Recursive Bayesian Estimation on Projective Spaces:

Theoretical Foundations and Practical Algorithms

by

Lauren N. Crider

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved September 2021 by the
Graduate Supervisory Committee:

Douglas Cochran, Chair
Brett Kotschwar
Louis Scharf
Thomas Taylor
Pavan Turaga

ARIZONA STATE UNIVERSITY

December 2021

ABSTRACT

This thesis develops geometrically and statistically rigorous foundations for multivariate analysis and bayesian inference posed on grassmannian manifolds. Requisite to the development of key elements of statistical theory in a geometric realm are closed-form, analytic expressions for many differential geometric objects, e.g., tangent vectors, metrics, geodesics, volume forms. The first part of this thesis is devoted to a mathematical exposition of these. In particular, it leverages the classical work of Alan James to derive the exterior calculus of differential forms on special grassmannians for invariant measures with respect to which integration is permissible.

Motivated by various multi-sensor remote sensing applications, the second part of this thesis describes the problem of recursively estimating the state of a dynamical system propagating on the Grassmann manifold. Fundamental to the bayesian treatment of this problem is the choice of a suitable probability distribution to *a priori* model the state. Using the Method of Maximum Entropy, a derivation of maximum-entropy probability distributions on the state space that uses the developed geometric theory is characterized. Statistical analyses of these distributions, including parameter estimation, are also presented. These probability distributions and the statistical analysis thereof are original contributions.

Using the bayesian framework, two recursive estimation algorithms, both of which rely on noisy measurements on (special cases of) the Grassmann manifold, are the devised and implemented numerically. The first is applied to an idealized scenario, the second to a more practically motivated scenario. The novelty of both of these algorithms lies in the use of the derived maximum-entropy probability measures as models for the priors. Numerical simulations demonstrate that, under mild assumptions, both estimation algorithms produce accurate and statistically meaningful outputs.

This thesis aims to chart the interface between differential geometry and statistical signal processing. It is my deepest hope that the geometric-statistical approach underlying this work facilitates and encourages the development of new theories and new computa-

tional methods in geometry. Application of these, in turn, will bring new insights and better solutions to a number of extant and emerging problems in signal processing.

ACKNOWLEDGMENTS

The fact that this thesis exists is due in no small part to the selfless support I have received from so many people over so many years. While I have no doubt the words written on this page will hardly do justice to my sense of admiration and gratitude, I would like to acknowledge those whose support, encouragement, and patient conversations are responsible for the best of what is here.

My advisor, Doug Cochran, deserves my deepest gratitude. It has been an extraordinary privilege to learn from and to work alongside him. I cannot imagine doing this – any of this – without his constant encouragement, his inexhaustible patience, his precise guidance, his unmatched sense of humor. Doug, thank you for being my greatest mentor, in every sense of the word.

Thank you also to the rest of my committee – Brett Kotschwar, Louis Scharf, Tom Taylor, and Pavan Turaga – for continually inspiring me, teaching me, and challenging me. You have each crafted the way I see mathematics and helped me to define my role in it. And to my tireless collaborators and informal committee members, Stephen Howard, Songsri Sirianunpiboon, and Bill Moran, thank you for your helpful discussions on various kinds of joint projects throughout the years.

So many of my fellow graduate students and cohorts also deserve thanks for making the School of Mathematical & Statistical Sciences so wonderful, but Joe Wells, Donnie Adams, Mary Cook, David Poletta, and Camille Moyer each merit special mention as inspirations, occasional commiserators, and true friends. Each of you, at different times and in different capacities, helped me along (maybe more than you realize), and I owe you a huge thank you. Of course to my favorite officemate Scott Jones, I thank you for the tremendously valuable time you spent in front of a board picking nits at my only half-baked, feeble arguments. And squarely last-but-not-least, the unwavering support of Rhonda Olson always helped to boost my not so rock-steady academic self-esteem. Thank you.

Finally, it is my family that has my ultimate love and gratitude. You are the fuel that

keeps me going.

***

The research presented in this thesis was generously supported by the National Science Foundation's Graduate Research Fellowship Program and by the United States Air Force Office of Scientific Research under grant number FA9550-18-1-0190.

TABLE OF CONTENTS

CHAPTER Page

APPENDIX Page

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

The focus of this dissertation is two-fold. First, it develops new mathematical theory for bayesian estimation and statistical analysis on real and complex Grassmann manifolds, where the geometrical aspects of the spaces are crucial considerations. Second (in fact, to some degree, in parallel), it devises and implements computational geometric-statistical algorithms that are applied to a specific class of extant and emerging problems in signal processing – namely the iterative estimation of particular subspaces of finite-dimensional vector spaces in dynamic scenarios.

1.1 Some Historical Remarks

It is said without great exaggeration that Grassmann invented linear algebra and, with none at all, that he showed how to properly apply it in geometry [9]. Though in a tragic tale, his visionary introduction to vectors, vector spaces, subspaces, bases, dimensions, exterior products, and a host of geometric algebra [16], was almost entirely ignored by contemporary mathematicians. Plücker introduced a special case of the concept of a space comprised of subspaces of a linear space [38], which was generalized by Grassmann to establish what only much later came to be known as a grassmannian. The definitions of vector spaces, their subspaces, and collections of their subspaces came into mathematics, in the sense of becoming widespread, around 1920, when Weyl's influential book [51] appeared.

Work that directly underpins the modern theory of manifolds was also emerging around the same time that Grassmann published his *Die Ausdehnungslehre*, perhaps most notably from Riemann who apparently coined the term *Mannigfaltigkeit* in his 1851 doctoral thesis [40]. The manifold nature of the grassmannian, along with its geometric and topological properties, structure as a homogeneous space and a quotient of Lie groups, construction as

a scheme, representation as an algebraic variety, and numerous other characteristics have been very well studied over the past century. Of particular importance to this thesis are the introduction of Pontryagin coordinates [39], which enable integration on grassmannians to be carried out in a single coordinate patch, and the pioneering work of James [21] in the early 1950s, which established foundations for statistics on grassmannians.

The early-to-mid twentieth century also saw the emergence of statistical estimation theory and prominent modern techniques. The method of maximum likelihood, first established by Fisher in 1912 [10], remains one of the most popular and useful estimation techniques of modern statistics; despite the bayesian emphasis in this thesis, we also draw on these maximum-likelihood techniques. Kolmogorov [28] in 1941 and Wiener [53] in 1942 independently developed a linear minimum mean-square estimation technique that received considerable attention and provided the foundation for the subsequent development of the celebrated Kalman filter [24]. The late 1950s brought significant attention to the development of bayesian statistics and the principle of maximum entropy, providing a context within which the work presented in thesis may be regarded as a direct extension of the Kalman filter.

Advances in sensor technologies and network sensor systems in the last quarter of the twentieth century and into the twenty-first century precipitated new interest in subspace-based linear models – and corresponding subspace-based methods – whose most natural settings are Grassmann manifolds. In particular, the past three decades have shown significant interest in applications in which a putative signal (that is known or unknown) of a fixed rank (that is also known or unknown!) is to be detected based on measurements collected from a single sensor, or possibly a suite of multiple sensors. Subspace-based linear models in signal analysis were initially popularized in array processing [29], and their more generic use in signal detection and estimation were pioneered by Scharf [41] and others in the 1980s. Notably, he and Friedlander established the *matched subspace detector*, which generalizes the inner product detection of the classical matched filter to a statistic obtained by projection of aggregated data into a higher dimensional signal subspace [42].

Complementing this work was the introduction by Gish and Cochran [14] of the *generalized coherence estimate*, which extended the popular two-channel magnitude-squared coherence statistic, based on an inner product, to multi-channel statistic based on a Gram matrix determinant. Both of these generalizations continue to see wide application.

The matched subspace detector, the generalized coherence estimate, and their generalizations and variants have clearly stated optimalities and invariances, as well as evocative geometrical interpretations. The matched filter – a special case of a matched subspace detector – admits interpretation as the length of the projection of aggregated sensor data onto the signal subspace. The generalized coherence estimate admits interpretation as the volume of a parallelotope or equivalently as the product of principal angles between subspaces of a vector space. In fact, many standard methods in statistical signal processing arose from multivariate statistical analysis and, though expressed in terms of matrix algebra, often entail geometric structure that is not explicitly elucidated.

This dissertation develops this geometric perspective, offering potential for broad new geometric insight and methodology to this same class of statistical signal processing problems.

1.2 A Motivating Application

This dissertation is focused on a specific class of subspace detection, estimation, and tracking problems, which has various multi-sensor remote sensing applications, including multistatic radar and electronic surveillance. Typical goals of these problems are to detect and characterize a putative unknown signal impinging on a network of distributed sensors. For concrete context, we consider the following practical application throughout this thesis.

Let us assume there are $k \geq 1$ linearly independent waveforms illuminating m spatially distributed sensors, which transduce impinging electromagnetic radiation (or acoustic sound pressure) to ultimately produce digital signals. The target typically manifests as a weighted sum of the waveforms, delayed and Doppler shifted according to the target location and motion, at each sensor. Because the k transmit waveforms are the same but

the weights (gains) are different at each receiver, they define a k-dimensional subspace of a vector space of dimension $n \gg k$, where n refers to the length of the sample sequences collected at each sensor for processing. As mentioned, the problem of detecting a common but unknown signal of known rank k using data collected at m spatially distributed sensors has been well studied in the signal processing literature [6] (and references therein). However due to time-varying illumination and the effects of presumed target motion, the k-dimensional subspace itself is time-varying. The problem is now to recursively estimate, or *track*, a sequence of k-dimensional subspaces from observations at the spatially distributed sensors. This task of detecting and estimating a sequence of subspaces of a time-varying system is what we refer to as *subspace tracking*.

We should note that estimating the *state* of a dynamical system is a well-studied problem in connection with numerous applications. When the *state space* is $\mathbb{R}^n$ and the dynamical and measurement models are linear with additive gaussian noise, the well-known and widely used Kalman filter provides an elegant solution that incorporates new measurements in each epoch into the state estimate. As the state of our dynamical system is an element of a Grassmann manifold, and we model the dynamics as rotations resulting from the action of an element of the (special) unitary group, our dynamical system is intrinsically nonlinear and the Kalman filter framework does not apply. bayesian analysis of stochastic dynamical systems on nonlinear manifolds entails development of suitable probability measures that play a role analogous to the gaussian distribution on $\mathbb{R}^n$, e.g. maximum-entropy probability measures. To our knowledge, the development of maximum-entropy probability measures on grassmannians is not known in the literature.

Recent work [47], which has served as a proverbial launching pad for the approach to iterative estimation put forth in this thesis, developed iterative estimation algorithms for the state of a dynamical system evolving both on $\mathbf{SO}(2)$ and $\mathbf{SO}(3)$, where the maximum-entropy distributions corresponding to a fixed first moment are the von Mises and the von Mises-Fisher matrix distributions, respectively. Their results demonstrate in simulations that their algorithms, based on bayesian recursion, appear to achieve an unprecedented

combination of fidelity and computational efficiency. The recursive estimators relying on an idealized measurement model that are set forth in Chapter 4 are direct extensions of the geometric-statistical approach in [47] to special cases of the grassmannian: real and complex projective space.

1.3 A Roadmap

The remainder of this thesis is organized as follows. Chapter 2 begins with a broad-based introduction to the mathematical, geometrical, statistical, and information theoretic foundations at play. While almost none of the material presented in this chapter is original, a goal of this thesis is to create an entirely self-contained, mathematically complete document accessible to members of both the differential geometry and statistical signal processing communities. And a sound understanding of the mathematical underpinnings will be essential to the ensuing chapters.

Chapters 3 and 4 comprise the bulk of the original work in this thesis. Chapter 3 develops the aspects of the theory of bayesian estimation of particular subspaces of finite-dimensional vector spaces in dynamic scenarios, specialized to the important, special case where the subspaces are one-dimensional. In particular, we construct differential forms for invariant measures on real and complex projective space, obtained from James' derivation of the differential form for the invariant measure on the Grassmann manifold [21]. It is in this chapter where we also evaluate the integrals of theses invariant measures, yielding normalized invariant measures – that is, *uniform distributions* – on $\mathbb{RP}^n$ and $\mathbb{CP}^n$. We further develop maximum-entropy probability distributions on $\mathbb{RP}^n$ and $\mathbb{CP}^n$ and include a brief statistical analysis, including the estimation of their parameters. These probability measures and the statistical analysis thereof are original contributions. Chapter 4 presents the application of the geometric-statistical theory, devising a bayesian framework for recursive estimation posed on real and complex projective space. For each of these spaces, an idealized scenario is considered first, followed by one more practically motivated.

Chapter 5 lays the groundwork for extension from complex projective space to the com-

plex Grassmann manifold. We obtain a few results in generality, but point out current obstructions for complete generalizations of the bayesian estimation techniques and statistical analyses presented in preceding chapters. While the primary application motivating this thesis is indeed the (recursive) estimation of particular time-varying subspaces of finite-dimensional vector spaces, e.g., points on a grassmannian, great care is taken so that the development of the differential geometric objects can be naturally applied to a variety of related problems in statistical signal processing, such as classification, coherence analysis, even manifold learning.

Finally, in Chapter 6, we conclude with a discussion of our contributions and propose opportunities to expand the scope of our geometric-statistical approach.

1.4 Some Notational Conventions

Scalar quantities are represented by regular, lowercase characters; vector quantities are boldface, lowercase characters; matrix quantities are boldface, uppercase characters. When the distinction is important, random quantities (variables, vectors, matrices) are in sans-serif font. Examples of scalar-valued random variables are $\mathsf{x}, \mathsf{y}, \alpha, \theta$, while $\mathbf{x}, \mathbf{y}, \boldsymbol{\alpha}, \boldsymbol{\theta}$ are vector-valued random variables. Examples of scalar sample values (and deterministic quantities) are x, y, α, θ, while examples of deterministic vector quantities are $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\theta}$.

While it is perhaps a bit nontraditional in the statistical signal processing community, we've attempted to adhere to denoting components of matrices and vectors with upper indices. For example, the i^{th} column of the matrix $\boldsymbol{X}$ is denoted $\boldsymbol{x}^i$; the $(i,j)^{\text{th}}$ element of the matrix $\boldsymbol{X}$ is denoted x^{ij}, or sometimes x^i_j.

Finally, boldface, calligraphic letters like $\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}}$ will denote the subspace spanned by the columns of the matrices $\boldsymbol{X}, \boldsymbol{Y}$.

Chapter 2

MATHEMATICAL PRELIMINARIES

The present chapter will sketch (without proofs) some of the principal concepts, results, and notational conventions of the core areas of mathematics on which this thesis is dependent: differential geometry, multivariate statistics, and bayesian estimation theory. While almost none of the material presented in this chapter is original, the hope is it will both encourage and facilitate the development of new geometric theories and computational methods fit for application to extant and emerging problems in statistical signal processing. Should the reader be interested in the deeper, perhaps more generalized, treatment of the geometric topics discussed in this chapter, some recommended reference texts include the exceptional and gentle introduction to differential geometry and smooth manifolds by Lee [32]; a more terse presentation of the same material by Spivak [46]; and Hall's leisurely introduction to matrix Lie groups and their Lie algebras [17]. For readers interested in a deeper introduction to statistical signal processing and signal detection and estimation theory, we recommend the elegant, theoretical presentation by Helstrom [19], an insightful geometric interpretation of much of the same material by Scharf [41], as well as the more pervasive texts by Kay [26, 27]. In all of what follows, additional references will be cited wherever possible.

2.1 Geometric Foundations

As our presentation of differential geometric constructs borrows from [32], notation here is mostly consistent with Lee's. Throughout this chapter, all our manifolds are assumed to be real[1], n-dimensional, smooth, Hausdorff, and second countable; and smooth always means C^∞, or infinitely differentiable. We write local coordinates of a manifold M on any open subset $U \subset M$ as $\left(x^1, \ldots, x^n\right)$. Since coordinates formally constitute a map

[1] We will touch on complex manifolds in a few places.

from U to $\mathbb{R}^n$, this is to say that we identify U with its image in $\mathbb{R}^n$ and we identify a point in U with its coordinate representation. We will denote the tangent space to a point p in M by T_p or T_pM, and the cotangent space to a point p in M by T_p^* or T_p^*M. The disjoint unions

$$\begin{aligned} TM &= \coprod_{p \in M} T_pM \\ T^*M &= \coprod_{p \in M} T_p^*M \end{aligned} \tag{2.1}$$

define the tangent and cotangent bundles of M, respectively. A vector field is a section of the tangent bundle; a covector field is a section of the cotangent bundle. We let $\mathfrak{X}(M)$ denote the set of all smooth vector fields on the manifold M. Smooth vector fields can be multiplied by smooth real-valued functions: if $f \in C^\infty(M)$ and $Y \in \mathfrak{X}(M)$, characterized by $p \mapsto Y_p$, then fY is a new vector field characterized by

$$p \mapsto (fY)_p = f(p)Y_p.$$

2.1.1 Tensors: Metrics and Differential Forms

In this section, we begin developing the technical machinery needed to measure geometric quantities – such as distances and angles – and to compute integrals over a manifold, which will be fundamental to the statistical analysis performed on particular manifolds presented in the latter half of this chapter.

For $k \in \mathbb{Z}^+$, we define a covariant k-tensor on T_pM to be an element of the k-fold tensor product $T_p^*M \otimes \cdots \otimes T_p^*M$, which we typically think of as a real-valued, multilinear function of k elements of T_pM:

$$\alpha : \underbrace{T_pM \times \cdots \times T_pM}_{k \text{ copies}} \to \mathbb{R}.$$

We denote the vector space of all covariant k-tensors on T_pM by $T^k\left(T_p^*M\right)$. A covariant k-tensor whose value is unchanged by interchanging any pair of elements is called a *symmetric tensor*; an *alternating tensor* is a covariant k-tensor whose value changes sign whenever any pair of elements are interchanged. As alternating tensors are the natural language of differential forms[2], we give special notation to the subspace comprising alternating co-

[2] Alternating covariant k-tensors are variously called exterior forms, multi-covectors, or k-covectors.

variant k-tensors: $\Lambda^k\left(T_p^*M\right)$. Analogous to the tangent and cotangent bundles on M, we have the bundle of covariant k-tensors on M, defined by

$$T^kM = \coprod_{p\in M} T^k(T_p^*M), \tag{2.2}$$

and the bundle of alternating covariant k-tensors on M, defined by

$$\Lambda^kT^*M = \coprod_{p\in M} \Lambda^k\left(T_p^*M\right). \tag{2.3}$$

It is smooth sections of these bundles that play critical roles in the research presented in this thesis through two applications. First, they'll serve as metrics, allowing us to compute distances and angles between points on a manifold thus to do Euclidean geometry; second, they'll serve as objects that can be integrated (in a coordinate-independent way) over a manifold, allowing us to compute standard integrals and thus to perform calculus and statistics by classical means.

Riemannian Metrics and Distances

In this section, we review the concepts of a riemannian metric, the Levi-Civita connection, and geodesics on a differentiable manifold. We begin here with a formal definition of a riemannian metric.

Definition 2.1. *Let M be a differentiable manifold. A **riemannian metric** on M is a smooth section of $T^2(M)$ – a 2-tensor field – denoted by g, which for all smooth vector fields X, Y in $\mathfrak{X}(M)$ and $p \in M$ satisfes*

(i) $g(X,Y) = g(Y,X)$*, and*

(ii) $g_p : T_pM \times T_pM \to \mathbb{R}$ *is positive definite.*

For every $p \in M$, the riemannian metric g provides an inner product on T_pM given by the nondegenerate symmetric bilinear form $g_p : T_pM \times T_pM \to \mathbb{R}$. This is the reason to use the standard notation $\langle X, Y\rangle_g$ to denote the real number $g_p(X,Y)$ for X, Y tangent vectors in T_pM. Below are just a few of the geometric constructions that we'll see in application.

Definition 2.2.

- *The **length** (or **norm**) of a tangent vector X in T_pM is defined to be*

$$|X|_g = \langle X, X \rangle_g^{\frac{1}{2}} = g_p(X, X)^{\frac{1}{2}}. \tag{2.4}$$

- *The **angle** between two (nonzero) tangent vectors X, Y in T_pM is the unique $\theta \in [0, \pi]$ satisfying*

$$\cos\theta = \frac{\langle X, Y \rangle_g}{|X|_g |Y|_g}. \tag{2.5}$$

- *Two tangent vectors X and Y are **orthogonal** if $\langle X, Y \rangle_g = 0$.*

- *Let $t \to \gamma(t)$, $t \in [a, b]$ be a curve segment in M. The **length** of γ is defined by the formula*

$$L_g(\gamma) = \int_a^b g_{\gamma(t)} \left(\dot{\gamma}(t), \dot{\gamma}(t)\right)^{1/2} dt = \int_a^b |\dot{\gamma}(t)|_g \, dt. \tag{2.6}$$

If M is connected, any two points p and q in M can be joined by a curve: $t \mapsto \gamma(t)$, $t \in [a, b]$ with $\gamma(a) = p$ and $\gamma(b) = q$. The infimum of the length of all curve segments joining p and q is the *distance* between p and q. Curves which locally yield the shortest distance between two points are called *geodesic curves* and are a main topic of discussion in Chapter 5.

While it's tempting to define geodesics only by their length minimizing property, this definition turns out to be technically difficult to work with in practice. So we will instead present the (equivalent) definition of a geodesic as a curve that is the generalization of a straight line in familiar Euclidean space. Rooted in this generalization is a covariant derivative operator, a tool which will allow for the differentiation of vector fields along curves in a coordinate-independent way. A covariant derivative is defined in terms of a connection on the manifold. While connections can certainly be defined on smooth manifolds in the abstract (see for example [31]), we'll focus only on what is known as the Levi-Civita connection to stay within the scope of this thesis.

Definition 2.3. *Given a smooth riemannian manifold M with riemannian metric g, the* ***Levi-Civita connection*** *is an $\mathbb{R}$-bilinear map*

$$\nabla : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M),$$

written $\nabla(X, Y) = \nabla_X Y$ for vector fields X and Y, such that the following four conditions are satisfied:

(i) $\nabla_{fX} Y = f \nabla_X Y$

(ii) $\nabla_X (fY) = (Xf)Y + f \nabla_X Y$

(iii) $\nabla_X \langle Y, Z \rangle = \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X Z \rangle$

(iv) $\nabla_X Y(f) - \nabla_Y X(f) = X(Yf) - Y(Xf)$

for all vector fields X, Y, Z in $\mathfrak{X}(M)$ and for all smooth functions $f \in C^\infty(M)$.

The vector field $\nabla_X Y$ – called the *covariant derivative of Y with respect to X* – and what is known as the Fundamental Lemma of Riemannian Geometry guarantees the existence (and uniqueness!) of ∇ on a riemannian manifold M. A proof, which includes explicit formulae for computing the connection in local coordinates, can be found in [31], specifically, chapter five. Literature refers to property (iv) in Definition 2.3 as *torsion free*. For smooth manifolds that are also matrix Lie groups equipped with a bi-invariant (riemannian) metric (discussed in Section 2.1.2), we have the following extremely useful and simple formula for the covariant derivative:

$$\nabla_X Y = \frac{1}{2} (XY - YX), \tag{2.7}$$

for vector fields X and Y that are invariant under all left translations. The covariant derivative is the essential tool we use to define geodesic curve in a smooth riemannian manifold.

Definition 2.4. *Let M be a riemannian manifold with riemannian metric g and Levi-Civita connection ∇. A smooth curve γ in M is a* ***geodesic*** *if*

$$\nabla_{\dot\gamma} \dot\gamma \equiv 0.$$

All geodesics are length minimizing; and all length-minimizing curves are geodesics. That is, Definition 2.4 is equivalent to a geodesic defined by the curve (locally) yielding the shortest distance between two points on a manifold. The reader interested in a rigorous proof of this equivalence is referred to Lee [31], specifically chapter six. Perhaps the most well-known example of geodesics are great circles on $\mathbf{S}^2$, depicted in Figure 2.1.

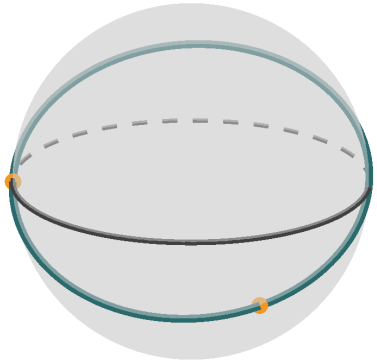

Figure 2.1: Geodesics on $\mathbf{S}^2$ are great circles.

Exterior Differential Forms

Just as the covariant derivative allows us to make sense of differentiation on smooth (riemannian) manifolds, we need an analogous tool that will allow us to make sense of integration over smooth manifolds (and submanifolds). Exterior differential forms are precisely the objects we need to be able to integrate over manifolds in a coordinate-independent way. We begin with the definition.

Definition 2.5. *Let M denote a smooth manifold. A **differential k-form**, or just a k-**form**, is a smooth section of the k^{th} exterior power of the cotangent bundle of M. This is a (smooth) tensor field whose value at each point is an alternating tensor. We denote the vector space of all smooth k-forms by $\Omega^k(M)$.*

To make sense of this definition (by writing it in local coordinates) let us develop a basis for $\Lambda^k M$. If $(x^1, \ldots, x^n)$ is the local coordinate representation of a point $p \in M$, then the

derivations

$$\left.\frac{\partial}{\partial x^1}\right|_p, \ldots, \left.\frac{\partial}{\partial x^n}\right|_p$$

form a basis for T_pM. In unwinding a few definitions, one can see that $\left.\partial/\partial x^i\right|_p$ is just the derivation that takes the i^{th} partial derivative of (the coordinate representation of) a function f at (the coordinate representation of) p. This basis for T_pM gives rise to a dual basis $(\varepsilon_p^1, \ldots, \varepsilon_p^n)$ for T_p^*M. However, application of the differential of a smooth function shows that the coordinate covector field ε_p^i is none other than the familiar dx_p^i. Now, given the basis (dx^i) for T^*M,

$$\left\{dx^I : I \text{ is increasing}\right\}$$

forms a basis for $\Lambda^k M$, where I is a multi-index of length k and dx^I is the covariant k-tensor that acts on vectors $v_1, \ldots, v_n$ by

$$dx^I(v_1, \ldots v_k) = \left(dx^{i_1} \wedge \cdots \wedge dx^{i_k}\right)(v_1, \ldots, v_k) = \det \begin{bmatrix} v_1^{i_1} & \cdots & v_k^{i_1} \\ \vdots & & \vdots \\ v_1^{i_k} & \cdots & v_k^{i_k} \end{bmatrix}. \tag{2.8}$$

In other words, $dx^I(v_1, \ldots, v_k)$ is the determinant of the $k \times k$ minor consisting of the rows $i_1, \ldots, i_k$ of a matrix whose columns are the components of the vectors $v_1, \ldots, v_k$ with respect to the coordinates on M. It follows that a k-form ω is written in local coordinates as

$$\omega = \sum_{\{I:1\le i_1<\cdots<i_k\le n\}} \omega_I dx^I = \sum_{\{I:1\le i_1<\cdots<i_k\le n\}} \omega_I dx^{i_1} \wedge \cdots \wedge dx^{i_k}, \tag{2.9}$$

where the coefficients ω_I are smooth functions of $x^1, \ldots x^n$. The exterior product (wedge product) of a k-form with a an ℓ-form is a $(k+\ell)$-form obtained by formal multiplication of the two forms using the associativity, bilinearity, and anticommutativity laws of the wedge product, as set forth in [32]. Paramount among the properties of the wedge product is the following proposition, a proof of which can be found in [32].

Proposition 2.1. *For any covector $\omega^1, \ldots, \omega^k$ and vectors $v_1, \ldots, v_k$,*

$$\omega^1 \wedge \cdots \wedge \omega^k(v_1, \ldots, v_k) = \det\left(\omega^i(v_j)\right).$$

As the determinant function has managed to weave its way into this discussion, it should be no surprise that the derivation of the invariant measure on the manifolds of particular interest in this thesis hinges on it. Specifically, Chapter 3 will make use of the following theorem[3], which can be proved using the proposition above.

Theorem 2.1. *Let V be an n-dimensional vector space with basis $\{x^1, \ldots, x^n\}$ and let T be a linear operator on V. Then*

$$T(x^1) \wedge \cdots \wedge T(x^n) = (\det T)\left(x^1 \wedge \cdots \wedge x^n\right).$$

Integration of Exterior Differential Forms

The preceding discussion began with a promise that differential forms are objects that can be integrated on manifolds in a coordinate-independent way. While it's certainly possible to define the integral of a differential k-form over a submanifold of dimension k, our applications require only integration of differential forms of maximum degree. We'll see through our promise, but restrict our definition to that of the integral of a differential n-form over a smooth manifold of the same dimension.

We begin by noting that, because there is only one increasing multi-index of length n, every n-form looks like

$$\omega = f dx^1 \wedge \cdots \wedge dx^n, \tag{2.10}$$

where f is a smooth function of $x^1, \ldots x^n$. If the domain of integration, D, is a subset of $\mathbb{R}^n$, the integral of the n-form is simply regarded as an ordinary volume integral over the domain and evaluated as such. It would be a tragedy to omit from this thesis the finest line in all of [32]: to compute the integral of an n-form, we just erase the wedges!

$$\int\limits_{D \subset \mathbb{R}^n} \omega = \int\limits_{D \subset \mathbb{R}^n} f dx^1 \wedge \cdots \wedge dx^n = \int\limits_{D \subset \mathbb{R}^n} f dx^1 \cdots dx^n. \tag{2.11}$$

The integral on the right is the ordinary Lebesgue integral of f on D. It is often the case, however, that we wish to integrate an n-form over a domain of integration on the manifold.

[3] We acknowledge that some authors will define the determinant as the unique scalar satisfying this relation.

For this, we divide the domain of integration into subdomains D_i, each contained in a coordinate chart (U_i, φ_i) on M. Then the pullback $\left(\varphi_i^{-1}\right)^* \omega$ is an n-form on the open subset $\varphi(D_i)$ in $\mathbb{R}^n$ and we define

$$\int\limits_{D\subset M} \omega = \sum_i \int\limits_{D_i} \omega = \sum_i \int\limits_{\varphi(D_i)} \left(\varphi_i^{-1}\right)^* \omega. \tag{2.12}$$

The right-most integral here can be computed by erasing the wedges, as discussed above. With integration defined over a single coordinate patch, one must use a partition of unity to define the integral over the entire manifold. And, finally, to integrate an integrable function on the manifold with respect to the differential form, we simply express it as a function of the coordinates $x^1, \ldots, x^n$ and include it under the integral sign.

We remark that in the general theory of integrating a differential form over a smooth manifold, a difficulty arises as to which sign should be assigned to the integrals over the subdomains, which is connected to the orientation of each domain, before taking the sum. As our purpose in this thesis is to only integrate differential forms over manifolds that admit atlases containing a single chart, we choose to avoid any further discussion of this subtlety.

2.1.2 Lie Groups, Lie Algebras, and the Exponential Mapping

This section addresses an incredibly important special class of differential manifolds: Lie groups. We begin with the definition.

Definition 2.6. *A **Lie group** is a smooth manifold G that is also a group in the algebraic sense, with the property that the multiplication map $m : G \times G \to G$ and the inversion map $i : G \to G$, given by*

$$m(g, h) = gh \qquad\qquad i(g) = g^{-1},$$

are both smooth.

As is customary, we denote the identity element of an arbitrary Lie group by the symbol e, except in specific examples in which there are more common notations (such as $\mathbb{I}$ or $\mathbb{I}_n$

for the identity matrix in a matrix Lie group). A crucial property of Lie groups is that each element $g \in G$ defines a diffeomorphism $L_g : G \to G$ called *left translation* given by

$$L_g(h) = gh$$

(the inverse is given by $L_{g^{-1}}$). Right translation, denoted R_g, is defined analogously. Left (or right) translation is used to diffeomorphically dance around a Lie group; namely from any point $g \in G$, one can move to $h \in G$ via $L_{g^{-1}h}$. Furthermore, the induced map on the tangent spaces $L_{g*} : T_eG \to T_gG$ is an isomorphism of vector spaces. If we can describe the tangent space T_eG to the group at the identity, L_{g*} will thus provide a description of the tangent space T_gG to the group at any point $g \in G$. Explicitly,

$$T_gG \cong \{L_{g*}X : X \in T_eG\}. \tag{2.13}$$

It is for this reason (and, well, for several other reasons) that we dedicate attention to the tangent space of a Lie group at the identity, which we'll soon see is the *Lie algebra* of G.

Of particular interest to the work presented in this thesis is "Her All-embracing Majesty $\mathbf{GL}(n)$," the general linear group consisting of n-by-n invertible matrices with entries from either $\mathbb{R}$ or $\mathbb{C}$, and a few subgroups of a certain sort [52]. Henceforth, the Lie theory presented will be restricted to closed subgroups of $\mathbf{GL}(n)$, i.e., matrix Lie groups, allowing for substantially more tangible definitions of the Lie algebra and the exponential mapping than in the case of a general Lie groups. We choose to generally focus only on real matrix Lie groups in this presentation; we'll highlight any key differences in definitions and theorems for complex matrix Lie groups, however.

The exponential of a matrix plays a crucial role in Lie theory: it enters into the definition of the Lie algebra and is the mechanism for passing information from the Lie algebra to the Lie group. The exponential map is defined for square matrices exactly as for numbers:

$$\exp(\boldsymbol{A}) = \sum_{m=0}^{\infty} \frac{\boldsymbol{A}^m}{m!}. \tag{2.14}$$

For a real (or complex) square matrix $\boldsymbol{A}$, the infinite series defined by $\exp(\boldsymbol{A})$ converges to

an invertible matrix of the same size. Paramount among the properties of the exponential map is the following relationship between the determinant of matrix and its trace.

Proposition 2.2. *For any n-by-n (real or complex) matrix $\boldsymbol{A}$,*

$$\det\left(\exp(\boldsymbol{A})\right) = \exp(\operatorname{trace}(\boldsymbol{A})).$$

A purely algebraic proof of Proposition 2.2 can be found in [17]; a clever geometric proof can be found in [32].

Important for our presentation of the Lie algebra is the relationship between the exponential map and one-parameter subgroups of a Lie group, that is, Lie group homomorphisms from $\mathbb{R}$ into G. Characterization of one-parameter subgroups of matrix Lie groups is straightforward.

Proposition 2.3.

1. *For every n-by-n complex matrix $\boldsymbol{A}$, $F(t) = \exp(t\boldsymbol{A})$ is the one-parameter subgroup of $\mathbf{GL}(n, \mathbb{R})$ generated by $\boldsymbol{A}$.*

2. *If F is a one-parameter subgroup of $\mathbf{GL}(n, \mathbb{R})$, there exists a unique n-by-n real matrix $\boldsymbol{A}$ such that*

$$F(t) = \exp(t\boldsymbol{A}).$$

We're ready for a definition of the Lie algebra of a Lie group.

Definition 2.7. *Let G be (a closed subgroup of) $\mathbf{GL}(n, \mathbb{F})$, where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. The **Lie algebra** of G, denoted $\mathfrak{g}$, is the set of all matrices $\boldsymbol{A}$ (invertible or not!) such that $\exp(t\boldsymbol{A})$ is in G for all real numbers t. That is,*

$$\mathfrak{g} \stackrel{\Delta}{=} \{\boldsymbol{A} \in \boldsymbol{M}(n, \mathbb{F}) : \exp(t\boldsymbol{A}) \in G \text{ for all } t \in \mathbb{R}\}. \tag{2.15}$$

This definition means that X is in $\mathfrak{g}$ if and only if the one-parameter subgroup generated by X lies in G. There is an abstract notion of a Lie algebra (that is not even necessarily associated to any Lie group). With an understanding of a Lie algebra in the abstract, it

is possible to alternatively define the Lie algebra of a Lie group to be the Lie algebra of all smooth left-invariant vector fields on the group. One can construct an explicit one-to-one correspondence between this algebra of smooth left-invariant vector fields on G and the one-parameter subgroups of G, thus equating the definitions. The interested reader is referred to Lee [32] (specifically chapters eight and twenty) for a thorough discussion.

Without considering the abstract notion of a Lie algebra, the result of the following theorem will show that that the Lie algebra, as defined in Definition 2.7, *is* the tangent space to G at the identity. This means that $\mathfrak{g}$ may alternatively be defined as the set of derivations of smooth curves through the identity in G.

Theorem 2.2. *Let G be (a closed subgroup of) $\mathbf{GL}(n, \mathbb{F})$, where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, and $\mathfrak{g}$ its Lie algebra. Then, a matrix $\boldsymbol{X}$ is in $\mathfrak{g}$ if and only if there exists a smooth curve γ in $\mathbf{M}(n, \mathbb{F})$ such that*

(i) $\gamma(t)$ *lies in* G *for all* $t \in \mathbb{R}$,

(ii) $\gamma(0) = \mathbb{I}$, *and*

(iii) $\left.\dfrac{d\gamma}{dt}\right|_{t=0} = \boldsymbol{X}$.

Thus, $\mathfrak{g}$ is the tangent space at the identity to G.

For proof of this theorem, we point the reader to [17]. Elementary exercises for a student in her typical graduate geometry coursework will identify the Lie algebras of classical Lie groups, displayed in Table 2.2.

