Analysis of Discrete Shapes Using Lie Groups

by

Mohamed Salahaldin Hefny

A thesis submitted to the
School of Computing
in conformity with the requirements for
the degree of Doctor of Philosophy

Queen’s University
Kingston, Ontario, Canada
January 2014

Copyright © Mohamed Salahaldin Hefny, 2014
Abstract

Discrete shapes can be described and analyzed using Lie groups, which are mathematical structures having both algebraic and geometrical properties. These structures, borrowed from mathematical physics, are both algebraic groups and smooth manifolds. A key property of a Lie group is that a curved space can be studied, using linear algebra, by local linearization with an exponential map.

Here, a discrete shape was a Euclidean-invariant computer representation of an object. Highly variable shapes are known to exist in non-linear spaces where linear analysis tools, such as Pearson’s decomposition of principal components, are inadequate. The novel method proposed herein represented a shape as an ensemble of homogenous matrix transforms. The Lie group of homogenous transforms has elements that both represented a local shape and acted as matrix operators on other local shapes. For the manifold, a matrix transform was found to be equivalent to a vector transform in a linear space. This combination of representation and linearization gave a simple implementation for solving a computationally expensive problem.

Two medical datasets were analyzed: 2D contours of femoral head-neck cross-sections and 3D surfaces of proximal femurs. The Lie-group method outperformed the established principal-component analysis by capturing higher variability with fewer components. Lie groups are promising tools for medical imaging and data analysis.
Acknowledgments

First of all, I thank my supervisor Dr. Randy Ellis for all the opportunities and challenges that served to organize and measure the best of my energies and skills.

I extend my thanks to Dr. Purang Abolmaesumi, Dr. Timothy Bryant, Dr. Gabor Fichtinger, Dr. David Pichora, Dr. John Rudan, and Dr. James Stewart for their excellent mentorship.

Thanks are also due to my examining committee Dr. Timothy Bryant, Dr. Richard Henriksen, Dr. Parvin Mousavi, Dr. David Rappaport, and Dr. Robert Rohling, for their time and valuable feedback.

Special thanks go to my fellows and collaborators Anna Belkova, Andrew Dickinson, Dan Giles, Dr. Andrew Giles, Dr. Manuela Kunz, Elodie Lugez, Brian Rasquinha, Paul St. John, Gabriel Venne, and Sima Zakani.

Funding support was provided in part by the Natural Sciences and Engineering Research Council of Canada, the International Society for Optics and Photonics, and the Ontario Ministry of Training, Colleges and Universities.

Finally, I would have not lead such a wonderful life or achievements without the way my parents, Salahaldin and Suad, brought me up, without the faith they had in me, and without all the support they gave throughout my life.
Statement of Originality

I state that this is an original body of work and that all references used have been cited.
# Table of Contents

Abstract ................................. i

Acknowledgments ......................... ii

Statement of Originality ................. iii

Table of Contents ........................ iv

List of Tables ............................ vii

List of Figures ........................... viii

Chapter 1:

<table>
<thead>
<tr>
<th>Introduction ......................................................... 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Lie Groups ..................................................... 2</td>
</tr>
<tr>
<td>1.2 Statistical Shape Analysis ..................................... 3</td>
</tr>
<tr>
<td>1.3 Relevant Applications ........................................ 5</td>
</tr>
<tr>
<td>1.4 Thesis Statement .............................................. 7</td>
</tr>
<tr>
<td>1.5 Contributions .................................................. 7</td>
</tr>
<tr>
<td>1.6 Subject Matter Organization ................................... 8</td>
</tr>
</tbody>
</table>
List of Tables

4.1 First 6 singular values of PCA for 2D femoral head-neck cross-sections. 62
4.2 First 6 singular values of Lie-group analysis for 2D femoral head-neck cross-sections. 62
4.3 First 6 singular values of PCA for 3D proximal femur surfaces. 70
4.4 First 6 singular values of Lie-group analysis for 3D proximal femur surfaces. 72
A.1 Examples of groups. 90
List of Figures

1.1 Dürer Face Transforms. ........................................... 4
2.1 Femoral head-neck contour imposed on a CT slice ............ 14
2.2 A proximal-femur triangulated mesh .......................... 15
4.1 PCA singular values and sums for femoral cross-sections. ....... 61
4.2 Lie-group singular values and cumulative sums for modes of variation of femoral head-neck cross-sections. ......................... 63
4.3 Scatter plots of two Lie-group parameters for 2D data. ......... 64
4.4 Scatter plots of three Lie-group parameters for 2D data. ......... 65
4.5 2D contour base shape variation using first component. ........ 66
4.6 2D contour base shape variation using second component. ....... 67
4.7 PCA singular values and sums for femur surfaces. ................ 69
4.8 Lie-group singular values and cumulative sums for modes of variation of proximal femur surfaces. .......................... 71
4.9 Scatter plots of two Lie-group parameters for 3D data. ......... 73
4.10 Scatter plots of three Lie-group parameters for 3D data. ......... 74
4.11 3D surface base shape variation using first component. ........ 75
4.12 3D surface base shape variation using second component. ....... 76
Chapter 1

Introduction

This thesis applies algebra and geometry to the statistical analysis of discrete shapes. Algebra is the branch of mathematics that studies symbolic computations; geometry is the study of shapes and space, including measurements. The intersection of these means that the power of symbolic representation can be used to study properties of spaces. For example, the algebraic description of the ordinary unit sphere is the set of all points that are at a distance 1 from the origin. The geometrical description of a sphere, following Euclid in his seventeenth definition, is the surface of revolution of a semi-circle; if the semi-circle has a unit radius then this procedure generates a unit sphere. Using ordinary Cartesian coordinates, it can easily be verified that a point on a geometrical unit sphere satisfies the algebraic constraint. This idea – that geometry and algebra can sometimes be applied to the same space of points – is in the branch of mathematics known as differential geometry. This thesis explores two high-dimensional geometrical spaces, having 6 and 12 dimensions, that described discrete curves in the plane and discrete surfaces in space.

Beginning with Pearson’s observations in 1901, statistical descriptions of vectors
have evolved into powerful analytical and descriptive tools that are commonly known as methods for analysis of principal components [61]. Relatively recently, since 2003, early steps have been made to bring principal-component analysis to high-dimensional spaces that also have algebraic structures. This dissertation will describe and empirically evaluate effective methods for performing such an analysis on spaces that are widely applicable in shape description.

An important mathematical structure, which is a smooth space of points that have a particularly simple algebraic structure, is called a Lie group. The concept of the Lie group is briefly introduced in the following section; the reader may also refer to the appendix for detailed mathematical descriptions and background.

\section{Lie Groups}

The basic idea that underlies the methods developed in this dissertation is the mathematical structure of a Lie group and its associated Lie algebra. A Lie group, named after the mathematician Sophus Lie, is a structure that has algebraic and topological properties. The algebraic properties are derived from groups and the topological properties are derived from smooth manifolds. A Lie algebra is constructed by linearizing a Lie group, done by expanding the group combinatorial operator about the coordinates of the group elements at any given group element. Such a linearization of the Lie group forms a new set of elements that are its Lie algebra. A Lie algebra is a linear vector space, so the group operation is reduced to being vector addition; it is this local linearization that allows the study of non-linear properties using simple linear tools.

A matrix Lie group is a Lie group of matrices that are also a smooth manifold.
The concept of the matrix manifold is century-old [72] but has seen increasing use in mathematics and physics. This concept is the foundation for many applications in mathematical physics, from elementary particles to cosmological principles [75]. Only in the last decade have Lie groups started to attract the attention of computer scientists, primarily in computer vision and image analysis. The use of a matrix manifold, in particular, makes it possible to use differential-geometry tools to analyze non-linear spaces.

In this dissertation, a novel usage of matrix Lie groups is introduced to analyze discrete shapes. The idea is to locally linearize Lie groups and apply methods from the realm of statistical shape analysis.

### 1.2 Statistical Shape Analysis

Statistical shape analysis is the study of random observations of shape [19]. The mathematical analysis of biological shapes, including human anatomy, is an ancient concept: one of the earliest remarkable works was the passage by the Roman architect and engineer Vitruvius about “geometry and human proportions”.

Based on this earlier writing, and empirical observations of two to three hundred people, Albrecht Dürer described the human body parts as fractions of the total height and included this description in his famous work “Four Books on Human Proportion” [60]. He also derived how those proportions can be modified geometrically. Figure 1.1 shows the Dürer face transform. The work of Dürer influenced D’arcy Thompson in writing his book “On Growth and Form” [82]. Thompson explored the use of simple mathematical transformations to describe the difference in form between related animals.
1.2.1 Shapes

Shape, in its simplest formulation, is the geometrical description of an object that is invariant to translation, rotation, and scaling transformations [44]. Although identifying similar shapes is an intuitively easy task for humans, it is not so for computers. The latter need a discrete representation of shape and a processing framework, with either prior knowledge or a mathematical model. Because the shapes used as empirical examples in this dissertation did not have simple mathematical descriptions, methods were needed from the realm of statistical shape models.

1.2.2 Statistical Shape Models

A statistical shape model is a mathematical structure representing the shape variability. Such a model uses measures of statistics that come from geometrical descriptions
of samples of a population. A highly influential development was the point distribution model by Cootes et al. [17], which has been widely used in many general applications. A point distribution model relies on principal component analysis [61, 39] and assumes that shapes can be described in a Euclidean space. Shapes that are better described in a non-linear space confound such analytical methods.

1.3 Relevant Applications

This work was motivated by close collaborations of the author with orthopedic surgeons in Kingston General Hospital. The high variability of bone shapes of patients is challenging and motivating to mathematically analyze. The initial anatomical target of the work described here was the proximal femur, which is the uppermost portion of the largest bone in the human body.

The proximal femur is known to be highly variable among normal human sub-populations according to genetic, environmental, and lifestyle factors. This variability is quantified by anthropometric measurements such as the femoral head offset, femoral head diameter, femoral head position, femoral neck diameter, and neck-shaft angle.

Bone shape analysis is a key tool for researchers in osteology and orthopedic surgery. Proper definition of the variations in normal anatomy, and patterns of abnormal morphologies, can assist in surgical planning and evaluation of outcomes. Accurate classification of groups based on race, sex, and activity is essential for analyzing the development and growth of our Homo Sapiens species.

Statistical shape models are widely used in 3D segmentation [38], including application to the proximal femur [78]. Because segmentation is an opportune application
of statistical shape models, three potential applications have been noted in the literature: implant design for computer assisted surgery; pathological variability for computer aided diagnosis; and bone morphology for forensic anthropology.

1.3.1 Computer Assisted Surgery

Understanding the shape of the proximal femur is important in implant design for orthopedic surgery. Orthopedic surgeons have long studied the geometry of the femur using anterior-posterior radiographs for precise fit of bone-implant with biological fixation [58]; as imaging technologies improved, 3D radiographs [71] were used. More recent studies have described the anatomy of the proximal femur in a normal population [83]. There has been intensive research effort in describing the proximal femur shape, with recent work observing wide variations with age and sex [13], and bone remodeling in adults within a short period of time based on activity [21]. Statistical shape modeling may also be a promising tool in patient-specific implant design.

1.3.2 Computer-Aided Diagnosis

Knowing the pathological variability of anatomy is a key part of computer aided diagnosis. For example, femoro-acetabular impingement is a condition of abnormal friction, or reduced range of motion, in the hip joint that is attributed to improper contact between the proximal femur and rim of the acetabulum. Impingement diagnosis and treatment has a growing interest in the orthopedic research community [37]. Nötzli et al. [59] were prominent in proposing the use of geometrical analysis of the contour of the femoral head-neck junction as a predictor for the risk of impingement. Although Nötzli’s method has been the standard method for the last decade, it is the
subject of recent debate [48, 7, 68, 6]. The use of statistical shape models may lead to advanced analysis and diagnosis.

1.3.3 Forensic Anthropology

Forensic anthropology studies, over the past twenty years, have shown that the femur varies according to race and gender [29]. Other studies have shown that sexual dimorphism, temporal homogeneity, and geographical homogeneity contribute to variability [86]. Such research may benefit from automated methods that enable processing large amounts of data, and that describe implicit features in the shape in addition to explicit measurements. Statistical shape analysis is a promising tool for quantitative anthropology.

1.4 Thesis Statement

This dissertation described the use of algebra and geometry to analyze discrete shapes. It uses the mathematics of matrix Lie groups to describe certain matrices as simultaneously elements and operators in the group. This enabled otherwise tedious computations to be performed rather simply.

1.5 Contributions

The thesis of this dissertation had five main contributions:

- Recognizing that certain matrices, called homogeneous transforms, were simultaneously operators and elements of Lie groups.
• Providing a new shape parametrization technique of $2D$ contours and $3D$ surfaces as homogeneous matrices.

• Deriving a novel theoretical foundation for analyzing discrete shapes.

• Developing a numerical framework for implementation of the derived theory.

• Validating an implementation using clinical $2D$ and $3D$ datasets.

Throughout this dissertation, four distinct fonts were used to clarify distinctions among closely related concepts:

Groups: Mathematical bold upper case (e.g., $G$).

Group elements: Italic lower case (e.g., $g$).

Matrices and Sets: Italic upper case (e.g., $G$).

Topologies: Mathematical calligraphy (e.g., $\mathcal{G}$).

1.6 Subject Matter Organization

The subject matter has a straightforward organization.

Chapter 2 reviews related literature. It gives an overview of shape analysis and statistical shapes models construction framework. A broad review of each phase of the framework is discussed.

Chapter 3 develops Lie-group analysis of discrete shapes. Detailed derivations of the method and numerical considerations are discussed.

Chapter 4 applies the implemented method to clinical datasets. The datasets are $2D$ contours and $3D$ surfaces, extracted from CT images of patients.
Chapter 5 concludes the dissertation by summarizing findings and contributions. It also discusses limitations and future work.

Appendix material includes relevant background theory for the development of the proposed method. The background, which covers aspects of algebra and geometry, assumes that the reader has a background in computer science.

The use of differential geometry is a promising avenue for understanding data in their natural spaces. Application to statistical analysis of discrete shapes is the spearhead that is discussed in this dissertation. Hence, statistical shape analysis literature is reviewed in the next chapter.
Chapter 2

Review of Related Literature

Statistical shape analysis is the study of random observations of shape [19]. Shape is the geometrical description of an object that is invariant to global translation, rotation, and scaling transformations [44]. A statistical shape model (SSM) is a mathematical structure representing shape variability. In this context, two objects have the same shape if they can be transformed to each other by applying Euclidean transformations. Object shapes are variously described as similar to known shapes or as a composition of a set of such shapes. For example, Gray’s Anatomy describes the proximal femur as having four parts of its shape: the head, neck, greater trochanter, and lesser trochanter [34].

This chapter is a broad review of statistical shape analysis and other literature related to the thesis work. For a detailed study of statistical shape analysis we refer the reader to Dryden and Mardia [19]; for a comprehensive review of 3D statistical shape models for segmentation we refer the reader to Heimann and Meinzer [38].

The rest of the chapter has a simple structure. Section 2.1 gives an overview of the general framework of statistical shape model construction. Section 2.2 covers major
data sources including acquisition technologies and shape extraction methodologies. Section 2.3 reviews shape representation details. An overview of shape alignment and required correspondence establishing methods are discussed in section 2.4. Variability analysis is discussed in details in section 2.5. The chapter concludes with a summary in Section 2.6.

2.1 A Framework for Statistical Shape Modeling

Construction of a statistical shape model has three major steps. First, shapes are represented in a common format. Second, shapes are aligned by removing global translations, rotations, and scaling. Third, statistical-variability analysis is applied to the population.

As one of many examples, the literature of SSM construction of proximal femurs has been active in the past ten years. Tang et al. [80] built a Point Distribution Model (PDM) for proximal and distal femurs using a hybrid 3D SSM that combined local and global models using a set of connected spheres. Zheng et al. [92] reconstructed a patient-specific proximal femur surface with a PDM and used it to calibrate 2D fluoroscopic images; they also used a multi-level PDM and intra operative sparse data to reconstruct a patient-specific model of the bone [91]. Kainmueller et al. [42] introduced a composite SSM of the femur and pelvis with a flexible hip joint for automatic segmentation purposes. Yokota et al. [88] developed a hierarchical multi-object SSM for pelvis and femur segmentation. Schumann et al. [74] conducted a validation study of reconstructing a surface model of the proximal femur from 2D X-ray radiographs and a SSM. Whitmarsh et al. [87] constructed a combined shape and bone mineral density statistical model. Nicolella et al. [56] constructed a combined
parametric proximal femur finite element model, based on a statistical shape and density model.

