Jacek Stankiewcz

Riemannian center of mass on Kendall's shape space

Bachelor's thesis in BMAT
Supervisor: Assoc. Prof. Dr. Ronny Bergmann
June 2024

Bachelor's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

NTNU
Norwegian University of
Science and Technology

Jacek Stankiewcz

Riemannian center of mass on Kendall's shape space

Bachelor’s thesis in BMAT
Supervisor: Assoc. Prof. Dr. Ronny Bergmann
June 2024

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

NTNU
Norwegian University of
Science and Technology

Riemannian center of mass on Kendall's shape space

A Bachelor's thesis

Jacek Stankiewicz

Supervisor: Assoc. Prof. Dr. Ronny Bergmann

June 17, 2024

Contents

1 Introduction

The mathematical study of shape theory offers a relatively novel insight into the commonly known idea of what a shape is. The field saw its emergence with at least three distinct entries into what we now call shape theory [5]. One of those origins, and one of the earliest, was in 1977 by Kendall [6]. His investigation was originally prompted by archaeological, astronomical, geological and ornithological considerations [6]. Those are still important applications today, however, the field has grown considerable in the last four decades. Many novel applications are found in medicine [11] or computer vision [14].

In particular, Kendall's shape space provides a framework for analyzing the shapes of objects by considering them as points in a high-dimensional topological space. A key property of Kendall's shape space, is the Riemannian structure that it has. That is, it is a differentiable manifold. This space is constructed by removing the effects of translation, scaling, and rotation. By doing this we focus only on the actual shape of the object, not any other physical properties.

One of the more important concepts within shape theory, is determining the mean of shapes [7]. Conceptually, determining the mean of a normal data set is extremely straight forward. However, extending the concept of a mean to the Kendall's shape space, requires a higher level of insight. The fact of the shape space being a Riemannian manifold, helps solve this issue. That is, we might define a distance function[2], of which the minimizer, is defined to be the Riemannian mean. Actually finding the minimum of this function, is not straight forward. The existence and uniqueness of a solution is not guaranteed. However under some specific conditions, and restrictions on the spread of data points, a solution can be guaranteed [3]. In the context of shape theory, this means that we need the shapes that we wish to find the average of, to be relatively similar.

Furthermore, as one might expect, finding an analytic solution to the problem is not the way to go. For this reason, numerical methods need to be used, specifically, gradient descent is a well known algorithm that can be utilized. However, since the Kendall's shape space was a manifold, the gradient descent has to be generalized to be defined on manifolds, in which case it is called a Riemannian gradient descent[4].

We consider the above mentioned topics in depth throughout this thesis. In the beginning, that is section 2, we introduce the notions related to manifolds. In section 3 we consider the notions related to optimization on manifolds, which mostly consists of generalizing the well known results from the Euclidean space. The Kendall's shape space is properly defined and discussed in section 4. After that we consider in which cases the Riemannian center of mass might exist, and discuss it in the context of Kendall's shape space. Lastly we consider some numerical experiments to test the proceeding theory. In that we consider firstly the space of all two dimensional triangles, and secondly we determine the mean of a set of outlines of chess pieces.

2 Preliminaries on Manifolds

Because this section merely gives a basic introduction on manifolds and quotient manifolds, no information here is new or specific to this thesis. Therefore, for this section, unless stated otherwise, the information originates (with some paraphrasing) from [4], mostly sections 3 and 9.

2.1 Embedded submanifolds

A manifold is a topological space that locally resembles Euclidean space near each point. Since a manifold is a topological space, to define it in most general terms requires the context of terms such as charts, open sets, etc. However, for the purpose of this text we need not the most general representation of a manifold. We concern ourselves with a subclass of manifolds that are the embedded submanifolds. We define those in terms of a mapping on the space that the manifold is embedded in.

Definition 2.1. *Let $\mathcal{E}$ be a linear space of dimension d. A non-empty subset $\mathcal{M}$ of $\mathcal{E}$ is a smooth embedded submanifold of $\mathcal{E}$ of dimension n if either:*

- $n = d$ *and $\mathcal{M}$ is open in $\mathcal{E}$ (also called the open submanifold).*
- $n = d - k$ *for some $k \geq 1$ and, for each $x \in \mathcal{M}$, there exists a neighbourhood U of x in $\mathcal{E}$ and a smooth function $h : U \to \mathbb{R}^k$ such that:*
    1. *if y is in U, then $h(y) = 0$ if and only if $y \in \mathcal{M}$; and*
    2. *rank $Dh(x) = k$.*

We call h a local defining function of $\mathcal{M}$ at x.

As mentioned, the manifold resembles the Euclidean space near each point, that linear approximation at each point is the tangent space. Elements of this space are called tangent vectors. It consists of all the velocities of the curves that initially pass through the given point. The tangent vectors describe the possible directions of movement at that point on the manifold. We can define the tangent space by:

Definition 2.2. *Let $\mathcal{M}$a subset of $\mathcal{M}$. For all $x \in M$, define the tangent space by:*

$$T_xM = \{c'(0) \,|\, c : I \to M \text{ is smooth, and } c(0) = x\}$$

We can also define the tangent bundle, which is the collection of all the tangent spaces of a manifold. Each element in the bundle consists of a pair, the point and the tangent vector at that point. This is to avoid confusion, as two different spaces can share a vector.

Definition 2.3. *The tangent bundle of a manifold $\mathcal{M}$ is the disjoint union of the tangent spaces of $\mathcal{M}$:*

$$TM = \{(x, v) \,|\, x \in M, v \in T_xM\}$$

Having now defined specific spaces, can we introduce the theory needed for understanding the functions on those spaces.

Definition 2.4. *A subset U of $\mathcal{M}$ is open (or closed) in $\mathcal{M}$ if U is the intersection of $\mathcal{M}$ with an open (or closed) subset of E.*

Definition 2.5. *A neighborhood of x in $\mathcal{M}$ is an open subset of $\mathcal{M}$ which contains x. By extension, a neighborhood of a subset of $\mathcal{M}$ is an open set of $\mathcal{M}$ which contains that subset.*

Definition 2.6. *Let $\mathcal{M}$ and $\mathcal{M}'$ be embedded submanifolds of E and E'. A map $F : \mathcal{M} \to \mathcal{M}'$ is smooth at $x \in \mathcal{M}$ if there exists a function $\bar{F} : U \to E'$ which is smooth on a neighborhood U of x in E and such that F and $\bar{F}$ coincide on $\mathcal{M} \cap U$, that is $F(y) = \bar{F}(y)$, for all $y \in M \cap U$. We call $\bar{F}$ a local smooth extension of F around x. The map F is smooth if it is smooth at all $x \in \mathcal{M}$.*

We can then introduce the differential on the manifold in terms of a curve on $\mathcal{M}$ passing through the given point:

2.2 Differentiation and Riemannian gradient

In this subsection we discuss the different notions related to differentiability. We begin by stating the simple definition of a differential of a function on a manifold.

Definition 2.7. *The differential of $F : \mathcal{M} \to \mathcal{M}'$ at the point $x \in \mathcal{M}$ is the linear map $DF(x) : T_x\mathcal{M} \to T_{F(x)}\mathcal{M}'$ defined by:*

$$DF(x)[v] = \frac{d}{dt}F(c(t))\bigg|_{t=0} = (F \circ c)'(0) = 0$$

where c is a smooth curve on $\mathcal{M}$ passing through x at $t = 0$ with velocity v.

In the Euclidean case, the existence of the gradient is limited by the properties of the function itself, that is, it being differentiable. When dealing with a manifold, the space itself must possess the structure such that a gradient can be defined. That requirement is for the manifold to be Riemannian. That is, it must be equipped with a Riemannian inner product.

