SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

On special orthogonal group and variable selection

av

Iris van Rooijen

2016 - No 9

On special orthogonal group and variable selection

Iris van Rooijen

Självständigt arbete i matematik 30 högskolepoäng, avancerad nivå

Handledare: Yishao Zhou

2016

Abstract

We will look at Euler angle representation of rotations, and the even subalgebras in Clifford algebra which form a double cover of the special orthogonal group. Some of their properties and metrics will be compared and some suggestions made on how they could be applied in Nordlings variable selection system for high dimensions. This will hopefully serve as an introduction to create interest in the subject and shed some light the difficulties that occur at different stages.

Contents

1 Introduction

In this report we want to explore the properties of SO(n), the special orthogonal group in n dimensions a.k.a. n-dimensional rotations, and their possible application on Nordlings variable selection system [Nor13].

Although rotations in (2 and) 3 dimensions have been extensively studied due to its application in robotics[Wil05], aerodynamics[Den] and many other fields, literature on rotations in higher dimensions quickly becomes much more scarce. However, there are already huge differences between rotations in 3 dimensions compared to rotations in 4 dimensions [Col90] [TPG15] [Wil05] [Lou01]. Rotations in higher dimensions seems even less explored.

Rotations can be represented in many different ways, such as a rotation matrix using Euler angles [Can96], or as (multiple) pairs of reflections using Clifford algebra [Wil05] [Den] [Lou01], or quaternions in 3 and 4 dimensions [TPG15] [Col90] [Lou01].

The Euler angle representations also have different ways of measuring the angle of rotation. Which could be Euclidean distance [Huy09], or using the Lie algebra of the Euler angle rotation matrix [PR97], or its Haar measure [Not][Tay].

As for Clifford algebra, the metrics found are restricted to quaternions in 3 dimensions [Huy09], or they apply only to vectors.

Hopefully this report will shed some light on the possibilities, difficulties and impossibilities that occur when trying to add a rotation matrix into Nordlings variable selection system. The subjects lifted in this report could be tools with different advantages when applying them to Nordlings variable selection system.

The first section of this thesis will deal with Nordlings variable selection system, to show why this report on rotations is motivated, and to narrow down on some more specific properties of rotations. The second section explores the different representations and metrics of rotation matrices, and some comparison between them. In the third section some applications of rotation matrices on Nordlings variable selection system are suggested. While the fourth section deals with problems that occur, and discusses how to continue research in this area.

2 About variable selection and linear regression

2.1 Defenitions of variable selection

Using linear regression it is possible to find relationships between data. Variable selection methods can be used to improve the linear regression method in several ways. According to Elisseeff [EG03] there are 3 types of objectives for variable selection, each of these fulfill a different role for the results.

Improving prediction performance
: Making a prediction based on all the features can lead to biasing (when many features say the same thing), or basing output on irrelevant data.

Computational
: Evidently, when decreasing the number of input variables for an algorithm, the algorithm will work faster.

Understanding of data
: Understanding a problem (or solution) in thousand dimensions is basically impossible. The fewer dimensions, the easier it might be to plot or visualize the underlying data and to understand the predictions.

The two goals for feature selection are: reconstructing one's data as well as possible (unsupervised), or being as good as possible at predicting (supervised).

They are often used as machine learning tecniques.

2.1.1 Optimisation versus robustness

There are two different goals of variable selection.

Optimisation: We are looking for the model which best explains the relation between input data and output data that we have got. However, even though we find a perfect match, this does not mean that we have necesarily found the correct relation between the input data and the output data.

In robust variable selection the aim is not to find the single model which best describes the input and output, but to find the set of models which are possible, and which are not, by looking at which properties must be considered to predict the outcome, and which properties can always be neglected.

2.1.2 Some different techniques

Before looking at a relation between the regressand and the regressor one can start reducing computations using one of the following techniques:

Ranking can be used as a pre-selection technique. It is used to sort out some special qualities either wanted or not wanted in further computation, e.g. one might throw all rows j (data from experiment j) with

only 0 entries, or only choose those 20 rows j with highest value in the first entry of the regressors. Here the number of experiments n is reduced.

Similarly one could throw (or only choose) properties which do (not) seem to effect the prediction. However a property which seems useless on itself, can be very useful together with others. Here the number of properties m is reduced.

Dimensionality reduction can also be used as a pre-selection technique, to see whether two properties seem to depend on each other very much, e.g. if one property denoted degrees Celsius and another degrees measured in Farenheit. Here the number of properties m is reduced. Another method is called *clustering*, in which several variables (experiments), e.g. with very similar data, are merged into one. Here the number of properties n is reduced.

It is worth noting that mathematically, clustering could seem like a good idea, but if the data should be explanatory one might want to use caution since it would become unclear exactly where the data originated from.

A SVM, Support vector machine, can be used to simplify linear regression and to find out which variables are important, by trying to find a linear function of the variables with smallest combined length to the data. Data which is too far from the function will then be removed, and the process repeated.

These different techniques often use stochastic theory, e.g. t-test, Paersons correlation etc., to decide whether some variable should be selected or not. This, however, might require some presumtions about the model, e.g. whether the data is spread with a normal distribution, which is not always desirable.

2.1.3 Validation of predictability or robustness

After computing a possible solution one usually wants to check its predictive properties. This could be done e.g. through checking how well future data fits ones predictions. However in most cases this might not be an option, yet one still has to be pretty certain that any prediction is correct. One way to solve this is to divide ones data into a set of training examples and a validation set. Another way to solve it is to 'create' new data using the data one has.

Choosing what fraction to use, or even which technique to use is an open problem. It could depend on e.g. whether some experiments have given the same results. Examples of techniques to use are: Bootstrapping and (leave-one-out) cross-validation.

2.2 Introduction to variable selection using linear regression

In variable selection using linear regression one has two sets of measured data represented as the two matrices regressor $\mathbf{X} \in \mathbb{R}^{n \times m}$, also called independent variable or input variable, and regressand $\mathbf{Y} \in \mathbb{R}^{n \times j}$, also called dependent variable or output variable.

The data in row i of $\mathbf{X}$ could e.g. correspond to the m property values measured in experiment $i \in [1, n]$, and so the data in each column k of $\mathbf{X}$ corresponds to the values the experiments had for property k. The matrix $\mathbf{X}$ is called a *regressor*, and its columns are *regressors*.

The data in row i of $\mathbf{Y}$ could e.g. corresponds to the measured outcome of some experiment $i \in [1, n]$. The matrix $\mathbf{Y}$ is called a *regressand*, and when $j = 1$ the equation 1 below is said to be univariate, else, when $j > 1$ it is called multivariate.

Note however that the data does not need to be structured this way. It is possible one measures all the data at once and one divides it into the sets $\boldsymbol{X}$ and $\boldsymbol{Y}$. These sets need not even be defined beforehand, but one can try dividing the data into different sets and compare the results and draw conclusions afterwards.

In *linear regression* one tries to find the matrix $\mathbf{A} \in \mathbb{R}^{m \times j}$ that solves the following equation:

$$\mathbf{XA} = \mathbf{Y} \tag{1}$$

Before using this equation one can first use the ranking technique eliminating those columns $\mathbf{X_k} = \mathbf{0}$, i.e. those variables about which we don't know how they will effect the outcome of $\mathbf{Y}$ because we have no experience to base that decision on. Unless otherwise stated, we will assume this is done for the remainder of the text.

Hence we have selected the columns $\mathbf{X_k} \neq \mathbf{0}$, and put them in the above equation (1). Assuming the equation does not contain contradictions, the following can be said about solving it: (1) If $n < m$, the equation can not be uniquely solved. (2) If $n = m + i$, where $i \geq 0$, the equation can only be uniquely solved if less than i rows of $\mathbf{X}$ are collinear, i.e. linearly dependent. If more than i rows of $\boldsymbol{X}$ are linearly dependent, Gauss elimination would put us case (1).

If we want to check whether columns $\mathbf{X_k}$ are *collinear* we check whether there is a vector $\boldsymbol{b} = (b_1, ..., b_m) \neq \mathbf{0}$ such that:

$$\sum_{k=1}^{m} \mathbf{X_k} b_k = 0 \tag{2}$$

In practice this means that one has no means of telling whether one or the other of two (or more) collinear columns, properties, is the one that predicts the outcome, or perhaps a combination of both.

2.3 Linear regression and uncertainty

Next will be some basic defenitions from Nordling [Nor13] to explain what happens when one introduces uncertainty to the linear regression system. The theorems and results, however, are defined later individually with an explanation of how a rotation matrix R (might) effect the results. I chose to present Nordlings results the following way: The number is the number in which it is found in his text or the page number of his text where the result can be found, but, for consistency, changed notation to match my own.

A first note on the uncertainty of linear regression is that one assumes that the regressor $\boldsymbol{X}$ and regressand $\boldsymbol{Y}$ are well defined and known. However, it is also possible one recieves a large set of data (columns) and need to pick which columns create the regressor and which columns create the regressand. In other words, the choice of columns for $\boldsymbol{X}$ and $\boldsymbol{Y}$ needs not be obvious.

In the Nordling system we will usually assume that the number of columns n is much larger than the number of rows m.

2.3.1 Properties of a rotation matrix

Recall, a rotation matrix $R \in \mathbb{R}^{n\times n}$ is an orthogonal matrix, and so has the following properties:

- $|R\overline{x}| = |\overline{x}|$ i.e. R preserves length.
- $\langle R\overline{x}|R\overline{y}\rangle = \langle \overline{x}|\overline{y}\rangle$ i.e. R preserves angles.
- Columns u_k of R and rows v_k of R are orthonormal, e.g. $u_i \cdot u_j = u_{i1} \cdot u_{j1} + ... + u_{in} \cdot u_{jn} = 0$.
- $R^T R = I = RR^T$ i.e. R's transpose is its inverse.

It will be used in the following equation

$$\overline{\boldsymbol{X} + \boldsymbol{u}}\boldsymbol{A} = R\boldsymbol{Y} + \boldsymbol{v}$$

whose terms will be explained throughout this section.

2.3.2 Short on Nordling's system

The basic idea is that in practice it is very unlikely that we find the exact values of the properties of the experiments j, i.e. $\boldsymbol{X_j}$ and $\boldsymbol{Y_j}$. Hence Nordling uses uncertainty measures to compensate for the inexactness. The technique itself is not new, however, he noted that deterministic uncertainty measures for the values in the regressor $\boldsymbol{X}$ had not been used before, though they are assumed in filed studying. Uncertainty can be deterministic or stochastic giving two slightly different answers. The stochastic version will be defined, however, most of the report handles the deterministic case.

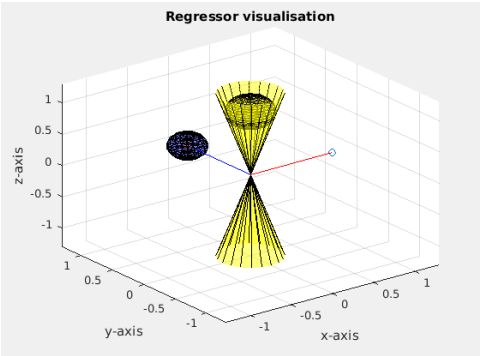


Figure 1: Some different representations of X_i
Here are some different examples of how to represent and look at a regressor. For example $X_1 = (1, 0, 0)$ (red) is a simple vector when one would disregard of any uncertainty, $X_2 = (0, 1, 0)$ (blue) can be seen as the measured vector with a(n $m - 1$-) sphere surrounding it within which all possible reallisations of X_2 are which can not be discarded. For $X_3 = (0, 0, 1)$ (yellow) two additional cones are added representing the points which can be reached by the possible reallisations.

The uncertainty of the regressor and regressand can be described as a ball or n-rectangle around the measured value.

The thought is to describe the *uncertainty of* $\mathbf{X_i}$ as a closed neighbourhood ball of radius u_i, after that we can create a so called uncertainty cone, representing all the possible points that can be explained using only $\mathbf{X_i}$.

Nordling Definition. (p.99 + p.109) *Given* $\mathbf{X_i} \in \mathbb{R}^m$*, row* i *of the regressor, measured with a given uncertainty* $u_i \in \mathbb{R}$*, the deterministic uncertainty set is the neighbourhood ball* $\mathcal{N}(\mathbf{X_i}, u_i)$ *of radius* u_i *around* $\mathbf{X_i}$*, where each vector within the neighbourhood* $\mathcal{N}(\mathbf{X_i}, u_i)$ *is a candidate to be the true value.*

I will denote a vector in $\mathcal{N}(\mathbf{X_i}, u_i)$ as $\overline{\mathbf{X_i} + \mathbf{u_i}}$, and $\overline{\mathbf{X} + \mathbf{u}}$ stands for a matrix where each row i is in the neighbourhood $\mathcal{N}(\mathbf{X_i}, u_i)$.

Nordling Definition. *Given a deterministic uncertainty set* $\mathcal{N}(\mathbf{X_i}, u_i)$*, the uncertainty cone* $\mathcal{C}(\boldsymbol{X_i})$ *of* $\boldsymbol{X_i}$ *is*

$$\mathcal{C}\boldsymbol{X_i} = \{t\mathbf{X_i}' | \mathbf{X_i}' \in \mathcal{N}(\mathbf{X_i}, u_i), t \in \mathbb{R}\}$$

Note: The vectors on the boundary of the cone can only be represented in one way, but a point closer to the 'center' of the cone is represented several times in $\mathcal{C}(\boldsymbol{X_i})$.

Note: Since t can be both positive and negative $\mathcal{C}(\boldsymbol{X_i}, u_i)$ will actually be shaped like two cones reflecting each other through the origin.

Nordling Definition. (p.101) *Given a regressor* $\mathbf{X} \in \mathbb{R}^{n \times m}$*, a significance level* α*, and uncertainty values* δ_{ij} *for* X_{ij}*. Let* $\Delta \in \mathbb{R}^{nm \times nm}$ *be the covariance matrix of* $\boldsymbol{X}$ *and let* $\Gamma = [u_{11}, u_{12}, ..., u_{1m}, u_{21}, ..., u_{nm}] \in \mathbb{R}^{1 \times nm}$ *be*

the vector containing all the uncertainties for all rows appended after each other. Then the stochastic uncertainty set $\mathcal{N}^s$ is defined as

$$\mathcal{N}^s(\mathbf{X_k}, \boldsymbol{u_k}) = \{\mathbf{X_k} + \mathbf{u_k} | \mathbf{u_k} \in \Gamma, \Gamma\Delta^{-1}\Gamma^T \leq \chi^{-2}(\alpha, nm)\} \tag{3}$$

When Δ is diagonal we get the stochastic uncertainty set $\mathcal{N}^s(\mathbf{X_i}, u_{ij})$ where $u_{ij} \leq \delta_{i1}\chi^{-2}(\alpha, nm)$ when $\delta_{i1} = \delta_{i2} = ... = \delta_{im}$. When Δ is diagonal, i.e.

$$\Delta = \text{diag}([\delta_{11}, \delta_{21}, ..., \delta_{n1}, \delta_{21}, ..., \delta_{nm}]) \tag{4}$$

we have $\Gamma\Delta\Gamma^T = \sum_{j=1}^{nm} u_j^2\delta_j$. And when $\delta_{1k} = \delta_{2k}, = ... = \delta_{nk} := \Delta_k$ we get $\Gamma\Delta\Gamma^T = \sum_{k=1}^{m} \Delta_k \sum_{i=1}^{n} u_{ik}^2$, so the neighbourhoods can be seen as m weighted balls with weight Δ_k, whose combined value is at most $\chi^{-2}(\alpha, nm)$.

However Δ does not need to satisfy this condition, and generaly does not. When it is diagonal the uncertainty will be weighted ellipsoids, and when it is not diagonal it is much more difficult to see how the uncertainty of the respective regressors effect each other.

In contrast to the deterministic case, the stochastic uncertainty of one row i depends on the 'realisations' of all other rows as well.

Here, a neighbourhood ball is created containing all the possible outcome which can not be rejected with significance level α. However, within this ball there are vectors which can be rejected as solution depending on the choice of the other $\mathbf{X_j} + \mathbf{u_j} \in \mathcal{N}^s(\mathbf{X_j}, u_j)$. But the closer to X_i one gets, the higher the probability that the vector can not be rejected. We could pick two (or more) $\mathbf{X_i} + \mathbf{u_i}$ and $\mathbf{X_j} + \mathbf{u_j}$ which would be in their respective neighourhood, however, combined, the uncertainty of the two (or more) regressors $\mathbf{X_i}$ and $\mathbf{X_j}$ is too large.