2.2 Shape Acquisition

Anatomical shapes, in 2D or 3D, are shapes extracted respectively from planar or volumetric medical images. Medical imaging modalities vary according to physics properties and interactions with tissue. Images are digitized as 2D or 3D arrays of intensity scalar values that are named pixels in 2D or voxels in 3D [32].

Fluoroscopy, a standard intra-operative bone-imaging technology for decades, is a 2D imaging technique that uses X-rays to reveal anatomical structures. Fluoroscopic X-ray images are formed in part by the atomic composition and the density of tissue [35]. A basic fluoroscope consists of a source that emits X-rays and a fluorescent screen by which the anatomy is imaged. Older fluoroscopes replaced a screen with an X-ray image intensifier and a recording camera; at the time of writing, solid-state X-ray detectors are entering common use.

With advancements in medical-computing devices, data acquisition has substantially shifted to 3D imaging. Computed tomography (CT) imaging is the 3D version of fluoroscopy. A CT scanner produces a virtual 3D anatomical image from a dense collection of digitized X-ray images that are acquired around a single rotation axis. For details about fundamentals of, and advances in CT, we refer the reader to Buzug [12].

Fluoroscopy and CT scanning have the drawback of using ionizing radiation. An alternative imaging modality is magnetic resonance imaging (MRI), in which images are acquired by detecting nuclear spin properties in the anatomy. Widely accepted as an imaging of anatomy and function, in the context of orthopedics and osteology
it is used for joint and spinal imaging. For principles and details of MRI, we refer the reader to Reiser et al. [70].

After the acquisition of a medical image, the anatomical region of interest is either manually or semi-automatically segmented. To the best of the author’s knowledge, fully automated segmentation of arbitrary anatomy is not in practice yet. One common manual segmentation method applies an intensity level of thresholding, followed by manual detailing for positive and negative outliers. Shapes are then constructed from the segmented image or volume. 2D shapes are often extracted from 2D images as iso-lines, or contours, using an algorithm such as marching squares [51]. Contours are generated as a set of points and a set of connecting edges. 3D shapes are extracted from 3D volumes as iso-surfaces using the seminal algorithm of marching cubes [49] or one of its variations [55]. Surfaces are generated as a set of points and a set of connecting triangles.

2.3 Shape Representation

Shape representation is a key part of the SSM construction framework and one of the critical design decisions that affects subsequent phases. A representation is a set of entities that identify a particular object in a coordinate frame. Three types are commonly used in medical applications: landmark representation, medial representation, and harmonic representation.
2.3.1 Landmark Representation

An object can be represented by a set of landmarks that have an established point correspondence over a population. Landmarks can be anatomical features identified by an expert, or might be geometrical features associated with a mathematical shape. In shape analysis, landmarks usually provide a relatively dense representation of shape. For example, a computer representation of a 2D contour is a set of connected points. Figure 2.1 shows a discretization of a head-neck cross-section of a human femur in a CT slice. In 3D, a surface might be extracted from a segmented CT volume as a triangulated mesh that consists of vertices and edges. Figure 2.2 shows a triangulated mesh of a proximal femur surface extracted from a segmented CT scan of a patient.

Figure 2.1: Femoral head-neck contour imposed on a CT slice

Points have been used to represent shapes in statistical analysis for over 30
years [19]. Some techniques have been reviewed by Kendall [43] and Bookstein [10]. Cootes et al. [16] introduced the term point distribution model (PDM), as an equivalent to landmarks, that has become the most widely used term. Many current shape models are based on PDM [38].

In this representation, training data have typically consisted of \( k \) objects. Each object was represented using \( m \) points \( p \in \mathbb{R}^n \), where \( n \) was either 2 or 3. Because each object was naturally in a \( \mathbb{R}^{n \times m} \) space, it was represented as a vector of high dimension. For the 3D case, an object \( S \) was commonly represented as

\[
S = (x_1 p, y_1 p, z_1 p, \ldots, x_m p, y_m p, z_m p)
\]

Distinct objects have distinct representation vectors, with natural correspondences between the points from which each representation vector is produced.
2.3.2 Medial Representation

A medial axis, or topological skeleton, is a shape representation that preserves geometrical and topological properties and enables reconstruction of the shape. As the name suggests, it is the axis that has equal distances to all boundaries of the shape. It was used in 1973 by Blum [9] to describe biological shapes. In mathematics, a medial axis has often been defined as a cut locus [45].

Pizer et al. [67] used medial axes, plus the radii of circles centered about each axis point, as vectors to reconstruct shape boundaries, or contours, in 2D. They extended the model to 3D and branded as a medial representation or m-rep [66]. In this framework, a 3D shape was represented by a connected set of points. Those points were the centers of the inscribed spheres of the shape and formed a continuous medial manifold $\mathcal{M}$. The points were modeled as regular lattice elements. Elements held position and radius information of the the inscribed spheres.

Each element was structured as a 4-tuple $m$ and named a medial atom. A medial atom has the structure

$$m = \{x, r, F, \theta\}$$

where $x \in \mathbb{R}^3$ was the center of the inscribed sphere, $r \in \mathbb{R}^+$ was the radius of the sphere, $F \in SO(3)$ was an orthonormal local frame, and $\theta \in [0, \pi)$ was the angulation relative to the frame $F$.

Joshi et al., from the same research group, used m-reps in a multi-scale framework for image segmentation [41]. The framework used Bayesian deformable templates to incorporate prior knowledge about shapes. The work of Pizer’s research group was amalgamated in software for medical image segmentation [65].
Yushkevich et al. introduced the continuous m-rep [90], which was an extension to the m-rep representation that modeled skeletons as continuous parametric manifolds. Terriberry and Gerig presented a similar model that accommodates for branch curves [81]. Branched curves enabled representation of complex medial axes with more than one medial sheet.

2.3.3 Harmonic Representation

Harmonic analysis, which is a 3D version of the Fourier transform, has been used to represent shapes. Staib et al. [79] used a Fourier decomposition to represent a signal as a weighted sum of orthonormal basis functions. Given a set of basis functions, each of which is $\phi_k(t)$, a 1D signal $f(t)$ can be represented as

$$f(t) = \inf \sum_{k=1}^{\infty} p_k \phi_k(t)$$

where the projections $p$ of $f(t)$, on the interval $(a,b)$, onto the basis functions are defined as

$$p_k = \int_a^b f(t) \phi_k(t) dt$$

Higher spatial variations are represented by higher order terms; thus, it is acceptable to truncate the sum for limited variation objects [79].

Matheny et al. [52] presented surface harmonics as an extension so as to model non-spherical topologies. Surface harmonics are intriguing for shape representation because they are orthogonal, complete, and ordered in spatial frequency. A 3D surface harmonic $V$ is represented in Cartesian coordinates as
\[
\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = 0
\]

Nikou et al. [57] described surfaces as amplitudes and vibration modes of a deformable model. Each point on a spherical surface mesh was modeled as a mass \( m \) that was connected to each neighbor point with a spring of stiffness \( k \). The governing equation was

\[
M\ddot{U} + C\dot{U} + KU = F
\]

where \( M \) was a mass matrix, \( K \) was a stiffness matrix, \( C \) was a damping matrix, and \( U \) was a nodal displacement from the original sphere mesh.

Davatzikos et al. [18] used wavelets to represent shape as separated features in multiple scales. They used the “mother” discrete wavelet

\[
\psi_{a,b} = \sqrt{2^{-a}}\psi(2^{-a}t - b)
\]

This wavelet used dilations by power of two and integer translations \( a, b \in \mathbb{Z} \).

Nain et al. [54] and Yu et al. [89] used spherical wavelets to represent shape. Spherical wavelets were scalar functions defined on a sphere [73]. Spherical wavelets combined both the power of multi-scale analysis and the power of spherical harmonics.

### 2.4 Shape Alignment

After representing the training data with a shape representation, filtering of Euclidean transformations is required. Procrustes analysis [31] is the celebrated method for aligning shapes to a unique coordinate frame by removing scaling, translation, and
rotation. This step requires establishing correspondence between landmarks. The rest of this section highlights common methods of establishing correspondence, followed by the Procrustes analysis.

### 2.4.1 Shape Correspondence

Establishing correspondences between samples is essential for a valid and reliable SSM construction. Manual landmarking is time consuming and prone to error even in 2D images; time and reproducibility concerns further increase in 3D. Registration, performed to automatically establish correspondence between images, has a huge body of literature. The reader can refer to Audette et al. [5], Brown et al. [11], Hajnaj et al. [35], Maintz et al. [50], van den Elsen et al. [85], and Zitova et al. [93] for general surveys in medical image registration. Heimann and Meinzer classified shape correspondence in the context of SSM construction into mesh-to-mesh, mesh-to-volume, volume-to-volume, parametrization-to-parametrization, and population-based registration [38].

A much-used method is the iterative closest point (ICP) of Besl and McKay [8], which can be adapted for mesh-to-mesh registrations. One of many alternatives to ICP is the *Softassign* Procrustes by Rangarajan et al. [69]. In general, a registration method is used to register given samples to a possibly arbitrarily selected reference shape. The generated SSM is, by construction, biased towards the reference shape. There have been many proposals to deal with this problem that were discussed by Heinmann [38]. Accurate registration is an active area of research that balances optimization similarity measures with transformation constraints, according to the application.
2.4.2 Procrustes Analysis

Procrustes analysis is a simple but efficient method for shape alignment. The method has typically been implemented as the application of four steps for each shape. First, the geometric centroid of the points is computed. This is the arithmetic mean of the coordinates of the points, i.e.,

$$\bar{S} = \frac{1}{m} \sum_{i=1}^{m} v_i$$

where $m$ is the number of points. Second, shapes are re-scaled to have equal size. This is done by a normalization process that finds a unit scale factor for the population, such as

$$S(v) = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (v_i - \bar{v})^2}$$

Third, shapes are translated by aligning centroids to the origin of the reference coordinate frame. Fourth, orientation is removed from all shapes. This can be done easily by extracting the orientation with an SVD algorithm.

This process was often applied iteratively to reach the optimal solution. The iterative process has been called generalized Procrustes analysis [33].

2.5 Shape Variability Analysis

The last phase of SSM construction is shape-variability analysis. Variability analysis has been used to reduce the dimensionality of a dataset while maintaining the variability of the data as much as possible [40].
This has been achieved by defining a new set of ordered uncorrelated variables, in which the first few variables captured most of the variation of the original dataset [40]. Those variables were called principal components. Principal component analysis (PCA) [61, 39] has long been the standard method for variability analysis, or dimensionality reduction, in SSM construction [38].

2.5.1 Principal Component Analysis

The PCA problem has usually been structured as a set of \( k \) shapes in a Euclidean vector space \( \mathbb{R}^n \) where \( n \) was either 2 or 3. Each shape consisted of \( m \) points with a defined point-to-point correspondence over shapes.

For PCA, shapes were often represented as landmark vectors \( \vec{v}_i \), each consisting of \( n \times m \) elements. The differences between shapes were typically modeled as point translations, with each vector element being a random variable with multiple observations \( k \). Here, the vector of random variables is written as \( \vec{v} \). If \( n \times m \) is very large, studying variances can be cumbersome.

A solution has been to find a set of uncorrelated linear functions of the form

\[
\alpha^j \vec{v} = \sum_{i=1}^{n \times m} \alpha_i^j v_i
\]

such that \( \alpha^j \vec{v} \) had a maximal variance, \( \alpha^j \) was a one-form, and \( \alpha_i^j \in \mathbb{R} \). The function \( \alpha_i^1 \vec{v} \) had the maximum variance and was named the first principal component. The goal of PCA was to find a number of components \( l << n \times m \) that accounted for most of the variations. In case \( l = 2 \), the data was approximately represented in 2D for easier visual analysis.

PCA was computed on mean-centered datasets. This was achieved by computing
the mean shape $\bar{\mu}$ as

$$\bar{\mu} = \frac{1}{k} \sum_{i=1}^{k} \vec{v}_i$$

where $k$ was number of shapes and $\vec{v}_i$ was a landmark shape instance.

The covariance matrix of the zero-mean data was computed as

$$C = \frac{1}{k - 1} \sum_{i=1}^{k} (\vec{v}_i - \bar{\mu})(\vec{v}_i - \bar{\mu})^T$$

Eigendecomposition was applied to find the eigenvalues and eigenvectors corresponding to the covariance matrix. The eigenvector corresponding to the largest eigenvalue was the first principal component and the eigenvector corresponding to the least eigenvalue was the last principal component [40]. Another alternative was to apply singular value decomposition (SVD) [20] to the shape landmark vectors directly; this method has been preferred because of its better numerical stability [38].

A shape could be reconstructed by the mean shape and the selected eigenvectors as

$$v_i' = \mu + \sum_{i=1}^{l} \beta_i \alpha_i$$

where $\beta_i \in \mathbb{R}$ was a weight of the mode variation.

A major problem with PCA or SVD, in this format, is that they assumed the data are adequately expressed as a Euclidean space; this was not necessarily the case with shape spaces [44]. Both being linear methods, PCA and SVD have tensor extensions [84]. Tensor, or multi-linear, methods used orthogonal transformations in the multidimensional space so there was an implicit assumption of inherent linearity of the data. Both PCA and SVD have had kernel extensions [1], which mapped the
nonlinear search space to a linear one. Defining the appropriate kernel remains an important open challenge for a novel dataset.

2.5.2 Principal Geodesic Analysis

The previous work most closely related to this thesis research was presented by Fletcher et al. as principal geodesic analysis (PGA), which was a generalization of PCA to manifolds. The method was initially applied to medial representations [24, 25, 26] and later to diffusion tensor imaging [22, 23]. Here, the method is detailed and then explained in terms of the applications.

The framework manipulated the basic statistical tools of mean, variance, principal components, and their projections. Those statistical concepts were extended from linear Euclidean space to non-linear manifolds. In other words, PGA described the general case of PCA.

Mean

The mean was the first statistical concept to be extended to manifolds. In Euclidean space, the mean point was defined as that which minimized the squared differences of all points in space, so

$$\mu = \arg \min_{x \in \mathbb{R}^n} \sum_{i=1}^{k} \|x - x_i\|^2$$

where $n$ was the Euclidean space dimension, $x_i$ was an element, and $k$ was the number of elements. This arithmetic difference does not necessarily apply to manifolds, for which there are two types of mean.

One way of computing a mean on a manifold has been to embed the manifold
into a Euclidean space and compute the Euclidean mean on the ambient space; the Euclidean mean was then projected to the manifold. The projection step is important because the Euclidean mean of the embedded points did not necessarily belong to the manifold. For example, if a sphere is embedded in $\mathbb{R}^3$, the mean of distinct points on the surface of the sphere falls inside the structure and not on the manifold. Such a mean has been termed the extrinsic mean on a manifold. Given an embedding $\Phi : \mathcal{M} \rightarrow \mathbb{R}^n$, the mean $\mu_\Phi$ has often been defined as

$$\mu_\Phi = \text{argmin}_{x \in \mathcal{M}} \sum_{i=1}^{k} \left\| \Phi(x) - \Phi(x_i) \right\|^2$$

The projection $\pi : \mathbb{R}^n \rightarrow \mathcal{M}$ mapped the mean back to the manifold and was defined as

$$\pi(x) = \text{argmin}_{y \in \mathcal{M}} \left\| \Phi(y) - x \right\|^2$$

This led to defining the extrinsic mean as

$$\mu_\Phi = \pi\left(\frac{1}{k} \sum_{i=1}^{k} \Phi(x_i)\right)$$

Another way to define a distance measure, or metric, on a manifold was to solve a minimization problem. The Riemannian metric $d(x, x_i)$ is a good example of an intrinsic mean of a manifold, as

$$\mu = \text{argmin}_{x \in \mathcal{M}} \sum_{i=1}^{k} d(x, x_i)^2$$
CHAPTER 2. REVIEW OF RELATED LITERATURE

Variance

According to Fréchet [28], the variance of a random variable in a metric space is given by

\[ \sigma^2 = \mathbb{E}[d(\mu, x)^2] \]

The variance \( \sigma^2 \), of a random variable \( x \) in a metric space, is usually defined as the expected squared distance from the mean \( \mu \). Fletcher et al. defined the variance of data samples \( x_i \), on a complete connected manifold \( M \), as

\[ \sigma^2 = \frac{1}{k} \sum_{i=1}^{k} d(\mu, x)^2 \]

where \( \mu \) was the intrinsic mean of the samples \( x_i \). Using the Riemannian metric, the variance was redefined as

\[ \sigma^2 = \frac{1}{k} \sum_{i=1}^{k} \| \log_{\mu}(x_i) \|^2 \]

The variance of data samples on a manifold \( M \) was then computed as the logarithmic mapping to the tangent space at the intrinsic mean of the samples.

