ABSTRACT

Title of Dissertation: HOMOTOPY CONTINUATION METHODS FOR PHASE RETRIEVAL
David J Bekkerman
Doctor of Philosophy, 2021

Dissertation Directed by: Professor Radu Balan
Department of Mathematics

In this dissertation, we discuss the problem of recovering a signal from a set of phaseless measurements. This type of problem shows up in numerous applications and is known for its numerical difficulty. It finds use in X-ray Crystallography, Microscopy, Quantum Information, and many others. We formulate the problem using a non-convex quadratic loss function whose global minimum recovers the phase of the measurement.

Our approach to this problem is via a Homotopy Continuation Method. These methods have found great use in solving systems of nonlinear equations in numerical algebraic geometry. The idea is to initialize the solution of a related system at a known global optimal, then continuously deform the criterion and follow the solution path until we find the minimum of the desired loss function. We analyze convergence properties and asymptotic results for these algorithms, as well as gather some numerical statistics. The main contribution of this thesis is deriving conditions for convergence of the algorithm and an asymptotic rate for when these conditions are satisfied. We also show that the algorithm achieves good numerical accuracy.

The dissertation is split into several chapters, and further divided by the real and complex case. Chapter 1 gives some background to Abstract Phase Retrieval and Homotopy Continuation Methods. Chapter 2 covers the nature of the algorithm (named the Golden Retriever), gives a summary and description of the theoretical results, and shows some numerical results. Chapter 3 covers the details of the derivation and results in the real case, and Chapter 4 covers the same for the complex case.

HOMOTOPY CONTINUATION METHODS FOR PHASE RETRIEVAL

by

David J Bekkerman

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2021

Advisory Committee:
Professor Radu Balan, Chair/Advisor
Professor Wojciech Czaja
Professor Leonid Koralov
Professor Min Wu
Professor Ramani Duraiswami, Dean's Representative

HOMOTOPY BOUND:

TALE OF A GOLDEN PHASE RETRIEVER

Dedication

For Jessica.

Acknowledgments

I would like to thank my advisor, Radu Balan, for introducing me to this problem, for always being willing to get in the trenches with me and work through the details, for mathematical and financial support, and for overall invaluable life advice.

To my friends in my office, Gilles, Dani, Zack, Jian-long, Charles and J.T. (listed up to permutation), you made the university feel like a home for me, and that alone made the commute worth it.

To friends outside of graduate school, I would like to thank Chinami and Yuichi for staying close through the years.

Thank you to my parents and siblings for all your support throughout my life. Finally thank you to Jessica, for being there for me when I needed you most.

Table of Contents

List of Figures

List of Mathematical Symbols

$\mathcal{F}$	Set of Frame Vectors, usually Finite Dimensional
$\mathbb{R}$	Real Numbers
$\mathbb{C}$	Complex Numbers
$\mathcal{N}(0, I)$	Standard Normal Distribution
$\mathcal{CN}(0, I)$	Complex Standard Normal Distribution
Tr	Trace of a Matrix
$\mathbb{S}_r^k$	Sphere of Radius r in $\mathbb{R}^k$
$\mathbb{E}$	The Expected Value of a Random Variable, typically over a Normal Distribution
m	Number of Frame Vectors
n	Dimension of the Frame Vectors
a_0	Lower Bound on a $2 \to 4$ Matrix Norm
b_0	Upper Bound on a $2 \to 4$ Matrix Norm
$R(x)$	Fundamental $n \times n$ Matrix associated to Real Phase Retrieval
I_k	The $k \times k$ Identity Matrix
$J(x, \lambda)$	J-criterion, non-convex criterion associated with phase retrieval
$F(x, \lambda)$	$\nabla_x J(x, \lambda)$
$Hess(x, \lambda)$	Shorthand for the Hessian of the $J - criterion$ at a point (x, λ)
$\varphi(\lambda)$	A suitable reference path
J	The $2n \times 2n$ symplectic matrix $\begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}$
$\Omega(x, \lambda)$	The Ω-criterion, non-convex criterion associated with complex phase retrieval

Chapter 1: Introduction

Phase retrieval is the problem of recovering the phase of a signal from the magnitudes (or the magnitudes squared) of its linear measurements. The applications range from X-ray Crystallography, where we want to discover the molecular structure of a crystal by using X-rays [1], to Quantum Tomography, where we want to recover a quantum state from a series of independent measurements on identical states [2]. It also finds applications in Speech Recognition [3]. In all these cases, the phase of the measurement is lost, and we want to recover the signal as best as possible up to this phase. To this end, several algorithms have been proposed. We propose and analyze the use of a Homotopy Continuation Algorithm to recover the signal. We named this method the Golden Retriever. Homotopy Continuation Methods have proven to be a useful tool in Numerical Algebraic Geometry, and we aim to use it to solve the system of polynomial equations that arise in Phase Retrieval. We study some of the convergence and analytical properties of such an algorithm. For a background on Homotopy Continuation, please see the section 1.3.

1.1 Overview

Phase retrieval began from trying to reconstruct a function from the magnitudes of its Fourier coefficients. It has now expanded to a branch called abstract Phase Retrieval, which states the problem more generally.

Let H be a Hilbert Space over either $\mathbb{R}$ or $\mathbb{C}$. Let I be a finite or countable index set. Define a set of vectors to be a frame set, $\mathcal{F} = \{f_1, f_2, ...\}$, indexed by I, if there exist positive constants A and B such that for every $v \in H$

$$A||v||^2 \leq \sum_{k \in I} |\langle v, f_k \rangle|^2 \leq B||v||^2 \tag{1.1}$$

We call A the lower frame bound of $\mathcal{F}$, and B the upper frame bound of $\mathcal{F}$. In the finite case, this turns out to be equivalent to $\mathcal{F}$ being a spanning set of vectors (see [4] for more information about frames).

For a frame set, $\mathcal{F} = \{f_i, i \in I\}$ we define two operators

$$\alpha_{\mathcal{F}}(x) = (|\langle x, f_i \rangle|)_{i \in I} \tag{1.2}$$

$$\beta_{\mathcal{F}}(x) = (|\langle x, f_i \rangle|^2)_{i \in I} \tag{1.3}$$

By linearity, it is clear that if $|c| = 1$, then $\alpha(cx) = \alpha(x)$ and $\beta(cx) = \beta(x)$. Therefore we define an equivalence relation where we say $x \sim y$ if there exists a constant c of magnitude 1 such that $x = cy$. If we quotient out by this relation, we say that the frame set $\mathcal{F}$ is **phase retrievable** if the corresponding map on the

quotient space is injective.

$$\alpha : \mathcal{H}/\sim \to \mathbb{R}_+^I \tag{1.4}$$

1.2 Injectivity

We now focus on the case when I is a finite index set. The infinite case can be found in many places, such as [5] or [6].

As with many of the results we present, we separate the real and complex cases.

Let $H = \mathbb{R}^n$ and $|I| = m$ and define

$$R(x) = \frac{1}{m} \sum_{i=k}^{m} |\langle x, f_k \rangle|^2 f_k f_k^T \tag{1.5}$$

This matrix plays an important role not just in the injectivity, but throughout many results presented later on as well.

The following theorem and proofs are taken from [7].

Theorem 1.2.1 (Complement Property [7])**.** *The following are equivalent*

1. *$\alpha_{\mathcal{F}}$ is an injective map on $\mathbb{R}^n/\sim$*

2. *For any disjoint partition of the frame set, $\mathcal{F} = \mathcal{F}_1 \cup \mathcal{F}_2$, either $\mathcal{F}_1$ spans H or $\mathcal{F}_2$ spans H*

Proof.

- (1) $\Rightarrow$ (2) Assume that there exists a subset $\mathcal{F}_1 \subset \{f_1, ..., f_m\}$ such that neither $\mathcal{F}_1$ or $\mathcal{F}_1^C$ spans H. Hence there exist vectors x, y such that $x \perp span(\mathcal{F}_1)$ and

$y \perp span(\mathcal{F}_1^C)$. Then a direct check shows $\alpha_{\mathcal{F}}(x+y) = \alpha_{\mathcal{F}}(x-y)$. We are left with showing that $x+y$ is not a multiple of $x-y$. If $x+y = x-y$, then $y = 0$ which we know is not possible, and if $x+y = y-x$, then $x = 0$, which is not possible. Therefore, we found two vectors which map to the same output, which are not in the same equivalence class, so α is not an injective map.

- (2) $\Rightarrow$ (1) Suppose that $\alpha_{\mathcal{F}}(x) = \alpha_{\mathcal{F}}(y)$, for $x, y \in H/\sim$. This means for all $0 \leq k \leq m$, $|\langle x, f_k\rangle| = |\langle y, f_k\rangle|$. We partition the set $\mathcal{F}$ into two subsets

$$\mathcal{F}_1 = \{f_k : \langle x, f_k\rangle = -\langle y, f_k\rangle\}$$

$$\mathcal{F}_2 = \{f_k : \langle x, f_k\rangle = \langle y, f_k\rangle\}$$

Note that $x+y \perp \mathcal{F}_1$ and $x-y \perp \mathcal{F}_2$. Assume that $span(\mathcal{F}_1) = H$, then $x+y = 0$ so $x = -y$, so they are in the same equivalence class. Similarly, if $span(\mathcal{F}_2) = H$, then $x-y = 0$ so $x = y$. Therefore, either way, the map $\alpha_{\mathcal{F}}$ is injective.

□

The following is an important injectivity result from [7].

Theorem 1.2.2 ([7]). *Let $\mathcal{F} = \{f_1, ..., f_m\} \subset H$ be m vectors and let H be a subset of $\mathbb{R}^n$. The following are equivalent*

1. *For any disjoint set of the frame vectors, $\mathcal{F} = \mathcal{F}_1 \cup \mathcal{F}_2$, either $\mathcal{F}_1$ spans H or $\mathcal{F}_2$ spans H.*

2. *For any 2 vectors $x, y \in H$, if $x \neq 0$ and $y \neq 0$ we have*

$$\sum_{k=1}^{m} |\langle x, f_k\rangle|^2 |\langle y, f_k\rangle|^2 > 0$$

3. *There exists a positive real constant $a_0 > 0$ such that for all $x, y \in H$*

$$\sum_{k=1}^{m} |\langle x, f_k\rangle|^2 |\langle y, f_k\rangle|^2 > a_0 ||x||^2 ||y||^2$$

4. *There exists a positive real constant $a_0 > 0$ such that for all $x \in H$*

$$R(x) \geq \frac{a_0}{m} ||x||^2 I$$

Proof.

- (1) $\Rightarrow$ (2) We prove this by contradiction. Assume we have two vectors $x, y \in H$ with $x, y \neq 0$, but $\sum_{k=1}^{m} |\langle x, f_k\rangle|^2 |\langle y, f_k\rangle|^2 = 0$. Then we have $\langle x, f_k\rangle \langle y, f_k\rangle = 0$ for all $1 \leq k \leq m$. Form the set $\mathcal{F}_1 = \{f_k | \langle x, f_k\rangle = 0\}$. Since x is orthogonal to all of $\mathcal{F}_1$, it is clear that $\mathcal{F}_1$ cannot span all of H. Similarly, we know that y is orthogonal to $\{\mathcal{F}\} \setminus \mathcal{F}_1 = \mathcal{F}_1^C$, so $\mathcal{F}_1^C$ cannot span all of H, therefore, the complement property is not satisfied.

- (2) $\Rightarrow$ (3) Since H is finite, the unit sphere $\mathbb{S}_1(H)$ is compact, and so is $\mathbb{S}_1(H) \times \mathbb{S}_1(H)$. Since the map

$$(x, y) \to \sum_{k=1}^{m} |\langle x, f_k\rangle|^2 |\langle y, f_k\rangle|^2$$

is continuous, by compactness we have that there exists a constant

$$a_0 = min_{(x,y)\in\mathbb{S}_1(H)\times\mathbb{S}_1(H)} \sum_{k=1}^{m} |\langle x, f_k\rangle|^2 |\langle y, f_k\rangle|^2 > 0$$

By homogeneity, we have that for $x, y \in H$, $x, y \neq 0$

$$\sum_{k=1}^{m} |\langle x, f_k\rangle|^2 |\langle y, f_k\rangle|^2 = ||x||^2||y||^2 \sum_{k=1}^{m} |\langle \frac{x}{||x||^2}, f_k\rangle|^2 |\langle \frac{y}{||y||^2}, f_k\rangle|^2 \geq a_0||x||^2||y||^2$$

If $x = 0$ or $y = 0$, then statement (2) is still satisfied.

- (3) $\Rightarrow$ (4) This follows from the property of quadratic forms

$$||R(x)|| = max_{||e||=1}\langle R(x)e, e\rangle = \frac{1}{m}\sum_{k=1}^{m} |\langle x, f_k\rangle|^2 |\langle e, f_k\rangle|^2$$

$$\geq \frac{1}{m} a_0 max_{||e||=1}||x||^2||e||^2 = \frac{a_0}{m}||x||^2$$

Therefore, we have that $R(x) \geq \frac{a_0}{m}||x||^2 I$

- (4) $\Rightarrow$ (1) We show the contrapositive. Assume (1) is not true, we aim to show (4) is not true either. Therefore there exists a partition $\mathcal{F} = \mathcal{F}_1 \cup \mathcal{F}_2$ such that $\mathcal{F}_1$ doesn't span H and $\mathcal{F}_2$ doesn't span H. Therefore there exists an $x \perp span(\mathcal{F}_1)$ and also a $y \perp span(\mathcal{F}_2)$, therefore for each f_k, either $\langle x, f_k\rangle = 0$ or $\langle y, f_k\rangle = 0$, Therefore $\langle x, f_k\rangle\langle y, f_k\rangle = 0 \Rightarrow \langle R(x)y, y\rangle = 0$

□

There are several things to note here. First is the following corollary, which

follows from the Complement Property.

Corollary 1.2.3. *If $H = \mathbb{R}^n$, then $\alpha_{\mathcal{F}}$ being injective on the quotient space is equivalent to any of the conditions of Theorem 1.2.2.*

Next, we note that $\alpha_{\mathcal{F}}$ being injective is equivalent to $\beta_{\mathcal{F}}$ being injective. For us, it is of particular interest to use the map $\beta_{\mathcal{F}}$ instead of $\alpha_{\mathcal{F}}$, because $\beta_{\mathcal{F}}$ is differentiable everywhere.

Now we state the theorem in the complex case. To do so, we need to begin with some definitions that follow from the process of realification (introduced in [8]). Define $J = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}$, and $\varphi_k = \begin{bmatrix} Re(f_k) \\ Im(f_k) \end{bmatrix} \in \mathbb{R}^{2n}$. Further, define

$$\Phi_k = \varphi\varphi^T + J\varphi\varphi^T J^T$$

again define the important $2n \times 2n$ matrix

$$\widetilde{\Gamma}(\xi) = \frac{1}{m}\sum_{k=1}^{m} \phi_k \xi\xi^T \phi_k^T \tag{1.6}$$

Now we state the following theorem, the proof of which can be found in [7].

Theorem 1.2.4. *([7]) The following statements are equivalent*

1. *$\beta_{\mathcal{F}}$ is injective on $\mathbb{C}^n/\sim$*

2. *For any $\xi \in \mathbb{R}^{2n}$, $\xi \neq 0$, $rank(\widetilde{\Gamma}(\xi)) = 2n - 1$*

3. For any $\xi \in \mathbb{R}^{2n}$, $\xi \neq 0$, there exists a constant $\alpha_0 > 0$ such that

$$\widetilde{\Gamma}(\xi) \geq \alpha_0 ||\xi||^2 P^{\perp}_{J\xi}$$

where $P^{\perp}_{J\xi} = I - J\xi\xi^T J^T$ is the orthogonal complement of the span of $J\xi$

1.3 Homotopy Continuation Methods

The technique we use for phase retrieval is known as a Homotopy Continuation Method. Homotopy Continuation Methods are a tool in Numerical Algebraic Geometry that can be used to solve a system of polynomial equations [9]. These methods can vary, sometimes being set up to give one solution, and sometimes all solutions. The advantage to these methods is that no approximations are needed to get solutions, so almost nothing needs to be known beforehand. These methods have been applied extensively to applications in Economics [10, 11], Mathematics[12], Engineering [13], and many other fields. See [14] for many more applications.

The overarching philosophy behind these algorithms is to find a solution to a simple problem, and then deform the simple problem into the desired complicated problem, and in the process deform the solution of the simple problem into the solution to the complicated problem. For the Phase Retrieval problem we will study, the simple problem will be an eigenvalue equation, and the complicated problem is minimizing the mean square error of a criterion, a non-convex quadratic loss function, associated with Phase Retrieval.

To formalize the notion of deformation, we review the concept of a homotopy.

Definition 1.3.0.1 (Homotopy [14]). *Let X, Y be two spaces and let I be a unit interval ($0 < t < 1$). Two maps $f : X \to Y$ and $g : X \to Y$ are called* ***homotopic*** *if there exists a continuous map*

$$H : X \times I \to Y$$

such that

$$H(x, 0) = f(x)$$

$$H(x, 1) = g(x)$$

for all $x \in X$.

To use this to solve a system of nonlinear polynomial equations (say over $\mathbb{R}^n$), we take two copies of $\mathbb{R}^n$, one for $t = 0$, denoted $\mathbb{R}^n \times \{0\}$, and the other for $t = 1$, denoted by $\mathbb{R}^n \times \{1\}$. Now we solve the problem in $\mathbb{R}^n \times \{0\}$, which is assumed to be easy by construction, and then trace the solution through the homotopy in $\mathbb{R}^n \times I$ to $\mathbb{R}^n \times \{1\}$ and with luck, find the solution to the system of equations there as well. [14]

Equivalent images to the following figures were originally drawn in [14].

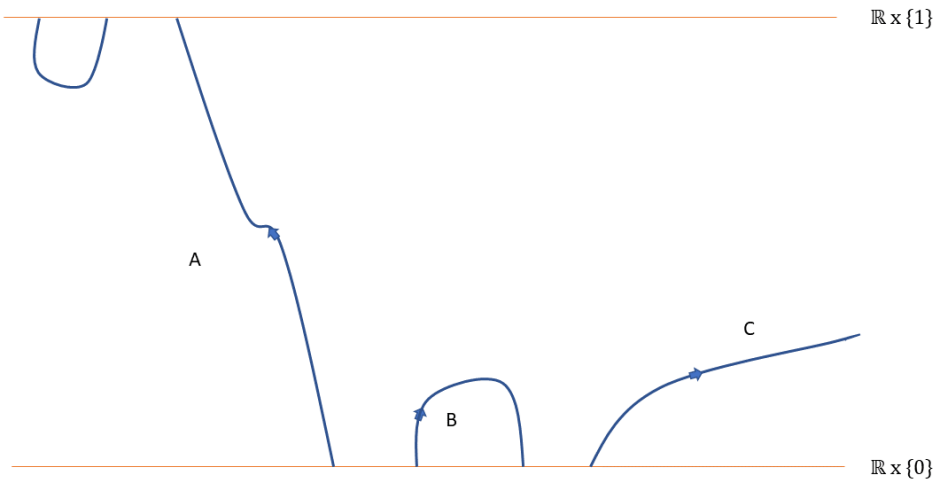

Figure 1.1: These are examples of homotopy paths the algorithm may make, with A yielding a traceable path to a solution in $\mathbb{R} \times \{1\}$

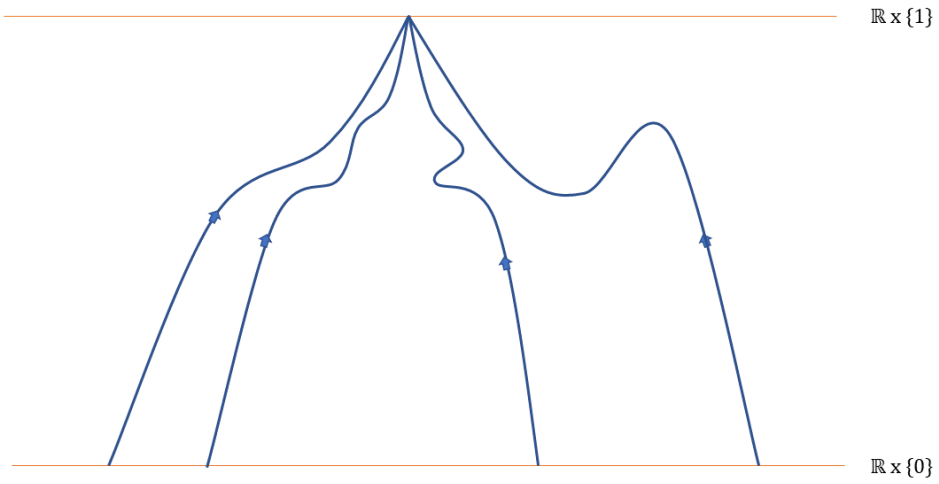

Figure 1.2: These are examples of homotopy paths which all lead to the same solution in $\mathbb{R} \times \{1\}$

In the first figure, we see three labeled paths, A, B, C. In path A, we have a solution to simple system in $\mathbb{R} \times \{0\}$ which converges to a solution to the complicated

system in $\mathbb{R} \times \{1\}$. Some paths, like path B, will never make it to $\mathbb{R} \times \{1\}$, and in this case, the path turns around and goes back to another solution at $\mathbb{R} \times \{0\}$. The third type of path C, goes off to infinity, and never converges to either $\mathbb{R} \times \{1\}$ or back to $\mathbb{R} \times \{0\}$. In the phase retrieval case, these types of paths are not possible by the boundedness properties we will show.

In the second figure, we see that all homotopy paths starting at $\{0\}$ converge to the same point in $\mathbb{R} \times \{1\}$. With the assumptions we will make, these type of paths will be impossible as well.

We now follow [9] to show how to apply these techniques to solve nonlinear systems. To apply these to solving systems of polynomial equations, say we want to solve $F(x) = 0$ with $x \in \mathbb{R}^n$. We do so by defining a smooth homotopy $H : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ such that $H(x, 0) = G(x)$ and $H(x, 1) = F(x)$, where $G : \mathbb{R}^n \to \mathbb{R}^n$ is a trivial smooth map having known zero points. One can choose a convex homotopy such as

$$H(x, \lambda) = \lambda F(x) + (1 - \lambda)G(x)$$

One can ask several immediate questions

1. When is it assured that a curve $c(s) \in H^{-1}(0)$ exists and is smooth?

2. If such a curve exists, when is it assured that it will intersect the target homotopy level $\lambda = 0$ in a finite length?

3. How can we numerically trace such a curve?

The first is answered by the implicit function theorem, if the Jacobian H' has

full rank $rank(H') = n$, then such a smooth curve will exist (at least locally).

Generally, for polynomial systems, the second requires the use of some bounding conditions to make sure the curve doesn't run to infinity before intersecting at $\lambda = 0$.

The third is usually used by a combination of a predicator and corrector step, such as an Euler Step, followed by a fixed point correction.

1.4 Overview of some Algorithms for Phase Retrieval

In this section, we take a look at existing algorithms for Phase Retrieval. In all of the cases, we assume that the frame set is Phase Retrievable.

1.4.1 PhaseLift

In the PhaseLift algorithm, it is assumed we have quadratic measurements of the form $y_k = \{|\langle x, f_k\rangle|^2\}$, and it recognizes this can be lifted up and interpreted as linear measurements on xx^*, so the quadratic constraints turn to linear constraints. Then one can note

$$|\langle x, f_k\rangle|^2 = Tr(x^* f_k f_k^* x) = Tr(f_k f_k^* x x^*) := Tr(F_k X)$$

Then the problem turns into one of matrix completion, given by solving the following

$$\begin{aligned} \text{find} \quad & X \\ \text{subject to} \quad & \mathcal{F}(X) = y \\ & X \geq 0 \\ & rank(X) = 1 \end{aligned} \iff \begin{aligned} \text{minimize} \quad & rank(X) \\ \text{subject to} \quad & \mathcal{F}(X) = y \\ & X \geq 0 \end{aligned}$$

In general the problem of Rank Minimization is known to be NP-hard, so the authors suggest instead to relax the constraints and solve a trace minimization problem, which can be done with a semidefinite program.

The problem would then be to solve

$$\begin{aligned} \text{minimize} \quad & Tr(X) \\ \text{subject to} \quad & \mathcal{F}(X) = y \\ & X \geq 0 \end{aligned}$$

After solving this SDP, if the solution has rank 1 (which would guarantee minimization to the Rank Minimization) one would then factorize it (through orthogonal diagonalization, for instance), and get the solution to the phase retrieval problem as well.

Detailed analysis of this algorithm can be found in [15].

1.4.2 Wirtinger Flow

Let $\boldsymbol{x} \in \mathbb{C}^n$. The problem is to recover $\boldsymbol{z}$ from m phaseless linear measurements. Let $\mathcal{F} = \{f_k\}_{k=1}^m$ be a finite frame which spans our vector space $\mathbb{C}^n$.

$$y_k = |\langle x, f_k \rangle|^2 + w_k \text{ for } k = 1, ..., m \tag{1.7}$$

where $w_k \sim \mathcal{CN}(0, \sigma_w^2)$.

The algorithm for Wirtinger Flow is a gradient descent algorithm which minimizes the loss function $\ell(x, y) = |(x - y)|^2$.

Hence, we want to minimize the loss function $I(x) = \frac{1}{2m} \sum_k \ell(y_k, |\langle x, f_k \rangle|^2)$, so we want to find

$$\arg\min_x I(x) = \arg\min_x \frac{1}{2m} \sum_{k=1}^{n} (y_k - |\langle x, f_k \rangle|)^2)^2 \tag{1.8}$$

After initialization, the update rules are given by a gradient descent procedure

$$x_{t+1} = x_t - \frac{\mu_{t+1}}{||x_0||^2} (\frac{1}{m} \sum_{r=1}^{m} (|\langle x_t, f_k \rangle|^2 - y_k) f_k f_k^* x_k) = x_t - \frac{\mu_{t+1}}{||x_0||^2} \nabla_x I(x)$$

For initialization, we note that we define

$$Y_0 = \frac{1}{m} \sum_{k=1}^{m} y_k f_k f_k^*$$

and let x_0 be the principal eigenvector of Y_0. Furthermore , set

$$\lambda^2 = n \frac{\sum_{k=1}^{m} y_k}{\sum_{k=1}^{m} ||f_k||^2}$$

and set $||x_0|| = \lambda$

It is worth noting, if the frame set is real then Y_0 is exactly R_0 given in the

real golden retriever algorithm, and in the complex case Y_0 is related to the Γ_0 in the complex golden retriever through the realification process.

Although the algorithm and the analysis are quite different, there are a lot of similarities between the Golden Retriever Algorithm and Wirtinger Flow.

Now there are many variants of Wirtinger Flow (see for instance [16], [17]) that deal with different loss functions. The original Wirtinger Flow converged with high probability when the number of frame vectors is of the order $m = O(nlogn)$. We are looking at the same loss function as Wirtinger Flow, but the proof strategy we employ is very different, as it will be a proof based on a perturbation analysis.

1.4.3 Approximate Message Passing

Let $\boldsymbol{x} \in \mathbb{C}^n$. The problem studied in Approximate Message Passing (AMP) is to recover $\boldsymbol{x}$ from m phaseless linear measurements of the form

$$y_k = |\sum_{i=1}^{n} A_{ki}x_i| + w_k \text{ for } k = 1, ..., m \tag{1.9}$$

where $w_k \sim \mathcal{CN}(0, \sigma_w^2)$.

Hence, the goal is to minimize:

$$\min_{\boldsymbol{x}} \sum_{k=1}^{n} (y_k - |(\boldsymbol{A}\boldsymbol{x})_k|)^2) + \frac{\mu_k}{2}\|\boldsymbol{x}\|_2^2 \tag{1.10}$$

Notice that a regularization term is included, $\frac{\mu_k}{2}\|\boldsymbol{x}\|_2^2$. This is known to reduce the variance of an estimator and because without the regularization term, the loss

function would be non-convex, this is expected to be useful even in the noiseless setting [18].

As with many problems in estimating probabilities , one approach is to attempt to make a probabilistic graphical model out of this (in the same way one can model speech recognition as a Hidden Markov Model).

To begin to do this, one needs to first have a PDF defined on a graphical model.

One can examine the minimization criterion $\min_{\boldsymbol{x}} \sum_{k=1}^{n} (y_k - |(\boldsymbol{A}\boldsymbol{x})_k|)^2) + \frac{\mu_k}{2}\|\boldsymbol{x}\|_2^2$ and construct a corresponding joint PDF

$$p(x) = \frac{1}{Z} \prod_{a=1}^{m} exp[-\beta(y_a - |(Ax)_a|)^2] \prod_{i=1}^{n} exp(-\beta \cdot \frac{\mu}{2} x_i^2) \tag{1.11}$$

Then the next step is to approximate this joint PDF. AMP accomplishes this in several steps [19]:

1. Derive the Belief-Propogation update rules for $p(x)$
2. Approximate the BP update rules
3. Find the Message update rules in the limit $\beta \to \infty$

The BP message update rules [20] act on the graphical model which in this case, a fully connected bi-partite graph with n vertices being for the x_i, and m others for the y_j. Given such a graphical model, the message update rules can be written (after some rearranging) in the following form

$$m^t_{a\to i}(x_i) = \int_{\boldsymbol{x}\setminus i} f(y_a, (Ax)_a) \prod_{j\neq i} dm^t_{j\to a}(x_j)$$

$$m^{t+1}_{i\to a}(x_i) = \prod_{b\neq a} m^t_{b\to i}(x_i) \cdot exp(-\beta\frac{\mu}{2}x_i^2)$$

$$\text{where } f(y,z) := exp(-\beta(y-|z|)^2)$$

After applying this to the AMP case and simplifying with asymptotic approximations, one gets the following algorithm which is called AMP.A

$$\boldsymbol{p}^t = \boldsymbol{A}\boldsymbol{x}^t - \frac{2}{\delta} g(\boldsymbol{p}^{t-1}, \boldsymbol{y})$$

$$\boldsymbol{x}^{t+1} = 2[-div_p(g_t)\cdot \boldsymbol{x}^t + \boldsymbol{A}^H g(\boldsymbol{p}^t, \boldsymbol{y})]$$

Here the functions are defined

- $div_p(g_t) = \frac{1}{m}\sum_{a=1}^{m} \frac{y_a}{2|p_a^t|} - 1$
- $g(p,y) = y\cdot\frac{p}{|p|} - p$

Now the analysis of the convergence for this system is governed by a dynamical system, which in the noiseless case under sufficient conditions on the asymptotic redundancy, converges. Details on this can be found [18].

Chapter 2: Overview of the Golden Retriever

In this chapter, we state the golden retriever algorithm as a PC (Predictor-Corrector) algorithm and we outline the main results of the thesis. The derivations and proofs for these results can be found in subsequent chapters. We also show some numerical results.

Let $\boldsymbol{x} \in \mathbb{C}^n$. The problem of phase retrieval for us is to recover $\boldsymbol{x}$ from m phaseless linear measurements. Let $\mathcal{F} = \{f_k\}_{k=1}^m$ be a finite frame which spans the vector space $\mathbb{C}^n$.

$$y_k = |\langle x, f_k \rangle|^2 + w_k \text{ for } k = 1, ..., m \tag{2.1}$$

where $w_k \sim \mathcal{CN}(0, \sigma_w^2)$.

Our objective function to minimize is a regularized quadratic loss function given by

$$\Omega(x, \lambda) = \frac{1}{4m} \sum_{k=1}^{m} (y_k - |\langle x, f_k \rangle|^2)^2 + \frac{\lambda}{2} \langle Qx, x \rangle \tag{2.2}$$

where Q is a hermitian positive definite matrix and λ is a real parameter ($\lambda \geq 0$), which we will use to homotope our solution to the desired $\lambda = 0$. Note that at $\lambda = 0$, the quadratic objective function we are minimizing is equivalent to the one in the Wirtinger Flow algorithm.

We also specialize the problem to the real case as and write it out explicitly as follows

Let $\boldsymbol{x} \in \mathbb{R}^n$. We wish to to recover $\boldsymbol{x}$ from m phaseless linear measurements. Let $\mathcal{F} = \{f_k\}_{k=1}^m$ be a finite frame which spans our vector space $\mathbb{R}^n$.

$$y_k = |\langle x, f_k \rangle|^2 + \sigma_k \text{ for } k = 1, ..., m \tag{2.3}$$

where $\sigma_k \sim \mathcal{N}(0, \sigma^2)$.

Hence, we want to minimize the same

$$J(x, \lambda) = \frac{1}{4m} \sum_{k=1}^{m} (y_k - |\langle x, f_k \rangle|^2)^2 + \frac{\lambda}{2} \langle Qx, x \rangle \tag{2.4}$$

where Q is now a symmetric positive definite matrix.

The minimization objective is not convex, so it may have many stationary points and local minima. Minimizing non-convex objectives such as this is in general known to be NP-hard. See [21] for an example of when convergence to a local minimum is known to be NP-hard.

It is worth noting that in many of the theoretical results, we specialize further to the noiseless case and we take $Q = I$.

2.1 The Real Golden Retriever Algorithm

To define the algorithm, we need to know the initialization, and we need to specify the update rules.

Algorithm 1: Real Golden Retriever Initialization

Input : Observations $\{y_k\}$, the frame set $\{f_k\}$, a positive symmetric semidefinite matrix Q, and an initial step size μ_0

Define e_1 to be the eigenvector corresponding to the largest eigenvalue (denoted λ_1) of

$$R_0 = \frac{1}{m}\sum_{k=1}^{m} y_k f_k f_k^T$$

Set

$$c = \sqrt{\frac{\mu_0 \langle Q e_1, e_1 \rangle}{\frac{1}{m}\sum_{k=1}^{m} (\langle e_1, f_k \rangle)^4}}$$

Set

$$x_0 = c e_1$$

Output: Initial parameters $(x_0, \lambda_1 - \mu_0)$

The update rules are split into two steps, the predictor and the corrector. The predictor is given by a linear step, and the corrector is a fixed point correction to get back to the path.

Algorithm 2: Real Golden Retriever Predictor Step

Input : Previous step (x_i, λ_i), the frame set $\{f_k\}$, a positive symmetric semidefinite matrix Q, and a step size μ_i and an $n+1$ vector of signs of the previous step sgn

Define the matrix

$$R(x) = \frac{1}{m}\sum_{k=1}^{m}\langle x, f_k\rangle^2 f_k f_k^T$$

Then form the $n \times (n+1)$ extended Hessian matrix

$$H_{ext}(x_i, \lambda_i) = \begin{bmatrix} 3R(x_i) + \lambda_i Q - R_0 & Qx_i \end{bmatrix}$$

Find a unit vector v is in $Null(H_{ext})$. Choose the index c to be the index $1, .., n+1$ largest in magnitude of v.

Now

$$(\tilde{\xi}_{t+1,0}, \tilde{\lambda}_{t+1,0}) = (\xi_t, \lambda_t) + \mu_i v$$

Choose the sign of v to be the one that matches the sign of the previous step at index c.

Output: Predictor parameters $(\tilde{x}_{t+1,0}, \tilde{\lambda}_{t+1,0})$

Now, since we took a step in a linear direction, we want to do a corrector step to get back onto the right path.

Algorithm 3: Real Golden Retriever (Newton) Corrector Step

Input : Predictor parameters $(\tilde{x}_{t+1,0}, \tilde{\lambda}_{t+1,0})$, the frame set $\{f_k\}$, the positive symmetric semidefinite matrix Q, and an error threshold Err

Define the matrix

$$R(x) = \frac{1}{m}\sum_{k=1}^{m}\langle x, f_k\rangle^2 f_k f_k^T$$

For $j = 0, 1, 2, ...,$ until a threshold of error

Form the $n \times (n+1)$ extended Hessian matrix

$$H_{ext}(\tilde{x}_{t+1,j}, \tilde{\lambda}_{t+1,j}) = \begin{bmatrix} 3R(\tilde{x}_{t+1,j}) + \tilde{\lambda}_{t+1,j}Q - R_0 & Q\tilde{x}_{t+1,j} \end{bmatrix}$$

Set $H_{ext}^{\dagger}$ to be the pseudoinverse of H_{ext} and then set

$$(\tilde{x}_{t+1,s+1}, \tilde{\lambda}_{t+1,s+1}) = (\tilde{x}_{t+1,s}, \tilde{\lambda}_{t+1,s}) - H_{ext}^{\dagger}[R(\tilde{x}_{t+1,s}) + \tilde{\lambda}_{t+1,s}Q - R_0]\tilde{x}_{t+1,s}$$

Terminate when the

$$||(\tilde{x}_{t+1,N+1}, \tilde{\lambda}_{t+1,N+1}) - (\tilde{x}_{t+1,N}, \tilde{\lambda}_{t+1,N})|| \leq Err$$

After convergence, we finally define

Output: Next step $(x_{t+1}, \lambda_{t+1}) = (\tilde{x}_{t+1,N}, \tilde{\lambda}_{t+1,N})$

After initialization, we continue doing the Predictor and Corrector steps until $\lambda = 0$ or $x = 0$, at which point the algorithm terminates.

There are a number of questions this algorithm brings up, including how to choose the step size, how to derive it, convergence analysis and asymptotics, etc.

2.2 Overview of Real Results

We outline the results here for the real case and some observations about these results. We start with the convergence results.

2.2.1 Real Convergence

We outline a convergence result for the Golden Retriever. Here we work in the noiseless case with $Q = I$. The analysis is based on a reference path. We can take the reference path to be anything, but we want it to start at $(0, \lambda_1)$ (the same point the Golden Retriever starts at), and end at $(z, 0)$, the global minimizer. With this reference path in mind, we want to see how far the Golden Retriever Homotopy Path can deviate from the reference path, and ensure that no other critical point can get close.

We define two conditions that the reference path $\varphi(\lambda)$ can satisfy.

Let $s_n(\lambda) = \lambda_n(Hess(\varphi(\lambda), \lambda))$, $b_0 = \max_{||e||=1} \langle R(e)e, e\rangle$ and $r(\lambda) = \frac{s_n(\lambda)}{6b_0||\varphi(\lambda)||}$

Condition 2.2.1 (Initialization Condition)**.** *Given a frame set, R_0, a suitable reference path $\varphi(\lambda)$, and the golden retriever path $x(\lambda)$, we say that φ satisfies the Initialization Condition if*

$$||x(\lambda) - \varphi(\lambda)|| < r(\lambda) \tag{2.5}$$

for some $0 < \lambda < \lambda_1$.

Condition 2.2.2 (Gradient Condition)**.** *Given a frame set, R_0 and a suitable reference path $\varphi(\lambda)$, we say that φ satisfies the Gradient Condition if*

$$||(R(\varphi) + \lambda I - R_0)\varphi|| < \frac{s_n(\lambda)^2}{12b_0||\varphi(\lambda)||} \tag{2.6}$$

for all $0 < \lambda < \lambda_1$

The remarkable theorem is the following.

Theorem 3.4.7. *If there exists a suitable reference path which satisfies the Initialization Condition and the Gradient Condition, then the Golden Retriever Homotopy Algorithm converges.*

Notice that the Initialization Condition involves the homotopy path, but the Gradient Condition is a condition which does not use the homotopy path directly, and thus can be checked without tracing the homotopy path.

The intuition behind this is the following: $r(\lambda)$ defines a radius for each λ, from which the homotopy path cannot cross, and no other critical point can enter. The Initialization Condition ensures that the homtopy path is inside this radius, and the Gradient Condition ensures that it never leaves this radius. Then it is possible to show that the only critical point it can converge to at $\lambda = 0$ is the global minimizer.

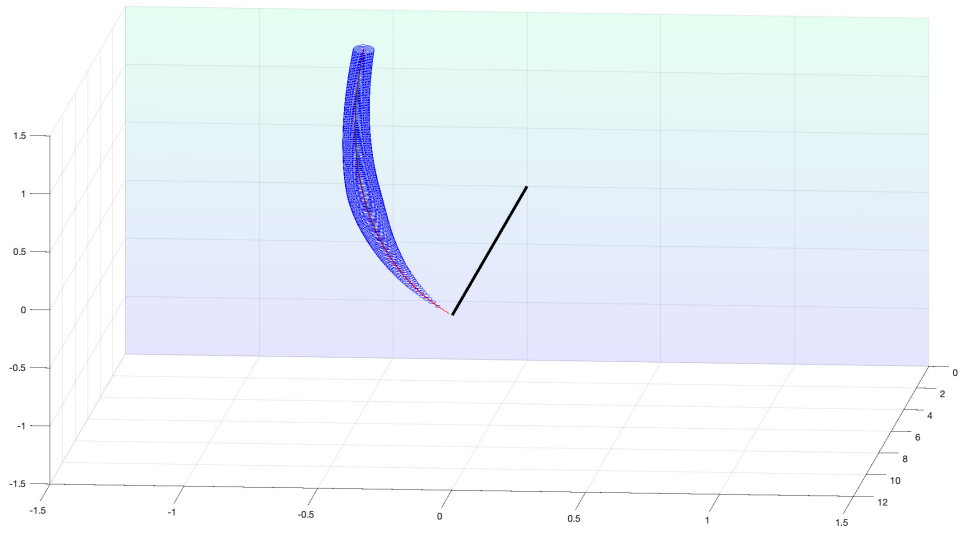

Figure 2.1: This is an example of the golden retriever satisfying both the Initialization Condition and the Gradient Condition. It was generated with $n = 2$, and $m = 5$.

In Figure 2.1, the red line is the Golden Retriever Homotopy path, the green line (barely visible), is a reference path, and the blue circles are the radius $r(\lambda)$. Because the Initialization Condition was satisfied, the red curve starts out inside the tube (called the leash), and because the Gradient Condition was satisfied, it never leaves the leash (and no other path enters), so it converges to the global minimizer.

It is important to note that this is not a requirement for convergence. Figure 2.2 shows that the red homotopy path leaves the leash (so it doesn't satisfy the Gradient Condition), yet it still converges to the global minimizer. Thus it is a sufficient, but not necessary condition for convergence.

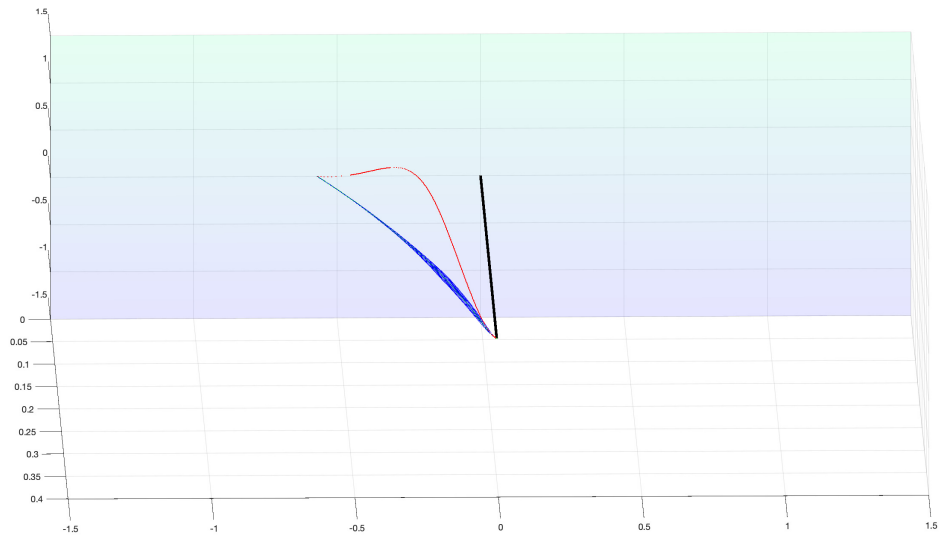

Figure 2.2: This is an example of the golden retriever satisfying the Initialization Condition but not the Gradient Condition. It still converges to the global minimizer, however. It was generated with $n = 2$, and $m = 5$.

Any reference path can be used (assuming it satisfies a few properties to make it suitable), but we study a specific reference path.

To define it, let g be the top eigenvector of R_0, normalized such that $g = \sqrt{\frac{\lambda_1}{\langle R(e_1)e_1, e_1\rangle}}e_1$. Define $\tau = 1 - \frac{\lambda}{\lambda_1}$. The reference path is now given by

$$\varphi_1(\lambda) = \sqrt{\tau}(\tau z + (1-\tau)g) \tag{2.7}$$

Notice that this is a convex combination of z and g, which is scaled by $\sqrt{\tau}$. We first state that this path satisfies the Initialization Condition.

Theorem 3.4.12. *For all $\tau > 0$ sufficiently small, $||x(\lambda) - \varphi_1(\lambda)|| < r(\lambda)$, i.e. $\varphi_1(\lambda)$ satisfies the Initialization Condition.*

Thus, $\varphi_1(\lambda)$ satisfies the Initialization Condition, so if it satisfies the Gradient Condition, we will have shown that it converges to the global minimizer.

We will investigate with what probability it satisfies the Gradient Condition based on the concentration of the matrix $R(x)$ about its mean. Assume that the frame set f_k is drawn from a standard normal distribution. Then we can say

Theorem 3.4.19 (Probabilistic Convergence Result)**.** *In the noiseless case with $Q = I$, let z be fixed and let f_k be drawn from a standard normal. Let m be sufficiently large, by which we mean $m \geq C \cdot n^3$, where the constant may be large, but independent of n. Then with probability greater than or equal to $1 - 5e^{-\gamma n} - \frac{4}{n^2} - \left(n^3 + 1\right)e^{-\frac{3n}{10}}$, $\varphi_1(\lambda)$ satisfies the Gradient Condition, and thus the algorithm converges to the global minimizer. Here $\gamma > log(9)$ is a universal constant.*

2.2.2 Path Verification

In this section, we state the derivation of a numerical certificate that can verify that one is staying on the same critical path. This can be used to get a numerical step size, μ_i, but it is not used in practice because it would slow down the algorithm considerably. However, the result is still interesting, and can be useful in debugging strange cases.

To state the result, we will need to first state some terminology and notation. Let (x_{old}, λ_{old}) be on the path of the algorithm. To get the next point on the path (x_{new}, λ_{new}), we want to make sure we didn't cross to a different path, so there is no other critical point $(x_{other}, \lambda_{other})$ in some hyperplane is the true point on the continuous path. Let $H_{ext,0}$ be the extended Hessian matrix at (x_{old}, λ_{old}) and $H_{ext,new}$ be the extended Hessian matrix at (x_{new}, λ_{new}). Let v be a normalized

vector in the null space of $H_{ext,new}$, and let c be the index of the largest entry in absolute value of v. Let$H_{red:c,0}$ denote $H_{ext,0}$ after deleting column c.

Define

$$b_1 = \frac{1}{m} U f^2$$

where U is the frame bound, and $f = \max_{k=1}^{m} ||f_k||$ (we can use the smaller constant b_0, as defined in the previous section, but computing that numerically can be difficult)

Now define

$$\rho(a, \lambda_a) = \min(\frac{1}{2}, \frac{s_n(H_{ext}(a, \lambda_a)))}{\sqrt{n+1}(b_1 + 3b_1||a|| + ||Q||)})$$

and

$$t_{min} = \min(\frac{\rho(x_{new}, \lambda_{new})}{2A}, -\frac{6b_1||x_0|| + ||Q||}{6Ab_1} + \sqrt{(\frac{6b_1||x_0|| + ||Q||}{6Ab_1})^2 + \frac{s_n(H_{red:c,0})}{6A^2 b_1}})$$

where $A = (2 + 2\frac{||H_{ext,0}||}{s_n(H_{red:c,0})})$

Now we state the Main Theorem from section 3.5.

Theorem 3.5.5. *Assume the Golden Retriever algorithm starts at a point (x_{old}, λ_{old}) which is a critical point. Let (x_{new}, λ_{new}) be a new critical point the algorithm decides and $(x_{other}, \lambda_{other})$ be any other critical point in the same coordinate at the index c, as defined above. Let D_1 denote the distance from (x_{old}, λ_{old}) to (x_{new}, λ_{new}).*

Assume the following two conditions are satisfied:

1. $D_1 < \frac{\rho(x_{new}, \lambda_{new})}{2}$

2. $t < t_{min}$

Then (x_{new}, λ_{new}) *is the point connected on the continuous homotopy path which goes through* (x_{old}, λ_{old}).

Theorem 3.5.5 gives us a numerical certificate we can check. In the implementation, it is now possible to choose a step size μ_i based on the previous one μ_{i-1}, take a step and see if the certificate certifies. If not, one can reduce the step size (update μ_i to $\mu_i/2$ for instance) and repeat the check.

Note that in practice this is still rarely used, and when used, it is a debugging parameter, as it can be significantly slower this way. It is more efficient in practice to pick a numerically feasible small step size. Also, to improve performance we usually bias the coordinate we move along to be λ for speed purposes, and the certificate is not compatible with the bias.

2.2.3 Oracle Convergence

The next theorem has a very interesting meaning to it. We no longer take $Q = I$. In general Homotopy Methods are not guaranteed to lead to the correct solution, they can turn around and go to a different eigenvalue at $x = 0$, or a different critical point at $\lambda = 0$. However, this theorem says it is always possible to initialize the system with a specific matrix Q that would guarantee convergence.

Theorem 3.6.3. *Let* z *be the minimizer to the optimization problem in* (2.2). *There*

exists a positive definite matrix Q_z such that the Golden Retriever Algorithm, initialized with Q_z, converges to z. Moreover, the trajectory of the homotopy path with Q_z projected onto $\lambda = 0$ follows a straight line.

This theorem means that with enough computing power, we could initialize the algorithm with several different choices of Q and run them in parallel. In principle, if one had more information about the location of the global minimizer, one could bias the Q matrix to give a higher probability of convergence.

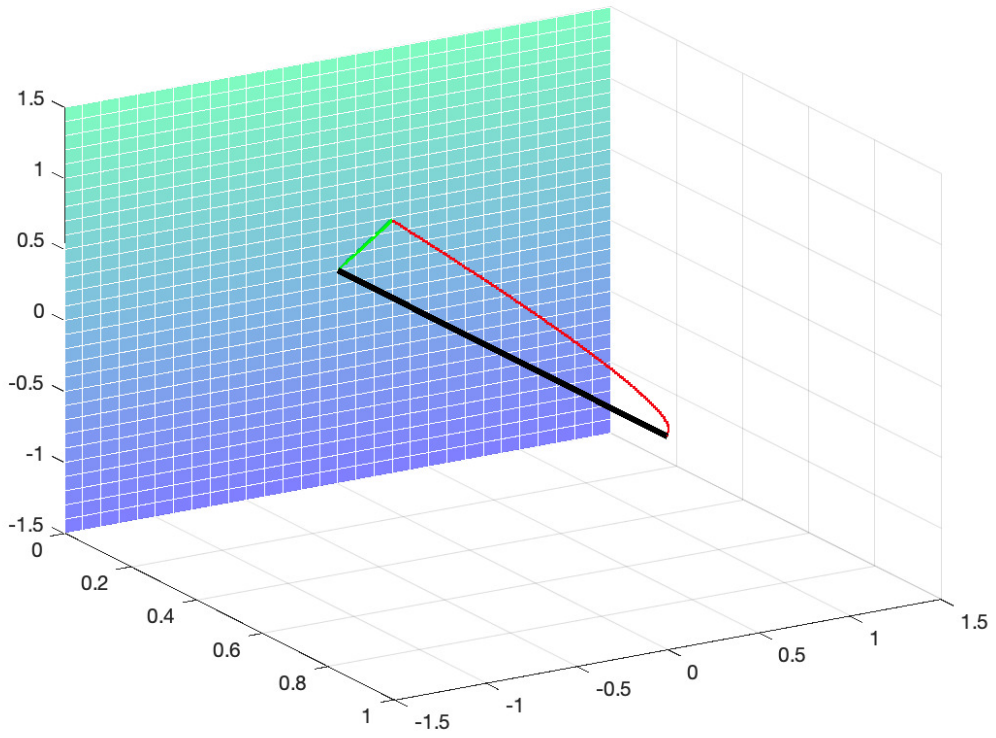

Figure 2.3: This figure was generated by running the Golden Retriever algorithm with the matrix Q_z instead of the identity. The red path is a parabola, while the projection onto the $\lambda = 0$ plane, the green path, shows that the x estimates follow a straight line to the solution. It was generated with $n = 2$, and $m = 5$.

2.3 The Complex Golden Retriever Algorithm

We again give the initialization. The complex case is done through the realification procedure so all vectors and matrices live in or act on $\mathbb{R}^{2n}$.