Lie Group	Lie Algebra
$\mathbf{GL}(n, \mathbb{C})$	$\mathfrak{gl}(n, \mathbb{C}) = \{\boldsymbol{A} \in \boldsymbol{M}(n, \mathbb{C})\}$
$\mathbf{SL}(n, \mathbb{C})$	$\mathfrak{sl}(n, \mathbb{C}) = \{\boldsymbol{A} \in \mathfrak{gl}(n, \mathbb{C}) : \operatorname{tr} \boldsymbol{A} = 0\}$
$\mathbf{U}(n)$	$\mathfrak{u}(n) = \left\{\boldsymbol{A} \in \mathfrak{gl}(n, \mathbb{C}) : \boldsymbol{A}^\dagger = -\boldsymbol{A}\right\}$

$\mathbf{SU}(n)$	$\mathfrak{su}(n) = \mathfrak{sl}(n) \cap \mathfrak{u}(n)$
$\mathbf{GL}(n, \mathbb{R})$	$\mathfrak{gl}(n, \mathbb{R}) = \{\boldsymbol{A} \in \boldsymbol{M}(n, \mathbb{R})\}$
$\mathbf{O}(n)$	$\mathfrak{o}(n) = \left\{\boldsymbol{A} \in \mathfrak{gl}(n, \mathbb{R}) : \boldsymbol{A}^\top = -\boldsymbol{A}\right\}$
$\mathbf{SO}(n)$	$\mathfrak{so}(n) = \mathfrak{o}(n)$

Table 2.2: Lie algebras of classical Lie groups.

Given a Lie group G and its Lie algebra $\mathfrak{g}$, we define the exponential mapping for G to be the map

$$\exp : \mathfrak{g} \to G. \tag{2.16}$$

That is, the exponential mapping for the Lie group G is the matrix exponential restricted to the Lie algebra $\mathfrak{g}$ of G. It's important to note that the exponential map is not always surjective onto the Lie group, as is quickly observed when, for example,

$$\exp : \mathfrak{gl}(n, \mathbb{R}) \to \mathbf{GL}(n, \mathbb{R}).$$

Indeed, since $\det(\exp(\boldsymbol{A})) = \exp(\operatorname{trace}(\boldsymbol{A}))$, every matrix of the form $\exp(\boldsymbol{A})$ has a positive determinant. Furthermore, the exponential mapping may not be one-to-one onto $\mathfrak{g}$. It is, however, *locally* one-to-one and onto, allowing it to serve as the crucial mechanism for passing information between the group and the Lie algebra. Furthermore, the exponential map from the Lie algebra to the Lie group is the natural language of geodesics on Lie groups equipped with a particular metric.

Definition 2.8. *Let G be a Lie group with Lie algebra $\mathfrak{g}$. A riemannian metric g on G is **left-invariant** if it is invariant under all left translations: $L_{p*}g = g$ for all $p \in G$. Similarly, g is **right-invariant** if it is invariant under all right translations, and **bi-invariant** if it is both left- and right-invariant.*

This is to alternatively say that a metric is left-invariant (respectively right-invariant) if and only if left (respectively right) translations are isometric with respect to the metric.

Unraveling the notation of the push-forward in Definition 2.8, the riemannian metric g is left-invariant precisely when

$$\langle u, v\rangle_a = \langle L_{p*|_a} u, L_{p*|_a} v\rangle_{L_p a} = \langle L_{p*|_a} u, L_{p*|_a} v\rangle_{pa} \tag{2.17}$$

for all p, a in G and for all u, v in T_pG. As we're working exclusively with Lie groups that are closed subgroups of $\mathbf{GL}(n, \mathbb{R})$ (or $\mathbf{GL}(n, \mathbb{C})$), left-translation is the restriction of a linear map whence, for any matrix $\boldsymbol{A}$, $L_{\boldsymbol{A}*} = L_{\boldsymbol{A}}$ whence $T_{\boldsymbol{A}}G \cong \boldsymbol{A}\mathfrak{g}$. The condition of left-invariance of a metric g becomes remarkably simple:

$$g(\boldsymbol{U}, \boldsymbol{V}) = g_{\boldsymbol{A}}(\boldsymbol{AU}, \boldsymbol{AV}),$$

or, equivalently,

$$g_{\boldsymbol{A}}(\boldsymbol{U}, \boldsymbol{V}) = g(\boldsymbol{A}^{-1}\boldsymbol{U}, \boldsymbol{A}^{-1}\boldsymbol{V}), \tag{2.18}$$

for any matrices $\boldsymbol{U}, \boldsymbol{V}$ in $\boldsymbol{A}\mathfrak{g}$ and any matrix $\boldsymbol{A}$ in G. Similarly, the condition of right-invariance of the metric is given by

$$g_{\boldsymbol{A}}(\boldsymbol{U}, \boldsymbol{V}) = g(\boldsymbol{U}\boldsymbol{A}^{-1}, \boldsymbol{V}\boldsymbol{A}^{-1}). \tag{2.19}$$

It is true that if G is commutative, a bi-invariant (riemannian) metric exists, and, if G is compact, a bi-invariant (riemannian) metric may be constructed from its Haar measure. However, for the Lie groups of interest in this thesis, we'll be able to explicitly write out the bi-invariant (riemannian) metric inherited from the Hilbert-Schmidt inner product on the vector space of all finite-dimensional matrices. Explicitly, we will make use of the following proposition.

Proposition 2.4. *Let G be a Lie group. Any inner product $\langle\cdot,\cdot\rangle_e$ defined on T_eG can be extended to a left-invariant riemannian metric on G.*

Proof. Define for each $g \in G$ and for all u, v in T_gG,

$$\langle u, v\rangle_g = \langle d\left(L_{g^{-1}}\right)_g u, d\left(L_{g^{-1}}\right)_g v\rangle_e.$$

Left-invariance follows from application of the push-forward to the composition of functions:

$$\begin{aligned}\langle d(L_s)_t u, d(L_s)_t v\rangle_{st} &= \langle d(L_{(st)^{-1}})_{st} d(L_s)_t u, d(L_{(st)^{-1}})_{st} d(L_s)_t v\rangle_e \\ &= \langle d(L_{(st)^{-1}} \circ L_s)_t u, d(L_{(st)^{-1}} \circ L_s)_t v\rangle_e \\ &= \langle d(L_{t^{-1}})_t u, d(L_{t^{-1}})_t v\rangle_e \\ &= \langle u, v\rangle_t. \end{aligned}$$

■

While the study of Lie groups equipped with a bi-invariant riemannian metric is rich, we record only the most noteworthy result for application in this thesis in the following theorem.

Theorem 2.3. *Let G be a Lie group equipped with a bi-invariant riemannian metric. The geodesics γ in G with $\gamma(0) = e$ are precisely the one-parameter subgroups of G.*

As Proposition 2.3 characterizes one-parameter subgroups in matrix Lie groups in terms of the exponential map, a geodesic starting at $\mathbb{I}$ in G is the curve $t \mapsto \exp(t\boldsymbol{X})$, where $\boldsymbol{X}$ is in $\mathfrak{g}$. Since left translations are isometries, all left translates of one-parameter subgroups are geodesics. Thus, for a fixed point $\boldsymbol{A}$ in G and a tangent vector $\boldsymbol{X}$ in $\mathfrak{g}$, the unique geodesic passing through $\boldsymbol{A}$ in the direction of $\boldsymbol{X}$ is written

$$\gamma(t) = L_{\boldsymbol{A}} \circ \exp(t\boldsymbol{X}) = \boldsymbol{A}\exp(t\boldsymbol{X}). \tag{2.20}$$

Lie Group Actions and Quotient Manifolds

Because many of the most important applications of Lie groups (especially in the work presented in this thesis!) involve them acting on other manifolds, this section will concentrate on properties of Lie group actions. For the remainder of this section, let G be a Lie group and let M be a (smooth) manifold.

Definition 2.9. *A **left-action** of a group G on M is a map $G \times M \to M$, often written*

as $(g, p) \mapsto g \cdot p$*, that satisfies*

$$\begin{aligned} g_1 \cdot (g_2 \cdot p) &= (g_1 g_2) \cdot p \\ e \cdot p &= p. \end{aligned} \tag{2.21}$$

A manifold endowed with a specific G*-action is called a* G***-space***.

For completeness, we include some basic terminology regarding Lie group actions.

Definition 2.10.

- *The action is* ***smooth*** *if it is a smooth map from* $G \times M$ *into* M.
- *For any point* $p \in M$*, the* ***orbit*** *of* p *under the action is the set of all images of* p *under elements of* G*; that is:*

$$G \cdot p \stackrel{\Delta}{=} \{g \cdot p : g \in G\}.$$

- *The action is* ***transitive*** *if for each pair of points* p, q *in* M *there exists an element* g *in* G *such that* $g \cdot p = q$.
- *For any point* $p \in M$ *the* ***stabilizer subgroup***[4] *of* p *is the set of elements* g *in* G *that fix* p*; that is:*

$$G_p \stackrel{\Delta}{=} \{g \in G : g \cdot p = p\}.$$

- *The action is* ***free*** *if the only element of* G *that fixes any point in* M *is the identity; that is, if* $g \cdot p = p$ *for some* p *in* M*, then* $g = e$.

Among the most interesting kinds of group actions are those in which a Lie group acts transitively. Examples of manifolds endowed with a transitive smooth action by a Lie group – *homogeneous spaces* – are the principal study of the work presented in this thesis. In the study of homogeneous spaces, the stabilizer subgroup plays a pivotal role, for when G acts on M, G_p – for any p in M – is a closed Lie subgroup of G. Allowing G_p to act on G by right translation allows for an application of the Quotient Manifold Theorem [32], which

[4] The stabilizer subgroup is sometimes called the isotropy subgroup.

guarantees that the coset space G/G_p – again, for any p in M – is a topological manifold endowed with a unique smooth structure such that the quotient map

$$\pi : G \to G/G_p \tag{2.22}$$

is a *smooth submersion*, i.e., its push-forward is surjective at each point. In the case, the orbit space G/G_p is a manifold in the honest sense of the word. When G is a Lie group equipped with a riemannian metric, the quotient map is in fact a smooth, *riemannian* submersion.

We now specialize to the case of a Lie group G acting transitively on a set M. For any fixed element p in M, there exists a one-to-one correspondence[5] between the set of (left-)cosets of G_p in G and M so that M is a quotient of G by G_p. Indeed, we define a map

$$F_p : G/G_p \to M,$$

by

$$gG_p \mapsto g \cdot p.$$

Since

$$gG_p = hG_p \Leftrightarrow g^{-1}h \in G_p \Leftrightarrow g^{-1}h \cdot p = p \Leftrightarrow g \cdot p = h \cdot p,$$

it follows that F_p is injective. In unraveling the definition of a transitive group action, we see F_p is also surjective and thus defines a bijection of sets identifying left multiplication in G/G_p with the action of G on M. When M is a smooth manifold, F_p is equivariant with respect to the group action and hence a diffeomorphism. It's important to note that because the two stabilizer subgroups G_p and G_q of any two points p and q in M are isomorphic (they are conjugate subgroups), it is enough to compute the stabilizer subgroup for a convenient p. We henceforth identify a point q in M with its coset representative qG_p in the coset space G/G_p and we refer to M (and G/G_p) as the *quotient*.

[5] This is a consequence of the so-called Orbit-Stabilizer Theorem. See [7].

As a directly relevant example, consider the transitive, smooth action of the orthogonal group $\mathbf{O}(n)$ on the real Grassmann manifold $\mathbf{G}(k,n)$ defined by left multiplication. That is,

$$\boldsymbol{R} \cdot \operatorname{span}\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\} = \operatorname{span}\{\boldsymbol{R}\boldsymbol{x}_1, \ldots, \boldsymbol{R}\boldsymbol{x}_k\}, \tag{2.23}$$

where the $\boldsymbol{x}_i$ are linearly independent, orthonormal n-vectors. As detailed fully in Chapter 5, the stabilizer subgroup of the element in $\mathbf{G}(k,n)$ that is the span of the first k standard basis vectors in $\mathbb{R}^n$ is isomorphic to

$$\mathbf{O}(k) \times \mathbf{O}(n-k), \tag{2.24}$$

from which a familiar presentation of the Grassmann manifold follows:

$$\mathbf{G}(k,n) \cong \mathbf{O}(n) / \left(\mathbf{O}(k) \times \mathbf{O}(n-k)\right). \tag{2.25}$$

We often refer to the quotient map π in Equation 2.22, which corresponds to a fixed p in M, as the projection of G onto M and, for every $q \in M$, the inverse image $\pi^{-1}(q) \subset G$ as the *fiber* above q. Because the projection of G onto M is a smooth, surjective submersion, if G possesses a riemannian metric, the tangent space to G at each point $g \in G$ decomposes into an orthogonal (with respect to the metric) direct sum

$$T_g G = \mathcal{V}_g \oplus \mathcal{H}_g, \tag{2.26}$$

where $\mathcal{V}_g \triangleq \ker d\pi_g$ is the *vertical space* and $\mathcal{H}_g \triangleq \mathcal{V}_g^{\perp}$ is the *horizontal space*. Via this direct sum, any tangent vector X in $T_g G$ may be written uniquely as

$$X = X^{\mathcal{V}} + X^{\mathcal{H}},$$

where $X^{\mathcal{V}}$ is vertical and $X^{\mathcal{H}}$ is horizontal. For each g in G,

$$d\pi_g : T_g G \to T_{\pi(g)}\left(G/G_p\right)$$

is a surjective linear map whose kernel is precisely the tangent space at g to the G_p-orbit of g under the action of G_p on G defined by right translation; that is,

$$\ker d\pi_g = T_g\left(G_p \star g\right).$$

By elementary algebra, $d\pi_g$ descends to an isomorphism

$$T_gG/T_g\left(G_p \star g\right) \cong T_{\pi(g)}\left(G/G_p\right).$$

Because this isomorphism is defined independently of any choices, we canonically identify each element in $T_{\pi(g)}\left(G/G_p\right)$ with an element of the quotient space $T_gG/\ker d\pi_g = T_gG/V_g$, which, in light of the direct sum decomposition in Equation 2.26, *is* the horizontal space at g. Thus, for a point $q \in M$, we can identify each element in the tangent space to M at the point q with an element of a horizontal space to G at a point in the fiber above q.

As this discussion suggests, horizontal tangent vectors play a critical role in many differential geometric objects on the quotient. Principal to our applications, is the dependence on horizontal tangent vectors in G to characterize metrics and geodesics in the quotient. The following theorem, which will conclude the presentation of the necessary geometric framework for this thesis, elucidates the dependence.

Theorem 2.4. *Let G be a Lie group equipped with a bi-invariant riemannian metric and let M be a G-space. If γ is a geodesic in G such that $\gamma'(0)$ is a horizontal vector, then γ is a horizontal geodesic in G, and the image of γ under the quotient map $\pi : G \to G/G_p$ – for any p in M – is a geodesic in M of the same length as γ.*

A proof of a generalized version of the theorem above can be found in [36], specifically chapter seven.

2.1.3 Complex Differentiable Manifolds

In many important applications of differential and riemannian geometry, most notably those involving the collection and analysis of real-world sensor data, we will encounter generalizations to complex manifolds. Our position is that a complex-analytic manifold of complex dimension n may be regarded as a $2n$-dimensional real manifold having certain additional properties. A key additional property is the requirement of the *complex-analytic* structure to comprise homeomorphic coordinate maps to the open unit disk in $\mathbb{C}^n$, such that

transition maps between any two coordinate maps is *bi-holomorphic*. This thesis does not dive deeply into the theory of complex-analytic manifolds, limiting our study to only the essential differences from real, differentiable manifolds impacting our work (of which there are few).

2.2 Foundations of Statistical Signal Theory

As discussed in Chapter 1, the work presented in this thesis is motivated by various multi-sensor remote sensing applications wherein a primary goal is to estimate a common unknown signal of known rank k measured by m spatially distributed sensors. The measurement collected at each sensor has the form of a complex n-vector, where $n \gg m$, and is regarded as an element of an ambient n-dimensional vector space. When the number of sensors exceeds the rank of the signal ($m > k$), the rank-k signal defines a particular k-dimensional subspace called the *signal subspace* that can be regarded as an element of $\mathbf{G}(k, n)$, the Grassmann manifold of all k-dimensional subspaces of an n-dimensional vector space. In practical problems, the sensor data are almost always corrupted by noise or other distortions which must be considered when seeking to identify the signal subspace. Much of modern signal theory is concerned with modeling of signal distortion probabilistically and developing statistical methods, such as hypothesis tests and parameter estimators, to understand the signal from the sensor measurements. The ubiquity of this type of problem in practice has led to development of a substantial corpus of theory and methods often referred to as *statistical signal theory* and *statistical signal processing*. The standard reference texts treating statistical signal theory and statistical signal processing recommended at the beginning of this chapter ([19, 26, 27, 41]) generally focus on the setting of a real or complex inner product space M. M is usually assumed to be finite-dimensional, which is practically justified by the limitations of realistic processing algorithms and achieved mathematically by approximately representing elements of infinite-dimensional Hilbert space using finitely many basis functions. These assumptions are leveraged to justify development of the theory for $M = \mathbb{R}^n$ or $M = \mathbb{C}^n$. While our goal is to

work with $M = \mathbf{G}(k, n)$, it is instructive to consider well-understood standard elements of statistical signal theory in $\mathbb{R}^n$ or $\mathbb{C}^n$ to motivate subsequent discussion of the setting where M is a non-linear manifold.

2.2.1 The Probability Background

The mathematical framework for statistical signal theory and statistical signal processing is provided by the theory of probability. Analogous to the presentation of the relevant theories of differential and riemannian geometry, this section serves to define some basic concepts of the theory of probability (rooted of course in the theory of measure and integration), to establish notation, and to state without proof some principal results that will be used throughout the remainder of this chapter and the document.

Probability theory begins with the concept of a *probability space*, typically denoted by the triplet $(\Omega, \mathcal{A}, \Pr)$. Adopting the terminology of elementary probability theory, the *sample space* Ω is a set of all possible outcomes $\omega \in \Omega$ of an experiment. Allowing 2^Ω to denote the set of all possible subsets of Ω, the *event space* $\mathcal{A}$ is a subset of 2^Ω that is closed under both complements and countable unions; that is $\mathcal{A}$ is a σ-algebra of subsets of Ω. It is the elements of $\mathcal{A}$ – called *events* – to which we assign probabilities.

Definition 2.1. *A pair $(\Omega, \mathcal{A})$ with $\mathcal{A}$ a σ-algebra of subsets of Ω is called a **measurable space.** Given a measurable space $(\Omega, \mathcal{A})$, a **measure** $\Pr$ is any countably additive, non-negative set function on $\mathcal{A}$. That is, $\Pr : \mathcal{A} \to \mathbb{R}^+$, having the properties:*

(i) $\Pr[A] \geq \Pr[\emptyset] = 0$ *for all* $A \in \mathcal{A}$, *and*

(ii) $\Pr\left[\bigcup A_j\right] = \sum \Pr[A_j]$ *for any countable collection of disjoint sets* A_j *in* $\mathcal{A}$.

When, in addition, $\Pr[\Omega] = 1$, *the measure* $\Pr$ *is a **probability measure** and the triplet* $(\Omega, \mathcal{A}, \Pr)$ *is a **probability space**.*

Fundamental to the study of statistical signal theory and statistical signal processing is the concept of *measurable* functions on probability spaces, which, in what has to be one of the stranger abuses of mathematical language are known as *random variables*.

Random Variables

Definition 2.2. *Let* $(\Omega, \mathcal{A}, \Pr)$ *be a probability space and* $(M, \mathcal{M})$ *a measurable space. A* ***random variable on*** M *is a function* $\mathsf{x} : \Omega \to M$ *such that the inverse image under* x *of all elements in* $\mathcal{M}$ *are events. That is:*

$$\mathsf{x}^{-1}(B) \triangleq \{\omega \in \Omega : \mathsf{x}(\omega) \in B\} \in \mathcal{A}, \text{ for all } B \in \mathcal{M}.$$

The codomain M *is called the* ***observation space*** *of the random variable* x.

Throughout this thesis, we will be interested in two cases for $(M, \mathcal{M})$: the first is the case in which $M = \mathbb{R}^n$ (and $\mathcal{M}$ is the usual Borel σ-algebra of subsets of $\mathbb{R}^n$); the second is the case in which $M = \boldsymbol{G}(k, n)$. For the present section, it will be instructive to focus solely on the former, where the theory is particularly well known. As a reminder, we adopt the useful convention of using fonts without serifs to denote random variables. Thus the vector-valued random variable x – termed an n-variate random variable – may take on the value $\boldsymbol{x} = \mathsf{x}(\omega)$.

As defined next, every random variable x induces a probability measure on its range, closely related to its cumulative distribution function.

Definition 2.3. *The* ***distribution*** *of a vector-valued random variable* x*, denoted* P_{x}*, is the probability measure on* $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ *such that*

$$P_{\mathsf{x}}(B) = \Pr\left[\{\omega \in \Omega : \mathsf{x}(\omega) \in B\}\right]$$

for any Borel set B*; it suffices*[6] *to require the same of all intervals* $(-\infty, \boldsymbol{x}] = (-\infty, x_1] \times \cdots \times (-\infty, x_n] \subset \mathbb{R}^n$.

Definition 2.4. *The* ***(cumulative) distribution function*** *of a vector-valued random variable* x*, denoted* F_{x}*, on* $\mathbb{R}^n$ *is given by*

$$\begin{aligned} F_{\mathsf{x}}(\boldsymbol{x}) &= \Pr\left[\{\omega : \mathsf{x}_1(\omega) \leq x_1, \mathsf{x}_2(\omega) \leq x_2, \ldots, \mathsf{x}_n(\omega) \leq x_n\}\right] \\ &= \Pr\left[\{\omega : \mathsf{x}(\omega) \leq \boldsymbol{x}\}\right], \end{aligned} \tag{2.27}$$

[6] The Borel σ-algebra on $\mathbb{R}^n$ is generated by all (products of) half rays $(-\infty, a_1] \times \cdots \times (-\infty, a_n]$. See [12].

where the last expression is common shorthand.[7] *In terms of the distribution, the cumulative distribution function*[8] *is*

$$F_{\mathbf{x}}(\boldsymbol{x}) = P_{\mathbf{x}}\left((-\infty, x_1] \times \cdots \times (-\infty, x_n]\right). \tag{2.28}$$

Definition 2.4 entails at once that $F_{\mathbf{x}}$ is a non-decreasing, right-continuous function such that

$$\lim_{\boldsymbol{x} \to -\boldsymbol{\infty}} F_{\mathbf{x}}(\boldsymbol{x}) = 0 \text{ and } \lim_{\boldsymbol{x} \to \boldsymbol{\infty}} F_{\mathbf{x}}(\boldsymbol{x}) = 1.$$

In fact, the distribution function is a complete characterization of the random variable. Likewise, the *probability density function*, $f_{\mathbf{x}}(\boldsymbol{x})$, which is related to the distribution function by

$$f_{\mathbf{x}}(\boldsymbol{x}) = \frac{dF_{\mathbf{x}}(\boldsymbol{x})}{d\boldsymbol{x}}, \tag{2.29}$$

is also a complete characterization.[9] This follows from the fact that a distribution function admitting a density can be expressed as

$$F_{\mathbf{x}}(\boldsymbol{x}) = \int_{-\infty}^{x_1} \cdots \int_{\infty}^{x_n} f_{\mathbf{x}}(\boldsymbol{x}) d\boldsymbol{x}, \tag{2.30}$$

for all $\boldsymbol{x} \in \mathbb{R}^n$. In terms of the distribution,

$$P_{\mathbf{x}}(B) = \int_{B} f_{\mathbf{x}}(\boldsymbol{x}) d\boldsymbol{x}, \tag{2.31}$$

for any Borel set $B \subset \mathbb{R}^n$. It follows too that the density $f_{\mathbf{x}}(\boldsymbol{x})$ must be an integrable, (almost-everywhere) non-negative function, with

$$\int_{\mathbb{R}^n} f_{\mathbf{x}}(\boldsymbol{x}) d\boldsymbol{x} = 1. \tag{2.32}$$

---

[7] We'll likely adopt the habit of omitting the argument of a random variable entirely, writing $\mathbf{x} = \mathbf{x}(\omega)$.

[8] As a consequence of the theory of Borel measures, the distribution function of a random variable uniquely determines the distribution of the random variable. See [12].

[9] Because it is always true for models used in practice, we assume in our treatment that densities always exist. However, it is worth keeping in mind that there certainly do exist random variables whose probability distributions are not differentiable (precisely those that are not absolutely continuous with respect to the probability measure).

In fact, any integrable, (almost-everywhere) non-negative function that integrates to unity over $\mathbb{R}^n$ defines a distribution by Equations 2.30 and 2.31, and hence a random variable on $\mathbb{R}^n$. While $\mathbb{R}^n$ will remain the sole focus of the developing treatment of introductory statistical signal theory in this section, it is crucial to note here that the generalization to an integrable, (almost-everywhere) non-negative function defined on $M = \boldsymbol{G}(k, n)$ (or on any non-linear manifold for the matter) that integrates to unity over M will likewise define a distribution, and hence a random variable on M, provided M is equipped with an atlas and an invariant measure with respect to which integration is permissible.

Expectations, Correlations, & Covariance Matrices

Often we are interested in partial characterizations of a random variable in the form of certain *expectations*.

Definition 2.5. *Let* $(\Omega, \mathcal{A}, \Pr)$ *be a probability space and* $\mathsf{x} = \mathsf{x}(\omega)$ *a vector-valued random variable on* $\mathbb{R}^n$ *with density function* $f_{\mathsf{x}}(\boldsymbol{x})$*. The* ***expected value*** *of* x*, denoted* $\mathrm{E}[\mathsf{x}]$*, is the value of the integral*

$$\mathrm{E}[\mathsf{x}] \triangleq \int_{\mathbb{R}^n} \boldsymbol{x} f_{\mathsf{x}}(\boldsymbol{x}) d\boldsymbol{x}, \tag{2.33}$$

should the integral exist.

Definition 2.5 entails at once that the expected value is a linear operator, in the sense that, for any random variables x and y on M, and any (finite) constants c_1, c_2,

$$\mathrm{E}[c_1\mathsf{x} + c_2\mathsf{y}] = c_1\mathrm{E}[\mathsf{x}] + c_2\mathrm{E}[\mathsf{y}],$$

provided both $\mathrm{E}[\mathsf{x}]$ and $\mathrm{E}[\mathsf{y}]$ exist. Furthermore, a measurable function of a random variable is itself a random variable whose expectation is often most readily computed via application of the following fundamental theorem of expectation.

Theorem 2.1. *Let* $(\Omega, \mathcal{A}, \Pr)$ *be a probability space and* $\mathsf{x} = \mathsf{x}(\omega)$ *a vector-valued random variable on* $\mathbb{R}^n$ *with density function* $f_{\mathsf{x}}(\boldsymbol{x})$*. Given a (measurable) function* $\boldsymbol{g}$ *of the random*

variable $\mathbf{x}$*, the resulting random variable* $\boldsymbol{g}(\mathbf{x})$ *has expected value*

$$\mathrm{E}[\boldsymbol{g}(\mathbf{x})] = \int_{\mathbb{R}^n} \boldsymbol{g}(\boldsymbol{x}) f_{\mathbf{x}}(\boldsymbol{x}) d\boldsymbol{x},$$

provided the integral exists.

Several expectations that are important characterizations of random variables are the *mean value* (or first moment)

$$m_{\mathbf{x}} \triangleq \mathrm{E}[\mathbf{x}],$$

the n-by-n *correlation matrix* (or second moment)

$$\mathrm{E}[\mathbf{x}\mathbf{x}^\top],$$

and the n-by-n *covariance matrix* (or second central moment)

$$\Lambda_{\mathbf{x}} \triangleq \mathrm{E}[(\mathbf{x} - m_{\mathbf{x}})(\mathbf{x} - m_{\mathbf{x}})^\top].$$

We should note that it is possible to regard $\mathbf{x}$ as a collection of n real-valued random variables $\mathsf{x}_1, \ldots, \mathsf{x}_n$ on $\mathbb{R}$, whence $f_{\mathbf{x}}(\boldsymbol{x}) = f_{\mathsf{x}_1,\ldots,\mathsf{x}_n}(x_1, \ldots, x_n)$ is termed a *joint probability density function*. The $(i,j)^{\text{th}}$ element of the covariance matrix $\Lambda_{\mathbf{x}}$ is

$$[\Lambda_{\mathbf{x}}]_{ij} = \mathrm{cov}(\mathsf{x}_i, \mathsf{x}_j) = \mathrm{E}[(\mathsf{x}_i - m_{\mathsf{x}_i})(\mathsf{x}_j - m_{\mathsf{x}_j})],$$

and when $i = j$,

$$\mathrm{cov}(\mathsf{x}_i, \mathsf{x}_i) = \mathrm{E}[(\mathsf{x}_i - m_{\mathsf{x}_i})^2]$$

is the *variance* of the real-valued random variable x_i.

Pairs of Random Variables

Let $(\Omega, \mathcal{A}, \Pr)$ be a probability space and let $\mathbf{x} = \mathbf{x}(\omega)$ and $\mathbf{y} = \mathbf{y}(\omega)$ be vector-valued random variables on $\mathbb{R}^n$ and $\mathbb{R}^\ell$, respectively. The joint behavior of $\mathbf{x}$ and $\mathbf{y}$ is described via their *joint (cumulative) distribution function* and their *joint probability density function* defined by

$$F_{\mathbf{x},\mathbf{y}}(\boldsymbol{x}, \boldsymbol{y}) \triangleq \Pr\left[\{\omega : \mathbf{x}(\omega) \leq \boldsymbol{x} \text{ and } \mathbf{y}(\omega) \leq \boldsymbol{y}\}\right] \tag{2.34}$$

and

$$f_{\mathbf{x},\mathbf{y}}(\boldsymbol{x},\boldsymbol{y}) \triangleq \frac{\partial^2 F_{\mathbf{x},\mathbf{y}}(\boldsymbol{x},\boldsymbol{y})}{\partial \boldsymbol{x} \partial \boldsymbol{y}}, \tag{2.35}$$

respectively.[10],[11] Equation 2.35 directly follows from the fact that a joint distribution function admitting a joint density can be expressed as[12]

$$F_{\mathbf{x},\mathbf{y}}(\boldsymbol{x},\boldsymbol{y}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} \int_{-\infty}^{y_1} \cdots \int_{-\infty}^{y_\ell} f_{\mathbf{x},\mathbf{y}}(\boldsymbol{x},\boldsymbol{y}) d\boldsymbol{x} d\boldsymbol{y}. \tag{2.36}$$

As previously mentioned, the distribution and density functions of the random variable $\mathbf{x}$ on $\mathbb{R}^n$ defined in Equations 2.27 and 2.29 may evidently be regarded as the joint distribution and the joint density functions of n random variables $\mathsf{x}_1, \ldots, \mathsf{x}_n$ on $\mathbb{R}$. It is a general principle that all properties of random variables that are relevant to probability theory can be expressed in terms of their joint distributions. The joint (cumulative) distribution function $F_{\mathbf{x},\mathbf{y}}(\boldsymbol{x},\boldsymbol{y})$ and the (cumulative) distribution functions $F_{\mathbf{x}}(\boldsymbol{x})$ and $F_{\mathbf{y}}(\boldsymbol{y})$ of each random variable $\mathbf{x}$ and $\mathbf{y}$ separately are related by

$$\begin{aligned} F_{\mathbf{x}}(\boldsymbol{x}) &= \lim_{\boldsymbol{y} \to \boldsymbol{\infty}} F_{\mathbf{x},\mathbf{y}}(\boldsymbol{x},\boldsymbol{y}) \\ F_{\mathbf{y}}(\boldsymbol{y}) &= \lim_{\boldsymbol{x} \to \boldsymbol{\infty}} F_{\mathbf{x},\mathbf{y}}(\boldsymbol{x},\boldsymbol{y}). \end{aligned} \tag{2.37}$$

Here, $F_{\mathbf{x}}(\boldsymbol{x})$ and $F_{\mathbf{y}}(\boldsymbol{y})$ are called the *marginal (cumulative) distribution functions* of $F_{\mathbf{x},\mathbf{y}}(\boldsymbol{x},\boldsymbol{y})$. The corresponding *marginal probability density functions* $f_{\mathbf{x}}(\boldsymbol{x})$ and $f_{\mathbf{y}}(\boldsymbol{y})$ of each random variable $\mathbf{x}$ and $\mathbf{y}$ separately can be computed via the joint probability density in Equation 2.35, without reference to the underlying distribution:

$$\begin{aligned} f_{\mathbf{x}}(\boldsymbol{x}) &= \int_{\mathbb{R}^\ell} f_{\mathbf{x},\mathbf{y}}(\boldsymbol{x},\boldsymbol{y}) d\boldsymbol{y} \\ f_{\mathbf{y}}(\boldsymbol{y}) &= \int_{\mathbb{R}^n} f_{\mathbf{x},\mathbf{y}}(\boldsymbol{x},\boldsymbol{y}) d\boldsymbol{x}. \end{aligned} \tag{2.38}$$

[10] We'll likewise only treat joint distribution functions that admit joint density functions.

[11] If $\mathbf{x}$ is a vector-valued random variable and $\boldsymbol{x}$ is a vector of the same order, $\mathbf{x} = \mathbf{x}(\omega) \leq \boldsymbol{x}$ is interpreted to mean that $\mathsf{x}_i \leq x_i$ for each respective element.

[12] We note that, for our purposes, the order of integration is interchangeable.

Geometrically, it is useful to visualize marginal densities $f_{\mathsf{x}}(\boldsymbol{x})$ and $f_{\mathsf{y}}(\boldsymbol{y})$ as (integrated) projections of the joint density $f_{\mathsf{x},\mathsf{y}}(\boldsymbol{x},\boldsymbol{y})$ defined on the product space $\mathbb{R}^n \times \mathbb{R}^\ell$ to a single factor.

The conditional density for x given $\mathsf{y} = \boldsymbol{y}$ is defined by

$$f_{\mathsf{x}}(\boldsymbol{x}|\mathsf{y}=\boldsymbol{y}) \triangleq \frac{f_{\mathsf{x},\mathsf{y}}(\boldsymbol{x},\boldsymbol{y})}{f_{\mathsf{y}}(\boldsymbol{y})}. \tag{2.39}$$

Since $f_{\mathsf{y}}(\boldsymbol{y}|\mathsf{x}=\boldsymbol{x})$ is defined analogously, we have

$$f_{\mathsf{x},\mathsf{y}}(\boldsymbol{x},\boldsymbol{y}) = f_{\mathsf{x}}(\boldsymbol{x}|\mathsf{y}=\boldsymbol{y})f_{\mathsf{y}}(\boldsymbol{y}) = f_{\mathsf{y}}(\boldsymbol{y}|\mathsf{x}=\boldsymbol{x})f_{\mathsf{x}}(\boldsymbol{x}),$$

from which *Bayes' Rule* (fundamental to the bayesian inference developed in ensuing chapters of this thesis!) readily follows:

$$f_{\mathsf{x}}(\boldsymbol{x}|\mathsf{y}=\boldsymbol{y}) = \frac{f_{\mathsf{y}}(\boldsymbol{y}|\mathsf{x}=\boldsymbol{x})f_{\mathsf{x}}(\boldsymbol{x})}{f_{\mathsf{y}}(\boldsymbol{y})} = \frac{f_{\mathsf{y}}(\boldsymbol{y}|\mathsf{x}=\boldsymbol{x})f_{\mathsf{x}}(\boldsymbol{x})}{\int_{\mathbb{R}^n} f_{\mathsf{y}}(\boldsymbol{y}|\mathsf{x}=\boldsymbol{x})f_{\mathsf{x}}(\boldsymbol{x})d\boldsymbol{x}}. \tag{2.40}$$

We remark that in cases where it will cause no confusion, we may omit the conditioning random variable from the conditional density function, i.e.,

$$f_{\mathsf{x}}(\boldsymbol{x}|\boldsymbol{y}) = f_{\mathsf{x}}(\boldsymbol{x}|\mathsf{y}=\boldsymbol{y}).$$

Definition 2.6. *Two vector-valued random variables* x *and* y *are* ***independent*** *if*

$$f_{\mathsf{x},\mathsf{y}}(\boldsymbol{x},\boldsymbol{y}) = f_{\mathsf{x}}(\boldsymbol{x})f_{\mathsf{y}}(\boldsymbol{y}). \tag{2.41}$$

A family of vector-valued random variables $\mathsf{x}_1,\ldots,\mathsf{x}_n$ *is* ***mutually independent*** *if*

$$f_{\mathsf{x}_1,\ldots,\mathsf{x}_n}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n) = \prod_{i=1}^{n} f_{\mathsf{x}_i}(\boldsymbol{x}_i). \tag{2.42}$$

In a straightforward application of the definition of conditional probabilities, Equation 2.41 is equivalent to the following condition for a pair of independent random variables x and y:

$$f_{\mathsf{x}}(\boldsymbol{x}|\mathsf{y}=\boldsymbol{y}) = f_{\mathsf{x}}(\boldsymbol{x}), \tag{2.43}$$

and Equation 2.42 is equivalent to the following condition for multiple independent random variables $\mathbf{x}_1, \ldots, \mathbf{x}_n$:

$$f_{\mathbf{x}_i}(\boldsymbol{x}_i | \{\mathbf{x}_j = \boldsymbol{x}_j, j \in \mathcal{J}\}) = f_{\mathbf{x}_i}(\boldsymbol{x}_i), \tag{2.44}$$

where $\mathcal{J}$ is a subset of distinct indices selected from 1 to n.

