step size of 10 decreases on average the number of iterations required to converge to any zero, the fractal patterns near the upper singular curve are destroyed.
Figure 4.5: Convergence regions of all zeros of (4.1) using Newton’s global method with $\max_{\text{cond}} = 10^5$, $\max_{\text{iter}} = 400$, $\text{tol} = 10^{-10}$ and $\Delta x_1 = \Delta x_2 = 5 \times 10^{-3}$. In addition, $\lambda_{\text{min}} = 1/32$ and $\max \{ \| \Delta x^{(\nu)} \|_2 \} = 10$. The location of the zeros are marked by the white *'s.
Figure 4.6: Convergence regions of all zeros of (4.1) over a larger area of interest using Newton’s local and global methods with \( \text{max}\_\text{cond} = 10^5, \text{max}\_\text{iter} = 400, \text{tol} = 10^{-10} \) and \( \Delta x_1 = \Delta x_2 = 5 \times 10^{-3} \). The location of the zeros are marked by the white *’s. As in Figure 4.5, \( \lambda_{\text{min}} = 1/32 \) but \( \max \{\|\Delta x^{(\nu)}\|_2\} \) has been reduced to 0.5.
Much of this is dependent on the value of \texttt{max \_step}. For example, Figure 4.6a shows the local convergence regions due to Newton’s method on an expanded area of interest. The fractal patterns are again very clear. If Newton’s global method is applied to this new area, this time with \texttt{max \_step} = 1/2, the noisy banded region we observed by the upper Jacobian singular curve in Figure 4.5 has disappeared. On the other hand, it is obvious that \texttt{max \_step} is now too small for such a large area as some of the nice convergence farther from the zeros is lost. It is interesting to note that the convergence “rings” surrounding the isolated zero are now elliptical as opposed to their previously rounded rectangular areas.

<table>
<thead>
<tr>
<th>METHOD</th>
<th>% TOTAL</th>
<th>THRES</th>
<th>% CONV ≤ THRES</th>
<th>% TOTAL ≤ THRES</th>
</tr>
</thead>
<tbody>
<tr>
<td>NWT</td>
<td>99.992</td>
<td>50</td>
<td>100</td>
<td>99.992</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>99.999</td>
<td>99.991</td>
</tr>
<tr>
<td>BROYG</td>
<td>99.790</td>
<td>50</td>
<td>99.661</td>
<td>99.451</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>98.723</td>
<td>98.515</td>
</tr>
<tr>
<td>BROYB</td>
<td>97.473</td>
<td>50</td>
<td>99.981</td>
<td>97.454</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>98.854</td>
<td>96.356</td>
</tr>
</tbody>
</table>

Table 4.1: Comparison of the total number of initial points that converge for each of the local methods when applied to (4.1). Additionally, the convergence percentage of convergent initial points and total initial points within a certain number of iterations are compared. All three methods displayed very good results within the given thresholds and also within the total number of iterations (400).
Chapter 5: Conte & de Boor’s Function

5.1 The Function

5.1.1 Nonlinear System

The last nonlinear function that the methods are applied to is taken from a book by Conte and De Boor [5]. Unfortunately, the derivation of this function is otherwise unknown. This function features two zeros, a local minimum, and a singularity along the line $x_1 = 0$. It is also the second function in this thesis that features nonlinear Jacobian singular curves.

\[ f_1(x) = x_1 + 3 \log |x_1| - x_2^2 \]
\[ f_2(x) = 2x_2^2 - x_1x_2 - 5x_1 + 1 \] (5.1)

The local minimum of at $(-0.40, -0.20)$ of approximately 4.55 is much closer to the zero at $(-1.52, 1.37)$ than the local minimum was to the zero in (2.1), which was the only other system we explored that contained a local minimum. Coupled with the singularity, (5.1) has the potential to cause many difficulties with the convergence of the methods.
5.1.2 Contour Plot

Figure 5.1: Contour plot of Equation 5.1. The zeros located approximately at \((-1.52, 1.37)\) and \((3.76, 2.78)\) are marked by the *’s. Additionally, the local minimum at \((-0.40, -0.20)\) is marked by the +. The singularity at \(x_1 = 0\) and the curves where the Jacobian is singular are denoted by the dotted lines. The immediate levels surrounding the zeros and minimum are denoted and the values increase by one for each successive contour.

5.2 Local Convergence Regions

The Jacobian matrix of this \(2 \times 2\) system is

\[
\mathcal{J}(x) = \begin{pmatrix}
1 + \frac{3}{x_1} & -2x_2 \\
\frac{4x_1 - x_2 - 5}{x_1} & -x_1
\end{pmatrix}. \tag{5.2}
\]
Since $J$ is a square matrix, it is singular if and only if $\det(J(x)) = 0$. Solving the resulting equation gives the curves as a function of $x_2$:

$$x_1 = \frac{2x_2^2 + 10x_2 + 3}{8x_2 - 1}, \quad x_2 \neq \frac{1}{8}. \quad (5.3)$$
Figure 5.2: Local convergence regions of all zeros of (5.1) using Newton’s method with $\text{max}\_\text{cond} = 10^5$, $\text{max}\_\text{iter} = 400$, $\text{tol} = 10^{-10}$ and $\Delta x_1 = \Delta x_2 = 0.01$. The location of the zeros are marked by the *’s.
Figure 5.3: Local convergence regions of all zeros of (5.1) using Broyden’s “good” method with $\text{max}_{\text{cond}} = 10^5$, $\text{max}_{\text{iter}} = 400$, $\text{tol} = 10^{-10}$ and $\Delta x_1 = \Delta x_2 = 0.01$. The location of the zeros are marked by the *'s.
Figure 5.4: Local convergence regions of all zeros of \((5.1)\) using Broyden's "bad" method with max-cond \(= 10^5\), max-iter \(= 400\), tol \(= 10^{-10}\) and \(\Delta x_1 = \Delta x_2 = 0.01\). The location of the zeros are marked by the *'s.

