ABSTRACT

Title of dissertation: GEOMETRIC STRUCTURES AND OPTIMIZATION ON SPACES OF FINITE FRAMES

Nathaniel Strawn, Doctor of Philosophy, 2011

Dissertation directed by: Professor John J. Benedetto
Professor Radu V. Balan
Department of Mathematics

A finite (μ, S)-frame variety consists of the real or complex matrices $F = [f_1 \cdots f_N]$ with frame operator $FF^* = S$, and which also satisfyies $\|f_i\| = \mu_i$ for all $i = 1, \ldots, N$. Here, S is a fixed Hermitian positive definite matrix and $\mu = [\mu_1 \cdots \mu_N]$ is a fixed list of lengths. These spaces generalize the well-known spaces of finite unit-norm tight frames. We explore the local geometry of these spaces and develop geometric optimization algorithms based on the resulting insights.

We study the local geometric structure of the (μ, S)-frame varieties by viewing them as intersections of generalized tori (the length constraints) with distorted Stiefel manifolds (the frame operator constraint). Exploiting this perspective, we characterize the nonsingular points of these varieties by determining where this intersection is transversal in a Hilbert-Schmidt sphere. A corollary of this characterization is a characterization of the tangent spaces of (μ, S)-frame varieties, which is in turn leveraged to validate explicit local coordinate systems. Explict bases for the tangent spaces are also constructed.

Geometric optimization over a (μ, S)-frame variety is performed by combining knowledge of the tangent spaces with geometric optimization of the frame operator distance over a product of spheres. Given a differentiable objective function, we project the full gradient onto the tangent space and then minimize the frame operator distance to obtain an approximate gradient descent algorithm. To partially validate this procedure, we demonstrate that the induced flow converges locally. Using Sherman-Morrision type formulas, we also describe a technique for constructing points on these varieties that can be used to initialize the optimization procedure. Finally, we apply the approximate gradient descent procedure to numerically construct equiangular tight frames, Grassmannian frames, and Welch bound equality sequences with low mutual coherence.

GEOMETRIC STRUCTURES AND OPTIMIZATION ON SPACES OF FINITE FRAMES

by

Nathaniel Kirk Strawn

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
2011

Advisory Committee:
Professor John J. Benedetto, Chair/Advisor
Professor Radu V. Balan, Co-Advisor
Professor Wojciech Czaja
Professor Kasso Okoudjou
Professor Ramalingam Chellappa

Dedication

To Grandma

Acknowledgments

First, I would like to thank my advisor, Dr. John J. Benedetto, for giving me countless opportunities (not all squandered) at the University of Maryland and at the Norbert Wiener Center for Harmonic Analysis and Applications. His expansive knowledge of Harmonic Analysis (and it's history) kept my horizons broad even as tunnel vision led me down the path of this thesis, and his genial nature kept me grounded throughout my hard-headed misadventures. I feel truly blessed to have him as my mathematical sire.

I am also immensely grateful for the many hours that my co-advisor, Dr. Radu V. Balan, spent lending me his expertise. I am perpetually inspired by his technical acumen, his expansive knowledge, his effortless recall, and his warm wit. I am also thankful for Dr. Wojciech Czaja, whose strategic planning and perspective were invaluable for moving forward on numerous projects at the Norbert Wiener Center, and for Dr. Mike Dellomo, who provided me with a wealth of practical knowledge. Furthermore, I would like to thank Dr. Kasso Okoudjou and Dr. Rama Chellappa for being warm and witty colleagues throughout my academic travels.

Many thanks go out to Dr. Matthew Fickus (my mathematical brother), Dustin Mixon (my mathematical godnephew), Dr. Pete Casazza (my mathematical fairy godfather), and Jameson Cahill (another mathematical godnephew) for taking an interest in my work and for engaging, ongoing collaborations. I am always in awe of their skill, ingenuity, and humor. I would also like to thank Dr. David R. Larson (my mathematical godfather) and Dr. Shidong Li (my mathematical brother) for

organizing the satellite conference where these collaborations began.

Whilst we pluck gossamers out of the mathematical aether, there are numerous people that make sure that reality does not fall apart. Celeste Relagado has constantly ensured that I (and many others) am on course for graduation, and she has been a patient audience as I air the grievances of fellow graduate students. Sharon Welton keeps my paychecks and insurance flowing, while also promoting my theatrical aspirations. Without Linette Berry and Sandy Allen, the Student Analysis Seminar could not exist in time and space. Without the warm help of Rachel Katz, the Spotlight Competition would be penniless. Without Liliana Gonzalez, graduates would starve during New Student Orientation Week and the Graduation Conference. Liz Wincek ensures that my travel reimbursements go through, and also that the administration offices have a fair helping of bawdy language. Without Rhyneta Gumbs, funding for our projects would be inaccessible. Finally, I am most fortunate to be leaving before the retirement of Fletcher Kinne, as he keeps the entire math building alive and functioning. I'll greatly miss his insights on sports and history, and I'll miss all of the Math Department staff as I move into the next phase of my life.

Dr. Emily King has at different times been my officemate, roommate, and best man, but she has always been a friend, colleague, and confidant. She is directly responsible for my attendance at the University of Maryland, and I shall greatly miss her as she moves to Germany to pursue her career. My other officemates in the "ninth gate" (Andrew Sanders, Kareem Sorathia, Aaron Skinner, Shelby Wilson, and Chrisitian Sykes) all contributed to intellectual growth, happiness, and

deliquency during my first two years. I am very lucky to have them as friends and colleagues.

I am also thankful for my colleagues at the Norbert Wiener Center. Dr. Chris Shaw's wry wit and incomparable organization were always crisp and refreshing. Dr. Matthew Hirn and Dr. Justin Flake showed me the ropes and provided me with direction during my first semester of work. Dr. Jens Christensen, Avner Halevy, Enrico Au Yueng, Kevin Duke, and Rongrong Wang have been engaging and supportive colleagues. Finally, I would like to acknowledge Joe Woodworth, whose work ethic puts all the graduate students to shame.

Many other colleagues have contributed their knowledge and effort to enrich my life and career. Jim Kelly taught me how to connect wirelessly through Linux, and helped me install Matlab on my computer when I did not have a CD drive. Sean Kelly graced me with several engaging conversations on the bus, and also taught me the zen of CSS so that my webpage is merely mediocre rather than an eyesore. Xu Wen Chen has been a tireless advocate for all the students, and has been a great help in organizing the Student Analysis Seminar. Domingo Ruiz has always been there to remind me that love is important. Jeffrey Frazier's lucid explanations greatly contributed to my success with the Topology/Geometry qualifier. Finally, I am greatly indebted to Joyce Hsiao for introducing me to my wife.

During my four years at the University of Maryland, I have been funded by teaching assistance and government projects. I would like to thank Dr. Chris Laskowski, Dr. Justin Wyss-Gallifent, Dr. Patrick Fitzpatrick, and Dr. Casey Cremins for being great primary instructors during my teaching assistantships. I also would like to thank the National Geospatial Intelligence Agency, Jack Yao of the National Institutes of Health, and the Laboratory for Telecommunications Sciences for engaging projects.

Without some key nudges in the proper direction, this thesis would not have come to fruition. I thank Andrew Sanders for reminding me about Lie brackets, Dr. Larry Washington for pointing me in the direction of the real analytic Inverse Function Theorem, Sean Rostami for convincing me that the Open Mapping Theorem holds in arbitrary dimension, and Linh Nguyen for presenting me with a simplified proof of the ellipsoidal condition.

My family has also been supportive throughout this long journey. I am very thank for my Uncle Steven and my Aunt Laura for giving me a sanctuary during the holidays. I am thankful for my cousins David, Julia, and Emily for their warm energy and generosity. I am thankful for my Mother and Father for their support during the entirety of my education. I am thankful for my little brother and sister Taylor and Natalie for letting me hone my tutoring skills on them so long ago, and I am very proud of what they have both accomplished. I am very grateful for my Grandmother, as she has always supported me with her love and resources.

My brother Devon deserves special commendation for his support. He has always been available to lend me applause and advice throughout my career, and he has taught me to truly evaluate my goals and to be satisfied with baby steps.

Lastly, I am eternally grateful for my lovely, intelligent, hardworking, and caring wife Anna. In her, I have found love and home.

Table of Contents

List of Figures

List of Notation

$\mathbb{E}$	the real $\mathbb{R}$ or complex $\mathbb{C}$ numbers
$\mathbb{E}^d$	d-dimensional real or complex Euclidean space
$\|\cdot\|^2$	sum of squares of matrix entries
$\mathcal{S}_{\mathbb{E}^d(c)}$	the sphere of radius c in $\mathbb{E}^d$
$[n]$	the n-set, $\{1,\ldots,n\}$
$M_{m\times n}(\mathbb{E})$	the space of m by n matrices
μ	list of lengths
S	a d by d Hermitian postive definite operator
$\mathbb{T}_{\mathbb{E}^d}(\mu)$	the generalized torus with radii in μ
$\mathrm{St}_{\mathbb{E}^d}(N)$	the Stiefel manifold of d orthonormal rows in $\mathbb{E}^N$
$\mathcal{F}_{\mathbb{E}}(\mu, S)$	the space of (μ, S)-frames
$\nabla_x^{\mathcal{M}}$	the gradient at x on the manifold $\mathcal{M}$
HPD	Hermitian positive definite

Chapter 1

Introduction

1.1 Background

Frames were originally introduced by Duffin and Schaeffer to generalize Fourier expansions [22], and general frames [50] furnish practitoners with expansion formulas akin to orthonormal expansions. Daubechies, Grossmann, and Meyer [16] renewed interest in infinite dimensional frames by exhibiting classes of frames with tractable reconstruction formulas. There is now a multitude of well-studied frames that have been constructed with various applications in mind; Gabor frames, wavelets [17, 18], curvelets [11], shearlets [31], and Fourier frames constitute a partial list of popular infinite dimensional frames. More recently, structured finite frames have become popular because of their applications in wireless telecommunications [42, 48], sigma-delta quantization [5, 7, 6, 4], coding theory [13, 24, 37], and sparse reconstruction [20, 33, 44].

The fundamental structural condition of interest for many of these applications is tightness. Finite tight frames satisfy $FF^* = cI_{d\times d}$, which is equivalent to the reconstruction formula

$$x = c \sum_{i\in[N]} \langle x, f_i\rangle f_i \tag{1.1.1}$$

for all $x \in \mathbb{E}^d$, and where $c > 0$ is fixed. The simple reconstruction formula provided by tightness motivated the work in [16], and is valuable for signal processing applications. Moreover, the least squares solution to $Fx = b$ is exactly $c^{-1}F^*b$ when F is a tight frame. Geometrically, it is well known that the spaces of tight frames are essentially the Stiefel manifolds, and Stiefel manifolds are well studied objects: a cell structure can be imposed to calculate homology and the cohomology ring, the tangent bundles are well understood, and local parameterizations are easily obtainable. Because of these local parameterizations, it is relatively easy to design tight frames for specific applications (see [1]).

Beyond tighness, there are several other structural conditions that are useful for applications. Other structured tight frames of interest are finite unit norm tight frames (FUNTFs), generalized Welch bound equality (WBE) sequences, equiangular tight frames, and Grassmannian frames.

Before this work, knowledge conerning the structure of the spaces of FUNTFs has only been superficial. By imposing the additional constraints $\|f_i\| = 1$ for all the columns of tight frame F, a complicated structure emerges. The work of Benedetto and Fickus [3] was the first to develop a characterization of FUNTFs in an elegant and intuitive manner. They characterized the FUNTFs as the minimizers of the *frame potential* over a product of unit spheres, which reflected the characteristic equidistribution exhibited by FUNTFs. Later, Dykema et al. [23] demonstrated that certain spaces of FUNTFs are manifolds and calculated the dimension of these manifolds.

Generalized WBE sequences are tight frames that satisfy $\|f_i\| = \mu_i$ for some positive sequence $\mu_1, \ldots, \mu_N$. These sequences arise in wireless telecommunications, where each f_i is the signature code of a user with average power μ_i. Constructions of these frames has been considered by numerous authors [29, 46, 19, 40], and existence of such sequences is characterized by a majorization condition (see [14]). The manifold structure of these varieties was first explored in [40], which generalized results from [23] and also developed construction techniques.

Equiangular tight frames are FUNTFs for which $|\langle f_i, f_j\rangle|$ is constant for all $i \neq j$, and Grassmannian frames are FUNTFs which minimize the *mutual coherence*,

$$\max_{i\neq j} |\langle f_i, f_j\rangle|.$$

Theorem 4.1 of Donoho et al. [21] established a general connection between dictionaries with low mutual coherence and successful sparse reconstruction via basis pursuit [15]. The existence of equiangular tight frames has been considered in a number of settings [8, 41, 43], and they were characterized as minimizers of the 4^{th}-order frame potential by Oktay [34]. Grassmannian frames exist by a simple compactness argument, but constructing Grassmannian frames is a very difficult problem. Strohmer and Heath demonstrated that explicit constructions can be carried out in certain cases in [42]. However, general constructions have not been found and the geometry of Grassmannian frames is also completely unknown.

1.2 Summary of results

We study the geometry of algebraic varieties of (μ, S)-frames, and show that their local geometry is tractable. That is, their nonsingular points can be characterized, expressions for the tangent spaces at nonsingular points can be written down, and explicit local parameterizations can be constructed. Moreover, we exploit the explicit form of the tangent spaces to develop an efficient, approximate gradient descent procedure over finite (μ, S)-frame varieties.

In Chapter 2, we consider a general (μ, S)-frame variety as the intersection of generalized torus with a warped Stiefel manifold, and we characterize the points at which this intersection is transversal in the ambient Hilbert-Schmidt sphere (Theorem 2.1.4). These points are exactly the nonorthodecomposable (μ, S)-frames. We then characterize the tangent space at a nonsingular point as the intersection of the point's tangent spaces on the generalized torus and warped Stiefel manifold, and utilize this characterization to construct an explicit basis for the tangent space.

Chapter 3 focuses on the construction of explicit, locally well-defined analytic coordinate patches around nonsingular points on a (μ, S)-frame variety. Chapter 3 begins with an example which motivates the general construction of such parameterizations. While the example demonstrates that coordinates can be constructed formally, it does not ensure that the coordinates are locally well-defined. By proving the hypotheses of the Real-Analytic Inverse Function Theorem (Theorems 3.2.6 and 3.3.1), we demonstrate that the coordinates are well-defined. The chapter concludes with the full explicit derivations of the coordinate systems for both the real and complex case. These parameterizations are a technical manifestation of the intuition that one can choose a basis from a (μ, S)-frame, and articulate (in a small neighborhood) the remaining vectors on their respective spheres while the basis reacts to ensure that the frame retains the same frame operator.

We develop and apply a geomeric optimization procedure in Chapter 4. This procedure is powered by efficient, geometric coordinate descent and gradient descent (explored by Casazza and Fickus [12] in the FUNTF case) algorithms for minimizing the distance of XX^* from a target frame operator in the Hilbert-Schmidt norm, or the *frame operator distance* (FOD). For both the geometric coordinate descent and geometric gradient descent algorithms, we derive simple, explicit expressions for step sizes in order to acquire efficient implementations. By projecting gradients of an objective function directly onto the tangent space of a (μ, S)-frame variety, we are able to leverage efficient minimization of the FOD to obtain first-order approximations to geodesics at regular points of the (μ, S)-frame variety. Based on the bases constructed in Chapter 2, we describe iterative and direct methods for computing these projections. We also address the issue of constructing starting points for initializing our optimization procedure using techniques from [40].The chapter concludes with two applications: numerical construction of Grassmannian frames and construction of WBE sequences with low mutual coherence. The natural objective function for these applications is the $2p^{\text{th}}$-order frame potential because it is an approximation of the mutual coherence. We empirically show that minimizing the $2p^{\text{th}}$-order frame potential yields frames with low mutual coherence. These numerically constructed frames are ideal candidates for applications in coding theory and sparse reconstruction.

Code for the Matlab package Framelab is provided in the Appendix. This package implements all of the numerical procedures detailed in Chapter 4, and is freely available at the the author's webpage.

1.3 Notation

Let $\mathbb{E}^d = \mathbb{R}^d$ or $\mathbb{C}^d$ denote real or complex d-dimensional space endowed with the (symmetric or Hermitian) inner product $\langle \cdot, \cdot \rangle$. For any vector or matrix, we let $\| \cdot \|^2$ denote the sum of squared absolute-values of the entries. In the case of a matrix, this is the square of the Fröbenius or Hilbert-Schmidt norm. Let $\mathcal{S}_{\mathbb{E}^d}(c) = \{x \in \mathbb{E}^d : \|x\| = c\}$ denote the sphere of radius $c > 0$, let $[n] = \{1, \ldots, n\}$ denote the n-set.

The set of all m by n matrices is denoted $M_{m\times n}(\mathbb{E})$, and we let X^* denote the conjugate transpose of $X \in M_{m\times n}(\mathbb{E})$. Note that the transpose $X^T = X^*$ if $\mathbb{E} = \mathbb{R}$. We shall often employ the notation $x^T y$ and $x^* y$ in place of $\langle x, y \rangle$. Given an $X \in M_{m\times n}(\mathbb{E})$, an $A \subset [m]$, and a $B \subset [n]$, we let

$$X_{A\times B} = \begin{bmatrix} x_{a_1 b_1} & \cdots & x_{a_1 b_{|B|}} \\ \vdots & \ddots & \vdots \\ x_{a_{|A|} b_1} & \cdots & x_{a_{|A|} b_{|B|}} \end{bmatrix} \tag{1.3.1}$$

denote the $|A|$ by $|B|$ matrix obtained by deleting the rows of X with indices not in A, and then deleting the columns with indices not in B. Here, we have used $|A|$ and $|B|$ to denote the cardinality of A and B respectively. We also let X_B denote the matrix obtained by deleting the columns of X that are not in B. For a square matrix, $S \in M_{d\times d}(\mathbb{E})$, we let $\mathrm{tr}(S)$ denote the trace of S, and we use $\mathrm{diag}(S) \in \mathbb{E}^d$ to signify the matrix obtained by setting the off-diagonal entries of S equal to zero. We use $\mathbb{1}_N$ to denote a column vector with N unit entries. A *finite frame* for $\mathbb{E}^d$ is a collection of vectors (which shall be referred to interchangeably with the matrix

$F = [f_1 \ldots f_N] \in M_{d\times N}(\mathbb{E}))$ satisfying

$$A\|x\|^2 \le \sum_{n\in[N]} |\langle x, f_n\rangle|^2 \le B\|x\|^2 \text{ for all } x \in \mathbb{E}^d \tag{1.3.2}$$

for some constants $0 < A \le B < \infty$. If such constants exist, then we let A and B denote the sharpest constants satisfying (1.3.2). If $\|f_n\| = 1$ for all $n \in [N]$, then the frame is called *unit-norm*. If $A = B$, the frame is called a *tight frame*. If F is a finite unit-norm tight frame, we say that F is a FUNTF for brevity.

Now, F is called the *synthesis operator* of the frame and F^* is the *analysis operator*. The matrices $S = FF^*$ and F^*F are the *frame operator* and *Grammian* of F.

For a sequence of strictly positive numbers, $\mu \in \mathbb{R}^N_+$, we let

$$\mathbb{T}_{\mathbb{E}^d}(\mu) = \prod_{i\in[N]} \mathcal{S}_{\mathbb{E}^d}(\mu_i) = \{F = [f_1 \ldots f_N] \in M_{d\times N}(\mathbb{E}) : \|f_i\| = \mu_i \text{ for all } i \in [N]\}$$

denote the generalized torus with radii μ. Given a Hermitian (or, symmetric) positive definite (HPD, or SPD) operator $S \in M_{d\times d}(\mathbb{E})$, let $\sqrt{S}$ denote its canonical square root, and let

$$\sqrt{S} \cdot \mathrm{St}_{\mathbb{E}^d}(N) = \{\sqrt{S}F \in M_{d\times N}(\mathbb{E}) : FF^* = I_{d\times d}\} \tag{1.3.3}$$

denote the $\sqrt{S}$-transformed Stiefel manifold, where $I_{d\times d}$ denotes the d by d identity matrix. If $c = \mathrm{tr}(S)^{1/2} = \|\mu\|_2$, then it is straightforward to check that $\mathbb{T}_{\mathbb{E}^d}(\mu)$ and $\sqrt{S} \cdot \mathrm{St}_{\mathbb{E}^d}(N)$ are both submanifolds of the Hilbert-Schmidt sphere of radius c,

$$\mathcal{S}_{\mathbb{E}^{d\times N}}(c) = \left\{ F \in M_{d\times N}(\mathbb{E}) : \|F\|_{\mathrm{HS}} = \sqrt{\sum_{i\in[d]} \sum_{j\in[N]} |f_{ij}|^2} = c \right\}. \tag{1.3.4}$$

Let $N \ge d$. Given a $\mu \in \mathbb{R}^N_+$, and a d by d HPD operator S, the (μ, S)-frames are

$$\mathcal{F}_{\mathbb{E}}(\mu, S) = \mathbb{T}_{\mathbb{E}^d}(\mu) \cap \sqrt{S} \cdot \mathrm{St}_{\mathbb{E}^d}(N). \tag{1.3.5}$$

That is $F \in \mathcal{F}_{\mathbb{E}}(\mu, S)$ if and only if F belongs to the generalized torus with radii μ when viewed as a collection of columns, and F is a transformation of an orthonormal system when viewed as a collection of rows. As frames, they are the frames with frame operator S, and with member lengths given by μ.

For any C^1 function $\varphi : \mathbb{E}^{d\times N} \to \mathbb{R}$, and any smooth embedded manifold $\mathcal{M} \hookrightarrow \mathbb{E}^{d\times N}$, we let

$$\nabla_x^{\mathcal{M}} \varphi \tag{1.3.6}$$

denote the gradient of φ at x along $\mathcal{M}$. Note that this is the orthogonal projection of the full gradient $\nabla_x \varphi$ onto the tangent space $T_x\mathcal{M} \subset \mathbb{E}^{d\times N}$. We use $\exp_x^{\mathcal{M}} : T_x\mathcal{M} \to \mathcal{M}$ to denote the the exponentiation map.

1.4 Preliminaries

Our general reference for matrix analysis is Horn and Johnson [26], and the general reference for differential geometry is Spivak [39].

If

$$\sum_{n\in[N]} \mu_n^2 = \sum_{n\in[d]} \lambda_n(S),$$

Theorem 2.1 of [14] essentially states that $\mathcal{F}_{\mathbb{E}}(\mu, S)$ is not empty if and only if

$$\max_{\substack{I\subset[N]\\|I|=k}} \sum_{n\in I} \mu_n^2 \leq \sum_{n\in[k]} \lambda_n(S) \text{ holds for all } k \in [d], \tag{1.4.1}$$

where $\lambda_1(S) \geq \lambda_2(S) \geq \cdots \geq \lambda_d(S) > 0$ are the eigenvalues of S. In general, we shall say that $\mu \in \mathbb{R}_+^N$ and HPD $S \in M_{d\times d}(\mathbb{E})$ satisfy the "usual conditions" if

$$N \geq d, \; \sum_{i\in[N]} \mu_i^2 = \sum_{i\in[d]} \lambda_i = c, \text{ and (1.4.1) all hold true.} \tag{1.4.2}$$

We shall often invoke the usual conditions to avoid vacuous assertions.

The Grammian stores all of the correlation information in the frame, and it can be shown that the frame operator satisfies

$$x = \sum_{n\in[N]} \left\langle x, S^{-1} f_n \right\rangle f_n \text{ for all } x \in \mathbb{E}^d. \tag{1.4.3}$$

For a FUNTF, the frame operator is $S = \frac{N}{d} I_{d\times d}$ and (1.4.3) reduces to the simple form

$$x = \frac{d}{N} \sum_{n\in[N]} \langle x, f_n\rangle f_n \text{ for all } x \in \mathbb{E}^d. \tag{1.4.4}$$

1.4.1 Dimension calculations

By counting the defining constraints, it is not difficult to see that

$$\dim(\mathbb{T}_{\mathbb{E}^d}(\mu)) = \begin{cases} (d-1)N & \text{if } \mathbb{E} = \mathbb{R} \\ (2d-1)N & \text{if } \mathbb{E} = \mathbb{C} \end{cases} \tag{1.4.5}$$

and

$$\dim(\mathcal{S}_{\mathbb{E}^{d\times N}}(c)) = \begin{cases} dN-1 & \text{if } \mathbb{E} = \mathbb{R} \\ 2dN-1 & \text{if } \mathbb{E} = \mathbb{C} \end{cases} \tag{1.4.6}$$

when both are viewed as real manifolds. The dimension of the Stiefel manifold as a real manifold is (see [27])

$$\dim(\sqrt{S}\cdot \mathrm{St}_{\mathbb{E}^d}(N)) = \begin{cases} \sum_{n\in[d]}(N-n) & \text{if } \mathbb{E} = \mathbb{R} \\ \sum_{n\in[d]}(2N-2n+1) & \text{if } \mathbb{E} = \mathbb{C} \end{cases} \tag{1.4.7}$$

since $\mathrm{St}_{\mathbb{R}^d}(N) \cong O(N)/O(N-d)$ and $\mathrm{St}_{\mathbb{C}^d}(N) \cong U(N)/U(N-d)$. Here, $O(k)$ and $U(k)$ are the orthogonal and unitary groups respectively, and the Implicit Function

Theorem can be used to show that

$$\dim(O(k)) = \dim(\{A \in M_{k\times k}(\mathbb{R}) : AA^T = I_{k\times k}\}) = \sum_{n\in[k-1]} n \tag{1.4.8}$$

$$\dim(U(k)) = \dim(\{A \in M_{k\times k}(\mathbb{E}) : AA^* = I_{k\times k}\}) \tag{1.4.9}$$

$$= \sum_{n\in[k-1]} n + \sum_{n\in[k]} n = \sum_{n\in[k]} (2n-1). \tag{1.4.10}$$

Chapter 2

Nonsingular points and tangent spaces of (μ, S)-frame varieties

2.1 The intersection of $\mathbb{T}_{\mathbb{E}^d}(\mu)$ and $\sqrt{S} \cdot \mathrm{St}_{\mathbb{E}^d}(N)$ in $\mathcal{S}_{\mathbb{E}^{d\times N}}(c)$

We shall now state and prove when (μ, S)-frame varieties are locally the transversal intersection of $\mathbb{T}_{\mathbb{E}^d}(\mu)$ and $\sqrt{S} \cdot \mathrm{St}_{\mathbb{E}^d}(N)$ in $\mathcal{S}_{\mathbb{E}^{d\times N}}(c)$. Our reference for transversal intersection is [9]. For μ and S satisfying the usual conditions, fix $F = [f_1 \cdots f_N] \in \mathcal{F}(\mu, S)$. It is not difficult to see that the tangent spaces of $\mathbb{T}_{\mathbb{E}^d}(\mu)$ and $\sqrt{S} \cdot \mathrm{St}_{\mathbb{E}^d}(N)$ at F satisfy

$$T_F\mathbb{T}_{\mathbb{E}^d}(\mu) = \{X \in M_{d\times N}(\mathbb{E}) : \mathrm{Re}\,\langle x_n, f_n\rangle = 0 \text{ for all } n \in [N]\}, \tag{2.1.1}$$

and

$$T_F\sqrt{S} \cdot \mathrm{St}_{\mathbb{E}^d}(N) = \{X \in M_{d\times N}(\mathbb{E}) : X = FZ \text{ for some } Z = -Z^*\} \tag{2.1.2}$$

respectively. This last equality follows from the fact that the Lie algebra of $SU(N)$ ($SO(N)$) is the space of skew-Hermitian (skew symmetric) matrices and that $SU(N)$ ($SO(N)$) has a transitive right-action on the connected components of $\sqrt{S} \cdot \mathrm{St}_{\mathbb{E}^d}(N)$ given by $F \mapsto FU$. In addition, we have that

$$T_F\mathcal{S}_{\mathbb{E}^{d\times N}}(c) = \left\{ X = [x_1 \cdots x_N] \in M_{d\times N}(\mathbb{E}) : \sum_{n\in[N]}^{N} \mathrm{Re}\,\langle x_n, f_n\rangle = 0 \right\}. \tag{2.1.3}$$

As shall become overwhelmingly apparent, the regularity of $\mathcal{F}_{\mathbb{E}}(\mu, S)$ at F is intimately connected to the structure of the nonzero entries in the Gram operator,

F^*F. The full Gram operator is cumbersome to deal with, so we replace it by a simple combinatorial object that captures the structure of the nonzero entries.

Definition 2.1.1. The *correlation network* of a frame F with N elements is the undirected graph $\gamma(F) = (V, E)$, where $V = [N]$ and $(i, j) \in E$ if and only if $\langle f_i, f_j \rangle$ is nonzero.

Remark 2.1.2. Intuitively, an edge (i, j) is in the correlation network of F if the elements f_i and f_j are correlated.

Example 1. For the F defined in Section 2.4,

$$[\langle f_i, f_j \rangle]_{(i,j)\in[3]^2} = F^T F = \begin{bmatrix} 1 & \sqrt{2}/2 & 0 \\ \sqrt{2}/2 & 1 & 1/2 \\ 0 & 1/2 & 1 \end{bmatrix}.$$

We conclude that $\gamma(F) = (\{1, 2, 3\}, \{(1, 2), (2, 3)\})$, since f_1, f_3 is the only orthogonal (uncorrelated) pair.

We shall also need to recall Definition 4.8 of [23].

Definition 2.1.3. A frame F is said to be *orthodecomposable* if it can be split into two nontrival subcollections, F_1 and F_2 satisfying $F_1^* F_2 = 0$. That is, span F_1 and span F_2 are nontrivial orthogonal subspaces.