Definition 2.8. *An inner product on $T_x\mathcal{M}$ is a bilinear, symmetric, positive definite function $\langle \cdot, \cdot \rangle_x : T_x\mathcal{M} \times T_x\mathcal{M} \to \mathbb{R}$. It induces a norm for tangent vectors: $\|u\|_x = \sqrt{\langle u, u \rangle_x}$. A metric on $\mathcal{M}$ is a choice of inner product $\langle \cdot, \cdot \rangle_x$ for each $x \in \mathcal{M}$.*

Definition 2.9. *A metric $\langle \cdot, \cdot \rangle_x$ on $\mathcal{M}$ is a Riemannian metric if it varies smoothly with x, in the sense that for all smooth vector fields V, W on $\mathcal{M}$, the function $x \mapsto \langle V(x), W(x) \rangle_x$ is smooth from $\mathcal{M}$ to $\mathbb{R}$.*

Thus, we call a manifold Riemannian if it has a Riemannian metric. Then, the gradient of a smooth function on a Riemannian manifold $\mathcal{M}$ is defined as follows:

Definition 2.10. *Let $f : \mathcal{M} \to \mathbb{R}$ be a smooth function on a Riemannian manifold $\mathcal{M}$. The Riemannian gradient of f is the vector field ∇f on $\mathcal{M}$ uniquely defined by:*

$$\forall (x, v) \in T\mathcal{M},\ Df(x)[v] = \langle v, \nabla f(x) \rangle_x$$

Introducing the Hessian would prove less straight forward than the gradient. Recall the domain and codomain of the differential in its definition 2.7. As it turns out, there is no guarantee that applying this definition to determine the Hessian of a function would result in a tangent vector[4, p 81]. Thus, as the understanding of the Hessian is not strictly necessary for the remainder of the text we will omit the detailed definition.

2.3 Quotient Manifolds

A quotient set is a set together with an equivalence relation. Elements that are equivalent under the equivalence relation in the original set, are one and the same element in the quotient set.

Definition 2.11. *A binary relation $\sim$ on a set X is said to be an equivalence relation, if and only if it is reflective, symmetric, and transitive. That is, for all $a, b, c \in X$ we must have that:*

- $a \sim a$
- *If $a \sim b$ we must have $b \sim a$*
- *If $a \sim b$ and $b \sim c$ then $a \sim c$*

The set of all elements of a set that are equivalent to a given element a is called the equivalence class of a, denoted by $[a] = \{x \in X : x \sim a\}$. The the quotient set is the collection of all given equivalence classes, denoted by $X/\sim = \{[x] : x \in X\}$. The canonical projection $\pi : X \to X/\sim : x \mapsto \pi(x) = [x]$ is the mapping that sends the elements of a set to its quotient set.

Consider now the situation where the set we want to quotient is a manifold $\mathcal{M}$. In general, the resulting quotient, which we denote by $\overline{\mathcal{M}} = \mathcal{M}/\sim$ will not be a manifold. We thus wish to determine when $\overline{\mathcal{M}}$ will itself be a differential manifold, which we call a quotient manifold. We quantify this initially by the following definition.

Definition 2.12. *The quotient set $\overline{\mathcal{M}} = \mathcal{M}/\sim$ equipped with a smooth structure is a quotient manifold of $\mathcal{M}$ if the canonical projection projection π is smooth and its differential $\mathrm{D}\pi(x) : T_x\mathcal{M} \to T_{[x]}\overline{\mathcal{M}}$ has rank $\dim \overline{\mathcal{M}}$ for all $x \in \mathcal{M}$.*

This definition, however, has little practical use in determining if the quotient of a manifold is itself a manifold. This comes from the difficulties related to the canonical projection. There is no guarantee that a global definition of it exists, and even if it does, checking its smoothness might be difficult or impossible. We can however limit the the scope of quotients that we consider. That is, we only consider those in which the equivalence relation is induced by a group action of a Lie group. Then, the quotient manifold theorem 2.1 will prove to be enough for our applications. It however, requires introduction of a few new concepts, we list the definitions below:

Definition 2.13. *Let $\mathcal{G}$ be both a group and a manifold. If the product map*

$$\mathrm{prod} : \mathcal{G} \times \mathcal{G} \to \mathcal{G} : (g, h) \mapsto \mathrm{prod}(g, h) = gh$$

and the inverse map:

$$\mathrm{inv} : \mathcal{G} \to \mathcal{G} : g \mapsto \mathrm{inv}(g) = g^{-1}$$

are smooth, then $\mathcal{G}$ is a Lie group.

Definition 2.14. *Given a Lie group $\mathcal{G}$ and a manifold $\mathcal{M}$, a left group action is a map $\theta : \mathcal{G} \times \mathcal{M} \to \mathcal{M}$ such that:*

- *for all $x \in \mathcal{M}$, $\theta(e, x) = x$ (Identity)*
- *for all $g, h \in \mathcal{G}$ and $x \in \mathcal{M}$, $\theta(g \cdot h, x) = \theta(g, \theta(h, x))$ (compatibility)*

We say that the group action is smooth if θ is smooth as a map on the product manifold $\mathcal{G} \times \mathcal{M}$ to the manifold $\mathcal{M}$.

Definition 2.15. *A group action θ is free, if for all x, acting on x with any group element which is not the identity results in a point different from x. That is, if for all $x \in \mathcal{M}$, $\theta(g, x) = x \Rightarrow g = e$*

Definition 2.16. *A left group action θ is proper if:*

$$v : \mathcal{G} \times \mathcal{M} \to \mathcal{M} \times \mathcal{M} : (g, x) \mapsto v(g, x) = (\theta(g, x), x)$$

Is a proper map, that is, all compact subsets of $\mathcal{M} \times \mathcal{M}$ map to compact subsets of $\mathcal{G} \times \mathcal{M}$ through v^{-1}.

Definition 2.17. *The orbit of $x \in \mathcal{M}$ by the left action θ of $\mathcal{G}$ is the set $\mathcal{G}x = \{\theta(g, x) : g \in \mathcal{G}\}$. This induces an equivalence relation $\sim$ on $\mathcal{M}$:*

$$x \sim y \iff y = \theta(g, x) \text{ for some } g \in \mathcal{G}$$

Theorem 2.1 (Quotient Manifold Theorem)**.** *[10, p 255] Suppose $\mathcal{G}$ is a Lie group acting smoothly, freely, and properly on a smooth manifold $\mathcal{M}$. Then the orbit space $\mathcal{M}/\mathcal{G}$ is a topological manifold of dimension equal to* $\dim \mathcal{M} - \dim \mathcal{G}$*, and has a unique smooth structure with the property that the canonical projection $\pi : \mathcal{M} \to \mathcal{M}/\mathcal{G}$ is a smooth submersion.*

Note that a smooth map $F : M \to N$ is a smooth submersion if its differential is surjective, or equivalently if $\operatorname{rank} F = \dim N$ [10, p 78].

Having discussed when a quotient manifold is a manifold, we wish to determine the tangent vectors in the quotient manifold space. By the 2.12 the intuitive way to achieve this is by considering the differential of the canonical projection $\mathrm{D}\pi(x) : T_x\overline{\mathcal{M}} \to T_{[x]}\mathcal{M}$. The problem with this is however, that the mapping is not one to one. Therefore, we must restrict the domain to make it be one to one. This is achieved by the introduction of the vertical and horizontal spaces.