Algorithm 4: Complex Golden Retriever Initialization

Input : Observations $\{y_k\}$, the frame set $\{f_k\}$, a positive definite hermitian matrix Q, and a step size μ_0

Define the following quantities

$$J = \begin{bmatrix} 0 & -I_n \\ I_n & 0 \end{bmatrix} \in \mathbb{R}^{2n\times 2n} \qquad S = \begin{bmatrix} Real(Q) & -Imag(Q) \\ Imag(Q) & Real(Q) \end{bmatrix} \in \mathbb{R}^{2n\times 2n}$$

$$\varphi_k = \kappa(f_k) = \begin{bmatrix} Real(f_k) \\ Imag(f_k) \end{bmatrix} \in \mathbb{R}^{2n} \qquad \Phi_k = \varphi_k\varphi_k^T + J\varphi_k\varphi_k^T J^T \in \mathbb{R}^{2n\times 2n}$$

Define η_1 to be an eigenvector corresponding to the largest eigenvalue (denoted λ_1, which will have multiplicity 2) of

$$\Gamma_0 = \frac{1}{m}\sum_{k=1}^{m} y_k\Phi_k$$

Set

$$c = \sqrt{\frac{\mu_0\langle S\eta_1, \eta_1\rangle}{\frac{1}{m}\sum_{k=1}^{m}(\eta_1^T\Phi_k\eta_1)^2}}$$

Output: Initial parameters $(\xi_0, \lambda_1 - \mu_0) = (c\eta_1, \lambda_1 - \mu_0)$

As with the real case, the update rules are split into two steps, the predictor and the corrector. The predictor is given by a linear step, and the corrector is a

fixed point correction to get back to the path.

Algorithm 5: Complex Golden Retriever Predictor Step

Input : Previous step (x_i, λ_i), the frame set $\{f_k\}$, a positive symmetric semidefinite matrix Q, and a step size μ_i and an $n+1$ vector of signs of the previous step sgn

Define the matrices

$$\Gamma(\xi) = \frac{1}{m}\sum_{k=1}^{m} \xi^T \Phi_k \xi \Phi_k, \quad \widetilde{\Gamma}(\xi) = \frac{1}{m}\sum_{k=1}^{m} \Phi_k \xi \xi^T \Phi_k$$

Then form the $2n \times (2n+1)$ extended Hessian matrix

$$H_{ext}(\xi_i, \lambda_i) = \begin{bmatrix} \Gamma(\xi_i) + 2\widetilde{\Gamma}(\xi_i) + \lambda_i S - \Gamma_0 & S\xi_i \end{bmatrix}$$

Find a unit vector v is in $Null(H_{ext})$, and $v \perp J\xi_i$. Choose the index c to be the index $1, .., n+1$ largest in magnitude of v.

Now

$$(\tilde{\xi}_{t+1,0}, \tilde{\lambda}_{t+1,0}) = (\xi_t, \lambda_t) + \mu_i v$$

Choose the sign of v to be the one that matches the sign of the previous step at index c.

Output: Predictor parameters $(\tilde{\xi}_{t+1,0}, \tilde{\lambda}_{t+1,0})$

Algorithm 6: Complex Golden Retriever (Newton) Corrector Step

Input : Predictor parameters $(\tilde{\xi}_{t+1,0}, \tilde{\lambda}_{t+1,0})$, the frame set $\{f_k\}$, the positive symmetric semidefinite matrix Q, and an error threshold Err

Define the matrices

$$\Gamma(\xi) = \frac{1}{m}\sum_{k=1}^{m} \xi^T \Phi_k \xi \Phi_k, \quad \widetilde{\Gamma}(\xi) = \frac{1}{m}\sum_{k=1}^{m} \Phi_k \xi \xi^T \Phi_k$$

For $j = 0, 1, 2, ...,$ until a threshold of error

Form the $2n \times (2n+1)$ extended Hessian matrix

$$H_{ext}(\tilde{\xi}_{t+1,j}, \tilde{\lambda}_{t+1,j}) = \begin{bmatrix} \Gamma(\tilde{\xi}_{t+1,j}) + 2\widetilde{\Gamma}(\tilde{\xi}_{t+1,j}) + \tilde{\lambda}_{t+1,j} S - \Gamma_0 & S\tilde{\xi}_{t+1,j} \end{bmatrix}$$

Set $H_{ext}^{\dagger}$ to be the pseudoinverse of H_{ext} and set

$$(\tilde{\xi}_{t+1,s+1}, \tilde{\lambda}_{t+1,s+1}) = (\tilde{\xi}_{t+1,s}, \tilde{\lambda}_{t+1,s}) - H_{ext}^{\dagger}[\Gamma(\tilde{\xi}_{t+1,s}) + \tilde{\lambda}_{t+1,s} S - \Gamma_0]\tilde{\xi}_{t+1,S}$$

Terminate when the

$$||(\tilde{\xi}_{t+1,N+1}, \tilde{\lambda}_{t+1,N+1}) - (\tilde{\xi}_{t+1,N}, \tilde{\lambda}_{t+1,N})|| \leq Err$$

After convergence, we finally define

Output: Next step $(\xi_{t+1}, \lambda_{t+1}) = (\tilde{\xi}_{t+1,N}, \tilde{\lambda}_{t+1,N})$

After initialization, we continue doing the Predictor and Corrector steps, and

the algorithm terminates when either $\lambda = 0$ or $\xi = 0$.

Again we need to know how to choose the step size, how to derive it, convergence analysis and asymptotics, etc.

2.4 Overview of Complex Results

2.4.1 Complex Convergence

As with the real case, we can analyze the convergence of the Golden Retriever algorithm. We work in the noiseless case, with $S = I_{2n}$. This analysis is also based on a reference path which starts at $(0, \lambda_1)$ and ends at $(z, 0)$.

We define, like in the real case, two conditions that the reference path $\varphi(\lambda)$ can satisfy.

Let $s_{2n-1}(\lambda) = \lambda_{2n-1}(Hess(\varphi(\lambda), \lambda))$, $\beta = \max_{||e||=1} \langle \Gamma(e)e, e \rangle$ and

$$r(\lambda) = \frac{-12\beta||\varphi(\lambda)|| + \sqrt{144\beta^2||\varphi(\lambda)||^2 + 72\beta s_{2n-1}(\lambda)}}{36\beta} \tag{2.8}$$

Finally define

$$\rho_2(\lambda) = \frac{-B + \sqrt{B^2 + 4AC}}{2A} \quad , \quad \rho_1(\lambda) = \frac{\rho_2(\lambda)}{s_{2n-1}(\lambda)} \tag{2.9}$$

where

$$A = 972\beta^2$$

$$B = 864\beta^3||\varphi(\lambda)||^3 + 648\beta^2||\varphi(\lambda)||s_{2n-1}(\lambda)$$

$$C = 36\beta^2 s_{2n-1}(\lambda)^2 ||\varphi(\lambda)||^2 + 24\beta s_{2n-1}(\lambda)^3$$

Condition 2.4.1 (Initialization Condition)**.** *Given a frame set,* Γ_0*, a suitable reference path* $\varphi(\lambda)$*, and the golden retriever path* $\xi(\lambda)$*, we say that* φ *satisfies the Initialization Condition if*

$$||\xi(\lambda) - \varphi(\lambda)|| < \rho_1(\lambda) \tag{2.10}$$

for some $0 < \lambda < \lambda_1$*.*

Condition 2.4.2 (Gradient Condition)**.** *Given a frame set,* Γ_0 *and a suitable reference path* $\varphi(\lambda)$*, we say that* φ *satisfies the Gradient Condition if*

$$||(\Gamma(\varphi) + \lambda I - \Gamma_0)\varphi|| < \rho_2(\lambda) \tag{2.11}$$

for all $0 < \lambda < \lambda_1$

Notice the difference in the expressions in the complex conditions from the real conditions. This comes from difficulties arising with the phase ambiguity in the complex case.

The conditions give rise to an equivalent theorem as in the real case.

Theorem 4.5.9. *If there exists a suitable reference path which satisfies the Initialization Condition and the Gradient Condition, then the Complex Golden Retriever Homotopy Algorithm converges to a global minimizer.*

Notice that the Initialization Condition involves the homotopy path, but the

Gradient Condition is a condition on the reference path alone.

The intuition behind this is similar to the real case. Assuming the Initializiation Condition and the Gradient Condition are satisfied, then $r(\lambda)$ defines a radius for each λ, from which the homotopy path cannot cross. The Initialization Condition ensures that the homtopy path is inside this radius, and the Gradient Condition ensures that it never leaves this radius. Using this, it is still possible to show that the only critical point it can converge to at $\lambda = 0$ is a global minimizer.

Then, we define a suitable reference path

$$\varphi_1(\lambda) = U(\lambda)\sqrt{\tau}(\tau\zeta + (1-\tau)\eta) \tag{2.12}$$

where $\tau = 1 - \frac{\lambda}{\lambda_1}$ and $U(\lambda)$ is an certain alignment matrix.

It is possible to show that $\varphi_1(\lambda)$ always satisfies the Initialization Condition. Thus we are left checking whether it satisfies the Gradient Condition.

Again, after using concentration of $\Gamma(\xi)$ about its mean, we can conclude the following theorem, by showing when $\varphi_1(\lambda)$ satisfies the Gradient Condition.

Theorem 2.4.3. *In the noiseless case with $Q = I$, fix a nonzero $\zeta \in \mathbb{R}^{2n}$ to be the realification of the generating signal. Assume f_k are distributed i.i.d. complex normal, with a sufficiently high number of samples. That means that $m \geq Cn^3$, where the constant may be large but independent of n. Then with probability at least $1 - \frac{13}{n^2} - 10e^{-\gamma n} - \left(n^3+1\right)e^{-\frac{3n}{10}}$, $\varphi_1(\lambda)$ satisfies the Gradient Condition, and thus the algorithm converges to a global minimizer. Here $\gamma > log(9)$ is a universal constant.*

2.4.2 Path Verification

As with the real case, we derived a certifier that issues a numerical certificate which, if verified, guarantees that one are on the correct homotopy path.

To state the theorem, we will need to state some terminology and notation. Let $(\xi_{old}, \lambda_{old})$ be a critical point on the the homotopy path. To get the next point on the path $(\xi_{new}, \lambda_{new})$, we want to make sure we didn't step to another critical point which came close (in some hyperplane) to the homotopy path, so we want to make sure there is no other critical point $(\xi_{other}, \lambda_{other})$ close enough to be the true critical point smoothly connected to $(\xi_{old}, \lambda_{old})$. Let $H_{ext,0}$ be the extended Hessian matrix at $(\xi_{old}, \lambda_{old})$ and $H_{ext,new}$ be the extended Hessian matrix at $(\xi_{new}, \lambda_{new})$. Let v be a normalized vector in the null space of $H_{ext,new}$, and let c be the index of the largest entry in absolute value of v. $H_{red:c,0} = \begin{bmatrix} \Gamma(\xi_{old}) + 2\widetilde{\Gamma}(\xi_{old}) + \lambda S - \Gamma_0 & S\xi_{old} \\ (J\xi_{old})^T & 0 \end{bmatrix}$ after deleting column c.

Define $\beta_1 = \frac{1}{m} U f^2$ where U is the upper frame bound, and $f = \max_{k=1}^{m} ||f_k||$ (one can use β instead of β_1 everywhere, but that makes it difficult to compute).

Now define

$$\rho(a, \lambda_a) = \min(\frac{1}{2}, \frac{s_{2n}(H_{ext}(a, \lambda_a)))}{\sqrt{2n+1}(\beta_1 + 3\beta_1||a|| + ||S||)})$$

and

$$t_{min} = min(\frac{\rho(\xi_{new}, \lambda_{new})}{2A}, -\frac{6\beta_1||\xi_0|| + ||S|| + 1}{6A\beta_1} + \sqrt{(\frac{6\beta_1||\xi_0|| + ||S|| + 1}{6A\beta_1})^2 + \frac{s_{2n}(H_{red:c,0})}{6A^2\beta_1}})$$

where $A = (2 + 2\frac{||H_{ext,0}||}{s_{2n}(H_{red:c,0})})$

Theorem 4.6.5. *Assume our algorithm starts at a point $(\xi_{old}, \lambda_{old})$ which is a critical point. Let $(\xi_{new}, \lambda_{new})$ be a new critical point the algorithm decides and $(\xi_{other}, \lambda_{other})$ be any other critical point in the same coordinate at the index c, as defined above. Let D_1 denote the distance from $(\xi_{old}, \lambda_{old})$ to $(\xi_{new}, \lambda_{new})$.*

Assume the following two conditions are satisfied:

1. $D_1 < \frac{\rho(\xi_{new}, \lambda_{new})}{2}$
2. $t < t_{min}$

Then $(\xi_{new}, \lambda_{new})$ is the point connected on the continuous homotopy path which goes through $(\xi_{old}, \lambda_{old})$.

Theorem 4.6.5 gives us a numerical certificate we can check. In the implementation, it is now possible to choose a step size μ_i based on the previous one μ_{i-1}, take a step and see if the certificate certifies. If not, one can reduce the step size (update μ_i to $\mu_i/2$ for instance) and repeat the check.

Note that this is usually used as a debugging parameter, as it can be significantly slower this way. Also, we usually bias the coordinate we move along to be λ for speed purposes, so it won't work together with this bias.

2.4.3 Oracle Convergence

The next theorem says it is always possible to initialize the system with a specific matrix positive definite symmetric matrix S in $\mathbb{R}^{2n\times 2n}$ (or equivalently a positive definite hermitian matrix Q in $\mathbb{C}^{n\times n}$) that would guarantee convergence.

Theorem 4.7.3. *Let ζ be the minimizer to the optimization problem in (2.2). There exists a positive definite matrix S_z such that the Golden Retriever Algorithm, initialized with S_z, converges to S. Moreover, the trajectory of the homotopy path with S_z, projected onto $\lambda = 0$, follows a straight line.*

This theorem means that with enough computing power, we could initialize the algorithm with several different choices of S and run them in parallel. In principle, if one had more information about the location of a global minimizer, one could bias the S matrix to give a higher probability of convergence.

2.5 Numerical Results

2.5.1 Numerical Experiments

We ran the golden retriever and gathered statistics on the convergence of the algorithm. We present results for the noiseless case, a trial is declared a success if the relative error was less than 10^{-5} from the global minimizer. The noisy case is more difficult since we don't have access to the global minimizer. For trials here, we recommend declaring success based on a success criterion which is a sum of the error tolerance 10^{-5} and a term involving the Cramer-Rao Lower Bound (see [22]

for details about the Cramer-Rao Lower bound in Phase Retrieval). This can then compared to the distance from the generating signal. The Cramer-Rao Lower Bound for the real case is given by

$$\frac{\sigma^2}{4m} Tr(R(z)^{-1}) \tag{2.13}$$

In the complex case, one can use the lower bound

$$\frac{\sigma^2}{4m} Tr(\widetilde{\Gamma}(\zeta)^{+}) \tag{2.14}$$

This expression is be derived in [22], where we use that $J\zeta$ is in the null space of $\widetilde{\Gamma}(\zeta)^{+}$.

For the noiseless real case, we looked at the Gaussian Case when $n = 128$ with $m = 1.1n, 1.2n, ...4.1n$ and gathered statistics about how often it converged to the global minimizer in 100 trials.

Figure 2.3 shows the empirical success probabilities in the real noiseless case $SNR = \infty$. We plot the redundancy to the number of success. Here success is defined by having an error less than 10^{-5}.

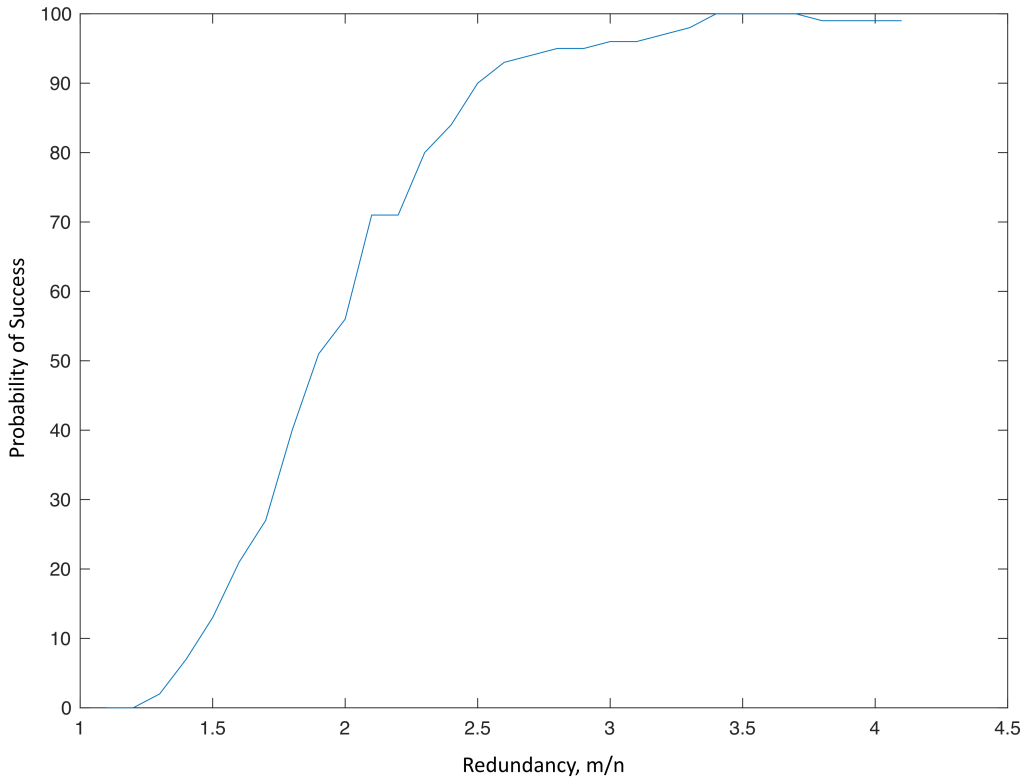

Figure 2.4: Real noiseless case, $n = 128$, $SNR = \infty$

We see from this that with a small number of samples ($m \approx 3.4$) already yields very high probability of success.

In the complex case, we looked at the Gaussian Case when $n = 128$ and $m = 2.6n, 1.9n, 2n, , ..., 4.5n$ and gathered statistics about how often it converged to a global minimizer, a non-global local minimum, a saddle point, or another eigenvalue at $x = 0$.

For the noiseless case, the criteria for convergence was whether the error was less than 10^{-5}. We plot the redundancy to the number of success.

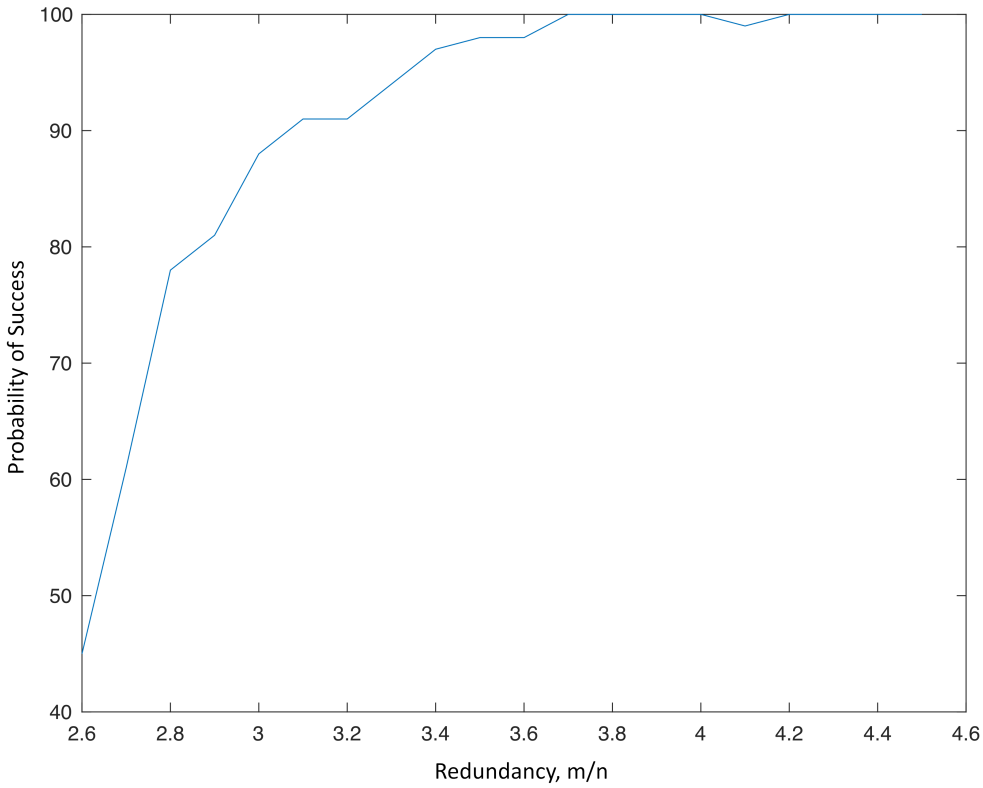

Figure 2.5: Complex noiseless case, $n = 128$, $SNR = \infty$

To compare with Wirtinger Flow, we note that for high enough redundancy, it seems that if either Wirtinger Flow or the Golden Retriever converges, then the other converges with high probability.

For lower redundancies, it seems as if Golden Retriever succeeds more often than Wirtinger Flow does.

To illustrate this, the following graphs compares Wirtinger Flow directly to the Golden Retriever.

The first graph is for the real Golden Retriever, and we compare it at redundancies from 1 to 4 against Wirtinger Flow. In these, the default parameters were

chosen, except the the number of iterations for Wirtinger Flow was increased to $100,000$.

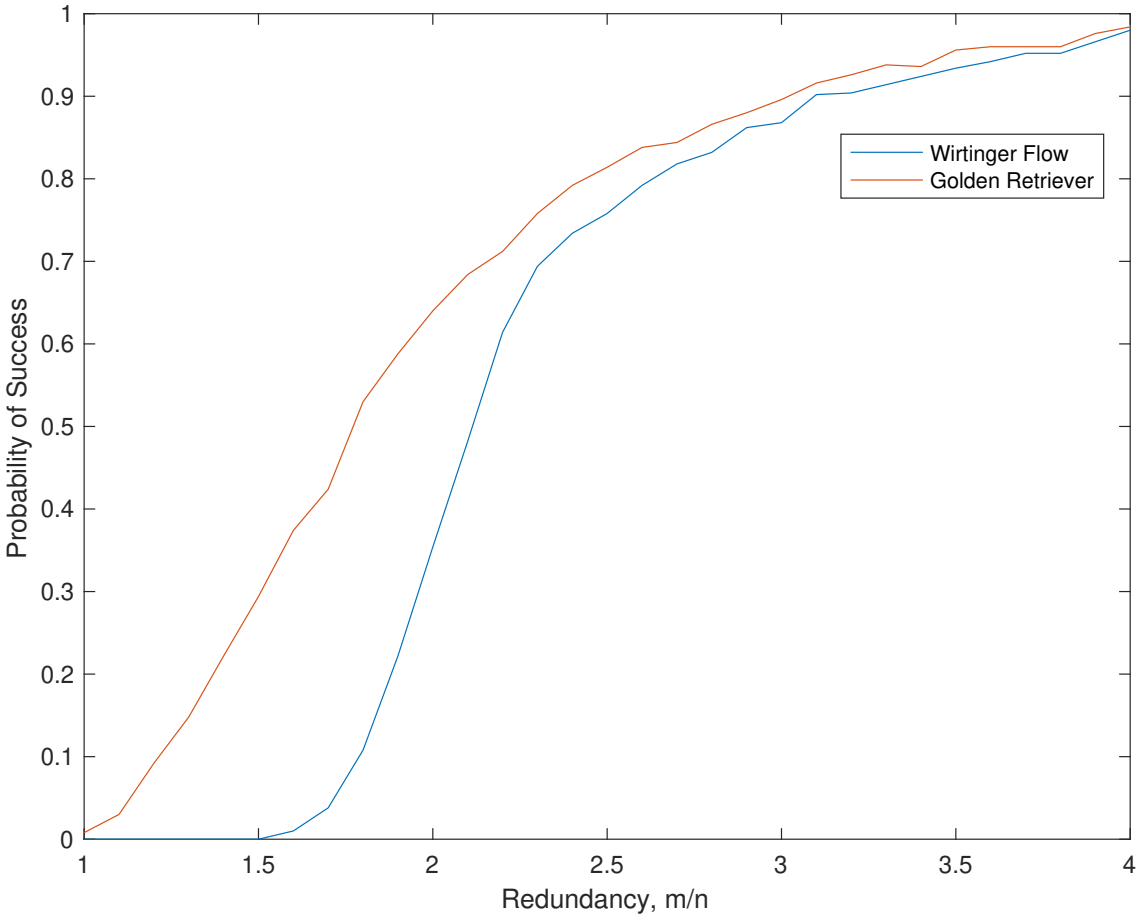

Figure 2.6: The following figure illustrates, in the real case, the Golden Retriever empirical success and Wirtinger Flows empirical success on the same graph. This was generated with $n = 30$, and 500 trials for each redundancy.

The next graph is for the complex Golden Retriever, and we compare it at redundancies from 1.5 to 4.5 against Wirtinger Flow. In these, the default parameters were chosen, except the the number of iterations for Wirtinger Flow was increased to $100,000$.

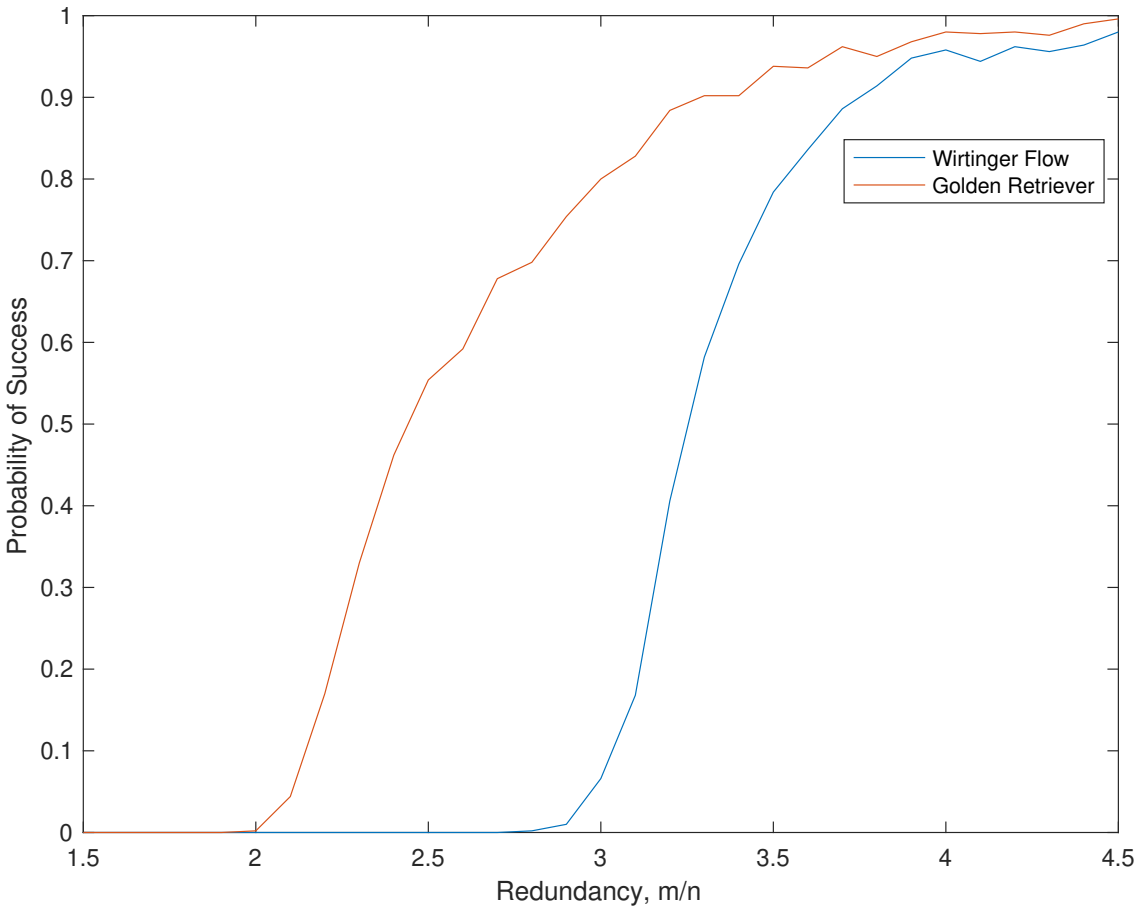

Figure 2.7: The following figure illustrates, in the complex case, the Golden Retriever empirical success and Wirtinger Flows empirical success on the same graph. This was generated with $n = 30$, and 500 trials for each redundancy.

We can see that the Golden Retriever outperforms Wirtinger Flow at low redundancies.

2.5.2 Computational Complexity

We estimate the computational complexity of the Golden Retriever algorithm. We will examine this with $Q = I$. We start with the space complexity. Storage is a known issue for many phase retrieval algorithms (see [23]). For reducing the required storage, we will not store any matrices, and instead sacrifice time complexity in

computing these on the fly. We also will assume that we don't store the frame vectors, and we can either compute them as needed (in constant time) or retrieve them from an oracle if needed. Thus the space complexity is going to be $O(n)$.

The time complexity is a little more complex. The predictor step requires computing the null space of a $n \times (n+1)$ extended Hessian matrix, H_e. This is equivalent to computing the null space of a symmetric positive definite $H_e^T H_e$. To do this, we deform the matrix by adding δI to look at the matrix $H_e^T H_e + \delta I$ and are looking at eigenvalue corresponding to the smallest eigenvector. This can be done with only matrix-vector multiplications and vector additions using a conjugate gradient algorithm outlined in [24]. Each matrix vector computation can be done in $O(mn)$ steps, and if we assume we do the conjugate gradient for κ_1 steps, that brings each predictor step to a complexity of $O(\kappa_2 mn)$.

For the corrector step, we need to solve a linear system consisting of $H_e v = F(x, \lambda)$, where $F(x, \lambda) = \nabla_x J(x, \lambda)$. Again we look at the equation $H_e^T H_e v = H_e^T F(x, \lambda)$. Computing $F(x, \lambda)$ can be done in $O(mn)$ steps, and so can computing $H_e^T F(x, \lambda)$. The conjugate gradient would also only involve matrix-vector multiplications, so if κ_2 denotes the number of fixed point corrections done, the complexity of the corrector step is $O(\kappa_2 mn)$.

Therefore, if we set a bound on the number of fixed points iterations and conjugate gradient steps, and if N is the total number of iterations in the golden retriever, the complexity of the golden retriever comes out to $O(Nmn)$. The same computations and complexities hold in the complex case as well.

2.6 Future Work

Here we identify some directions for future research. We have shown that in the noiseless case, the Golden Retriever converges with some probability if the number of samples is of the order $m = O(n^3)$. This is based on finding upperbounds and getting control on some constants which can be done in $O(n^3)$. If these upperbounds were tightened, or the assumptions removed, then the Golden Retriever would converge with a lower sampling requirement of $O(nlogn)$. In addition, many variants of Wirtinger Flow have been created which have sampling size of order $m = O(n)$, so the number of measurements is of the same order as the signal. In many of these, the loss function has been altered from the mean square error, to a less smooth absolute error. That is, the criterion to minimize may look like

$$K(x, \lambda) = \frac{1}{2m} \sum_{k=1}^{m} \left| y_k - |\langle x, f_k \rangle|^2 \right| + \frac{\lambda}{2} \langle Qx, x \rangle$$

We postulate that a similar homotopic algorithm with this absolute error loss function would bring the convergence rate of the retriever to $m = O(n)$.

At the same time, one could also play with the regularization term, change it from a quadratic regularization term to a different order term, such as a linear term.

Another area of future work is to optimize the analysis of the leash developed in the convergence result. In the derivation it is a sufficient result, where one side of an inequality is minimized and the other maximized which provides a sufficient bound necessary for convergence. However, a more careful analysis may lead to finer

results. This may prove particularly effective in removing the assumptions on b_0, as this is what drives the rate up to $O(n^3)$

Another possibility is the investigation into different reference paths. Perhaps a different suitable reference path may enjoy nicer convergence results.

Chapter 3: Real Case

In this chapter, we analyze details about the Golden Retriever algorithm in the real case. We begin with the derivation of the algorithm.

3.1 Derivation of the Golden Retriever in the Real Case

First we start off by rewriting the minimization criterion, then we show some properties satisfied by solutions to the system, and finally we derive the Golden Retriever Algorithm.

3.1.1 Preliminaries

First we look at the minimization criterion

$$J(x, \lambda; \mathcal{F}, Q, y) = \frac{1}{4m} \sum_{k=1}^{m} (y_k - |\langle x, f_k \rangle|^2)^2 + \frac{\lambda}{2} \langle Qx, x \rangle \tag{3.1}$$

Usually we will suppress the dependence on the frame set, the measurements, as well as the symmetric positive definite matrix Q by denoting the criterion as $J(x, \lambda)$.

We would like to rewrite the criterion into a form more manageable to work

with. To his end, we expand the form to get:

$$J(x,\lambda) = \frac{1}{4m}\sum_{k=1}^{m}(y_k^2 - 2y_k|\langle x, f_k\rangle|^2 + |\langle x, f_k\rangle|^4) + \frac{\lambda}{2}\langle Qx, x\rangle$$
$$= \frac{1}{4m}\sum_{k=1}^{m} y_k^2 + \frac{1}{4m}\sum_{k=1}^{m}|\langle x, f_k\rangle|^4 + \frac{\lambda}{2}\langle Qx, x\rangle - \frac{1}{2m}\sum_{k=1}^{m} y_k|\langle x, f_k\rangle|^2$$

Now we write some terminology:

$$R_0 = \frac{1}{m}\sum_{k=1}^{m} y_k f_k f_k^T, \quad R(x) = \frac{1}{m}\sum_{k=1}^{m}|\langle x, f_k\rangle|^2 f_k f_k^T \tag{3.2}$$

With these, we can simplify the criterion using the following proposition.

Proposition 3.1.1. *With the above notation, we can simplify the criterion into the following form*

$$J(x,\lambda) = \frac{1}{4}\langle R(x)x, x\rangle + \frac{1}{2}\langle(\lambda Q - R_0)x, x\rangle + \frac{1}{4m}\sum_{k=1}^{m} y_k^2 \tag{3.3}$$

Proof. First look at $R(x)x = \frac{1}{m}\sum_{k=1}^{m}|\langle x, f_k\rangle|^2 f_k f_k^T x = \frac{1}{m}\sum_{k=1}^{m}|\langle x, f_k\rangle|^2\langle x, f_k\rangle f_k = \frac{1}{m}\sum_{k=1}^{m}(\langle x, f_k\rangle)^3 f_k$

Therefore $\langle R(x)x, x\rangle = \langle\frac{1}{m}\sum_{k=1}^{m}|\langle x, f_k\rangle|^3 f_k, x\rangle = \frac{1}{m}\sum_{k=1}^{m}|\langle x, f_k\rangle|^2\langle x, f_k\rangle\langle f_k, x\rangle = \frac{1}{m}\sum_{k=1}^{m}|\langle x, f_k\rangle|^4$.

Similarly $R_0 x = \frac{1}{m}\sum_{k=1}^{m} y_k f_k f_k^T x = \frac{1}{m}\sum_{k=1}^{m} y_k\langle x, f_k\rangle f_k$. Therefore, $\langle R_0 x, x\rangle = \frac{1}{m}\sum_{k=1}^{m} y_k|\langle x, f_k\rangle|^2$

□

Proposition 3.1.2. *For a fixed λ, we have*

$$F(x,\lambda) := \nabla_x J(x,\lambda) = R(x)x + (\lambda Q - R_0)x \tag{3.4}$$

Proof. Let us first examine $\nabla_x \frac{1}{4}\langle R(x)x, x\rangle$. We have that

$$\frac{1}{4}\nabla_x(\langle R(x)x, x\rangle) = \frac{1}{4}\nabla_x(\frac{1}{m}\sum_{k=1}^{m}|\langle x, f_k\rangle|^4) = \frac{1}{4m}\sum_{k=1}^{m}\nabla_x[(\langle x, f_k\rangle)^4]$$
$$= \frac{4}{4m}\sum_{k=1}^{m}(\langle x, f_k\rangle)^3\nabla_x(\langle x, f_k\rangle) = \frac{1}{m}\sum_{k=1}^{m}(\langle x, f_k\rangle)^3 f_k = R(x)x$$

In addition, we know for a constant symmetric matrix A, $\nabla_x(\langle Ax, x\rangle) = 2Ax$. Therefore, since $\lambda Q - R_0$ is a constant symmetric matrix, we have that $\nabla_x(\frac{1}{2}\langle(\lambda Q - R_0)x, x\rangle) = (\lambda Q - R_0)x$

Since the last term is constant with respect to x, we have our result. □

Proposition 3.1.3.

$$Hess(J(x,\lambda)) = 3R(x) + \lambda Q - R_0 \tag{3.5}$$

Proof. First of all, we have that $\nabla_x[(\lambda Q - R_0)x] = (\lambda Q - R_0)$, so what we want to find is $\nabla_x(R(x)x)$

To do this, as we established before, $R(x)x = \frac{1}{m}\sum_{k=1}^{m}(\langle x, f_k\rangle)^3 f_k$, and if we use

the vector calculus identity that $\nabla_x(cA) = A \otimes \nabla_x(c) + c\nabla_x(A)$, we get the following

$$\begin{aligned}\nabla_x(\frac{1}{m}\sum_{k=1}^{m}(\langle x, f_k\rangle)^3 f_k) &= \frac{3}{m}\sum_{k=1}^{m}(\langle x, f_k\rangle)^2 f_k \otimes f_k \\ &= \frac{3}{m}\sum_{k=1}^{m}(\langle x, f_k\rangle)^2 f_k f_k^T = 3R(x)\end{aligned}$$

□

To summarize the results we got so far, we showed that for the criterion

$$J(x,\lambda) = \frac{1}{4m}\sum_{k=1}^{m}(y_k - |\langle x, f_k\rangle|^2)^2 + \frac{\lambda}{2}\langle Qx, x\rangle \tag{3.6}$$

$$= \frac{1}{4}\langle R(x)x, x\rangle + \frac{1}{2}\langle(\lambda Q - R_0)x, x\rangle + \frac{1}{4m}\sum_{k=1}^{m} y_k^2 \tag{3.7}$$

We have the following gradient:

$$\nabla_x J(x,\lambda) = (R(x) + \lambda Q - R_0)x \tag{3.8}$$

And we have the following hessian matrix:

$$Hess(J(x,\lambda)) = 3R(x) + \lambda Q - R_0 \tag{3.9}$$

Now we look at the the case where x, λ are parameterized by another parameter t, and look at the extended Gradient and the extended hessian.

Proposition 3.1.4. *If $x = x(t), \lambda = \lambda(t)$, then the extended hessian of both x and*

λ can be written

$$Hess_{ext} = \begin{bmatrix} (3R(x) + \lambda Q - R_0) & Qx \end{bmatrix} \tag{3.10}$$

Proof. The proof follows from the taking the derivative of the gradient with respect to t

$$\begin{aligned} \frac{d}{dt}(\nabla_x J(x(t), \lambda(t))) &= \frac{d}{dt}(R(x(t))x(t) + \lambda(t)Qx(t) - R_0 x(t)) \\ &= \nabla_x(\nabla_x J)\frac{dx}{dt} + \frac{\partial}{\partial \lambda}(\nabla_x J) \\ &= Hess_J(x, \lambda) \cdot \frac{dx}{dt} + Qx\frac{d\lambda}{dt} \\ &= \begin{bmatrix} (3R(x) + \lambda Q - R_0) & Qx \end{bmatrix} \cdot \begin{bmatrix} \frac{dx}{dt} \\ \\ \frac{d\lambda}{dt} \end{bmatrix} \\ &= Hess_{ext} \cdot \begin{bmatrix} \frac{dx}{dt} \\ \\ \frac{d\lambda}{dt} \end{bmatrix} \end{aligned}$$

□

Now to handle expansions in $R(x)$, we define associated bilinear matrices.

Definition 3.1.4.1. *Let $x, y \in \mathbb{R}^n$. Define $R(x, y) = \frac{1}{m}\sum_{k=1}^{m}\langle x, f_k\rangle\langle y, f_k\rangle f_k f_k^T$.*

We can summarize some of the properties of these matrices in the following proposition.

Proposition 3.1.5. *Let $x, y, z \in \mathbb{R}^n$. Then we have the following properties:*

1. $R(x, x) = R(x)$
2. $R(x + y) = R(x) + R(y) + 2R(x, y)$

3. $R(x,y) = R(y,x)$

4. $R(x,y)z = R(y,z)x = R(z,x)y$

5. $R(x,y)y = R(y)x$

Proof. These are easy computations

1. $R(x,x) = \frac{1}{m}\sum_{k=1}^{m}\langle x, f_k\rangle\langle x, f_k\rangle f_k f_k^T = \frac{1}{m}\sum_{k=1}^{m}\langle x, f_k\rangle^2 f_k f_k^T = R(x)$

2. $R(x+y) = \frac{1}{m}\sum_{k=1}^{m}\langle (x+y), f_k\rangle^2 f_k f_k^T = \frac{1}{m}\sum_{k=1}^{m}\langle x, f_k\rangle^2 f_k f_k^T + \frac{1}{m}\sum_{k=1}^{m}\langle y, f_k\rangle^2 f_k f_k^T + \frac{2}{m}\sum_{k=1}^{m}\langle x, f_k\rangle\langle y, f_k\rangle f_k f_k^T = R(x) + R(y) + 2R(x,y)$

3. $R(x,y) = \frac{1}{m}\sum_{k=1}^{m}\langle x, f_k\rangle\langle y, f_k\rangle f_k f_k^T = \frac{1}{m}\sum_{k=1}^{m}\langle y, f_k\rangle\langle x, f_k\rangle f_k f_k^T = R(y,x)$

4. $R(x,y)z = \frac{1}{m}\sum_{k=1}^{m}\langle x, f_k\rangle\langle y, f_k\rangle\langle z, f_k\rangle f_k$, now we can permute them in any order.

5. By (4), we have $R(x,y)y = R(y,y)x$, which by (1) gives us $R(y,y)x = R(y)x$.

□

Theorem 3.1.6. *Let $x, y \in \mathbb{R}^n$. Define $b_0 = \max_{||e||=1}\langle R(e)e, e\rangle$ Then the following properties hold.*

1. $||R(x) - R(y)|| \leq b_0||x-y|| \cdot ||x+y||$

2. $||R(x)|| \leq b_0||x||^2$

3. $||R(x,y)|| \leq b_0||x|| \cdot ||y||$

Proof. 1.

$$
\begin{aligned}
||R(x) - R(y)|| &= \max_{||e||=1} |\langle (R(x) - R(y))e, e\rangle| \\
&= \frac{1}{m} |\max_{||e||=1} \sum_{k=1}^{m} \langle x, f_k\rangle^2 \langle e, f_k\rangle^2 - \langle y, f_k\rangle^2 \langle e, f_k\rangle^2| \\
&\leq \frac{1}{m} \max_{||e||=1} \sum_{k=1}^{m} (\langle e, f_k\rangle^2) |\langle x, f_k\rangle^2 - \langle y, f_k\rangle^2|
\end{aligned}
$$

Now we use the Cauchy Schwarz Inequality twice to split the summation into three parts. To do this, let $x - y = ||x - y|| \cdot u$ and $x + y = ||x + y|| \cdot v$, where u, v are unit vectors. Then we have

$$
\begin{aligned}
||R(x) - R(y)|| &\leq \frac{1}{m} \max_{||e||=1} \sum_{k=1}^{m} \langle x - y, f_k\rangle \langle x + y, f_k\rangle \langle e, f_k\rangle^2 \\
&= ||x - y|| \cdot ||x + y|| \cdot \frac{1}{m} \max_{||e||=1} \sum_{k=1}^{m} \langle u, f_k\rangle \langle v, f_k\rangle \langle e, f_k\rangle^2 \\
&\leq ||x - y|| \cdot ||x + y|| \cdot \frac{1}{m} \max_{||e||=1} \sum_{k=1}^{m} (\langle u, f_k\rangle^2 \langle v, f_k\rangle^2)^{\frac{1}{2}} (\sum_k \langle e, f_k\rangle^4)^{\frac{1}{2}} \\
&\leq ||x - y|| \cdot ||x + y|| \cdot \frac{1}{m} \max_{||e||=1} \sum_{k=1}^{m} (\langle u, f_k\rangle^4)^{\frac{1}{4}} (\langle v, f_k\rangle^4)^{\frac{1}{4}} (\sum_k \langle e, f_k\rangle^4)^{\frac{1}{2}} \\
&= ||x - y|| \cdot ||x + y|| \cdot \max_{||e||=1} \sum_{k=1}^{m} (\frac{1}{m}\langle u, f_k\rangle^4)^{\frac{1}{4}} (\frac{1}{m}\langle v, f_k\rangle^4)^{\frac{1}{4}} (\frac{1}{m} \sum_k \langle e, f_k\rangle^4)^{\frac{1}{2}} \\
&\leq ||x - y|| \cdot ||x + y|| \cdot b_0^{\frac{1}{4}} \cdot b_0^{\frac{1}{4}} \cdot b_0^{\frac{1}{2}} \\
&= b_0 ||x - y|| \cdot ||x + y||
\end{aligned}
$$

2. Set $y = 0$ in (1) and we get the result

3. To show this one, we note for $v \in \mathbb{R}^n$, we have

$$\langle R(x,y)v, v\rangle = \frac{1}{m}\sum_{k=1}^{n}\langle x, f_k\rangle\langle y, f_k\rangle\langle v, f_k\rangle^2$$

We can rearrange and we get $\langle R(x,y)v, v\rangle = \langle R(v)x, y\rangle$ Therefore, we have

$$||R(x,y)|| = \max_{||v||=1}|\langle R(x,y)v, v\rangle| = \max_{||v||=1}|\langle R(v)x, y\rangle|$$

$$\leq \max_{||v||=1}||R(v)x|| \cdot ||y|| \leq \max_{||v||=1}||R(v)|| \cdot ||x|| \cdot ||y||$$

$$\leq b_0||x|| \cdot ||y||$$

□

If we look at the proof, we see that the constant b_0 is optimal, because if all the vectors, $x+y, x-y, e$ were equal, all the inequalities would be equalities, and such an equality is possible.

What is further interesting is that this is a natural distance on the quotient space $\mathbb{R}^n/\sim$, (it is well defined on representatives) and we'll see such distance also plays a role in the complex case.

Corollary 3.1.7. *Let $x, y \in \mathbb{R}^n$. Then*

$$||Hess(x,\lambda) - Hess(y,\lambda)|| \leq 3b_0||x-y||(||x-y|| + 2||y||)$$

Proof. It is easy to verify that $||Hess(x,\lambda) - Hess(y,\lambda)|| = 3||R(x) - R(y)||$, so by Theorem 3.1.6, we get that $||Hess(x,\lambda) - Hess(y,\lambda)|| \leq 3b_0||x-y|| \cdot ||x+y||$

Therefore, using the triangle inequality we have that $||Hess(x,\lambda) - Hess(y,\lambda)|| \leq 3b_0||x-y||(||x-y|| + 2||y||)$ as desired. □

3.1.2 Boundedness

We aim to show that if $x \neq 0$ is a critical point of the J criterion, then it is bounded within a parabolic region. Such a critical point x with $x \neq 0$ satisfies

$$(R(x) + \lambda Q - R_0)x = 0$$

Therefore, taking the inner product of that expression with x, we get

$$\langle R(x)x, x\rangle + \langle(\lambda Q - R_0)x, x\rangle = 0$$

$$\langle R(x)x, x\rangle = \langle(R_0 - \lambda Q)x, x\rangle$$

On the one hand we have

$$\langle(R_0 - \lambda Q)x, x\rangle \leq \lambda_{max}(R_0 - \lambda Q)||x||^2 \tag{3.11}$$

On the other hand we have

$$\langle R(x)x, x\rangle \geq \frac{a_0}{m}||x||^4 \tag{3.12}$$

Putting these together, we get

$$||x||^2 \leq m\frac{\lambda_{max}(R_0 - \lambda Q)}{a_0} \tag{3.13}$$

so there is a specifically parabolic form to the bound and the trajectories are bounded.

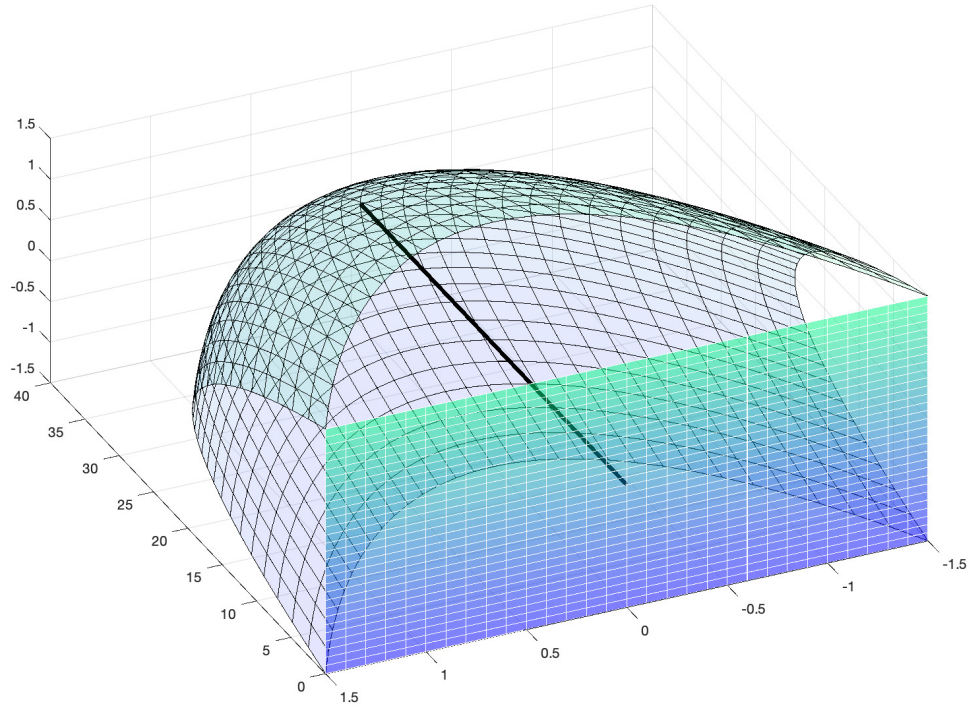

Figure 3.1: Boundedness Restriction on the Golden Retriever

In the case that $Q = I$, we get $\lambda_{max}(R_0 - \lambda I) = \lambda_1 - \lambda$, so

$$||x||^2 \leq m\frac{\lambda_1 - \lambda}{a_0} \tag{3.14}$$

(where $\lambda_1 = \lambda_{max}(R_0)$).

For general Q, instead of equation 3.11, we can say

$$\langle (R_0 - \lambda Q)x, x\rangle \leq \lambda_{max}(Q^{-\frac{1}{2}} R_0 Q^{-\frac{1}{2}} - \lambda I)||Q^{\frac{1}{2}}x||^2 \tag{3.15}$$

from which we now see that

$$||x||^2 \leq \frac{m\big(\lambda_{max}(Q^{-1}R_0) - \lambda\big)||Q||}{a_0} \tag{3.16}$$

3.1.3 Sufficiency

Let $z \in \mathbb{R}^n$ be fixed. Define $\{y_k = |\langle z, f_k\rangle|^2 + \nu_k\}_{k=1\ldots m}$ where $\nu_k \sim N(0, \sigma^2)$ are i.i.d. measurements.

Proposition 3.1.8. *R_0 is a sufficient statistic for z, if the noise is drawn from a normal.*

Proof. We aim to use the Fisher–Neyman factorization theorem. To do this, we take the PDF

$$p(y; z) = \frac{1}{(\sqrt{2\pi}\sigma)^m} exp\{\frac{-1}{2\sigma^2}\sum_{k=1}^{m}(y_k - \langle z, f_k\rangle^2)^2\} \tag{3.17}$$

Therefore, by taking the logarithm, we get

$$\begin{aligned} log(p(y; z)) &= \frac{-1}{2\sigma^2}\sum_{k=1}^{m} y_k^2 - mlog(\sqrt{2\pi}\sigma) + \frac{1}{\sigma^2}\sum_{k=1}^{m} y_k\langle z, f_k\rangle^2 - \frac{1}{2\sigma^2}\langle z, f_k\rangle^4 \\ &= \frac{-1}{2\sigma^2}\sum_{k=1}^{m} y_k^2 - mlog(\sqrt{2\pi}\sigma) + \frac{m}{\sigma^2}\langle R_0 z, z\rangle - \frac{m}{2\sigma^2}\langle R(z)z, z\rangle \end{aligned}$$

Now we can factor

$$p(y;z) = f_0(y)g(R_0,z)$$

where both f_0 and g are nonnegative functions defined by

$$f_0 = \frac{1}{(\sqrt{2\pi}\sigma)^m} exp(-\frac{1}{2\sigma^2}\sum_{k=1}^{m} y_k^2)$$

$$g(R_0,z) = exp(-\frac{m}{2\sigma^2}\langle (R(z) - 2R_0)z, z\rangle)$$

Therefore, the factorization theorem applies and R_0 is a sufficient statistic for z. $\square$

3.1.4 Assumptions

There are several assumptions we make for this algorithm, which usually will happen in the generic case, or at least with high probability.