Similarly to the deterministic case, we create the cone the following way (Note: the cones are not defined ly in TN):

Nordling Definition. *Given a stochastic uncertainty set $\mathcal{N}^s(\mathbf{X_i}, u_i)$, the stochastic uncertainty cone $C^s(\boldsymbol{X_i})$ is*

$$C^s(\boldsymbol{X_i}) = \{t\mathbf{X_i}' | \mathbf{X_i}' \in \mathcal{N}(\mathbf{X_i}, u_i), t \in \mathbb{R}\}$$

In the stochastic case, the size of the cones will also depend on each other, since the uncertainty sets depend on each other.

Nordling also provides a different way to represent uncertainty, which would be more rectangular. We will define it mostely to illustrate the difference in difficulty when rotating the rectangular uncertainty compared to the circular one.

The other way to describe the uncertainty of the measured values is by giving each value separate uncertainty, resulting in a rectangular uncertainty space. Here too we will give an example using the regressor, however this technique applies just as well to the regressand.

For the regressor $\mathbf{X}$, the *uncertainty* of $\mathbf{X_i}$ as a closed neighbourhood hyperrectangle of m dimensions, with uncertainty lengths described as vector $\mathbf{v_i} = 2(v_{i1}, v_{i2}, ..., v_{im})$. After that we can create an uncertainty cone, representing all the possible points that can be explained using only $\mathbf{X_i}$.

Nordling Definition. (p.99) *Given $\mathbf{X_i} \in \mathbb{R}^m$, row i of the regressor, measured with a given uncertainty $\mathbf{v_i} \in \mathbb{R}^m$, the deterministic uncertainty set is the neighbourhood hyperrectangle $\mathcal{N}(\mathbf{X_i}, \mathbf{v_i})$ of lengths $\mathbf{v_i}$ around $\mathbf{X_i}$, where each vector within the neighbourhood $\mathcal{N}(\mathbf{X_i}, \mathbf{v_i})$ is a possible candidate to be the true value.*

Nordling Definition. *Given a deterministic uncertainty set $\mathcal{N}^s(\mathbf{X}, v)$, the deterministic uncertainty cone C_i of X_i is*

$$C_i = \{t\mathbf{X} | \mathbf{X} \in \mathcal{N}(\mathbf{X_i}, v_i), t \in \mathbb{R}\}$$

Now we have a good way to describe the uncertainty of the rows of the regressors and regressands. Similarly one could represent the columns of $\overline{\boldsymbol{X} + \boldsymbol{u}}$ and $\overline{\boldsymbol{Y} + \boldsymbol{v}}$ this way. We would like to find out which columns of the regressor are necesary to describe a column in the regressand.

Given these definitions of uncertainty the definition of a valid *feasible solution* is the following:

Nordling Definition. 5.5.1. *A parameter matrix $\mathbf{A} \triangleq [A_1^T, ..., A_j^T, ..., A_n^T]^T$ is feasible if*

$$\sum_j A_j \overline{\mathbf{X} + \mathbf{u}} = \overline{\mathbf{Y} + \mathbf{v}}$$

In other words a solution $\mathbf{A}$ is feasible if any combination of the cones $\overline{\mathbf{X} + \mathbf{u}}$ can intersect the hyperrectangle $\overline{\mathbf{Y} + \mathbf{v}}$. A point $\overline{\mathbf{Y} + \mathbf{v}}$ in the hyperrectangle has a solution if there exist $\overline{\mathbf{X} + \mathbf{u}}$ and $\boldsymbol{A}$ such that $\overline{\boldsymbol{X} + \boldsymbol{u}}\boldsymbol{A} = \boldsymbol{Y}$.

If a row j of $\boldsymbol{A}$ equals $\mathbf{0}$, that means column j of regressor $\overline{\boldsymbol{X} + \boldsymbol{u}}$ is not needed to get the values in $\overline{\boldsymbol{Y} + \boldsymbol{v}}$. In other words, property j is not needed to explain the values in the regressand.

As for the version with rotation matrix, the definition with the unknown $c \in \mathbb{R}$ and $\mathbf{R} \in \mathbb{R}^{m \times m}$ looks as follows:

Definition 1. *A parameter vector $\mathbf{A} \in \mathbb{R}^n$ and the parameters $c \in \mathbb{R}$ and rotation matrix $\mathbf{R} \in \mathbb{R}^{m \times m}$ are feasible if*

$\sum_{j \in \mathcal{V}} \boldsymbol{A_j} \overline{\boldsymbol{X_j} + \boldsymbol{u_j}} = c\mathbf{R}\overline{\boldsymbol{Y} + \boldsymbol{v}}$ *for some consistent* $\overline{\boldsymbol{X} + \boldsymbol{u_j}} \in \mathcal{U}_{X_j}^{\alpha} \subseteq \mathbb{R}^m$ *and*

$$\overline{\boldsymbol{Y} + \boldsymbol{v}} \in \mathcal{U}_Y^{\alpha} \subseteq \mathbb{R}^m$$

In other words a solution is feasible if there is a rotation matrix $\mathbf{R}$, and a scaling vector c, such that the regressand Y can be rotated, and scaled, into the space of X.

Another way to describe feasible solution is by first defining the *practical span*, i.e. all the possible points that can be reached by the uncertainties of $\mathbf{X}$.

Nordling Definition. (5.5.11.) *The practical span of the set of uncertain vectors in the matrix* $\mathbf{X} = [X_1, ..., X_n]$ *is*

$$pspan\mathbf{X} \triangleq \{\sum_{i=1}^{n} a_i\overline{\mathbf{X_i} + \mathbf{u_i}} | a_i \in \mathbb{R}, \overline{\mathbf{X_i} + \mathbf{u_i}} \in \mathcal{N}(\mathbf{X_i}, u_i)\} \tag{5}$$

Giving the following definition of feasible solution:

Nordling Definition. (5.5.2) *A solution practically exists if and only if* $\overline{\boldsymbol{Y} + \boldsymbol{v}} \in$ *pspan* $\overline{\boldsymbol{X} + \boldsymbol{u}}$ *for* $\overline{\boldsymbol{X} + \boldsymbol{u}} \in \mathcal{N}(\boldsymbol{X}, \boldsymbol{u}) \subseteq \mathbb{R}^m$.

This definition coincides with Nordlings definition of *practical uniqueness* (Defenition 5.5.12., Theorem 5.5.3.)[Nor13] when the uncertainty of the regressand is compact. When we introduce a rotation matrix into the system, we can easily see that we can always rotate the regressand $\boldsymbol{Y}$ into or out of the *pspace* created by any set of regressors $\boldsymbol{X_i}$. This should illustrate the need of restrictions on the rotation matrix.

Some other of Nordlings definitions are not effected by a rotation matrix, or only partially. We look at independence and collinearity of the rows $\mathbf{X} + \mathbf{u}$ as well as when a regressor $\mathbf{X_i}$ is neglectable, and how a rotation matrix would effect it.

Nordling Definition. (5.5.13) *The matrix* $\mathbf{X} = [X_1, ..., X_i, ..., X_n]$ *is practically (linearly) independent* $\forall\overline{\mathbf{X_i} + \mathbf{u_i}} \in \mathcal{N}(X_i, u_i) \subseteq \mathbb{R}^m$ *the trivial solution* $\mathbf{B} = \mathbf{0}$ *is the only solution of*

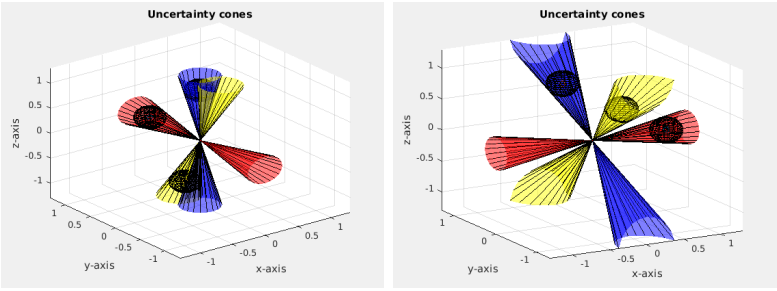
$$\sum_{i=1}^{n} B_i\overline{\mathbf{X_i} + \mathbf{u_i}} = \mathbf{0} \tag{6}$$

Since R is a rotation matrix, it will not change the internal structure between the regressors. For any column ϕ_k we will get a matrix with (sums of the entries) entries $R_{ij}\theta_k\phi_{jk}$ for each row j. Now we can easily see that if $\theta = \mathbf{0}$ then $R_{ij}\theta_k\phi_{jk} = 0\ \forall i, j, k$, and similarly $\theta \neq \mathbf{0}$ then there is at least one $R_{ij} \neq 0 \leftarrow R_{ij}\theta_k \neq 0\ \forall i, jk$,.

For *collinearity* Nordling has the following definition

Nordling Definition. (5.5.14) *The matrix* $\mathbf{X} = [\boldsymbol{X}_1^T, ..., \boldsymbol{X}_i^T, ..., \boldsymbol{X}_n^T]^T$ *is practically collinear, or practically (linearly) dependent, if* $\forall\phi_k \in \mathcal{U}_{\phi_k}^{\alpha}$ *of some row* $\boldsymbol{X}_i$ *with* $i \in \{1, 2, ..., n\}$ *s.t.* $\exists\mathbf{A} = [A_1, ..., A_i, ..., A_n]^T \neq \mathbf{0}$ *to*

$$\sum_{i=1}^{n} A_i\boldsymbol{X}_i = \mathbf{0}$$



(a) collinear regressors (b) Independent regressors

Figure 2: Example of collinear and independent regressors
The cones in a) illustrate example 1 and are collinear to each other. The blue $\boldsymbol{X_2}$ and yellow $\boldsymbol{X_3}$ intersect, however $\boldsymbol{X_1}$ also lies on the y-z-plane with the same uncertainty and is collinear with the other two, but this is harder to see. It would be necesary to get a better knowledge of *pspace*, which will be explained in a later section. In b) all the regressors are independent, X1 (yellow, (0, -cos(pi/4), sin(pi/4)), X2(blue (0, cos(pi/4), sin(pi/4)) and X3 (red, (1,0,0)). All regressors have an uncertainty of 0.2.

Similar reasoning can be used to conclude that collinearity remains unchanged when the regressand is rotated. In fact the definitions of independence and collinearity are independent of the regressand $\boldsymbol{Y}$.

Example 1: Let $\boldsymbol{X_1} = (0, 1, 0)$, $\boldsymbol{X_2} = (0, 0, 1)$ and $\boldsymbol{X_3} = (0, \cos(\pi + 0.3), \sin(\pi + 0.3))$, all with uncertainty 0.2 as seen in figure 2 (a). Any point in the uncertainty set of $\boldsymbol{X_3}$ can be written as $k(A, \cos\theta, \sin\theta)$ for $|A| \leq |0.2|$, $k \in \mathbb{R}$. Let $s_1, s_2, s = \pm 1$ such that $s_1 sign(\cos\theta) = s = s_2 sign(\sin\theta) = sign(A)$.

$$(\cos(\theta))(\boldsymbol{X_1} + s_1(0.2, 0, 0)) + (\sin(\theta))(\boldsymbol{X_2} + s_2(0.2, 0, 0)) = \\ ((\cos\theta + \sin\theta)s0.2, \cos(\theta), \sin(\theta))$$

Hence we can see that the uncertainty regressor $\boldsymbol{X_3}$ is collinear with $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$. However this might not be completely obvious from figure 2 (a). In a later section we will introduce a way to visualize *pspace*, and in the appendix is another visualization is suggested.

However, if u_3 is suffictiently large, while the uncertainty of the other two regressors remains the same, $\boldsymbol{X_3}$ might not be collinear with $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$, while $\boldsymbol{X_1}$ would still be collinear with $\boldsymbol{X_2}$ and $\boldsymbol{X_3}$ for example.

Example 2: Now let $\boldsymbol{X_1} = (1, 0, 0)$, $\boldsymbol{X_2} = (0, \cos(\frac{\pi}{4}), \sin(\frac{\pi}{4}))$ and $\boldsymbol{X_3} = (0, -\cos(\frac{\pi}{4}), \sin(\frac{\pi}{4}))$, and let the regressors have uncertainty $u = (u_1, u_2, u_3), v = (v_1, v_2, v_3), w = (w_1, w_2, w_3)$ respectively, of length ≤ 0.2. To show that they are independent we want to show that B, C are a solution to $\overline{\boldsymbol{X_1} + \boldsymbol{u}} + B\overline{\boldsymbol{X_2} + \boldsymbol{v}} + C\overline{\boldsymbol{X_3} + \boldsymbol{w}} = \boldsymbol{0}$ only when $B = C = 0$. We get the

equations

$$(1+u_1)+Bv_1+Cw_1=0$$
$$u_2+B(\cos(\frac{\pi}{4})+v_2)+C(-\cos(\frac{\pi}{4})+w_2)=0$$
$$u_3+B(\sin(\frac{\pi}{4})+v_3)+C(\sin(\frac{\pi}{4})+w_3)=0$$

Adding the last two eqations, using that $\cos(\frac{\pi}{4})=\sin(\frac{\pi}{4})\approx 0.70711$ with each other gives

$$(u_2+u_3)+B(2\cdot\cos(\frac{\pi}{4})+v_2+v_3)+C(w_2+w_3)=0$$

However, we could have also subtracted the two equations, giving instead

$$(u_2-u_3)+B(v_2-v_3)+C(2\cdot\cos(\frac{\pi}{4})+w_2-w_3)=0$$

Changing B and C such that one equation is valid, will make the other equation invalid. Using the constraints on the length of u we also have $|u_2+u_3|\leq 2\sqrt{0.02}$, and similarly for v and w.

Next we take a look at Nordlings definition of neglectability, and how a rotation of the regressand would effect it.

Nordling Definition. (p.130) *A regressor* $\mathbf{X_i}$ *is neglectable if* $\mathbf{0}\in\mathcal{N}(\mathbf{X_i},u_i)$, *and* $\overline{\boldsymbol{X^*+u^*}}\boldsymbol{A}=\mathbf{Y}$ *where* $\overline{\boldsymbol{X^*+u^*}}$ *is the original regressor matrix without row i with uncertainty.*

This gives rise to the question whether there could be a case where many regressors are separately neglegable, but at least one of them is needed to solve the equation 2.3.2. If we would solve this by taking away one at the time, the result could depending on indexing, which in turn could lead to different solutions for the same set of data.

It might also effect the stochastic case in strange ways. Either we are calculating with an uncertainty of a regressor that is neglected, or the uncertainty of some other regressor could expand resulting in some contradictions, and hence no 'ranking'.

Since the rotation matrix preserves length, the property $\mathbf{0}\in\mathcal{N}(\mathbf{X_i},u_i)$ remains uneffected, as for the condition $\mathbf{X}^*\mathbf{A}=\mathbf{Y}$ we can divide it into two cases 1) any $\overline{\boldsymbol{X_i}+u_i}$ is independent i.e. can not be covered by any set uncertainty cones $\mathcal{C}(\boldsymbol{X_j},u_j)\ \forall j\neq i$, or 2) all $\overline{\mathbf{X_i}+u_i}$ are covered by some uncertainty cone $\mathcal{C}(\boldsymbol{X_j},u_j)\ j\neq i$, i.e. $\boldsymbol{X_i}$ is collinear.

If any part of $\mathbf{X_i}$ is independent, that part $\overline{\mathbf{X_i}+u_i}$ was needed to span a dimension within which $\mathbf{Y}$ was not present. However, with the rotation matrix one can always rotate $\mathbf{Y}$ such that it ends up in a dimension where $\overline{\mathbf{X_i}+u_i}$ is needed to explain it.

In the collinear case, all $\overline{\mathbf{X_i} + u_i}$ can be expressed by the other rows/cones, hence we can always find a version where it is not needed. However, when $\mathbf{0} \in \mathcal{N}(\mathbf{X_i}, u_i)$, its uncertainty cone covers at least half of $\mathbb{R}^{n \times m}$. It is then rather unlikely that the other cones cover the other half (unless there are more regressors with uncertainty containing $\mathbf{0}$, in which case a solution could depend on which regressors one chooses to neglect first).