2.2.2 *Statistical Signal Theory: Estimation*

The starting point of estimation theory, as formulated in the preceding section, is a random, observable variable $\mathbf{x}$, whose distribution admits a density function *parameterized* by an unknown variable $\boldsymbol{\theta}$. Given the conditional probability density $f_{\mathbf{x}}(\boldsymbol{x}|\boldsymbol{\theta})$, the theory of parameter estimation seeks a strategy that, on the basis of observing $\mathbf{x}$, associates a random variable $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}(\boldsymbol{x})$ as an estimate of $\boldsymbol{\theta}$. It is often convenient to consider situations in which $\boldsymbol{\theta}$ takes on only a single value, which is regarded as deterministic but unknown. In this case, it is common to invoke estimators that produce a deterministic *point estimate* from the datum $\boldsymbol{x}$. Even if the estimate is deterministic in this sense, the estimator $\widehat{\boldsymbol{\theta}}(\boldsymbol{x})$, as a function of the random variable $\mathbf{x}$, is still a random variable.

In this thesis, we generally maintain a bayesian perspective of estimation, and regard parameter estimates as random variables whose distributions result from combining prior information and observed data according to Bayes' Rule. In a few instances, we also employ maximum-likelihood estimation techniques, treating parameters as deterministic but unknown values as discussed above. As mentioned at the start of this chapter, some standard reference texts for a reader who might be interested in a deeper, perhaps more exhaustive, treatment of estimation theory (and the intimately related theory of detection) are [19, 26, 27, 41].

It should be noted that in this treatment, we consider only the estimation of variables that are static, i.e., that are constant in time. The estimation of dynamic variables is considered in Chapter 4.

A Bayesian Approach

A bayesian approach to parameter estimation is, at least in principle, remarkably simple. First, the vector of parameters to be estimated is regarded as a vector-valued random (though unobservable) variable $\boldsymbol{\theta}$ and we presume the conditional density function $f_{\mathbf{x}}(\boldsymbol{x}|\boldsymbol{\theta} = \boldsymbol{\theta})$ fully characterizes the relationship between $\boldsymbol{\theta}$ and the observable random variable $\mathbf{x}$. The marginal probability density function $f_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ – called herein a *prior probability density function* – is too presumed to be a specified, known density function expressing the state of knowledge or ignorance about $\boldsymbol{\theta}$ before $\mathbf{x}$ is observed. Then, given this prior density, the conditional density, and a realization of $\mathbf{x}$, it is possible to compute the density $f_{\boldsymbol{\theta}}(\boldsymbol{\theta}|\mathbf{x} = \boldsymbol{x})$ via application of Bayes' Rule, as in Equation 2.40. This density – called the *posterior probability density function* of $\boldsymbol{\theta}$ – modifies the prior, summarizing what can be said about $\boldsymbol{\theta}$ on the basis of the assumptions made *and the data observed*. It is from this posterior density that inferences about the unknown parameter are made. In what follows, we sometimes refer to the prior density and the posterior density functions simply as the "prior" and the "posterior," respectively.

Frequently, the distribution function of $\mathbf{x}$ admits a density function dependent not only upon the unknown parameter of interest $\boldsymbol{\theta}$ but also on a further so-called *nuisance*, or *incidental* parameter $\boldsymbol{\varphi}$. In the bayesian approach, overall inferences about $\boldsymbol{\theta}$ are completely determined by the posterior distribution of $\boldsymbol{\theta}$, obtained by *marginalizing* the joint posterior density of $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$ over Φ:

$$f_{\boldsymbol{\theta}}(\boldsymbol{\theta}|\boldsymbol{x}) = \int_{\Phi} f_{\boldsymbol{\theta},\boldsymbol{\varphi}}(\boldsymbol{\theta}, \boldsymbol{\varphi}|\boldsymbol{x}) d\boldsymbol{\varphi}, \tag{2.45}$$

where Φ denotes the appropriate region of $\boldsymbol{\varphi}$.

In view of Equation 2.39, we may write the joint posterior density as the product of the conditional (posterior) density of $\boldsymbol{\theta}$ given $\boldsymbol{\varphi}$ and the marginal (posterior) density of $\boldsymbol{\varphi}$ as (omitting the decorative subscripts on the densities for concision)

$$f(\boldsymbol{\theta}, \boldsymbol{\varphi}|\boldsymbol{x}) = f(\boldsymbol{\theta}|\boldsymbol{\varphi}, \boldsymbol{x}) f(\boldsymbol{\varphi}|\boldsymbol{x}),$$

from which it follows the posterior density of $\boldsymbol{\theta}$ may be written

$$f(\boldsymbol{\theta}|\boldsymbol{x}) = \int_{\Phi} f(\boldsymbol{\theta}|\boldsymbol{\varphi}, \boldsymbol{x}) f(\boldsymbol{\varphi}|\boldsymbol{x}) d\boldsymbol{\varphi}, \tag{2.46}$$

in which the marginal (posterior) distribution $f(\boldsymbol{\varphi}|\boldsymbol{x})$ of the parameter $\boldsymbol{\varphi}$ acts as a weight function multiplying the conditional distribution $f(\boldsymbol{\theta}|\boldsymbol{\varphi}, \boldsymbol{x})$ of the parameter of interest.

Alternatively, using Bayes' Rule, we can express the posterior density of $\boldsymbol{\theta}$, conditioned on $\mathbf{x}$ as

$$f(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{f(\boldsymbol{x}|\boldsymbol{\theta}) f(\boldsymbol{\theta})}{\int f(\boldsymbol{x}|\boldsymbol{\theta}) f(\boldsymbol{\theta}) d\boldsymbol{\theta}}, \tag{2.47}$$

where

$$f(\boldsymbol{x}|\boldsymbol{\theta}) = \int_{\Phi} f(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{\varphi}) f(\boldsymbol{\varphi}|\boldsymbol{\theta}) d\boldsymbol{\varphi}. \tag{2.48}$$

If we furthermore assume that $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$ are independent, then Equation 2.48 reduces to

$$f(\boldsymbol{x}|\boldsymbol{\theta}) = \int_{\Phi} f(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{\varphi}) f(\boldsymbol{\varphi}) d\boldsymbol{\varphi}. \tag{2.49}$$

From this perspective, the nuisance parameter $\boldsymbol{\varphi}$ is first marginalized out of the conditional density $f(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{\varphi})$ and then the posterior density is found as usual via application of Bayes' Rule.

We note that the focus of this bayesian approach to estimation presented here is not on determining a single value for the unknown parameter $\boldsymbol{\theta}$ but instead on determining, from a realization of $\mathbf{x}$, its posterior distribution. When a single value for $\boldsymbol{\theta}$ – termed a *point estimate* – is desired, a popular choice is the value of $\boldsymbol{\theta}$ for which its posterior distribution $f_{\boldsymbol{\theta}|\mathbf{x}}(\boldsymbol{\theta}|\boldsymbol{x})$ is maximum, assuming of course a unique maximum exists. This point estimate is known as the *maximum a posteriori estimate* and is often denoted $\widehat{\boldsymbol{\theta}}_{\text{MAP}}$:

$$\widehat{\boldsymbol{\theta}}_{\text{MAP}} \triangleq \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmax}}\, f_{\boldsymbol{\theta}}(\boldsymbol{\theta}|\mathbf{x} = \boldsymbol{x}), \tag{2.50}$$

where Θ is the set of all permissible values of $\boldsymbol{\theta}$.

Evidently requisite to a bayesian approach to estimation is the choice of a prior probability density, or, equivalently, the prior distribution of the unknown parameter $\boldsymbol{\theta}$. Often when

choosing a prior, our knowledge lies somewhere between complete ignorance (e.g.., the use of a *noninformative prior*) and strong prior knowledge (i.e., the use of a *conjugate prior*). Often we'd like to assign a prior probability distribution that is as noninformative as possible, given partial information constraints. This desiderata baselines maximum-entropy probability distributions, which are an integral part to the research presented hereto and, hence, deserve specific attention in this chapter.

Prior Probabilities and the Principle of Maximum Entropy The principle of maximum entropy is a powerful and versatile tool for assigning a probability distribution on the basis of knowing only partial information about the distribution, e.g., its moments, while introducing no new assumptions. It is based on Jaynes' argument [22] that the probability density function consistent with the known information and introduces no unwarranted information is the distribution with maximum (differential) entropy [43]. Any probability density function satisfying the constraints which has smaller entropy will contain more (Shannon) information (less uncertainty), and thus says something stronger than what we are assuming.

Many common probability distributions naturally arise as maximum-entropy distributions under moment constraints. For example the gaussian distribution is the maximum-entropy distribution of a continuous real-valued random variable whose second moment is fixed, and the von Mises-Fisher distribution is the maximum-entropy distribution of a random variable on $\mathbf{S}^n$ whose first moment is fixed. Given that the Shannon differential entropy of a probability measure $f(\boldsymbol{x})$ in a probability space Ω is defined (following from [43]) by

$$H(f) \triangleq -\int_{\Omega} f(\boldsymbol{x}) \log f(\boldsymbol{x}) d\boldsymbol{x},$$

the set up for the maximum-entropy problem looks like

$$\underset{f(\boldsymbol{x})}{\text{maximize}} \; -\int_{\Omega} f(\boldsymbol{x}) \log f(\boldsymbol{x}) d\boldsymbol{x} \;\; \text{subject to} \int_{\Omega} h_i(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x} = c_i, \tag{2.51}$$

for all constraints h_i, $i = 0, 1, \ldots, m$. One constraint is always $h_0(\boldsymbol{x}) = 1$ and $c_0 = 1$; that is, $f(\boldsymbol{x})$ must be a proper probability density. We solve the constrained optimization problem via introducing Lagrangian multipliers, λ_i, of variational calculus. The Lagrangian is

$$\begin{aligned}\mathcal{L}(f(\boldsymbol{x}), \lambda_0, \{\lambda_i\}) = -\int_\Omega f(\boldsymbol{x}) \log f(\boldsymbol{x}) d\boldsymbol{x} + \lambda_0 \left[\int_\Omega f(\boldsymbol{x}) - 1\right] \\ + \sum_{i=1}^{m} \lambda_i \left[\int_\Omega f(\boldsymbol{x}) h_i(\boldsymbol{x}) d\boldsymbol{x} - c_i\right].\end{aligned} \tag{2.52}$$

A solution, if it exists, will be a critical point of this Lagrangian. We have

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial f(\boldsymbol{x})} &= 0 = -\log f(\boldsymbol{x}) - 1 + \lambda_0 + \sum_i \lambda_i h_i(\boldsymbol{x}) \\ \frac{\partial \mathcal{L}}{\partial \lambda_0} &= 0 = \int_\Omega f(\boldsymbol{x}) d\boldsymbol{x} - 1 \\ \frac{\partial \mathcal{L}}{\partial \lambda_i} &= 0 = \int_\Omega f(\boldsymbol{x}) h_i(\boldsymbol{x}) d\boldsymbol{x} - c_i.\end{aligned}$$

For λ_0 and λ_i, $i = 1, \ldots, m$, such that the constraints are satisfied, we can solve the first equation for $f(\boldsymbol{x})$ and retrieve the solution

$$f(\boldsymbol{x}) = \exp\left\{-1 + \lambda_0 + \sum_{i=1}^{m} \lambda_i h_i(\boldsymbol{x})\right\}.$$

Because $\exp\{-1 + \lambda_0\}$ is a constant, we see that the maximum-entropy probability density, subject to the constraints, has the form

$$f(\boldsymbol{x}) \propto \exp\left\{\sum_{i=1}^{m} \lambda_i h_i(\boldsymbol{x})\right\}. \tag{2.53}$$

It is precisely this formulation that leads to the assertion of the maximum-entropy characterizations of the gaussian and von Mises-Fisher distributions stated above.

The Method of Maximum Likelihood

In light of Bayes' Rule, the maximum *a posteriori* estimate in Equation 2.50 is given by the vector $\boldsymbol{\theta} \in \Theta$ for which the product $f_{\mathbf{x}}(\boldsymbol{x}|\boldsymbol{\theta} = \boldsymbol{\theta}) f_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ is maximum. Though for many observation models arising in practice, assigning a prior distribution to the parameter regarded

as a random variable $\boldsymbol{\theta}$ might be a formidable, if not impossible, task. For such models, a prevailing approach is to take, as point estimates, any values $\boldsymbol{\theta}$ in Θ that maximize only the conditional density of $\boldsymbol{\theta}$, given an observation $\boldsymbol{x}$ of $\mathbf{x}$. Since $f_{\mathbf{x}}(\boldsymbol{x}|\boldsymbol{\theta} = \boldsymbol{\theta})$ as a function of $\boldsymbol{\theta}$ is often termed the *likelihood function*, these point estimates are called *maximum-likelihood estimates*. Maximizing the likelihood function defines the method of maximum-likelihood and, if we assume a unique global maximum exists, defines the maximum-likelihood estimate, denoted $\widehat{\boldsymbol{\theta}}_{\text{ML}}$:

$$\widehat{\boldsymbol{\theta}}_{\text{ML}} \triangleq \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmax}}\, f_{\mathbf{x}}(\boldsymbol{x}|\boldsymbol{\theta} = \boldsymbol{\theta}), \tag{2.54}$$

where, again, Θ is set of all permissible values of $\boldsymbol{\theta}$.

We remark that the maximum-likelihood estimate yields the same point estimate as the maximum *a posteriori* estimate when $\boldsymbol{\theta}$ is assumed *a priori* to be uniformly distributed in its range (i.e., when the prior probability density is constant on its range), assuming, of course, it's possible to construct a uniform distribution on Θ.

Although the method of maximum likelihood is not based on any clearly defined optimum considerations, it has been very successful in leading to satisfactory procedures in many practical problems. For wide classes of problems, maximum-likelihood procedures actually possess various optimum properties in the asymptotic regime. Most notably, as the number of independent realizations of $\mathbf{x}$ increases without bound, sequences of maximum-likelihood estimates converge to the true parameter value (in probability). This is a property known as *consistency*. Proof of consistency and other aspects of the behavior of maximum-likelihood estimates are treated fully in [19, 33].

Maximum-Likelihood Estimation: An Important Example Suppose that the n-vector $\boldsymbol{x}$ is a realization of a gaussian distributed, vector-valued random variable $\mathbf{x}$ on $\mathbb{C}$, with the mean equal to $\boldsymbol{Sa}$ and the covariance equal to $\sigma^2\mathbb{I}_n$.[13] Presumably, the columns

[13]We consider complex random variables as pairs of real random variables comprising their real and imaginary parts. The distribution of a complex random variable is interpreted as the joint distribution of two real-valued random variables.

of the n-by-k matrix $\boldsymbol{S}$ form an orthonormal basis in $\mathbb{C}^n$ and the elements of the vector $\boldsymbol{a}$ are unknown complex values. We assume σ^2 is *known*. The probability density function for $\mathbf{x}$, conditioned on $\mathbf{S} = \boldsymbol{S}$ and $\mathsf{a} = \boldsymbol{a}$ is given by

$$f_{\mathbf{x}}(\boldsymbol{x}|\mathbf{S} = \boldsymbol{S}, \mathbf{a} = \boldsymbol{a}) = \left(\pi\sigma^2\right)^{-n} \exp\left\{-\frac{1}{\sigma^2}\left(\boldsymbol{x} - \boldsymbol{S}\boldsymbol{a}\right)^\dagger \left(\boldsymbol{x} - \boldsymbol{S}\boldsymbol{a}\right)\right\}, \tag{2.55}$$

where $\dagger$ denotes the hermitian transpose. In assuming the absence of a suitable prior for $\mathbf{a}$, computing its maximum-likelihood estimate is an elementary exercise for a student in her typical graduate electrical engineering coursework:

$$\widehat{\boldsymbol{a}}_{\text{ML}} = \boldsymbol{S}^\dagger \boldsymbol{x},$$

that is, the maximum-likelihood estimate for each component a^j, $j = 1, \ldots, k$, is the scalar projection of the data vector $\boldsymbol{x}$ in the direction of the subspace defined by the columns of $\boldsymbol{S}$. It's worth noting that the maximum-likelihood estimate for $\boldsymbol{a}$ is unique due to the unimodality of the assumed gaussian distribution of $\mathbf{x}$. In assuming too the absence of a suitable prior for $\mathbf{S}$, we can compute its maximum-likelihood estimate, which in contrast to the maximum-likelihood estimate for $\boldsymbol{a}$, is significantly more rigorous. We provide a summary here, though the estimate is presented in detail in [44]. We start with substituting the maximum-likelihood estimate of $\boldsymbol{a}$ into the likelihood function, yielding a *generalized* likelihood:

$$\begin{aligned} f(\boldsymbol{S}|\mathbf{x} = \boldsymbol{x}) &= \left(\pi\sigma^2\right)^{-n} \exp\left\{-\frac{1}{\sigma^2}\left(\boldsymbol{x}^\dagger\boldsymbol{x} - \boldsymbol{x}^\dagger\boldsymbol{S}\boldsymbol{S}^\dagger\boldsymbol{x}\right)\right\} \\ &= \left(\pi\sigma^2\right)^{-n} \exp\left\{-\frac{1}{2\sigma^2}\left[\operatorname{tr}(\boldsymbol{W}) - \operatorname{tr}(\boldsymbol{W}\boldsymbol{P}_{\boldsymbol{S}})\right]\right\}, \end{aligned} \tag{2.56}$$

where $\boldsymbol{P}_{\boldsymbol{S}} = \boldsymbol{S}\boldsymbol{S}^\dagger$ is the projection onto the subspace defined by the columns of $\boldsymbol{S}$ and $\boldsymbol{W} = \boldsymbol{x}\boldsymbol{x}^\dagger$. As σ^2 is assumed to be known, it can be omitted from the maximization. As we'll see in Chapters 3 and 5, rank-k projection operators are in one-to-one correspondence with points on the Grassmann manifold of all k-dimensional subspaces of an n-dimensional vector space. Hence maximizing the generalized likelihood in Equation 2.56 for the maximum-likelihood estimate of $\boldsymbol{S}$ is achieved by maximizing $\operatorname{tr}(\boldsymbol{W}\boldsymbol{P}_{\boldsymbol{S}})$ over $\mathbf{G}(k, n)$. Letting $\boldsymbol{W} =$

$\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^\dagger$, where $\boldsymbol{U}$ is a unitary matrix and $\boldsymbol{D}$ a diagonal matrix of non-increasing, real eigenvalues, we observe

$$\begin{aligned}\underset{\boldsymbol{P_S}\in\mathbf{G}(k,n)}{\operatorname{argmax}}\ \operatorname{tr}(\boldsymbol{W}\boldsymbol{P_S}) &= \underset{\boldsymbol{P_S}\in\mathbf{G}(k,n)}{\operatorname{argmax}}\ \operatorname{tr}(\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^\dagger\boldsymbol{P_S}) \\ &= \underset{\boldsymbol{P}\in\mathbf{G}(k,n)}{\operatorname{argmax}}\ \operatorname{tr}(\boldsymbol{D}\boldsymbol{P}).\end{aligned}$$

A clever application of the Schur-Horn Theorem [20], which allows for the determination of the possible diagonal entries of a hermitian matrix with a fixed set of eigenvalues, gives the unique maximum of $\operatorname{tr}(\boldsymbol{D}\boldsymbol{P})$ when

$$\boldsymbol{P} = \boldsymbol{P_0} = \begin{bmatrix} \mathbb{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

It follows that $\operatorname{tr}(\boldsymbol{W}\boldsymbol{P_s})$ is maximized when

$$\boldsymbol{P_s} = \boldsymbol{U}\boldsymbol{P_0}\boldsymbol{U}^\dagger.$$

The maximum-likelihood estimate of $\boldsymbol{s}$ is therefore given by

$$\widehat{\boldsymbol{s}}_{\mathrm{ML}} = \boldsymbol{U}\begin{bmatrix} \mathbb{I}_k \\ \mathbf{0} \end{bmatrix}, \tag{2.57}$$

which results in the k eigenvectors that correspond to the k largest eigenvalues of $\boldsymbol{x}\boldsymbol{x}^\dagger$. Figure 2.2 below shows an example of five realizations of $\mathbf{x}$ in $\mathbb{R}^2$, plotted as crosses (x), generated from a one-dimensional, real subspace plotted in red (dashed). The maximum-likelihood estimate of the subspace from the data is plotted in blue (solid).

We note here that the reader will find a resemblance in the assumptions for this example the mathematical model fit for the motivating application underpinning this thesis. As detailed more fully in Chapter 4, noisy sensor data, generated from a particular k-dimensional subspace, will present as (independent and identically distributed) realizations of a gaussian random vector whose mean is at least partially described by the subspace. When we represent an orthonormal basis for the *unknown* subspace with a matrix $\mathbf{S}$, the setup is

squarely matched to this example. Rather than drawing on its maximum-likelihood estimate, we regard $\mathbf{S}$ as a random variable on the (complex) Grassmann manifold and treat estimation using a bayesian framework, where a suitable prior distribution defined on the manifold is a crucial factor. Specializing to the important, special case where the unknown subspace is one-dimensional, the next chapter develops the requisite geometric objects and uses them to derive maximum-entropy probability densities on projective space that will be used for prior densities in the context of bayesian inference in Chapter 4.

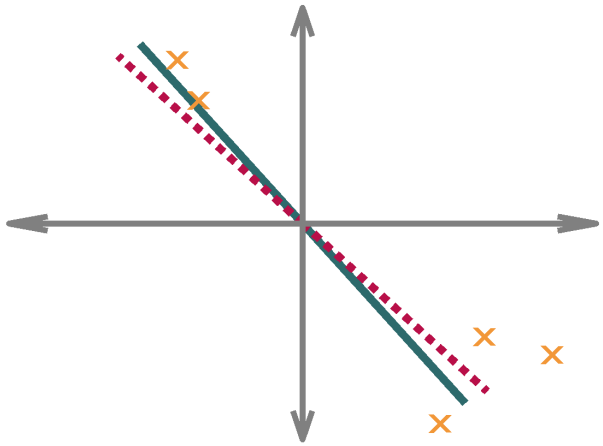

Figure 2.2: Maximum-likelihood estimate of a one-dimensional, real subspace in $\mathbb{R}^2$ from data.

Chapter 3

PROJECTIVE SPACES

While the class of subspace-based signal processing problems we are interested in involve k-dimensional subspaces of an ambient n-dimensional vector space, i.e., elements of the Grassmann manifold, the theory presented in the next two chapters is specialized to the important, special case where the signal subspace is one-dimensional (the ambient dimension n remains general) and the Grassmann manifold is known as projective space. The first half of the present chapter is dedicated to real projective space; the latter highlights some key differences in the extension to complex projective space.

3.1 Real Projective Space

We begin with a formal, general definition.

Definition 3.1. ***Real projective*** n***-space*** *(also referred to as either* ***projective*** n***-space*** *or* ***projective space****) is the set of all one-dimensional linear subspaces in* $\mathbb{R}^{n+1}$*. We denote this space* $\mathbb{RP}^n$*.*

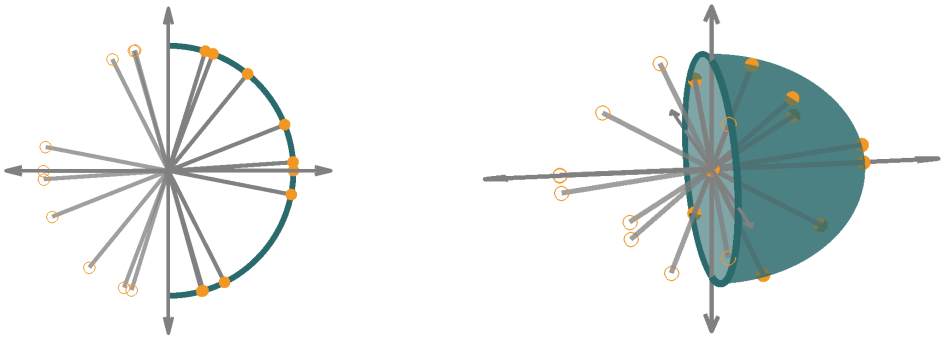

Figure 3.1: Illustrations of real projective spaces $\mathbb{RP}^1$ in $\mathbb{R}^2$ (left) and $\mathbb{RP}^2$ in $\mathbb{R}^3$ (right).

Often $\mathbb{RP}^n$ is presented as the unit n-sphere $\mathbf{S}^n$ in $\mathbb{R}^{n+1}$ modulo the identification of antipodal points, as illustrated for the special cases $\mathbb{RP}^1$ and $\mathbb{RP}^2$ in Figure 3.1. Indeed,

much intuition for integrating and recursively estimating on $\mathbb{RP}^n$ comes from implementing a naïve approach on n-spheres and orthogonal groups.

Following the much more general treatment of defining the Grassmann manifold in Chapter 5, we present local coordinates on $\mathbb{RP}^n$ from three distinct perspectives, each of which will become useful as we move to develop differential geometric objects and statistical analyses throughout this chapter and the document.

An Orthonormal Basis Perspective For $\boldsymbol{x} \in \mathbb{RP}^n$, let $\boldsymbol{x}_{\boldsymbol{z}}$ be the unit vector of the form

$$\boldsymbol{x}_{\boldsymbol{z}} = \begin{bmatrix} 1 \\ \boldsymbol{z} \end{bmatrix} \left(1 + \boldsymbol{z}^\top \boldsymbol{z}\right)^{-1/2}, \tag{3.1}$$

where $\boldsymbol{z} = \begin{bmatrix} z^1 & z^2 & \cdots & z^n \end{bmatrix}^\top$ is an element of $\mathbb{R}^n$. It is important to note that this representation, which we identify with the point $\boldsymbol{x} \in \mathbb{RP}^n$, is an $(n+1)$-dimensional vector whose first coordinate is positive and that these coordinates implicitly characterize an atlas comprising a single chart covering almost all of $\mathbb{RP}^n$; the set of excluded points has measure zero. There is a corresponding orthonormal basis for the orthogonal complement of $\boldsymbol{x}_{\boldsymbol{z}}$ in $\mathbb{R}^{n+1}$, comprising the columns of the $(n+1)$-by-n matrix we denote $\boldsymbol{x}_{\boldsymbol{z}}^\perp$. In coordinates,

$$\boldsymbol{x}_{\boldsymbol{z}}^\perp = \begin{bmatrix} -\boldsymbol{z}^\top \\ \mathbb{I}_n \end{bmatrix} \left(\mathbb{I}_n + \boldsymbol{z}\boldsymbol{z}^\top\right)^{-1/2}. \tag{3.2}$$

An Orthogonal Projector Perspective Associated uniquely to each one-dimensional subspace of $\mathbb{R}^n$ is a rank-one orthogonal projection operator. This one-to-one correspondence gives rise to the following proposition.

Proposition 3.1. *Let n be fixed positive integer. Let $\mathbb{P}^n$ define the set of all rank-one orthogonal projection operators on $\mathbb{R}^n$. That is,*

$$\mathbb{P}^n \triangleq \left\{ \boldsymbol{P} \in \mathbb{R}^{n \times n} \middle| \boldsymbol{P}^\top = \boldsymbol{P}^2 = \boldsymbol{P},\ \text{rank}(\boldsymbol{P}) = 1 \right\}. \tag{3.3}$$

There exists a bijection between real projective space $\mathbb{RP}^n$ and $\mathbb{P}^n$.

For proof of this proposition, we turn an interested reader to [18], wherein a natural bijection between $\mathbb{RP}^n$ and $\mathbb{P}^n$ is constructed. Moreover, with respect to a designated topology, Helmke proves his bijection is a diffeomorphism.

The standard form representing a one-dimensional subspace of $\mathbb{R}^n$ in Equation 3.1 has the following associated standard form for a rank-one, orthogonal projection operator:

$$\boldsymbol{P}_{\boldsymbol{x}_{\boldsymbol{z}}} = \boldsymbol{x}_{\boldsymbol{z}}\boldsymbol{x}_{\boldsymbol{z}}^{\top} = \left(1 + \boldsymbol{z}^{\top}\boldsymbol{z}\right)^{-1} \begin{bmatrix} 1 & \boldsymbol{z}^{\top} \\ \boldsymbol{z} & \boldsymbol{z}\boldsymbol{z}^{\top} \end{bmatrix}.$$

A Homogeneous Perspective One last presentation of projective space we'll include is as a homogeneous manifold by letting $\mathbf{O}(n+1)$ act transitively on $\mathbb{RP}^n$. A consequence of the so-called Orbit-Stabilizer Theorem (see [7]) yields the following one-to-one correspondence:

$$\mathbb{RP}^n \cong \mathbf{O}(n+1)/\left(\mathbf{O}(1) \times \mathbf{O}(n)\right). \tag{3.4}$$

A special case of the Quotient Manifold Theorem (see [32]) will ensure that, with this identification, $\mathbb{RP}^n$ is indeed a smooth manifold. To the astute reader (or to a reader who has read ahead!), this presentation is a specialization of the real Grassmann manifold $\mathbf{G}(k,n)$ to the case $k = 1$. A detailed mathematical justification for the representation in Equation 3.4 is included in Chapter 5. While the quotient manifold presentation may seem a bit unwieldy, it gives rise to a differentiable structure so that $\mathbb{RP}^n$ – defined only as a set in Definition 3.1 – is a true differentiable manifold of dimension n. The quotient manifold presentation will too avail ourselves of the particularly well-understood geometry of Lie groups and homogeneous manifolds, which may be used to develop specific geometric objects (e.g., tangent vectors, metrics, and geodesics). In fact, Chapter 5 will exploit this quotient space representation for more general grassmannians to introduce a riemannian metric and rigorously develop formulae for tangent vectors and geodesics.

3.1.1 The Differential Form for the Invariant Measure

For the remaining treatment of $\mathbb{RP}^n$ in this chapter, we return to the coordinate system defined by Equations 3.1 and 3.2: for $\boldsymbol{x} \in \mathbb{RP}^n$, we let $\boldsymbol{x}_{\boldsymbol{z}}$ be the unit vector of the form

$$\boldsymbol{x}_{\boldsymbol{z}} = \begin{bmatrix} 1 \\ \boldsymbol{z} \end{bmatrix} \left(1 + \boldsymbol{z}^\top \boldsymbol{z}\right)^{-1/2}, \tag{3.1 revisited}$$

where $\boldsymbol{z} = \begin{bmatrix} z^1 & z^2 & \cdots & z^n \end{bmatrix}^\top$ is an element of $\mathbb{R}^n$. And we let the corresponding orthonormal basis for the orthogonal complement of $\boldsymbol{x}_{\boldsymbol{z}}$ in $\mathbb{R}^{n+1}$ comprise the columns of the $(n+1)$-by-n matrix denoted $\boldsymbol{x}_{\boldsymbol{z}}^\perp$:

$$\boldsymbol{x}_{\boldsymbol{z}}^\perp = \begin{bmatrix} -\boldsymbol{z}^\top \\ \mathbb{I}_n \end{bmatrix} \left(\mathbb{I}_n + \boldsymbol{z}\boldsymbol{z}^\top\right)^{-1/2}. \tag{3.2 revisited}$$

Observing the fact that real projective n-space is a special case of the real Grassmann manifold avails ourselves of the classical work of James [21]. By analogy with his derivation of the invariant measure for the real Grassmann manifold, the invariant measure on $\mathbb{RP}^n$ can be defined for points $\boldsymbol{x}_{\boldsymbol{z}}$ in a neighborhood of a point $\boldsymbol{x} \in \mathbb{RP}^n$:

$$\omega_n(\boldsymbol{x}_{\boldsymbol{z}}) \triangleq \bigwedge_{i=1}^{n} \boldsymbol{b}_i^\top d\boldsymbol{x}_{\boldsymbol{z}}. \tag{3.5}$$

In this expression, $\boldsymbol{x}_{\boldsymbol{z}}$ is a unit vector in the subspace defined by $\boldsymbol{x}$; $d\boldsymbol{x}_{\boldsymbol{z}}$ is the total derivative of $\boldsymbol{x}_{\boldsymbol{z}}$ (i.e., the differential); and $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ are the orthonormal $(n+1)$-dimensional column vectors comprising the matrix $\boldsymbol{x}_{\boldsymbol{z}}^\perp$. This is a system of vectors that can only be constructed locally. For the proof that Equation 3.5 is indeed a differential form that defines the invariant measure, we refer the reader to James [21], specifically section four.

Using the coordinate system in Equations 3.1 and 3.2 and introducing $\alpha = \left(1 + \boldsymbol{z}^\top \boldsymbol{z}\right)$ for concision,

$$d\boldsymbol{x}_{\boldsymbol{z}} = \frac{\partial \boldsymbol{x}_{\boldsymbol{z}}}{\partial \boldsymbol{z}} d\boldsymbol{z} = \alpha^{-3/2} \begin{bmatrix} -\boldsymbol{z}^\top \\ \alpha\mathbb{I}_n - \boldsymbol{z}\boldsymbol{z}^\top \end{bmatrix} \begin{bmatrix} dz^1 \\ \vdots \\ dz^n \end{bmatrix},$$

from which it follows

$$\omega_n(\boldsymbol{x}) = \alpha^{-3n/2}\boldsymbol{b}_1^\top \begin{bmatrix} -\boldsymbol{z}^\top \\ \alpha\mathbb{I}_n - \boldsymbol{z}\boldsymbol{z}^\top \end{bmatrix} \begin{bmatrix} dz^1 \\ \vdots \\ dz^n \end{bmatrix} \wedge \cdots \wedge \boldsymbol{b}_n^\top \begin{bmatrix} -\boldsymbol{z}^\top \\ \alpha\mathbb{I}_n - \boldsymbol{z}\boldsymbol{z}^\top \end{bmatrix} \begin{bmatrix} dz^1 \\ \vdots \\ dz^n \end{bmatrix}. \tag{3.6}$$

Let $\boldsymbol{t}_i$ be the n-dimensional (row) vector given by

$$\boldsymbol{t}_i = \boldsymbol{b}_i^\top \begin{bmatrix} -\boldsymbol{z}^\top \\ \alpha\mathbb{I}_n - \boldsymbol{z}\boldsymbol{z}^\top \end{bmatrix} = \begin{bmatrix} t_i^1 & t_i^2 & \cdots t_i^n \end{bmatrix},$$

so that, using Theorem 2.1,

$$\begin{aligned} \omega_n(\boldsymbol{x}) &= \alpha^{-3n/2} \sum_{j=1}^n t_1^j dz_j \wedge \cdots \wedge \sum_{j=1}^n t_n^j dx_j \\ &= \alpha^{-3n/2} \det \begin{bmatrix} t_1^1 & \cdots & t_n^1 \\ \vdots & \ldots & \vdots \\ t_1^n & \cdots & t_n^n \end{bmatrix} dz^1 \wedge \cdots \wedge dz^n. \end{aligned} \tag{3.7}$$

By elementary linear algebra,

$$\begin{aligned} \boldsymbol{T} &= \begin{bmatrix} -\boldsymbol{z}^\top \left(\mathbb{I}_n + \boldsymbol{z}\boldsymbol{z}^\top\right)^{-1/2} \\ \left(\mathbb{I}_n + \boldsymbol{z}\boldsymbol{z}^\top\right)^{-1/2} \end{bmatrix}^\top \begin{bmatrix} -\boldsymbol{z}^\top \\ \alpha\mathbb{I}_n - \boldsymbol{z}\boldsymbol{z}^\top \end{bmatrix} \\ &= \left(\mathbb{I}_n + \boldsymbol{z}\boldsymbol{z}^\top\right)^{-1/2} \boldsymbol{z}\boldsymbol{z}^\top + \left(\mathbb{I}_n + \boldsymbol{z}\boldsymbol{z}^\top\right)^{-1/2} \left(\alpha\mathbb{I}_n - \boldsymbol{z}\boldsymbol{z}^\top\right) \\ &= \alpha \left(\mathbb{I}_n + \boldsymbol{z}\boldsymbol{z}^\top\right)^{-1/2}. \end{aligned}$$

Because the spectrum of the matrix $\boldsymbol{z}\boldsymbol{z}^\top$ is $\left\{\boldsymbol{z}^\top\boldsymbol{z}, 0, \ldots, 0\right\}$, the spectrum of the matrix $\mathbb{I}_n + \boldsymbol{z}\boldsymbol{z}^\top$ is $\left\{1 + \boldsymbol{z}^\top\boldsymbol{z}, 1, \ldots, \ 1\right\}$, and it follows that

$$\begin{aligned} \omega_n(\boldsymbol{x}) &= \alpha^{-3n/2} \det \left(\alpha \left(\mathbb{I}_n + \boldsymbol{z}\boldsymbol{z}^\top\right)^{-1/2}\right) dz^1 \wedge \cdots \wedge dz^n \\ &= \alpha^{-n/2} \det \left(\mathbb{I}_n + \boldsymbol{z}\boldsymbol{z}^\top\right)^{-1/2} dz^1 \wedge \cdots \wedge dz^n \\ &= \left(1 + \boldsymbol{z}^\top\boldsymbol{z}\right)^{-n/2} \left(1 + \boldsymbol{z}^\top\boldsymbol{z}\right)^{-1/2} dz^1 \wedge \cdots \wedge dz^n \\ &= \left(1 + \boldsymbol{z}^\top\boldsymbol{z}\right)^{-(n+1)/2} dz^1 \wedge \cdots \wedge dz^n. \end{aligned} \tag{3.8}$$

Integrating the Differential Form

It is apparently well known that the volume of real projective n-space is equal to one half the volume of the n-sphere, but the derivation of the volume via direction integration of an exterior differential form over the entire space is, to our knowledge, not readily available in the literature. We derive the volume explicitly via the following direct transformation to *hyperspherical coordinates*. Given

$$\boldsymbol{x}_{\boldsymbol{z}} = \begin{bmatrix} 1 \\ \boldsymbol{z} \end{bmatrix} \left(1 + \boldsymbol{z}^\top \boldsymbol{z}\right)^{-1/2},$$

with $\boldsymbol{z} = \begin{bmatrix} z^1 & \cdots & z^n \end{bmatrix}^\top$, consider the transformation

$$\begin{aligned} z^k &= \tan\theta^1 \prod_{i=1}^{k} \sin\theta^i \cos\theta^{k+1} \qquad k < n \\ z^n &= \tan\theta^1 \prod_{i=1}^{n} \sin\theta^i \end{aligned} \tag{3.9}$$

with $\theta^1 \in (-\pi/2, \pi/2)$ and $\theta^i \in (0, \pi)$ for $i > 1$. As is worked out in detail in Appendix B.1, the transformation law for the volume form on $\mathbb{RP}^n$ is given by

$$\left(1 + \boldsymbol{z}^\top \boldsymbol{z}\right)^{-(n+1)/2} dz^1 \wedge \cdots \wedge dz^n = \left| \prod_{j+1}^{n} \sin^{n-j}\left(\theta^j\right) \right| d\theta^1 \wedge \cdots \wedge d\theta^n. \tag{3.10}$$

The volume of $\mathbb{RP}^n$ is thus computed as (see Appendix A.1.1)

$$\begin{aligned} \mathrm{Vol}(\mathbb{RP}^n) &= \int_{\mathbb{RP}^n} \omega_n(\boldsymbol{x}_{\boldsymbol{z}}) \\ &= \int_{\mathbb{R}^n} \left(1 + \boldsymbol{z}^\top \boldsymbol{z}\right)^{-(n+1)/2} dz^1 dz^2 \cdots dz^n \\ &= \int_0^{\pi} \cdots \int_0^{\pi} \int_{-\pi/2}^{\pi/2} \left| \prod_{j=1}^{n} \sin^{n-j}(\theta^j) \right| d\theta^1 d\theta^2 \cdots d\theta^n \\ &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \left|\sin^{n-1}\left(\theta^1\right)\right| d\theta^1 \prod_{j=2}^{n} \int_0^{\pi} \sin^{n-j}\left(\theta^j\right) d\theta^j \end{aligned}$$

$$
\begin{aligned}
&= \frac{\sqrt{\pi}\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)} \prod_{j=2}^{n} \frac{\sqrt{\pi}\Gamma\left(\frac{1+n-j}{2}\right)}{\Gamma\left(\frac{2+n-j}{2}\right)} \\
&= \frac{\pi^{n/2}\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)\Gamma\left(\frac{n}{2}\right)} \\
&= \frac{\pi^{\frac{n+1}{2}}}{\Gamma\left(\frac{n+1}{2}\right)},
\end{aligned}
$$

which is exactly half the volume of the n-sphere, as expected.