Geodesic Submanifolds

A geodesic is the generalization of straight lines from a linear space to a manifold, defined as the locally shortest path between two points. The geodesics of a submanifold are not necessarily the geodesics of the mother manifold. For example, the geodesics of a sphere \( S^2 \) are great circles but the geodesics of the mother space \( \mathbb{R}^3 \) are straight lines.
An important concept for PGA is to define geodesic submanifolds. A submanifold \( \mathcal{P} \) was termed a geodesic submanifold \( p \in \mathcal{P} \) if all geodesics of \( \mathcal{P} \) that passed through \( p \) were also geodesics of the mother manifold \( \mathcal{M} \). Because the variance was defined with respect to the mean \( \mu \), it was necessary to define submanifolds with respect to the mean as well.

**Projection**

The projection of a point \( x \in \mathcal{M} \) on the geodesic submanifold \( \mathcal{P} \) has been defined as a point \( y \in \mathcal{P} \) for which the Riemannian distance \( d(x, y)^2 \) is minimal. This has commonly been formulated as the minimization

\[
\pi_{\mathcal{P}}(x) = \text{argmin}_{y \in \mathcal{P}} \{d(x, y)^2\}
\]

Existence and uniqueness of a solution is guaranteed on only a small neighborhood around the mean [26]. This can be approximated on the tangent space \( T_p \mathcal{P} \) by

\[
\log_p(\pi_{\mathcal{P}}(x)) \approx \sum_{i=1}^{k} \langle v_i \log_p(x) \rangle
\]

where \( v_1, \ldots, v_k \) were an orthonormal basis for \( T_p \mathcal{P} \).

**Algorithm**

The purpose of Fletcher’s work was to define principal geodesic submanifolds that covers most of the variance in a data set. The method assumed that the data were localized around the intrinsic mean \( \mu \) such that there existed a neighborhood \( \mathcal{U} \subset T_\mu \mathcal{M} \) for which the geodesic submanifolds \( \exp_\mu(\mathcal{U}) \) were well defined. In other words,
the projections of the data were localized around the identity of the tangent space $T_\mu \mathcal{M}$.

The principal geodesic submanifolds were defined as the projections of the orthonormal tangent vectors $v_1, \ldots, v_l \in T_\mu \mathcal{M}$ that spanned the tangent space. The first principal geodesic was defined as

$$v_1 = \arg\max_{\|v\|=1} \sum_{i=1}^{k} \| \log_\mu(\pi_H(x_i)) \|_2^2$$

where $H = \exp_\mu(\text{span}\{v\}) \cap U$. The rest of the principal geodesics were defined as

$$v_q = \arg\max_{\|v\|=1} \sum_{i=1}^{k} \| \log_\mu(\pi_H(x_i)) \|_2^2$$

where $H = \exp_\mu(\text{span}\{v_1, \ldots, v_{q-1}\}) \cap U$.

It can be seen that the PGA method is a straightforward application of PCA to the logarithmic mapping of the data to the tangent space, then projecting vector computations in the tangent space back to the manifold using the exponential mapping. PGA can be expressed concisely:

**Principal Geodesic Analysis (PGA)**

**Input:** $x_1, \ldots, x_k \in \mathcal{M}$

**Outputs:** principal directions $v_q \in T_\mu \mathcal{M}$

variances $\lambda_q \in \mathbb{R}$

$\mu = \text{intrinsic mean of } \{x_i\}$

$u_i = \log_\mu(x_i)$

$S = \frac{1}{k} \sum_{i=1}^{k} u_i u_i^T$

$\{v_q, \lambda_q\} = \text{eigenvectors and eigenvectors of } S$
CHAPTER 2. REVIEW OF RELATED LITERATURE

Application of PGA to Medial Representation

PGA was first applied to medial representations by Fletcher et al. [24, 25, 26]. In this section, medial representations are discussed as shape descriptors from which the manifold of those representations is deduced. Finally, the mapping between the manifold and its tangent space is described.

M-rep Manifold

An M-rep consists of 4 components. The position $x$ is a member of $\mathbb{R}^3$, which is a Lie group under vector addition. The radius $r$ is a member of $\mathbb{R}^+$, which is a Lie group under real multiplication. The frame $F$ is a member of the special orthogonal group $SO(3)$, which is a Lie group under matrix multiplication. The angle $\theta$ is a member of $SO(2)$, which is a Lie group under matrix multiplication as well. Because the product of Lie groups forms a Lie group, the medial group $M$ can be defined as

$$\mathbb{M} = \mathbb{R}^3 \times \mathbb{R}^2 \times SO(3) \times SO(2)$$

In other words, each medial atom is a point on the Lie group manifold.

Each M-rep consists of a lattice of $m \times k$ atoms forming the space $\mathcal{M}^{m \times n}$. Also by using product groups, M-reps form a Lie group where each M-rep is a point in the manifold of a Lie group.

Application of PGA to Diffusion Tensors

PGA was later applied to diffusion tensors by Fletcher et al. [22, 23]. In this section, diffusion tensors are defined as physical and mathematical structures from which the manifold of tensors is deduced.
**Diffusion Tensors**

A diffusion tensor is a rank-2 tensor that describe the forces that cause diffusion of water molecules. One way of describing a diffusion tensor is as a symmetric positive-definite matrix $D$ that has the form

$$
D = \begin{bmatrix}
    d_{xx} & d_{xy} & d_{xz} \\
    d_{xy} & d_{yy} & d_{yz} \\
    d_{xz} & d_{yz} & d_{zz}
\end{bmatrix}
$$

By construction, $D$ is symmetric so $D = D^T$. A constraint is that $D$ must be positive-definite that $x^T D x > 0$ where $x \neq 0 \in \mathbb{R}^n$. Such a matrix can be used to describe shape and orientation of an ellipse, and also to compute eigenvectors and eigenvalues.

**Diffusion-Tensor Manifold**

The general linear group $\text{GL}(n, \mathbb{R})$, the group of $n \times n$ real matrices, is a Lie group closed under matrix multiplication. The subgroup $\text{GL}^+(n, \mathbb{R})$ is a set of $n \times n$ real matrices with positive determinants, which is a Lie group by definition. Applying polar decomposition [76] to any of such matrix $M \in \text{GL}^+(n, \mathbb{R})$ produces two matrices, a symmetric positive definite matrix $Q \in \mathbb{P}(n)$ and a rotation matrix $R \in \text{SO}(n)$. Because $\text{SO}(n)$ is also a Lie group, the quotient space $\mathbb{P}(n)$ is also a Lie group. $\mathbb{P}(n)$ is defined mathematically as

$$
\mathbb{P}(n) \cong \text{GL}^+(n, \mathbb{R})/\text{SO}(n)
$$

The set of diffusion-tensor matrices is thus a Lie group, to which PGA can be applied.
2.5.3 Exact PGA

PGA, as described above, was an approximation of the data samples on the tangent space that assumed localization of the data around an intrinsic mean. Sommer et al. [77] introduced exact-PGA that performed intrinsic computations on the manifold instead of using the simplifying mapping to the tangent space. The work concluded that PGA performed well on some data but did not provide acceptable approximations for some data sets. It was also noticed that the exact-PGA computation was very expensive, compared to the cost of PGA that used the tangent-vector space.

It is important to note that applying PGA directly to landmarks will reduce to PCA, because the differences are pure translations. PGA was successful for m-reps because of the rotational components and for diffusion-tensor images because of the positive-definite constraints.

2.5.4 Other Manifold Statistics Techniques

Pennec and Arsigny et al. developed tools and metrics for applying statistics on Riemannian manifolds. Those tools extended measurements on real variables or vector spaces to Riemannian manifolds.

Basic Tools

Pennec et al. [62, 64, 63] used a random primitive as an equivalent to a random variable on the manifold. Based on this idea, the main statistical tools were defined.

For the random primitive $x$, the variance was defined as

$$\sigma_x^2 = E \text{ dist}(y, x)^2$$
where \( y \) was a fixed primitive. On a Riemannian manifold, the variance was the expected Riemannian distance between the random primitive and a fixed primitive. This fixed primitive was usually the mean.

The set of mean primitives that minimized the distance was defined as

\[
E[x] = \arg\min_{y \in \mathcal{M}} E\text{\,dist}(y, x)^2
\]

Because the mean was defined via a minimization process, uniqueness was not guaranteed. The mean was considered as a local minimum in a neighborhood rather than being a unique global mean.

For the mean \( \bar{x} \) of a random primitive \( x \), the covariance was defined as

\[
\Sigma_{xx} = \text{cov}_{\bar{x}}(x) = E[\bar{x} \bar{x}^T]
\]

### 2.5.5 Log-Euclidean Metric

Arsigny et al. [3, 4] provided Log-Euclidean metrics as tools to perform fast computations of smooth manifold spaces using the exponential mapping to vector spaces. This was based on finding a one-to-one mapping between symmetric matrices and vector spaces. This enabled the use of vector addition and scalar multiplication to symmetric matrices using exponential mapping. The logarithmic multiplication \( \odot \) was defined as

\[
S_1 \odot S_2 \equiv (\log(S_1) + \log(S_2))
\]

The logarithmic scalar multiplication \( \otimes \) was defined as
\[ \lambda \odot S \equiv \exp(\lambda \log(S)) = S^{\lambda} \]

The norm on symmetric matrices was defined as

\[ d(S_1, S_2) = \| \log(S_1) - \log(S_2) \| \]

The mean was defined as

\[ \mu = \exp(\sum_{i=1}^{N} p_i \log(S_i)) \]

They later applied the Log-Euclidean framework to performing computations on diffusion tensor imaging [4]. The method used differential geometry to perform simple computations on the Euclidean tangent space of the DTI tensor space. Later, they enhanced the method and developed a fast polyaffine transform algorithm for efficient computation [2]. Commowick et al. [14, 15] combined both the polyaffine and Log-Euclidean frameworks as a new registration framework. The method was applied to bone and brain-structure registration.

2.5.6 A Critique of Closely Related Work

PGA is a generalization of PCA from vector spaces to manifolds. The work of Fletcher et al. provided a general framework for using exponential mapping to linearize manifolds. Although the PGA framework has been around for ten years, the paucity of literature suggests that it has not been much used for statistical shape models. This is attributable to two main reasons. First, the method was introduced for m-reps,
whereas most researchers use point distributions because the latter are easier to construct. Second, the implementation was presented as an optimization problem; this entails an expensive computation that does not guarantee an optimal solution. A further complication was that Fletcher’s solution required a localization of data that was not easy to verify empirically. Although Fletcher et al. derived m-reps as Lie groups, they dealt with the implementation as a general computation of a manifold geodesic.

Sommer et al. presented exact PGA as a more accurate solution than PGA. Exact PGA computed the exact intrinsic distance in the manifold without the linearization on the tangent space. As they indicated in their paper, their solution was much more expensive to compute than was PGA; in one example, it took PGA 5 seconds to compute whereas it took exact-PGA 10 minutes to compute. Their results showed that the accuracy difference was not always significant, with a negligible difference in accuracy if the data were well localized.

PGA was, overall, more efficient than the Exact-PGA but less accurate when applied to some datasets. The method was also applied to diffusion tensors. This latter application required manipulating the data to fit the criteria of symmetric positive-definite matrices, an alternative that is not always possible for other in empirical data.

Arsigny et al. introduced the log-Euclidean as a deviation from the trend of applying Riemannian geometry to process data in non-linear spaces. The Log-Euclidean framework was an easier, and less expensive, alternative to PGA. It used matrices as Lie groups and the matrix exponential as a linearization tool. The Log-Euclidean framework was first developed for symmetric matrices and applied to diffusion-tensor
images, because those matrices are symmetric and positive-definite.

The Log-Euclidean method was later extended to rigid and affine transformation matrices that modeled local deformations. The ensemble of transformations, named poly-rigid and poly-affine, was used to model global non-rigid deformations. A poly-affine model was used for registration of MRI images, with a linearization that relied on a pure matrix-logarithm implementation. This implementation had limitations of accuracy and required, like PGA, that the data be localized; they required that the data have very small variations from the mean, which is not always guaranteed in empirical datasets. Use of their method has not been reported by other researchers.

2.6 Summary

Statistical shape model construction is a three-step framework consisting of shape representation, shape alignment, and shape variability analysis. In orthopedics and osteology, CT imaging has been a major source of medical images. Landmark representation has been the dominant shape representation in medical imaging, in part because of the methods used to extract boundaries and surfaces form planar and volumetric data. Medial representation, harmonics, and tensor representations have been used – but not in major ways – for representation.

Shape correspondence has often been approximated using ICP or Softassign for mesh-to-mesh registration. Shape alignment was usually done by Procrustes analysis. Dimensionality reduction, which is a main topic of this thesis work, was usually done using PCA that reduced the dataset into uncorrelated orthogonal principal components. Use of a few principal components covered most of the data variation and allowed easier visual analysis of the data. A difficulty with PCA has been the
assumption that the data are best described as being in a Euclidean space, which has not been the case in most shapes derived from medical data.

PGA was used to accommodate for rotational and scaling components but it did not apply to landmark representations. As the review of recent literature has suggested, PCA is still the dominant dimensionality reduction technique for SSM construction in general, and for SSM of bony anatomy in particular. The purpose of this work is to address the lack of simple and accurate implementations of methods analyzing shapes in their natural spaces.
Chapter 3

Theoretical Methods

This chapter describes the mathematical methods developed in this thesis. It relies heavily on the background, introduced in the appendix, to develop an understanding and derive computations. An abstract view of the method is introduced in Section 3.1. A derivation of the method is discussed, using the 2D contours, in Section 3.2. After that, the derivation is extended to 3D surfaces in Section 3.3. The chapter concludes with a summary in Section 3.4.

3.1 Overview

The general linear group $\mathbb{GL}(n, \mathbb{R})$, or just $\mathbb{GL}(n)$, is the group of real-valued non-singular $n \times n$ matrices. It is a Lie group that it is also a smooth manifold $\mathcal{M}$ of $n^2$. A particularly useful subgroup, here denoted as $\mathbb{H}(n) \subset \mathbb{GL}(n)$, can be defined by imposing constraints on the mother group.

Homogeneous coordinates are a convenient notation for representing affine transformations of vectors. A homogeneous vector with $n$ entries, conventionally, has a
final entry of either 0 or 1 (to respectively transform an entity vectorially or point-wise). To ensure that a homogeneous vector transforms correctly, the matrix operator must be constrained so that

$$M_{(n,i)} = \begin{cases} 
0 & \text{if } i < n \\
1 & \text{if } i = n 
\end{cases}$$

This defines the set of homogeneous matrices for which the last row has 0 values except for the final entry, which has the value of 1. This is easily verified to be a subgroup: the identity matrix is an element of $\mathbb{H}(n)$; the product of two elements is in $\mathbb{H}(n)$; and the inverse both exists and is in the subgroup.

The submanifold $\mathcal{H}$ associated with this Lie group has the dimension of $n^2 - n$ because the last row of the matrix is constant. The submanifold inherits properties from the mother manifold, chiefly that it is smooth.