The location of the zeros are marked by the *'s.
The local convergence plots reinforced some of the patterns that were observed in previous examples. Figure 5.2 shows the Jacobian singular curves separating regions where Newton’s method exhibited fast convergence from areas where the method struggled. Figure 5.4 shows Broyden’s “bad” method once again struggling with initial points picked close to or on nonlinear Jacobian singular curves.

Over the approximate area \( \{(x_1, x_2) \mid 0 < x_1 \leq 2, \ 0 \leq x_2 \leq 5\} \), all the methods had surprisingly poor results even though neither zero was far from this region. In addition, the convergence of initial points for \( x_1 < 0 \) was very poor. Except for small regions surrounding each zero, none of the local methods displayed good results.

<table>
<thead>
<tr>
<th>METHOD</th>
<th>% TOTAL</th>
<th>THRES</th>
<th>% CONV ≤ THRES</th>
<th>% TOTAL ≤ THRES</th>
</tr>
</thead>
<tbody>
<tr>
<td>NWT</td>
<td>94.538</td>
<td>50</td>
<td>95.248</td>
<td>90.045</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>81.494</td>
<td>77.043</td>
</tr>
<tr>
<td>BROYG</td>
<td>81.401</td>
<td>50</td>
<td>73.689</td>
<td>59.984</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>67.206</td>
<td>54.706</td>
</tr>
<tr>
<td>BROYB</td>
<td>64.972</td>
<td>50</td>
<td>87.158</td>
<td>56.628</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>77.511</td>
<td>50.361</td>
</tr>
</tbody>
</table>

Table 5.1: Comparison of the total number of initial points that converge for each of the local methods when applied to (5.1). Additionally, the convergence percentage of convergent initial points and total initial points within a certain number of iterations are compared. The majority of the initial points that converge for Broyden’s “bad” method converge in 50 iterations, so that initially, it outperforms Broyden’s “good” method. However, a higher percentage of all initial points have converged for Broyden’s “good” method at this threshold and ultimately, Broyden’s “good” method does much better. As has been seen in all the previous nonlinear systems, Newton’s method does better than Broyden’s methods on all counts.
5.3 Global Convergence Regions

Figure 5.5: Newton’s global convergence regions showing the sensitivity of the convergence regions to the maximum step size with $\lambda_{\min} = 1/32$ and the local parameters unchanged for both plots.
We ran Newton’s global method on the area of interest first with \texttt{max\_step} = 10 and then reducing \texttt{max\_step} to 3.7. Figure 5.5 shows the resulting convergence regions. The convergence benefited from the reduced maximum step size, especially for $x_1 > 0$. The convergence problems that Newton’s local method had with initial points in the small region between the individual convergence regions of the two zeros were eradicated with the smaller step size. In addition, there was some improvement in the area for which $x_1 < 0$. 
Chapter 6: Conclusion: “Good” vs. “Bad”

Up to this point, we’ve seen both Newton and Broyden methods applied to a variety of two-dimensional nonlinear systems of equations. First, we will mention some of the arguments for and against the use of Broyden’s “bad” method. Second, the results of the numerical analysis of the methods will be summarized. Finally, we will give our verdict on whether one Broyden method is truly better than the other.

6.1 Arguments For and Against Broyden’s “Bad” Method

There has been much debate concerning Broyden’s original secant method. Is one formulation really numerically and practically better than the other one to the point of naming them “good” and “bad”? In his original paper (1965) [4] detailing his new method, Broyden himself dismissed the “bad” formulation as inferior, stating

By similar arguments used for [the “good method] it can be shown that if the equations to be solved are linear, then \([-J_{\nu+1}^{-1}\] is not a worse approximation to \([-J^{-1}\] than \([-J\nu{-1}\]. Since, however, this method appears in practice to be unsatisfactory, it will be discussed no further at this stage . . . While the program was being developed, it became increasingly apparent that [the “bad” method] was quite useless. In every case tried it rapidly reached a state where successive step vectors \([p_\nu]\) and \([p_{\nu+1}]\) were identical and further progress thus became impossible, and for this reason it is omitted from the following case histories.

Spedicato and Huang (1997) performed an extensive computational exploration [13] of Newton’s method along with six quasi-Newton methods, which included both
Broyden methods. They tested the methods on 31 different families of nonlinear systems of dimensions 10, 50, and 100. The initial points per system was the standard one $x_s^{(0)}$ given by the original reference and the point 50 times farther away, $50x_s^{(0)}$. The authors determined the effectiveness of a given method by the total number of solved problems, the speed of the method, and the size of the nonconvergence region. Based on this criteria, of the three methods we considered, Newton’s method was named the best method and Broyden’s “bad” method was deemed the worst.

In addition to these higher dimensional systems, Spedicato and Huang conducted an additional experiment by applying Newton’s and Broyden’s “good” methods on a couple of two-dimensional systems, namely (3.1) and (4.1), for the purpose of examining the structure of the regions of convergence and/or nonconvergence. However, the authors did not test Broyden’s “bad” method on either of these two systems because they deemed the method’s performance to be too poor. The results of Chapter 3 present evidence that suggests this may have been a premature conclusion. In this second experiment, there is no mention of whether or not a global or local method is used. By comparing our results with the authors’ original convergence results, the methods used seem to be local. In light of this, and based on our results in Chapter 3 showing that all three methods were very sensitive even on a smaller area of interest, it seems unfair to label Broyden’s “bad” method as significantly worse than either Newton’s or Broyden’s “good” method.

