# Optimization over Grassmann Manifolds

Author: Milos Vukadinovic

Advisor: Peter Dalakov

Department of Mathematics,
American University in Bulgaria

May 13, 2022

**Abstract**

We present optimization methods for functions whose domain lies on Grassmann Manifold. Such functions are ubiquitous because data with subspace-structured features, orthogonality, or low-rank constraints is naturally expressed using Grassmann manifold. We consider two different representations of the Grassmann manifold: a quotient of the Stiefel manifold and a set of projectors. We then develop Grassmannian gradient descent and Grassmannian Newton method on these representations. We demonstrate the efficiency of Grassmannian algorithms by optimizing Rayleigh quotient and conclude that our algorithms converge faster, generalize better and perform as well as the best-known problem-specific algorithms.

# Contents

# Chapter 1

# Introduction

Constrained optimization is the process of optimizing an objective function with respect to some variables with the presence of constraints on those variables. In our paper, we consider geometric constraints, which express that the solution to the optimization problem lies on a manifold. Specifically, we consider problems where the solution lies on the Grassmann manifold - $G(p, n)$ a set of $p$ dimensional subspaces in a higher $n$-dimensional space. Such problems are ubiquitous because data with subspace-structured features, orthogonality constraints, or low-rank constraints can be naturally expressed using the Grassmann manifold. For example, symmetric eigenvalue problems, nonlinear eigenvalue problems, electronic structure computations, and signal processing can all be optimized over $G(p, n)$. In our paper, we first show how the Grassmann manifold can be considered as a quotient manifold and as a set of projectors. Then, we develop Gradient Descent and Newton's algorithm on the Grassmann manifold and study its applications to eigenvalue and eigenvector computations.

A framework for algorithms involving these constraints was first introduced by Edelman et al [EAS98] . They used a quotient manifold representation to develop the Newton algorithm, which inspired a line of works that improve the algorithm or find a different application [HL] [MKP20] [Joh21] [TFBJ18]. A set of projectors approach was introduced by Helmke et al [HHT07], and used to develop Newton algorithms on Grassmannand Lagrange–Grassmann manifolds. While the mentioned representations (set of projectors and quotient manifold) are employed in our paper, they are certainly not exhaustive. For example, Lai et al [LLY20] represent the Grassmann manifold as symmetric orthogonal matrices of trace $2k - n$. Extensive resources for learning about Riemannian optimization are books by Absil [AMS08] and Boumal [Bou22] Moreover, there are some programming frameworks for R, Python, Julia, and Matlab that implemented Grassmannian optimization such as GrassmannOptim [ACW12], and ManOpt [TKW16]. The recent success of geometric deep learning [BBCV21] [MBBV18], shows the importance of exploiting the underlying geometric structure to improve learning. Some attempts to construct a Grassmannian DNN have already been made [ZZHJH18] [HWVG18], but there are many challenges to overcome for them to appear in the industry.

There is a huge amount of digital data that we can use to extract valuable insights and predictions. Any time-series data like stock price, electrocardiogram data, or video data can be considered as points on the Grassmann manifold. Thus the development of Grassmannian optimization algorithms will help us make discoveries, reduce data size, and do it faster than classical optimization algorithms. Moreover, by exploring the optimization algorithms on the Grassmann manifold, we will be taking a first step towards defining a Grassmannian deep neural network. By embedding the geometry in a neural network we will develop models with better accuracy and robustness.

Our contribution is the following. We present the optimization algorithms on the Grassmann manifold and show that they have lower time complexity than classical Euclidean algorithms. Next, we show how exploiting the underlying geometry of data can benefit optimization. Finally, we set the ground for the future research developments in Grassmannian Deep Learning.

We will start by describing a sphere as a smooth manifold and showing that the sphere $S^2$ is equivalent to $S(1,3)$. Then, we will take a look at an example of the Grassmannian which is a sphere with identified antipodal points $G(1,3)$. We will use this example to set the ground for the general $G(p,n)$ case. Then in Chapter 3, we will prove that the Grassmannian is a smooth manifold and show two of its representations. In the same chapter, we will derive all equations needed to create a mathematical setup for optimization algorithms. In chapter 4 we presented three different optimization algorithms and provided pseudocode for them. Finally, in Chapter 5 we demonstrate all presented algorithms on the problem of minimizing the Rayleigh quotient.

# Chapter 2

# $G(1,3)$ As Manifold

## 2.1 Preliminaries

We begin by recalling briefly some standard definitions.

**Definition 2.1.1.** *$(X, \mathcal{T})$ is a* ***locally Euclidean*** *topological space of dimension $n$ if $\forall\ p \in X\ \exists\ v \subseteq X$ which is homeomorphic to an open in $\mathbb{R}^n$.*

**Definition 2.1.2.** *A* ***topological manifold*** *of dimension $n$ is a locally Euclidean space of dimension $n$, which is Hasudorff and has a countable base for its topology.*

**Definition 2.1.3.** *A* ***smooth atlas*** *is a collection of charts $\{(v_\alpha, \varphi_\alpha)\}_\alpha$, $\bigcup_\alpha v_\alpha = X$, s.t. for any two charts $(v_\alpha, \varphi_\alpha)$, and $(v_\beta, \varphi_\beta)$ the transition map $\varphi_{\alpha\beta} = \varphi_\alpha \circ \varphi_\beta^{-1} : \varphi_\beta(v_\alpha \cap v_\beta) \to \varphi_\alpha(v_\alpha \cap v_\beta)$ is infinitely differentiable ($C^\infty$). An atlas is* ***maximal*** *if it is not contained in another atlas.*

It is not hard to see that each atlas is contained in a unique maximal atlas.

**Definition 2.1.4.** *A* ***smooth manifold*** *is a manifold together with a maximal smooth atlas on it. The choice of maximal atlas is also called a* ***smooth structure*** *on the manifold.*

Practically, one never works directly with the chosen maximal atlas. In fact, it is usually preferable to work with an atlas that contains as few charts as possible. To specify the smooth structure (maximal atlas) it is enough to choose one particular smooth atlas – and that determines (implicitly) a maximal atlas.

For different problems different atlasses may be appropriate. To see that two smooth atlasses belong to (determine) the same maximal atlas, one needs to verify that their union is again a smooth atlas. In such a case one also says that the two atlasses are equivalent. This determines an equivalence relation on the set of all atlasses, and the choice of maximal atlas corresponds to a choice of an equivalence class of smooth atlasses.

Let us reiterate that, for a given topological space $X$, being a topological manifold is a **property** – $X$ either is or is not one, while being a smooth manifold is an **additional data** on $X$ – the data of a maximal atlas (smooth structure). Some spaces admit such an extra structure, while others do not. Those topological manifolds that admit smooth structure in fact admit infinitely many smooth structures.

Occasionally in this text we are going to abuse the terminology and say, e.g., *$X$ is a smooth manifold*, which will mean *$X$ can be equipped with a smooth structure* – and a particular choice of smooth structure will be given.

## 2.2 Example: the 2-sphere as a smooth manifold

**Proposition 2.2.1.** *Sphere $S^2$ is a smooth manifold.*

First let's state a lemma that will help us prove the proposition.

**Lemma 2.2.1.** *Let $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$. Suppose the partial derivatives $\frac{\partial f_i}{\partial x_j}$ of $f$ all exist and are continuous in a neighbourhood of a point $x \in U$. Then $f$ is differentiable at $x$.*

The proof for this lemma can be found in any vector calculus textbook [MT88]. Now we proceed with the proof of the proposition.

*Proof.* Let's denote a sphere as:

$$M = S^2 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_1^2 + x_2^2 + x_3^2 = 1\} \tag{2.1}$$

Also, we define a disk with a center at $(x_0, y_0)$ and radius $\epsilon$ as:

$$D_\epsilon(x_0, y_0) = \{(x, y) \in \mathbb{R}^2 : (x - x_0)^2 + (y - y_0)^2 < \epsilon^2\}$$

Thus we can cover the sphere with 6 charts and 6 functions, described as follows:

$$u_i^+ = S^2 \cap \{x_i > 0\}, \quad u_i^- = S^2 \cap \{x_i < 0\}$$

$$\varphi_i^+ : u_i^+ \to D_1(0,0), \quad \varphi(x_1, x_2, x_3) = (\ldots \hat{x}_i \ldots)$$

$$\varphi_i^- : u_i^- \to D_1(0,0), \quad \varphi(x_1, x_2, x_3) = (\ldots \hat{x}_i \ldots)$$

Then the atlas covering the sphere is:

$$A_1 = \{(u_i^\pm, \varphi_i^\pm) : i \in I\}, \quad I = \{1, 2, 3\}$$

For example $\varphi_3^+ : u_3^+ \to \mathbb{R}^2, \quad \varphi_3^+(x_1, x_2, x_3) = (x_1, x_2, \hat{x_3}) = (x_1, x_2)$

We first show that a function $\varphi_3^+ : u_3^+ \to \mathbb{R}$ is injective.
Assume $\exists\, P_1$ and $P_2$ s.t. $P_1 = (p_1, p_2, p_3)$, and $P_2 = (p_1', p_2', p_3')$, $P_1 \neq P_2$, and $f(P_1) = f(P_2)$ implies that $(p_1, p_2) = (p_1', p_2')$ Because $P_1$ and $P_2$ are the points on the sphere:

$$p_3' = \pm\sqrt{1 - p_1^2 - p_2^2} \quad \pm p_3' = \pm\sqrt{1 - (p_1')^2 - (p_2')^2}$$

Since $u_3^+$ has only positive numbers for a third coordinate

$$p_3 = p_3' \implies (p_1, p_2, p_3) = (p_1', p_2', p_3') \implies P_1 = P_2$$

We can conclude that $\varphi_3^+$ is injective.

To show that $\varphi_i^\pm$ is injective, we similarly assume that $\exists\, P_1 P_2$, s.t. $P_1 \neq P_2$ and $\varphi_i^\pm(P_1) = \varphi_i^\pm(P_2)$. Let $K = \{1, 2, 3\}/i$. Then $\forall\, k \in K, a_k = a_k' \quad a_i = \pm\sqrt{1 - \sum(a_k^2)} = a_i'$. Since $a_i$ and $a_i'$ are limited to only positive or only negative values $P_1 = P_2$, and $\varphi_i^\pm$ is injective.

To prove surjectivity, consider $(\varphi_i^\pm)^{-1}$ componentwise. It sends an arbitary point $(b_1, b_2) \in D_1(0,0)$ to a point $(a_1, a_2, a_3) \in S^2$, or componentwise

$$a_j = \begin{cases} b_j & j < i \\ \sqrt{1 - b_1^2 - b_2^2} & j = i \\ b_{j-1} & j > i \end{cases}$$

Since each component in both $S^2$ and the disk varies from 0 to 1, each point in the codomain is mapped to, and we have surjectivity for $\varphi_i^\pm$.