1. The frame set $\mathcal{F}$ is phase-retrievable.

2. For a fixed λ, the set of critical points of $J(x,\lambda)$ is isolated.

3. The top eigenvalue of $Q^{-1}R_0$ has a one dimensional eigenspace.

4. Assume $x \neq 0$ and (x,λ) is a critical point of the J-criterion, so $F(x,\lambda) = 0$. Then we assume the extended hessian, $Hess_{ext}$ has full rank (rank$(Hess_{ext}) = n$) at (x,λ).

Conditions 1 and 2 ensure that it is reasonable to try to recover the signal. In fact, these conditions are not independent of eachother. If Condition 1 is not

true, then Condition 2 need not hold either, in the noiseless case. To see this, recall that if $\mathcal{F}$ is not phase-retrievable it is possible (by Theorem 1.2.2) for the matrix $R(z)$ not being strictly positive definite. Assume it is not. Then we can look at the $\nabla_x J(x,\lambda)|_{\lambda=0} = R(x)x - R(z)x = 0$. Clearly at $x = z$, the gradient is zero. However, the hessian is given by $Hess(x,\lambda)|_{x=z,\lambda=0} = 2R(z)$, which has rank strictly less than n, thus the critical point at $(z,0)$ is degerate and is not isolated.

Condition 3 will ensure that the initialization of the algorithm is well defined.

Condition 4 ensures that there is no really degenerate cases, such as bifurcations of the path, or exploding to a hypersurface, etc.

Condition 4 implies condition 2 as well. To see this, if condition 2 is not true, it is possible to find a continuous path in the fixed λ hyperplane which gives rise to a null vector of the hessian. This lifts up to another null vector of the $Hess_{ext}$, so condition 4 would not be true either. We leave condition 2, as it is important to emphasize it.

3.1.5 Initialization

Now, in the spirit of Homotopy Continuation, we would like to find solutions to (2.2) when λ is reasonably large, and get a sense on how large λ should be to get simple solutions. After this the algorithm will be to homotope the solutions to $\lambda = 0$. Therefore, from the largest eigenvalue, under our assumptions there a well defined path we can follow. By the theory developed by (among others) Rabonowitz ([25, 26]), the path can only end at $x = 0$, at an eigenvalue different from the one it

started at, or at infinity. By boundedness, the case where the path goes to infinity must cross the $\lambda = 0$ hypersurface. Thus we continue following the path until we reach either $x = 0$ or until we reach $\lambda = 0$.

Both cases are possible and do show up in numerical simulations. If one were to do these numerically, and the path ends at $x = 0$, one would then have to choose a different matrix Q in an attempt to reach $\lambda = 0$.

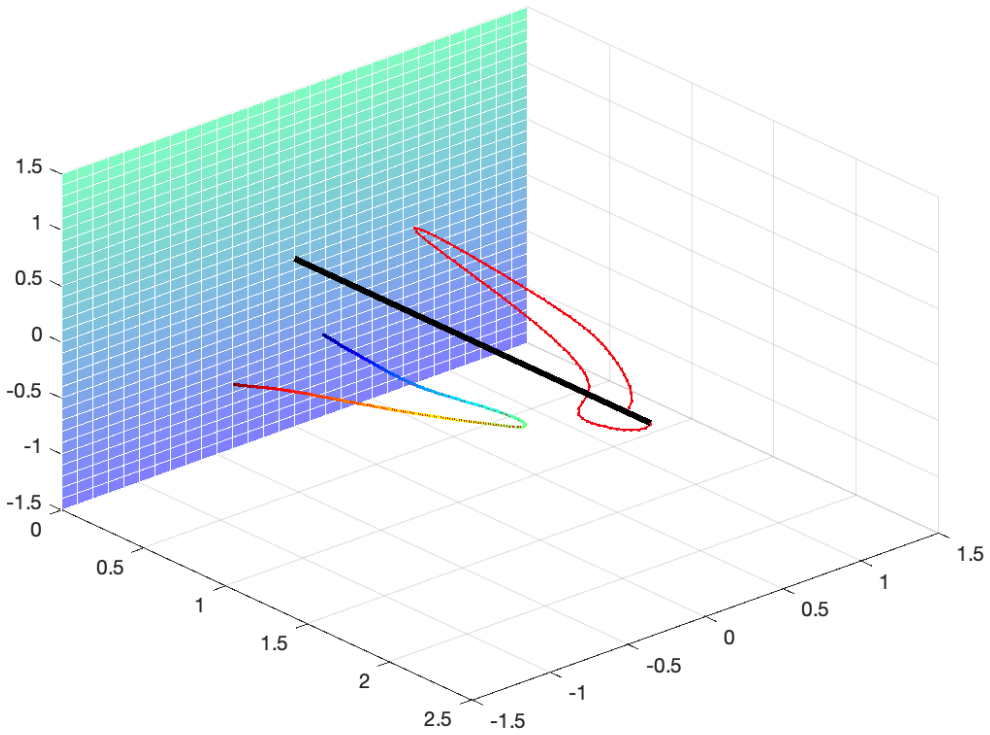

Figure 3.3: This is an example of the golden retriever turning back and ending up at the second largest eigenvalue. At the same time, the critical path from the global minimizer is shown. It was generated with $n = 5$, and $m = 20$.

Figures 3.2 and 3.3 are examples of the algorithm turning around and going to the second largest eigenvalue. In the second case, also displayed is the homotopy

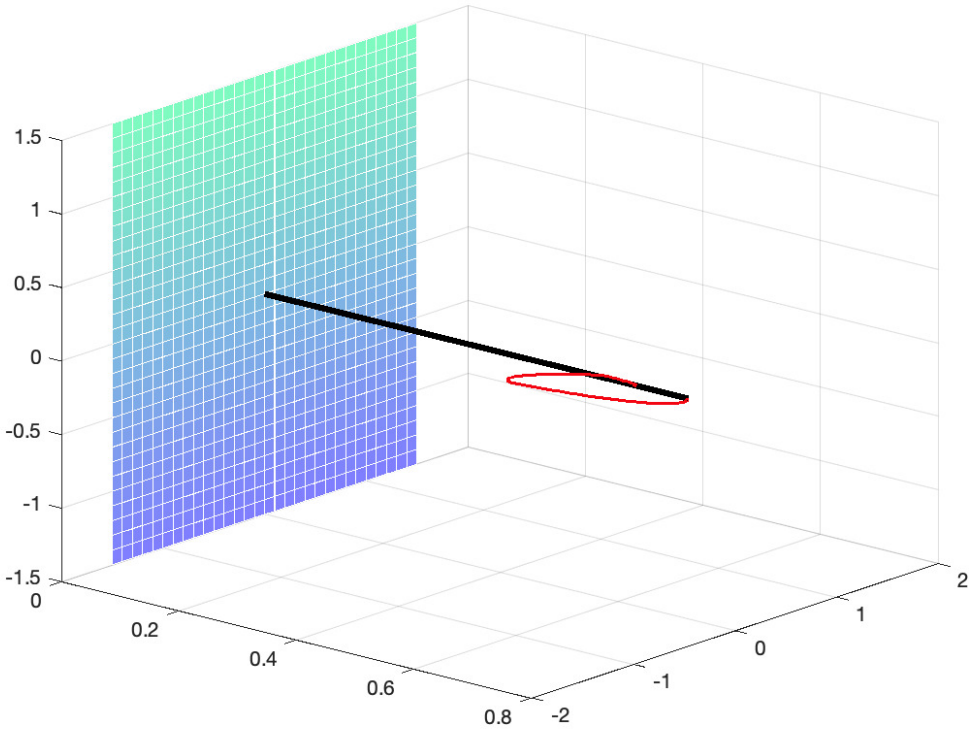

Figure 3.2: This is an example of the golden retriever turning back and ending up at the second largest eigenvalue. It was generated with $n = 5$, and $m = 20$.

path going from the true solution (in red) to another critical point at $\lambda = 0$ (in blue).

Recall that $R(x)$ is a positive semidefinite matrix, and since we assume that the frame set is phase retrievable, we can assume it is positive definite so long as $x \neq 0$.

Proposition 3.1.9. *Assume* $\lambda \geq eig_{max}(Q^{-1}R_0) = eig_{max}(Q^{-\frac{1}{2}}R_0Q^{-\frac{1}{2}})$, *then* $x = 0$ *is the solution to the optimization problem in (2.2).*

Proof. Since $R(x) \geq 0$, it is clear from (3.7) that if $(\lambda Q - R_0)$ is positive semidefinite, then $J(x, \lambda) \geq 0$ so a solution to the optimization problem in (2.2) is given by $x = 0$.

To solve for $(\lambda Q - R_0) \geq 0$, we want $\lambda Q \geq R_0$. This happens if and only if $\lambda I \geq Q^{-\frac{1}{2}}R_0Q^{-\frac{1}{2}}$. By rearranging, this implies that $(\lambda I - Q^{-\frac{1}{2}}R_0Q^{-\frac{1}{2}})$ should be positive definite. Since $Q^{-\frac{1}{2}}R_0Q^{-\frac{1}{2}}$ is symmetric, and by assumption, $\lambda > eig_{max}(Q^{-\frac{1}{2}}R_0Q^{-\frac{1}{2}})$, then $(\lambda I - Q^{-\frac{1}{2}}R_0Q^{-\frac{1}{2}})$ is positive definite, which implies that $(\lambda Q - R_0) \geq 0$.

To show that this is the same as $\lambda > eig_{max}(Q^{-1}R_0)$, note that if e is the eigenvector corresponding to the largest eigenvalue of $(Q^{-1}R_0)$, then e satisfies the equation

$$\begin{aligned} 0 &= det(eI - Q^{-1}R_0) \\ &= det(Q^{\frac{1}{2}})det(eI - Q^{-1}R_0)det(Q^{-\frac{1}{2}}) \\ &= det(eI - Q^{-\frac{1}{2}}R_0Q^{-\frac{1}{2}}) \end{aligned}$$

and since the process is reversible, these have the same eigenvalues. □

Denote λ_1 as the largest eigenvalue of $(Q^{-1}R_0)$. From our philosophy, we know how to solve our system at $\lambda = \lambda_1$, and it is achieved when $x = 0$, so this will be the reference point for λ in our algorithm.

Now we know we can initialize $\lambda = \lambda_1$, and $x = 0$, we want to know which direction to step into. This is equivalent to initializing the algorithm at $\lambda = \lambda_1 - \epsilon$ for small ϵ and determining how to initialize x.

To initialize such a vector x, we look at the ball centered at $(x, \lambda) = (0, \lambda_1 - \epsilon)$ with a sufficiently small radius. Since $R(x) \approx 0$ if $x \approx 0$, and it is a quadratic term, we neglect this term and instead solve the dominant linear terms in $F(x, \lambda) = 0$. This implies that $(R(x) + \lambda Q - R_0)x = 0 \Rightarrow (\lambda Q - R_0)x = 0$.

Thus we have

$$(\lambda Q - R_0)x = 0 \Rightarrow \lambda Q x = R_0 x \Rightarrow \lambda x = Q^{-1}R_0 x$$

So we have that to satisfy this equation, it suffices to be an eigenvector for $Q^{-1}R_0$. Denote this eigenvector e_{max}.

In other words, we initialize our algorithm such that

$$x = c \cdot e_{max} \quad , \quad \lambda = \lambda_1 - \epsilon$$

Now we need to find the constant c we initialize with.

To answer this, we want a constant c which minimizes the J criterion at $\lambda_1 - \epsilon$

$$\arg\min_c J(c \cdot e_{max}, \lambda_1 - \epsilon)$$

Expanding what this means, we get that

$$\begin{aligned}
&\arg\min_c (c^4 \frac{1}{4m} \sum_{k=1}^{m} (\langle e_{max}, f_k \rangle)^4 + c^2 \frac{1}{2} \langle (\lambda_1 Q - R_0) e_{max}, e_{max} \rangle - c^2 \frac{\epsilon}{2} \langle Q e_{max}, e_{max} \rangle + \frac{1}{4m} \sum_{k=1}^{m} y_k^2) \\
&= \arg\min_c (c^4 \frac{1}{4m} \sum_{k=1}^{m} (\langle e_{max}, f_k \rangle)^4 + c^2 \frac{1}{2} \langle (\lambda_1 Q - R_0) e_{max}, e_{max} \rangle - c^2 \frac{\epsilon}{2} \langle Q e_{max}, e_{max} \rangle) \\
&= \arg\min_c (c^4 \frac{1}{4m} \sum_{k=1}^{m} (\langle e_{max}, f_k \rangle)^4 - c^2 \frac{\epsilon}{2} \langle Q e_{max}, e_{max} \rangle)
\end{aligned}$$

This is a quadratic in c^2, thus after solving for c we we get

$$c = \sqrt{\frac{\epsilon \cdot \langle Q e_{max}, e_{max} \rangle}{\frac{1}{m} \sum_{k=1}^{m} (\langle e_{max}, f_k \rangle)^4}}$$

3.1.6 Update rules

The next thing we have to do is check how we update the algorithm. This is divided into two steps. This is done by a predictor-corrector method, which are a class of well studied methods and are commonly used for Homotopy Continuation Methods. Another way to think of these is as doing an Euler Step followed by a Fixed Point Iteration.

Step 1: The Predictor

The goal of the predictor step is to make an Euler Step in the direction of the homotopy path. Therefore, we want a new point (x, λ) that roughly follows the path $\nabla_x(J)^{-1}(0)$ which is smoothly connected to the $(0, \lambda_1)$. If we parameterize the path by t, so $x = x(t)$ and $\lambda = \lambda(t)$, we want to step in the direction based on the slope of the curve at the current point $(x(t), \lambda(t))$.

Therefore, we want to step into the direction of the tangent of this curve, which is given by $\begin{bmatrix} \frac{dx}{dt} \\ \\ \frac{d\lambda}{dt} \end{bmatrix}$.

By differentiating the equation

$$F(x(t), \lambda(t)) = 0$$

we can find the tangent by computing the derivative

$$\frac{d}{dt}F(x(t), \lambda(t)) = 0 \tag{3.18}$$

From the work we did earlier we know

$$\frac{d}{dt}F(x(t), \lambda(t)) = Hess_{ext}(x(t), \lambda(t)) \begin{bmatrix} \frac{dx}{dt} \\ \\ \frac{d\lambda}{dt} \end{bmatrix} = 0 \tag{3.19}$$

Therefore the direction we want to step in the same direction as the vector in the null space of $Hess_{ext}(x(t), \lambda(t))$.

To summarize this, in the predictor step, we compute the extended hessian matrix $Hess_{ext}$, find the vector in the null space which matches sign in the largest

coordinate with the sign of the coordinate in the previous step (to make sure the path is moving in the correct direction), and then make a choice in step size. Unfortunately, it likely going to step away from the path, so we need a corrector algorithm to get us back on the path.

Step 2: The Corrector

In this, we want to find a point (x, λ) which is a solution to the gradient being zero but is as close as possible to the point in the Predictor Step. We can use the Newton Step. If $\begin{bmatrix} x_{old} \\ \lambda_{old} \end{bmatrix}$ was our old estimate, we can update it with a correction of the form

$$\begin{bmatrix} x_{new} \\ \lambda_{new} \end{bmatrix} = \begin{bmatrix} x_{old} \\ \lambda_{old} \end{bmatrix} - Hess^{+}_{ext} F(x_{old}, \lambda_{old}) \tag{3.20}$$

Where $Hess^{+}_{ext}$ is the pseudo-inverse of the extended hessian.

The Newton corrector step is well studied, and under suitable conditions on the extended hessian, is guaranteed to converge to a critical point after a number of corrector steps. See Chapter 3 of [9], specifically Theorem 3.4.1 in for a full treatment on the subject.

One can also modify the extended hessian $Hess_{ext}$ to ensure that the new critical point stays in a specific hyperplane by adding appropriate rows to the matrix.

3.2 Expected System

In this section we analyze the expected system if a single frame vector, $f \sim \mathcal{N}(0, I_n)$.

Lemma 3.2.1. *For f_i, f_j, f_k, f_l elements of a vector from any frame vector f, then we have the following:*

$$\mathbb{E}_{f\sim\mathcal{N}(0,I_n)}(f_i f_j f_k f_l) == \begin{cases} 1 & \textit{if the indices match in distinct pairs} \\ 3 & \textit{if } i=j=k=l \\ 0 & \textit{otherwise} \end{cases}$$

Proof. In case one, without loss of generality, say $i = j, k = l, i \neq k$, then $\mathbb{E}(f_i f_j f_k f_l) = \mathbb{E}(f_i f_j)\mathbb{E}(f_k f_l) = \mathbb{E}(f_i^2)\mathbb{E}(f_k^2) = 1$

Let $x = f_i$. In case two, it reduces to finding

$$\mathbb{E}(x^4) = \int_{-\infty}^{\infty} x^4 \exp -\frac{x^2}{2} dx = 3 \cdot \int_{-\infty}^{\infty} x^2 \exp -\frac{x^2}{2} dx = 3\mathbb{E}(x^2) = 3$$

The second equality is by integration by parts.

Case three is obvious, as there is a distinct index, independent from the others whose expectation is 0.

□

Proposition 3.2.2. $\mathbb{E}(R(x)) = ||x||^2 I + 2xx^T$

Proof. Let us examine $\mathbb{E}(R(x))$. By definition, this is: $\mathbb{E}(\frac{1}{m}\sum_{k=1}^{m} |\langle x, f_k\rangle| f_k f_k^T)$

$$\mathbb{E}(\frac{1}{m}\sum_{k=1}^{m} |\langle x, f_k\rangle|^2 f_k f_k^T)$$
$$= \frac{1}{m}\sum_{k=1}^{m} \mathbb{E}(|\langle x, f_k\rangle|^2 f_k f_k^T)$$

So we reduced it to finding the expected value of $M = |\langle x, f_k \rangle| f_k f_k^T$.

To do this, we look at the (i, j) component, we have that

$$\begin{aligned} M_{(i,j)} &= |\langle x, f_k \rangle|^2 (f_k f_k^T)_{(i,j)} \\ &= (\sum_{l=1}^{n} x_l (f_l))^2 f_i f_j) \\ &= (\sum_{p=1}^{n} \sum_{q=1}^{n} x_p x_q f_p f_q f_i f_j) \end{aligned}$$

Therefore $\mathbb{E}(M)_{(i,j)}$ reduces to finding $\mathbb{E}(f_p f_q f_i f_j)$.

When $i \neq j$ (off diagonal terms of M) we have

$$\begin{aligned} \mathbb{E}(M)_{(i,j)} &= \sum_{p=1}^{n} \sum_{q=1}^{n} x_p x_q \mathbb{E}(f_p f_q f_i f_j) \\ &= 2 x_i x_j \mathbb{E}(f_i^2)^2 = 2 x_i x_j \end{aligned}$$

When $i = j$ (diagonal terms of M) we have

$$\begin{aligned} \mathbb{E}(M)_{(i,i)} &= \sum_{p=1}^{n} \sum_{q=1}^{n} x_p x_q \mathbb{E}(f_p f_q f_i f_i) \\ &= \sum_{p=1, p \neq i}^{n} x_p^2 \mathbb{E}(f_i^2)^2 + x_i^2 \mathbb{E}(f_i^4) \\ &= ||x||^2 - x_i^2 + 3 x_i^2 \\ &= ||x||^2 + 2 x_i^2 \end{aligned}$$

Therefore we have found that

$$\mathbb{E}(M) = ||x||^2 I_n + 2xx^T \tag{3.21}$$

Therefore, since this doesn't depend on the dimension m, we see that

$$\mathbb{E}(R(x)) = \frac{1}{m}\sum_{k=1}^{m}\mathbb{E}(M) = \frac{1}{m}\cdot m\mathbb{E}(M) = \mathbb{E}(M)$$

Therefore we have that

$$\mathbb{E}(R(x)) = ||x||^2 I_n + 2xx^T \tag{3.22}$$

□

Corollary 3.2.3. *Let z denote the true signal. Then in the noiseless case, we have:*

$$\mathbb{E}(R_0) = ||z||^2 I + 2zz^T \tag{3.23}$$

Corollary 3.2.4. $\mathbb{E}(F(x,\lambda)) = ((||x||^2) - ||z||^2 + \lambda)I) + 2xx^T - 2zz^T)x$

We can solve the system of equations given above, what we call the expected system. To do so, note that the spectrum of $\mathbb{E}(R_0)$ is given by $\{3||z||^2, ||z||^2, ..., ||z||^2\}$, so $\lambda_1(\mathbb{E}(R_0)) = 3||z||^2$. Now if we guess that $x = kz$, for some scaling function k,

we can derive

$$0 = ((k^2||z||^2) - ||z||^2 + \lambda)I) + 2k^2zz^T - 2zz^T)z$$
$$= ((k^2 - 1)||z||^2 + \lambda + 2(k^2 - 1)||z||^2)z$$

Setting the coefficient of z equal to 0 gives us

$$0 = 3k^2 + \frac{\lambda}{3||z||^2} - 3$$
$$k^2 = 1 - \frac{\lambda}{3||z||^2}$$

So we get that $k = \sqrt{1 - \frac{\lambda}{\lambda_1(\mathbb{E}(R_0))}}$. Therefore, a solution to our expected system is given by

$$x(\lambda) = \left(\sqrt{1 - \frac{\lambda}{\lambda_1(\mathbb{E}(R_0))}}\right)z \tag{3.24}$$

Corollary 3.2.5. $\mathbb{E}(Hess(x, \lambda)) = (3||x||^2 - ||z||^2)I + 6xx^T - 2zz^T + \lambda Q$

The next theorem establishes the concentration of $R(x)$ about its mean, which will prove invaluable in proving the convergence of the algorithm. We first need to state some Lemmas involving the normal distribution.

Lemma 3.2.6. *[23] Let $v \sim \mathcal{N}(0, I_m)$. Then for any $\epsilon_0 > 0$ there exists an upper bound $C(\epsilon_0)$ such that for $m \geq C(\epsilon_0)$, each of the following hold with probability at*

least $1 - \frac{1}{m^2}$

$$\frac{1}{m}\sum_{k=1}^{m} v_k^2 - 1 < \epsilon_0$$
$$\frac{1}{m}\sum_{k=1}^{m} v_k^4 - 3 < \epsilon_0$$
$$\frac{1}{m}\sum_{k=1}^{m} v_k^6 - 15 < \epsilon_0$$
$$\max_{1\le k\le m} |v_k| \le \sqrt{10log(m)}$$

Furthermore, such probabilities are achieved with

$$C(\epsilon_0) = C_c = \max\{4.76 \times 10^{13}, \frac{9}{4}\frac{10395^2}{\epsilon_0^2}, \frac{71^2}{\epsilon_0^2} log(\frac{1}{\epsilon_0})^2\}$$

Proof. We will start by showing the second inequality. The first and third can be done similarly. We want to find $\mathbb{P}(\sum_{k=1}^{m} v_k^4 > (3+\epsilon_0)m)$. Denote this probability as P. Define a constant L, which we will specify later, but is allowed to depend on m and ϵ_0.

We know that

$$P = \mathbb{P}(\sum_{k=1}^{m} v_k^4 > (3+\epsilon_0)m) = \mathbb{P}(\sum_{k=1}^{m} v_k^4 > (3+\epsilon_0)m \;\Big|\; |v_1|,...,|v_m| \le L)\mathbb{P}(|v_1|,...,|v_m| \le L)$$
$$+\mathbb{P}(\sum_{k=1}^{m} v_k^4 > (3+\epsilon_0)m) \;\Big|\; |v_1| \ge L \vee ... \vee |v_m| \le L) \cdot \mathbb{P}(|v_1| \ge L \vee ... \vee |v_m| \le L)$$
$$\le \mathbb{P}(\sum_{k=1}^{m} v_k^4 > (3+\epsilon_0)m \;\Big|\; |v_1|,...,|v_m| \le L) \cdot 1 + 1 \cdot \mathbb{P}(|v_1| \ge L \vee ... \vee |v_m| \le L)$$
$$\le \mathbb{P}(\sum_{k=1}^{m} v_k^4 > (3+\epsilon_0)m \;\Big|\; |v_1|,...,|v_m| \le L) \cdot 1 + m \cdot \mathbb{P}(|v_1| \ge L)$$

Now to bound these terms, we start with the second and recall that since $erfc(z) \leq e^{-z^2}$ ([27]), we get that $\mathbb{P}(|v_1| \geq L) \leq e^{\frac{-L^2}{2}}$ (see Proposition A.3.1). Now for the first inequality, we can bound it by Bernstein's inequality so after centering we see that

$$\mathbb{P}(\sum_{k=1}^{m}(v_k^4 - 3) > \epsilon_0 m \quad \Big| \quad |v_1|, ..., |v_m| \leq L) \leq exp(\frac{-\frac{1}{2}\epsilon_0^2 m}{105 + \frac{1}{3}L^4\epsilon_0}) \tag{3.25}$$

Therefore $P \leq exp(\frac{-\frac{1}{2}\epsilon_0^2 m}{105+\frac{1}{3}L^4\epsilon_0}) + mexp(-\frac{L^2}{2})$, and choosing $L = m^{\frac{1}{8}}$ shows that $P \leq \frac{1}{m^2}$ for sufficiently high $C(\epsilon_0)$.

The expressions for the other two cases are very similar. For v_k^6 we get $P_6 \leq exp(\frac{-\frac{1}{2}\epsilon_0^2 m}{10395+\frac{1}{3}L^6\epsilon_0}) + mexp(-\frac{L^2}{2})$

Similarly, for v_k^2 we get $P_2 \leq exp(\frac{-\frac{1}{2}\epsilon_0^2 m}{3+\frac{1}{3}L^2\epsilon_0}) + mexp(-\frac{L^2}{2})$

To estimate the $C(\epsilon_0)$ needed, we will look at the v_k^6 case and we examine the sufficient bounds $mexp(\frac{-m^{1/6}}{2}) \leq \frac{1}{2m^2}$, and $exp(\frac{-\frac{1}{2}\epsilon_0^2 m}{10395+\frac{1}{3}L^6\epsilon_0}) \leq \frac{1}{2m^2}$.

The first one is true for any $m \geq 4.76 \times 10^{13}$ by a direct check. For the second term, is is sufficient for

$$exp(\frac{-\frac{1}{2}\epsilon_0^2 m}{\sqrt{m}\epsilon_0}) \leq \frac{1}{2m^2}$$

so long as $10395 \leq \frac{2}{3}\sqrt{m}\epsilon_0$. or $m \geq \frac{9}{4}\frac{10395^2}{\epsilon_0^2}$. This simplifies to $exp(-\frac{1}{2}\epsilon_0\sqrt{m}) \leq \frac{1}{2m^2}$. Define $x = \sqrt{m}\epsilon_0$, so this is equivalent to $exp(-\frac{1}{2}x) \leq \frac{\epsilon_0^4}{2x^4}$, so we rewrite

$$2x^4 exp(-\frac{1}{2}x) \leq \epsilon_0^4$$

It is sufficient to let $2x^4 exp(-\frac{1}{4}x) \leq 1$ and $exp(-\frac{1}{4}x) \leq \epsilon_0^4$. By a direct check, the first one is satisfied for $x \geq 71$ or equivalently $m \geq \frac{71^2}{\epsilon_0^2}$, and the second is satisfied for $x \geq 16log(\frac{1}{\epsilon_0})$, or equivalently $m \geq \frac{256}{\epsilon_0^2} log(\frac{1}{\epsilon_0})^2$. Therefore both of these are satisfied if $m \geq \frac{71^2}{\epsilon_0^2} log(\frac{1}{\epsilon_0})^2$ for $C(\epsilon_0) = C_c = \max\{4.76 \times 10^{13}, \frac{9}{4}\frac{10395^2}{\epsilon_0^2}, \frac{71^2}{\epsilon_0^2} log(\frac{1}{\epsilon_0})^2\}$, we get the probabilities needed, and we can see that the constants for the v_k^4 and v_k^2 cases would be smaller.

For the last inequality, we want to know what is the probability that $\mathbb{P}(|v_k|^2 \geq 10\epsilon_0 m) = \mathbb{P}(|v_k| \geq \sqrt{10\epsilon_0 m})$. By the $erfc(z)$ inequality given above, we know that $\mathbb{P}(|v_k| \geq \sqrt{10\epsilon_0 m}) \leq exp(\frac{-10log(m)}{2}) = m^{-5}$. Now we apply the union bound and see that $\mathbb{P}(\max_k |v_k| \geq \sqrt{10\epsilon_0 m}) \leq m \cdot m^{-5} = m^{-4} \leq \frac{1}{m^2}$ □

Theorem 3.2.7. *[23] Assume $f_k \sim \mathcal{N}(0, I)$ and $||x|| = 1$. Choose $\epsilon > 0$ and $\gamma > log(9)$. There exists a function $C(\epsilon, \gamma) > 0$, independent of n, such that for every $m \geq C(\epsilon, \gamma) n\, log(n)$, $||R(x) - \mathbb{E}[R(x)]|| \leq \epsilon$ with probability $1 - 5e^{-\gamma n} - \frac{4}{n^2}$. Let $\epsilon_0 = \frac{\epsilon}{8}$ and $\delta_0 = \frac{\epsilon}{12}$. Furthermore, let $C_c = \max\{4.76 \times 10^{13}, \frac{9}{4}\frac{10395^2}{\epsilon_0^2}, \frac{71^2}{\epsilon_0^2} log(\frac{1}{\epsilon_0})^2\}$ as in the preceding lemma, $C_0 = \max\{\sqrt{40/30}\frac{\sqrt{\gamma}}{\delta_0}, 16\frac{\gamma}{\delta_0}\}$, $C_1 = 2\sqrt{\frac{\gamma}{\delta_0}}$, then a sufficient upperbound for C would be $C(\epsilon, \gamma) = \max\{C_c, 16C_1^2, 80C_0, 1600C_0^2\}$*

Proof. We follow the proof of Lemma 7.4 in [23]. By unitary invariance, we let $x = e_1$, the first canonical basis vector. Let $||y|| = 1$ and we write $y = (y(1), \widetilde{y})$ and $f_k = (v_k, \widetilde{f_k})$. We examine the quantity

$$\begin{aligned} I_0(y) &= |y^T(R(e_1) - (I + 2e_1 e_1^T))y| \\ &= |\frac{1}{m}\sum_{k=1}^{m} v_k^2 \langle y, f_k\rangle^2 - (1 + 2y(1)^2)| \end{aligned}$$

Now notice that $\langle y, f_k\rangle = y(1)v_k + \langle \widetilde{y}, \widetilde{f}_k\rangle$, so $\langle y, f_k\rangle^2 = y(1)^2 v_k^2 + 2y(1)v_k\langle \widetilde{y}, \widetilde{f}_k\rangle\langle \widetilde{y}, \widetilde{f}_k\rangle^2$

This together with the fact that $y(1)^2 + ||\widetilde{y}|| = 1$ gives

$$I_0(y) = |\frac{1}{m}\sum_{k=1}^{m} v_k^4 y(1)^2 + 2v_k^3 y(1)\langle \widetilde{y}, \widetilde{f}_k\rangle + v_k^2\langle \widetilde{y}, \widetilde{f}_k\rangle^2 - 1 - 2y(1)^2|$$
$$= |\frac{1}{m}\sum_{k=1}^{m}(v_k^4 - 3)y(1)^2 + \frac{2}{m}\sum_{k=1}^{m} v_k^3 y(1)\langle \widetilde{y}, \widetilde{f}_k\rangle + \frac{1}{m}\sum_{k=1}^{m} v_k^2\langle \widetilde{y}, \widetilde{f}_k\rangle^2 - ||\widetilde{y}||^2|$$

Splitting the last term using the triangle inequality gives

$$I_0(y) \leq |\frac{1}{m}\sum_{k=1}^{m}(v_k^4 - 3)|y(1)^2 + |\frac{1}{m}\sum_{k=1}^{m}(v_k^2 - 1)| \cdot ||\widetilde{y}||^2$$
$$+2|\frac{1}{m}\sum_{k=1}^{m} v_k^3 y(1)\langle \widetilde{y}, \widetilde{f}_k\rangle| + |\frac{1}{m}\sum_{k=1}^{m} v_k^2(\langle \widetilde{y}, \widetilde{f}_k\rangle^2 - ||\widetilde{y}||^2)|$$
$$\leq 2\epsilon + 2|\frac{1}{m}\sum_{k=1}^{m} v_k^3 y(1)\langle \widetilde{y}, \widetilde{f}_k\rangle| + |\frac{1}{m}\sum_{k=1}^{m} v_k^2(\langle \widetilde{y}, \widetilde{f}_k\rangle^2 - ||\widetilde{y}||^2)|$$

Now for the last term, we can apply Hoeffding's inequality (Proposition 5.10 in [28]) gives us that for any constants δ_0 and γ, $m \geq C_1(\delta_0, \gamma)\sqrt{n(\sum_{k=1}^{m} v_k^6)}$ we have

$$|\frac{1}{m}\sum_{k} v_k^3 y(1)\langle \widetilde{y}, \widetilde{f}_k\rangle| \leq \delta_0 |y(1)| \cdot ||\widetilde{y}|| \leq \delta_0$$

holds with probability at least $1 - 3e^{-2\gamma n}$. Here one can choose $C_1 = 2\sqrt{\frac{\gamma}{\delta_0}}$ (see Proposition A.3.2).

For the final term, we apply the Bernstein-type inequality (Proposition 5.16 in [28]) which asserts: for any positive δ_0, γ, there exist constants

$m \geq C_0(\delta_0, \gamma)\big(\sqrt{n\sum_{k=1}^m v_k^4} + n \cdot max_{k=1,..,m}|v_k|^2\big)$, such that

$$|\frac{1}{m}\sum_{k=1}^{m} v_k^2(\langle \widetilde{y}, \widetilde{f}_k\rangle^2 - ||\widetilde{y}||^2)| \leq \delta_0||\widetilde{y}||^2 \leq \delta_0$$

holds with probability at least $1-2e^{-2\gamma n}$. Here one can choose $C_0 = \max\{\sqrt{40/3}\frac{\sqrt{\gamma}}{\delta_0}, 16\frac{\gamma}{\delta_0}\}$ (see Proposition A.3.3).

Therefore, for any unit norm vector y, $I_0(y) \leq 2\epsilon_0 + 3\delta_0$ holds with probability at least $1 - 5e^{-2\gamma n}$. By Lemma 5.4 in [28], we can bound the operator norm via an ϵ-net argument, so

$$||R(x) - \mathbb{E}[R(x)]|| = \max_{y\in\mathbb{S}_1^{n-1}} I_0(y) \leq 2\max_{y\in\mathcal{N}} I_0(y) \leq 4\epsilon_0 + 6\delta_0 = \epsilon$$

where $\mathcal{N}$ is a 1/4 - net of $\mathbb{S}_1^{n-1}$.

Therefore, using that the cardinality of a $\frac{1}{4}$ - net can be achieved by 9^n points, and by applying the union bound it follows that the theorem holds with probability $1 - 5e^{-\gamma n}$ so long as $m \geq \max\{C_1\sqrt{n\sum_{k=1}^m f_k(1)^6}, C_0(\sqrt{n\sum_{k=1}^m f_k(1)^4} + n\max_k f_k(1)^2)\}$

Write $C_c = \max\{4.76 \times 10^{13}, \frac{9\cdot 64}{4}\frac{10395^2}{\epsilon^2}, \frac{71^2\cdot 64}{\epsilon^2} log(\frac{8}{\epsilon_0})^2\}$ as in the lemma.

The theorem holds with the probability $1 - 5e^{-\gamma n} - \frac{4}{n^2}$ when $m \geq Cnlog(n)$, for C sufficiently large.

To find a C sufficiently large, we examine each of those terms. First we look at $m \geq C_1\sqrt{n\cdot\sum_k f_k(1)^6}$. By lemma 3.2.6, we get that for $m \geq C_c\sum_k f_k(1)^6 < m(15+\epsilon_0) < 16m$ so $m \geq C_1\sqrt{n\cdot 16m}$. This implies that $m \geq 16C_1^2 n$.

Examining the other term, we want $m \geq C_0\sqrt{n\sum_k f_k(1)^4} + C_0 \max_k f_k(1)^2$. Again, using the lemma, we see that it is sufficient if $\frac{m}{2} \geq C_0\sqrt{n\sum_k f_k(1)^4}$ and $\frac{m}{2} \geq C_0 n \max_k f_k(1)^2$. We examine the first of these terms.

$$\frac{m}{2} \geq C_0\sqrt{n\sum_k f_k(1)^4} \geq C_0\sqrt{4nm}$$

$$m \geq 16C_0^2 n$$

Using the lemma, the other term gives us

$$\frac{m}{2} \geq C_0 \max_k f_k(1)^2 \geq C_0 10 log(m)$$

$$m \geq 20C_0 nlog(m)$$

now we take $m = C_2 nlogn$ and we get

$$m \geq 20C_0 nlog(C_2 nlogn) = 20C_0 nlog(n) + 20C_0 nlog(C_2) + 20C_0 nlog(log(n))$$

Bounding $log(log(n))$ by $log(n)$, we get

$$m \geq 40C_0 nlog(n) + 20C_0 nlog(C_2)$$

Again bounding each by $\frac{m}{2}$ gives us $\frac{C_2 nlog(n)}{2} \geq 20C_0 nlog(C_2)$ which is satisfied if $C_2 \geq 40C_0 log(C_2)$. Bounding $log(C_2)$ by $\sqrt{C_2}$, this is satisfied if $C_2 \geq 1600C_0^2$. On

the other hand, we need $m \geq 40C_0 nlog(n)$ so we want

$$C_2 nlogn \geq 80C_0 nlogn$$

so we need $C_2 \geq 80C_0$. Putting everything together, it suffices to take $m \geq Cnlog(n)$, with $C = \max\{1600 \cdot C_0^2, 80C_0, 16C_1^2, C_c\}$

□

3.3 Analysis of the minimum distance between critical points

It will be important that the zero of the gradient of $J(x, \lambda)$ is isolated within some radius around it. To this end, we want to find an estimate for the minimum distance to the next zero.

Theorem 3.3.1. *Let z denote the global minimizer of $J(x, \lambda)$ at $\lambda = 0$. Let z' denote any other critical point at $\lambda = 0$. Then $||z - z'|| \geq \frac{2}{3}\sqrt{\frac{\lambda_n(R(z))}{b_0}}$*

Proof. Since z denotes the global minimizer at $\lambda = 0$ of the J-Criterion, z satisfies

$$F(z, 0) = R(z)z - R_0 z = 0$$

$$R(z)z = R_0 z$$

Now consider the another point $z' = z + te$, with $||e|| = 1$. Therefore, we can write

a polynomial for t by the evaluating the gradient at the point z'

$$F(z',0) = R(z')z' - R_0 z'$$

$$= R(z+te)(z+te) - R_0(z+te)$$

$$= (R(z) + R(te) + 2R(z,te))z + (R(z) + R(te) + 2R(z,te))(te) - R_0 z - R_0(te)$$

$$= (R(z) - R_0)z + t^2 R(e)z + 2tR(z,e)z + t(R(z) - R_0)e + t^3 R(e)e + 2t^2 R(z,e)e)$$

Now we know $R(z,e)z = \frac{1}{m}\sum_{k=1}^{m}\langle z, f_k\rangle^2\langle e, f_k\rangle f_k = \frac{1}{m}\sum_{k=1}^{m}\langle z, f_k\rangle^2 f_k f_k^T e = R(z)e$. Similarly, we have: $R(z,e)e = R(e)z$.

Therefore, the above simplifies to the following

$$= (R(z) - R_0)e + t(3R(z) - R_0)e + 3t^2 R(e)z + t^3 R(e)e$$

$$= t(3R(z) - R_0)e + 3t^2 R(e)z + t^3 R(e)e$$

Where the second equality follows from the fact that z is a critical point of the gradient.

Now note that $3R(z) - R_0 = Hess(z,0)$, and that this is necessarily positive definite by the global minimality of z. Therefore, we will write $H_z = 3R(z) - R_0$. In the noiseless case, we note that $R(z) = R_0$, so $H_z = 2R(z)$.

We can now look at the following polynomial expression

$$P(t) = \langle F(z+te,\lambda), e\rangle = t[\langle H_z e, e\rangle + 3t\langle R(e)z, e\rangle + t^2\langle R(e)e, e\rangle] \tag{3.26}$$

We write $P(t) = tQ(t)$. Now we want to locate zeros of the gradient. It is clear that that can only happen if $P(t) = 0$. $t = 0$ is a solution which corresponds to the critical point at z, so we want to look at roots of the polynomial $Q(t)$.

$Q(t)$ is a convex polynomial which is positive at $t = 0$ (since H_z is positive semidefinite), so we can approximate the root by: (setting $z = z_0||z||$)

$$|t_0| \geq \frac{Q(0)}{|Q'(0)|} = \frac{\langle H_z e, e\rangle}{3|\langle R(e)z, e\rangle|}$$
$$= \frac{1}{3}\frac{\langle H_z e, e\rangle}{|\langle R(e)z, e\rangle|}$$

In the noiseless case we have:

$$|t_0| \geq \frac{2}{3}||z||\frac{\langle R(z_0)e, e\rangle}{|\langle R(e)z_0, e\rangle|}$$

So what remains is to bound this quantity over all possible directions e using properties of the frame and the magnitude of z.

We want to find a lower bound for this. We do this for the noiseless case and do so by rewriting the denominator as something larger which has the numerator as a factor. So we start with the denominator:

$$|\langle R(e)z_0, e\rangle| = |\langle R(e)^{\frac{1}{2}}e, R(e)^{\frac{1}{2}}z_0\rangle| \leq ||R(e)^{\frac{1}{2}}e|| \cdot ||R(e)^{\frac{1}{2}}z_0||$$

$$= \sqrt{\langle R(e)e, e\rangle}\sqrt{\langle R(e)z_0, z_0\rangle} = \sqrt{\langle R(e)e, e\rangle}\sqrt{\langle R(z_0)e, e\rangle} \leq \sqrt{b_0}\sqrt{\langle R(z_0)e, e\rangle}$$

Now we can rewrite the whole expression using the following

$$|t_0| \geq \frac{2}{3\sqrt{b_0}}||z||\sqrt{\langle R(z_0)e, e\rangle} \geq \frac{2}{3\sqrt{b_0}}\sqrt{\lambda_n(R(z_0))}||z||$$
$$= \frac{2}{3\sqrt{b_0}}\sqrt{\lambda_n(R(z))}$$

This gives us the inequality we want. We also note that this is $\geq \frac{2}{3}\sqrt{\frac{a_0}{b_0}}||z||$, but this proves to be less useful. □

3.4 Real Convergence Analysis

We look to the expected system to provide a means of analyzing the convergence properties of the system. If we assume we can parameterize the homotopy path by λ, we want a curve $(\varphi(\lambda), \lambda)$, known as the reference path, from which we will measure how much our golden retriever path $(x(\lambda), \lambda)$ deviates from. An interesting choice to try is the curve gotten from solving the expected system, denoted by φ_0, but it turns out this doesn't have the right theoretical properties near $\lambda = \lambda_1$. Instead, we will analyze the curve gotten from taking the convex combination of φ_0 and the eigenvector of R_0 corresponding to λ_1. We show that if the homotopy path doesn't deviate too far from this reference curve, then it converges to the global optimal. Then we say, for sufficiently large m, that the conditions needed for this deviation will be satisfied.

The main idea is to define the reference path that goes from $\lambda = \lambda_1$ to $\lambda = 0$, then for each λ to define the radius of a sphere in $\mathbb{R}^n$ such that no other critical point

is on the sphere. Over all λ, this changing radius forms a tube with no critical points on the boundary. If this condition is satisfied for all λ, then the critical path defined by the golden retriever stays inside this tube and no other critical point enters the tube. If the radius is smaller than the distance to the nearest critical point at $\lambda = 0$, then the homtopy path is forced to converge to the global minimizer.

Definition 3.4.0.1. *We call a reference path $\varphi(\lambda)$* ***suitable*** *if it satisfies the following conditions.*

- *It is a smooth path parameterized by λ for $0 \leq \lambda \leq \lambda_1$*
- *$\varphi(\lambda_1) = 0$, and $\varphi(\lambda)$ is nonzero for $\lambda < \lambda_1$.*
- *$\varphi(0) = z$, the global minimizer*

First we assume we are given a suitable reference path $\varphi(\lambda)$.

The following lemma shows how the gradient varies when it is perturbed.

Lemma 3.4.1. *Assume $x_1 = x_2 + \delta$. Then we have*

$$F(x_1, \lambda) = F(x_2, \lambda) + Hess(x_2, \lambda)\delta + 3R(\delta)x_2 + R(\delta)\delta \tag{3.27}$$

Proof. This is a direct computation, it is also true in the noisy case and for general Q.

$F(x_1, \lambda) = R(x_1)x_1 + \lambda Q x_1 - R_0 x_1$, and since $R(x_1) = R(x_2 + \delta) = R(x_2) + R(\delta) + 2R(x_2, \delta)$, we get

$$F(x_1,\lambda) = \big(R(x_2) + R(\delta) + 2R(x_2,\delta)\big)(x_2+\delta) + \lambda Q(x_2+\delta) - R_0(x_2+\delta)$$

$$= \big(R(x_2)x_2 + \lambda Q x_2 - R_0 x_2\big) + \big(3R(x_2)\delta + \lambda Q\delta - R_0\delta\big) + 3R(\delta)x_2 + R(\delta)\delta$$

$$= F(x_2,\lambda) + Hess(x_2,\lambda)\delta + 3R(\delta)x_2 + R(\delta)\delta$$

□

Now we define a boundary region, which for each $0 \leq \lambda \leq \lambda_1$ is given by a sphere of radius $r(\lambda)$. The following theorem will give us a criterion to check there are no critical points on the this region, which we name the leash of the retriever.

Theorem 3.4.2. *Define the radius* $r(\lambda) = \frac{s_n(\lambda)}{6b_0||\varphi(\lambda)||}$. *If the following condition is satisfied* $||F(\varphi,\lambda)|| \leq \frac{s_n(\lambda)^2}{12b_0||\varphi(\lambda)||}$, *then no other critical points are on the sphere of radius* $r(\lambda)$, *centered at* $\varphi(\lambda)$.

Proof. A sufficient condition to ensure that there are no critical points on the boundary of the sphere is to require

$$\langle F(s,\lambda), s-\varphi(\lambda)\rangle > 0 \quad \forall s \in \mathbb{S}^n_{r(\lambda)}(\varphi(\lambda)) \tag{3.28}$$

If we let $\delta = s - \varphi$, and we use lemma 3.4.1, we see that this is equivalent to

$$\Big\langle F(\varphi,\lambda) + (Hess(\varphi,\lambda) + R(\delta))\delta + 3R(\delta)\varphi, \delta \Big\rangle > 0 \tag{3.29}$$

Now if we make the substitution $\delta = ru$, where r is a positive radius, and $u \in$

$\mathbb{S}^n_1(\varphi(\lambda))$ then an sufficient condition becomes

$$\left\langle F(\varphi,\lambda) + r(Hess(\varphi,\lambda) + r^2R(u))u + 3r^2R(u)\varphi, u \right\rangle > 0 \tag{3.30}$$

Therefore, expanding this out we see that

$$r^3\langle R(u)u,u\rangle + r\langle Hess(\varphi)u,u\rangle > -\langle F(\varphi,\lambda),u\rangle - 3r^2\langle R(u)\varphi,u\rangle \tag{3.31}$$

A lower bound for the left hand side is given by $r^3a_4^4 + rs_n(Hess(\varphi)) > rs_n(Hess(\varphi))$ and an upper bound for the right hand side is given by $||F(\varphi,\lambda)|| + 3r^2||R(u)||\cdot||\varphi|| \leq ||F(\varphi,\lambda)|| + 3r^2b_0 \cdot ||\varphi||$.

Therefore, a sufficient condition for this is

$$rs_n(Hess(\varphi)) > ||F(\varphi,\lambda)|| + 3r^2b_0 \cdot ||\varphi|| \tag{3.32}$$

Define the polynomial $P_2(r) = 3r^2b_0 \cdot ||\varphi|| - rs_n(Hess(\varphi)) + ||F(\varphi,\lambda)||$, this sufficient condition is satisfied in the region of this polynomial where it is negative. This is satisfied with two roots if the discriminant is positive, or equivalently if $s_n^2(Hess(\varphi)) - 12b_0||\varphi|| \cdot ||F(\varphi,\lambda)|| > 0$, or rewritten if

$$\frac{s_n^2(Hess(\varphi))}{12b_0||\varphi|| \cdot ||F(\varphi,\lambda)||} > 1 \tag{3.33}$$

so if

$$\frac{s_n^2(Hess(\varphi))}{12b_0||\varphi||} > ||F(\varphi,\lambda)|| \tag{3.34}$$

To ensure that $P_2(r(\lambda))$ is negative, we can take $r(\lambda)$ to be the vertex of the quadratic $P_2(r)$, which is given by

$$r(\lambda) = \frac{s_n(Hess(\varphi))}{6b_0||\varphi||} \tag{3.35}$$

□

This theorem inspires us to have a condition on the gradient to ensure that there are no other critical points which cross the boundary of the leash.

Condition 3.4.3 (Gradient Condition)**.** *Given a frame set, R_0 and a suitable reference path $\varphi(\lambda)$, we say that φ satisfies the Gradient Condition if $||F(\varphi,\lambda)|| < \frac{s_n(\lambda)^2}{12b_0||\varphi(\lambda)||}$ for all $0 < \lambda < \lambda_1$.*

Definition 3.4.3.1. *We define the fundamental constant $\rho_2 = \frac{s_n(\lambda)^2}{12b_0||\varphi(\lambda)||}$, which shows up in the Gradient Condition.*

The other condition that needs to be satisfied is the Initialization Condition, that the golden retriever path $x(\lambda)$ is initialized within the tube.

Condition 3.4.4 (Initialization Condition)**.** *Given a frame set, R_0, a suitable reference path $\varphi(\lambda)$, and the golden retriever path $x(\lambda)$, we say that φ satisfies the Initialization Condition if $||x(\lambda) - \varphi(\lambda)|| < r(\lambda)$ for some $0 < \lambda < \lambda_1$.*

We now show that leash doesn't contain the origin for any $\lambda < \lambda_1$.

Lemma 3.4.5. *Given a suitable reference path $\varphi(\lambda)$, then $||\varphi(\lambda)|| > r(\lambda)$ if $\lambda < \lambda_1$. Therefore for such a reference path $\varphi(\lambda)$, the radius $r(\lambda)$ is disjoint from the origin for all $\lambda < \lambda_1$.*

Proof. We want to show that $||\varphi(\lambda)|| > r(\lambda) = \frac{s_n(\lambda)}{6b_0||\varphi(\lambda)||}$. This is equivalent to showing that

$$6b_0||\varphi(\lambda)||^2 > s_n(\lambda) \tag{3.36}$$

Now we examine $s_n(\lambda) = \lambda_n(Hess(\varphi, \lambda)) = \lambda_n(3R(\varphi) + \lambda I - R_0)$. Since $\lambda < \lambda_1$, $= \lambda I - R_0$ is of mixed signature, so $\lambda_n(\lambda I - R_0) < 0$. Using Weyl's perturbation theorem ([29]), we see that an upper bound on $\lambda_n(3R(\varphi) + \lambda I - R_0)$ is given by $3||R(\varphi)||$, by thinking of $3R(\varphi)$ as the perturbation. Thus, $s_n(\lambda) \leq 3||R(\varphi)|| \leq 3b_0||\varphi(\lambda)||^2$. Therefore, we have shown that

$$s_n(\lambda) \leq 3b_0||\varphi(\lambda)||^2 < 6b_0||\varphi(\lambda)||^2 \tag{3.37}$$

and so we are done. □

The consequence of these conditions is that if we find a path $\varphi(\lambda)$ which satisfies the Initialization Condition (for sufficiently small λ) and satisfies the Gradient Condition for all $0 < \lambda < \lambda_1$, then the homotopy path $x(\lambda)$ must stay inside the tube, which is disjoint from the origin, and so must end up at $\lambda = 0$. The next lemma says that the distance between critical points at the $\lambda = 0$ is strictly larger than the radius $r(0)$. This ensures that the homotopy path cannot end up at any point other than the global minimizer, thus forcing $x(0) = z$.