Some have suggested that the computational differences derive from the fundamental difference between the definition of Broyden’s “good” and “bad” methods. In their numerical methods text, Dennis and Schnabel (1983) offer their opinion on the subject [8]:
Although we don’t pretend to understand the lack of computational success with Broyden’s bad update, it is interesting that Broyden’s good update comes from minimizing the change in the affine model, subject to satisfying the secant equation, while Broyden’s bad update is related to minimizing the change in the solution to the model. Perhaps the fact that we use information about the function, and not about its solution, in forming the secant model, is a reason why the former approach is more desirable.

Even before Spedicato and Huang’s paper, there were arguments in favor of the merits of Broyden’s “bad method. In a paper in 1991, E. Kvaalen [10] claimed that Broyden’s original reasons for dismissing his “bad” method were invalid and that Broyden’s “bad” method has been tried on several types of problems (including those given by Broyden and has encountered no such difficulty . . . Broyden gives the results of ten sample problems, but explicitly states only six of them. Each of these six has been retried with [Broyden’s “bad” method] and found to converge in the same number of iterations as with [Broyden’s “good” method] . . . Only in one problem with starting point very far from the solution, did [Broyden’s “bad” method] fail, while [Broyden’s “good” method] succeeded after many iterations. But with the original starting point, [Broyden’s “bad” method] succeeded, and with less iterations.

Broyden’s “bad” method has also been shown to be particularly effective in the solution of systems of stiff ODEs [3] and has more recently been under additional investigation [1].
6.2 Summary of Numerical Experiments

We focus on three main areas of numerical investigation: the efficiency of each method via an examination of the percentage of points that converge during the run, the behavior near nonlinear Jacobian singular curves, and the degree to which the final Broyden approximations to the Jacobian are accurate.

6.2.1 Efficiency of Methods

In the previous chapters we tabulated some convergence results of each method per problem. We calculated the percentage of total initial points that converged iterations as well as the percentage of convergent points that converged within some small threshold iteration values. The total percentage of initial points that will ultimately converge for a given method is dependent on \( \text{max\_cond} \), \( \text{max\_iter} \), and in some ways, \( \Delta x_1 \) and \( \Delta x_2 \).

Of course, the choice of \( \text{max\_cond} \), while reasonable, is still somewhat of an arbitrary choice. Additionally, for all runs of the three methods in the previous chapters, \( \text{max\_iter} \) was chosen to be large enough that an accurate representation of the percentage of convergent initial points was obtained for the chosen value of \( \text{max\_cond} \). Unfortunately, this value was usually not optimal. In light of this, we calculated the minimum value of \( \text{max\_iter} \) for which the percentage of convergent initial points was unchanged. We also calculated the number of iterations needed to ensure \( \approx 99\% \) of convergent points converge. These results are presented in Tables 6.1 - 6.4.

Remark. These tables only examine the “optimal” value of \( \text{max\_iter} \) required for each method to maintain the same percentage of convergent points as calculated with the original value of \( \text{max\_iter} \). Therefore, although at first glance the numbers in
Table 6.1 seem to suggest the Broyden’s “bad” method outperforms the other two methods, Table 2.1 showed that 95.306% and 54.322% of initial points converged for Newton and Broyden’s “good” methods, respectively, as opposed to only 41.153% for Broyden’s “bad” method.

It is evident from the data that the initial choice of \( \text{max\_iter} \) was almost always too high and a significant number of iterations are wasted in the application of Newton’s Method. However, there was one pattern in particular that appeared in all four tables. In each nonlinear system, Broyden’s “bad” method took far less iterations for the majority (90%) of its convergent points to converge than Broyden’s “good” method.

<table>
<thead>
<tr>
<th>METHOD</th>
<th>MAX_ITER</th>
<th>ITER</th>
<th>% CONV IN ITER</th>
</tr>
</thead>
<tbody>
<tr>
<td>NWT</td>
<td>200</td>
<td>200</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>103</td>
<td>99.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>47</td>
<td>89.99</td>
</tr>
<tr>
<td>BROYG</td>
<td>200</td>
<td>200</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>128</td>
<td>99.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>48</td>
<td>90.14</td>
</tr>
<tr>
<td>BROYB</td>
<td>200</td>
<td>134</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>41</td>
<td>99.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>19</td>
<td>89.67</td>
</tr>
</tbody>
</table>

Table 6.1: Comparison of the efficiency of each local method when applied to (2.1). The last column indicates the percentage of initial points that converged when using the original value of \( \text{max\_iter} \) (second column) if instead \( \text{max\_iter} = \text{iter} \). Broyden’s “bad” method shows a significantly faster initial convergence. This can also be seen graphically in the detail plots in Figure 2.6.
Table 6.2: Comparison of the efficiency of each local method when applied to (3.1). The last column indicates the percentage of initial points that converged when using the original value of max_iter (second column) if instead max_iter = iter. Broyden’s “bad” method shows a significantly faster initial convergence and coupled with the total percentage of points that converge (10.565% to 61.613% for Broyden’s “good” and Broyden’s “bad”, respectively) solidifies its significantly better performance on this system.

<table>
<thead>
<tr>
<th>METHOD</th>
<th>MAX_ITER</th>
<th>ITER</th>
<th>% CONV IN ITER</th>
</tr>
</thead>
<tbody>
<tr>
<td>NWT</td>
<td>400</td>
<td>33</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30</td>
<td>99.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>26</td>
<td>88.63</td>
</tr>
<tr>
<td>BROYG</td>
<td>400</td>
<td>355</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>157</td>
<td>99.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>111</td>
<td>90.05</td>
</tr>
<tr>
<td>BROYB</td>
<td>400</td>
<td>180</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>70</td>
<td>99.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>51</td>
<td>89.32</td>
</tr>
</tbody>
</table>

Table 6.3: Comparison of the efficiency of each local method when applied to (4.1). The last column indicates the percentage of initial points that converged when using the original value of max_iter (second column) if instead max_iter = iter. Both Broyden methods are comparable when applied to this system, although Newton’s Method is a step above.