Each function $\varphi_i^{\pm}$ is continuous (and in fact differentiable) since it is the restriction to $u_i^{\pm}$ of a linear map $\mathbb{R}^3 \to \mathbb{R}^2$, and $S^2$ is equipped with the subspace topology. The maps $\left(\varphi_i^{\pm}\right)^{-1}$ are continuous (and in fact, differentiable), since they are continuous (in fact, differentiable) as maps $D_1(0,0) \to \mathbb{R}^3$ by Lemma 2.2.1. Therefore $\varphi_i^{\pm}$ are homeomorphisms.
The sphere $S^2$ is Hausdorff and second countable since it is a subspace of $\mathbb{R}^3$.
We now show that $A$ is a smooth atlas. First we consider the transition function

$$\varphi_3 \circ \varphi_1^{-1} : \phi_1(u_1^+ \cap u_3^+) \longrightarrow \phi_3(u_1^+ \cap u_3^+)$$

$$(x,y) \mapsto (\sqrt{x^2+y^2}, x, y) \mapsto (\sqrt{x^2+y^2}, x).$$

Both $\phi_1(u_1^+ \cap u_3^+)$ and $\phi_3(u_1^+ \cap u_3^+)$ are open half-disks in $\mathbb{R}^2$:

$$\phi_1(u_1^+ \cap u_3^+) = \{(x_1,x_2) \in D_1(0,0), x_2 > 0\},$$

$$\phi_3(u_1^+ \cap u_3^+) = \{(x_1,x_2) \in D_1(0,0), x_1 > 0\},$$

and so both transition functions are infinitely differentiable, by Lemma 2.2.1. Notice that on the boundary of $D_1(0,0)$ differentiability is lost.

Similarly, we can show that all transition functions are $C^\infty$. This proves that the sphere is a smooth manifold. □

In the next section we provide another (equivalent) atlas that uses fewer charts.

## 2.3 Sphere as a smooth manifold using stereographic projection

*Proof.* Stereographic projection 2.3 is a map that will allow us to endow the sphere with smooth structure by using 2 charts.

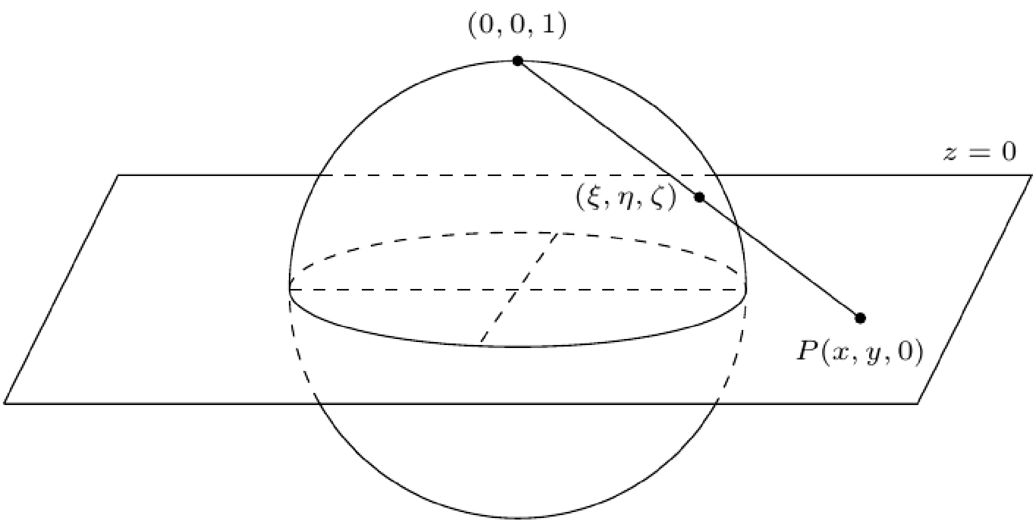



Figure 2.1: Stereographic Projection Visualized

$$v^+ = S^2 \setminus \{(0,0,-1)\} \quad v^- = S^2 \setminus \{(0,0,1)\}$$

We think of map $\phi^-$ as drawing a line through the point on the north pole $(0,0,1)$ and $(x,y,z)$, then the output is the point where the line intersects $z$ plane. Similarly for $\phi^+$, we project from the south pole $(0,0,-1)$. We get the maps explicitly by parametrizing the line:

$$l : (0,0,1) + t((x_1,x_2,x_3) - (0,0,1)) = (x_1 t, x_2 t, x_3 t - t + 1)$$

$l$ intersects $x_3 = 0$ when $x_3 t - t + 1 =$, thus $t = \frac{1}{(1-x_3)}$ It follows that the point of intersection is $(\frac{x_1}{1-x_3}, \frac{x_2}{1-x_3}, 0)$ Similarly we can explicitly find $\phi^+$

$$\phi_+ : v^+ \to \mathbb{R}^2, \quad (x_1, x_2, x_3) \to (\frac{x_1}{1+x_3}, \frac{x_2}{1+x_3})$$

$$\phi_- : v^- \to \mathbb{R}^2, \quad (x_1, x_2, x_3) \to (\frac{x_1}{1-x_3}, \frac{x_2}{1-x_3})$$

We can find the inverses in the similar way. Consinder points on the $z = 0$ plane in $\mathbb{R}^3$ $(\alpha, \beta, 0)$. We parametrize the line through this point and north pole:

$$l : (0, 0, 1) + t((\alpha, \beta, 0) - (0, 0, 1)) = (\alpha t, \beta t, -t + 1)$$

We need to see when are these points going to be on the sphere , i.e.

$$(\alpha t)^2 + (\beta t)^2 + (-t + 1)^2 = 1$$

$$\alpha^2 t^2 + \beta^2 t^2 + t^2 - 2t + 1 = 1$$

$$t^2(\alpha^2 + \beta^2 + 1 - \frac{2}{t}) = 0$$

$$\therefore t = \frac{2}{\alpha^2 + \beta^2 + 1}$$

$$\phi_+^{-1} : \mathbb{R}^2 \to v^+ \quad (\alpha, \beta) \to (\frac{2\alpha}{1 + \alpha^2 + \beta^2}, \frac{2\beta}{1 + \alpha^2 + \beta^2}, -\frac{\alpha^2 + \beta^2 - 1}{\alpha^2 + \beta^2 + 1})$$

$$\phi_-^{-1} : \mathbb{R}^2 \to v^- \quad (\alpha, \beta) \to (\frac{2\alpha}{1 + \alpha^2 + \beta^2}, \frac{2\beta}{1 + \alpha^2 + \beta^2}, \frac{\alpha^2 + \beta^2 - 1}{\alpha^2 + \beta^2 + 1})$$

Since we explicitly found an inverse, the function is a bijection.
And it's easy to see that $\phi$ is continuous by lemma 2.2.1 (denom never zero) Thus we equipped the sphere with the following atlas:

$$A_2 = \{(v^{\pm}, \phi_{\pm})\}$$

We can check that the transition maps from $A_1$ to $A_2$ are smooth.
First we check for $\varphi_3^+ \circ \phi_+^{-1}$ where

$$\varphi_3^+ : S^2 \cap x_3 > 0 = u_3^+ \to D_1(0, 0) \subset \mathbb{R}^2$$

$$\phi_+^{-1} : \mathbb{R}^2 \to v^+ = S^2 \setminus \{(0, 0, -1)\}$$

Now, since we require that the domain of $\varphi_3^+$ is positive, the following is how our transition function will look like

$$\varphi_3^+ \circ \phi_+^{-1} : D_1(0, 0) \to D_1(0, 0), \quad \varphi_3^+(\phi_+^{-1}(x_1, x_2)) = (\frac{2x_1}{1 + x_1^2 + x_2^2}, \frac{2x_2}{1 + x_1^2 + x_2^2})$$

$$\phi_+^{-1} \circ \varphi_3^+ : D_1(0, 0) \to D_1(0, 0), \quad \phi_+(\varphi_3^+((x_1, x_2))^{-1}) = (\frac{x_1}{1 + \sqrt{1 - x_1^2 - x_2^2}}, \frac{x_1}{1 + \sqrt{1 - x_1^2 - x_2^2}})$$

It is smooth componentwise and thus smooth. Next, we list the domain and the image for the rest of transition maps.

$$\varphi_3^- \circ \phi_+^{-1} : \mathbb{R}^2 \setminus \bar{D}_1(0, 0) \to D_1(0, 0) \setminus (0, 0)$$

$$\varphi_2^+ \circ \phi_+^{-1} : \mathbb{R}^2 \cap \{x_2 > 0\} \to D_1(0,0)$$
$$\varphi_1^+ \circ \phi_+^{-1} : \mathbb{R}^2 \cap \{x_1 > 0\} \to D_1(0,0)$$
$$\varphi_2^- \circ \phi_+^{-1} : \mathbb{R}^2 \cap \{x_2 < 0\} \to D_1(0,0)$$
$$\varphi_1^- \circ \phi_+^{-1} : \mathbb{R}^2 \cap \{x_1 < 0\} \to D_1(0,0)$$
$$\varphi_3^+ \circ \phi_-^{-1} : \mathbb{R}^2 \setminus \bar{D}_1(0,0) \to D_1(0,0) \setminus (0,0)$$
$$\varphi_3^- \circ \phi_-^{-1} : D_1(0,0) \to D_1(0,0)$$
$$\varphi_2^+ \circ \phi_+^{-1} : \mathbb{R}^2 \cap \{x_2 > 0\} \to D_1(0,0)$$
$$\varphi_1^+ \circ \phi_+^{-1} : \mathbb{R}^2 \cap \{x_1 > 0\} \to D_1(0,0)$$
$$\varphi_2^- \circ \phi_+^{-1} : \mathbb{R}^2 \cap \{x_2 < 0\} \to D_1(0,0)$$
$$\varphi_1^- \circ \phi_+^{-1} : \mathbb{R}^2 \cap \{x_1 < 0\} \to D_1(0,0)$$

Therefore, we proved that a sphere $S^2$ is a smooth manifold of dimension 2. □

The two smooth atlasses – the one from this section and the one described earlier – are in fact equivalent, i.e., determine the same maximal atlas.

## 2.4 $G(1,3)$ as quotient manifold

As mentioned in the introduction, we define the Grassmannian of $p$-dimensional subspaces in $\mathbb{R}^n$ as the set

$$G(p,n) = \{V \subseteq \mathbb{R}^n \text{ vector subspace, } \dim V = p\} .$$

Each $p$-dimensional subspace $V \subseteq \mathbb{R}^n$ is spanned by $p$ linearly independent vectors in $\mathbb{R}^n$, and such a $p$-tuple of vectors can be thought of as an $n \times p$ matrix of rank $p$. Let $F(n,p) \subseteq \mathrm{Mat}_{n\times p}(\mathbb{R})$ be the set of all such matrices, i.e., all "$p$-frames" in $\mathbb{R}^n$. There are infinitely many different matrices that determine the same subspace $V$ – since there are infinitely many choices of basis. In section 3.1 we show that there is a bijection between $G(p,n)$ and a suitable quotient of $F(n,p)$ – and hence $G(p,n)$ can be equipped with a natural topology. It is in fact a topological manifold and admits a smooth structure, that we describe in section 3.1.