Recall that the distance to the nearest critical point is lower bounded by $\rho_c = \frac{2}{3}\sqrt{\frac{\lambda_n(R(z))}{b_0}}$.

Lemma 3.4.6. *$\rho_c \geq r(0)$, in other words, $\frac{2}{3}\sqrt{\frac{\lambda_n(R(z))}{b_0}} \geq \frac{s_n(0)}{6b_0||\varphi(0)||}$*

Proof. Recall that $\varphi(0) = z$ an $s_n(0) = \lambda_n(2R(z))$ in the noiseless case. Thus we want to show

$$\frac{2}{3}\sqrt{\frac{\lambda_n(R(z))}{b_0}} \geq r(0) = \frac{2\lambda_n(R(z))}{6b_0||z||} \tag{3.38}$$

Rewriting this, it is equivalent to requiring

$$\sqrt{\lambda_n(R(z))} \leq 2\sqrt{b_0}||z|| \tag{3.39}$$

which then gives

$$\lambda_n(R(z)) \leq 4b_0||z||^2 \tag{3.40}$$

Since $\lambda_n(R(z)) = ||z||^2\lambda_n(R(\frac{z}{||z||})) \leq b_0||z||^2$, thus we get $\rho_c \geq r(0)$. □

What we have shown is that

Theorem 3.4.7. *If there exists a suitable reference path which satisfies the Initialization Condition and the Gradient Condition, then the Golden Retriever Homotopy Algorithm converges.*

Now we choose a suitable reference path which will satisfy the Initialization Condition. After this, we will work towards finding the probability that it satisfies the Gradient Condition.

Definition 3.4.7.1. *We define a specific suitable reference path, but in terms of a parameter* $\tau = 1 - \frac{\lambda}{\lambda_1}$. *Define* $\varphi_1(\lambda) = \sqrt{\tau}(\tau z + (1-\tau)g)$, *where* z *is the global minimizer and* $g = \sqrt{\frac{\lambda_1}{\langle R(e_1)e_1, e_1\rangle}}e_1$, *and* e_1 *is the normalized eigenvector of* R_0 *associated to* λ_1.

For the path $\varphi_1(\lambda)$, we aim to show that it satisfies the Initialization Condition. We do so by showing that the asymptotic rate as $\tau \to 0$ ($\lambda \to \lambda_1$) of $r(\lambda)$ is bounded below by $\tau^{\frac{1}{2}}$. We then argue that the asymptotic rate of $||x(\lambda) - \varphi_1(\lambda)||$ as $\tau \to 0$ is bounded above in the order of $\tau^{\frac{3}{2}}$. Therefore, for τ sufficiently small, we get $\varphi_1(\lambda)$ satisfies the Initialization Condition.

To establish this, we first establish asymptotics of two quantities that show up often: $||\varphi_1(\lambda)||$ and $s_n(\lambda)$.

Lemma 3.4.8. *There exists a $\tau_0 > 0$ such that $0.9\tau^{\frac{1}{2}}\sqrt{\frac{\lambda_1}{\langle R(e_1)e_1,e_1\rangle}} \leq ||\varphi_1(\lambda)|| \leq 1.1\tau^{\frac{1}{2}}\sqrt{\frac{\lambda_1}{\langle R(e_1)e_1,e_1\rangle}}$ for all $0 < \tau < \tau_0$*

Proof. Recall that $\varphi_1(\lambda) = ||\varphi_1(\lambda_1(1-\tau))||$. This is then, in turn,

$$||\sqrt{\tau}\tau z + \sqrt{\tau}(1-\tau)\sqrt{\frac{\lambda_1}{\langle R(e_1)e_1,e_1\rangle}}e_1|| = ||\tau^{\frac{1}{2}}\sqrt{\frac{\lambda_1}{\langle R(e_1)e_1,e_1\rangle}}e_1 + \tau^{\frac{3}{2}}(z - \sqrt{\frac{\lambda_1}{\langle R(e_1)e_1,e_1\rangle}}e_1)||$$

On one hand this is less than $\tau^{\frac{1}{2}}\sqrt{\frac{\lambda_1}{\langle R(e_1)e_1,e_1\rangle}} + \tau^{\frac{3}{2}}||z - \sqrt{\frac{\lambda_1}{\langle R(e_1)e_1,e_1\rangle}}e_1|| \leq 1.1\tau^{\frac{1}{2}}\sqrt{\frac{\lambda_1}{\langle R(e_1)e_1,e_1\rangle}}$, so long as $0.1\tau^{\frac{1}{2}}\sqrt{\frac{\lambda_1}{\langle R(e_1)e_1,e_1\rangle}} \geq \tau^{\frac{3}{2}}||z - \sqrt{\frac{\lambda_1}{\langle R(e_1)e_1,e_1\rangle}}e_1||$. If z is aligned with e_1, this is always true, otherwise, we see that this is true so long as $\tau \leq \frac{0.1\sqrt{\frac{\lambda_1}{\langle R(e_1)e_1,e_1\rangle}}}{||z - \sqrt{\frac{\lambda_1}{\langle R(e_1)e_1,e_1\rangle}}e_1||}$ On the other hand, $||\varphi_1(\tau)|| \geq \tau^{\frac{1}{2}}\sqrt{\frac{\lambda_1}{\langle R(e_1)e_1,e_1\rangle}} - \tau^{\frac{3}{2}}||z - \sqrt{\frac{\lambda_1}{\langle R(e_1)e_1,e_1\rangle}}e_1|| \geq 0.9\tau^{\frac{1}{2}}\sqrt{\frac{\lambda_1}{\langle R(e_1)e_1,e_1\rangle}}$ so long as $0.1\tau\sqrt{\frac{\lambda_1}{\langle R(e_1)e_1,e_1\rangle}} \geq \tau^{\frac{3}{2}}||z - \sqrt{\frac{\lambda_1}{\langle R(e_1)e_1,e_1\rangle}}e_1||$, which is the same condition as before.

Therefore, for all $0 < \tau < \tau_0 := \frac{0.1\sqrt{\frac{\lambda_1}{\langle R(e_1)e_1,e_1\rangle}}}{||z - \sqrt{\frac{\lambda_1}{\langle R(e_1)e_1,e_1\rangle}}e_1||}$, we have the desired inequalities. □

Now we examine the asymptotics of $s_n(\lambda) = Hess(\varphi_1(\lambda), \lambda)$

Lemma 3.4.9. *There exists a $\tau_1 > 0$ such that $s_n(\lambda) = s_n(\lambda_1(1-\tau)) \geq \lambda_1\tau$ for all $0 < \tau < \tau_1$*

Proof. First note that, if we assume the top eigenvalue of R_0 is distinct (which happens with high probability and the spectrum$(\mathbb{E}(R_0)) = \{3||z||^2, ||z||^2, ..., ||z||^2\}$), we can bound

$$R_0 = \sum_{k=1}^{n} \lambda_k e_k e_k^T \leq \lambda_2 I + (\lambda_1 - \lambda_2) e_1 e_1^T$$

If we examine $Hess(\varphi) = 3R(\varphi) + \lambda I - R_0$, and substitute φ in terms of z and g, we get

$$\begin{aligned} Hess(\varphi_1) &= 3\tau R(\tau z + (1-\tau)g) + \lambda_1(1-\tau)I - R_0 \\ &= 3\tau R(g) + \lambda_1(1-\tau)I - R_0 + O(\tau^2) \\ &\geq (\lambda_1 - \lambda_2)I - (\lambda_1 - \lambda_2)e_1 e_1^T + \tau(3R(g) - \lambda_1 I) - O(\tau^2) \end{aligned}$$

Define $M = (\lambda_1 - \lambda_2)I - (\lambda_1 - \lambda_2)e_1 e_1^T + \tau(3R(g) - \lambda_1 I)$, we will show M is positive definite for τ sufficiently small. First we check $\langle Me_1, e_1\rangle$, and substituting $g = \sqrt{\frac{\lambda_1}{\langle R(e_1)e_1, e_1\rangle}} e_1$ we get

$$\begin{aligned} \langle Me_1, e_1\rangle &= (\lambda_1 - \lambda_2) + (3\frac{\langle R(e_1)e_1, e_1\rangle}{\langle R(e_1)e_1, e_1\rangle} - 1)\tau\lambda_1 - (\lambda_1 - \lambda_2) \\ &= 2\tau\lambda_1 > 0 \end{aligned}$$

Next, we take a direction $x \perp e_1$, $||x|| = 1$

$$\langle Mx, x\rangle = (\lambda_1 - \lambda_2) + (3\frac{\langle R(e_1)x, x\rangle}{\langle R(e_1)e_1, e_1\rangle} - 1)\tau\lambda_1$$

Which for τ sufficiently small is greater than 0, since there is a gap between λ_1 and λ_2. Finally we look at a linear combination of x and e_1, so define $\tilde{x} = cos(\theta)e_1 + sin(\theta)x$, and look at $\langle M\tilde{x}, \tilde{x}\rangle$

$$\langle M\tilde{x}, \tilde{x}\rangle = (\lambda_1 - \lambda_2) + (3\frac{\langle R(e_1)\tilde{x}, \tilde{x}\rangle}{\langle R(e_1)e_1, e_1\rangle} - 1)\tau\lambda_1 - (\lambda_1 - \lambda_2)cos^2(\theta)$$

$$= (\lambda_1 - \lambda_2)sin^2(\theta) + (3cos^2(\theta) - 1 + 6cos(\theta)sin(\theta)\frac{\langle R(e_1)e_1, x\rangle}{\langle R(e_1)e_1, e_1\rangle} + 3sin^2(\theta)\frac{\langle R(e_1)x, x\rangle}{\langle R(e_1)e_1, e_1\rangle})\tau\lambda_1$$

Define $\alpha = \frac{\langle R(e_1)e_1, x\rangle}{\langle R(e_1)e_1, e_1\rangle}$ and $\beta = \frac{\langle R(e_1)x, x\rangle}{\langle R(e_1)e_1, e_1\rangle} > 0$. Additionally, note that $|\alpha| \leq \beta$, since $\alpha = \frac{\langle R(e_1)e_1, x\rangle}{\langle R(e_1)e_1, e_1\rangle} = \frac{\langle R(e_1)^{\frac{1}{2}}e_1, R(e_1)^{\frac{1}{2}}x\rangle}{\langle R(e_1)e_1, e_1\rangle}$ and $|\alpha| = \frac{|\langle R(e_1)^{\frac{1}{2}}e_1, R(e_1)^{\frac{1}{2}}x\rangle|}{\langle R(e_1)e_1, e_1\rangle} \leq \frac{\langle R(e_1)e_1, e_1\rangle\langle R(e_1)x, x\rangle}{\langle R(e_1)e_1, e_1\rangle}$

Therefore, substituting these bounds in, we see that

$$\langle M\tilde{x}, \tilde{x}\rangle = (\lambda_1 - \lambda_2 + 3\beta\tau\lambda_1 - 3\tau\lambda_1)sin^2(\theta) + 3sin(2\theta)\alpha\tau\lambda_1 + 2\tau\lambda_1$$

$$= \frac{1}{2}(\lambda_1 - \lambda_2 + 3(1-\beta)\tau\lambda_1) + 3\alpha\tau\lambda_1 sin(2\theta) - \frac{\lambda_1 - \lambda_2 + 3\beta\tau\lambda_1}{2}cos(\theta) + 2\tau\lambda_1$$

$$\geq \frac{1}{2}(\lambda_1 - \lambda_2 + 3(1-\beta)\tau\lambda_1) + 2\tau\lambda_1 - \sqrt{\frac{1}{4}(\lambda_1 - \lambda_2 - 3(1-\beta)\tau\lambda_1)^2 + (3\alpha\tau\lambda_1)^2}$$

$$\geq \frac{1}{2}(\lambda_1 - \lambda_2 + 3(1-\beta)\tau\lambda_1) + 2\tau\lambda_1 - \frac{1}{2}(\lambda_1 - \lambda_2 + 3(1-\beta)\tau\lambda_1) - \frac{(3\alpha\tau\lambda_1)^2}{(\lambda_1 - \lambda_2 + 3(1-\beta)\tau\lambda_1)}$$

$$= 2\tau\lambda_1 - \frac{(3\alpha\tau\lambda_1)^2}{(\lambda_1 - \lambda_2 + 3(1-\beta)\tau\lambda_1)}$$

Now we see that this implies that $\lambda_n(M) \geq 1.5\lambda_1\tau$ for all small τ. From here, we use Weyl's inequalities ([29]), by taking the hessian to be a perturbation of

M. Thus, if we take $Hess(\varphi_1) = M + R$, we get that since $\lambda_n(R) \leq d\tau^2$, then $\lambda_n(Hess(\varphi_1)) \geq \lambda_n(M) - d\tau^2 \geq 1.5\lambda_1\tau - d\tau^2 \geq \lambda_1\tau$ for all sufficiently small τ. Thus we get there exists a τ_1 such that $\lambda_n(Hess(\varphi_1, \lambda)) = s_n(\lambda) \geq \lambda_1\tau$ for all $0 < \tau < \tau_1$. □

Now we can put these lemmas together and find the asymptotic rate of the radius of the leash $r(\lambda)$.

Lemma 3.4.10. *For all λ sufficiently close to λ_1, $r(\lambda) \geq \frac{\sqrt{\lambda_1\langle R(e_1)e_1,e_1\rangle}}{6.6b_0}\tau^{\frac{1}{2}}$*

Proof. Since $r(\lambda) = \frac{s_n(\lambda)}{6b_0||\varphi_1(\lambda)||}$. Lower bounding $s_n(\lambda)$ by $\lambda_1\tau$ and upper bounding $||\varphi_1(\lambda)||$ by $1.1\sqrt{\frac{\lambda_1}{\langle R(e_1)e_1,e_1\rangle}}\tau^{\frac{1}{2}}$ gives us the result. □

Now that we established an asymptotic lower bound on the radius $r(\lambda)$, we look at the other term in the Intitialization Condition.

Lemma 3.4.11. *There exists a $\tau_2 > 0$ and a positive constant C such that for all $0 < \tau < \tau_2$, $||x(\lambda) - \varphi_1(\lambda)|| \leq C\tau^{\frac{3}{2}}$*

Proof. Let's decompose $x(\lambda) = c\varphi_1(\lambda) + \tau^{\frac{1}{2}}\varphi_1^{\perp}(\lambda)$, such that $\langle\varphi_1, \varphi_1^{\perp}\rangle = 0$.

If we write $\varphi_1(\lambda) = \sqrt{\tau}(\tau z + (1-\tau)g)$, we get the following expression for the

gradient

$$0 = \frac{1}{\tau^{\frac{1}{2}}} F(x(\lambda), \lambda) = \frac{1}{\tau^{\frac{1}{2}}} F(c\varphi_1 + \varphi_1^{\perp}, \lambda) =$$

$$= c^3\tau(1-\tau)^3 R(g)g + (c^3\tau^4 - c\tau)R(z)z + \tau R(\varphi_1^{\perp})\varphi_1^{\perp} +$$

$$3c^3\tau^2(1-\tau)^2 R(g)z + 3c^2\tau(1-\tau)^2 R(g)\varphi_1^{\perp} +$$

$$(3c^3\tau^3(1-\tau) - c\tau(1-\tau))\lambda_1 g + (3c^2\tau^3 - 1)R(z)\varphi_1^{\perp} + 3c\tau(1-\tau)R(\varphi_1^{\perp})g +$$

$$3c\tau^2 R(\varphi_1^{\perp})z + 6c^2\tau^2(1-\tau)R(g,z)\varphi_1^{\perp} +$$

$$c\tau(1-\tau)\lambda_1 z + (1-\tau)\lambda_1\varphi_1^{\perp}$$

Now let $\{g_k\}$ be a basis of eigenvectors of R_0, we normalize them to be $\{e_k\}$, and note that $g = constant \cdot e_1$. Also define $v = \varphi_1^{\perp}/||\varphi_1^{\perp}||$.

Note that taking the inner product of the expression above with e_k, and definining $-\tau T_k$ to be the coefficient of every term that has a coefficient of at least degree τ gives us $n-1$ equations of the form

$$\lambda_1\langle v, e_k\rangle||\varphi_1^{\perp}|| - \lambda_k||\varphi_1^{\perp}||\langle v, e_k\rangle = \tau(M)$$

$$(\lambda_1 - \lambda_k)\langle v, e_k\rangle||\varphi_1^{\perp}|| = \tau(T_k)$$

Now summing the squares over $k = 2, .., n$, we get

$$\sum_{k=2}^{n}(\lambda_1 - \lambda_k)^2\langle v, e_k\rangle^2||\varphi_1^{\perp}||^2 \leq \tau^2 T$$

where $T = \sum_k T_k^2$. We can get a lower bound by using the smallest gap $\lambda_1 - \lambda_1$ to get

$$(\lambda_1 - \lambda_2)^2 ||\varphi_1^\perp||^2 \sum_{k=2}^{n} \langle v, e_k \rangle^2 \leq \tau^2 T$$

Summing over the whole range gives us

$$(\lambda_1 - \lambda_2)^2 (1 - \langle v, e_1 \rangle^2) ||\varphi_1^\perp||^2 \leq \tau^2 T$$

$$\sqrt{(1 - \langle v, e_1 \rangle)^2} ||\varphi_1^\perp|| \leq \tau (\frac{\sqrt{T}}{(\lambda_1 - \lambda_2)})$$

We will briefly examine $\langle v, e_1 \rangle$. Note that

$$0 = \langle v, \varphi_1 \rangle = \sqrt{\tau}\tau \langle v, z \rangle + \sqrt{\tau}(1 - \tau)\langle v, g \rangle$$

$$= \sqrt{\tau}\tau \langle v, z \rangle + \sqrt{\tau}(1 - \tau)||g||\langle v, e_1 \rangle$$

So simplifying, we get

$$\langle v, e_1 \rangle^2 = \frac{\tau^2 \langle z, v \rangle^2}{(1 - \tau)^2 ||g||^2} \leq \frac{\tau^2 ||z||^2}{(1 - \tau)^2 ||g||^2}$$

Substituting this back in, we get

$$||\varphi_1^\perp|| \sqrt{1 - \frac{\tau^2 ||z||^2}{(1 - \tau^2)||g||^2}} \leq \tau (\frac{\sqrt{T}}{(\lambda_1 - \lambda_2)})$$

$$||\varphi_1^\perp|| \leq C_1 \tau \text{ for all } \tau \text{ sufficiently small}$$

Now going back to the equation, if we take the inner product with g itself (and use the fact that $||\varphi_1^{\perp}|| \leq C_1\tau$), if we look at all terms with a coefficient of τ or less, and collect all terms with a coefficient of τ^2 or more and label it $-\tau^2 N$, we get

$$c^3\tau\langle R(g)g, g\rangle - c\tau\lambda_1\langle g, g\rangle = \tau^2 N$$

$$c^3\langle R(g)g, g\rangle - c\lambda_1\langle g, g\rangle = \tau N$$

If we substitute in $g = \sqrt{\frac{\lambda_1}{\langle R(e_1)e_1, e_1\rangle)}}e_1$, we get

$$c^3\frac{\lambda_1^2}{\langle R(e_1)e_1, e_1\rangle} - c\frac{\lambda_1^2}{\langle R(e_1)e_1, e_1\rangle} = \tau N$$

$$(c^3 - c) = \tau(N\frac{\langle R(e_1)e_1, e_1\rangle}{\lambda_1})$$

$$(c-1)(c+1)c = \tau(N\frac{\langle R(e_1)e_1, e_1\rangle}{\lambda_1})$$

These three paths, $c = 0, 1, -1$ each correspond to a different path of the solution. For $c = 0$, this corresponds to staying at $x = 0$, and the $c = \pm 1$ correspond to the inability to distinguish phase between the paths. By analyticity of the roots, we get that on the path corresponding to $c = 1$, $|c - 1| = O(\tau)$.

Putting these results together, we get that $x(\lambda) = c\varphi_1 + \tau^{\frac{1}{2}}\varphi_1^{\perp}$, therefore

$$||x(\lambda) - \varphi_1(\lambda)|| = ||(c-1)\varphi_1 + \tau^{\frac{1}{2}}\varphi_1^{\perp}||$$

$$\leq |(c-1)| \cdot ||\varphi_1|| + \tau^{\frac{1}{2}}||\varphi_1^{\perp}|| \leq C\tau^{\frac{3}{2}}(\text{ for all } \tau \text{ sufficiently small})$$

□

The consequences of the above lemma are immediate.

Theorem 3.4.12. *For all $\tau > 0$ sufficiently small, $||x(\lambda) - \varphi_1(\lambda)|| < r(\lambda)$, i.e. $\varphi_1(\lambda)$ satisfies the Initialization Condition.*

Proof. This is just looking at the order, $r(\lambda)$ stays above something of order $\tau^{\frac{1}{2}}$ as $\tau \to 0$ while $||x(\lambda) - \varphi_1(\lambda)|| \leq C\tau^{\frac{3}{2}}$. Thus for all sufficiently small τ, we get $||x(\lambda) - \varphi_1(\lambda)|| < r(\lambda)$ □

Now we have shown that $\varphi_1(\lambda)$ satisfies the Initialization Condition, we know that if it satisfies the Gradient Condition, i.e. if $||F(\varphi_1, \lambda)|| < \frac{s_n(\lambda)^2}{12b_0||\varphi_1(\lambda)||}$ for all $0 < \lambda < \lambda_1$, then the algorithm converges to the global minimizer.

Our next goal is to understand when $\varphi_1(\lambda)$ satisfies the Gradient Condition. We study this probabilistically. The main idea is to realize that in the expected system, g aligns with z exactly, so if we treat g as a perturbation of z, then we can rewrite the Gradient Condition as a condition on the perturbation. Then we show that for sufficiently high m, the size of the perturbation decreases, and the Gradient Condition is true with high probability.

Thus we define the perturbation $p = g - z$. We first rewrite the gradient $F(\varphi_1, \lambda)$ in terms of p (and $\tau = \lambda_1(1 - \frac{\lambda}{\lambda_1})$).

In this part, we will make the following assumptions:

- $b_0 > 2$
- $\lambda_n(R(z)) \geq \frac{1}{2}||z||^2$

Later we will see that these hold with high probability.

Lemma 3.4.13. *The gradient for $\varphi_1(\lambda)$ can be written as follows*

$$F(\varphi_1(\lambda),\lambda) = \tau^{\frac{3}{2}}\Big(\tau^3\big(-R(p)p\big) + \tau^2\big(3R(p)z + 3R(p)p\big)$$
$$+\tau\big(\lambda_1 p - 3R(z)p - 6R(p)z - 3R(p)p\big)$$
$$+\big(-\lambda_1 p + 3R(z)p + 3R(p)z + R(p)p\big)\Big)$$

Proof. First note that $\varphi_1 = \sqrt{\tau}(\tau z + (1-\tau)g)$. For $g = z + p$, we get $\varphi_1 = \sqrt{\tau}(\tau z + (1-\tau)z + (1-\tau)p) = \sqrt{\tau}(z + (1-\tau)p)$.

Now if we look at

$$F(\varphi_1,\lambda) = \sqrt{\tau}(R(\sqrt{\tau}(z + (1-\tau)p)) + \lambda I - R_0)(z + (1-\tau)p)$$

We can simplify this expression (with a bit of careful bookkeeping) to get

$$F(\varphi_1,\lambda) = \sqrt{\tau}\Big(\tau^4\big(-R(p)p\big) + \tau^3\big(3R(p)z + 3R(p)p\big)$$
$$+\tau^2\big(\lambda_1 p - 3R(z)p - 6R(p)z - 3R(p)p\big)$$
$$+\tau\big(-\lambda_1 p + 3R(z)p + 3R(p)z + R(p)p\big)\Big)$$

Factoring out the extra term of τ, we get our result. □

Now we can find an upper bound on the norm of this gradient.

Lemma 3.4.14. $||F(\varphi_1,\lambda)|| < \tau^{\frac{3}{2}}(4b_0||z||^2||p|| + 3b_0||p||^2||z|| + b_0||p||^3)$

Proof. Examining the terms in $||p||$ of the gradient, we see that $F(\varphi_1,\lambda) = \tau^{\frac{3}{2}}\big((\tau -$

$1)\lambda_1 p + 3(1-\tau)R(z)p + +3(1-2\tau+\tau^2)R(p)z + (1-3\tau+3\tau^2-\tau^3)R(p)p$, so we get that

$$||F(\varphi_1,\lambda)|| \leq \tau^{\frac{3}{2}}\big(\lambda_1||p|| + 3b_0||z||^2||p|| + 3b_0||p||^2||z|| + b_0||p||^3\big)$$

Now we use the fact that $\lambda_1 = \lambda_1(R(z)) = ||R(z)|| = b_0||z||^2$ to say that

$$||F(\varphi_1,\lambda)|| \leq \tau^{\frac{3}{2}}\big(4b_0||z||^2||p|| + 3b_0||p||^2||z|| + b_0||p||^3)\big) \tag{3.41}$$

□

Now that we have bounded $||F(\varphi_1,\lambda)||$ from above, we bound $\rho_2(\lambda) = \frac{s_n(\lambda)^2}{12b_0||\varphi_1(\lambda)||}$ from below, to get a sufficient condition for satisfying the Gradient Condition.

Lemma 3.4.15. *If $||p|| < \frac{1}{5b_0}||z||$, we have $\rho_2(\lambda) \geq \tau^{\frac{3}{2}}\frac{\lambda_n^2(R(z))}{14b_0||z||}$*

Proof. We start with a lower bound on the hessian.

$$Hess(\varphi_1) = 3\tau R(z+(1-\tau)p) + \lambda_1(1-\tau)I - R(z)$$

$$\geq (\tau-1)R(z) + 2\tau R(z) + \lambda_1(1-\tau)I - \tau O(||p||)$$

$$\geq (\tau-1)\lambda_1 I + 2\tau R(z) + \lambda_1(1-\tau)I - \tau O(||p||)$$

$$= 2\tau R(z) - \tau O(||p||)$$

Therefore, we know that $s_n(\lambda) \geq 2\tau\lambda_n - \tau O(||p||)$. Thus for $||p||$ sufficiently small (one can check that it is satisfied if $||p|| \leq \frac{1}{6b_0}||z||$), we get $s_n(\lambda) \geq \tau\lambda_n$. Also $||\varphi_1|| \leq \sqrt{\tau}||z+(1-\tau)p|| \leq \sqrt{\tau}(||z|| + (1-\tau)||p||)$. Since $0 < \tau < 1$, we get

$||\varphi_1|| \leq \sqrt{\tau}||z|| + ||p|| < \frac{7}{6}||z||$, so long as $||p|| < \frac{1}{6}||z||$

So $\rho_2(\lambda) = \frac{s_n(\lambda)^2}{12b_0||\varphi||}$ gives us

$$\begin{aligned}\rho_2(\lambda) = \frac{s_n^2}{12b_0||\varphi||} &> \frac{\tau^2\lambda_n^2}{\sqrt{\tau}14b_0||z||} \\ &= \tau^{\frac{3}{2}}\frac{\lambda_n^2}{14b_0||z||}\end{aligned}$$

For $||p|| < \frac{1}{5b_0}||z||$.

□

From the previous two lemmas, we see that a sufficient condition for satisfying the Gradient Condition ($\lambda_n > 0.5||z||^2$) is for $\tau^{\frac{3}{2}}\big(4b_0||z||^2||p|| + 3b_0||p||^2||z|| + b_0||p||^3\big)\big) < \tau^{\frac{3}{2}}\frac{||z||^4}{56b_0||z||}$, which is satisfied if

$$4b_0^2\frac{||p||}{||z||} + 3b_0^2\frac{||p||^2}{||z||^2} + b_0^2\frac{||p||^3}{||z||^3} \leq \frac{1}{56}$$

This is satisfied if

$$||p|| \leq \min(\frac{1}{6}, \frac{1}{448b_0^2})||z|| \tag{3.42}$$

Since $\frac{1}{448b_0^2} < \frac{1}{6}$ and $\frac{1}{448b_0^2} < \frac{1}{5b_0}$, under our assumptions, thus it is satisfied if

$$||p|| \leq \frac{1}{448b_0^2}||z|| \tag{3.43}$$

We define $r_{crit} = \frac{1}{448b_0^2}||z||$, we get that if our assumptions are true and $||p|| < r_{crit}$, then the Gradient Condition is satisfied and the algorithm converges.

Now we want to use the difference in $||R(z) - \mathbb{E}(R(z))||$ to get an upper bound

for $||p||$. We work on this in two steps. Since $p = g - z$, we first do an estimate on $||e_1 - z_0||$, where $e_1 = \frac{g}{||g||}$ and $z_0 = \frac{z}{||z||}$. Then we work with the normalization terms.

Theorem 3.4.16. $||e_1 - z_0|| \leq \frac{2^{\frac{3}{2}}||R_0 - \mathbb{E}(R_0)||_{op}}{2||z||^2}$

This is a consequence of the famous Davis–Kahan $sin(\Theta)$ theorem. A proof of it can be found in [30].

Theorem 3.4.17. $||p|| \leq ||z|| \cdot ||e_1 - z_0||(\frac{b_0}{\langle R(e_1)e_1, e_1\rangle} + 1)$

Proof. Define $g' = ||z||e_1$. Then

$$||p|| = ||g - z|| = ||g - g' + g' - z||$$
$$\leq ||g - g'|| + ||g' - z||$$

Now we want to estimate each of these terms. The term $||g' - z|| = ||z|| \cdot ||e_1 - z_0||$.

For the term $||g - g'|| = \big| \sqrt{\frac{\lambda_1}{\langle R(e_1)e_1, e_1\rangle}} - ||z|| \big|$. We write $\lambda_1 = \langle R(z)e_1, e_1\rangle$ and examine the fraction

$$\frac{\langle R(z)e_1, e_1\rangle}{\langle R(e_1)e_1, e_1\rangle}$$
$$= ||z||^2 \frac{\langle (R(z_0) - R(e_1))e_1, e_1\rangle + \langle R(e_1)e_1, e_1\rangle}{\langle R(e_1)e_1, e_1\rangle}$$
$$= ||z||^2 \big(\frac{\langle (R(z_0) - R(e_1))e_1, e_1\rangle}{\langle R(e_1)e_1, e_1\rangle} + 1 \big)$$
$$\leq ||z||^2 \big(\frac{||R(z_0) - R(e_1)||}{\langle R(e_1)e_1, e_1\rangle} + 1 \big)$$

Substituting this back into the expression, and using the fact that $\sqrt{1+\epsilon} < 1 + \frac{\epsilon}{2}$

we see that

$$\begin{aligned}||g - g'|| &\leq ||z|| \cdot \left(\sqrt{\frac{||R(z_0) - R(e_1)||}{\langle R(e_1)e_1, e_1\rangle} + 1} - 1\right) \\ &\leq ||z|| \frac{||R(z_0) - R(e_1)||}{2\langle R(e_1)e_1, e_1\rangle} \\ &\leq ||z|| \frac{||R(z_0) - R(e_1)||}{2\langle R(e_1)e_1, e_1\rangle}\end{aligned}$$

We know that $||R(z_0) - R(e_1)|| \leq 2b_0||e_1 - z_0||$, so we see that

$$||g - g'|| \leq \frac{||z|| \cdot b_0 \cdot ||e_1 - z_0||}{\langle R(e_1)e_1, e_1\rangle}$$

Putting it together, we see that

$$\begin{aligned}||p|| &\leq ||g - g'|| + ||g' - z|| \\ &\leq \frac{||z||b_0||e_1 - z_0||}{\langle R(e_1)e_1, e_1\rangle} + ||z|| \cdot ||e_1 - z_0|| \\ &= ||z|| \cdot ||e_1 - z_0||(\frac{b_0}{\langle R(e_1)e_1, e_1\rangle} + 1)\end{aligned}$$

□

Assume that $\lambda_n(R(e_1)) > 0.5$. We can now simplify the argument in the previous theorem by saying it is sufficient if $|z|| \cdot ||e_1 - z_0||(\frac{b_0}{\langle R(e_1)e_1, e_1\rangle} + 1) \leq \frac{1}{448b_0^2}||z||$, which is equivalent to

$$||e_1 - z_0||(\frac{b_0}{\langle R(e_1)e_1, e_1\rangle} + 1) \leq \frac{1}{448b_0^2}$$

If we assume $\lambda_n(R(z_0)) \geq 0.5$, then we know that since

$$||R(z_0) - R(e_1)|| \leq 2b_0||e_1 - z_0||$$

if $||e_1 - z_0|| \leq \frac{0.1}{2b_0}$, then $||R(z_0) - R(e_1)|| \leq 0.1$, so $\lambda_n(R(e_1)) \geq 0.4$. Thus under this assumption, we get a sufficient condition for the Gradient Condition is

$$||e_1 - z_0|| \leq \min\{\frac{1}{20b_0}, \frac{1}{448b_0^2}\frac{0.4}{b_0 + 0.4}\} \tag{3.44}$$

Since under our assumptions, $b_0 > \lambda_1(R(z_0) > 2.9$, we get that being less than the second term always implies being less than the first, so the sufficient condition can be written as

$$||e_1 - z_0|| \leq \frac{1}{448b_0^2}\frac{0.4}{b_0 + 0.4} \tag{3.45}$$

To get some control on the smallest eigenvalue, we use the concentration of the $R(e)$ about its mean to estimate $\lambda_n(R(z))$.

Lemma 3.4.18. *Let $C(\delta)$ be an upper bound and γ be a universal bound as defined in the Concentration Lemma. Then for $m \geq C(0.1)nlog(n)$, $\lambda_n(R(z)) \geq 0.9||z||^2$ with probability $1 - \frac{4}{n^2} - 5e^{-\gamma n}$*

Proof. By the concentration of expectation (Theorem 3.2.7), there exists a $C > 0$ such for $m \geq Cnlog(n)$, $||R(e) - \mathbb{E}[R(e)]|| \leq 0.1$. Since $\lambda_1(\mathbb{E}[R(e)]) = 3$, we get that $\lambda_1(R(e)) \leq 3 + 0.1 = 3.1$. Similarly, $\lambda_n(\mathbb{E}[R(e)]) = 1$, so $\lambda_n(R(z)) = ||z||^2\lambda_n(R(\frac{z}{||z||})) \geq ||z||^2(1 - 0.1) = 0.9||z||^2$ □

Putting this together with what we had before, we get the following theorem

Theorem 3.4.19. *Assume that we are in the noiseless case with the frame vectors drawn from a standard normal. Fix a nonzero $z \in \mathbb{R}^n$ to be the generating signal. Let $\gamma > log(9)$ be a universal constant. Then there exists an upper bound C sufficiently large (but independent of n) such that if $m \geq \max\{Cnlog(n), 64n^3\}$, then the golden retriever algorithm converges with probability $1 - \frac{4}{n^2} - 5e^{-\gamma n} - \left(n^3 + 1\right)e^{-\frac{3n}{10}}$.*

Proof. Define $\delta = \min\{0.1, \frac{1}{2240b_0^3}\}$, as specified above. We will take $C = C(\delta, \gamma)$ from the concentration theorem (Theorem 3.2.7)

Now we note that that the assumptions are true if $||R(z_0) - \mathbb{E}R(z_0)|| \leq 0.1$, and thus under this assumption, a sufficient condition for the Gradient Condition is if

$$2||R(z_0) - \mathbb{E}[R(z_0)]|| \leq \frac{1}{448b_0^2}(\frac{0.4}{b_0 + 0.4}) \leq \frac{2}{4480b_0^3} \tag{3.46}$$

It is possible to give bounds on b_0, specifically we can show that if $m \geq 64n^3$, then $b_0 < 64$ with probability $1 - \left(n^3 + 1\right)e^{-\frac{3n}{10}}$ (see lemma A.4.1).

□

3.5 Following the Retriever: Real Certifier

In this section, we want to provide theoretical guarantees that we are staying on the same path after an Euler step and correction. This analysis will provide a certificate which can be numerically verified to ensure one is following the correct path. This can be used in numerical applications to determine a step size, but the

size of the requirement does not make it practical to do this. However, it can still prove useful in certain debugging scenarios.

The idea behind the proof is to look at a cross section with one of the coordinate directions and find an upper bound for how far the distance the path can go in a single step, and then make sure there is no other critical point that is within the upper bound's distance.

As in Corollary 3.1.7, let $b_0 = \max_{||e||=1} \langle R(e)e, e\rangle \leq b_1 = B(\max_k ||f_k||^2)$, where B is the frame bound.

Here we recall that b_0 is a $||\cdot||_{2\to 4}$ matrix norm, (norm of the frame analysis operator acting from $(\mathbb{R}^n, ||\cdot||_2) \to (\mathbb{R}^m, ||\cdot||_4)$) which is NP-hard to compute in general (see [31]), so we may use the upper bound b_1 in place of b_0 in all numerical computations and the results still apply.

Now let $Hess_{ext}(x, \lambda)$ denote the $n \times (n+1)$ extended hessian, and let c to be the index of the maximum absolute value component of its null vector computed at the next point of the Golden Retriever algorithm (usually, this will mean we are marching along the $c'th$ column). Let q denote the $c'th$ column of the extended hessian. Furthermore, let $Hess_{red:c}$ denote the $n \times n$ reduced hessian, the hessian without column c. We start at a coordinate $X_0 = \begin{bmatrix} x_0 \\ \\ \lambda_0 \end{bmatrix}$ and we parameterize by the distance moved along the c'th entry. Notationally, whenever a quantity is computed at the original point, it will use the notation $(\cdot)_{,0}$. So the algorithm moves to a new point $X(t) = \begin{bmatrix} x(t) \\ \\ \lambda(t) \end{bmatrix}$, and it satisfies in the difference of the c'th coordinates $X_c(t) - X_{c,0} = t$.

Let $D(t) = ||X(t) - X_0||$ and q denote column c of the hessian (the column being removed). Note that

$$D\frac{dD}{dt} = \frac{1}{2}\frac{d}{dt}D^2 = \langle \frac{dX}{dt}, X(t) - X_0 \rangle \tag{3.47}$$

By examining the components of $\frac{dX}{dt}$, we know that $(\frac{dX}{dt})_c = 1$, thus breaking off that component we see from the equation $Hess_{ext}\frac{dX}{dt} = 0$ we get

$$Hess_{red:c}\frac{dX_{red:c}}{dt} + 1 \cdot q = 0 \tag{3.48}$$

Therefore we have $\frac{dX_{red:c}}{dt} = -Hess_{red:c}q$, since we chose c in a way that allows us to assume $Hess_{red:c}$ is invertible. (Our general assumption is the $Hess_{ext}$ is always full rank).

Then we have the following lemmas.

Lemma 3.5.1. $|\frac{dD}{dt}| \leq \frac{||q||}{s_n(Hess_{red:c})} + 1$

Proof.

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}D^2 = \langle \frac{dX}{dt}, X - X_0 \rangle &= \langle \frac{dX_{red:c}}{dt}, X_{red:c} - X_{red:c,0} \rangle + t \\ &= -\langle Hess_{red:c}^{-1}q, X_{red:c} - X_{red:c,0} \rangle + t \\ D|\frac{dD}{dt}| = |\langle Hess_{red}^{-1}q, X_{red} - X_{red,0} \rangle + t| &\leq ||Hess_{red:c}^{-1}q|| \cdot ||X_{red:c} - X_{red:c,0}|| + |t| \\ &\leq ||Hess_{red:c}^{-1}q||D + D \leq \frac{1}{s_n(Hess_{red:c})}||q||D + D\end{aligned}$$

Dividing by D gives us the desired result. □

Lemma 3.5.2. *Assume* $||Hess_{ext} - Hess_{ext,0}|| \leq \frac{s_n(Hess_{red:c,0})}{2}$. *Then* $D(t) \leq (2 + 2\frac{||Hess_{ext,0}||}{s_n(Hess_{red,0})})t$

Proof. First, we note that

$$||q|| \leq ||Hess_{ext}||_{op} \leq ||Hess_{ext} - Hess_{ext,0}|| + ||Hess_{ext,0}|| \tag{3.49}$$

Since by assumption, $||Hess_{ext} - Hess_{ext,0}|| \leq \frac{s_n(Hess_{red:c,0})}{2}$, we get

$$||q|| \leq \frac{s_n(Hess_{red:c,0})}{2} + ||Hess_{ext,0}|| \tag{3.50}$$

Also, by Weyl's inequalities ([29]), we know that

$$s_n(Hess_{red:c}) \geq s_n(Hess_{red:c,0}) - ||Hess_{ext} - Hess_{ext,0}|| \geq \frac{1}{2}s_n(Hess_{red:c,0}) \tag{3.51}$$

(because adding a row can only increase the norm, and by using the assumption) so it follows that

$$\begin{aligned} |\frac{dD}{dt}| \leq \frac{||q||}{s_{min(Hess_{red:c})}} + 1 &\leq \frac{||q||}{s_n(Hess_{red:c,0}) - ||Hess_{ext} - Hess_{ext,0}||} + 1 \\ &\leq \frac{2(\frac{s_n(Hess_{red:c,0})}{2} + ||Hess_{ext,0}||)}{s_n(Hess_{red:c,0})} + 1 \\ &= 2 + 2\frac{||Hess_{ext,0}||}{s_n(Hess_{red:c,0})} \end{aligned}$$

Now if we examine the integral:

$$|\int_0^T \frac{dD}{dt}dt| \leq \int_0^T (2 + 2\frac{||Hess_{ext,0}||}{s_n(Hess_{red:c,0})})dt = (2 + 2\frac{||Hess_{ext,0}||}{s_n(Hess_{red:c,0})})T \tag{3.52}$$

On the other hand, we can evaluate it directly and we get

$$|\int_0^T \frac{dD}{dt}dt| = |\int_0^{D(T)} 1dD| = D(T) \tag{3.53}$$

Therefore, as desired, we get that

$$D(t) \leq (2 + 2\frac{||Hess_{ext,0}||}{s_n(Hess_{red:c,0})})t \tag{3.54}$$

□

Now we want to see for which condition on t will guarantee that $||Hess_{ext} - Hess_{ext,0}|| \leq \frac{s_n(Hess_{red:c,0})}{2}$.

Define the constants

$$A = (2 + 2\frac{||Hess_{ext,0}||}{s_n(Hess_{red:c,0})}) \tag{3.55}$$

and

$$t_+ = -\frac{6b_0||x_0|| + ||Q||}{6Ab_0} + \sqrt{(\frac{6b_0||x_0|| + ||Q||}{6Ab_0})^2 + \frac{s_n(Hess_{red:c,0})}{6A^2b_0}} \tag{3.56}$$

Lemma 3.5.3. *For $t < t_+$, $||Hess_{ext} - Hess_{ext,0}|| \leq \frac{s_n(Hess_{red:c,0})}{2}$ and $D(t) \leq$*

$(2 + 2\frac{||Hess_{ext,0}||}{s_n(Hess_{red:c,0})})t$.

Proof. We want to check the condition that

$$||Hess_{ext} - Hess_{ext,0}|| \leq \frac{s_n(Hess_{red:c,0})}{2} \tag{3.57}$$

Let us begin by finding an upper bound on the left hand side.

$$||Hess_{ext} - Hess_{ext,0}|| \leq ||Hess - Hess_0|| + ||Q|| \cdot ||x - x_0||$$

$$\leq 3b_0(2||x_0|| + D)D + 2D||Q||$$

The last inequality is given by the estimate on the difference of the hessians found in Corollary 3.1.7.

Now, for a sufficient condition, we would want

$$3b_0(2||x_0|| + D)D + D||Q|| \leq \frac{s_n(Hess_{red:c,0})}{2} \tag{3.58}$$

This is a quadratic in terms of D, and solving for D, we get

$$D^2 + (\frac{6b_0||x_0|| + ||Q||}{3b_0})D - \frac{s_n(Hess_{red:c,0})}{6b_0} \leq 0 \tag{3.59}$$

Now if we substitute the bound we want to derive on D, we can find the zeros and will get

$$((2+2\frac{||Hess_{ext,0}||}{s_n(Hess_{red:c,0})})t)^2+(\frac{6b_0||x_0|| + ||Q||}{3b_0})(2+2\frac{||Hess_{ext,0}||}{s_n(Hess_{red:c,0})})t-\frac{s_n(Hess_{red:c,0})}{6b_0} = 0$$

Since $A = (2 + 2\frac{||Hess_{ext,0}||}{s_n(Hess_{red:c,0})})$, we have the equation

$$A^2t^2 + (\frac{6b_0||x_0|| + ||Q||}{3b_0})At - \frac{s_n(Hess_{red:c,0})}{6b_0} = 0 \tag{3.60}$$

After solving, we get $t < t_+$, where t_+ is defined as above is the positive root of this quadratic.

Now, we still need to justify why it is a sufficient to substitute the upperbound for $D(t)$.

We define the following parameter

$$t_2 = \sup\{s : \forall t \in [0, s], D(t) \leq (2 + 2\frac{||Hess_{ext,0}||}{s_n(Hess_{red:c,0})})t, \quad s \leq t_+\}$$

We claim that $t_2 = t_+$.

First note that the set above is nonempty because $s = 0$ satisfies the conditions.

Therefore, by construction, if $t_2 < t_+$, then for all $t \in [0, t_2]$, $D(t) \leq (2 + 2\frac{||Hess_{ext,0}||}{s_n(Hess_{red:c,0})})t$, and for every $t' > t_2$, there exists a $t'' \in (t_2, t')$ such that $D(t'') > (2 + 2\frac{||Hess_{ext,0}||}{s_n(Hess_{red:c,0})})t''$.

We claim that from this we can derive a contradiction, so that $t_2 \geq t_+$. In fact, we claim that there exists a $\epsilon > 0$ such that for all $t \in [t_2, t_2 + \epsilon]$ satisfies $D(t) \leq (2 + 2\frac{||Hess_{ext,0}||}{s_n(Hess_{red:c,0})})t$.

By the existence and uniqueness of the differential equation, we can define the

solution on $[t_2, t_2 + \eta]$ for some $\eta > 0$. Also note that

$$D(t_2) \leq (2 + 2\frac{||Hess_{ext,0}||}{s_n(Hess_{red:c,0})})t_2 < (1 + 2\frac{||Hess_{ext,0}||}{s_n(Hess_{red:c,0})})t_2$$

Therefore we know since

$$|\frac{dD}{dt}| \leq \frac{2||q||}{s_n(Hess_{red,0})} + 1 \leq (2 + 2\frac{||Hess_{ext,0}||}{s_n(Hess_{red:c,0})}) \tag{3.61}$$

is still satisfied at t_2.

Examining this, take a $\epsilon < \eta$, so the solution to the differential equation is defined on $[t_2, t_2 + \epsilon]$.

Therefore, if we take $t \in [t_2, t_2 + \epsilon]$

$$|\int_{t_2}^{t} \frac{dD}{dt}dt| \leq \int_{t_2}^{t_2+\epsilon} (2 + 2\frac{||Hess_{ext,0}||}{s_n(Hess_{red:c,0})})dt = (2 + 2\frac{||Hess_{ext,0}||}{s_n(Hess_{red:c,0})})(t - t_2)$$

On the other hand, we can evaluate it directly and we get

$$|\int_{t_2}^{t} \frac{dD}{dt}dt| = |\int_{D(t_2)}^{D(t)} 1 d\varphi| = D(t) - D(t_2)$$

Therefore, since the upper bound is satisfied at t_2, we find

$$D(t) - D(t_2) \leq (2 + 2\frac{||Hess_{ext,0}||}{s_n(Hess_{red:c,0})})(t - t_2)$$
$$D(t) \leq (2 + 2\frac{||Hess_{ext,0}||}{s_n(Hess_{red:c,0})})(t - t_2) + D(t_2)$$
$$D(t) \leq (2 + 2\frac{||Hess_{ext,0}||}{s_n(Hess_{red:c,0})})t$$

Therefore, for all $t < \epsilon$, the upper bound condition is satisfied, which provides the contradiction that $t_2 < t_+$ can be the maximum (or supremum) of the set, so the maximum must be t_+ itself.

Therefore to summarize, so long as the chain of equalities (3.61) is satisfied, which it is for all $||Hess_{ext} - Hess_{ext,0}||$ or $t < t_+$, we get the upper bound. As soon as $t > t_+$, the chain of inequalities is broken and the upperbound no longer needs to hold true.

□

Now that we have found an upper bound on the distance, we need to make sure there is no other critical point within this distance. The following theorem will be useful in showing this.

Theorem 3.5.4. *Let $X = (a, \lambda_a)$ be a critical point of $J(x, \lambda)$ and let $\rho = \min(\frac{1}{2}, \frac{s_n(Hess_{ext}(a,\lambda_a)))}{\sqrt{n+1}(b_0+3b_0||a||+||Q||)})$.*

Then there is no other critical point X_1 such that $X_1(c) = X(c)$ and $||X_1(c) - X(c)|| \leq \rho$. In other words, ρ serves as a lower bound for a distance to the nearest critical point on the same $X_1(c) = X(c)$ hyperplane.

Proof. Let $e = (b, \lambda_b)$ be a unit vector with $e_c = 0$.

Standard computations show that

$$F((a, \lambda_a) + t(b, \lambda_b)) = t^3 R(b)b + t^2(3R(b)a + \lambda_b Qb) + t(Hess_{ext}(a, \lambda_a) \begin{bmatrix} b \\ \\ \lambda_b \end{bmatrix})$$

Define $v_1 = R(b)b$, $v_2 = 3R(b)a + \lambda_b Qb$, and $v_3 = (Hess_{ext}(a, \lambda_a) \begin{bmatrix} b \\ \\ \lambda_b \end{bmatrix})$

Now define the function

$$G(t) = ||F((a, \lambda_a) + t(b, \lambda_b))|| \tag{3.62}$$

We want to obtain some estimates on the roots from this function.

$$G(t) = ||F((a, \lambda_a) + t(b, \lambda_b))||$$

$$= |t| \cdot ||(t^2 v_1 + t v_2 + v_3)||$$

$$\geq |t| \cdot (||v_3|| - |t|||v_2|| - |t|^2||v_1||)$$

So what we want to do is find the nearest root of

$$(||v_3|| - |t|||v_2|| - |t|^2||v_1||)$$

If the root is farther than $\frac{1}{2}$ (so $t > \frac{1}{2}$), then $\frac{1}{2}$ is a lower bound on the root. Otherwise, the root is closer than $\frac{1}{2}$, so the slope of the function is dominated by

the slope at $\frac{1}{2}$. The slope at $\frac{1}{2}$ is given by $M = -||v_1|| - ||v_2||$.

Now if we go back and estimate the zero of the line passing through $(0, a)$ with slope $M = -||v_1|| - ||v_2||$, we get that the root is at $x_{root} = \frac{||v_3||}{||v_1||+||v_2||}$.

Therefore $\frac{||v_3||}{||v_1||+||v_2||}$ is a bound on the closest root.

So now we want to minimize this bound over all directions e not in the null direction.

First we will minimize $||v_3||$.

$$||v_3||^2 = ||(Hess_{ext}(a, \lambda_a) \begin{bmatrix} b \\ \\ \lambda_b \end{bmatrix})||^2$$

$$= \langle (Hess_{ext}(a, \lambda_a) \begin{bmatrix} b \\ \\ \lambda_b \end{bmatrix}), (Hess_{ext}(a, \lambda_a) \begin{bmatrix} b \\ \\ \lambda_b \end{bmatrix}) \rangle$$

Now let $v = \begin{bmatrix} dx \\ \\ d\lambda \end{bmatrix}$ be the null vector of the extended hessian, and decompose $\begin{bmatrix} b \\ \\ \lambda_b \end{bmatrix} = c_1 v + w$, where $w \in span(v)^{\perp}$. We examine this norm in a little more detail.

$$||v_3||^2 = \langle (Hess_{ext}(a, \lambda_a) \begin{bmatrix} b \\ \\ \lambda_b \end{bmatrix}), (Hess_{ext}(a, \lambda_a) \begin{bmatrix} b \\ \\ \lambda_b \end{bmatrix}) \rangle$$

$$= \langle (Hess_{ext}(a, \lambda_a)w), (Hess_{ext}(a, \lambda_a)w \rangle$$

$$= \langle (Hess_{ext}(a, \lambda_a)^T)(Hess_{ext}(a, \lambda_a)w, w \rangle$$

$$= \lambda_n(Hess_{ext}(a, \lambda_a)^T Hess_{ext}(a, \lambda_a))||w||^2$$

$$\geq s_n^2(Hess_{ext}(a, \lambda_a)) \min_{||e||=1, e_c=0} ||proj_{span(v^\perp)}(b, \lambda_b)||^2$$

Note that the null vector v here is normalized so that $v_c = 1$, as c is chosen so that v_c it is the largest component.