<table>
<thead>
<tr>
<th>METHOD</th>
<th>MAX_ITER</th>
<th>ITER</th>
<th>% CONV IN ITER</th>
</tr>
</thead>
<tbody>
<tr>
<td>NWT</td>
<td>400</td>
<td>33</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13</td>
<td>99.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>90.42</td>
</tr>
<tr>
<td>BROYG</td>
<td>400</td>
<td>400</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>29</td>
<td>99.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15</td>
<td>91.35</td>
</tr>
<tr>
<td>BROYB</td>
<td>400</td>
<td>91</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>26</td>
<td>99.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15</td>
<td>89.39</td>
</tr>
</tbody>
</table>
Table 6.4: Comparison of the efficiency of each local method when applied to (5.1). The last column indicates the percentage of initial points that converged when using the original value of max_iter (second column) if instead max_iter = iter. As discussed in Chapter 4, all methods struggled with this nonlinear system but the results here highlight how much slower Broyden’s “good” method converged compared to Newton and Broyden’s “bad” methods.

### 6.2.2 Behavior Near Jacobian Singular Curves

Two of the examples presented earlier, namely Spedicato & Huang’s Function #2 (4.1) and Conte & de Boor’s function (5.1), feature nonlinear Jacobian singular curves in the areas of interest. Most of the shortcomings of Broyden’s “bad” methods on these two problems occurred for points on or near these curves. By plotting the individual convergence regions, it is easier to observe the difficulties the Broyden’s “bad” method has near the Jacobian singular curves. Figure 6.1 shows the individual convergence regions corresponding to each zero of (4.1) when Newton’s local method is used. The intensity of each particular color denotes the number of iterations a point requires to converge to the corresponding zero, with a deeper intensity translating to a longer convergence time. The Newton convergence plot has fractal behavior along...
the singular Jacobian curves that separate the individual convergence regions. One particular area is outlined in the figure below and explored in Figure 6.2.

Figure 6.1: Convergence regions corresponding to each zero of (4.1) using Newton’s local method with max iter = 30 and $\Delta x_1 = \Delta x_2 = 2.5 \times 10^{-3}$. The zeros are denoted by the *'s. The region outlined by the black box in the lower left region of the plot is enlarged in Figure 6.2a.
(a) $\Delta x_1 = 1 \times 10^{-4}$, $\Delta x_2 = 1 \times 10^{-4}$

(b) $\Delta x_1 = 2 \times 10^{-6}$, $\Delta x_2 = 1 \times 10^{-5}$

Figure 6.2: Detail convergence regions corresponding to each zero of (4.1) using Newton’s local method with max_iter = 30 and the corresponding $\Delta x_1$ and $\Delta x_2$ values labeled on each subfigure. The region outlined by the black box in Figure 6.2a is enlarged and the $x_1$ axis is rescaled to form Figure 6.2b and the region outlined by the black box in the lower portion of Figure 6.2b is enlarged and the $x_2$ axis rescaled to form Figure 6.2c.
Figure 6.2: (continued) The region outlined by the black box in Figure 6.2c is enlarged and the $x_1$ axis rescaled to form Figure 6.2d. The self-similarity in the overall convergence plot is clearly displayed both in the previous and current figure.
Figure 6.3: Convergence regions corresponding to each zero of (4.1) using Broyden’s local method. For both plots, $\text{max\_iter} = 30$ and $\Delta x_1 = \Delta x_2 = 2.5 \times 10^{-3}$. The zeros are denoted by the *’s.
Figure 6.4: Convergence regions corresponding to each zero of (5.1) using (6.4a) Newton’s local method and (6.4b) Broyden’s “bad” local method. For both plots, \( \text{max}_{\text{iter}} = 30 \) and \( \Delta x_1 = \Delta x_2 = 0.01 \). The zeros are denoted by the *'s.
The fractal behavior seen in Figure 6.1 disappears when the Broyden methods are applied to (4.1). The struggles of Broyden’s “bad” method near the Jacobian singular curves are visible, especially near the upper curve. There are areas of the convergence regions corresponding to the zeros at \((-1,0)\) and \((0,-1)\) that suddenly become convergent to the zero at \((1+\sqrt{5})/2, (1+\sqrt{5})/2\). Additionally, there are clear nonconvergent areas along each curve. While Broyden’s “good” method does not have this convergence area “leakage” across the upper singular curve, there are interior areas, e.g., the bottom left of Figure 6.3a, that switch convergence. There are no such areas in Figure 6.3b. The Conte & de Boor system (5.1) also contains nonlinear Jacobian singular curves and the corresponding “bad” Broyden individual convergence regions also show the same instability near the singular curves. Figure 6.4 shows the contrast in the Newton and “bad” Broyden convergence regions. While the particular choice for \(\text{max\_iter}\) is far below the optimal number of iterations for Broyden’s “bad” method for this system (see Table 6.4), enough points have converged to compare its behavior to that of Newton’s method.