The orthodecomposable frames are generally where complications arise in the analysis of the local structure. For example, the orthodecomposable finite unit-norm frames are the points on the product of spheres that are critical points of the frame force [3], but which do not necessarily minimize the frame potential [12].

We can now state the main theorem of this section, which relates regularity of $\mathcal{F}_{\mathbb{E}}(\mu, S)$ at F with the correlation network $\gamma(F)$ and the orthodecomposability of F.

Theorem 2.1.4. *Let $\mu \in \mathbb{R}_+^N$ and HPD $S \in M_{d\times d}(\mathbb{E})$ satisfy the usual conditions, set $c = \|\mu\|_2$, and suppose $F \in \mathcal{F}_{\mathbb{E}}(\mu, S)$. Then the following are equivalent:*

(i) $T_F\mathbb{T}_{\mathbb{E}^d}(\mu) + T_F\sqrt{S} \cdot St_{\mathbb{E}^d}(N) = T_F\mathcal{S}_{\mathbb{E}^{d\times N}}(c)$;

(ii) *For all $A \in T_F\mathcal{S}_{\mathbb{E}^{d\times N}}(c)$, there is a skew-Hermitian $Z = [z_{ij}]$ which is a solution to the system*

$$Re\,\langle f_i, a_i\rangle = Re \sum_{j\in[N]} z_{ji}\,\langle f_i, f_j\rangle \text{ for all } i \in [N]; \tag{2.1.4}$$

(iii) *F is not orthodecomposable;*

(iv) *$\gamma(F)$ is connected.*

Proof. The equivalence of (iii) and (iv) is trivial. In this proof, we first demonstrate the equivalence of (i) and (ii), then show that (ii) implies (iii), and conclude with proving that (iv) implies (ii).

First we show (i)⇔(ii). Suppose $A = Y + FZ$, where $A = [a_1 \cdots a_N] \in T_F\mathcal{S}_{\mathbb{E}^{d\times N}}(c)$, $Y = [y_1 \cdots y_N] \in T_F\mathbb{T}_{\mathbb{E}^d}(\mu)$, and skew-Hermitian $Z = [z_{ij}]$. Then

$$a_i = y_i + \sum_{j\in[N]} z_{ji} f_j \text{ for all } i \in [N],$$

and hence

$$\langle f_i, a_i\rangle = \langle f_i, y_i\rangle + \sum_{j\in[N]} z_{ji}\,\langle f_i, f_j\rangle \text{ for all } i \in [N].$$

Taking the real part of this and invoking 2.1.1, we have

$$\operatorname{Re}\langle f_i, a_i\rangle = \operatorname{Re}\sum_{j\in[N]} z_{ji}\langle f_i, f_j\rangle \text{ for all } i \in [N].$$

Thus, it is clear that (i)⇒(ii). On the other hand, for any $A \in T_F\mathcal{S}_{\mathbb{E}^{d\times N}}(c)$, (ii) supplies us with a Z satisfying (2.1.4) and we can set $Y = A - FZ \in T_F\mathbb{T}_{\mathbb{E}^d}(\mu)$ to acquire $A = Y + FZ$. This observation finalizes the equivalence of (i) and (ii).

We now show that (ii)⇒(iii) by demonstrating the contrapositive. If F is orthodecomposable, then $FP = [F_1\ F_2]$ for some permutation matrix P, and for some nontrivial F_1 and F_2 with $F_1^*F_2 = 0$. Set

$$A = [A_1\ A_2] = [\operatorname{tr}(F_1^*F_1)^{-1}F_1 \quad -\operatorname{tr}(F_2^*F_2)^{-1}F_2],$$

and note that $\operatorname{Re}\operatorname{tr}(P^*F^*A) = 0$, so $A \in T_{FP}\mathcal{S}_{\mathbb{E}^{d\times N}}(c)$. We have that

$$1 = \operatorname{tr}(F_1^TA_1) = \sum_{f_i\in F_1}\langle f_i, a_i\rangle,$$

and thus (2.1.4) fails for A because of the necessary condition

$$\begin{aligned}
\sum_{f_i\in F_1}\operatorname{Re}\langle f_i, a_i\rangle &= \sum_{f_i\in F_1}\operatorname{Re}\sum_{j\in[N]} z_{ji}\langle f_i, f_j\rangle\\
&= \sum_{f_i\in F_1}\operatorname{Re}\sum_{f_j\in F_1} z_{ji}\langle f_i, f_j\rangle\\
&= \operatorname{Re}\sum_{f_i\in F_1}\sum_{f_j\in F_1} z_{ji}\langle f_i, f_j\rangle \qquad (2.1.5)\\
&= 0.
\end{aligned}$$

This last equality holds since (2.1.5) is the real part of the Hilbert-Schmidt product between a Hermitian matrix $(F_1F_1^*)$ and a skew-Hermitian matrix (Z).

We now establish (iv)⇒(ii). Since $\gamma(F)$ is connected, we can extract a rooted spanning tree $T_0 = (V_0, E_0)$ with $E_0 = [N]$, where we denote the root as i_*. We now construct a nested sequence of trees $T_0 \supset T_1 \supset \ldots \supset T_K = (\{i_*\}, \emptyset)$ by inductively pruning extruding leaves of the tree at each step. More formally, we define the sequence of trees $T_n = (V_n, E_n)$ for $n \geq 1$ inductively by setting $V_n = V_{n-1} \setminus \partial T_{n-1}$ and $E_n = E_{n-1} \setminus \overline{\partial} T_{n-1}$, where ∂T_{n-1} and $\overline{\partial} T_{n-1}$ denote the leaves and pendant edges of T_{n-1} respectively. Note that only finitely many T_n are nontrivial since G is finite, and we let $T_K = (\{i_*\}, \emptyset)$ denote the last nontrivial tree obtained in this process.

We now have the structure necessary to construct a Z solving (2.1.4) for a fixed A. Let A be given, and set $z_{ji} = -z_{ij}^* = 0$ if $(i, j) \notin E_0$. For each $i \in \partial T_0$, there is exactly one parent $j = j(i)$ of i. By definition, we have that $\left\langle f_i, f_{j(i)} \right\rangle \neq 0$, and so we set $z_{j(i),i} = -z_{i,j(i)}^* = \langle f_i, a_i \rangle / \left\langle f_i, f_{j(i)} \right\rangle$. This solves (2.1.4) for all $i \in \partial T_0$.

Applying this strategy inductively, we must solve (2.1.4) for all $i \in \partial T_n$, which reduces to

$$\operatorname{Re} \langle f_i, a_i \rangle = \operatorname{Re} \sum_{j \in \mathrm{Adj}_i} z_{ji} \langle f_i, f_j \rangle = \operatorname{Re} z_{j(i),i} \left\langle f_i, f_{j(i)} \right\rangle + \operatorname{Re} \sum_{j \in \mathrm{Chd}_i} z_{ji} \langle f_i, f_j \rangle ,$$

where Adj_i consists of all the vertices adjacent to i in T_0 and Chd_i consists of all the children of i. By induction, z_{ji} is known for all $j \in \mathrm{Chd}_i$ and we may set

$$z_{j(i),i} = \frac{\langle f_i, a_i \rangle - \sum_{j \in \mathrm{Chd}_i} z_{ji} \langle f_i, f_j \rangle}{\left\langle f_i, f_{j(i)} \right\rangle}.$$

This process continues until $n = K$. At this stage, all z_{ij} are fixed and we compute

$$\begin{aligned} \operatorname{Re} \sum_{i \in [N]} z_{i,i_*} \langle f_{i_*}, f_i \rangle &= -\operatorname{Re} \sum_{i \in \mathrm{Chd}_{i_*}} z_{i_*,i}^* \langle f_{i_*}, f_i \rangle \\ &= -\operatorname{Re} \sum_{i \in \mathrm{Chd}_{i_*}} \left(\frac{\langle f_i, a_i \rangle - \sum_{j \in \mathrm{Chd}_i} z_{j,i} \langle f_i, f_j \rangle}{\langle f_i, f_{i_*} \rangle} \right)^* \langle f_{i_*}, f_i \rangle . \end{aligned}$$

Now note that $\mathrm{Chd}_{i_*} = V_{K-1} \setminus \{i_*\} = \partial T_{K-1}$, so

$$\begin{aligned}
\mathrm{Re} \sum_{i\in[N]} z_{i,i_*} \langle f_{i_*}, f_i \rangle &= -\mathrm{Re} \sum_{i\in V_{K-1}\setminus\{i_*\}} \langle f_i, a_i \rangle + \mathrm{Re} \sum_{i\in\partial T_{K-1}} \sum_{j\in\mathrm{Chd}_i} z_{ji} \langle f_i, f_j \rangle \\
&= -\mathrm{Re} \sum_{i\in V_{K-1}\setminus\{i_*\}} \langle f_i, a_i \rangle \\
&\quad -\mathrm{Re} \sum_{i\in\partial T_{K-1}} \sum_{j\in\mathrm{Chd}_i} \langle f_j, a_j \rangle - \sum_{k\in\mathrm{Chd}_j} z_{kj} \langle f_j, f_k \rangle \\
&= -\mathrm{Re} \sum_{i\in V_{K-2}\setminus\{i_*\}} \langle f_i, a_i \rangle + \mathrm{Re} \sum_{i\in\partial T_{K-2}} \sum_{j\in\mathrm{Chd}_i} z_{ji} \langle f_i, f_j \rangle .
\end{aligned}$$

Inductively expanding and contracting this sum, we obtain

$$\begin{aligned}
\mathrm{Re} \sum_{i\in[N]} z_{i,i_*} \langle f_{i_*}, f_i \rangle &= -\mathrm{Re} \sum_{i\in V_1\setminus\{i_*\}} \langle f_i, a_i \rangle + \mathrm{Re} \sum_{i\in\partial T_1} \sum_{j\in\mathrm{Chd}_i} z_{ji} \langle f_i, f_j \rangle \\
&= -\mathrm{Re} \sum_{i\in V_1\setminus\{i_*\}} \langle f_i, a_i \rangle - \mathrm{Re} \sum_{i\in\partial T_1} \sum_{j\in\mathrm{Chd}_i} \left(\frac{\langle f_j, a_j \rangle}{\langle f_j, f_i \rangle} \right)^* \langle f_i, f_j \rangle \\
&= -\mathrm{Re} \sum_{i\in[N]\setminus\{i_*\}} \langle f_i, a_i \rangle \\
&= \mathrm{Re} \langle f_{i_*}, a_{i_*} \rangle
\end{aligned}$$

since $\mathrm{Re}\,\mathrm{tr}(F^*A) = \mathrm{Re} \sum_{i\in[N]} \langle f_i, a_i \rangle = 0$. Thus, the constructed Z solves (2.1.4), (iv)⇒(ii), and the proof is complete. □

2.2 The tangent space at a nonsingular point of $\mathcal{F}_{\mathbb{E}}(\mu, S)$

In the event that the sum

$$T_F \mathbb{T}_{\mathbb{E}^d}(\mu) + T_F \sqrt{S} \cdot \mathrm{St}_{\mathbb{E}^d}(N) = \{x + y : x \in T_F \mathbb{T}_{\mathbb{E}^d}(\mu), y \in T_F \sqrt{S} \cdot \mathrm{St}_{\mathbb{E}^d}(N)\}$$

equals $T_F \mathcal{S}_{\mathbb{E}^{d\times N}}(c)$, we have the following short exact sequence:

$$0 \to T_F \mathbb{T}_{\mathbb{E}^d}(\mu) \cap T_F \sqrt{S} \cdot \mathrm{St}_{\mathbb{E}^d}(N) \longrightarrow T_F \mathbb{T}_{\mathbb{E}^d}(\mu) \oplus T_F \sqrt{S} \cdot \mathrm{St}_{\mathbb{E}^d}(N) \longrightarrow T_F \mathcal{S}_{\mathbb{E}^{d\times N}}(c) \to 0,$$

where the second arrow is the map $X \mapsto (X, -X)$ and the third arrow is the map $(X, Y) \mapsto X + Y$. This implies

$$\dim(T_F \mathbb{T}_{\mathbb{E}^d}(\mu) \cap T_F \sqrt{S} \cdot \mathrm{St}_{\mathbb{E}^d}(N)) - \dim(T_F \mathbb{T}_{\mathbb{E}^d}(\mu) \oplus T_F \sqrt{S} \cdot \mathrm{St}_{\mathbb{E}^d}(N)) + \dim(T_F \mathcal{S}_{\mathbb{E}^{d\times N}}(c)) = 0,$$

which in turn implies

$$\begin{aligned}\dim(T_F \mathbb{T}_{\mathbb{E}^d}(\mu) \cap T_F \sqrt{S} \cdot \mathrm{St}_{\mathbb{E}^d}(N)) &= \dim(T_F \mathbb{T}_{\mathbb{E}^d}(\mu)) + \dim(T_F \sqrt{S} \cdot \mathrm{St}_{\mathbb{E}^d}(N)) \\ &\quad - \dim(T_F \mathcal{S}_{\mathbb{E}^{d\times N}}(c)). \qquad (2.2.1)\end{aligned}$$

Corollary 2.2.1. *Under the assumptions in Theorem 2.1.4, and assuming that $F \in \mathcal{F}_{\mathbb{E}}(\mu, S)$ is not orthodecomposable, we have*

$$\begin{aligned}T_F \mathcal{F}_{\mathbb{E}}(\mu, S) &= T_F \mathbb{T}_{\mathbb{E}^d}(\mu) \cap T_F \sqrt{S} \cdot \mathrm{St}_{\mathbb{E}^d}(N) \qquad (2.2.2)\\ &= \{X = FZ \in M_{d\times N}(\mathbb{E}) : Z = -Z^* \in M_{N\times N}(\mathbb{E}), \mathrm{Re}\,\mathrm{diag}(F^*X) = 0\}\end{aligned}$$

and

$$\dim(T_F \mathcal{F}_{\mathbb{E}}(\mu, S)) = \begin{cases} (d-1)N + \sum_{n\in[d]} (N-n) - (dN-1) & \text{if } \mathbb{E} = \mathbb{R} \\ (2d-1)N + \sum_{n\in[d]} (2N-2n+1) - (2dN-1) & \text{if } \mathbb{E} = \mathbb{C} \end{cases} \qquad (2.2.3)$$

as a vector space over $\mathbb{R}$.

Proof. Note that the dimension of a manifold equals the dimension of the tangent space at any point on the manifold. Now, the "$\subset$" direction of (2.2.2) is clear since $\mathcal{F}_{\mathbb{E}}(\mu, S)$ is contained in both $\mathbb{T}_{\mathbb{E}^d}(\mu)$ and $\sqrt{S} \cdot \mathrm{St}_{\mathbb{E}^d}(N)$. On the other hand, F not orthodecomposable is equivalent to (i) of Theorem 2.1.4, which combines with (2.2.1) to imply

$$\begin{aligned}
\dim(\mathcal{F}_{\mathbb{E}}(\mu, S) \cap \mathcal{B}) &= \dim(\mathbb{T}_{\mathbb{E}^d}(\mu)) + \dim(\sqrt{S} \cdot \mathrm{St}_{\mathbb{E}^d}(N)) - \dim(\mathcal{S}_{\mathbb{E}^{d\times N}}(c)) \\
&= \dim(T_F \mathbb{T}_{\mathbb{E}^d}(\mu)) + \dim(T_F \sqrt{S} \cdot \mathrm{St}_{\mathbb{E}^d}(N)) - \dim(T_F \mathcal{S}_{\mathbb{E}^{d\times N}}(c)) \\
&= \dim(T_F \mathbb{T}_{\mathbb{E}^d}(\mu) \cap T_F \sqrt{S} \cdot \mathrm{St}_{\mathbb{E}^d}(N)).
\end{aligned}$$

for some open ball $\mathcal{B} \subset M_{d\times N}(\mathbb{E})$. Thus, $T_F \mathcal{F}_{\mathbb{E}}(\mu, S)$ and $T_F \mathbb{T}_{\mathbb{E}^d}(\mu) \cap T_F \sqrt{S} \cdot \mathrm{St}_{\mathbb{E}^d}(N)$ have the same dimension and we may conclude their equality. Moreover, combining (1.4.5), (1.4.6), and (1.4.7) with (2.2.1) yields (2.2.3). □

It should be noted that this dimension calculation is a generalization of the one encountered in [23] and equivalent to the one performed in [40].

2.3 Explicit bases for $T_F \mathcal{F}_{\mathbb{E}}(\mu, S)$

In order to construct an explicit basis for $T_F \mathcal{F}_{\mathbb{E}}(\mu, S)$, we must first extract a non-orthodecomposable basis for $\mathbb{E}^d$. The next proposition demonstrates that this is possible at the nonsingular points.

Proposition 2.3.1. *Suppose μ and S satisfy the usual conditions. Then $F = [f_1 \cdots f_N] \in \mathcal{F}_{\mathbb{E}}(\mu, S)$ is not orthodecomposable if and only if there is a d-set $A \subset [N]$ such that F_A is a non-orthodecomposable basis for $\mathbb{E}^d$*

Proof. First suppose that F contains a non-orthodecomposable basis. Since this basis cannot be split into nontrivial mutually orthogonal collections, any splitting of F into two mutually orthogonal sets must be trivial. Thus, F is not orthodecomposable.

If F is not orthodecomposable, then we build A inductively beginning with $A = \{i\}$ for any $i \in [N]$. Suppose we have added indices to A so that F_A is not orthodecomposable and also linearly independent. If F_A is a basis for $\mathbb{E}^d$, then we stop. Otherwise, there is an index j such that $F_{A\cup\{j\}}$ is not orthodecomposable and linearly dependent (if not, then F_A and $F_{[N]\setminus A}$ is a nontrivial splitting of F into mutually orthogonal collections). Once $|A| = d$, this process ceases and F_A is a non-orthodecomposable basis. $\square$

We shall now describe the process for constructing a basis for $T_F\mathcal{F}_{\mathbb{E}}(\mu, S)$. Let A be the set of indices from Proposition 2.3.1. Since $\gamma(F)$ restricted to the indices in A is connected by Theorem 2.1.4 let $T \subset \gamma(F)$ be a rooted spanning tree on A. Let T_0 be the graph containing just the root of T and inductively define T_{k+1} to be the subtree of T containing all of the children of T_k. Since T is finite, we let n denote the first k such that $T_k = T$, and we shall use ν_k to denote the vertices in T_{n-k+1} but not in T_{n-k} for all $k \in [n]$. For each $i \in A$ that is not the root, there is a smallest k such that i is a vertex in T_k, and we let $\rho(i)$ be the unique parent of i in T_k.

The basis that we shall construct is most neatly defined as a union of two disjoint collections indexed by the sets

$$\Lambda_1 = \begin{cases} ([N] \setminus A) \times [d-1] & \text{if } \mathbb{E} = \mathbb{R} \\ ([N] \setminus A) \times [2d-1] & \text{if } \mathbb{E} = \mathbb{C} \end{cases}$$

and

$$\Lambda_2 = \{(i, j, 0) \in A \times A \times \{0\} : i < j \text{ and } (i, j) \notin E(T)\}$$

if $\mathbb{E} = \mathbb{R}$, and

$$\Lambda_2 = \{(i, j, \epsilon) \in A \times A \times \{0, 1\} : i < j \text{ and } (i, j) \notin E(T), \text{ or } i = j \text{ and } \epsilon = 1\}$$

if $\mathbb{E} = \mathbb{C}$. We now proceed to describe the procedure for constructing members indexed by Λ_1.

For each $i \in [N] \setminus A$, we fix an orthonormal basis $\{y_j^{(i)}\}_{j \in [2^{\delta_{\mathbb{E},\mathbb{C}}} d-1]}$ of $T_{f_i} \mathcal{S}_{\mathbb{E}^d}(\mu_i)$, where $\delta_{\mathbb{E},\mathbb{C}} = 1$ if $\mathbb{E} = \mathbb{C}$ and equals zero otherwise. For every $\alpha = (i, j) \in \Lambda_1$, we shall construct an N by N skew-symmetric (or skew-Hermitian) $Z^\alpha = [z_{kl}^\alpha]$ so that $V^\alpha = FZ^\alpha$ is a member of the basis that we seek to construct.

First, we fix the i^{th} column and row of Z^α:

$$Z_{A \times \{i\}}^\alpha = -(Z_{\{i\} \times A}^\alpha)^* = F_A^{-1} y_j^{(i)} \tag{2.3.1}$$

and

$$Z_{([N] \setminus A) \times \{i\}}^\alpha = -(Z_{\{i\} \times ([N] \setminus A)}^\alpha)^* = 0. \tag{2.3.2}$$

For all $k \in [N] \setminus A$ with $k \neq i$, we set

$$Z_{[N] \times \{k\}}^\alpha = -(Z_{\{k\} \times [N]}^\alpha)^* = 0. \tag{2.3.3}$$

All that remains is to fix $Z_{A \times A}^\alpha$, and we first set $z_{kl}^\alpha = -(z_{lk}^\alpha)^* = 0$ for all $k < l$ such that $(k, l) \in A \times A$ is not an edge in T. For each $k \in \nu_m$, we inductively define

$$z_{\rho(k),k}^\alpha = -(z_{k,\rho(k)}^\alpha)^* = -\left(\operatorname{Re} \sum_{l \in [N] \setminus \{k, \rho(k)\}} \langle f_l, f_k \rangle \, z_{lk} \right) / \left\langle f_{\rho(k)}, f_k \right\rangle \tag{2.3.4}$$

as $m \in [n]$ increases. At each step of this induction, the sum is well defined because the values of z_{kl}^α have all been fixed by the previous step. Also, $\left\langle f_k, f_{\rho(k)} \right\rangle \neq 0$ since

$(k, \rho(k))$ or $(\rho(k), k)$ is an edge in $T \subset \gamma(F)$. We conclude our construction of Z^α by ensuring that it has zero diagonal.

Proposition 2.3.2. *With Z^α the skew-symmetric (skew-Hermitian) matrix constructed above, $V^\alpha = FZ^\alpha$ satisfies*

(i) the i^{th} column of V^α is $y_j^{(i)}$;

(ii) for all $k \in [N] \setminus (A \cup \{i\})$, the k^{th} column of V^α is 0;

(iii) $V^\alpha \in T_F \mathcal{F}_{\mathbb{E}}(\mu, S)$.

Proof. Because of (2.3.1) and (2.3.2), we have that the i^{th} column of $V^\alpha = FZ^\alpha$ is

$$FZ^\alpha_{[N]\times\{i\}} = F_A Z^\alpha_{A\times\{i\}} + F_{[N]\setminus A} Z^\alpha_{([N]\setminus A)\times\{i\}} \tag{2.3.5}$$

$$= F_A F_A^{-1} y_j^{(i)} + F_{[N]\setminus A} 0 \tag{2.3.6}$$

$$= y_j^{(i)}. \tag{2.3.7}$$

This shows that (i) holds. Our choice in (2.3.3) ensure that V^α satisfies property (ii), and we now proceed to demonstrate that (iii) holds.

Since $\gamma(F)$ is connected, Corollary 2.2.1 implies that

$$T_F \mathcal{F}_{\mathbb{E}}(\mu, S) = T_F \sqrt{S} \cdot \mathrm{St}_{\mathbb{E}^d}(N) \cap T_F \mathbb{T}_{\mathbb{E}^d}(\mu). \tag{2.3.8}$$

Clearly,

$$V^\alpha = FZ^\alpha \in T_F \sqrt{S} \cdot \mathrm{St}_{\mathbb{E}^d}(N)\{X = FZ \in M_{d\times N}(\mathbb{E}) : Z = -Z^*\},$$

so (2.3.8) implies that we only need to show

$$V^\alpha = [v_1^\alpha \cdots v_N^\alpha] \in T_F \mathbb{T}_{\mathbb{E}^d}(\mu). \tag{2.3.9}$$

For $k = i$, we have that $v_i^\alpha = y_j^i \in T_{f_i}\mathcal{S}_{\mathbb{E}^d}(\mu_i)$ and hence

$$\mathrm{Re}\,\langle v_k^\alpha, f_k\rangle = \mathrm{Re}\left\langle y_j^{(i)}, f_i\right\rangle = 0.$$

If $k \in [N] \setminus A$ and $k \neq i$, then $v_k^\alpha = 0$ and

$$\mathrm{Re}\,\langle v_k^\alpha, f_k\rangle = 0.$$

For each $m \in [n]$, and each $k \in \nu_m$ we have

$$\begin{aligned}
\mathrm{Re}\,\langle v_k^\alpha, f_k\rangle &= \mathrm{Re}\left\langle \sum_{l\in[N]\setminus\{k\}} f_l z_{lk}^\alpha, f_k\right\rangle = \mathrm{Re}\sum_{l\in[N]\setminus\{k\}} \langle f_l, f_k\rangle\, z_{lk}^\alpha && (2.3.10)\\
&= \mathrm{Re}\left\langle f_{\rho(k)}, f_k\right\rangle z_{\rho(k),k}^\alpha + \mathrm{Re}\sum_{l\in[N]\setminus\{k,\rho(k)\}} \langle f_l, f_k\rangle\, z_{lk}^\alpha && (2.3.11)\\
&= -\mathrm{Re}\sum_{l\in[N]\setminus\{k,\rho(k)\}} \langle f_l, f_k\rangle\, z_{lk}^\alpha + \mathrm{Re}\sum_{l\in[N]\setminus\{k,\rho(k)\}} \langle f_l, f_k\rangle\, z_{lk}^\alpha && (2.3.12)\\
&= 0 && (2.3.13)
\end{aligned}$$

by (2.3.4). Finally, if k is the root of T, then

$$\mathrm{Re}\,\langle v_k^\alpha, f_k\rangle = 0$$

since we have verified that $\mathrm{Re}\,\langle v_l^\alpha, f_l\rangle = 0$ for $l \in [N] \setminus \{k\}$ and Z^α skew-Hermitian (orthogonal to F^*F Hermitian) implies

$$\sum_{k\in[N]} \mathrm{Re}\,\langle v_k^\alpha, f_k\rangle = \mathrm{Re}\sum_{k\in[N]}\sum_{l\in[N]} z_{lk}^\alpha\, \langle f_k, f_l\rangle^* = \mathrm{Re}\,\langle Z^\alpha, F^*F\rangle = 0$$

□

For each $\alpha = (i, j, \epsilon) \in \Lambda_2$, we now construct a skew-symmetric(Hermitian) matrix Z^α so that $W^\alpha = FZ^\alpha$ is a member of the second collection. For this collection, we immediately set

$$z_{kl}^\alpha = -(z_{lk}^\alpha)^* = 0$$

for all $(k, l) \notin A \times A$, and all $(k, l, \epsilon) \in \Lambda_2 \setminus \{\alpha\}$ with $k < l$. We set

$$z_{ij}^{\alpha} = -(z_{ij}^{\alpha})^* = \sqrt{-1}^{\epsilon}, \tag{2.3.14}$$

and inductively define

$$z_{\rho(k),k}^{\alpha} = -(z_{k,\rho(k)}^{\alpha})^* = -\left(\mathrm{Re} \sum_{l \in [N] \setminus \{k, \rho(k)\}} \langle f_l, f_k \rangle \, z_{lk}\right) / \left\langle f_{\rho(k)}, f_k \right\rangle \tag{2.3.15}$$

for $k \in \nu_m$ as $m \in [n]$ increases. At this point, all z_{kl}^{α} have been defined except for diagonal elements, and the undefined diagonal elements are now set to zero. The proof of the following proposition follows from an argument similar to the one used to prove Proposition 2.3.2.

Proposition 2.3.3. *With Z^{α} the skew-symmetric (skew-Hermitian) matrix constructed above, $W^{\alpha} = FZ^{\alpha}$ satisfies*

(i) for all $k \in [N] \setminus A$, the k^{th} column of W^{α} is 0;

(ii) $W^{\alpha} \in T_F \mathcal{F}_{\mathbb{E}}(\mu, S)$.

We conclude this subsection by verifying that $\{V^{\alpha}\}_{\alpha \in \Lambda_1} \cup \{W^{\alpha}\}_{\alpha \in \Lambda_2}$ is a basis for the tangent space.

Proposition 2.3.4. *Under the assumptions of this subsection, $\Omega = \{V^{\alpha}\}_{\alpha \in \Lambda_1} \cup \{W^{\alpha}\}_{\alpha \in \Lambda_2}$ is a basis for $T_F \mathcal{F}_{\mathbb{E}}(\mu, S)$.*

Proof. We know that each member of Ω is in $T_F \mathcal{F}_{\mathbb{E}}(\mu, S)$ by Propositions 2.3.2 and 2.3.3. Since the number of elements in Ω coincides with the dimension of $T_F \mathcal{F}_{\mathbb{E}}(\mu, S)$, if Ω is a linearly independent collection, then it is a basis. We now show that Ω is a linearly independent collection.