To estimate this, define $\tilde{e}$ to be e without the c'th component, and $\tilde{v}$ as v without the c'th component. Now note we are trying to minimize the projection onto the complement of v, so we get

$$\min_{||e||=1, e_{c=0}} ||e - \frac{\langle e, v\rangle}{||v||^2}v||^2 = ||\tilde{e} - \frac{\langle \tilde{e}, \tilde{v}\rangle}{1+||\tilde{v}||^2}\tilde{v}||^2 + (\frac{\langle \tilde{e}, \tilde{v}\rangle}{1+||\tilde{v}||^2}v_c)^2$$

$$= \langle \tilde{e} - \frac{\langle \tilde{e}, \tilde{v}\rangle}{1+||\tilde{v}||^2}\tilde{v}, \tilde{e} - \frac{\langle \tilde{e}, \tilde{v}\rangle}{1+||\tilde{v}||^2}\tilde{v}\rangle + (\frac{\langle \tilde{e}, \tilde{v}\rangle}{1+||\tilde{v}||^2})^2$$

$$= 1 - 2\frac{\langle \tilde{e}, \tilde{v}\rangle^2}{1+||\tilde{v}||^2} + \frac{\langle \tilde{e}, \tilde{v}\rangle^2}{(1+||\tilde{v}||^2)^2}||\tilde{v}||^2 + (\frac{\langle \tilde{e}, \tilde{v}\rangle}{1+||\tilde{v}||^2})^2$$

$$= 1 + \frac{\langle \tilde{e}, \tilde{v}\rangle^2}{1+||\tilde{v}||^2} - 2\frac{\langle \tilde{e}, \tilde{v}\rangle^2}{1+||\tilde{v}||^2}$$

$$= 1 - \frac{\langle \tilde{e}, \tilde{v}\rangle^2}{1+||\tilde{v}||^2}$$

$$\geq 1 - \frac{||\tilde{v}||^2}{1+||\tilde{v}||^2} \quad \text{by Cauchy Schwarz}$$

$$= \frac{1}{||\tilde{v}||^2+1} \geq \frac{1}{n+1} \quad \text{since 1 is the largest component in a size } n \text{ vector}$$

Therefore, we see that

$$||v_3|| \geq s_n(Hess_{ext}(a, \lambda_a)) \cdot \frac{1}{\sqrt{n+1}} \tag{3.63}$$

Now we want to maximize the denominator, so we want to maximize both $||v_1||$ and $||v_2||$

$$||v_1|| = ||R(b)b|| \leq ||b||^3||R(\frac{b}{||b||})|| \leq ||R(\frac{b}{||b||})|| \leq b_0 \tag{3.64}$$

And for $||v_2||$, we can find a bound similarly, again using that $||b|| \leq 1$

$$||v_2|| = ||3R(b)a + \lambda_b Qb|| \leq 3||R(b)a|| + ||Q|| \tag{3.65}$$

$$\leq 3||R(b)|| \cdot ||a|| + ||Q|| \leq 3b_0||a|| + ||Q|| \tag{3.66}$$

Therefore, the root

$$t_{root} \geq \min\{\frac{1}{2}, \frac{||v_1||}{||v_2|| + ||v_3||}\} \geq \min(\frac{1}{2}, \frac{s_n(Hess_{ext}(a, \lambda_a)))}{\sqrt{n+1}(b_0 + 3b_0||a|| + ||Q||)}) \tag{3.67}$$

□

Theorem 3.5.5. *Assume our algorithm starts at a point (x_{old}, λ_{old}) which is a critical point. Let (x_{new}, λ_{new}) be a new critical point the algorithm decides and $(x_{other}, \lambda_{other})$ be any other critical point in the same coordinate at the index c, as defined above. Let D_1 denote the distance from (x_{old}, λ_{old}) to (x_{new}, λ_{new}). Define $t_1 = \frac{\rho(x_{new}, \lambda_{new})}{2A}$ and $t_{min} = min(t_1, t_+)$*

Assume the following two conditions are satisfied:

1. $D_1 < \frac{\rho(x_{new}, \lambda_{new})}{2}$
2. $t < t_{min}$

Then (x_{new}, λ_{new}) is the point connected on the continuous homotopy path which goes through (x_{old}, λ_{old}).

Proof. Assume $(x_{other}, \lambda_{other})$ is a critical point on the path instead of (x_{new}, λ_{new}). Define $UB(t) = (2 + 2\frac{||Hess_{ext,0}||}{s_n(Hess_{red:c,0})})t$. Then, since $t < t_{min} < t_+$, we have that

$$||(x_{other}, \lambda_{other}) - (x_{old}, \lambda_{old})|| \leq UB(t) \tag{3.68}$$

Since $t < t_1$, we get that $UB(t) < \frac{\rho(x_{new},\lambda_{new})}{2}$, so

$$||(x_{other}, \lambda_{other}) - (x_{old}, \lambda_{old})|| \leq \frac{\rho(x_{new}, \lambda_{new})}{2} \tag{3.69}$$

Also, since $D_1 < \frac{\rho(x_{new},\lambda_{new})}{2}$ we get that

$$\begin{aligned} ||(x_{new}, \lambda_{new}) - (x_{other}, \lambda_{other})|| &\leq ||(x_{new}, \lambda_{new}) - (x_{old}, \lambda_{old})|| + ||(x_{old}, \lambda_{old}) - (x_{other}, \lambda_{other})|| \\ &< \frac{\rho(x_{new}, \lambda_{new})}{2} + \frac{\rho(x_{new}, \lambda_{new})}{2} = \rho(x_{new}, \lambda_{new}) \end{aligned}$$

but this is a contradiction, because any other critical point must be a distance further that ρ away from (x_{new}, λ_{new}).

Therefore, the only possible point on this level set that is the point passing through the continuous path is (x_{new}, λ_{new}). □

3.6 Oracle Convergence

Given an Oracle, we can ask the following question: Does there exist a positive semidefinite matrix Q such that the Golden Retriever algorithm converges to the exact solution?

The answer is yes. In fact, we can make the algorithm converge to any critical point we want.

Lemma 3.6.1. *Let v be a critical point of $J(x, \lambda)$ and let R_0 be the matrix in the condition, there exists a positive definite symmetric matrix $Q \neq R_0$ satisfying the following properties:*

- $Qv = R_0 v$
- *The determinant of the pencil, $det(\lambda Q - R_0)$ has generalized eigenvalues which satisfy: $\lambda \leq 1$*
- *The generalized eigenvalue around $\lambda = 1$ has corresponding eigenvector v, and is distinct*

Proof. The proof is by construction.

Define the following rank one matrix.

$$Q_1 = \frac{R_0 v v^T R_0}{\langle R_0 v, v \rangle} \tag{3.70}$$

Note that $Q_1 v = R_0 v$. Since Q_1 is a rank one matrix, so we want to make it full rank.

Set

$$Q = Q_1 + \mu(I - \frac{vv^t}{||v||^2}) \tag{3.71}$$

Note that we still have $Qv = R_0 v$.

At this point, we constructed a family of symmetric matrices Q such that $Qv = R_0 v$. The only thing left to do is to make 1 be the largest generalized eigenvalue here.

To do so, let $\mu_1 = \lambda_{max}(R_0 - Q)$. The claim is that for any $\mu > \mu_1$, that Q would satisfy the pencil criterion.

To justify this, $Q = Q_1 + \mu Q_0$, what we need is $Q \geq R_0$ (because $\lambda Q - R_0$

would be positive definite for $\lambda \geq 1$)

$$Q = Q_1 + \mu Q_0 \geq R_0$$

$$\mu Q_0 \geq R_0 - Q_1$$

$$\mu I \geq R_0 - Q_1$$

$$1 \geq \frac{1}{\mu}(R_0 - Q_1)$$

The reason the identity appears is because $Q_0 = (I - \frac{vv^t}{||v||^2})$. We already know that for v, the generalized eigenvalue is $\lambda = 1$. We now get Q_0 acts as the identity on the orthogonal complement.

Therefore, for $\mu > \lambda_{max}(R_0 - Q_1)$, we have that $\lambda \leq 1$.

Therefore a Q with the listed properties exists and is constructible.

□

Theorem 3.6.2. *There exists a matrix Q such that if the Golden Retriever is initialized with the given Q, then the algorithm converges to the critical point v.*

Proof. Let Q be as in the previous lemma. we examine the path $\nabla_x J(x, \lambda) = (R(x) + \lambda Q - R_0)x$ where $x = cv$.

$$0 = (R(cv) + \lambda Q - R_0)cv$$

$$0 = c^3 R(v)v + c\lambda Q v - cR_0 v$$

$$0 = c^2 R(v)v + \lambda Q v - R_0 v$$

Now since $R(v)v = R_0 v$ (v is a critical point at $\lambda = 0$), and $Qv = R_0 v$, and $R_0 v \neq 0$ (since R_0 is positive definite) then we have the following:

$$0 = c^2 R(v)v + \lambda Qv - R_0 v$$

$$0 = (c^2 + \lambda - 1)R_0 v$$

$$c^2 = (1 - \lambda)$$

$$c = \sqrt{1 - \lambda}$$

The last line is effectively choosing one of the two equivalent paths. Therefore, if we initialize the algorithm with the given Q matrix, and initialize the direction along the principal eigenvector, v, we have that the algorithm will follow the critical path: $(x(\lambda), \lambda) = \left((\sqrt{1 - \lambda})v, \lambda\right)$ □

The theorem above, when applied to $v = z$, the global minimizer, shows that there exists a Q which guarantees that the algorithm converges. This gives us the following theorem as a corollary.

Theorem 3.6.3. *Let z be the minimizer to the optimization problem in (2.2). There exists a positive definite matrix Q_z such that the Golden Retriever Algorithm, initialized with Q_z, converges to z. Moreover, the trajectory of the homotopy path with Q_z, when projected onto $\lambda = 0$, follows a straight line.*

However, it is worth noting that to construct such a Q, we will need to know z, so Q can only be given by an oracle.

Chapter 4: Complex Case

4.1 Background

We define the following equivalence classes:

$$\widehat{\mathbb{C}^n} = \{\hat{x}, x \in \mathbb{C}^n\} \tag{4.1}$$

$$\hat{x} = \{xe^{i\theta}, \theta \in \mathbb{R}\} \tag{4.2}$$

Now say we are given a frame, $\mathcal{F} = \{f_1, ..., f_m\} \subset \mathcal{C}^n$ Then we can also define the following function:

$$\beta(\hat{x}) = (|\langle x, f_k \rangle|^2)_{k=1}^m \tag{4.3}$$

We want to adjust our results in the real case, and one way to quotient out the phase ambiguity is to do so through realification. Therefore, We want to understand what happens with the beta map once we apply the realification procedure to it.

Definition 4.1.0.1 (Realification and Complexification)**.** *$j : \mathbb{C}^n \to \mathbb{R}^{2n}$ defined by*

$$x \in \mathbb{C}^n \to \xi = \begin{pmatrix} real(x) \\ \\ imag(x) \end{pmatrix} = j(x) \in \mathbb{R}^{2n}$$

is called the realification map.

The map $k : \mathbb{R}^{2n} \to \mathbb{C}^n$ with $k = j^{-1}$ is called the complexification map.

The following are basic properties of realification which are straight forward to check.

Lemma 4.1.1 (Basic Properties of Realification)**.** *Let $x \in \mathbb{C}^n$*

1. *j is a $\mathbb{R}$-linear map*

2. *$||x|| = ||j(x)||$*

In the context of phase retrieval we have

$$x \in \mathbb{C}^n \to \xi = \begin{pmatrix} real(x) \\ \\ imag(x) \end{pmatrix} = j(x) \in \mathbb{R}^{2n}$$

Similarly we have

$$f_k \in \mathbb{C}^n \to \varphi_k = \begin{pmatrix} real(f_k) \\ \\ imag(f_k) \end{pmatrix} = j(f_k) \in \mathbb{R}^{2n}$$

We also define the $(2n \times 2n)$ symplectic matrix $J = \begin{pmatrix} 0 & -I_n \\ \\ I_n & 0 \end{pmatrix}$. This is an im-

portant matrix and plays the role of a higher dimensional complex unit and has many similarities with the number i and the complex plane's identification with $\mathbb{R}^2$. Specifically, $j(ix) = Jj(x)$ and more generally let $U(\theta) = cos(\theta)I_{2n} + isin(\theta)J$, and let $u = e^{i\theta}$, then $j(ux) = U(\theta)j(x)$

If we now examine what is the inner product and expand, we see that

$$\langle x, f_k\rangle = \langle real(x) + i \cdot imag(x), real(f_k) + i \cdot imag(f_k)\rangle$$

$$= \langle real(x), real(f_k)\rangle + i \cdot \langle imag(x), real(f_k)\rangle - i \cdot \langle real(x), imag(f_k)\rangle + \langle imag(x), imag(f_k)\rangle$$

$$= \langle \xi, \varphi_k\rangle + i \cdot \langle \xi, J\varphi_k\rangle$$

Therefore, we have that

$$|\langle x, f_k\rangle|^2 = |\langle \xi, \varphi_k\rangle|^2 + |\langle \xi, J\varphi_k\rangle|^2$$

$$= \xi^T \varphi_k \varphi_k^T \xi + \xi^T J\varphi_k (J\varphi_k)^T \xi^T$$

$$= \xi^T (\varphi_k \varphi_k^T + J\varphi_k \varphi_k^T J^T)\xi$$

Now if we define the matrix $\Phi_k = (\varphi_k \varphi_k^T + J\varphi_k \varphi_k^T J^T)$, we get

$$|\langle x, f_k\rangle|^2 = \xi^T \Phi_k \xi \tag{4.4}$$

Proposition 4.1.2. *With Φ_k defined as above, we have that $rank(\Phi_k) \leq 2$*

Proof. First note that $J^T = -J$. Then we can see that

$$\langle \varphi_k, J\varphi_k \rangle = \langle J^T \varphi_k, \varphi_k \rangle = -\langle J\varphi_k, \varphi_k \rangle = -\langle \varphi_k, J\varphi_k \rangle$$

Therefore we get orthogonality

$$\langle \varphi_k, J\varphi_k \rangle = 0 \tag{4.5}$$

Since φ_k and $J\varphi_k$ are orthogonal and thus linearly independent, we know Φ_k is the sum of two rank-1 matrices with linearly independent components.

Therefore $rank(\Phi_k) = 0$ iff $\varphi_k = 0 (\leftrightarrow f_k = 0)$, and $rank(\Phi_k) = 2$ otherwise.

□

Proposition 4.1.3. *Φ_k commutes with J, so $\Phi_k J = J\Phi_k$.*

Proof.

$$\Phi_k J = (\varphi_k \varphi_k^T + J\varphi_k \varphi_k^T J^T)J = -\varphi_k \varphi_k^T J^T + J\varphi_k \varphi_k^T$$

$$= J(J\varphi_k \varphi_k^T J^T + \varphi_k \varphi_k^T)$$

$$= J\Phi_k$$

□

4.2 Derivation of the golden retriever in the Complex Case

We follow a similar idea as how we started in the real case. We will start by adjusting the criterion through realification to map everything into $\mathbb{R}^{2n}$.

4.2.1 Preliminaries

We want to minimize the criterion

$$\Omega(x, \lambda; \mathcal{F}, Q, y) = \frac{1}{4m} \sum_{k=1}^{m} (y_k - |\langle x, f_k \rangle|^2)^2 + \frac{\lambda}{2} \langle Qx, x \rangle \tag{4.6}$$

Here, $x \in \mathbb{C}^n$, $\lambda \in \mathbb{R}$, $\mathcal{F} \subset \mathbb{C}^n$, $y \in \mathbb{R}^n$, and Q is a positive semidefinite hermitian matrix.

Once again, we expand to get

$$\Omega(x, \lambda) = \frac{1}{4m} \sum_{k=1}^{m} (y_k^2 - 2y_k |\langle x, f_k \rangle|^2 + |\langle x, f_k \rangle|^4) + \frac{\lambda}{2} \langle Qx, x \rangle \tag{4.7}$$

$$= \frac{1}{4m} \sum_{k=1}^{m} y_k^2 + \frac{1}{4m} \sum_{k=1}^{m} |\langle x, f_k \rangle|^4 + \frac{\lambda}{2} \langle Qx, x \rangle - \frac{1}{2m} \sum_{k=1}^{m} y_k |\langle x, f_k \rangle|^2 \tag{4.8}$$

Let ξ be the realification of x, and let φ_k be the realification of f_k. Also let $\Phi_k = (\varphi_k \varphi_k^T + J \varphi_k \varphi_k^T J^T)$, with $J = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}$

We introduce the following definitions

$$\Gamma_0 = \frac{1}{m} \sum_{k=1}^{m} y_k \Phi_k, \quad \widetilde{\Gamma}(\xi) = \frac{1}{m} \sum_{k=1}^{m} \Phi_k \xi \xi^T \Phi_k, \quad \Gamma(\xi) = \frac{1}{m} \sum_{k=1}^{m} \xi^T \Phi_k \xi \Phi_k \tag{4.9}$$

Then we can rewrite the criterion with the following

Proposition 4.2.1. *With the above notation, we can simplify the criterion into the following form*

$$\Omega(x,\lambda) = \Omega(\xi,\lambda) = \frac{1}{4}\langle\widetilde{\Gamma}(\xi)\xi,\xi\rangle + \frac{1}{2}\langle(\lambda S - \Gamma_0)\xi,\xi\rangle + \frac{1}{4m}\sum_{k=1}^{m} y_k^2$$

$$= \frac{1}{4}\langle\Gamma(\xi)\xi,\xi\rangle + \frac{1}{2}\langle(\lambda S - \Gamma_0)\xi,\xi\rangle + \frac{1}{4m}\sum_{k=1}^{m} y_k^2 \tag{4.10}$$

For the symmetric matrix $S = \begin{bmatrix} Q_R & -Q_I \\ Q_I & Q_R \end{bmatrix}$ *Where* $Q_R = real(Q)$ *and* $Q_I = imag(Q)$

Proof. First note that $\Gamma(\xi)\xi = \widetilde{\Gamma}(\xi)\xi$, so the two expressions above are equivalent. Therefore, it suffices to prove the first one.

Let us examine the term $\frac{1}{4m}\sum_{k=1}^{m}|\langle x, f_k\rangle|^4$. By (4.4), we know that

$$\begin{aligned}\frac{1}{4m}\sum_{k=1}^{m}|\langle x, f_k\rangle|^4 &= \frac{1}{4m}\sum_{k=1}^{m}(\xi^T\Phi_k\xi)^2\\ &= \frac{1}{4m}\sum_{k=1}^{m}\xi^T\Phi_k\xi\xi^T\Phi_k\xi\\ &= \frac{1}{4m}\sum_{k=1}^{m}\xi^T(\Phi_k\xi\xi^T\Phi_k)\xi\\ &= \frac{1}{4m}\sum_{k=1}^{m}\langle(\Phi_k\xi\xi^T\Phi_k)\xi,\xi\rangle\\ &= \frac{1}{4}\langle(\frac{1}{m}\sum_{k=1}^{m}\Phi_k\xi\xi^T\Phi_k)\xi,\xi\rangle\\ &= \frac{1}{4}\langle\widetilde{\Gamma}(\xi)\xi,\xi\rangle\end{aligned}$$

Similarly, it also follows from (4.4) that

$$\begin{aligned}-\frac{1}{2m}\sum_{k=1}^{m}y_k|\langle x, f_k\rangle|^2 &= -\frac{1}{2,}\sum_{k=1}^{m}y_k\xi^T\Phi_k\xi\\ &= -\frac{1}{2m}\sum_{k=1}^{m}\xi^Ty_k\Phi_k\xi\\ &= -\frac{1}{2}\xi^T(\frac{1}{m}\sum_{k=1}^{m}y_k\Phi_k)\xi\\ &= -\frac{1}{2}\xi^T\Gamma_0\xi\\ &= -\frac{1}{2}\langle\Gamma_0\xi,\xi\rangle\end{aligned}$$

Now we just need to examine what happens to $\langle Qx, x\rangle$ in terms of ξ.

To do this, note that since Q is hermitian, $\langle Qx, x\rangle$ is real (and positive, since Q is positive definite).

Now let $x = a + ib$, $Q = Q_R + iQ_I$.

Then we can compute

$$\langle Qx, x\rangle = x^* Qx$$

$$= (a^T - ib^T)(Q_R + iQ_I)(a + ib)$$

$$= a^T Q_R a - a^T Q_I b + b^T Q_R b + b^T Q_I a$$

The last line is because we know the inner product must be real, so we only need to keep track of the real terms.

Now if we examine the following

$$\begin{bmatrix} a^T & b^T \end{bmatrix} \begin{bmatrix} Q_R & 0 \\ 0 & Q_R \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a^T & b^T \end{bmatrix} \begin{bmatrix} Q_R a & 0 \\ 0 & Q_R b \end{bmatrix}$$

$$= a^T Q_R a + b^T Q_R b$$

Similarly, it is easy to see that

$$\begin{bmatrix} a^T & b^T \end{bmatrix} \begin{bmatrix} 0 & -Q_I \\ Q_I & 0 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = -a^T Q_I b + b^T Q_R a$$

So therefore, if we define the matrix

$$S = \begin{bmatrix} Q_R & 0 \\ 0 & Q_R \end{bmatrix} + \begin{bmatrix} 0 & -Q_I \\ Q_I & 0 \end{bmatrix} = \begin{bmatrix} Q_R & 0 \\ 0 & Q_R \end{bmatrix} + \begin{bmatrix} Q_I & 0 \\ 0 & Q_I \end{bmatrix} J$$
$$= \begin{bmatrix} Q_R & -Q_I \\ Q_I & Q_R \end{bmatrix}$$

If we write $\xi = \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} real(x) \\ imag(x) \end{bmatrix}$, then by construction, we get $\langle Qx, x\rangle = \langle S\xi, \xi\rangle$

It is also the true Q being Hermitian implies S is symmetric since for a Hermitian matrix Q, $Q_I^T = -Q_I$ and $Q_R^T = Q_R$ (since $Q_R + iQ_I = Q = Q^* = Q_R^T - iQ_I^T$, and then equating real and imaginary parts).

□

Proposition 4.2.2. $\nabla_\xi \Omega(x, \lambda) = (\Gamma(\xi) + \lambda S - \Gamma_0)\xi$

Proof. Note that $\nabla_\xi[\frac{1}{2}\langle(\lambda S - \Gamma_0)\xi, \xi\rangle] = (\lambda S - \Gamma_0)\xi$, so all we have to compute is $\nabla_\xi[\frac{1}{4}\langle\widetilde{\Gamma}(\xi)\xi, \xi\rangle]$.

Note that

$$\begin{aligned}
\frac{1}{4}\nabla_\xi\langle\widetilde{\Gamma}(\xi)\xi,\xi\rangle &= \frac{1}{4}\langle\frac{1}{m}\sum_{k=1}^{m}\Phi_k\xi\xi^T\Phi_k\xi,\xi\rangle \\
&= \frac{1}{4}\nabla_\xi\frac{1}{m}\sum_{k=1}^{m}\xi^T\Phi_k\xi\xi^T\Phi_k\xi \\
&= \frac{1}{4m}\sum_{k=1}^{m}\nabla_\xi(\xi^T\Phi_k\xi)^2 \\
&= \frac{1}{4m}\sum_{k=1}^{m}2\xi^T\Phi_k\xi\cdot\nabla_\xi\langle\Phi_k\xi,\xi\rangle \\
&= \frac{1}{4m}\sum_{k=1}^{m}2\xi^T\Phi_k\xi\cdot 2\Phi_k\xi \\
&= \frac{1}{m}\sum_{k=1}^{m}\xi^T\Phi_k\xi\Phi_k\xi \\
&= \frac{1}{m}\sum_{k=1}^{m}\Phi_k\xi\xi^T\Phi_k\xi \\
&= \widetilde{\Gamma}(\xi)\xi
\end{aligned}$$

Now since $\widetilde{\Gamma}(\xi)\xi = \Gamma(\xi)\xi$, we get our result. □

Proposition 4.2.3. $Hess_\xi(\Omega(\xi,\lambda)) = 2\widetilde{\Gamma}(\xi) + \Gamma(\xi) + \lambda S - \Gamma_0$

Proof. We are looking at the following computation (and simplify things by using Lemma A.1.5 from the Appendix)

$$\begin{aligned}
&\nabla_\xi[\xi^T\Phi_k\xi\Phi_k\xi] \\
&= \Phi_k\xi\otimes\nabla_\xi\xi^T\Phi_k\xi + \xi^T\Phi_k\xi\nabla_\xi\Phi_k\xi \\
&= \Phi_k\xi\cdot(2\Phi_k\xi)^T + \xi^T\Phi_k\xi\Phi_k \\
&= 2\Phi_k\xi\xi^T\Phi_k + \xi^T\Phi_k\xi\Phi_k
\end{aligned}$$

Therefore we have

$$Hess_{\xi}(\Omega(\xi,\lambda)) = \frac{1}{m}\sum_{k=1}^{m}(2\Phi_k\xi\xi^T\Phi_k + \xi^T\Phi_k\xi\Phi_k) + \lambda S - \Gamma_0$$

$$= 2\widetilde{\Gamma}(\xi) + \Gamma(\xi) + \lambda S - \Gamma_0$$

□

4.2.2 Boundedness

We show that if $\xi \neq 0$ is a critical point of the Ω criterion, then it is bounded by a parabola. Such a critical point ξ with $\xi \neq 0$ satisfies

$$(\Gamma(\xi) + \lambda Q - \Gamma_0)\xi = 0$$

Therefore, taking the inner product of that expression with ξ, we get

$$\langle\Gamma(\xi)\xi,\xi\rangle + \langle(\lambda S - \Gamma_0)\xi,\xi\rangle = 0$$

$$\langle\Gamma(\xi)\xi,\xi\rangle = \langle(\Gamma_0 - \lambda S)\xi,\xi\rangle$$

We know that

$$\langle(\Gamma_0 - \lambda S)\xi,\xi\rangle \leq \lambda_{max}(\Gamma_0 - \lambda S)||\xi||^2 \tag{4.11}$$

$$\langle\Gamma(\xi)\xi,\xi\rangle \geq \alpha_0||\xi||^4 \tag{4.12}$$

Putting these together, we get

$$||\xi||^2 \leq \frac{\lambda_{max}(\Gamma_0 - \lambda S)}{\alpha_0}$$

so there is a specifically parabolic form to the bound and the trajectories are bounded.

In the case that $S = I$, then $\lambda_{max}(\Gamma_0 - \lambda I) = \lambda_1 - \lambda$, so

$$||\xi||^2 \leq \frac{\lambda_1 - \lambda}{\alpha_0} \tag{4.13}$$

where $\lambda_1 = \lambda_{max}(\Gamma_0)$.

For general S, we can look, instead of at equation 4.11, at

$$\langle(\Gamma_0 - \lambda S)\xi, \xi\rangle \leq \lambda_{max}(S^{-\frac{1}{2}}\Gamma_0 S^{-\frac{1}{2}} - \lambda I)||S^{\frac{1}{2}}\xi||^2 \tag{4.14}$$

From which we see that

$$||\xi||^2 \leq \frac{(\lambda_{max}(S^{-1}\Gamma_0) - \lambda)||S||}{\alpha_0} \tag{4.15}$$

4.2.3 Sufficiency

We show that Γ_0 is a sufficient statistic, an analogue for R_0 being a sufficient statistic in the real case. The proof closely follows that for the real case as well.

Theorem 4.2.4. $\Gamma_0 = \frac{1}{m} y_k \Phi_k$ *is a sufficient statistic for* y*, if the noise is drawn*

from a normal.

Proof. This proof is essentially the same proof as was used in the real case, but modified with the notation of realification.

As before, we will use the Fisher–Neyman factorization theorem. Let $z \in \mathbb{C}^n$ be fixed. Define $\{y_k = |\langle z, f_k\rangle|^2 + \nu_k\}_{k=1\ldots m}$ where $\nu_k \sim \mathcal{CN}(0, \sigma^2)$ are i.i.d. measurements. We examine the PDF

$$p(y; z) = \frac{1}{(\sqrt{2\pi}\sigma)^m} exp\{\frac{-1}{2\sigma^2}\sum_{k=1}^{m}(y_k - |\langle z, f_k\rangle|^2)^2\}$$

Therefore, by taking the logarithm, we get

$$log(p(y; z)) = \frac{-1}{2\sigma^2}\sum_{k=1}^{m} y_k^2 - mlog(\sqrt{2\pi}\sigma) + \frac{1}{\sigma^2}\sum_{k=1}^{m} y_k\langle z, f_k\rangle^2 - \frac{1}{2\sigma^2}|\langle z, f_k\rangle|^4$$

Now we use the face that $\langle \Gamma_0\zeta, \zeta\rangle = \zeta^T \frac{1}{m}\sum_{k=1}^{m} y_k\Phi_k\zeta = \frac{1}{m}\sum_{k=1}^{m} y_k\zeta^T\Phi_k\zeta = \frac{1}{m}\sum_{k=1}^{m} y_k|\langle z, f_k\rangle|^2$.

Similarly, $\langle \Gamma(\zeta)\zeta, \zeta\rangle = |\langle z, f_k\rangle|^4$ (we can equivalently use $\widetilde{\Gamma}(\zeta)$ here)

Therefore, using these, we get

$$log(p(y; z)) = \frac{-1}{2\sigma^2}\sum_{k=1}^{m} y_k^2 - mlog(\sqrt{2\pi}\sigma) + \frac{m}{\sigma^2}\langle \Gamma_0 z, z\rangle - \frac{m}{2\sigma^2}\langle \Gamma(\zeta)z, z\rangle$$

Now we can factor

$$p(y; z) = f_0(y)g(\Gamma_0, z)$$

where both f_0 and g are nonnegative functions defined by

$$f_0 = \frac{1}{(\sqrt{2\pi}\sigma)^m} exp(-\frac{1}{2\sigma^2}\sum_{k=1}^{m} y_k^2)$$

$$g(R_0, z) = exp(-\frac{m}{2\sigma^2}\langle(\Gamma(\zeta) - 2\Gamma_0)z, z\rangle)$$

Therefore, the factorization theorem applies and Γ_0 is a sufficient statistic for z.

□

4.2.4 Properties of the Hessian and Gradient

We list some properties of the Hessian and Gradient in this section, as these properties differ from the real case.

Let $H = Hess_\xi(\Omega(\xi, \lambda))$, $Hess_{ext} = [H, S\xi]$, $d(\xi, \lambda) = \nabla_\xi \Omega(\xi, \lambda)$

Proposition 4.2.5. *Assume all of the notations above:*

Then the following are true:

1. $\Gamma(\xi)J = J\Gamma(\xi)$
2. $\Gamma(J\xi) = \Gamma(\xi)$
3. $\widetilde{\Gamma}(J\xi) = J^T\widetilde{\Gamma}(\xi)J$
4. $Jd(\xi, \lambda) = d(J\xi, \lambda)$
5. $\Gamma(\xi) = \widetilde{\Gamma}(\xi) + \widetilde{\Gamma}(J\xi)$

Note that $HJ \neq JH$, since $\widetilde{\Gamma}(\xi)$ and J do not commute as operators.

Proof. 1.

$$\begin{aligned}\Gamma(\xi)J &= \frac{1}{m}\sum_{k=1}^{m}\xi^T\Phi_k\xi\Phi_k J \\ &= \frac{1}{m}\sum_{k=1}^{m}(\xi^T\Phi_k\xi)J\Phi_k \\ &= \frac{1}{m}\sum_{k=1}^{m}J(\xi^T\Phi_k\xi)\Phi_k \\ &= J\frac{1}{m}\sum_{k=1}^{m}\xi^T\Phi_k\xi\Phi_k \\ &= J\Gamma(\xi)\end{aligned}$$

2. $\Gamma(J\xi) = \frac{1}{m}\sum_{k=1}^{m}\xi^T J^T\Phi_k J\xi\Phi_k = \frac{1}{m}\sum_{k=1}^{m}\xi^T\Phi_k(J^TJ)\xi\Phi_k = \frac{1}{m}\sum_{k=1}^{m}\xi^T\Phi_k\xi\Phi_k = \Gamma(\xi)$

3. $\widetilde{\Gamma}(J\xi) = \frac{1}{m}\sum_{k=1}^{m}\Phi_k J\xi(J\xi)^T\Phi_k = J\frac{1}{m}\sum_{k=1}^{m}\Phi_k\xi\xi^TJ^T\Phi_k = -J\frac{1}{m}\sum_{k=1}^{m}\Phi_k\xi\xi^TJ\Phi_k = J^T\frac{1}{m}\sum_{k=1}^{m}\Phi_k\xi\xi^T\Phi_k J = J^T\widetilde{\Gamma}(\xi)J$

4. First we know that $JS = SJ$, since $S = \begin{bmatrix} Q_R & Q_I \\ Q_I & Q_R \end{bmatrix}$

$$\begin{aligned}Jd(\xi,\lambda) &= J(\Gamma(\xi) + \lambda S - \Gamma_0)\xi \\ &= (\Gamma(\xi) + \lambda S - \Gamma_0)J\xi \\ &= (\Gamma(J\xi) + \lambda S - \Gamma_0)J\xi \\ &= d(J\xi,\lambda)\end{aligned}$$

5. It is sufficient to show that $\xi^T\Phi_k\xi\Phi_k = \Phi_k\xi\xi^T\Phi_k + \Phi_k J\xi(J\xi)^T\Phi_k$ for all k.

Therefore, label Φ_k as Φ.

We will equate these by taking quadratic forms on both sides against a vector e. First note that $e^T\xi^T\Phi\xi\Phi e = (\xi^T\Phi\xi)(e^T\Phi e)$. On the right hand side, we see that $e^T\Phi[\xi\xi^T + J\xi(J\xi)^T]e = \langle\Phi[\xi\xi^T + J\xi(J\xi)^T]\Phi e, e\rangle = \langle[\xi\xi^T + J\xi(J\xi)^T]\Phi e, \Phi e\rangle$. Now splitting the sum, we get $\langle\xi\xi^T\Phi e, \Phi e\rangle + \langle J\xi(J\xi)^T\Phi e, \Phi e\rangle = \langle\xi^T\Phi e, \xi^T\Phi e\rangle + \langle(J\xi)^T\Phi e, (J\xi)^T\Phi e\rangle = (\xi^T\Phi e)^2 + (\xi^T J^T\Phi e)^2$

Now we use that $\Phi = \varphi\varphi^T + J\varphi(J\varphi)^T$ and $J^T\Phi = \varphi(J\varphi)^T - (J\varphi)\varphi^T$.

Substituting these in on the left hand side, we see that $(\langle\varphi, \xi\rangle^2 + \langle J\varphi, \xi\rangle^2)(\langle\varphi, e\rangle^2 + \langle J\varphi, e\rangle^2)$.

On the right hand side, the substitution gets us $[(\langle\varphi, \xi\rangle\langle\varphi, e\rangle) + \langle J\varphi, \xi\rangle\langle J\varphi, e\rangle]^2 + [\langle\varphi, \xi\rangle\langle J\varphi, e\rangle - \langle J\varphi, \xi\rangle\langle\varphi, e\rangle]^2$.

Setting $a = \langle\varphi, \xi\rangle, b = \langle J\varphi, \xi\rangle, c = \langle\varphi, e\rangle, d = \langle J\varphi, e\rangle$, equating the left hand side with the right hand side, we want to show that $(a^2 + b^2)(c^2 + d^2) = (ac + bd)^2 + (ad - bc)^2$, but this is the Brahmagupta–Fibonacci identity, so the statement is true.

□

We now define and examine the properties of the complex bilinear cross matrices. These play the analogues to the real bilinear cross term $R(\cdot, \cdot)$.

Let $a, b \in \mathbb{R}^{2n}$. We write

$$\Gamma(a,b) = \frac{1}{m}\sum_{k=1}^{m} a^T \Phi_k b \Phi_k$$
$$\widetilde{\Gamma}(a,b) = \frac{1}{m}\sum_{k=1}^{m} \Phi_k a b^T \Phi_k$$

Note that $\Gamma(a,b) = \Gamma(b,a)$ but in general $\widetilde{\Gamma}(a,b) \neq \widetilde{\Gamma}(b,a)$. Also note that $\Gamma(a,a) = \Gamma(a)$ and $\widetilde{\Gamma}(a,a) = \widetilde{\Gamma}(a)$.

We can summarize some of the properties of the complex bilinear cross matrices in the following proposition.

Proposition 4.2.6. *Let $a, b, u \in \mathbb{R}^{2n}$. Then we have the following properties*

1. $\Gamma(a+b) = \Gamma(a) + \Gamma(b) + 2\Gamma(a,b)$
2. $\widetilde{\Gamma}(a+b) = \widetilde{\Gamma}(a) + \widetilde{\Gamma}(b) + \widetilde{\Gamma}(a,b) + \widetilde{\Gamma}(b,a)$
3. $\widetilde{\Gamma}(a,b)u = \widetilde{\Gamma}(a,u)b = \Gamma(b,u)a = \Gamma(u,b)a$ *(Trilinear relations)*
4. $\Gamma(a,b)a = \widetilde{\Gamma}(a)b$

Proof. 1. Each term in the summation of $m\Gamma(a+b)$ is $(a+b)^T\Phi_k(a+b)\Phi_k$. Expanding this, we get $a^T\Phi_k a\Phi_k + b^T\Phi_k b\Phi_k + 2a^T\Phi_k b\Phi_k$, so putting this back together with the summation and the scaling by $\frac{1}{m}$, we get $\Gamma(a)+\Gamma(b)+2\Gamma(a,b)$.

2. This proof is identical to (1).

3. Each term in the summation of $m\widetilde{\Gamma}(a,b)u = \Phi_k a b^T \Phi_k u = \Phi_k a(b^T\Phi_k u) = (b^T\Phi_k u)\Phi_k a$ which when putting it together with the summation is $\Gamma(b,u)a$.

Therefore $\widetilde{\Gamma}(a,b)u = \Gamma(b,u)a$ which by symmetry is the same as $\Gamma(u,b)a$. Now if we exchange all the positions of b and u, we get this is the same as $\widetilde{\Gamma}(a,u)b$ and the result is shown.

4. By the trilinear relations in (3), if we let $u = a$, then we get $\Gamma(a,b)a = \widetilde{\Gamma}(a,a)b = \widetilde{\Gamma}(a)b$

□

Proposition 4.2.7. *For ξ, λ such that $d(\xi,\lambda) = 0$, we have $dim(Null(Hess_{ext})) \geq 2$*

Proof. We first show that for such a pair (ξ,λ), $J\xi$ is in the null space of H.

First we begin by noting

$$H(J\xi) = (2\widetilde{\Gamma}(\xi) + \Gamma(\xi) + \lambda S - \Gamma_0)J\xi$$

$$= 2\widetilde{\Gamma}(\xi)J\xi + (\Gamma(\xi) + \lambda S - \Gamma_0)J\xi$$

Now by proposition 4.2.5, we have that $\Gamma(\xi)J = J\Gamma(\xi)$. We also know that $SJ = JS$, and $\Gamma_0 J = J\Gamma_0$.

Therefore we have

$$2\widetilde{\Gamma}(\xi)J\xi + (\Gamma(\xi) + \lambda S - \Gamma_0)J\xi = 2\widetilde{\Gamma}(\xi)J\xi + J(\Gamma(\xi) + \lambda S - \Gamma_0)\xi$$

$$= 2\widetilde{\Gamma}(\xi)J\xi$$

Since the gradient is assumed to be 0.

Now, we are left with showing that: $\widetilde{\Gamma}(\xi)J\xi = 0$.

To show this, we have that

$$\begin{aligned}
\widetilde{\Gamma}(\xi)J\xi &= \frac{1}{m}\sum_{k=1}^{m}\Phi_k\xi\xi^T\Phi_k J\xi \\
&= \frac{1}{m}\sum_{k=1}^{m}\Phi_k\xi\langle\xi,\Phi_k J\xi\rangle \\
&= \frac{1}{m}\sum_{k=1}^{m}\Phi_k\xi\langle\xi,J\Phi_k\xi\rangle \\
&= -\frac{1}{m}\sum_{k=1}^{m}\Phi_k\xi\langle J\xi,\Phi_k\xi\rangle \\
&= -\frac{1}{m}\sum_{k=1}^{m}\Phi_k\xi\langle\Phi_k J\xi,\xi\rangle \\
&= -\frac{1}{m}\sum_{k=1}^{m}\Phi_k\xi\langle\xi,\Phi_k J\xi\rangle \\
&= -\widetilde{\Gamma}(\xi)J\xi
\end{aligned}$$

Therefore we have that $\widetilde{\Gamma}(\xi)J\xi = 0$.

Therefore we have 1 vector in the null space of H, which is $J\xi$. To extend this to a vector in the null space of the extended hessian, we can just extend by 0, so our first vector in the null space would be: $\begin{bmatrix} J\xi \\ \\ 0 \end{bmatrix}$.

Now there are 2 cases. First is if $nullity(H) > 1$, in which case, we have another vector in the null space, so we can extend this by 0 as well, and have another null equation to our extended hessian.

Otherwise, the $nullity(H) = 1$, and we want to find another vector which is a solution to the null equation of the extended Hessian.

To find another vector, note we want:

$$Hess_{ext}\begin{bmatrix}\eta\\ y\end{bmatrix} = 0 \Rightarrow \begin{bmatrix}H & S\xi\end{bmatrix}\begin{bmatrix}\eta\\ y\end{bmatrix} = 0$$

Therefore we have that $H\eta + yS\xi = 0 \Rightarrow H\eta = -yS\xi$.

Therefore, we want $S\xi \in \text{Range}(H)$. To ensure such a inclusion exists, recall that since H is a symmetric matrix, we have that Range(H) =null$(H^T)^\perp$ =null$(H)^\perp$.

Now note that we can show it is orthogonal to our 1 dimensional null space by the following:

$$\begin{aligned}\langle J\xi, S\xi\rangle &= -\langle \xi, JS\xi\rangle\\ &= -\langle \xi, SJ\xi\rangle\\ &= -\langle S\xi, J\xi\rangle\end{aligned}$$

Therefore $\langle J\xi, S\xi\rangle = 0$. Now this implies that $S\xi \in$Range(H).

Now take η to be the vector such that $H\eta = -S\xi$. Then then by the above construction $\begin{bmatrix}\eta\\ 1\end{bmatrix}$ is another vector in the null space of $Hess_{ext}$. □

Now we have several very important bounds that will prove very useful in the analysis of the system.

Theorem 4.2.8. *Let $U(\theta) = cos(\theta)I + sin(\theta)J$. Let $\beta = \max_{||e||=1}\langle\Gamma(e)e, e\rangle$ Then the following properties hold:*

1. $||\Gamma(\xi_1) - \Gamma(\xi_2)|| \leq \beta \min_\theta ||\xi_1 - U(\theta)\xi_2|| \cdot ||\xi_1 + U(\theta)\xi_2||$ *for all* $\xi_1, \xi_2 \in \mathbb{R}^{2n}$

2. $||\widetilde{\Gamma}(\xi)|| \leq ||\Gamma(\xi)||$ *for all* $\xi \in \mathbb{R}^{2n}$

3. $||\Gamma(\xi)|| \leq \beta||\xi||^2$ *for all* $\xi \in \mathbb{R}^{2n}$

4. $||\widetilde{\Gamma}(\xi)|| \leq \beta||\xi||^2$ *for all* $\xi \in \mathbb{R}^{2n}$

5. $||\widetilde{\Gamma}(\xi_1) - \widetilde{\Gamma}(\xi_2)|| \leq \beta||\xi_1 - \xi_2|| \cdot ||\xi_1 + \xi_2||$ *for all* $\xi_1, \xi_2 \in \mathbb{R}^{2n}$

6. $||\Gamma(\xi_1, \xi_2)|| \leq \beta||\xi_1|| \cdot ||\xi_2||$ *for all* $\xi_1, \xi_2 \in \mathbb{R}^{2n}$

7. $||\widetilde{\Gamma}(\xi_1, \xi_2)|| \leq \beta||\xi_1|| \cdot ||\xi_2||$ *for all* $\xi_1, \xi_2 \in \mathbb{R}^{2n}$

In the following proofs, let x_1 be the complexification of ξ_1 and x_2 be the complexification of ξ_2. Also let ν be a unit vector in $\mathbb{R}^{2n}$, let e be the complexification of the vector ν. Also let $u = e^{i\theta}$ so that ux_2 is the realification of the vector $U\xi_2$.

Proof. 1. First recall that $\Gamma(\xi) = \Gamma(U\xi)$ for all $\xi \in \mathbb{R}^{2n}$.Now we see that $||\Gamma(\xi_1) - \Gamma(\xi_2)|| = ||\Gamma(\xi_1) - \Gamma(U\xi_2)||$. Now let ν be a unit vector in $\mathbb{R}^{2n}$, so let us examine $\nu^T(\Gamma(\xi_1) - \Gamma(U\xi_2))\nu$.

$$\begin{aligned}
\nu^T(\Gamma(\xi_1) - \Gamma(U\xi_2))\nu &= \frac{1}{m}\sum_k (\xi_1^T\Phi_k\xi_1 - \xi_2^tU^T\Phi_kU\xi_2)\nu^T\Phi_k\nu \\
&\leq \frac{1}{m}\sum_k |(\xi_1^T\Phi_k\xi_1 - \xi_2^tU^T\Phi_kU\xi_2)|\nu^T\Phi_k\nu \\
&= \frac{1}{m}\sum_k (|\langle x_1, f_k\rangle|^2 - |\langle ux_2, f_k\rangle|^2) \cdot |\langle e, f_k\rangle|^2 \\
&= \frac{1}{m}\sum_k |\langle x_1 - ux_2, f_k\rangle||\langle x_1 + ux_2, f_k\rangle| \cdot |\langle e, f_k\rangle|^2
\end{aligned}$$

Now this reduces to the proof in the real case, from which we know that

$$\max_{||\nu||=1}(\nu^T(\Gamma(\xi_1)-\Gamma(U\xi_2))\nu) \leq \beta||x_1-ux_2||\cdot||x_1+ux_2|| = \beta||\xi_1-U(\theta)\xi_2||\cdot||\xi_1+U(\theta)\xi_2||$$

This is true for all θ, so we get

$$||\Gamma(\xi_1) - \Gamma(\xi_2)|| \leq \beta \min_{\theta} ||\xi_1 - U(\theta)\xi_2|| \cdot ||\xi_1 + U(\theta)\xi_2||$$

2. First note this follows from the fact that $\Gamma(\xi) = \widetilde{\Gamma}(\xi) + \widetilde{\Gamma}(J\xi)$, but we can also show this directly. We recall that if A is a real positive semidefinite symmetric matrix, the Cauchy Schwarz inequality for positive semidefinite hermitian forms implies that $\langle Ax, y\rangle^2 \leq \langle Ax, x\rangle\langle Ay, y\rangle$ for all x, y.

   Also note that Φ_k is a symmetric positive semidefinite matrix.

   Let us start by examining

$$||\widetilde{\Gamma}(\xi)|| = \max_{||e||=1} e^T\widetilde{\Gamma}(\xi)e$$

   To use this, after expanding, we examine the k'th term of the summation

$$e^T\Phi_k x x^T\Phi_k e = \langle\Phi_k e, x\rangle^2$$

$$\leq \langle\Phi_k x, x\rangle\langle\Phi_k e, e\rangle$$

   Now we note that the k'th term in $e^T\Gamma(\xi)e = \langle\Phi_k x, x\rangle\langle\Phi_k e, e\rangle$, thus we have

$||\widetilde{\Gamma}(\xi)|| \leq ||\Gamma(\xi)||$ for all ξ

3. This follows from (1) by substituting $y = 0$ into the inequality.

4. This follows from (1) and (3).

5. To start we notice that $\widetilde{\Gamma}(\xi_1) - \widetilde{\Gamma}(\xi_2) = \frac{1}{m}\sum_{k=1}^{m} \Phi_k(\xi_1\xi_1^T - \xi_2\xi_2^T)\Phi_k$

   Define $u = \xi_1 - \xi_2$ and $v = \xi_1 + \xi_2$. Then $\widetilde{\Gamma}(\xi_1) - \widetilde{\Gamma}(\xi_2) = \frac{1}{m}\sum_{k=1}^{m} \Phi_k \frac{1}{2}(uv^T + uv^t)\Phi_k$

   Thus $\widetilde{\Gamma}(\xi_1) - \widetilde{\Gamma}(\xi_2) = \frac{1}{2m}\sum_{k=1}^{m} \Phi_k uv^T\Phi_k + \frac{1}{2m}\sum_{k=1}^{m} \Phi_k vu^T\Phi_k$

   Now we know that

$$
\begin{aligned}
||\widetilde{\Gamma}(\xi_1) - \widetilde{\Gamma}(\xi_2)|| &= \frac{1}{2}\max_{||a||=1} |\frac{1}{m}\sum_{k=1}^{m}\langle \Phi_k uv^T\Phi_k a, a\rangle + \frac{1}{m}\sum_{k=1}^{m}\langle \Phi_k vu^T\Phi_k a, a\rangle| \\
&= \max_{||a||=1} |\frac{1}{m}\sum_{k=1}^{m}\langle \Phi_k u, a\rangle\langle \Phi_k v, a\rangle| \\
&\leq \max_{||a||=1}[(\frac{1}{m}\sum_{k=1}^{m}\langle \Phi_k u, a\rangle)^2]^{\frac{1}{2}} \cdot \max_{||a||=1}[(\frac{1}{m}\sum_{k=1}^{m}\langle \Phi_k v, a\rangle)^2]^{\frac{1}{2}} \\
&= ||\frac{1}{m}\sum_{k=1}^{m}\Phi_k uu^T\Phi_k||^{\frac{1}{2}} \cdot ||\frac{1}{m}\sum_{k=1}^{m}\Phi_k vv^T\Phi_k||^{\frac{1}{2}} \\
&\quad ||\widetilde{\Gamma}(u)||^{\frac{1}{2}}||\widetilde{\Gamma}(v)||^{\frac{1}{2}} \\
&\quad \leq \beta||u|| \cdot ||v|| \\
&\quad = \beta||\xi_1 - \xi_2|| \cdot ||\xi_1 + \xi_2||
\end{aligned}
$$

6. Since for $v \in \mathbb{R}^{2n}$ we have $\langle \Gamma(\xi_1, \xi_2)v, v\rangle = \frac{1}{m}\sum_{k=1}^{m} \xi_1^T\Phi_k\xi_2 v^T\Phi_k v = \langle \Gamma(v)\xi_1, \xi_2\rangle$,

thus

$$||\Gamma(\xi_1,\xi_2)|| = \max_{||v||=1} |\langle \Gamma(\xi_1,\xi_2)v, v\rangle| = \max_{||v||=1} |\langle \Gamma(v)\xi_1, \xi_2\rangle|$$

$$\leq \max_{||v||=1} ||\Gamma(v)\xi_1|| \cdot ||\xi_2|| \leq max_{||v||=1}||\Gamma(v)|| \cdot ||\xi_1|| \cdot ||\xi_2|| \leq \beta||\xi_1|| \cdot ||\xi_2||$$

7. For $v \in \mathbb{R}^{2n}$, $\langle \widetilde{\Gamma}(\xi_1,\xi_2)v, v\rangle = \frac{1}{m}\sum_{k=1}^{m}(v^T\Phi_k\xi_1)(\xi_2^T\Phi_k v) = \frac{1}{m}\sum_{k=1}^{m}\xi_2^T\Phi_k v v^T\Phi_k\xi_1 = \langle \widetilde{\Gamma}(v)\xi_1, \xi_2\rangle$ Therefore we have

$$||\widetilde{\Gamma}(\xi_1,\xi_2)|| = \max_{||v||=1} |\langle \widetilde{\Gamma}(\xi_1,\xi_2)v, v\rangle| = \max_{||v||=1} |\langle \widetilde{\Gamma}(v)\xi_1, \xi_2\rangle|$$

$$\leq \max_{||v||=1} ||\widetilde{\Gamma}(v)\xi_1|| \cdot ||\xi_2|| \leq max_{||v||=1}||\widetilde{\Gamma}(v)|| \cdot ||\xi_1|| \cdot ||\xi_2|| \leq \beta||\xi_1|| \cdot ||\xi_2||$$

□

Define $\beta = \max_{||e||=1}\langle \Gamma(e)e, e\rangle$, we will write many of the results that require bounds in terms of β.