Definition 2.18. *For a quotient manifold $\overline{\mathcal{M}} = \mathcal{M}/\sim$, the vertical space at $x \in \mathcal{M}$ is the subspace*

$$V_x = \ker \mathrm{D}\pi(x)$$

If $\mathcal{M}$ is Riemannian, we call the orthogonal complement of V_x the horizontal space at x:

$$H_x = (V_x)^\perp = \{u \in T_x\mathcal{M} : \langle u, v \rangle_x = 0 \text{ for all } v \in V_x\}.$$

With this, it is easy to see that the map:

$$\mathrm{D}\pi(x)\big|_{H_x} : H_x \to T_{[x]}\mathcal{M} \tag{2.1}$$

is a bijection. The last useful concept related to quotient manifolds, is the horizontal lift. This mapping lets us consider the tangent vectors of the quotient manifolds, which is useful in some applications later on. The definition of the horizontal lift is:

Definition 2.19. *Consider a point $x \in \mathcal{M}$ and a tangent vector $\xi \in T_{[x]}\mathcal{M}$. The horizontal lift of ξ at x is the (unique) horizontal vector $u \in H_x$ such that $\mathrm{D}\pi(x)[u] = \xi$. We write:*

$$u = \left(\mathrm{D}\pi(x)\big|_{H_x}\right)^{-1}[\xi] = \mathrm{hor}_x(\xi)$$

It is important to note, that similarly to using the definition of the quotient manifold to determine if a set is a quotient manifold, using the definition of the horizontal lift to find an expression for it, might be difficult. However, the specifics of determining it depend strongly on the manifold being discussed, thus we delay the further discussion of the lift until section 4.2.

2.4 Distance on a manifold

We wish to consider the mean on a manifold. For this reason we will require defining what the distance function is on a manifold space, and when a manifold is a metric space. As determining those concepts on a manifold is relatively straightforward, we mainly recount chapter 10.1 from [4]. We recount the basic definition of a metric space.

Definition 2.20. *[4, p. 253] A distance on a set $\mathcal{M}$ is a function* $\text{dist} : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ *such that, for all $x, y, z \in \mathcal{M}$,*

1. $\text{dist}(x, y) = \text{dist}(y, x)$;
2. $\text{dist} \geq 0$ *and* $\text{dist}(x, y) = 0$ *if and only if* $x = y$;
3. $\text{dist}(x, z) \leq \text{dist}(x, y) + \text{dist}(y, z)$.

Equipped with a distance, $\mathcal{M}$ is a metric space.

In the Euclidean case the most basic distance that might come to mind, could be the length of a straight line connecting two points in the space. However, in a manifold, we are yet to define an equivalnet notion to a straight line. As a preliminary to this we need to define a curve segment on a manifold.

Definition 2.21. *[4, p. 253] A curve segment on a manifold $\mathcal{M}$ is a continuos map $c : [a, b] \to \mathcal{M}$, where $a \leq b$ are real. A curve segment $c : [a, b] \to \mathcal{M}$ is:*

1. *smooth if c can be extended to a smooth map $\tilde{c} : I \to \mathcal{M}$ on a neighbourhood I of $[a, b]$, in which case $c'(a)$ and $c'(b)$ denote $\tilde{c}'(a)$ and $\tilde{c}'(b)$, respectively;*
2. *regular if it is smooth and $c'(t) \neq 0$ for all $t \in [a, b]$*
3. *piecewise smooth (resp., piecewise regular) if there exists a finite set of times $a = t_0 < t_1 < \cdots < t_{k-1} < t_k = b$ such that the restriction $c|_{[t_{i-1}, t-i]}$ are smooth (resp., regular) curve segments for $i = 1, \ldots, k$.*

In particular, piecewise regular curves are piecewise smooth. We say a curve segment $c : [a, b] \to \mathcal{M}$ connects x to y if $c(a) = x$ and $c(b) = y$.

Definition 2.22. *Let $\mathcal{M}$ be a Riemannian Manifold. Let $c : [a, b] \to \mathcal{M}$ be a piecewise smooth curve, we define the Riemannian distance by:*

$$\text{dist}(x, y) = \inf_c \int_a^b \|c'(t)\|_{c(t)} dt$$

The following theorem lets us determine when the manifold equipped with the Riemannian distance is a metric space. For atlas topology, see [10].

Theorem 2.2. *[4, p. 254] If $\mathcal{M}$ is connected (meaning each pair of points is connected by a curve segment), Riemannian distance defined in definition 2.22 defines a distance. Equipped with this distance, $\mathcal{M}$ is a metric space whose metric topology coincides with its atlas topology.*

If the minimum in 2.22 is attained we call c a minimizing curve. We can then note the following theorem.

Theorem 2.3. *Every minimizing curve admits a constant-speed parametrization such that it is a geodesic, called a minimizing geodesic.*

Until now we have omitted the important notion of a geodesic. Informally, we understand a geodesic as a generalization of the notion of a straight line connecting two points. The formal definition and discussion of a geodesic is not necessary for the content of this text. We may understand a geodesic in the informal sense as an acceleration free curve, and by Theorem 2.3, as the shortest curve connecting two points. For more details, see for example, [9], [4], or [1].

3 Optimization on manifolds

Conceptually, the general idea does not change between optimization on a Euclidean space, and a manifold. We are still discussing the problems of the form:

$$\min_{x \in \mathcal{M}} f(x) \tag{3.1}$$

In which we call $f(x)$ the objective function. We are attempting to find a global solution, also called a global minimizer, that is, a point $x \in \mathcal{M}$ such that $f(x) \leq f(y)$ for all $y \in \mathcal{M}$. Of course, such a problem might be difficult if not impossible to solve. In which case we would instead look for a local minimizer. That is, a point $x \in \mathcal{M}$ such that $f(x) \leq f(y)$ for all y in a neighbourhood of $\mathcal{M}$.

Because optimization algorithm define sequences of points on a manifold $\mathcal{M}$, we must define what convergence is. We use the following definition.

Definition 3.1. *[4, p 52] Consider a sequence S of points $x_0, x_1, x_2, \ldots$ on a manifold $\mathcal{M}$. Then:*

1. *A point $x \in \mathcal{M}$ is a limit of S if, for every neighborhood U of x in $\mathcal{M}$, there exists an integer K such that $x_K, x_{K+1}, x_{K+2}, \ldots$ are in U. If x is the limit, we write $\lim_{k\to\infty} x_k = x$ or $x_k \to x$ and we say the sequence converges to x.*

2. *A point $x \in \mathcal{M}$ is an accumulation point of S if it is the limit of a subsequence of S, that is, if every neighborhood U of x in $\mathcal{M}$ contains an infinite number of elements of S.*

Further, we provide a definition of a first order necessary condition for a smooth manifold, which can be used to identify a minimum.

Definition 3.2. *[4, p 53] A point $x \in \mathcal{M}$ is critical (or stationary) for a smooth function $f : \mathcal{M} \to \mathbb{R}$ if*

$$(f \circ c)'(0) \geq 0$$

for all smooth curves c on $\mathcal{M}$ such that $c(0) = x$.

Additionally, if the manifold is Riemannian, we obtain a first order necessary condition that is equivalent to the functions in Euclidean optimization, that is:

Theorem 3.1. *[4, p 54] Let $f : \mathcal{M} \to \mathbb{R}$ be smooth on a Riemannian manifold $\mathcal{M}$. Then, x is a critical point of f if and only if* $\operatorname{grad} f(x) = 0$.

Arguably, the most common class of iterative methods for solving minimization problems are line search methods, where the basic idea is to choose a direction and determine an optimal step size along this direction, in order to reduce the objective function. Of those, the gradient descent is one of the simplest variants. In the Euclidean case, the iterates are given by $x_{k+1} = x_k + \alpha_k \operatorname{grad} f(x_k)$ with x_k, α_k, $\operatorname{grad} f(x_k)$ denoting respectively, the k-th iteration, step-size, and the gradient of the objective function. We wish to generalize this idea to work on a manifold.