Let $A, B \in \mathbb{H}(n)$ be two homogeneous matrices. Because the matrices are in a Lie group, they are in the manifold $\mathcal{H}$ as points. The matrix $B$ is dragged on the manifold to the matrix $A$ by group multiplication by some matrix $C$, so

$$A = CB$$

The matrix $C$, a transformation from the matrix $B$ to $A$, is also an element of the Lie group $\mathbb{H}(n)$ and can be found by trivial matrix algebra as

$$C = AB^{-1}$$

A critical matter for this thesis is: a homogeneous matrix is both a transformation and a point in a manifold. These two properties arise because $\mathbb{H}(n)$ is a Lie group: $C$ is a transformation of $B$ when taken together with the group operation of matrix multiplication; it is also a point in a smooth manifold $\mathcal{H}$. Computer scientists are
typically familiar with a matrix as an operator but may be less familiar with it as a point in a smooth Hausdorff topological space. However, because homogeneous matrices are a smooth manifold, the rich geometry that includes curves and tangents can be applied to them. The relationship between points and tangents is the exponential map.

Because a homogeneous matrix is invertible, by virtue of being in the Lie group $\text{GL}(n)$, it has a unique matrix logarithm that will be shown below to be easily computed. The matrix logarithm allows us to readily transfer between the manifold of the Lie group $\mathbb{H}(n)$ and its tangent bundle. For any matrix $C \in \mathbb{H}(n)$, there is a unique matrix $K$ such that

$$\log(C) = K$$
$$\exp(K) = e^K = C$$

(3.1)

Mathematically, the relationships between the matrices $C$ and $K$ of Equation 3.1 are usually expressed as the matrix version of the familiar Taylor series expansions of the logarithmic and exponential functions. The matrix exponential of $K$ is

$$\exp(K) = e^K = \sum_{i=0}^{\infty} \frac{K^i}{i!}$$
$$= I + K + \frac{K^2}{2!} + \frac{K^3}{3!} + \cdots$$

(3.2)

Computationally, far better methods than a direct use of the Taylor series have long been known [53].

The matrix $K$ can be thought of as a vector in the tangent space of the identity $I$, projecting to the manifold point $C$ via the exponential map. It is easy to see that the non-constant entries of the matrix $K$ can be “unravelled” into a vector $\vec{k}$, of dimension $n^2 - n$, by a convention such as column-major ordering. Although $C$
in non-singular, the matrix $K$ will be singular because its final row will have all zero entries.

Suppose next that there is a set of homogeneous matrices $\{A_i\}$ and a single matrix $B$ that is of interest. Together, these generate a new set of matrices $\{C_i\}$ as

$$C_i = A_iB^{-1}$$

Computing the logarithm of each element of the set $\{C_i\}$ generates a set of vectors $\{K_i\}$. Those vectors are members of a vector space, namely the tangent space at the identity element $T_I\mathcal{H}$. Linear statistics, such as PCA, can be applied to such vectors because they are in a single consistent space.

A key concept in this thesis is that certain discrete shapes can be understood as sets of discrete points in a manifold. To better understand this concept, it will be developed first for simple 2D contours and then for simple 3D surfaces.

### 3.2 Contours in Two Dimensions

This section derives Lie group analysis of discrete shapes using simple 2D shapes. Here, a 2D shape instance is a non-overlapping and non-intersecting contour. Colloquially it is a discretization of a closed non-intersecting curve, i.e., the boundary of a simple polygon.

A planar affine transformation maps a column vector $\vec{p}$ to a vector $\vec{p}'$ as

$$\vec{p}' = M\vec{p} + \vec{t}$$

where $M$ is a [non-singular] $2 \times 2$ transformation matrix and $\vec{t}$ is a 2D translation.
vector.

Suppose that \( m \) points are represented, in some coordinate frame, as vectors \( \vec{p}_j \).
A shape matrix \( P_{2 \times m} \) can be constructed from the points by having the columns of \( P \) be the vector representations of the points. For example, to apply the transformation to 3 points of a shape using an integer \( j \) to index into the shape, the matrix \( jP \) could be constructed as

\[
jP = \begin{bmatrix}
  x_{j-1}p & x_{j+1}p & \bar{x}p \\
  y_{j-1}p & y_{j+1}p & \bar{y}p \\
  1 & 1 & 1
\end{bmatrix}
\]

where the ordering of the points is not important but must be consistent.

To apply a global affine transformation to the shape \( jP \), the vector \( \vec{p} \) in Equation 3.3 can be replaced with the matrix \( jP \) from Equation 3.4. A notationally convenient alternative is to represent each 2D vector in its homogeneous form, so that a single homogeneous matrix can be applied to the ensemble of points. A naive construction would be to represent each triple of 3 points, each column being the homogeneous coordinates of the corresponding point, such a matrix would have the construction

\[
jP = \begin{bmatrix}
  x_{j-1}p & x_{j+1}p & \bar{x}p \\
  y_{j-1}p & y_{j+1}p & \bar{y}p \\
  1 & 1 & 1
\end{bmatrix}
\]

3.2.1 A 2D Triple of Points as a Homogeneous Matrix

By inspection, a matrix such as the one in Equation 3.5 is “nearly” a homogeneous matrix; if the first two entries of its last row were zero, then it would have the correct construction. One way that this can be accomplished is to replace the first column of
the matrix in Equation 3.5 with the difference between the first and third columns, and likewise with the second column. This would produce a new matrix \( \hat{P} \) that is constructed as

\[
\hat{P} = \begin{bmatrix}
(x_j - 1 - x_j p & (x_j p - x_j) & x_j p \\
(y_j - 1 - y_j p & (y_j p - y_j) & y_j p \\
0 & 0 & 1
\end{bmatrix}
\] (3.6)

For the matrix \( \hat{P} \) of Equation 3.6 to be in the Lie group \( \mathbb{H}(3) \) it must be non-singular.

A matrix that is structured like \( \hat{P} \) can be block-partitioned, and factored, as

\[
\hat{P} = \begin{bmatrix}
[M] & \vec{t} \\
[0 0] & 1
\end{bmatrix} = \begin{bmatrix}
[I] & \vec{t} \\
[0 0] & 1
\end{bmatrix} \begin{bmatrix}
[M] & \vec{0} \\
[0 0] & 1
\end{bmatrix} = \hat{T} \hat{M}
\] (3.7)

Taking the determinant by expanding along the last row, such a matrix \( \hat{P} \) is non-singular if and only if the sub-matrix \( M \) is non-singular, i.e., iff \( \det(M) \neq 0 \). Geometrically, this is guaranteed if and only if the points used to construct the homogeneous matrix are non-collinear. There is an elegant duality between the geometry of a triple of planar points and the homogeneous matrix used to represent the triple.

A discrete 2D contour that is represented as an ordered sequence of triples of points can thus be represented as an ordered sequence of homogeneous matrices. Each such matrix is in the Lie group \( \mathbb{H}(3) \) so it lies in a smooth manifold \( \mathcal{H} \). Such a simple 2D shape is a set of points that can be manipulated using tools from differential geometry.

### 3.2.2 2D Homogeneous Matrices Using Exponentials

For this subsection, assume that a matrix that is constructed as in Equation 3.7 is non-singular. With this constraint, the sub-matrix \( M \) is a member of the general
linear Lie group of real $2 \times 2$ matrices, that is, $M \in \mathbb{GL}(2)$. $M$ can be factored in many ways, one of which is the QL decomposition (which is a trivial variant of the better known QR decomposition that always avoids shearing the first row of its operand). This would express the matrix as $M = RL$, where $R$ is orthogonal and $L$ is lower triangular; explicitly,

$$M = \begin{bmatrix} \cos(\eta) & -\sin(\eta) \\ \sin(\eta) & \cos(\eta) \end{bmatrix} \begin{bmatrix} l_{(1,1)} & 0 \\ l_{(2,1)} & l_{(2,2)} \end{bmatrix}$$

$R$ is an orthogonal matrix so it is a rotation matrix. The matrix $L$ can be further factored into the product of a diagonal matrix and a strictly lower matrix, so it can be written as

$$L = EH = \begin{bmatrix} l_{(1,1)} & 0 \\ l_{(2,1)} & l_{(2,2)} \end{bmatrix} = \begin{bmatrix} l_{(1,1)} & 0 \\ 0 & l_{(2,2)} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ h & 1 \end{bmatrix}$$

Thus, the matrix $M$ can be re-written as $M = REH$. The affine matrix \( \hat{P} \), written in homogeneous coordinates, is therefore factored as

$$\hat{P} = \hat{T} \hat{R} \hat{E} \hat{H}$$

(3.8)

The exponential of the matrix $\hat{P}$, using the matrix $\hat{K}$, is

$$\hat{P} = e^{[\hat{K}]}$$

Taking care to preserve the order of the factors of Equation 3.8, because they do not commute, expresses the matrix $\hat{K}$ as the sum

$$\hat{K} = \hat{W} + \hat{S} + \hat{\Lambda} + \hat{\Gamma}$$

(3.9)
Using the Lie product, and simplifying, yields
\[ e^{\hat{K}} = e^{(\hat{W} + \hat{S} + \hat{\Lambda} + \hat{\Gamma})} = \lim_{m \to \infty} \left( e^{[\hat{W}/m]} e^{[\hat{S}/m]} e^{[\hat{\Lambda}/m]} e^{[\hat{\Gamma}/m]} \right)^m = e^{[\hat{W}]} e^{[\hat{S}]} e^{[\hat{\Lambda}]} e^{[\hat{\Gamma}]} \]

It is easy to verify the constructions of the matrices in Equation 3.9 by taking their exponentials using Equation 3.2. Beginning with the translation matrix, its logarithm of \( \hat{W} \) is constructed as
\[
\hat{W} = \begin{bmatrix} 0 & 0 & w_1 \\ 0 & 0 & w_2 \\ 0 & 0 & 0 \end{bmatrix}
\]
Because the matrix \( \hat{W} \) is nilpotent of degree 2, its exponential series is truncated after the second term so
\[
\hat{T} = e^{[\hat{W}]} = I + \hat{W} = \begin{bmatrix} 1 & 0 & w_1 \\ 0 & 1 & w_2 \\ 0 & 0 & 1 \end{bmatrix}
\]
The logarithm of a rotation matrix, from Cayley’s formula, is a skew-symmetric matrix. For planar transformations, the logarithm and its unit-determinant correlate are
\[
\hat{S} = \begin{bmatrix} 0 & -\eta & 0 \\ \eta & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \hat{Z} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \Rightarrow S = \eta \hat{Z}
Because of the powers of this $\hat{Z}$ matrix conveniently reduce, the Taylor expansion of the exponential matrix $\hat{S}$ is

$$\hat{R} = e[\hat{S}] = e^\eta[\hat{Z}] = e^\eta \hat{Z} - \eta^2 \frac{\hat{Z}^2}{2!} - \eta^3 \frac{\hat{Z}^3}{3!} + \ldots$$

$$= \cos(\eta)I + \sin(\eta)\hat{Z}$$

$$= \begin{bmatrix}
\cos(\eta) & -\sin(\eta) & 0 \\
\sin(\eta) & \cos(\eta) & 0 \\
0 & 0 & 1 
\end{bmatrix}$$

The logarithm of the scaling matrix is the diagonal matrix

$$\hat{\Lambda} = \begin{bmatrix}
\varepsilon_1 & 0 & 0 \\
0 & \varepsilon_2 & 0 \\
0 & 0 & 0 
\end{bmatrix}$$

Because the exponential of a diagonal matrix is the matrix of the exponentials of the diagonal entries, the scaling matrix is

$$\hat{E} = e[\hat{\Lambda}] = \begin{bmatrix}
e^{\varepsilon_1} & 0 & 0 \\
0 & e^{\varepsilon_2} & 0 \\
0 & 0 & e^0 
\end{bmatrix} = \begin{bmatrix}
e^{\varepsilon_1} & 0 & 0 \\
0 & e^{\varepsilon_2} & 0 \\
0 & 0 & 1 
\end{bmatrix}$$

The logarithm of the shear matrix is

$$\hat{\Gamma} = \begin{bmatrix}
0 & 0 & 0 \\
h & 0 & 0 \\
0 & 0 & 0 
\end{bmatrix}$$
Because the matrix $\Gamma$ is nilpotent of degree 2, its exponential is simply

$$e^{[\hat{\Gamma}]} = I + \hat{\Gamma} = \begin{bmatrix} 1 & 0 & 0 \\ h & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Using sum of these logarithm matrices, the matrix $\hat{K}$ has the construction

$$\hat{K} = \begin{bmatrix} \varepsilon_x & -\eta & t_x \\ \eta + h & \varepsilon_y & t_y \\ 0 & 0 & 0 \end{bmatrix} \tag{3.11}$$

where $\eta$ is a rotation angle, the two $\varepsilon$ are scaling factors, $h$ is a shear factor, and $t$ is the translation.

The Lie algebra of the Lie group $\mathbb{H}(3)$ is the product space of its factors. The algebra of planar rotation is $\mathbb{S}^1$; the algebra of independent scale factors that can take on any real values is $\mathbb{R}^2$; the algebra of a single shear factor that can take on any real value is $\mathbb{R}^1$; and the algebra of independent translation factors that can take on any real values is $\mathbb{R}^2$. The Lie algebra of 2D homogeneous transformations is therefore

$$\mathfrak{h}(3) = \mathbb{S}^1 \times \mathbb{R}^5$$

which has a dimension of 6. The tangent vector of the matrix $\hat{M}$ is a vector $\vec{k}$; here, it will be written as

$$\vec{k} = [\eta, \varepsilon_x, \varepsilon_y, h, t_x, t_y]^T \tag{3.12}$$

There are two important consequences of the derivations in this section. The first, already stated, is that a non-collinear triple of planar points can be represented as a non-singular homogeneous matrix. The second is that any such matrix can be
linearized using matrix logarithms, so linear statistics (such as the SVD and certain types of PCA) can be applied to the tangent space of homogeneous matrices. In mathematical terms, this section has shown how to linearize the Lie group $\mathbb{H}(3)$.

### 3.2.3 A Numerical Implementation of 2D Contours

Contours can be discretized as ordered sets of planar vertices, written as $(x_P, y_P) \in \mathbb{R}^2$. Each set has the same integer cardinality $m$; planar implementation requires that $m$ be divisible by 3, so there need to be $m = 3k$ vertices in each contour. One contour is selected as the base contour $B$ and the others as a set of altered contours $A_1, A_2, ..., A_n$, with $n \in \mathbb{Z}^+$. A traversal of the base contour $B$ is “unraveled” into a vector $\vec{b}$ and each altered contour $A_i$ is similarly unraveled into a vector $\vec{a_i}$. Vertex correspondence, between the base shape and the altered shapes, is implied by the order of the points and the traversal direction.

A 2D contour is segmented into a set of $k$ distinct sets of three consecutive points, henceforth simply called “triples”. Each triple is parametrized as a square homogeneous matrix of the construction as in Equation 3.6. The derivation is detailed in Subsection 3.2.1.

For each corresponding pair, the homogeneous affine transformation matrix $j\hat{C}_i$, from the base shape to the altered shape is computed by multiplying the altered matrix by the inverse of the base matrix, so that

$$j\hat{C}_i = j\hat{A}_i(j\hat{B})^{-1}$$

Each matrix $j\hat{C}_i$ was decomposed into its four basic transformation matrices: translation $j\hat{T}_i$, rotation $j\hat{R}_i$, shear $j\hat{S}_i$, and scaling $j\hat{H}_i$, as
\[ j\hat{C}_i = j\hat{T}_{ij} \hat{R}_{ij} \hat{S}_{ij} \hat{H}_i \]

Factoring each translation matrix is straightforward because the translation matrix \( \hat{T}_{ij} \) is extracted by stripping off the last column, having terms \((t_x, t_y)\). The factorization is

\[ j\hat{C}_i = j\hat{T}_{ij} \hat{M}_i \]

with a general \(2 \times 2\) matrix \( j\hat{M}_i \) embedded in the homogeneous matrix \( j\hat{M}_i \). The next part of the decomposition extracted a rotation matrix \( j\hat{R}_i \) which is required to be orthogonal and positive-definite. Each such matrix \(M\) is decomposed using a trivial variation of the QR decomposition; it was such that

\[ M = RL \]

where \( R \) is orthogonal and \( L \) is lower triangular. The QR decomposition algorithm is known to be efficient, stable, and produce unique factors. One simple implementation is as the well known Gram-Schmidt algorithm:

**Gram-Schmidt algorithm**

**input:** \( M \)

**output:** \( R, L \)

\[
\begin{align*}
R &= [0]; \\
L &= [0]; \\
\text{for index} &= 1 \text{ to (number of columns)} \\
T &= M(:,\text{index}); \\
L(:,\text{index}) &= R^*T;
\end{align*}
\]
\[ T = T - R \cdot L(:,\text{index}); \]
\[ L(\text{index, index}) = \text{norm}(T); \]
\[ R(:,\text{index}) = T / L(\text{index, index}); \]
end for
end

The orthogonal part of the decomposition \( R \) is a rotation matrix because its determinant is +1. The lower-triangular factor \( L \) is further decomposed into a shear factor \( H \) having at most one non-zero off-diagonal element, and a scale factor \( S \) that is a diagonal matrix.