3.1.2 A Probability Distribution

The distribution developed and studied in this section plays a key role in bayesian inference on real projective space. Its importance lies in a number of characteristics, but principally in the fact that it has the same statistical role as the multivariate normal distribution in n-dimensional Euclidean space or the von Mises and von Mises-Fisher distributions on the n-sphere.

Theorem 3.1. *For a fixed non-negative, real number κ and a fixed point $\boldsymbol{\mu}$ in $\mathbb{RP}^n$, the continuous probability density function on $\mathbb{RP}^n$ with first moment equal to $\rho\boldsymbol{\mu} \in \mathbb{R}^{n+1}$ which maximizes entropy among all such densities (on $\mathbb{RP}^n$ with first moment $\rho\boldsymbol{\mu}$) is given by*

$$
f_{\mathbf{x}}(\boldsymbol{x}|\boldsymbol{\mu}, \kappa) = c_n(\kappa) e^{\kappa|\boldsymbol{\mu}^\top \boldsymbol{x}|}, \tag{3.11}
$$

where $c_n(\kappa)$ is chosen so that $e^{\kappa|\boldsymbol{\mu}^\top \boldsymbol{x}|}$ integrates, with respect to the volume form, to unity over the whole space and κ so that $\mathrm{E}[\boldsymbol{x}] = \rho(\kappa)\boldsymbol{\mu}$.

Proof. The constrained optimization problem to maximize the (Shannon) differential entropy on real projective space, subject to a constrained first moment is given by[1]

$$
\begin{aligned}
&\underset{f_{\mathbf{x}}(\boldsymbol{x})}{\text{maximize}} \; -\int_{\mathbb{RP}^n} f_{\mathbf{x}}(\boldsymbol{x}) \log f_{\mathbf{x}}(\boldsymbol{x}) d\boldsymbol{x}, \\
&\text{subject to} \int_{\mathbb{RP}^n} f_{\mathbf{x}}(\boldsymbol{x}) d\boldsymbol{x} = 1 \text{ and } \int_{\mathbb{RP}^n} \boldsymbol{x} f_{\mathbf{x}}(\boldsymbol{x}) d\boldsymbol{x} = \rho(\kappa)\boldsymbol{\mu}.
\end{aligned}
$$

[1] Against better judgement, we use $d\boldsymbol{x}$ and $\omega_n(\boldsymbol{x})$ interchangeably to represent the invariant measure on $\mathbb{RP}^n$.

Solving this optimization problem is tantamount to solving the following:

$$\begin{aligned}&\underset{f_{\mathbf{x}}(\boldsymbol{x})}{\text{maximize}} \; -\int_{\mathbb{RP}^n} f_{\mathbf{x}}(\boldsymbol{x}) \log f_{\mathbf{x}}(\boldsymbol{x}) d\boldsymbol{x}, \\ &\text{subject to} \int_{\mathbb{RP}^n} f_{\mathbf{x}}(\boldsymbol{x}) d\boldsymbol{x} = 1 \text{ and } \int_{\mathbb{RP}^n} \left|\boldsymbol{\mu}^\top \boldsymbol{x}\right| f_{\mathbf{x}}(\boldsymbol{x}) d\boldsymbol{x} = \rho(\kappa),\end{aligned}$$

where the absolute value bars in the second constraint ensure the algebraic manipulation introduces no extraneous solutions on the quotient manifold. Applying Equation 2.53 to the constraints above results in the form of the maximum-entropy probability distribution on $\mathbb{RP}^n$, subject to a fixed first moment, given by

$$f_{\mathbf{x}}(\boldsymbol{x}|\boldsymbol{\mu}, \kappa) = c_n(\kappa) e^{\kappa|\boldsymbol{\mu}^\top \boldsymbol{x}|},$$

as desired. ■

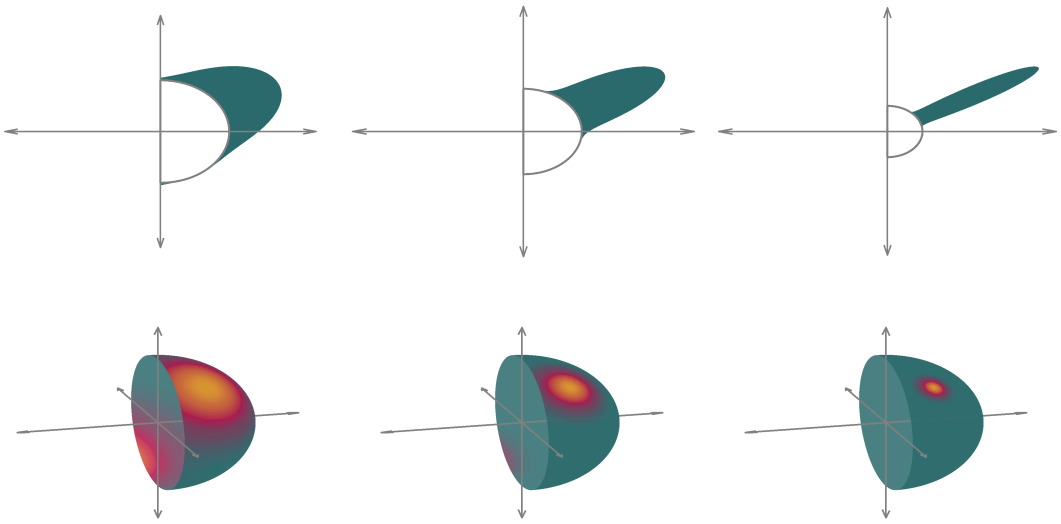

Figure 3.2: Illustration of the maximum-entropy probability distributions on $\mathbb{RP}^1$ (top) and on $\mathbb{RP}^2$ (bottom) with increasing concentration: from left to right, $\kappa = 5, 20, 100$.

The density $f_{\mathbf{x}}(\boldsymbol{x}|\boldsymbol{\mu}, \kappa)$ is parameterized by the *mean direction* $\boldsymbol{\mu} \in \mathbb{RP}^n$, and the *concentration factor* $\kappa \in \mathbb{R}^+$, so called because it characterizes how strongly elements on $\mathbb{RP}^n$ drawn according to $f_{\mathbf{x}}(\boldsymbol{x}|\boldsymbol{\mu}, \kappa)$ are concentrated about the mean direction $\boldsymbol{\mu}$. Larger values of κ imply strong concentration about the mean direction. In particular, as κ tends to 0, the distribution reduces to the uniform distribution on $\mathbb{RP}^n$, and as κ tends to infinity, the

density tends to a point density. Figure 3.2 illustrates this probability density for the cases $n = 1$ and $n = 2$ for a fixed fixed values of $\boldsymbol{\mu}$ in each space and increasing values of κ. As noted above, this maximum-entropy density on $\mathbb{RP}^n$ is analogous to the widely studied von Mises and von Mises-Fisher distributions on $\mathbf{S}^n$ [23, 34, 35].

To derive the value of the normalizing constant $c_n(\kappa)$, we use the fact that

$$\int_{\mathbb{RP}^n} c_n(\kappa) e^{\kappa |\boldsymbol{\mu}^\top \boldsymbol{x}|} d\boldsymbol{x} = 1,$$

where $d\boldsymbol{x}$ represents the differential form for the invariant measure on $\mathbb{RP}^n$. To evaluate this integral, let us first make the transformation $\boldsymbol{y} = \boldsymbol{Q}\boldsymbol{x}$, where the first component of $\boldsymbol{y}$ is equal to $\boldsymbol{\mu}^\top \boldsymbol{x}$ and $\boldsymbol{Q}$ is orthogonal (the first row of $\boldsymbol{Q}$ will be $\boldsymbol{\mu}^\top$). The matrix $\boldsymbol{Q}$ is chosen to be orthogonal so that $d\boldsymbol{x} = d\boldsymbol{y}$. We rewrite the equation above as

$$\frac{1}{c_n(\kappa)} = \int_{\mathbb{RP}^n} e^{\kappa |\boldsymbol{\mu}^\top \boldsymbol{x}|} d\boldsymbol{x} = \int_{\mathbb{R}^n} e^{\kappa \left(1 + \boldsymbol{z}^\top \boldsymbol{z}\right)^{-\frac{1}{2}}} \frac{d\boldsymbol{z}}{\left(1 + \boldsymbol{z}^\top \boldsymbol{z}\right)^{(n+1)/2}}.$$

Under a second transformation to hyperspherical coordinates,

$$\begin{aligned}
\frac{1}{c_n(\kappa)} &= \int_0^{\pi} \cdots \int_0^{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} e^{\kappa \cos\theta^1} \left| \prod_{j=1}^{n} \sin^{n-j}(\theta^j) \right| d\theta^1 d\theta^2 \cdots d\theta^n \\
&= \int_0^{\pi} \cdots \int_0^{\pi} \prod_{j=2}^{n} \sin^{n-j}\left(\theta^j\right) d\theta^2 \cdots d\theta^n \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} e^{\kappa \cos\theta^1} \left|\sin^{n-1}\left(\theta^1\right)\right| d\theta^1 \\
&= \prod_{j=2}^{n} \int_0^{\pi} \sin^{n-j}\left(\theta^j\right) d\theta^j \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} e^{\kappa \cos\theta^1} \left|\sin^{n-1}\left(\theta^1\right)\right| d\theta^1 \\
&= 2 \prod_{j=2}^{n} \int_0^{\pi} \sin^{n-j}\left(\theta^j\right) d\theta^j \int_0^{\pm\frac{\pi}{2}} e^{\kappa \cos\theta^1} \sin^{n-1}\left(\theta^1\right) d\theta^1.
\end{aligned}$$

After application of results proved in Appendix A, simplification yields

$$\frac{1}{c_n(\kappa)} = \pi^{(n+1)/2} \left(\frac{2}{\kappa}\right)^{(n-1)/2} \left[I_{\frac{n-1}{2}}(\kappa) + L_{\frac{n-1}{2}}(\kappa) \right],$$

so that

$$c_n(\kappa) = \frac{\left(\frac{\kappa}{2}\right)^{\frac{n-1}{2}}}{\pi^{\frac{n+1}{2}} \left[I_{\frac{n-1}{2}}(\kappa) + L_{\frac{n-1}{2}}(\kappa) \right]}, \tag{3.12}$$

where $I_\nu(x)$ is the modified Bessel function of the first kind and $L_\nu(x)$ is the related modified Struve function [1]. This value of $c_n(\kappa)$ verifies that Equation 3.11 is indeed a proper probability density.

To see the relationship between ρ and κ, we simply use the fact that

$$c_n(\kappa) \int_{\mathbb{RP}^n} \left| \boldsymbol{\mu}^\top \boldsymbol{x} \right| e^{\kappa |\boldsymbol{\mu}^\top \boldsymbol{x}|} d\boldsymbol{x} = \rho.$$

Applying the same two coordinate transformations as in the derivation of the normalizing constant above will yield

$$\rho(\kappa) = \frac{\frac{2}{n} + \left(\frac{2}{\kappa}\right)^{\frac{n-1}{2}} \sqrt{\pi} \Gamma\left(\frac{n}{2}\right) \left[I_{\frac{n+1}{2}}(\kappa) + L_{\frac{n+1}{2}}(\kappa) \right]}{\left(\frac{2}{\kappa}\right)^{\frac{n-1}{2}} \sqrt{\pi} \Gamma\left(\frac{n}{2}\right) \left[I_{\frac{n-1}{2}}(\kappa) + L_{\frac{n-1}{2}}(\kappa) \right]}. \tag{3.13}$$

Using the results regarding the limiting behavior of $I_\nu(x)$ and $L_\nu(x)$ recorded in Appendix A, together with elementary algebra,

$$\begin{aligned} \lim_{\kappa \to 0^+} \rho(\kappa) &= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi} \Gamma\left(\frac{n+2}{2}\right)}, \\ \lim_{\kappa \to \infty} \rho(\kappa) &= 1, \end{aligned} \tag{3.14}$$

so that ρ is a real-valued, monotonically increasing, and bounded function of κ with values between $\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi} \Gamma\left(\frac{n+2}{2}\right)}$ and 1.

Maximum-Likelihood Estimators

This section presents a detailed derivation of the maximum-likelihood estimates for the parameters of the maximum-entropy distribution just introduced. Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ comprise a finite set of samples drawn independently from the distribution characterized by the probability density

$$f_{\mathbf{x}}(\boldsymbol{x} | \boldsymbol{\mu}, \kappa) = c_n(\kappa) e^{\kappa |\boldsymbol{\mu}^\top \boldsymbol{x}|},$$

where $\boldsymbol{\mu} \in \mathbb{RP}^n$ and $\kappa \in \mathbb{R}^+$ are fixed parameters, and $c_n(\kappa)$ is the normalizing constant given by Equation 3.12. Under the assumption of independence, we may write the joint

likelihood of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ as

$$f_{\mathbf{x}_1,\ldots,\mathbf{x}_m}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m|\boldsymbol{\mu}, \kappa) = c_n^m(\kappa) \exp\left\{\kappa \sum_{i=1}^{m} |\boldsymbol{\mu}^\top \boldsymbol{x}_i|\right\}. \tag{3.15}$$

Since maximizing $f_{\mathbf{x}_1,\ldots,\mathbf{x}_m}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m|\boldsymbol{\mu}, \kappa)$ in Equation 3.15 is equivalent to maximizing its log, the maximum-likelihood estimates $\widehat{\boldsymbol{\mu}}_{\text{ML}}$ and $\widehat{\kappa}_{\text{ML}}$ result from maximizing

$$\ell\left(\boldsymbol{x}_1, \ldots \boldsymbol{x}_m|\boldsymbol{\mu}, \kappa\right) = m \log c_n(\kappa) + \kappa \sum_{i=1}^{m} |\boldsymbol{\mu}^\top \boldsymbol{x}_i|. \tag{3.16}$$

As only the last term of Equation 3.16 is dependent on $\boldsymbol{\mu}$, the maximum-likelihood estimate $\widehat{\boldsymbol{\mu}}_{\text{ML}}$ of $\boldsymbol{\mu}$ is evidently

$$\widehat{\boldsymbol{\mu}}_{\text{ML}} = \underset{\boldsymbol{p}\in\mathbb{RP}^n}{\operatorname{argmax}} \sum_{i=1}^{m} \left|\boldsymbol{p}^\top \boldsymbol{x}_i\right|, \tag{3.17}$$

which defines $\widehat{\boldsymbol{\mu}}_{\text{ML}}$ to be (monotonically related to) a so-called Karcher mean of the samples $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ on $\mathbb{RP}^n$ [25]. To find the maximum-likelihood estimate of κ, we solve

$$0 = \frac{\partial}{\partial \kappa} \ell(\boldsymbol{x}_1, \ldots \boldsymbol{x}_m; \widehat{\boldsymbol{\mu}}_{\text{ML}}, \kappa) = \frac{m c_n'(\kappa)}{c_n(\kappa)} + \sum_{i=1}^{m} \left|\widehat{\boldsymbol{\mu}}_{\text{ML}}^\top \boldsymbol{x}_i\right|.$$

Using the formulae for the derivatives of the special functions in the normalizing constant (see Appendix A),

$$-\frac{c_n'(\kappa)}{c_n(\kappa)} = \frac{I_{\frac{n-3}{2}}(\kappa) + L_{\frac{n-3}{2}}(\kappa)}{I_{\frac{n-1}{2}}(\kappa) + L_{\frac{n-1}{2}}(\kappa)} - \frac{(n-1)}{\kappa}.$$

Since computing $\widehat{\kappa}_{\text{ML}}$ thus involves an implicit equation that is a ratio of special functions, it is not possible to obtain an analytic solution, and we must take recourse to numerical or asymptotic methods to obtain an approximation.

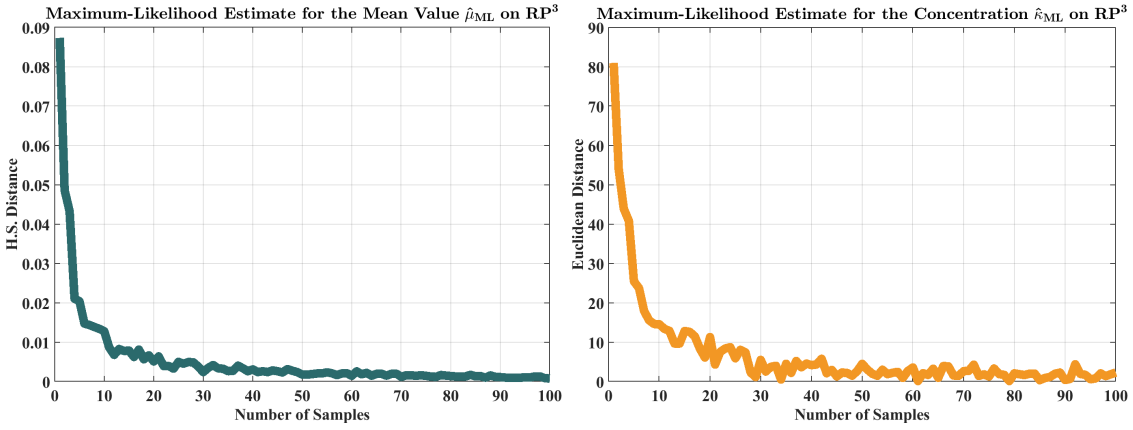

Figure 3.3: The maximum-likelihood estimates of $\boldsymbol{\mu}$ (left) and κ (right) of the maximum-entropy probability density function $f_{\mathbf{x}}(\boldsymbol{x}|\boldsymbol{\mu}, \kappa)$ defined on $\mathbb{RP}^3$.

Figure 3.3 suggests consistency of the maximum-likelihood estimates $\widehat{\boldsymbol{\mu}}_{\text{ML}}$ and $\widehat{\kappa}_{\text{ML}}$: empirically the estimates appear to converge to the true parameter values, as the number of independent realizations of $\mathbf{x}$ increases. The distance metric used here on $\mathbb{RP}^n$ is the Hilbert-Schmidt distance metric, defined by the Hilbert-Schmidt norm on the space of n-by-n orthogonal projection operators:

$$d_{\text{H.S.}}(\widehat{\boldsymbol{\mu}}_{\text{ML}}, \boldsymbol{\mu}) = \|\boldsymbol{P}_{\widehat{\boldsymbol{\mu}}_{\text{ML}}} - \boldsymbol{P}_{\boldsymbol{\mu}}\|_{\text{H.S.}} .$$

Explicit distance functions on grassmannians will be discussed in detail in Chapter 5.

3.2 Complex Projective Space

Important applications for subspace-based methods in data analysis arise in telecommunications and radio-frequency remote sensing. These applications typically involve narrowband signals modulating a high-frequency sinusoidal carrier wave. Complex-valued functions or complex vectors are natural representations for such signals for processing at the receiver [19] (specifically, chapter three), so complex vector spaces are of considerable practical interest in this context.

Much of what is presented here about complex projective space is directly analogous to the case of real projective space. Derivation of the invariant measure and its integration entail some differences that warrant particular attention, however.

For completeness, we begin with a formal, general definition.

Definition 3.1. ***Complex projective*** n***-space*** *(also referred to as* ***complex projective space****) is the set of all one-dimensional linear subspaces in* $\mathbb{C}^{n+1}$*. We denote this space* $\mathbb{CP}^n$*.*

Coordinates on complex projective space are formulated analogously to the real case presented above.

An Orthonormal Basis Perspective For $\boldsymbol{x} \in \mathbb{CP}^n$, let $\boldsymbol{x}_{\boldsymbol{z}}$ be the unit vector of the form

$$\boldsymbol{x}_{\boldsymbol{z}} = \begin{bmatrix} 1 \\ \boldsymbol{z} \end{bmatrix} \left(1 + \boldsymbol{z}^\dagger \boldsymbol{z}\right)^{-1/2}, \tag{3.18}$$

where $\boldsymbol{z} = \begin{bmatrix} z^1 & z^2 & \cdots & z^n \end{bmatrix}^\top$ is a now an element of $\mathbb{C}^n$ and $\dagger$ denotes the hermitian transpose. It is important to note that this representation is an $(n+1)$-dimensional, complex vector whose first coordinate is positive *and real-valued*. As with the case of real projective space, these coordinates implicitly characterize an atlas comprising a single chart covering almost all of $\mathbb{CP}^n$; the set of excluded points has measure zero. There is a corresponding orthonormal basis for the orthogonal complement of $\boldsymbol{x}_{\boldsymbol{z}}$ in $\mathbb{C}^{n+1}$, comprising the columns of the $(n+1)$-by-n matrix we denote $\boldsymbol{x}_{\boldsymbol{z}}^\perp$. In coordinates,

$$\boldsymbol{x}_{\boldsymbol{z}}^\perp = \begin{bmatrix} -\boldsymbol{z}^\dagger \\ \mathbb{I}_n \end{bmatrix} \left(\mathbb{I}_n + \boldsymbol{z}\boldsymbol{z}^\dagger\right)^{-1/2}. \tag{3.19}$$

A Hermitian Projector Perspective In following through the different perspectives of coordinate representations of $\mathbb{CP}^n$, we may identify to each one-dimensional subspace of $\mathbb{C}^n$ a rank-one, hermitian projection operator on $\mathbb{C}^{n+1}$. Using the coordinates in Equation 3.18, these coordinates are given by

$$\boldsymbol{P}_{\boldsymbol{z}} = \boldsymbol{x}_{\boldsymbol{z}} \boldsymbol{x}_{\boldsymbol{z}}^\dagger = \left(1 + \boldsymbol{z}^\dagger \boldsymbol{z}\right)^{-1} \begin{bmatrix} 1 & \boldsymbol{z}^\dagger \\ \boldsymbol{z} & \boldsymbol{z}\boldsymbol{z}^\dagger \end{bmatrix}.$$

A Homogeneous Perspective And finally, letting $\mathbf{U}(n+1)$ act (transitively) on $\mathbb{CP}^n$ gives rise to the homogeneous perspective of complex n-space:

$$\mathbb{CP}^n \cong \mathbf{U}(n+1) / \left(\mathbf{U}(1) \times \mathbf{U}(n)\right). \tag{3.20}$$

As in the case of real projective space, a specialized case of the Quotient Manifold Theorem ensures that, with this identification, $\mathbb{CP}^n$ is indeed a smooth manifold.

3.2.1 The Differential Form for the Invariant Measure

Let us return to the coordinate system defined by Equations 3.18 and 3.19. For $\boldsymbol{x} \in \mathbb{CP}^n$, $\boldsymbol{x}_{\boldsymbol{z}}$ denotes the unit vector of the form

$$\boldsymbol{x}_{\boldsymbol{z}} = \begin{bmatrix} 1 \\ \\ \boldsymbol{z} \end{bmatrix} \left(1 + \boldsymbol{z}^\dagger \boldsymbol{z}\right)^{-1/2}, \tag{3.18 revisited}$$

where $\boldsymbol{z} = \begin{bmatrix} z^1 & z^2 & \cdots & z^n \end{bmatrix}^\top$ is an element of $\mathbb{C}^n$. The corresponding orthonormal basis for the orthogonal complement of $\boldsymbol{x}_{\boldsymbol{z}}$ in $\mathbb{C}^{n+1}$ comprises the columns of the $(n+1)$-by-n complex matrix given by

$$\boldsymbol{x}_{\boldsymbol{z}}^{\perp} = \begin{bmatrix} -\boldsymbol{z}^\dagger \\ \\ \mathbb{I}_n \end{bmatrix} \left(\mathbb{I}_n + \boldsymbol{z}\boldsymbol{z}^\dagger\right)^{-1/2}. \tag{3.19 revisited}$$

By analogy again with James' invariant measure for real projective space given in Equation 3.5, the invariant measure on $\mathbb{CP}^n$ can be defined for points $\boldsymbol{x}_{\boldsymbol{z}}$ in a neighborhood of a point $\boldsymbol{x} \in \mathbb{CP}^n$:

$$\omega_n(\boldsymbol{x}_{\boldsymbol{z}}) \triangleq \left(\frac{i}{2}\right)^n \left(\bigwedge_{j=1}^{n} \boldsymbol{b}_j^\dagger d\boldsymbol{x}_{\boldsymbol{z}} \wedge \bigwedge_{j=1}^{n} d\boldsymbol{x}_{\boldsymbol{z}}^\dagger \boldsymbol{b}_j \right). \tag{3.21}$$

In this expression, $\boldsymbol{x}_{\boldsymbol{z}}$ is a unit vector in the subspace defined by $\boldsymbol{x}$; $d\boldsymbol{x}_{\boldsymbol{z}}$ is the total derivative of $\boldsymbol{x}_{\boldsymbol{z}}$ (i.e., the differential); and $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ are the orthonormal column vectors spanning the orthogonal complement of $\boldsymbol{x}$ in $\mathbb{C}^{n+1}$, as specified in Equation 3.19. The factor $\left(\frac{i}{2}\right)$ accounts for the fact that decomposing any form into its real and imaginary components, say x and y, respectively, yields the following relation:

$$(x + iy) \wedge (x - iy) = -2i(x \wedge y).$$

Having already computed $\bigwedge_{j=1}^{n} \boldsymbol{b}_j^\dagger d\boldsymbol{x}_{\boldsymbol{z}}$ in Equations 3.6 to 3.8, it evidently follows that

$$\omega_n(\boldsymbol{x}_{\boldsymbol{z}}) = \left(\frac{i}{2}\right)^n \left(1 + \boldsymbol{z}^\dagger \boldsymbol{z}\right)^{-(n+1)} d\boldsymbol{z} \wedge d\overline{\boldsymbol{z}}. \tag{3.22}$$

Integrating the Differential Form

Given

$$\boldsymbol{x}_{\boldsymbol{z}} = \begin{bmatrix} 1 \\ \\ \boldsymbol{z} \end{bmatrix} \left(1 + \boldsymbol{z}^{\dagger}\boldsymbol{z}\right)^{-1/2},$$

where $\boldsymbol{z} = \begin{bmatrix} z^1 & z^2 & \cdots & z^n \end{bmatrix}^{\top}$ is an element of $\mathbb{C}^n$, consider the transformation

$$\begin{aligned} z^k &= e^{i\varphi^k} \tan\theta^1 \prod_{j=2}^{k} \sin\theta^j \cos\theta^{k+1}, \quad k < n \\ z^n &= e^{i\varphi^n} \tan\theta^1 \prod_{j=2}^{n} \sin\theta^j \end{aligned}$$

with $\theta^j \in (0, \pi/2)$ and $\varphi^j \in (0, 2\pi)$ for $j = 1, \ldots, n$. Introducing $\boldsymbol{z} = \boldsymbol{u} + i\boldsymbol{v}$, where $\boldsymbol{u}$ and $\boldsymbol{v}$ are in $\mathbb{R}^n$, the transformation law for the volume form on $\mathbb{CP}^n$ is given by (see Appendix B)

$$\begin{aligned} \left(\frac{i}{2}\right)^n (1 + \boldsymbol{z}^{\dagger}\boldsymbol{z})^{-(n+1)} d\boldsymbol{z} \wedge d\overline{\boldsymbol{z}} &= \left(1 + \boldsymbol{u}^{\top}\boldsymbol{u} + \boldsymbol{v}^{\top}\boldsymbol{v}\right)^{-(n+1)} du^1 \wedge \cdots \wedge du^n dv^1 \wedge \cdots \wedge dv^n \\ &= \left| \prod_{j=1}^{n} \sin^{2(n-j)+1}\theta^j \cos\theta^j \right| d\theta^1 \wedge \cdots \wedge d\theta^n d\varphi^1 \wedge \cdots \wedge d\varphi^n, \end{aligned} \tag{3.23}$$

whence the volume of $\mathbb{CP}^n$ is computed as (see Appendix A)

$$\begin{aligned} \mathrm{Vol}(\mathbb{CP}^n) &= \int_{\mathbb{CP}^n} \omega_n(\boldsymbol{x}_{\boldsymbol{z}}) \\ &= \int_{\mathbb{R}^{2n}} \left(1 + \boldsymbol{u}^{\top}\boldsymbol{u} + \boldsymbol{v}^{\top}\boldsymbol{v}\right)^{-(n+1)} du^1 dv^1 \cdots du^n dv^n \\ &= \int_0^{2\pi} \cdots \int_0^{2\pi} \prod_{i=1}^{n} d\varphi^i \int_0^{\pi/2} \cdots \int_0^{\pi/2} \sin^{2n-1}\theta^1 \cos\theta^1 d\theta^1 \prod_{i=2}^{n} \sin^{2(n-i)+1}\theta^i \cos\theta^i d\theta^i \\ &= (2\pi)^n \cdot \int_0^{\pi/2} \sin^{2n-1}\theta^1 \cos\theta^1 d\theta^1 \prod_{i=2}^{n} \int_0^{\pi/2} \sin^{2(n-i)+1}\theta^i \cos\theta^i d\theta^i \\ &= (2\pi)^n \cdot \frac{1}{2n} \cdot \frac{1}{2^{n-1}(n-1)!} \\ &= \frac{\pi^n}{n!}. \end{aligned}$$

As in the case of $\mathbb{RP}^n$, the volume of $\mathbb{CP}^n$ is an apparently well-known result, but justification often relies on invoking theorems about volumes of the product spaces and quotient manifolds [4]. An explicit proof does not appear to be readily available in the literature.

3.2.2 A Probability Distribution

We now seek a class of maximum-entropy probability distributions on $\mathbb{CP}^n$ to support bayesian estimation on this manifold. As in the case of $\mathbb{RP}^n$, treated in Section 3.1.2, a variational approach is used to obtain a distribution whose (differential) entropy is maximum, with respect to a known constraint.

Theorem 3.1. *For a fixed non-negative, real number κ and a fixed point $\boldsymbol{\mu}$ in $\mathbb{CP}^n$, the continuous probability density function on $\mathbb{CP}^n$ with first moment equal to $\rho\boldsymbol{\mu} \in \mathbb{C}^{n+1}$ which maximizes entropy among all such densities (on $\mathbb{CP}^n$ with first moment $\rho\boldsymbol{\mu}$) is given by*

$$f_{\mathbf{x}}(\boldsymbol{x}|\boldsymbol{\mu}, \kappa) = c_n(\kappa)e^{\kappa|\boldsymbol{\mu}^\dagger \boldsymbol{x}|}, \tag{3.24}$$

where $c_n(\kappa)$ is chosen so that $e^{\kappa|\boldsymbol{\mu}^\dagger \boldsymbol{x}|}$ integrates, with respect to the volume form, to unity over the whole space and κ so that $\mathrm{E}[\boldsymbol{x}] = \rho(\kappa)\boldsymbol{\mu}$

The proof of this theorem is essentially identical to the one given for Theorem 3.1. In fact, the density functions for classes of maximum-entropy distributions useful to us on real and complex projective spaces appear to be indistinguishable. Just as for the real case, the density $f_{\mathbf{x}}(\boldsymbol{x}|\boldsymbol{\mu}, \kappa)$ on $\mathbb{CP}^n$ is parameterized by the *mean direction* $\boldsymbol{\mu} \in \mathbb{CP}^n$ and the *concentration factor* $\kappa \in \mathbb{R}^+$, so-called because it characterizes how strongly elements on $\mathbb{CP}^n$ are concentrated about the mean direction $\boldsymbol{\mu}$. The key differences though, lie in the values of their normalizing constants and the relationships between the fixed ρ and the density parameter κ, which manifest only subtly in the analysis of this distribution. We focus here first on the derivation of the value of $c_n(\kappa)$ and the express relationship between ρ and κ for the complex case.

To derive the value of the normalizing constant $c_n(\kappa)$ in this case, we use the fact that

$$\int_{\mathbb{CP}^n} c_n(\kappa)e^{\kappa|\boldsymbol{\mu}^\dagger \boldsymbol{x}|} d\boldsymbol{x} = 1,$$

where $d\boldsymbol{x}$ represents the differential form for the invariant measure on $\mathbb{CP}^n$. As in the case of $\mathbb{RP}^n$, to evaluate this integral, we first make the transformation $\boldsymbol{y} = \boldsymbol{Q}\boldsymbol{x}$, where the first

component of $\boldsymbol{y}$ is equal to $\boldsymbol{\mu}^\dagger \boldsymbol{x}$ and $\boldsymbol{Q}$ is unitary (the first row of $\boldsymbol{Q}$ will be $\boldsymbol{\mu}^\dagger$). The matrix $\boldsymbol{Q}$ is chosen to be unitary so that $d\boldsymbol{x} = d\boldsymbol{y}$. Introducing again $\boldsymbol{z} = \boldsymbol{u} + i\boldsymbol{v}$, where $\boldsymbol{u}$ and $\boldsymbol{v}$ are in $\mathbb{R}^n$, we rewrite the equation above as

$$\frac{1}{c_n(\kappa)} = \int\limits_{\mathbb{CP}^n} e^{\kappa |\boldsymbol{\mu}^\dagger \boldsymbol{x}|} d\boldsymbol{x} = \left(\frac{i}{2}\right)^n \int\limits_{\mathbb{R}^{2n}} e^{\kappa\left(1+\boldsymbol{z}^\dagger \boldsymbol{z}\right)} \left(1 + \boldsymbol{z}^\dagger \boldsymbol{z}\right)^{-(n+1)} d\boldsymbol{z} \wedge d\overline{\boldsymbol{z}}.$$

Under a second transformation to complex hyperspherical coordinates,

$$\begin{aligned}
\frac{1}{c_n(\kappa)} &= \int_0^{2\pi} \cdots \int_0^{2\pi} d\varphi^1 \cdots d\varphi^n \int_0^{\frac{\pi}{2}} \cdots \int_0^{\frac{\pi}{2}} e^{\kappa \cos\theta^1} sin^{2n-1}\theta^1 \cos\theta^1 \prod_{j=2}^{n} \sin^{2(n-j)+1}\theta^j \cos\theta^j d\theta^1 \cdots d\theta^n \\
&= (2\pi)^n \int_0^{\frac{\pi}{2}} e^{\kappa \cos\theta^1} \sin^{2n-1}\theta^1 \cos\theta^1 d\theta^1 \prod_{j=2}^{n} \int_0^{\frac{\pi}{2}} \sin^{2(n-j)+1}\theta^j \cos\theta^j d\theta^j \\
&= (2\pi)^n \frac{1}{2^{n-1}(n-1)!} \int_0^{\frac{\pi}{2}} e^{\kappa \cos\theta^1} \sin^{2n-1}\theta^1 \cos\theta^1 d\theta^1,
\end{aligned}$$

where the last equality makes use of results derived in Appendix A. We continue via integration by parts:

$$\frac{1}{c_n(\kappa)} = (2\pi)^n \frac{1}{2^{n-1}(n-1)!} \left[\frac{1}{2n} + \frac{\kappa}{2n} \int_0^{\frac{\pi}{2}} e^{\kappa \cos\theta^1} \sin^{2n+1}\theta^1 d\theta^1\right].$$

After application of results derived in Appendix A, simplification yields

$$c_n(\kappa) = \frac{1}{\frac{\pi^n}{n!} + \left(\frac{2}{\kappa}\right)^{n-\frac{1}{2}} (\pi)^{n+\frac{1}{2}} \left[I_{n+\frac{1}{2}}(\kappa) + L_{n+\frac{1}{2}}(\kappa)\right]}, \tag{3.25}$$

where $I_\nu(x)$ and $L_\nu(x)$ are the now familiar modified Bessel function of the first kind and modified Struve function. This normalizing constant serves in the complex case the role in $c_n(\kappa)$ defined by Equation 3.12. To see the relationship between ρ and κ, we again use the fact that

$$c_n(\kappa) \int\limits_{\mathbb{CP}^n} \left|\boldsymbol{\mu}^\dagger \boldsymbol{x}\right| e^{\kappa |\boldsymbol{\mu}^\dagger \boldsymbol{x}|} d\boldsymbol{x} = \rho.$$

Applying the same two coordinate transformations as in the derivation of the normalizing constant above will yield the following (rather hideous) expression:

$$\rho(\kappa) = \frac{\kappa + \left(\frac{2}{\kappa}\right)^{n+\frac{1}{2}} \sqrt{\pi}\Gamma(n+2)\left[\kappa I_{n+\frac{3}{2}}(\kappa) + I_{n+\frac{1}{2}}(\kappa) + \kappa L_{n+\frac{3}{2}}(\kappa) + L_{n+\frac{1}{2}}(\kappa)\right]}{2n + 2 + \left(\frac{2}{\kappa}\right)^{n+\frac{1}{2}} \sqrt{\pi}\kappa\Gamma(n+2)\left[I_{n+\frac{1}{2}}(\kappa) + L_{n+\frac{1}{2}}(\kappa)\right]}. \tag{3.26}$$

Using the results regarding the limiting behavior of $I_\nu(x)$ and $L_\nu(x)$ recorded in Appendix A, together will elementary algebra,

$$\begin{aligned}\lim_{\kappa\to 0^+}\rho(\kappa) &= \frac{\sqrt{\pi}\Gamma(n+2)}{2\,(n+1)\,\Gamma\left(\frac{3}{2}+n\right)}\\ \lim_{\kappa\to\infty}\rho(\kappa) &= 1,\end{aligned} \tag{3.27}$$

so that ρ is a real-valued, monotonically increasing, and bounded function of κ with values between $\frac{\sqrt{\pi}\Gamma(n+2)}{2\,(n+1)\,\Gamma\left(\frac{3}{2}+n\right)}$ and 1.