We could create the following definition:

Definition 2. *A regressor* $\mathbf{X_i}$ *is neglectable if* $\mathbf{0} \in \mathcal{N}(\mathbf{X_i}, u_i)$*, and* $\forall \overline{\mathbf{X_i} + \mathbf{u_i}} \in \mathcal{N}_i, \exists B \neq \mathbf{0} s.t \overline{\mathbf{X}^* + \mathbf{u}^*} B = 0$*, where* $\overline{\mathbf{X}^* + \mathbf{u}^*}$ *is the matrix* $\overline{\mathbf{X} + \mathbf{u}} \in \mathcal{N}$ *without row* i*.*

If one were to optimise the rotation matrix in some way, e.g. by choosing the rotation with shortest rotation distance, one might be able to neglect a few more regressors. This is one reason why we will explore the properties of rotations and spheres in the next section.

Other definitions in Nordlings system will be completely useless if one does not put any constraint on the rotation matrix. We here mention *parameter classification* to further illustrate constraints on the rotation matrix could be necessary to be able to draw certain conclusions.

Nordling Definition. (5.5.7.) *For some column* k *of* A*, a parameter* a_j *in a solution* $A_k = \mathbf{a} = [a_1, ..., a_j, ..., a_m]^T$*, with respect to column* k *of the regressand* $\boldsymbol{Y}$*, is*

1. ***practically non-zero*** *if* $\forall \mathbf{a}, a_j \neq 0$*,*
2. ***practically positive*** *if* $\forall \mathbf{a}, a_j > 0$*,*
3. ***practically negative*** *if* $\forall \mathbf{a}, a_j < 0$*,*
4. ***practically zero*** *if* $\exists \mathbf{a}, a_j = 0$*.*

It is easy to see that *practically positive* and *practically negative* parameters can never be found, since we can always rotate $\boldsymbol{Y}$ 180 degrees to its antipod i.e. $-\boldsymbol{Y}$.

Now we look at what it means when a parameter a_j is *zero*. For the column $A_k = \mathbf{a}$ there is at least one parameter a_j that is equal to zero. One can immediately include any k such that there exists a regressor with uncertianty, $\boldsymbol{X} + \boldsymbol{u}$, where column k of $\boldsymbol{X} + \boldsymbol{u}$ is collinear with some other columns of $\boldsymbol{X} + \boldsymbol{u}$.

Next, looking at 'independent' columns of $\boldsymbol{X} + \boldsymbol{u}$. This means that column j in the regressor matrix $\boldsymbol{X}$ is always needed to explain column $\boldsymbol{Y_k}$ of the regressand matrix.

Suppose we have a regressor matrix $\boldsymbol{X}$ such that $\boldsymbol{X_{ij}} = 1$ and $\boldsymbol{X_{sj}} = \boldsymbol{X_{it}} = 0$, $\forall s \neq i, t \neq j$, with the accommpanying regressand matrix $\boldsymbol{Y}$

such that $\boldsymbol{Y_{ik}} = 1$ and $\boldsymbol{Y_{sk}} = 0, \forall s \neq i$. It is easy to see that, unless the uncertainty is very large, the parameter a_{jk} is selectable.

However, if we rotate $\boldsymbol{Y}$ such that its basis vectors change place, e.g.

$$R = \begin{pmatrix} \mathbf{0} & & 1 \\ & \iddots & \\ 1 & 0 & \mathbf{0} \end{pmatrix} \tag{7}$$

a_{jk} is no longer selectable, which means many parameters become practically zero, or, in a similar fashion, non-zero.

3 Rotation

The following sections provide tools for representing the Rotation matrix, as well as computing the distance between two points on a sphere.

We begin by describing the basic properties of a rotation matrix.

Definition: 1. *The special orthogonal group is defined by*

$$SO(n) = \{A | A \in GL_n, A^{-1} = A^T, det(A) = 1\}$$

This might not seem like a very intuitive way to describe the rotations, however we shall see that this is exactly the group we are looking for. We begin by showing that it really is a group.

Theorem: 1. *The special orthogonal group*

$$SO(n) = \{A | A \in GL_n, A^{-1} = A^T, det(A) = 1\}$$

is a group under matrix multiplication.

Proof. We have det(E)=1, hence identity is in it, det(A)=det(A^T)=1, hence all the inverse elements are in it and also det(A)=det(B)=1 which means det(AB)=1, hence it is closed under multiplication. □

Note, it is in fact a subgroup of the orthogonal group $O(n)$ for which $\det(A) = \pm 1$, more specifically, SO(n) is the subgroup of $O(n)$ which does not contain reflections, ($\det(A) = -1$).

To convince us that a matrix $A \in SO(n)$ has the properties we expect a rotation matrix to have, we want to show that A has the following properties:

i) preserves length of vectors

ii) preserves angles between vectors

Proof. Let $v, w \in \mathbb{R}^n$.

i) We need to show that $||vA|| = ||v||$. We have that $||vA||^2 = ||vA(vA)^T|| = ||vAA^Tv^T|| = ||v||^2$, hence the length is preserved.

ii) We have

$$\cos\theta = \frac{v \cdot w}{||v||\ ||w||} = \frac{vA \cdot wA}{||vA||\ ||wA||} = \frac{wA(vA)^T}{||v||\ ||w||} = \frac{v \cdot w}{||v||\ ||w||} \tag{8}$$

This means that the rotation matrix is orthonormal, i.e. the length of the vectors in the columns and rows equals 1. □

3.1 Representation of the rotation matrix

To be able to use a rotation matrix in computations, one would want to have a good representation of it. Different representation can have different advantages, e.g. computational, visional etc.

3.1.1 Euler angles

The representation most simple to understand is the use of Euler angles, i.e. rotating in one plane at the time. For 2 dimensions, this is pretty straight forward, since we only need to rotate around the origin (one axis).

Euler angles in SO(2)

Theorem: 2. *A representation of the rotation in two dimensions is of the form:*

$$A = \begin{pmatrix} \cos(\theta) & \text{-}\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \tag{9}$$

Proof. Note that, putting $\theta \in [0, 2\pi]$, and computing with modulo $\pi/2$ for θ, will give exactly one θ for each point on the circle.

Now we wish to show that every such matrix $A \in \mathbb{R}^{2\times 2}$ is in SO(2).

We see that $\det(A) = \cos^2\theta + \sin^2\theta = 1$ and that $A^T = A^{-1}$ since

$$\begin{aligned} AA^T &= \begin{pmatrix} \cos(\theta) & \text{-}\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ \text{-}\sin(\theta) & \cos(\theta) \end{pmatrix} \\ &= \begin{pmatrix} \cos^2(\theta) + \sin^2(\theta) & \cos(\theta)\sin(\theta) - \cos(\theta)\sin(\theta) \\ \cos(\theta)\sin(\theta) - \cos(\theta)\sin(\theta) & \sin^2(\theta) + \cos^2(\theta) \end{pmatrix} \\ &\qquad \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{aligned} \tag{10}$$

Now we need to show that any matrix in SO(2) can be represented as a rotation matrix $A \in \mathbb{R}^{2\times 2}$. We do this by finding a base for the rotation matrix. The first vector of this base is $\mathbf{u}$=($\cos\theta$, $\sin\theta$), which parametrises the unit circle around the origin. The vectors orthogonal to $\mathbf{u}$ are $\mathbf{v}_1$ = (-$\sin\theta$, $\cos\theta$) and $\mathbf{v}_2$ = ($\sin\theta$,-$\cos\theta$), of which a A is the matrix with $\mathbf{u}$ and $\mathbf{v}_1$ as columns. We can also see that a matrix with $\mathbf{u}$ and $\mathbf{v}_2$ as columns will have determinant -$\cos^2\theta$-$\sin^2\theta$=-1, which is not in SO(2). □

Now we can easily compute a rotation of vector $\mathbf{v} \in \mathbb{R}^2$ as

$$A\mathbf{v} = \begin{pmatrix} \cos(\theta) & \text{-}\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} \cos(\theta)v_1 - \sin(\theta)v_2 \\ \sin(\theta)v_1 + \cos(\theta)v_2 \end{pmatrix}$$

Plugging in $\mathbf{v} = (1, 0)$, and $\theta = \pi/2$ it is easy (and not suprising) to see that $A\mathbf{v} = (0, 1)$.

Theorem: 3. *SO(2) is abelian.*

Proof. Let

$$A = \begin{pmatrix} \cos(\theta) & \text{-}\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}, B = \begin{pmatrix} \cos(\phi) & \text{-}\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{pmatrix}$$

Then

$$\begin{aligned} AB &= \begin{pmatrix} \cos(\theta) & \text{-}\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} \cos(\phi) & \text{-}\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{pmatrix} \\ &= \begin{pmatrix} \cos(\theta)\cos(\phi) - \sin(\theta)\sin(\phi) & -\cos(\theta)\sin(\phi) - \sin(\theta)\cos(\phi) \\ \sin(\theta)\cos(\phi) + \cos(\theta)\sin(\phi) & -\sin(\theta)\sin(\phi) + \cos(\theta)\cos(\phi) \end{pmatrix} \\ &= \begin{pmatrix} \cos(\theta+\phi) & \text{-}\sin(\theta+\phi) \\ \sin(\theta+\phi) & \cos(\theta+\phi) \end{pmatrix} \\ &= \begin{pmatrix} \cos(\phi)\cos(\theta) - \sin(\phi)\sin(\theta) & -\cos(\phi)\sin(\theta) - \sin(\phi)\cos(\theta) \\ \sin(\phi)\cos(\theta) + \cos(\phi)\sin(\theta) & -\sin(\phi)\sin(\theta) + \cos(\phi)\cos(\theta) \end{pmatrix} \\ &= \begin{pmatrix} \cos(\phi) & \text{-}\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{pmatrix} \begin{pmatrix} \cos(\theta) & \text{-}\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} = BA \end{aligned}$$

□

However, as we shall see, SO(2) is the only one which is abelian.

Euler angles in SO(3) One way to now represent a rotation in SO(n) is to break it down to a concatenation of rotations in smaller (2) dimensions. For SO(3), these would be rotations around each of the axises, keeping the axis in question in place. Note the similarity between Euler angle representation in 2 dimensions for each of the three rotations:

$$\begin{aligned} &A = A_1A_2A_3 = \\ &\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & \text{-}\sin(\alpha) \\ 0 & \sin(\alpha) & \cos(\alpha) \end{pmatrix} \begin{pmatrix} \cos(\beta) & 0 & \text{-}\sin(\beta) \\ 0 & 1 & 0 \\ \sin(\beta) & 0 & \cos(\beta) \end{pmatrix} \begin{pmatrix} \cos(\gamma) & \text{-}\sin(\gamma) & 0 \\ \sin(\gamma) & \cos(\gamma) & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \cos(\beta)\sin(\gamma) & -\cos(\beta)\sin(\gamma) & -\sin(\beta) \\ -\sin(\alpha)\sin(\beta)\cos(\gamma) + \cos(\alpha)\sin(\gamma) & \sin(\alpha)\sin(\beta)\sin(\gamma) + \cos(\alpha)\cos(\gamma) & -\sin(\alpha)\cos(\beta) \\ \cos(\alpha)\sin(\beta)\cos(\gamma) + \sin(\alpha)\sin(\gamma) & -\cos(\alpha)\sin(\beta)\cos(\gamma) + \sin(\alpha)\cos(\gamma) & \cos(\alpha)\cos(\beta) \end{pmatrix} \end{aligned}$$

Here A_1 represents a rotation around the x-axis, A_2 a rotation around the y-axis and A_3 a rotation around the z-axis. In this case, given a matrix A

with entries a_{ij} for row i and column j, we can compute $\beta = -\sin^{-1}(a_{13})$. Next we use $a_{23} = -\sin(\alpha)\cos(\beta)$, getting

$$\sin\alpha = -\frac{a_{23}}{\cos\beta} = -\frac{a_{23}}{\sqrt{1-(\sin\beta)^2}} = -\frac{a_{23}}{\sqrt{1-(-a_{13})^2}}$$

to compute $\alpha = -\sin^{-1}(a_{23}/\sqrt{1-a_{13}^2})$, and with similar computations we get $\gamma = -\sin^{-1}(a_{12}/\sqrt{1-a_{13}^2})$.

Though for other Euler angle representations, say $A = A_3A_2A_1$, this will not be valid as we shall see in theorem 5. However the same technique can be used to find the values for those representations.

For convenience one would like every point on the sphere to be represented in exactly one way. This can be done with the help of the following constraints: $\alpha, \gamma \in [-\pi, \pi), \beta \in [-\pi/2, \pi/2)$ and computing with modulo.

Theorem: 4. *An Euler angle representation A is an element of SO(3).*

Proof. It is easy to see that the determinant of A_i is 1, the inverse of A_i is A_i^T for $i \in [1,3]$, hence $\det(A)$=$\det(A_1)\det(A_2)\det(A_3)$=1, and $A^{-1} = (A_1A_2A_3)^{-1} = A_3^TA_2^TA_1^T = A^T$. Similarly one can show that the Euler angle representations from a group (for each seperate Euler angle representation). This shows that this Euler angle representation indeed are elements of SO(3). □

Theorem: 5. *SO(n) for $n > 2$ is not abelian*

Proof. The proof will be shown by an example in 3 dimensions, which can be extended to higher dimensions analogously. Let

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & \text{-sin }(\alpha) \\ 0 & \sin(\alpha) & \cos(\alpha) \end{pmatrix}, B = \begin{pmatrix} \cos(\alpha) & 0 & \text{-sin }(\alpha) \\ 0 & 1 & 0 \\ \sin(\alpha) & 0 & \cos(\alpha) \end{pmatrix}$$

We get

$$AB = \begin{pmatrix} \cos(\alpha) & 0 & \text{-sin }(\alpha) \\ -\sin^2(\alpha) & \cos(\alpha) & \text{-sin }(\alpha)\cos(\alpha) \\ \cos(\alpha)\sin(\alpha) & \sin(\alpha) & \cos^2(\alpha) \end{pmatrix}$$

While

$$BA = \begin{pmatrix} \cos(\alpha) & -\sin^2(\alpha) & \text{-sin }(\alpha)\cos(\alpha) \\ 0 & \cos(\alpha) & \text{-sin }(\alpha) \\ \sin(\alpha) & \cos(\alpha)\sin(\alpha) & \cos(\alpha) \end{pmatrix}$$

Hence A and B do not commute, and so $SO(3)$ is not abelian. For higher dimensions the results will be similar, since we can have 3-dimensional rotation in higher dimensions. □

This means that a representation in Euler angles is not unique. In fact, the representations $A = A_1A_2A_3$ will seldom be the same as e.g. $A^* = A_3A_2A_1$, even though these Euler angle representation are equally valid.

Theorem: 6. *Euler's rotation theorem: If A is an element of $SO(3)$ where $A \neq I$, then A has a one-dimensional eigenspace, which is the axis of rotation.*

As we shall see, this axis of rotation will only exist in 3 dimensions.

Euler angles in SO(4) Like a rotation in SO(3) we can represent rotation in SO(4) with Euler angles, consisting of a composition of rotations, one within each plane. However, unlike in three dimensions the 2-dimensional rotations will not occur around an axis. A rotation in SO(4) consists of the following 6 rotations [Tri09]:

$$\begin{bmatrix} \cos(\alpha_1) & -\sin(\alpha_1) & 0 & 0 \\ \sin(\alpha_1) & \cos(\alpha_1) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} \cos(\alpha_2) & 0 & -\sin(\alpha_2) & 0 \\ 0 & 1 & 0 & 0 \\ \sin(\alpha_2) & 0 & \cos(\alpha_2) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} \cos(\alpha_3) & 0 & 0 & -\sin(\alpha_3) \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \sin(\alpha_3) & 0 & 0 & \cos(\alpha_3) \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\alpha_4) & -\sin(\alpha_4) & 0 \\ 0 & \sin(\alpha_4) & \cos(\alpha_4) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\alpha_5) & 0 & -\sin(\alpha_5) \\ 0 & 0 & 1 & 0 \\ 0 & \sin(\alpha_5) & 0 & \cos(\alpha_5) \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos(\alpha_6) & -\sin(\alpha_6) \\ 0 & 0 & \sin(\alpha_6) & \cos(\alpha_6) \end{bmatrix}$$

This can be seen as choosing the plane spanned by the 2 vectors e_i and e_j, with $i, j \in [1, n]$. The number of rotations for each n is then $\binom{n}{2} = \frac{(n-1)n}{2}$, which gives an increase of $\mathcal{O}(n)$.

In order to avoid having multiple representations of the 'same' rotation, i.e. those rotations which end up on the same point, we might want constraints on the angles $\alpha_1, \alpha_n \in [0, 2\pi), \alpha_2, \alpha_{n-1} \in [0, \pi)$.