3.1 Moving on a manifold

In Euclidean space, we increment by addition with the current iterate to produce the following. As vectors are not defined in the usual sense on a manifold, the situation becomes more complex in this scenario. One might assume that we might iterate along a tangent space, moving along the tangent vector's direction. However, doing just that is not a possibility, as even an infinitesimal movement in a tangent vectors direction will result in a point that is not on the manifold [4, p. 33]. This issue is resolved by the means of a retraction mapping, which allows the movement in a tangent vectors direction without leaving the manifolds structure.

Definition 3.3. *[4, p. 39] A retraction on a manifold $\mathcal{M}$ is a smooth map*

$$R : T\mathcal{M} \to \mathcal{M} : (x, v) \mapsto R_x(v)$$

such that for all $(x, v) \in T\mathcal{M}$, we have

1. $R_v(0) = x$,
2. $DR_v(0)\colon T_x\mathcal{M} \to T_x\mathcal{M}$ *is the identity map:* $DR_v(0)[v] = v$.

This lets us define the simplest form of a Riemannian gradient descent [4, p 55], where the $k+1$ iterate is given by:

$$x_{k+1} = R_{x_k}(-\alpha_k \operatorname{grad} f(x_k)) \tag{3.2}$$

Note that since retractions are not unique, one might naturally try to find which retraction might be "better" than others. We may note one "special" type of retraction that is the exponential mapping. The reasons for the expoential mapping being important depend on theory not introduced in this text, in large part the notion of geodesics. For this reason, for a deeper insight in why the exponential mapping is essential, see, for example, [9], [4], or [1]. We still define it. First consider a maximal geodesic [4, p. 256]. Here, maximal means that the interval I is as large as possible.

$$\gamma_v : I \to \mathcal{M}, \quad \text{with} \quad \gamma_v(0) = x \quad \text{and} \quad \gamma_v'(0) = v$$

With this we define the exponential map as:

Definition 3.4. *[4, p. 256]*
Consider the following subset of the tangent bundle:

$$\mathcal{O} = \{(x, v) \in T\mathcal{M} : \gamma_v \text{ is defined on an interval containing } [0,1]\}$$

The exponential map $\exp : \mathcal{O} \to \mathcal{M}$ *is defined by*

$$\exp(x, v) = \exp_x(v) = \gamma_v(1)$$

The restriction $\exp_x$ *is defined on* $\mathcal{O}_x = \{v \in T_x\mathcal{M}(x, v) \in \mathcal{O}\}$

3.2 Step-size Selection

The step-size selection does not vary conceptually between optimization on an Euclidean space and a manifold. We have the three basic options for step-size α_k selection [4, p. 55]:

1. Fixed step-size: $\alpha_k = \alpha$ for all k.
2. Optimal step-size: α_k minimizes $g(t) = f(R_{x_k}(-t \operatorname{grad} f(x_k)))$ exactly.
3. Backtracking: starting with a guess $t_0 > 0$, iteratively reduce it by a factor as $t_i = \tau t_{i-1}$ with $\tau \in (0,1)$ until t_i is deemed acceptable, and set $\alpha_k = t_i$.

We discuss the third case, backtracking line search, further. The line-search is ran until the Armijo-Goldstand condition is satisfied. With some constant r (often 10^{-4}), the condition is:

$$f(x) - f(R_x(-\alpha \operatorname{grad} f(x))) \geq \alpha \|\operatorname{grad} f(x)\|^2 \tag{3.3}$$

With which, we can state the Backtracking line-line search algorithm.[4, p 59]

Algorithm 1 Backtracking line-search

Require: $\tau, r \in (0,1)$; for example, $\tau = \frac{1}{2}$ and $r = 10^{-4}$
Require: $x \in \mathcal{M}, \bar{\alpha} > 0$
  Set $\alpha \leftarrow \bar{\alpha}$
  while $f(x) - f(R_x(-\alpha \operatorname{grad} f(x))) < r\alpha \|\operatorname{grad} f(x)\|^2$ **do**
    Set $\alpha \leftarrow \tau\alpha$
  end while
  Output: α

As it turns out, the line search given by the previous algorithm, might in specific conditions guarantee convergence of the gradient descent algorithm. Before stating the theorem guaranteeing this we state two required conditions. The conditions and following theorem given by [4]

Condition 1: There exists $f_{\text{low}} \in \mathbb{R}$ such that $f(x) \geq f_{\text{low}}$ for all $x \in \mathcal{M}$

Condition 2: For a given subset S f the tangent bundle $T\mathcal{M}$, there exists a constant $L > 0$, such that, for all $(x, s) \in S$

$$f(R_x(s)) \leq f(x) + \langle \operatorname{grad} f(x), s \rangle + \frac{L}{2}\|s\|^2 \tag{3.4}$$

Theorem 3.2. *[4, p 61] Let f be a smooth function satisfying Condition 1 on a Riemannian manifold $\mathcal{M}$. For a retraction R, let $f \circ R$ satisfy Condition 2 on a set $S \subseteq T\mathcal{M}$ with constant L. Let $x_0, x_1, x_2, \ldots$ be the iterates generated by Riemannian gradient descent (equation 3.2) with backtracking line-search (Algorithm 1) using fixed parameters $\tau, r \in (0,1)$ and initial step-sizes $\bar{\alpha}_0, \bar{\alpha}_1, \bar{\alpha}_2, \ldots$ If for every k the set $\{(x_k, -\alpha \operatorname{grad} f(x_k)) : \alpha \in [0, \bar{\alpha}_k]\}$ is in S and if $\liminf_{k\to\infty} \bar{\alpha}_k > 0$, then*

$$\lim_{k\to\infty} \|\operatorname{grad} f(x_k)\| = 0.$$

Furthermore, for all $K \geq 1$, there exists k in $0, \ldots, K-1$ such that

$$\|\operatorname{grad} f(x_k)\| \leq \sqrt{\frac{f(x_0) - f_{low}}{r \min(\bar{\alpha}_0, \ldots, \bar{\alpha}_{K-1}, \frac{2\tau(1-r)}{L})} \frac{1}{\sqrt{K}}}.$$

4 Kendall's shape space

Kendall's shape space introduces a way to mathematically define what a shape is. Each object in the shape space consists of k points in $\mathbb{R}^m$, that are centred and normalized. We describe the basic structure of the shape space as defined in [5]. Consider $X \in \mathbb{R}^{m\times k}$, and denote the ith column by x_i^*. Recall that a centroid of a finite set of k points in $\mathbb{R}^n$ is $x_c = (\sum_{i=1}^k x_i)/k$. We remove the effects of translation by mapping each x_i^* to $x_i^* - x_c^*$ where x_c^* is the centroid of the k points. This results in an $m \times (k-1)$ matrix. We then remove the effects of scaling by requiring that natural quadratic measure is equal to unity; $\sqrt{\|x_1^* - x_c^*\|^2 + \cdots + \|x_k^* - x_c^*\|^2} = 1$. Applying those two effects on the original matrix produces the pre-shape space. We note that this is equivalent to $S^{m(k-1)-1}(1)$; the unit sphere of dimension $m(k-1)-1$, which is a known embedded submanifold of $\mathbb{R}^{m(k-1)}$. We denote this pre-shape-space sphere by $\mathcal{S}_m^k$.