6.2.3 Final Jacobian Approximations

In Chapter 2, we briefly gave an example showing that \(J^{(\text{good})}_{\nu}\) does not necessarily approximate \(J(\bar{x})\) well. Similar behavior can be seen in the plots for (4.1). While Broyden’s “good” method has significantly less struggles near the Jacobian singular curves than Broyden’s “bad” method, Figure 6.3 suggests that some of the initial points in the interior of the convergence regions take longer to converge.
Figure 6.5: Function and error convergence plots for (4.1) using two different initial points. The plots show the results using Newton’s Method (solid with * markers), Broyden’s “good” method (dot-dashed with ◦ markers), and Broyden’s “bad” method (dashed with Δ markers). The third plot tracks the condition number of \( J(x^{(p)}) \) or \( J_\nu \) at each iteration.
If we consider, $x^{(0)} = (2, -2)$, the final Jacobian approximations at $\bar{x} = (0, -1)$ are

$$J(\bar{x}) = \begin{pmatrix} 0 & -1 \\ -1 & -2 \end{pmatrix}, \quad J_{\nu}^{(good)} = \begin{pmatrix} 0.457 & 0.360 \\ -1.45 & -3.33 \end{pmatrix}, \quad J_{\nu}^{(bad)} = \begin{pmatrix} -0.048 & -1.14 \\ -0.960 & -1.89 \end{pmatrix}. \quad (6.1)$$

From a relative error standpoint, $J_{\nu}^{(bad)}$ is a significantly better approximation to $J(\bar{x})$. However, if we only look at the error and function convergence plots (Figure 6.5a) of the three methods, both Broyden methods behave similarly.

Figure 6.5b again shows similar behavior for both Broyden methods when $x^{(0)} = (-1.6, 1.9)$, but again the final Jacobian approximations show a better result for Broyden’s “bad” method:

$$J(\bar{x}) = \begin{pmatrix} -2 & -1 \\ -1 & 0 \end{pmatrix}, \quad J_{\nu}^{(good)} = \begin{pmatrix} -2.87 & -1.30 \\ -0.026 & 0.337 \end{pmatrix}, \quad J_{\nu}^{(bad)} = \begin{pmatrix} -1.96 & -0.985 \\ -1.06 & -0.025 \end{pmatrix}. \quad (6.2)$$

Oddly enough, if $x^{(0)} = (0, 0)$, both Broyden methods have bad, but identical, final approximations to $J(\bar{x})$

$$J(\bar{x}) = \begin{pmatrix} -1.236 & -1 \\ -1 & -1.236 \end{pmatrix}, \quad J_{\nu}^{(good)} = J_{\nu}^{(bad)} = \begin{pmatrix} -0.618 & -1.618 \\ -1.618 & -0.618 \end{pmatrix}. \quad (6.3)$$

### 6.3 Conclusions

Combining these numerical results leaves a mixed bag of results. When running all the methods for long enough, Broyden’s “good” method will generally outperform Broyden’s “bad” method by the sheer number of initial points that converge. However, many times Broyden’s “bad” method will have better convergence for a smaller number of iterations. Additionally, for a larger area of interest, there were times when Broyden’s “bad” method had a better convergence percentage. We also
observed the instability of convergence regions of Broyden’s “bad” method near non-linear Jacobian singular curves. In the case where we prevent the iterations from continuing if the maximum condition number of the Jacobian approximation $\mathbf{J}_\nu$ is exceeded, both methods tend to typically fail in different ways. Iterates that fail to converge when using Broyden’s “good” method are stuck in a loop without $\text{max\_cond}$ ever being exceeded while Broyden’s “bad” method will cause $\text{max\_cond}$ to increase rapidly past the threshold. From the small sample size of initial points we selected to explore, we observed that many times similar Broyden convergence behavior does not imply a similar final Jacobian approximation, although more often Broyden’s “bad” method proved to be better. Ultimately, over local regions Broyden’s “good” method seems to have a slight edge over Broyden’s “bad” method but not enough to warrant the contrast in names. However, due to overall performance and the easy extension to a global method, Newton’s method is definitely the best of the three methods we explored.
Appendix A: Proof of Superlinear Convergence of Broyden’s Method

A.1 The Method

Recall that Broyden’s secant method can be represented by the following iteration over $\nu \in \mathbb{N}[0, \infty)$ and the corresponding matrix update:

\[
\begin{align*}
\mathbf{x}^{(\nu+1)} &= \mathbf{x}^{(\nu)} - \mathbf{J}_\nu^{-1}f(\mathbf{x}^{(\nu)}) \\
\mathbf{J}_{\nu+1} &= \mathbf{J}_\nu + \frac{(\Delta f^{(\nu)} - \mathbf{J}_\nu \Delta \mathbf{x}^{(\nu)}) \Delta \mathbf{x}^{(\nu)^T}}{\| \Delta \mathbf{x}^{(\nu)} \|^2_2}
\end{align*}
\] (A.1)

A.2 Useful Lemmas

Lemma A.2.1. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a continuously differentiable function in the open convex set $D \subset \mathbb{R}^n$, and $\mathcal{J} \in \text{Lip}_\gamma(D)$. For any $\mathbf{u}, \mathbf{v} \in D$,

\[
\| f(\mathbf{v}) - f(\mathbf{u}) - \mathcal{J}(\mathbf{x})(\mathbf{v} - \mathbf{u}) \| \leq \gamma \| \mathbf{v} - \mathbf{x} \| + \| \mathbf{u} - \mathbf{x} \| \| \mathbf{v} - \mathbf{u} \|. 
\] (A.2)

Proof. By Newton, we have

\[ f(\mathbf{v}) - f(\mathbf{u}) = \int_\mathbf{u}^{\mathbf{v}} \mathcal{J}(\mathbf{z}) \, d\mathbf{z}. \]
Then
\[
\begin{align*}
\mathbf{f}(v) - \mathbf{f}(u) - \mathcal{J}(\mathbf{x})(v - u) &= \int_u^v \left( \mathcal{J}(z) - \mathcal{J}(\mathbf{x}) \right) dz \\
&= \int_0^1 \left( \mathcal{J}(u - t(v - u)) - \mathcal{J}(\mathbf{x}) \right)(v - u) dt
\end{align*}
\]

after the substitution \( z = u - t(v - u) \). Applying the triangle inequality and simplifying,
\[
\| \mathbf{f}(v) - \mathbf{f}(u) - \mathcal{J}(\mathbf{x})(v - u) \| \leq \int_0^1 \| (\mathcal{J}(u - t(v - u)) - \mathcal{J}(\mathbf{x})) (v - u) \| dt \\
\leq \| v - u \| \int_0^1 \gamma \| u - t(v - u) - x + (t x - t x) \| dt \\
= \gamma \| v - u \| \int_0^1 \| t(u - x) + (1 - t)(v - x) \| dt \\
\leq \gamma \| v - u \| \int_0^1 \left( \| t\| \| u - x \| + \| 1 - t\| \| v - x \| \right) dt \\
= \gamma \| v - u \| \left( \frac{1}{2} \| u - x \| + \frac{1}{2} \| v - x \| \right)
\]

The proofs of the next two lemmas are from construction and equally trivial so they will not be shown here.