Note, plugging in $\beta = -\pi/2$ to a rotation representation $A \in SO(3)$, we get:

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

This is called a *Gimbal lock*, meaning that the same rotation can be reached whether we rotate the x-axis or the z-axis. It is easy to see that

any similar representation for $n > 3$ will also result in one or more Gimbal locks. It would be interesting to know what a Gimbal lock means for the uncertainty. It could perhaps be compared to some kind of (collinearity), since in both cases, we don't know how much of one or the other is needed. If only the overall length of the rotation is interesting, then Gimbal lock does not mean much, however, if axises have meaning, then Gimbal lock could mean something.

Theorem 1. *In n dimensions one can have at most* $\lfloor n/3 \rfloor$ *Gimbal lock(uncertainty)s in one rotation.*

Proof. One Gimbal lock effects three neighbouring angles, hence without overlapping we could have a rotation such that $\theta_i = \pi/2 \, \forall i = 2+3j, i \in [1, n]$, which would result in $\lfloor n/3 \rfloor$ Gimbal locks. Now we need to show that no overlapping can exist.

Suppose $\theta_i = \pi/2$, and $\theta_{i-1} + \theta_{i+1} = k$. Now we choose $\theta_{i-1} = \theta/2$. There would not arise a new Gimbal lock around $i - 1$, since θ_i is already set to $\pi/2$, changing its value would eliminate the first Gimbal lock. □

However this might not be very relevant in this optimisation case, since a regressor $\boldsymbol{X}$ (and regressand) will always have uncertainty, and an oposite $-\boldsymbol{X}$, such that we will have the smallest distance to $\pm\overline{\boldsymbol{X} + \boldsymbol{u}} < \pi/2$ at all times for 3 dimensions.

3.1.2 Generalized Euler theorem of rotations, SO(n) and S^{n-1}

As we might have already guessed we can construct functions which map rotation matrices to Euler angles and expand Eulers rotation theorem, which only works for 3 dimensions. This will be useful both for the Haar-measure in a later section, but also to get a better understanding of what a specific rotation looks like.

Given an Euler angle representation matrix $A \in \mathbb{R}^n$ with angles $\boldsymbol{\theta} = \theta_1, \theta_2, ..., \theta_{n-1}$, we want to have a map $\sigma^{n-1}(\boldsymbol{\theta}) : [0, 2\pi] \times [0, \pi]^{n-2} \to S^{n-1} \in \mathbb{R}^n$.

As we know we could describe the points on the unit circle when given an angle θ as

$$p = \begin{pmatrix} \sin\theta \\ \cos\theta \end{pmatrix} \in S^1$$

As for the 2-sphere S^2, mathematicians and physicists frequently use [wol] (Spherical Coordinates)

$$p = \begin{pmatrix} \sin\theta_1 \sin\theta_2 \\ \cos\theta_1 \sin\theta_2 \\ \cos\theta_2 \end{pmatrix} \in S^2$$

Where, in mathematics, θ_1 is usually called the azimuthal angle, here on the y-x-plane, and θ_2, often denoted ϕ, is called the polar angle, being the angle from the z-axis. Though the notation can vary between and amongst mathematical and physical litterature [wol](Spherical Coordinates).

Theorem: 7. *[Can96] We can construct a map* $\sigma^n : [0, 2\pi] \times [0, \pi]^{n-1} \to S^n$ *inductively, letting* $\sigma^1(\theta) = (\sin\theta, \cos\theta)^T$, *and for* $\boldsymbol{\theta}_n = (\theta_1, ...\theta_n) = (\boldsymbol{\theta}_{n-1}, \theta_n)$ *define*

$$\sigma^n(\boldsymbol{\theta_n}) = \begin{pmatrix} \sin\theta_n \sigma^{n-1}(\boldsymbol{\theta}_{n-1}) \\ \cos\theta_n \end{pmatrix} \in S^n \tag{11}$$

Proof. To show that σ^n is a indeed a map from $[0, 2\pi] \times [0, \pi]^{n-1}$ to S^n we show that $||\sigma^n|| = 1$ and that for every point $p \in S^n$ we can find $\boldsymbol{\theta}$ such that $\sigma^n(\boldsymbol{\theta}) = p$. This will be done inductively. First, we see that $||\sigma^1(\theta)|| = \cos^2\theta + \sin^2\theta = 1$. Now, suppose $||\sigma^{n-1}(\boldsymbol{\theta}_{n-1})|| = 1$. Then $||\sigma^n(\boldsymbol{\theta}_n)|| = ||\sin^2\theta_n(\sigma^{n-1}(\boldsymbol{\theta}_{n-1})) + \cos^2\theta_n|| = ||\sin^2\theta_n \cdot 1 + \cos^2\theta_n|| = 1$.

Second, let $p = (p_1, ..., p_{n+1}) \in S^n \subset \mathbb{R}^{n+1}$. We want to show that $\forall p \exists \boldsymbol{\theta}_{n-1}$ such that $\sqrt{1-p_{n+1}}\sigma^{n-1}(\boldsymbol{\theta}_{n-1}) = (p_1, ..., p_n)$. Since $p \in S^n$ we have that $-1 \leq p_{n+1} \leq 1$ and hence $\sqrt{1-p_{n+1}} \leq 1$ which means that $\exists\theta_n \in [0, \pi]$ such that $\cos\theta_n = p_{n+1}$ and $\sin\theta_n = \sqrt{1-p_{n+1}}$.

This shows that $\forall p \in S^n \ \exists\boldsymbol{\theta}$ such that $\sigma^n(\boldsymbol{\theta}) = p$. □

From this we can see that $p \in S^n$ is independent of which point $p' \in S^{n-1}$ we use as starting point.

Now we define the orthonormal base $\boldsymbol{\omega}$ in which we could express the rotation matrix later as [Can96]:

$$\begin{aligned} \omega_1(\theta_1) &:= \sigma^{n-1}(\theta_1 + \pi/2, \pi/2, ..., \pi/2) \\ \omega_k(\theta_1, ..., \theta_k) &:= \sigma^{n-1}(\theta_1, ..., \theta_{k-1}, \theta_k + \pi/2, \pi/2, ..., \pi/2) \\ \omega_{n-1}(\theta_1, ..., \theta_n) &= \sigma^{n-1}(\theta_1, ..., \theta_{n-1}, \theta_n + \pi/2) \\ \omega_n(\theta_1, ..., \theta_n) &= \sigma^{n-1}(\theta_1, ..., \theta_n) \end{aligned} \tag{12}$$

Let $\boldsymbol{\omega}_k(\boldsymbol{\theta}_n) = ((\omega_k)_1, ..., (\omega_k)_i, ..., (\omega_k)_n)$. We can see that for $k \leq n_1 - 2 < n_2 - 2$, the $\omega_k(\theta_1, ..., \theta_k)$ for n_1 and n_2, are up to a number of $\pi/2$ at the end. Letting ω_k^j be the k'th vector of the base of size j, we can also see that

$$\omega_{n-1}^n(\theta_1, ..., \theta_{n-1}) = \begin{pmatrix} \cos(\theta_{n-1})\omega_{n-1}^{n-1}(\theta_1, ..., \theta_{n-2}) \\ -\sin\theta_{n-1} \end{pmatrix} \tag{13}$$

and

$$\omega_n^n(\theta_1, ..., \theta_{n-1}) = \begin{pmatrix} \sin(\theta_{n-1})\omega_{n-1}^{n-1}(\theta_1, ..., \theta_{n-2}) \\ \cos\theta_{n-1} \end{pmatrix} \tag{14}$$

using equation (11) from the definition of σ. We will need these results to show a later theorem.

Theorem: 8. [Can96] *Given vectors* $\boldsymbol{\omega}_k$ $k \in [1, n]$ *we can create the matrix* $M^n(\boldsymbol{\theta}) := (\boldsymbol{\omega_1}, ..., \boldsymbol{\omega_n})$ *which creates an orthonormal base in* S^{n-1}*, where* $\boldsymbol{\theta} = (\theta_1, ..., \theta_{n-1})$*.*

Proof. (Sketch) To show that $M^n(\boldsymbol{\theta})$ creates an othonormal base for S^{n-1} we need to show that its columns are of length 1, and that they are orthogonal to each other, i.e. $\omega_k \cdot \omega_j = 0 \; \forall k \neq j$, which is shown inductively, using e.g. equation (11). The proof that $\det M^n = 1$ is also shown inductively. For details the reader is deferred to [Can96]. □

To illustrate how the proof works, we give instead two examples. The first is the base case, two dimensions, and how to expand it to 3 dimensions.
Example: Base case For two dimensions we have $\omega_1(\theta) = \sigma^1(\theta + \pi/2) = (\sin(\theta+\pi/2), \cos(\theta+\pi/2)) = (\cos\theta, -\sin\theta)$ and $\omega_2(\theta) = \sigma^1(\theta) = (\sin\theta, \cos\theta)$, and hence $\boldsymbol{\omega}_1 = ((\omega_1(\theta))_1, (\omega_2(\theta))_1) = (\cos\theta, \sin\theta)$ and $\boldsymbol{\omega}_2 = (-\sin\theta, \cos\theta)$. Which gives

$$M^2(\theta) = \begin{pmatrix} \boldsymbol{\omega_1} \\ \boldsymbol{\omega_2} \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

We have already seen in a previous section that $M^2(\theta)$ is an orthonormal basis for the rotation matrix, and by the same reasoning it is also one for S^1.
Example: 3 dimensions We have $\omega_1(\theta_1, \theta_2) = \sigma^2(\theta_1 + \pi/2, \pi/2, \pi/2) = (\sin(\theta_1+\pi/2)\sin(\pi/2), \cos(\theta_1+\pi/2)\sin(\pi/2), \cos(\pi/2)) = (\cos\theta_1, -\sin\theta_1, 0)$, $\omega_2(\theta_1, \theta_2) = (\sin\theta_1\cos\theta_2, \cos\theta_1\cos\theta_2, -\sin\theta_2)$ and $\omega_3(\theta_1, \theta_2) = (\sin\theta_1\sin\theta_2, \cos\theta_1\sin\theta_2, \cos\theta_2)$. Which gives

$$M^3(\theta_1, \theta_2) = \begin{pmatrix} \cos\theta_1 & \sin\theta_1\cos\theta_2 & \sin\theta_1\sin\theta_2 \\ -\sin\theta_1 & \cos\theta_1\cos\theta_2 & \cos\theta_1\sin\theta_2 \\ 0 & -\sin\theta_2 & \cos\theta_2 \end{pmatrix}$$

Some basic computations will convince us that the columns are of length one. Letting $\alpha_i = \theta_i(+\pi/2)$, we have $||\sigma^2|| = ||(\sin\alpha_2\sigma^1(\alpha_1), \cos\alpha_2)|| = ||(\sin\alpha_2 \cdot 1, \cos\alpha_2)|| = 1$. Next, showing that the columns are orthogonal, we see that

$$\omega_1 \cdot \omega_2 = \cos\theta_1\sin\theta_1\cos\theta_2 + -\sin\theta_1\cos\theta_1\cos\theta_2 + 0(-\sin\theta_2) = (\cos\theta_1\cos\theta_2)(\sin\theta_1 + (-\sin\theta_1)) + 0 = 0$$

Here $(\cos\theta_1\cos\theta_2)(\sin\theta_1 + (-\sin\theta_1))$ can also be written as $(\omega_1^2 \cdot \omega_2^2)\cos\theta_2$, which shows how to inductively extend it to higher dimensions. For the other columns a similar technique is possible.

Next we would like to know the smallest possible parametrisation of M^n,

i.e. how many independent angles are needed to uniquely construct a point on the $n-1$-sphere.

As it turns out, M^n can in turn be described going through one plane at the time. We will use this to find the smallest parametrisation of M^n. We start by showing the following theorem.

Theorem: 9. *Define a rotation in SO(n) on a plane k as*

$$P_k^n(\theta_k^{n-1}) := \left(\begin{array}{c|cc|c} I_{k-1} & \multicolumn{2}{c|}{\mathbf{0}} & \mathbf{0} \\ \hline \mathbf{0} & \cos\theta_k^{n-1} & \sin\theta_k^{n-1} & \mathbf{0} \\ & -\sin\theta_k^{n-1} & \cos\theta_k^{n-1} & \\ \hline \mathbf{0} & \multicolumn{2}{c|}{\mathbf{0}} & I_{n-(k+1)} \end{array}\right)$$

Then a rotation matrix M^n can be decomposed into planar rotation such that $M^n(\theta) = \prod_{k=1}^{n} P_k^n(\theta)$.

Proof. The proof is given inductively. The base case, $n = 2$, is clear since $M^2(\theta) = \prod_{k=1}^{1} P_k^2(\theta) = P^2(\theta)$. Now assume it holds for M^{n-1}, we want to show it will then hold for M^n. We get

$$\prod_{k=1}^{n-1} P_k^n(\theta_k^{n-1}) = \prod_{k=1}^{n-2} \left(\begin{array}{c|c} P_k^n(\theta_k^{n-1}) & \mathbf{0} \\ \hline \mathbf{0} & I_1 \end{array}\right) \cdot P_{n-1}^n(\theta_{n-1}^{n-1}) =$$

$$\left(\begin{array}{c|c} M^{n-1}(\theta_k^{n-1}) & \mathbf{0} \\ \hline \mathbf{0} & I_1 \end{array}\right) \cdot \left(\begin{array}{c|cc} I_{n-2} & \multicolumn{2}{c}{\mathbf{0}} \\ \hline \mathbf{0} & \cos\theta_{n-1}^{n-1} & \sin\theta_{n-1}^{n-1} \\ & -\sin\theta_{n-1}^{n-1} & \cos\theta_{n-1}^{n-1} \end{array}\right) =$$

$$\left(\begin{array}{ccc|c} \omega_1^{n-1} & \cdots & \omega_{n-1}^{n-1} & \mathbf{0} \\ \hline 0 & \cdots & 0 & I_1 \end{array}\right) \cdot \left(\begin{array}{c|cc} I_{n-2} & \multicolumn{2}{c}{\mathbf{0}} \\ \hline \mathbf{0} & \cos\theta_{n-1}^{n-1} & \sin\theta_{n-1}^{n-1} \\ & -\sin\theta_{n-1}^{n-1} & \cos\theta_{n-1}^{n-1} \end{array}\right) =$$

$$\left(\begin{array}{ccccc} \omega_1^{n-1} & \cdots & \omega_{n-2}^{n-1} & \omega_{n-1}^{n-1}\cos\theta_{n-1}^{n-1} & \sin\theta_{n-1}^{n-1}\omega_{n-1}^{n-1} \\ \hline 0 & \cdots & 0 & -\sin\theta_{n-1}^{n-1} & \cos\theta_{n-1}^{n-1} \end{array}\right) = (\omega_1^n, \cdots, \omega_n^n)$$

The last equality uses equations (13) and (14) and since

$$(\omega_1^n, \cdots, \omega_n^n) = M^n(\theta_n^{n-1})$$

we are finished. □

This means that M^n is parametrised by $n-1$ different angles. However, M^n is the base for the single point $p \in S^{n-1}$. More specifically, e.g. for 3 M^3 could only express rotations of an angle θ_1 around the z-axis, followed by a rotation θ_2 around the x-axis. It can only describe a subset of the rotation matrices. Hence we need some extra rotation to express SO(3), and SO(n).

We introduce the function $\Omega^n : [0, 2\pi]^{n-1} \times [0, \pi]^{(n-1)(n-2)/2} \to$SO($n$) such that $\Omega^n = \prod_{k=2}^{n} M_{n-k+2}^n$[Can96], where $M_{n-k+2}^n = \left(\begin{smallmatrix} M^{n-k+2} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I} \end{smallmatrix}\right)$. We

want to show that for an orthonormal base $\boldsymbol{x_1}, ..., \boldsymbol{x_n} \in \mathbb{R}^n$ we have $\Omega^n \cdot \boldsymbol{e}_k = \boldsymbol{x_k}\ \forall k \in \{1, n\}$.

Since the last column of Ω^n is only effected by the last column of M^n, we get the first (n'th) base vector from the following:

$$\Omega^n \cdot \boldsymbol{e}_n = M^n(\boldsymbol{\theta}_{n-1})\boldsymbol{e}_n = \sigma^{n-1}(\boldsymbol{\theta}_{n-1}) = \boldsymbol{x_n} = p \in S^{n-1} \tag{15}$$

In other words the last column of Ω describes a point $p \in S^{n-1}$. Furthermore, since M^n is orthonormal, we have that $(M^n)^T \cdot \boldsymbol{x_n} = \boldsymbol{e_n}$.