Suppose

$$\sum_{\alpha\in\Lambda_1} a_\alpha V^\alpha + \sum_{\alpha\in\Lambda_2} b_\alpha W^\alpha = 0, \tag{2.3.16}$$

where $\{a_\alpha\}_{\alpha\in\Lambda_1} \cup \{b_\alpha\}_{\alpha\in\Lambda_2} \in \mathbb{R}$. For each $i \in [N] \setminus A$, this means that

$$\sum_{\alpha\in\Lambda_1} a_\alpha v_i^\alpha + \sum_{\alpha\in\Lambda_2} b_\alpha w_i^\alpha = 0.$$

Part (i) and (ii) of Proposition 2.3.2 and part (i) of Proposition 2.3.3 then imply that

$$\sum_{j\in[2^{\delta_{\mathbb{E},\mathbb{C}}}d-1]} a_{(i,j)} y_j^{(i)} = 0.$$

Since the $y^{(i)_j}$ were chosen to be orthonormal, they are linearly independent and we conclude that $a_{(i,j)} = 0$ for all $j \in [2^{\delta_{\mathbb{E},\mathbb{C}}}d-1]$. Since i was arbitrary, $a_\alpha = 0$ for all $\alpha \in \Lambda_1$ and (2.3.16) reduces to

$$\sum_{\alpha\in\Lambda_2} b_\alpha W^\alpha = 0.$$

This in turn reduces to

$$\sum_{\alpha\in\Lambda_2} b_\alpha F_A Z_{A\times A}^\alpha = 0.$$

by part (i) of Proposition 2.3.3. By our choice of A, F_A is invertible, and this equation becomes

$$\sum_{\alpha\in\Lambda_2} b_\alpha Z_{A\times A}^\alpha = 0.$$

By construction, $\{Z_{A\times A}^{(i,j,\epsilon)}\}_{\epsilon\in\{0,1\}}$ are the only Z^α in this collection that have $z_{ij}^\alpha \neq 0$ if $i < j$. This implies that

$$b_{(i,j,0)} Z^{(i,j,0)} + b_{(i,j,1)} Z^{(i,j,1)} = 0.$$

Since $z_{ij}^{(i,j,0)}$ is real and $z_{ij}^{(i,j,1)}$ is purely imaginary, we conclude that $b_{(i,j,\epsilon)} = 0$ for $\epsilon = 0, 1$. Since i and j were arbitrary, and $b_{(i,i,1)} = 0$ for all $i \in A$ follows similarly, we conclude that $b_\alpha = 0$ for all $\alpha \in \Lambda_2$. This concludes the proof of linear independence. $\square$

Chapter 3

Explicit analytic coordinate systems on $\mathcal{F}_{\mathbb{E}}(\mu, S)$

3.1 A motivating example

As we briefly mentioned in the introduction, the (μ, S)-frame varieties should consist of frames that locally split into a collection that one can freely articulate, and a basis. The motion of the basis compensates for the motion of the collection that can be freely articulated, but the complication that arises is that the basis contributes to the degrees of freedom. This does not occur in $\mathbb{R}^2$, and this is one reason why (μ, S)-frame varieties are easily characterized if $d = 2$ and $\mathbb{E} = \mathbb{R}$ (see Theorem 3.3 of [3]). For $d > 2$, things quickly become more complicated.

In this example, we demonstrate how coordinates can be obtained for a space of bases in $\mathbb{R}^3$ with fixed lengths and a fixed frame operator. This is the simplest nontrivial case, but our approach requires a decent amount of effort. The benefit is that the approach of this example works in general, with minor modifications.

We consider the case $\mathbb{E} = \mathbb{R}$, $N = d = 3$, $\mu = [1\ 1\ 1]^T$,

$$F = \begin{bmatrix} 1 & \sqrt{2}/2 & 0 \\ 0 & \sqrt{2}/2 & \sqrt{2}/2 \\ 0 & 0 & \sqrt{2}/2 \end{bmatrix}, \text{ and } S = FF^T = \begin{bmatrix} 3/2 & 1/2 & 0 \\ 1/2 & 1 & 1/2 \\ 0 & 1/2 & 1/2 \end{bmatrix}.$$

Application of Corollary 2.2.1 shows that

$$\dim(\mathcal{F}_{\mathbb{R}}(\mu, S)) = 1$$

as a real algebraic variety. Consequently, we seek formal parameterizations of $\mathcal{F}_{\mathbb{R}}(\mu, S)$ with the form

$$F(t) = \begin{bmatrix} x_1(t) & y_1(t) & z_1(t) \\ x_2(t) & y_2(t) & z_2(t) \\ t & y_3(t) & z_3(t) \end{bmatrix}, F(0) = F.$$

The constraints are $\mathrm{diag}(F^T(t)F(t)) = [1\ 1\ 1]^T$ and

$$F(t)F(t)^T = S \Longleftrightarrow F(t)^T S^{-1} F(t) = F(t)^T \begin{bmatrix} 1 & -1 & 1 \\ -1 & 3 & -3 \\ 1 & -3 & 5 \end{bmatrix} F(t) = I_{3\times 3}.$$

We proceed inductively through the columns of $F(t)$. The constraints that only involve the first column are the normality condition and the condition imposed by $S_{11} = 1$,

$$\begin{aligned} x_1^2 + x_2^2 + t^2 &= 1 \\ x_1^2 + 3x_2^2 + 5t^2 - 2x_1x_2 + 2x_1t - 6x_2t &= 1. \end{aligned}$$

Viewing these two multinomials as polynomials in x_2 with coefficients in x_1 and t, we have

$$\begin{aligned} x_2^2 + (x_1^2 + t^2 - 1) &= 0 \\ 3x_2^2 + (-2x_1 - 6t)x_2 + (x_1^2 + 5t^2 + 2x_1t - 1) &= 0. \end{aligned}$$

To perform elimination on this system, we need to invoke the following proposition (which is a simple application of Gaussian elimination):

Proposition 3.1.1. Suppose $\alpha_i, \beta_i \in \mathbb{R}$ for $i = 0, 1, 2$ and $\alpha_2, \beta_2 \neq 0$. The quadratics $p = \alpha_2\xi^2 + \alpha_1\xi + \alpha_0$ and $q = \beta_2\xi^2 + \beta_1\xi + \beta_0$ have a mutual zero if and only if the Bézout determinant (see [36]) satisfies

$$\mathrm{Bz}(p, q) = (\alpha_2\beta_1 - \alpha_1\beta_2)(\alpha_1\beta_0 - \alpha_0\beta_1) - (\alpha_2\ beta_0 - \alpha_0\beta_2)^2 = 0. \qquad (3.1.1)$$

Applying this propostion to the last two quadratics, we can eliminate x_2 to obtain

$$\begin{aligned} 0 &= [(1)(-2x_1 - 6t) - 0(3)][0(x_1^2 + 5t^2 + 2x_1t - 1) - (x_1^2 + t^2 - 1)(-2x_1 - 6t)] \\ &\quad -[(1)(x_1^2 + 5t^2 + 2x_1t - 1) - (3)(x_1^2 + t^2 - 1)]^2 \\ &= 8x_1^4 + 16tx_1^3 + (36t^2 - 12)x_1^2 + (32t^3 - 16t)x_1 + (40t^4 - 28t^2 + 4). \end{aligned}$$

Solving for x_1 in terms of t (using a computer algebra package!), we obtain the four possible solutions:

$$x_1(t) = \pm\sqrt{1 - 2t^2}, -t \pm \frac{1}{2}\sqrt{-6t^2 + 2}$$

The condition $x_1(0) = 1$ leaves us with just one possible solution:

$$x_1(t) = \sqrt{1 - 2t^2}$$

and we readily verify that this implies $x_2(t) = t$. Having solved for the first column, we consider the contraints that have not been satisfied, but which only depend on the first and second columns:

$$\begin{aligned} y_1^2 + y_2^2 + y_3^2 &= 1 \\ y_1^2 + 3y_2^2 + 5y_3^2 - 2y_1y_2 + 2y_1y_3 - 6y_2y_3 &= 1 \\ x^T S^{-1} y = \sqrt{1 - 2t^2}y_1 - \sqrt{1 - 2t^2}y_2 + (\sqrt{1 - 2t^2} + 2t)y_3 &= 0 \qquad (3.1.2) \end{aligned}$$

Locally, we know that $x_1(t) \neq 0$, so we may solve the third equation for y_1 to obtain

$$y_1 = y_2 - (1 + 2t/\sqrt{1-2t^2})y_3.$$

This allows us to elimnate y_1 from the first two equations, and we may view these new equations as quadratics in y_2 with coefficients in y_3 and t:

$$\begin{aligned} 2y_2^2 + [(-2 - 4t/\sqrt{1-2t^2})y_3]y_2 + [(2 + 4t/\sqrt{1-2t^2} + 4t^2/(1-2t^2))y_3^2 - 1] &= 0 \\ 2y_2^2 + [-4y_3]y_2 + [(4 + 4t^2/(1-2t^2))y_3^2 - 1] &= 0 \end{aligned}$$

We now solve for y_3 in terms of t so that the Bézout determinant of this system vanishes, and we obtain only three solutions,

$$y_3(t) = 0, \pm\frac{1}{2}\sqrt{2-4t^2}.$$

Since $y_3(0) = 0$, we are left with the solution $y_3(t) = 0$. Substitution into (3.1.2) immediately implies that $y_1(t) = y_2(t)$ for all t, so we must conclude that $y_1(t) = y_2(t) = \sqrt{2}/2$ for all t.

We now solve for the final column, z. Noting that conditions on y are also imposed upon z, we see that

$$z_3(t) = 0, \pm\frac{1}{2}\sqrt{2-4t^2}.$$

However, $z_3(0) = \sqrt{2}/2$, so we have that $z_3(t) = \frac{1}{2}\sqrt{2-4t^2}$. A similar line of reasoning reveals that

$$z_2(t) = \pm\sqrt{2}/2, \pm\frac{1}{2}\sqrt{2-4t^2}.$$

Invoking the orthogonality condition,

$$y^T S^{-1} z = \sqrt{2}z_2 - \sqrt{2}z_3 = 0$$

we may eliminate the constant solutions, and $z_2(0) = \sqrt{2}/2$ leaves us with $z_2(t) = \frac{1}{2}\sqrt{2-4t^2}$. Using the spherical condition $z_1^2 + z_2^2 + z_3^2 = 1$ and the orthogonality condition $x^T S^{-1} z = 0$, we obtain $z_1(t) = -\sqrt{2}t$. Thus, the final solution is

$$F(t) = \begin{bmatrix} \sqrt{1-2t^2} & \sqrt{2}/2 & -\sqrt{2}t \\ t & \sqrt{2}/2 & \frac{1}{2}\sqrt{2-4t^2} \\ t & 0 & \frac{1}{2}\sqrt{2-4t^2} \end{bmatrix}.$$

This parameterization is relatively simple because the first and third columns form an orthonormal basis of $\text{span}\{x(0), z(0)\}$ for all t. If we had observed this at the beginning of the example, the parameterizations would follow very quickly. However, a generic frame does not contain an orthonormal basis and the approach of this example is generically effective.

3.2 Existence of structured, local coordinate systems on $\mathcal{F}_{\mathbb{R}}(\mu, S)$

In this section, we demonstrate the existence of structured, locally well-defined charts around a generic $F \in \mathcal{F}_{\mathbb{R}}(\mu, S)$. An essential ingredient for the existence proofs of this section is the extraction of a "good" basis B from F. After B is extracted, we can shuffle/rotate $\mathcal{F}_{\mathbb{R}}(\mu, S)$ on the Hilbert-Schmidt sphere so that an orthogonal projection onto a relatively simple subspace of $M_{d\times N}(\mathbb{R})$ has the transformed version of F as a regular point of the orthogonal projection (that is, the Jacobian applied to the respective tangent spaces is onto).

3.2.1 The case $N = d$

If $N = d$, then Corollary 2.2.1 tells us that the intrinsic dimension of a generic, nonempty $\mathcal{F}_{\mathbb{R}}(\mu, S)$ is

$$\begin{aligned}
\dim(\mathcal{F}_{\mathbb{R}}(\mu, S)) &= (d-1)d + \sum_{n\in[d]}(d-n) - (d^2-1) \\
&= d^2 - d + \sum_{n\in[d-1]} n - d^2 + 1 \\
&= -d + 1 + \sum_{n\in[d-1]} n \\
&= \sum_{n\in[d-2]} n,
\end{aligned}$$

which leads us to search for local coordinates around $F \in \mathcal{F}_{\mathbb{R}}(\mu, S)$ that depend solely upon the entries below the subdiagonal. Ultimately, this is not what holds generally, but the partial result that we obtain is more easily digested than the general result.

First, we establish some notation. Let

$$\Delta_{\mathbb{R}} = \{L \in M_{d\times d}(\mathbb{R}) : l_{ij} = 0 \text{ if } i \leq j+1\}$$

denote the space of lower triangular matrices with zeros on the diagonal and subdiagonal, and note that

$$\dim(\Delta_{\mathbb{R}}) = \sum_{n\in[d-2]} n$$

since there are $d-2$ free variables in the first column, $d-3$ in the second column, and so forth. Our goal is to show the existence of a local parameterization with a form analogous to

$$
\begin{aligned}
\Phi(L) &= \Phi\left(\begin{bmatrix} 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ l_{31} & 0 & \cdots & 0 & 0 & 0 & 0 \\ l_{41} & l_{42} & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ l_{d-1,1} & l_{d-1,2} & \cdots & l_{d-1,d-3} & 0 & 0 & 0 \\ l_{d1} & l_{d2} & \cdots & l_{d,d-3} & l_{d,d-2} & 0 & 0 \end{bmatrix}\right) \\
&= \begin{bmatrix} \phi_{11}(L) & \phi_{12}(L) & \cdots & \phi_{1,d-3}(L) & \phi_{1,d-2}(L) & \phi_{1,d-1}(L) & \phi_{1d}(L) \\ \phi_{21}(L) & \phi_{22}(L) & \cdots & \phi_{2,d-3}(L) & \phi_{2,d-2}(L) & \phi_{2,d-1}(L) & \phi_{2d}(L) \\ l_{31} & \phi_{32}(L) & \cdots & \phi_{3,d-3}(L) & \phi_{3,d-2}(L) & \phi_{3,d-1}(L) & \phi_{3d}(L) \\ l_{41} & l_{42} & \cdots & \phi_{4,d-3}(L) & \phi_{4,d-2}(L) & \phi_{4,d-1}(L) & \phi_{4d}(L) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ l_{d-1,1} & l_{d-1,2} & \cdots & l_{d-1,d-3} & \phi_{d-1,d-2}(L) & \phi_{d-1,d-1}(L) & \phi_{d-1,d}(L) \\ l_{d1} & l_{d2} & \cdots & l_{d,d-3} & l_{d,d-2} & \phi_{d,d-1}(L) & \phi_{dd}(L) \end{bmatrix} \\
&= [\phi_1(L)\ \phi_2(L)\ \cdots \phi_d(L)],
\end{aligned}
$$

where $\Phi : \mathcal{N} \to \mathcal{F}_{\mathbb{R}}(\mu, S)$ with $\mathcal{N} \subset \Delta_{\mathbb{R}}$ a neighborhood of 0. We now state a partial result whose proof we shall use as a scaffolding for the more general results to come. For the following proposition, recall that a Hamiltonian path in a graph is a chain that visits each vertex exactly once.

Proposition 3.2.1. *Let $\mu \in \mathbb{R}^d_+$, suppose $S \in M_{d\times d}(\mathbb{R})$ is SPD, and assume that μ and S satisfy the usual conditions. If $\gamma(F)$ of $F \in \mathcal{F}_{\mathbb{R}}(\mu, S)$ contains a Hamiltonian* *path, then there is an orthogonal matrix $Q \in O(d)$ and a d by d permutation matrix P such that the orthogonal projection*

$$\pi : T_{Q^T F P^T} \mathcal{F}_{\mathbb{R}}(P\mu, Q^T S Q) \to \Delta_{\mathbb{R}}$$

is injective.

We then obtain the following corollary due to the Real-Analytic Inverse Function Theorem (see [30]) coupled with a counting argument.

Corollary 3.2.2. *If the conditions of Theorem 3.2.1 are satisfied, then there is a unique, locally well-defined, real-analytic inverse of π, $\Phi : \Delta_{\mathbb{R}} \to \mathcal{F}_{\mathbb{R}}(P\mu, Q^T S Q)$.*

Remark 3.2.3. For Φ given above, $Q\Phi(L)P$ is a parameterization about $F \in \mathcal{F}_{\mathbb{R}}(\mu, S)$.

proof of Proposition 3.2.1. First, note that F is not orthodecomposable since existence of a Hamiltonian path implies that $\gamma(F)$ is connected. Moreover, for any orthogonal Q and any permutation matrix P, we have that

$$(Q^T F P^T)^T (Q^T F P^T) = P F^T Q Q^T F P^T = P F^T F P^T,$$

and hence $\gamma(Q^T F P^T)$ is obtained by permuting the nodes of $\gamma(F)$. Consequently, $Q^T F P^T \in \mathcal{F}_{\mathbb{R}}(P\mu, Q^T S Q)$ is not orthodecomposable.

Now, we construct P and Q. Fix a terminal vertex of the Hamiltonian path, i_1, and let i_k denote the k^{th} ancestor of i_1 in the Hamiltonian path. The map $i_k \mapsto k$ is then a permutation of $[d]$, and we let P be the matrix realization of this permutation acting on rows. Let Q denote the unique (since FP^T is of full rank) orthogonal matrix obtained from the QR-decomposition of FP^T.

We shall now work exclusively with the nonorthodecomposable, upper triangular matrix $R = Q^T F P^T$, noting that $G = R^T R$ has nonzero entries along the subdiagonal by our choice of P. Also, set $\nu = P\mu$ and $S' = Q^T S Q$. Let $\{e_i\}_{i \in [d]}$ denote the standard orthonormal basis, and set $E_k = \text{span}\{e_i\}_{i \in [k]}$. It is trivial to establish that each E_k is an invariant subspace of R^{-1}, and we shall use this fact momentarily.

Suppose $\pi(X) = 0$, where $X = [x_1 \cdots x_d] \in T_R \mathcal{F}(\nu, S')$,

$T_R \mathcal{F}(\nu, S') = \{X = RZ \in M_{d \times d}(\mathbb{R}) : Z = -Z^T \text{ and } \text{diag}(GZ) = 0\}$,

and note that $x_k = R z_k \in E_{k+1}$ for each $k < d$ because $\pi(X) = 0$. In particular, we have that $x_1 \in E_2$.

Now, $z_1 \in E_1^\perp$ since it is the first column in a skew symmetric matrix. Additionally, $z_1 = R^{-1} x_1$ and hence $z_1 \in E_2$ since E_2 is an invariant subspace of R^{-1}. Combining these two facts, we have that $z_1 = z_{21} e_2$. Now, the condition $\text{diag}(GZ) = 0$ implies

$$0 = e_1^T G Z e_1 = z_{21} e_1^T G e_2 = z_{21} g_{12}$$

By construction, $g_{12} \neq 0$, and we have that $z_1 = 0$ and $x_1 = R z_1 = 0$.

Proceeding inductively, suppose that the first $k \leq d - 2$ columns of X are zero. By invertibility of R, we must have that the first k columns of Z are also zero and hence $z_{k+1} \in E_{k+1}^\perp$. Since $x_{k+1} \in E_{k+2}$, it follows that $z_{k+1} = R^{-1} x_{k+1} \in E_{k+2}$. Consequently, $z_{k+1} = z_{k+2,k+1} e_{k+2}$ and

$$0 = e_{k+1}^T G Z e_{k+1} = z_{k+2,k+1} e_{k+1}^T G e_{k+2} = z_{k+2,k+1} g_{k+1,k+2}$$

implies $z_{k+1} = 0$ since $g_{k+1,k+2} \neq 0$ by construction. Thus, $x_{k+1} = Rz_{k+1} = 0$.

By induction, the first $d-1$ columns of X are zero and hence the first $d-1$ columns of Z are also zero. Since Z is skew symmetric, we conclude that $Z = 0$ and $X = RZ = 0$. Thus, π is injective and the proof is complete. $\square$

Remark 3.2.4. If $d > 3$, then we are not ensured the existence of a Hamiltonian path in the correlation network of a nonsingular point in $\mathcal{F}_{\mathbb{E}}(\mu, S)$. On the other hand, an argument in the spirit of this proposition also carries through if the correlation network contains a vertex which shares an edge with every other vertex, but if $d > 4$, then we are not ensured the existence of such a vertex or a Hamiltonian path in the correlation network of a nonsingular point in $\mathcal{F}_{\mathbb{E}}(\mu, S)$.

Moreover, shuffling around Δ_R and finding a clever choice of Q and P cannot extend this argument since any connected graph can be a correlation network. This means that, in general, application of the inductive argument ends on a column z_k of Z for which the information in x_k is insufficient to annihilate all but one entry of z_k.

The key to extending the previous proposition is to see the application of Q as the action of

$$\mathbf{Q}_* = [Q\, Q \cdots Q] \in \prod_{i \in [d]} O(d) = O^d(d)$$

on $\mathbb{T}_{\mathbb{E}^d}(P\mu)$, which is denoted $\mathbf{Q}_* \star X$. Thus, to generalize the proposition, we exploit the fact that $\mathbb{T}_{\mathbb{E}^d}(P\mu)$ is a homogeneous space for $O^d(d)$. For a general $\mathbf{Q} = [Q_1 \cdots Q_d] \in O^d(d)$,

$$\mathbf{Q} \star X = [Q_1 x_1 \cdots Q_d x_d]$$

and we then have that

$$\mathbf{Q}_* \star X = [Qx_1 \cdots Qx_d] = QX.$$

We shall also use $\mathbf{Q}^T$ to denote $[Q_1^T \cdots Q_d^T]$. From this vantage point, the central argument of the preceding proposition can be applied almost verbatim to obtain the general result.

Theorem 3.2.5. *Let $\mu \in \mathbb{R}_+^d$, suppose $S \in M_{d\times d}(\mathbb{R})$ is SPD, assume that μ and S satisfy the usual conditions, and suppose that $\gamma(F)$ is connected for $F \in \mathcal{F}_{\mathbb{R}}(\mu, S)$. Then there is a $\mathbf{Q} \in O^d(d)$ and a d by d permutation matrix P such that the orthogonal projection*

$$\pi : \mathbf{Q}^T \star T_{FP^T}\mathcal{F}_{\mathbb{R}}(P\mu, S) \to \Delta_{\mathbb{R}}$$

is injective.

Proof. We first construct P while performing some bookkeeping that shall prove useful later in the proof. Let T be a spanning tree in the correlation network of F (which exists because $\gamma(F)$ is connected). Inductively choose i_{k+1} to be a leaf of T_{k+1}, where $T_{k+1} = T_k \setminus \{i_k\}$ and $T_1 = T$. After extracting this sequence, for $k \leq d-1$, set $\alpha_k = l$, where i_l is the only vertex sharing an edge with i_k in T_k. Note that $k < \alpha_k$. The sequence $\{i_k\}_{k\in[d]}$ induces the permutation $i_k \mapsto k$, and P is the matrix realization of this permutation acting on column entries. By construction, we have that the $(k, \alpha_k)^{\text{th}}$ entry of PF^TFP^T is nonzero for all $k \in [d]$. Without loss of generality, we replace FP^T with F, $P\mu$ with μ, and PF^TFP^T with G for the remainder of the proof.

Now, we construct $\mathbf{Q}$. For $k \leq d$, set $P_k = I_{d\times d}$ if $k+1 = \alpha_k$ or if $k = d$; otherwise, let $P_k = P_k^T$ be the matrix realization of the transposition $(k+1\ \alpha_k)$ acting on row entries. Let $Q_k R_k = FP_k^T$ denote the QR-factorization of FP_k^T. We set $\mathbf{Q} = [Q_1, \dots, Q_d]$. Let $\{e_i\}_{i\in[d]}$ be the standard orthonormal basis in $\mathbb{R}^d$, set $E_k = \operatorname{span}\{e_i\}_{i\in[k]}$, and note that E_l is an invariant subspace of R_k^{-1} for all $k, l \in [d]$.

Since F is a nonsingular point of $\mathcal{F}_{\mathbb{R}}(\mu, S)$, $T_F\mathcal{F}_{\mathbb{R}}(\mu, S)$ is well-defined and

$$\mathbf{Q}^T \star T_F\mathcal{F}_{\mathbb{R}}(\mu, S) = \{X = \mathbf{Q}^T \star (FZ) \in M_{d\times d}(\mathbb{R}) : Z = -Z^T, \operatorname{diag}(GZ) = 0\}.$$

By setting $w_k = P_k z_k$, we have that

$$\mathbf{Q}^T \star (FZ) = [Q_1^T F z_1 \cdots Q_d^T F z_d] = [R_1 P_1 z_1 \cdots R_d P_d z_d] = [R_1 w_1 \cdots R_d w_d].$$

Also note that $z_k, w_k \in \operatorname{span}\{e_k\}^\perp$ because Z is skew symmetric.

Suppose $\pi(X) = 0$ for $X = [x_1 \cdots x_d] \in \mathbf{Q}^T \star T_F\mathcal{F}_{\mathbb{R}}(\mu, S)$. Then $x_k \in E_{k+1}$ for $k \in [d-1]$, and hence $w_k = R_k^{-1} x_k \in E_{k+1}$ for all $k \in [d-1]$. Since $w_1 \in E_1^\perp$, we have that $w_1 = z_{\alpha_1,1} e_2$ and $z_1 = P_1 w_1 = z_{\alpha_1,1} e_{\alpha_1}$. The constraint $\operatorname{diag}(GZ) = 0$ then implies

$$0 = e_1^T G Z e_1 = z_{\alpha_1,1} e_1^T G e_{\alpha_1} = z_{\alpha_1,1} g_{1,\alpha_1}.$$

We conclude that $z_1 = 0$ since $g_{1,\alpha_1} \neq 0$, and thus $x_1 = R_1 P_1 z_1 = 0$. Proceeding inductively, if the first $k \leq d-2$ columns of X are zero, then the first k columns of Z are zero, and hence $w_{k+1}, z_{k+1} \in E_{k+1}^\perp$ by skew-symmetry of Z. Since $x_{k+1} \in E_{k+2}$, $w_{k+1} = R_{k+1}^{-1} x_{k+1} \in E_{k+2}$, and then $w_{k+1} = z_{\alpha_{k+1},k+1} e_{k+2}$. Thus, $z_{k+1} = z_{\alpha_{k+1},k+1} e_{\alpha_{k+1}}$ and

$$0 = e_{k+1}^T G Z e_{k+1} = z_{\alpha_{k+1},k+1} e_{k+1}^T G e_{\alpha_{k+1}} = z_{\alpha_{k+1},k+1} g_{k+1,\alpha_{k+1}}$$

implies $z_{k+1} = 0$ since $g_{k+1,\alpha_{k+1}} \neq 0$ by construction. We conclude that $x_{k+1} = R_{k+1}P_{k+1}z_{k+1} = 0$.

Following this induction up to $k+1 = d-1$, we have that the first $d-1$ columns of X are zero. This implies that the first $d-1$ columns of Z are zero, and hence $Z = 0$ by skew-symmetry. Thus, $X = 0$ and π is injective. $\square$

3.2.2 The case $N > d$

We begin by noting that the dimension of a generic $\mathcal{F}_{\mathbb{R}}(\mu, S)$,

$$\begin{aligned} \dim(\mathcal{F}_{\mathbb{R}}(\mu, S)) &= (d-1)N + \sum_{i \in [d]} (N-i) - (Nd-1) \\ &= (d-1)(N-d) + \sum_{i \in [d-2]} i \end{aligned}$$

follows from (2.2.3). This quantity leads us to suspect that we can obtain a parameterization of the form $\Phi(\Theta, L) = [\Gamma(\Theta)\ B(\Theta, L)]$, where $L \in \Delta_{\mathbb{R}}$,

$$\Theta \in \Omega_{\mathbb{R}} = \{X \in M_{d\times(N-d)}(\mathbb{R}) : x_{1i} = 0, \forall i \in [N-d]\},$$

$$\Gamma(\Theta) = \begin{bmatrix} \phi_{11}(\theta_1) & \phi_{11}(\theta_2) & \cdots & \phi_{1,N-d}(\theta_{N-d}) \\ \theta_{21} & \theta_{22} & \cdots & \theta_{2,N-d} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{d1} & \theta_{d2} & \cdots & \theta_{d,N-d} \end{bmatrix},$$

and where $B(\Theta, L)$ has the form

$$\begin{bmatrix} \phi_{1,N-d+1}(\Theta,L) & \phi_{1,N-d+2}(\Theta,L) & \cdots & \phi_{1,N-3}(\Theta,L) & \phi_{1,N-2}(\Theta,L) & \phi_{1,N-1}(\Theta,L) & \phi_{1N}(\Theta,L) \\ \phi_{2,N-d+1}(\Theta,L) & \phi_{2,N-d+2}(\Theta,L) & \cdots & \phi_{2,N-3}(\Theta,L) & \phi_{2,N-2}(\Theta,L) & \phi_{2,N-1}(\Theta,L) & \phi_{2N}(\Theta,L) \\ l_{31} & \phi_{3,N-d+2}(\Theta,L) & \cdots & \phi_{3,N-3}(\Theta,L) & \phi_{3,N-2}(\Theta,L) & \phi_{3,N-1}(\Theta,L) & \phi_{3N}(\Theta,L) \\ l_{41} & l_{42} & \cdots & \phi_{4,N-3}(\Theta,L) & \phi_{4,N-2}(\Theta,L) & \phi_{4,N-1}(\Theta,L) & \phi_{4N}(\Theta,L) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ l_{d-1,1} & l_{d-1,2} & \cdots & l_{d-1,d-3} & \phi_{d-1,N-2}(\Theta,L) & \phi_{d-1,N-1}(\Theta,L) & \phi_{d-1,N}(\Theta,L) \\ l_{d1} & l_{d2} & \cdots & l_{d,d-3} & l_{d,d-2} & \phi_{d,N-1}(\Theta,L) & \phi_{dN}(\Theta,L) \end{bmatrix}.$$

Note that $\Gamma(\Theta)$ and $B(\Theta, L)$ are $(N-d)$ by d and d by d arrays, respectively. Our suspicion is now justified by the main theorem of this subsection.

Theorem 3.2.6. *Let $\mu \in \mathbb{R}_+^N$ and SPD $S \in M_{d\times d}(\mathbb{R})$ satisfy the usual conditions, and suppose that $\gamma(F)$ is connected for a fixed $F \in \mathcal{F}_{\mathbb{R}}(\mu, S)$. Then there is a $\mathbf{Q} \in O^N(d)$ and an $N \times N$ permutation matrix P such that the orthogonal projection*

$$\pi : \mathbf{Q}^T \star T_{FP^T}\mathcal{F}_{\mathbb{R}}(P\mu, S) \to \Omega_{\mathbb{R}} \oplus \Delta_{\mathbb{R}}$$

is injective.