The logarithm of each transformation matrix \( \hat{C}_i \) is computed by computing the logarithm of the parts. The sum of the logarithms of the parts produced a matrix \( \hat{K}_i \), each of which is constructed as

\[
\hat{K} = \begin{bmatrix}
l_x & -\eta & t_x \\
\eta + h & t_y & t_y \\
0 & 0 & 0
\end{bmatrix}
\]

where \( \eta \) is the rotation angle, the two \( l \) values are scale components, \( h \) is a shear component, and \( t \) is a translation. For each such matrix \( \hat{K}_i \), there is an associated vector \( \vec{k}_i \) constructed as

\[
\vec{k} = [\eta, l_x, l_y, h, t_x, t_y]^T.
\]

All such vectors \( \vec{k}_i \) for a given index \( j \) are assembled into a column, and all such columns are assembled in a matrix \( D \). The matrix is then decomposed using the singular value decomposition \( SVD \) as

\[
D = U \Sigma V^T
\]
3.3 Surfaces in Three Dimensions

This section extends the Lie group analysis from 2D contours to 3D surfaces. Here, a 3D shape instance is a non-overlapping and non-intersecting surface. Colloquially it is a discretization of a genus-0 surface.

A spatial affine transformation maps a column vector $\vec{p}$ to a vector $\vec{p}'$ as in Equation 3.3. In spatial coordinates, $M$ is a [non-singular] $3 \times 3$ transformation matrix and $\vec{t}$ is a 3D translation vector.

Suppose that $m$ points are represented, in some coordinate frame, as vectors $\vec{p}_j$. A shape matrix $P_{3 \times m}$ can be constructed from the points by having the columns of $P$ be the vector representations of the points. For example, to apply the transformation to 4 points of a shape using an integer $j$ to index into the shape, the matrix $jP$ could be constructed as

$$
\begin{bmatrix}
\tilde{x}_{j+1}P & \tilde{x}_{j+2}P & \tilde{x}_{j+3}P & \tilde{x}_jP \\
\tilde{y}_{j+1}P & \tilde{y}_{j+2}P & \tilde{y}_{j+3}P & \tilde{y}_jP \\
\tilde{z}_{j+1}P & \tilde{z}_{j+2}P & \tilde{z}_{j+3}P & \tilde{z}_jP
\end{bmatrix}
$$

(3.13)

where the ordering of the points is not important but must be consistent.

To apply a global affine transformation to the shape $jP$, the vector $\vec{p}$ in Equation 3.3 can be replaced with the matrix $jP$ from Equation 3.13. A notationally convenient alternative is to represent each 3D vector in its homogeneous form, so that a single homogeneous matrix can be applied to the ensemble of points. A naive construction would be to represent each quadrilateral of 4 points, each column being the homogeneous coordinates of the corresponding point, such a matrix would have
the construction

\[
\begin{bmatrix}
  x_{j+1}^p & x_{j+2}^p & x_{j+3}^p & x_j^p \\
  y_{j+1}^p & y_{j+2}^p & y_{j+3}^p & y_j^p \\
  z_{j+1}^p & z_{j+2}^p & z_{j+3}^p & z_j^p \\
  1 & 1 & 1 & 1
\end{bmatrix}
\]  

(3.14)

3.3.1 A 3D Quadruple of Points as a Homogeneous Matrix

By inspection, a matrix such as the one in Equation 3.14 is “nearly” a homogeneous matrix; if the first three entries of its last row were zero, then it would have the correct construction. One way that this can be accomplished is to replace the first column of the matrix in Equation 3.14 with the difference between the first and fourth columns, and likewise with the second and third columns. This would produce a new matrix \( \hat{P} \) that is constructed as

\[
\hat{P} = \begin{bmatrix}
  (x_{j+1}^p - x_j^p) & (x_{j+2}^p - x_j^p) & (x_{j+3}^p - x_j^p) & x_j^p \\
  (y_{j+1}^p - y_j^p) & (y_{j+2}^p - y_j^p) & (y_{j+3}^p - y_j^p) & y_j^p \\
  (z_{j+1}^p - z_j^p) & (z_{j+2}^p - z_j^p) & (z_{j+3}^p - z_j^p) & z_j^p \\
  0 & 0 & 0 & 1
\end{bmatrix}
\]  

(3.15)

For the matrix \( \hat{P} \) of Equation 3.15 to be in the Lie group \( \mathbb{H}(4) \) it must be non-singular. A matrix that is structured like \( \hat{P} \) can be block-partitioned, and factored, as

\[
\hat{P} = \begin{bmatrix}
  [M] & \tilde{t} \\
  [0 0 0] & 1
\end{bmatrix} = \begin{bmatrix}
  [I] & \tilde{t} \\
  [0 0 0] & 1
\end{bmatrix} \begin{bmatrix}
  [M] & \tilde{0} \\
  [0 0 0] & 1
\end{bmatrix} = \hat{T}\hat{M}
\]  

(3.16)

Taking the determinant by expanding along the last row, such a matrix \( \hat{P} \) is non-singular if and only if the sub-matrix \( M \) is non-singular, i.e., if \( \det(M) \neq 0 \). Geometrically, this is guaranteed if and only if the points used to construct the homogeneous
matrix are non-coplanar. There is an elegant duality between the geometry of a quadrilateral of spatial points and the homogeneous matrix used to represent the quadruple.

A discrete $3D$ surface that is represented as an ordered sequence of quadruples of points can thus be represented as an ordered sequence of homogeneous matrices. Each such matrix is in the Lie group $\mathbb{H}(4)$ so it lies on a smooth manifold $\mathcal{H}$. Such a simple $3D$ shape is a set of points that can be manipulated using tools from differential geometry.

### 3.3.2 3D Homogeneous Matrices Using Exponentials

For this subsection, assume that a matrix that is constructed as in Equation 3.16 is non-singular. With this constraint, the sub-matrix $M$ is a member of the general linear Lie group of real $3 \times 3$ matrices, that is, $M \in \mathbb{GL}(3)$. $M$ can be factored in many ways, one of which is the QL decomposition (which is a trivial variant of the better known QR decomposition but has physical meaning). This would express the matrix as $M = RL$, where $R$ is orthogonal and $L$ is lower triangular.

$R$ is an orthogonal matrix, either a rotation matrix or a reflected rotation matrix. The matrix $L$ can be further factored into the product of a diagonal matrix and a strictly lower matrix, so it can be written as

$$ L = EH = \begin{bmatrix}
  l_{1,1} & 0 & 0 \\
  l_{2,1} & l_{2,2} & 0 \\
  l_{3,1} & l_{3,2} & l_{3,3}
\end{bmatrix} = \begin{bmatrix}
  l_{1,1} & 0 & 0 \\
  0 & l_{2,2} & 0 \\
  0 & 0 & l_{3,3}
\end{bmatrix} \begin{bmatrix}
  1 & 0 & 0 \\
  h_{xy} & 1 & 0 \\
  h_{xz} & h_{yz} & 0
\end{bmatrix} $$

So, the matrix $M$ can be re-written as $M = REH$. The affine matrix $\hat{P}$, written in homogeneous coordinates, is therefore factored as
\[
\hat{P} = \hat{T}\hat{R}\hat{E}\hat{H} \tag{3.17}
\]

The exponential of the matrix $\hat{P}$ matrix, using the matrix $\hat{K}$, is
\[
\hat{P} = e^{[\hat{K}]} \tag{3.8}
\]

Taking care to preserve the order of the factors of Equation 3.8, because they do not commute, expresses the matrix $\hat{K}$ as the sum
\[
\hat{K} = \hat{W} + \hat{S} + \hat{\Lambda} + \hat{\Gamma} \tag{3.18}
\]

Using the Lie product, and simplifying, yields
\[
e^{\hat{K}} = e^{([\hat{W}] + [\hat{S}] + [\hat{\Lambda}] + [\hat{\Gamma}])} = \lim_{m \to \infty} \left( e^{[\hat{W}]/m}e^{[\hat{S}]/m}e^{[\hat{\Lambda}]/m}e^{[\hat{\Gamma}]/m} \right)^m = e^{[\hat{W}]e^{[\hat{S}]e^{[\hat{\Lambda}e^{[\hat{\Gamma}]}}}} \tag{3.19}
\]

It is easy to verify the constructions of the matrices in Equation 3.9 by taking their exponentials using Equation 3.2. Beginning with the translation matrix, its logarithm of $\hat{W}$ is constructed as
\[
\hat{W} = \begin{bmatrix} 0 & 0 & 0 & w_1 \\ 0 & 0 & 0 & w_2 \\ 0 & 0 & 0 & w_3 \\ 0 & 0 & 0 & 0 \end{bmatrix}
\]

Because the matrix $\hat{W}$ is nilpotent of degree 2, its exponential series is truncated
after the second term so

\[ \hat{T} = e^{[\hat{W}]} = I + \hat{W} = \begin{bmatrix} 1 & 0 & 0 & w_1 \\ 0 & 1 & 0 & w_2 \\ 0 & 0 & 1 & w_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \]

The logarithm of a rotation matrix, from Cayley’s formula, is a skew-symmetric matrix

\[ \hat{S} = \begin{bmatrix} 0 & -\eta_z & \eta_y & 0 \\ \eta_z & 0 & -\eta_x & 0 \\ -\eta_y & \eta_x & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \]

The Taylor expansion of the exponential matrix \( \hat{S} \) is

\[ \hat{R} = \begin{bmatrix} c(\eta_y)c(\eta_z) & s(\eta_x)s(\eta_y)c(\eta_z) - c(\eta_z)s(\eta_x) & s(\eta_x)s(\eta_z) + c(\eta_z)s(\eta_y)c(\eta_x) & 0 \\ c(\eta_y)s(\eta_z) & c(\eta_x)c(\eta_z) + s(\eta_x)s(\eta_y)s(\eta_z) & c(\eta_x)s(\eta_y)s(\eta_z) - s(\eta_x)c(\eta_z) & 0 \\ -s(\eta_y) & s(\eta_x)c(\eta_y) & c(\eta_x)c(\eta_y) & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \]

The logarithm of the scaling matrix is the diagonal matrix

\[ \hat{\Lambda} = \begin{bmatrix} \varepsilon_1 & 0 & 0 & 0 \\ 0 & \varepsilon_2 & 0 & 0 \\ 0 & 0 & \varepsilon_3 & 0 \end{bmatrix} \]

Because the exponential of a diagonal matrix is the matrix of the exponentials of the
diagonal entries, the scaling matrix is

\[
\hat{E} = e^{[\hat{\Lambda}]} = \begin{bmatrix}
e^{\varepsilon_1} & 0 & 0 & 0 \\
0 & e^{\varepsilon_2} & 0 & 0 \\
0 & 0 & e^{\varepsilon_3} & 0 \\
0 & 0 & 0 & e^0
\end{bmatrix}
\]

The logarithm of the shear matrix is

\[
\hat{\Gamma} = \begin{bmatrix}
0 & 0 & 0 & 0 \\
h_{xy} & 0 & 0 & 0 \\
h_{zx} & h_{yz} & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}
\]

Because the matrix \( \Gamma \) is nilpotent of degree 2, its exponential is simply

\[
e^{[\hat{\Gamma}]} = I \hat{\Gamma} = \begin{bmatrix}
1 & 0 & 0 & 0 \\
h_{xy} & 1 & 0 & 0 \\
h_{zx} & h_{yz} & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]

Using sum of these logarithm matrices, the matrix \( \hat{K} \) has the construction

\[
\hat{K} = \begin{bmatrix}
\varepsilon_x & (\eta_z) & (\eta_y) & t_x \\
(\eta_z + h_{xy}) & \varepsilon_y & (\eta_z) & t_y \\
(\eta_y + h_{zx}) & (\eta_x + h_{yz}) & \varepsilon_z & t_z \\
0 & 0 & 0 & 0
\end{bmatrix}
\]

where \( \eta \) is a rotation angle, \( \varepsilon \) is a scaling factor, \( h \) is a shear factor, and \( t \) is the translation.

The Lie algebra of the Lie group \( \mathbb{H}(4) \) is the product space of its factors. The algebra of spatial rotation is \( \mathbb{S}^3 \); the algebra of independent scale factors that can
take on any real values is $\mathbb{R}^3$; the algebra of shear factors that can take on any real value is $\mathbb{R}^3$; and the algebra of independent translation factors that can take on any real values is $\mathbb{R}^3$. The Lie algebra of 3D homogeneous transformations is therefore

$$\mathfrak{h}(4) = \mathbb{S}^3 \times \mathbb{R}^9$$

which has a dimension of 12. The tangent vector of the matrix $\hat{M}$ is a vector $\vec{k}$; here, it will be written as

$$\vec{k} = [\eta_x, \eta_y, \eta_z, \varepsilon_x, \varepsilon_y, \varepsilon_z, h_{xy}, h_{zx}, h_{yz}, t_x, t_y, t_z]^T$$

(3.21)

There are two important consequences of the derivations in this section. The first, already stated, is that a non-coplanar quadrilateral of spatial points can be represented as a non-singular homogeneous matrix. The second is that any such matrix can be linearized using matrix logarithms, so linear statistics (such as the SVD and certain types of PCA) can be applied to the tangent space of homogeneous matrices. In mathematical terms, this section has shown how to linearize the Lie group $\mathbb{H}(4)$.

### 3.3.3 A Numerical Implementation of 3D Surfaces

Here, simple surfaces are discretized as quadrilateral meshes consisting of ordered sets of spatial $m$ vertices, written as $(x^P, y^P, z^P) \in \mathbb{R}^3$, and associated sets of $k$ quadrilateral faces, with $m$ and $k \in \mathbb{Z}^+$ and $k \ll m$. A face is a set of four points forming a quadrilateral, henceforth simply “quadruples”. Quadruples intersect each neighbor in one or two vertices. One of the surfaces is selected as the base surface $B$ and the others are used as a set of altered surfaces $A_1, A_2, \ldots, A_n$, with $n \in \mathbb{Z}^+$. A traversal of the base surface $B$ is unraveled into a vector $\vec{b}$ and each altered surface
A_i is unraveled into a vector \( \vec{a}_i \). Vertex correspondence, between the base shape and the altered shapes, is implied by the order of points and the traversal direction.

The method preserves the structure by parameterizing the shape as a set of homogeneous matrices. The matrices are constructed by treating points as vectors with respect to a reference point.

A 3D surface is segmented into a set of \( k \) “quadruples”. Each quadruple was represented as a square homogeneous matrix constructed as Equation 3.15. The derivation is detailed in Subsection 3.3.1.

The rest of the implementation was similar to the 2D except that it accommodates for 12DOF instead of 6DOF. The tangent for the 3D case, represented as a matrix, was constructed as

\[
\begin{bmatrix}
\sigma_x & (-\eta_z + h_{xy}) & (\eta_y + h_{xz}) & t_x \\
(\eta_z + h_{xy}) & \sigma_y & (-\eta_z + h_{yz}) & t_y \\
(-\eta_y + h_{zx}) & (\eta_x + h_{yz}) & \sigma_z & t_z \\
0 & 0 & 0 & 0
\end{bmatrix}
\]

and the corresponding tangent vector \( \vec{k} \) was

\[
\vec{k} = [\eta_x, \eta_y, \eta_z, \sigma_x, \sigma_y, \sigma_z, h_{xy}, h_{xz}, h_{yz}, t_x, t_y, t_z]^T
\]

### 3.4 Summary

This chapter presented theoretical methods of analyzing discrete shapes using Lie groups. The method models global non-rigid transformations of 2D and 3D shapes as local affine transformations. This by parametrization of the shape as piece-wise
homogeneous transformation matrices. The constructed matrices form a Lie group $\mathbb{H}(n, \mathbb{R})$ that is a subgroup of the general linear group of invertible real matrices $\mathbb{GL}(n, \mathbb{R})$. As a Lie group, it is a group in the algebraic sense and a smooth manifold in the topological sense. This allows linearization of the manifold at the identity matrix point using the matrix logarithm. The linearization is reversed using the exponential of the vector on the tangent space of the manifold. This is a powerful framework that allows processing complex non-linear spaces with linear tools.