For the other vectors we have that $(M^n)^T \cdot \boldsymbol{x_k}$ is in the span of $\boldsymbol{e_1}, ..., \boldsymbol{e_{n-1}}$, and are therefore orthogonal to $\boldsymbol{e_n}$ [Can96]. We can see that we can get the other base vectors inductively, by first computing M^n from the angles we get from $p \in S^{n-1}$. then multiply Ω^n with $(M^n)^{-1}$ and repeating the process. Each point $p_k \in S^k$ for M^n_{k+1} will be of length one.

Finally we need to show that each rotation matrix $R \in \mathrm{SO}(n)$ can be expressed as $\Omega^n(\theta^1, \boldsymbol{\theta}^2, ..., \boldsymbol{\theta}^{n-1})$. For each dimension $n - k + 2$ we find the base M^n_{n-k+2} inductively using that Ω is a product of matrices where at any step $k(\geq 2)$ the lower right quadrant is the identity matrix of size $k - 2$.

We can find the point $p \in S^n$ corresponding to the first base vector as

$$p = \vec{\sigma}^{n-1}(\boldsymbol{\theta}^{n-1}) = (\sigma_1^{n-1}, ..., \sigma_{n-1}^{n-1})^T = R_n \tag{16}$$

where R_n is column n of rotation matrix R. To get the other base vectors we first compute the angle $\theta_n = \cos^{-1} R_{n,n}$. We use this to compute the other Euler angles of p inductively, knowing that $\sigma^n(\boldsymbol{\phi}_n) = \binom{\sin\phi_n \sigma^{n-1}(\boldsymbol{\theta}_{n-1})}{\cos\theta_n}$.

This procedure gives the parameters $\boldsymbol{\theta}^{n-k-1} \in [0, 2\pi] \times [0, \pi]^{n-k-2}$ for M^{n-k}. In other words, for any step we have $(n - k) - 1$ planar rotations, each on a seperate plane.

From the theorem (9) we have

$$\Omega^n(\theta, \boldsymbol{\theta}^2, ..., \boldsymbol{\theta}^{n-1}) = \prod_{k=2}^{n} M^n_{n-k+2}(\boldsymbol{\theta}^{n-k+1}) = \prod_{k=2}^{n} \prod_{j=1}^{n-k+1} P_j^n(\boldsymbol{\theta}_j^{n-k+1})$$

These are $\binom{n}{2}$ planar rotations, and the rotations can occur multiple times on the same plane. This will give a different example of Euler angle representation for $n = 3, 4$ than we had in the previous section, where we had one rotation on each plane. The order and choice of planar rotation can vary, although there will of course never be two consecutive rotations in the same plane.

Hence we have found a way to find the base of the rotation matrix independently of dimensions n, and thus generalizing the Euler theorem of rotation. Note, however, that it takes quite a lot of computations to find the base. For each step k, one needs to compute the angles $\theta_1, ..., \theta_k$ for σ^k, after which one needs to compute the other k columns in M^n, whose transpose is then multiplied wich the matrix that is left. There are n such steps.

3.1.3 Quaternions $\mathbb{H}$, $\mathbb{H} \times \mathbb{H}$ and groups SO(3) and SO(4)

Let a quaternion $\mathbf{q} \in \mathbb{H} = \{q_0 + q_1 i + q_2 j + q_3 k | q_i \in \mathbb{R}\}$, with the following properties:

(i) $i^2 = j^2 = k^2 = -1$

(ii) $ij = k$, $jk = i$, $ki = j$

(iii) $ji = -k$, $kj = -i$, $ik = -j$

(iv) $\overline{q} = q_0 - q_1 i - q_2 j - q_3 k$

Now define q_0 as the scalar pats of $\mathbf{q}$ and $\vec{q} = q_1 i + q_2 k + q_3$ as its vector part. Let $\mathbf{q}, \mathbf{p} \in \mathbb{H}$. It is easy to see that the set of quaternions form a group under componentwise addition. As for multiplication we get:

$$\begin{aligned}\mathbf{qp} = (q_0p_0 - q_1p_1 - q_2p_2 - q_3p_3) + (q_0p_1 + p_1q_0 + q_2p_3 - q_3p_2)i + \\ + (q_0p_2 + q_2p_0 + q_3p_1 - q_1p_3)j + (q_0p_3 + q_4 + p_0 + q_1p_2 - q_2p_1)k = \\ (q_0p_0 - \vec{q} \cdot \vec{p}) + (q_0\vec{p} + p_0\vec{q} + \vec{q} \times \vec{p}) \quad (17)\end{aligned}$$

and it is easy to check that e.g. $(-q_0p_1 - q_1p_0 + q_2p_3 - q_3p_2) = (q_0(-p_1) + (-q1)p_0 + (-q_2)(-p_3) - (q_3)(-p_2)$ and hence (note the change in order)

$$\overline{\mathbf{qp}} = \overline{\mathbf{p}}\ \overline{\mathbf{q}} \quad (18)$$

We can also see that $\mathbf{q}\overline{\mathbf{q}} = q_0^2 + q_1^2 + q_2^2 + q_3^2 = \overline{\mathbf{q}}\mathbf{q}$ and hence that

$$|\mathbf{q}| = \sqrt{\mathbf{q}\overline{\mathbf{q}}} \quad (19)$$

Combining equations (18) and (19) we get

$$|\mathbf{qp}|^2 = \mathbf{qp}\overline{\mathbf{qp}} = \mathbf{q}\ \mathbf{p}\overline{\mathbf{p}}\ \overline{\mathbf{q}} = \mathbf{q}|\mathbf{p}|\overline{\mathbf{q}} = |\mathbf{p}|^2|\mathbf{q}|^2 \quad (20)$$

Now we define the unit quaternions by letting $|\vec{q}| = 1$ and $q_0 = \cos(\theta)$, then we can rewrite $\mathbf{q} \in \mathbb{H}$ as

$$\mathbf{q} = \cos(\theta) + \vec{q}\sin(\theta) \quad (21)$$

It is easy to see that $|\mathbf{q}| = 1$ and that these unit quaternions form the 3-sphere S^3 in $\mathbb{R}^4$. [Tay]

Theorem: 10. [Tri09] *The 3-sphere $S^3 = \{\mathbf{q} \in \mathbb{H} | \ |\mathbf{q}| = 1\}$ is a non-abelian group under quaternion multiplication.*

Proof. It is also easy to see that the unit quaternions are closed under multiplication since $|\mathbf{qp}| = |\mathbf{q}||\mathbf{p}| = 1 \cdot 1 = 1$, contains the identity $\mathbf{q} = 1$ and $\forall \mathbf{q} \in S^3$ we have the inverse $\mathbf{q}^{-1} = \overline{\mathbf{q}}/|\overline{q}|^2 = \overline{\mathbf{q}} \in S^3$ since $|\overline{\mathbf{q}}| = \overline{\mathbf{q}}\mathbf{q} = 1$.

It is easy to see that the quaternions are associative but rarely commute, and hence the group S^3 is not abelian. □

The next theorem is very useful since it is sometimes easier to slow something in S^3 than it is in SO(3), and it provides a way to move between them.

Theorem: 11. *There is a 2-1-homomorphism from S^3 to SO(3).*

Proof. We want to show that each unit quaternion $\mathbf{q} \in \mathbb{R}^4$ can be mapped to rotations in $A \in SO(3) \in \mathbb{R}^{3\times 3}$ such that every rotation A has exactly 2 quaternions mapped to it.

Let $\mathbf{r} = (0, x, y, z) \in \mathbb{R}^4$, or equivalently $\mathbf{r} = (x, y, z) \in \mathbb{R}^3$. Let $\mathbf{q} = \cos(\theta) + \vec{q}\sin(\theta)$. Define $f[\mathbf{q}](\mathbf{r})$ as multiplication of a unit quaternion with r , i.e. $f[\mathbf{q}](\mathbf{r}) = \mathbf{q}\mathbf{r}\overline{\mathbf{q}}$. It is easy to see that $|\mathbf{q}\mathbf{r}\overline{\mathbf{q}}| = |\mathbf{q}||\mathbf{r}||\overline{\mathbf{q}}| = |\mathbf{r}|$, hence this multiplication with unit quaternions is length preserving, as a rotation should be. To see that it will be in $\mathbb{R}^3$ we have

$$\begin{aligned}\mathbf{q}\mathbf{r}\overline{\mathbf{q}} = (\cos\theta + \vec{q}\sin\theta)\mathbf{r}(\cos\theta - \vec{q}\sin\theta) = \\ [\cos^2\theta x + 2\cos\theta\sin\theta(zq_2 - yq_3) - \\ \sin^2\theta(q_1(-xq_1 - yq_2 - zq_3) + q_2(xq_2 - yq_1) - q_3(zq_1 - xq_3))]i \\ [\cos^2\theta y + 2\cos\theta\sin\theta(xq_3 - zq_1) - \\ \sin^2\theta(q_2(-xq_1 - yq_2 - zq_3) + q_3(yq_3 - zq_2) - q_1(xq_2 - yq_1))]j \\ [\cos^2\theta z + 2\cos\theta\sin\theta(yq_1 - xq_2) - \\ \sin^2\theta(q_3(-xq_1 - yq_2 - zq_3) + q_1(zq_1 - xq_3) - q_2(yq_3 - zq_2))]k \quad (22)\end{aligned}$$

which is of the same form as $\mathbf{r}$.

From (22) we can create the following matrix:

$$A = \begin{pmatrix} \cos^2\theta + \sin^2\theta(q_1^2 - q_2^2 - q_3^2) & 2(-\cos\theta\sin\theta q_3 + \sin^2\theta q_1 q_2) & 2(\cos\theta\cos\theta q_2 + \sin^2\theta q_1 q_3) \\ 2(\cos\theta\sin\theta q_3 + \sin^2\theta q_1 q_2) & \cos^2\theta + \sin^2\theta(q_2^2 - q_1^2 - q_3^2) & -2(\cos\theta\cos\theta q_1 + \sin^2\theta q_2 q_3) \\ -2(\cos\theta\cos\theta q_2 + \sin^2\theta q_1 q_3) & 2(\cos\theta\cos\theta q_1 + \sin^2\theta q_2 q_3) & \cos^2\theta + \sin^2\theta(q_1^2 - q_2^2 - q_3^2) \end{pmatrix} \quad (23)$$

It can be shown that $\det A = 1$ and $A^T A = 1$, showing that $f[\mathbf{q}] \in$SO(3), and hence that $f[\mathbf{q}] : S^3 \to$SO(3). Given $\mathbf{q_1}, \mathbf{q_2} \in S^3$ we have:

$$f[\mathbf{q_1}\mathbf{q_2}](\mathbf{r}) = \mathbf{q_1}\mathbf{q_2}\mathbf{r}\overline{\mathbf{q_1}\mathbf{q_2}} = \mathbf{q_1}\mathbf{q_2}\mathbf{r}\overline{\mathbf{q_2}}\ \overline{\mathbf{q_1}} = (f[\mathbf{q_1}] \circ f[\mathbf{q_2}])(\mathbf{r}) \quad (24)$$

Showing that $f[\mathbf{q}]$ is a homomorphism. To show that it is surjective, we try to represent a rotation around the x-axis, i.e. $\vec{q} = (1, 0, 0) = i$ in the matrix

(23), and get:

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos^2\theta - \sin^2\theta & -2\cos\theta\sin\theta \\ 0 & 2\cos\theta\sin\theta & \cos^2\theta - \sin^2\theta \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(2\theta) & -\sin(2\theta) \\ 0 & \sin(2\theta) & \cos(2\theta) \end{pmatrix} \quad (25)$$

Which turns out to be a rotation of 2θ around the x-axis. Similar results can be shown for rotation around the y-axis and z-axis. And since any rotation in SO(3) is a combination of rotations around the x-axis, y-axis and z-axis, and $\mathbf{q_1 q_2 q_3} \in S^3$, it is therefor also surjective.

To show that it is two-to-one, we show that ± 1 are only two elements in the kernel. The elements in the kernel are those unit quaternions $\mathbf{q}$ such that $f(\mathbf{r}) = \mathbf{q}\mathbf{r}\overline{\mathbf{q}} = \mathbf{r}\ \forall \mathbf{r} = xi + yj + zk$. But then $\mathbf{qr} = \mathbf{rq}$, hence $\mathbf{q}$ is real and of length 1, which means the kernel is $\mathbf{q} = \pm 1$. □

One might want to think of the homomorphism as glueing the two point $\mathbf{q}, -\mathbf{q} \in S^3$ together.

Truly, when one computes the *Rodrigues rotation formula* with angle 2θ around $\overline{\mathbf{q}}$, one gets matrix (23). For example the first element, using the trigonometric identities and the fact that $|\overline{\mathbf{q}}| = 1$, is [wol](Rodrigues' Rotation Formula)

$$a_{11} = \cos(2\theta) + q_1(1 - \cos(2\theta)) = (\cos^2\theta - \sin^2\theta) + q_1(1 - \cos^2\theta + \sin^2\theta) = (\cos^2\theta - \sin^2\theta(q_1^2 + q_2^2 + q_3^2)) + q_1(2\sin^2\theta) = \cos^2\theta + \sin^2\theta(q_1 - q_2 - q_3) \quad (26)$$

There are also a few physical experiments one can do to convince oneself there is a difference between rotating something 360 vs. 720 degrees. For example the Dirac string trick and the Philippine wine glass trick [Den]

Note that if we compare the spherical coordinates for a point $p \in S^3$ as being $(\cos\theta_1, \sin\theta_1\cos\theta_2, \sin\theta_1\sin\theta_2\cos\theta_3, \sin\theta_1\sin\theta_2\sin\theta_3)$, with quaternions, we get

$$p = \cos\theta_1 + \sin\theta_1\cos\theta_2 + \sin\theta_1\sin\theta_2\cos\theta_3 + \sin\theta_1\sin\theta_2\sin\theta_3 = \cos\theta_1 + \sin\theta_1(\cos\theta_2 + \sin\theta_2\cos\theta_3 + \sin\theta_2\sin\theta_3) = \cos\theta_1 + \sin\theta_1\vec{q} \quad (27)$$

since $(\cos\theta_2 + \sin\theta_2\cos\theta_3 + \sin\theta_2\sin\theta_3)^2 = 1$.

The quaternions $\mathbf{q} = \cos\theta + \vec{q}\sin\theta$ can also be expressed as $\mathbf{q} = e^{\theta\vec{q}}$ shown through power series expansion [WW05]:

$$e^{\theta\vec{q}} = \sum_{n=0}^{\infty} \frac{(\theta\vec{q})^n}{n!} = 1\cos\theta + \vec{q}\sin\theta = \mathbf{q} \quad (28)$$

It is not possible to uniquely determine θ and $\vec{q}$ given only unit quaternion $\mathbf{q}$.

Since $\mathbb{H}$ is not commutative this will effect the operations on $e^{\theta\vec{q}}$, e.g. $e^{\theta\vec{q_1}}e^{\theta\vec{q_2}}$ is not necesserily the same as $e^{\theta\vec{q1}+\theta\vec{q_2}}$ nor as $e^{\theta\vec{q_2}}e^{\theta\vec{q_1}}$.

Since SO(2) is isomorphic to the unit circle, and we have previously shown that there is a 2-1-homomorphism between SO(3) and the unit sphere, one might think that SO(4) could be represented with some other sphere. However it turns out to be wrong. As can be shown by investigating the eigenvalues and eigenspaces of SO(4), it is rather two simultaneous rotations each on a plane non-commutative to the other [Tri09].

In fact, the elements on one plane will move with an angle of say α, and the elements on the other plane (which does not share any points with the first) move by an angle β, then all the other points will rotate in an angle between α and β [Col90]

Theorem: 12. [Tri09][WW05] *Let $\boldsymbol{q}, \boldsymbol{p}, \boldsymbol{r} \in \mathbb{H}$, and let $f_{qp} : \mathbb{H}\times\mathbb{H} \to SO(4)$ such that $f_{qp} = \boldsymbol{qrp} = \boldsymbol{x}$, then $\boldsymbol{x} \in SO(4)$ and f_{qp} is a two-to-one, surjective homomorphism.*

The proof can be given in a similar way as for the SO(3) case.

3.1.4 Clifford algebra, an extension of quaternions to higher dimensions

Since this report deals with rotation in dimensions much higher than 3 or 4, we need a way to extend the concept of quaternions into those dimensions. We get the Clifford algebra, which is an extension of (amongst other things) the real numbers, the complex numbers as well as the quaternions and geometric algebra. We can express the rotations as multiple (pairs of) reflections about some plane [Lou01].