Proof. We shall first show that it is possible to extract a basis of $\mathbb{R}^d$, B, from F such that $\gamma(B)$ is connected. To verify this fact, we begin with any vector, f_{i_1} in F. Proceeding inductively, suppose the first $n-1 \leq d-1$ vectors have been chosen so that they are linearly independent and have a connected correlation network. If there is no $f_{i_n} \notin \operatorname{span}\{f_{i_k}\}_{k\in[n-1]}$ with $\langle f_{i_k}, f_{i_n}\rangle \neq 0$ for some $k \in [n-1]$, then $F = F_1 \cup F_2$ where $F_1 = \{f_i \in F : f_i \in \operatorname{span}\{f_{i_k}\}_{k\in[n-1]}\} \neq \emptyset$ and $F_2 = \{f_i \in F : f_i \in \operatorname{span}\{f_{i_k}\}_{k\in[n-1]}^{\perp}\} \neq \emptyset$. Thus, F is orthodecomposable under this assumption, which contradicts the fact that $\gamma(F)$ connected. Thus, there is an $f_{i_n} \notin \operatorname{span}\{f_{i_k}\}_{k\in[n-1]}$ with $\langle f_{i_k}, f_{i_n}\rangle \neq 0$. These properties of f_{i_n} imply that $\{f_{i_k}\}_{k\in[n]}$ is a linearly independent set with a connected correlation network. Thus, induction supplies us with a basis $B = \{f_{i_k}\}_{k\in[d]} \subset F$ that has a connected correlation network.

With B in hand, we can construct P. Let T be a spanning tree in the correlation network of B. Inductively choose i_{k_n} to be a leaf of T_n, where $T_n = T_{n-1}\backslash\{i_{k_{n-1}}\}$ and $T_1 = T$. After extracting this sequence, for $n \leq d-1$, set $\alpha_{N-d+n} = N-d+m$, where i_{k_m} is the only vertex sharing an edge with i_{k_n} in T_n. Note that $N-d+n <$

α_{N-d+n}. Let σ be any permutation of $[N]$ satisfying $i_{k_n} \mapsto N-d+n$, and set P the matrix realization of this permutation acting on column entries. By construction, we have that the $(N-d+n, \alpha_{N-d+n})^{\text{th}}$ entry of PF^TFP^T is nonzero for all $n \in [d]$. Without loss of generality, we replace FP^T with $F = [\Gamma\, B]$, $P\mu$ with μ, and PF^TFP^T with G for the remainder of the proof. Note that our new B is a basis.

Now, we construct $\mathbf{Q}$. For $k \in [N-d]$, let $Q_k \in SO(d)$ be the unique rotation taking $\mu_k e_1 \mapsto f_k$ that fixes $\operatorname{span}\{f_k, e_1\}^\perp$. For $k \in [N-d+1, N-1]$, set $P_k = I_{d\times d}$ if $k+1 = \alpha_k$; otherwise, let $P_k = P_k^T$ be the matrix realization of the transposition $(k+1 \;\; \alpha_k)$ acting on row entries. If $k = N$, set $P_k = I_{d\times d}$. For each $k \in [N-d+1, N]$, let $Q_k R_k = BP_k^T$ denote the QR-factorization of BP_k^T. We set $\mathbf{Q} = [Q_1, \ldots, Q_N]$. Let $\{e_i\}_{i\in[d]}$ be the standard orthonormal basis in $\mathbb{R}^d$, set $E_k = \operatorname{span}\{e_i\}_{i\in[k]}$, and note that E_l is an invariant subspace of R_k^{-1} for all $k, l \in [d]$.

Since F is a nonsingular point of $\mathcal{F}_{\mathbb{R}}(\mu, S)$, $T_F\mathcal{F}_{\mathbb{R}}(\mu, S)$ is well-defined and

$$\mathbf{Q}^T \star T_F\mathcal{F}_{\mathbb{R}}(\mu, S) = \{X = \mathbf{Q}^T \star (FZ) \in M_{d\times N}(\mathbb{R}) : Z = -Z^T, \operatorname{diag}(GZ) = 0\}.$$

For any $X \in \mathbf{Q}^T \star T_F\mathcal{F}(\mu, S)$ and any $W = -W^T$ with $FW = 0$, note that $X = \mathbf{Q} \star (FZ) = \mathbf{Q} \star (F(Z+W))$. Thus, we may always alter Z by such a W without changing X. Moreover, $FW = 0$ implies $GW = 0$, and hence the diagonal of $G(Z+W)$ remains zero.

Suppose $\pi(X) = 0$ for $X = [x_1 \cdots x_N] = \mathbf{Q}^T \star (FZ) \in \mathbf{Q}^T \star T_F\mathcal{F}(\mu, S)$. For each $k \in [N-d]$, this implies that $x_k = c_k e_1 = Q_k^T F z_k$, and thus

$$\frac{c_k}{\mu_k} f_k = c_k Q_k e_1 = F z_k.$$

However,

$$0 = e_k^T G Z e_k = f_k^T F z_k = \frac{c_k}{\mu_k} f_k^T f_k = c_k \mu_k$$

implies $x_k = 0$ since $\mu_k \neq 0$. Moreover, $Fz_k = 0$ if $k \in [N-d]$. We now claim that Z can be chosen to have the form

$$Z = \begin{bmatrix} 0_{[N-d]\times[N-d]} & 0_{[N-d]\times[N-d+1,N]} \\ 0_{[N-d+1,N]\times[N-d]} & Z_{[N-d+1,N]\times[N-d+1,N]} \end{bmatrix}. \tag{3.2.1}$$

without loss of generality. To show this, suppose

$$Z = \begin{bmatrix} Z_{[N-d]\times[N-d]} & Z_{[N-d]\times[N-d+1,N]} \\ Z_{[N-d+1,N]\times[N-d]} & Z_{[N-d+1,N]\times[N-d+1,N]} \end{bmatrix},$$

set

$$W = \begin{bmatrix} Z_{[N-d]\times[N-d]} & Z_{[N-d]\times[N-d+1,N]} \\ Z_{[N-d+1,N]\times[N-d]} & -B^{-1}\Gamma Z_{[N-d]\times[N-d]}\Gamma^T B^{-T} \end{bmatrix},$$

and note that $\Gamma Z_{[N-d]\times[N-d]} + B Z_{[N-d+1,N]\times[N-d]} = 0$ since the first $N-d$ rows of FZ are zero. Since B is invertible, we have that $Z_{[N-d+1,N]\times[N-d]} = -B^{-1}\Gamma Z_{[N-d]\times[N-d]}$. Consequently,

$$\begin{aligned} (FW)_{[d]\times[N-d+1,N]} &= \Gamma Z_{[N-d]\times[N-d+1,N]} - B(B^{-1}\Gamma Z_{[N-d]\times[N-d]}\Gamma^T B^{-T}) \\ &= \Gamma Z_{[N-d]\times[N-d+1,N]} - \Gamma Z_{[N-d]\times[N-d]}\Gamma^T B^{-T} \\ &= \Gamma Z_{[N-d]\times[N-d+1,N]} + \Gamma Z^T_{[N-d+1,N]\times[N-d]} \\ &= \Gamma Z_{[N-d]\times[N-d+1,N]} - \Gamma Z_{[N-d]\times[N-d+1,N]} \\ &= 0, \end{aligned}$$

and hence both blocks of FW vanish. Thus, we may replace Z with $Z - W$ (which has the form of (3.2.1)) since $FW = 0$ and $W = -W^T$.

Finally, applying the concluding argument in the proof of Theorem 3.2.5 (now adapted to the correlation structure of B) demonstrates that $Z_{[N-d+1,N]\times[N-d+1,N]} = 0$. This is possible because of our initial choice of permutation, which required the connectivity of the correlation network of B. We conclude that $X = 0$, and hence π is injective. $\square$

Corollary 3.2.7. *If the conditions of Theorem 3.2.6 are satisfied, then there is a unique, locally well-defined, real-analytic inverse of* π, $\Phi : \Omega_{\mathbb{R}} \oplus \Delta_{\mathbb{R}} \to \mathbf{Q}^T \star \mathcal{F}_{\mathbb{R}}(P\mu, S)$.

Remark 3.2.8. If Φ is as in the above corollary, then $(\mathbf{Q} \star \Phi(\Theta, L))P$ is a parameterization around $F \in \mathcal{F}_{\mathbb{R}}(\mu, S)$.

3.3 Existence of structured, local coordinate systems on $\mathcal{F}_{\mathbb{C}}(\mu, S)$

The complex case and the real case are very similar, but there is right action of the N-torus on a given $F \in \mathcal{F}_{\mathbb{E}}(\mu, S)$ that embeds an N-torus into $\mathcal{F}_{\mathbb{E}}(\mu, S)$. This action is accomplished by taking the product

$$F \cdot D(\vartheta),$$

where $\vartheta \in \mathbb{R}^N$ and $D(\vartheta)$ is a diagonal matrix with $\text{diag}(D(\vartheta)) = [e^{i\vartheta_1} \cdots e^{i\vartheta_N}]^T$. This is also apparent in the following dimension calculation:

$$\begin{aligned}
\dim(\mathcal{F}_{\mathbb{C}}(\mu, S) &= (2d-1)N + \sum_{i\in[d]}(2N-2i+1) - (2Nd-1) \\
&= \sum_{i\in[d]}(N-i) + \sum_{i\in[d]}(N-i+1) - (N-1) \\
&= \sum_{i\in[2,d]}(N-i) + \sum_{i\in[d]}(N-i+1) \\
&= \sum_{i\in[2,d]}(N-d)+(d-i) + \sum_{i\in[d]}(N-d)+(d-i+1) \\
&= (d-1)(N-d) + \sum_{i\in[2,d]}(d-i) + d(N-d) + \sum_{i\in[d]}(d-i+1) \\
&= (2d-1)(N-d) + \sum_{i\in[d-2]} i + \sum_{i\in[d]} i \\
&= N + (2d-2)(N-d) + \sum_{i\in[d-2]} i + \sum_{i\in[d-1]} i \qquad (3.3.1)
\end{aligned}$$

Note that (3.3.1) collapses to $d + \sum_{i\in[d-2]} i + \sum_{i\in[d-1]} i$ if $N = d$. Now, the leading term, N, reflects the embedding of the N-torus. Since the degrees of freedom arising from these phase changes are easily parameterized, we first determine how to remove them from consideration.

As in the real case, we shall define a structured subspace that only supports the coordinates. Let

$$\Omega_{\mathbb{C}} = \{X \in M_{d\times(N-d)}(\mathbb{C}) : x_{1i} = 0 \text{ for all } i \in [N-d]\},$$

$$\Delta_{\mathbb{C}} = \{X \in M_{d\times d}(\mathbb{C}) : x_{ij} = 0 \text{ if } i \leq j+1\},$$

and

$$\Sigma_\beta = \{X \in M_{d\times d}(\mathbb{C}) : x_{ij} = 0 \text{ if } i \neq j+1 \text{ and } \text{Re}(x_{i+1,i}\beta_i) = 0, \forall i \in [d-1]\},$$

where $\beta \in \mathbb{C}^{d-1}$ has nonzero entries. We shall view $\Omega_{\mathbb{C}}$, Δ_C, and Σ_β as vector spaces over $\mathbb{R}$, and a simple counting argument shows that that

$$\Omega_{\mathbb{C}} \oplus (\Delta_{\mathbb{C}} + \Sigma_\beta) = \{[X\,Y] \in M_{d\times N}(\mathbb{C}) : X \in \Omega_{\mathbb{C}} \text{ and } Y \in \Delta_{\mathbb{C}} + \Sigma_\beta\}$$

has dimension exactly N less than (3.3.1) as a vector space over $\mathbb{R}$. Lastly, we let

$$\Pi_{d\times N} = \{X \in M_{d\times N}(\mathbb{C}) : \mathrm{Im}(x_{1i}) = 0 \forall i \in [N-d]; \mathrm{Im}(x_{i,N-d+i}) = 0, \forall i \in [d]\}.$$

Theorem 3.3.1. *Let $\mu \in \mathbb{R}_+^N$ and HPD $S \in M_{d\times d}(\mathbb{C})$ satisfy the usual conditions, and suppose $\gamma(F)$ is connected for a given $F \in \mathcal{F}_{\mathbb{C}}(\mu, S)$. Then there is a $\mathbf{Q} \in U^N(d)$, an $N \times N$ permutation matrix P, and a $\beta \in \mathbb{C}^{d-1}$ with nonzero entries such that*

(i) $\mathbf{Q}^* \star (FP^*) \in \Pi_{d\times N}$;

(ii) $\mathbf{Q}^* \star \mathcal{F}_{\mathbb{C}}(P\mu, S)$ *intersects* $\Pi_{d\times N}$ *transversally at* $\mathbf{Q}^* \star (FP^*)$

(iii) $T_{\mathbf{Q}^*\star(FP^*)}\mathbf{Q}^* \star \mathcal{F}_{\mathbb{C}}(P\mu, S) \cap \Pi_{d\times N}$ *is the real subspace*

$$\{\mathbf{Q}^* \star (FP^*Z) \in M_{d\times N}(\mathbb{C}) : Z = -Z^*, \mathrm{Re}\,\mathrm{diag}(PF^*FP^*Z) = \mathrm{diag}(Z) = 0\},$$

which we denote $\widetilde{T}_F$;

(iv) *and the orthogonal projection*

$$\pi : \widetilde{T}_F \to \Omega_{\mathbb{C}} \oplus (\Delta_{\mathbb{C}} + \Sigma_\beta)$$

is injective.

Proof. The construction of P proceeds exactly as in Theorem 3.2.6. What is different is that we now keep track of the values $\beta_i = f^*_{N-d+i} f_{N-d+\alpha_i} \neq 0$. First, we construct an index set $I = \{i_k\}_{k\in[d]} \subset [N]$ such that

(a) the columns of F with indices in I form a basis of $\mathbb{C}^d$ with a connected correlation network;

(b) and for each $k \in [d-1]$, there is an α_k ($k < \alpha_k \in [d]$) such that $f_{i_k}^* f_{i_{\alpha_k}} \neq 0$.

Using the process in Theorem 3.2.6, we can construct I satisfying (a). By considering a spanning tree for the correlation network of this basis, and inductively removing it's leaves, we can reorder I to satisfy (b) as well (this induction is detailed in the proof of the previous theorem). P is then chosen so that $i_k \mapsto N - d + k$, Without loss of generality, we assume that P is the identity for the remainder of the proof by replacing FP^* with $F = [\Gamma\ B]$. Here, $B \in M_{d\times d}(\mathbb{C})$ is the basis that we have extracted.

The construction of $\mathbf{Q} \in \prod_{i\in[N]} U(d)$ also mimics the construction contained in the proof of Theorem 3.2.6. Again, for $k \in [N-d]$, we choose $Q_k \in SU(d)$ to be the unique unitary matrix which leaves $\operatorname{span}\{f_k, e_1\}^\perp$ undisturbed and which sends f_k to $\mu_k e_1$. For $k \in [N-d+1, N-1]$, let P_k be the matrix realization of the transposition $(k-N+d+1\ \alpha_{k-N+d})$ if $k-N+d+1 < \alpha_{k-N+d}$. Otherwise, $P_k = I_{d\times d}$ and $P_N = I_{d\times d}$. We then choose Q_k corresponding to the unique QR-factorization $Q_k R_k = BP_k$ such that the diagonal of R_k consists of positive real entries. Finally, we choose Q_N as in the unique QR-factorization $Q_N R_N = B$, and we set $\mathbf{Q} = [Q_1 \cdots Q_N]$. Clearly, we have constructed $\mathbf{Q}$ so that (i) is satisfied.

We now show that (ii) and (iii) hold. Let $X \in M_{d\times N}(\mathbb{C})$, and set Z to be the

$N \times N$ diagonal matrix with k^{th} diagonal entry

$$z_{kk} = \begin{cases} i\text{Im}(x_{1k})/\mu_k & k \in [N-d] \\ i\text{Im}(x_{k-N+d,k})/r^{(k)}_{k-N+d,k-N+d} & k \in [N-d+1, N] \end{cases},$$

where $r^{(k)}_{ij}$ is the ij^{th} entry of R_k. Then $Z = -Z^*$, $\text{Re}\,\text{diag}(F^*FZ) = 0$, $X - \mathbf{Q}^* \star (FZ) \in \Pi_{d\times N}$, and it follows that $T_F\mathbf{Q}^* \star \mathcal{F}_{\mathbb{C}}(\mu, S) + \Pi_{d\times N} = M_{d\times N}(\mathbb{C})$. Moreover, based upon the construction of Z it is easy to deduce that any $Z = -Z^*$ with nonzero diagonal yields a $\mathbf{Q}^* \star (FZ)$ that is not contained in $\Pi_{d\times N}$. Since the purely diagonal $Z = -Z^*$ also give rise to an N-dimensional subspace, the equivalence in (iii) holds.

Finally, we demonstrate that (iv) holds. Note that, for any $X \in \widetilde{T}_F$ and any $W = -W^*$ with $FW = 0$, note that $X = \mathbf{Q}^* \star (FZ) = \mathbf{Q}^* \star (F(Z+W))$. Thus, we may always alter Z by such a W without changing X.

Suppose $\pi(X) = 0$ for $A = [X\,Y] = [x_1 \cdot x_{N-d}\, y_1 \cdots y_d] \in \mathbf{Q}^* \star (FZ) \in \widetilde{T}_F$. For each $k \in [N-d]$, this implies that $x_k = c_k e_1 = Q_k^* F z_k$ for some real c_k, and thus

$$\frac{c_k}{\mu_k} f_k = c_k Q_k e_1 = F z_k.$$

However,

$$0 = \text{Re}(e_k^* GZe_k = f_k^* F z_k) = \text{Re}\frac{c_k}{\mu_k} f_k^* f_k = \text{Re}(c_k \mu_k) = c_k \mu_k$$

implies $x_k = 0$ since $\mu_k \neq 0$. Moreover, $Fz_k = 0$ if $k \in [N-d]$. We now claim that Z can be chosen to satisfy

$$Z = \begin{bmatrix} 0_{[N-d]\times[N-d]} & 0_{[N-d]\times[N-d+1,N]} \\ 0_{[N-d+1,N]\times[N-d]} & \widetilde{Z} \end{bmatrix}$$

without loss of generality. Subsitution of conjugate transposition with transposition in the argument used Theorem 3.2.6 substantiates this claim.

Finally, we apply a modified version of the usual induction argument to the remaining columns of A, Y. We have that $\widetilde{\pi}(Y) = 0$ for the orthogonal projection

$$\widetilde{\pi} : T_{\widetilde{\mathbf{Q}}^* \star B} \widetilde{\mathbf{Q}}^* \star \mathcal{F}_{\mathbb{C}}(\widetilde{\mu}, BB^*) \cap \Pi_{d \times d} \to \Delta_{\mathbb{C}} + \Sigma_{\beta},$$

where $\widetilde{\mathbf{Q}}$ and $\widetilde{\mu}$ are the last d entries of $\mathbf{Q}$ and μ. Additionally, $Y = \mathbf{Q}^* \star (B\widetilde{Z})$ with $\widetilde{Z} = -\widetilde{Z}^*$, $\operatorname{Re}\operatorname{diag}(B^*B\widetilde{Z}) = 0$, and $\operatorname{diag}(Z) = 0$.

Let $\{e_i\}_{i\in[d]}$ denote the standard orthonormal basis for $\mathbb{C}^d$, and set $E_k = \operatorname{span}_{\mathbb{C}}\{e_i\}_{i\in[k]}$. We have that $y_1 = Q_1^* B z_1 = R_1 P_1 z_1 = R_1 w_1$. Since $z_1 \in E_1^{\perp}$ and R_1^{-1} is triangular with positive diagonal entries, $w_1 \in \operatorname{span}\{e_2\}$. Furthermore,

$$0 = \operatorname{Im}(y_{21}\beta_1) = \operatorname{Im}(r_{22}^{(1)} w_{21}\beta_1)$$

implies $\operatorname{Im}(w_{21}\beta_1) = 0$ since $r_{22}^{(1)}$ is positive and real. It follows that $z_1 = z_{\alpha_1,1} e_{\alpha_1}$ with $\operatorname{Im}(z_{\alpha_1,1}\beta_1) = 0$. The final condition

$$0 = \operatorname{Re}(e_1^* B^* B Z e_1) = \operatorname{Re}(z_{\alpha_1,1} f_{N-d+1}^* f_{N-d+\alpha_1}) = \operatorname{Re}(z_{\alpha_1,1}\beta_1)$$

forces $z_{\alpha_1,1}\beta_1 = 0$, and thus $z_{\alpha_1,1} = 0$ since $\beta_1 \neq 0$. Consequently, $z_1 = y_1 = 0$. We may now proceed with the usual induction argument, noting that the augmentation of the projection by Σ_β forces the complex version of $z_{\alpha_k,k}\beta_k = 0$. This proves (iv), finishing the proof. □

Corollary 3.3.2. *If the conditions of Theorem 3.2.6 are satisfied, then there is a unique, locally well-defined, real-analytic inverse of* π, $\Phi : \Omega_{\mathbb{C}} \oplus (\Delta_{\mathbb{C}} + \Sigma_\beta) \to \mathbf{Q}^* \star \mathcal{F}_{\mathbb{C}}(P\mu, S) \cap \Pi_{d\times N}$.

Remark 3.3.3. If Φ is as in the above corollary, then $(\mathbf{Q} \star \Phi(\Theta, L))PD(\vartheta)$ is a parameterization around $F \in \mathcal{F}_{\mathbb{C}}(\mu, S)$.

3.4 Explicit construction of the coordinate systems on $\mathcal{F}_{\mathbb{R}}(\mu, S)$

In this section, we construct explict local inversions of the π in Theorem 3.2.6. The first $N-d$ columns of the inverse, Φ, are easily determined, and the last d columns can be solved inductively. Solving each of the final d columns can be reduced to a quartic, which means that coordinate systems on $\mathcal{F}_{\mathbb{R}}(\mu, S)$ are solvable by radicals.

The intuition behind the coordinate systems is that (locally) $N-d$ members of the frame may be freely articulated in their respective spheres while the remaining d members adjust to compensate for these articulations. The remaining d members can also be articulated, but their subframe operator is determined by the position of the free $N-d$ members.

3.4.1 Parameterizing the intersection of two ellipsoids

Given SPD $A, B \in M_{d\times d}(\mathbb{R})$, we shall exhibit formal parametermizations of the system

$$\begin{bmatrix} y \\ l \end{bmatrix}^T A \begin{bmatrix} y \\ l \end{bmatrix} = 1, \quad \begin{bmatrix} y \\ l \end{bmatrix}^T B \begin{bmatrix} y \\ l \end{bmatrix} = 1,$$

where $y = [y_1\ y_2]^T$ and $l = [l_3 \cdots l_d]^T$. These equations expand to

$$a_{11}y_1^2 + a_{22}y_2^2 + 2a_{12}y_1y_2 + 2l^Ta_1y_1 + 2l^Ta_2y_2 + l^TA'l - 1 = 0$$

$$b_{11}y_1^2 + b_{22}y_2^2 + 2b_{12}y_1y_2 + 2l^Tb_1y_1 + 2l^Tb_2y_2 + l^TB'l - 1 = 0,$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} & a_1^T \\ a_{12} & a_{22} & a_2^T \\ a_1 & a_2 & A' \end{bmatrix}, \text{ and } B = \begin{bmatrix} b_{11} & b_{12} & b_1^T \\ b_{12} & b_{22} & b_2^T \\ b_1 & b_2 & B' \end{bmatrix}.$$

When these equations are viewed as quadratics in y_2, we have

$$a_{22}y_2^2 + (2a_{12}y_1 + 2l^Ta_2)y_2 + (a_{11}y_1^2 + 2l^Ta_1y_1 + l^TA'l - 1) = 0$$

$$b_{22}y_2^2 + (2b_{12}y_1 + 2l^Tb_2)y_2 + (b_{11}y_1^2 + 2l^Tb_1y_1 + l^TB'l - 1) = 0$$

This system has a solution if and only if the Bézout determinant vanishes. That is,

$$\begin{aligned} 0 \;=\; & [a_{22}(2b_{12}y_1 + 2l^Tb_2) - (2a_{12}y_1 + 2l^Ta_2)b_{22}] \\ & \times[(2a_{12}y_1 + 2l^Ta_2)(b_{11}y_1^2 + 2l^Tb_1y_1 + l^TB'l - 1) \\ & -(a_{11}y_1^2 + 2l^Ta_1y_1 + l^TA'l - 1)(2b_{12}y_1 + 2l^Tb_2)] \\ & -[a_{22}(b_{11}y_1^2 + 2l^Tb_1y_1 + l^TB'l - 1) - (a_{11}y_1^2 + 2l^Ta_1y_1 + l^TA'l - 1)b_{22}]^2. \end{aligned}$$

This expands to a quartic in y_1, which has four formal roots. Each formal root provides us with a solution to y in the original quadratic system, and we let $\rho_\varepsilon(l, A, B)$ denote the formal solutions, where $\varepsilon \in \{0, 1\}^2$. Thus, we obtain formal coordinates for the intersection of two ellipsoids,

$$\begin{bmatrix} \rho_\varepsilon(l, A, B) \\ l \end{bmatrix}.$$

3.4.2 Elimination for two quadratic constraints and a linear system

We seek solutions to the system

$$\begin{bmatrix} x \\ y \\ l \end{bmatrix}^T A \begin{bmatrix} x \\ y \\ l \end{bmatrix} = 1, \quad \begin{bmatrix} x \\ y \\ l \end{bmatrix}^T B \begin{bmatrix} x \\ y \\ l \end{bmatrix} = 1,$$

and

$$C \begin{bmatrix} x \\ y \\ l \end{bmatrix} = \begin{bmatrix} C_{[k-1]^2} & C_{[k-1]\times[k,k+1]} & C_{[k-1]\times[k+2,d]} \end{bmatrix} \begin{bmatrix} x \\ y \\ l \end{bmatrix} = 0.$$

If $C_{[k-1]^2}$ is invertible, then $x = -C^{-1}_{[k-1]^2} C_{[k-1]\times[k,k+1]} y - C^{-1}_{[k-1]^2} C_{[k-1]\times[k+2,d]} l$. In this case, we may eliminate x from the quadratics to obtain

$$\begin{bmatrix} y \\ l \end{bmatrix}^T \widetilde{A}(A,C) \begin{bmatrix} y \\ l \end{bmatrix} = 1, \quad \begin{bmatrix} y \\ l \end{bmatrix}^T \widetilde{B}(B,C) \begin{bmatrix} y \\ l \end{bmatrix} = 1.$$

For a fixed l, there are four formal solutions for y, $\rho_\varepsilon(l, \widetilde{A}(A,C), \widetilde{B}(B,C))$. Thus, we have formal coordinates of the form

$$\begin{aligned} \psi_\varepsilon(l, A, B, C) \quad &= \begin{bmatrix} -C^{-1}_{[k-1]^2} C_{[k-1]\times[k,k+1]} & -C^{-1}_{[k-1]^2} C_{[k-1]\times[k+2,d]} \\ I_{2\times 2} & 0_{2\times(d-k-1)} \\ 0_{(d-k-1)\times 2} & I_{(d-k-1)\times(d-k-1)} \end{bmatrix} \\ &\times \begin{bmatrix} \rho_\varepsilon(l, \widetilde{A}(A,C), \widetilde{B}(B,C))) \\ l \end{bmatrix}. \end{aligned}$$

If $k = d-1$, we have

$$\psi'_\varepsilon(A,B,C) = \begin{bmatrix} -C^{-1}_{[d-2]^2} C_{[d-2]\times[d-1,d]} \\ I_{2\times 2} \end{bmatrix} \rho_\varepsilon(l, \widetilde{A}(A,C), \widetilde{B}(B,C))).$$

3.4.3 Coordinates for $\mathcal{F}_{\mathbb{R}}(\mu, S)$

Suppose $\mu \in \mathbb{R}_+^N$ and $S \in M_{d\times d}(\mathbb{R})$ satisfy the usual constraints, and suppose $F \in \mathcal{F}_{\mathbb{R}}(\mu, S)$ is a nonsingular point. Given the $\mathbf{Q} = [\mathbf{Q}_{[N-d]}\ \mathbf{Q}_{[N-d+1,N]}] \in O^N(d)$ and the $N \times N$ permutation matrix P ensured by Theorem 3.2.6, we construct explicit coordinates about

$$\mathbf{Q}^T \star (FP^T) \in \mathbf{Q}^T \star \mathcal{F}_{\mathbb{R}}(P\mu, S).$$

This coordinate system has the form $\Phi(\Theta, L) = [\Gamma(\Theta) B(\Theta, L)]$, where $(\Theta, L) \in \Omega_{\mathbb{R}} \oplus \Delta_{\mathbb{R}}$, $\Gamma(\Theta) \in M_{d\times(N-d)}(\mathbb{R})$, and $B(\Theta, L) \in M_{d\times d}(\mathbb{R})$. For any such Φ, $(\mathbf{Q} \star \Phi(\Theta, L))P$ is a coordinate system about $F \in \mathcal{F}_{\mathbb{R}}(\mu, S)$.

First, we shall describe the procedure for determining the coordinate functions. Afterwards, we shall demonstrate that this process produces a valid coordinate system that is well-defined in a suitable neighborhood.