We remark that since the density function corresponding to the maximum-entropy distribution (with respect to a fixed first moment) on $\mathbb{CP}^n$ is identical to the form of the density function corresponding to the maximum-entropy distribution (with respect to a fixed first moment) on $\mathbb{RP}^n$, the maximum-likelihood estimates of its parameters too take the same form. The maximum-likelihood estimate for the concentration parameter κ is an implicit function involving a ratio of the normalizing constant and its derivative, however. For the complex case,

$$-\frac{c_n'(\kappa)}{c_n(\kappa)} = \frac{\sqrt{\pi}\left(\frac{2}{\kappa}\right)^{n+1}\Gamma(n+1)\left(-2nI_{n+\frac{1}{2}}(\kappa)+\kappa I_{n-\frac{1}{2}}(\kappa)-2nL_{n+\frac{1}{2}}(\kappa)+\kappa L_{n-\frac{1}{2}}(\kappa)\right)}{\sqrt{\pi}\left(\frac{2}{\kappa}\right)^{n-\frac{1}{2}}\Gamma(n+1)\left(I_{n+\frac{1}{2}}(\kappa)+L_{n+\frac{1}{2}}(\kappa)\right)+1}.$$

As in the real case, we take recourse to numerical methods to obtain an approximation for $\widehat{\kappa}_{\text{ML}}$, given that

$$-\frac{mc_n\left(\widehat{\kappa}_{\text{ML}}\right)}{c_n\left(\widehat{\kappa}_{\text{ML}}\right)} = \sum_{i=1}^{m}\left|\widehat{\boldsymbol{\mu}}_{\text{ML}}^{\dagger}\boldsymbol{x}_i\right|.$$

Chapter 4

RECURSIVE BAYESIAN ESTIMATION ON PROJECTIVE SPACES

With the requisite geometric-statistical objects developed on special grassmannians in the the preceding chapter in hand, we now devise and implement a bayesian framework for *recursive* estimation posed on these manifolds. Practical motivation lies in the context of a particular multi-sensor remote sensing application, wherein a primary goal, as mentioned in Section 2.2, is to estimate a common unknown signal of known rank k measured by $m > k$ spatially distributed sensors. Formally, we assume the signal is characterized by $k \geq 1$ linearly independent waveforms. As depicted in Figure 4.1, the waveforms emanate from an omnidirectional source, illuminating $m > k$ spatially distributed sensors, which transduce impinging electromagnetic radiation (or acoustic sound pressure).

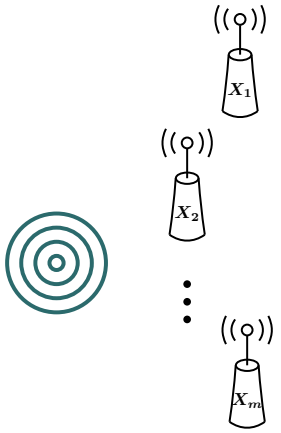

Figure 4.1: Data in the form of complex n-vectors arise from sampling k linearly independent waveforms (i.e., a rank-k signal) illuminating $m > k$ spatially distributed sensors. The rank-k signal defines a particular k-dimensional subspace, regarded in this setup as a realization of a random variable on $\mathbf{G}_{\mathbb{C}}(k, n)$.

Data received at each of the m sensors are assumed to have been filtered to a band of interest, suitably sampled, and appropriately adjusted in time delay and Doppler according to the putative target location and motion. This pre-processing provides m complex n-vectors,

which, when suitably arranged, can be described by the n-by-m complex data matrix

$$\boldsymbol{X} = \boldsymbol{S}\boldsymbol{A} + \boldsymbol{\nu} \tag{4.1}$$

whose elements $\boldsymbol{x}^{nm}$ are the samples of the noisy rank-k signal collected at the m^{th} sensor. The k-dimensional signal subspace is defined by $\boldsymbol{S} \in \mathbb{C}^{n\times k}$, whose columns are orthonormal complex n-vectors, and the element a^{km} of the matrix $\boldsymbol{A} \in \mathbb{C}^{k\times m}$ is a complex gain of the component signal received at sensor m and in the subspace corresponding to the k^{th} column of $\boldsymbol{S}$. As is common (and rather accurate to many real-world systems), we assume the columns of the additive noise matrix, denoted here by $\boldsymbol{\nu}$, are realizations of independent and identically distributed complex gaussian vectors with zero mean and covariance matrix equal to a scalar multiple of the identity matrix. With the exception of the properties described here, both $\boldsymbol{A}$ and $\boldsymbol{S}$ are treated as unknown variables.

We introduce additional dynamical assumptions on the k linearly independent waveforms defining $\boldsymbol{S}$ in this setup, presumably due to time-varying illumination or the effects of possible target motion, so that $\boldsymbol{S}$ itself is time-varying, at least in princible. Decorating Equation 4.1 with subscripts,

$$\boldsymbol{X}_t = \boldsymbol{S}_t\boldsymbol{A}_t + \boldsymbol{\nu}_t \tag{4.2}$$

denotes the received data in discrete time epochs, indexed by $t = 1, 2, \ldots$. When $\boldsymbol{S}_t$ is regarded as a realization of a random variable $\mathbf{S}_t$ on $\mathbf{G}_{\mathbb{C}}(k, n)$ with an assigned (unknown) distribution, receipt of new measurements comprising $\boldsymbol{X}_t$ each time epoch can be used to update an estimate of $\boldsymbol{S}_t$. These probabilistic formulations are ideally suited for a bayesian approach to estimation. We begin this chapter with the presentation of a general overview of the topic of recursive estimation, from a bayesian perspective, before applying it to this setup.

4.1 A Bayesian Framework for Recursive Estimation

Adopting traditional terminology, the dynamic random variable to be estimated is called the *state* and its observation space the *state space*. A typical recursive estimation algo-

rithm consists of essentially two stages: a prediction stage and an update stage. The prediction stage uses a dynamical model to propagate the probability density of the state forward from one measurement time to the next, and the update stage uses a measurement, with a probabilistic assumption, to modify the prediction probability density of the state. From a bayesian perspective, the objective is to recursively calculate some degree of belief in the state at a discrete time $t = 1, 2, \ldots$ given measurements up to time t. Throughout this discussion (and this chapter!) we let $\mathbf{s}_t$ and $\boldsymbol{z}_{1:t}$ denote the state at time t and the measurements up to time t, respectively. We seek to construct the probability density $f_{\mathbf{s}_t}(\boldsymbol{s}_t|\,\boldsymbol{z}_{1:t})$, with the assumption that the initial density of the state is available.[1] If we suppose the density $f_{\mathbf{s}_t}(\boldsymbol{s}_{t-1}|\boldsymbol{z}_{1:t-1})$ at time $t-1$ is available, the prediction stage then involves using the system dynamics to obtain the prior probability density of the state at time t via the Chapman-Kolmogorov equation

$$f_{\mathbf{s}_t}(\boldsymbol{s}_t|\boldsymbol{z}_{1:t-1}) = \int f_{\mathbf{s}_t}(\boldsymbol{s}_t|\boldsymbol{s}_{t-1})\, f_{\mathbf{s}_{t-1}}(\boldsymbol{s}_{t-1}|\boldsymbol{z}_{1:t-1})\, d\boldsymbol{s}_{t-1}. \tag{4.3}$$

At time t, a measurement $\boldsymbol{z}_t$ becomes available, and is assimilated into the prior probability density via Bayes' Rule:

$$f_{\mathbf{s}_t}(\boldsymbol{s}_t|\boldsymbol{z}_{1:t}) = \frac{f_{\mathbf{z}_t}(\boldsymbol{z}_t|\boldsymbol{s}_t)\, f_{\mathbf{s}_t}(\boldsymbol{s}_t|\boldsymbol{z}_{1:t-1})}{f_{\mathbf{z}_t}(\boldsymbol{z}_t|\boldsymbol{z}_{1:t-1})}, \tag{4.4}$$

where the normalizing constant

$$f_{\mathbf{z}_t}(\boldsymbol{z}_t|\boldsymbol{z}_{1:t-1}) = \int f_{\mathbf{z}_t}(\boldsymbol{z}_t|\boldsymbol{s}_t)\, f_{\mathbf{s}_t}(\boldsymbol{s}_t|\boldsymbol{z}_{1:t-1})\, d\boldsymbol{s}_t$$

depends on the likelihood function $f_{\mathbf{z}_t}(\boldsymbol{z}_t|\boldsymbol{s}_t)$ defined by the measurement model. The recurrence relations in Equations 4.3 and 4.4 form the basis for the optimal bayesian solution of the iterative estimation problem and, under particular model assumptions, yield the familiar Kalman filter. It's crucial to note that Equations 4.3 and 4.4 require integration over the state space, which, for the estimation problems of principal interest to this thesis, are Grassmann manifolds. The present chapter specializes the problem to real and complex

[1] In our simulations, the initial density is the uniform density on the state space.

projective spaces, applying the geometric-statistical objects introduced in the preceding chapter. Figure 4.2 illustrates the prediction and update steps of a desired iterative estimator of a state evolving on real projective space.

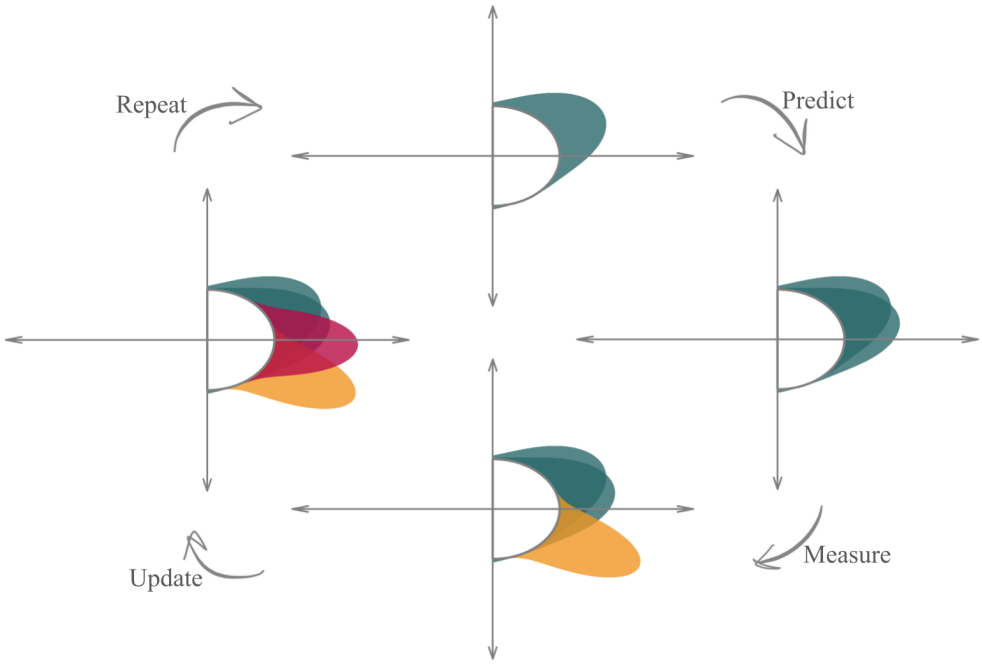

Figure 4.2: A recursive bayesian estimation procedure on $\mathbb{RP}^1$.

4.2 An Idealized Model

Before moving to recursively estimating the state of a dynamical system evolving in time on $\mathbb{RP}^n$ using measurements modeled by the practically motivated scenario set forth at the start of the chapter, we focus on a much simpler problem in which the measurement model is, in a very particular sense, *idealized*.

We model the statistics with the maximum-entropy distribution (with respect to a fixed first moment) on real projective space, as derived in Section 3.1.2. For this idealized model, we further assume the measurements at each time step are also distributed according to this distribution, but perhaps characterized by a different mean or concentration parameter value. A bayesian formalism, as described above, is used. Now because the forms of the density functions corresponding to the maximum-entropy distributions on real and com-

plex projective space are identical, the bayesian approach to estimation described here for real projective space extends directly to the complex case. We devote our presentation to real projective space, however. Numerical simulations demonstrate that, under mild assumptions, our recursive estimator on $\mathbb{RP}^n$ can be found to achieve successful estimation. While original, the work presented in this section is a direct extension of the geometric-statistical approach in [47].

4.2.1 The Prediction

On the space $\mathbb{RP}^n$, we are interested in recursively estimating the state $\mathbf{s}_t \in \mathbb{RP}^n$ of a dynamical system at time t. We assume that, at time $t-1$, the posterior distribution of the state has been computed using all of the measurements collected at discrete times $1, 2, \ldots, t-1$. We further assume that, at time $t-1$, we can describe the posterior distribution of the state $\mathbf{s}_{t-1}$ in terms of the maximum-entropy distribution (with respect to a fixed first moment) with mean parameter $\widehat{s}_{t-1}$ and concentration parameter κ_{t-1}. We write the posterior density at time $t-1$ as

$$f_{\mathbf{s}_t}(\boldsymbol{s}_{t-1}^+) = c_n(\kappa_{t-1}^+) \exp\left\{\kappa_{t-1}^+ \left|(\widehat{\boldsymbol{s}}_{t-1}^+)^\top \boldsymbol{s}_{t-1}^+\right|\right\}.$$

The dynamical model in this research assumes that the state at time $t-1$ changes via an action of a special orthogonal matrix (the group of orthogonal matrices acts transitively on $\mathbb{RP}^n$ by left multiplication) to yield the state at time t. That is, the dynamics are governed by the equation

$$\mathbf{s}_t = \boldsymbol{R}_t \mathbf{s}_{t-1},$$

where $\boldsymbol{R}_t$ is a deterministic and *known* element of $\mathbf{SO}(n+1)$. Because we assume $\boldsymbol{R}_t$ is known (for each t), the prior probability distribution of the state at time t is characterized by

$$f_{\mathbf{s}_t}(\boldsymbol{s}_t) = c_n(\kappa_{t-1}) \exp\left\{\kappa_{t-1} \left|(\boldsymbol{R}_t \widehat{\boldsymbol{s}}_{t-1})^\top \boldsymbol{s}_t\right|\right\}, \tag{4.5}$$

and hence the mean and concentration parameters at time t are, respectively,

$$\widehat{\boldsymbol{s}}_t = \boldsymbol{R}_t \widehat{\boldsymbol{s}}_{t-1}$$

$$\kappa_t = \kappa_{t-1}.$$

When the state space is $\mathbb{CP}^n$, the only difference lies in the assumptions of the dynamics. The state $\mathbf{s}_{t-1}$ at time $t-1$ changes via an action of a special *unitary* matrix to yield the state $\mathbf{s}_t$ at time t. That is, the dynamics are governed by the equation

$$\mathbf{s}_t = \boldsymbol{U}_t \mathbf{s}_{t-1},$$

where $\boldsymbol{U}_t$ is a deterministic and known element of $\boldsymbol{SU}(n+1)$. The prior probability density is thus

$$f_{\mathbf{s}_t}(\boldsymbol{s}_t) = c_n(\kappa_{t-1}) \exp\left\{\kappa_{t-1} \left| (\boldsymbol{U}_t \widehat{\boldsymbol{s}}_{t-1})^\dagger \, \boldsymbol{s}_t \right|\right\}, \tag{4.6}$$

and hence the mean and concentration parameters at time t are, respectively,

$$\widehat{\boldsymbol{s}}_t = \boldsymbol{U}_t \widehat{\boldsymbol{s}}_{t-1}$$

$$\kappa_t = \kappa_{t-1}.$$

4.2.2 A Measurement and the Bayesian Update

A measurement is collected in each time epoch and also follows the maximum-entropy distribution on $\mathbb{RP}^n$, as promised. At time t, $\boldsymbol{s}_t$ is the mean and η_t is the concentration parameter of the distribution of the measurement. The measurement likelihood is thus

$$f_{\mathbf{x}_t}(\boldsymbol{x}_t | \boldsymbol{s}_t) = c_n(\eta_t) \exp\left\{\eta_t \left| \boldsymbol{s}_t^\top \boldsymbol{x}_t \right|\right\}.$$

We want to find a probability density for the update (the posterior probability density) which belongs to the same class of maximum-entropy probability densities as the prior. We begin calculating the posterior by application of Bayes' Rule:

$$\begin{aligned} f_{\mathbf{s}_t}(\boldsymbol{s}_t | \boldsymbol{x}_t) &\propto f_{\mathbf{x}_t}(\boldsymbol{x}_t | \boldsymbol{s}_t) f_{\mathbf{s}_t}(\boldsymbol{s}_t) \\ &\propto \exp\left\{\kappa_t \left| \widehat{\boldsymbol{s}}_t^\top \boldsymbol{s}_t \right|\right\} \exp\left\{\eta_t \left| \boldsymbol{s}_t^\top \boldsymbol{x}_t \right|\right\} \\ &= \exp\left\{\kappa_t \left| \widehat{\boldsymbol{s}}_t^\top \boldsymbol{s}_t \right| + \eta_t \left| \boldsymbol{s}_t^\top \boldsymbol{x}_t \right|\right\}. \end{aligned} \tag{4.7}$$

Because this density does not always belong to the desired class of maximally entropic densities on the state space, we approximate the density in Equation 4.7 by a maximum-entropy

distribution (with respect to a fixed first moment) on $\mathbb{RP}^n$ with matching first moment. Presently, this moment-matching procedure is performed numerically.

4.2.3 A Numerical Example

A simple numerical example illustrates the theory developed for the idealized case, when the measurement collected at each epoch is a realization of a random variable whose distribution follows the maximum-entropy distribution on $\mathbb{RP}^1$, with mean equal to the true subspace at time t and concentration parameter $\eta = 3$. Plotted in Figure 4.3 are the angles (ranging from $-\pi/2$ to $\pi/2$) corresponding to elements of $\mathbb{RP}^1$. In each epoch, the angle of the subspace is rotated by a fixed element of $\mathbf{SO}(2)$ and the estimation is performed by the recursive estimation routine described here.

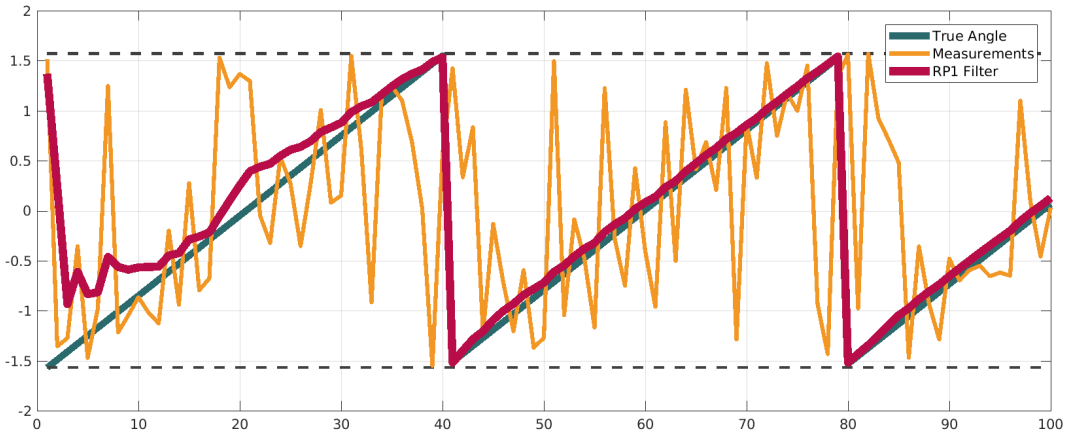

Figure 4.3: Results of an idealized estimation routine on $\mathbb{RP}^1$.

4.3 A Sensing Model

As demonstrated in the presentation, the development of the bayesian estimators of a one-dimensional subspace evolving on real projective space, when the measurement model is idealized, is essentially identical to its complex generalization. The same is not true when the measurements model the practical multistatic sensor system set up at the beginning of this chapter. We focus our attention first on the real case.

4.3.1 On Real Projective Space

The Prediction

We hold on to the initial assumptions made of the state and dynamics in the preceding section: the posterior distribution of the state at time $t-1$ is assumed to follow the maximum-entropy distribution (with respect to a fixed moment) with mean parameter $\widehat{\boldsymbol{s}}_{t-1}$ and concentration parameter κ_{t-1}; the dynamical model assumes the state at time $t-1$ changes via an action of a deterministic and known special orthogonal matrix to yield the state at time t. We begin with the prior probability density of the state at time t.

$$f_{\mathbf{s}_t}(\boldsymbol{s}_t) = c_n(\kappa_{t-1}) \exp\left\{ \kappa_{t-1} \left| (\boldsymbol{R}_t \widehat{\boldsymbol{s}}_{t-1})^\top \, \boldsymbol{s}_t \right| \right\}, \tag{4.5 revisited}$$

and hence the mean and concentration parameters at time t are, respectively,

$$\widehat{\boldsymbol{s}}_t = \boldsymbol{R}\widehat{\boldsymbol{s}}_{t-1}$$

$$\kappa_t = \kappa_{t-1}.$$

Measurements and the Bayesian Update

At each time epoch, we collect now m samples in $\mathbb{R}^{n+1}$ of a noisy signal whose subspace is defined by the state $\mathbf{s}_t \in \mathbb{RP}^n$. According to the measurement model set forth at the start of this chapter, the measurements are modeled as the $(n+1)$-by-m data matrix

$$\boldsymbol{X}_t = \boldsymbol{s}_t \boldsymbol{a}_t + \boldsymbol{\nu_t},$$

where the elements of the m-vector $\boldsymbol{a}$ are the amplitudes of the component of the signal in the m^{th} measurement and $\boldsymbol{\nu}$ is the additive noise matrix whose columns are realizations of identically and independently distributed as gaussian random vectors with zero mean and covariance matrix $\sigma^2 \mathbb{I}_{n+1}$. In this development, we only ever treat σ^2 as a *known quantity* so that we have control over the signal-to-noise ratio in our numerical simulations.

As each column of $\boldsymbol{X}_t$ represents a measurement at time t, we can also write

$$\boldsymbol{x}_t^i = \boldsymbol{s}_t a_t^i + \boldsymbol{\nu}_t^i, \; i = 1, \ldots, m$$

to denote the i^{th} measurement in $\mathbb{R}^{n+1}$. As each $\boldsymbol{x}_t^i$ lives in $\mathbb{R}^{n+1}$, we define $\boldsymbol{z}_t^i \in \mathbb{RP}^n$ to be the radial projection of the $\boldsymbol{x}_t^i$ with positive first component.[2] Figure 4.4 illustrates $m = 5$ measurements in $\mathbb{R}^2$ generated from a one-dimensional subspace, plotted as crosses (×), radially projected onto $\mathbb{RP}^1$, plotted as dots (●).

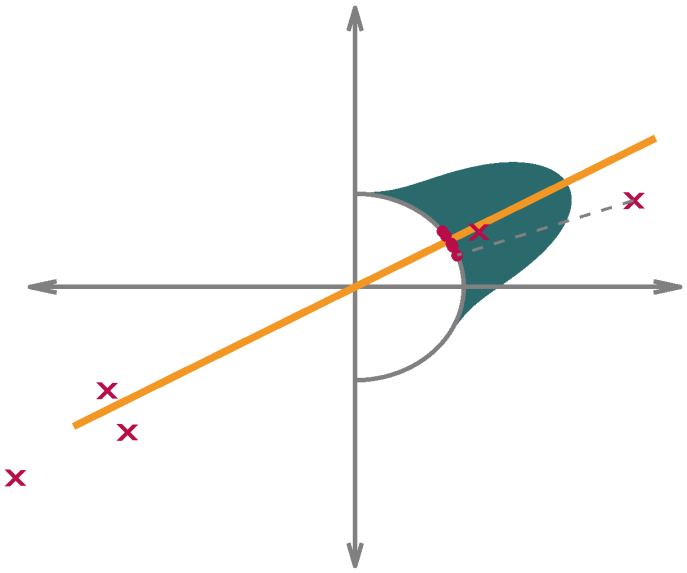

Figure 4.4: Measurements in $\mathbb{R}^2$ generated from a point in $\mathbb{RP}^1$.

For each $i = 1, \ldots, m$ the likelihood function of the i^{th} measurement at time t, $\boldsymbol{z}_t^i$, is

$$f_{\mathbf{z}_t^i}(\boldsymbol{z}_t^i | \boldsymbol{s}_t, a_t^i, \sigma^2) = \int_{\langle \boldsymbol{z}_t^i \rangle} f_{\mathbf{x}_t^i}(\boldsymbol{x}_t^i | \boldsymbol{s}_t, a_t^i, \sigma^2) d\boldsymbol{x}_t^i,$$

where the integral is the Lebesgue line integral over the span of $\boldsymbol{z}_t^i$. We temporarily suppress subscripts and superscripts to simplify notation and curb confusion. As $\boldsymbol{x}$ is a realization of gaussian random vector $\mathbf{x}$ with mean $\boldsymbol{s}a$ and covariance matrix $\sigma^2 \mathbb{I}_{n+1}$ by assumption, the density function corresponding to the distribution of $\mathbf{x}$ is

$$f_{\mathbf{x}}(\boldsymbol{x} | a, \boldsymbol{s}, \sigma^2) = (2\pi)^{-\frac{n+1}{2}} \exp\left\{ -\frac{1}{2\sigma^2} (\boldsymbol{x} - \boldsymbol{s}a)^\top (\boldsymbol{x} - \boldsymbol{s}a) \right\}.$$

[2] Situations where such a projection does not exist are negligible.

Consider the change of coordinates

$$\boldsymbol{x} = \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^n \\ x^{n+1} \end{bmatrix} \mapsto \begin{bmatrix} r\cos\theta^1 \\ r\sin\theta^1\cos\theta^2 \\ \vdots \\ r\sin\theta^1\sin\theta^2\cdots\sin\theta^{n-1}\cos\theta^n \\ r\sin\theta^1\sin\theta^2\cdots\sin\theta^n \end{bmatrix}, \tag{4.8}$$

where $\theta^1 \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, $\theta^i \in (0, \pi)$, for $i = 2, \ldots, n$ and $r \in (-\infty, \infty)$. Figure 4.4 emphasizes the need for the domain of r to include the entire real line when we restrict the domain of θ^1 to be $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. Underpinning these coordinates are the hyperspherical coordinates on $\mathbb{RP}^n \subset \mathbb{R}^{n+1}$. Indeed, using these coordinates for $\boldsymbol{x}$ will yield the following for $\boldsymbol{z}$, again, equal to the radial projection of $\boldsymbol{x}$ with a positive first component:

$$\boldsymbol{z} = \begin{bmatrix} \cos\theta^1 \\ \sin\theta^1\cos\theta^2 \\ \vdots \\ \sin\theta^1\sin\theta^2\cdots\sin\theta^{n-1}\cos\theta^n \\ \sin\theta^1\sin\theta^2\cdots\sin\theta^n \end{bmatrix} \in \mathbb{RP}^n.$$

The Jacobian accompanying the transformation in Equation 4.8 is

$$\mathcal{J}_n = r^n \prod_{j=2}^{n+1} \sin^{n+1-j}\theta^{j-1},$$

from which it follows the joint probability density of $\mathsf{r}, \theta^1, \ldots, \theta^n$ is given by

$$\begin{aligned} f_{\mathsf{r},\theta^1,\ldots,\theta^n}\left(r, \theta^1, \ldots, \theta^n | a, \boldsymbol{s}\right) = (2\pi)^{-\frac{n+1}{2}} |r^n| \exp\left\{-\frac{1}{2}\left(\boldsymbol{x}_\theta - a\boldsymbol{s}\right)^\top \left(\boldsymbol{x}_\theta - a\boldsymbol{s}\right)\right\} \\ \cdot \left|\prod_{j=2}^{n+1} \sin^{n+1-j}\theta^{j-1}\right|. \end{aligned} \tag{4.9}$$

The likelihood function of z_t^i, the i^{th} measurement at time t, is thus computed by marginalizing out r in Equation 4.9. That is (after some simplification),

$$\begin{aligned} f_{\mathsf{z}_t^i}\left(\boldsymbol{z}_t^i \middle| \boldsymbol{s}, a, \sigma^2\right) = (2\pi)^{-\frac{n+1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left(\|a\|^2 - \langle \boldsymbol{z}_t^i, \boldsymbol{s}a\rangle^2\right)\right\} \\ \cdot \int_{\mathbb{R}} |r^n| \exp\left\{-\frac{1}{2\sigma^2}\left(r - \langle \boldsymbol{z}_t^i, \boldsymbol{s}a\rangle\right)^2\right\} dr. \end{aligned} \tag{4.10}$$

For n even, say $n = 2\ell$,

$$f_{\mathbf{z}_t^i}\left(\boldsymbol{z}_t^i \middle| \boldsymbol{s}, a, \sigma^2\right) = 2^{\ell+\frac{1}{2}} e^{-\frac{\alpha^2}{2\sigma^2}} \sigma^{2\ell+1} \Gamma\left(\ell + \frac{1}{2}\right) {}_1F_1\left(\ell + \frac{1}{2}; \frac{1}{2}; \frac{\alpha^2}{2\sigma^2}\right), \tag{4.10a}$$

and for n odd, say $n = 2\ell + 1$,

$$f_{\mathbf{z}_t^i}\left(\boldsymbol{z}_t^i \middle| \boldsymbol{s}, a, \sigma^2\right) = 2^{\ell+1} \sigma^{2(\ell+1)} \Gamma(\ell+1) {}_1F_1\left(-\ell - \frac{1}{2}; \frac{1}{2}; -\frac{\alpha^2}{2\sigma^2}\right), \tag{4.10b}$$

where Γ is the Gamma function and ${}_1F_1$ is the confluent hypergeometric function of the first kind [1]. As these results comprise special functions, Table 4.1 gives a more favorable form of the measurement likelihood function on $\mathbb{RP}^n$ for select (small) n. In each expression, $\alpha = \alpha_t^i = \langle \boldsymbol{z}_t^i, \boldsymbol{s}a \rangle$ and $\operatorname{erf}(\cdot)$ is the standard error function.

n	$f\left(\boldsymbol{z}_t^i \middle\| \boldsymbol{s}, a, \sigma^2\right)$
1	$2\sigma^2 e^{-\frac{\alpha^2}{2\sigma^2}} + \sqrt{2\pi\sigma^2}\alpha \operatorname{erf}\left(\frac{\alpha}{\sqrt{2\sigma^2}}\right)$
2	$\sqrt{2\pi\sigma^2}\left(\alpha^2 + \sigma^2\right)$
3	$\sqrt{2\pi\sigma^2}\left(\alpha^2 + \sigma^2\right) + \left(\alpha^2 + 3\sigma^2\right)\sqrt{2\pi\sigma^2}\alpha \operatorname{erf}\left(\frac{\alpha}{\sqrt{2\sigma^2}}\right)$
4	$\sqrt{2\pi\sigma^2}\left(\alpha^4 + 6\alpha^2\sigma^2 + 3\sigma^2\right)$

Table 4.1: Measurement likelihood functions on $\mathbb{RP}^n$, for select n.

Since introducing a bayesian prior to marginalize out of the likelihood function the nuisance parameter a_t^i presents a formidable calculation, we instead formulate its maximum-likelihood estimate using the maximum-likelihood estimate of $\boldsymbol{s}_t$, as given in Chapter 2. To be clear, we replace a_t^i in Equation 4.10 with $\widehat{a}_t^i$, where

$$\widehat{a}_t^i = \widehat{\boldsymbol{s}}_{\text{ML}}^\top \boldsymbol{x}_t^i.$$

We ultimately fuse together all m measurements to arrive at our likelihood function defined on $\mathbb{RP}^n$:

$$f_{\mathbf{z}_t}(\boldsymbol{z}_t | \boldsymbol{s}_t) = \frac{1}{m} \sum_{i=1}^{m} f_{\mathbf{z}_t^i}(\boldsymbol{z}_t^i | \boldsymbol{s}_t, \sigma^2). \tag{4.11}$$

Similar to the update step presented in Section 4.2.2, the posterior probability density at time t is computed by an application of Bayes' Rule:

$$f_{\mathbf{s}_t}(\boldsymbol{s}_t|\boldsymbol{z}_t) = \frac{f_{\mathbf{z}}(\boldsymbol{z}_t|\boldsymbol{s}_t) f_{\mathbf{s}_t}(\boldsymbol{s}_t)}{\int_{\mathbb{RP}^n} f_{\mathbf{z}_t}(\boldsymbol{z}_t|\boldsymbol{s}_t) f_{\mathbf{s}_t}(\boldsymbol{s}_t) d\boldsymbol{s}_t}. \tag{4.12}$$

It's important to note that this posterior density is certainly not in the class of maximum-entropy distributions on $\mathbb{RP}^n$. We therefore approximate the distribution in Equation 4.12 by a maximum-entropy distribution on $\mathbb{RP}^n$ with matching first moment. As with the moment-matching procedure offered in the idealized scenario, this is computed numerically.

A Numerical Example

A numerical example illustrates the theory developed for the tracker on $\mathbb{RP}^4$, when the measurements collected at each epoch are noisy samples of a signal whose subspace is defined by the true state at time t. The parameters σ^2 and $\boldsymbol{a}_t$ are adjusted to control the signal-to-noise ratio (SNR) for each measurement.

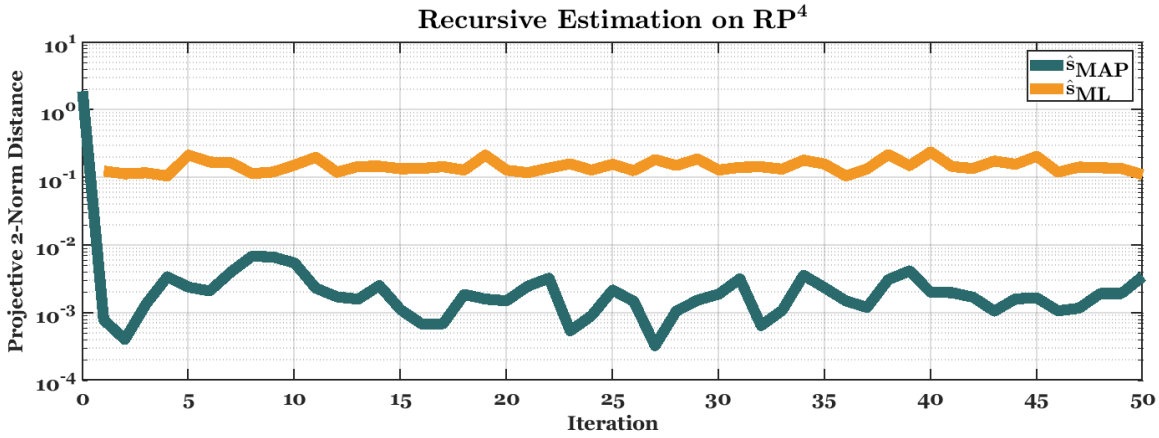

Figure 4.5: Tracking on $\mathbb{RP}^4$ with measurements in $\mathbb{R}^5$ that are samples of a noisy signal whose subspace is defined by an element of $\mathbb{RP}^4$.