As an extension of quaternions, Clifford algebra will turn out to be a way to simplify computations for rotations in higher dimensions. These simplifications are called spinors which are a double covering of SO(n). These spinors also express some great difference between SO(n) and the way one would think about rotations.

Clifford algebra was first developed by William Kingdon Clifford (1845-1879), combining quaternions with the outer product of Hermann Grassmann. However the contemporary Vector algebra became more widely known, and it was not untill later that Clifford algebra recieved more attention as it was found to have application in both physics and computer science [Wil05].

Definition: 2. [Wil05] *Given a real vector space V, the* Clifford algebra Cl_V *is the associative algebra freely generated as an $\mathbb{R}$-algebra by V satisfying*

$$x^2 = -|x|^2 \ , \ \forall x \in V$$

where $|x|$ is a quadratic form.

In this report we only need to care about those cases when $V \in \mathbb{R}^n$ (denoted Cl_n) and $|x| = (\sum_{i=1}^{n} x_i^2)^{1/2}$, for $x = (x_1, x_2, ..., x_n)$.

The objects in the Clifford algebra Cl_n, where $V \in \mathbb{R}^n$ with basis $e_1, e_2, ..., e_n$, are defined as follows:

Grade	**Object**	**Products** of basis elements
0	Scalar	1
1	Vector	$e_1, e_2, ..., e_n$
2	Bivector	$e_1e_2, e_1e_3, ..., e_{n-1}e_n$
3	Trivector	$e_1e_2e_3, ..., e_{n-2}e_{n-1}e_n$
$\vdots$	$\vdots$	$\vdots$
n	Blade	$e_1e_2...e_n$

Like a vector, the bivectors, trivectors and blades have orientations. For example, a bivector defines both a plane and an orientation on that plane. This makes thee visualization of an n-blade quite difficult, as it will be a piece of n-space with orientation but without any specific shape. For example, a bivector $3e_1e_2$ will be a part of a plane spanned by e_1 and e_2 with an area of 3 units, and an orientation such that the top of vector e_1 meets the bottom of e_2 if it were a rectangle. However, it might just as well be a circular shape.

A vector v with coordinates $(v_1, ..., v_n)$ in Clifford algebra will then look like $v = \sum_{i=1}^{n} v_i e_i$ for basis $e_1, ..., e_n$. The basis for Cl_n, will consist of 1 element of grade zero, n elements of grade one, $\binom{n}{k}$ elements of grade k, and hence a total of 2^n elements.

We define the addition as follows. Given objects a_i, b_i of grade i and objects a_j, b_j of grade j we define *addition* $a_i + b_i$ as a new element of grade i. For example addition of scalars simply means adding reals together, and addition of vectors can be seen as attaching two vectors tip to tail.

There are two different ways to add e.g. two bivectors, i.e. parallelograms, to each other. 1) When they have at least one vector in common. These vectors will then cancel each other out, creating a new bivector. One could think of attaching one parallelogram to the other, where the new bivector goes between the sides opposite to those that are attached to each other. For example $e_1e_2 + e_1e_3$. 2) When they do not have any vectors in common, they will be a pair of bivectors. This can occur in dimensions 4 and higher, e.g. where one bivector is e_1e_2 and the other is e_3e_4. We can see that $e_1e_2 + e_3e_4 = \frac{1}{2}(e_1 + e_3)(e_2 + e_4) + \frac{1}{2}(e_1 - e_3)(e_2 - e_4)$. This shows that, in 4 dimensions and higher, there will be bivectors which do not represent a plane.

In contrast to vector algebra, we can perform addition of objects of different grade. Hence in Clifford algebra we could have an element C, denoted *clif*, as $C = a_0 + a_1$ where a_0 is a scalar and a_1 is a vector. The

part of C that has grade i is denoted then $\langle C \rangle_i$. When it only consists of parts of the same grade it is called homogenous.

There are various different *multiplication* rules defined on Clifford algebra. The Clifford product AB, the dot product $A \cdot B$ and the wedge product $A \wedge B$ such that $AB = (A \cdot B) + (A \wedge B)$. They can be used to calculate the rotations, but for readability their explanation is left in the appendix.

When V is the Euclidean space of n dimensions, the Clifford algebra Cl_n is generated by the basis vectors e_i such that $e_i^2 = -1$ and $e_i e_j = -e_j e_i$, for $i \neq j$, $i, j \in \{2, n\}$[wol](Clifford Algebra). Which is a result of the Clifford quadratic form relation $x^2 = -|x|^2$, as it gives

$$(ae_i + be_j)^2 = a^2 e_i^2 + b^2 e_j^2 + abe_i e_j + abe_j e_i \text{ and } (ae_i + be_j)^2 = -(a^2 + b^2) \tag{29}$$

The elements in Cl_2 are of the form $A = a_0 1 + a_1 e_1 + a_2 e_2 + a_3 e_1 e_2$, i.e. it is a real, linear space, in 4 dimensions, whose basis elements are $1, e_1, e_2, e_1 e_2$.

A rotation in Clifford algebra will be multiple (pairs of) reflections, hence to show how to express a rotation in Clifford algebra, we start by looking at reflection.

Example: reflection in Cl_2. Let $A = a_1 e_1 + a_2 e_2$ and $B = b_1 e_1 + b_2 e_2$ be vectors in Cl_2. To express a refleciton of A through B we first find $A_{||}$ its projection onto, and $A_\perp$ its length from, B. Since $B^{-1} = B/|B|$, we get:

$$A_{||} = |A| \cos\theta \frac{B}{|B|} = (A \cdot B) B^{-1}$$

$$A_\perp = |A||B| \sin\theta = (A \wedge B) B^{-1} \tag{30}$$

We can then describe A as $A = A_{||} + A_\perp$, and A's reflection on B as

$$A' = A_{||} - A_\perp = (A \cdot B) B^{-1} - (A \wedge B) B^{-1} = ((A \cdot B) - (A \wedge B)) B^{-1} = BAB^{-1} \tag{31}$$

A second reflection, through C, where the angle between B and $C = \phi$, will rotate A at an angle 2ϕ. It is worth noting that these computations only work when A and B are both vectors, and will not work in higher dimensions since the wedge and dot product would work differently in those dimensions.

Example: SO(2), rotation in Cl_2^+. Let the even subalgebra of Cl_2 be the elements $Cl_2^+ \cong \mathbb{C}$ that are of the form $A = a_0 s + a_3 e_1 e_2$, i.e. only even products of basis elements, where $e_i e_j$ represents i. A rotation in 2 dimensions is then represented the following way. Let $R = a + be_1 e_2$ and $R' = a - be_1 e_2$, and a vector $v = (v_1, v_2)$. We find its rotation

$$v' = \begin{pmatrix} v_1' \\ v_2' \end{pmatrix} = \begin{pmatrix} a^2 - b^2 & -2ab \\ 2ab & a^2 - b^2 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = R \cdot v \cdot R' \tag{32}$$

Now letting $a = \cos(\theta/2)$ and $b = \sin(\theta/2)$, this becomes the familiar Euler angle rotation matrix for 2 dimensions, see equation (9).

The pair of reflections that is used in rotation of two dimensions is a reflection through the plane of A followed by a reflection through the plane of B, such that the angle between A and B is ϕ.

Example: quaternions and SO(3). We look at the even subalgebra Cl_3^+, with orthogonal basis e_1, e_2, e_3, and let $i = e_2e_3$, $j = e_3e_1$ and $k = e_1e_2$, given by the quadratic form just as in the previous example. For example we have $i^2 = (e_2e_3)^2 = e_2e_3(-e_3e_2) = -e_2e_3e_3e_2 = -(-1)(-1) = -1$.

Let $A = a_1e_1 + a_2e_2 + a_3e_3$ be vector perpendicular to the bivector $A' = -Ae_1e_2e_3 = a_1e_2e_3 + a_2e_3e_1 + a_3e_1e_2$ through which we want to reflect a vector $v = v_1e_1 + v_2e_2 + v_3e_3$. First we make a few observations.

- Vector v is 3-dimensional, like the one in equation 22, and consists only of complex parts. The vector A and bivector A' are also 3-dimensional.
- Vector A is perpendicular to bivector A'.
- The reflection of v through A is the negation of its reflection through A'.

Similarly to the 2-dimensional case, we get the reflection of v through A as AvA^{-1}. Hence v's reflection through the plane described by A' is $-AvA^{-1}$. A second reflection, through plane described by B will give a rotation around an axis which is perpendicular to both A and B [Lou01].

One can then get Rodrigues' rotation formula, by expressing the axis of rotation k as the norm of the plane of rotation spanned by A and B, [Lou01]

$$k = \frac{A \times B}{|A|\ |B| \sin \alpha} \tag{33}$$

If we want to express a rotation of vector $v = v_1e_1 + v_2e_2 + v_3e_3$ with quaternions, we let $R = q_0 + q_1e_2e_3 + q_2e_3e_1 + q_3e_1e_2$ and $R^{-1} = R' = q_0 - q_1e_2e_3 - q_2e_3e_1 - q_3e_1e_2$. A rotation in 3 dimensions is then represented as RvR^{-1}.

Let $r_1 = \sqrt{0.5} + \sqrt{0.5}e_1e_2$ be a rotation around the z-axis and $r_2 = \sqrt{0.5} + \sqrt{0.5}e_3e_2$ a rotation around the x-axis, then we get the rotation

$$r = r_1r_2 = 0.5 + 0.5e_1e_2 + 0.5e_3e_2 + 0.5e_1e_3 \tag{34}$$

which corresponds to

$$(\sqrt{0.5} + \sqrt{0.5}k)(\sqrt{0.5} + \sqrt{0.5}(-i)) = 0.5 + 0.5k + 0.5(-i) + 0.5(-j) \tag{35}$$

Geometrically the element r would consist of a point 0.5, and bivectors $0.5e_1e_2$, $0.5e_3e_2$ and $0.5e_1e_3$. The bivectors are attached to each other since they share a basis vector, and form a plane with dual vector $0.5(e_1+e_2-e_3)$.

Example: quaternions in Cl_4^+. If we try to do a similar rotation in 4

dimensions the multiplication could be on two bivectors that have no basis vectors in common. Say $r_1 = \sqrt{0.5} + \sqrt{0.5}e_1e_2$ and $r_2 = \sqrt{0.5} + \sqrt{0.5}e_3e_4$, being rotations in two orthogonal planes, giving

$$r = r_1 r_2 = 0.5 + 0.5e_1e_2 + 0.5e_3e_4 + 0.5e_1e_2e_3e_4 \tag{36}$$

We see that for both 3 and 4 dimensions, multiplication of scalars and bivictors gives new scalars and bivectors, and in 4 dimensions also a blade of grade 4. This blade of 4 dimensions is an element which can not be represented by one quaternion alone, which is why we need two sequential quaternions (on each side) to be able to do these kinds of equations.

Also, unlike for 3 dimensions we can have an element consisting of two bivectors, e.g. $e_1e_2 + e_3e_4$ and is hence homogenous but not a blade since they have no dimensions in common with which to attach the vectors to each other.

This is a very important property of SO(4), and shows why rotation in 4 dimensions are more complicated than rotations in 3 or 2 dimensions. They need no longer be a product of two vectors, as in 3 dimensions.

Let $A = a_0 + a_1e_1e_2 + a_2e_2e_3 + a_3e_3e_1$ and $B = b_0e_1e_2e_3e_4 + b_1e_1e_4 + b_2e_2e_4 + b_3e_34_e$ be two unit quaternions, and let $v = v_0 + v_1e_1 + v_2e_2 + v_3e_3$ be a vector we wish to rotate. We can then represent a rotation as AvB^{-1}.

A rotation matrix R rotating vector v in 4 dimensions, Rv, can be decomposed into unit quaternions q and p which are left and right-isoclinic respectively. Meaning that v is multiplied with q on the left, and p on the right, giving qvp. This decomposition is called Cayley's factorisation of 4R rotations [TPG15].

The first effective method for computing Cayley's factorisation was made by van Elfrinkhof [vE97]. Unfortunately, since it was written in Dutch, it did not gain much attention [TPG15].

For a unit quaternions $q = q_0 + q_1i + q_2j + q_3k$ we can get the matrix

$$Q = \begin{pmatrix} q_0 & -q_1 & -q_2 & -q_3 \\ q_1 & q_0 & -q_3 & q_2 \\ q_2 & q_3 & q_0 & -q_1 \\ q_3 & -q_3 & q_2 & q_1 \end{pmatrix} \tag{37}$$

We can construct a similar matrix P for unit quaternion $p = p_0 + p_1 + p_2 + p_3$. It will turn out that the rotation matrix $R = Q \cdot P$. This result was accieved using the fact that unit quaternions have length 1, and that hence the associate matrix R^a of R can also be expressed as:

$$R^a = \begin{pmatrix} q_0p_0 & q_0p_1 & q_0p_2 & q_0p_3 \\ q_1p_0 & q_1p_1 & q_1p_2 & q_1p_3 \\ q_2p_0 & q_2p_1 & q_2p_2 & q_2p_3 \\ q_3p_0 & q_3p_1 & q_3p_2 & q_3p_3 \end{pmatrix} = \begin{pmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{pmatrix} \cdot \begin{pmatrix} p_0 & p_1 & p_2 & p_3 \end{pmatrix} \tag{38}$$

Rotations in 4 dimensions will require more than two pairs of reflections to represent a rotation[Wil05]. For each pair (A_i, B_i) of reflection, the plane $A \times B$ is the plane of rotation. This makes sence, rotations in 4 dimensions are rotations in two planes simultaneously[TPG15], which can also be seen when inspecting the eigenvalues and eigenspaces of the rotation matrix [Tri09].

Higher dimensions The number of pairs required to represent a unique rotation in n dimensions is $\lfloor n/2 \rfloor$ [Han]. The number of even clifs also quickly become large when n grows. For example, a rotation in 5 dimensions is represented by a clif with 1 scalar, $\binom{5}{2} = 60$ bivectors and 5 blades of 4 dimensions. It will also become more difficult to understand the reflection, since we will now reflect a point through 4-dimensional space.

Embedding of Cl_{n-1} into Cl_n. We want only the part of the clifford algebra with even number of reflections to represent rotations. Hence, we are only interested in the even subalgebra of cl_n. However, it turns out that there is a close relation between Cl_{n-1} and the even subalgebra of Cl_n.

This can be seen when looking at $\mathbb{C}$, as we can express it both as the Clifford algebra Cl_1 where $e_1 = i$, and as Cl_2^+, as we have done for the rotations. Similarly, Cl_2 can in fact be seen as representing the quaternions with $e_1 = i$, $e_2 = j$ and $e_1 e_2 = k$, for which we used Cl_3^+ before.

The reason we can not use these 'normal' Cl_1 is that, we need to reflect through a plane. For Cl_1 we have the only reflection $x \to -x$ along a line, since the vector in Cl_1 is only one dimensional.

There is an algebra isomorphism $Cl_{n-1} \to Cl_n^+$ with a function $e_i \mapsto e_i e_n$

The following table describes isomorphisms between Clifford algebras and some other groups and was discovered 1908 by Élie Cartan [Bae01]

n	0	1	2	3	4	5	6	7	8+n
Cl_n^+	$\mathbb{R}$	$\mathbb{C}$	$\mathbb{H}$	$\mathbb{H} \oplus \mathbb{H}$	$\mathbb{H}[2]$	$\mathbb{C}[4]$	$\mathbb{R}[8]$	$\mathbb{R}[8] \otimes \mathbb{R}[8]$	$Cl_{n-8}^+ \otimes \mathbb{R}[16]$

The number of bi-vectors for n dimensions in the Clifford algebra is $\binom{n}{2}$, i.e. the number of planar rotations for n dimensions. [Wil05].

Spinors One subsection of the field of Clifford algebra are the Spinors. These can represent rotation in such a way that they are sensitive to the path and magnitude of the rotation made. They can be extended to any dimension n. However, some of the features the spinors have in 2, 3 or 4 dimensions, can be lost in higher dimensions, which could still make it quite difficult to compute with them. However, there are many applications of spinors, and hence many different programs which can compute with clifford algebra and spinors.

Spinors are those object which have to turn 4π degrees to turn a full turn, at 2π degrees they are their negative. We already saw this when in the section about quaternions, and given some example that they can be

found in every-day life. Even though they are more often used to explain e.g. quantum physics [Wil05]

There is a surjective 2-1-homomorphism from the spin group Spin(n)→ SO(n)with kernel $\{\pm 1\}$. [Lou01]

Spinors also have a Lie algebra, and their double covering connection to SO(n) can be shown through their respective Lie algebra's.