**Lemma A.2.2.** Let \( \mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m \) be a continuously differentiable function in the open convex set \( D \subset \mathbb{R}^n \), \( \mathcal{J} \in \text{Lip}_\gamma(D) \), and \( \mathcal{J}(\mathbf{x})^{-1} \) exists. Then there exists \( \varepsilon > 0 \), \( 0 < \alpha < \beta \) such that
\[
\alpha \| v - u \| \leq \| \mathbf{f}(v) - \mathbf{f}(u) \| \leq \beta \| v - u \| \tag{A.3}
\]
for all \( \mathbf{v}, \mathbf{u} \in D \) for which \( \max \{ \| v - \mathbf{x} \|, \| u - \mathbf{x} \| \} \leq \varepsilon \).

**Lemma A.2.3** (Bounded Divergence). Let \( D \subseteq \mathbb{R}^n \) be an open convex set containing \( \mathbf{x}^{(\nu)}, \mathbf{x}^{(\nu+1)} \), with \( \mathbf{x}^{(\nu)} \neq \mathbf{x} \). Let \( \mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n \), \( \mathcal{J}(\mathbf{x}) \in \text{Lip}_\gamma(D) \), \( \mathbf{J_\nu} \in \mathbb{R}^{n \times n} \), with
\( J \), defined by Equation A.1b. Furthermore, if \( \bar{x} \in D \) and \( J(x) \) obeys the weaker Lipschitz condition \( \| J(x) - J(\bar{x}) \| \leq \gamma \| x - \bar{x} \| \) for all \( x \in D \), then for either the Frobenius or \( \ell_2 \) matrix norm,

\[
\| J_{\nu+1} - J(\bar{x}) \| \leq \| J_{\nu} - J(\bar{x}) \| + \frac{7}{2} \left( \| x^{(\nu+1)} - \bar{x} \|_2 + \| x^{(\nu)} - \bar{x} \|_2 \right) \tag{A.4}
\]

**Definition A.2.4.** A linear operator \( P \) on a vector space \( V \) is a projector \( \iff P^2 = P \).

**Lemma A.2.5.** Let \( u \in \mathbb{R}^n \) be nonzero and \( E \in \mathbb{R}^{n \times n} \). Then

\[
\left\| E \left( I - \frac{uu^T}{u^Tu} \right) \right\|_F \leq \left\| E \left( I - \frac{uu^T}{u^Tu} \right) \right\|_F^2 - \frac{1}{2} \left( \left\| Eu \right\|_2 \right)^2 \tag{A.6}
\]

**Proof.** By Definition A.2.4, \( I - uu^T/u^Tu \) is an Euclidean projection matrix and so is \( uu^T/u^Tu \). Since

\[
\left( I - \frac{uu^T}{u^Tu} \right)^T \left( \frac{uu^T}{u^Tu} \right) = \left( uu^T - \frac{u(u^Tu)u^T}{u^Tu} \right) = \left( uu^T - uu^T \right) = 0,
\]

the projection matrices are orthogonal and by the Pythagorean Theorem,

\[
\left\| E \left( I - \frac{uu^T}{u^Tu} \right) \right\|_F^2 = \left\| E \right\|_F^2 + \left\| \frac{uu^T}{u^Tu} \right\|_F^2.
\]

By properties of the \( \ell_2 \) and Frobenius matrix norms, the last term can be simplified to

\[
\left\| \frac{uu^T}{u^Tu} \right\|_F = \frac{\left\| (Eu)u^T \right\|_F}{\| u \|_2^2} = \frac{\left\| (Eu)u^T \right\|_2}{\| u \|_2^2} = \frac{\left\| Eu \right\|_2}{\| u \|_2^2} = \frac{\left\| Eu \right\|_2}{\| u \|_2}.
\]

since \( \| u^T \|_2 = \| u \|_2 \). After rearranging the terms,

\[
\left\| E \left( I - \frac{uu^T}{u^Tu} \right) \right\|_F^2 = \left\| E \right\|_F^2 - \left( \frac{\| Eu \|_2}{\| u \|_2} \right)^2,
\]

which leads to the desired result. In addition, \( \sqrt{\alpha^2 - \beta^2} \leq \alpha - \beta^2/2\alpha \), so Equation A.6 is implied from Equation A.5. \( \square \)
A.3 Theorems for \( q \)-Superlinear Convergence

The first theorem in this section is taken from Dennis & Moré [7] and gives an alternative formulation of \( q \)-superlinear convergence of \( \{x^{(\nu)}\} \) to \( \bar{x} \) which will greatly simplify the proof of the second theorem.

**Theorem A.3.1** (Dennis & Moré (1974)). Let \( f : \mathbb{R}^n \to \mathbb{R}^n \) be defined in the open convex set \( D \subseteq \mathbb{R}^n \), with \( J \in \text{Lip}_\gamma(D) \), \( J(\bar{x}) \) nonsingular and \( \bar{x} \in D \).