Corollary 4.2.9. *Let ξ_1 and ξ_2 be in $\mathbb{R}^{2n}$ such that they are aligned. That is, let $\delta = \xi_1 - \xi_2 = \xi_1 - U(\theta)\xi_2$ for θ that best aligns them (so $\min_\theta ||\xi_1 - U(\theta)\xi_2|| = ||\xi_1 - \xi_2||$). Then*

$$||Hess(\xi_1,\lambda) - Hess(\xi_2,\lambda)|| \leq 3\beta||\delta|| \cdot (||\delta|| + 2||\xi_2||)$$

Proof.

$$Hess(\xi_1, \lambda) - Hess(\xi_2, \lambda) = \Gamma(\xi_1) + 2\widetilde{\Gamma}(\xi_1) + \lambda S - \Gamma_0 - \Gamma(\xi_2) - 2\widetilde{\Gamma}(\xi_2) - \lambda S + \Gamma_0$$

$$= \Gamma(\xi_1) - \Gamma(\xi_2) + 2\widetilde{\Gamma}(\xi_1) - 2\widetilde{\Gamma}(\xi_2)$$

Therefore, taking norms, we see that

$$||Hess(\xi_1, \lambda) - Hess(\xi_2, \lambda)|| = ||\Gamma(\xi_1) - \Gamma(\xi_2) + 2\widetilde{\Gamma}(\xi_1) - 2\widetilde{\Gamma}(\xi_2)||$$

$$\leq ||\Gamma(\xi_1) - \Gamma(\xi_2)|| + 2||\widetilde{\Gamma}(\xi_1) - \widetilde{\Gamma}(\xi_2)||$$

$$\leq 3\beta||\xi_1 - \xi_2|| \cdot ||\xi_1 + \xi_2||$$

$$\leq 3\beta||\xi_1 - \xi_2|| \cdot ||\xi_1 + \xi_2||$$

$$\leq 3\beta||\xi_1 - \xi_2|| \cdot (||\xi_1 - \xi_2|| + ||2\xi_2||)$$

$$\leq 3\beta||\delta|| \cdot (||\delta|| + 2||\xi_2||)$$

□

4.2.5 Assumptions

There are several assumptions we make for this algorithm, which usually will happen in the generic case.

1. The frame set $\mathcal{F}$ is phase-retrievable.

2. For a fixed λ, the set of critical points of $\Omega(\xi, \lambda)$ is isolated in the quotient

space $\mathbb{R}^{2n}/\sim$, where $\xi_1 \sim \xi_2$ if for some $0 < \theta < 2\pi$, $\xi_1 = U(\theta)\xi_2$, for $U(\theta) = cos(\theta)I_{2n} + sin(\theta)J$

3. The top eigenvalue of $S^{-1}\Gamma_0$ has a two dimensional eigenspace.

4. For any critical point of the Ω-criterion, (ξ, λ), we assume that the extended Hessian $Hess_{ext}$ is of rank $2n - 1$.

Conditions 1 and 2 ensure that it is reasonable to try to recover the signal. Condition 3 will ensure that the initialization algorithm is well defined. Condition 4 ensures that the homotopy path is well defined and smooth.

As with the real case, the conditions are not independent of eachother, but it is important to emphasize each of them.

4.2.6 Initialization

In the next sections, we explore how to initialize the Complex Golden Retriever, which is very similar to the real case.

First we notice that from the Ω-criterion (4.10), we get that for $\lambda \geq \lambda_1(S^{-1}\Gamma_0)$, the minimization occurs when $\xi = 0$, since $\Gamma(\xi)$ is positive definite, and $(\lambda S - \Gamma_0)$ is as well. So, like the real case, we can initialize the algorithm at $(\lambda_1 - \epsilon)$. Next we need to determine how to initialize ξ.

To initialize ξ, we turn to the gradient equation 4.2.2, and since $\xi \approx 0$, we ignore the cubic term in ξ, $\Gamma(\xi)\xi$ (as a higher order error term), so we see

$$0 = \nabla_\xi \Omega(\xi, \lambda) = (\Gamma(\xi) + \lambda S - \Gamma_0)\xi \approx (\lambda S - \Gamma_0)\xi$$

We look at $0 = \lambda S\xi - \Gamma_0\xi$, and note that it satisfies $\lambda\xi = S^{-1}\Gamma_0\xi$, for ϵ, small therefore, this approaches the eigenvector of $S^{-1}\Gamma_0$ corresponding to the eigenvalue at λ_1. Therefore, we will initialize the algorithm at $(\xi = ce_1, \lambda_1 - \epsilon)$.

To finish the initialization, we need to determine the scaling parameter c. We choose the one which minimizes the Ω-criterion.

Therefore we want to find

$$\arg\min_{\xi\in\mathbb{R}^{2n}} \Omega(ce_1, \lambda_1 - \epsilon) = \frac{1}{4}c^4\langle\Gamma(e_1)e_1, e_1\rangle - \frac{\epsilon}{2}c^2\langle Se_1, e_1\rangle + \frac{1}{4m}\sum_k y_k^2$$

This is a quadratic equation in c^2 which is minimized at $c^2 = \frac{\frac{\epsilon}{2}\langle Se_1,e_1\rangle}{\frac{1}{2}\langle\Gamma(e_1)e_1,e_1\rangle} = \epsilon\frac{\langle Se_1,e_1\rangle}{\langle\Gamma(e_1)e_1,e_1\rangle}$

Therefore, we choose $c = \sqrt{\frac{\epsilon\langle Se_1,e_1\rangle}{\langle\Gamma(e_1)e_1,e_1\rangle}}$. This determines the initialization of the algorithm.

4.2.7 Update Rules

As with the real case, the update rules are split into two steps, the predictor step and the corrector step.

Step 1: The Predictor

The goal of the predictor step is to make an Euler Step in the direction of the homotopy path. Therefore, we want a new point (ξ, λ) that roughly follows the path $\nabla_x(\Omega)^{-1}(0)$ which is smoothly connected to the $(0, \lambda_1)$. If we parameterize the path by t, so $\xi = \xi(t)$ and $\lambda = \lambda(t)$, we want to step in the direction based on the slope of the curve at the current point $(\xi(t), \lambda(t))$.

Therefore, we want to step into the direction of the tangent of this curve,

which is given by $\begin{bmatrix} \frac{d\xi}{dt} \\ \\ \frac{d\lambda}{dt} \end{bmatrix}$.

By differentiating the equation

$$F(\xi(t), \lambda(t)) = 0$$

we can find the tangent by computing the derivative

$$\frac{d}{dt} F(\xi(t), \lambda(t)) = 0 \tag{4.16}$$

From the work we did earlier we know

$$\frac{d}{dt} F(\xi(t), \lambda(t)) = Hess_{ext}(\xi(t), \lambda(t)) \begin{bmatrix} \frac{d\xi}{dt} \\ \\ \frac{d\lambda}{dt} \end{bmatrix} = 0 \tag{4.17}$$

Therefore the direction we want to step in the same direction as a vector in the null space of $Hess_{ext}(\xi(t), \lambda(t))$. Note that unlike the real case, the $dim(Null(Hess_{ext})) \geq 2$, but with equality under our assumptions. One of the null vectors is given by the phase ambiguity $\begin{bmatrix} J\xi \\ \\ 0 \end{bmatrix}$. This we pick the vector in the $Null(Hess_{ext})$ perpendicular to $\begin{bmatrix} J\xi \\ \\ 0 \end{bmatrix}$.

To summarize this, in the predictor step, we compute the extended Hessian matrix $Hess_{ext}$, find the vector in the null space, perpendicular to $\begin{bmatrix} J\xi \\ \\ 0 \end{bmatrix}$, that

matches sign in the largest coordinate with the sign of the coordinate in the previous step (to make sure the path is moving in the correct direction), and then make a choice in step size. Unfortunately, it likely going to step away from the path, so we need a corrector algorithm to get us back on the path.

Step 2: The Corrector

In this, we want to find a point (ξ, λ) which is a solution to the gradient being zero but is as close as possible to the output in the Predictor Step. We can use the Newton Step. If $\begin{bmatrix} \xi_{old} \\ \lambda_{old} \end{bmatrix}$ was our old estimate, we can update it with a correction of the form

$$\begin{bmatrix} \xi_{new} \\ \lambda_{new} \end{bmatrix} = \begin{bmatrix} \xi_{old} \\ \lambda_{old} \end{bmatrix} - H_{ext}^{+} F(\xi_{old}, \lambda_{old}) \tag{4.18}$$

Where H_{ext}^{+} is the pseudo-inverse of the extended Hessian, with an extra row added to ensure that it is orthogonal to the vector $\begin{bmatrix} J\xi_{old} \\ 0 \end{bmatrix}$. Therefore, it is the pseudo-inverse of the following matrix:

$$H_{ext}(\xi, \lambda) = \begin{bmatrix} Hess(\xi, \lambda) & S\xi \\ (J\xi)^T & 0 \end{bmatrix} \tag{4.19}$$

The Newton corrector step is well studied, and under suitable conditions on the extended Hessian, is guaranteed to converge to a critical point after a number of corrector steps. See Chapter 3 of [9], specifically Theorem 3.4.1 in for a full treatment on the subject.

4.3 Expected System

Assume that φ is a random variable distributed as a standard normal distribution, i.e. $\varphi \sim \mathcal{N}(0, I_{2n})$. Also let ζ be the realification of the generating signal z.

Lemma 4.3.1. $\mathbb{E}_\varphi[\sum_{k=1}^N \sum_{l=1}^N (\xi_k \eta_l \varphi_k \varphi_l \varphi_i \varphi_j)] = \langle \xi, \eta \rangle I_N + \eta \xi^T + \xi \eta^T$

Proof. We first look at the case where $i = j$. We have two subcases, one where $k = l = i = j$, in which case we get $3\xi_i \eta_i$ (where 3 comes from $\mathbb{E}(\varphi_i^4)$). If $k = l \neq i = j$, then we have $\sum_{k \neq i} \xi_k \eta_k = \langle \xi, \eta \rangle - \xi_i \eta_i$.

Therefore the expected value along the diagonal $i = j$, gives us $\langle \xi, \eta \rangle + 2\xi_i \eta_i$.

For the case where $i \neq j$, we have either that $i = k, j = l$ or that $i = l, j = k$, so we get two terms: $\xi_k \eta_l + \xi_l \eta_k$.

Therefore, we see we can put both terms together with $\eta \xi^T + \xi \eta^T + \langle \xi, \eta \rangle I_N$

□

Corollary 4.3.2. $\mathbb{E}_\varphi[\sum_{k=1}^N \sum_{l=1}^N (\xi_k \xi_l \varphi_k \varphi_l \varphi_i \varphi_j)] = ||\xi||^2 I_N + 2\xi \xi^T$

Proposition 4.3.3.

$$\mathbb{E}_\varphi(\Gamma(\xi)) = 4||\xi||^2 I_{2n} + 4\xi \xi^T + 4(J\xi)(J\xi)^T \tag{4.20}$$

Proof. Recall that:

$$\Gamma(\xi) = \frac{1}{m} \sum_{k=1}^m \xi^T \Phi_k \xi \Phi_k$$

Since each of the summands are independently and identically distributed, it is sufficient to consider $\mathbb{E}(\xi^T \Phi \xi \Phi)$ for a randomly chosen Φ.

Note that $\Phi = \varphi\varphi^T + (J\varphi)(J\varphi)^T$, so this is a sum of four terms

$$\mathbb{E}(\xi^T \Phi \xi \Phi)$$

$$= \mathbb{E}(\xi^T \varphi\varphi^T \xi \varphi\varphi^T + \xi^T \varphi\varphi^T \xi J\varphi(J\varphi)^T$$

$$+\xi^T (J\varphi)(J\varphi)^T \xi \varphi\varphi^T + \xi^T (J\varphi)(J\varphi)^T \xi (J\varphi)(J\varphi)^T)$$

Call these M_1, M_2, M_3, M_4 respectively. Now we check each of these

$$(M_1)_{i,j} = (\xi^T \varphi\varphi^T \xi \varphi\varphi^T)_{i,j}$$

$$= (\xi^T \varphi\varphi^T \xi \varphi\varphi^T)\varphi_i\varphi_j$$

$$= \langle \varphi^T\xi, \varphi^T\xi \rangle \varphi_i\varphi_j$$

$$= \sum_{s=1}^{2n} (\varphi_s\xi_s)^2 \varphi_i\varphi_j$$

$$= \sum_{k=1}^{2n} (\varphi_k\xi_k) \sum_{l=1}^{2n} (\varphi_l\xi_l)\varphi_i\varphi_j$$

$$= \sum_{k=1}^{2n}\sum_{l=1}^{2n} (\xi_k\xi_l\varphi_k\varphi_l\varphi_i\varphi_j)$$

By lemma 4.3.1 above, we get: $||\xi||^2 I_{2n} + 2\xi\xi^T$.

Therefore we have

$$\mathbb{E}_\varphi(M_1) = ||\xi||^2 I_{2n} + 2\xi\xi^T$$

Now we consider $M_2 = \xi^T \varphi\varphi^T \xi J\varphi (J\varphi)^T$. The key observation to simplify things is to note that $J^T M_2 J = \xi^T \varphi\varphi^T \xi\varphi\varphi^T = M_1$, so we get $\mathbb{E}_\varphi[M_2] = \mathbb{E}_\varphi[JM_1J^T] = J\mathbb{E}_\varphi[M_1]J^T$

Substituting in we see that $\mathbb{E}_\varphi[M_2] = ||\xi||^2 I + 2(J\xi)(J\xi)^T$

For M_3, we define $\eta = J\varphi$, and then we rewrite $M_3 = \xi^T (J\varphi)(J\varphi)^T \xi\varphi\varphi^T = \xi^T \eta\eta^T \xi (J\eta)(J\eta)^T$

Since the distribution of η is the same as the distribution of φ we get

$$\mathbb{E}_\varphi(M_3) = \mathbb{E}_\eta(M_3) = ||\xi||^2 I + 2(J\xi)(J\xi)^T$$

Finally for M_4, again since the distribution is the same, we can rewrite

$$M_4 = \xi^T \eta\eta^T \xi\eta\eta^T \Rightarrow \mathbb{E}_\eta[M_4] = \mathbb{E}_\varphi[M_1] = ||\xi||^2 I + 2\xi\xi^T$$

So putting it all together, we see that

$$\begin{aligned}\mathbb{E}_\varphi(\Gamma(\xi)) &= \mathbb{E}_\varphi(M_1) + \mathbb{E}_\varphi(M_2) + \mathbb{E}_\varphi(M_3) + \mathbb{E}_\varphi(M_4) \\ &= 4||\xi||^2 I_{2n} + 4\xi\xi^T + 4(J\xi)(J\xi)^T\end{aligned}$$

□

Corollary 4.3.4. *In the noiseless case, we have:* $\mathbb{E}_\varphi(\Gamma_0) = 4||\zeta||^2 I_{2n} + 4\zeta\zeta^T + 4(J\zeta)(J\zeta)^T$

Proof. Note that in the noiseless case $\Gamma_0 = \Gamma(\zeta)$, therefore we can apply the above

proposition. □

Proposition 4.3.5.

$$\mathbb{E}_\varphi(\widetilde{\Gamma}(\xi)) = 2||\xi||^2 I_{2n} + 6\xi\xi^T - 2(J\xi)(J\xi)^T \tag{4.21}$$

Proof. Note that, similar to the case for $\Gamma(\xi)$, we can split it up into four cases

$$\mathbb{E}_\varphi(\widetilde{\Gamma}(\xi)) = \mathbb{E}_\varphi(M_1 + M_2 + M_3 + M_4)$$

Where

$$M_1 = \varphi\varphi^T\xi\xi^T\varphi\varphi^T$$

$$M_2 = \varphi\varphi^T\xi\xi^T(J\varphi)(J\varphi)^T = \varphi\varphi^T\xi\xi^T J\varphi\varphi^T J^T$$

$$M_2 J = \varphi\varphi^T\xi\eta^T\varphi\varphi^T$$

$$\text{Where } \eta = J^T\xi$$

$$M_3 = (J\varphi)(J\varphi)^T\xi\xi^T\varphi\varphi^T$$

$$M_4 = (J\varphi)(J\varphi)^T\xi\xi^T(J\varphi)(J\varphi)^T$$

For M_1, we have the following

$$\begin{aligned}(M_1)_{i,j} &= \sum_{k=1}^{2n}(\varphi\varphi^T\xi\xi^T)_{i,k}(\varphi\varphi^T)_{k,j} \\ &= \sum_{k=1}^{2n}\sum_{l=1}^{2n}(\varphi\varphi^T)_{i,l}(\xi\xi^T)_{l,k}(\varphi\varphi^T)_{k,j} \\ &= \sum_{k=1}^{2n}\sum_{l=1}^{2n}\xi_k\xi_l\varphi_k\varphi_l\varphi_i\varphi_j\end{aligned}$$

Which is the same form as M_1 from $\Gamma(\xi)$ (or from the real case), therefore we know

$$\mathbb{E}_\varphi(M_1) = ||\xi||^2 I_{2n} + 2\xi\xi^T$$

For M_2, it can be reduced to a similar expression with the same kind of trick, we can reduce it further.

Define $\eta = J^T\xi$, then we can write

$$M_2 = \varphi\varphi^T\xi\xi^T(J\varphi)(J\varphi)^T = \varphi\varphi^T\xi\xi^T J\varphi\varphi^T J^T$$

$$M_2 J = \varphi\varphi^T\xi\eta^T\varphi\varphi^T$$

As with M_1, this reduces to the following

$$(M_2J)_{i,j} = \sum_{k=1}^{2n}\sum_{l=1}^{2n}\xi_k\eta_l\varphi_k\varphi_l\varphi_i\varphi_j$$

By lemma 4.3.1, we have that this is exactly $\eta\xi^T + \xi\eta^T + \langle\xi,\eta\rangle I_{2n}$. Since $\eta = J^T\xi$, which is orthogonal to ξ, this is just $J^T\xi\xi^T + \xi(J^T\xi)^T$

Therefore putting things together, we have

$$\mathbb{E}_\varphi(M_2) = \mathbb{E}_\varphi(M_2 J)J^T$$

$$= (\eta\xi^T + \xi\eta^T + \langle\xi,\eta\rangle I_{2n})J^T$$

$$= (J^T\xi\xi^T + \xi(J^T\xi)^T + \langle\xi, J^T\xi\rangle I_{2n})J^T$$

$$= J^T\xi\xi^T J^T + \xi\xi^T JJ^T$$

$$= -J\xi\xi^T J^T + \xi\xi^T JJ^T$$

$$= \xi\xi^T - (J\xi)(J\xi)^T$$

Note that $M_3 = M_2^T$

$$\mathbb{E}_\varphi(M_3) = \mathbb{E}_\varphi(M_2^T) = \mathbb{E}_\varphi(M_2)^T$$

$$= (\xi\xi^T - (J\xi)(J\xi)^T)^T$$

$$= \xi\xi^T - (J\xi)(J\xi)^T$$

So M_3 has the same expectation as M_2.

For M_4, we notice that $J\varphi$ has the same distribution as φ, so it would have the same expectation. Therefore: $\mathbb{E}(M_1) = \mathbb{E}(M_4)$.

Putting the four matrices together, we get

$$\mathbb{E}_\varphi(\widetilde{\Gamma}(\xi)) = \mathbb{E}_\varphi(M_1 + M_2 + M_3 + M_4) = 2||\xi||^2 I_{2n} + 6\xi\xi^T - 2(J\xi)(J\xi)^T$$

□

Corollary 4.3.6. *In the noiseless case:* $\mathbb{E}_\varphi(\nabla_\xi \Omega(\xi,\lambda)) = 8||\xi||^2\xi + \lambda S\xi - 4||\zeta||^2\xi - 4\langle\xi,\zeta\rangle\zeta - 4\langle\xi, J\zeta\rangle J\zeta$

As in the real case, let us now examine the expected system with $S = I$, which we define to be $\mathbb{E}_\varphi(\nabla_\xi \Omega(\xi,\lambda)) = 0$. To do this, we presuppose that the solution is of the form $\xi = k\zeta$ and we simplify to the equation to

$$\begin{aligned} 0 &= 8||\xi||^2\xi + \lambda\xi - 4||\zeta||^2\xi - 4\langle\xi,\zeta\rangle\zeta - 4\langle\xi, J\zeta\rangle J\zeta \\ &= k(8k^2||\zeta||^2\zeta + \lambda\zeta - 4||\zeta||^2\zeta - 4||\zeta||^2\zeta) \end{aligned}$$

Therefore, dividing through by k gives us

$$\begin{aligned} 0 &= (8k^2||\zeta||^2\zeta + \lambda\zeta - 4||\zeta||^2\xi - 4||\zeta||^2\zeta) \\ &= (8k^2||\zeta||^2 + \lambda - 8||\zeta||^2)\zeta \end{aligned}$$

Setting the scaling to be zero, we see

$$\begin{aligned} 0 &= 8k^2||\zeta||^2 + \lambda - 8||\zeta||^2 \\ k^2 &= 1 - \frac{\lambda}{8||\zeta||^2} \end{aligned}$$

Since $\mathbb{E}(\Gamma_0) = 4||\zeta||^2 + 4\zeta\zeta^T + 4J\zeta(J\zeta)^T$ we see that the spectrum of $\mathbb{E}(\Gamma_0) = \{8||\zeta||^2, 8||\zeta||^2, 4||\zeta||^2, ..., 4||\zeta||^2\}$. Therefore, we get that $k = \sqrt{1 - \frac{\lambda}{\lambda_1(\mathbb{E}(\Gamma_0))}}$

So the solution to the gradient system is given by

$$\xi(\lambda) = \left(\sqrt{1 - \frac{\lambda}{\lambda_1(\mathbb{E}(\Gamma_0))}}\right)\zeta \tag{4.22}$$

Corollary 4.3.7. *In the noiseless case, we have:*

$$\mathbb{E}_\varphi(Hess(\xi,\lambda)) = (8||\xi||^2 - 4||\zeta||^2)I_{2n} + 16\xi\xi^T - 4\zeta\zeta^T - 4(J\zeta)(J\zeta)^T + \lambda S \tag{4.23}$$

Proof.

$$\mathbb{E}_\varphi(Hess(\xi,\lambda)) = \mathbb{E}_\varphi(2\widetilde{\Gamma}(\xi) + \Gamma(\xi) + \lambda S - \Gamma_0)$$

$$= (4||\xi||^2 I_{2n} + 12\xi\xi^T - 4(J\xi)(J\xi)^T) + (4||\xi||^2 I_{2n} + 4\xi\xi^T + 4(J\xi)(J\xi)^T) + \lambda S$$

$$-(4||\zeta||^2 I_{2n} + 4\zeta\zeta^T + 4(J\zeta)(J\zeta)^T)$$

$$= (8||\xi||^2 - 4||\zeta||^2)I_{2n} + 16\xi\xi^T - 4\zeta\zeta^T - 4(J\zeta)(J\zeta)^T + \lambda S$$

□

Lemma 4.3.8. *Let $v, w \sim \mathcal{N}(0, I_m)$. Then for any $\epsilon > 0$ there exists a C which depends on ϵ such that for $m \geq C(\epsilon)$, each of the following hold with probability at*

least $1 - \frac{1}{m^2}$

$$|\frac{1}{m}\sum_{k=1}^{m} v_k^2 - 1| < \epsilon \qquad |\frac{1}{m}\sum_{k=1}^{m} w_k^2 - 1| < \epsilon$$
$$|\frac{1}{m}\sum_{k=1}^{m} v_k^4 - 3| < \epsilon \qquad |\frac{1}{m}\sum_{k=1}^{m} w_k^4 - 3| < \epsilon$$
$$|\frac{1}{m}\sum_{k=1}^{m} v_k^6 - 15| < \epsilon \qquad |\frac{1}{m}\sum_{k=1}^{m} w_k^6 - 15| < \epsilon$$
$$\max_{1\le k\le m} |v_k| \le \sqrt{10log(m)} \qquad \max_{1\le k\le m} |w_k| \le \sqrt{10log(m)}$$
$$|\frac{1}{m}\sum_{k=1}^{m} v_k^2 w_k^2 - 1| \le \epsilon$$
$$|\frac{1}{m}\sum_{k=1}^{m} v_k^3 w_k| < \epsilon \qquad |\frac{1}{m}\sum_{k=1}^{m} w_k^3 v_k| < \epsilon$$

Furthermore, conditional on the above probabilities, for $m \ge C(\epsilon)$*, the above also hold with probability at least* $1 - \frac{1}{n}$

$$|\frac{1}{m}\sum_{k=1}^{m}(w_k^3 + v_k^2 w_k)^2| < 500 \quad |\frac{1}{m}\sum_{k=1}^{m}(v_k^3 + w_k^2 v_k)^2| < 500$$

Proof. The first 8 inequalities are identical to the real case, and so they is handled in Lemma 3.2.6. The next 3 can be handled in much the same way, after conditioning on one of the variables.

For the term $|\frac{1}{m}\sum_{k=1}^{m}(w_k^3 + v_k^2 w_k)^2| < 500$ We first expand and see we want an upper bound on $\frac{1}{m}\sum_{k=1}^{m}(w_k^6 + 2v_k^2 w_k^4 + v_k^4 w_k^2) \le \frac{3}{m}\sum_{k=1}^{m} w_k^8 + \frac{3}{2m}\sum_{k=1}^{m} v_k^8 + \frac{3}{2m}\sum_{k=1}^{m} w_k^2 + \frac{1}{m}\sum_{k=1}^{m} v_k^2$. The last inequality hold because $2w_k^4 v_k^2 \le w_k^8 + v_k^4$ and $v_k^4 w_k^2 \le \frac{1}{2}(v_k^8 + w_k^4)$ and $w_k^6 \le w_k^8 + w_k^2$.

Thus since $\frac{1}{m}\sum_{k=1}^{m} v_k^2 \le 1+\epsilon$ and $\frac{1}{m}\sum_{k=1}^{m} v_k^8 \le 105+\epsilon$, $\frac{1}{m}\sum_{k=1}^{m}(w_k^3+v_k^2 w_k)^2 \le$

$475 + 7\epsilon$. Therefore, we can bound it by 500 and for $m \geq C(\epsilon)$, it holds true with probability $1 - \frac{1}{m^2}$, given that the other bounds are true.

□

Theorem 4.3.9. *Assume $\varphi_k \sim \mathcal{N}(0, I_{2m})$ and $||\xi|| = 1$. Let $\epsilon > 0$ be a constant and $C(\epsilon)$ be a sufficiently large constant that is allowed to depend on ϵ. Let $m > C(\epsilon)nlog(n)$. Then $||\Gamma(\xi) - \mathbb{E}[\Gamma(\xi)]|| \leq \epsilon$ with probability $1 - \frac{13}{n^2} - 10e^{-\gamma n}$*

Proof. The proof is very similar to the proof of Theorem 3.2.7. By unitary invariance, let $\xi = e_1$, the standard unit vector in $\mathbb{R}^{2n}$. Let y be a unit norm vector and let us examine $I_0(y) = y^T\Gamma(e_1)y - y^T\mathbb{E}[\Gamma(e_1)]y$.

Let $v_k = \varphi_k(1)$ and $w_k = \varphi_k(n+1)$, then standard computations show that

$$y^T\Gamma(e_1)y = \frac{2}{m}\sum_{k=1}^{m}(v_k^2 + w_k^2)(y(1)^2v_k^2 + y(n+1)^2w_k^2 + 2y(1)y(n+1)v_kw_k$$
$$+2y(1)v_k\langle\widetilde{y}, \widetilde{\varphi_k}\rangle + 2y(n+1)w_k\langle\widetilde{y}, \widetilde{\varphi_k}\rangle + \langle\widetilde{y}, \widetilde{\varphi_k}\rangle^2)$$

Since $\mathbb{E}[\Gamma(e_1)] = 4I + 4e_1e_1^T + 4(Je_1)(Je_1)^T$, we get that

$$y^T\mathbb{E}[\Gamma(e_1)]y = 8y(1)^2 + 8y(n+1)^2 + 4||\widetilde{y}||^2$$

Putting these together, we see that

$$I_0(y) = \frac{2}{m}\sum_{k=1}^{m}(v_k^4-3)y(1)^2 + \frac{2}{m}\sum_{k=1}^{m}(w_k^4-3)y(1)^2 + \frac{2}{m}\sum_{k=1}^{m}(v_k^2w_k^2-1)(y(1)^2+y(n+1)^2)$$
$$+\frac{4}{m}\sum_{k=1}^{m}y(1)y(n+1)v_k^3w_k + \frac{4}{m}\sum_{k=1}^{m}y(1)y(n+1)v_kw_k^3 + \frac{4}{m}\sum_{k=1}^{m}\langle\widetilde{y},\widetilde{\varphi_k}\rangle y(1)(v_k^3+v_kw_k^2)$$
$$+\frac{4}{m}\sum_{k=1}^{m}\langle\widetilde{y},\widetilde{\varphi_k}\rangle y(n+1)(w_k^3+v_k^2w_k) + \frac{2}{m}\sum_{k=1}^{m}\langle\widetilde{y},\widetilde{\varphi_k}\rangle^2v_k^2 - ||\widetilde{y}||^2 + \frac{2}{m}\sum_{k=1}^{m}\langle\widetilde{y},\widetilde{\varphi_k}\rangle^2w_k^2 - ||\widetilde{y}||^2$$

Now we estimate each of these terms. Let $delta_0 = \frac{\epsilon}{16}$ and $\epsilon_0 = \frac{\epsilon}{28}$. We start with the last 2 terms, each of which is of the form $\frac{2}{m}\sum_{k=1}^{m}\langle\widetilde{y},\widetilde{\varphi_k}\rangle^2v_k^2 - ||\widetilde{y}||^2 = \frac{2}{m}\sum_{k=1}^{m}v_k^2(\langle\widetilde{y},\widetilde{\varphi_k}\rangle^2-||\widetilde{y}||^2)+\frac{2}{m}\sum_{k=1}^{m}(v_k^2-1)||\widetilde{y}||^2$. The second part we estimate from the inequalities in Lemma 3.2.6, and the other part is identical to the Bernstein Type Inequality used in the proof of Theorem 3.2.7. Therefore for $m \geq C_0(\sqrt{n\sum_k\varphi_k^6} + n\max_k|v_k|^2)$

$$|\frac{1}{m}\sum_{k=1}^{m}\langle\widetilde{y},\widetilde{\varphi_k}\rangle^2v_k^2 - ||\widetilde{y}||^2| \leq \delta_0||\widetilde{y}||^2 \leq \delta_0$$

holds with probability $1-2e^{-2\gamma n}$. A similar thing holds with the w_k term, thus with probability $1-4e^{-2\gamma n}$ both of these hold.

To estimate the remaining term $\frac{4}{m}\sum_{k=1}^{m}\langle\widetilde{y},\widetilde{\varphi_k}\rangle y(n+1)(w_k^3+v_k^2w_k)$, we want to use the Hoeffding inequality, as we did in the real case, but we need an upper bound on the l_2 norm of the coefficients, so we need an upper bound on $\frac{1}{m}\sum_{k=1}^{m}(w_k^3+v_k^2w_k)^2$.

We fortunately get that from the lemma, so we can apply Hoeffding's inequality with a bound of 500.

Thus for $m \geq C_1\sqrt{nV}$, where $V = \sum_{k=1}^{m}(w_k^3 + v_k^2 w_k)^2$, we have

$$\frac{1}{m}\sum_{k=1}^{m}\langle\widetilde{y},\widetilde{\varphi_k}\rangle y(1)(v_k^3 + v_k w_k^2) \leq \delta_0|y(1)| \cdot ||\widetilde{y}|| \leq \delta_0$$

Conditioning on the bounds in the lemma (with an upper bound of ϵ_0) , we get for $m \geq \max\{C_0(\sqrt{n\sum_k \varphi_k^6} + n\max_{k=1,\ldots,n}|v_k|^2 + n\max_{k=1,\ldots,n}|w_k|^2), C_1(\sqrt{nV} + \sqrt{nW})\}$, (where $W = \sum_{k=1}^{m}(v_k^3 + w_k^2 v_k)^2$) with probability at least $1 - 10e^{-2\gamma n}$, $I_0(y) \leq 14\epsilon_0 + 8\delta_0 = \epsilon$

A similar argument to the real case (involving the $\mathcal{N} - net$ which forces $\gamma > log(9)$) now shows that with probability at least $1 - \frac{13}{n^2} - 10e^{-\gamma n}$, the theorem holds true. □

4.4 Analysis of the minimum distance between critical points

Here we show an analogous result to the real case. However, as with many results in the complex case, special care must be given to the distance because unlike the real case, the critical points are not isolated, but they come with the the continuous path that corresponds to the invariance of the Ω-criterion to the continuous change in phase.

Let ζ be a zero of the gradient of Ω, so $F(\zeta, \lambda) = 0$. Further, let η be a unit vector in the orthogonal complement of the subspace spanned by $J\zeta$.

We start with some computations.

Proposition 4.4.1. $\langle F(\zeta + t\eta), \eta\rangle = t[\langle Hess(\zeta, \lambda)\eta, \eta\rangle + t\langle(\Gamma(\eta) + 2\widetilde{\Gamma}(\eta))\zeta, \eta\rangle +$

$t^2\langle\Gamma(\eta)\eta,\eta\rangle]$

Proof. We begin by examining the Gamma term.

$$\Gamma(\zeta+t\eta)=\Gamma(\zeta)+t^2\Gamma(\eta)+2t\Gamma(\zeta,\eta)$$

Therefore, we see that

$$\Gamma(\zeta+t\eta)(\zeta+t\eta)=(\Gamma(\zeta)+t^2\Gamma(\eta)+2t\Gamma(\zeta,\eta))(\zeta+t\eta)$$

$$=\Gamma(\zeta)\zeta+t(2\Gamma(\zeta,\eta)\zeta+\Gamma(\zeta)\eta)+t^2(\Gamma(\eta)\zeta+2\Gamma(\zeta,\eta)\eta)+t^3(\Gamma(\eta)\eta)$$

$$=\Gamma(\zeta)\zeta+t(2\widetilde{\Gamma}(\zeta)\eta+\Gamma(\zeta)\eta)+t^2(\Gamma(\eta)\zeta+2\widetilde{\Gamma}(\eta)\zeta)+t^3(\Gamma(\eta)\eta)$$

Now putting this together in the full gradient, we get

$$F(\zeta+t\eta,\lambda)=\Gamma(\zeta+t\eta)(\zeta+t\eta)+\lambda(\zeta+t\eta)-\Gamma_0(\zeta+t\eta)$$

$$=(\Gamma(\zeta)\zeta+\lambda\zeta-\Gamma_0\zeta)+t(2\widetilde{\Gamma}(\zeta)\eta+\Gamma(\zeta)\eta+\lambda\eta-\Gamma_0\eta)+t^2(\Gamma(\eta)\zeta+2\widetilde{\Gamma}(\eta)\zeta)+t^3(\Gamma(\eta)\eta)$$

$$=F(\zeta,\lambda)+t((2\widetilde{\Gamma}(\zeta)\eta+\Gamma(\zeta)\eta+\lambda\eta-\Gamma_0\eta)+t(\Gamma(\eta)\zeta+2\widetilde{\Gamma}(\eta)\zeta)+t^2(\Gamma(\eta)\eta))$$

$$=F(\zeta,\lambda)+t(Hess(\zeta,\lambda)\eta)+t(\Gamma(\eta)\zeta+2\widetilde{\Gamma}(\eta)\zeta)+t^2(\Gamma(\eta)\eta))$$

Now taking inner product with η, and noting that $F(\zeta,\lambda)=0$, since it is a critical point gives us our result.

□

Theorem 4.4.2. *Let ρ_{crit} denote the minimum distance from ζ, a global minimizer at $\lambda=0$ to the nearest critical point in the orthogonal complement of $J\xi$. Then*

$\rho_{crit} \geq \frac{2}{3}\sqrt{\frac{s_{2n-1}(\widetilde{\Gamma}(\zeta))}{\beta}}$

Proof. Define the polynomial $Q(t) = [\langle Hess(\zeta, \lambda)\eta, \eta\rangle + t\langle(\Gamma(\eta) + 2\widetilde{\Gamma}(\eta))\zeta, \eta\rangle + t^2\langle\Gamma(\eta)\eta, \eta\rangle]$. Therefore we have by Proposition 4.4.1 that

$$\langle F(\zeta + t\eta), \eta\rangle = tQ(t)$$

$Q(t)$ is convex, and positive at $t = 0$ (since ζ is a global minimizer), therefore a lower bound for the distance to the root can be given through the tangent line

$$t_0 \geq \frac{Q(0)}{Q'(0)}$$

Therefore, expanding this out, we get that

$$t_0 \geq \frac{Q(0)}{|Q'(0)|} = \frac{\langle Hess(\zeta, 0)\eta, \eta\rangle}{|\langle\Gamma(\eta) + 2\widetilde{\Gamma}(\eta)\zeta, \eta\rangle|}$$

At a global minumum ζ, $Hess(\zeta, 0) = 2\widetilde{\Gamma}(\zeta)$

Also note that since $(\Gamma(\eta) + 2\widetilde{\Gamma}(\eta))$ is symmetric

$$\langle(\Gamma(\eta) + 2\widetilde{\Gamma}(\eta))\zeta, \eta\rangle = \langle(\Gamma(\eta) + 2\widetilde{\Gamma}(\eta))\eta, \zeta\rangle$$

$$= 3\langle\widetilde{\Gamma}(\eta)\zeta, \eta\rangle$$

Therefore, the expression can be simplified to

$$t_0 \geq \frac{\langle Hess(\zeta,0)\eta,\eta\rangle}{|\langle \Gamma(\eta)+2\widetilde{\Gamma}(\eta)\zeta,\eta\rangle|} = \frac{2\langle\widetilde{\Gamma}(\zeta)\eta,\eta\rangle}{3|\langle\widetilde{\Gamma}(\eta)\zeta,\eta\rangle|}$$

Now let us once more examine the denominator. We will upper bound it in terms of the numerator.

$$|\langle\widetilde{\Gamma}(\eta)\zeta,\eta\rangle| = |\langle\widetilde{\Gamma}^{\frac{1}{2}}(\eta)\zeta,\widetilde{\Gamma}^{\frac{1}{2}}(\eta)\eta\rangle| \leq ||\widetilde{\Gamma}^{\frac{1}{2}}(\eta)\zeta|| \cdot ||\widetilde{\Gamma}^{\frac{1}{2}}(\eta)\eta||$$
$$= \sqrt{\langle\widetilde{\Gamma}(\eta)\eta,\eta\rangle}\sqrt{\langle\widetilde{\Gamma}(\eta)\zeta,\zeta\rangle} = \sqrt{\langle\widetilde{\Gamma}(\eta)\eta,\eta\rangle}\sqrt{\langle\widetilde{\Gamma}(\zeta)\eta,\eta\rangle}$$

Recall from Theorem 4.2.8 that $\langle\widetilde{\Gamma}(\eta)\eta,\eta\rangle \leq \beta||\eta||^2 = \beta$. Therefore we get

$$t_0 \geq \frac{2\langle\widetilde{\Gamma}(\zeta)\eta,\eta\rangle}{3|\langle\widetilde{\Gamma}(\eta)\zeta,\eta\rangle|}$$
$$\geq \frac{2\langle\widetilde{\Gamma}(\zeta)\eta,\eta\rangle}{3\sqrt{\beta}\sqrt{\langle\widetilde{\Gamma}(\zeta)\eta,\eta\rangle}} = \frac{2\sqrt{\langle\widetilde{\Gamma}(\zeta)\eta,\eta\rangle}}{3\sqrt{\beta}}$$

Now if we minimize over all possible η, we get

$$\rho_{crit} \geq \min_{\substack{||\eta||=1 \\ \eta\perp J\zeta}} \frac{2\sqrt{\langle\widetilde{\Gamma}(\zeta)\eta,\eta\rangle}}{3\sqrt{\beta}}$$
$$= \frac{2\sqrt{s_{2n-1}(\widetilde{\Gamma}(\zeta))}}{3\sqrt{\beta}}$$

□

4.5 Complex Convergence Analysis

In this section, we show an analogous result to the convergence result to the real case. Because of the ambiguities present in the complex case, many of the real results don't carry over exactly to the complex case, so other methods need to be employed.

Similar to the real case, we compare the golden retriever's homotopy path to a fixed reference path.

Definition 4.5.0.1. *We call a reference path $\varphi(\lambda)$* ***suitable*** *if it satisfies the following conditions.*

- *It is a smooth path parameterized by λ for $0 \leq \lambda \leq \lambda_1$.*
- *$\varphi(\lambda_1) = 0$, and $\varphi(\lambda)$ is nonzero for $\lambda < \lambda_1$.*
- *$\varphi(0) = z$, a global minimizer.*
- *It is aligned with the Golden Retriever Homotopy Path, in the sense that for each $0 \leq \lambda \leq \lambda_1$, $||\xi(\lambda) - \varphi(\lambda)|| = \min_\theta ||\xi(\lambda) - U(\theta)\varphi(\lambda)||$, where $U(\theta) = cos(\theta)I + sin(\theta)J$.*

First we assume we are given a suitable reference path $\varphi(\lambda)$.

We work towards defining a similar region as the leash in the real case. We begin with a proposition.

Proposition 4.5.1. $F(\xi, \lambda) - F(\varphi, \lambda) = (Hess(\varphi)\delta + \Gamma(\delta)\delta) + (\Gamma(\delta) + 2\tilde{\Gamma}(\delta))\varphi$

Proof. This is by direct computation. We compute

$$F(\xi,\lambda) - F(\varphi,\lambda) = (\Gamma(\xi) + \lambda I - \Gamma_0)\xi - (\Gamma(\varphi) + \lambda I - \Gamma_0)\varphi$$

$$= \Gamma(\varphi+\delta)(\varphi+\delta) + \lambda(\varphi+\delta) - \Gamma_0(\varphi+\delta) - \Gamma(\varphi) - \lambda\varphi + \Gamma_0\varphi$$

By the identities on Γ and $\widetilde{\Gamma}$, we have

$$= \Gamma(\varphi)\varphi + \Gamma(\delta)\varphi + 2\widetilde{\Gamma}(\varphi)\delta + \Gamma(\varphi)\delta + \Gamma(\delta)\delta + 2\widetilde{\Gamma}(\delta)\varphi + \lambda\varphi + \lambda\delta - \Gamma_0\varphi - \Gamma_0\delta$$

$$-\Gamma(\varphi) - \lambda\varphi + \Gamma_0\varphi$$

$$= (\Gamma(\varphi)\delta + 2\widetilde{\Gamma}(\varphi)\delta + \lambda\delta - \Gamma_0\delta) + \Gamma(\delta)\delta + (\Gamma(\delta) + 2\widetilde{\Gamma}(\delta))\varphi$$

$$= Hess(\varphi)\delta + \Gamma(\delta)\delta + (\Gamma(\delta) + 2\widetilde{\Gamma}(\delta))\varphi$$

□

In the same way, by exchanging roles of ξ and φ, we get $F(\varphi,\lambda) - F(\xi,\lambda) = -((Hess(\xi)\delta + \Gamma(\delta)\delta) + (\Gamma(\delta) + 2\tilde{\Gamma}(\delta))\xi)$

Now we are can rearrange to get a bound on the difference.

Therefore, we get

$$||F(\varphi,\lambda) - F(\xi,\lambda)|| = ||((Hess(\xi)\delta + \Gamma(\delta)\delta) + (\Gamma(\delta) + 2\tilde{\Gamma}(\delta))\xi)||$$

Call $s_{2n-1}(\lambda) = \lambda_{2n-1}(Hess(\varphi(\lambda),\lambda))$.

Definition 4.5.1.1. *Given a suitable path $\varphi(\lambda)$, define the following auxiliary con-*

stants

$$A = 27 \cdot 36\beta^2$$

$$B = 4 \cdot 6^3\beta^3||\varphi(\lambda)||^3 + 18 \cdot 36\beta^2||\varphi(\lambda)||s_{2n-1}(\lambda)$$

$$C = 36\beta^2 s_{2n-1}(\lambda)^2||\varphi(\lambda)||^2 + 24\beta s_{2n-1}(\lambda)^3$$

Define the following important constants.

$$\rho_2(\lambda) = \frac{-B + \sqrt{B^2 + 4AC}}{2A} \tag{4.24}$$

$$\rho_1(\lambda) = \frac{\rho_2(\lambda)}{s_{2n-1}(\lambda)} \tag{4.25}$$

$$r(\lambda) = \frac{-12\beta||\varphi(\lambda)|| + \sqrt{144\beta^2||\varphi(\lambda)||^2 + 72\beta s_{2n-1}(\lambda)}}{36\beta} \tag{4.26}$$

Now we state two conditions, which are the complex case analogues to the Initialization Condition and the Gradient Condition in the real case.

The first condition is the Initialization Condition, that the golden retriever path $\xi(\lambda)$ is sufficiently close to $\varphi(\lambda)$.

Condition 4.5.2 (Initialization Condition)**.** *Given a frame set, Γ_0, a suitable reference path $\varphi(\lambda)$, and the golden retriever path $\xi(\lambda)$, we say that φ satisfies the Initialization Condition at a point λ' if $||\xi(\lambda') - \varphi(\lambda')|| < \rho_1(\lambda')$ for $0 < \lambda' < \lambda_1$.*

Condition 4.5.3 (Gradient Condition)**.** *Given a frame set, Γ_0 and a suitable reference path $\varphi(\lambda)$, we say that φ satisfies the Gradient Condition if $||F(\varphi, \lambda)|| < \rho_2(\lambda)$ for all $0 < \lambda < \lambda_1$*

Theorem 4.5.4. *If the Initialization Condition is true for some* $0 < \lambda' < \lambda_1$ *and the Gradient Condition is also true, then* $||\xi(\lambda) - \varphi(\lambda)|| < r(\lambda)$ *for all* $0 \leq \lambda < \lambda'$.

The proof of this theorem revolves around the analysis about a cubic polynomial. Let $\delta(\lambda) = \xi(\lambda) - \varphi(\lambda)$. Define the cubic polynomial

$$Q_3(t) = \lambda_{2n-1}(Hess(\varphi))t - 6\beta t^3 - 6\beta||\varphi||t^2 - \rho_2$$

with $t = ||\delta||$.

Lemma 4.5.5. *If the Gradient Condition is satisfied, then* $Q_3(t) < 0$

Proof. By assumption on the hessian, the smallest eigenvalue of the $Hess(\xi)$ is 0 with an associated eigenspace spanned by $J\xi$. We know that $\xi \perp J\xi$, and by $\varphi(\lambda)$ being suitable, $\varphi \perp J\xi$. Therefore, $\delta = \xi - U\varphi$ also satisfies $\delta \perp J\xi$.

Since δ is in the complement of the eigenspace corresponding to $\lambda_{2n}(Hess(\xi))$, we get

$$||F(\varphi, \lambda) - F(\xi, \lambda)|| \geq \lambda_{2n-1}(Hess(\xi))||\delta|| - ||\Gamma(\delta) + 2\tilde{\Gamma}(\delta)|| \cdot ||\xi||$$

If we use Weyl's inequalities ([29]), and bound $||\xi||$ by $||\xi - \varphi + \varphi|| \leq ||\delta|| + ||\varphi||$, we get

$$||F(\xi, \lambda) - F(\varphi, \lambda)|| \geq \lambda_{2n-1}(Hess(\varphi))||\delta|| - ||Hess(\xi) - Hess(\varphi)|| \cdot ||\delta||$$
$$-||\Gamma(\delta) + 2\tilde{\Gamma}(\delta)|| \cdot (||\delta|| + ||\varphi||)$$

By Corollary 4.2.9, we know $||Hess(\xi) - Hess(\varphi)|| \leq 3\beta||\delta||(||\delta|| + 2||\varphi||)$, we therefore get the cubic equation in $||\delta||$.

Since we know $t = ||\delta||$ and by the Gradient Condition $\rho_2 > ||F(\xi, \lambda) - F(\varphi, \lambda)||$, we get

$$\rho_2 > \lambda_{2n-1} Hess(\varphi)t - 6\beta t^3 - 6\beta||\varphi||t^2$$

We write down the cubic polynomial and we see

$$Q_3(t) = \lambda_{2n-1} Hess(\varphi)t - 6\beta t^3 - 6\beta||\varphi||t^2 - \rho_2 < 0$$

□

Lemma 4.5.6. *Assume* $\rho_2(\lambda) < \frac{-B+\sqrt{B^2+4AC}}{2A}$, *where* $A = 27 \cdot 36\beta^2, B = 4 \cdot 6^3\beta^3||\varphi||^3 + 18 \cdot 36\beta^2||\varphi||s_{2n-1}, C = 36\beta^2 s_{2n-1}^2||\varphi||^2 + 24\beta s_{2n-1}^3$. *Then* $Q_3(t)$ *has* 3 *positive roots.*

Proof. The proof is in the discriminant. Recall that a cubic polynomial has 3 real roots if and only if the discriminant is positive. If we write $Q_3(t) = at^3 + bt^3 + ct + d$, then the discriminant inequality is given by $b^2c^2 - 4ac^3 - 4b^3d - 27a^2d^2 + 18abcd > 0$ [32]

Gathering the terms for $d = \rho_2$, we see that this would be equivalent to having $-A\rho_2^2 - B\rho_2 + C > 0$, with A, B, C as defined above. Solving for the quadratic gives us the desired result. □

Now we use the properties of $Q_3(t)$ to show that if the Initialization Condition

and the Gradient Condition were both true, then the the difference between $\varphi_1(\lambda)$ and $\xi(\lambda)$, which we denoted t, will always be less than $r(\lambda)$. This is similar to the bound the leash provided in the real case, but in this case is a root of a polynomial bounding it away from $r(t)$.

Proof of Theorem 4.5.4. Assume the Gradient Condition holds, which by the previous lemma means that $Q_3(t)$ has 3 roots (and by the shape of it, 2 positive roots).

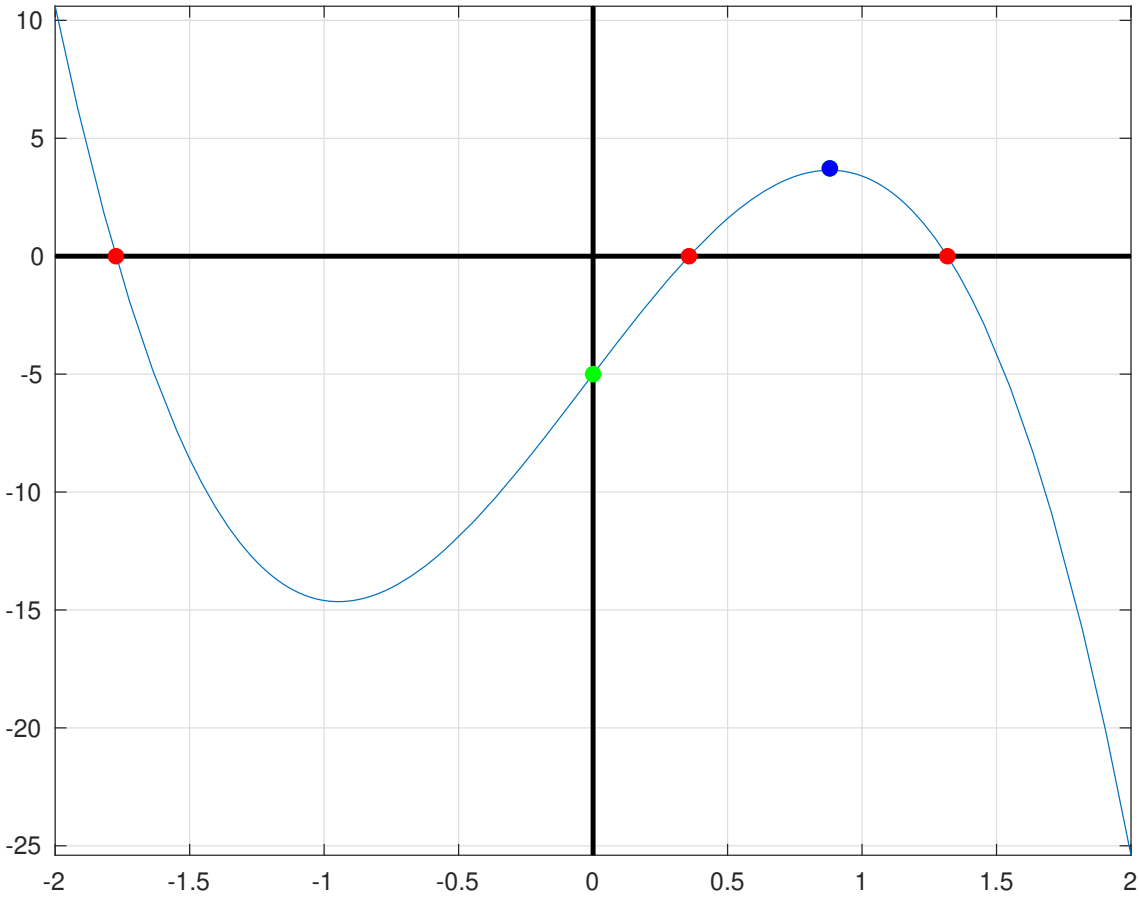

Figure 4.1: $Q_3(t)$ with 3 roots marked in red. The y coordinate of the green point is $-\rho_2(\lambda)$ and the x coordinate of the blue point is $r(\lambda)$

Since the Gradient Condition holds for all $0 \leq \lambda < \lambda_1$, this means that the polynomial will always have 3 roots, for all λ. If the $\delta(\lambda) = t$ is initialized in the region to the left of the first root, since the roots change continuously when the

coefficients of $Q_3(t)$ change continuously, then δ can never go past the first root. This implies that it can never go past the local maximum of the cubic, so we define the coordinate of the local maximum to be $r(\lambda)$. This provides the leash from which the algorithm cannot escape.