4.1 The pre-shape sphere

Having introduced the pre-shape as a sphere, we turn to defining the appropriate spaces and objects on it. Consider an arbitrary unit sphere $S^{d-1} = \{x \in \mathbb{R}^d : \|x\| = 1\}$. We obtain the tangent space of S^{d-1} by considering its local defining function $h(x) = \|x\| - 1 = x^T x - 1$. The differential and the kernel of the differential are respectively $Dh(x) = 2x$ and $\operatorname{rank} Dh(x) = 1$. From which we see that the tangent space of S^{d-1} at x is:

$$T_x S^{d-1} = \{v \in \mathbb{R}^d : x^T v = 0\} \tag{4.1}$$

The sphere becomes a Riemannian manifold by being equipped with the standard metric, $\langle u, v\rangle_u = u^T v$. With this can we consider the exponential mapping on the sphere. As the derivation of the expression requires some theory that we have omitted, we state the exponential mapping as defined in [4, Example 5.37], see the same example for the details of the derivation.

$$\exp_x(v) = \cos(\|v\|)x + \frac{\sin(\|v\|)}{\|v\|}v \tag{4.2}$$

Note that this is smooth over the entire tangent bundle, with the extension $\frac{\sin(t)}{t} = 1$ at $t = 0$. Then for the inverse of the exponential map we can follow the derivation from

[4, Example 10.21]. Let $y = \exp_x(v)$, and note that $x^T x = 1, x^T v = 0$, this gives that $x^T y = \cos(\|v\|)$. Then define a new vector u as:

$$u = y - (x^T y)x = \mathrm{Proj}_x(y) = \frac{\sin(\|v\|)}{\|v\|}v$$

Normalize u and restrict its domain to tangent vectors v whose norm is less than π, this gives that $\frac{\sin(\|v\|)}{\|u\|} = 1$. Lastly, we see that $x^T y = \cos(\|v\|)$ has the unique solution $\|v\| = \arccos(x^T y)$, where $\arccos : [-1, 1] \to [0, \pi]$ is the principle inverse of cos. In total this results in the inverse of the exponential map being:

$$\exp_x^{-1}(y) = \arccos(x^T y) \frac{y - (x^T y)x}{\|y - (x^T y)x\|} \tag{4.3}$$

Note also that the distance on the sphere is defined by:

$$\mathrm{dist}(x, y) = \arccos(x^T y) \tag{4.4}$$

4.2 Quotient of the pre-shape

We say that two pre-shapes have the same shape if one can be rotated into the other. Thus, the pre-shape space under the constraint of two elements being equal if they can be rotated into each other results in the total shape space. The shape space is then denoted by Σ_m^k. Formally, the idea of two pre-shapes being equal is expressed in terms of the equivalency induced by the special orthogonal group $\mathrm{SO}(m)$. This group consists of all $m \times m$ orthogonal matrices with determinant equal to 1:

$$\mathrm{SO}(m) = \{Q \in \mathrm{GL}(m, \mathbb{K}) \mid Q^T Q = QQ^T = I, \det(Q) = 1\}.$$

Thus, for $x, y \in S_m^k$, we say that $x \sim y$ if there exists some $Q \in \mathrm{SO}(m)$ such that $Qx = y$. Furthermore, this allows the expression of the shape space as the quotient of the pre-shape sphere by the left group action of $\mathrm{SO}(m)$:

$$\Sigma_m^k \triangleq S_m^k / \mathrm{SO}(m)$$

We wish to invoke theorem 2.1 for the shape space to be a quotient manifold. $\mathrm{SO}(m)$ is the open subgroup of $\mathrm{O}(m)$, and is therefore also an embedded Lie subgroup of dimension $\frac{m(m-1)}{2}$ in $\mathrm{GL}(m, \mathbb{R})$. It is a compact group because it is a closed subset of $\mathrm{O}(m)$ [10, p 167]. It acts smoothly because it is given by a matrix multiplication. Thus, the group action is proper [4, p 214]. However, determining if the group action is free, requires a more detailed examination.

We can first note that there are two known cases for which the group action will be free everywhere. This stems from the fact that we can identify the shape space Σ_1^k with $S^{k-2}(1)$ for $k \geq 2$, and the shape space Σ_2^k with $\mathbb{CP}^{k-2}(4)$ for $k \geq 3$. Aside from $k = 2$ the group action must be free since those manifolds are smooth[8].

As shown in [8], for $m \geq 3$, singularities (points where the group action fails to be free) will always exist. Furthermore, such a singularity will always be of the form

$$R_m^T \,\mathrm{diag}(I_{m-2} R_2) R_m$$

where $R_2 (\neq I_2) \in SO(2)$, $R_m \in SO(m)$. Thus, barring the case where $k = 2$, and the singularities mentioned above, the group action will be free, and the shape space will be a quotient manifold.

Remark 4.1. *The dimension of the shape space Σ_m^k is* $\dim \Sigma_m^k = \frac{-m(m-2k+1)}{2} - 1$

This is a direct consequence of the Theorem 2.1and the dimension of $\mathcal{S}_m^k$ and $SO(m)$.

The main advantage the Theorem 2.1 provides, beyond assuring that the shape space is a quotient manifold, is the assurance that the canonical projection will be a Riemannian submersion. This will be essential in determining the exponential map and its inverse on the shape space. However, as a further preliminary to this, we need to determine what the horizontal and vertical spaces are on the pre shape space. We define the vertical and horizontal spaces to the shape space, as discussed in [5, p 109]:

$$\mathrm{Ver}_x = \{Ax | A \in \mathrm{Skew}(m)\}$$

Here, $x \in \mathcal{S}_m^k$ and Skew(m) is the space of skew-symmetric matrices of size m. Then the horizontal space at x is:

$$\begin{aligned} \mathrm{Hor}_x &= \{w \in T_x S_m^k \mid \mathrm{Tr}(Axw^T) = 0 \, \forall A \in \mathrm{Skew}(m)\} \\ &= \{w \in T_x S_m^k \mid xw^T \in \mathrm{Sym}(m)\} \end{aligned}$$

We may then use the result provided by [11], to determine the vertical, and thus horizontal component of any tangent vectors.

Lemma 4.1. *Fix $x \in \mathcal{S}_m^k$ and $w \in T_x\mathcal{S}_m^k$. Let ver$_x$, resp. hor$_x$ denote the restriction of vertical resp. horizontal projection to $T_x\mathcal{S}_m^k$.*

(a) ver$_x(w) = Ax$ if and only if A solves the Sylvester equation

$$Axx^t + xx^tA = wx^t - xw^t.$$

Moreover, the above equation has a unique skew-symmetric solution if rank$(x) \geq m - 1$.

With this, the horizontal component is given simply by:

$$\mathrm{hor}_x(w) = w - \mathrm{ver}_x(w) \tag{4.5}$$

With this can we use an important theorem given by [13].

Theorem 4.1. *[13] Let $\pi : \mathcal{M} \to B$ be a Riemannian submersion. If γ is a geodesic in $\mathcal{M}$ such that $\dot{\gamma}(0)$ is a horizontal vector, then $\dot{\gamma}$ is horizontal everywhere and $\pi \circ \gamma$ is a geodesic of B of the same length as γ.*

The last concept required to define the desired mappings on the shape space, is the notion of the optimal alignment. The optimal rotation R between any pre-shape x, y is unique in a subset U of $\mathcal{S}_m^k \times \mathcal{S}_m^k$. This allows the definition of the align map $\omega : U \to \mathcal{S}_m^k$. This notion comes from [12].

To summarize, we obtain the following mappings for the shape space, where the exponential mapping and its inverse are a direct consequence of Theorem 4.1. For any $x, y \in \mathcal{S}_m^k$ and $v \in T_x\mathcal{S}_m^k$, as stated in [12]:

$$\exp_{\Sigma,[x]}(d_x\pi v) = \pi\left(\exp_x\left(\mathrm{hor}_x(v)\right)\right), \tag{4.6}$$

$$\exp_{\Sigma,[x]}^{-1}([y]) = d_x\pi \exp_x^{-1}\left(\omega(x, y)\right), \tag{4.7}$$

$$d_\Sigma\left([x], [y]\right) = d(x, \omega(x, y)). \tag{4.8}$$

5 Riemannian centre of mass

In this section we introduce the theory behind the Riemannian centre of mass. Prior to considering the properties of the Riemannian L^2 center of mass we have to define the function itself.