Let us also define the Stiefel manifold as:

$$St(p,n) = \{A \in Mat_{n\times p}(\mathbb{R}) \mid \mathrm{rk}A = p,\ A^T A = I_k\} \subseteq \mathrm{Mat}_{n\times p}(\mathbb{R}). \tag{2.2}$$

This is in fact a compact embedded submanifold of $\mathrm{Mat}_{n\times p}(\mathbb{R}) \simeq \mathbb{R}^{np}$, and its points are *orthonormal* $p$-frames in $\mathbb{R}^n$. Every $p$-dimensional subspace of $\mathbb{R}^n$ has an orthonormal basis, so there is a set-theoretical (in fact, smooth!) map

$$St(p,n) \longrightarrow G(p,n),$$

sending each orthonormal frame to the plane that it spans. If $p > 1$, there are infinitely many different orthonormal frames that span the same $p$-dimensional subspace, and one can represent $G(p,n)$ as a quotient space of $St(p,n)$. In the literature, the Stiefel manifold is usually denoted $V_p(\mathbb{R}^n)$ or $V_{n,p}$.

Let us spell out these objects in the simple case $n = 3$, $p = 1$.

First, $F(3,1) = \mathbb{R}^3 \setminus \{\mathbf{0}\}$, the set of non-zero vectors in $\mathbb{R}^3$. Two non-zero vectors span the same line if they are related through rescaling by a non-zero number $\lambda \in \mathbb{R}^\times = \mathbb{R}\setminus\{0\}$, hence

$$G(1,3) = (\mathbb{R}^3\setminus\{0\})/\sim\, = (\mathbb{R}^3\setminus\{0\})/\,\mathbb{R}^\times = \mathbb{RP}^2.$$

Next, $St(1,3)$ is just the 2-sphere in $\mathbb{R}^3$. Indeed, $St(1,3) = \{A \in \mathbb{R}^{3\times 1} \mid rkA = 1,\ A^TA = 1\} = \{A \in \mathbb{R}^3 \mid A^TA = 1\} = \{(A_1, A_2, A_3) \in \mathbb{R}^3 \mid A_1^2 + A_2^2 + A_3^2 = 1)\}$ we see that this is exactly the equation of the sphere from 2.1. Thus, $S^2 = St(1,3)$.

Two unit vectors in $\mathbb{R}^3$ span the same line if and only if they differ by a sign, so if we introduce on $S^2$ the equivalence relation $\sim$ by declaring $u \sim v$ when $v = -u$ (or $u = v$), we get

$$G(1,3) = \mathbb{RP}^2 \simeq St(1,3)/\sim = S^2 \big/ \mathbb{Z}/2\mathbb{Z}.$$

I.e., if we quotient the Stiefel manifold with this relation, we get a sphere with antipodal points identified, namely:

$$St(1,3)/\sim = \{(A_1, A_2, A_3) \mid A_3 \geq 0\} \cup \{(A_1, A_2, 0) \mid A_2 \geq 0) \cup \{(A_1, 0, 0)\}$$

The Stiefel manifold $St(1,3) = S^2$ can be described as a quotient, $St(1,3) \simeq O_3(\mathbb{R})/O_2(\mathbb{R})$. Here we identify $O_2(\mathbb{R})$ as the subgroup of $O_3(\mathbb{R})$, consisting of matrices $\begin{pmatrix} 1 & 0 \\ 0 & Q \end{pmatrix}$, where $Q \in O_2(\mathbb{R})$. In this description, a unit vector $A \in St(1,3) = S^2$, corresponds to an equivalence class of $3 \times 3$ matrices, all having the same first column equal to $A$.

Correspondingly, $G(1,3) = \mathbb{RP}^2$ is identified with either $O_3/O_1 \times O_2$ or $S^2/O_1$.

## 2.5 $G(1,3)$ as a set of projectors

We can easily see that $G(1,3) \cong \mathbb{RP}^2$. Consider a set of projectors $X \subseteq Mat_{3\times 3}(\mathbb{R})$,

$$X = \{M \mid M^T = M,\ M^2 = M,\ TrM = 1\}$$

If we describe the embedding from $\mathbb{RP}^2$ to $X$, we will understand why we can consider a sphere with antipodal points identified as a set of projectors.

**Proposition 2.5.1.** *There is an embedding from $\mathbb{RP}^2$ to $X \subseteq Mat_{3\times 3}(\mathbb{R}) \cong \mathbb{R}^9$*

*Proof.* We know that $X$ consists of matrices that are symmetric, idempotent and whose eigenvalues add up to one. Spectral theorem tells us that a real symmetric matrix is diagonizable. We can also show that the eigenvectors of symmetric matrices, with distinct eigenvalues, are orthogonal. Indeed, let $x$ and $y$ be eigenvectors of a symmetric matrix $M$, with eigenvalues $\lambda$ and $\mu$, $\lambda \neq \mu$:

$$\lambda\langle x, y\rangle = \langle \lambda x, y\rangle = \langle Mx, y\rangle = \langle xM^T, y\rangle = \langle x, My\rangle = \langle x, \mu y\rangle = \mu\langle x, y\rangle$$

Therefore $(\lambda - \mu)\langle x, y\rangle = 0$, since $\lambda$ and $\mu$ are distinct $xy$=0, thus orthogonal. Next, we can show that the eigenvalues of $M$ can be only 0 and 1. Let $v$ be an eigenvector, of eigenvalue $\lambda$.

$$Mv = \lambda v$$

$$M^2v = M(\lambda v) = \lambda M(v) = \lambda^2 v$$

As $\bar{v} \neq 0$ $(\lambda^2 - \lambda)v = 0 \iff \lambda^2 - \lambda = 0$. Then solving for $\lambda^2 - \lambda = 0$, we get that $\lambda$ can only be 0 or 1. Finally, the fact that $Tr(M) = 1$ tells us that eigenvalues of $M$ are 0,0 and 1. Therefore $\dim\ker M = 2$ and $\dim Im(M) = 1$. This is telling us that there is a whole plane, that is sent to zero vector by $M$, and all vectors in the image are sitting on a line.
Thus, applying a matrix operator $M$ to a vector, is equivalent to projecting a vector to a line in $R^3$. So $M : R^3 \to R^3$ is the operator of orthogonal projection on the line $Im(M)$.
Now, let's explicitly define a map $\phi$, which to given line in $R^3$ assigns a corresponding matrix operator, that will orthogonally project all the vectors in $R^3$ to that line.

$$\phi : \mathbb{RP}^2 \to Mat_{3\times 3} \quad \phi([x : y : z]) \to A$$

To explicitly find $A$, note that we first need to find a unit vector along a line $[x:y:z] \in \mathbb{RP}^2$, we can do that by normalizing coordinates. $n = \frac{1}{\sqrt{x^2+y^2+z^2}}[x,y,z]^T$. Finally, to orthogonally project any $v \in R^3$ along $n$, we apply $(vn)n = v\, n \otimes n$. We can then define $\phi$ as follows:

$$\phi([x:y:z]) = \frac{1}{\sqrt{x^2+y^2+z^2}^2}[x,y,z]^T \otimes [x,y,z] = \frac{1}{x^2+y^2+z^2}\begin{pmatrix} x^2 & xy & xz \\ xy & y^2 & yz \\ xz & yz & z^2 \end{pmatrix}$$

Now, we redefine $\phi : \mathbb{RP}^2 \to \mathbb{R}^6$, because $Mat_{3\times 3} \supseteq Sym_{3\times 3} \simeq \mathbb{R}^6$.

**Definition 2.5.1.** *Immersion Let $X$ and $Y$ be smooth manifolds,* $\dim X = n$*,* $\dim Y = k$*. Let $f : X \to Y$ be a smooth map. We say that $f$ is a*

- *submersion, if $df_p$ is surjective $\forall p \in X$*
- *immersion, if $df_p$ is injective $\forall pinX$ equivalently if $rank D_p f = \dim M, M = f(X)$*

**Definition 2.5.2.** *Let $f : X \to Y$ be a smooth map of smooth manifolds. We say that $f$ is an embedding if*

- *$f$ is an injective immersion*
- *$X$ is homeomorphic to $f(X) \subset Y$ (equipped with the subspace topology)*

**Theorem 2.5.1.** *If $X$ is a compact smooth manifold, and injective immersion $f : X \to Y$ is an embedding.*

Next, we argue that $\phi : \mathbb{RP}^2 \to \mathbb{R}^6$ is an embedding ($\mathbb{R}^6$ because we are taking only non-symmetric lower triangular entries) Namely, we will show that the following function is an embedding.

$$\phi([x:y:z]) = \frac{1}{x^2+y^2+z^2}(x^2, xy, xz, y^2, yz, z^2)$$

First, we show that $\phi$ is well defined. Take two vectors $a$,$b \in [x:y:z]$ on the same line. If $a$ is given by $a = [a_1, a_2, a_3]$ then $b = [ka_1, ka_2, ka_3]$ for $k \in \mathbb{R}$. We need to show that $\phi([a]) = \phi([b])$.

$$\phi([a_1, a_2, a_3]) = \frac{(a_1^2, a_1a_2, a_2^2, a_2a_3, a_3^2)}{a_1^2 + a_2^2 + a_3^2}$$

$$\phi([ka, kb, kc]) = \frac{(k^2a_1^2, k^2a_1a_2, k^2a_2^2, k^2a_2a_3, k^2a_3^2)}{k^2a_1^2 + k^2a_2^2 + k^2a_3^2} = \frac{k^2(a_1^2, a_1a_2, a_2^2, a_2a_3, a_3^2)}{k^2(a_1^2 + a_2^2 + a_3^2)} = \frac{(a_1^2, a_1a_2, a_2^2, a_2a_3, a_3^2)}{a_1^2 + a_2^2 + a_3^2}$$

Now that we showed that $\phi$ is well-defined, we show that it is injective.
Assume that $\phi$ is not injective, then there are unit vectors $a = [a_1, a_2, a_3]$ and $b = [b_1, b_2, b_3]$ lying on different lines such that $\phi([a]) = \phi([b])$ In other words $a \in [x:y:z]$ and $b \in [x':y':z']$

$$\phi([a_1, a_2, a_3]) = (a_1^2, a_1a_2, a_2^2, a_2a_3, a_3^2)$$

$$\phi([b_1, b_2, b_3]) = (b_1^2, b_1b_2, b_2^2, b_2b_3, b_3^2)$$

From $\phi([a_1, a_2, a_3]) = \phi([b_1, b_2, b_3])$ we have that $b_1 = \pm a_1$, $b_2 = \pm a_2$, $b_3 = \pm a_3$, and we know that all $b_i$ have the same sign. Therefore we either have $b = [a_1, a_2, a_3]$ or $b = [-a_1, -a_2, -a_3]$ which both lie on the line $[x:y:z]$. So we have that $b \in [x:y:z]$ which is a contradiction.
We proved that $\phi$ is injective, and we know that $\mathbb{RP}^2$ is compact, so we proceed to proving that $\phi$ is an immersion, once we have that we can claim that $\phi$ is an embedding.
We can use the definition with local charts to prove it. Consider the following charts and maps.

$$u_0 = \{[x:y:z], x \neq 0\} \simeq \mathbb{R}^2$$

$$u_1 = \{[x:y:z], y \neq 0\} \simeq \mathbb{R}^2$$
$$u_2 = \{[x:y:z], z \neq 0\} \simeq \mathbb{R}^2$$
$$\psi_0 : RP^2 \to \mathbb{R}^2, \quad \psi_0([x:y:z]) = (\frac{y}{x}, \frac{z}{x}), \quad \psi_0^{-1}(s,t) = [1:s,t]$$
$$\psi_1 : RP^2 \to \mathbb{R}^2, \quad \psi_1([x:y:z]) = (\frac{x}{y}, \frac{z}{y}), \quad \psi_1^{-1}(s,t) = [s:1:t]$$
$$\psi_2 : RP^2 \to \mathbb{R}^2, \quad \psi_2([x:y:z]) = (\frac{x}{z}, \frac{y}{z}), \quad \psi_2^{-1}(s,t) = [s:t:1]$$