In addition, let \( \{J_\nu\} \in \mathbb{R}^{n \times n} \) and for some \( x^{(0)} \in D \), \( \{x^{(\nu)}\} \) remains in \( D \) such that \( x^{(\nu)} \neq \bar{x} \) and \( \lim_{\nu \to \infty} x^{(\nu)} = \bar{x} \). Then,

\[
\lim_{\nu \to \infty} \frac{\|J_\nu - J(\bar{x})\| \Delta x^{(\nu)}}{\|\Delta x^{(\nu)}\|} = 0 \iff \{x^{(\nu)}\} \to \bar{x} \ q\text{-superlinearly.} \tag{A.7}
\]

This theorem is useful for our purpose here only in the forward direction, and thus the proof of the reverse direction will not be included below.

**Proof.** ( \( \implies \) ): From Equation A.1a, we have

\[
0 = J_\nu \Delta x^{(\nu)} + f(x^{(\nu)})
\]

\[
= \left(J_\nu - J(\bar{x})\right) \Delta x^{(\nu)} + f(x^{(\nu)}) + J(\bar{x}) \Delta x^{(\nu)}.
\]

Then,

\[
\implies -f(x^{(\nu+1)}) = \left(J_\nu - J(\bar{x})\right) \Delta x^{(\nu)} + \left(-f(x^{(\nu+1)}) + f(x^{(\nu)}) + J(\bar{x}) \Delta x^{(\nu)}\right)
\]

\[
\implies \frac{\|f(x^{(\nu+1)})\|}{\|\Delta x^{(\nu)}\|} \leq \frac{\|J_\nu - J(\bar{x})\| \Delta x^{(\nu)}}{\|\Delta x^{(\nu)}\|} + \frac{\|f(x^{(\nu+1)}) + f(x^{(\nu)}) + J(\bar{x}) \Delta x^{(\nu)}\|}{\|\Delta x^{(\nu)}\|}
\]

\[
\leq \frac{\|J_\nu - J(\bar{x})\| \Delta x^{(\nu)}}{\|\Delta x^{(\nu)}\|} + \frac{\gamma}{2} \left(\|e^{(\nu+1)}\| + \|e^{(\nu)}\|\right)
\]

with the last equality resulting from Lemma A.2.1. But

\[
\lim_{\nu \to \infty} x^{(\nu)} = \bar{x} \implies \lim_{\nu \to \infty} \|x^{(\nu)} - \bar{x}\| = 0 \implies \lim_{\nu \to \infty} \|e^{(\nu)}\| = 0
\]

86
and
\[ \lim_{\nu \to \infty} \frac{\| (J_\nu - J(\bar{x})) \Delta x(\nu) \|}{\| \Delta x(\nu) \|} = 0. \]
Therefore,
\[ \lim_{\nu \to \infty} \frac{\| f(x^{(\nu+1)}) \|}{\| \Delta x(\nu) \|} = 0. \]

From Lemma A.2.2,
\[ 0 = \lim_{\nu \to \infty} \frac{\| f(x^{(\nu+1)}) \|}{\| \Delta x(\nu) \|} \geq \lim_{\nu \to \infty} \frac{\alpha \| e^{(\nu+1)} \|}{\| e^{(\nu+1)} - e(\nu) \|} \]
\[ \geq \lim_{\nu \to \infty} \frac{\alpha \| e^{(\nu+1)} \|}{\| e^{(\nu+1)} \| + \| e^{(\nu)} \|} \]
\[ = \lim_{\nu \to \infty} \frac{\alpha r_\nu}{r_\nu + 1} \]
where \( r_\nu = \| e^{(\nu+1)} \| / \| e^{(\nu)} \| \). Then we must have
\[ \lim_{\nu \to \infty} r_\nu = 0 \implies \lim_{\nu \to \infty} \frac{\| e^{(\nu+1)} \|}{\| e^{(\nu)} \|} = 0. \]

\[ \Box \]

**Theorem A.3.2.** Let \( f : \mathbb{R}^n \to \mathbb{R}^m \), \( f \in C^1(D) \), open convex set \( D \subseteq \mathbb{R}^n \). Assume there exists \( \bar{x} \in \mathbb{R}^n \), \( r, \beta > 0 \) such that \( N(\bar{x}, r) \subset D \), \( f(\bar{x}) = 0 \), \( J(\bar{x})^{-1} \) exists and \( \| J(\bar{x})^{-1} \| \leq \beta \), and \( J \in \text{Lip}_\gamma(N(\bar{x}, r)) \).

There exist \( \varepsilon, \delta > 0 \) such that if \( \| x^{(0)} - \bar{x} \|_2 \leq \varepsilon \) and \( \| J_0 - J(\bar{x}) \|_2 \leq \delta \), then the sequence \( \{ x^{(\nu)} \} \) generated by Equation A.1a converges \( q \)-superlinearly to \( \bar{x} \). If we only assume \( \{ J_\nu \} \) satisfies Equation A.4, \( \{ x^{(\nu)} \} \) converges at least \( q \)-linearly to \( \bar{x} \).