Despite the high level of abstraction of the concepts underlying the method, the implementation was simple and effective. The implementation relied on the matrix data structure and the matrix multiplication. The inverse of a matrix, the natural logarithm, and the exponential function were used in the implementation; these functions are included in most mathematical packages. The Gram-Schmidt method for $QR$ decomposition was re-implemented in a orthogonal-lower variant.
Chapter 4

Empirical Methods and Results

This chapter describes experiments that empirically evaluated the basic linear analytic method (principal components analysis, or PCA) and the proposed analytic method that used Lie groups. Both PCA and the Lie-group method were implemented in MATLAB (Mathworks, Natick, USA). The methods were applied to 2D contours and 3D surfaces extracted from computed tomography (CT) medical images.

CT scans were obtained from patients who went through computer-assisted hip resurfacing in Kingston General Hospital. All patients consented using their medical images in research. The data collection was approved by the Queen’s University General Research Ethics Board, which was the relevant Institutional Review Board (IRB). The images were acquired using a GE Lightspeed 16-slice CT scanner (General Electric, Milwaukee, USA) with a moving gantry; axial images had a 1.25mm slice thickness reconstructed from automatic tube-current modulation.

The femurs of the patients were segmented in the CT scans as part of their operative procedures for hip arthritis. For the 2D dataset, multi-planar reformats were performed on 83 scans to obtain the oblique axial plane of the proximal femur. For
the 3D dataset, an interpolation algorithm was performed on 50 scans to extract the surfaces of the proximal femur. All CT image processing were performed using Mimics (Materialise, Leuven, BE).

The data was preprocessed according to the following steps. First, the region of interest, the proximal femur cut right below the lesser trochanter, was extracted as a cube from the CT volume. Second, each slice of the selected cube was segmented by applying the threshold function in Mimics and then the bony area was extracted as a blob of grey values. Third, contours were extracted from the blobs using an edge-detection function in Matlab, with a required number of points as 33. Fourth, the contours were processed by determinant computation for consecutive triples of points to test for co-linearity. If co-linear points were detected, the middle point was translated by 0.5 mm. In the 2D experiment, the contours were directly used. In the 3D experiment, the contours extracted from a volume were used to construct a quadrilateral mesh by connecting consecutive contours with diagonal edges in order to avoid co-planarity.

Application of the method to 2D planar data is discussed in Section 4.1 and application to 3D spatial data is discussed in Section 4.2. The chapter concludes with a summary in Section 4.3.

4.1 Femoral Head-Neck Cross-Sections in 2D

A femoral head-neck contour was extracted from each CT slice by manually segmenting the femoral slice region and applying an edge detection algorithm. The data was undistorted real data that included uncertainty in the segmentation accuracy. Traditional shape analysis required a fixed number $m$ of corresponding vertices from
each contour. The choice of the number \( m \) of vertices could have been critical: the sampling had to be dense enough to represent the contour; it had to be sparse enough that it does not over-sample the continuous underlying shape; and it needed to satisfy requirements of Lie-group analysis. To meet the latter requirements, \( m \) had to be divisible by 3 and each triple of vertices could not be co-linear. After much preliminary work, the contours were resampled using \( m = 33 \) vertices that were approximately equally spaced on a representative example shape.

For PCA implementation, each contour was unraveled into the \( 66 \times 1 \) “position” vector \( \vec{p} \), which was structured as

\[
\vec{p} = [\vec{p}_1, \vec{p}_2, \ldots, \vec{p}_{33}]^T
\]

The set of all contours was structured as a \( 66 \times 83 \) matrix \( M \). The mean contour \( \vec{\mu} \) was computed along the rows of \( M \), then subtracted from each column to give a mean-centered dataset \( D \) with the same dimensions of \( M \). Following standard practice, \( D \) was factored using the SVD into orthogonal matrices \( U \) and \( V \), and a diagonal matrix \( \Sigma \) as

\[
D = U \Sigma V^T
\]

The singular values and their accumulated sum are plotted in Figure 4.1. It is clear from the plot that the dataset was challenging for PCA. The accumulation of components only approached the 90\% criterion after 24 components were used. Table 4.1 presents the values of the first 6 components, the ratio of those components to the mean value, the cumulative percentage of total variation, and the difference between components in sequence. Each of these criteria might be used to choose a
subset of components for subsequent analysis.

![Graph showing PCA singular values and cumulative sums for modes of variation of femoral head-neck cross-sections.](image)

Figure 4.1: PCA singular values and cumulative sums for modes of variation of femoral head-neck cross-sections.

For analysis using the Lie group $\mathbb{H}(3)$, a base contour $B$ was arbitrarily selected and parametrized as a set of 11 $\mathbb{H}(3)$ matrices $B_j$. The matrices $B_j$ were concatenated as the matrix $B$. A similar process was applied to each altered contour $A$ to generate the corresponding matrix $A_i$.

A matrix $C_i$ was computed by multiplying the inverse of each matrix $B_j$, from the left, with the corresponding matrices $A_i$, for $i = 1, \ldots, 11$. A tangent vector $k_i$ was computed for each matrix $C_i$. All the vectors corresponding to the contour transformation of index $i$ were concatenated as a column vector $\tilde{k}_i$. All such vectors were concatenated as columns in a matrix $K$. As in PCA, the matrix $K$ was factored...
Table 4.1: First 6 singular values of PCA for 2D femoral head-neck cross-sections.

<table>
<thead>
<tr>
<th>$PC_k$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_k$</td>
<td>668.35</td>
<td>149.63</td>
<td>96.49</td>
<td>75.32</td>
<td>66.60</td>
<td>55.77</td>
</tr>
<tr>
<td>$\bar{\sigma}_k$</td>
<td>25.30</td>
<td>5.66</td>
<td>3.65</td>
<td>2.85</td>
<td>2.52</td>
<td>2.11</td>
</tr>
<tr>
<td>$100 \times \frac{\sum_{j=1}^{k} \sigma_j}{\sum_{j=1}^{p} \sigma_j}$</td>
<td>38.33</td>
<td>46.92</td>
<td>52.45</td>
<td>56.77</td>
<td>60.59</td>
<td>63.79</td>
</tr>
<tr>
<td>$\sigma_{k-1} - \sigma_k$</td>
<td>-</td>
<td>518.72</td>
<td>53.14</td>
<td>21.17</td>
<td>8.72</td>
<td>10.82</td>
</tr>
</tbody>
</table>

using the SVD into an orthogonal matrix $U$, a diagonal matrix $\Sigma$, and an orthogonal matrix $V$ as

$$K = U \Sigma V^T$$

To reduce the potential that a poor choice of the base shape would skew results, shapes were exhaustively searched to find the best base shape $B$ that maximized the percentage recovered by the first few components. The singular values of the Lie-group analysis, and their accumulated sums, are plotted in Figure 4.2. It is clear from the plot that only 2 components were needed to recover 96.36% of the dataset. Table 4.2 presents the values of the first 6 components in similar manner to PCA.

Table 4.2: First 6 singular values of Lie-group analysis for 2D femoral head-neck cross-sections.

<table>
<thead>
<tr>
<th>$PC_k$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_k$</td>
<td>131263.47</td>
<td>36207.56</td>
<td>6041.54</td>
<td>31.77</td>
<td>20.83</td>
<td>17.24</td>
</tr>
<tr>
<td>$\bar{\sigma}_k$</td>
<td>4968.86</td>
<td>1370.61</td>
<td>228.70</td>
<td>1.20</td>
<td>0.79</td>
<td>0.65</td>
</tr>
<tr>
<td>$100 \times \frac{\sum_{j=1}^{k} \sigma_j}{\sum_{j=1}^{p} \sigma_j}$</td>
<td>75.52</td>
<td>96.36</td>
<td>99.83</td>
<td>99.85</td>
<td>99.86</td>
<td>99.87</td>
</tr>
<tr>
<td>$\sigma_{k-1} - \sigma_k$</td>
<td>-</td>
<td>95055.91</td>
<td>30166.02</td>
<td>6009.77</td>
<td>10.94</td>
<td>3.59</td>
</tr>
</tbody>
</table>

When only two components are used to describe a set of shapes, it is prudent to produce a simple scatter plot and visually inspect the results. Figure 4.3 presents the
parameters that correspond to the 83 individual 2D contours that were included in the study. As expected there was no distinct trend in the data.

A scatter plot of the parameters for the first three parameters is shown in Figure 4.4. No obvious pattern emerged from the addition of the third component.

To better visualize shape descriptors, singular values were scaled using weights of $-\sqrt{3}$ and $\sqrt{3}$ and used to reconstruct shapes for the purpose of illustration, a practice that is common in literature of registration and segmentation. The weights are selected to cover the maximum values indicating component variations.

The scaling was applied to both the first and second components. The results showed that most of the variation was controlled by the first component, with the
Figure 4.3: Scatter plots of the first two Lie-group parameters for the 2D data set. Each dot in the plot represents a shape reconstruction that described over 95% of the variation of a discrete contour.

The second component having less effect. Figures 4.5 and 4.6 shows the variation of the first two components respectively.
Figure 4.4: Scatter plots of the first three Lie-group parameters for the 2D data set. Each dot in the plot represents a shape reconstruction that described over 99% of the variation of a discrete contour.
Figure 4.5: 2D contour base shape variation using first component. The solid red line is the base shape $B$. The solid blue line with asterisk markers is, from left to right, the constructed shape $B - \sqrt{3}\sigma_1$, $B$, and $B + \sqrt{3}\sigma_1$. 
Figure 4.6: 2D contour base shape variation using second component. The solid red line is the base shape $\mathcal{B}$. The solid blue line with asterisk markers is, from left to right, the constructed shape $\mathcal{B} - \sqrt{3}\sigma^2$, $\mathcal{B}$, and $\mathcal{B} + \sqrt{3}\sigma^2$. 
4.2 Proximal Femur Surfaces in 3D

As in the 2D case, 3D shape analysis required a fixed number $m$ of corresponding vertices on each surface. The vertices also had to meet the requirements for Lie-group analysis, which were that the number $m$ of vertices had to be divisible by 4 and that each quadruple of vertices could not be co-planar.

In this experiment, CT scan segmentations of 50 patients were re-processed into quadrilateral surfaces that were then resampled using 2244 vertices. This was done by arbitrarily selecting one subject, manually segmenting the surface of the proximal femur into a quadrilateral mesh, and then applying a standard bootstrapping algorithm to establish correspondences among the vertices.

For principal-components analysis, each surface was unraveled into the $6732 \times 1$ “position” vector $\vec{p}$ that was structured as

$$
\vec{p} = [x_1 p, x_2 p, \ldots, x_{2244} p, y_1 p, y_2 p, \ldots, y_{2244} p, z_1 p, z_2 p, \ldots, z_{2244} p]^T
$$

The set of all surfaces was structured as the $6732 \times 50$ matrix $M$. The mean surface $\bar{\mu}$ was computed along the rows of $M$, then subtracted from each column to have a mean centered dataset $D$ with the same dimensions of $M$. Following standard practice, $D$ was factored using the SVD into orthogonal matrices $U$ and $V$, and a diagonal matrix $\Sigma$ as

$$
D = U \Sigma V^T
$$

The singular values and their accumulated sums are plotted in Figure 4.7. It is clear from the plot that the dataset was challenging for PCA. The accumulation of components only approached the 90% criterion after 23 components were used.
Table 4.3 presents the values of the first 6 components, the ratio of those components to the mean value, the cumulative percentage of total variation, and the difference between components in sequence. Each of these criteria might be used to choose a subset of components for subsequent analysis.

![Diagram showing PCA singular values and cumulative sums for modes of variation of proximal femur surfaces.]

Figure 4.7: PCA singular values and cumulative sums for modes of variation of proximal femur surfaces.

For analysis using the Lie group $\mathbb{H}(4)$, a base shape $B$ was arbitrarily selected and parametrized as a set of 561 matrices $jB$. The matrices $jB$ were concatenated as the matrix $B$. A similar process was applied to each altered surface $A$ to generate the corresponding matrix $A_i$.

A matrix $jC_i$ was computed by multiplying the inverse of each matrix $jB$, from the left, with the corresponding matrices $jA_i$, for $i = 1, \ldots, 561$. A tangent vector
Table 4.3: First 6 singular values of PCA for 3D proximal femur surfaces.

<table>
<thead>
<tr>
<th>$PC_k$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_k$</td>
<td>881.05</td>
<td>357.57</td>
<td>188.92</td>
<td>182.96</td>
<td>164.94</td>
<td>160.60</td>
</tr>
<tr>
<td>$\overline{\sigma}_k$</td>
<td>33.35</td>
<td>13.54</td>
<td>7.15</td>
<td>6.93</td>
<td>6.24</td>
<td>6.08</td>
</tr>
<tr>
<td>$100 \times \frac{\sum_{j=1}^{k} \sigma_j}{\sum_{j=1}^{p} \sigma_j}$</td>
<td>20.19</td>
<td>28.38</td>
<td>32.71</td>
<td>36.90</td>
<td>40.68</td>
<td>44.36</td>
</tr>
<tr>
<td>$\sigma_{k-1} - \sigma_k$</td>
<td>−</td>
<td>523.48</td>
<td>168.65</td>
<td>5.96</td>
<td>18.02</td>
<td>4.34</td>
</tr>
</tbody>
</table>

$\vec{\kappa}_i^j$ was computed for each matrix $jC_i$. All the vectors corresponding to the surface transformation of index $i$ were concatenated as a column vector $\vec{k}_i$. All such vectors were concatenated as columns in a matrix $K$. As in PCA, the matrix $K$ was factored using the SVD into an orthogonal matrix $U$, a diagonal matrix $\Sigma$, and an orthogonal matrix $V$ as

$$K = U \Sigma V^T$$

As in the 2D experiment, to reduce the potential that a poor choice of the base shape would skew results, shapes were exhaustively searched to find the best base shape $B$ that maximized the percentage recovered by the first few components. The singular values of the Lie-group analysis, and their accumulated sums, are plotted in Figure 4.8. It is clear from the plot that only 1 component was needed to recover 91.23% of the dataset. Table 4.4 presents the values of the first 6 components in similar manner to PCA.

As in the 2D experiment, a simple scatter plot was produced and visually inspected. Figure 4.9 presents the parameters that correspond to the 50 individual 3D surfaces that were included in the study. There clearly were clusters among the surface shapes; based on just the first component there may have been four clusters. At the time of submitting this dissertation for publication, no further information was
available regarding morphological or pathological implications of these data clusters. Further classification may be an interesting subject for future research.

A scatter plot of the parameters for the first three parameters is shown in Figure 4.10. This plot clearly illustrates that there are clusters and outliers of parameter values for the femur surfaces of these arthritic patients.

To better visualize shape descriptors, singular values were scaled using weights of $-\sqrt{3}$ and $\sqrt{3}$ and used to reconstruct shapes for the purpose of illustration, a practice that is common in literature of registration and segmentation.

The scaling was applied to both the first and second components. The results showed that most of the variation was controlled by the first component, with the
Table 4.4: First 6 singular values of Lie-group analysis for 3D proximal femur surfaces.