These local charts cover all the points in $\mathbb{RP}^2$, to prove that $\phi$ is an immersion we need to show the following, for all $p \in R^2$

$$rank(J(\phi \circ \psi_i^{-1}(s,t))) = 2$$

for $i \in 1,2,3$.
We check for $\psi_0$.

$$\phi \circ \psi_0^{-1}(s,t) = [1:s:t] \to (1, s, t, s^2, st, t^2)\frac{1}{1+s^2+t^2}$$

$$J(\phi \circ \psi_0^{-1}(s,t)) = \frac{1}{(1+s^2+t^2)^2}\begin{pmatrix} -2s & -2t \\ -s^2+t^2+1 & -2st \\ -2st & s^2-t^2+1 \\ 2s(t^2+1) & -2s^2t \\ t(-s^2+t^2+1) & s(s^2-t^2+1) \\ -2st^2 & 2t(s^2+1) \end{pmatrix}$$

To see that the rank is always 2 we can check the the determinant of minor $\Delta_{4,6} = 4st(s^2+t^2+1)$ which is only zero when $st = 0$. But when both $s = 0$ and $t = 0$ equal to zero, the determinant of the minor $\Delta_{2,3} = 1$, and if $\Delta_{2,3}$ is zero only if $s = 1$ and $t = 0$ or $s = 0$ and $t = 1$. But when that is the case $\Delta_{1,5} \neq 0$. In conclusion there will always be a $2 \times 2$ minor with non-zero determinant, which means that our matrix has rank 2. Similarly we can check that $rankJ(\phi_i) = 2$ ☐

We can conclude that $\phi$ is an immersion, and thus embedding to $\mathbb{R}^9$ and diffeomorphism to $\mathbb{R}^6$.

## 2.6 Gradient and Hessian on the sphere

Tangent space of the sphere is given by

$$T_x St(1,3) = \{v \in \mathbb{R}^3 \mid x^T v = 0\}$$

Normal space is given as

$$N_x St(1,3) = \{aX \mid a \in \mathbb{R}\}$$

We take the metric:

$$g_c(\Delta, \Delta) = \text{tr } \Delta^T (I - \frac{1}{2} X X^T) \Delta$$

We pick $\nabla$ to be the Levi-Civita connetion.
Let $f$ be a function that we want to calculate the gradient of, on the sphere. Consider $f$ as a restriction of a function defined on the higher space, i.e. $f$ is defined on a submanifold and $\bar{f}$ is defined on the whole manifold. In our case, the submanifold is a sphere $S^2$ and the higher manifold is $\mathbb{R}^3$ $f = \bar{f}|_{\mathcal{M}}$ Every vector $\Delta \in T_x\mathbb{R}^3$ admits a decomposition $\Delta = P_x\Delta + P_x^{\perp}\Delta$ where $P_x\Delta \in T_x\mathcal{M}$ and $P_x^{\perp}\Delta \in T_x^{\perp}\mathcal{M}$. Then the gradient is defined as:

$$\nabla f(x) = gradf(x) = P_x grad\bar{f}(x)$$

Using this identity, we can now realize Levi-Civita connection by:

$$\nabla_\eta gradf = P_x D(gradf(x))[\eta] = Hessf(x)\eta$$

We take the following projection $P_x = (I - xx^T)$.

# Chapter 3

# Grassmann Manifold

## 3.1 Grassmannian as a smooth manifold

Let us write $G(p,n) = \{$p-dimensional (vector) subspaces of $\mathbb{R}^{n\times p}\}$. A hyperplane $V \subseteq \mathbb{R}^{n\times p}$ is specified by $n \times p$ matrix $A = [\vec{a}_1, \vec{a}_2, \dots, \vec{a_p}] \in Mat_{n\times p}$ where $\{\vec{a}_1, \vec{a}_2, \dots, \vec{a}_p\}$ is a basis for V. I.e., given $A \in Mat_{n\times p}$, rk$A = p$, we get a p-hyperplane $V \subseteq \mathbb{R}^{n\times p}$ by $V =$span $\{\vec{a}_1, \vec{a}_2, \dots, \vec{a}_p\}$. Conversely, given any p-dim subspace $V \subseteq \mathbb{R}^{n\times p}$, there is a $n \times p$ matrix A with rk$(A) = p$, from which $V$ is obtained in the above way. Two matrices, $A$ and $B$, determine the same subspace $V \iff \exists g \in GL(p)$, such that $B = Ag$. $GL(p)$ stands for general linear group of degree $p$ over real field.
We thus have the following setup. Let the set of all 2-frames be

$$F(p,n) = \{A \mid rkA = p\} \subseteq Mat_{n\times p}(\mathbb{R}) \simeq \mathbb{R}^{n\times p}$$

and consider on it the equivalence relation

$$B \sim A \text{ if } \exists\, g \in GL(p), \text{ s.t. } B = Ag$$

We have described a bijection of sets

$$F(p,n)/\!\sim\, \simeq G(p,n).$$

We will show that $G(p,n)$ is a manifold equipped with a natural smooth structure. To achieve that we need to prove that:

- $G(p,n)$ has a countable base
- $G(p,n)$ is Hausdorff
- $G(p,n)$ is locally euclidean

**Proposition 3.1.1.** *$F(p,n)$ is an open subset of $\mathbb{R}^{p\times n}$*

**Lemma 3.1.1.** *The rank of an $m \times n$ matrix is $r \iff$ some $r \times r$ minor does not vanish, and every $(r+1)\times(r+1)$ minor vanishes.*

Since we know that for $M \in F(p,n)^{\complement}$, rk$M < p$ lemma 3.1.1 tells us all $p \times p$ minors of an arbitary element $A \in F(p,n)^{\complement}$ vanish. Let's denote the determinant of each minor of $A$ with $S_i,\ i \in (1, 2, \dots, \binom{n}{p})$ Then consider a continuous map $\psi : Mat_{n\times p} \to \mathbb{R}^{\binom{n}{p}}, \quad \psi(M) \to (S_1, S_2, \dots, S_{\binom{n}{p}})$. We can express $F(p,n)^{\complement} = \psi^{-1}(\vec{0})$ because all minors vanish (det=0). A point $\vec{0} \in \mathbb{R}^{\binom{n}{p}}$ is a closed set, and because continuity preserves the closedness, $F(p,n)^{\complement}$ is closed in $Mat_{n\times p}$, an since its complement is closed, $F(p,n)$ is an open subset of $Mat_{n\times p}(\mathbb{R})$

**Proposition 3.1.2.** *$\sim$ is an open equivalence relation on $F(p,n)$*

In other words we need to show that the map $\pi : F(p,n) \to F(p,n)/\sim$ is an open map. Then $\pi$ is a quotient map and is equipped with $F(p,m)/\sim$ is equipped with quotient topology.

**Lemma 3.1.2.** *A subset of a quotient space is open if and only if its preimage under the canonical projection map is open in the original topological space.*

Let $U$ be an open in $F(p,n)$. Then for every $g \in GL(p)$ the set $Ug = \{xg | x \in U\}$ is an open subset of $F(p,n)$. Therefore $\pi^{-1}\pi(U) = \bigcup_{g \in G} Ug$ is an open in $F(p,n)$ because the union of open sets is open. And by 3.1.2 $\pi(U) = [U]$ is open in $G(p,n)$. $\pi$ is a canonical quotient map, and $F(p,n)/\sim$ is open in $\mathbb{R}^{n\times p}$.

**Lemma 3.1.3.** *if $\beta = \{\beta_\alpha\}_\alpha$ is a base for a topology $\mathcal{T}$ on a topological space $S$, and if $f : S \to X$ is an open map, then the collection $\{f(\beta_\alpha)\}_\alpha$ is a base for the topology on $X$.*

*Proof.* Let $V$ be an open in $X$ and $y \in V$. Choose $x \in f^{-1}(y)$. Since $f^{-1}(V)$ is open there is a basis element $U \in \beta$ s.t. $x \in U \subset f^{-1}(V)$ which implies that $y \in f(U) \subset V$. Since $y$ is arbitary, and $f(U) \subset f(\beta)$ the collection $\{f(\beta_\alpha)\}_\alpha$ is a base for the topology on $X$. □

**Proposition 3.1.3.** *$G(p,n)$ has a second countable base.*

*Proof.* We know that $F(p,n)$ has a second countable base since it is a subspace of $\mathbb{R}^{n\times p}$. Thus by lemma 3.1.3, we have that the base of $G(p,n)$ is second countable. □

**Proposition 3.1.4.** *The graph of the equivalence relation on $F(p,n)$ is a closed subset of $F(p,n) \times F(p,n)$. i.e. $R = \{(A,B) \in F(p,n) \times F(p,n) \mid A = Bg\}$ is closed.*

*Proof.* We can consider $R$ as a set of matrices $[AB] = [\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_p, \vec{b}_1, \vec{b}_2, \ldots, \vec{b}_p]$ of rank $p$. Lemma 3.1.1 tells us that every $(p+1)\times(p+1)$ minor of an element in $R$ must vanish. Consider the map that assigns to $(A,B)$ the values of all $(p+1)\times(p+1)$ minors

$$\psi : F(p,n) \times F(p,n) \to \mathbb{R}^{\binom{n}{3}(p+2)}$$

Since $\phi$ is continuous ( as all of its components are polynomials ) and $R = \psi^{-1}(0)$ , then R is closed. □

**Example 3.1.1.** For example take $G(2,4)$, then $\phi : Mat_{4\times 2} \to \mathbb{R}^{16}$

**Proposition 3.1.5.** *$G(p,n)$ is Hausdorff.*

*Proof.* Because $R$ is closed in $F(p,n) \times F(p,n)$, $(F(p,n) \times F(p,n)) \setminus R = R^{\complement}$ is open. $\implies$ $\forall (x,y) \in R^{\complement}$ there is a basic open set $u \times v$ containing $(x,y)$ s.t. $(u \times v) \cap R = \emptyset$ $\implies$ $\forall x, y$ s.t. $(x,y) \notin R, \exists$ u around x and v around y s.t. $u \cap v = \emptyset$ Thus for any two points $[x] \neq [y] \in F(p,n)/\sim$ there exist disjoint neighborhood of $x$ and $y$ and $F(2,4)/\sim$ which is exactly the definition of Hausdorff property. □

**Proposition 3.1.6.** *$G(p,n)$ is locally euclidean.*

*Proof.* Now that we have Hausdorff property and secound countable basis, we need to prove that every point lying on a manifold has a neighbourhood that is homeomorphic to an open in $\mathbb{R}^n$. Then we can claim that $G(p,n)$ is a manifold.
First we define charts. Take $A \in Mat_{n\times p}$ denote by $A_k$, ($k \in$ all possible picks of p from the set $[1,\ldots,n]$) the $p \times p$ minor, formed by the $k_1$th $\ldots k_p$th rows of $A$. The set

$$U_k = \{A \mid \det(A_k) \neq 0\} \subset F(p,n)$$

is open, because its complement is closed. We also have that $\forall g \in GL(p)$ if $A \in U_k$ then $Ag \in U_k$. Indeed, because $\det(Ag) = \det(A)\det(g)$, $\det((Ag)_{i,j}) \neq 0$ which means $Ag$ will belong to a set $U_k$ Next, define

$$V_k = U_k/\sim = \pi(U_k) \subset G(p,k)$$

The set $V_k$ is open since the equivalence relation is open. i.e. $\pi$ is an open map.
$U_k$ has a canonical representative $A \sim \widehat{AA_k^{-1}}$. $\widehat{\cdot}$ discards all the rows whose index is in $k$. Similarly $V_k$ has a canonical representative: $[A] \sim [\widehat{AA_k^{-1}}]$

**Example 3.1.2.** Following the previous example consider $[A] \in G(2,4)$. If a minor $A_{2,4}$ is invertible we have that $[A] \sim [\widehat{AA_{2,4}^{-1}}] = \widehat{\begin{bmatrix} * & * \\ 1 & 0 \\ * & * \\ 0 & 1 \end{bmatrix}} = \begin{bmatrix} * & * \\ * & * \end{bmatrix}$. Since charts $\bigcup U_k$ cover $F(p,n)$, charts $V_k$ cover $G(p,n)$ (because $\pi$ is open).