The true, veiled subspace is rotated by a deterministic and known element of $\mathbf{SO}(5)$ and the recursive estimation algorithm described here is performed. At each time epoch, the maximum *a posteriori* point estimate $\widehat{\boldsymbol{s}}_{t_{\text{MAP}}}$ of the unknown subspace is computed from its posterior distribution, as well as the maximum-likelihood point estimate $\widehat{\boldsymbol{s}}_{t_{\text{ML}}}$ from the

sensor measurements in $\mathbb{R}^5$. Both point estimates are compared to the true subspace at time t. Plotted on a log scale in Figure 4.5 are the distances between the two point estimates and the truth, defined by the projection 2-norm, i.e., by the sine of the principal angle between the two subspaces.[3] The distance between the recursive maximum *a posteriori* point estimate $\widehat{\boldsymbol{s}}_{t_{\text{MAP}}}$ and the truth decreases with each iteration of bayesian estimation algorithm, while the distance between the maximum-likelihood point estimate $\widehat{\boldsymbol{s}}_{t_{\text{ML}}}$ remains effectively static and significantly inferior to the recursive maximum *a posteriori* point estimate. There are $m = 10$ measurements observed at each time epoch, and the SNR on each sensor is set to approximately 0 dB.

4.3.2 On Complex Projective Space

The Prediction

Recall the initial assumptions in recursively estimating the state $\mathbf{s}_t \in \mathbb{CP}^n$ of a dynamical system at time t: the posterior distribution of the state has been computed using all of the measurements collected at discrete times $1, 2, \ldots, t-1$ and follows the maximum-entropy distribution (with respect to a fixed moment) with parameters $\widehat{\boldsymbol{s}}_{t-1}$ and κ_{t-1}; the dynamical model is completely characterized by a deterministic and known special unitary matrix; and the prior distribution of the state at time t is described by the density

$$f_{\mathbf{s}_t}(\boldsymbol{s}_t) = c_n(\kappa_{t-1}) \exp\left\{\kappa_{t-1} \left|(\boldsymbol{U}_t \widehat{\boldsymbol{s}}_{t-1})^\dagger \, \boldsymbol{s}_t\right|\right\}, \qquad \text{(4.6 revisited)}$$

Measurements and the Bayesian Update

At each time epoch, we collect now m samples in $\mathbb{C}^{n+1}$ of a noisy signal whose subspace is defined by the state $\mathbf{s}_t \in \mathbb{CP}^n$. The measurements are modeled as the $(n+1)$-by-m complex data matrix

$$\boldsymbol{X}_t = \boldsymbol{s}_t \boldsymbol{a}_t + \boldsymbol{\nu_t},$$

[3] As previously mentioned, explicit distance functions on grassmannians will be discussed in detail in Chapter 5.

where the m-vector, $\boldsymbol{a}$, are complex amplitudes of the component of the signal in the m^{th} measurement and $\boldsymbol{\nu}$ is the additive noise matrix whose columns are realizations of identically and independently distributed as complex gaussian random vectors with zero mean and covariance matrix $\sigma^2\mathbb{I}_{n+1}$. Just as in the real case, because each column of $\boldsymbol{X}_t$ represents a measurement at time t, we can also write

$$\boldsymbol{x}_t^i = \boldsymbol{s}_t a_t^i + \boldsymbol{\nu}_t^i,\ i = 1, \ldots, m$$

to denote the i^{th} measurement in $\mathbb{C}^{n+1}$. As each $\boldsymbol{x}_t^i$ lives in $\mathbb{C}^{n+1}$, we define $\boldsymbol{z}_t^i \in \mathbb{CP}^n$ to be the radial projection of the $\boldsymbol{x}_t^i$ whose first component is positive *and real*. Rather than changing coordinates and computing a closed-form expression for the likelihood function $f_{\boldsymbol{z}_t^i}(\boldsymbol{z}_t^i|\boldsymbol{s}, a, \sigma^2)$ of the i^{th} measurement at time t, as we did for the real case, we use the m independent and identically distributed data points $\boldsymbol{z}_t^1, \boldsymbol{z}_t^2, \ldots \boldsymbol{z}_t^m$ to compute maximum-likelihood estimates of parameters of an *assumed* maximum-entropy probability density function, justified by the fact that application of Bayes' Rule for computation of the posterior distribution requires this approximation anyway. For completeness,

$$f_{\boldsymbol{z}_t^i}(\boldsymbol{z}_t^i|\boldsymbol{s}, a, \sigma^2) = c_n(\widehat{\kappa}_{t_{\text{ML}}}) \exp\left\{\widehat{\kappa}_{t_{\text{ML}}} \left|\widehat{\boldsymbol{\mu}}_{t_{\text{ML}}}^\dagger \boldsymbol{z}_t^i\right|\right\}, \tag{4.13}$$

where $\widehat{\boldsymbol{\mu}}_{t_{\text{ML}}}$ and $\widehat{\kappa}_{t_{\text{ML}}}$ denote the maximum-likelihood estimates of the mean value and concentration parameters of the maximum-entropy distribution on $\mathbb{CP}^n$, respectively. With a density of this form, the bayesian update is computed exactly as in Section 4.2.

A Numerical Example

A numerical example illustrates the theory developed for the tracker on $\mathbb{CP}^2$, when the measurements collected at each epoch are noisy samples of a signal whose subspace is defined by the true state at time t. As in the real case, the parameters σ^2 and $\boldsymbol{a}_t$ are adjusted to control the signal-to-noise ratio (SNR) for each measurement.

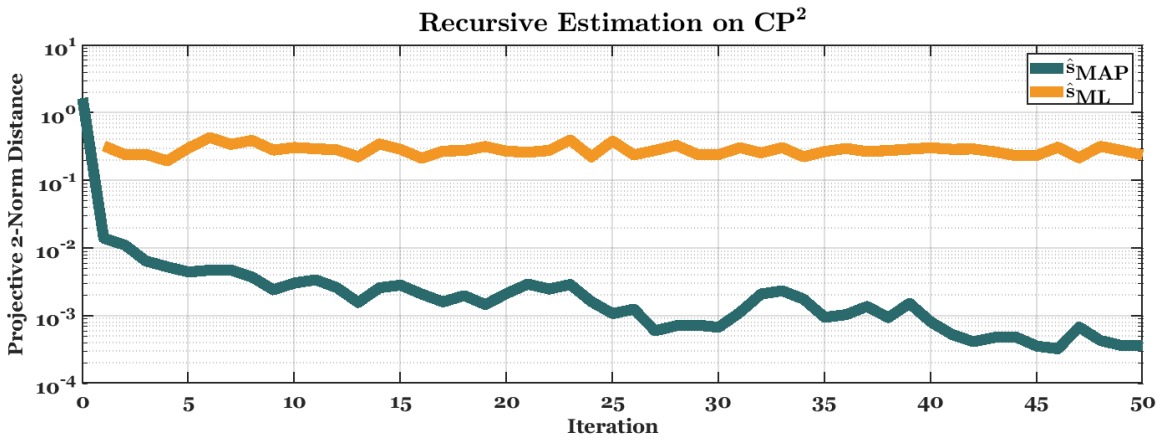

Figure 4.6: Tracking on $\mathbb{CP}^2$ with measurements in $\mathbb{C}^3$ that are samples of a noisy signal whose subspace is defined by an element of $\mathbb{CP}^2$.

Analogous to the numerical example presented in Section 4.3.1, the true, veiled subspace is rotated by a deterministic and known element of $\mathbf{SU}(3)$ and the recursive estimation algorithm described here is performed. At each time epoch, the maximum *a posteriori* point estimate $\widehat{s}_{t_{\text{MAP}}}$ of the unknown subspace is computed from its posterior distribution, as well as the maximum-likelihood point estimate $\widehat{s}_{t_{\text{ML}}}$ from the sensor measurements in $\mathbb{C}^3$. Both point estimates are compared to the true subspace at time t. Plotted on a log scale in Figure 4.6 are the distances between the two point estimates and the truth, defined by the projection 2-norm, i.e., by the sine of the principal angle between the two subspaces. The distance between the recursive maximum *a posteriori* point estimate $\widehat{s}_{t_{\text{MAP}}}$ and the truth decreases with each iteration of bayesian estimation algorithm, while the distance between the maximum-likelihood point estimate $\widehat{s}_{t_{\text{ML}}}$ remains effectively static and significantly inferior to the recursive maximum *a posteriori* point estimate. There are $m = 10$ measurements observed at each time epoch, and the SNR is set to approximately 0 dB on each sensor.

We remark that while the recursive estimation algorithm set forth in this chapter allows for direct implementation in principle, the numerical methods used to generate the results in Figures 4.5 and 4.6 are evidently computationally feasible for only considerably modest n. We address particular numerical limitations of our current methods specifically in Chapter 6.

Chapter 5

TOWARD GRASSMANN MANIFOLDS

The motivating multi-sensor remote sensing application underpinning this thesis and fully characterized at the start of Chapter 4 requires the extension from the case wherein the signal subspace is one-dimensional to the case wherein the signal subspace is k-dimensional ($k \geq 1$). The proper setting is thus the general, complex Grassmann manifold. This chapter lays the groundwork for this ultimate extension, including a self-contained derivation of the differential form for the invariant measure on $\mathbf{G}_{\mathbb{C}}(k,n)$. This is a result originally obtained in [44], however. A complete extension to $\mathbf{G}_{\mathbb{C}}(k,n)$ is presently hamstrung by the challenges that arise in deriving a maximum-entropy distribution to *a priori* model the subspace, as we did for the cases $\mathbb{RP}^n$ and $\mathbb{CP}^n$.

While the dynamical model underlying the recursive estimation algorithms posed on $\mathbb{RP}^n$ and $\mathbb{CP}^n$ in the preceding chapter does not explicitly assume the subspace evolves along a (piecewise) geodesic, this is a most natural generalization that holds significant practical merit. In service of generalizing our dynamical model to be governed by geodesic motions in $\mathbf{G}_{\mathbb{C}}(k,n)$, this chapter begins by providing a collection of self-contained derivations of closed-form analytic expressions for tangent vectors, metrics, and geodesics which can be specialized to the case of complex projective space. Rigorous justification of these sometimes classical results are not readily available in the literature.

In this chapter we focus on the complex Grassmann manifold; most results and formulae carry through *mutatis mutandis* for the real Grassmann manifold.

Definition 5.1. *Let n, k be fixed integers, with $n \geq 0$ and $0 \leq k \leq n$. The **Grassmann manifold**, also referred to as the **grassmannian**, is the set of all k-dimensional linear subspaces of $\mathbb{R}^n$. We denote this space $\mathbf{G}(k,n)$. The set of all k-dimensional linear subspaces of $\mathbb{C}^n$ is defined to be the **complex Grassmann manifold**, or the **complex***

***grassmannian**, and is denoted* $\mathbf{G}_{\mathbb{C}}(k, n)$.

While this definition presents the real and complex Grassmann manifolds as sets, we'll soon see that, as very natural generalizations of projective spaces, they are smooth, compact manifolds of real dimension $k(n-k)$ and $2k(n-k)$, respectively. As the chapter progresses, we will often omit the word "complex" from "complex Grassmann manifold" and "complex grassmannian," except when highlighting a particular distinction.

A point $\boldsymbol{\mathcal{X}}$ on the Grassmann manifold $\mathbf{G}_{\mathbb{C}}(k, n)$ is a linear subspace, which may be specified by an arbitrary basis of k linearly independent, complex n-vectors comprising an n-by-k complex matrix:

$$\boldsymbol{X} = \begin{bmatrix} \vert & & \vert \\ \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_k \\ \vert & & \vert \end{bmatrix}.$$

Partitioning $\boldsymbol{X}$ into submatrices $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ of size k-by-k and $(n-k)$-by-k, respectively, allows us to define the following form for an element of the grassmannian:[1]

$$\boldsymbol{X}' = \begin{bmatrix} \mathbb{I}_k \\ \boldsymbol{Z} \end{bmatrix},$$

where $\boldsymbol{Z} = \boldsymbol{X}_2 \boldsymbol{X}_1^{-1}$, which is then orthonormalized to give the following standard representation of a element in $\mathbf{G}_{\mathbb{C}}(k, n)$.

$$\boldsymbol{X}_{\boldsymbol{Z}} = \begin{bmatrix} \mathbb{I}_k \\ \boldsymbol{Z} \end{bmatrix} \left(\mathbb{I}_k + \boldsymbol{Z}^\dagger \boldsymbol{Z} \right)^{-1/2}. \tag{5.1}$$

Equation 5.1 presents *local coordinates* on the grassmannian from *an orthonormal basis perspective*. We may write

$$\boldsymbol{\mathcal{X}} = \mathrm{col}(\boldsymbol{X}) = \mathrm{col}(\boldsymbol{X}_{\boldsymbol{Z}})$$

[1] We acknowledge that this standard form is viable only for subspaces for which X_1 is invertible! The set of excluded subspaces has measure zero.

to emphasize that the point $\boldsymbol{\mathcal{X}}$ on the grassmannian is the column space of $\boldsymbol{X}$, which is of course equivalent to the column space of $\boldsymbol{X}_{\boldsymbol{Z}}$. When $\boldsymbol{Z} = \mathbf{0}$, $\boldsymbol{\mathcal{X}}$ is the subspace spanned by the first k standard basis vectors of $\mathbb{C}^n$; that is,

$$\boldsymbol{X}_{\mathbf{0}} = \begin{bmatrix} \mathbb{I}_k \\ \mathbf{0} \end{bmatrix}. \tag{5.2}$$

The element $\boldsymbol{X}_{\mathbf{0}}$ will play an important role in the presentation of the grassmannian as a homogeneous manifold.

Associated uniquely to each k-dimensional subspace $\boldsymbol{\mathcal{X}}$ of $\mathbb{C}^n$ is a rank-k, hermitian projection operator $\boldsymbol{P}_{\boldsymbol{\mathcal{X}}}$ in $\mathbb{C}^{n\times n}$, which for any basis $\boldsymbol{X}$ such that $\mathrm{col}(\boldsymbol{X}) = \boldsymbol{\mathcal{X}}$ is given by

$$\boldsymbol{P}_{\boldsymbol{\mathcal{X}}} = \boldsymbol{X}\left(\boldsymbol{X}^{\dagger}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\dagger}.$$

This one-to-one correspondence gives rise to the following proposition.

Proposition 5.1. *Let n, k be fixed integers, with $n \geq 0$ and $0 \leq k \leq n$. Let $\mathbb{P}(k,n)$ be the set of all rank-k, hermitian projection operators on $\mathbb{C}^n$. That is,*

$$\mathbb{P}(k,n) \triangleq \left\{ \boldsymbol{P} \in \mathbb{C}^{n\times n} \middle| \boldsymbol{P}^{\dagger} = \boldsymbol{P}^2 = \boldsymbol{P},\ \mathrm{rank}(\boldsymbol{P}) = k \right\}. \tag{5.3}$$

There exists a bijection between the Grassmann manifold $\mathbf{G}_{\mathbb{C}}(k,n)$ and $\mathbb{P}(k,n)$.

We turn an interested reader to [18] for the construction of a natural bijection between $\mathbf{G}_{\mathbb{C}}(k,n)$ and $\mathbb{P}(k,n)$ and a proof that, with respect to a designated topology, it defines a diffeomeorphism. Regarding the Grassmann manifold as the set of all rank-k, hermitian projection operators has particular advantages: notably a characterization of an embedded submanifold of the manifold consisting of all n-by-n hermitian matrices where the differentiable and riemannian structures are inferred. Our geometric development of the Grassmann manifold hinges instead on a particular Lie group action, inheriting much of the differential and riemannian structure of the Lie group.

The standard form for a basis of a k-dimensional subspace in Equation 5.1 has the following associated standard form for a rank-k, hermitian projection operator:

$$\boldsymbol{P}_{\boldsymbol{X}_{\boldsymbol{Z}}} = \boldsymbol{X}_{\boldsymbol{Z}} \boldsymbol{X}_{\boldsymbol{Z}}^{\dagger} = \begin{bmatrix} \left(\mathbb{I}_k + \boldsymbol{Z}^{\dagger}\boldsymbol{Z}\right)^{-1} & \left(\mathbb{I}_k + \boldsymbol{Z}^{\dagger}\boldsymbol{Z}\right)^{-1} \boldsymbol{Z}^{\dagger} \\ \boldsymbol{Z}\left(\mathbb{I}_k + \boldsymbol{Z}^{\dagger}\boldsymbol{Z}\right)^{-1} & \boldsymbol{Z}\left(\mathbb{I}_k + \boldsymbol{Z}^{\dagger}\boldsymbol{Z}\right)^{-1} \boldsymbol{Z}^{\dagger} \end{bmatrix}, \tag{5.4}$$

which presents local coordinates on the grassmannian from *a hermitian projector perspective*. We introduce

$$\boldsymbol{P}_{\boldsymbol{0}} = \begin{bmatrix} \mathbb{I}_k & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}, \tag{5.5}$$

which is the matrix representation of the canonical projection onto the subspace spanned by $\boldsymbol{X}_{\boldsymbol{0}}$ in Equation 5.2. The element $\boldsymbol{P}_{\boldsymbol{0}}$ plays the role of $\boldsymbol{X}_{\boldsymbol{0}}$ in the presentation of the grassmannian as a homogeneous manifold.

Having chosen the standardized orthonormal basis comprising the columns of $\boldsymbol{X}_{\boldsymbol{Z}}$ for the k-dimensional subspace $\boldsymbol{\mathcal{X}}$ of $\mathbb{C}^n$, there exists a corresponding standardized orthonormal basis for the orthogonal complement $\boldsymbol{\mathcal{X}}^{\perp}$ of $\boldsymbol{\mathcal{X}}$ in $\mathbb{C}^n$ comprising the columns of the n-by-$(n-k)$ matrix we denote $\boldsymbol{X}_{\boldsymbol{Z}}^{\perp}$. In coordinates,

$$\boldsymbol{X}_{\boldsymbol{Z}}^{\perp} = \begin{bmatrix} -\boldsymbol{Z}^{\dagger} \\ \mathbb{I}_{n-k} \end{bmatrix} \left(\mathbb{I}_{n-k} + \boldsymbol{Z}\boldsymbol{Z}^{\dagger}\right)^{-1/2}. \tag{5.6}$$

Together, $\boldsymbol{X}_{\boldsymbol{Z}}$ and $\boldsymbol{X}_{\boldsymbol{Z}}^{\perp}$ define a set of n-by-n unitary matrices

$$\boldsymbol{U}_{\boldsymbol{Z}} = \begin{bmatrix} \boldsymbol{X}_{\boldsymbol{Z}} & \boldsymbol{X}_{\boldsymbol{Z}}^{\perp} \end{bmatrix} = \begin{bmatrix} \left(\mathbb{I}_k + \boldsymbol{Z}^{\dagger}\boldsymbol{Z}\right)^{-1/2} & -\boldsymbol{Z}^{\dagger}\left(\mathbb{I}_{n-k} + \boldsymbol{Z}\boldsymbol{Z}^{\dagger}\right)^{-1/2} \\ \boldsymbol{Z}\left(\mathbb{I}_k + \boldsymbol{Z}^{\dagger}\boldsymbol{Z}\right)^{-1/2} & \left(\mathbb{I}_{n-k} + \boldsymbol{Z}\boldsymbol{Z}^{\dagger}\right)^{-1/2} \end{bmatrix}, \tag{5.7}$$

parameterized by the matrix $\boldsymbol{Z}$. In terms of $\boldsymbol{U}_{\boldsymbol{Z}}$,

$$\boldsymbol{P}_{\boldsymbol{X}_{\boldsymbol{Z}}} = \boldsymbol{U}_{\boldsymbol{Z}} \boldsymbol{P}_{\boldsymbol{0}} \boldsymbol{U}_{\boldsymbol{Z}}^{\dagger}.$$

5.1 The Geometry of the Grassmann Manifold

While it is certainly possible to equip the set of k-dimensional subspaces of $\mathbb{C}^n$ (and the set of rank-k, hermitian projection operators) with a differentiable structure (e.g., Plücker

coordinates [38]), our goal is to instead present the Grassmann manifold as a homogeneous manifold, using the complementary coordinate representations introduced in Equations 5.1 and 5.6, and leverage the geometric theory developed in Chapter 2, which guarantees the existence of a differentiable structure on $\mathbf{G}_{\mathbb{C}}(k,n)$. The presentation of the grassmannian as a homogeneous manifold will help too in describing particular geometric objects, such as tangent vectors, metrics, and geodesics, as we will see.

From the orthonormal basis perspective, let $\mathbf{U}(n)$ act on the grassmannian (defined as a set of k-dimensional subspaces of $\mathbb{C}^n$) via left multiplication; that is for a k-dimensional subspace $\boldsymbol{\mathcal{X}}$ in $\mathbb{C}^n$, let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ be k linearly independent, orthonormal, complex n-vectors spanning $\boldsymbol{\mathcal{X}}$. The group action is given by

$$\boldsymbol{U} \cdot \operatorname{span}\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\} = \operatorname{span}\{\boldsymbol{U}\boldsymbol{x}_1, \ldots, \boldsymbol{U}\boldsymbol{x}_k\}. \tag{5.8}$$

Elementary linear algebra assures that any two orthonormal bases for $\mathbb{C}^n$ are related via a unitary matrix, from which it follows that this action is evidently transitive. We refer to this group action as the *natural action* of $\mathbf{U}(n)$.

From the hermitian projector perspective, the natural action of $\mathbf{U}(n)$ on $\mathbb{P}(k,n)$ manifests as conjugation; that is, for a rank-k, hermitian projection matrix P,

$$\boldsymbol{U} \cdot \boldsymbol{P} = \boldsymbol{U}\boldsymbol{P}\boldsymbol{U}^{\dagger}. \tag{5.9}$$

While perhaps not as straightforward, the natural action on $\mathbb{P}(k,n)$ is transitive. Indeed, given two elements, $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$, in $\mathbb{P}(k,n)$, they must decompose as

$$\boldsymbol{P}_1 = \boldsymbol{V}_1\boldsymbol{P}_{\mathbf{0}}\boldsymbol{V}_1^{\dagger} \text{ and } \boldsymbol{P}_2 = \boldsymbol{V}_2\boldsymbol{P}_{\mathbf{0}}\boldsymbol{V}_2^{\dagger},$$

where $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ are unitary matrices, and $\boldsymbol{P}_{\mathbf{0}}$ is as in Equation 5.5. For $\boldsymbol{V} = \boldsymbol{V}_2\boldsymbol{V}_1^{\dagger} \in \mathbf{U}(n)$,

$$\boldsymbol{V}\boldsymbol{P}_1\boldsymbol{V}^{\dagger} = \boldsymbol{V}_2\boldsymbol{V}_1^{\dagger}\boldsymbol{P}_1\boldsymbol{V}_1\boldsymbol{V}_2^{\dagger} = \boldsymbol{V}_2\boldsymbol{V}_1^{\dagger}\boldsymbol{V}_1\boldsymbol{P}_{\mathbf{0}}\boldsymbol{V}_1^{\dagger}\boldsymbol{V}_1\boldsymbol{V}_2^{\dagger} = \boldsymbol{V}_2\boldsymbol{P}_{\mathbf{0}}\boldsymbol{V}_2^{\dagger} = \boldsymbol{P}_2,$$

ultimately verifying that this action is transitive. The natural actions of $\mathbf{U}(n)$ on $\mathbf{G}_{\mathbb{C}}(k,n)$ and of $\mathbf{U}(n)$ on $\mathbb{P}(k,n)$ give rise to the following familiar result.

Theorem 5.1. *The grassmannian manifold is diffeomorphic to the quotient manifold*

$$\mathbf{U}(n)/\left(\mathbf{U}(k)\times\mathbf{U}(n-k)\right). \tag{5.10}$$

It follows, in particular, that $\mathbf{G}_{\mathbb{C}}(k,n)$ *(similarly,* $\mathbb{P}(k,n)$*) is a compact and connected smooth manifold of real dimension* $2k(n-k)$.

Proof. This result follows precisely from the discussion presented in Section 2.1.2. The stabilizer subgroup of the point defined by $\boldsymbol{X}_{\mathbf{0}}$ (in Equation 5.2) by the natural action of $\mathbf{U}(n)$ on $\mathbf{G}_{\mathbb{C}}(k,n)$ is given by the group of unitary operators that leave the subspaces spanned by the columns of $[\mathbb{I}_k\ \mathbf{0}]^{\top}$ and $[\mathbf{0}_k\ \mathbb{I}_{n-k}]^{\top}$ invariant; this group is clearly isomorphic to $\mathbf{U}(k)\times\mathbf{U}(n-k)$. A special case of the Quotient Manifold Theorem (see [32]) guarantees that the coset space $\mathbf{U}(n)/\mathrm{Stab}_{\boldsymbol{X}_{\mathbf{0}}}$ is a topological manifold endowed with a unique differentiable structure such that the quotient map is a smooth submersion. Because the natural action is transitive, the one-to-one correspondence

$$\beta:\mathbf{U}(n)/\left(\mathbf{U}(k)\times\mathbf{U}(n-k)\right)\longrightarrow\mathbf{G}_{\mathbb{C}}(k,n), \tag{5.11}$$

defined by

$$\beta\left([\boldsymbol{X}]\right)=\boldsymbol{X}\cdot\boldsymbol{X}_{\mathbf{0}}$$

is a smooth diffeomorphism and the result follows. ∎

In the proof of Theorem 5.1, the form of the stabilizer subgroup uses the orthonormal basis representation of a particular element in the grassmannian. The same form can be obtained using the rank-k, hermitian projector representation of a particular element, namely $\boldsymbol{P}_{\mathbf{0}}$ in Equation 5.5, but it is simply not as immediate.

It should be noted that there is yet another way of viewing the Grassmann manifold as a quotient. Instead of taking equivalence classes in the unitary group, one can take the quotient of the *non-compact* Stiefel manifold by the general linear group over $\mathbb{C}$. The geometric treatment under this approach can be found in Absil [2].

For the remainder of the geometric development of the Grassmann manifold, we will identify $\mathbf{U}(n)/\left(\mathbf{U}(k)\times\mathbf{U}(n-k)\right)$ with $\mathbf{G}_{\mathbb{C}}(k,n)$ so that any k-dimensional linear subspace of $\mathbb{C}^n$ will take the form

$$[\boldsymbol{X}] = \boldsymbol{X}\begin{bmatrix} \boldsymbol{U}_k & \mathbf{0} \\ \mathbf{0} & \boldsymbol{U}_{n-k} \end{bmatrix}, \tag{5.12}$$

where $\boldsymbol{U}_k$ is in $\mathbf{U}(k)$, $\boldsymbol{U}_{n-k}$ is in $\mathbf{U}(n-k)$, and $\boldsymbol{X}$ is in $\mathbf{U}(n)$. It is often more convenient to define the unitary matrix $\boldsymbol{X}$ by

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_{\boldsymbol{Z}} & \boldsymbol{X}_{\boldsymbol{Z}}^{\perp} \end{bmatrix},$$

as in Equation 5.7.

When the specification of a basis bears no importance, we will use $\boldsymbol{\mathcal{X}}$ to denote an element of $\mathbf{G}(k,n)$; when the specification is important, we will often identify $\boldsymbol{\mathcal{X}}$ with either the n-by-k matrix $\boldsymbol{X}_{\boldsymbol{Z}}$ or the n-by-n matrix $[\boldsymbol{X}]$. Table 5.2 provides a quick reference to the various representations of an element $\boldsymbol{\mathcal{X}} \in \mathbf{G}_{\mathbb{C}}(k,n)$ we've so far discussed.

Coordinate Representations of $\boldsymbol{\mathcal{X}} \in \mathbf{G}_{\mathbb{C}}(k,n)$	
$\boldsymbol{X}$	$\operatorname{col}(\boldsymbol{X}) = \boldsymbol{\mathcal{X}}$
$\boldsymbol{X}_{\boldsymbol{Z}}$	$\begin{bmatrix} \mathbb{I}_k \\ \boldsymbol{Z} \end{bmatrix}\left(\mathbb{I}_k + \boldsymbol{Z}^{\dagger}\boldsymbol{Z}\right)^{-1/2}$
$[\boldsymbol{X}]$	$\begin{bmatrix} \boldsymbol{X}_{\boldsymbol{Z}} & \boldsymbol{X}_{\boldsymbol{Z}}^{\perp} \end{bmatrix}\begin{bmatrix} \boldsymbol{U}_k & \mathbf{0} \\ \mathbf{0} & \boldsymbol{U}_{n-k} \end{bmatrix}$
$\boldsymbol{P}_{\boldsymbol{\mathcal{X}}}$	$\boldsymbol{X}(\boldsymbol{X}^{\dagger}\boldsymbol{X})^{-1}\boldsymbol{X}^{\dagger}$
$\boldsymbol{P}_{\boldsymbol{X}_{\boldsymbol{Z}}}$	$\boldsymbol{X}_{\boldsymbol{Z}}\boldsymbol{X}_{\boldsymbol{Z}}^{\dagger}$

Table 5.2: Various coordinate representations of $\boldsymbol{\mathcal{X}} \in \mathbf{G}_{\mathbb{C}}(k,n)$.

5.1.1 The Tangent Space

In this section we will completely characterize the tangent space at a point to the Grassmann manifold vis-á-vis the tangent space at a point to the unitary group. As discussed in Chapter 2, the tangent space at the identity element $\mathbb{I}_n$ in $\mathbf{U}(n)$ coincides with the Lie algebra of $\mathbf{U}(n)$, known to be the set of skew-hermitian matrices; that is,

$$T_{\mathbb{I}}\mathbf{U}(n) = \mathfrak{u}(n) \triangleq \left\{ \boldsymbol{A} \in \mathfrak{gl}(n, \mathbb{C}) : \boldsymbol{A}^{\dagger} = -\boldsymbol{A} \right\}.$$

Via left-translation, the tangent space to $\mathbf{U}(n)$ at the point $\boldsymbol{U}$, for any $\boldsymbol{U}$ in $\mathbf{U}(n)$, is

$$T_{\boldsymbol{U}}\mathbf{U}(n) = \boldsymbol{U}\mathfrak{u}(n) = \{\boldsymbol{U}\boldsymbol{A} : \boldsymbol{A} \in \mathfrak{u}(n)\}.$$

Making use of Proposition 2.4 in Chapter 2, the Hilbert-Schmidt inner product defined on the finite-dimensional vector space of all matrices of a fixed dimension by

$$\langle \boldsymbol{U}, \boldsymbol{V} \rangle_{\text{H.S.}} \triangleq \operatorname{tr}\left(\boldsymbol{U}^{\dagger}\boldsymbol{V}\right) \tag{5.13}$$

induces a unique bi-invariant riemannian metric on $\mathbf{U}(n)$ when restricted to skew-symmetric matrices.[2] It follows from (a special case of) the Quotient Manifold Theorem, the standard quotient map

$$\pi : \mathbf{U}(n) \to \mathbf{U}(n)/\left(\mathbf{U}(k) \times \mathbf{U}(n-k)\right) \tag{5.14}$$

is a smooth, surjective submersion, whence, for any $\boldsymbol{U} \in \mathbf{U}(n)$, the tangent space to $\mathbf{U}(n)$ at the point $\boldsymbol{U}$ decomposes into an orthogonal (with respect to this riemannian metric) direct sum of its vertical and horizontal components. That is,

$$\boldsymbol{U}\mathfrak{u}(n) = \mathcal{V}_{\boldsymbol{U}} \oplus \mathcal{H}_{\boldsymbol{U}},$$

where $\mathcal{V}_{\boldsymbol{U}} = \ker d\pi_{\boldsymbol{U}}$ and $\mathcal{H}_{\boldsymbol{U}} = \mathcal{V}_{\boldsymbol{U}}^{\perp}$. Referring to Section 2.1.2,

$$d\pi_{\boldsymbol{U}} : T_{\boldsymbol{U}}\mathbf{U}(n) \longrightarrow T_{\pi(\boldsymbol{U})}\left(\mathbf{U}(n)/\left(\mathbf{U}(k) \times \mathbf{U}(n-k)\right)\right)$$

[2] With no such restriction, the Hilbert-Schmidt inner product induces a bi-invariant metric on $\mathbf{GL}(n)$.

is a surjective linear map whose kernel is precisely the tangent space at $\boldsymbol{U}$ to the orbit of $\boldsymbol{U}$ under the action of $\mathbf{U}(k) \times \mathbf{U}(n-k)$ on $\mathbf{U}(n)$ defined by right translation. That is, when $\boldsymbol{U} = \mathbb{I}$,

$$\begin{aligned} V_{\mathbb{I}} = \ker d\pi_{\mathbb{I}} &= T_{\mathbb{I}} \left((\mathbf{U}(n) \times \mathbf{U}(n-k)) \star \mathbb{I} \right) \\ &= T_{\mathbb{I}} \left(\mathbf{U}(n) \times \mathbf{U}(n-k) \right) \\ &= \mathfrak{u}(n) \oplus \mathfrak{u}(n-k). \end{aligned}$$

It follows that the vertical component of the tangent space to $\mathbf{U}(n)$ at the point $\boldsymbol{U}$, for any $\boldsymbol{U}$ in $\mathbf{U}(n)$, is

$$\mathcal{V}_{\boldsymbol{U}} = \boldsymbol{U}\mathcal{V}_{\mathbb{I}} = \left\{ \boldsymbol{U} \begin{bmatrix} \boldsymbol{V}_k & \mathbf{0} \\ \mathbf{0} & \boldsymbol{V}_{n-k} \end{bmatrix} \middle| \, \boldsymbol{V}_k \in \mathfrak{u}(k), \ \boldsymbol{V}_{n-k} \in \mathfrak{u}(n-k) \right\},$$

and its orthogonal complement, with respect to the riemannian metric induced by the Hilbert-Schmidt inner product, is the horizontal component of the tangent space to $\mathbf{U}(n)$ at the point $\boldsymbol{U}$, given by

$$\mathcal{H}_{\boldsymbol{U}} \triangleq \boldsymbol{U}\mathcal{V}_{\mathbb{I}}^{\perp} = \left\{ \boldsymbol{U} \begin{bmatrix} \mathbf{0} & -\boldsymbol{B}^{\dagger} \\ \boldsymbol{B} & \mathbf{0} \end{bmatrix} \middle| \, \boldsymbol{B} \in \mathbb{C}^{(n-k)\times k} \right\}.$$

As discussed in Section 2.1.2, for a point $[\boldsymbol{X}]$ in $\mathbf{G}_{\mathbb{C}}(k,n)$, we identify each element in the tangent space to $\mathbf{G}_{\mathbb{C}}(k,n)$ at the point $[\boldsymbol{X}]$ with an element of a horizontal space to $\mathbf{U}(n)$ at a point in the fiber above $[\boldsymbol{X}]$. We write

$$T_{[\boldsymbol{X}]}\mathbf{G}_{\mathbb{C}}(k,n) = \mathcal{H}_{\boldsymbol{X}} = \left\{ \boldsymbol{X} \begin{bmatrix} \mathbf{0} & -\boldsymbol{B}^{\dagger} \\ \boldsymbol{B} & \mathbf{0} \end{bmatrix} \middle| \, \boldsymbol{B} \in \mathbb{C}^{(n-k)\times k} \right\}. \tag{5.15}$$

5.1.2 Geodesics

In this section, we derive from the theoretical framework built in Chapter 2, specifically Section 2.1.2, explicit formulae for a geodesic curve on the Grassmann manifold. As we've previously mentioned, the formulae derived here can be found in the literature. In

fact, formulae for geodesics on the Grassmann manifold exist in a variety of forms [2, 8] (and references cited therein), but the explicit development, exploitation, and application of the theoretical framework supporting an approach to solving specific dynamical estimation problems on Lie groups and homogeneous spaces is so fundamental to this thesis that we include our own derivations.

As detailed in Section 2.1.2, the horizontal tangent space of $\mathbf{U}(n)$ plays a critical role in the characterization of metrics and geodesics in the quotient space

$$\mathbf{G}_{\mathbb{C}}(k,n) \cong \mathbf{U}(n)/\left(\mathbf{U}(k) \times \mathbf{U}(n-k)\right).$$

Because the unitary group is equipped with a bi-invariant riemannian metric (induced by the Hilbert-Schmidt inner product defined on $\mathfrak{u}(n)$), Proposition 2.3 and Theorem 2.3 guarantee that geodesics γ in $\mathbf{U}(n)$, with $\gamma(0) = \mathbb{I}$, are precisely the one-parameter subgroups of $\mathbf{U}(n)$, which are characterized exactly by the exponential map. Thus, for a fixed point $\boldsymbol{U}$ in $\mathbf{U}(n)$ and a tangent vector $\boldsymbol{X}$ in $\mathfrak{u}(n)$, the unique geodesic passing through $\boldsymbol{U}$ at $t = 0$ in the direction of $\boldsymbol{X}$ is written

$$\gamma(t) = \boldsymbol{U} \exp(t\boldsymbol{X}). \tag{5.16}$$

The formula describing these horizontal geodesics in $\mathbf{U}(n)$, together with Theorem 2.4 underscore the derivation of one possible characterization of geodesics on the Grassmann manifold in the following theorem.

Theorem 5.2. *For $[\boldsymbol{X}]$ in $\mathbf{G}_{\mathbb{C}}(k,n)$, the unit speed geodesics in $\mathbf{G}_{\mathbb{C}}(k,n)$ passing through the point $[\boldsymbol{X}]$ (at time $t = 0$) are of the form*

$$\gamma : [a,b] \to \mathbf{G}_{\mathbb{C}}(k,n),\ \gamma(t) = \boldsymbol{X} \exp\left(t\boldsymbol{\Omega}\right) \begin{bmatrix} \mathbb{I}_k \\ \mathbf{0} \end{bmatrix}, \tag{5.17}$$

where $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_{\boldsymbol{Z}} & \boldsymbol{X}_{\boldsymbol{Z}}^{\perp} \end{bmatrix}$ is defined in Equation 5.7 and

$$\boldsymbol{\Omega} = \begin{bmatrix} \mathbf{0} & -\boldsymbol{B}^{\dagger} \\ \boldsymbol{B} & \mathbf{0} \end{bmatrix}, \tag{5.18}$$

for some $\boldsymbol{B} \in \mathbb{C}^{(n-k)\times k}$.