3.1.5 Final notes on representations

Eigenvalues and the different representation. There is a difference between rotations in even compared to odd dimensions. These can be understood by looking at the Eigenvalues and Eigenspaces of the Euler angle representations [Tri09]. The n eigenvalues, scalars, λ for a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ have the property that $\boldsymbol{AX} = \lambda \boldsymbol{X}$ for some vector $\boldsymbol{X} \in \mathbb{R}^n,\ X \neq \boldsymbol{0}$, called the eigenvector. Also for all $\boldsymbol{r} \in \mathbb{R}^n$ that $|\boldsymbol{Ar}| = |\lambda \boldsymbol{r}| = |\boldsymbol{r}|$, since $\boldsymbol{A} \in$SO(n).

This way we can create the characteristic polynomial of $\boldsymbol{A}$ as

$$p(\lambda) = \det[\lambda I - A] = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0 = \prod_{i=1}^{n}(\lambda - \lambda_i) \quad (39)$$

Next, $\det A = \lambda_1 \cdots \lambda_n = 1$ since A is a rotation matrix, where each λ_i is a root, with multiplicity. This means that the only real roots are $\lambda_i = \pm 1$, also if $\lambda_i = a + bi$ is a root, then there is a root $\lambda_j = a - bi$. These roots will define the planes in clifford algebra.

In odd dimensions at least one of the roots $\lambda_i = +1$. For this eigenvalue we can compute the eigenspace. This eigenspace will be of one dimension defining the vector which will not be effected by the rotation A. For example in 3 dimensions, the eigenspace of $\lambda_i = 1$ will be the axis of rotation [Eulers rotation theorem]. When more roots are 1, they will define the space which remains uneffected by the rotation [Tri09] [Lou01].

For the roots $\lambda_i = a + bi$ and $\lambda_j = a - bi$ we can compute the eigenspace which will define the plane and angle of rotation [Lou01]. For example for 2 dimensions the roots for the characteristic polynomial are $\lambda = \cos\theta \pm i \sin\theta$, which corresponds to a rotation of θ. For 3 dimensions the roots of the characteristic polynomial are $\lambda_1 = 1$ and $\lambda = \cos\theta \pm i \sin\theta$. This corresponds to a rotation of θ around the eigenspace of $\lambda_1 = 1$. [Tri09]

Hence one can use eigenvalues and eigenspace of matrix A to convert Euler angles to quaternions. Also, knowing the rotation axis and angle of rotation one can use the Rodrigues rotation formula to create a matrix A [wol](Rodrigues Rotation Formula).

For all rotations $\boldsymbol{A}$ in odd dimensions one can always find points p, other than the origin, which are not effected by the rotation A. However, for even dimensions there are rotations where the only point not effected by the rotation, is the origin. Also $-I \in$SO($2n$), while $-I \notin$SO($2n+1$).

Advantages and disadvantages. One of the main reasons to choose Clifford algebra before Euler angles could be that Clifford algebra is much more coordinate independent, which is seen e.g. in the fact that bivectors have area but no specific shape. Otherwise the choice between Euler angles or Clifford algebra relies partially on interpretation versus speed.

The Euler angles computations are easier to interpret, as we can compute the angle we rotate for each plane of rotation. They work the same way independently of the number n of dimensions. It is more difficult to interpret what the numbers in a Clifford rotation correspond to. On top of that, it will become more and more difficult to interpret the larger n becomes.

It could be argued, that a rotation in high dimensions is difficult to interpret independently of representation method. Even though the rotation itself is of angle θ, to understand it, we need to break it up into multiple rotations, in the Euler angle case it will be $\binom{n}{2}$. As mentioned in the first section, interpretation is important to understand the variable selection that is made.

On the other hand, computing rotations with Euler angles is slow [EG03]. In recent years toolboxes have been constructed to facilitate computation with Clifford algebra [Dor01] [Den]. Note, however, that if one has to keep in mind that a conversion between the different reprentations can be quite expensive, and might be needed to take into consideration when it has to be done a lot.

However, the toolbox in [Den] only works in 3 dimensions, while the toolbox in [Dor01] claims

> Programs written are magically insensitive to dimensionality of the embedding space, or of the objects they act on.

But this toolbox seems to still be in a developing phase, since most of his latest work does not seem to go much beyond 3 dimensions, and concern e.g. quaterions.

The toolbox *Clical* seems to be able to handle higher dimensions [LMV87]. They also mention other toolboxes which work with e.g. matlab (made by Dorst and others), which can compute with Clifford algebra but unfortunately does not deal with higher dimensions.

There does seem to be research on combining clifford algebra with Neural networks, Support (multi-)vector maschines, classification and other variable selection techniques [LH14] [EJ01] [EN10]. However, due to lack of access I have not been able to check how relevent they are to this paper.

Lie algebra. It turns out that $\mathrm{SO}(n)$ is in fact a Lie group, i.e. a group which is also a smooth manifold. This means that it has a Lie algebra, which can facilitate computations. The spinors also turn out to have a Lie algebra, which will be a double cover of the Lie algebra of $\mathrm{SO}(n)$. See the appendix

for more details.

Even versus odd dimensions. We would like to end this section with some notes on the difference between rotating in even and odd dimensions. As we have already seen before we can only find points uneffected by the rotation when n is odd. We quickly mention some other differences and refer the reader to some further reading.

- Difference between O(n) and SO(n) in even dimensions, maximal tori and weyl groups.
- Wiki Orthogonal group: for odd dimensions, the lie algrebra is so(2r+1) where r is rank. For even dimensions the rank of the lie algebra is so(2r).
- Reflection through a point is: Orientation preserving in even dimensions, orientation reversing in odd dimensions. for O(n), reflection not for special orthogonal group.

3.2 Metrics for rotation

3.2.1 Metric properties

Recall the properties of a metric:

Let S be the space within which one has elements whose distance to each other one wants to measure. A *metric* or *distance function* $\phi : S \times S \to \mathbb{R}$ is defined with the following properties:

- $\phi(x, y) = 0 \Leftrightarrow x = y$
- $\phi(x, y) = \phi(y, x)\ \forall x, y \in S$
- $\phi(x, z) \leq \phi(x, y) + \phi(y, z)\ \forall x, y, z \in S$

Also we define a metric ϕ to be *left invariant* if

- $\phi(x, y) = \phi(zx, zy)\ \forall x, y, z \in S$

Similarly ϕ is *right invariant* if $\phi(x, y) = \phi(xz, yz)\ \forall x, y, z \in S$, and when ϕ is both left and right invariant it is called *bi-invariant.*

3.2.2 Euclidean distance

Let the Euclidean distance from point $p_1 \in \mathbb{R}^n$ to point $p_2 \in \mathbb{R}^n$ be the norm $||p_1 - p_2||_i = \sqrt[i]{\sum_n (p_1 - p_2)_i}$ for $i = 2$. However, since for a particular order of rotation, $R(\pi, \pi, 0) = R(0, 0, \pi)$, this can not be a distance function since $\phi((\pi, \pi, 0), (0, 0, \pi)) = 0$ despite $(\pi, \pi, 0) \neq (0, 0, \pi)$. This could be remedied by adding the conditions $\alpha, \gamma \in [-\pi, \pi)$ and $\beta \in [-\pi/2, \pi/2)$, and we get:

$\phi_1 : \mathrm{SO}(n) \times \mathrm{SO}(n) \to \mathbb{R}$ *s.t.*

$$\phi_1 = ||p_1 - p_2||_2 = \sqrt[2]{\sum_n (p_1 - p_2)_2} \quad (40)$$

for $p_i = (\theta_1, ..., \theta_n) \in \mathbb{R}^n,\ i \in [1, 2],\ \theta_1 \in [0, 2\pi),\ \theta_j \in [0, \pi), j \in [2, n]$.

Now lets see what happens in case of a Gimbal lock. Let

$$R(\alpha, \beta, \gamma) \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} =$$

$$\begin{bmatrix} p_1 \mathrm{cos}\beta\mathrm{cos}\gamma - p_2\mathrm{cos}\beta\mathrm{sin}\gamma + p_3\mathrm{sin}\beta \\ p_1(\mathrm{cos}\alpha\mathrm{sin}\gamma - \mathrm{sin}\alpha\mathrm{sin}\beta\mathrm{cos}\gamma) + p_2(\mathrm{cos}\alpha\mathrm{cos}\gamma - \mathrm{sin}\alpha\mathrm{sin}\beta\mathrm{sin}\gamma) - p_3\mathrm{sin}\alpha\mathrm{cos}\beta \\ p_1(\mathrm{sin}\alpha\mathrm{sin}\gamma - \mathrm{cos}\beta\mathrm{sin}\beta\mathrm{cos}\gamma) + p_2(\mathrm{sin}\alpha\mathrm{cos}\gamma + \mathrm{cos}\alpha\mathrm{sin}\beta\mathrm{sin}\gamma) + p_3\mathrm{cos}\alpha\mathrm{sin}\beta \end{bmatrix} \quad (41)$$

$$\phi((\alpha_1, \beta_1, \gamma_1), (\alpha_2, \beta_2, \gamma_2)) = \sqrt{(\alpha_1 - \alpha_2)^2 + (\beta_1 - \beta_2)^2 + (\gamma_1 - \gamma_2)^2} \quad (42)$$

$\phi((\pi, \pi, 0), (0, 0, \pi)) = \sqrt{3\pi^2}$ while on the other hand we have that $R(\pi, \pi, 0)\,[p_1, p_2, p_3]^T - R(0, 0, \pi)\,[p_1, p_2, p_3]^T = [-p_1 + p_1, -p_2 + p_2, 0 + 0]^T = [0, 0, 0]^T$

Now we look more closely at what happens in the case of a gimbal lock, i.e. when $\beta = \pi/2$ and $\alpha_1 + \gamma_1 = \theta = \alpha_2 + \gamma_2$ which gives rise to the same rotation. We get the following distance:

$$\sqrt{(\alpha_1 - \alpha_2)^2 + (\pi/2 - \pi/2)^2 + ((\theta - \alpha_1) - (\theta - \alpha_2))^2} = \sqrt{2(\alpha_1 + 2\alpha_1\alpha_2 + \alpha_2^2)} = \sqrt{2}(\alpha_1 - \alpha_2) \quad (43)$$

Which is independent of the total angle θ and, moreover, when $\alpha_1 \neq \alpha_2$, the equation is not equal to 0, even though the rotations are the same. This shows that the Euclidean metric can depend on the representation of a specific point.

One might wonder why this trouble occurs. Looking at the definition of Euclidean distance, and that of Euclidean rotation, it is easy to see that while the Euclidean distance is indifferent of the order, i.e. it remains the same whether we compute $\phi((\alpha_1, \beta_1, \gamma_1), (\alpha_2, \beta_2, \gamma_2))$ or say $\phi((\gamma_1, \beta_1, \alpha_1), (\gamma_2, \beta_2, \alpha_2))$, whilst the Euclidean rotation, with a few specific exception, depends very heavily on the order in which the rotations in the different planes are made.

Hence, while the matrix multiplicity would give two different answers, Euclidean measure fails to see a difference, or vice versa. This means that

small distances can have large value and vice versa [Huy09]. Similar problems occur in higher dimensions. One might want a metric which keeps in mind the order of multiplicity with matrices, but which somehow cancels such things out.

This might seem rather discouraging for the use of Euclidean angles. Fortunately there will turn out to be other ways to compute the distance for rotations represented with Euclidean angles. First we take a look at how to do things with quaternions and Clifford algebra.

On a final note on the defense of the Euclidean distance, if we have established a base in which to represent all rotations and we are only interested in finding those points p which are of a specific Euclidean distance d from a point $q \in S^n$, the Euclidean distance is a reliable tool. However, as soon as we would like to compare distances with each other, since a twice a Euclidean distance in e.g. 2 dimensions, does not mean that we have twice the distance over the circle i.e. twice the angle.

3.2.3 Quaterion metrics in 3 dimensions

In three dimensions on can construct metrics using unit quaternions.

When using *quaternions* we measure the distance between two points $\mathbf{q_1}, \mathbf{q_2} \in S^3$, however we would like two points opposit to each other on the 3-sphere to be equal, see 3.1.3. This gives $\phi_2 : S^3 \times S^3 \to \mathbb{R}^+$ such that [Huy09][LaV06]:

$$\phi_2(\mathbf{q}_1, \mathbf{q}_2) = \min\{||\mathbf{q}_1 - \mathbf{q}_2||, ||\mathbf{q}_1 + \mathbf{q}_2||\} \tag{44}$$

where $||\cdot||$ denotes the Euclidean norm.

Note how this is not a metric of S^3, and the quaternions, since for one $\phi(\mathbf{q}, -\mathbf{q}) = 0$ even though they are two different points on S^3, unlike in SO(3). That ϕ_2 has the second metric property is clear, and to show the third property we have:

$$\begin{aligned}
&\phi_2(\mathbf{q}_1, \mathbf{q}_2) + \phi_2(\mathbf{q}_2, \mathbf{q}_3) = \\
&\quad \min\{||\mathbf{q}_1 - \mathbf{q}_2||, ||\mathbf{q}_1 + \mathbf{q}_2||\} + \min\{||\mathbf{q}_2 - \mathbf{q}_3||, ||\mathbf{q}_2 + \mathbf{q}_3||\} \\
&\quad\quad \min\{||\mathbf{q}_1 - \mathbf{q}_2|| + ||\mathbf{q}_2 - \mathbf{q}_3||, ||\mathbf{q}_1 - \mathbf{q}_2|| + ||\mathbf{q}_2 + \mathbf{q}_3||, \\
&\quad\quad\quad ||\mathbf{q}_1 + \mathbf{q}_2|| + ||\mathbf{q}_2 - \mathbf{q}_3||, ||\mathbf{q}_1 + \mathbf{q}_2|| + ||\mathbf{q}_2 + \mathbf{q}_3||\} \geq \\
&\quad\quad\quad\quad \min\{||\mathbf{q}_1 - \mathbf{q}_3||, ||\mathbf{q}_1 + \mathbf{q}_3||\}
\end{aligned} \tag{45}$$

Note also the similarity between this metric and the Euler distance, as we with this metric measure the distance between the two points on S^3, instead of an angle.

Unfortunately it becomes quite difficult to see how to extend this metric to Clifford algebra of higher dimensions in a meaningfull way. This could be since the blades of higher grade, i.e. the parts of e.g. $e_i e_j$ or $e_i e_j e_k e_l$

have an area or volume, but no specific shape. This means we would need a metric which is valid for length, area and volume.

The metric which could be applied to bivectors depends on being able to rewrite the bivectors as simple vectors. The even subalgebra Cl_n^2 has as orthogonal basis the unit bivectors $\boldsymbol{i_k} = e_n e_k$, $k \in [1, n-1]$. Since $e_i e_k = -e_k e_i$ and $e_k e_k = -1$ they create a so called negative definite metric [Lou01]. For a vector $v \in \mathbb{R}^n$ such that $v = v_1 e_1 + ... + v_n e_1$ we get the new expression $w = v_1 e_1 e_n + ... + v_{n-1} e_{n-1} e_n + y$ where $y = v_n$ is a scalar, hence w is called a paravector.

There exists another metric for quaternions. This method measures the angle using the geometric interpretation of the dotproduct *cdot* of two vectors $v_1, v_2 \in \mathbb{R}^n$, where θ is an angle:

$$v_1 \cdot v_2 = ||v_1||\ ||v_2|| \cos(\theta)\ [= \cos(\theta)] \tag{46}$$

where the second equality is valid when we deal with unit quaternions.

Similarly to the previous metric one would need to identify $\mathbf{q}$ with $-\mathbf{q}$ on S^3 for the quaterions to represent SO(3), giving $\phi_3 : S^3 \times S^3 \to \mathbb{R}^+$ such that

$$\phi_3(\mathbf{q}_1, \mathbf{q}_2) = \cos^{-1} |\mathbf{q}_1 \cdot \mathbf{q}_2| \tag{47}$$

where $\cdot$ is the dotproduct of vectors.

It is easy to see that ϕ_3 has the properties of a metric since $\cos\theta = 0$ only when $v_1 = \pm v_2$, $v_1 \cdot v_2 = v_2 \cdot v_1$ and $v_1 \cdot v_2 + v_2 \cdot v_3 \geq v_1 \cdot v_3$.

Note that this too is a metric on SO(3), but not on S^3 (and can therefor not easily be expanded to higher dimensions). It also is clear that, since rotation preserves angles between v_1 and v_2, this metric is bi-invariant.