Now we define homeomorphisms between charts $V_k$ and opens in $R^{p\times(n-p)}$ as follows:

$$\phi_k : V_k \subset G(p,n) \to Mat_{(n-p)\times p}(\mathbb{R}) \simeq \mathbb{R}^{(n-p)\times p}, \quad \phi_k([A]) = \widehat{AA_k^{-1}}$$

We can show that $\phi$ is well defined.
Let $A, A' \in [A]$ we will show that $\phi$ is well defined. Equivalently $\phi_k(A) = \phi_k(A'])$ Since $A$ and $A'$ are in the same class, we have that $A' = Ag, \quad g \in GL(p)$, $\phi_k(A) = AA_k^{-1}$.

$$\phi_k(A') = \phi_k(Ag) = Ag((Ag)_k)^{-1} = Ag(A_kg)^{-1} = Agg^{-1}A_k^{-1} = AIA_k^{-1} = AA_k^{-1} = \phi_k(A)$$

$\phi$ is continuous because matrix multiplication is continous. Next, we can see that $\phi$ is surjective and $\phi^{-1}$ is continuous by explicitly defining inverse.

$$\phi_k^{-1}(\begin{pmatrix} — \alpha_1 — \\ \vdots \\ — \alpha_{n-p} — \end{pmatrix}) = \begin{pmatrix} 1_1 \\ \vdots \\ 1_p \\ \alpha_1 \\ \vdots \\ \alpha_{n-p} \end{pmatrix}$$

Finally, to show that $\phi$ is a homeomorphism, we have left to shot that $\phi$ is injective.
Assume that there $\phi_k$ is not injective then there are $A \in [A]$ and $B \in [B]$ such that there is **no** $g \in GL(p)$ for which $Ag = B$. i.e. $AA_k^{-1} = BB_k^{-1} \iff AA_k^{-1}B_k = B$ but $A_k^{-1}B_k \in GL(p)$ thus we reach contradiction. Therefore $\phi_k$ is homeomorphism and we proved that $G(p,n)$ is locally Euclidean. □

**Example 3.1.3.** Let $A = \begin{bmatrix} 2 & 6 \\ 1 & 3 \\ 2 & 1 \\ 4 & 3 \end{bmatrix}, \quad [A] \in V_{3,4}$

$$AA_{3,4}^{-1} = \begin{pmatrix} 2 & 6 \\ 1 & 3 \\ 2 & 1 \\ 4 & 3 \end{pmatrix} \begin{pmatrix} \frac{3}{2} & -\frac{1}{2} \\ -2 & 1 \end{pmatrix} = \begin{pmatrix} -9 & 5 \\ -\frac{9}{2} & \frac{5}{2} \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$$

the above multiplication is continuous by 3.1.2 and we can exlcude rows 3 and 4 so that we get result in $R^4$. Then the restriction to $R^4$ is also continuous.

$$\phi_{3,4}([A]) = \begin{pmatrix} -9 & 5 \\ -\frac{9}{2} & \frac{5}{2} \end{pmatrix}$$

Next the inverse map $\phi_{3,4}(\beta)^{-1}$ $\beta \in Mat_{2\times 2} = \phi_{3,4}(A_{1,2}A_{3,4}^{-1}g) = [A]$ for some matrix $A$, such that $A_{1,2} = \beta$. But if we pick $g = A_{3,4}$ then $\phi_{3,4}(\beta) = \begin{bmatrix} \beta_{1,1} & \beta_{1,2} \\ \beta_{2,1} & \beta_{2,2} \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$ More generally

$$\phi_{i,j}^{-1} : \mathbb{R}^4 \to v_{i,j} \subset G(2,4) \quad \phi_{i,j}^{-1}(\beta) \to \begin{bmatrix} \beta \\ I_{2\times 2} \end{bmatrix} = [\alpha]$$

Such that $\alpha[i :] = \beta[1 :]$, $\alpha[j :] = \beta[2 :]$ , $\alpha[(I \setminus \{i,j\})[1]] = I[1 :]$ and $\alpha[(I \setminus \{i,j\})[2]] = I[2 :]$

**Example 3.1.4.** $\phi_{3,4}^{-1}(\alpha) = [A]$ as defined in 3.1.3

$$\phi_{3,4}^{-1}\begin{pmatrix} -9 & 5 \\ -\frac{9}{2} & \frac{5}{2} \end{pmatrix} = \begin{bmatrix} -9 & 5 \\ -\frac{9}{2} & \frac{5}{2} \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

We can confirm that $\alpha = \begin{bmatrix} -9 & 5 \\ -\frac{9}{2} & \frac{5}{2} \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $A = \begin{bmatrix} 2 & 6 \\ 1 & 3 \\ 2 & 1 \\ 4 & 3 \end{bmatrix}$ span the same subspace. Because if we take $g = \begin{pmatrix} 2 & 1 \\ 4 & 3 \end{pmatrix}$ then $\alpha g = A$

Since $\bigcup U_{i,j}$ covers $F(2,4)$, $\cup v_{i,j}$ covers $G(2,4)$ Finally, we check transition maps.

$$\phi_{1,2}([A])^{-1} = A_{3,4}A_{1,2}^{-1}, \quad \phi_{1,2}^{-1}(u) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ v_{1,1} & v_{1,2} \\ v_{2,1} & v_{2,2} \end{pmatrix}$$

$$\phi_{2,4}([A]) = A_{1,3}A_{2,4}^{-1}, \quad \phi_{2,4}^{-1}(v) = \begin{pmatrix} v_{1,1} & v_{1,2} \\ 1 & 0 \\ v_{2,1} & v_{2,2} \\ 0 & 1 \end{pmatrix}$$

$$\phi_{2,4}\circ\phi_{1,2}^{-1}(v) = \begin{pmatrix} v_{1,1} & v_{1,2} \\ v_{2,1} & v_{2,2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ v_{1,1} & v_{1,2} \end{pmatrix}\begin{pmatrix} 0 & 1 \\ v_{2,1} & v_{2,2} \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ v_{1,1} & v_{1,2} \end{pmatrix}\begin{pmatrix} v_{2,2} & -1 \\ -v_{2,1} & 0 \end{pmatrix}\frac{1}{-v_{2,1}} = -\frac{1}{v_{2,1}}\begin{pmatrix} v_{2,2} & -1 \\ v_{1,1}v_{2,2} - v_{1,2}v_{2,1} & -v_4 \end{pmatrix}$$

Now let's check the transition map $\phi_{3,4} \circ \phi_{2,3}^{-1}$

$$\phi_{3,4} \circ \phi_{2,3}^{-1}(v) = \begin{pmatrix} v_{1,1} & v_{1,2} \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ v_{2,1} & v_{2,2} \end{pmatrix} = -\frac{1}{v_{2,1}}\begin{pmatrix} v_{1,1}v_{2,2} + v_{1,2}v_{2,1} & -v_{1,1} \\ v_{2,2} & -1 \end{pmatrix}$$

**Proposition 3.1.7.** *$G(p,n)$ can be equipped with the structure of a $p(n-p)$ dimensional smooth manifold.*

The proof for the proposition follows from propositions 3.1.6, 3.1.5, 3.1.3, and by checking that transition maps are infinitely differentiable.

## 3.2 Grassmann manifold as a quotient manifold

We will describe the Grassmann manifold as a quotient of the Stiefel manifold with respect to the orthogonal group.

$$\begin{array}{ccc} St(p,n) & \hookrightarrow & F(p,n) \\ & \overset{p}{\searrow} & \downarrow \pi \\ & & G(p,n) \end{array}$$

Map $p$ is surjective because every subspace has an orthonormal basis. I.e. starting with any basis we can construct an orthonormal one via Gram-Schmidt algorithm. Now, if we redefine the map $p$ such that $p : St(p,n)/_{O(p)} \to G(p,n)$, where $O(p)$ is the orthogonal group of $2-frames$, we will have a bijection. To see why is it possible to quotient over $O(p)$ instead $GL(p)$ consider $A, B \in St(p,n)$ and consider that $A$ and $B$ are in the same subspace (go to the same point under equivalence relation) $A = Bg, g \in GL(p)$. We know that $A^TA = I$, when we substitute we get $(Bg)^TBg = I \implies g^TB^TBg = I \implies g^Tg = I$ which tells us that $g$ has to be an element of the orthogonal group. Therefore

$$G(p,n) = F(p,n)/_{Gl_p(\mathbb{R})} \quad G(p,n) = St(p,n)/_{O(p)}$$

We also have the diffeomorphism $St(p,n) \simeq O_n(\mathbb{R})/\,O_{n-p}(\mathbb{R})$.

Here we identify $O_{n-p}(\mathbb{R})$ as the subgroup of $O_n(\mathbb{R})$, consisting of matricess $\begin{pmatrix} 1 & 0 \\ 0 & Q \end{pmatrix}$, where $Q \in O_{n-p}(\mathbb{R})$. In this description, a matrix $A \in St(p,n)$, with mutually orthogonal and orthonormal column, corresponds to an equivalence class of $n \times n$ matrices, all having the same first $n \times p$ block equal to $A$.