Proof. We will provide a sketch proof here as we have provided rigorous justification for all technical intricacies throughout this thesis. Let $\gamma(t)$ be a geodesic in $\mathbf{G}_{\mathbb{C}}(k,n)$ with $\gamma(0) = [\boldsymbol{X}]$. Geodesics in $\mathbf{G}_{\mathbb{C}}(k,n)$ are made explicit via the corresponding geodesics in $\mathbf{U}(n)$, since $\mathbf{G}_{\mathbb{C}}(k,n)$ can be identified with the quotient space $\mathbf{U}(n)/\left(\mathbf{U}(k)\times\mathbf{U}(n-k)\right)$. The geodesics in $\mathbf{U}(n)$, passing through a point $\boldsymbol{X} = \begin{bmatrix}\boldsymbol{X_Z} & \boldsymbol{X_Z^{\perp}}\end{bmatrix} \in \mathbf{U}(n)$, are the one-parameter subgroups of the type $\alpha(t) = \boldsymbol{X}\exp(t\boldsymbol{\Omega})$ for any $\boldsymbol{\Omega}$ in $\mathfrak{u}(n)$. By Theorem 2.4, the geodesic α in $\mathbf{U}(n)$ projects down to a geodesic γ in $\mathbf{G}_{\mathbb{C}}(k,n)$ if and only if $\boldsymbol{\Omega}$ belongs to the orthogonal complement of the vertical space in $\mathfrak{u}(n)$; that is, if and only if $\boldsymbol{\Omega}$ is of the form

$$\boldsymbol{\Omega} = \begin{bmatrix} \mathbf{0} & -\boldsymbol{B}^{\dagger} \\ \boldsymbol{B} & \mathbf{0} \end{bmatrix},$$

for some $\boldsymbol{B} \in \mathbb{C}^{(n-k)\times k}$. Finally, the projection of α to $\mathbf{G}_{\mathbb{C}}(k,n)$ gives γ, using the diffeomorphism β in Equation 5.11,

$$\begin{aligned}\gamma(t) = \beta\left(\alpha(t)\right) &= \boldsymbol{X}\exp\left(t\begin{bmatrix} \mathbf{0} & -\boldsymbol{B}^{\dagger} \\ \boldsymbol{B} & \mathbf{0} \end{bmatrix}\right)\cdot\begin{bmatrix} \mathbb{I}_k \\ \mathbf{0} \end{bmatrix} \\ &= \operatorname{span}\left(\boldsymbol{X}\exp\left(t\begin{bmatrix} \mathbf{0} & -\boldsymbol{B}^{\dagger} \\ \boldsymbol{B} & \mathbf{0} \end{bmatrix}\right)\begin{bmatrix} \mathbb{I}_k \\ \mathbf{0} \end{bmatrix}\right),\end{aligned}$$

as desired. ∎

We note that $\gamma(t)$ in Equation 5.17 is the span of an n-by-k matrix whose columns define a k-dimensional subspace in $\mathbb{C}^n$. The formula for a geodesic that is a rank-k, hermitian projection operator is given by the following (perhaps more common) formula:

$$\gamma(t) = \exp\left(t\boldsymbol{\Omega}\right)\boldsymbol{P}_{\boldsymbol{X_Z}}\exp\left(-t\boldsymbol{\Omega}\right), \tag{5.19}$$

where $\boldsymbol{P}_{\boldsymbol{X_Z}}$ is given in Equation 5.4 and $\boldsymbol{\Omega}$ is given in Equation 5.18.

Computing the Length of a Geodesic

Heretofore, Theorem 5.1 has allowed us to define geodesics on the Grassmann manifold as images of horizontal geodesics in the unitary group. In this section, though, we will appeal

to the stronger result offered by Theorem 5.1: that if $\widetilde{\gamma}(t)$ is a horizontal geodesic in $\mathbf{U}(n)$, not only is its image under the quotient map a geodesic in $\mathbf{G}_{\mathbb{C}}(k,n)$, but its image under the quotient map is a geodesic in $\mathbf{G}_{\mathbb{C}}(k,n)$ of the same length as $\widetilde{\gamma}$.

Let $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ be two points in $\mathbf{G}_{\mathbb{C}}(k,n)$ and let $\gamma(t)$ be a geodesic connecting $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$. Using the result of Theorem 5.2, the geodesic connecting $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ in $\mathbf{G}_{\mathbb{C}}(k,n)$ has the form

$$\gamma(t) = \boldsymbol{X} \exp(t\boldsymbol{\Omega}) \begin{bmatrix} \mathbb{I}_k \\ \mathbf{0} \end{bmatrix},$$

where $\boldsymbol{\Omega}$ is a skew-hermitian, block diagonal matrix of the form

$$\boldsymbol{\Omega} = \begin{bmatrix} \mathbf{0} & -\boldsymbol{B}^\dagger \\ \boldsymbol{B} & \mathbf{0} \end{bmatrix}, \ \boldsymbol{B} \in \mathbb{C}^{(n-k)\times k}$$

and where $\boldsymbol{X}$ is a unitary matrix whose first k columns span the subspace defined by $\boldsymbol{\mathcal{X}}$. The horizontal geodesic in $\mathbf{U}(n)$ whose image is $\gamma(t)$ has the form[3]

$$\widetilde{\gamma}(t) = \boldsymbol{X} \exp(t\boldsymbol{\Omega})$$

and the length of γ can be computed by computing the length of $\tilde{\gamma}$ using the riemannian metric on $\mathbf{U}(n)$ induced by the Hilbert-Schmidt inner product restricted to the vector space of skew-hermitian matrices. Observing first that

$$\frac{d}{dt}\widetilde{\gamma}(t) = \boldsymbol{X}\boldsymbol{\Omega}\exp(t\boldsymbol{\Omega}) = \boldsymbol{X}\exp(t\boldsymbol{\Omega})\,\boldsymbol{\Omega},$$

it follows that

$$\begin{aligned} L(\gamma) = L(\widetilde{\gamma}) &= \int_0^1 \left\langle \frac{d}{dt}\widetilde{\gamma}(t), \frac{d}{dt}\widetilde{\gamma}(t) \right\rangle^{\frac{1}{2}} dt \\ &= \int_0^1 \operatorname{tr}\left(\boldsymbol{\Omega}^\dagger \exp(-t\boldsymbol{\Omega})\, \boldsymbol{U}_{\boldsymbol{Z}}^\dagger \boldsymbol{U}_{\boldsymbol{Z}} \exp(t\boldsymbol{\Omega})\, \boldsymbol{\Omega} \right)^{\frac{1}{2}} dt \\ &= \operatorname{tr}\left(\boldsymbol{\Omega}^\dagger \boldsymbol{\Omega} \right)^{\frac{1}{2}}. \end{aligned}$$

Exploiting the matrix structure of $\boldsymbol{\Omega}$ gives

$$L(\gamma) = \sqrt{2}\operatorname{tr}\left(\boldsymbol{B}^\dagger \boldsymbol{B}\right)^{\frac{1}{2}} = \sqrt{2}\|\boldsymbol{B}\|_{\text{H.S.}}. \tag{5.20}$$

[3]The curve $\tilde{\gamma}$ is often termed the "horizontal lift" of γ.

5.1.3 Distance Functions

In this section, we will present an overview of viable metrics that may be used to compute the distance between two points on the Grassmann manifold. In particular, we will formulate the riemannian distance metric as the metric induced by the two-norm of the principal angles between subspaces. Throughout this discussion, $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ will denote two k-dimensional subspaces of $\mathbb{C}^n$; $\boldsymbol{X}$ and $\boldsymbol{Y}$ will be n-by-k matrices whose columns form orthonormal bases for $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$, respectively; and $\boldsymbol{P_X} = \boldsymbol{X}\boldsymbol{X}^\dagger$ and $\boldsymbol{P_Y} = \boldsymbol{Y}\boldsymbol{Y}^\dagger$ will be the rank-k, hermitian projection operators onto the subspaces $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$, again, respectively.

On a riemannian manifold, the riemannian distance between a pair of points is defined, as in Chapter 2, to be the infimum of the lengths of all curve segments joining the pair. Because a geodesic joining $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ is only locally length minimizing, its length may not realize the riemannian distance between its end points, e.g., the long great circle route between two non-antipodal points on $\mathbf{S}^2$ is a geodesic but is not the shortest geodesic between the points on $\mathbf{S}^2$. In this thesis, we assume two points $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ are sufficiently close so that the length of the geodesic joining $\boldsymbol{\mathcal{X}}$ to $\boldsymbol{\mathcal{Y}}$ does in fact realize the greatest lower bound of the lengths of all possible geodesics connecting $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$.

We let $\gamma : [0, 1] \to \mathbf{G}_{\mathbb{C}}(k, n)$ denote the unit-speed geodesic in $\mathbf{G}_{\mathbb{C}}(k, n)$ emanating from $\boldsymbol{\mathcal{X}}$ and reaching $\boldsymbol{\mathcal{Y}}$ in unit time. Using the result of Theorem 5.2,

$$\boldsymbol{\mathcal{Y}} = \operatorname{span}\left(\begin{bmatrix} \boldsymbol{X_Z} & \boldsymbol{X_Z^\perp} \end{bmatrix} \exp\left(t\boldsymbol{\Omega}\right) \begin{bmatrix} \mathbb{I}_k \\ \mathbf{0} \end{bmatrix} \right),$$

where $\boldsymbol{\Omega}$ is a skew-hermitian, block diagonal matrix of the form

$$\boldsymbol{\Omega} = \begin{bmatrix} \mathbf{0} & -\boldsymbol{B}^\dagger \\ \boldsymbol{B} & \mathbf{0} \end{bmatrix}, \ \boldsymbol{B} \in \mathbb{C}^{(n-k)\times k}.$$

The riemannian distance between $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ is computed as

$$d_{\text{geo}}\left(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}}\right) = \left\|\boldsymbol{\Omega}\right\|_{\text{H.S.}} = \sqrt{2}\left\|\boldsymbol{B}\right\|_{\text{H.S.}}. \tag{5.21}$$

Principal Angles Between Subspaces

Let $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ be two points on the real Grassmann manifold, $\mathbf{G}_{\mathbb{C}}(k,n)$. When $k=1$ and $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ are one-dimensional subspaces of $\mathbb{R}^n$, there is a canonical notion of the distance between $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ defined by the (acute) angle between two n-dimensional vectors whose spans define $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$, respectively. That is,

$$d_{\text{ang}}\left(\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{Y}}\right) \triangleq \arccos\left|\boldsymbol{X}^{\dagger}\boldsymbol{Y}\right|.$$

This notion of a distance metric defined by the acute angle between two (unit) vectors spanning one-dimensional subspaces can be recursively extended to define the *principal angles* between k-dimensional subspaces [15][4].

Definition 5.1. *Let $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ be k-dimensional subspaces of $\mathbb{C}^n$. The **principal angles** (sometimes called **canonical angles**) $\theta_i \in \left[0,\frac{\pi}{2}\right]$, $i=1,\ldots,k$ between $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ are defined recursively by*

$$\cos\theta_i \triangleq \max_{\boldsymbol{x}\in\boldsymbol{\mathcal{X}}}\max_{\boldsymbol{y}\in\boldsymbol{\mathcal{Y}}}\left|\boldsymbol{X}^{\dagger}\boldsymbol{Y}\right| = \left|\boldsymbol{x}_i^{\dagger}\boldsymbol{y}_i\right| \tag{5.22}$$

subject to $\|\boldsymbol{x}\|=\|\boldsymbol{y}\|=1$, $\boldsymbol{x}^{\dagger}\boldsymbol{x}_j=0$ and $\boldsymbol{y}^{\dagger}\boldsymbol{y}_j=0$ for $j=1,\ldots,i-1$.

Importantly, the angles defined by Equation 5.22 satisfy the ordering $0\leq\theta_1\leq\cdots\leq\theta_k\leq\frac{\pi}{2}$. The vector $\boldsymbol{\theta}=[\theta_1\ \cdots\ \theta_k]$ of principal angles between $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ induce several distance metrics on the Grassmann manifold, one of which is equivalent to the geodesic distance expressed in Equation 5.21 as a result of the following proposition [15].

Proposition 5.1. *Let $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ be in $\mathbf{G}_{\mathbb{C}}(k,n)$ and let $\boldsymbol{X}$ and $\boldsymbol{Y}$ denote n-by-k matrices whose columns form orthonormal bases for $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$, respectively. Consider the singular value decomposition of $\boldsymbol{X}^{\dagger}\boldsymbol{Y}$, given as*

$$\boldsymbol{X}^{\dagger}\boldsymbol{Y}=\boldsymbol{U}\boldsymbol{C}\boldsymbol{V}^{\dagger},$$

[4]This reference provides definitions, theorems, and proofs for the real case; the complex case considered here follows from the real case in the obvious way.

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are in $\mathbf{U}(k)$ and $\boldsymbol{C}$ is a k-by-k diagonal matrix $\boldsymbol{C} = \operatorname{diag}(c_1, \ldots c_k)$ of non-negative singular values. Ordering $c_1 \geq c_2 \geq \cdots \geq c_k$, the principal angles $\theta_1, \ldots, \theta_k$ associated with the subspaces $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ satisfy

$$\cos\theta_i = c_i,\ i = 1, \ldots, k.$$

Recognizing the Hilbert-Schmidt norm of a matrix in Equation 5.21 as the sum of the squares of its singular values, and the singular values of the submatrix $\boldsymbol{B}$ of the direction matrix $\boldsymbol{\Omega}$ as the principal angles of the subspaces $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ from Proposition 5.1, the riemannian distance between $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ in $\mathbf{G}_{\mathbb{C}}(k, n)$ is related to the following distance metric induced by their principal angles.

$$d_{\text{geo}}\left(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}}\right) = \sqrt{2}\left\|\boldsymbol{\theta}\right\|_2, \tag{5.23}$$

where $\boldsymbol{\theta}$ is the k-vector of principal angles between $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$.

Furthermore, the singular values of the difference $\boldsymbol{P_X} - \boldsymbol{P_Y}$ are precisely the sines of the principal angles between $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$, which leads us to another possible characterization of the distance between $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$:

$$d_{\text{H.S.}}\left(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}}\right) = \left\|\boldsymbol{P_X} - \boldsymbol{P_Y}\right\|_{\text{H.S.}} = \left\|\sin\boldsymbol{\theta}\right\|_2. \tag{5.24}$$

It is important to note that, since $\sin\theta \leq \theta$ for all θ in $\left(0, \frac{\pi}{2}\right)$, $\left\|\sin\boldsymbol{\theta}\right\|_2 \leq \left\|\boldsymbol{\theta}\right\|_2$, so that this Hilbert-Schmidt distance function is monotonically related to the riemannian distance function.

One final possible characterization of the distance between k-dimensional subspaces we include is the projection two-norm, which is defined to be the spectral norm of the difference between projection matrices $\boldsymbol{P_X}$ and $\boldsymbol{P_Y}$. The distance induce by this norm is called the *subspace distance* in [15] and has been widely adapted in the engineering and image processing literature. It is straightforward to see (a proof is detailed in [15]) that this projection two-norm is equal to the supremum norm of the sine of the principal angles between $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$. That is,

$$d_{\text{P2}}\left(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}}\right) = \left\|\boldsymbol{P_X} - \boldsymbol{P_Y}\right\|_2 = \left\|\sin\boldsymbol{\theta}\right\|_\infty. \tag{5.25}$$

Table 5.3 provides a quick reference to the distance functions on the Grassmann manifold discussed here, as functions of the principal angles between the two subspaces.

Metric Name	Mathematical Expression
Geodesic	$d_{\text{geo}}(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}}) = \sqrt{2}\,\|\boldsymbol{\theta}\|_2$
Hilbert-Schmidt	$d_{\text{H.S.}}(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}}) = \|\sin\boldsymbol{\theta}\|_2$
Projection 2-norm	$d_{\text{P2}}(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}}) = \|\sin\boldsymbol{\theta}\|_\infty$

Table 5.3: Distance functions on the grassmannian, as functions of principal angles between subspaces.

5.2 The Differential Form for the Invariant Measure

For the remainder of this chapter, we return to the coordinate system defined by Equation 5.1. For each $\boldsymbol{\mathcal{X}} \in \mathbf{G}_{\mathbb{C}}(k,n)$, let $\boldsymbol{X}_{\boldsymbol{Z}}$ be the n-by-k-dimensional matrix of the form

$$\boldsymbol{X}_{\boldsymbol{Z}} = \begin{bmatrix} \mathbb{I}_k \\ \boldsymbol{Z} \end{bmatrix} \left(\mathbb{I}_k + \boldsymbol{Z}^\dagger \boldsymbol{Z}\right)^{-1/2}. \tag{5.1 revisited}$$

where $\boldsymbol{Z} \in \mathbb{C}^{(n-k)\times k}$. The corresponding orthonormal basis for the orthogonal complement of $\boldsymbol{\mathcal{X}}$ in $\mathbb{C}^n$ comprises the columns of the following n-by-$(n-k)$ matrix:

$$\boldsymbol{X}_{\boldsymbol{Z}}^{\perp} = \begin{bmatrix} -\boldsymbol{Z}^\dagger \\ \mathbb{I}_{n-k} \end{bmatrix} \left(\mathbb{I}_{n-k} + \boldsymbol{Z}\boldsymbol{Z}^\dagger\right)^{-1/2}. \tag{5.6 revisited}$$

Together, $\boldsymbol{X}_{\boldsymbol{Z}}$ and $\boldsymbol{X}_{\boldsymbol{Z}}^{\perp}$ define a set of n-by-n unitary matrices, indexed by $\boldsymbol{Z}$:

$$\boldsymbol{U}_{\boldsymbol{Z}} = \begin{bmatrix} \boldsymbol{X}_{\boldsymbol{Z}} & \boldsymbol{X}_{\boldsymbol{Z}}^{\perp} \end{bmatrix} = \begin{bmatrix} \left(\mathbb{I}_k + \boldsymbol{Z}^\dagger\boldsymbol{Z}\right)^{-1/2} & -\boldsymbol{Z}^\dagger\left(\mathbb{I}_{n-k} + \boldsymbol{Z}\boldsymbol{Z}^\dagger\right)^{-1/2} \\ \boldsymbol{Z}\left(\mathbb{I}_k + \boldsymbol{Z}^\dagger\boldsymbol{Z}\right)^{-1/2} & \left(\mathbb{I}_{n-k} + \boldsymbol{Z}\boldsymbol{Z}^\dagger\right)^{-1/2} \end{bmatrix}. \tag{5.7 revisited}$$

By analogy with James' invariant measure for the real grassmannian, the differential form for the invariant measure can be defined for points $\boldsymbol{X}_{\boldsymbol{Z}}$ in a neighborhood of a point $\boldsymbol{\mathcal{X}} \in \mathbf{G}_{\mathbb{C}}(k,n)$:

$$\omega_n^k(\boldsymbol{X}_{\boldsymbol{Z}}) \triangleq \left(\frac{i}{2}\right)^{k(n-k)} \left(\bigwedge_{i=1}^{k} \bigwedge_{j=1}^{n-k} \boldsymbol{b}_j^\dagger d\boldsymbol{x}_{\boldsymbol{z}i} \wedge \bigwedge_{i=1}^{k} \bigwedge_{j=1}^{n-k} d\boldsymbol{x}_{\boldsymbol{z}i}^\dagger \boldsymbol{b}_j \right). \tag{5.26}$$

In this expression, the n-vectors $\boldsymbol{x}_{\boldsymbol{z}i}$, for $i = 1, \ldots, k$, comprise the columns in the matrix $\boldsymbol{X}_{\boldsymbol{Z}}$ defining an orthonormal basis for the subspace $\boldsymbol{\mathcal{X}}$; the n-vectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{n-k}$ are the orthonormal n-vectors comprising the matrix $\boldsymbol{X}_{\boldsymbol{Z}}^{\perp}$; and the factor $\left(\frac{i}{2}\right)$ accounts for the fact that decomposing any form into its real and imaginary components, say x and y, respectively, yields the following relation:

$$(x + iy) \wedge (x - iy) = -2i(x \wedge y).$$

Observing that $\bigwedge_{i=1}^{k} \bigwedge_{j=1}^{n-k} \boldsymbol{b}_j^{\dagger} d\boldsymbol{x}_{\boldsymbol{z}i}$ in Equation 5.26 can be interpreted as the exterior product of the $k(n-k)$ elements in the northeast k-by-$(n-k)$ block of the matrix resulting from the product $\boldsymbol{U}_{\boldsymbol{Z}}^{\dagger} d\boldsymbol{U}_{\boldsymbol{Z}}$, we leverage elementary linear algebra in the computation of the invariant volume form on $\mathbf{G}_{\mathbb{C}}(k, n)$ in contrast to the direct derivations for $\mathbb{RP}^n$ and $\mathbb{CP}^n$.

We write $\boldsymbol{U}_{\boldsymbol{Z}}$ as the following product of matrices

$$\begin{aligned}
\boldsymbol{U}_{\boldsymbol{Z}} &= \begin{bmatrix} \mathbb{I}_k & -\boldsymbol{Z}^{\dagger} \\ \boldsymbol{Z} & \mathbb{I}_{n-k} \end{bmatrix} \begin{bmatrix} \left(\mathbb{I}_k + \boldsymbol{Z}^{\dagger}\boldsymbol{Z}\right)^{-1/2} & \mathbf{0} \\ \mathbf{0} & \left(\mathbb{I}_{n-k} + \boldsymbol{Z}\boldsymbol{Z}^{\dagger}\right)^{-1/2} \end{bmatrix} \\
&= \begin{bmatrix} \mathbb{I}_k & -\boldsymbol{Z}^{\dagger} \\ \boldsymbol{Z} & \mathbb{I}_{n-k} \end{bmatrix} \begin{bmatrix} \boldsymbol{Q}_1(\boldsymbol{Z}) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Q}_2(\boldsymbol{Z}) \end{bmatrix}
\end{aligned}$$

to facilitate the computation of $d\boldsymbol{U}_{\boldsymbol{Z}}$:

$$d\boldsymbol{U}_{\boldsymbol{Z}} = \begin{bmatrix} \mathbf{0} & -d\boldsymbol{Z}^{\dagger} \\ d\boldsymbol{Z} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{Q}_1(\boldsymbol{Z}) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Q}_2(\boldsymbol{Z}) \end{bmatrix} + \begin{bmatrix} \mathbb{I}_k & -\boldsymbol{Z}^{\dagger} \\ \boldsymbol{Z} & \mathbb{I}_{n-k} \end{bmatrix} \begin{bmatrix} d\boldsymbol{Q}_1(\boldsymbol{Z}) & \mathbf{0} \\ \mathbf{0} & d\boldsymbol{Q}_2(\boldsymbol{Z}) \end{bmatrix}.$$

It follows that

$$\begin{aligned}
\boldsymbol{U}_{\boldsymbol{Z}}^{\dagger} d\boldsymbol{U}_{\boldsymbol{Z}} &= \begin{bmatrix} \boldsymbol{Q}_1(\boldsymbol{Z}) & \boldsymbol{Q}_1(\boldsymbol{Z})\boldsymbol{Z}^{\dagger} \\ -\boldsymbol{Q}_1(\boldsymbol{Z})\boldsymbol{Z} & \boldsymbol{Q}_2(\boldsymbol{Z}) \end{bmatrix} \begin{bmatrix} d\boldsymbol{Q}_1(\boldsymbol{Z}) & -d\boldsymbol{Z}^{\dagger}\boldsymbol{Q}_2(\boldsymbol{Z}) - \boldsymbol{Z}^{\dagger} d\boldsymbol{Q}_2(\boldsymbol{Z}) \\ d\boldsymbol{Z}\boldsymbol{Q}_1(\boldsymbol{Z}) + \boldsymbol{Z} d\boldsymbol{Q}_1(\boldsymbol{Z}) & d\boldsymbol{Q}_2(\boldsymbol{Z}) \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{Q}_1(\boldsymbol{Z})\boldsymbol{Z}^{\dagger} d\boldsymbol{Z}\boldsymbol{Q}_1(\boldsymbol{Z}) + \boldsymbol{Q}_1^{-1}(\boldsymbol{Z}) d\boldsymbol{Q}_1(\boldsymbol{Z}) & -\left(\boldsymbol{Q}_1(\boldsymbol{Z}) d\boldsymbol{Z}^{\dagger}\boldsymbol{Q}_2(\boldsymbol{Z})\right) \\ \boldsymbol{Q}_2(\boldsymbol{Z}) d\boldsymbol{Z}\boldsymbol{Q}_1(\boldsymbol{Z}) & \boldsymbol{Q}_2(\boldsymbol{Z})\boldsymbol{Z} d\boldsymbol{Z}^{\dagger}\boldsymbol{Q}_2(\boldsymbol{Z}) + \boldsymbol{Q}_2^{-1}(\boldsymbol{Z}) d\boldsymbol{Q}_2(\boldsymbol{Z}) \end{bmatrix}.
\end{aligned}$$

Toward the computation of the invariant measure, we compute the exterior product of the $k(n-k)$ elements in the northeast block:

$$\bigwedge_{i=1}^{k} \bigwedge_{j=1}^{n-k} \left[-\boldsymbol{Q}_1(\boldsymbol{Z}) d\boldsymbol{Z}^{\dagger}\boldsymbol{Q}_2(\boldsymbol{Z})\right]_{ij} = (-1)^{k(n-k)} \bigwedge_{i=1}^{k(n-k)} \operatorname{vec}\left(\boldsymbol{Q}_2(\boldsymbol{Z}) d\boldsymbol{Z}\boldsymbol{Q}_1(\boldsymbol{Z})\right)_i$$

$$
\begin{aligned}
&= (-1)^{k(n-k)} \bigwedge_{i=1}^{k(n-k)} \boldsymbol{Q}_1(\boldsymbol{Z}) \otimes \boldsymbol{Q}_2(\boldsymbol{Z}) \mathrm{vec}\left(d\boldsymbol{Z}\right)_i \\
&= (-1)^{k(n-k)} \det\left(\boldsymbol{Q}_1(\boldsymbol{Z}) \otimes \boldsymbol{Q}_2(\boldsymbol{Z})\right) \bigwedge_{i=1}^{k(n-k)} \mathrm{vec}\left(d\boldsymbol{Z}\right)_i \\
&= (-1)^{k(n-k)} \det\left(\left(\mathbb{I}_k + \boldsymbol{Z}^\dagger \boldsymbol{Z}\right)^{-\frac{1}{2}} \otimes \left(\mathbb{I}_{n-k} + \boldsymbol{Z}\boldsymbol{Z}^\dagger\right)^{-\frac{1}{2}}\right) \bigwedge_{i=1}^{k} \bigwedge_{j=1}^{n-k} dz_{ij} \\
&= (-1)^{k(n-k)} \det\left(\mathbb{I}_k + \boldsymbol{Z}^\dagger \boldsymbol{Z}\right)^{-\frac{k}{2}} \det\left(\mathbb{I}_{n-k} + \boldsymbol{Z}\boldsymbol{Z}^\dagger\right)^{-\frac{(n-k)}{2}} \bigwedge_{i=1}^{k} \bigwedge_{j=1}^{n-k} dz_{ij}.
\end{aligned}
$$

Here, $\mathrm{vec}(\boldsymbol{A})$ denotes the vectorization of the matrix $\boldsymbol{A}$, formed by stacking each column of $\boldsymbol{A}$ into a single column vector, and the symbol $\otimes$ denotes the Kronecker product of two matrices. Finally, applying (a generalization of) the Matrix Determinant Lemma yields

$$
\bigwedge_{i=1}^{k} \bigwedge_{j=1}^{n-k} \left[-\boldsymbol{Q}_1(\boldsymbol{Z}) d\boldsymbol{Z}^\dagger \boldsymbol{Q}_2(\boldsymbol{Z})\right]_{ij} = (-1)^{k(n-k)} \det\left(\mathbb{I}_k + \boldsymbol{Z}^\dagger \boldsymbol{Z}\right)^{-\frac{n}{2}} \bigwedge_{i=1}^{k} \bigwedge_{j=1}^{n-k} dz_{ij},
$$

whence

$$
\omega_n^k(\boldsymbol{X}_{\boldsymbol{Z}}) = 2^{-k(n-k)} i^{k^2(n-k)^2} \det\left(\mathbb{I}_k + \boldsymbol{Z}^\dagger \boldsymbol{Z}\right)^{-n} \bigwedge_{i=1}^{k} \bigwedge_{j=1}^{n-k} dz_{ij} \wedge d\overline{z}_{ij}. \tag{5.27}
$$

As $\mathbb{CP}^n$ *is* $\mathbf{G}_{\mathbb{C}}(1, n+1)$, this special case of Equation 5.27 becomes

$$
\omega_{n+1}^1\left(\boldsymbol{x}_{\boldsymbol{z}}\right) = \left(\frac{i}{2}\right)^n \left(1 + \boldsymbol{z}^\dagger \boldsymbol{z}\right)^{-(n+1)} d\boldsymbol{z} \wedge d\overline{\boldsymbol{z}},
$$

which matches exactly Equation 3.22, the differential form for the invariant measure on $\mathbb{CP}^n$ derived in Chapter 3.

We highlight too that the differential form for the invariant measure on the real Grassmann manifold is defined for points $\boldsymbol{X}_{\boldsymbol{Z}}$ in a neighborhood of a point $\boldsymbol{\mathcal{X}} \in \mathbf{G}(k, n)$ by

$$
\begin{aligned}
\omega_n^k\left(\boldsymbol{X}_{\boldsymbol{Z}}\right) &\triangleq \bigwedge_{i=1}^{k} \bigwedge_{j=1}^{n-k} \boldsymbol{b}_j^\top d\boldsymbol{x}_{\boldsymbol{z}i} \\
&= \det\left(\mathbb{I}_k + \boldsymbol{Z}^\top \boldsymbol{Z}\right)^{-\frac{n}{2}} \bigwedge_{i=1}^{k} \bigwedge_{j=1}^{n-k} dz_{ij}.
\end{aligned} \tag{5.28}
$$

Here, $\boldsymbol{x}_{\boldsymbol{z}i}$ are the k orthonormal n-vectors comprising the columns of the matrix

$$
\boldsymbol{X}_{\boldsymbol{Z}} = \begin{bmatrix} \mathbb{I}_k \\ \boldsymbol{Z} \end{bmatrix} \left(\mathbb{I}_k + \boldsymbol{Z}^\top \boldsymbol{Z}\right)^{-1/2},
$$

with $\boldsymbol{Z}$ now in $\mathbb{R}^{(n-k)\times k}$, and $\boldsymbol{b}_j$ are the $n-k$ orthonormal n-vectors comprising the columns of the matrix

$$\boldsymbol{X}_{\boldsymbol{Z}}^{\perp} = \begin{bmatrix} -\boldsymbol{Z}^{\top} \\ \mathbb{I}_{n-k} \end{bmatrix} \left(\mathbb{I}_{n-k} + \boldsymbol{Z}\boldsymbol{Z}^{\top}\right)^{-1/2},$$

defining a basis for the orthogonal complement of $\boldsymbol{\mathcal{X}}$ in $\mathbb{R}^n$. Equation 5.28 specializes to

$$\omega_{n+1}^1(\boldsymbol{x}_{\boldsymbol{z}}) = \left(1 + \boldsymbol{z}^{\top}\boldsymbol{z}\right)^{-\frac{n+1}{2}} dz^1 \wedge \cdots \wedge dz^n,$$

which matches exactly Equation 3.8, the differential form for the invariant measure on $\mathbb{RP}^n$ derived in Chapter 3.

5.2.1 Integrating the Differential Form

As in Chapter 3, evaluation of the integral of the invariant measures over the entire manifold yields the normalized invariant measure, that is, the uniform distribution on the Grassmann manifold. In the case of the real Grassmann manifold, we have

$$\begin{aligned} \mathrm{Vol}(\mathbf{G}(k,n)) &= \int_{\mathbf{G}(k,n)} \omega_n^k(\boldsymbol{X}_{\boldsymbol{Z}}) \\ &= \int_{\mathbb{R}^{k(n-k)}} \det\left(\mathbb{I}_k + \boldsymbol{Z}^{\top}\boldsymbol{Z}\right)^{-\frac{n}{2}} \prod_{i=1}^{k}\prod_{j=1}^{n-k} dz_{ij}, \end{aligned} \tag{5.29}$$

and in the case of the complex Grassmann manifold we have

$$\begin{aligned} \mathrm{Vol}(\mathbf{G}_{\mathbb{C}}(k,n)) &= \int_{\mathbf{G}_{\mathbb{C}}(k,n)} \omega_n^k(\boldsymbol{X}_{\boldsymbol{Z}}) \\ &= \int_{\mathbb{R}^{k(n-k)}} 2^{-k(n-k)} i^{k^2(n-k)^2} \det\left(\mathbb{I}_k + \boldsymbol{Z}^{\dagger}\boldsymbol{Z}\right)^{-n} \prod_{i=1}^{k}\prod_{j=1}^{n-k} dz_{ij}. \end{aligned} \tag{5.30}$$

In the special cases $\mathbf{G}(1,n+1)$ and $\mathbf{G}_{\mathbb{C}}(1,n+1)$ addressed specifically in Chapter 3, we were able to directly compute these integrals via transformations to hyperspherical coordinatizations, which are presently unique to these cases. It should be noted that expressions for the solutions to Equations 5.29 and 5.30 are known in the literature, but, as demonstrated even by James, results are obtained not by the direct integration of forms but rather by invoking

theorems about volumes of the product spaces and quotient spaces. For completeness,

$$\mathrm{Vol}(\mathbf{G}(k,n)) = \frac{\prod_{\ell=n-k+1}^{n} A_\ell}{\prod_{\ell=1}^{k} A_\ell},$$

and

$$\mathrm{Vol}(\mathbf{G}_{\mathbb{C}}(k,n)) = \frac{\prod_{\ell=n-k+1}^{n} A_{2\ell-1}}{\prod_{\ell=1}^{k} A_{2\ell-1}},$$

where A_ℓ is the volume of the unit sphere in $\mathbb{R}^\ell$:

$$A_\ell = \frac{2\pi^{n/2}}{\Gamma\left(\frac{n}{2}\right)}.$$

While the integrals in Equations 5.29 and 5.30 are undeniably amenable to numerical integration techniques, for reasons discussed in more detail in the next chapter, this is not fully satisfactory for application of our bayesian estimators.

Chapter 6

CONCLUSIONS

Over the past three decades, linear subspace data models and corresponding linear subspace-based methods of signal processing have become standard tools in statistical signal processing. With ongoing advances in technologies that enable multi-channel architectures in telecommunications and sensing, the importance of such models and methods continues to grow. An appealing feature of the linear subspace perspective is that it often admits geometric views of problems and algorithms that can provide valuable insight. The literature is rich with with examples of enlightening geometric interpretations of results that were initially obtained from purely statistical formulations and analysis, and there are also many documented cases in which approaches have arisen from geometric reasoning and later derived using more classical statistical techniques.

Interest in geometrical structures beyond linear manifolds has recently surged among researchers in signal and information engineering [37]. Examples of topics that underpin recent research in signal analysis include: *information geometry*, initially popularized by Amari and his collaborators [3], and *nonlinear dimensionality reduction*, which incorporates the modeling assumption that data in a high-dimensional ambient space lie on (or near) a manifold of much lower dimension [30]. Settings involving identification and characterization of unknown subspaces in high-dimensional vector spaces using noisy measurement data, as exemplified in this thesis, have been of considerable interest for multistatic radar and multiple-input/multiple-output (MIMO) wireless communications over the past decade. We have already noted that grassmannians are the natural setting for such problems.

This thesis has taken a geometric-statistical perspective on a class of estimation problems posed on the Grassmann manifold. This led to the parallel development and imple-

mentation of new mathematical theory and computational algorithms framed by bayesian recursion on projective spaces and to formulations of such problems in the more general grassmannian setting. Our specific contributions include: (1) derivation and statistical analysis of maximum-entropy probability distributions on both real and complex projective space; (2) design of two recursive bayesian estimation algorithms relying on noisy measurements on both real and complex projective space; (3) numerical implementations of these algorithms, which empirically demonstrate that, under mild assumptions, both achieve precise estimation. These contributions serve the goal of developing a geometric theory of signal processing that is as broad in its scope and as precise in its methodologies as multivariate analysis and matrix algebra. This thesis shows promise that this goal can be fully achieved, and the forthcoming discussions show robust opportunities for extensions and generalizations.

6.1 Ongoing Work and New Directions

An earnest effort is made in this section to write for posterity, for the work presented in this thesis and the ensuing discussions should serve to guide the next mathematician as she launches her own research.

6.1.1 Ultimate Extensions to Grassmannians

The demonstrated success of the geometric-statistical approach to the recursive estimation of one-dimensional subspaces of n-dimensional real and complex vector spaces holds great promise for the extension to general k-dimensional subspaces of n-dimensional real and complex vector spaces. As suggested by the motivating practical application underpinning much of this research, ultimate interest in this further application of our approach to general real and complex grassmannians carries increasingly great practical importance. But, our work toward the general grassmannian, as presented in Chapter 5, is but a seedling.