We begin by defining $\Gamma(\Theta) = [\phi_1(\theta_1) \cdots \phi_{N-d}(\theta_{N-d})]$, where $\nu = P\mu$ and

$$\phi_k(\theta_k) = \begin{bmatrix} \sqrt{\nu_k^2 - \|\theta_k\|_2^2} \\ \theta_k \end{bmatrix}, \text{ for all } k \in [N-d].$$

We then set $\mathcal{B}(\Theta) = (S - (\mathbf{Q}_{[N-d]}\Gamma(\Theta))(\mathbf{Q}_{[N-d]}\Gamma(\Theta))^T)^{-1}$, and search for solutions to the system

$$\phi_i^T(\Theta, L)\phi_i(\Theta, L) = \nu_i^2$$
$$\phi_i^T(\Theta, L)Q_i^T\mathcal{B}(\Theta)Q_j\phi_j(\Theta, L) = \delta_{ij},$$

for all $i, j \in [N-d+1, N]$, and where δ_{ij} is the Kronecker delta. As foreshadowed by the existence proofs, we shall construct solutions in an inductive, column-by-column manner. First we set

$$\phi^{\varepsilon_1}_{N-d+1}(\Theta, L) = \begin{bmatrix} \rho_{\varepsilon_1}(l_1, \nu^{-2}_{N-d+1} I_{d\times d}, Q^T_{N-d+1}\mathcal{B}(\Theta)Q_{N-d+1}) \\ l_1 \end{bmatrix},$$

and then we set

$$\begin{aligned} \phi^{\varepsilon_{k-N+d}}_k(\Theta, L) &= \psi_{\varepsilon_{k-N+d}}(l_{k-N+d}, \nu_k^{-2} I_{d\times d}, Q_k^T \mathcal{B}(\Theta)Q_k, \\ &\qquad (\mathbf{Q}_{[N-d+1,k-1]} \star \Phi(\Theta, L)_{[N-d+1,k-1]})^T \mathcal{B}(\Theta)) \end{aligned}$$

inductively for all $k \in [N-d+2, N-2]$. For the final two columns, we set

$$\begin{aligned} \phi^{\varepsilon_{d-1}}_{N-1} &= \psi'_{\varepsilon_{d-1}}(\nu^{-2}_{N-1} I_{d\times d}, Q^T_{N-1}\mathcal{B}(\Theta)Q_{N-1}, \\ &\qquad (\mathbf{Q}_{[N-d+1,N-2]} \star \Phi(\Theta, L)_{[N-d+1,N-2]})^T \mathcal{B}(\Theta)) \\ \phi^{\varepsilon_d}_N &= \psi'_{\varepsilon_d}(\nu_N^{-2} I_{d\times d}, Q_N^T \mathcal{B}(\Theta)Q_N, (\mathbf{Q}_{[N-d+1,N-2]} \star \Phi(\Theta, L)_{[N-d+1,N-2]})^T \mathcal{B}(\Theta)), \end{aligned}$$

where ψ'_ε are defined at the end of the preceding section. Now that we have a several formal solutions, we proceed to demonstrate that exactly one of these solutions is the local inverse of π.

Theorem 3.4.1. *Suppose μ, S, and F are given and satisfy the usual conditions. Then there is a unique $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_d) \in (\{0,1\}^2)^d$ such that*

(i) $\Phi_\varepsilon(0,0) = \mathbf{Q}^T \star (FP^T)$;

(ii) $\Phi_\varepsilon(\Theta, L) = [\phi_1(\theta_1) \cdots \phi_{N-d}(\theta_{N-d})\ \phi^{\varepsilon_1}_{N-d+1}(\Theta, L) \cdots \phi^{\varepsilon_d}_N(\Theta, L)]$ *is well-defined in a neighborhood of* $[0_{[N-d]} 0_{[N-d+1,N]}] \in \Omega_{\mathbb{R}} \oplus \Delta_{\mathbb{R}}$;

(iii) *and* Φ_ε *is a chart on* $\mathbf{Q}^T \star \mathcal{F}_{\mathbb{R}}(P\mu, S)$.

Proof. First, we note that we can (locally) carry out the inductive process described in this section. For the first $N-d$ columns, we only require that $\mathcal{B}(\Theta)$ exists. By continuity of the functions in consideration, and since $\mathcal{B}(0)$ is invertible, this can be ensured locally. Now, inductively set

$$\begin{aligned} C^{(k)} &= [C^{(k)}_{[k-1]^2}(\Theta, L)\; C^{(k)}_{[k-1]\times[d-k+1]}] \\ &= (\mathbf{Q}_{[N-d+1,k-1]} \star \Phi(\Theta, L)_{[N-d+1,k-1]})^T \mathcal{B}(\Theta)) \end{aligned}$$

for all $k \in [N-d+2, N-1]$. There is a neighborhoood of $[0_{[N-d]} 0_{[N-d+1,N]}] \in \Omega_{\mathbb{R}} \oplus \Delta_{\mathbb{R}}$ where $C^{(k)}_{[k-1]^2}(\Theta, L)$ is invertible for each $k \in [N-d+2, N-1]$ since the first $k-1$ columns of

$$(\mathbf{Q}_{[N-d+1,k-1]} \star \Phi(0,0)_{[N-d+1,k-1]})^T$$

form a square upper triangular matrix with strictly positive entries on the diagonal. Since $\mathcal{B}(\Theta)$ is also locally invertible around Θ, the claim follows. Since inveriblity of $C^{(k)}_{[k-1]^2}(\Theta, L)$ is the only hypothesis required to do the elimination of Section 5.1.2, we conclude that the inductive process (locally) produces all possible solutions to the system of constraints.

Now, each ε defines a branch of the inverse of the orthogonal projection

$$\pi : \mathbf{Q}^T \star \mathcal{F}_{\mathbb{R}}(P\mu, S) \to \Omega_{\mathbb{R}} \oplus \Delta_{\mathbb{R}},$$

and since $\widetilde{F} = \mathbf{Q}^T \star (FP^T) \in \mathbf{Q}^T \star \mathcal{F}_{\mathbb{R}}(P\mu, S)$ solves $\pi(\widetilde{F}) = 0$, there is at least one ε satisfying (i). Now, $\widetilde{F}$ is a regular point of π, so there is a neighborhood $\mathcal{N}$ of $[0_{[N-d]}\ 0_{[N-d+1,N]}] \in \Omega_{\mathbb{R}} \oplus \Delta_{\mathbb{R}}$ such that π^{-1} is well-defined, real-valued, and analytic by the Real-Analytic Inverse Function Theorem. For each $[\Theta\ L] \in \mathcal{N}$, there is an ε such that $\pi^{-1}([\Theta\ L]) = \Phi_\varepsilon(\Theta, L)$ since the $\{\Phi_\varepsilon(\Theta, L)\}_{\varepsilon\in(\{0,1\})^d}$ form an exhaustive list of the possible solutions to the inversion at that point. If there are two ε satisfying (i), there must be a neighborhood in $\mathcal{N}$ such that they disagree since they have distinct analytic expansions (or else we may apply the multivariable Open Mapping Theorem to obtain equality). Using this fact in an induction process gives us that there is exactly one ε satisfying (i) such that Φ_ε agrees with π^{-1} in some subneighborhood of $\mathcal{N}$. By the Open Mapping Theorem, this implies $\Phi_\varepsilon = \pi^{-1}$ everywhere in $\mathcal{N}$. Now, restricting $\mathcal{N}$ to satisfy the requirements of the first paragraph, we have that both (ii) and (iii) hold. $\square$

3.5 Explicit construction of the coordinate systems on $\mathcal{F}_{\mathbb{C}}(\mu, S)$

3.5.1 The intersection of two complex ellipsoids

Given HPD $A, B \in M_{d\times d}(\mathbb{C})$ and $\beta \in \mathbb{C}$ with $\beta \neq 0$, we shall construct formal parametermizations of the system

$$\begin{bmatrix} y \\ l \end{bmatrix}^* A \begin{bmatrix} y \\ l \end{bmatrix} = 1, \quad \begin{bmatrix} y \\ l \end{bmatrix}^* B \begin{bmatrix} y \\ l \end{bmatrix} = 1,$$

where

$$y = \begin{bmatrix} y_1 \\ y_2\widetilde{\beta} + i\lambda\widetilde{\beta} \end{bmatrix} \in \mathbb{C}^2$$

with $y_1, y_2, \lambda \in \mathbb{R}$, $\widetilde{\beta} = \beta^*/|\beta|$, and $l = [l_3 \cdots l_d]^T \in \mathbb{C}^{d-2}$. These equations expand to

$$\begin{aligned}
a_{11}y_1^2 + a_{22}y_2^2 + 2\mathrm{Re}(a_{12}\widetilde{\beta}^*)y_1y_2 + 2\mathrm{Re}(l^*a_1 - i\lambda a_{12}\widetilde{\beta}^*)y_1 & \\
+2\mathrm{Re}(l^*a_2\widetilde{\beta})y_2 + a_{22}\lambda^2 + l^*A'l - 1 &= 0 \\
b_{11}y_1^2 + b_{22}y_2^2 + 2\mathrm{Re}(b_{12}\widetilde{\beta}^*)y_1y_2 + 2\mathrm{Re}(l^*b_1 - i\lambda b_{12}\widetilde{\beta}^*)y_1 & \\
+2\mathrm{Re}(l^*b_2\widetilde{\beta})y_2 + b_{22}\lambda^2 + l^*B'l - 1 &= 0
\end{aligned}$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} & a_1^* \\ a_{12}^* & a_{22} & a_2^* \\ a_1 & a_2 & A' \end{bmatrix}, \text{ and } B = \begin{bmatrix} b_{11} & b_{12} & b_1^* \\ b_{12}^* & b_{22} & b_2^* \\ b_1 & b_2 & B' \end{bmatrix}.$$

When these equations are viewed as quadratics in y_2, we have

$$\begin{aligned}
a_{22}y_2^2 + 2\mathrm{Re}[a_{12}\widetilde{\beta}^*y_1 + l^*a_2\widetilde{\beta}]y_2 & \\
+[a_{11}y_1^2 + 2\mathrm{Re}(l^*a_1 - i\lambda a_{12}\widetilde{\beta}^*)y_1 + (a_{22}\lambda^2 + l^*A'l - 1)] &= 0 \\
b_{22}y_2^2 + 2\mathrm{Re}[b_{12}\widetilde{\beta}^*y_1 + l^*b_2\widetilde{\beta}]y_2 & \\
+[b_{11}y_1^2 + 2\mathrm{Re}(l^*b_1 - i\lambda b_{12}\widetilde{\beta}^*)y_1 + (b_{22}\lambda^2 + l^*B'l - 1)] &= 0.
\end{aligned}$$

One can see that these expressions soon become unwieldy, so we introduce the auxilary functions $p_k(z_1, \lambda, l, A, B)$ and $q_k(z_1, \lambda, l, A, B)$ for $k = 0, 1, 2$ which replace the coefficients of the previous polynomials. We then work with the replacement system

$$p_2z_2^2 + p_1z_2 + p_0 = 0$$

$$q_2z_2^2 + q_1z_2 + q_0 = 0$$

This system has a solution if and only if the Bézout determinant vanishes. That is,

$$[p_2q_1 - p_1q_1][p_1q_0 - p_0q_1] - [p_2q_0 - p_0q_2]^2 = 0.$$

This expands to a quartic in y_1 with real coefficients dependent upon λ, l, A, and B. This quartic yields four formal roots in terms of the coefficient functions. Each formal root provides us with a solution to y in the original quadratic system, and we let $\rho_\varepsilon(\lambda, l, A, B)$ denote the formal solutions, where $\varepsilon \in \{0,1\}^2$. Thus, we obtain formal coordinates for the intersection of two ellipsoids,

$$\begin{bmatrix} \rho_\varepsilon(\lambda, l, A, B) \\ l \end{bmatrix}.$$

3.5.2 Elimination for two quadratic constraints and a linear system

For HPD A, B and $\beta \neq 0$, we seek solutions to the system

$$\begin{bmatrix} x \\ y \\ l \end{bmatrix}^* A \begin{bmatrix} x \\ y \\ l \end{bmatrix} = 1, \quad \begin{bmatrix} x \\ y \\ l \end{bmatrix}^* B \begin{bmatrix} x \\ y \\ l \end{bmatrix} = 1,$$

and

$$C \begin{bmatrix} x \\ y \\ l \end{bmatrix} = \begin{bmatrix} C_{[k-1]^2} & C_{[k-1]\times[k,k+1]} & C_{[k-1]\times[k+2,d]} \end{bmatrix} \begin{bmatrix} x \\ y \\ l \end{bmatrix} = 0,$$

where $x \in \mathbb{C}^{k-1}$,

$$y = \begin{bmatrix} y_1 \\ y_2\widetilde{\beta} + i\lambda\widetilde{\beta} \end{bmatrix} \in \mathbb{C}^2$$

with $y_1, y_2, \lambda \in \mathbb{R}$, $\widetilde{\beta} = \beta^*/|\beta|$, and $l \in \mathbb{C}^{d-k+1}$. If $C_{[k-1]^2}$ is invertible, then $x = -C_{[k-1]^2}^{-1} C_{[k-1]\times[k,k+1]} y - C_{[k-1]^2}^{-1} C_{[k-1]\times[k+2,d]} l$. In this case, we may eliminate x from the quadratics to obtain

$$\begin{bmatrix} y \\ l \end{bmatrix}^* \widetilde{A}(A,C) \begin{bmatrix} y \\ l \end{bmatrix} = 1, \begin{bmatrix} y \\ l \end{bmatrix}^* \widetilde{B}(B,C) \begin{bmatrix} y \\ l \end{bmatrix} = 1.$$

For a fixed l, there are four formal solutions for y,

$$\rho_\varepsilon(\lambda, l, \widetilde{A}(A,C), \widetilde{B}(B,C)).$$

Thus, we have formal coordinates of the form

$$\begin{aligned} \psi_\varepsilon(\lambda, l, A, B, C) &= \begin{bmatrix} -C_{[k-1]^2}^{-1} C_{[k-1]\times[k,k+1]} & -C_{[k-1]^2}^{-1} C_{[k-1]\times[k+2,d]} \\ I_{2\times 2} & 0_{2\times(d-k-1)} \\ 0_{(d-k-1)\times 2} & I_{(d-k-1)\times(d-k-1)} \end{bmatrix} \\ &\quad \times \begin{bmatrix} \rho_\varepsilon(\lambda, l, \widetilde{A}(A,C), \widetilde{B}(B,C)) \\ l \end{bmatrix}. \end{aligned}$$

If $k = d-1$, we have

$$\psi'_\varepsilon(\lambda, A, B, C) = \begin{bmatrix} -C_{[d-2]^2}^{-1} C_{[d-2]\times[d-1,d]} \\ I_{2\times 2} \end{bmatrix} \rho_\varepsilon(\lambda, \widetilde{A}(A,C), \widetilde{B}(B,C)),$$

and if $k = d$ we have

$$\begin{aligned} \psi'_\varepsilon(A, B, C) &= \begin{bmatrix} -C_{[d-1]^2}^{-1} C_{[d-1]\times\{d\}} \\ 1 \end{bmatrix} \\ &\quad \Bigg/ \begin{bmatrix} -C_{[d-1]^2}^{-1} C_{[d-1]\times\{d\}} \\ 1 \end{bmatrix}^* A \begin{bmatrix} -C_{[d-1]^2}^{-1} C_{[d-1]\times\{d\}} \\ 1 \end{bmatrix} \end{aligned}$$

3.5.3 Coordinates for $\mathcal{F}_{\mathbb{C}}(\mu, S)$

Suppose $\mu \in \mathbb{R}_+^N$ and $S \in M_{d\times d}(\mathbb{C})$ satisfy the usual constraints, and suppose $F \in \mathcal{F}_{\mathbb{C}}(\mu, S)$ is a nonsingular point. Given the $\mathbf{Q} = [\mathbf{Q}_{[N-d]}\ \mathbf{Q}_{[N-d+1,N]}] \in U^N(d)$, the $N \times N$ permutation matrix P, and the sequence $\beta \in \mathbb{C}^{d-1}$ ensured by Theorem 3.3.1, we construct explicit coordinates about

$$\mathbf{Q}^* \star (FP^*) \in \mathbf{Q}^* \star \mathcal{F}_{\mathbb{C}}(P\mu, S) \cap \Pi_{d\times N}.$$

This coordinate system has the form $\Phi(\Theta, L) = [\Gamma(\Theta)\ B(\Theta, L)]$, where $(\Theta, L) \in \Omega_{\mathbb{C}} \oplus (\Delta_{\mathbb{C}} + \Sigma_\beta)$, $\Gamma(\Theta) \in M_{d\times(N-d)}(\mathbb{C})$, and $B(\Theta, L) \in M_{d\times d}(\mathbb{C})$. For any such Φ, $(\mathbf{Q} \star \Phi(\Theta, L))PD(\vartheta)$ is a coordinate system about $F \in \mathcal{F}_{\mathbb{C}}(\mu, S)$.

First, we shall describe the procedure for determining the coordinate functions. Afterwards, we shall demonstrate that this process produces a valid coordinate system that is well-defined in a suitable neighborhood.

We begin by defining $\Gamma(\Theta) = [\phi_1(\theta_1) \cdots \phi_{N-d}(\theta_{(N-d)})]$, where

$$\phi_k(\theta_k) = \begin{bmatrix} \sqrt{\nu_k^2 - \|\theta_k\|_2^2} \\ \\ \theta_k \end{bmatrix} \text{ for all } k \in [N-d],$$

and $\nu = P\mu$. We then set $\mathcal{B}(\Theta) = (S - (\mathbf{Q}_{[N-d]}\Gamma(\Theta))(\mathbf{Q}_{[N-d]}\Gamma(\Theta))^*)^{-1}$, and search for solutions to the system

$$\phi_i^*(\Theta, L)\phi_i(\Theta, L) = \nu_i^2$$

$$\phi_i^*(\Theta, L)Q_i^*\mathcal{B}(\Theta)Q_j\phi_j(\Theta, L) = \delta_{ij},$$

for all $i, j \in [N-d+1, N]$. We now set

$$\phi_{N-d+1}^{\varepsilon_1}(\Theta, L) = \begin{bmatrix} \rho_{\varepsilon_1}(\lambda_1, l_1, \nu_{N-d+1}^{-2} I_{d\times d}, Q_{N-d+1}^* \mathcal{B}(\Theta) Q_{N-d+1}) \\ l_1 \end{bmatrix},$$

and then we set $C_k(\Theta, L) = (\mathbf{Q}_{[N-d+1,k-1]} \star \Phi(\Theta, L)_{[N-d+1,k-1]})^* \mathcal{B}(\Theta)$ and

$$\phi_k^{\varepsilon_{k-N+d}}(\Theta, L) = \psi_{\varepsilon_{k-N+d}}(\lambda_{k-N+d}, l_{k-N+d}, \nu_k^{-2} I_{d\times d}, Q_k^* \mathcal{B}(\Theta) Q_k, C_k(\Theta, L))$$

inductively for all $k \in [N-d+2, N-2]$. For the final two columns, we set

$$\begin{aligned} \phi_{N-1}^{\varepsilon_{d-1}} &= \psi'_{\varepsilon_{d-1}}(\lambda_{d-1}, \nu_{N-1}^{-2} I_{d\times d}, Q_{N-1}^* \mathcal{B}(\Theta) Q_{N-1}, C_{N-1}(\Theta, L)) \\ \phi_N^{\varepsilon_d} &= \psi'_{\varepsilon_d}(\nu_N^{-2} I_{d\times d}, Q_N^* \mathcal{B}(\Theta) Q_N, C_{N-1}(\Theta, L)) \end{aligned}$$

Now, we have just constructed several formal coordinate systems. The main theorem of this section demonstrates that exactly one of these formal solutions is actually the local inverse of π.

Theorem 3.5.1. *Suppose μ, S, and F are given and satisfy the usual conditions. Then there is a unique $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_d) \in (\{0,1\}^2)^d$ such that*

(i) $\Phi_\varepsilon(0,0) = \mathbf{Q}^* \star (FP^*)$;

(ii) $\Phi_\varepsilon(\Theta, L) = [\phi_1(\theta_1) \cdots \phi_{N-d}(\theta_{N-d})\ \phi_{N-d+1}^{\varepsilon_1}(\Theta, L) \cdots \phi_N^{\varepsilon_d}(\Theta, L)]$ *is well-defined in a neighborhood of* $[0_{[N-d]} 0_{[N-d+1,N]}] \in \Omega_{\mathbb{C}} \oplus (\Delta_{\mathbb{C}} + \Sigma_\beta)$;

(iii) *and Φ_ε is a chart on $\mathbf{Q}^* \star \mathcal{F}_{\mathbb{C}}(P\mu, S) \cap \Pi_{d\times N}$.*

Proof. This argument of this proof is essentially the same as the argument of Theorem 3.4.1. First, we note that we can (locally) carry out the inductive process described in this section. For the first $N-d$ columns, we only require that $\mathcal{B}(\Theta)$ exists. By continuity of the functions in consideration, and since $\mathcal{B}(0)$ is invertible, this can be ensured locally. Now, inductively set

$$C^{(k)} = [C^{(k)}_{[k-1]^2}(\Theta, L)\, C^{(k)}_{[k-1]\times[d-k+1]}] = (\mathbf{Q}_{[N-d+1,k-1]} \star \Phi(\Theta, L)_{[N-d+1,k-1]})^* \mathcal{B}(\Theta)$$

for all $k \in [N-d+2, N-1]$. There is a neighborhoood of $[0_{[N-d]}\, 0_{[N-d+1,N]}] \in \Omega_{\mathbb{C}} \oplus (\Delta_{\mathbb{C}} + \Sigma_\beta)$ where $C^{(k)}_{[k-1]^2}(\Theta, L)$ is invertible for each $k \in [N-d+2, N-1]$ since the first $k-1$ columns of

$$(\mathbf{Q}_{[N-d+1,k-1]} \star \Phi(0,0)_{[N-d+1,k-1]})^*$$

form a square upper triangular matrix with strictly positive entries on the diagonal. Since $\mathcal{B}(\Theta)$ is also locally invertible around Θ, the claim follows. Since inveriblity of $C^{(k)}_{[k-1]^2}(\Theta, L)$ is the only hypothesis required to do the elimination of Section 5.2.2, we conclude that the inductive process (locally) produces all possible solutions to the system of constraints.

Each ε defines a branch of the inverse of the orthogonal projection

$$\pi : \mathbf{Q}^* \star \mathcal{F}_{\mathbb{C}}(P\mu, S) \cap \Pi_{d\times N} \to \Omega_{\mathbb{C}} \oplus (\Delta_{\mathbb{C}} + \Sigma_\beta),$$

and since $\widetilde{F} = \mathbf{Q}^* \star (FP^*) \in \mathbf{Q}^* \star \mathcal{F}_{\mathbb{C}}(P\mu, S) \cap \Pi_{d\times N}$ solves $\pi(\widetilde{F}) = 0$, there is at least one ε satisfying (i). Now, $\widetilde{F}$ is a regular point of π, so there is a neighborhood $\mathcal{N}$ of $[0_{[N-d]}\, 0_{[N-d+1,N]}] \in \Omega_{\mathbb{C}} \oplus (\Delta_{\mathbb{C}} + \Sigma_\beta)$ such that π^{-1} is well-defined, real-valued, and analytic by the Real-Analytic Inverse Function Theorem. For each $[\Theta\ L] \in \mathcal{N}$, there is an ε such that $\pi^{-1}([\Theta\ L]) = \Phi_\varepsilon(\Theta, L)$ since the $\{\Phi_\varepsilon(\Theta, L)\}_{\varepsilon\in(\{0,1\})^d}$ form an exhaustive list of the possible solutions to the inversion at that point. If there are two ε satisfying (i), there must be a neighborhood in $\mathcal{N}$ such that they disagree since they have distinct analytic expansions (or else we may apply the multivariable Open Mapping Theorem to obtain equality). Using this fact in an induction process gives us that there is exactly one ε satisfying (i) such that Φ_ε agrees with π^{-1} in some subneighborhood of $\mathcal{N}$. By the Open Mapping Theorem, this implies $\Phi_\varepsilon = \pi^{-1}$ everywhere in $\mathcal{N}$. Now, restricting $\mathcal{N}$ to satisfy the requirements of the first paragraph, we have that both (ii) and (iii) hold. $\square$

Chapter 4

Optimization

In this chapter, we develop machinery for performing approximate gradient descent over $\mathcal{F}_{\mathbb{E}}(\mu, S)$. The two main tools that make this efficient and plausible are

1. fast geometric minimization of the frame operator distance (FOD);

2. methods for projecting onto tangent spaces of $\mathcal{F}_{\mathbb{E}}(\mu, S)$.

After developing this machinery, we describe the full algorithm in detail and demonstrate its convergence.

4.1 Minimization of the frame operator distance (FOD)

We shall consider the program

$$\text{minimize } \|S - FF^*\| \text{ subject to } F \in \mathbb{T}_{\mathbb{E}^d}(\mu). \tag{4.1.1}$$

where $\mu \in \mathbb{R}_+^N$ and $S \in M_{d\times d}(\mathbb{E})$. In addition, we require that μ and S satisfy the usual conditions. It then follows that the minimum in (4.1.1) is zero and that this minimum is attained exactly when $FF^* = S$. We conclude that $\mathcal{F}_{\mathbb{E}}(\mu, S)$ is completely characterized as the space of solutions to (4.1.1).

4.1.1 Relationship with the frame potential

Suppose $S = \frac{N}{d} I_{d\times d}$ and $\mu = \mathbb{1}_N$. Armed with the knowledge that $F \in \mathbb{T}_{\mathbb{E}^d}(\mu)$ implies $\mathrm{tr}(FF^*) = N$, we have

$$\begin{aligned}
\|S - FF^*\|^2 &= \|\frac{N}{d} I_{d\times d}\|^2 - 2\mathrm{Re}\left\langle \frac{N}{d} I_{d\times d}, FF^* \right\rangle + \|FF^*\|^2 && (4.1.2)\\
&= \frac{N^2}{d} - 2\frac{N^2}{d} + \sum_{i=1}^{N}\sum_{j=1}^{N} |\langle f_i, f_j\rangle|^2 && (4.1.3)\\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} |\langle f_i, f_j\rangle|^2 - \frac{N^2}{d}. && (4.1.4)
\end{aligned}$$

The summation term is Benedetto and Fickus's frame potential, so (4.1.1) is equivalent to minimizing the frame potential. An important consequence of this relationship is demonstrated by the following theorem.

Theorem 4.1.1 (Benedetto, Fickus)**.** *For a given d and N with $d \leq N$, consider the frame potential*

$$FP : \mathbb{T}_{\mathbb{E}^d}(\mathbb{1}_N) \to [0, \infty);\ \{f_i\}_{i\in[N]} \mapsto \sum_{i\in[N]}\sum_{j\in[N]} |\langle f_i, f_j\rangle|^2 . \tag{4.1.5}$$

Then

a. Every local minimizer of the frame potential is also a global minimizer.

b. If $N \leq d$, the minimum value of the frame potential is N and the minimizers are precsiely the orthonormal sequences in $\mathbb{E}^d$.

c. If $N \geq d$, the minimum value of the frame potential is N^2/d and the minimizers are precisely the FUNTFs for $\mathbb{E}^d$.

Thus, in this particular instance, the minima of the frame operator distance satisfy $FF^* = \frac{N}{d} I_{d\times d}$. Having such a characterization is highly desirable from a numerical perspective, and generalizing Benedetto and Fickus's result to the case of (μ, S)-frames would provide a beneficial theoretical guarantee for the frame operator distance. Generalizing this result to (4.1.1) is nontrival, but there are several heuristic reasons that lead one to believe that this is true in general. We shall not take up this problem in this paper, but we look forward to settling it later.

Conjecture 4.1.2. *Suppose $N \geq d$. If d by d HPD S and $\mu \in \mathbb{R}_+^N$ satisfy the usual conditions, then local minimizers of (4.1.1) are also global minimizers in the generic case.*

It should also be noted that (4.1.1) can be generalized to the fusion frame case, but we shall not delve into this matter presently.

4.1.2 Relationship with the Rayleigh Quotient

By singling out one f_n in F (the n^{th} column), we obtain

$$\begin{aligned}
\|S - FF^*\|^2 &= \|S - \sum_{i=1}^{N} f_i f_i^*\|^2 = \|S - \sum_{i\neq n} f_i f_i^* - f_n f_n^*\|^2 \\
&= \|S - \sum_{i\neq n} f_i f_i^*\|^2 - 2\mathrm{Re}\left\langle S - \sum_{i\neq n} f_i f_i^*, f_n f_n^* \right\rangle + \|f_n f_n^*\|^2 \\
&= \|S - \sum_{i\neq n} f_i f_i^*\|^2 - 2 f_n^*(S - \sum_{i\neq n} f_i f_i^*) f_n + \|f_n\|^4. \qquad (4.1.6)
\end{aligned}$$

If only f_n is varied in its sphere, the only active term is $-2f_n^*(S - \sum_{i\neq n} f_i f_i^*) f_n$. This term is essentially a Rayleigh quotient, which is minimal when f_n is a "top" eigenvector of $S - \sum_{i\neq n} f_i f_i^*$.

If we allowed f_n to vary in $\mathbb{E}^d$, the gradient of this term would be

$$-4(S - \sum_{i \neq n} f_i f_i^*) f_n \tag{4.1.7}$$

and projection onto $T_{f_n}\mathcal{S}(\mu_n)$ yields

$$-4(S - \sum_{i \neq n} f_i f_i^*) f_n + 4 \left[f_n^* (S - \sum_{i \neq n} f_i f_i^*) f_n \right] f_n / \|f_n\|^2. \tag{4.1.8}$$

The full gradient matrix of the FOD over $\mathbb{T}_{\mathbb{E}}(\mu)$ at F is then $G = [g_1 \cdots g_N]$, where g_n is given by (4.1.8). For the FUNTF case, Casazza and Fickus [12] have demonstrated that a gradient descent can be performed using this G, and they also determine a reasonable step size for each update. This full gradient descent requires a full update of the gradient at each step. In the next section, we shall see that the columnwise gradient descent updates can be computed explicitly using nothing more than the quadratic formula, so updating the gradient requires few operations.