## 3.3 Grassmann manifold as a set of projectors

Given

$$X = \{M \mid M^2 = M = M^T, trM = p\} \subset M_{n \times n}$$

We will prove that there is an embedding of $G(p,n)$ to $X$. Based on the section 2.5 , we hypothesize that the embedding is given by

$$\phi :\ G(p,n) \to X \quad \phi(A) = A(A^TA)^{-1}A^T$$

**Proposition 3.3.1.** *$\phi$ is well defined*

*Proof.* Take $A \in G(p,n)$ and $B \in G(p,n)$ s.t. $B = Ag$. We know that $\phi(A) = A(A^TA)^{-1}A^T$ then

$$\phi(B) = Ag((Ag)^TAg)^{-1}(Ag)^T = Ag(g^TA^TAg)^{-1}(Ag)^T = Agg^{-1}(g^TA^TA)^{-1}g^TA^T = A(A^TA)^{-1}A^T \tag{3.1}$$

□

**Proposition 3.3.2.** *$\phi$ is injective*

*Proof.* Assume that a function is not injective, then $\exists A, B \in G(p,n)$ s.t. $\phi(A) \neq= \phi(B)g$ for any $g$ in $GL(\mathbb{R})$ equivalently $A(A^TA)^{-1}A^T = B(B^TB)^{-1}B^T$, we use the fact that any $n \times p$ matrix $A$ can be decomposed as $A = QR$ where $Q$ is of shape $p \times p$ and $R$ is of shape $n \times p$ then

$$A(A^TA)^{-1}A^T = (QR)((QR)^TQR)^{-1}(QR)^T = QR(R^TQ^TQR)^{-1}R^TQ^T =$$

$$QR(R^TR)^{-1}R^TQ^T = QRR^{-1}(R^T)^{-1}R^TQ^T = QQ^T$$

But we know that there exists $Q$ such that $\exists\ g$ s.t. $Q = Q'$. Therefore we get that $A = QR \quad B = Q'R'$ and $\phi(A) = QQ^T = \phi(B) = Q'Q'^T$ but since $Q$ is orthogonal we know that $\exists\, g\ Q = Q'g \implies A = Bg$. we reach the contradiction and prove that $\phi$ is injective. ☐

**Proposition 3.3.3.** *$\phi$ is differentiable*

*Proof.*

$$\phi'(A) = (A(A^TA)^{-1}A)' = A'(A^TA)^{-1}A - A(A^TA)^{-1}(A'A^T)(A^TA)^{-1}A -$$

$$A(A^TA)^{-1}(A(A^T)')(A^TA)^{-1}A^T + A(A^TA)^{-1}(A^T)'$$

Since we know that $A$ is differentiable this equation shows us that $\phi(A)$ is differentiable.

☐

**Proposition 3.3.4.** *$P$ is a projection matrix to the subspace $A$, if given a vector $u$ that lies in the subspace, and $v$ that is perpendicular to the subspace $A$, $Pu = u$ and $Pv = 0$. Show that $\phi(A)$ is a projector.*

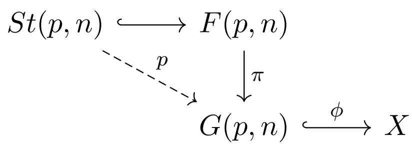
*Proof.* First we show that given a vector that already lies on $A$, the vector won't change. Let $u = Av$, then $\phi(u) = A(A^TA)^{-1}A^TAv = Av = u$ Given a vector orthogonal to $A$ projection will go to zero. Take arbitary $u$ such that

$$u = u^{\perp} + u^{\parallel} \quad u^{\parallel} \in ImA \implies u^{\parallel} = Av$$

Then $u^{\perp} = u - u^{\parallel}$. Now to show that $\phi(u^{\perp}) = 0$

$$\phi(u - u^{\parallel}) = \phi(u) - \phi(u^{\parallel}) = A(A^TA)^{-1}A^Tu - A(A^TA)^{-1}A^TAv = Av - Av = 0$$

We will in fact show that $X$ can be identified with $St(p,n)/\sim$ which is (as we showed in 3.2 ) $G(p,n)$ ☐




In this setup $A \in St(p,n)$ $\phi(A) = A(A^TA)^{-1}A^T = AA^T$

$$D\phi(X)[V] = \lim_{t\to 0} \frac{\phi(X+tV) - \phi(X)}{t} = \lim_{t\to 0} \frac{(X+tV)(X+tV)^T - XX^T}{t} =$$

$$\lim_{t\to 0} \frac{(X+tV) - (X^T + (tV)^T) - XX^T}{t} = \lim_{t\to 0} \frac{XX^TX(tV)^T + tVX^T + tV(tV)^T - XX^T}{t} =$$

$$\lim_{t\to 0} \frac{tXV^T + tVX^T + t^2VV^T}{t} = XV^T + VX^T$$

Take $V = \frac{1}{2}XB \quad B \in Sym(p)$

$$\frac{1}{2}XA^TX^T + \frac{1}{2}XAX^T = XAX^T$$

It can be checked that $XAX^T \in X$ In other words for any matrix $XAX^T \in X$ there exists a matrix $V \in \mathbb{R}^{n\times p}$, such that $Dh(X)[V] = XAX^T$. Thus, $\phi$ is a defining function for $G(p,n)$ making it an embedded submanifold.

## 3.4 Tangent Space

To define a tangent and a normal space we need the metric. When working on the Stiefel manifold the canonical metric is introduced with the purpose to restrict the orthogonal group metric to the horizontal space the canonical metric is introduced with the purpose to restrict the orthogonal group metric to the horizontal space. Canonical metric on Stiefel is given as:

$$g_c(\Delta_1, \Delta_2) = \text{tr } \Delta^T(I - \frac{1}{2}AA^T)\Delta$$

However the canonical metric on Grassman manifold is equivalent to the Euclidean metric

$$g_c(\Delta_1, \Delta_2) = \text{tr } \Delta_1^T(I - \frac{1}{2}YY^T)\Delta_2 = \text{tr } \Delta_1^T\Delta_2 = g_e(\Delta_1, \Delta_2)$$

Thus we proceed with such choice of canonical metric.

**Proposition 3.4.1.** *The tangent space of $G(p, n)$ is given by all the commutators* $[P, \Omega] = P\Omega - \Omega P \quad \Omega \in \mathfrak{so}_n$

*Proof.* Consider the map $\delta : O(n) \to G(p, n), \quad \delta(T) = TP_0T^T$. So that we fix $P_0$ to satisfy the following three conditions:

1. $P^T = P \quad (TP_0T^T)^T = TP_0^TT^T$
2. $P^2 = P \quad (TP_0T^T)(TP_0T^T) = TP_0T^T$
3. $tr(TP_0T^T) = tr(P_0T^TT) = trP_0 = k$

Here these three rules are saying that we can get any projector $P_n = TP_0T^T$ Note that $\delta$ is a submersion and therefore it induces a surjective map on tangent spaces. The tangent space of $O(n)$ at the $n \times n$ identity matrix $I$ is $T_xO(n) = \{\Omega \in \mathfrak{so}_n\}$ Note that $\Omega \in \mathfrak{so}_n$ means that omega is skew-symmetric $\Omega^T = -\Omega$. Now if we take a derivative

$$D_\delta : TxO(n) \to T_{P_o}G(p, n), \quad \Omega \to P_0\Omega - \Omega P_0$$

□

**Example 3.4.1.** Tangent space in $G(1, 3)$ Take $P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ find all $T_pX = \{[P, \Omega] \mid \Omega \in \mathfrak{so}_n\}$ $\Omega = \begin{bmatrix} \Omega_1 & -\Omega_2 & -\Omega_3 \\ \Omega_2 & \Omega_4 & -\Omega_5 \\ \Omega_3 & \Omega_5 & \Omega_6 \end{bmatrix}$ $\Omega P = \begin{bmatrix} \Omega_1 & 0 & 0 \\ \Omega_2 & 0 & 0 \\ \Omega_3 & 0 & 0 \end{bmatrix}$ $P\Omega = \begin{bmatrix} \Omega_1 & -\Omega_2 & -\Omega_3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ Thus, the elements in the tangent spaces look like: $P\Omega - \Omega P = \begin{bmatrix} 0 & -\Omega_2 & -\Omega_3 \\ \Omega_2 & 0 & 0 \\ \Omega_3 & 0 & 0 \end{bmatrix}$

**Example 3.4.2.** Tangent space in $G(2, 3)$ $P \in G(2, 3) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ $\Omega = \begin{bmatrix} \Omega_1 & -\Omega_2 & -\Omega_3 \\ \Omega_2 & \Omega_4 & -\Omega_5 \\ \Omega_3 & \Omega_5 & \Omega_6 \end{bmatrix}$ $P\Omega = \begin{bmatrix} \Omega_1 & -\Omega_2 & -\Omega_3 \\ \Omega_2 & \Omega_4 & -\Omega_5 \\ 0 & 0 & 0 \end{bmatrix}$ $\Omega P = \begin{bmatrix} \Omega_1 & -\Omega_2 & 0 \\ \Omega_2 & \Omega_4 & 0 \\ \Omega_3 & \Omega_5 & 0 \end{bmatrix}$ Thus, the elements in the tangent space look like: $P\Omega - \Omega P = \begin{bmatrix} 0 & 0 & -\Omega_3 \\ 0 & 0 & -\Omega_5 \\ -\Omega_3 & -\Omega_5 & 0 \end{bmatrix}$

**Example 3.4.3.** Tangent space in $G(2,4)$ $P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ $\Omega = \begin{bmatrix} \Omega_1 & -\Omega_2 & -\Omega_3 & -\Omega_4 \\ \Omega_2 & \Omega_5 & -\Omega_6 & -\Omega_7 \\ \Omega_3 & \Omega_6 & \Omega_8 & -\Omega_9 \\ \Omega_4 & \Omega_7 & \Omega_9 & \Omega_{10} \end{bmatrix}$
$P\Omega = \begin{bmatrix} \Omega_1 & -\Omega_2 & -\Omega_3 & -\Omega_4 \\ \Omega_2 & \Omega_5 & -\Omega_6 & -\Omega_7 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ $\Omega P = \begin{bmatrix} \Omega_1 & -\Omega_2 & 0 & 0 \\ \Omega_2 & \Omega_5 & 0 & 0 \\ \Omega_3 & \Omega_6 & 0 & 0 \\ \Omega_4 & \Omega_7 & 0 & 0 \end{bmatrix}$ $[P,\Omega] = \begin{bmatrix} 0 & 0 & -\Omega_3 & -\Omega_4 \\ 0 & 0 & -\Omega_6 & -\Omega_7 \\ -\Omega_3 & -\Omega_6 & 0 & 0 \\ -\Omega_4 & -\Omega_7 & 0 & 0 \end{bmatrix}$

Now, based on our examples, we can see that: $P_0 = \begin{bmatrix} I_p & 0 \\ 0 & 0 \end{bmatrix}$ where $I_p$ is $p \times p$ identity matrix. For $G(p,n)$ we have the result $\begin{bmatrix} 0_p & A^T \\ A & 0_{n-p} \end{bmatrix}$ $\Omega = \begin{bmatrix} A & -B \\ B & C \end{bmatrix}$ where $A^T = -A$ and it's shape is $p \times p$ and $C^T = -C$ and it's shape is $(n-p) \times (n-p)$ $P\Omega = \begin{bmatrix} A & -B^T \\ 0 & 0 \end{bmatrix}$ $\Omega P = \begin{bmatrix} A & 0 \\ B & 0 \end{bmatrix}$ So the tangent space looks like: $[\Omega, P] = \begin{bmatrix} 0 & -B^T \\ B & 0 \end{bmatrix}$ where $B$ has the shape $p \times (n-p)$ And because we considered this under equivalence relation $Q \in O(p)$, the tangent of $G(p,n)$ is described as $T_xG(p,n) = \{\Delta \mid \Delta = Q \begin{bmatrix} 0 & -B^T \\ B & 0 \end{bmatrix}\}$

## 3.5 Normal Space

$$N_xG(p,n) = (T_xG(p,n))^{\perp} = \{U \in \mathbb{R}^{n \times p} : \langle U, V \rangle = 0 \text{ for all V } \in T_xG(n,p)\}$$

$$N_xG(p,n = \{U \in \mathbb{R}^{n \times p} \mid U^T Q \begin{bmatrix} 0 & -B^T \\ B & 0 \end{bmatrix} = 0\})$$

From here we can see that $U = Q \begin{bmatrix} A & 0 \\ 0 & C \end{bmatrix}$ Therefore

$$N_xG(p,n) = \{Q \begin{bmatrix} A & 0 \\ 0 & C \end{bmatrix} \mid Q \in O(p), A \in \mathfrak{so}(p), C \in \mathfrak{so}(n-p)\}$$

## 3.6 Geodesic

The orthogonal group geodesic is given as

$$Q(t) = Q(0) \exp t \begin{pmatrix} 0 & -B^T \\ B & 0 \end{pmatrix}$$

It has a horizontal tangent at every point along the curve $Q(t)$

$$\dot{Q}(t) = Q(t) \begin{pmatrix} 0 & -B^T \\ B & 0 \end{pmatrix}$$

Thus Grassmann geodesics = $[Q(t)]$ The following theorem will be useful for computing the geodesic formula.