<table>
<thead>
<tr>
<th>$PC_k$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_k$</td>
<td>290.12</td>
<td>4.71</td>
<td>3.52</td>
<td>1.13</td>
<td>1.10</td>
<td>1.03</td>
</tr>
<tr>
<td>$\frac{\sigma_k}{\bar{\sigma}}$</td>
<td>10.98</td>
<td>0.18</td>
<td>0.13</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>$100 \times \frac{\sum_{j=1}^{k-1} \sigma_j}{\sum_{j=1}^{6} \sigma_j}$</td>
<td>91.23</td>
<td>92.71</td>
<td>93.82</td>
<td>94.18</td>
<td>94.52</td>
<td>94.84</td>
</tr>
<tr>
<td>$\sigma_{k-1} - \sigma_k$</td>
<td>–</td>
<td>285.40</td>
<td>1.20</td>
<td>2.39</td>
<td>0.04</td>
<td>0.06</td>
</tr>
</tbody>
</table>

second component having less effect. Figures 4.11 and 4.12 show the variation of the first two components respectively.
Figure 4.9: Scatter plots of the first two Lie-group parameters for the 3D data set. Each dot in the plot represents a shape reconstruction that described over 92% of the variation of a discrete surface.
Figure 4.10: Scatter plots of the first three Lie-group parameters for the $2D$ data set. Each dot in the plot represents a shape reconstruction that described over 93% of the variation of a discrete surface.
Figure 4.11: 3D surface base shape variation using first component. The red dots are the base shape $B$. The blue dots are, from left to right, the constructed shape $B - \sqrt{3}\sigma_1$, $B$, and $B + \sqrt{3}\sigma_1$. 
Figure 4.12: 3D surface base shape variation using second component. The red dots are the base shape $B$. The blue dots are, from left to right, the constructed shape $B - \sqrt{3}\sigma_2$, $B$, and $B + \sqrt{3}\sigma_2$. 
4.3 Summary

This chapter described empirical comparisons of traditional PCA and the novel Lie-group analysis of discrete shapes. Data, in $2D$ and $3D$, were extracted from CT scans of patients who later underwent computer-assisted surgery for early hip arthritis. The first dataset had 83 contours of femoral head-neck cross-sections and the second had 50 surfaces of proximal femurs.

For both data sets, the Lie-group analysis outperformed PCA in descriptive compactness. Between 2 to 3 components were required by the new analysis to cover 90% of the population, whereas PCA required 24 components. The results also showed that there were classification patterns in plotting the first two components of the new analysis that may be topics for future research.
Chapter 5

Conclusions

This thesis was an exploration of one way to analyze discrete shapes using Lie groups. The basic mathematics – Lie groups – is a century-old concept that has been the foundation for many physics theories at scales from elementary particles to cosmology. Lie groups have only captured attention of computer scientists, mainly in computer vision and medical image analysis, in the last ten years.

5.1 Discussion

This section provides a summary and discussion of the thesis. It will briefly examine the underlying theory, related frameworks, the methods, and the significance of the work described in this dissertation.

5.1.1 The Underlying Theory

The power of a Lie group is that it is a structure in two seemingly distinct areas of mathematics, namely algebra and geometry. Algebra is the branch of mathematics
that studies systems of relations using symbols; geometry is the branch of mathematics that studies properties, relationships, and measurements of points and shapes in space. Their combination enables the symbolic analysis, or generalization, of shapes. Because a Lie group is both an abstract group and a smooth manifold, many algebraic operations have a geometrical interpretation and *vice versa*. A prominent example in this dissertation was that the process of following a geodesic curve in the topological space of the manifold was equivalent to performing a symbolic operation on the homogeneous matrix that represented a point in the manifold. This double nature allowed considerable simplification of preceding work, which attempted to perform computations within a manifold that were numerical approximations; here, the exact solution was found by straightforward matrix operations that have long been known in linear algebra.

### 5.1.2 Related Frameworks

Discrete shapes are a computer representation of objects. One way of modeling shapes is by applying statistics to a population of shapes. The most common statistical shape analysis method is the point distribution model where shapes are represented as points or landmarks. Principal component analysis, commonly abbreviated as PCA, can be applied to some populations of shapes to find lower dimensional representations. PCA manipulates shapes as points in a Euclidean space; specifically, it uses vector addition as an operator to map one shape to the other. PCA is adequate for many general applications of data that are in a Euclidean, or linear space. Medical images and anatomical shapes are, however, better described as being in some non-Euclidean space for which PCA is not an effective description.
Non-Euclidean, or non-linear, spaces can be modeled as smooth manifolds. Two bodies of work have previously studied generalizations of PCA to smooth manifolds: principal geodesic analysis (PGA) and Log-Euclidean framework. PGA was applied to both m-reps and symmetric positive-definite matrices that represented diffusion-tensor images, each of which are a Lie group. PGA approximated geodesic distances by the logarithmic mapping to the tangent space at the identity, with PCA then applied to the tangent vectors in that space. Exact PGA is another version that actually computes the geodesic on the manifold, but for which there is a trade-off between accuracy and efficiency. The Log-Euclidean framework was applied to symmetric positive-definite matrices representing diffusion-tensor images, and to rigid and affine transformations of local regions representing magnetic-resonance images. The Log-Euclidean method was mainly used for registration. As the name indicates, it uses the logarithmic mapping of matrices to perform simpler computations on the associated tangent space.

5.1.3 The Method and its Significance

Similar to PGA and Log-Euclidean frameworks, the method of this thesis used the power of Lie groups but in a different sense. This new method recognizes that the Lie group of homogeneous transforms inherits a special property, which is that the element matrices are also operators on the group; in other words, a first homogeneous matrix is mapped to a second homogeneous matrix by multiplication with a third matrix. This third matrix is an element of the matrix Lie group and, geometrically, is equivalent to the geodesic between the two points on the manifold. Shape vertices were parametrized as homogeneous matrices, which implies that an operator matrix is
an affine transformation of the original vertices to new vertices sense. The logarithmic
mapping of a matrix that operates on a given set of vertices is a vector in the tangent
space of the point on the manifold that corresponds to the set of vertices; this tangent
vector is a transformation element in a higher-dimensional Euclidean space. This
allows computations, such as linear operations and statistics, on the tangent space. In
the work reported in this dissertation, the singular-value decomposition was applied to
compute the most significant components but other linear computations can doubtless
be applied.

This theoretical development is an important finding in the importance of data
representation, and how this representation can simplify computations without loss
of accuracy. A homogeneous transform is a geodesic on the manifold of homogeneous
matrices and it has an equivalent vector space where associated linear computations
are exact, rather than being approximated. This insight into data representation,
and manifold structure and properties, can open doors for analyzing different types
of data in the era of big data.

5.1.4 The Empirical Results and their Significance

The Lie-group method was applied to two clinical datasets: 2D contours of femoral
head-neck cross-sections and 3D surfaces of proximal femurs.Those two datasets are
of particular significance because of their high variability and their use in computer-
assisted surgical planning. The method outperformed PCA in compactness, needing
only 2 to 3 components to capture variations of the data whereas PCA needed com-
ponents of 20 or more to achieve the same representational richness. This work
incidentally revealed a clear clustering in the 3D data, the anatomical implications of
which were unclear at the time that this dissertation was submitted.

The compactness results show promise for a leap in the automation of registration and segmentation, with fewer components needed to be adjusted for data fitting. This can significantly facilitate surgical planning. The clustering results can influence at least two areas of research, namely computer-aided diagnosis and forensic anthropology. This influence could be investigated by conducting further studies with healthy control samples versus pathological samples for diagnosis, or with different groups of people for forensic anthropology.

5.1.5 Recommendations

In conclusion, processing data as Lie groups is still a young field of research for computer scientists that has excellent promise. An important step to understanding data, and their underlying manifold, is to devise a suitable representation that simplifies computations without loss of accuracy. The examples described in this dissertation were Lie groups of homogeneous transforms, and their parametrization of discrete shapes as homogeneous matrices. This work relied heavily on the power of the logarithmic and exponential mappings between the manifold and the associated tangent space. This insight could influence other areas of computing because only the medical-imaging community has, to date, shown interest in Lie groups. One recommendation that comes from this research is that computer scientists be introduced to differential geometry, particularly because (at the time of writing) virtually all textbooks are written for practitioners in mathematics and physics.
5.2 Contributions

This thesis had five main contributions:

- Recognizing that certain matrices, called homogeneous transforms, were simultaneously operators and elements of Lie groups.

- Providing a new shape parametrization technique of $2D$ contours and $3D$ surfaces as homogeneous matrices.

- Deriving a novel theoretical foundation for analyzing discrete shapes.

- Developing a numerical framework for implementation of the derived theory.

- Validating an implementation using clinical $2D$ and $3D$ datasets.

5.3 Limitations and Future Work

This thesis is the first application of Lie groups to the geometry of discrete shapes. Three main future development directions are:

- A procedure for improved selection of base shapes is needed. The implementation described here was based on either arbitrary selection or a semi-exhaustive search on the dataset. The challenge is that there was no obvious closed-form way to compute the mean in the manifolds under scrutiny; the mean may not exist, or may not be unique. An alternative may be to implement the mean as an optimization problem, with the understanding that that optimization may be computationally expensive and that there is a risk of local-minima solutions.
• Effective matrix processing is highly dependent on numerical stability. Shape parameterizations required tiny numerical adjustments to deal with collinear and coplanar points, respectively, in 2D contours and 3D surfaces. Incorporating matrix parametrization into contour or surface construction techniques might resolve this problem. Another approach might be to develop stable processing techniques for singular matrices.

• Further validation of the method using other anatomical and non-anatomical shapes is prudent. Consideration of how to use Lie groups in analysis of other kinds of data may also be promising.

Lie-group analysis may spark a renaissance of data analysis that will provide accurate and efficient data understanding.
Appendix: Mathematical Background

The main idea presented in this dissertation was the analysis of shapes in non-linear spaces by using familiar and easily implemented software tools. The analysis was done by a process of linearization of the non-linear space. For this purpose, a powerful tool that is derived from two main branches of mathematics – algebra and geometry – was used: the theory of Lie groups and Lie algebras. Lie groups are simultaneously groups and smooth manifolds. Because the foundation of this theory is not covered in most current computer-science curricula, this chapter introduces the essential concepts for understanding the topic and following the development of the method.

This chapter begins with algebra, then moves to geometry and their unification. Section A.1 covers elementary group theory by defining groups, basic group properties, and related concepts. Section A.2 introduces the topology, topological spaces, and their properties. Section A.3 uses concepts from topology to present topological manifolds and smooth manifolds. Section A.4 builds upon the understanding of manifolds to define tangent spaces of manifolds as associated linear spaces. Section A.5 starts delving into the core of the Lie groups and their structures and properties. Section A.6 describes the linearization of Lie groups, which is the key concept underlying
this thesis. The chapter concludes with a summary in Section A.7.

A.1 Group Theory

Group theory is the basis of abstract algebra. A group structure consists of a set of elements $G$ and a map $G \times G$ into $G$ that satisfy four fundamental axioms. The material in this section is based on texts by Hall [36] and Gilmore [30], but any standard book on group theory should cover these concepts.

Definition A.1.1. A group $G$ is a set of elements $\{g, h, k, \ldots\}$, with a combinatorial binary operation $\circ$ that satisfies four axioms:

- **Closure:** $\forall g, h \in G, (g \circ h) \in G$
- **Associativity:** $\forall g, h, i \in G, (g \circ h) \circ i = g \circ (h \circ i)$
- **Identity:** $\exists e \in G, \text{ such that } \forall g \in G, e \circ g = g = g \circ e$
- **Inverse:** $\forall g \in G, \exists g^{-1} \in G, \text{ such that } g \circ g^{-1} = e = g^{-1} \circ g$

The closure property implies that the group operation, also called multiplication or product, does not generate new elements outside the group. If the operands are from the group then the product is in the group.

The associativity asserts that there are no precedence rules in the group. The group has an operation but the parentheses does not affect the product. It is important to note that, in general, the order of multiplication matters unless it is explicitly stated that the group under consideration is commutative. A commutative group is called an Abelian group and defined as:
Definition A.1.2. An **Abelian** group is a group that satisfies the axiom

- **Commutativity**: \( \forall g, h \in G, \ g \circ h = h \circ g \)

The identity property defines a special element in the group that maps every element in the group to itself. Because the uniqueness of such an element is not specified in the definition, it can be stated as a proposition:

**Proposition A.1.1.** *The identity element \( e \in G \) is unique.*

Existence of an identity element is closely allied with the existence of an inverse for each element. This inverse element maps any element to the identity element. The inverse of any element in the group is unique:

**Proposition A.1.2.** *The inverse \( g^{-1} \) of an element \( g \in G \) is unique.*

The definition of a group states that every element has an inverse. Even the identity element has an inverse, which is a proposition:

**Proposition A.1.3.** *The inverse of the identity is the identity such that \( e^{-1} = e \).*

Another important property is that the inverse relation is one-to-one:

**Proposition A.1.4.** *The inverse of the inverse of an element is the element such that \( (g^{-1})^{-1} = g \).*

The inverse is distributive over the group product:

**Proposition A.1.5.** *The inverse of the product is the commutated product of the inverse of each element such that \( (g \circ h)^{-1} = h^{-1} \circ g^{-1} \).*
It may be useful to consider some simple examples of groups. To define a group, a set of elements must be specified along with the group operation. The four axioms also need to be verified to qualify the set and operation as a group.

**Example A.1.1.** *The trivial group \( \{e\} \) and any operator \(*\).*

The trivial group \( \{e\} \) consists of the set of the identity element \( e \), with a group multiplication operation \(*\). It is clear that all group properties hold for this group.

**Example A.1.2.** *The integers \( \mathbb{Z} \) and the arithmetic operator \(+\).*

The group of integers \( \mathbb{Z} \) consists of the set of all integers and the group operation of basic addition \(+\). Closure holds in \( \mathbb{Z} \) because the addition of any two integers is an integer. Associativity holds because of how the addition operation works on elements of the group. The identity element is 0, which maps each element to itself under addition. Finally, the existence of inverse of each positive integer is its negative integer, and the opposite. Note that the set of this group is countably infinite.

**Example A.1.3.** *The real numbers \( \mathbb{R} \) and the arithmetic operator \(+\).*

The group of real numbers \( \mathbb{R} \) consists of the set of all real numbers \( \mathbb{R} \) with addition operation \(+\). The four axioms follows the same arguments used with integers. Note that the set of this group is uncountably infinite.

**Example A.1.4.** *The \( n \)D Euclidean vectors \( \mathbb{R}^n \) and vector addition \(+\).*

The \( n \)D Euclidean vectors \( \mathbb{R}^n \) is a group structure that is used in this dissertation. It consists of a set of \( n \)D vectors with the vector-addition operation \(+\). The addition of two vectors produces a vector of the same dimensionality. Vector addition is associative, as are real numbers. It has a unique identity: a vector of zeros in the
same dimension of the group. Each element has a unique inverse that is the element-wise arithmetic inverses of the real numbers in its entries.

**Example A.1.5.** The unit circle $S^1$ in the complex plane with complex number multiplication.

The group of complex numbers $\mathbb{C}$ consists of the set of all complex numbers $\mathbb{C}$ with multiplication operation $\ast$. The unit circle group $S^1$ is a subset of $\mathbb{C}$ with the absolute value of 1. $\mathbb{C}$ is an abelian group so as the subgroup $S^1$. The unit circle can be parametrized by the rotation angle $\theta$ such that

$$z = e^{i\theta} = \cos(\theta) + i\sin(\theta)$$

for all $\theta$.

**Example A.1.6.** The set of $n \times n$ non-singular matrices and the operation of matrix multiplication.

This group has been so extensively studied that it has a distinct name, the general linear group abbreviated as $GL(n, \mathbb{R})$. It is the set of $n \times n$ non-singular matrices (i.e., having a non-zero determinant) composed of real entries, and the matrix multiplication operation. Multiplication of non-singular matrices produces a non-singular matrix. Matrix multiplication is associative. The identity is a matrix with the same dimension with ones in the diagonal and zeros elsewhere. Finally, the inverse of a given non-singular matrix is its matrix inverse, which is non-singular and thus is in the group.

These examples of groups are summarized in table A.1.

A new group can be created from existing groups. One such construction is a subgroups:
Table A.1: Examples of groups.