First we show that if the Initialization Condition is satisfied for some $0 < \lambda' < \lambda_1$, then $\delta(\lambda')$ is between 0 and the first positive root of $Q_3(t)$. The first positive root of the cubic can be estimated by a tangent line approximation, which is given by $\rho_1(\lambda) = \frac{\rho_2(\lambda)}{s_{2n-1}(\lambda)}$. By convexity, this is an underestimate for the root, so if $||\delta||$ was initialized before ρ_1, then it is initialized before the first positive root of $Q_3(t)$.

Next it is an easy calculation that the vertex between the positive roots of the cubic is given by $r(\lambda)$. □

Now first show that the origin is disjoint from the leash for all $\lambda < \lambda_1$

Theorem 4.5.7. *Let $\varphi(\lambda)$ be any suitable reference path. Then $r(\lambda) < ||\varphi(\lambda)||$, for all $\lambda < \lambda_1$ thus the origin is not contained in the leash for any $\lambda < \lambda_1$.*

Proof. We start with an estimate

$$s_{2n-1}(\lambda) < 3\beta||\varphi(\lambda)||^2$$

To show this, we note that for $s_{2n-1}(\lambda) = \lambda_{2n-1}(\Gamma(\varphi) + 2\widetilde{\Gamma}(\varphi) + \lambda I - \Gamma_0)$, and for $\lambda < \lambda_1$, $\lambda I - \Gamma_0$ is of mixed signature, with repeated eigenvalues, thus $\lambda_{2n-1}(\lambda I 0 \Gamma_0)$ is negative. Thus, treating this as a perturbation, we get that

$$s_{2n-1}(\lambda) \leq ||\Gamma(\varphi) + 2\widetilde{\Gamma}(\varphi)|| \leq 3\beta||\varphi||^2$$

Thus, we have that

$$r(\lambda) = \frac{-12\beta||\varphi(\lambda)|| + \sqrt{144\beta^2||\varphi(\lambda)||^2 + 72\beta s_{2n-1}(\lambda)}}{36\beta} \leq \frac{\sqrt{72\beta s_{2n-1}(\lambda)}}{36\beta}$$

$$\leq \frac{\sqrt{72 \cdot 3\beta^2||\varphi||^2}}{36\beta} = \frac{\sqrt{216}}{36}||\varphi|| < ||\varphi||$$

□

Theorem 4.5.8. *With the notation used above,* $r(0) < \rho_{crit} = \frac{2}{3}\frac{\sqrt{s_{2n-1}}}{\sqrt{\beta}}$*, the critical distance to the nearest critical point.*

Proof. $Q_3'(t) = -18\beta t^2 - 12\beta||\varphi||t + s_{2n-1}(t) = 0$

The positive root is given by $r(\lambda) = \frac{-12\beta||\varphi|| + \sqrt{144\beta^2||\varphi||^2 + 4\cdot 18\beta s_{2n-1}(\lambda)}}{2\cdot 18\cdot\beta}$

Using the identity $\sqrt{A^2 + B} < A + \frac{B}{2A}$ for $A, B > 0$, we get

$$r(\lambda) < \frac{s_{2n-1}(\lambda)}{12\beta||\varphi||}$$

So putting this together with ρ_{crit}, and knowing that $||\varphi(0)|| = ||\zeta||$ since

$$r(\lambda) \leq \frac{s_{2n-1}(0)}{12\beta||\zeta||}$$

We can now compare

$$\frac{s_{2n-1}(0)}{12\beta||\zeta||} \leq \frac{2}{3}\sqrt{\frac{s_{2n-1}(0)}{\beta}}$$

$$s_{2n-1}(0) \leq 8^2\beta||\zeta||^2$$

Since we know $s_{2n-1}(\lambda) < 3\beta||\varphi(\lambda)||^2$, we get that $s_{2n-1}(0) \leq 3\beta||\zeta||^2$, so since $64 > 3$, this inequality is true. □

What we have shown is the following theorem.

Theorem 4.5.9. *If there exists a suitable path which satisfies the Initialization Condition and the Gradient Condition, then the homotopy path converges to a global minimizer ζ.*

So now we define a reference path $\varphi_1(\lambda)$. In the complex case, this is done in two steps.

First we define the parameter $\tau = 1 - \frac{\lambda}{\lambda_1}$. Let η_1 be an normalized eigenvector corresponding to the eigenvalue λ_1 of Γ_0. Then we define $\eta = \left(\sqrt{\frac{\lambda_1}{\langle\Gamma(\eta_1)\eta_1,\eta_1\rangle}}\right)\eta_1$. Then our first choice for a reference path is the equivalent to the real case, $\varphi_0 = \sqrt{\tau}(\tau\zeta + (1-\tau)\eta)$.

The issue is that φ_0 is not a suitable reference path because it is not aligned with $\xi(\lambda)$. To fix this, we define the following.

Definition 4.5.9.1. *Let $U(t)$ be a unitary matrix which aligns the vectors such that $\varphi_1(\lambda) = U(\lambda)\varphi_0(\lambda)$ and $\varphi_1 \perp J\xi(\lambda)$*

Now $\varphi_1(\lambda)$ is a suitable reference path.

We want to show that for $\varphi_1(\lambda)$ the Initialization Condition is satisfied. We begin with an asymptotic analysis of $||\varphi_1(\lambda)||$

Lemma 4.5.10. *For $\varphi_1(\lambda) = U(\lambda)\sqrt{\tau}(\tau\zeta + (1-\tau)\eta)$, there exists a positive constant τ_1 such that for all $0 < \tau < \tau_1$, we have $0.9\tau^{\frac{1}{2}}\sqrt{\frac{\lambda_1}{\langle\Gamma(\eta_1)\eta_1,\eta_1\rangle}} < ||\varphi_1(\lambda)|| < 1.1\tau^{\frac{1}{2}}\sqrt{\frac{\lambda_1}{\langle\Gamma(\eta_1)\eta_1,\eta_1\rangle}}$*

Proof. Note that $||\varphi_1|| = ||\varphi_0||$ since $U(\lambda)$ is unitary. It follows that we can argue, in the same way as the real case

$$||\sqrt{\tau}\tau\zeta+\sqrt{\tau}(1-\tau)\sqrt{\frac{\lambda_1}{\langle\Gamma(\eta_1)\eta_1,\eta_1\rangle}}\eta_1|| = ||\tau^{\frac{1}{2}}\sqrt{\frac{\lambda_1}{\langle\Gamma(\eta_1)\eta_1,\eta_1\rangle}}\eta_1+\tau^{\frac{3}{2}}(\zeta-\sqrt{\frac{\lambda_1}{\langle\Gamma(\eta_1)\eta_1,\eta_1\rangle}}\eta_1)||$$

On one hand this is less than $\tau^{\frac{1}{2}}\sqrt{\frac{\lambda_1}{\langle\Gamma(\eta_1)\eta_1,\eta_1\rangle}}+\tau^{\frac{3}{2}}||\zeta-\sqrt{\frac{\lambda_1}{\langle\Gamma(\eta_1)\eta_1,\eta_1\rangle}}\eta_1|| \leq 1.1\tau^{\frac{1}{2}}\sqrt{\frac{\lambda_1}{\langle\Gamma(\eta_1)\eta_1,\eta_1\rangle}}$, so long as $0.1\tau^{\frac{1}{2}}\sqrt{\frac{\lambda_1}{\langle\Gamma(\eta_1)\eta_1,\eta_1\rangle}} \geq \tau^{\frac{3}{2}}||\zeta - \sqrt{\frac{\lambda_1}{\langle\Gamma(\eta_1)\eta_1,\eta_1\rangle}}\eta_1||$. If ζ is aligned with η_1, this is always true, otherwise, we see that this is true so long as $\tau \leq \frac{0.1\sqrt{\frac{\lambda_1}{\langle\Gamma(\eta_1)\eta_1,\eta_1\rangle}}}{||\zeta-\sqrt{\frac{\lambda_1}{\langle\Gamma(\eta_1)\eta_1,\eta_1\rangle}}\eta_1||}$ On the other hand, $||\varphi_1(\tau)|| \geq \tau^{\frac{1}{2}}\sqrt{\frac{\lambda_1}{\langle\Gamma(\eta_1)\eta_1,\eta_1\rangle}} - \tau^{\frac{3}{2}}||\zeta - \sqrt{\frac{\lambda_1}{\langle\Gamma(\eta_1)\eta_1,\eta_1\rangle}}\eta_1|| \geq 0.9\tau^{\frac{1}{2}}\sqrt{\frac{\lambda_1}{\langle\Gamma(\eta_1)\eta_1,\eta_1\rangle}}$ so long as $0.1\tau\sqrt{\frac{\lambda_1}{\langle\Gamma(\eta_1)\eta_1,\eta_1\rangle}} \geq \tau^{\frac{3}{2}}||\zeta - \sqrt{\frac{\lambda_1}{\langle\Gamma(\eta_1)\eta_1,\eta_1\rangle}}\eta_1||$, which is the same condition as before.

Therefore, for all $0 < \tau < \tau_0 := \frac{0.1\sqrt{\frac{\lambda_1}{\langle\Gamma(\eta_1)\eta_1,\eta_1\rangle}}}{||\zeta-\sqrt{\frac{\lambda_1}{\langle\Gamma(\eta_1)\eta_1,\eta_1\rangle}}\eta_1||}$, we have the desired inequalities. □

Now we show the asymptotic rate of $s_{2n-1}(\lambda)$ as τ approaches 0.

Lemma 4.5.11. *There exists a $\tau_2 > 0$ such that, $s_{2n-1}(\lambda) > \lambda_1\tau$ for all $0 < \tau < \tau_2$*

Proof. Now we compute the Hessian at the point φ_1, but note that

$$Hess(\varphi_1) = Hess(U\varphi_0) = UHess(\varphi_0)U^T$$

and this has the same eigenvalues as the $Hess(\varphi_0)$, so we will just examine the eigenvalues of $Hess(\varphi_0)$.

Note that since Γ_0 has repeated eigenvalues, it's spectrum looks like $\{\lambda_1, \lambda_1, \lambda_2, \lambda_2, ...\lambda_n, \lambda_n\}$. Also, the eigenvectors are related by multiplication by J. Therefore,

$$\Gamma_0 \leq \lambda_2 I_{2n} + (\lambda_1 - \lambda_2)\eta_1\eta_1^T + (\lambda_1 - \lambda_2)(J\eta_1)(J\eta_1)^T$$

Using this identity, we can now look at the Hessian

$$\begin{aligned} Hess(\varphi_0) &= \tau\Gamma(\tau\zeta + (1-\tau)\eta) + 2\tau\widetilde{\Gamma}(\tau\zeta + (1-\tau)\eta) + \lambda_1(1-\tau)I - \Gamma_0 \\ &= \tau\Gamma(\eta) + 2\tau\widetilde{\Gamma}(\eta) + \lambda_1(1-\tau)I - \Gamma_0 + O(\tau^2) \\ &\geq (\lambda_1 - \lambda_2)I - (\lambda_1 - \lambda_2)(\eta_1\eta_1^T + (J\eta_1)(J\eta_1)^T) + \tau(\Gamma(\eta) + 2\widetilde{\Gamma}(\eta) - \lambda_1 I) - O(\tau^2) \end{aligned}$$

Now define the matrix $M = (\lambda_1 - \lambda_2)I - (\lambda_1 - \lambda_2)(\eta_1\eta_1^T + (J\eta_1)(J\eta_1)^T) + \tau(\Gamma(\eta) + 2\widetilde{\Gamma}(\eta) - \lambda_1 I)$

First note that $\langle M\eta_1, \eta_1\rangle = 2\tau\lambda_1 \geq 0$. Second, note that since $\widetilde{\Gamma}(\eta)J\eta_1 = 0$, we get $\langle MJ\eta_1, J\eta_1\rangle = 0$ but this is a direction that we don't need to worry about since we want to find the second smallest eigenvalue and thus will only look at critical points that will be perpendicular to the eigenvector corresponding to the smallest

eigenvalue, which at $\lambda = \lambda_1$ is $J\eta_1$.

Therefore, take a direction x such that $||x|| = 1$ and $\langle x, \eta_1 \rangle = 0$ and $\langle x, J\eta_1 \rangle = 0$. Then

$$\langle Mx, x \rangle = (\lambda_1 - \lambda_2) + \tau(\langle \Gamma(\eta)x, x \rangle + 2\langle \widetilde{\Gamma}(\eta)x, x \rangle - \lambda_1)$$

So for τ sufficiently small, this is positive definite.

Now take $\tilde{x} = cos(\theta)\eta_1 + sin(\theta)x$. Then

$$\langle M\tilde{x}, \tilde{x} \rangle = (\lambda_1 - \lambda_2) - (\lambda_1 - \lambda_2)cos^2(\theta) + \tau[\langle \Gamma(\eta)\tilde{x}, \tilde{x} \rangle + 2\langle \widetilde{\Gamma}(\eta)\tilde{x}, \tilde{x} \rangle - \lambda_1]$$

Now after rearranging terms, and knowing that $\eta = \sqrt{\frac{\lambda_1}{\langle \Gamma(\eta_1)\eta_1, \eta_1 \rangle}}\eta_1$, and setting $\alpha = \frac{\langle \Gamma(\eta_1)x, x \rangle + 2\langle \widetilde{\Gamma}(\eta_1)x, x \rangle}{\langle \Gamma(\eta_1)\eta_1, \eta_1 \rangle}$ and $\gamma = \frac{\langle \Gamma(\eta_1)\eta_1, x \rangle}{\langle \Gamma(\eta_1)\eta_1, \eta_1 \rangle}$ we get

$$\langle M\tilde{x}, \tilde{x} \rangle = (\lambda_1 - \lambda_2)sin^2(\theta) + \tau\lambda_1[3cos^2(\theta) + 6\gamma cos(\theta)sin(\theta) + \alpha sin^2(\theta) - 1]$$

$$= (\lambda_1 - \lambda_2 - 3\tau\lambda_1 + \tau\lambda_1\alpha)sin^2(\theta) + 3\gamma\lambda_1\tau sin(2\theta) + 2\tau\lambda_1$$

$$= -\frac{(\lambda_1 - \lambda_2 - 3\tau\lambda_1 + \tau\lambda_1\alpha)}{2}cos(2\theta) + 3\gamma\lambda_1\tau sin(2\theta) + 2\tau\lambda_1 + \frac{(\lambda_1 + \lambda_2 - 3\tau\lambda_1 + \tau\lambda_1\alpha)}{2}$$

$$\geq \frac{(\lambda_1 - \lambda_2 - 3\tau\lambda_1 + \tau\lambda_1\alpha)}{2} + 2\tau\lambda_1 - \sqrt{(\frac{\lambda_1 - \lambda_2 - 3\tau\lambda_1 + \tau\lambda_1\alpha}{2})^2 + (3\gamma\lambda_1\tau)^2}$$

$$\geq \frac{(\lambda_1 - \lambda_2 - 3\tau\lambda_1 + \tau\lambda_1\alpha)}{2} + 2\tau\lambda_1 - (\frac{\lambda_1 - \lambda_2 - 3\tau\lambda_1 + \tau\lambda_1\alpha}{2}) - \frac{(3\gamma\lambda_1\tau)^2}{(\lambda_1 - \lambda_2 - 3\tau\lambda_1 + \tau\lambda_1\alpha)}$$

$$= 2\tau\lambda_1 - \frac{(3\gamma\lambda_1\tau)^2}{(\lambda_1 - \lambda_2 - 3\tau\lambda_1 + \tau\lambda_1\alpha)}$$

Therefore, we get that for sufficiently small τ, $\lambda_{2n-1}(M) \geq 1.5\lambda_1\tau$ for all small τ. From here, we use Weyl's inequalities ([29]), by taking the Hessian to be a perturbation of M. Thus, if we take $Hess(\varphi_1) = M + R$, we get that since $\lambda_{2n-1}(R) \leq d\tau^2$,

then $\lambda_{2n-1}(Hess(\varphi_1)) \geq \lambda_{2n-1}(M) - d\tau^2 \geq 1.5\lambda_1\tau - d\tau^2 \geq \lambda_1\tau$ for all sufficiently small τ. Thus we get there exists a τ_1 such that $\lambda_{2n-1}(Hess(\varphi_1, \lambda)) = s_{2n-1}(\lambda) \geq \lambda_1\tau$ for all $0 < \tau < \tau_1$. □

Now we can put these lemmas together and find the asymptotic rate of the radius of the leash $r(\lambda)$.

Lemma 4.5.12. *For all λ sufficiently close to λ_1 (i.e. τ sufficiently small) we get* $r(\lambda) > \frac{\lambda_1}{24\beta}\tau^{\frac{1}{2}}$

Proof. Examining the expression for $r(\lambda)$, we use that $\sqrt{x^2+y} \geq x + \frac{y}{2x} - \frac{y^2}{8x^2}$, for $y > 0$, to get that

$$r(\lambda) \geq \frac{1}{36\beta}(3\frac{s_{2n-1}(\lambda)}{||\varphi_1||} - \frac{36}{8}\frac{s_{2n-1}^2(\lambda)}{||\varphi_1||})$$

Now using the bounds that $s_{2n-1}(\lambda) > \lambda_1\tau$ and $||\varphi_1(\lambda)|| < 1.1\sqrt{\frac{\lambda_1}{\langle\Gamma(\eta_1)\eta_1,\eta_1\rangle}}\tau^{\frac{1}{2}}$, we get

$$r(\lambda) > \frac{1}{\beta}(\frac{\lambda_1}{13.2}\tau^{\frac{1}{2}}) - \frac{1}{9.68}\langle\Gamma(\eta_1)\eta_1,\eta_1\rangle\tau$$

Now for sufficiently small τ (where the $\tau^{\frac{1}{2}}$ term dominates the τ term), we get

$$r(\lambda) > \frac{\lambda_1}{24\beta}\tau^{\frac{1}{2}}$$

□

Now that we established an asymptotic lower bound on the radius $r(\lambda)$, we look at the other term in the Intitialization Condition.

Lemma 4.5.13. *There exists a $\tau_2 > 0$ and a positive constant C such that for all $0 < \tau < \tau_2$, $||\xi(\lambda) - \varphi_1(\lambda)|| \leq C\tau^{\frac{3}{2}}$*

Proof. The proof is by decomposing ξ into a component along φ_1 and an orthogonal component. Denote $\varphi_0^{\perp}$ to be an orthogonal component to φ_0 such that $\langle \varphi_0, \varphi_0^{\perp} \rangle = 0$. Then define $\varphi_1^{\perp} = U(\lambda)\varphi_0^{\perp}$ and note that this is orthogonal to φ_1. Now we can decompose $\xi(\lambda) = c\varphi_1 + \tau^{\frac{1}{2}}\varphi_1^{\perp}$. Then we look at the map

$$0 = F(\xi(\lambda), \lambda) = F(c\varphi_1(\lambda) + \varphi_1^{\perp})$$

By factoring out, we get

$$0 = F(c\varphi_1(\lambda) + \varphi_1^{\perp}) = UF(c\varphi_0(\lambda) + \varphi_0^{\perp})U^T$$

. Therefore, we get

$$0 = F(c\varphi_0(\lambda) + \varphi_0^{\perp})$$

Expanding out we get the equation

$$F(c\varphi_0(\lambda) + \varphi_0^{\perp}) = \Gamma(c\varphi_0(\lambda) + \varphi_0^{\perp})(c\varphi_0(\lambda) + \varphi_0^{\perp}) + \lambda(c\varphi_0(\lambda) + \varphi_0^{\perp}) - \Gamma_0(c\varphi_0(\lambda) + \varphi_0^{\perp})$$

We can expand this out and after a little bit of work, we get

$$\frac{1}{\tau^{\frac{1}{2}}}F(c\varphi_0(\lambda)+\varphi_0^{\perp}) = c^3\tau^4\Gamma(\zeta)\zeta + 2c^3\tau^3(1-\tau)\widetilde{\Gamma}(\zeta)\eta + c^3\tau^2(1-\tau)^2\Gamma(\eta)\zeta + c^3\tau^3(1-\tau)\Gamma(\zeta)\eta$$

$$+2c^3\tau^2(1-\tau)^2\widetilde{\Gamma}(\eta)\zeta + c^3\tau(1-\tau)^3\Gamma(\eta)\eta + 2c^2\tau^3\widetilde{\Gamma}(\zeta)\varphi_0^{\perp}$$

$$+2c^2\tau(1-\tau)^2\widetilde{\Gamma}(\eta)\varphi_0^{\perp} + 2c^2\tau^2(1-\tau)\widetilde{\Gamma}(\zeta,\eta)\varphi_0^{\perp} + 2c^2\tau^2(1-\tau)\widetilde{\Gamma}(\eta,\zeta)\varphi_0^{\perp}$$

$$+c\tau^2\Gamma(\varphi_1^{\perp})\zeta + c\tau(1-\tau)\Gamma(\varphi_0^{\perp})\eta + c^2\tau^3\Gamma(\zeta)\varphi_0^{\perp}$$

$$+c^2\tau(1-\tau)^2\Gamma(\eta)\varphi_0^{\perp} + 2c^2\tau^2(1-\tau)\Gamma(\zeta,\eta)\varphi_0^{\perp} + c\tau^2\widetilde{\Gamma}(\varphi_0^{\perp})\zeta + c\tau(1-\tau)\widetilde{\Gamma}(\varphi_0^{\perp})\eta + \tau\Gamma(\varphi_0^{\perp})\varphi_0^{\perp}$$

$$+\lambda_1(1-\tau)c\tau\zeta + \lambda_1(1-\tau)^2c\eta + \lambda_1(1-\tau)\varphi_0^{\perp} - c\tau\Gamma_0\zeta - c(1-\tau)\Gamma_0\eta - \Gamma_0\varphi_0^{\perp}$$

Now let us look at all the τ^0 terms in the expression and simplify them using the fact that $\Gamma_0 = \Gamma(\zeta)$ and η is an eigevector for Γ_0 of eigenvalue λ_1.

$$\lambda_1 c\eta + \lambda_1\varphi_0^{\perp} - c\Gamma_0\eta - \Gamma_0\varphi_0^{\perp}$$

$$= \lambda_1 c\eta + \lambda_1\varphi_0^{\perp} - \lambda_1 c\eta - \Gamma_0\varphi_0^{\perp}$$

$$= \lambda_1\varphi_0^{\perp} - \Gamma_0\varphi_0^{\perp}$$

Now let $\{\eta_1, \eta_2, ..., \eta_{2n-1}, \eta_{2n}\}$ be an eigenbasis for Γ_0, where η_1 is in the direction of η and η_2 is in the direction of $J\eta$. Since

$$\frac{1}{\tau^{\frac{1}{2}}}F(c\varphi_0(\lambda)+\varphi_0^{\perp}) = 0$$

if we label all the terms with a coefficient of τ by $-\tau M$, then we can solve

$$\lambda_1 \varphi_0^\perp - \Gamma_0 \varphi_0^\perp = \tau M$$

Now projecting onto an eigenspace $\{\eta_k\}$ for η_k corresponding to eigenvalues $\lambda_2, ..., \lambda_n$, we get $2n-2$ equations of the form

$$(\lambda_1 - \lambda_k)\langle \varphi_0^\perp, \eta_k \rangle = \tau \langle M, \eta_k \rangle := \tau M_k$$

so we can bound below by the difference with λ_2, and letting $\varphi_0^\perp = ||\varphi_0^\perp|| v$, we get

$$(\lambda_1 - \lambda_2)\langle v, \eta_k \rangle ||\varphi_0^\perp|| \leq \tau M_k$$

So summing the square of both sides, we get

$$(\lambda_1 - \lambda_2)^2 ||\varphi_0^\perp||^2 \sum_{k=3}^{2n} \langle v, \eta_k \rangle^2 \leq \tau^2 M_S$$

, where $M_S = \sum_k M_k$. Therefore , we get

$$||\varphi_0^\perp|| \sqrt{1 - \langle v, \eta_1 \rangle - \langle v, J\eta_1 \rangle} \leq \tau \sqrt{\frac{M_S}{(\lambda_1 - \lambda_2)^2}}$$

This shows that $||\varphi_0^\perp|| = O(\tau)$

Now we want to find the order of $|c-1|$. To do so, let us look at all the terms in the expression with orders less than or equal to τ^1. Denote $-\tau^2 N$ the all terms

with a coefficient of at least τ^2 and note that here we use the fact that $||\varphi_0|| = O(\tau)$ to get

$$c^3\tau\Gamma(\eta)\eta + \lambda_1 c\tau\zeta + \lambda_1 c\eta - 2c\tau\lambda_1\eta + \lambda_1\varphi_0^\perp - c\tau\Gamma_0\zeta - c\Gamma_0\eta + c\tau\Gamma_0\eta - \Gamma_0\varphi_0^\perp = \tau^2 N$$

Simplifying a little bit, we get that

$$c^3\tau\Gamma(\eta)\eta + \lambda_1 c\tau\zeta - c\tau\lambda_1\eta + \lambda_1\varphi_0^\perp - c\tau\Gamma_0\zeta - \Gamma_0\varphi_0^\perp = \tau^2 N$$

Now we take the inner product of the expression with η itself, and we get

$$c^3\tau\langle\Gamma(\eta)\eta,\eta\rangle + \lambda_1 c\tau\langle\zeta,\eta\rangle - c\tau\lambda_1\langle\eta,\eta\rangle + \lambda_1\langle\varphi_0^\perp,\eta\rangle - c\tau\langle\Gamma_0\zeta,\eta\rangle - \langle\Gamma_0\varphi_0^\perp,\eta\rangle = \tau^2\langle N,\eta\rangle := \tau^2 N_\eta$$

Simplifying one using the fact that Γ_0 is symmetric, we get

$$c^3\tau\langle\Gamma(\eta)\eta,\eta\rangle + \lambda_1 c\tau\langle\zeta,\eta\rangle - c\tau\lambda_1\langle\eta,\eta\rangle + \lambda_1\langle\varphi_0^\perp,\eta\rangle - c\tau\langle\Gamma_0\zeta,\eta\rangle - \langle\Gamma_0\varphi_0^\perp,\eta\rangle = \tau^2 N_\eta$$

$$c^3\tau\langle\Gamma(\eta)\eta,\eta\rangle + \lambda_1 c\tau\langle\zeta,\eta\rangle - c\tau\lambda_1\langle\eta,\eta\rangle + \lambda_1\langle\varphi_0^\perp,\eta\rangle - c\tau\lambda_1\langle\zeta,\eta\rangle - \lambda_1\langle\varphi_0^\perp,\eta\rangle = \tau^2 N_\eta$$

$$c^3\tau\langle\Gamma(\eta)\eta,\eta\rangle - c\tau\lambda_1\langle\eta,\eta\rangle = \tau^2 N_\eta$$

Therefore, dividing by τ, we get that

$$c^3\langle\Gamma(\eta)\eta,\eta\rangle - c\lambda_1\langle\eta,\eta\rangle = \tau N_\eta$$

Now substituting $\eta = \sqrt{\frac{\lambda_1}{\langle \Gamma(\eta_1)\eta_1, \eta_1 \rangle}}\eta_1$, we get

$$c^3 \frac{\lambda_1^2}{\langle \Gamma(\eta_1)\eta_1, \eta_1 \rangle} - c\frac{\lambda_1^2}{\langle \Gamma(\eta_1)\eta_1, \eta_1 \rangle} = \tau N_\eta$$

Therefore, we get

$$c^3 - c = \tau \frac{N_\eta \langle \Gamma(\eta_1)\eta_1, \eta_1 \rangle}{\lambda_1^2}$$

So $c^3 - c = O(\tau)$. Therefore, there are three paths, $c = 1, c = -1$ and $c = 0$, which are the three homotopy paths perpendicular to $J\eta_1$. If we, without loss of generality, choose one of the nonzero homotopy paths, say $c = 1$, we get $|c - 1| = O(\tau)$.

Now note that since U is chosen such that ξ is orthogonal to $J\varphi_1$ we can decompose ξ into its components and examine the expression

$$||\xi(\lambda) - \varphi_1(\lambda)|| = ||c\varphi_1 + \tau^{\frac{1}{2}}\varphi_1^{\perp} - \varphi_1|| = ||(c-1)\varphi_1 + \tau^{\frac{1}{2}}\varphi_1^{\perp}||$$

Now using the triangle inequality, we get see that

$$\leq |c-1| \cdot ||\varphi_1(\lambda)|| + \tau^{\frac{1}{2}}||\varphi_1^{\perp}|| \leq C(\tau^{\frac{3}{2}})$$

τ sufficiently small

□

The consequences of the above lemma are immediate.

Theorem 4.5.14. *For all $\tau > 0$ sufficiently small, $||\xi(\lambda) - \varphi_1(\lambda)|| < r(\lambda)$, i.e. $\varphi_1(\lambda)$ satisfies the Initialization Condition.*

Proof. This is just looking at the order, $r(\lambda)$ is bounded below by the order of $\tau^{\frac{1}{2}}$ as $\tau \to 0$ while $||\xi(\lambda) - \varphi_1(\lambda)|| \leq C\tau^{\frac{3}{2}}$. Thus for all sufficiently small τ, we get $||\xi(\lambda) - \varphi_1(\lambda)|| < r(\lambda)$ □

Now we have shown that $\varphi_1(\lambda)$ satisfies the Initialization Condition, we know that if it satisfies the Gradient Condition, i.e. if $||F(\varphi_1, \lambda)|| < \rho_2(\lambda)$ for all $0 < \lambda < \lambda_1$, then the algorithm converges to a global minimizer.

Our next goal is to understand when $\varphi_1(\lambda)$ satisfies the Gradient Condition. As in the real case, we study this probabilistically. The main idea is to realize that in the expected system, η aligns with ζ, so if we treat η as a perturbation of ζ, then we can rewrite the Gradient Condition as a condition on the perturbation. Then we show that for sufficiently high m, the size of the perturbation decreases, and the Gradient Condition is true with high probability.

Thus we define the perturbation $p = \eta - \zeta$. We first rewrite the gradient $F(\varphi_1, \lambda)$ in terms of p (and $\tau = \lambda_1(1 - \frac{\lambda}{\lambda_1})$).

In this part, we will make the following assumptions:

- $\beta > 7$
- $\lambda_{2n}(\Gamma(\zeta)) \geq 3||\zeta||^2$
- $\lambda_{2n-1}(\widetilde{\Gamma}(\zeta)) \geq ||\zeta||^2$

Later we will see that these hold with high probability.

Lemma 4.5.15. *The gradient for $\varphi_0(\lambda)$ can be written as follows*

$$F(\varphi_0,\lambda) = \tau^{\frac{3}{2}}\Big(\tau^3\big(-\Gamma(p)p\big) + \tau^2\big(\Gamma(p)\zeta + 2\widetilde{\Gamma}(p)\zeta + 3\Gamma(p)p\big)$$
$$+\tau\big(\lambda_1 p - \Gamma(\zeta)p - 2\widetilde{\Gamma}(\zeta)p - 2\Gamma(p)\zeta - 4\widetilde{\Gamma}(p)\zeta) - 3\Gamma(p)p\big)$$
$$+\big(-\lambda_1 p + \Gamma(\zeta)p + 2\widetilde{\Gamma}(\zeta)p + \Gamma(p)\zeta + 2\widetilde{\Gamma}(p)\zeta + \Gamma(p)p\big)\Big)$$

Proof. First note that $\varphi_0 = \sqrt{\tau}(\tau\zeta + (1-\tau)\eta)$. For $\eta = \zeta + p$, we get $\varphi_0 = \sqrt{\tau}(\tau\zeta + (1-\tau)\zeta + (1-\tau)p) = \sqrt{\tau}(\zeta + (1-\tau)p)$.

Now if we look at

$$F(\varphi_0,\lambda) = \sqrt{\tau}(\Gamma(\sqrt{\tau}(\zeta + (1-\tau)p)) + \lambda I - \Gamma_0)(\zeta + (1-\tau)p)$$

We can simplify this expression to get

$$F(\varphi_0,\lambda) = \sqrt{\tau}\Big(\tau^4\big(-\Gamma(p)p\big) + \tau^3\big(\Gamma(p)\zeta + 2\widetilde{\Gamma}(p)\zeta + 3\Gamma(p)p\big)$$
$$+\tau^2\big(\lambda_1 p - \Gamma(\zeta)p - 2\widetilde{\Gamma}(\zeta)p - 2\Gamma(p)\zeta - 4\widetilde{\Gamma}(p)\zeta) - 3\Gamma(p)p\big)$$
$$+\tau\big(-\lambda_1 p + \Gamma(\zeta)p + 2\widetilde{\Gamma}(\zeta)p + \Gamma(p)\zeta + 2\widetilde{\Gamma}(p)\zeta + \Gamma(p)p\big)\Big)$$

□

We can use this to get estimates on $||F(\varphi_1,\lambda)||$ because

$$||F(\varphi_1,\lambda)|| = ||U(\lambda)F(\varphi_0,\lambda)|| = ||F(\varphi_0,\lambda)|| \tag{4.27}$$

Lemma 4.5.16. $||F(\varphi_1, \lambda)|| \leq \tau^{\frac{3}{2}}(4\beta||\zeta||^2||p|| + 3\beta||p||^2||\zeta|| + \beta||p||^3)$

Proof. To show this, we note that from the previous lemma, we get

$$F(\varphi_0, \lambda) = \tau^{\frac{3}{2}}\Bigg((\tau-1)\lambda_1 p + (1-\tau)\Gamma(\zeta)p + 2(1-\tau)\widetilde{\Gamma}(\zeta)p$$
$$+(1-2\tau+\tau^2)\Gamma(p)\zeta + 2(1-2\tau+\tau^2)\widetilde{\Gamma}(p)\zeta$$
$$+(1-3\tau+3\tau^2-\tau^3)\Gamma(p)p\Bigg)$$

Thus, we get

$$||F(\varphi_1, \lambda)|| \leq \tau^{\frac{3}{2}}\Bigg(\lambda_1||p|| + 3\beta||\zeta||^2||p|| + 3\beta||p||^2||\zeta|| + \beta||p||^3\Bigg) \tag{4.28}$$

Since $\lambda_1 = \lambda_1(R(z)) \leq \beta||z||^2$, we get

$$||F(\varphi_1, \lambda)|| \leq \tau^{\frac{3}{2}}\Bigg(4\beta||\zeta||^2||p|| + 3\beta||p||^2||\zeta|| + \beta||p||^3\Bigg) \tag{4.29}$$

□

Now that we have bounded $||F(\varphi_1, \lambda)||$ from above, we bound $\rho_2(\lambda)$ from below, to get a sufficient condition for satisfying the Gradient Condition.

We begin with a lemma on $\rho_2(\lambda)$. Recall the definition of ρ_2 with $\rho_2(\lambda) = \frac{-B+\sqrt{B^2+4AC}}{2A}$, where $A = 27 \cdot 36\beta^2, B = 4 \cdot 6^3\beta^3||\varphi||^3 + 18 \cdot 36\beta^2||\varphi||s_{2n-1}, C = 36\beta^2 s_{2n-1}^2(\lambda)||\varphi||^2 + 24\beta s_{2n-1}^3$.

Lemma 4.5.17. *With A, B, C as above, $\rho_2(\lambda) \geq \min\{(\frac{\sqrt{2}-1}{2})\frac{B}{A}, \frac{3C}{4B}\}$*

Proof. Let us look at a general expression $\sqrt{x^2+y^2}$. We will restrict it so x, and y are also positive. We know that $\sqrt{x^2+y^2} \geq \sqrt{2}\min\{x,y\}$. So therefore, we know that $\rho_2 = \frac{-B+\sqrt{B^2+4AC}}{2A} > \frac{(\sqrt{2}-1)}{2}\frac{B}{A}$ if $B < 2\sqrt{AC}$, or in other words if $B^2 < 4AC$. If $B^2 > 4AC$, we need a different bound.

We recall the bound $\sqrt{x^2+y} \geq x + \frac{y}{2x} - \frac{y^2}{8x^2}$. This bound can either be derived from the taylor expansion, or simply checked directly. If we look at what this means for our bound, we get $\rho_2 > \frac{C}{B} - \frac{AC^2}{B^3}$. This bound always holds, but we want to know when this gives us a meaningful bound, so we want $\frac{C}{B} - \frac{AC^2}{B^3} > 0$ which happens if and only if $B^2 > AC$. Therefore for the case $B^2 < 4AC$, we use the first bound, for the case $B^2 > 4AC$, we will use the second, and note that if $B^2 > 4AC$, the second bound can be bounded further $\frac{C}{B} - \frac{AC^2}{B^3} > \frac{C}{B} - \frac{C}{4B} = \frac{3C}{4B}$.

Therefore, we have that $\rho_2 \geq \min\{(\frac{\sqrt{2}-1}{2})\frac{B}{A}, \frac{3C}{4B}\}$ □

Lemma 4.5.18. *Under the assumptions above, also assume* $||p|| < \frac{||\zeta||}{6\beta}$*, then we have* $\rho_2(\lambda) \geq \tau^{\frac{3}{2}}||\zeta||^3\frac{\sqrt{2}-1}{2} \cdot \frac{1}{1215\beta}$

Proof. We begin using a bound on $||\varphi_1(\lambda)|| = \sqrt{\tau}||\zeta + (1-\tau)p||$. By using the triangle inequality and the reverse triangle inequality, we get upper and lower bounds (assuming $||p|| < \frac{||\zeta||}{2}$)

$$\frac{1}{2}\sqrt{\tau}||\zeta|| < ||\varphi_1(\lambda)|| < \frac{3}{2}\sqrt{\tau}||\zeta|| \tag{4.30}$$

Now we want upper and lower bounds on $s_{2n-1}(\lambda)$. We look at the Hessian.

$$Hess(\varphi_0) = \tau\Gamma(\zeta + (1-\tau)p) + 2\tau\widetilde{\Gamma}(\zeta + (1-\tau)p) + \lambda_1(1-\tau)I - \Gamma(\zeta)$$

$$\geq (\tau - 1)\Gamma(\zeta) + 2\tau\widetilde{\Gamma}(\zeta) + \lambda_1(1-\tau)I - \tau O(||p||)$$

Since $\Gamma(\zeta)$ is positive definite, we can bound $(\tau-1)\Gamma(\zeta)$ below by $(\tau-1)\lambda_1 I$ and we get

$$Hess(\varphi_1, \lambda) \geq 2\tau\widetilde{\Gamma}(\zeta) - \tau O(||p||)$$

Now we note $s_{2n}(\lambda) \geq -\tau(||p||)$, because $\widetilde{\Gamma}(\zeta)$ is positive definite. We can therefore bound the sum

$$s_{2n}(\lambda) + s_{2n-1}(\lambda) \geq \lambda_{2n-1}(2\tau\widetilde{\Gamma}(\zeta) - \tau O(||p||)) - \tau O(||p||)$$

Which means that

$$s_{2n-1}(\lambda) \geq 2\tau\lambda_{2n-1}(\widetilde{\Gamma}(\zeta)) - 2\tau O(||p||)$$

Thus, since $s_{2n-1}(\lambda) \geq s_{2n}(\lambda)$, we get $2s_{2n-1}(\lambda) \geq s_{2n-1}(\lambda) + s_{2n}(\lambda) \geq 2\tau\lambda_{2n-1}(\widetilde{\Gamma}(\zeta)) - 2\tau O(||p||)$ so we get

$$s_{2n-1}(\lambda) \geq \tau\lambda_{2n-1}(\widetilde{\Gamma}) - \tau O(||p||) \tag{4.31}$$

So for $||p||$ sufficiently small (and it turns out $||p|| \leq \frac{||\zeta||}{6\beta}$ suffices), we have

$$s_{2n-1}(\lambda) \geq \frac{\tau}{2}\lambda_{2n-1}(\widetilde{\Gamma}) \tag{4.32}$$

On the other hand, we can bound $s_{2n-1}(\lambda)$ above since we know $||Hess(\varphi_1(\lambda), \lambda)|| \leq 3\beta||\varphi_1(\lambda)||^2$, we get

$$s_{2n-1}(\lambda) \leq \frac{81}{4}\tau||\zeta||^2 \tag{4.33}$$

Now we can turn our attention to bounding the quantities in $\rho_2(\lambda)$ using lemma 4.5.17, We see that since $\frac{\sqrt{2}-1}{2} < \frac{3}{4}$, we get

$$\rho_2(\lambda) \geq \frac{\sqrt{2}-1}{2}\min\{\frac{B}{A}, \frac{C}{B}\} \tag{4.34}$$

We analyze each of these cases separately.

We note that we can ignore one of the terms in B and bound using the estimates we found above to get

$$\frac{B}{A} \geq \frac{4 \cdot 6^3 \cdot \beta^3||\varphi_1||^3}{27 \cdot 36\beta^2} \geq \tau^{\frac{3}{2}}\frac{1}{9}||\zeta||^3\beta \tag{4.35}$$

Similarly we can bound

$$\frac{C}{B} \geq \frac{36\beta^2 s_{2n-1}^2||\varphi_1||^2}{4 \cdot 6^3\beta^3||\varphi_1||^3 + 18 \cdot 36\beta^2||\varphi_1||s_{2n-1}}$$

Upperbounding and lowerbounding $s_{2n-1}(\lambda)$ and $||\varphi_1(\lambda)||$ using our estimates as

appropriate gives us

$$\frac{C}{B} \geq \tau^{\frac{3}{2}} ||\zeta||^3 \frac{\lambda_{2n-1}(\widetilde{\Gamma}(\zeta_0))}{729\beta + 486} \tag{4.36}$$

Now note that $\min\{\frac{\beta}{9}, \frac{\lambda_{2n-1}(\widetilde{\Gamma}(\zeta_0))}{729\beta+486}\} = \frac{\lambda_{2n-1}(\widetilde{\Gamma}(\zeta_0))}{729\beta+486}$ and since we are assuming that $\lambda_{2n-1}(\widetilde{\Gamma}(\zeta_0)) > 1$ and $\beta > 7 \Rightarrow 729\beta + 486 \leq 1215\beta$, we get the result. □

From the previous two lemmas, we see that (under our assumptions and assuming $||p|| \leq \frac{1}{6\beta}||\zeta||$, so $||p|| \leq ||\zeta||$)a sufficient condition for satisfying the Gradient Condition under our assumptions is

$$4\beta \frac{||p||}{||\zeta||} + 3\beta(\frac{||p||}{||\zeta||})^2 + \beta(\frac{||p||}{||\zeta||})^3 \leq \frac{\sqrt{2}-1}{2} \cdot \frac{1}{1215\beta} \tag{4.37}$$

Thus a sufficient condition for the Gradient Condition, is

$$||p|| \leq \frac{\sqrt{2}-1}{16} \frac{1}{1215\beta^2} ||\zeta|| \tag{4.38}$$

Since $\sqrt{2} - 1 \geq \frac{4}{10}$, we get

$$||p|| \leq \frac{1}{48600\beta^2} ||\zeta|| \tag{4.39}$$

Now we note that since $\beta > 7$, then this automatically implies that $||p|| \leq ||\zeta||$ and $||p|| \leq \frac{1}{6\beta}||\zeta||$. Thus we define $r_{crit} = \frac{1}{48600\beta^2}||\zeta||$. Thus we see that if $\lambda_{2n-1}(\widetilde{\Gamma}(\zeta_0)) > 1$ is satisfied, $\beta > 7$, and $||p|| < r_{crit}$, then the Gradient Condition is satisfied and the algorithm converges to a global minimizer.

Now we want to use the difference in $||\Gamma(\zeta) - \mathbb{E}(\Gamma(\zeta))||$ to get an upper bound for $||p||$. We work on this in two steps. Since $p = \eta - \zeta$, we first do an estimate

on $||\eta_1 - \zeta_0||$, where $\eta_1 = \frac{\eta}{||\eta||}$ and $\zeta_0 = \frac{\zeta}{||\zeta||}$. Then we work with the normalization terms.

Theorem 4.5.19. $||\eta_1 - \zeta_0|| \leq \frac{2^{\frac{5}{2}}||\Gamma_0 - \mathbb{E}(\Gamma_0)||_{op}}{2||\zeta||^2}$

η_1 is a normalized eigenvector of Γ_0. Similarly, ζ_0 is an eigenvector of $\mathbb{E}(\Gamma_0)$. The result is now a consequence of the famous Davis–Kahan $sin(\Theta)$ theorem. A proof of it can be found in [30].

Theorem 4.5.20. $||p|| \leq ||\zeta|| \cdot ||\eta_1 - \zeta_0||(\frac{\beta}{\langle \Gamma(\eta_1)\eta_1, \eta_1 \rangle} + 1)$

Proof. Define $\eta' = ||\zeta||\eta_1$. Then

$$\begin{aligned} ||p|| &= ||\eta - \zeta|| \cdot ||\eta - \eta' + \eta' - \zeta|| \\ &\leq ||\eta - \eta'|| + ||\eta' - \zeta|| \end{aligned}$$

Now we want to estimate each of these terms. The term $||\eta' - \zeta|| = ||\zeta|| \cdot ||\eta_1 - \zeta_0||$.

For the term $||\eta - \eta'|| = \big| \sqrt{\frac{\lambda_1}{\langle \Gamma(\eta_1)\eta_1, \eta_1 \rangle}} - ||\zeta|| \big|$. We write $\lambda_1 = \langle \Gamma(\zeta)\eta_1, \eta_1 \rangle$ and examine the fraction

$$\begin{aligned} &\frac{\langle \Gamma(\zeta)\eta_1, \eta_1 \rangle}{\langle \Gamma(\eta_1)\eta_1, \eta_1 \rangle} \\ &= ||\zeta||^2 \frac{\langle (\Gamma(\zeta_0) - \Gamma(\eta_1))\eta_1, \eta_1 \rangle + \langle \Gamma(\eta_1)\eta_1, \eta_1 \rangle}{\langle \Gamma(\eta_1)\eta_1, \eta_1 \rangle} \\ &= ||\zeta||^2 \big(\frac{\langle (\Gamma(\zeta_0) - R(\eta_1))\eta_1, \eta_1 \rangle}{\langle \Gamma(\eta_1)\eta_1, \eta_1 \rangle} + 1 \big) \\ &\leq ||\zeta||^2 \big(\frac{||\Gamma(\eta_0) - \Gamma(\eta_1)||}{\langle \Gamma(\eta_1)\eta_1, \eta_1 \rangle} + 1 \big) \end{aligned}$$

Substituting this back into the expression, and using the fact that $\sqrt{1+\epsilon} < 1 + \frac{\epsilon}{2}$

we see that

$$\begin{aligned}||\eta - \eta'|| &\leq ||\zeta|| \cdot \left(\sqrt{\frac{||\Gamma(\zeta_0) - \Gamma(\eta_1)||}{\langle \Gamma(\eta_1)\eta_1, \eta_1\rangle} + 1} - 1\right) \\ &\leq ||\zeta|| \frac{||\Gamma(\zeta_0) - \Gamma(\eta_1)||}{2\langle \Gamma(\eta_1)\eta_1, \eta_1\rangle} \\ &\leq ||\zeta|| \frac{||\Gamma(\zeta_0) - \Gamma(\eta_1)||}{2\langle \Gamma(\eta_1)\eta_1, \eta_1\rangle}\end{aligned}$$

We know that $||\Gamma(\zeta_0) - \Gamma(\eta_1)|| \leq \beta||\eta_1 - \zeta_0|| \cdot ||\eta_1 + \zeta_0|| \leq 2\beta||\eta_1 - \zeta_0||$, so we see that we get $||\Gamma(\zeta_0) - \Gamma(\eta_1)|| \leq 2\beta||\eta_1 - \zeta_0||$. Therefore

$$||\eta - \eta'|| \leq \frac{||\zeta|| \cdot \beta \cdot ||\eta_1 - \zeta_0||}{\langle \Gamma(\eta_1)\eta_1, \eta_1\rangle}$$

Putting it together, we see that

$$\begin{aligned}||p|| &\leq ||\eta - \eta'|| + ||\eta' - \zeta|| \\ &\leq \frac{||\zeta||\beta||\eta_1 - \zeta_0||}{\langle \Gamma(\eta_1)\eta_1, \eta_1\rangle} + ||\zeta|| \cdot ||\eta_1 - \zeta_0|| \\ &= ||\zeta|| \cdot ||\eta_1 - \zeta_0||(\frac{\beta}{\langle \Gamma(\eta_1)\eta_1, \eta_1\rangle} + 1)\end{aligned}$$

□

What this shows is is that a sufficient condition for the Gradient Condition to be true is

$$||\eta_1 - \zeta_0|| \leq \frac{1}{48600\beta^2}(\frac{1}{\frac{\beta}{\langle \Gamma(\eta_1)\eta_1, \eta_1\rangle} + 1}) \tag{4.40}$$

Now we can estimate $||\Gamma(\eta_1) - \Gamma(\zeta_0)|| \leq 2\beta||\eta_1 - \zeta_0||$, so if $||\eta_1 - \zeta_0|| \leq \frac{0.1}{2\beta}$ then

$||\Gamma(\eta_1) - \Gamma(\zeta_0)|| \leq 0.1$, so $\lambda_n(\Gamma(\eta_1)) \geq 0.4$. Therefore, we get

$$||\eta_1 - \zeta_0|| \leq \frac{1}{48600\beta^2}(\frac{0.4}{2\beta}) \tag{4.41}$$

Therefore, assuming $\beta > 7$, $\lambda_{2n}(\Gamma(\zeta_0)) > 3$, $\lambda_{2n-1}(\widetilde{\Gamma}(\zeta_0)) > 0.5$ we get a sufficient condition for convergence is

$$||\eta_1 - \zeta_0|| \leq \frac{1}{243000\beta^3} \tag{4.42}$$

Combining this with Theorem 4.5.19 gives the sufficient condition for convergence

$$||\Gamma(\zeta_0) - \mathbb{E}\Gamma(\zeta_0)|| \leq \frac{1}{2^{\frac{3}{2}} \cdot 243000\beta^3} \tag{4.43}$$

which gives us a sufficient condition of

$$||\Gamma(\zeta_0) - \mathbb{E}\Gamma(\zeta_0)|| \leq \frac{1}{687308\beta^3} \tag{4.44}$$

Now we want to know when the assumptions are satisfied. This is given to us by the following lemma.