Definition 5.1. *[3]*
The (global) Riemannian L^2 center of mass or mean of the data set $\{x_i\}_{i=1}^N \subset \mathcal{M}$ with respect to weights $0 \leq w_i \leq 1$ $\left(\sum_{i=1}^N w_i = 1\right)$ is defined as the minimizer(s) of

$$f(x) = \frac{1}{2}\sum_{i=1}^{N} w_i d^2(x, x_i)$$

in $\mathcal{M}$. We denote the center by $\bar{x}_p$.

Note that often, the weights are assumed to be uniformly distributed, in which case we obtain the following expression for the Riemannian mean. Unless stated otherwise, this is the expression that we will consdier for the remainder of the text.

$$f(x) = \frac{1}{2n}\sum_{i=1}^{N} d^2(x, x_i) \tag{5.1}$$

5.1 Convexity, existence, and uniqueness

When considering an optimization problem, the first question that arises might be about the convexity of the set and the function. The definitions of convex functions and sets are conceptually similar to those in the Euclidean case. The definitons for a convex function and set are respectively:

Definition 5.2. *[3] Let A be an open subset of $\mathcal{M}$ such that every two points in A can be connected by at least one geodesic of $\mathcal{M}$ such that this geodesic lies entirely in A. Assume that $f : A \to \mathbb{R}$ is a continuous function. Then f is called (strictly) convex if the composition $f \circ \gamma : [0, 1] \to \mathbb{R}$ is (strictly) convex for any geodesic $\gamma : [0, 1] \to A$. We say that f is globally (strictly) convex if it is (strictly) convex in $\mathcal{M}$.*

Definition 5.3. *[3] A set $A \subset \mathcal{M}$ is called strongly convex if any two points in A can be connected by a unique minimizing geodesic in $\mathcal{M}$ and the geodesic segment lies entirely in A.*

The question of the global convexity of 5.1 is answered by the result proved by [16]; that is, the only globally convex function on a compact Riemannian manifold is the constant function. Thus, as long as we are dealing with a compact manifold, there is no hope of having 5.1 be convex. We therefore turn to determining a subset of the manifold that is strongly convex.

Consider an arbitrary open ball on the manifold. Let $o \in \mathcal{M}$ be some point on the manifold, and r be the radius; then the open ball is denoted by $B(o, r)$. We wish to determine when this ball is strongly convex. The theorem by [15] provides the necessary condition. It however, relies on two concepts yet to be introduced in this text: the injectivity radius, and the upper bound of the sectional curvature of the manifold.

Note that a diffeomorphism is a bijective map $F : U \to V$, where U, V are open sets and such that both F and F^{-1} are smooth [4, p 26].

Definition 5.4. *[4, p 257] The injectivity radius of a Riemannian manifold $\mathcal{M}$ at a point x, denoted by $\text{inj}(x)$, is the supremum over radii $r > 0$ such that $\exp_x$ is defined and is a diffeomorphism on the open ball*

$$B(x, r) = \{v \in T_x\mathcal{M} : \|v\|_x < r\}.$$

The notion of curvature is too great to be defined in the scope of this text. Thus, we will omit the precise definition. Instead we only state strictly informally that it is, like the name suggests, in a measure of how much a surface bends in different directions. For the purpose of this text we will assume that the upper bound of curvature is denoted by Δ and the lower bound by δ.

We can then state the theorem by [15]:

Theorem 5.1. *[15]*
Suppose r satisfies

1. $r \leq \frac{1}{2} inj(x),\ x \in B(o, r)$,
2. $r \leq \frac{1}{2}\frac{\pi}{\sqrt{\Delta}}$,

Then $r(x) = d(x, o)$ is smooth and convex on $B(o, r)$, and any two points in $B(o, r)$ are joined by a unique segment that lies in $B(o, r)$.

The last statement of the theorem is equivalent to the definition of strict convexity. Thus, we obtain the constraint that if

$$\rho < r_{cx} \triangleq \frac{1}{2} \min\left\{\text{inj}M, \frac{\pi}{\sqrt{\Delta}}\right\}, \tag{5.2}$$

then the open ball $B(o, \rho)$ is strictly convex. For the remainder of the text we will refer to r_{cx} as convexity radius.

Having defined a strictly convex subset of the manifold, we can assure the existence and uniqueness of the solution to 5.1, provided that all data points lie within this convex ball. This fact comes from the following theorem, proven in [2]:

Theorem 5.2. *Consider $\{x_i\}_{i=1}^N \subset B(o, \rho)$ and assume $0 \leq w_i \leq 1$ with $\sum_{i=1}^N w_i = 1$. If $\rho \leq r_{cx}$, then the Riemannian L^2 center of mass $\bar{x}$ is unique, is inside $B(o, \rho)$, and is the unique zero of the gradient vector field ∇f in $\bar{B}(o, \rho)$.*

5.2 Riemannian center of mass for Kendall's shape space

As a preliminary to the discussion of the numerical solutions for 5.1, we begin by stating the actual gradient of the function. By [3], the gradient of 5.1 is:

$$\text{grad}\, f(x) = -\frac{1}{2n}\sum_{i=1}^{N} \exp_x^{-1} x_i \tag{5.3}$$

Having now derived an area of the manifold in which the solution is guaranteed to exist, we may consider the discussion about actually determining it. In the previous section, the only assumption that we made was for the manifold to be complete, and Riemannian. As it turns out, determining conditions under which an algorithm can be constructed such that the centre of mass can actually be determined, requires a further restriction on the radius of the ball in which the points must reside.

The possible conditions one can achieve vary greatly between the type of manifold being considered, see [3]. Therefore, we restrict our discussion to the Kendall's shape space. Even then, considering all the shape spaces might prove difficult. The reason for this lies first and foremost, in the sectional curvature of the shape spaces in specific dimensions, that is m. As previously mentioned, Σ_1^k and Σ_2^k can be identified with some familiar spaces for $k \geq 3$. For an arbitrary shape space Σ_m^k, the curvature might be unbounded [5, p 207], making the r_{cx} infinitesimal, in which case the center of mass might not exist. Determining the curvature for an arbitrary space might be possible, as discussed in [8]. However, this would require understanding of the curvature beyond the content of this thesis. For this reason, we will limit ourselves to the discussion concerning Σ_2^k, which is also the shape space with a more significantly developed theory (see for example [5], or [7]).

We begin by noting that the space is a compact and connected, and that the sectional curvature is bounded below by 1, and above by 4, so it is not constant. Furthermore, we note that because of the domain we choose for the exponential map, the injectivity radius is $\pi/2$ [7]. From this we obtain the convexity radius $r_{cx} = \pi/4$.

The theory behind which convexity radius guarantees convergence has been developed by [3]. As it is based on an estimate of the Hessian, we repeat the the required expression for the estimate $c_\kappa(l)$, given by [3].

$$c_\kappa(l) = \begin{cases} 1, & \kappa \geq 0, \\ \sqrt{|\kappa|}l \coth(\sqrt{|\kappa|}l), & \kappa < 0. \end{cases} \tag{5.4}$$

We additionally state the basic gradient descent algorithm, as it is required for the formulation of the required theorem.

Algorithm 2 Gradient descent for finding the Riemannian L^2 center of mass.

[3]

Require: $\{x_i\}_{i=1}^N \subset B(o, \rho) \subset \mathcal{M}$ and weights $\{w_i\}_{i=1}^N$ $(0 \leq w_i \leq 1, \sum_{i=1}^N w_i = 1)$. Choose $x^0 \in \mathcal{M}$.