4.1.3 Gradient descent update of $-2f^*Af$ over a sphere

We now derive the gradient descent update to $R(f) = -2f^*Af$ for A a d by d HPD matrix and $\|f\|$ fixed. This update was alluded to by Arias, Edelman, and Smith in [2]. The gradient is $\widetilde{g} = -4Af + 4(f^*Af)f/\|f\|^2$, so we perform a line search along the geodesic

$$\gamma(t) = \sqrt{1 - t^2} f + tg, \; t \in [-1, 1], \tag{4.1.9}$$

where $g = \|f\|\widetilde{g}/\|\widetilde{g}\|$. The composition then expands into

$$(R \circ \gamma)(t) = -2(\sqrt{1-t^2}f + tg)^* A(\sqrt{1-t^2}f + tg) \tag{4.1.10}$$

$$= -2(1-t^2)f^*Af - 4t\sqrt{1-t^2}\mathrm{Re} f^*Ag - 2t^2 g^*Ag \tag{4.1.11}$$

$$= 2(f^*Af - g^*Ag)t^2 - 4t\sqrt{1-t^2}\mathrm{Re} f^*Ag - 2f^*Af \tag{4.1.12}$$

Taking a derivative, we have

$$(R \circ \gamma)'(t) = 4(f^*Af - g^*Ag)t - 4((1-2t^2)/\sqrt{1-t^2})\mathrm{Re} f^*Ag \tag{4.1.13}$$

and setting this expression equal to zero eventually reduces to

$$t^4 - t^2 + (\mathrm{Re} f^*Ag)^2/[(f^*Af - g^*Ag)^2 + 4(\mathrm{Re} f^*Ag)^2] = 0. \tag{4.1.14}$$

The solutions to this quadratic in t^2 are

$$t = \pm\sqrt{\frac{1}{2} \pm \frac{1}{2}\sqrt{\frac{(f^*Af - g^*Ag)^2}{(f^*Af - g^*Ag)^2 + 4(\mathrm{Re} f^*Ag)^2}}}, \tag{4.1.15}$$

and it is easy to deduce that the solution which minimizes $R \circ \gamma$ is

$$t_{\min} = -\sqrt{\frac{1}{2} - \frac{1}{2}\sqrt{\frac{(f^*Af - g^*Ag)^2}{(f^*Af - g^*Ag)^2 + 4(\mathrm{Re} f^*Ag)^2}}} \tag{4.1.16}$$

4.1.4 Sequential, columnwise minimization of the FOD

Implementation of a columnwise gradient descent scheme is now straightforward:

1. choose n such that the norm of (4.1.8) is maximal;
2. update f_n with $\sqrt{1-t_{\min}^2}f_n + t_{\min}g_n$;
3. and repeat.

We have the following pseudocode:

Algorithm 1 minFOD initialized with $F \in \mathbb{T}_{\mathbb{E}}(\mu)$, $\mathbb{E}$, μ, $d \times d$ HPD S, tolerance ε

while $\|S - FF^*\| > \varepsilon$ **do**

$n \leftarrow \arg\max_{n\in[N]} \|(S - FF^*)f_n - [f_n^*(S - FF^*)f_n]f_n/\|f_n\|^2\|$

$g \leftarrow [(S - FF^*)f_n - [f_n^*(S - FF^*)f_n]f_n/\mu_n^2$

$g \leftarrow \mu_n g/\|g\|$

$t \leftarrow -\sqrt{\frac{1}{2} - \frac{1}{2}\sqrt{\frac{(f_n^*Af_n - g^*Ag)^2}{(f_n^*Af_n - g^*Ag)^2 + 4(\mathrm{Re} f_n^*Ag)^2}}}$

$f_n \leftarrow \sqrt{1 - t^2} f_n + tg$

end while

4.1.5 A step size for the full gradient descent

The coordinate descent method detailed above is fast, but can result in very abrupt transitions. Consequently, the coordinate descent method can deliver an initial F to a member of $\mathcal{F}_{\mathbb{E}}(\mu, S)$ that deviates dramatically from the member we would obtain by moving F along the gradient flow. We shall often require smoother transitions since our optimization method over $\mathcal{F}_{\mathbb{E}}(\mu, S)$ requires that we compute accurate minimizers of the FOD for the perturbed members of $\mathcal{F}_{\mathbb{E}}(\mu, S)$. To this end, we derive a step size for the full gradient descent in the spirit of Casazza and Fickus [12].

Note that geodesics along $\mathbb{T}_{\mathbb{E}^d}(\mu)$ starting at $F = [f_1 \cdots f_N]$ have columns of the form

$$f_i(t) = \cos(\omega_i t) f_i + \sin(\omega_i t) g_i, \tag{4.1.17}$$

where $\mathrm{Re}\,\langle f_i, g_i\rangle = 0$ and $\|g_i\| = \|f_i\|$. Thus,

$$\dot{f}_i(t) = \omega_i \left[-\sin(\omega_i t) f_i + \cos(\omega_i t) g_i\right] \tag{4.1.18}$$

and $\ddot{f}_i(t) = -\omega_i^2 f_i(t)$.

Theorem 4.1.3. *Let $F \in \mathbb{T}_{\mathbb{E}^d}(\mu)$ for some $\mu \in \mathbb{R}_+^N$, set*

$$\gamma_n = -\left[S - FF^* - \frac{f_i^*(S - FF^*)f_i}{\mu_i^2} I_{d\times d}\right] f_i$$

and $\omega_i = \|\gamma_i\|/\mu_i$ for $i \in [N]$, and set $g_i = \gamma_i/\omega_i$ if $\omega_i \neq 0$ and $g_i = 0$ otherwise. Suppose further that μ and S satisfy the usual conditions. Then, for all $t \in \mathbb{R}$, we have

$$\|S - F(t)F(t)^*\|^2 \leq \|S - FF^*\|^2 - 4\|\Gamma\|^2 t + At^2, \tag{4.1.19}$$

where

$$A = \left(2(\lambda_1(S) - \lambda_d(S)) + 4\sum_{i\in[N]} \mu_i^2\right) \|\Gamma\|^2, \tag{4.1.20}$$

and $\Gamma = [\gamma_1 \cdots \gamma_N]$.

Proof. For a C^2 function, we have

$$\varphi(t) \leq \varphi(0) + \dot{\varphi}(0)t + \frac{1}{2}\max_{s\in\mathbb{R}} |\ddot{\varphi}(s)| t^2$$

by Taylor's theorem. It is immediately clear that the constant term is $\|S - FF^*\|$ for our $\varphi(t) = \|S - F(t)F(t)^*\|^2$. Moreover,

$$\begin{aligned}
\dot{\varphi}(0) &= 2\mathrm{Re}\left\langle \sum_{i\in[N]} \omega_i(g_i f_i^* + f_i g_i^*), S - FF^* \right\rangle & (4.1.21)\\
&= 4\sum_{i\in[N]} \omega_i \mathrm{Re}\, g_i^*(S - FF^*)f_i = 4\sum_{i\in[N]} \omega_i \mathrm{Re}\, g_i^* \gamma_i & (4.1.22)\\
&= 4\sum_{i\in[N]} \gamma_i^* \gamma_i = 4\|\Gamma\|^2. & (4.1.23)
\end{aligned}$$

For the second order term, we compute (while suppressing many evaluations at t)

$$\begin{aligned}
\ddot{\varphi}(t) &= 2\mathrm{Re}\left\langle \sum_{i\in[N]} \ddot{f}_i f_i^* + f_i \ddot{f}_i^* + 2\dot{f}_i \dot{f}_i^*, S - F(t)F(t)^* \right\rangle + 2\| \sum_{i\in[N]} \dot{f}_i f_i^* + f_i \dot{f}_i^* \|^2 \qquad (4.1.24)\\
&= 4\mathrm{Re}\left\langle \sum_{i\in[N]} \dot{f}_i \dot{f}_i^* - \omega_i^2 f_i f_i, S - F(t)F(t)^* \right\rangle \qquad (4.1.25)\\
&\quad +4\left\| \sum_{i\in[N]} \dot{f}_i f_i^* \right\|^2 + 4\mathrm{Re}\left\langle \sum_{i\in[N]} \dot{f}_i f_i^*, \sum_{i\in[N]} f_i \dot{f}_i^* \right\rangle \qquad (4.1.26)
\end{aligned}$$

The term in (4.1.25) becomes

$$4(\sum_{i\in[N]} \dot{f}_i^* S \dot{f}_i - \omega_i^2 f_i^* S f_i) - 4\sum_{i\in[N]}\sum_{j\in[N]} |\dot{f}_i^* f_j|^2 + 4\left\langle \sum_{i\in[N]} \omega_i^2 f_i f_i^*, \sum_{i\in[N]} f_i f_i^* \right\rangle$$

which is bounded by

$$4(\lambda_1(S) - \lambda_d(S)) \sum_{i\in[N]} \mu_i^2 \omega_i^2 + 4(\sum_{i\in[N]} \mu_i^2) \sum_{i\in[N]} \mu_i^2 \omega_i^2 - 4\sum_{i\in[N]}\sum_{j\in[N]} |\dot{f}_i^* f_j|^2 \qquad (4.1.27)$$

since $\lambda_d(S)\|x\|^2 \leq x^* S x \leq \lambda_d(S)\|x\|^2$ for all $x \in \mathbb{E}^d$, and

$$\left\langle \sum_{i\in[N]} \omega_i^2 f_i f_i^*, \sum_{i\in[N]} f_i f_i^* \right\rangle \leq \sum_{i\in[N]} \omega_i^2 \|f_i f_i^*\| \| \sum_{j\in[N]} f_j f_j^* \| \leq \sum_{i\in[N]} \omega_i^2 \mu_i^2 \sum_{j\in[N]} \mu_j^2$$

by Cauchy-Schwarz. By three more applications of Cauchy-Schwarz, we also have that (4.1.26) is bounded by

$$4\sum_{i\in[N]}\sum_{j\in[N]} |\dot{f}_i^* f_j|^2 + 4(\sum_{i\in[N]} \mu_i^2) \sum_{i\in[N]} \mu_i^2 \omega_i^2. \qquad (4.1.28)$$

Combining (4.1.27) and (4.1.28), and dividing by 2, we arrive at the expression for A, and the proof is complete. $\square$

Note that minimum of the quadratic inequality in this theorem occurs at

$$t = \frac{1}{\lambda_1(S) - \lambda_d(S) + 2\sum_{i\in[N]} \mu_i^2}, \qquad (4.1.29)$$

and we may use this as our step size for the full gradient descent.

4.1.6 Complexity and convergence of the minFOD algorithm

Each iteration of this algorithm requires the computation of $(S-FF^*)F$, which is a parallel $O(dN^2)$ operation. The update of $S - FF^*$ requires $O(d^2)$ multiplications. The remaining operations are negligible, so the computational complexity of minFOD is $O(dN^2 + d^2)$ per iteration.

Newton-like Rayleigh quotient iterations exhibit cubic convergence [35], but at each stage we are only performing one iteration of a linearly convergent gradient descent. Nevertheless, each step of this procedure is ensured to lower the FOD as long as the gradient is nonzero because of the decoupling evident in (4.1.6). Thus, it is certain that minFOD converges to a critical point. The convergence rate of general block coordinate descent is still an open problem, but there are partial results in the literature [32, 47]. However, we can be certain that the iteration converges to a critical point. If Conjecture 4.1.2 holds, then the only stable critical points are the global minimizers.

Preliminary empirical results indicate that this method often converges linearly and that the procedure converges to a global minimizer in the generic case. Figure 1 illustrates a typical convergence profile obtained from minFOD. The solid lines are for geometric coordinate descent, and the dashed lines come from the full geometric gradient descent. The left plot illustrates the convergence of the FOD to zero, and the right illustrates the linear convergence of iterates. Note that the full geometric gradient descent converges more slowly, but its convergence profile is smoother.

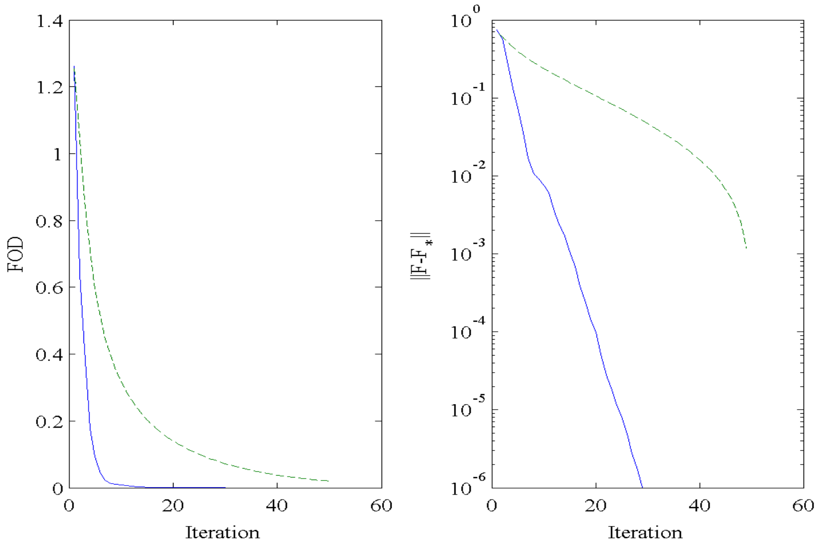

Figure 4.1: Convergence of minFOD to a FUNTF of 5 elements in $\mathbb{R}^3$.

4.2 Optimization over $\mathcal{F}_{\mathbb{E}}(\mu, S)$

4.2.1 Projection of search directions onto $T_F\mathcal{F}_{\mathbb{E}}(\mu, S)$

For this entire section, F is assumed to be a regular point of $\mathcal{F}_{\mathbb{E}}(\mu, S)$ and hence $T_F\mathcal{F}_{\mathbb{E}}(\mu, S)$ is well-defined.

While the orthogonal projection onto $T_F\mathcal{F}_{\mathbb{E}}(\mu, S)$ is complicated, the orthogonal projection onto $T_F\mathbb{T}_{\mathbb{E}^d}(\mu)$ is simple:

$$P(X) = X - F \operatorname{Re} \operatorname{diag}(F^*X) \operatorname{diag}(F^*F)^{-1} \tag{4.2.1}$$

The orthogonal projection onto $T_F\sqrt{S} \cdot \mathrm{St}_{\mathbb{E}^d}(N)$ is more complicated, but is quite simple in at least one notable case.

Proposition 4.2.1. *Suppose $S = cI_{d\times d}$ for $c > 0$, and fix $F \in \mathcal{F}_{\mathbb{E}}(\mu, S)$. Then,*

$$Q(X) = X - \frac{1}{2}(FX^* + XF^*)S^{-1}F \tag{4.2.2}$$

is the orthogonal projection onto $T_F\sqrt{S}\cdot St_{\mathbb{E}^d}(N)$.

Proof. We have that

$$\begin{aligned}
Q^2(X) &= Q(X) - \frac{1}{2}(FQ(X)^* + Q(X)F^*)S^{-1}F \\
&= Q(X) - \frac{1}{2}[FX^* + XF^* \\
&\quad -\frac{1}{2}FF^*S^{-1}(FX^* + XF^*) - \frac{1}{2}(FX^* + XF^*)S^{-1}FF^*]S^{-1}F \\
&= Q(X)
\end{aligned}$$

since $FF^* = S$. Thus, $Q^2 = Q$ and the following calculation shows that $Q^T = Q$ (we are viewing the tangent space as a real vector space):

$$\begin{aligned}
\operatorname{Re}\langle Q(X), Y\rangle &= \operatorname{Re}\langle X, Y\rangle - \frac{1}{2}\operatorname{Re}\left\langle (FX^* + XF^*)S^{-1}F, Y\right\rangle \\
&= \operatorname{Re}\langle X, Y\rangle - \frac{1}{2}\operatorname{Re}\operatorname{tr}(FX^*S^{-1}FY^*) - \frac{1}{2}\operatorname{Re}\operatorname{tr}(XF^*S^{-1}FY^*) \\
&= \operatorname{Re}\langle X, Y\rangle - \\
&\quad \frac{1}{2c}\operatorname{Re}\sum_{i=1}^{N}\sum_{j=1}^{N}\langle f_i, x_j\rangle\langle f_i, y_j\rangle - \frac{1}{2}\operatorname{Re}\operatorname{tr}(XF^*S^{-1}FY^*) \\
&= \operatorname{Re}\langle X, Y\rangle - \\
&\quad \frac{1}{2c}\operatorname{Re}\sum_{i=1}^{N}\sum_{j=1}^{N}\langle x_j, f_i\rangle\langle y_j, f_i\rangle - \frac{1}{2}\operatorname{Re}\operatorname{tr}(XF^*S^{-1}FY^*) \\
&= \operatorname{Re}\langle X, Y\rangle - \frac{1}{2c}\operatorname{Re}\operatorname{tr}(XF^*YF^*) - \frac{1}{2}\operatorname{Re}\operatorname{tr}(XF^*S^{-1}FY^*) \\
&= \operatorname{Re}\left\langle X, Y - \frac{1}{2}(FY^* + YF^*)S^{-1}F\right\rangle \\
&= \operatorname{Re}\langle X, Q(Y)\rangle
\end{aligned}$$

□

If S is not a multiple of the identity, the orthogonality of the right hand side of (4.2.2) fails. In this case, we may construct Q by forming an orthonormal basis of

$$T_F\sqrt{S}\cdot \mathrm{St}_{\mathbb{E}^d}(N) \quad = \quad \{X \in M_{d\times N}(\mathbb{E}) : X = FZ,\ Z = -Z^*\}. \tag{4.2.3}$$

To construct this orthonormal basis, we simply apply the Gram-Schmidt procedure to the basis defined in Proposition 4.2.2.

Proposition 4.2.2. *Let S be a d by d HPD matrix, and suppose $F \in \sqrt{S}\cdot St_{\mathbb{E}^d}(N)$ for $d \leq N$. Fix $A \subset [N]$ so that F_A is a basis. For bookkeeping purposes, define*

$$\Lambda_1 = (A \times ([N] \setminus A)) \cup \{(i,j) \in A \times A : i < j\} \text{ and } \Lambda_2 = \Lambda_1 \cup \{(i,i) \in A \times A\}.$$

Then

$$\{V_{ij} = F(\mathbb{1}_{ij} - \mathbb{1}_{ji}) : (i,j) \in \Lambda_1\} \tag{4.2.4}$$

is a basis of $T_F\sqrt{S}\cdot St_{\mathbb{R}^d}(N)$ (assuming S and F are real valued) and

$$\{V_{ij}\}_{(i,j)\in\Lambda_1} \cup \{W_{ij} = \sqrt{-1}\cdot F(\mathbb{1}_{ij} + \mathbb{1}_{ji}) : (i,j) \in \Lambda_2\} \tag{4.2.5}$$

is a basis of $T_F\sqrt{S}\cdot St_{\mathbb{C}^d}(N)$.

Proof. Note that the V_{ij} have the form FZ with Z skew-symmetric (and skew-Hermitian), and that the W_{ij} have the same form but with Z skew-Hermitian. Thus, these collections belong to the respective tangent spaces.

We shall now show that $\{V_{ij}\}_{(i,j)\in\Lambda_1}$ and $\{W_{ij}\}_{(i,j)\in\Lambda_2}$ are linearly independent sets in their respective spaces. Suppose that

$$\sum_{(i,j)\in\Lambda_1} a_{ij}V_{ij} + \sum_{(i,j)\in\Lambda_2} b_{ij}W_{ij} = 0 \tag{4.2.6}$$

with the a_{ij} and b_{ij} real-valued. Breaking this into the real and imaginary components, we have

$$\sum_{(i,j)\in\Lambda_1} a_{ij}V_{ij} = 0 \text{ and } \sum_{(i,j)\in\Lambda_2} b_{ij}W_{ij} = 0. \tag{4.2.7}$$

Since the k^{th} column of V_{ij} is zero for all $j \neq k$, the contributions to the k^{th} column in this sum only come from $\{V_{ik}\}_{i\in A}$. However, the k^{th} columns of $\{V_{ik}\}_{i\in A}$ come from the columns of F_A and are thus linearly independent. We conclude that $a_{ik} = 0$ for $i \in A$. Since k was arbitrary, $a_{ij} = 0$ for all $(i,j) \in \Lambda_1$. Similarly, $b_{ij} = 0$ for all $(i,j) \in \Lambda_2$. A counting argument now finishes the proof. □

Applying the Gram-Schmidt process to this basis, we obtain an orthonormal basis $\{U_i\}_{i\in\Lambda}$ for $T_F\sqrt{S}\cdot \mathrm{St}_{\mathbb{E}^d}(N)$. Then

$$Q(X) = \sum_{i\in\Lambda} (\mathrm{Re}\,\langle X, U_i\rangle)U_i$$

is the orthogonal projection from $M_{d\times N}(\mathbb{E})$ to $T_F\sqrt{S}\cdot \mathrm{St}_{\mathbb{E}^d}(N)$.

In order to perform geometric optimization, we need to project search directions onto $T_F\mathcal{F}_{\mathbb{E}}(\mu, S)$. By an eigenvalue argument, it can be shown that

$$\lim_{n\to\infty} (QP)^n = R,$$

where R is the orthogonal projection onto $T_F\mathbb{T}_{\mathbb{E}^d}(\mu) \cap T_F\sqrt{S}\cdot \mathrm{St}_{\mathbb{E}^d}(N)$. If the conditions of the preceding theorem hold, then R is the orthogonal projection onto $T_F\mathcal{F}_{\mathbb{E}}(\mu, S)$. While the convergence of this alternating projection method is only linear, it can be implemented efficiently.

Under certain circumstances (such as when $F \in \mathcal{F}_{\mathbb{E}}(\mu, S)$ is close to being singular) it is more efficient to directly compute the orthogonal projection

$R : M_{d\times N}(\mathbb{E}) \to T_F\mathcal{F}_{\mathbb{E}}(\mu, S)$. In this case, we construct R by taking the basis provided in 2.3.4, applying the Gram-Schmidt procedure to this basis, and then R is the sum of dyadic products of the resulting orthonormal basis.

4.2.2 Description of the optimization algorithm

We now describe how the FOD minimization algorithm and the alternating projection method can be exploited to perform an approximate gradient descent on $\mathcal{F}_{\mathbb{E}}(\mu, S)$. Let $\varphi \in C^1(M_{d\times N}(\mathbb{E}))$. We begin by describing how one iteration of the theoretical geometric gradient descent algorithm is performed at a nonsingular point $F \in \mathcal{F}_{\mathbb{E}}(\mu, S)$:

1. the restricted gradient, $\Phi = \nabla_F^{\mathcal{F}_{\mathbb{E}}(\mu,S)}\varphi$ is computed;
2. a line search is performed on the geodesic $\exp_F^{\mathcal{F}_{\mathbb{E}}(\mu,S)}(-t\Phi)$.

This first step is now easily carried out numerically with the machinery we have developed; our alternating projection procedure allows us to compute $\nabla_F^{\mathcal{F}_{\mathbb{E}}(\mu,S)}\varphi$ by projecting the full gradient $\nabla_F\varphi$ onto $T_F\mathcal{F}_{\mathbb{E}}(\mu, S)$. Carrying out the second step numerically using our machinery is slightly more complicated. Instead of performing a line search on a geodesic in $\mathcal{F}_{\mathbb{E}}(\mu, S)$, we perform a line search on the FOD minimization applied along a geodesic in $\mathbb{T}_{\mathbb{E}^d}(\mu)$. That is, we perform the line search on the path

$$\text{minFOD}(\exp_F^{\mathbb{T}_{\mathbb{E}^d}(\mu)}(-tX), S, \varepsilon) \tag{4.2.8}$$

and the update is

$$F' = \arg\min_{t\geq 0} \varphi(\text{minFOD}(\exp_F^{\mathbb{T}_{\mathbb{E}^d}(\mu)}(-tX), S, \varepsilon)). \tag{4.2.9}$$

The primary reason for exponentiating along the generalized torus is that the paths are easy to compute and the FOD minimization is efficient. The following psuedocode summarizes the procedure.

Algorithm 2 minFEmuS initialized with φ, F, $\mathbb{E}$, μ, S, and ε

$\Phi \leftarrow \nabla_F^{\mathbb{T}_{\mathbb{E}^d}} \varphi$

$\Phi \leftarrow \text{projTFEmuS}(F, \mathbb{E}, \mu, S)$

while $\|\Phi\| > \varepsilon$ **do**

  $F \leftarrow \text{linesearch}(\varphi(\text{minFOD}(\exp_F^{\mathbb{T}_{\mathbb{E}^d}}(-t\Phi), S, \varepsilon)))$

  $\Phi \leftarrow \nabla_F^{\mathbb{T}_{\mathbb{E}^d}} \varphi$

  $\Phi \leftarrow \text{projTFEmuS}(F, \mathbb{E}, \mu, S)$

end while

In this algorithm, projTFEmuS is a realization of one of the projection methods detailed in Section 3.3. For our purposes, we use the golden ratio line search algorithm.

4.2.3 Convergence of the approximate gradient flow

Suppose $\varphi : M_{d\times N}(\mathbb{E}) \to \mathbb{R}$ is our objective function and that $F \in \mathcal{F}_{\mathbb{E}}(\mu, S)$ is not a critical point of $\varphi|_{\mathcal{F}_{\mathbb{E}}(\mu,S)}$. Then $\Phi = \nabla_F^{\mathcal{F}_{\mathbb{E}}(\mu,S)} \varphi$ is nonzero and we define the smooth curve $\gamma(t) = \exp_F^{\mathbb{T}_{\mathbb{E}^d}(\mu)}(t\Phi)$ for $t \in [0, T]$. Here, T is chosen so that $\gamma(t)$ avoids self intersection. Since γ is smooth, it is an integral curve of some vector field

Y.

The frame operator distance to S induces a gradient flow on $\mathbb{T}_{\mathbb{E}^d}(\mu)$, and we let X denote this gradient flow. Since $\mathcal{F}_{\mathbb{E}}(\mu, S)$ are the global minimizers of the frame operator distance to S on $\mathbb{T}_{\mathbb{E}^d}(\mu)$, X vanishes on $\mathcal{F}_{\mathbb{E}}(\mu, S)$. Moreover, the gradient flow of each scalar component of X is orthogonal to $T_F\mathcal{F}_{\mathbb{E}}(\mu, S)$, and hence is orthogonal to Y_F. Formally, $X_F = 0$, a brief ascension into local coordinates and Einstein summation implies

$$(Y(X))_F = \left(Y^\alpha \frac{dX^\beta}{dx^\alpha} \frac{d}{dx^\beta} \right)_F = 0, \tag{4.2.10}$$

and therefore

$$[X, Y]_F = (X(Y) - Y(X))_F = 0. \tag{4.2.11}$$

When we evolve the curve $\gamma(t)$ along the flow X, $X_F = 0$ implies that $\gamma(0) = F$ remains stationary and $[X, Y]_F = 0$ implies that tangent at $\gamma(0)$ also remains stationary. Thus, we have

$$\nabla_F^{\mathcal{F}_{\mathbb{E}}(\mu,S)} \varphi = \nabla_F^{\mathcal{F}_{\mathbb{E}}(\mu,S)} (\text{minFOD} \circ \gamma), \tag{4.2.12}$$

and so φ is locally strictly decreasing along the curve $\text{minFOD}(\gamma(t))$. We conclude that the optimization procedure produces a sequence F_i such that $\varphi(F_i)$ is strictly decreasing, and hence the procedure converges to a critical point.

While [38] indicates that geometric gradient descent converges linearly near an isolated minimum, our line searches do not occur over geodesics. This makes the analysis of the local convergence difficult, but empirical evidence indicates that linear convergence is still maintained. Figure 2 depicts the convergence profile of a sequence ending in a numerical minimizer of the 4^{th}-order frame potential over the space of 6-member FUNTFs in $\mathbb{R}^3$. The resulting frame is a numerical approximation to an equiangular tight frame. Solid lines are for the implementation of minFEmuS with geometric coordinate gradient descent, and dashed lines are for the implementation with full geometric gradient descent. The left plot depicts the decay profile of fp4, and the right plot illustrates the convergence profile. Both methods recover an equiangular tight frame.

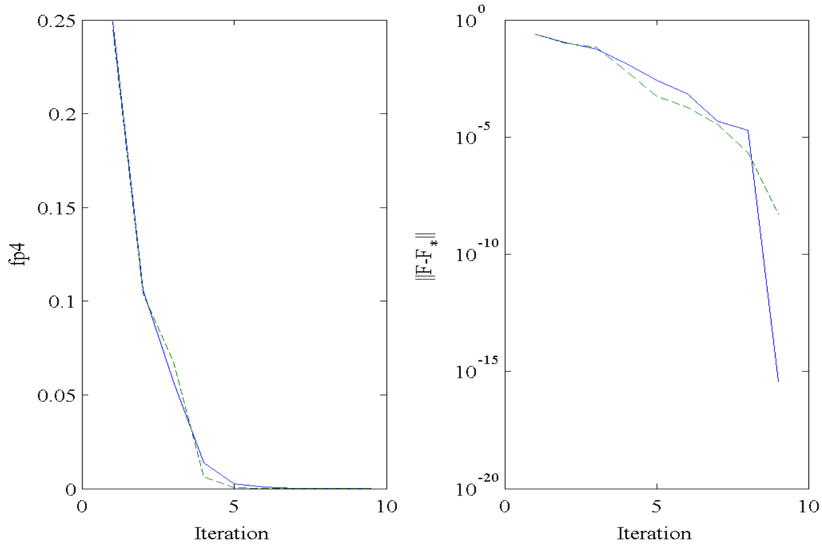

Figure 4.2: Hilbert-Schmidt distances from the final numerical minimum of the 4^{th}-order frame potential on 6-member FUNTFs in $\mathbb{R}^3$.