Lemma 4.5.21. *Let $C(0.1)$ be an upper bound and γ be a universal bound as defined in the Concentration Theorem. Then for $m \geq C(0.1)nlog(n)$, we have $\beta > 7.9$ and $\lambda_{2n}(\Gamma(\zeta)) \geq 3.9||\zeta||^2$ with probability $1 - \frac{13}{n^2} - 10e^{-\gamma n}$.*

Proof. By the concentration of expectation, there exists a $C > 0$ such for $m \geq Cnlog(n)$, $||\Gamma(e) - \mathbb{E}[\Gamma(e)]|| \leq 0.1$. Since $\lambda_1(\mathbb{E}[\Gamma(e)]) = 8$, we get that $\beta \geq$

$\lambda_1(\Gamma(\zeta_0)) \geq 7.9$. Similarly, $\lambda_{2n}(\mathbb{E}[\Gamma(e)]) = 4$, so $\lambda_{2n}(\Gamma(\zeta)) = ||\zeta||^2\lambda_{2n}(\Gamma(\frac{\zeta}{||\zeta||})) \geq ||\zeta||^2(4-0.1) = 3.9||\zeta||^2$. □

We will often use the bounds $\beta > 7$ and $\lambda_{2n}(\Gamma(\zeta)) > 3||\zeta||^2$, and $\lambda_{2n-1}(\widetilde{\Gamma}(\zeta)) > ||\zeta||^2$

Theorem 4.5.22 (Sufficient Convergence Result)**.** *Let* $\delta_T = \min\{0.1, \frac{1}{687308\beta^3}\}$ *Then if* $||\Gamma(\zeta_0) - \mathbb{E}[\Gamma(\zeta_0)]|| \leq \delta_T$ *and* $\lambda_{2n-1}(\widetilde{\Gamma}(\zeta)) > ||\zeta||^2$ *then the algorithm converges to a global minimizer.*

Proof. Since the algorithm converges to a global minimizer if $||p|| \leq r_{crit}$, or equivalently equation 4.44, we see $||\Gamma(\zeta_0) - \mathbb{E}[\Gamma(\zeta_0)]|| \leq \delta_T$ implies both the assumptions and equation 4.44. Thus this is a sufficient condition for the Gradient Condition to hold. □

Theorem 4.5.23. *Assume we are in the noiseless case, where* f_k *are drawn from a complex standard normal. Fix a nonzero* $\zeta \in \mathbb{R}^{2n}$ *to be the realification of the generating signal. Choose a universal constant* $\gamma > log(9)$*. Assume there are a sufficiently high number of samples. That means that* $m \geq \max\{Cnlogn, 64n^3\}$*. Then the algorithm converges to a global minimizer with probability at least* $1 - \frac{13}{n^2} - 10e^{-\gamma n} - \left(2n^3+1\right)e^{-\frac{3n}{5}}$

Proof. Let $\delta = \frac{\delta_T}{||\zeta||^2}$ be as above. Take $C = C(\delta, \gamma)$ from the concentration theorem, Theorem 4.3.9. Now we apply this to Theorem 4.5.22. We also need $\lambda_{2n-1}(\widetilde{\Gamma}(\zeta)) > ||\zeta||^2$, but this follows from a equivalent Concentration Theorem on $\widetilde{\Gamma}(\zeta)$ as Theorem 4.3.9 for $\Gamma(\zeta)$. A proof of it can be seen as a concentration of the Hessian in

[23]. Note that the $\widetilde{\Gamma}(\zeta)$ concentration implies the concentration of $\Gamma(\zeta)$ (since $\Gamma(\zeta) = \widetilde{\Gamma}(\zeta) + \widetilde{\Gamma}(J\zeta)$). Now the last step is give bounds on β, specifically we can show that if $m = O(n^3)$, then $\beta < M$ with high probability. Now it follows that $\beta < M$ with the same probability argument that $b_0 < M$ in the real case, but n becomes $2n$ from the $\mathcal{N} - net$ argument, and m becomes $2m$ from considering the real and imaginary parts. □

4.6 Following the Retriever: Complex Certifier

In this section, we give a numerical certificate that can be checked at each step to certify that the next point in the algorithm is on the same path as the previous point. This gives as adaptive step size which guarantees one is following the correct path.

The idea behind the proof is to look at a cross section with one of the coordinate directions and find an upper bound for how far the distance the path can go in a single step, and then make sure there is no other critical point that is within the upper bound's distance.

To start, say we begin at the point $X_0 = \begin{bmatrix} \xi_0 \\ \\ \lambda_1 \end{bmatrix}$ and we move to a new point along column c, parameterized by t, such that $X_c(t) - X_{c,0} = t$.

Let $D = ||X(t) - X_0||$. Note that

$$D\frac{dD}{dt} = \frac{1}{2}\frac{d}{dt}D^2 = \langle \frac{dX}{dt}, X(t) - X_0 \rangle$$

Since we assume the new point is on the gradient path, it follows that since

$$\nabla_\xi \Omega(\xi, \lambda) = \Gamma(\xi)\xi + \lambda S\xi - \Gamma_0 \xi = 0$$

then taking the derivative with respect to t, we get

$$Hess(\xi, \lambda)\frac{d\xi}{dt} + S\xi\frac{d\lambda}{dt} = 0 \tag{4.45}$$

We also impose two more condition on the derivative, which is that $\frac{d\xi}{dt} \perp J\xi$ and $\frac{dX_c(t)}{dt} = 1$.

If we define the $(2n+1) \times (2n+1)$ orthogonal extended hessian matrix by

$$H_{ext}(\xi, \lambda) = \begin{bmatrix} Hess(\xi, \lambda) & S\xi \\ (J\xi)^T & 0 \end{bmatrix} \tag{4.46}$$

Now we note that the conditions given above are imply

$$H_{ext}(\xi, \lambda) \begin{bmatrix} \frac{d\xi}{dt} \\ \\ \frac{d\lambda}{dt} \end{bmatrix} = H_{ext}(\xi, \lambda)\frac{dX}{dt} = 0$$

Furthermore, define $H_{red:c}$ to be the $(2n+1) \times 2n$ matrix gotten by removing column c from H_{ext}, call that column q. Furthermore, define $X_{red:c}(t)$ to be $X(t)$ after removing row c.

Lemma 4.6.1. *With the definitions above, we can bound* $|\frac{dD}{dt}| \leq \frac{||q||}{s_{2n}(H_{red:c})} + 1$

Proof. From the way that c is chosen, it follows that $rank(H_{red:c}) = 2n$.

From the equation above, and the condition on the derivative, we get that

$$H_{red:c}\frac{dX_{red:c}}{dt} + 1 \cdot q = 0 \tag{4.47}$$

Since q is in the column space of $H_{red:c}$, we can project everything onto the columns of $H_{red:c}$ and get

$$H_{red:c}^T H_{red:c}\frac{dX_{red:c}}{dt} = -H_{red:c}^T q \tag{4.48}$$

Therefore, we get that

$$\frac{dX_{red:c}}{dt} = -(H_{red:c}^T H_{red:c})^{-1} H_{red:c}^T q = -H_{red:c}^{\dagger} q \tag{4.49}$$

Where $H_{red:c}^{\dagger}$ is the pseudoinverse of $H_{red:c}$;

The pseudoinverse satisfies the following equality for v

$$H_{red:c}v + q = 0 \tag{4.50}$$

and we know that the equation has a solution because q is in the column space of $H_{red:c}$

Therefore, since the rank is $2n$, we get that

$$s_{2n}(H_{red:c})||v|| \leq ||q|| \tag{4.51}$$

from which we get that

$$||H^{\dagger}_{red:c}q|| \leq \frac{||q||}{s_{2n}(H_{red:c})} \tag{4.52}$$

Therefore, we get that

$$\begin{aligned} D|\frac{dD(t)}{dt}| &= |\frac{1}{2}\frac{d}{dt}D^2| = |\langle \frac{dX}{dt}, X(t) - X_0\rangle| \\ &= |\langle \frac{dX_{red:c}}{dt}, X_{red:c}(t) - X_{red:c,0}\rangle + t| \\ &\leq |\langle -H^{\dagger}_{red:c}q, X_{red:c}(t) - X_{red:c,0}\rangle| + |t| \\ &\leq ||H^{\dagger}_{red:c}q|| \cdot ||X_{red:c}(t) - X_{red:c,0}|| + |t| \\ &\leq \frac{||q||}{s_{2n}(H_{red:c})}D + D \end{aligned}$$

Dividing by D gives us the desired result.

□

Now we follow the steps for the real certifier.

Lemma 4.6.2. *Assume* $||H_{ext} - H_{ext,0}|| \leq \frac{s_{min}(H_{red:c,0})}{2}$. *Then* $D(t) \leq (2+2\frac{||H_{ext,0}||}{s_{2n}(H_{red:c,0})})t$

Proof. First we note find some upper bounds on $||q||$.

By Weyl's inequalities ([29]), we know

$$||q|| \leq ||H_{ext}||op \leq ||H_{ext} - H_{ext,0}|| + ||H_{ext,0}|| \tag{4.53}$$

Therefore, by our assumptions, we have that

$$||q|| \leq \frac{s_{min}(H_{red:c,0})}{2} + ||H_{ext,0}|| \tag{4.54}$$

Also by Weyl's inequalities, we have that

$$s_{2n}(H_{red:c}) \geq s_{2n}(H_{red:c,0}) - ||H_{red:c} - H_{red:c,0}|| \tag{4.55}$$

Since

$$||H_{red:c} - H_{red:c,0}|| \leq ||H_{ext} - H_{ext,0}|| \tag{4.56}$$

We get

$$s_{2n}(H_{red:c}) \geq s_{2n}(H_{red:c,0}) - ||H_{ext} - H_{ext,0}|| \geq s_{2n}(H_{red:c,0}) - \frac{s_{2n}(H_{red:c,0})}{2} = \frac{s_{2n}(H_{red:c,0})}{2}$$

Putting these together, we get that

$$\begin{aligned} |\frac{dD}{dt}| \leq \frac{||q||}{s_{2n}(H_{red:c})} + 1 &\leq \frac{2(\frac{s_{min}(H_{red:c,0})}{2} + ||H_{ext,0}||) + 1}{s_{min}(H_{red:c,0})} + 1 \\ &= (2 + 2\frac{||H_{ext,0}||}{s_{2n}(H_{red:c})}) \end{aligned}$$

Now if we examine the integral

$$|\int_0^T \frac{dD}{dt} dt| \leq \int_0^T (2 + 2\frac{||H_{ext,0}||}{s_{2n}(H_{red:c,0})}) dt = (2 + 2\frac{||H_{ext,0}||}{s_{2n}(H_{red:c,0})}) T$$

On the other hand, we can evaluate it directly and we get

$$|\int_0^T \frac{dD}{dt} dt| = |\int_0^{D(T)} 1 dD| = D(T)$$

Therefore, as desired, we get that

$$D(t) \leq (2 + 2\frac{||H_{ext,0}||}{s_{2n}(H_{red:c,0})})t \tag{4.57}$$

□

Now we want to provide a condition on t for when the assumption $||H_{ext} - H_{ext,0}|| \leq \frac{s_{2n}(H_{red:c,0})}{2}$ is true. Define $A = (2 + 2\frac{||H_{ext,0}||}{s_{2n}(H_{red:c,0})})$, and

$$t_+ = -(\frac{6\beta||\xi_0|| + ||S|| + 1}{6\beta A}) + \sqrt{(\frac{6\beta||\xi_0|| + ||S|| + 1}{6\beta A})^2 + \frac{s_{2n}(H_{red:c,0})}{6\beta A^2}} \tag{4.58}$$

Lemma 4.6.3. *For $t < t_+$, $||H_{ext} - H_{ext,0}|| \leq \frac{s_{2n}(H_{red:c,0})}{2}$*

Proof. Let us first examine $||H_{ext} - H_{ext,0}||$. We know that

$$||H_{ext} - H_{ext,0}|| = || \begin{bmatrix} Hess(\xi(t), \lambda(t)) - Hess(\xi_0, \lambda_0) & S(\xi(t) - \xi_0) \\ (J(\xi(t) - \xi_0))^T & 0 \end{bmatrix} ||$$

$$\leq ||Hess(\xi(t), \lambda(t)) - Hess(\xi_0, \lambda_0)|| + ||S|| \cdot ||\xi(t) - \xi_0|| + ||\xi(t) - \xi_0||$$

Now we can use the bound on the difference of the Hessians (derived in Corollary 4.2.9) to bound

$$||Hess(\xi(t), \lambda(t)) - Hess(\xi_0, \lambda_0)|| \leq 3\beta(2||\xi_0|| + D)D \tag{4.59}$$

This gives us a upperbound on the difference of the extended Hessians

$$||H_{ext} - H_{ext,0}|| \leq 3\beta(2||\xi_0|| + D)D + (||S|| + 1)D \tag{4.60}$$

So we want a value of t for which

$$3\beta(2||\xi_0|| + D)D + (||S|| + 1)D \leq \frac{s_{2n}(H_{red:c})}{2} \tag{4.61}$$

Substituting the bound for $D(t) = At$, for $A = (2 + 2\frac{||H_{ext,0}||}{s_{2n}(H_{red:c,0})})$ we get that

$$\begin{aligned} D^2 + (\frac{6\beta||\xi_0|| + ||S|| + 1}{3\beta})D - \frac{s_{2n}(H_{red:c,0})}{6\beta} \leq 0 \\ A^2t^2 + (\frac{6\beta||\xi_0|| + ||S|| + 1}{3\beta})At - \frac{s_{2n}(H_{red:c,0})}{6\beta} \leq 0 \\ = t^2 + (\frac{6\beta||\xi_0|| + ||S|| + 1}{3\beta A})t - \frac{s_{2n}(H_{red:c,0})}{6\beta A^2} \leq 0 \end{aligned}$$

Therefore, for t less than the positive root of this quadratic gives us the bound we would like. Therefore, define t_+ to be the first root, so

$$t_+ = -(\frac{6\beta||\xi_0|| + ||S|| + 1}{6\beta A}) + \sqrt{(\frac{6\beta||\xi_0|| + ||S|| + 1}{6\beta A})^2 + \frac{s_{2n}(H_{red:c,0})}{6\beta A^2}} \tag{4.62}$$

and for $t < t_+$, the condition is satisfied. To justify the substitution, we note that the same proof used in the real case also works here. □

What we have shown so far is that for $t < t_+$, $D(t) \leq (2 + 2\frac{||H_{ext,0}||}{s_{2n}(H_{red:c,0})})t$. Now that we have found an upper bound on the distance, we need to make sure there is

no other critical point within this distance. The following theorem will be useful in showing this.

Theorem 4.6.4. *Let $X = (a, \lambda_a)$ be a critical point of $\Omega(x, \lambda)$, set $Hess_{ext}(a, \lambda_a) = \begin{bmatrix} Hess(a, \lambda_a) & Sa \end{bmatrix}$ and define*

$$\rho(a, \lambda_a) = \min(\frac{1}{2}, \frac{s_{2n-1}(Hess_{ext}(a, \lambda_a))}{\sqrt{2n+1}(\beta||a||^3 + 3\beta||b||^2||a|| + ||S||)}) \tag{4.63}$$

Then there is no other critical point X_1 such that $X_1(c) = X(c)$ and $||X_1(c) - X(c)|| \leq \rho$. In other words, ρ serves as a lower bound for a distance to the nearest critical point on the same $X_1(c) = X(c)$ hyperplane.

Proof. Let (a, λ_a) be a critical point, and let (b, λ_b) be a unit vector such that $(b, \lambda_b)_c = 0$, and that $(b, \lambda_b) \perp (Ja, 0)$. Assume that $(a, \lambda_a) + r(b, \lambda_b)$ is another critical point, for some scalar r

We first expand out

$$F(a + tb, \lambda_a + t\lambda_b)$$

Standard computations show that

$$F(a + tb, \lambda_a + r\lambda_b) = F(a, \lambda_a) + r(\begin{bmatrix} Hess(a, \lambda_a) & Sa \end{bmatrix} \begin{bmatrix} b \\ \\ \lambda_b \end{bmatrix})$$

$$+r^2(\Gamma(b)a + 2\widetilde{\Gamma}(b)a) + \lambda_b Sb + r^3(\Gamma(a)a)$$

Since (a, λ_a) is a critical point, $F(a, \lambda_a) = 0$, so

$$0 = F(a+rb, \lambda_a+r\lambda_b) = r(\begin{bmatrix} Hess(a,\lambda_a) & Sa \end{bmatrix} \begin{bmatrix} b \\ \\ \lambda_b \end{bmatrix}) + r^2(\Gamma(b)a + 2\widetilde{\Gamma}(b)a) + \lambda_b Sb + r^3(\Gamma(a)a)$$

Let $rG(r) = ||F(a + rb, \lambda_a + r\lambda_b)|| = 0$, and we want to find the smallest nonzero value for r.

Define $v_1 = \Gamma(a)a, \quad v_2 = (\Gamma(b)+2\widetilde{\Gamma}(b))a+\lambda_b Sb$ and $\quad v_3 = \begin{bmatrix} Hess(a,\lambda_a) & Sa \end{bmatrix} \begin{bmatrix} b \\ \\ \lambda_b \end{bmatrix}$

Now let's get estimates on each. We have

$$||v_1|| = ||\Gamma(a)a|| \leq ||a||^3 ||\Gamma(\frac{a}{||a||})|| \leq \beta ||a||^3 \tag{4.64}$$

Similarly we get

$$||v_2|| = (\Gamma(b) + 2\widetilde{\Gamma}(b))a + \lambda_b Sb|| \leq ||(\Gamma(b) + 2\widetilde{\Gamma}(b))a|| + ||S|| \leq 3\beta||a|| \cdot ||b||^2 + ||S|| \tag{4.65}$$

Lastly, we want to show that

$$||v_3|| \geq \frac{s_{2n-1}(Hess_{ext})}{\sqrt{2n+1}} \tag{4.66}$$

To do so, note that

$$||v_3||^2 = ||(Hess_{ext}(a, \lambda_a) \begin{bmatrix} b \\ \\ \lambda_b \end{bmatrix})||^2$$

$$= \langle (Hess_{ext}(a, \lambda_a) \begin{bmatrix} b \\ \\ \lambda_b \end{bmatrix}), (Hess_{ext}(a, \lambda_a) \begin{bmatrix} b \\ \\ \lambda_b \end{bmatrix}) \rangle$$

Now let $v = \begin{bmatrix} \frac{d\xi}{dt} \\ \\ \frac{d\lambda}{dt} \end{bmatrix}$ be the null vector of the extended Hessian, and decompose $\begin{bmatrix} b \\ \\ \lambda_b \end{bmatrix} = c_1 v + w$, where $w \in span(v)^{\perp}$.

$$||v_3||^2 = \langle (Hess_{ext}(a, \lambda_a) \begin{bmatrix} b \\ \\ \lambda_b \end{bmatrix}), (Hess_{ext}(a, \lambda_a) \begin{bmatrix} b \\ \\ \lambda_b \end{bmatrix}) \rangle$$

$$= \langle (Hess_{ext}(a, \lambda_a) w), (Hess_{ext}(a, \lambda_a) w \rangle$$

$$= \langle (Hess_{ext}(a, \lambda_a)^T)(Hess_{ext}(a, \lambda_a) w, w \rangle$$

$$= \lambda_{2n-1}(Hess_{ext}(a, \lambda_a)^T Hess_{ext}(a, \lambda_a))||w||^2$$

$$\geq s^2_{2n-1}(Hess_{ext}(a, \lambda_a)) \min_{||e||=1, e_c=0} ||proj_{span(v^{\perp})}(b, \lambda_b)||^2$$

Note that the null vector v here is normalized so that $v_c = 1$, as c is chosen so that v_c it is the largest component.

To estimate this, define $\tilde{e}$ to be e without the c'th component, and $\tilde{v}$ as v without the c'th component. Now note we are trying to minimize the projection

onto the complement of v, so we get

$$\begin{aligned}
\min_{||e||=1, e_{c=0}} ||e - \frac{\langle e, v\rangle}{||v||^2} v||^2 &= ||\tilde{e} - \frac{\langle \tilde{e}, \tilde{v}\rangle}{1+||\tilde{v}||^2}\tilde{v}||^2 + (\frac{\langle \tilde{e}, \tilde{v}\rangle}{1+||\tilde{v}||^2} v_c)^2 \\
&\leq \langle \tilde{e} - \frac{\langle \tilde{e}, \tilde{v}\rangle}{1+||\tilde{v}||^2}\tilde{v}, \tilde{e} - \frac{\langle \tilde{e}, \tilde{v}\rangle}{1+||\tilde{v}||^2}\tilde{v}\rangle + (\frac{\langle \tilde{e}, \tilde{v}\rangle}{1+||\tilde{v}||^2})^2 \\
&= 1 - 2\frac{\langle \tilde{e}, \tilde{v}\rangle^2}{1+||\tilde{v}||^2} + \frac{\langle \tilde{e}, \tilde{v}\rangle^2}{(1+||\tilde{v}||^2)^2}||\tilde{v}||^2 + (\frac{\langle \tilde{e}, \tilde{v}\rangle}{1+||\tilde{v}||^2})^2 \\
&= 1 + \frac{\langle \tilde{e}, \tilde{v}\rangle^2}{1+||\tilde{v}||^2} - 2\frac{\langle \tilde{e}, \tilde{v}\rangle^2}{1+||\tilde{v}||^2} \\
&= 1 - \frac{\langle \tilde{e}, \tilde{v}\rangle^2}{1+||\tilde{v}||^2} \\
&\geq 1 - \frac{||\tilde{v}||^2}{1+||\tilde{v}||^2} \quad \text{by Cauchy Schwarz} \\
= \frac{1}{||\tilde{v}||^2+1} \geq \frac{1}{2n+1} &\quad \text{since 1 is the largest component in a size } 2n \text{ vector}
\end{aligned}$$

Therefore, we see that

$$||v_3|| \geq s_{2n-1}(Hess_{ext}(a, \lambda_a)) \cdot \frac{1}{\sqrt{2n+1}} \tag{4.67}$$

Using these three bounds, we want to estimate the nearest root of

$$||v_3|| - |r|||v_2|| - |r|^2||v_1|| \tag{4.68}$$

If the root is farther than $\frac{1}{2}$ (so $r > \frac{1}{2}$), then $\frac{1}{2}$ is a lower bound on the root. Otherwise, the root is closer than $\frac{1}{2}$, so the slope of the function is dominated by the slope at $\frac{1}{2}$ (since it is a quadratic with negative slope at 0). The slope at $\frac{1}{2}$ is given by $M = -||v_1|| - ||v_2||$.

Now if we go back and estimate the zero of the line passing through $(0, a)$ with slope $M = -||v_1|| - ||v_2||$, we get that the root is at $r_{root} = \frac{||v_3||}{||v_1||+||v_2||}$.

Therefore $\frac{||v_3||}{||v_1||+||v_2||}$ is a bound on the closest root.

Now since $0 = G(r) \geq ||v_1|| - |r| \cdot ||v_2|| - |r|^2||v_3||$, so we know that $r \geq min(\frac{1}{2}, \frac{||v_3||}{||v_1||+||v_2||})$, so from our bounds we get

$$\frac{||v_3||}{||v_1|| + ||v_2||} \geq \frac{s_{2n-1}(Hess_{ext})}{\sqrt{2n+1}(\beta||a||^3 + 3\beta||b||^2||a|| + ||S||)} \tag{4.69}$$

So we got the desired bounds.

□

Theorem 4.6.5. *Assume our algorithm starts at a point $(\xi_{old}, \lambda_{old})$ which is a critical point. Let $(\xi_{new}, \lambda_{new})$ be a new point the algorithm decides and $(\xi_{other}, \lambda_{other})$ be any other critical point. Let D_1 denote the distance from $(\xi_{old}, \lambda_{old})$ to $(\xi_{new}, \lambda_{new})$ and D_2 denote the distance from $(\xi_{old}, \lambda_{old})$ to $(\xi_{other}, \lambda_{other})$. Let $t_1 = \frac{\rho(\xi_{new}, \lambda_{new})}{2(2+2\frac{||H_{ext,0}||}{s_{2n}(H_{red:c,0})})}$ and $t_{max} = min(t_1, t_+)$, and $UB(t) = (2 + 2\frac{||Hess_{ext,0}||}{s_{2n}(H_{red:c,0})})t$, and $\rho(\cdot)$ be the expression defined in the previous theorem.*

Assume the following two conditions are satisfied:

1. $D_1 < \frac{\rho(\xi_{new}, \lambda_{new})}{2}$

2. $t < t_{max}$

Then $(\xi_{new}, \lambda_{new})$ is the point connected on the continuous path defined by the zero of the gradient passing through $(\xi_{old}, \lambda_{old})$.

Proof. Assume $(\xi_{other}, \lambda_{other})$ is a critical point on the path instead of $(\xi_{new}, \lambda_{new})$. Then, since $t < t_{max} < t_+$, we have that

$$||(\xi_{other}, \lambda_{other}) - (\xi_{old}, \lambda_{old})|| \leq UB(t) \tag{4.70}$$

Since $t < t_1$, we get that $UB(t) < \frac{\rho(\xi_{new}, \lambda_{new})}{2}$, so

$$||(\xi_{other}, \lambda_{other}) - (\xi_{old}, \lambda_{old})|| \leq \frac{\rho(\xi_{new}, \lambda_{new})}{2} \tag{4.71}$$

Also, since $D_1 < \frac{\rho(\xi_{new}, \lambda_{new})}{2}$ we get that

$$\begin{aligned} ||(\xi_{new}, \lambda_{new}) - (\xi_{other}, \lambda_{other})|| &\leq ||(\xi_{new}, \lambda_{new}) - (\xi_{old}, \lambda_{old})|| + ||(\xi_{old}, \lambda_{old}) - (\xi_{other}, \lambda_{other})|| \\ &< \frac{\rho(\xi_{new}, \lambda_{new})}{2} + \frac{\rho(\xi_{new}, \lambda_{new})}{2} = \rho(\xi_{new}, \lambda_{new}) \end{aligned}$$

but this is a contradiction, because any other critical point must be a distance further that $\rho(\xi_{new}, \lambda_{new})$ away from $(\xi_{new}, \lambda_{new})$.

Therefore, the only possible point on this level set that is the point passing through the continuous path is $(\xi_{new}, \lambda_{new})$. □

4.7 Oracle Convergence

Given an Oracle, we can ask the following question, equivalent to the question in the real case: Does there exist a positive hermitian semidefinite matrix Q such that the Golden Retriever algorithm converges to the exact solution?

The answer, again, is yes. As before, we can make the algorithm converge to any critical point we want.

Lemma 4.7.1. *Let v be a critical point of $\Omega(\xi, \lambda)$ and let Γ_0 be the matrix in the condition, there exists a positive definite hermitian matrix Q with realification $S \neq \Gamma_0$ satisfying the following properties:*

- $Sv = \Gamma_0 v$
- *The determinant of the pencil, $det(\lambda S - \Gamma_0)$ has generalized eigenvalues which satisfy: $\lambda \leq 1$*
- *The generalized eigenvalue around $\lambda = 1$ has a corresponding eigenvector v, and has dimension 2*

Proof. The proof is by construction.

Define:

$$S_1 = \frac{\Gamma_0\big(vv^T + (Jv)(Jv)^T\big)\Gamma_0}{\langle \Gamma_0 v, v\rangle} \tag{4.72}$$

Note that $S_1 v = \Gamma_0 v$ (since $\langle J\Gamma_0 v, v\rangle = 0$). S_1 is a rank two matrix, so we want to make it full rank. Note that S_1 is symmetric, and is the realification of some hermitian matrix Q_1 since $S_1 J = J S_1$.

Set

$$S = S_1 + \mu(I - \frac{vv^t}{||v||^2} - \frac{(Jv)(Jv)^t}{||v||^2}) \tag{4.73}$$

Note that $Sv = \Gamma_0 v$ still.

At this point, we constructed a family of symmetric matrices S such that

$Sv = \Gamma_0 v$. The only thing left to do is to make 1 be the largest generalized eigenvalue here.

To do so, let $\mu_1 = \lambda_{max}(\Gamma_0 - S)$. The claim is that for any $\mu > \mu_1$, that S would satisfy the pencil criterion.

To justify this, $S = S_1 + \mu S_0$, what we need is $S \geq \Gamma_0$ (because $\lambda S - \Gamma_0$ would be positive definite for $\lambda \geq 1$)

$$S = S_1 + \mu S_0 \geq \Gamma_0$$

$$\mu S_0 \geq \Gamma_0 - S_1$$

$$\mu I \geq \Gamma_0 - S_1$$

$$1 \geq \frac{1}{\mu}(\Gamma_0 - S_1)$$

The reason the identity appears is because $S_0 = (I - \frac{vv^t}{||v||^2} - \frac{(Jv)(Jv)^t}{||v||^2})$. We already know that for v, the generalized eigenvalue is $\lambda = 1$. We now get S_0 acts as the identity on the orthogonal complement.

Therefore, for $\mu > \lambda_{max}(\Gamma_0 - S_1)$, we have that $\lambda \leq 1$.

Therefore a S with the listed properties exists and is constructible.

□

Theorem 4.7.2. *There exists a matrix S_v such that if the Golden Retriever is initialized with the given S_v, then the algorithm converges to the critical point v.*

Proof. Let S be as in the previous lemma. we examine the path $\nabla_x \Omega(\xi, \lambda) =$

$(\Gamma(\xi) + \lambda S - \Gamma_0)\xi$ where $\xi = cv$.

$$0 = (\Gamma(cv) + \lambda S - \Gamma_0)cv$$

$$0 = c^3\Gamma(v)v + c\lambda Sv - c\Gamma_0 v$$

$$0 = c^2\Gamma(v)v + \lambda Sv - \Gamma_0 v$$

Now since $\Gamma(v)v = \Gamma_0 v$ (v is a critical point at $\lambda = 0$), and $Sv = \Gamma_0 v$, and $\Gamma_0 v \neq 0$ (since $\Gamma_0 v = \Gamma(v)v$ and $\Gamma(v)$ is positive definite) then we have the following

$$0 = c^2\Gamma(v)v + \lambda Sv - \Gamma_0 v$$

$$0 = (c^2 + \lambda - 1)\Gamma_0 v$$

$$c^2 = (1 - \lambda)$$

$$c = \sqrt{1 - \lambda}$$

The last line is effectively choosing one of the two equivalent paths. Therefore, if we initialize the algorithm with the given S matrix, and initialize the direction along the principal eigenvector, v, we have that the algorithm will follow the critical path: $(\xi(\lambda), \lambda) = \left((\sqrt{1 - \lambda})v, \lambda\right)$ □

The theorem above, when applied to $v = z$, a global minimizer, shows that there exists a S which guarantees that the algorithm converges. This gives us the following theorem as a corollary.

Theorem 4.7.3. *Let ζ be the minimizer to the optimization problem in (2.2). There*

exists a positive definite matrix S_z such that the Golden Retriever Algorithm, initialized with S_z, converges to S. Moreover, the trajectory of the homotopy path with S_z, when projected onto $\lambda = 0$, follows a straight line.

However, it is worth noting that to construct such a Q, we will need to know z, so Q can only be given by an oracle.

Appendix A: Useful Identities and Derivations

A.1 Useful Identities

To derive the equations we used, there are several vector calculus identities we needed. Here we give a brief derivation of those identities. These identities can be found in many places (such as [33]).

Lemma A.1.1. $\nabla_x \langle x, f \rangle = f$

Proof. $\nabla_x(\langle x, f \rangle) = \nabla_x(x_1 f_1 + ... + x_m f_n)$.

Therefore: $\nabla_x(\langle x, f \rangle)_i = f_i \Rightarrow \nabla_x(\langle x, f \rangle) = f$ □

Corollary A.1.2. $\nabla_x \langle Ax, f \rangle = A^T f$

Proof. $\nabla_x \langle Ax, f \rangle = \nabla_x \langle x, A^T f \rangle$. Now we can apply the above lemma. □

Lemma A.1.3. $\nabla_x \langle Ax, x \rangle = (A^T + A)x$

Proof. Note that $(Ax)_i = \sum_k A_{ik} x_k$, then we have that $\langle Ax, x \rangle = \sum_i \sum_k A_{ik} x_k x_i$

Therefore, we have that:

$$\nabla_x \langle Ax, x\rangle_m = \frac{\partial}{\partial x_m} \sum_i \sum_k A_{ik} x_k x_i$$

$$= \sum_i \sum_k A_{ik} \frac{\partial}{\partial x_m}(x_k x_i)$$

$$= \sum_i \sum_k A_{ik} \frac{\partial}{\partial x_m}(x_k) x_i + \sum_i \sum_k A_{ik} x_k \frac{\partial}{\partial x_m}(x_i)$$

$$= \sum_i A_{im} x_i + \sum_k A_{mk} x_k =$$

$$= \sum_i A^T_{mi} x_i + \sum_k A_{mk} x_k$$

$$= (A^T x)_m + (Ax)_m = [(A^T + A)x]_m$$

Since this is true for each fixed m, it follows that $\nabla_x \langle Ax, x\rangle = (A^T + A)x$ □

We immediately get the following corollary.

Corollary A.1.4. *If A is symmetric, then $\nabla_x \langle Ax, x\rangle = 2Ax$*

Lemma A.1.5. $\nabla_x(c(x)v) = v \otimes \nabla_x(c(x)) + c(x)\nabla_x v$

Proof. First we need to define what it means to take a gradient of a vector field. In rectangular coordinates, the gradient of a vector field $\nabla_x f = \frac{\partial f^i}{\partial x^j} e_i \otimes e_k$ (see [33]). Note that in general, one can put in a metric tensor component, but for us g^{jk} is the metric tensor components for usual Euclidean space, so $g^{jk} = \delta^{jk}$.

Therefore, we have the following:

$$\nabla_x(c(x)v) = \sum_i \sum_j \frac{\partial c(x) v_i}{\partial x_j} e_i \otimes e_j$$

$$= \sum_i \sum_j v_i \frac{\partial (c(x))}{\partial x_j} e_i \otimes e_j + c(x) \sum_i \sum_j \frac{\partial (v_i)}{\partial x_j} e_i \otimes e_j$$

$$= \sum_i v_i e_i \otimes \sum_j \frac{\partial (c(x))}{\partial x_j} e_j + c(x) \sum_i \sum_j \frac{\partial (v_i)}{\partial x_j} e_i \otimes e_j$$

$$= v \otimes \nabla_x (c(x)) + c(x) \nabla_x v$$

□

Lemma A.1.6. $\nabla_x(Ax) = A$

Proof. Note we have that $(Ax)_i = \sum_k A_{ik} x_k$.

Now again, we have that, by the definition of the gradient of a vector above:

$$\nabla_x(Ax) = \sum_i \sum_j \frac{\partial (Ax)_i}{\partial x_j} e_i \otimes e_j$$

$$= \sum_i \sum_j \frac{\partial \sum_k A_{ik} x_k}{\partial x_j} e_i \otimes e_j$$

$$= \sum_i \sum_j A_{ij} e_i \otimes e_j$$

$$= A$$

□

A.2 Simple Properties of Matrices

Lemma A.2.1. *If $\lambda I - A$ is positive definite, then $\lambda > eig_{max}(A)$*

Proof. Since $\lambda I - A$ is positive definite, then $x^T(\lambda I - A)x > 0$ for all $x \neq 0$, so take $x = v_{max}(A)$, the eigenvector corresponding to the largest eigenvalue, e.

Then $0 < v^T(\lambda I - A)v = \lambda v^T v - v^T Av = \lambda v^T v - v^T ev = (\lambda - e)v^T v = (\lambda - e)||v||^2$.

Therefore $0 < \lambda - e$, and we get $\lambda > e$. □

Lemma A.2.2. *If A is symmetric and $\lambda > eig_{max}(A)$, then $\lambda I - A$ is positive definite.*

Proof. Since A is a symmetric matrix, then being positive definite is equivalent to every eigenvalue being positive.

A vector v is an eigenvector of $(\lambda I - A)$ if and only if it is an eigenvector for A since if $(\lambda I - A)v = \lambda v - Av = cv$, we can rearrange to have $Av = (\lambda - c)v$. Let e be the corresponding eigenvalue for A ($e = \lambda - c$). Then the eigenvalue for $(\lambda I - A)$ is given by $\lambda - e$, which is minimized when e is the largest eigenvalue of A. Since $\lambda > eig_{max}(A)$, we have that $(\lambda I - A)$ is positive definite.

□

A.3 Constants in Concentration Lemma

In this section, we sketch some of the probabilistic results used in the Concentration Theorems in the Thesis.

Proposition A.3.1. *If* $erfc(z) \leq e^{-z^2}$, *we get that* $\mathbb{P}(|v_1| \geq L) \leq e^{\frac{-L^2}{2}}$

Proof. Recall that $erfc(z) = 1 - \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$, so if we now examine

$$\mathbb{P}(|v_1| \geq L) = 1 - \mathbb{P}(|v_1| \leq L) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-L}^{L} e^{-\frac{x^2}{2}} dx \tag{A.1}$$

$$= 1 - \frac{\sqrt{2}}{\sqrt{\pi}} \int_0^L e^{-\frac{x^2}{2}} dx \tag{A.2}$$

Now making the substitution $x = \sqrt{2}t$, we get

$$1 - \frac{2}{\sqrt{\pi}} \int_0^{\frac{L}{\sqrt{2}}} e^{-t^2} dt = erfc(\frac{L}{\sqrt{2}}) \leq e^{\frac{-L^2}{2}} \tag{A.3}$$

□

In the next propositions, we will show the upperbounds which were claimed in Theorem 3.2.7

Proposition A.3.2. *A sufficient upper bound for Hoeffding's inequality in Theorem 3.2.7 is given by* $C_1 = 2\sqrt{\frac{\gamma}{\delta_0}}$

Proof. Let $X_k = \langle \widetilde{f}_k, \widetilde{y} \rangle$, and $a_k = v_k^3$. Then $\mathbb{E}[X_k] = 0$ and $Var(X_k) = ||\widetilde{y}||^2$. Therefore, $X_k \sim \mathcal{N}(0, ||\widetilde{y}||)$. Therefore, $Var(\sum_{k=1}^m a_k X_k) = \sum_{k=1}^m v_k^6 ||\widetilde{y}||^2$. There-

fore,

$$\mathcal{P} = \mathbb{P}(|\sum_{k=1}^{m} a_k X_k| \geq m\delta_0||\widetilde{y}||) = \mathbb{P}\bigg(\frac{|\sum_{k=1}^{m} a_k X_k|}{\sqrt{\sum_{k=1}^{m} v_k^6||\widetilde{y}||^2}} \geq \frac{m\delta_0}{\sqrt{\sum_{k=1}^{m} v_k^6}}\bigg)$$

Applying the tail bound for the cdf of a normal distribution gives us

$$\mathcal{P} \leq 2exp(\frac{-1}{2}\frac{m^2\delta_0^2}{\sum_{k=1}^{m} v_k^6})$$

. Now choosing $m = C_1\sqrt{n\sum_{k=1}^{m} v_k^6}$, we get

$$\mathcal{P} \leq 2exp(\frac{-1}{2}C_1^2 n\delta_{0^2}) \leq 3exp(\frac{-1}{2}C_1^2 n\delta_{0^2})$$

To ensure this is less than $3exp(-2\gamma n)$, it is sufficient to take $\gamma = \frac{1}{4}C_1^2\delta_0^2$, so $C_1 = 2\frac{\sqrt{\gamma}}{\delta_0}$ □

Proposition A.3.3. *A sufficient upper bound for Bernsteins's inequality in Theorem 3.2.7 is given by* $C_0 = \max\{\sqrt{40/3}\sqrt{\frac{\gamma}{\delta_0}}, 16\frac{\gamma}{\delta_0}\}$

Proof. We do a direct computation of the probability. Let $X_k = (\langle\widetilde{y}, f_k\rangle^2 - ||\widetilde{y}||^2)$. Then we want to compute

$$\mathcal{P} = \mathbb{P}(\sum_{k=1}^{m} v_k^2 X_k \geq m\delta_0||\widetilde{y}||^2) = \mathbb{P}(exp(\lambda\sum_{k} v_k X_k) \geq exp(m\lambda\delta_0||\widetilde{y}||^2))$$

Now, by Markov's inequality, this is less than

$$inf_{\lambda>0} e^{-m\delta_0||\widetilde{y}||^2} \mathbb{E}[e^{\lambda \sum_{k=1}^m v_k^2 X_k}] = inf_{\lambda>0} e^{-m\delta_0||\widetilde{y}||^2} \prod_{k=1}^m \mathbb{E}[e^{\lambda v_k^2 X_k}]$$

Let $\mu = \lambda v_k^2$, then a computation shows Computing

$$\mathbb{E}[e^{\lambda v_k^2 X_k}] = \frac{e^{-\mu||\widetilde{y}||^2}}{\sqrt{1-2\mu||\widetilde{y}||^2}}$$

Therefore we get that

$$\mathcal{P} \leq E(\lambda) := inf_{\lambda>0} \frac{exp(-m\lambda\delta_0||\widetilde{y}||^2 - \lambda \sum_k v_k^2 ||\widetilde{y}||^2)}{\prod_{k=1}^m (1-2\lambda v_k^2||\widetilde{y}||^2)^{\frac{1}{2}}}$$

Choose a $0 < c_0 < 1$, then define $\beta = \frac{1}{2} + \frac{1}{3}\frac{c_0}{1-c_0}$. Now

$$\frac{1}{1-x} \leq e^{x+\beta x^2} \qquad \text{for } 0 \leq x \leq c_0 < 1$$

Therefore

$$\frac{1}{\sqrt{1-x}} \leq e^{\frac{x}{2} + \frac{5x^2}{12}} \quad \text{for } 0 \leq x \leq \frac{1}{2}$$

Therefore

$$E(\lambda) \leq exp\{-\lambda||\widetilde{y}||^2(m\delta_0 + \sum_{k=1}^m v_k^2) + \lambda||\widetilde{y}||^2 \sum_k v_k^2 + \frac{5}{12}\sum_k 4\lambda^2 v_k^4 ||\widetilde{y}||^4\}$$

$$= exp\{-m\delta_0||\widetilde{y}||^2\lambda + \frac{5||\widetilde{y}||^4}{3}\sum_k v_k^4\lambda^2\}$$

Define $Q(\lambda)$ to be the exponent

$$Q(\lambda) = -m\delta_0||\widetilde{y}||^2\lambda + \frac{5||\widetilde{y}||^4}{3}\sum_k v_k^4\lambda^2$$

The root of the quadratic in λ is given by $\lambda_1 = \frac{3}{10}\frac{m\delta_0}{\sum_k v_k^4||\widetilde{y}||^2}$. We also need $2\lambda v_k^2||\widetilde{y}||^2 \leq \frac{1}{2}$, so we get that we must have $\lambda \leq \frac{1}{4||\widetilde{y}||^2 v_k^2} \leq \frac{1}{4||\widetilde{y}||^2 \max_k v_k^2} := \lambda_2$

Thus we can take

$$\lambda = \min\{\frac{1}{4||\widetilde{y}||^2 \max_k v_k^2}, \frac{3}{10}\frac{m\delta_0}{||\widetilde{y}||^2 \sum_k v_k^4}\}$$

Now we examine the upper bound for $E(\lambda)$ with these parameters.

We have $Q(\lambda_1) = -\frac{3}{20}\frac{m^2\delta_0^2}{\sum_k v_k^4}$ and if $\lambda_2 < \lambda_1$ then $\frac{\sum_k v_k^4}{\max_k v_k^2} \leq \frac{12}{10}m\delta_0$, thus $Q(\lambda_2) = \frac{m\delta_0}{4\max_k v_k^2} - \frac{5}{48}\frac{\sum_k v_k^4}{\max_k v_k^2} \geq \frac{m\delta_0}{8\max_k v_k^2}$

Therefore, we have

$$\mathcal{P} \leq E(\lambda) \leq max\left(exp\{\frac{-3}{20}\frac{m^2\delta_0^2}{\sum_k v_k^4}\}, exp\{\frac{-m\delta_0}{8\max_k v_k^2}\}\right)$$

Now if we match the exponents to $-2\gamma n$, we get that $C_0 = \max\{\sqrt{40/3}\frac{\sqrt{\gamma}}{\delta_0}, 16\frac{\gamma}{\delta_0}\}$ $\square$

A.4 Probabilistic Bounds on b_0

In this section, we show a lemma which gives an upper bound on b_0.

Lemma A.4.1. *For $m \geq 64n^3$, the probability that $b_0 > 64$ is less than $(m+1)exp(-(2-log(5))m^{\frac{1}{3}})$*

Proof. By definition,

$$b_0 = \max_{||x||=1} \sum_{k=1}^{m} \langle x, f_k \rangle^4$$

Define a positive constant t. We are looking for an upper bound on the probability

$$\mathbb{P}\{b_0 > t\} \tag{A.4}$$

We are looking for

$$(b_0 m)^{\frac{1}{4}} = ||T||_{2\to 4} = \sup_{||x||=1} ||Tx||_4 \tag{A.5}$$

Let $\mathcal{N}$ be an r-net in $\mathbb{R}^n$. Thus for some $x_0 \in \mathbb{S}^{n-1}$

$$||T||_{2\to 4} = \max_{||x||=1} ||Tx||_4 = ||Tx_0|| \tag{A.6}$$

This in turn is equal to

$$||T(x_0 - \tilde{x}) + T\tilde{x}||_4 \leq ||T\tilde{x}||_4 + ||T(x_0 - \tilde{x})||_4 \tag{A.7}$$

$$\leq \max_{x\in\mathcal{N}} ||Tx||_4 + ||T||_{2\to 4} \cdot r \tag{A.8}$$

Thus implies, after rearranging

$$||T||_{2\to 4} \leq \frac{1}{1-r} (\max_{x\in\mathcal{N}} ||Tx||_4) \tag{A.9}$$

Thus we get

$$b_0 \leq (\frac{1}{1-r})^4 \max_{x\in\mathcal{N}} (\frac{1}{m} \sum_{k=1}^{m} \langle x, f_k \rangle^4) \tag{A.10}$$

Thus $\mathbb{P}\{b_0 > t\} \leq \mathbb{P}\{\exists x \in \mathcal{N} : \frac{1}{m}\sum_{k=1}^{m}\langle x, f_k\rangle^4 > (1-r)^4 t\}$ This in turn is less than

$$\leq |\mathcal{N}|\mathbb{P}\{\frac{1}{m}\sum_{k=1}^{m} v_k^4 > (1-r)^4 t\} \leq (1+\frac{2}{r})^n \mathbb{P}\{\sum_{k} v_k^4 > m(1-r)^4 t\} \tag{A.11}$$

Now we do a similar bound to what we did in the Concentration Lemma, where we limit the quantity by some upper bound L, then use the upper bound on $erfc(z)$ and Bernstein's Inequality to get

$$\mathbb{P}\{b_0 > t\} \leq \left(me^{\frac{-L^2}{2}} + exp\{-\frac{1}{2}\frac{m\big((1-r)^4 t - 3\big)^2}{105 + \frac{1}{3}L^4\big((1-r)^4 t - 3\big)}\}\right)(1+\frac{2}{r})^n \tag{A.12}$$

Now choose $r = \frac{1}{2}$ and assume $\frac{2}{3}L^4(\frac{t}{16} - 3) \geq 105$. Then we get

$$\mathbb{P}\{b_0 > t\} \leq \left(me^{\frac{-L^2}{2}} + exp\{-\frac{1}{2}\frac{m\big(\frac{t}{16} - 3\big)}{L^4}\right)5^n \tag{A.13}$$

Choosing $L = \big(m(\frac{t}{16} - 3)\big)^{\frac{1}{6}}$ gives us

$$\mathbb{P}\{b_0 > t\} \leq (m+1)exp\{-\frac{1}{2}\big(m[(\frac{t}{16}) - 3]\big)^{\frac{1}{3}}\}5^n \tag{A.14}$$

When $t = 64$, and $\frac{t}{16} - 3 = 1$, we get

$$\mathbb{P}\{b_0 > 64\} \leq (m+1)exp\{-\frac{m^{\frac{1}{3}}}{2}\}5^n \tag{A.15}$$

If $m \geq 64n^3$, this gives us the bound in the statement of the lemma. Note that we made an assumption that $L^4 \geq \frac{3}{2} \cdot 105 \Rightarrow m^{\frac{2}{3}} \geq \frac{3}{2} \cdot 105$, so $m \geq 1977$ (if $m = 64n^3$,

this would mean $n \geq 4$). $\square$

Bibliography

[1] Wooshik Kim and Monson H Hayes. The phase retrieval problem in x-ray crystallography. In *[Proceedings] ICASSP 91: 1991 International Conference on Acoustics, Speech, and Signal Processing*, pages 1765–1768. IEEE, 1991.

[2] Michael Kech and Michael Wolf. From quantum tomography to phase retrieval and back. In *2015 International Conference on Sampling Theory and Applications (SampTA)*, pages 173–177. IEEE, 2015.

[3] Lawrence Rabiner. Fundamentals of speech recognition, vol 14. *Fundamentals of speech recognition*, 1993.

[4] Peter G Casazza et al. The art of frame theory. *Taiwanese Journal of Mathematics*, 4(2):129–201, 2000.

[5] Philipp Grohs, Sarah Koppensteiner, and Martin Rathmair. The mathematics of phase retrieval. *arXiv preprint arXiv:1901.07911*, 2, 2019.

[6] Jameson Cahill, Peter Casazza, and Ingrid Daubechies. Phase retrieval in infinite-dimensional hilbert spaces. *Transactions of the American Mathematical Society, Series B*, 3(3):63–76, 2016.

[7] Radu Balan. Reconstruction of signals from magnitudes of redundant representations: The complex case. *Foundations of Computational Mathematics*, 16(3):677–721, 2016.

[8] Radu Balan. Reconstruction of signals from magnitudes of redundant representations: The complex case. *Foundations of Computational Mathematics*, 16(3):677–721, 2016.

[9] Eugene L Allgower and Kurt Georg. *Introduction to numerical continuation methods*. SIAM, 2003.

[10] Steve Smale. A convergent process of price adjustment and global newton methods. *Journal of Mathematical Economics*, 3(2):107–120, 1976.

[11] John Piggott, John Whalley, et al. Uk tax policy and applied general equilibrium analysis. *Cambridge Books*, 2009.

[12] JC Alexander and James A Yorke. The homotopy continuation method: numerically implementable topological procedures. *Transactions of the American Mathematical Society*, 242:271–284, 1978.

[13] S Kishore Kumar, William I Thacker, and Layne T Watson. Magnetohydrodynamic flow and heat transfer about a rotating disk with suction and injection at the disk surface. *Computers & fluids*, 16(2):183–193, 1988.

[14] W Forster. Homotopy methods. In *Handbook of global optimization*, pages 669–750. Springer, 1995.

[15] Emmanuel J Candes, Thomas Strohmer, and Vladislav Voroninski. Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics*, 66(8):1241–1274, 2013.

[16] Yuxin Chen and Emmanuel J Candès. Solving random quadratic systems of equations is nearly as easy as solving linear systems. *arXiv preprint arXiv:1505.05114*, 2015.

[17] Huishuai Zhang and Yingbin Liang. Reshaped wirtinger flow for solving quadratic system of equations. *Advances in Neural Information Processing Systems*, 29:2622–2630, 2016.

[18] Junjie Ma, Ji Xu, and Arian Maleki. Optimization-based amp for phase retrieval: The impact of initialization and $\backslash \ell_{\{}2\}$ regularization. *IEEE Transactions on Information Theory*, 65(6):3600–3629, 2019.

[19] Arian Maleki. Approximate message passing algorithms for compressed sensing. *a degree of Doctor of Philosophy, Stanford University*, 2011.

[20] Jonathan S Yedidia, William T Freeman, and Yair Weiss. Understanding belief propagation and its generalizations. *Exploring artificial intelligence in the new millennium*, 8:236–239, 2003.

[21] Katta G Murty and Santosh N Kabadi. Some np-complete problems in quadratic and nonlinear programming. *Mathematical Programming*, 39(2):117–129, 1987.

[22] Radu Balan and David Bekkerman. The cramer-rao lower bound in the phase retrieval problem. In *2019 13th International conference on Sampling Theory and Applications (SampTA)*, pages 1–5. IEEE, 2019.

[23] Emmanuel J Candes, Xiaodong Li, and Mahdi Soltanolkotabi. Phase retrieval via wirtinger flow: Theory and algorithms. *IEEE Transactions on Information Theory*, 61(4):1985–2007, 2015.

[24] YT Feng and DRJ Owen. Conjugate gradient methods for solving the smallest eigenpair of large symmetric eigenvalue problems. *International Journal for Numerical Methods in Engineering*, 39(13):2209–2229, 1996.

[25] Paul H Rabinowitz. Some global results for nonlinear eigenvalue problems. *Journal of functional analysis*, 7(3):487–513, 1971.

[26] Michael G Crandall and Paul H Rabinowitz. Bifurcation from simple eigenvalues. *Journal of Functional Analysis*, 8(2):321–340, 1971.

[27] Marco Chiani, Davide Dardari, and Marvin K Simon. New exponential bounds and approximations for the computation of error probability in fading channels. *IEEE Transactions on Wireless Communications*, 2(4):840–845, 2003.

[28] Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010.

[29] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

[30] Yi Yu, Tengyao Wang, and Richard J Samworth. A useful variant of the davis–kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2015.

[31] Daureen Steinberg. Computation of matrix norms with applications to robust optimization. *Research thesis, Technion-Israel University of Technology*, 2, 2005.

[32] Ronald S Irving. *Integers, polynomials, and rings: a course in algebra*. Springer Science & Business Media, 2003.

[33] Piaras Kelly. Solid mechanics part iii: Foundations of continuum mechanics. solid mechanics lecture notes, 2013.