The original proof also showed that Broyden’s method has at least \( q \)-linear convergence. For the sake of continuity, we will not take the time to include that here. However, there are several takeaways from the proof. It can be shown that
\[ \| e^{(\nu+1)} \| \leq \frac{1}{2} \| e^{(\nu)} \|. \] In fact, the proof ensures that \( \{ x^{(\nu)} \} \) converges to \( \bar{x} \) with
\[
\sum_{\nu=0}^{\infty} \| e^{(\nu)} \| < \infty.
\]
It also can be shown that \( \| E_\nu \|_F \leq 2\delta \) for \( \nu \in \mathbb{N}(0, \infty) \) and
\[
\sum_{\nu=0}^{\infty} \| e^{(\nu)} \| \leq 2\varepsilon.
\]

\textbf{Proof.} \((q\text{-superlinear convergence}): \) From Broyden’s update, we have
\[
J_{\nu+1} - J(\bar{x}) = J_\nu - J(\bar{x}) + \frac{(\Delta f^{(\nu)} - J_\nu \Delta x^{(\nu)}) \Delta x^{(\nu)^T}}{\Delta x^{(\nu)^T} \Delta x^{(\nu)}}
\]
\[
= J_\nu - J(\bar{x}) + \frac{(\mathcal{J}(\bar{x}) \Delta x^{(\nu)} - J_\nu \Delta x^{(\nu)}) \Delta x^{(\nu)^T}}{\Delta x^{(\nu)^T} \Delta x^{(\nu)}}
\]
\[
+ \frac{(\Delta f^{(\nu)} - \mathcal{J}(\bar{x}) \Delta x^{(\nu)}) \Delta x^{(\nu)^T}}{\Delta x^{(\nu)^T} \Delta x^{(\nu)}}
\]
\[
= (J_\nu - J(\bar{x})) \left[ I - \frac{\Delta x^{(\nu)} \Delta x^{(\nu)^T}}{\Delta x^{(\nu)^T} \Delta x^{(\nu)}} \right] + \frac{(\Delta f^{(\nu)} - \mathcal{J}(\bar{x}) \Delta x^{(\nu)}) \Delta x^{(\nu)^T}}{\Delta x^{(\nu)^T} \Delta x^{(\nu)}}
\]

Then for either the Frobenius norm or the \( \ell_2 \) matrix norm, we have
\[
\| J_{\nu+1} - J(\bar{x}) \| \leq \| J_\nu - J(\bar{x}) \| \left\| I - \frac{\Delta x^{(\nu)} \Delta x^{(\nu)^T}}{\Delta x^{(\nu)^T} \Delta x^{(\nu)}} \right\|_F + \frac{\left\| (\Delta f^{(\nu)} - \mathcal{J}(\bar{x}) \Delta x^{(\nu)}) \Delta x^{(\nu)^T} \right\|_F}{\| \Delta x^{(\nu)} \|_F^2}.
\]
(A.8)

Now define \( E_\nu = J_\nu - J(\bar{x}) \). From the equation above,
\[
\| E_{\nu+1} \| \leq \| E_\nu \| \left\| I - \frac{\Delta x^{(\nu)} \Delta x^{(\nu)^T}}{\Delta x^{(\nu)^T} \Delta x^{(\nu)}} \right\|_F + \frac{\left\| (\Delta f^{(\nu)} - \mathcal{J}(\bar{x}) \Delta x^{(\nu)}) \Delta x^{(\nu)^T} \right\|_F}{\| \Delta x^{(\nu)} \|_F^2}.
\]
(A.9)

From Lemma A.2.1, it is easy to see
\[
\| \Delta f^{(\nu)} - \mathcal{J}(\bar{x}) \Delta x^{(\nu)} \|_2 \leq \frac{\gamma}{2} \left( \| e^{(\nu+1)} \|_2 + \| e^{(\nu)} \|_2 \right) \| \Delta x^{(\nu)} \|_2
\]
by setting \( v = x^{(\nu+1)} \), \( u = x^{(\nu)} \), and \( x = \bar{x} \). Combining this with \( \| e^{(\nu+1)} \| \leq \frac{1}{2} \| e^{(\nu)} \| \) and Lemma A.2.5, gives
\[
\| E_{\nu+1} \|_F \leq \| E_\nu \|_F - \frac{\| E_\nu \|_F \| \Delta x^{(\nu)} \|_2^2}{2 \| E_\nu \|_F \| \Delta x^{(\nu)} \|_2^2} + \frac{3\gamma}{4} \| e^{(\nu)} \|_2.
\]

88
Equivalently, the inequality can be rewritten as

$$\frac{\|E_\nu \Delta x^{(\nu)}\|_2^2}{\|\Delta x^{(\nu)}\|_2^2} \leq 2\|E_\nu\|_F \left[\|E_\nu\|_F - \|E_{\nu+1}\|_F + \frac{3\gamma}{4}\|e^{(\nu)}\|_2\right].$$  \hspace{1cm} (A.10)

Summing both sides of the inequality for $\nu = 0, 1, \ldots, i$, where $i \in \mathbb{N}[0, \infty)$, and simplifying, we get

$$\sum_{\nu=0}^{i} \frac{\|E_\nu \Delta x^{(\nu)}\|_2^2}{\|\Delta x^{(\nu)}\|_2^2} \leq 2(2\delta) \left[\sum_{\nu=0}^{i} \left(\|E_\nu\|_F - \|E_{\nu+1}\|_F\right) + \frac{3\gamma}{4}\sum_{\nu=0}^{i}\|e^{(\nu)}\|_2\right]$$

$$\leq 4\delta \left[\|E_0\|_F - \|E_{i+1}\|_F + \frac{3\gamma\varepsilon}{2}\right]$$

$$\leq 4\delta \left[\|E_0\|_F + \frac{3\gamma\varepsilon}{2}\right] \leq 4\delta \left[2\delta + \frac{3\gamma\varepsilon}{2}\right]$$

Since this inequality holds true for any $i$,

$$\sum_{\nu=0}^{\infty} \frac{\|E_\nu \Delta x^{(\nu)}\|_2^2}{\|\Delta x^{(\nu)}\|_2^2} < \infty$$

and thus

$$0 = \lim_{\nu \to \infty} \frac{\|E_\nu \Delta x^{(\nu)}\|_2}{\|\Delta x^{(\nu)}\|_2} = \lim_{\nu \to \infty} \frac{\|J_\nu - J(\bar{x})\Delta x^{(\nu)}\|_2}{\|\Delta x^{(\nu)}\|_2}$$

so by Theorem A.3.1 we are done. \qed