<table>
<thead>
<tr>
<th>Group</th>
<th>Elements</th>
<th>Operation</th>
<th>Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>The trivial group</td>
<td>{e}</td>
<td>*</td>
<td>e</td>
</tr>
<tr>
<td>The integers</td>
<td>\mathbb{Z}</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>The reals</td>
<td>\mathbb{R}</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>The (nD) Euclidean vectors</td>
<td>\mathbb{R}^n</td>
<td>+</td>
<td>[0]_{n\times1}</td>
</tr>
<tr>
<td>The unit circle</td>
<td>\mathbb{S}^1</td>
<td>*</td>
<td>1 + 0i</td>
</tr>
<tr>
<td>The general linear</td>
<td>GL(n, \mathbb{R})</td>
<td>\times</td>
<td>I</td>
</tr>
</tbody>
</table>

Definition A.1.3. A **subgroup** \(H\) of a group \(G\) is a subset of the elements that satisfy the axioms:

- \(e \in H\)
- \(\forall h \in H, h^{-1} \in H\)
- \(\forall h, j \in H, h \circ j \in H\)

It is implicit that the subgroup has the same group operation \(\circ\) as the original group. The axioms of the subgroup assert that the selected subset of elements forms a group, so group axioms hold. The first axiom means that the identity is an element of each subgroup; because the original group has only one identity, and the generated subgroup must have an identity, the identity of the subgroup must be the same identity of the group. The second axiom assures that the inverse of each element in the subgroup is included in the subgroup, which also assures that the inverse property holds in the subgroup. The third axiom assures group closure within the subgroup, so that products of subgroup elements are in the subgroup. Note that associativity is naturally inherited from the original group.

Each group has at least two subgroups, the trivial group and the group itself. Two non-trivial subgroups can be derived from the previous examples of groups.
Example A.1.7. *The even integers.*

Even integers are a subgroup of the group of integers. The identity 0 is even, so it is in the subgroup. The inverses of even integers are even integers. Finally, addition of even integers produce even integers.

Example A.1.8. *The set of \( n \times n \) non-singular matrices with unit determinants.*

The special linear \( \mathbb{SL}(n, \mathbb{R}) \) is the set of \( n \times n \) non-singular matrices, each of which has determinant 1, plus the matrix-multiplication operation. The determinant of the identity matrix \( I \) is 1 so it is in the subgroup. The determinant of the inverse is the reciprocal of the determinant, so the inverse of a unit-determinant matrix has a unit determinant. Third, the determinant of the product is the product of determinants, so the product of two unit-determinant matrices has a unit determinant.

Another way of generating subgroups is to take the intersection of existing subgroups, which is a proposition:

**Proposition A.1.6.** *The intersection \( (H \cap J) \), where \( H \subset G \) and \( J \subset G \), is also a subgroup of \( G \).*

Another way of generating new groups is the Cartesian product, or outer product, of existing groups.

**Definition A.1.4.** *A Cartesian product \( G \times H \) of groups \( G \) and \( H \) is the set of ordered pairs \( (g_i, h_i) \) with the group operation defined as*

\[
(g_i, h_i) \circ (g_j, h_j) = (g_i \circ g_j, h_i \circ h_j)
\]

From this definition, the identity element is defined as the pair \( (e_G, e_H) \) where \( e_G \) is the identity element of the group \( G \), and \( e_H \) is the identity element of the group \( H \). The other group properties hold at the associated entry of the pair.
Example A.1.9. A cylinder is a Cartesian product $S^1 \times \mathbb{R}^1$, where $S^1$ is a unit circle and $\mathbb{R}^1$ is a line.

Example A.1.10. A torus is a Cartesian product $S^1 \times S^1$, where $S^1$ is a unit circle.

Groups can be mapped to each other, one important such mapping being homomorphism.

Definition A.1.5. A homomorphism is a map $\Phi : G \rightarrow H$ that satisfies the property

$$\forall g, h \in G, \Phi(g \circ h) = (\Phi(g) \circ \Phi(h))$$

Homomorphism preserves closure in the mapping. It follows that the identity and inverses are preserved by the mapping, which are propositions:

Proposition A.1.7. A homomorphism $\Phi : G \rightarrow H$ maps the identity element of $G$ to the identity element of $H$.

Proposition A.1.8. A homomorphism $\Phi : G \rightarrow H$ maps the inverse of an element $g^{-1}$ to the inverse of the mapping of $\Phi^{-1}(g)$ in $H$.

A special case of homomorphism is isomorphism:

Definition A.1.6. An isomorphism is a one-to-one and onto homomorphism.

A.2 Algebraic Topology

Algebraic topology is the use of abstract algebra to study topological spaces. The concepts defined in this section are abstract, and appreciated more when used in the
subsequent section to define manifolds as topological spaces, so only definitions are introduced and brief discussions are as provided needed. The material is this section is based on two texts by Lee [46, 47]. Standard texts in topology should cover these concepts.

Topology usually treats the empty set $\emptyset$ as an open set. With this understanding, a topology can be simply defined.

**Definition A.2.1.** A *topology* $\mathcal{T}$ is a collection of open subsets $\{X_i : X_i \subset X \text{ for } i = 1, \ldots, n \}$, that satisfy three axioms:

- $\mathcal{T}$ and $\emptyset$ are open sets
- $\bigcup X_i$ is open
- $\bigcap X_i$ is open

A topology is, basically, a set of open subsets of a defined set, including the empty set $\emptyset$. The union of sets of those subsets is open. The intersection of any set of such subsets is also open.

**Definition A.2.2.** A *topological space* is the pair $(X, \mathcal{T})$ of a set $X$ and the associated topology $\mathcal{T}$.

When the topology is understood from the context, the set is used to identify the topology. Examples of topological spaces may help to increase the reader’s understanding of topological spaces introduced.

**Example A.2.1.** A discrete space, such as a finite set or the set of integers.

A discrete set $X$ is a topological space because we can treat a finite set as an open set.
Example A.2.2. A metric space.

A metric space is a set with a distance measure $d(x, y) \forall x, y \in X$, also called a metric, for which:

- Positive: $d(x, y) = 0$ if $x = y$; otherwise, $d(x, y) > 0$
- Symmetric: $d(x, y) = d(y, x)$
- Triangle inequality: $d(x, z) \leq d(x, y) + d(y, z)$

In this context, elements of sets are points in spaces. An important concept is the neighborhood of a point.

Definition A.2.3. A neighborhood of point $p \in X$ is an open set $Y$ such that $p \in Y$.

One feature of topological spaces are that they are mapped to each other through continuous mappings.

Definition A.2.4. A continuous map $\Psi : X \to Y$ is a map of a topological space $X$ to a topological space $Y$ such that, if $Z \subset Y$ and open, then $\Psi^{-1}(Z) \subset X$ and open.

This definition of continuous mapping does not imply bijection, which must be separately defined.

Definition A.2.5. A homeomorphism is a bijective continuous map.

The definition of a topological space, so far, is very general and flexible. A specific example of restricted topological space is the Hausdorff space defined as follows.
Definition A.2.6. A **Hausdorff space** is a topological space for which \( \forall p, q \in X, \exists U, V \subseteq X \) such that \( U \cap V = \emptyset \), \( p \in U \) and \( q \in V \).

Another useful term is the basis of a topological space.

Definition A.2.7. A **basis** for a topology \( X \) is a collection of subsets \( B \) such that

- \( X = \bigcup X_i \)
- \( \forall x \in X_i \cap X_j \exists X_k \subseteq B_i \cap B_j \)

A concept, used briefly in subsequent definitions, is that of countability.

Definition A.2.8. A **first countable** topological space is one for which, at each point, there is a countable neighborhood basis.

Definition A.2.9. A **second countable** topological space is one for which there is a countable basis for its topology.

One of the key properties of a topological space is whether or not it is locally Euclidean.

Definition A.2.10. A **locally Euclidean** space is a topological space for which each point has a neighborhood homeomorphic to an open subset of \( \mathbb{R}^n \).

**A.3 Smooth Manifolds**

A special type of topological spaces is the topological manifold, in particular a smooth manifold. The material in this section is derived from Schutz [75], Lee [46, 47], and Frankel [27].
Definition A.3.1. A topological manifold $M$ is a topological space that satisfies three properties:

- $M$ is a Hausdorff space
- $M$ is second-countable
- $M$ is locally Euclidean

A subspace of a manifold is also a manifold because it inherits the three properties. Some well known sets are also topological manifolds.

Example A.3.1. The Euclidean space $\mathbb{R}^n$.

A Euclidean space $\mathbb{R}^n$ is a topological manifold of dimension $n$.

Example A.3.2. The unit n-sphere $S^n$.

The $S^n$ is the set of unit vectors in $\mathbb{R}^{n+1}$, which inherits the property of being a topological manifold from the mother group.

Example A.3.3. The product manifolds $M_1 \times \cdots \times M_k$

A manifold can be defined as a product of manifolds. The product manifold $M_1 \times \cdots \times M_m$ is a topological manifold of dimension $n_1 + \cdots + n_m$, where $m_i$ is the dimension of the manifold $M_i$.

Example A.3.4. The n-torus $T$

The n-torus $T = S^1 \times \cdots \times S^1$ is a topological manifold that is a product manifold of the unit spheres $S^1$. 
As a topological space, a manifold may be mapped to another topological space; there are plainly subsets of a manifold. A chart is a subset of a topological manifold that is locally Euclidean. In other words, there is a homeomorphism that maps the subset to a subset of $\mathbb{R}^n$.

**Definition A.3.2.** A chart of a manifold $\mathcal{M}$ is pair $(U, \Psi)$ of an open subset $U \subset \mathcal{M}$ and homeomorphism $\Psi : U \to \tilde{U}$, where $\tilde{U} \subset \mathbb{R}^n$.

A collection of such charts, the union of which covers a given manifold, is called an atlas. The atlas is used to map the topological manifold to a subset of $\mathbb{R}^n$.

**Definition A.3.3.** An atlas $\mathcal{A}$ for a manifold $\mathcal{M}$ is a collection of charts $(U, \Psi)$ that cover $\mathcal{M}$.

A smooth atlas $\mathcal{A}$ is an atlas with a smooth mapping.

**Definition A.3.4.** A smooth mapping $f : \mathcal{M}^n \to \mathbb{R}^m$ is function that for all $p \in \mathcal{M}$ there is a chart $(U, \Psi)$ such that $f \circ \Psi^{-1}$ is smooth on the open subset $\tilde{U} = \phi(U) \subset \mathbb{R}^n$.

The definition of a smooth function does not indicate whether or not it is bijective. A bijective smooth function is a diffeomorphism.

**Definition A.3.5.** A diffeomorphism is a smooth bijective map with a smooth inverse.

The previous definitions lead to defining a smooth manifold, which is an interesting structure with many applications.

**Definition A.3.6.** A smooth manifold is a pair of $(\mathcal{M}, \mathcal{A})$ of a topological manifold $\mathcal{M}$ and a smooth atlas $\mathcal{A}$. 
Some familiar sets are smooth manifolds.

**Example A.3.5.** *The Euclidean space $\mathbb{R}^n$.***

The Euclidean space $\mathbb{R}^n$ is a smooth manifold with dimension $n$. This manifold is called standard smooth structure.

**Example A.3.6.** *The general linear $\text{GL}(n, \mathbb{R})$.***

The general linear group $\text{GL}(n, \mathbb{R})$ is a subset of an $n^2$ dimensional vector space, and is smooth manifold of dimension $n^2$.

### A.4 Tangent Spaces

Tangent spaces are vector spaces that exist at each point of a smooth manifold.

**Definition A.4.1.** A *tangent space* $T_p\mathcal{M}$ is a vector space at point $p \in \mathcal{M}$.

A tangent space of a manifold contains all the directions that a tangent vector can pass through the manifold at a certain point. In other words, it is the set of all tangent vectors at point $p$.

**Definition A.4.2.** A *tangent vector* $\vec{v}_p$ is an element of the the tangent space $T_p\mathcal{M}$.

The set of all tangent spaces is defined as a tangent bundle.

**Definition A.4.3.** A *tangent bundle* $\mathcal{T}\mathcal{M} = \bigsqcup_{p \in \mathcal{M}} T_p\mathcal{M}$ is the disjoint union of tangent spaces at all points $p$ on the manifold $\mathcal{M}$.

A tangent bundle $\mathcal{T}\mathcal{M}$ is a $2n$ dimensional smooth manifold with smooth map $\pi : \mathcal{T}\mathcal{M} \to \mathcal{M}$. Another important concept is the definition of a vector field.
**Definition A.4.4.** A vector field is a smooth map $Y : \mathcal{M} \to \mathcal{T}\mathcal{M}$.

If $p \in \mathcal{M}$ and $X \in \mathcal{T}\mathcal{M}$, then there is a vector field $\hat{X}_p = X$.

### A.5 Lie Groups

A Lie group, named after the mathematician Sophus Lie, is a structure that has algebraic and topological properties. The algebraic properties are derived from groups and the topological properties are derived from smooth manifolds. The material in this section was derived Gilmore [30] and Hall [36].

A Lie group is simultaneously a group and a smooth manifold.

**Definition A.5.1.** A Lie Group $\mathcal{G}_L$ is a group $\mathcal{G}$ and a smooth manifold $\mathcal{M}$ such that the group operation $g \circ h \to k$ is a smooth map $\mathcal{G}_L \circ \mathcal{G}_L \to \mathcal{G}_L$, and the inverse $g \to g^{-1}$ is a smooth map $\mathcal{G}_L \to \mathcal{G}_L$.

The unification of group and manifold properties into a Lie group arises from the smoothness requirement imposed on both the group operation and the inverse property [30]. The manifold $\mathcal{M}$ parametrizes group elements $\{g, h, k, ...\}$ and operation $\circ$ through a smooth group composition map. The existence of the inverse is parametrized as a smooth map. A Lie group is homogeneous in that all points are “the same”: points can be mapped to each other through the group operation. Although topologically all points are the same, the identity element is special in the group algebra.
A.6 Lie Algebras

Lie algebras are constructed by linearizing Lie groups. The linearization is done by expanding the group combinatorial operator about the coordinates of the group elements at any given group element. The linearization of the Lie group forms a new set of elements that are the Lie algebra. A Lie algebra is a linear vector space, so the group operation is reduced to being vector addition.

**Definition A.6.1.** The **commutator**, or bracket, of two elements \( g, h \in G \) is defined as

\[
[g, h] = g^{-1}h^{-1}gh
\]

The commutator is equal to the group identity if and only if \( G \) is abelian.

**Definition A.6.2.** The **Jacobi identity** is defined as

\[
[g, [h, k]] + [h, [k, g]] + [k, [g, h]] = 0, \forall g, h, k \in G
\]

The Jacobi identity defines the order of evaluation of a binary operation behavior. In contrast with associativity, the order of evaluation is important.

**Definition A.6.3.** A **Lie algebra** \( g \) is a linearization of a Lie group \( G_L \) at an element of the group that satisfies three properties:

- The elements of a Lie algebra form a vector space
- Elements are closed under commutator
- Elements satisfy the Jacobi identity
The dimension of the manifold that parametrizes the Lie group is the same dimension of the Lie algebra. Because the Lie algebra is a vector space, it has a basis and an inner-product operator. Exponential mapping is the medium of communicating information between the manifold and the tangent space, or the Lie group and the Lie algebra.

For example, for every $n \times n$ real matrix $M$, the exponential $e^M$ is a continuous function of $M$ that uniformly converges and defined by the Taylor series

$$e^M = \sum_{k=0}^{\infty} \frac{M^k}{k!}$$

The identity matrix $I$ on the manifold is mapped to the origin of the vector space such that

$$e^0 = I$$

An important concept is the Lie product.

**Definition A.6.4.** The Lie product is defined for the matrices $X, Y \in \mathbb{GL}(n, \mathbb{R})$ as

- $e^{X+Y} = \lim_{m \to \infty} \left( e^{X/m} e^{Y/m} \right)^m$

The defined Lie product is the general form. For commutative matrices the following equation applies.

$$e^{X+Y} = (e^X e^Y)$$
A.7 Summary

A Lie group is a powerful mathematical structure that permits the processing of a complex non-linear space by working on a linear space, which is the associated Lie algebra. The power of a Lie group is that it is both an algebraic and a topological structure: a Lie group is a group and a smooth manifold. The smooth manifold is linearizable on the tangent spaces of each point on the structure. The tangent space at the identity is the Lie algebra. The exponential mapping is used as the medium to transfer elements between the Lie group and the Lie algebra.
Bibliography


[34] Gray H. *Anatomy, Descriptive and Surgical*. John W. Parker and Son, 1858.