The novelty of the bayesian framework set forth in Chapter 4 for recursive estimation of $\mathbb{RP}^n$ and $\mathbb{CP}^n$ lies in the use of the derived maximum-entropy probability measures as

models for prior probability distributions of an unknown one-dimensional subspace. The derivation of the maximum-entropy probability measures hinges explicitly on particular coordinatizations of $\mathbb{RP}^n$ and $\mathbb{CP}^n$ that grant the ability to explicitly integrate by classical means. The generalization of these hyperspherical and complex hyperspherical coordinatizations to coordinatizations for general grassmannians remains ongoing.

6.1.2 Robust Numerical Integration Methods

While our framework for bayesian recursion indeed allows for direct implementation in principle, current numerical methods are evidently only computationally feasible for considerably modest n. Table 6.2 demonstrates the computational inefficiency and numerical instability of the admittedly unsophisticated numerical techniques presently implemented. Computational time drastically increases and integration accuracy decreases for numerical integration performed with respect to a uniformly discretized domain. As each iteration requires the numerical computation of two integrals, current methods impose heavy runtime demands and offer unreliable results.

	Gridpoints	10^3	10^4	10^5	10^6	10^7	10^8
$\mathbb{RP}^1$	Accuracy	10^{-7}	10^{-11}	10^{-12}	10^{-14}	10^{-15}	10^{-16}
	Runtime (s)	0.00	0.00	0.01	0.05	1	9.35
$\mathbb{RP}^2$	Accuracy	10^{-3}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
	Runtime (s)	0.00	0.00	0.01	0.08	1.01	10.38
$\mathbb{RP}^3$	Accuracy	0.01	10^{-3}	10^{-4}	10^{-4}	10^{-5}	10^{-6}
	Runtime (s)	0.00	0.00	0.01	0.11	1.63	16.15
$\mathbb{RP}^4$	Accuracy	0.05	0.01	0.00	10^{-3}	10^{-4}	10^{-5}
	Runtime (s)	0.00	0.00	0.02	0.19	2.62	27.06
$\mathbb{RP}^5$	Accuracy	0.38	0.04	0.01	10^{-3}	10^{-3}	10^{-4}
	Runtime (s)	0.00	0.00	0.03	0.23	3.65	**34.78**

$\mathbb{RP}^6$	Accuracy	0.23	0.09	0.04	0.01	10^{-3}	10^{-3}
	Runtime (s)	0.00	0.00	0.02	0.45	3.75	**41.39**
$\mathbb{RP}^7$	Accuracy	1.02	0.56	0.11	0.02	0.01	0.01
	Runtime (s)	0.01	0.01	0.03	0.4	6.13	**36.23**
$\mathbb{RP}^8$	Accuracy	2.01	1.75	0.03	0.15	0.04	0.04
	Runtime (s)	0.01	0.04	0.04	0.20	3.8	**149.24**

Table 6.2: Runtime (in seconds) and accuracy to compute *one* integral taken over a uniformly discretized domain whose number of *total* gridpoints increases from 10^3 (left) to 10^8 (right).

There are two computational bottlenecks in integrating the maximum-entropy probability density over $\mathbb{RP}^n$, which continue to manifest in all instances of implemented estimation procedures. First, the normalizing constants of the densities are explicit functions of modified Bessel and Struve functions, both of which are known to present computational difficulties. Although both $I_\nu(z)$ and $L_\nu(z)$ have simple power series expansions that are everywhere convergent, they exhibit approximately periodic behavior which makes the direct use of these power series impractically slow and numerically unstable. Algorithms that make use of their backward recurrence relations for numerical computation exist, but have not been explored. Second, the numerical integration techniques presently implemented use Newton–Cotes formulas (e.g., the trapezoidal rule and Simpson's rule from elementary calculus) that approximate the values of these multi-dimensional integrals with respect to uniformly discretized domains. While we have considered it to be beyond the scope of this research, there certainly exist a host of numerical integration techniques from the numerical analysis literature that would all substantially improve our geometric-computational methods. Implementation and analysis of these techniques should serve as the subject of future research.

6.1.3 *Changes in Dynamical Assumptions*

In both recursive estimation algorithms developed on real and complex projective space in Chapter 4, we chose to model the system dynamics as rotations resulting from an action of elements of the (special) orthogonal and unitary groups. There is incredible practical and theoretical interest in the assumption that a k-dimensional subspace of an n-dimensional vector space evolves on $\mathbf{G}(k,n)$ (or on $\mathbf{G}_{\mathbb{C}}(k,n)$) along a (piecewise) geodesic. In fact, embryonic efforts of the work presented in this dissertation maintained this assumption. Motivated in part by the work of Suvorova et al. [47], we devised a rudimentary but heuristically justified recursive estimation algorithm fit for implementation on the orthogonal and unitary groups, where the system dynamics were governed by geodesics in these spaces. Quick success appeared to hold great promise in descending from $\mathbf{U}(n)$ to the complex grassmannian, regarded of course as the quotient space $\mathbf{U}(n)/\left(\mathbf{U}(k)\times\mathbf{U}(n-k)\right)$. Indeed, many of the differential geometric objects on the complex grassmannian (derivations of formulae for tangent spaces, metrics, geodesics) as developed in Chapter 5 were the fruits of this approach. The inherit nature of the riemannian structure of the grassmannian presented as a primary roadblock in our approach: even modest perturbations in the posterior state estimate at time $t-1$ may have unwieldy effects in the prior state estimate at time t. Indeed, to arrive a prior state estimate at time t, propagation of the posterior state estimate at time $t-1$ along a *known* geodesic requires the general notion of *parallel transport*. Concretely, given one (unit-speed) geodesic curve joining points $\boldsymbol{P}$ and $\boldsymbol{Q}$ and another joining $\boldsymbol{P}$ and $\widehat{\boldsymbol{P}}$, parallel transporting the tangent vector of the curve joining $\boldsymbol{P}$ and $\widehat{\boldsymbol{P}}$ from $\boldsymbol{P}$ to $\boldsymbol{Q}$ results in a tangent vector at $\boldsymbol{Q}$ which generates a (unit-speed) geodesic whose end point is $\widehat{\boldsymbol{Q}}$. Evidently, the length of this geodesic constructed to join $\boldsymbol{Q}$ to $\widehat{\boldsymbol{Q}}$ is different than the length of the geodesic joining $\boldsymbol{P}$ to $\widehat{\boldsymbol{P}}$. Though the difference can be measured by the Riemann curvature tensor [5], the result remains unsatisfactory for application of recursive estimation. While we understand its intricacies, the theoretically principled treatment of the riemannian structure of the grassmannian in the context of recursive estimation is left

as an open problem.

Returning to a dynamical model governed by a transitive group action, we largely focused on actions of the (special) orthogonal and unitary groups on real and complex projective space by *deterministic* elements when there is opportunity to introduce probabilistic formulations. For example, one might assume the rotation matrix defining the dynamical model itself is a random variable on $\mathbf{O}(n)$ (or $\mathbf{U}(n)$) that is perhaps distributed according to the von Mises-Fisher matrix distribution, which can be completely characterized by *known* parameter values. The key stumbling block is that such a stochastic assumption would complicate significantly the Chapman-Kolmogorov equation used to obtain a prior probability density of the state at time t from its posterior density at time $t-1$, possibly presenting as an integral not amenable to direct computation by classical means. Should the convolution of the maximum-entropy density on projective space with the von Mises-Fisher matrix distribution on the orthogonal (or unitary) group be analytically feasible to compute, it will almost certainly result in a prior distribution that would need to be approximated by a maximum-entropy distribution on projective space in order to maintain the recursive nature of estimation presented here. We leave the precise characterization and implementation of a probabilistic formulation of the dynamical model as an open problem.

Even further extensions of this work bear their own intriguing research questions: how robust is the bayesian framework for recursive estimation to *deterministically* incorrect dynamical assumptions? Can our proposed method for recursive estimation in the abstract spaces be amenable to problems formulated in a bayesian fashion with risk functions that involve a decision at each epoch? How applicable is the geometric-statistical machinery developed in this thesis to the topic of registration on manifolds? We view this dissertation as only the beginning. With clear opportunities to both expand the scope of the framework and to apply it to a number of practical applications, there is no end of work in sight.

REFERENCES

[1] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, 1964.

[2] P. A. Absil, R. Mahony, and R. Sepulchre. Riemannian geometry of Grassmann manifolds with a view on algorithmic computation. *Acta Applicandae Mathematica*, 80:199–220, 2004.

[3] S.-I. Amari and H. Nagaoka. *Methods of Information Geometry*. AMS Translations of Mathematical Monographs. American Mathematical Society, 2007.

[4] L. J. Boya, E. C. G. Sudarshan, and T. Tilma. Volumes of compact manifolds. *Reports on Mathematical Physics*, 52(3):401–422, 2003.

[5] É. Cartan. *Geometry of Riemannian Spaces: Lie Groups; History, Frontiers and Applications Series*. Math Sci Press, 1983.

[6] D. Cochran, H. Gish, and D. Sinno. A geometric approach to multiple-channel signal detection. *IEEE Transactions on Signal Processing*, 43(9):2049–2057, 1995.

[7] D. S. Dummit and R. M. Foote. *Abstract Algebra*. Wiley, 2004.

[8] A. Edelman, T. A. Arias, and S. T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications*, 20(2):303–353, 1999.

[9] D. Fearnley-Sander. Hermann Grassmann and the creation of linear algebra. *The American Mathematical Monthly*, 86(10):809–817, 1979.

[10] R. A. Fisher. On an absolute criterion for fitting frequency curves. *Messenger of Matheics*, 41:155–160, 1912.

[11] R. A. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London, A*, 222:309–368, 1922.

[12] G. B. Folland. *Real Analysis: Modern Techniques and Their Applications*. Wiley, 1999.

[13] D. R. Fuhrmann, A. Srivastava, and H. Moon. Subspace tracking via rigid body dynamics. In *Proceedings of the IEEE Workshop on Statistical Signal and Array Processing*, pages 578–581, 1996.

[14] H. Gish and D. Cochran. Generalized coherence. In *International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pages 2745–2748, 1988.

[15] G. H. Golub and C. F. Van Loan. *Matrix Computations (3rd Ed.)*. Johns Hopkins University Press, Baltimore, MD, USA, 1996.

[16] H. Grassmann. *Die Ausdehnungslehre von 1844*. 1862.

[17] B. C. Hall. *Lie Groups, Lie Algebras, and Representations*. Graduate Texts in Mathematics. Springer, 2015.

[18] U. Helmke and J. Moore. *Optimization and Dynamical Systems*. Springer, 1996.

[19] C. W. Helstrom. *Elements of Signal Detection and Estimation*. Prentice-Hall, Inc., 1994.

[20] A. Horn. Doubly stochastic matrices and the diagonal of a rotation matrix. *American Journal of Mathematics*, 76(3):620–630, 1954.

[21] A. T. James. Normal multivariate analysis and the orthogonal group. *The Annals of Mathematical Statistics*, 25(1):40–75, 1954.

[22] E. T. Jaynes. Information theory and statistical mechanics. *Physical Review*, 106(4):620–630, 1957.

[23] A. M. Kagan, V. Yu, and C. R. Rao. *Characterization Problems in Mathematical Statistics*. Wiley, New York, 1973.

[24] R. E. Kalman and R. S. Bucy. New results in linear filtering and prediction theory. *Transactions of the ASME, SER. D, J. Basic Engineering*, pages 95–108, 1961.

[25] H. Karcher. Riemannian center of mass and so called Karcher mean, 2014.

[26] S. M. Kay. *Fundamentals of Statistical Signal Processing, Volume 1: Estimation Theory*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.

[27] S. M. Kay. *Fundamentals of Statistical Processing, Volume 2: Detection Theory*. Prentice-Hall signal processing series. Pearson Education, 2009.

[28] A. N. Kolmogorov, W. L. Doyle, and Ivan Selin. *Interpolation and Extrapolation of Stationary Random Sequences*. RAND Corporation, Santa Monica, CA, 1962.

[29] H. Krim and M. Viberg. Two decades of array signal processing research. *IEEE Signal Processing Letters*, 13(4):67–94, 1996.

[30] J. A. Lee and M. Verleysen. *Nonlinear Dimensionality Reduction*. Information Science and Statistics Series. Springer, 2007.

[31] J. M. Lee. *Riemannian Manifolds: An Introduction to Curvature*. Graduate Texts in Mathematics. Springer, 1997.

[32] J. M. Lee. *Introduction to Smooth Manifolds*. Graduate Texts in Mathematics. Springer, 2003.

[33] E. L. Lehmann and J. P. Romano. *Testing Statistical Hypotheses*. Springer Texts in Statistics. Springer, 2005.

[34] K. V. Mardia. *Statistics of Directional Data*. Probability and Mathematical Statistics. Academic Press, 1972.

[35] K. V. Mardia. Characterizations of directional distributions. In G. P. Patil, S. Kotz, and J. K. Ord, editors, *A Modern Course on Statistical Distributions in Scientific Work*, pages 365–385. Springer Netherlands, Dordrecht, 1975.

[36] B. O'Neill. *Semi-Riemannian Geometry With Applications to Relativity*. ISSN. Elsevier Science, 1983.

[37] A. Pezeshki, L. Scharf, and D. Cochran. Geometry in signal processing and machine learning. Report from an NSF-sponsored workshop, 2017.

[38] J. Plücker. On a new geometry of space. *Transactions of the Royal Society of London*, 1865.

[39] L. Potryagin. Homologies in compact Lie groups. *Matematicheskii Sbornik, Novaya Seriya*, 6(48):389–422, 1939.

[40] B. Riemann. *Grundlagen für eine allgemeine Theorie der Functionen einer veränderlichen complexen Grösse*. PhD thesis, University of Göttingen, Göttingen, Germany, 1851.

[41] L. L. Scharf. *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Addison-Wesley series in electrical and computer engineering. Addison-Wesley Publishing Company, 1991.

[42] L. L. Scharf and B. Friedlander. Matched subspace detectors. *IEEE Transactions on Signal Processing*, 8:2146–2157, 1994.

[43] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.

[44] S. Sirianunpiboon, S. D. Howard, and D. Cochran. Multiple-channel detection of signals having known rank. In *2013 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 6536 – 6540, 2013.

[45] S. T. Smith. *Geometric Optimization Methods for Adaptive Filtering*. PhD thesis, Harvard University., Cambridge, MA, 1993.

[46] M. Spivak. *A Comprehensive Introduction to Differential Geometry*. Publish or Perish, Inc., 1970.

[47] S. Suvorova, S. D. Howard, and B. Moran. Tracking rotations using maximum entropy distributions. *IEEE Transactions on Aerospace and Electronic Systems*, 57(5):2953 – 2968, 2021.

[48] N. M. Temme. *Special Functions: An Introduction to the Classical Functions of Mathematical Physics*. Wiley, 2011.

[49] P. Turaga, A. Veeraraghavan, A. Srivastava, and R. Chellappa. Statistical computations on Grassmann and Stiefel manifolds for image and video-based recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(11):2273–2286, 2011.

[50] H. L. Van Trees. *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. Wiley, 2004.

[51] H. Weyl. *Raum, Zeit, and Matrie*. Berlin, 1923.

[52] H. Weyl. *The Classical Groups: Their Invariants and Representations*. Princeton University Press, 1966.

[53] N. Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. The MIT Press, 1964.

APPENDIX A

SPECIAL FUNCTIONS AND INTEGRALS

The primary objective of this appendix is to present a sampling of useful results regarding special mathematical functions within the limits of what appears in the work presented throughout this thesis. Almost every result presented hereto is lifted from the distinguished Handbook of Mathematical Functions prepared by Abromowitz and Stegun [1]. Results from any other source are cited appropriately.

A.1 The Gamma Function

The Gamma function, first introduced by Euler in the early 1700s, is a common extension of the factorial function to complex numbers. If the real part of a complex number z is strictly positive, the Gamma function is defined by the following integral

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx, \tag{A.1}$$

which converges absolutely.

Theorem 1.1. $\Gamma(z+1) = z\Gamma(z)$.

Proof. Integrating by parts $\Gamma(z+1)$ as defined in Equation A.1, will give

$$\begin{aligned}
\Gamma(z+1) &= \int_0^\infty x^z e^{-x} dx \\
&= -x^z e^{-x}\Big|_0^\infty + \int_0^\infty z x^{z-1} e^{-x} dx \\
&= z \int_0^\infty x^{z-1} e^{-x} dx \\
&= z\Gamma(z).
\end{aligned}$$

■

Given this recurrence relation, together with the value of

$$\Gamma(1) = \int_0^\infty e^{-x} dx = 1,$$

it follows that for all positive integers n,

$$\Gamma(n) = 1 \cdot 2 \cdot 3 \cdots (n-1) = (n-1)! \tag{A.2}$$

An important special case that comes up quite often is the value of $\Gamma\left(\frac{1}{2}\right)$, which can be computed directly from the definition:

$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty \frac{e^{-x}}{\sqrt{x}} dx = 2\int_0^\infty e^{-u^2} du = \sqrt{\pi}. \tag{A.3}$$

Alternatively, we may make use of the following result for non-negative integer values of n:

$$\Gamma\left(\frac{1}{2}+n\right)=\frac{(2n-1)!!}{2^n}\sqrt{\pi}, \tag{A.4}$$

where $n!!$ denotes the double factorial of n and, when $n=0$, $n!!=1$.[1] As Appendix A.1.1 will make use of this result and the double factorial function itself, let us record a few of its important identities here.

- For an even non-negative integer $n=2k$, with $k\geq 0$,

$$n!!=2^k k!. \tag{A.5a}$$

- For an odd positive integer $n=2k-1$, with $k\geq 1$,

$$n!!=\frac{(2k)!}{2^k k!}=\frac{(2k-1)!}{2^{k-1}(k-1)!}. \tag{A.5b}$$

A.1.1 The $\sin^n(x)dx$ Integral

The normalizing constants of the maximum-entropy distributions presented in this thesis notably rely on the value of the integral of the function $\sin^n(x)dx$. While the result can be found in several different sources, this section presents its own derivation.

Theorem 1.2. *For $n\geq 0$,*

$$\int_0^\pi \sin^n\theta d\theta=\frac{\sqrt{\pi}\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n+2}{2}\right)}.$$

Proof. Via integration by parts, we have

$$\begin{aligned}
I_n &= \int_0^\pi \sin^n\theta d\theta \\
&= -\sin^{n-1}\theta\cos\theta\Big|_0^\pi + (n-1)\int_0^\pi \sin^{n-2}\theta\cos^2\theta d\theta \\
&= (n-1)\int_0^\pi \sin^{n-2}\theta d\theta - (n-1)\int_0^\pi \sin^n\theta d\theta.
\end{aligned}$$

From which it follows $nI_n=(n-1)I_{n-2}$, for all integers $n\geq 2$ whence we have two sets of formulae for I_n, depending on whether n is odd or even:

$$\begin{aligned}
I_{2k} &= \frac{2k-1}{2k}\cdot\frac{k-3}{2k-2}\cdots\frac{1}{2}I_0=\frac{(2k-1)!!}{(2k)!!}I_0 \\
I_{2k+1} &= \frac{2k}{2k+1}\cdot\frac{2k-2}{2k-1}\cdots\frac{2}{3}I_1=\frac{(2k)!!}{(2k+1)!!}I_1
\end{aligned} \tag{A.6}$$

[1] The double factorial should never be confused with the factorial function iterated twice!

Equations A.5a and A.4, together with the values of $I_0 = \pi$ and $I_1 = 2$, give us

$$I_{2k} = \frac{\sqrt{\pi}\Gamma\left(k+\frac{1}{2}\right)}{\Gamma\left(k+1\right)},$$

and

$$I_{2k+1} = \frac{\sqrt{\pi}\Gamma\left(k+1\right)}{\Gamma\left(\frac{1}{2}+k+1\right)},$$

from which it follows, regardless of the parity of n,

$$I_n = \frac{\sqrt{\pi}\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n+2}{2}\right)}, \tag{A.7}$$

as desired. ■

A.2 The Modified Bessel Function of the First Kind

The standard definition for the modified Bessel function of the first kind and of order ν is

$$I_\nu(\kappa) = \sum_{r=0}^{\infty} \frac{1}{r!\Gamma\left(r+\nu+1\right)} \left(\frac{\kappa}{2}\right)^{2r+\nu}. \tag{A.8}$$

Its integral representation is usually given by

$$I_\nu(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \cos(\nu\theta) e^{\kappa\cos\theta} d\theta, \tag{A.9}$$

which is equivalent to [48]

$$I_\nu(\kappa) = \frac{2^{-\nu}\kappa^\nu}{\Gamma\left(\nu+1/2\right)\Gamma\left(1/2\right)} \int_0^{\pi} e^{\kappa\cos\theta} \sin^{2\nu}\theta d\theta. \tag{A.10}$$

Substituting $t = \cos\theta$ in Equation A.10 yields the following integral representation, which is suitable for numerical integration procedures when $\nu > -\frac{1}{2}$:

$$I_\nu(\kappa) = \frac{2^{-(\nu-1)}\kappa^\nu}{\Gamma\left(\nu+1/2\right)\Gamma\left(1/2\right)} \int_0^{1} \cosh(\kappa t)(1-t^2)^{\nu-1/2} dt. \tag{A.11}$$

It is immediate from the series representation in Equation A.8, and the standard formulas $\Gamma\left(\frac{3}{2}\right) = \frac{\sqrt{\pi}}{2}$ and $z\Gamma(z) = \Gamma(z+1)$ that, as $\kappa \to 0$,

$$I_\nu(\kappa) \sim \frac{\kappa^\nu}{2^\nu\Gamma\left(\nu+1\right)}, \tag{A.12}$$

for $\nu > -1$. Similarly, as $\kappa \to \infty$,

$$I_\nu(\kappa) \sim \frac{e^\kappa}{\sqrt{2\pi\kappa}} \left(1 - \frac{4\nu^2-1}{8\kappa} + \frac{\left(4\nu^2-1\right)\left(4\nu^2-9\right)}{128\kappa^2}\right), \tag{A.13}$$

for $\nu \in \mathbb{R}$. The modified Bessel function $I_\nu(\kappa)$ satisfies the following relations, where I'_ν is the derivative of $I'_\nu(\kappa)$, with respect to κ:

$$I'_\nu = I_{\nu-1} - \frac{\nu}{\kappa} I_\nu, \tag{A.14a}$$

$$I'_\nu = I_{\nu+1} + \frac{\nu}{\kappa} I_\nu, \text{ and} \tag{A.14b}$$

$$2I'_\nu = I_{\nu-1} + I_{\nu+1}. \tag{A.14c}$$

A.3 Struve Functions

Related to the modified Bessel function of the first kind is the modified Struve function defined by

$$L_\nu(\kappa) = \sum_{r=0}^{\infty} \frac{1}{\Gamma\left(r + \frac{3}{2}\right) \Gamma\left(r + \nu + \frac{3}{2}\right)} \left(\frac{\kappa}{2}\right)^{2r+\nu+1}. \tag{A.15}$$

Analogous to the integral representation for the modified Bessel function of the first kind in Equation A.11 there exists the following integral representation for the modified Struve function, which is suitable for numerical integration procedures when $\nu > -\frac{1}{2}$:

$$L_\nu(\kappa) = \frac{2^{-(\nu-1)} \kappa^\nu}{\Gamma\left(\nu + 1/2\right) \Gamma\left(1/2\right)} \int_0^1 \sinh(\kappa t)(1 - t^2)^{\nu - 1/2} dt. \tag{A.16}$$

The limiting behavior of the modified Struve function is incredibly similar to the limiting behavior of the modified Bessel function of the first kind. As $\kappa \to 0$,

$$L_\nu(\kappa) \sim \frac{\kappa^{\nu+1}}{\sqrt{\pi} 2^\nu \Gamma\left(\nu + \frac{3}{2}\right)} \left(1 + \frac{\kappa^2}{3(2\nu + 3)}\right), \tag{A.17}$$

for $\nu > -\frac{3}{2}$. And, as $\kappa \to \infty$,

$$L_\nu(\kappa) \sim \frac{e^\kappa}{\sqrt{2\pi\kappa}} \left(1 - \frac{4\nu^2 - 1}{8\kappa} + \frac{\left(4\nu^2 - 1\right)\left(4\nu^2 - 9\right)}{128\kappa^2}\right), \tag{A.18}$$

for $\nu \in \mathbb{R}$. The modified Struve function $L_\nu(\kappa)$ satisfies the following relations, where L'_ν is the derivative of $L_\nu(\kappa)$, with respect to κ:

$$L_{\nu-1} - L_{\nu+1} = \frac{2\nu}{\kappa} L_\nu + \frac{(\kappa/2)^\nu}{\Gamma\left(\frac{1}{2}\right) \Gamma\left(\nu + \frac{3}{2}\right)}, \text{ and} \tag{A.19a}$$

$$L_{\nu-1} + L_{\nu+1} = 2L'_\nu + \frac{(\kappa/2)^\nu}{\Gamma\left(\frac{1}{2}\right) \Gamma\left(\nu + \frac{3}{2}\right)} \tag{A.19b}$$

Rearranging terms in Equation A.19b and applying Equation A.19a yields a simplified expression for the derivative of $L_\nu(\kappa)$:

$$L'_\nu = L_{\nu-1} - \frac{\nu}{\kappa} L_\nu. \tag{A.20}$$

Of principal importance to the derivation of normalizing constants for probability densities defined on real and complex projective space is the sum of a modified Bessel function (of the first kind) and a modified Struve function:

$$I_\nu(\kappa) + L_\nu(\kappa) = \frac{2^{-(\nu-1)}\kappa^\nu}{\Gamma\left(\nu + 1/2\right)\Gamma\left(1/2\right)} \int_0^1 e^{\kappa t}(1-t^2)^{\nu-1/2} dt. \tag{A.21}$$

With the substitution $t = \cos\theta$, Equation A.21 has an equivalent form given by

$$I_\nu(\kappa) + L_\nu(\kappa) = \frac{2^{-(\nu-1)}\kappa^\nu}{\Gamma\left(\nu + 1/2\right)\Gamma\left(1/2\right)} \int_0^{\pm\frac{\pi}{2}} e^{\kappa\cos\theta} \sin^{2\nu}\theta d\theta. \tag{A.22}$$

APPENDIX B

COORDINATE TRANSFORMATIONS ON PROJECTIVE SPACES

The notion of coordinates plays an important role in manifold theory. Throughout the document, we presented and worked with two useful coordinate systems for both real and complex projective spaces: standard coordinates from the perspective of orthonormal bases and what we termed *hyperspherical coordinates*. This appendix computes the determinant of the Jacobian matrices corresponding to the transformations between the two coordinate systems on $\mathbb{RP}^n$ and on $\mathbb{CP}^n$.

B.1 Hyperspherical Coordinates on Real Projective Space

As in much of the document, let $\boldsymbol{x}_{\boldsymbol{z}}$ be the $(n+1)$-dimensional unit vector of the form

$$\boldsymbol{x}_{\boldsymbol{z}} = \begin{bmatrix} 1 \\ \boldsymbol{z} \end{bmatrix} \left(1 + \boldsymbol{z}^\top \boldsymbol{z}\right)^{-1/2}, \tag{B.1}$$

with $\boldsymbol{z} = \begin{bmatrix} z^1 & z^2 & \cdots z^n \end{bmatrix}^\top \in \mathbb{R}^n$. In these coordinates, the differential form for the invariant measure on $\mathbb{RP}^n$ is given by

$$\omega^n(\boldsymbol{x}_{\boldsymbol{z}}) = \left(1 + \boldsymbol{z}^\top \boldsymbol{z}\right)^{-(n+1)/2} dz_1 \wedge dz_2 \wedge \cdots \wedge dz_n. \tag{B.2}$$

Consider the transformation

$$\begin{aligned} z_k &= \tan\theta_1 \prod_{i=1}^{k} \sin\theta_i \cos\theta_{k+1} \qquad k < n \\ z_n &= \tan\theta_1 \prod_{i=1}^{n} \sin\theta_i \end{aligned} \tag{B.3}$$

with $\theta_1 \in (-\pi/2, \pi/2)$ and $\theta_i \in (0, \pi)$ for $i > 1$. The transformation law for a differential n-form [32] involves the absolute value of the determinant of the $n \times n$ Jacobian matrix, whose element in the i^{th} row and j^{th} column is

$$\mathcal{J}_n^{ij} = \frac{\partial z_i}{\partial \theta_j}.$$

Because only z_{n-1} and z_n in Equation B.3 are functions of θ_n, calculating the determinant of the Jacobian is most easily done via cofactor expansion along the n^{th} row. Exploiting the matrix structure of the Jacobian in this cofactor expansion yields the following recurrence relation for the equation of the determinant:

$$\det \mathcal{J}_n = (-1)^n \left(\frac{\partial z_{n-1}}{\partial \theta_n} \sin\theta_n \det \mathcal{J}_{n-1} - \frac{\partial z_n}{\partial \theta_n} \cos\theta_n \det \mathcal{J}_{n-1} \right). \tag{B.4}$$

Using this recurrence in a simple proof by mathematical induction will confirm that

$$\det \mathcal{J}_n = \frac{1}{\cos^{n+1}(\theta_1)} \prod_{j=1}^{n} \sin^{n-j}(\theta_j). \tag{B.5}$$

Observing that, with respect to this coordinate transformation,

$$\left(1 + \boldsymbol{z}^\top \boldsymbol{z}\right)^{(n+1)/2} = \cos^{n+1}\left(\arctan\left(\sqrt{\boldsymbol{z}^\top \boldsymbol{z}}\right)\right) = \cos^{n+1}(\theta_1),$$

the differential forms for the invariant measure in the two coordinate systems are related via

$$\left(1+\boldsymbol{z}^\top\boldsymbol{z}\right)^{-(n+1)/2} dz_1\wedge dz_2\wedge\cdots\wedge dz_n = \left|\prod_{j=1}^{n}\sin^{n-j}(\theta_j)\right| d\theta_1\wedge d\theta_2\wedge\cdots\wedge d\theta_n. \tag{B.6}$$

With this coordinate transformation, we can explicitly calculate

$$\begin{aligned}
\mathrm{Vol}(\mathbb{RP}^n) &= \int_{\mathbb{RP}^n}\omega^n(\boldsymbol{x}_{\boldsymbol{z}}) \\
&= \int_{\mathbb{R}^n}\left(1+\boldsymbol{z}^\top\boldsymbol{z}\right)^{-(n+1)/2} dz_1dz_2\cdots dz_n \\
&= \int_0^{\pi}\cdots\int_0^{\pi}\int_{-\pi/2}^{\pi/2}\left|\prod_{j=1}^{n}\sin^{n-j}(\theta_j)\right| d\theta_1d\theta_2\cdots d\theta_n \\
&= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\left|\sin^{n-1}(\theta_1)\right| d\theta_1\prod_{j=2}^{n}\int_0^{\pi}\sin^{n-j}(\theta_j)\,d\theta_j \\
&= \frac{\sqrt{\pi}\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)}\prod_{j=2}^{n}\frac{\sqrt{\pi}\Gamma\left(\frac{1+n-j}{2}\right)}{\Gamma\left(\frac{2+n-j}{2}\right)} \\
&= \frac{\pi^{n/2}\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)\Gamma\left(\frac{n}{2}\right)} \\
&= \frac{\pi^{\frac{n+1}{2}}}{\Gamma\left(\frac{n+1}{2}\right)},
\end{aligned}$$

which is exactly half the volume of the n-sphere, as expected.

B.2 Complex Hyperspherical Coordinates on Complex Projective Space

Analogously, we begin with the $(n+1)$-dimensional complex unit vector of the form

$$\boldsymbol{x}_{\boldsymbol{z}} = \begin{bmatrix} 1 \\ \boldsymbol{z} \end{bmatrix}\left(1+\boldsymbol{z}^\dagger\boldsymbol{z}\right)^{-1/2}, \tag{B.7}$$

with $\boldsymbol{z} = \begin{bmatrix} z_1 & z_1 & \cdots z_n \end{bmatrix}^\top$ in $\mathbb{C}^n$. In these coordinates, the differential form for the invariant measure on $\mathbb{CP}^n$ is given by

$$\begin{aligned}
\omega^n(\boldsymbol{x}_{\boldsymbol{z}}) &= \left(\frac{i}{2}\right)^n\left(1+\boldsymbol{z}^\dagger\boldsymbol{z}\right)^{-(n+1)}\bigwedge_{i=1}^{n}dz_i\bigwedge_{i=1}^{n}d\bar{z}_i \\
&= \left(1+\boldsymbol{u}^\top\boldsymbol{u}+\boldsymbol{v}^\top\boldsymbol{v}\right)^{-(n+1)}\bigwedge_{i=1}^{n}du_i\wedge dv_i,
\end{aligned}$$

where we have introduced $\boldsymbol{z} = \boldsymbol{u} + i\boldsymbol{v}$ for $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^n$. We consider the transformation

$$z_k = e^{i\varphi_k} \tan\theta_1 \prod_{i=2}^{k} \sin\theta_i \cos\theta_{k+1}, \quad k < n$$
$$z_n = e^{i\varphi_n} \tan\theta_1 \prod_{i=2}^{n} \sin\theta_i.$$

In terms of the real and imaginary parts of each z_k, the transformation is

$$u_k = \cos\varphi_k \tan\theta_1 \prod_{i=2}^{k} \sin\theta_i \cos\theta_{k+1}, \quad k < n$$
$$u_n = \cos\varphi_n \tan\theta_1 \prod_{i=2}^{n} \sin\theta_i$$
$$v_k = \sin\varphi_k \tan\theta_1 \prod_{i=2}^{k} \sin\theta_i \cos\theta_{k+1}, \quad k < n$$
$$v_n = \sin\varphi_n \tan\theta_1 \prod_{i=2}^{n} \sin\theta_i$$

As in the case of the hyperspherical coordinate transformation above, the transformation law for a differential n-form involves the absolute value of the determinant of the Jacobian matrix, which for the case of this transformation to *complex* hyperspherical coordinates is a $2n$-by-$2n$ matrix whose structure is given by

$$\mathcal{J}_n = \begin{bmatrix} \frac{\partial u_1}{\partial\theta_1} & \cdots & \frac{\partial u_n}{\partial\theta_1} & \frac{\partial v_1}{\partial\theta_1} & \cdots & \frac{\partial v_n}{\partial\theta_1} \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial u_1}{\partial\theta_n} & \cdots & \frac{\partial u_n}{\partial\theta_n} & \frac{\partial v_1}{\partial\theta_n} & \cdots & \frac{\partial v_n}{\partial\theta_n} \\ \frac{\partial u_1}{\partial\varphi_1} & \cdots & \frac{\partial u_n}{\partial\varphi_1} & \frac{\partial v_1}{\partial\varphi_1} & \cdots & \frac{\partial v_n}{\partial\varphi_1} \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial u_1}{\partial\varphi_n} & \cdots & \frac{\partial u_n}{\partial\varphi_n} & \frac{\partial v_1}{\partial\varphi_n} & \cdots & \frac{\partial v_n}{\partial\varphi_n} \end{bmatrix}.$$

Exploiting this matrix structure will confirm that

$$\det \mathcal{J}_n = \frac{\sin^{2n-1}\theta_1}{\cos^{2n+1}\theta_1} \prod_{i=2}^{n} \sin^{2(n-i)+1}\theta_i \cos\theta_i.$$

Observing that, with respect to this coordinate transformation,

$$\left(1 + \boldsymbol{u}^\top\boldsymbol{u} + \boldsymbol{v}^\top\boldsymbol{v}\right)^{-(n+1)} = \cos^{2(n+1)}\left(\arctan\left(\sqrt{\boldsymbol{u}^\top\boldsymbol{u} + \boldsymbol{v}^\top\boldsymbol{v}}\right)\right) = \cos^{2(n+1)}\left(\theta_1\right),$$

the differential forms for the invariant measure in the two coordinate systems are related via

$$\begin{aligned}\left(1+\boldsymbol{u}^\top\boldsymbol{u}+\boldsymbol{v}^\top\boldsymbol{v}\right)^{-(n+1)}\,du_1\wedge dv_1\wedge\cdots\wedge du_n\wedge dv_n\\ =\left|\prod_{j=1}^{n}\sin^{2(n-j)+1}\theta_j\cos\theta_j\right|d\theta_1\wedge\cdots\wedge d\theta_n\wedge d\varphi_1\wedge\cdots\wedge d\varphi_n.\end{aligned}\tag{B.8}$$

Using this transformation, we can explicitly calculate

$$\begin{aligned}\mathrm{Vol}(\mathbb{CP}^n)=\int_{\mathbb{CP}^n}\omega(\boldsymbol{x})&=\int_{\mathbb{R}^{2n}}\left(1+\boldsymbol{u}^\top\boldsymbol{u}+\boldsymbol{v}^\top\boldsymbol{v}\right)^{-(n+1)}du_1dv_1\cdots du_ndv_n\\ &=\int_0^{2\pi}\cdots\int_0^{2\pi}\prod_{i=1}^{n}d\varphi_i\\ &\quad\cdot\int_0^{\pi/2}\cdots\int_0^{\pi/2}\sin^{2n-1}\theta_1\cos\theta_1d\theta_1\prod_{i=2}^{n}\sin^{2(n-i)+1}\theta_i\cos\theta_id\theta_i\\ &=(2\pi)^n\cdot\int_0^{\pi/2}\sin^{2n-1}\theta_1\cos\theta_1d\theta_1\prod_{i=2}^{n}\int_0^{\pi/2}\sin^{2(n-i)+1}\theta_i\cos\theta_id\theta_i\\ &=(2\pi)^n\cdot\frac{1}{2n}\cdot\frac{1}{2^{n-1}(n-1)!}\\ &=\frac{\pi^n}{n!},\end{aligned}$$

as expected.