4.3 Direct construction of initial points on $\mathcal{F}_{\mathbb{E}}(\mu, S)$

Now that we have an optimization algorithm over $\mathcal{F}_{\mathbb{E}}(\mu, S)$, we turn our attention to the construction of starting points for initializing the optimization algorithm. This section focuses on a direct, sequential method for constructing members of $\mathcal{F}_{\mathbb{E}}(\mu, S)$. We shall construct $F = [f_1 \cdots f_N] \in \mathcal{F}_{\mathbb{E}}(\mu, S)$ column by column. For low dimensional examples of this construction technique, see [40].

Let $\mu' \in \mathbb{R}_+^{N-1}$ denote the truncation of μ obtained by omitting the first entry. Our first task is to characterize the set of all $f \in \mathbb{E}^d$ such that

$$\mathcal{F}_{\mathbb{E}}(\mu', S - ff^*) \neq \emptyset \tag{4.3.1}$$

since this is equivalent to failure of the construction.

4.3.1 The ellipsoidal condition for positivity of $S - ff^*$

One immediate necessary condition of (4.3.1) is that $S - ff^*$ must be positive semidefinite, and the f satisfying this condition can be characterized exactly by an ellipsoidal condition.

Proposition 4.3.1. *Let $S \in M_{d\times d}(\mathbb{E})$ be Hermitian postive semidefinite, and suppose $f \in \mathbb{E}^d$. Denote the Moore-Penrose pseudoinverse of S by $S^\dagger$. Then*

(i) $S - ff^*$ *is Hermitian positive semidefinite if any only if* $f \in \operatorname{range}(S)$ *and* $f^* S^\dagger f \leq 1$.

(ii) $\operatorname{rank}(S - ff^*) = \operatorname{rank}(S) - 1$ *if and only if* $f \in \operatorname{range}(S)$ *and* $f^* S^\dagger f = 1$.

Proof. Note that $x^*(S - ff^*)x = -|\langle x, f\rangle|^2 \leq 0$ for any $x \in \text{range}(S)^\perp$, so $f \in \text{range}(S)$ is a necessary condition, and we may assume that S has full rank without loss of generality. Now, $S - ff^*$ has at most one negative eigenvalue by the interlacing inequalities for eigenvalues, and hence $S - ff^*$ has all nonnegative eigenvalues if and only if $\det(S - ff^*) \geq 0$. Using multilinearity of the determinant and Cramer's rule, it is easy to deduce the Sherman-Morrison determinant formula

$$\det(S - ff^*) = \det(S)(1 - f^* S^{-1} f). \tag{4.3.2}$$

Applying this formula and noting that $\det(S) > 0$, we have $f^* S^{-1} f \leq 1$ and $S - ff^*$ positive semidefinite are equivalent. This proves part *(i)*, and part *(ii)* follows by noting that this case is equivalent to $\det(S - ff^*) = 0$ by the interlacing eigenvalue inequalities. ☐

Part *(ii)* of this proposition addresses another necessary condition: if there are only N entries in μ, then $S - ff^*$ must have rank less than $N - 1$. A combinatorial proof of this result was established in [40].

4.3.2 Sufficient conditions for ensuring majorization

Having determined when $S - ff^*$ is positive semidefinite, we know that (4.3.1) holds if and only if μ' and $S - ff^*$ satisfy the usual conditions. Thus, for a given f, it is easy to check (4.3.1) by computing the eigenvalues of $S - ff^*$ and determining if the majorization condition holds, but this approach is highly inefficient. Instead, we content ourselves with some easily-computed sufficient condtions. Given S, let

P_k denote the orthogonal projetion ont the subspace spanned by the eigenvectors of S corresponding to the k largest eigenvalues of S, and set $P_k^{\perp} = I_{d\times d} - P_k$.

Theorem 4.3.2. *Suppose μ and S satisfy the usual conditions, that $f \in \mathrm{range}(S)$, and $f^*S^{\dagger}f \le 1$. Then 4.3.1 holds if for each $k = 1, \cdots, d-1$ one of the following holds:*

(i) $\|P_k f\|^2 \le c_k$;

(ii) *(i) fails, but* $\|P_k f\|^2 > 0$ *and* $\rho_k \le \lambda_{k+1}(S) + c_k$;

(iii) *(i) and (ii) both fail, but* $c_k > 0$ *and*

$$\|P_k f\|^2/c_k + f^* P_k^{\perp}[S + (c_k - \rho_k)I_{d\times d})]^{\dagger} P_k^{\perp} f \le 1. \tag{4.3.3}$$

Here,

$$c_k = \sum_{i=1}^{k} \lambda_k(S) - \max_{\substack{A\subset[N-1]\\|A|=k}} \sum_{i\in A} \mu_i' \tag{4.3.4}$$

and $\rho_k = \|\sqrt{S}P_k f\|^2/\|P_k f\|^2$

For a proof of this theorem, see [40]. For uniform-norm frames, (4.3.1) holds trivially as long as the ellipsoidal condition is satisfied.

4.3.3 Sherman-Morrison formulas and efficient eigensystem updates

In order to apply the ellipsoidal condition and Theorem (4.3.2) as we choose vectors, we need to compute $(S - ff^*)^{\dagger}$ and the eigensystem of $S - ff^*$ efficiently. To compute $(S - ff^*)^{\dagger}$ efficiently when $f \in \mathrm{range}(S)$ and $f^*S^{\dagger}f < 1$, we apply the

Sherman-Morrison formula:

$$(S - ff^*)^\dagger = S^\dagger - S^\dagger f f^* S^\dagger / (1 - f^* S^\dagger f)$$

If $f^* S^\dagger f = 1$, then we must apply the formula

$$(S - ff^*)^\dagger \;=\; S^\dagger - \frac{S^{2\dagger} f f^* S^\dagger + S^\dagger f f^* S^{2\dagger}}{f^* S^{2\dagger} f} + \frac{(f^* S^{3\dagger} f) S^\dagger f f^* S^\dagger}{(f^* S^{2\dagger} f)^2}.$$

This formula follows very readily when one realizes that $S^\dagger f$ is in the kernel of $S - ff^*$ in this case. Bunch et al. [10] describe an efficient numerical method for the symmetric rank-one eigensystem update.

4.3.4 Description of an initialization algorithm

Here, we describe a procedure for constructing a frame in $\mathcal{F}_{\mathbb{E}}(\mu, S)$ from an arbitrary $F = [f_1 \cdots f_N] \in \mathbb{T}_{\mathbb{E}^d}(\mu)$ so that the the resulting frame, $\widetilde{F}$ so that many of the columns are equal to, or close to corresponding columns in F. This is accomplished by constructing a set of indices $A \subset [N]$ for which $\widetilde{f_i} = f_i$ for all $i \in A$, and then applying a restricted minFOD to the columns with indices in $[N] \setminus A$

First, we set $A = \emptyset$. In the first step of the algorithm, we check that conditions of Proposition 4.3.1 and Theorem 4.3.2 hold for f_1 given μ and S. If they hold, then $A = A \cup \{1\}$ and we set $\widetilde{S} = S - f_1 f_1^*$. If they fail to hold, nothing is changed. This process continues inductively through the columns. That is, at the n^{th} step, we check to see that the conditions hold for f_n given μ and $\widetilde{S}$, and update $A = A \cup \{n\}$ and $\widetilde{S} = S - f_n f_n^*$.

Algorithm 3 initFEmuS initialized with F, $\mathbb{E}$ μ, S, ε

$A \leftarrow \emptyset$

for $n \in [N]$ **do**

  if chkMAJ$(\mu, S, f_n) =$ true **then**

    $A \leftarrow A \cup n$

    $S \leftarrow S - f_n f_n^*$

  end if

end for

$F_{[N]\setminus A} \leftarrow \text{minFOD}(F_{[N]\setminus A}, \mathbb{E}, \mu, S, \varepsilon)$

4.4 Applications to Grassmannian frames and WBE sequences

4.4.1 Grassmannian frames

Grassmannian frames are FUNTFs that also are also minimizers of the off-diagonal infinity norm of the Gram matrix. They are the solutions to the program

$$F = \arg \min_{F \in \mathcal{F}_{\mathbb{E}}(\mu,S)} \max_{i \neq j} | \langle f_i, f_j \rangle |. \tag{4.4.1}$$

These very special frames have been studied extensively, and have applications in coding theory and communications [42]. Now, the objective function in (4.4.1) is not differentiable, so we may attempt to replace with the differentiable approximation

$$\left(\sum_{i \neq j} | \langle f_i, f_j \rangle |^p \right)^{1/p} \tag{4.4.2}$$

for $p \in (2, \infty)$. This quantity is the p^{th}-*order frame potential.* A theoretical indication of the validity of this approximation is given by the result of Oktay [34]:

Theorem 4.4.1. *Let $d < N$, $1 < p < \infty$, and let $\{x_i\}_{i=1}^N$ be a set of unit norm vectors in $\mathbb{E}^d$. Then,*

$$\sum_{i \neq j} |\langle x_i, x_j \rangle|^{2p} \geq N(N-1) \left(\frac{N-d}{d(N-1)} \right)^p . \tag{4.4.3}$$

Furthermore, the lower bound is achieved if and only if $\{x_i\}_{i=1}^N$ is an equiangular tight frame.

It is natural to question whether this elegant result extends to Grassmannian frames in some form, but currently there are no theoretical guarantees that minimizers of (4.4.2) are even close to minimizers of (4.4.1).

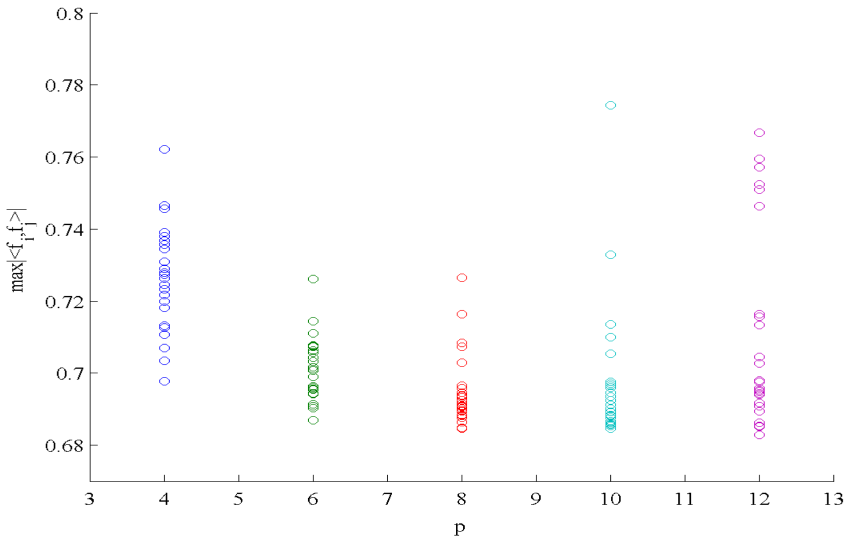

Figure 4.3: Maximum values of $\max_{i \neq j} |\langle f_i, f_j \rangle|$ for numerical minimizers of the p^{th}-order frame potential when $p = 4, 6, 8, 10, 12$.

Figure 3 summarizes the results obtained by minimizing the p^{th}-order frame potential over the 8-member FUNTFs in $\mathbb{R}^5$. Note that the range of $\max_{i \neq j} |\langle f_i, f_j \rangle|$ becomes lower as p increases, but becomes wide as p increases. The reason that this occurs is that the numerical accuracy of the gradient computation becomes unstable as p increases. The graph indicates that we should fix $p = 8$ or 10. In Figure 4, numerically computed minimum values of the 4^{th}-order frame potential are compared with the corresponding $\max_{i \neq j} | \langle f_i, f_j \rangle |$. There is obviously a strong correlation between these two quantities, and we infer that global minimizers of (4.4.2) are nearly global minimizers of (4.4.1). It should be noted that our method computes an equiangular tight frame whenever one exists.

It is clear from Figure 4 that numerous local minima exist for the 4^{th}-order frame potential. However, the figure also shows how a large percentage of these local minima occur near the minimum value. There are no theoretical guarantees that this is near the true minimum value, but this empirical data is compelling.

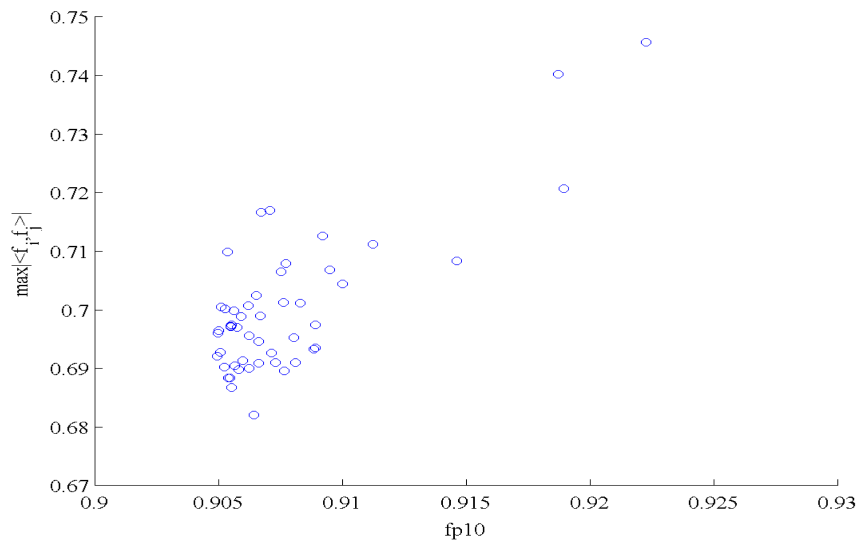

Figure 4.4: Numerical minimums of the 4^{th}-order frame potential versus values of $\max_{i \neq j} | \langle f_i, f_j \rangle |$.

4.4.2 WBE sequences with maximal separation

Direct algorithms for constructing Welch bound equality sequences were discussed in Tropp et al. in [45]. In the while noise, unequal norm case, a WBE sequence is exactly a member of $\mathcal{F}_{\mathbb{E}}(\mu, cI_{d\times d})$, where $\mu \in \mathbb{R}_+^N$ is the list of average powers of each user in the uplink and $c = \sum_{i=1}^N \mu_i/d$. In this section, we apply our optimization framework to numerically construct WBE sequences that are also maximally separated. WBE sequences with maximal separation minimize interference between users.

Assuming that μ and S satisfy the usual conditions, $\mathcal{F}_{\mathbb{E}}(\mu, S)$ is non-empty and we may seek

$$F = \arg \min_{F\in\mathcal{F}_{\mathbb{E}}(\mu,S)} \max_{i\neq j} |\langle f_i, f_j\rangle|. \tag{4.4.4}$$

The objective function in this minimization program is not differentiable, so we instead seek

$$F = \arg \min_{F\in\mathcal{F}_{\mathbb{E}}(\mu,S)} \left(\sum_{i\neq j} |\langle f_i, f_j\rangle|^4\right)^{1/4}. \tag{4.4.5}$$

As an experiment, we compute a random μ that is majorized by $c\mathbb{1}_d$. Here, we have set $N = 13$ and $d = 8$. We then minimize the objective function described above and plot these minimums versus the maximum absolute correlation of the system. Our results from our experiment are summarized in Figure 5. Again, numerous minima exist.

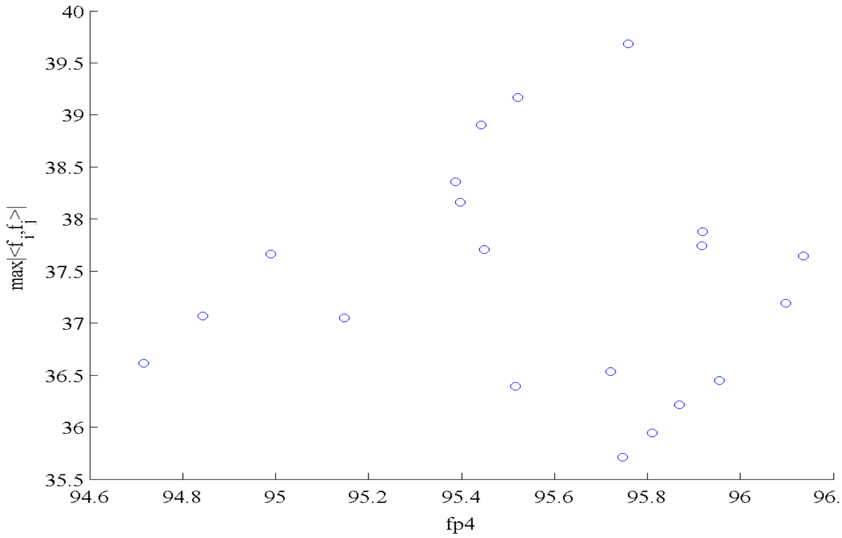

Figure 4.5: Numerical minimums of the 4^{th}-order frame potential versus values of $\max_{i \neq j} |\langle f_i, f_j \rangle|$.

Chapter 5

Conclusion and future work

This work is the culmination of an eight year journey into finite frame theory, and represents a leap forward for the problem of designing frames for applications. Nevertheless, the gradient descent optimization procedure that we have elucidated is relatively elementary and we would like to use the local parameterizations derived in Chapter 3 to perform sophisticated optimization. Moreover, this research all began as an attempt to develop an understanding of finite frames that could be exploited to bring down the Kadison-Singer problem [28], but there is still not a clear path to the resolution of this problem.

5.1 Beyond gradient descent optimization

In theory, the local coordinate systems derived here can be used to perform direct optimization over $\mathcal{F}_{\mathbb{E}}(\mu, S)$ without having to minimize the FOD at every step. The reason that we have not described such a procedure is that there is no theory that quantitatively ensures the validity of the coordinate systems. All we know is that they are valid locally. In the future, we hope to overcome this technical obstacle.

Problem 5.1.1. *Find a computable lower bound for the radius of a ball in which the coordinate systems from Theorems 3.4.1 and 3.5.1 are valid.*

Related to this issue is the Paulsen problem, which is currently being investigated with gusto.

5.2 The Kadison-Singer problem

The original formulation of this problem is in the language of C^*-algebras.

Kadison-Singer Problem 5.2.1. *Does every pure state on an atomic maximal, abelian self-adjoint subalgebra of $\mathcal{B}(\mathcal{H})$ extend uniquely to a pure state on $\mathcal{B}(\mathcal{H})$?*

Nik Weaver introduced an equivalent formulation of the Kadison-Singer conjecture [49] that is more accessible from a frame-theoretic perspective.

Theorem 5.2.2 (Weaver)**.** *The following are equivalent:*

1. *The Kadison-Singer Problem has a positive solution.*

2. *There is some natural number r so that there exists universal constants $K \geq 4$ and $\varepsilon > \sqrt{K}$ such that the following holds: Let $\{f_i\}_{i=1}^N$ be a frame for $\mathbb{E}^d$ satisfying $\|f_i\| = 1$ for all $i \in [N]$ and suppose*

$$\sum_{i=1}^{N} |\langle x, f_i \rangle|^2 \leq K$$

for every unit vector $x \in \mathbb{E}^d$. Then there exists a partition $\{I_j\}_{j=1}^r$ of $[N]$ such that

$$\sum_{i \in I_j} |\langle x, f_i \rangle|^2 \leq K - \varepsilon$$

for all $j \in [r]$ and all unit vectors $x \in \mathbb{E}^d$.

This formulation suggests that an understanding of spaces of frames may provide an avenue of attack that could bring down the Kadison-Singer problem. Unfortunately, the geometry that we have uncovered does not have an immediate relationship with the behavior of a frame's subsets. The primary obstacle to sovling the Kadison-Singer problem from Weaver's perspective is the fact that choosing a partition that optimizes ε is most certainly an NP-hard combinatorial optimization problem. Thus, finding some way to quantify ε is a daunting challenge.

One potential approach that seems promising, if formidable, is to determine how well "good frames" cover spaces of frames. That is, we consider maps from

$$\mathcal{F}_{\mathbb{E}}(\mu, CI_{d\times d} - \varepsilon) \oplus \mathcal{F}_{\mathbb{E}}(\nu, CI_{d\times d} + \varepsilon) \longrightarrow \mathcal{F}_{\mathbb{E}}(\mu \oplus \nu, 2CI_{d\times d})$$

of the form

$$(F_1, F_2) \mapsto [F_1\ F_2].$$

Here, ε is a HPD perturbation matrix. Of course, we can also consider factorizing these maps through a permutation of the columns. One would then like to quantify the range of ε that ensures that this system of permuted embeddings covers the target space. Computing this quantity most likely requires an estimate of the volume of a general $\mathcal{F}_{\mathbb{E}}(\mu, S)$ variety, which is currently unavailable. As such, we leave this as an open problem.

Problem 5.2.3. *Find lower and upper bounds on the volume of $\mathcal{F}_{\mathbb{E}}(\mu, S)$.*

Appendix: Framelab

This appendix contains all of the code for Framelab 1.0, a Matlab implementation of the optimization methods detailed in Chapter 4. The code listed here includes

- *minFOD.m* - the implementation of the frame operator distance minimization techniques from Section 4.1;
- *projTFEmuS.m* - the implementation of the orthogonal projection methods discussed in Section 4.2.1;
- *minFEmuS.m* - the implementation of the approximate gradient descent procedure developed in Section 4.2;
- *initFEmuS.m* - the implementation of the initial construction method discussed in Section 4.3;
- *chkMAJ.m* - an algorithm for checking the majorization conditions.

minFOD.m

```
% Title: Minimization of the Frame Operator Distance (FOD)

% Author: Nate Strawn

% Description: Performs a fast, sphere by sphere minimization of the

% FOD.

% Last Revision: 8.18.2010
```

```
% Inputs:
%   F: d by N, synthesis operator, d<=N.
%   mu: N by 1, list of lengths of the columns of F.
%   S: d by d, the target Frame Operator.
%   parameters: struct,
%       .tol: numerical tolerance and threshold parameter;
%       .maxITER: maximum number of iterations for the minimization;
%       .type: either 'full' or 'partial' depending upon the descent
% type.

% Outputs:
%   F: d by N, a numerical approximation to a (mu,S)-frame.
%   info: struct,
%       .ITER: count of the iterations;
%       .hist: list of all the iterates;
%       .histFOD: list of the FOD values for each iterate.

% Auxiliary variables:
%   ITER: integer, number of iterations so far.
%   grad: d by N, gradient of FOD^2 at F.
%   gradNorms: 1 by N, column norms of grad.
%   g: d by 1, the partial search direction.
%   G: d by N, the full search direction.
```

```
%   maxEntry: scalar, largest value of gradNorms.

%   maxIndex: integer, index of maxEntry in gradNorms.

%   T: d by d, dynamic variable used to represent S-FF'.

%   FOD: scalar, ||S-FF^\ast||.

%   A,B: scalar, auxilary coefficients for determining the step size.

%   t: scalar, the step size for the search.


function [F,info] = minFOD(F,mu,S,parameters)


if (nargin==3)

    parameters.tol=1e-6;

    parameters.maxITER=1000;

    parameters.type='full';

end


% Ensure that the column norms correspond to the entries of \mu

F=F*diag(sqrt(sum(abs(F).^2)).^(-1))*diag(mu);


% Initialize the frame operator difference

T = S-F*F';

T = (T+T')/2;


% Calculate initial FOD to S
```

```
FOD=norm(T,'fro');


% Intialize the information structure

ITER=1;


if (nargout == 2)

    info.hist=cell(1,parameters.maxITER);

    info.histFOD=zeros(1,parameters.maxITER);

    info.hist{ITER}=F;

    info.histFOD(ITER)=FOD;

end


% Minimization Step

if (strcmp(parameters.type,'partial'))

    while (parameters.tol<FOD && ITER<parameters.maxITER)

        % Calculate the gradient of the square FOD at F

        grad = -T*F;

        grad = grad ...

- F*diag(real(sum(conj(F).*grad)))*diag(mu.^(-2));

        gradNorms = sum(abs(grad).^2);


        % Find the entry of largest gain

        [maxEntry, maxIndex] = max(gradNorms);
```

```
        f=F(:,maxIndex);
        g=mu(maxIndex)*grad(:,maxIndex)/sqrt(maxEntry);
        T=T+f*f';

        % Compute auxilary coefficients
        A = (real(f'*T*f-g'*T*g))^2;
        B = 4*real(f'*T*g)^2;

        % Calculate the minimizing value of t
        t = -sqrt(1/2 - 1/2*sqrt(A/(A+B)));

        % Update F, the frame operator difference, and the FOD
        F(:,maxIndex) = sqrt(1-t^2)*f+t*g;
        F(:,maxIndex) = ...
mu(maxIndex)*F(:,maxIndex)/norm(F(:,maxIndex));
        T=T-F(:,maxIndex)*F(:,maxIndex)';
        FOD=norm(T,'fro');

        % Update the gradient
        ITER=ITER+1;
        if (nargout == 2)
            info.ITER=ITER;
            info.hist{info.ITER}=F;
```

```
            info.histFOD(info.ITER)=FOD;
        end
    end
end

if (strcmp(parameters.type,'full'))
    d=size(S,1);

    % If the condition number of S is sufficiently large, the spec-
    % tral gap is computed by applying a Rayleigh quotient iteration.
    % Otherwise, the spectral gap is assumed to be zero. We compute
    % the appropriate stepsize.
    if (parameters.tol <= cond(S)-1)
        t=1/(topeig(S,randn(d,1),parameters.tol)...
            +topeig(-S,randn(d,1),parameters.tol)+2*sum(mu.^2));
    else
        t=1/(2*sum(mu.^2));
    end

    % Compute the modified step sizes.
    t=t./mu;
    while (parameters.tol<FOD && ITER<parameters.maxITER)
        % Calculate the gradient of the square FOD at F
```

```
  grad = -T*F;
  grad = grad ...
- F*diag(real(sum(conj(F).*grad)))*diag(mu.^(-2));
  gradNorms = sqrt(sum(abs(grad).^2));
  annihilated = (gradNorms < parameters.tol);

  % Compute the rotation coefficients.
  A = cos(gradNorms.*t);
  B = sin(gradNorms.*t);

  % Compute the full search direction.
  G = grad*diag(mu./(gradNorms+annihilated))...
*diag(1-annihilated);

  % Update F.
  F=F*diag(A)-G*diag(B);

  % Update T.
  T = S-F*F';
  T = (T+T')/2;
  FOD=norm(T,'fro');

  % Update the gradient
```

```
        ITER=ITER+1;

        if (nargout == 2)

            info.ITER=ITER;

            info.hist{info.ITER}=F;

            info.histFOD(info.ITER)=FOD;

        end

    end

end

end


% Rayleigh quotient iteration for the top eigenvalue

function a=topeig(S,x,tol)

    d=size(S,1);

    a=(x'*S*x)/(x'*x);

    A=S-a*eye(d);

    residual=Inf;

    while (tol <= residual && 1e-12 < rcond(A))

        x=A\x;

        x=x/norm(x);

        temp=x'*S*x;

        residual=abs(temp-a);

        a=temp;

        A=S-a*eye(d);
```

```
    end
end
```