**Theorem 3.6.1.** *If* $Y(t) = Qe^{t\begin{pmatrix} 0 & -B^T \\ B & 0 \end{pmatrix}} I_{n,p}$ *with* $Y(0) = Y$ *and* $\dot{Y}(0) = H$*, then*

$$Y(t) = \begin{pmatrix} YV & U \end{pmatrix} \begin{pmatrix} \cos \Sigma t \\ sin \Sigma t \end{pmatrix} V^T$$

## 3.7 Parallel Transport

**Theorem 3.7.1.** *Let $H$ and Delta be tangent vectors to the Grassmann manifold at $Y$. Then the parallel translation of $\Delta$ along the geodesic in the direction $\dot{Y}(0) - H$ is*

$$\tau\Delta(t) = \left( \begin{pmatrix} YV & U \end{pmatrix} \begin{pmatrix} -\sin \Sigma t \\ \cos \Sigma t \end{pmatrix} U^T + (I - UU^T) \right) \Delta$$

## 3.8 Gradient

The gradient of $F$ at $[Y]$ is defined to be the tangent vector $\nabla F$ such that

$$trF_Y^T \Delta = g_c(\nabla F, \Delta) = tr(\nabla F)^T \Delta \tag{3.2}$$

For all tangent vectors $\Delta$ at $Y$.
Solving the equation 3.2 for $\nabla F$ such that $Y^T(\nabla F) = 0$ we get

$$\nabla F = F_Y - YY^T F_Y \tag{3.3}$$

## 3.9 Hessian

Hessian is defined as

$$HessF(\Delta_1, \Delta_2) = F_{YY}(\Delta_1, \Delta_2) - \text{tr } (\Delta_1^T \Delta_2 Y^T F_Y)$$

For Newton's method, we must determine $\Delta = -Hess^{-1}G$, which for the Grassmann manifold is expressed as the linear problem:

$$F_{YY}(\Delta) - \Delta(Y^T F_Y) = -G$$

# Chapter 4

# Optimization Algorithms

Classical Gradient Descent is defined as follows:

1. $\Delta x_k = \frac{d}{dx_k} f(x_k)$
2. $x_{k+1} = x_k - lr \cdot \Delta x_k$

It computed the gradient of the function, and then in the next steps move in the direction of gradient. When the min/max is sufficiently close, it stops.
Newton's root finding method is given in the following two steps

1. $\Delta x_k = -\frac{f(x_k)}{f'(x_k)}$
2. $x_{k+1} = x_k + \Delta x_k$

We perform optimization using Newton's method by applying it to the derivative of twice differentiable function $f$ to find the critical points.
Now we proceed by defining Gradient Descent and Newton's method on Grassmann manifolds. We perform optimization with Newton's root-finding method by applying it to the derivative of the twice differentiable function $f$ to find the critical points.

## 4.1 Gradient Descent

Our objective is to minimize $F : G(p, n) \to \mathbb{R}$.
In the given algorithm $\delta$ stands for learning rate and $Q = \big(U, V\big)$

**Algorithm 1** Gradient Descent method for minimizing $F(Y)$ on $G_1(p, n)$

```
// Input: F(·) and the initial choice of Y such that Y^T Y = I_p
// Output: First p columns of Q whose span is the minimal subspace
procedure MINIMIZE
    while ||B|| < ε do                                   ▷ We define a stopping criteria
        Compute the directional derivative B and get the tangent Δ

        Update Q_{k+1} = Q_k exp{δΔ} such that f(U_{k+1}) > f(U_t)
    return Q[:, : p]
```

## 4.2 Newton 1

We have $F : G_1(p,n) \to \mathbb{R},\ \ F(Y) = F(YQ),\ Y \in G_1(p,n),\ Q \in O(p), Y^TY = I_p$

**Algorithm 2** Newton's method for minimizing $F(Y)$ on $G_1(p,n)$

// Input: $F(\cdot)$ and the initial choice of $Y$ such that $Y^TY = I_p$
// Output: $Y$ for which $F(Y)$ gives the minimum value
**procedure** MINIMIZE
  **while** $numSteps--$ **do** ▷ We have to choose the number of steps, or define some stopping criteria
    $G = F_Y - YY^TF_Y$
    $\Delta = -Hess^{-1}G$ such that $Y^T\Delta = 0$ and $F_{YY}(\Delta) - \Delta(Y^TF_Y) = -G$

    Move from $Y$ in the direction $\Delta$ to $Y(1)$ using the formula
    $Y(t) = YV\cos(\Sigma t)V^T + U\sin(\sigma t)V^T$ ▷ $U\Sigma V^T$ is SVD of $\Delta$
  **return** $Y$

## 4.3 Newton 2

For this one we define $F : Sym(n) \to \mathbb{R}$ and $f : G_2(p,n) \to \mathbb{R}$ s.t. $f = F|_{G(p,n)}$. $ad_p(X) = [P, X] = PX - XP$ $M_Q$ subscript means that we are taking only $Q$ part from the $QR$ decomposition of $M$

**Algorithm 3** Newton's method for minimizing $F(M)$ on $G_2(p,n)$

// Input: $F(\cdot)$ and the initial choice of $M$ such that $M^T = M, M^2 = M, TrM = p$
// Output: $M$ for which $F(M)$ gives the minimum value
**procedure** MINIMIZE
  **while** $numSteps--$ **do** ▷ We have to choose the number of steps, or define some stopping criteria
    Solve
    $ad^2_M Hess_F(M)(ad_M\Omega) - ad_M ad_{\nabla F(M)} ad_M\Omega = -ad^2_M \nabla_F(M)$
    for $\Omega \in skew\_sym(n)$

    Solve
    $M = \Theta^T \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \Theta$ ▷ $\Theta$ is orthonormal
    for $\Theta \in SO_n$

    $M = \Theta^T(\Theta(I - ad^2_M\Omega)\Theta^T)_Q \Theta M \Theta^T (\Theta(I - ad^2_M\Omega)\Theta^T)^T_Q \Theta$
  **return** $M$

# Chapter 5

# Minimize Rayleigh Quotient

The Rayleigh quotient for a given symmetric matrix $M$ and a nonzero vector $x$ is defined as

$$R(M, x) = \frac{x^T M x}{x^T x}$$

**Theorem 5.0.1.** *For any given symmetric matrix $M \in \mathbb{R}^{n \times n}$*

$$max_{x \in \mathbb{R}^n : x \neq 0} \frac{x^T M x}{x^T x} \text{ (when } x = \text{"largest" eigenvector of } M)$$

$$min_{x \in \mathbb{R}^n : x \neq 0} \frac{x^T M x}{x^T x} \text{ when } x = \text{"smallest" eigenvector of } M$$

*Proof.* Let $M = Q \Lambda Q^T$ be the spectral decomposition, where $Q = [q_1, \ldots, q_n]$ is orthogonal and $\Lambda = diag(\lambda_1, \ldots, \lambda_n)$ is diagonal with sorted diagonals from large to small. Then for any unit vector $x$,

$$x^T M x x^T (Q \Lambda Q^T) x = (x^T Q) \Lambda (Q^T x) = y^T \Lambda y$$

where $y = Q^T x$ is also a unit vector:

$$||y||^2 = y^T y = (Q^T x)^T (Q^T x) = x^T Q Q^T x = x^T x = 1$$

So the original optimization problem becomes:

$$max_{y \in \mathbb{R}^n : ||y|| = 1} y^T \Lambda y \textbf{ (Lambda diagonal)}$$

To solve this problem write $y = (y_1, \ldots, y_n)^T$. It follows that:

$$y^T \Lambda y = \sum_{i=1}^{n} \lambda_i y_i^2$$

Because $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_n$, when $y_1^2 = 1, y_2^2 = \cdots = y_n^2 = 0$ the objective function attains its minimum value $y^T \Lambda y = \lambda_1$ In terms of the original variable $x$, the maximizer is

$$x^* = Q y^* = Q(\pm e_q) = \pm q_1$$

In conclusion, when $x = \pm q_1$ (largest eigenvector), $x^T M x$ attaints its maximum value $\lambda_1$ (largest eigenvalue) □

In the next subsections we will focus on **Computing the eigenvectors and eigenvalues of a symmetric matrix by minimizing rayleigh quotient**

## 5.1 Gradient Descent on the sphere

We have the following setup:
Compute $min_{x \in S^n} \frac{1}{2} x^T M x$
The cost function $f : S^n \to \mathbb{R}$ is the restriction of $\bar{f} = \frac{1}{2} x^T M x$ from $\mathbb{R}^n to S^n$
Tangent spaces are given by $T_x S^n = \{v \in \mathbb{R}^n : x^T v = 0\}$
To make $S^n$ into a Riemmannian submanifold of $\mathbb{R}^n$ we take a dot product $\langle u, v \rangle = u^T v$
Projection to $T_x S^n$: $Proj_x(z) = z - (x^T z)x$
Gradient of $\bar{f} = \nabla \bar{f}(x) = Mx$
Gradient of $f$: $gradf(x) = Proj_x(\nabla \bar{f}(x)) = Mx - (x^T M x)x$ Thus algorithm becomes

**Algorithm 4** Gradient Descent method for minimizing the Rayleigh quotient

// Input: The initial choice of $Y$ such that $Y^T Y = I_p$ and the choice of learning rate $lr$
// Output: $Y$ for which of $F(Y)$ gives the dominant eigenvalue
**procedure** MINIMIZE
    **for** $i$ in $num\_steps$ **do** ▷ Define a number of steps or a stopping criteria
        if $i$ mod $100 == 0$ then $lr = lr/100$ ▷ Learning rate decay
        $Y = Y - lr \cdot \nabla f(X)$
    **return** $Y$

**Example 5.1.1.** Find the dominant eigenvector and eigenvalue of $A = \begin{pmatrix} 181 & 101 & 146 \\ 101 & 74 & 103 \\ 146 & 103 & 146 \end{pmatrix}$ by using gradient decent on Rayleigh quotient.