1: **if** $\operatorname{grad} f_p(x^k) = 0$ **then** stop; **else** set

$$x^{k+1} = \exp_{x^k}(-t_k \operatorname{grad} f_p(x^k)),$$

where $t_k > 0$ is an "appropriate" step-size and $\operatorname{grad} f_p(\cdot)$ is defined in 5.1.

2: goto step 2.

Theorem 5.3. *[3] Assume that $\bar{x}_2$ is the L^2 center of mass of $\{x_i\}_{i=1}^N \subset B(o, \rho) \subset \mathcal{M}$, where $\rho \leq \frac{1}{3} r_{cx}$. Define $t_{\delta,\rho} = \frac{1}{H_{B(o,3\rho)}}$, where $H_{B(o,3\rho)} = c_\delta(4\rho)$ and c_κ is defined in 5.4. In Algorithm 2 assume that $x^0 \in B(o, \rho)$ and for every $k \geq 0$ choose $t_k = t$, where $t \in (0, 2t_{\delta,\rho})$. Then we have the following: The algorithm is well defined for all $k \geq 0$, and each iterate of the algorithm continuously stays in $B(o, 3\rho)$, $f_2(x^{k+1}) \leq f_2(x^k)$ for $k \geq 0$ (with equality only if $x^k = \bar{x}_2$), and $x^k \to \bar{x}_2$ as $k \to \infty$. Moreover, if x^0 coincides with o, then $\rho \leq \frac{1}{2} r_{cx}$ is enough to guarantee the convergence, in which case each iterate of the algorithm continuously stays in $B(o, 2\rho)$ and we can take $t_{\delta,\rho} = \frac{1}{H_{B(o,2\rho)}}$, where $H_{B(o,2\rho)} = c_\delta(3\rho)$.*

We may additionally note the following theorem, as it might apply to some specific shape spaces.

Theorem 5.4. *[3] Assume that $\mathcal{M}$ is either a manifold of constant nonnegative curvature $\Delta \geq 0$ or a 2-dimensional manifold with nonnegative curvature upper bounded by $\Delta \geq 0$. Let $p = 2$ and $\{x_i\}_{i=1}^N \subset B(o, \rho)$, where $\rho \leq r_{cx}$. In Algorithm 2, choose an initial point $x^0 \in B(o, \rho)$ and a constant step-size $t_k = t$, where $t \in (0, 1]$. Then we have that the algorithm is well defined for every $k \geq 0$, each iterate continuously stays in $B(o, \rho)$, $f_2(x^{k+1}) \leq f_2(x^k)$ with equality only if $x_k = \bar{x}_2$, and $x^k \to \bar{x}_2$ as $k \to \infty$.*

6 Numerical Experiments

In this section we present a series of numerical experiments. The purpose of this is to illustrate the efficiency and reliability of the previously discussed theory. We firstly consider the space of triangles in a two dimensional space, that is Σ_2^3. Then we will consider some specific data that outlines a set of chess figures. There, by the nature of the how we sample our data, it is possible to investigate how different sampling frequencies might affect the result.

6.1 The Space of Triangles

We consider the space of space of shapes of three points in two dimensions, Σ_2^3, which clearly is the space of triangles. Recall that by remark 4.1, we have that $\dim \Sigma_2^3 = 2$. Because of this, and the previously mentioned values for the curvature of Σ_2^k being non-negative, it is clear that theorem 5.4 applies in this case.

We use the space of triangles to test with a large number of data points, how different values of step size might affect convergence. Random shapes of triangles were used for all of the tests. Note the simple fact that if every two elements in a set have a distance of less than $2r$ apart, then they must all lie inside the open ball of radius r. As such we constructed a set of random triangles that are all inside of the convexity radius r_{cx}, by generating random 2×3 matrices, that are centered and normalized, and rejecting those that were more than $2r_{cx}$ apart from every other triangle in the set.

The data used for the numerical experiment was a random sample of 4 shapes inside of the convexity radius of $\pi/4$. The test was performed 50 times, after which the average of the data was taken. By the Theorem 5.4 the possible constant step sizes that guarantee convergence were the values between 0 and 1. However, we perform the test on the step-sizes up to 4. The reason for this is to obtain some insight into what happens to the center of mass beyond the convexity radius that assures convergence. As the small step-sizes rapidly approach a very high number of iterations, we limit the maximum number of iterations to a 1000. The results are presented in 1. The red line depicts the rate of convergence for the gradient descent with backtracking line search, as mentioned in Section 3.2.

There are two main things that can be observed by the Figure 1. We see that if we only consider the step-sizes that do guarantee convergence, then 1 would be the optimal choice, in terms of rate of convergence. However, one can see from the graph that the algorithm converges relatively quickly, for values up to about a step-size of 2. We can note that this is the maximal step-size range given by 5.3, but only if the convexity radius were to be $\pi/12$. For this reason we cannot guarantee that the quick convergence means that a global solution is being reached, but it could be an indication of that happening. This could indicate that at least in this specific case, the restriction given by the spread of the data-points, is not as strict in practice as it is in theory. The random nature of our data prevents us from checking if the same center is reached with a step-size of 1 and 2.

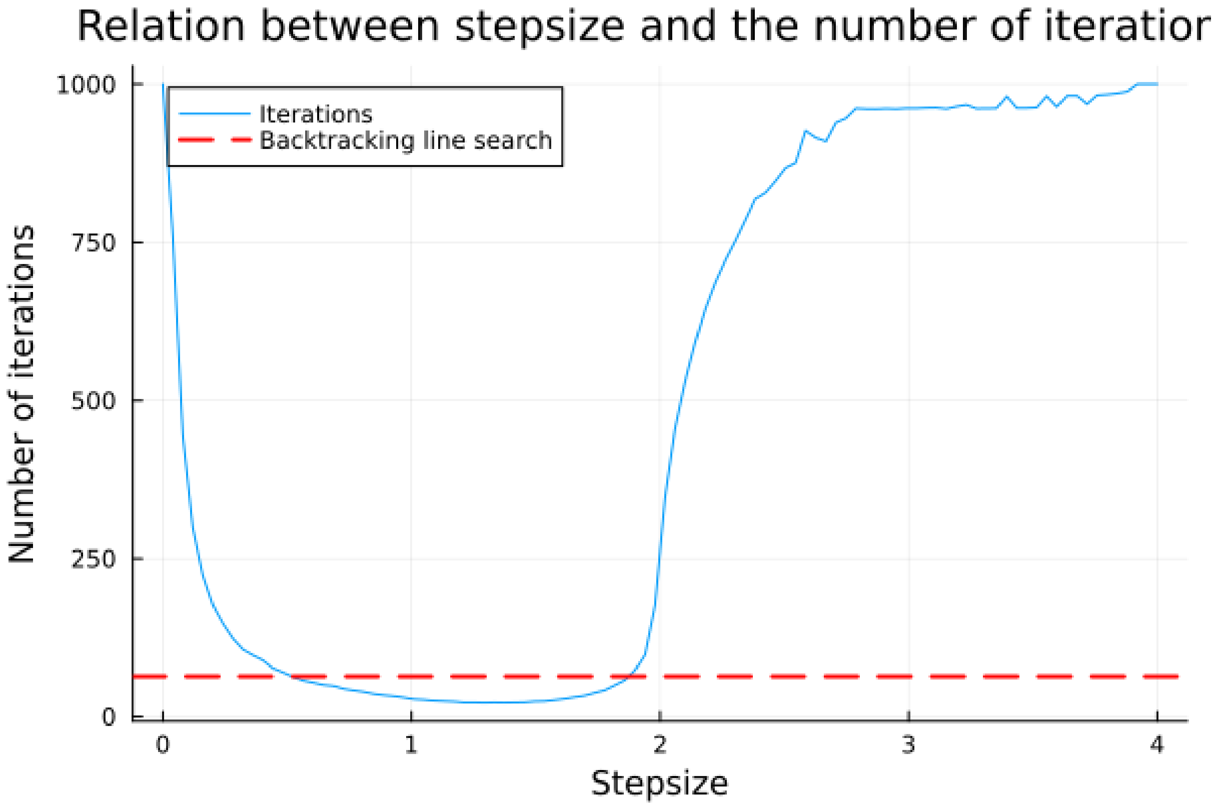
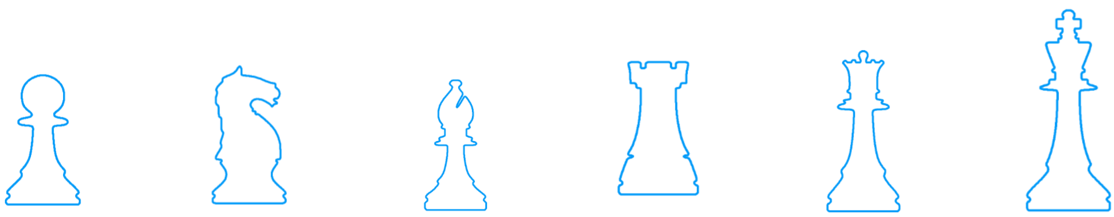