projTFEmuS.m

```
% Title: (\mu,S)-frame variety tangent space projection
% Author: Nate Strawn
% Description: This function computes the projection of a matrix X
% onto the tangent space of a (\mu,S)-frame variety at a point F.
% Last Revision: 8.18.2010

% Inputs:
%   X: d by N, the matrix to be projected onto T_F \F(\mu,S).
%   F: d by N, the point on a warped Stiefel manifold.
%   E: string, the field of interest, either 'R' or 'C'.
%   mu: 1 by N, norms of the columns of F.
%   S: d by d, F*F'.
%   parameters: struct,
%       .tol: small tolerance parameter;
%       .method: either 'iterative' or 'direct' depending upon the
%            desired projection method.

% Outputs:
```

```
%   X: d by N, the projection of the input.

% Auxiliary variables:
%   Xnew: d by N, holds the update for computing the residual.
%   Q: d*N by d*N or 2*d*N by 2*d*N, projection onto tangent space of
%       the warped Stiefel manifold.
%   R: d*N by d*N or 2*d*N by 2*d*N, projection onto tangent space of
%       the (\mu,S)-frame variety.
%   V: d by d, holder for an orthonormal basis adapted to columns of
% F.
%   Y: d by N, holder for basis members used to compute Q and R.
%   B: struct, information concerning the 'good' basis extracted from
%  F.
%   charge: N by 1, stores the values distributed on the correlation
%       network by the current state of Y.
%   z: scalar, update constant used to zero-out charge an entry at a
%    time.
%   residual: scalar, the distance between X and Xnew determines when
% the alternating projection stops.

function X=projTFEmuS(X,F,E,mu,S,parameters)

if (nargin == 5)
```

```
    parameters.tol=10^(-6);
    parameters.method='direct';
end

[d,N]=size(F);
residual=Inf;

% Test to see how far S is from a multiple of the identity.  If it is
% close to being near the identity, then we perform alternating pro-
% jections. If it is not close, then we perform the full projection.
if (strcmp(parameters.method,'iterative')==1)
if (cond(S)-1<parameters.tol)
    while (residual >= parameters.tol)
        Xnew=X-(X*F'+F*X')*F/(2*S(1,1));
        Xnew=Xnew-F*diag(real(diag(F'*Xnew)))*diag(mu.^(-2));
        residual=norm(X-Xnew,'fro');
        X=Xnew;
    end
else
    % First, we construct the projection Q.
    [B,quit]=goodbasis(F,parameters.tol);
    if (strcmp(quit,'failure')==1)
        error('F is a singular point');
```

```
end

if (strcmp(E,'R')==1)

    % Construct Q

    Q=zeros(d*N,d*N);

    for i=B.A

        for j=setdiff(1:N,B.A)

            Y=zeros(d,N);

            Y(:,j)=F(:,i);

            Y(:,i)=-F(:,j);

            Q=updateprojection(Q,Y,E);

        end

    end

    for i=B.A(1:d-1)

        for j=B.A(find(B.A==i)+1:d)

            Y=zeros(d,N);

            Y(:,j)=F(:,i);

            Y(:,i)=-F(:,j);

            Q=updateprojection(Q,Y,E);

        end

    end


    % Perform alternating projections.

    while (residual >= parameters.tol)
```

```
        Xnew=Q*reshape(X,d*N,1);
        Xnew=reshape(Xnew,d,N);
        Xnew=Xnew-F*diag(real(diag(F'*Xnew)))*diag(mu.^(-2));
        residual=norm(X-Xnew,'fro');
        X=Xnew;
    end
elseif (strcmp(E,'C')==1)
    Q=zeros(2*d*N,2*d*N);
    for i=B.A
        for j=setdiff(1:N,B.A)
            Y=zeros(d,N);
            Y(:,j)=F(:,i);
            Y(:,i)=-F(:,j);
            Q=updateprojection(Q,Y,E);
            Y=zeros(d,N);
            Y(:,j)=sqrt(-1)*F(:,i);
            Y(:,i)=sqrt(-1)*F(:,j);
            Q=updateprojection(Q,Y,E);
        end
    end
    for i=B.A(1:d-1)
            Y=zeros(d,N);
            Y(:,i)=sqrt(-1)*F(:,i);
```

```
            Q=updateprojection(Q,Y,E);
        for j=B.A(find(B.A==i)+1:d)
            Y=zeros(d,N);
            Y(:,j)=F(:,i);
            Y(:,i)=-F(:,j);
            Q=updateprojection(Q,Y,E);
            Y=zeros(d,N);
            Y(:,j)=sqrt(-1)*F(:,i);
            Y(:,i)=sqrt(-1)*F(:,j);
            Q=updateprojection(Q,Y,E);
        end
    end
    Y=zeros(d,N);
    Y(:,B.A(d))=sqrt(-1)*F(:,B.A(d));
    Q=updateprojection(Q,Y,E);
    while (residual >= parameters.tol)
        Xnew=Q*[reshape(real(X),d*N,1);reshape(imag(X),d*N,1)];
  Xnew=reshape(Xnew(1:d*N),d,N)...
+sqrt(-1)*reshape(Xnew(d*N+1:2*d*N),d,N);
        Xnew=Xnew-F*diag(real(diag(F'*Xnew)))*diag(mu.^(-2));
        residual=norm(X-Xnew,'fro');
        X=Xnew;
    end
```

```
    end
end
elseif (strcmp(parameters.method,'direct')==1)
    B=goodbasis(F,parameters.tol);
    B.Ac=setdiff(1:N,B.A);
    B.inv=inv(F(:,B.A));

    %Compute the first part of the basis
    if (strcmp(E,'R')==1)
        R=zeros(d*N,d*N);
        for i=B.Ac
            % Compute a basis for the perp span of each f_i
            temp=[1 zeros(1,d-1)]'+F(:,i)/mu(i);
            temp=temp/norm(temp);
            V=eye(d)-2*temp*temp';
            for j=2:d
                Z=zeros(N,N);
                Z(B.A,i)=B.inv*V(:,j);
                Z=Z-Z';
                Y=F*Z;
                charge=sum(Y(:,B.A).*F(:,B.A)');
                Y=distribute(Y,F,B,charge);
                R=updateprojection(R,Y,E);
```

```
        end
    end


    %Construct the second part of the basis.
    for i=B.A(1:d-1)
        for j=B.A(find(B.A==i)+1:d)
            if (B.Adj(i,j) == 0)
                Y=zeros(d,N);
                Y(:,i)=F(:,j);
                Y(:,j)=-F(:,i);
                charge=zeros(d,1);
                charge(i)=F(:,i)'*F(:,j);
                charge(j)=-conj(charge(i));
                Y=distribute(Y,F,B,charge);
                R=updateprojection(R,Y,E);
            end
        end
    end
    X=reshape(R*reshape(X,d*N,1),d,N);
elseif (strcmp(E,'C')==1)
    R=zeros(2*d*N,2*d*N);
    for i=B.Ac
        % Compute a basis for the perp span of each f_i
```

```
            u=[1 zeros(1,d-1)]';

            fhat=F(:,i)/mu(i);

            ip=fhat(1);

            temp=fhat-ip*u;

            V=eye(d)+((u-conj(ip)*fhat)*temp'...

/(1-abs(ip)^2)-eye(d))*(u*u'+temp*temp'/(1-abs(ip)^2));

            Y=zeros(d,N);

            Y(:,i)=sqrt(-1)*F(:,i);

            R=updateprojection(R,Y,E);

            for j=2:d

                for xi=0:1

                Z=zeros(N,N);

                Z(B.A,i)=(sqrt(-1)^xi)*B.inv*V(:,j);

                Z=Z-Z';

                Y=F*Z;

                charge=real(sum(Y(:,B.A).*conj(F(:,B.A)))');

                Y=distribute(Y,F,B,charge);

                R=updateprojection(R,Y,E);

                end

            end

        end


        %Construct the second part of the basis.
```

```
for i=B.A(1:d-1)
            Y=zeros(d,N);
            Y(:,i)=sqrt(-1)*F(:,i);
            R=updateprojection(R,Y,E);
    for j=B.A(find(B.A==i)+1:d)
        if (B.Adj(i,j) == 0)
            for xi=0:1
            Y=zeros(d,N);
            Y(:,i)=(sqrt(-1)^xi)*F(:,j);
            Y(:,j)=-(sqrt(-1)^(-xi))*F(:,i);
            charge=zeros(d,1);
            charge(i)=real(F(:,i)*F(:,j));
            charge(j)=-real(conj(charge(i)));
            Y=distribute(Y,F,B,charge);
            R=updateprojection(R,Y,E);
            end
        end
    end
end


Y=zeros(d,N);
Y(:,B.A(d))=sqrt(-1)*F(:,B.A(d));
R=updateprojection(R,Y,E);
```

```
        X=[reshape(real(X),d*N,1); reshape(imag(X),d*N,1)];
        X=R*X;
        X=reshape(X(1:d*N),d,N)+sqrt(-1)*reshape(X(d*N+1:2*d*N),d,N);
    end
end
end


% This function performs updates to a projection P given a new matrix
% P.
function P=updateprojection(P,Y,field)
    [d,N]=size(Y);
    if (strcmp(field,'R')==1)
        Y=reshape(Y,d*N,1);
        Y=Y-P*Y;
        Y=Y/norm(Y);
        P=P+Y*Y';
    elseif (strcmp(field,'C')==1)
        Y=[reshape(real(Y),d*N,1);
reshape(imag(Y),d*N,1)];
        Y=Y-P*Y;
        Y=Y/norm(Y);
        P=P+Y*Y';
    else
```

```
        error('Field must be R or C');
    end
end

% This function completes the matrix Y so that it is in the tanget
% space of the (\mu,S)-variety at F.
function Y=distribute(Y,F,B,charge)
    for k=1:B.gen
        for l=B.nu{k}
            leaf_index=find(B.A==l);
            parent_index=find(B.A==B.alpha(leaf_index));
            z=charge(leaf_index)/(F(:,parent_index)'*F(:,l));
            Y(:,l)=Y(:,l)-conj(z)*F(:,parent_index);
            Y(:,parent_index)=Y(:,parent_index)+z*F(:,l);
            charge(parent_index)=charge(parent_index)...
  +charge(leaf_index);
            charge(leaf_index)=0;
        end
    end
end

% This function extracts the column indices of a 'good' basis from
% the matrix F. By 'good', we mean that the basis has a connected
```

```
% correlation network. The function also returns pertinant informa-
% tion about the connectivity of the basis.
% Last Revision: 7.13.2010

% Inputs:
%   F: d by N, a frame matrix.
%   thresh: scalar, the threshold used to determine the adjacency
%  matrix.

% Outputs:
%   B: struct, all the info about the 'good' basis;
%       .A: 1 by d, the list of indices;
%       .alpha: 1 by d, the parent of each index;
%       .nu: cell, holds the different generations of the tree on
%  B.A;
%       .Adj: N by N, the adjacency matrix of the correlation network
%  of F.
%   quit: string, termination indicator.

% Auxiliary variables:
%   mu: 1 by N, the vector of lengths of columns of F.
%   P: d by d, the projection onto the span of the basis elements so
%  far.
```

```
function [B,quit]=goodbasis(F,thresh)
[d,N]=size(F);

% Compute the thresholded adjacency matrix
B.Adj=F'*F;
mu=sqrt(diag(B.Adj));
B.Adj=(abs(diag(mu.^(-1))*(B.Adj-diag(mu.^2))...
*diag(mu.^(-1)))>thresh);

% Attempt to extract a nonorthodecomposable basis from F
B.A=1;
P=F(:,1)*F(:,1)'/mu(1)^2;
quit='false';
while (strcmp(quit,'false') == 1)
    nbrs=neighbors(B.A,1:N,B.Adj);
    if (size(nbrs,1)==0)
        quit='failure';
    else
        [junk,best]=min(sum(abs(P*F(:,nbrs)...
      *diag(mu(nbrs).^(-1)).^2)));
        B.A=union(B.A,nbrs(best));
    end
```

```
    if (size(B.A,2)==d)
        quit='success';
    end
end

% Having successfully extracted the indices of a basis, we turn to
% constructing connectivity structures from the correlation network.
    if (strcmp(quit,'success')==1)
        %Count how many generations are in the tree rooted at 1
        B.gen=0;
        nbrs=1;
        while (size(nbrs,2) ~= d)
            nbrs=union(nbrs,neighbors(nbrs,B.A,B.Adj));
            B.gen=B.gen+1;
        end

        % Now construct nu
        B.nu=cell(1,B.gen);
        nbrs=1;
        B.alpha=zeros(1,d);
        for i=fliplr(1:B.gen)
            B.nu{i}=neighbors(nbrs,B.A,B.Adj);
            for j=1:size(B.nu{i},2)
```

```
                temp=neighbors(B.nu{i}(j),B.A,B.Adj);

                B.alpha(logical(B.A==B.nu{i}(j)))=temp(1);

            end

            nbrs=union(nbrs,B.nu{i});

        end

    end

end


% This function computes the neighbors of active in domain\active

% given the adjacency matrix.


function n=neighbors(active,domain,adjacencyMTX)

    D=setdiff(domain,active);

    [n,junk]=find(adjacencyMTX(D,active)==1);

    n=unique(D(n));

end
```

minFEmuS.m

```
% Title: Minimization of an objective function over a (\mu,S)-frame

%  variety

% Author: Nate Strawn

% Description: This function finds a local minimum of an objective
```

```
% function on a (\mu,S)-frame variety.
% Last Revision: 8.18.2010


% Inputs:
%   objfcn: function handle, the objective function to be minimized;
%       returns a value as well as a gradient.
%   F: d by N, the initial guess matrix, and the holder for the
% matrices found in resulting search, vectors are columns.
%   E: string, either 'R' or 'C' depending upon whether real or
%   complex frames are desired.
%   mu: 1 by N, the lengths of the columns of F.
%   S: d by d, the frame operator of F, S=F*F'.
%   parameters: struct,
%       .tol: tolerance used to decide termination of the
% optimization, default is 1e-6;
%       .maxITER: maximum number of iterations allowed, default is
% 1000;
%       .method: either 'iterative' or 'direct' depending upon the
%           projection strategy employed, default is 'iterative'.


% Outputs:
%   F: d by N, the computed matrix with vectors in the columns.
%   info: struct,
```

```
%       .ITER: the current iteration;
%       .hist: cell, the list of iterates;
%       .histfcn: the list of values of objfcn applied to the
% iterates.
%       .histres: the distances between each pair of successive
% iterates.

% Auxiliary variables:
%   Y: d by N, the test matrix for the golden section line search.
%   C,R: scalars, values used in the golden section search.
%   residual: scalar, the distance between successive iterates.
%       the search ends when this is below parameters.tol.
%   value: scalar, the value of objfcn at X.
%   grad: d by N, the gradient of objfcn at X.
%   t: 1 by 4, array of bracket values.
%   k: integer, the exponential factor for the initial bracket in the
%       golden section search.
%   value1,value2: scalars, the values of objfcn inside the bracket.
%   tau: scalar, time elapsed for each iteration.

function [F,info]=minFEmuS(objfcn,F,E,mu,S,parameters)
if (nargin == 5)
    parameters.tol=1e-9;
```

```
    parameters.maxITER=1000;

    parameters.method='direct';

    parameters.type='full';

end

% Bring F close to the (\mu,S)-frame variety

F=minFOD(F,mu,S,parameters);

% Initialize the information structure

if (nargout == 2)

    info.histfcn=zeros(1,parameters.maxITER);

    info.hist=cell(1,parameters.maxITER);

    info.histres=zeros(1,parameters.maxITER);

    info.ITER=1;

    info.hist{info.ITER}=F;

    info.histfcn(info.ITER)=objfcn(F);

end


% Initialize the residual

ITER=0;

residual.F=Inf;

residual.grad=0;

disp(sprintf('\n    ...

  Iteration\t Value\t\t Residual\t Gradient\t Time'));

disp(sprintf('====================================================...
```

```
=====================’));
while (residual.F > parameters.tol && ITER<parameters.maxITER)

% Apply the golden section search
    tau=cputime;
    [F,residual]=muSlinesearch(objfcn,F,E,mu,S,parameters);
    tau=cputime-tau;
    ITER=ITER+1;

% Store data
    if (nargout == 2)
        info.histres(ITER)=residual.F;
        info.hist{ITER}=F;
        info.histfcn(ITER)=objfcn(F);
    end

% Display the iteration data
    disp(sprintf('\t%d\t %4.5f\t %4.5f\t %4.5f\t %4.5f',...
        ITER, objfcn(F), residual.F, residual.grad, tau));
end

if (nargout == 2)
    info.ITER=ITER;
```

```
end

end


function [F,residual]=muSlinesearch(objfcn,F,E,mu,S,parameters)

% Define the values for the Golden Section search

    C = (3-sqrt(5))/2;

    R = 1-C;


% Initialize the bracket variables

    k=0;

    t=parameters.tol;


% Evaluate the objective function at F

    [value,grad]=objfcn(F);


% Project the gradient onto the tangent space of the (mu,S)-frame

% variety

    grad=projTFEmuS(grad,F,E,mu,S,parameters);

    residual.grad=norm(grad,'fro');


% Compute the column norms of the gradient and replace the gradient

% with it's column-ameliorated version.

    gradnorms=sqrt(sum(grad.^2));
```

```
    grad=grad*diag((gradnorms.*(gradnorms >= parameters.tol)...
        +(gradnorms < parameters.tol)).^(-1))*diag(mu);

% Compute the first point in the line search along the product of
% spheres geodesic.
    Y=F*diag(cos(gradnorms*t))-grad*diag(sin(gradnorms*t));

% Once the point along the geodesic is computed, we use minFOD to
% bring it back to the (mu,S)-frame variety, and evaluate the object-
% ive function at this point.
    Y=minFOD(Y,mu,S,parameters);
    value1=objfcn(Y);

% Initialize the bracket: first consider the step at (2^0)*tol.  If
% this is lower than at t=0, then consider steps inductively at
% (2^k)*tol. Once the process stops at k+1, the bracket is
% 0,tol*2^k,tol*2^(k+1).  If this value is higher, then we back track
% until we get a lower value.
    if (value1<value)
    % If value1 is below value, we proceed to construct the initial
    % bracket.
        while (value1<value)
            k=k+1;
```

```
        t=(2^k)*parameters.tol;
        Y=F*diag(cos(gradnorms*t))-grad*diag(sin(gradnorms*t));
        Y=minFOD(Y,mu,S,parameters);
        % value=value1 ensures that this doesn't go off to
% infinity
        % value=value1;
        value1=objfcn(Y);
    end
    t=[0 (2^(k-1))*parameters.tol (2^k)*parameters.tol];

% We now apply the golden section search technique.
    if (abs(t(3)-t(2)) > abs(t(2)-t(1)))
        t=[t(1) t(2) t(2)+C*(t(3)-t(2)) t(3)];
    else
        t=[t(1) t(2)-C*(t(2)-t(1)) t(2) t(3)];
    end
    Y=F*diag(cos(gradnorms*t(2)))-grad*diag(sin(gradnorms*t(2)));
    Y=minFOD(Y,mu,S,parameters);
    value1=objfcn(Y);
    Y=F*diag(cos(gradnorms*t(3)))-grad*diag(sin(gradnorms*t(3)));
    Y=minFOD(Y,mu,S,parameters);
    value2=objfcn(Y);
    k = 1;
```

```
    while (abs(t(4)-t(2)) > parameters.tol*(t(2)+t(3)) &&...
            t(4)>parameters.tol^16),
        if (value2<value1)
            t(1) = t(2);
            t(2) = t(3);
            t(3) = R*t(2) + C*t(4);
            value1 = value2;
            Y=F*diag(cos(gradnorms*t(3)))...
-grad*diag(sin(gradnorms*t(3)));
            Y=minFOD(Y,mu,S,parameters);
            value2=objfcn(Y);
        else
            t(4) = t(3);
            t(3) = t(2);
            t(2) = R*t(3) + C*t(1);
            value2 = value1;
            Y=F*diag(cos(gradnorms*t(2)))...
-grad*diag(sin(gradnorms*t(2)));
            Y=minFOD(Y,mu,S,parameters);
            value1=objfcn(Y);
        end
        k = k+1;
    end
```

```
        if (value1<value2)

            Y=F*diag(cos(gradnorms*t(2)))...

      -grad*diag(sin(gradnorms*t(2)));

            Y=minFOD(Y,mu,S,parameters);

        else

            Y=F*diag(cos(gradnorms*t(3)))...

      -grad*diag(sin(gradnorms*t(3)));

            Y=minFOD(Y,mu,S,parameters);

        end

    % Compute the residual and update F

        residual.F=norm(Y-F,'fro');

        F=Y;

    else


    % The initial point of the line search yields a point which is

    % close, but which has larger value.  Thus, the search ends.

        residual.F=0;

        F=Y;

    end

end
```

initFEmuS.m

```
% Title: Initialize (mu,S)-frame
% Author: Nate Strawn
% Description: Produces a member of a (mu,S)-frame with the goal of
% keeping many columns of F the same.
% Last Revision: 8.18.2010

% Inputs:
%   F: d by N, original frame input.
%   mu: 1 by d, list of vector lengths.
%   S: d by d, target frame operator.

% Outputs:
%   F: d by N, the (mu,S)-frame.

% Auxiliary variables:
%   U: d by d, the list of eigenvectors of S.
%   lambda: 1 by d, the list of eigenvalues of S.

function F=initFEmuS(F,mu,S)
N=size(mu,2);
[U,lambda]=eig(S);
```

```
U=fliplr(U);
lambda=fliplr(diag(lambda)');
A=[];
for i=1:N
    f=U'*F(:,i);
    if(chkMAJ(mu(setdiff(1:N,A)),lambda,f,find(setdiff(1:N,A)==i)))
        A=setunion(A,i);
        S=S-f*f';
        [U,lambda]=eig(S);
        U=fliplr(U);
        lambda=fliplr(diag(lambda)');
    end
end
F(:,setdiff(1:N,A))=minFOD(F(:,setdiff(1:N,A)),mu,S);
```

chkMAJ.m

```
% Title: Check majorization conditions
% Author: Nate Strawn
% Description: Either checks that majorization holds for two sequen-
% ces, or checks sufficient conditions for majorization after removal
% of a rank one operator.
% Last Revision: 8.18.2010
```

```
% Inputs:
%   mu: 1 by N, unordered list of column lengths.
%   lambda: 1 by d, ascending list of eigenvalues for a target
% operator.
%   f: d by 1, candidate vector to be checked
%   index: integer, index of the candidate vector.

% Outputs:
%   b: logical, 0 if the condition fails, 1 if the condition holds

% Auxiliary variables:
%   k: integer, an index counter.
%   mu_sum, lambda_sum, f_sum, sqrtSf_sum, c: scalar, holders for the
%       quantities updated through the loop.

function b=chkMAJ(mu,lambda,f,index)
b=1;
d=size(lambda,2);
if (nargin == 2)
    if (abs(sum(mu.^2)-sum(lambda)) > 1e-6)
        b=0;
    end
```

```
    if (b==1)
        mu=sort(mu.^2,'descend');
        k=1;
        mu_sum=0;
        lambda_sum=0;
        while (b == 1 && k < d)
            mu_sum=mu_sum+mu(k);
            lambda_sum=lambda_sum+lambda(k);
            if (mu_sum > lambda_sum)
                b=0;
            end
            k=k+1;
        end
    end
else
    if ((f'./lambda)*f > 1)
       b=0;
    end
    if (b == 1)
        N=size(mu,2);
        mu=sort(mu([1:index-1 index+1:N]).^2,'descend');
        k=1;
        mu_sum=0;
```

```
        lambda_sum=0;
        f_sum=0;
        sqrtSf_sum=0;
        while(b == 1 && k < d)
            mu_sum=mu_sum+mu(k);
            lambda_sum=lambda_sum+lambda(k);
            f_sum=f_sum+abs(f(k))^2;
            sqrtSf_sum=sqrtSf_sum+lambda(k)*abs(f(k))^2;
            c=lambda_sum-mu_sum;
            if (f_sum > c)
                rho=sqrtSf_sum/f_sum;
                if (f_sum < 1e-9 || rho > lambda(k+1)+c)
                    if (c < 1e-9 ||...
                        f_sum/c+(f(k+1:d)'./(lambda(k+1:d)+c-rho))...
  *f(k+1:d)>1)
                        b=0;
                    end
                end
            end
            k=k+1;
        end
    end
end
```

Bibliography

[1] P. A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds.* Princeton, NJ: Princeton University Press, 2008.

[2] T. Arias, A. Edelman, and S. Smith. Curvature in Conjugate Gradient Eigenvalue Computation with Applications to Materials and Chemistry Calculations. SIAM Applied Linear Algebra Conference, Philadelphia, PA:233–238, 1994.

[3] J. J. Benedetto and M. Fickus. Finite Normalized Tight Frames. *Advances in Computational Mathematics*, 18(2–4):357–385, 2003.

[4] J.J. Benedetto and O. Oktay. PCM - Sigmadelta comparison and sparse representation quantization. *Proceedings of the Conference on Information Sciences and Systems*, Princeton, NJ, 737–742, 2008.

[5] J. J. Benedetto, O. Oktay, and A. Tangboondouangjit. Complex sigma-delta quantization algorithms for finite frames. *Radon transforms, geometry, and wavelets*, Contemporary Mathematics, 464:27–49. Providence, RI: American Mathematical Society, 2008.

[6] J. J Benedetto, A. M. Powell, and O. Yılmaz. Second-order sigma-delta ($\Sigma\Delta$) quantization of finite frame expansions. *Applied and Computational Harmonic Analysis*, 20(1):126–148, 2006.

[7] J. J. Benedetto, A. M. Powell, and O. Yılmaz. Sigma-Delta ($\Sigma\Delta$) quantization and finite frames. *IEEE Transactions on Information Theory*, 52(5):1990–2005, 2006.

[8] B. G. Bodmann, V. I. Paulsen, and M. Tomforde Equiangular tight frames from complex Seidel matrices containing cube roots of unity. *Linear Algebra and its Applications*, 430(1):396–417, 2009.

[9] G. E. Bredon. *Topology and geometry.* New York, NY: Springer-Verlag, 1997.

[10] J. R. Bunch, C. P. Nielsen, and D. C. Sorensen. Rank-one modification of the symmetric eigenproblem. *Numerische Mathematik*, 31(1):31–48, 1978.

[11] E. J. Candès and D. L. Donoho. Curvelets and reconstruction of images from noisy Radon data. *Wavelet Applications in Signal and Image Processing VIII*, A. Aldroubi, A. F. Laine, M. A. Unser eds., Proc. SPIE 4119, 2000.

[12] P. G. Casazza and M. Fickus. Gradient descent of the frame potential. *Proceedings of Sampling Theory and Applications*, 21:83–112, 2009.

[13] P. G. Casazza and J. Kovacevic. Equal-Norm Tight Frames with Erasures. Advances in Computational Mathematics, 18(2–4):387–430, 2003.

[14] P. G. Casazza and M. T. Leon. Existence and Construction of Finite Frames with a Given Frame Operator. *International Journal of Pure and Applied Mathematics*, 2010.

[15] S. S. Chen. *Basis Pursuit*, Stanford, CA: Stanford University, 1995.

[16] I. Daubechies, A. Grossmann, and Y. Meyer. Painless nonorthogonal expansions. *Journal of Mathematical Physics*, 27(5):1271–1283, 1986.

[17] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 41(7):909–996, 1988.

[18] I. Daubechies. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, 1992.

[19] I. S. Dhillon, R. W. Heath Jr., M. Sustik, and J. A. Tropp. Generalized finite algorithms for constructing Hermitian matrices with prescribed diagonal and spectrum. *SIAM Journal of Matrix Analysis and Applications*, 27(1):61–71, 2005.

[20] D. L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization. *Proceedings of the National Academy of Science*. 100(5): 21972202, 2003.

[21] D. L. Donoho, M. Elad, V. N. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Transactions on Information Theory*, 52(1):6–18, 2006.

[22] R. J. Duffin and A. C. Schaeffer. A class of nonharmonic Fourier series. *Transactions of the American Mathematical Society*. 72:341–366, 1952.

[23] K. Dykema and N. Strawn. Manifold structure of spaces of spherical tight frames. *International Journal of Pure and Applied Mathematics*, 28(2):217-256, 2006.

[24] V. K. Goyal, M. Vetterli, and N. T. Thao. Quantized overcomplete expansions in $\mathbb{R}^N$: analysis, synthesis, and algorithms. *IEEE Transactions on Information Theory*, 44(1):16–31, 1998.

[25] D. Han and D. R. Larson Frames, bases and group representations. *Memoirs of the American Mathematical Society*, 147(697), 2000

[26] R. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge, UK: Cambridge University Press, 1985.

[27] I. M. James. *The Topology of Stiefel Manifolds.* Cabridge, UK: Cambridge University Press, 1976.

[28] R. Kadison and I. Singer. Extensions of pure states. *American Journal of Mathematics*, 81:547–564, 1959.

[29] K. A. Kornelson and D. R. Larson. Rank-one decomposition of operators and construction of frames. *Wavelets, frames and operator theory*, Contemporary Mathematics, 345:203–214. Providence, RI: American Mathematical Society, 2004.

[30] S. G. Krantz and H. R. Parks. *The implicit function theorem: History, theory, and applications.* Boston, MA: Birkhäuser, 2002.

[31] D. Labate, W. Lim, G. Kutyniok, and G. Weiss. Sparse multidimensional representation using shearlets. Wavelets XI (San Diego, CA, 2005), 254–262, SPIE Proc. 5914, SPIE, Bellingham, WA, 2005.

[32] Z. Q. Luo and P. Tseng. On the convergence of the coordinate descent method for convex differentiable minimization. *Journal of Optimization Theory and Applications*, 72(1):7–35, 1992.

[33] S. G. Mallat and Z. Zhang. Matching Pursuits with Time-Frequency Dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, 1993.

[34] O. Oktay. *Frame Quantization Theory and Equiangular Tight Frames.* College Park, MD: University of Maryland, 2007.

[35] A. M. Ostrowski. On the convergence of the Rayleigh quotient iteration for the computation of the characteristic roots and vectors, III. *Archive for Rational Mechanics and Analysis*, 3(1):325–340, 1959.

[36] , V. Pan and D. Bini. *Polynomial and matrix computations.* Basel, Switzerland: Birkäuser, 1994.

[37] M. Püschel and J. Kovacevic. Real, Tight Frames with Maximal Robustness to Erasures. *IEEE Data Compression Conference*, 63–72, 2005.

[38] S. Smith. *Geometric optimization methods for adaptive filtering.* Cambridge, MA: Harvard University, 1993.

[39] M. Spivak. *A Comprehensive Introduction to Differential Geometry I*, Houston, TX: Publish or Perish Inc., 1999.

[40] N. Strawn. *Geometry and Constructions of Finite Frames*, College Station, TX: Texas A&M University, 2007.

[41] T. Strohmer. A note on equiangular tight frames. *Linear Algebra and its Applications*, 429(1):326–330, 2008.

[42] T. Strohmer and R. W. Heath Jr.. Grassmannian frames with applications to coding and communication. *Applied and Computational Harmonic Analysis*, 14(3):257–275, 2003.

[43] M. Sustik and J. A. Tropp and I. S. Dhillon and R. W. Heath Jr.. On the existence of equiangular tight frames. *Linear Algebra and its Applications*, 426(2-3):619–635, 2007.

[44] J. A. Tropp Greed is good: Algorithmic results for sparse approximation. *IEEE Transactions on Information Theory*, 50(10):2231-2242,2004.

[45] J. A. Tropp, I. S. Dhillon, and R. W. Heath Jr.. Finite-step algorithms for constructing optimal CDMA signature sequences. *IEEE Transactions on Information Theory*, 50(11):2916–2921, 2004.

[46] J. A. Tropp, I. S. Dhillon, R. W. Heath Jr. and T. Strohmer. Designing structured tight frames via an alternating projection method. *IEEE Transactions on Information Theory*, 51(1):188–209, 2005.

[47] P. Tseng and S. Yun. A coordinate gradient descent method for nonsmooth separable minimization. *Journal of Mathematical Programming*, 117(1-2):387–423, 2009.

[48] S. Waldron. Generalized Welch bound equality sequences are tight frames. *IEEE Transactions on Information Theory.* 49(9):2307–2309, 2003.

[49] N. Weaver. The Kadison-Singer Problem in discrepancy theory. *Discrete Mathematics*. 278:227–239, 2004.

[50] R. M. Young. *An introduction to nonharmonic Fourier series.* San Diego, CA: Academic Press Inc., 2001.