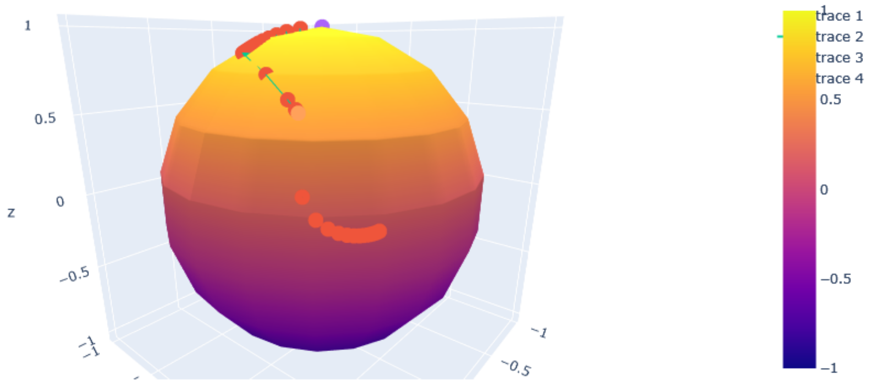



Figure 5.1: Convergence of the GD algorithm on Rayleigh quotient example

## 5.2 Newton 1

Given the function $\bar{f} = x^T M x \quad grad\bar{f} = 2Mx - 2(x^T M x) = 2(I - xx^T)Ax = 2PMx$

$$D(gradf) = 2M - 4xMx\eta + 2x^T Mx$$

Reminder $g_c(\Delta, \Delta) = tr\Delta^T(I - \frac{1}{2}YY^T)\Delta$

$$g_c(D(gradf), \eta) = g_c(2M - 2x^T Mx + 4Mxx^T, \eta) = 2MP\eta - 2\eta x^T Mx$$

$$Pg_c(D(gradf), \eta) = 2MP\eta - 2\eta x^T Mx$$

Therefore we have a Newton iteration:

$$P_x MP_x\eta - \eta xMx = -PMx$$

$P$ is a projection, $R$ is a retraction.

**Algorithm 5** Newton's method for minimizing the Rayleigh quotient

// Input: The initial choice of $Y$ such that $Y^TY = I_p$
// Output: $Y$ for which $F(Y)$ gives the minimum value
**procedure** MINIMIZE
    **while** $numSteps--$ **do** ▷ define some stopping criteria
        $M = P(Y)MP(Y) - x^T Ax$
        $y = -P(Y)Ax$
        $\eta = solve(M, y)$
        $Y = R(Y, \eta)$
    **return** $Y$

**Example 5.2.1.** Find eigenvectors and eigenvalues of $A = \begin{pmatrix} 181 & 101 & 146 \\ 101 & 74 & 103 \\ 146 & 103 & 146 \end{pmatrix}$ using Newton 1

OUTPUT: $[-0.74148822, 0.44835462, 0.49917267]$

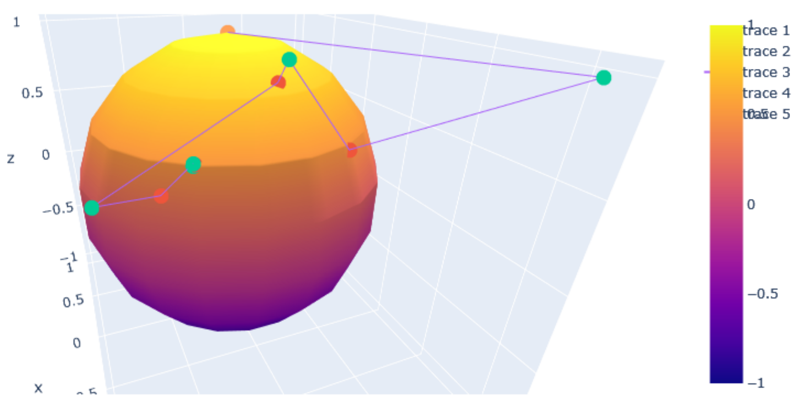



Figure 5.2: Convergence of the Newton 1 algorithm on Rayleigh quotient example

## 5.3 Newton 2

For the Rayleigh quotient the equation that we need to solve in the first step of 4.3 becomes:

$$-ad_{P_j} ad_A ad_{P_j} \Omega_j = -ad^2_{P_j} A = \Theta_j (ad_{P_j} ad_A ad_{P_j} \Omega_j)\Theta_j^T = \Theta_j (ad^2_{P_j} A)\Theta_j^T$$

$$P_j = \Theta_j^T \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \Theta_j$$

$$ad_{\begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix}} ad_{\Theta_j A \Theta_j^T} ad_{\begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix}} \begin{bmatrix} 0 & Z_j \\ -Z_j^T & 0 \end{bmatrix} = ad^2_{\begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix}} (\Theta_j A \Theta_j^T)$$

for $Z_j \in \mathbb{R}^{m\times(n-m)}$. Denoting

$$\Theta_j A \Theta_j^T = \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^T & A_{22} \end{bmatrix}$$

so we just have to solve the Sylvester equation

$$A_{11} Z_j - Z_j A22 = A_{12}$$

**Algorithm 6** Newton's method for minimizing Rayleigh quotient on $G_2(1,3)$

```
// The initial choice of Θ ∈ SO(n)
// Output: Θ whose first p are the eigenvector
procedure MINIMIZE
    while numSteps −− do                                    ▷ define some stopping criteria
        Compute [A11 A12; A12^T A22] = Θ_j A Θ_j^T
        Solve the Sylverster equation A11 Z_j − Z_j A22 = A12 for Z_j ∈ R^{m×(n−m)}

        Compute Θ_{j+1}^T = Θ_j^T [I_m Z_j; −Z_j^T I_{n−m}]_Q and P_{j+1} = Θ_{j+1}^T [I_m 0; 0 0] Θ_{j+1}
    return M
```

**Example 5.3.1.** Find eigenvectors and eigenvalues of $A = \begin{pmatrix} 181 & 101 & 146 \\ 101 & 74 & 103 \\ 146 & 103 & 146 \end{pmatrix}$ using Newton 2

OUTPUT: $[-0.74364137, 0.43442584, 0.5082044]$

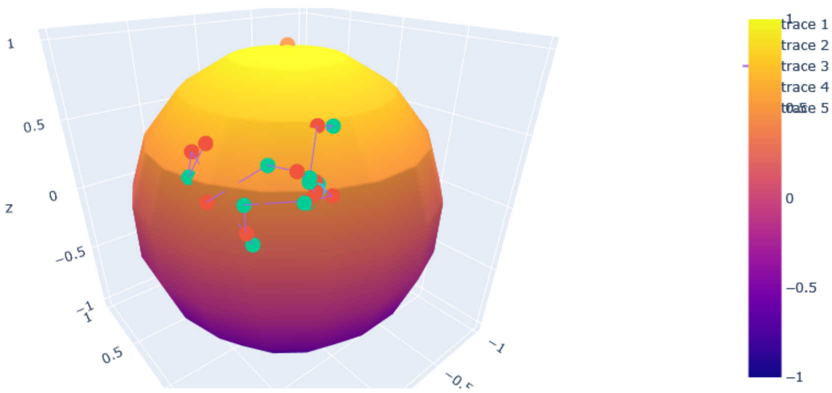



Figure 5.3: Convergence of the Newton 2 algorithm on Rayleigh quotient example

# Chapter 6

# Notation

| Symbol | Matrix Definition | Name |
|---|---|---|
| $F(p,n)$ | $\{A \in Mat_{n\times p} \mid rkA = p\}$ | 2 frames |
| $St(p,n)$ | $\{A \in F(2,4) \mid A^T A = I_k\}$ | Stiefel Manifold |
| $GL(p)$ | $\{A \in Mat_{n\times n} \mid detA \neq 0\}$ | General Linear Group |
| $O(p)$ | $\{Q \in GL(p) \mid Q^T Q = I\}$ | Orthogonal group |
| $\mathfrak{so}_n$ | $\{\Omega \in \mathbb{R}^{n\times n} \mid \Omega^T = -\Omega\}$ | Real skew-symmetric matrices |
| $Sym(p)$ | $\{A \in \mathbb{R}^{p\times p} \mid A^T = A\}$ | Symmetric matrices |
| $SO(n)$ | $\{Q \in O(n) \mid \det(Q) = 1\}$ | Special Orthogonal Group |

# Bibliography


[ACW12] Kofi Placid Adragni, R. Dennis Cook, and Seongho Wu. GrassmannOptim: An R Package for Grassmann Manifold Optimization. *Journal of Statistical Software*, 50:1–18, July 2012.

[AMS08] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, NJ, 2008.

[BBCV21] Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Velickovic. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *CoRR*, abs/2104.13478, 2021.

[Bou22] Nicolas Boumal. An introduction to optimization on smooth manifolds. To appear with Cambridge University Press, Apr 2022.

[EAS98] Alan Edelman, T. A. Arias, and Steven T. Smith. The Geometry of Algorithms with Orthogonality Constraints. *arXiv:physics/9806030*, June 1998. arXiv: physics/9806030.

[HHT07] Uwe Helmke, Knut Hüper, and Jochen Trumpf. Newton's method on Gra{\ss}mann manifolds. *arXiv:0709.2205 [math]*, September 2007. arXiv: 0709.2205.

[HL] Jihun Hamm and Daniel D Lee. Grassmann Discriminant Analysis: a Unifying View on Subspace-Based Learning. page 10.

[HWVG18] Zhiwu Huang, Jiqing Wu, and Luc Van Gool. Building Deep Networks on Grassmann Manifolds. *arXiv:1611.05742 [cs]*, January 2018. arXiv: 1611.05742.

[Joh21] Kerstin Johnsson. Optimization over Grassmann manifolds. page 9, 2021.

[LLY20] Zehua Lai, Lek-Heng Lim, and Ke Ye. Simpler Grassmannian optimization. *arXiv:2009.13502 [math]*, September 2020. arXiv: 2009.13502.

[MBBV18] Jonathan Masci, Davide Boscaini, Michael M. Bronstein, and Pierre Vandergheynst. Geodesic convolutional neural networks on Riemannian manifolds. *arXiv:1501.06297 [cs]*, June 2018. arXiv: 1501.06297.

[MKP20] Xiaofeng Ma, Michael Kirby, and Chris Peterson. The flag manifold as a tool for analyzing and comparing data sets. *arXiv:2006.14086 [cs, math]*, June 2020. arXiv: 2006.14086.

[MT88] Jerrold E. Marsden and Anthony. Tromba. *Vector calculus /*. W.H. Freeman,, New York :, 3rd ed. edition, c1988. Includes index.

[TFBJ18] Nilesh Tripuraneni, Nicolas Flammarion, Francis Bach, and Michael I. Jordan. Averaging Stochastic Gradient Descent on Riemannian Manifolds. *arXiv:1802.09128 [cs, math, stat]*, June 2018. arXiv: 1802.09128.

[TKW16] James Townsend, Niklas Koep, and Sebastian Weichwald. Pymanopt: A python toolbox for optimization on manifolds using automatic differentiation. *Journal of Machine Learning Research*, 17(137):1–5, 2016.

[ZZHJH18] Jiayao Zhang, Guangxu Zhu, Robert W. Heath Jr., and Kaibin Huang. Grassmannian Learning: Embedding Geometry Awareness in Shallow and Deep Learning. *arXiv:1808.02229 [cs, eess, math, stat]*, August 2018. arXiv: 1808.02229.