Figure 1: Relation between stepsize and the number of iterations

6.2 Chess Pieces



(a) Pawn (b) Knight (c) Bishop (d) Rook (e) Queen (f) King

Figure 2: Shapes outlining the chess figures

We now consider a data set consisting of the outlines of the six chess pieces[1]. When sampled with a thousand datapoints ($k = 1000$), the outlines can be seen in Figure 2. Note that despite the appearance of some figures being larger than others, mathematically,

[1]The figures were sampled by the Bézier curves provided in this Github repository https://github.com/estennw/srvpy

all of the figures have been scaled and normalized to be of equal size. We wish to apply Theorem 5.3 to the set of shapes. For this, we have two possibilities of a convexity radius guaranteeing the convergence. The case when the convex ball is centered at any point on the manifold, in which we obtain the convexity radius of $\pi/12$, or the case where the center of the convex ball is the initial iterate, in which the radius would be $\pi/8$. The small number of shapes allows us to manually check whether the figures are in an "appropriate" distance from each other. For this purpose we note that $\frac{\pi}{12} \approx 0.2618$ and $\frac{\pi}{6} \approx 0.5236$. Thus, the uniqueness of the center of mass, is guaranteed, if the shapes are less than 0.5236 apart. Table 1 shows the numerical distances between the six chess figures, rounded to four decimal places.

Table 1: Distances Between Chess Figures

	Pawn	Knight	Bishop	Rook	Queen	King
Pawn	0	0.2600	0.8310	0.9499	0.1377	0.1819
Knight	0.2600	0	0.8843	0.9950	0.2813	0.3016
Bishop	0.8310	0.8843	0	0.2125	0.7260	0.6944
Rook	0.9499	0.9950	0.2125	0	0.8378	0.8039
Queen	0.1377	0.2813	0.7260	0.8378	0	0.0858
King	0.1819	0.3016	0.6944	0.8039	0.0858	0

From the table it is easy to see that we can observe two distinct groups of chess figures based on their proximities. The first group consists of the Pawn, Knight, Queen, and King. The second group consists of the Bishop and the Rook. Thus, it is not possible to obtain a unique global mean of all of the shapes together. If we however, only consider the shapes that are in the same group as defined previously, then the global mean can be attained. Figure 3 and 4 shows the mean shape of the respectively, second, and first group, the third image shows the resulting mean if all the images are taken into consideration. One might notice that the figures are crooked, note that because of the nature of Kendall's shape space, the rotation of a shape does not matter. Despite the figures appearing similar to

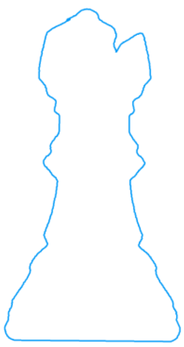


Figure 3: Mean of the Bishop and Rook

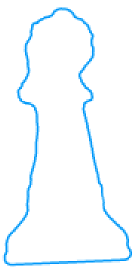


Figure 4: Mean of the Pawn, Knight, King, and, Queen

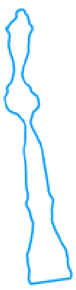


Figure 5: Mean of all six chess pieces

the human eye, they were relatively far apart from each other as shapes. The reason for this could lie in the way we sample our data. The data is generated by evenly sampling a Bézier curve to generate an array of points. In the discrete Kendall's shape space, we require the points of individual shapes to be "approximately" near each other. In the shape of the bishop, there is a "slit" at the top of it. As the slit is only on one side, all the data points might be shifted, relative to other shapes. This could be a possible explanation for why this shape is so far apart from the other shapes. This also shows the limitation of the Kendall's shape space, at leat in the discrete case, the data points might appear very similar to the human eye, while being mathematically completely different.

7 Summary

Throughout this text we have given an introduction into the theory of manifolds. The given theory was defined in the context of optimization on manifolds. The function that we optimized was the function defining the Riemannian center of mass. In our case we specifically considered how this functioned on Kendall's shape space, which consists of the shapes modulo, translation, scaling, and rotation. This allowed us to defined a specific range, in which the center of mass does exist, and is a unique minimum to the objective function that we considered. In the end we performed successfully numerical experiments on the Kendall's shape space, which gave us a mean shape of a set of data points.

References

[1] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008.

[2] Bijan Afsari. Riemannian l^p center of mass: Existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society*, 139(2):655–673, 2011. Article electronically published on August 27, 2010.

[3] Bijan Afsari, Roberto Tron, and René Vidal. On the convergence of gradient descent for finding the riemannian center of mass. *SIAM Journal on Control and Optimization*, 51(3):2230–2260, 2013.

[4] Nicolas Boumal. *An Introduction to Optimization on Smooth Manifolds*. Cambridge University Press, 2023.

[5] T. K. Carne H. Le D.G. Kendall, D. Barden. *Shape and Shape Theory*. John Wiley and Sons, 1999.

[6] D. G. Kendall. The diffusion of shape. *Advances in Applied Probability*, 9(3):428–430, 1977.

[7] Huiling Le. Locating fréchet means with application to shape spaces. *Advances in Applied Probability*, 33(2):324–338, 2001.

[8] Huiling Le and David G. Kendall. The riemannian structure of euclidean shape spaces: A novel environment for statistics. *The Annals of Statistics*, 21(3):1225–1271, 1993.

[9] John M. Lee. *Riemannian Manifolds An Introduction to Curvature*. 1997.

[10] John M. Lee. *Introduction to Smooth Manifolds, Second Edition*. Springer, 2003.

[11] Esfandiar Nava-Yazdani, Hans-Christian Hege, T. J. Sullivan, and Christoph von Tycowicz. Geodesic analysis in kendall's shape space with epidemiological applications. February 24 2020.

[12] Alain Trouve Xavier Pennec Nicolas Guigui, Elodie Maignant. Parallel transport on kendall shape spaces.

[13] B O'Neill. *Semi-Riemannian Geometry With Applications to Relativity*. Academic Press, 1983.

[14] Martha Paskin, Daniel Baum, Mason N. Dean, and Christoph von Tycowicz. A kendall shape space approach to 3d shape estimation from 2d landmarks, 2022.

[15] Peter Petersen. *Riemannian Geometry*, volume 171 of *Graduate Texts in Mathematics*. Springer New York, NY, 2 edition, 2006.

[16] Shing-Tung Yau. Non-existence of continuous convex functions on certain riemannian manifolds. *Mathematische Annalen*, 207:269–270, 1974.

NTNU

Norwegian University of Science and Technology