Examples for 3 dimensions: Let $\mathbf{n} = (n_1, n_2, n_3) \in \mathbb{R}^3$ be a unit vector around which we will rotate, i.e. $|\mathbf{n}| = 1$, and let $\mathbf{q_1} = x + |\sqrt{1-x^2}|(n_1 i + n_2 j + n_3 k)$, $\mathbf{q_2} = y + |\sqrt{1-y^2}|(n_1 i + n_2 j + n_3 k) \in \mathbb{H}$ such that $|\mathbf{q_1}| = |\mathbf{q_2}| = 1$. Then we get

$$\begin{aligned}\phi_2(\mathbf{q_1}, \mathbf{q_2}) &= \min_{\pm}\left(\sqrt{(x \pm y)^2 + ((1-x^2) \pm (1-y^2))|n|}\right) = \\ &\min\left(\sqrt{(x+y)^2 + (2-x^2-y^2)}, \sqrt{(x-y)^2 + (y-x)^2}\right) = \\ &\min\left(\sqrt{2xy+2}, \sqrt{2}(x-y)\right)\end{aligned} \tag{48}$$

While, using $\cos(a-b) = \cos a \cos b + \sin a \sin b$,

$$\begin{aligned}&\phi_3(\mathbf{q_1}, \mathbf{q_2}) = \cos^{-1}(|xy + |\sqrt{1-x^2}\sqrt{1-y^2}|\ |n|\ |) = \\ &\cos^{-1}(|xy + |\sqrt{(1-x^2)(1-y^2)}|\ |) = [x = \cos(\theta_1/2),\ y = \cos(\theta/2)] = \\ &\cos^{-1}(|\cos((\theta_1 - (sign(xy))\theta_2)/2)|) = (\theta_1 - (sign(xy))\theta_2)/2\end{aligned} \tag{49}$$

Which is fortunate since $\theta_1/2$ and $\theta_2/2$ are the angles of rotation for $\mathbf{q_1}$ and $\mathbf{q_2}$.

Next, let $\mathbf{q_1} = \cos(\pi/4) + |\sin(\pi/4)|(1+0+0)$ and $\mathbf{q_2} = \cos(\pi/4) + |\sin(\pi/4)|(0+1+0)$ be a unit quaternions with rotation of $\pi/2$ around an axis. Then we get

$$\begin{aligned}\phi_2(\mathbf{q_1}, \mathbf{q_2}) = \\ \min\left(\sqrt{(2\cos(\pi/4))^2 + \sin^2(\pi/4) + \sin^2(\pi/4)}, \sqrt{\sin^2(\pi/4) + \sin^2(\pi/4)}\right) = \\ \min\left(\sqrt{2 + 2\cos^2(\pi/4)}, \sqrt{2\sin^2(\pi/4)}\right) = \min\left(\sqrt{3}, 1\right) = 1 \quad (50)\end{aligned}$$

While,

$$\phi_3(\mathbf{q_1}, \mathbf{q_2}) = \cos^{-1}(\cos^2(\pi/4)) = \cos^{-1}((\frac{\sqrt{2}}{2})^2) = \cos^{-1}(\frac{1}{2}) = \pi/3 \quad (51)$$

This could give a better understanding of how the two different metrics work and why one would prefer one before the other when working with angles and quaternions.

It can be shown that ϕ_2 and ϕ_3 are boundedly equivalent and that when two metrics ϕ_i and ϕ_j are boundedly equivalent and one of them is bi-invariant, then so is the other [Huy09]. Hence ϕ_2 is also bi-invariant.

Though a dotproduct exists in Clifford algebra, it is only well defined for vectors [Lou01] [Den], attempts seem to have been made to generalise the dotproduct to clifs of higher grade in [Bak16], who points out that other extended dot products might exist. However we might not be able to use these in either case, since the geometric interpretation might not be the same.

The dot product $v \cdot K = \frac{1}{2}(vK + (-1)^{k-1}Kv)$, for vector v and blade K of grade k, proposed by [Bak16] can not multiply bivector with bivector. However, we know we can rewrite quaternions as $q = q_0 + q_1e_2e_3 + q_2e_3e_1 + q_3e_1e_2$ and perform dot-product on them.

3.2.4 Geodesics - a metric using Lie algebra

Given any surface S, with $a, b \in S$, a geodesic is the shortest path from a to b across the surface S. This surface S can be e.g. a plane or a sphere.

In [PR97] (cited in [Huy09]) the metric is given as $\phi_4 : \mathrm{SO}(3) \times \mathrm{SO}(3) \to \mathbb{R}$ such that

$$\phi_4(R_1, R_2) = ||\log(R_1R_2^t)|| \quad (52)$$

Here we use that $\mathrm{SO}(n)$ are smooth manifolds, and hence Lie groups (denoted $\mathfrak{SO}(n)$). This means we can create their Lie algebra $\mathfrak{so}(n)$ with the exponentional map on the Lie group, i.e. for $R \in SO(n)$ we can find its Lie

algebra representative $r = e^R \in \mathfrak{so}(n)$. The elements $r \in \mathfrak{so}(n)$ turn out to be the skew $n \times n$ matrices.

In (52) log is the inverse of the exponential map from the Lie algebra, which is surjective since $\mathfrak{SO}(n)$ is compact. Since r is skew-symmetric r^2 is symmetric, meaning that $||r|| = \frac{1}{2}\sqrt{Tr(r^2)}$, where Tr denotes the trace of a matrix.

The Lie algebra $\mathfrak{so}(3)$ can be shown to be of the from:

$$[\mathbf{r}] = \begin{pmatrix} 0 & -r_3 & r_2 \\ r_3 & 0 & -r_1 \\ -r_2 & r_1 & 0 \end{pmatrix} \tag{53}$$

Using such notation the Lie bracket of

$$[\mathbf{r_1}, \mathbf{r_2}] = [\mathbf{r_1}][\mathbf{r_2}] - [\mathbf{r_2}][\mathbf{r_1}] = [\mathbf{r_1} \times \mathbf{r_2}] \tag{54}$$

It can be shown that $\forall \mathbf{R}\ \exists \mathbf{r}\ e^{[\mathbf{r}]} = \mathbf{R}$. Let γ be the curve such that $\gamma(s) : \mathbb{R} \ni s \mapsto e^{s[r]}$ will be a geodesic in the bi-invariant metric on SO(n) using taylor expension. Hence it will give the shortest path between $\gamma(0) = I$ and $\gamma(1) = e^{[r]} = R$.

To get the shortest path between R_1 and R_2 we just note that it is the same as the shortest path between I and $R_1^T R_2$.

A geodesic can be computed using the Riemann integral on the Lie Algebra.

Theorem: 13. *If* $\mathbf{R} \in$ *Lie group* $\mathfrak{SO}(n)$, $\mathbf{r} \in$ *Lie algebra* $\mathfrak{so}(n)$ *then* $\mathbf{RrR^{-1}} \in$ *Lie algebra* $\mathfrak{so}(n)$.

Hence the square geodesic distance between I and $R \in$SO(n) is

$$\int_0^1 \langle \gamma(t)^{-1} \boldsymbol{r} \gamma(t), \gamma(t)^{-1} \boldsymbol{r} \gamma(t) \rangle dt = \langle \boldsymbol{r}, \boldsymbol{r} \rangle = Tr(\boldsymbol{r}^2) \tag{55}$$

by in variance under conjugacy.

In 3 dimensions the algorithm can be shown to be as fast as ϕ_3 when using quaternions, i.e. ϕ_3 and ϕ_4 have the same computational complexity [Huy09]. However the cross product and the logarithm will become more complicated when $n > 3$.

3.3 Haar measure

3.3.1 Introduction

It is very conventient to have an invariant measure for rotations. They can be created, for topological compact groups (such as SO(n)) and use integration.

Consider a box (hyperrectangle) $B = ((a_1, b_1), (a_2, b_2), ..., (a_n, b_n))$ in $\mathbb{R}^n$. The length of side i can be computed with Lebesque measure we are all familiar with $b_i - a_i$, giving the volume

$$V(B) = \sum_{i=0}^{n} (b_i - a_i) \tag{56}$$

With this measure, we can move the box around in $\mathbb{R}^n$ as much as we like, without the volume ever changing. Which should come as no suprise. In other words:

$$V(B) = V(t + B) = \sum_{i=0}^{n} ((t_i + b_i) - (t_i + a_i)), t \in \mathbb{R}^n \tag{57}$$

This means the Lebesque measure is invariant to a translation t. Moreover, it is both left and right invariant, i.e. *bi-invariant* or unimodular, which means $V(t + B) = V(B) = V(B + t)$.

Theorem: 14. *The Haar measure μ of a compact group G, with operation $\circ$ is bi-invariant.*

$$\forall g \in G, \mu(f \circ g) = \mu(f) = \mu(g \circ f)$$

3.3.2 Definition

Let $f : SO(n) \to \mathbb{R}$ be a continuous function, and let $C = \{f\}$. For all $f \in C$, let $|f| = \max_{x \in SO(n)} |f(x)|$ be its norm defining a metric $d(f_1, f_2) = |f_1 - f_2|$ on $C = \{f | f : SO(n) \to \mathbb{R}\}$. The norm is complete and defines a topology on the space [Not].

Let μ be a measure on SO(n) such that $\mu : C(SO(n), \mathbb{R} \to \mathbb{R}$, where $\mu(f)$ can also be denoted as $\int_{SO(n)} f d\mu$ or $\int_{SO(n)} f(x) d\mu(x)$. Let

- μ is linear: $\mu(a_1 f_1 + a_2 f_2) = a_1 \mu(f_1) + a_2 \mu(f_2)$.
- μ is continuous: $\forall \epsilon \exists \delta$ s.t. $|f| < \delta \Rightarrow |\mu(f)| < \epsilon$.

Theorem: 15. *Suppose G is a compact group. Then there exists a unique real measure μ on G such that*

- *μ is invariant on G, ie*

$$\int_G (gf)\ d\mu = \int_G f\ d\mu$$

  for all $g \in G$, measures $f \in C = \{f : G \to \mathbb{R}$.

- *μ is normalized so that G has volume 1, ei*

$$\int_G 1\ d\mu = 1$$

- *measure f is strictly positive, ei*

$$f(x) \geq 0 \text{ for all } x \Rightarrow \int f \, d\mu \geq 0$$

with equality if $f = 0$, i.e. $f(g) = 0$ for all g.

- $|\int_G f \, d\mu| \leq |f|$.

3.3.3 Haar measure for SO(n)

A lot of litterature on Haar measure is rather theoretical and concentrates on the long technical proof showing that (on any smooth group) a Haar measure always exists and is unique [Not][Tay].

One way to construct a Haar measure for SO(n) is to use its connection to S^{n-1}[Bel]. Recall from 3.1.2 the following properteis of σ describing a point on the sphere:

(i) $\sigma^n(\boldsymbol{\theta}_n) = \sin(\theta_n)\sigma^{n-1}(\boldsymbol{\theta}_{n-1}) + \cos\theta_n \vec{\boldsymbol{e}}_{n+1}$.

(ii) $||\sigma^n(\boldsymbol{\theta}_n)|| = 1$ since it is a point on the $n+1$-sphere.

To find the partial derivative of σ^n we first look at the case θ_n and θ_{n-1}:

$$\frac{\partial \sigma^n}{\partial \theta_n} = \frac{\partial \sin\theta_n(\sigma^{n-1}(\boldsymbol{\theta}_{n-1}))}{\partial \theta_n} + \frac{\partial \cos\theta_n \vec{\boldsymbol{e}}_{n+1}}{\partial \theta_n} \tag{58}$$

$$\frac{\partial \sigma^n}{\partial \theta_{n-1}} = \frac{\partial \sin\theta_n(\sigma^{n-1}(\boldsymbol{\theta}_{n-1}))}{\partial \theta_{n-1}} + \frac{\partial \cos\theta_n \vec{\boldsymbol{e}}_{n+1}}{\partial \theta_{n-1}} = \sin\theta_n \left(\frac{\partial \sin\theta_{n-1}(\sigma^{n-1}(\boldsymbol{\theta}_{n-1}))}{\partial \theta_{n-1}} + \frac{\partial \cos\theta_{n-1} \vec{\boldsymbol{e}}_n}{\partial \theta_{n-1}} \right) + 0 \tag{59}$$

Hence it is easy to see that the partial derivative of σ^n for case θ_i becomes:

$$\frac{\partial \sigma^n(\boldsymbol{\theta}_n)}{\partial \theta_i} = \sin\theta_n \cdots \sin\theta_{i+1} \frac{\partial \sigma^i(\boldsymbol{\theta}_i)}{\partial \theta_i} \tag{60}$$

This makes it easy to see that the length (or weight) of such a partial derivative is

$$||\frac{\partial \sigma^n(\boldsymbol{\theta}_n)}{\partial \theta_i}|| = \sin\theta_n \cdots \sin\theta_{i+1} ||\frac{\partial \sigma^i(\boldsymbol{\theta}_i)}{\partial \theta_i}|| = \sin\theta_n \cdots \sin\theta_{i+1} \sqrt{\cos^2\theta_i ||\sigma^{i-1}(\boldsymbol{\theta}_{i-1})|| + \sin^2\theta_i} = \sin\theta_n \cdots \sin\theta_{i+1} \cdot 1 \tag{61}$$

Hence the volume form in n dimensions, with respect to σ^n, is

$$dV = (\sin\theta_2)(\sin^2\theta_3) \cdots (\sin^{n-1}\theta_n) d\theta_1 \cdots d\theta_n \tag{62}$$

where $\theta_1 \in [0, 2\pi]$ and $\theta_i \in [0, \pi]$ $\forall i \in [2, n]$.

Now we would like the corresponding Haar measure for Euler angles. We use $\Omega(\theta^1, \boldsymbol{\theta}^2, ..., \boldsymbol{\theta}^n)$ from 3.1.2 to find the Haar measure for SO(n). We get

$$dV = \left(\prod_{1 \leq i \leq j \leq n-1} \sin^{i-1} \theta_{ij} \right) d\theta_{11} ... d\theta_{n-1,n-1} \tag{63}$$

Where $\theta_1^i \in [0, 2\pi]$ and $\theta_j^i \in [0, \pi]$ for $2 \leq j \leq i$.

Now we can compute the volume for SO(3) as

$$\begin{aligned} V^3 = \int_{\theta_{1,1}} \int_{\theta_{1,2}} \int_{\theta_{2,2}} & \sin \theta_{2,2} d\theta_{1,1} d\theta_{1,2} d\theta_{2,2} \\ & = [2\pi - 0][2\pi - 0][-\cos(\pi) - (-\cos(0))] = 8\pi^2 \end{aligned} \tag{64}$$

However, since Haar measure for a group must equal 1, we have to divide the integration with $8\pi^2$. We can now define the Haar measure, represented with Euler angles, in SO(n) as

$$\frac{1}{V^n} \int_{\theta_{1,1}} \int_{\theta_{1,2}} \int_{\theta_{2,2}} \cdots \int_{\theta_{n-1,n-1}} \left(\prod_{1 \leq i \leq j \leq n-1} \sin^{i-1} \theta_{ij} \right) d\theta_{11} ... d\theta_{n-1,n-1} \tag{65}$$

For SO(2) the Euclidean distance (along the circle) and the Haar measure coincide (though Haar measure multiplicity $1/2\pi$ to get volume 1 for full circle.) The Haar measure is unique up to multiplication by a constant.

This gives a measure of how large a neighbourhood of an element $X \in$ SO(n) is in terms of Euler angles.

4 The deterministic rotation matrix

To see what the rotation matrix in Nordlings uncertainty system could look like, we would first need to take a closer look at *pspace*, the space covered by $\overline{\boldsymbol{X}+\boldsymbol{u}}$. Otherwise it will not be possible to show we have actually found a rotation which rotates the regressand into the *pspace*, let alone proving whether it is the most optimised or not.

However, this first step will already turn out to be quite a tough nut to crack. We will start with giving a different representation of the uncertainty cone and see how it can be used to more efficiently find distances between the projection of two cones onto the sphere.

4.1 The uncertainty cone

4.1.1 Definition of uncertainty cone through angle

We want a definition of the uncertainty cone which would depend on some angle θ instead of a length u_i in $\mathbf{u} = (u_1, u_2, ..., u_n)$. Recall that, for any measured $\mathbf{X}$ with uncertainty $\mathbf{u}$, the elements in the uncertainty cone of $\mathbf{X}$ are:

$$\{tX : t \in \mathbb{R}, X = R_{\theta'}(X) \in \mathcal{N}(\mathbf{X}, u)\} \tag{66}$$

Here U_x is the uncertainty of x as described in TN. We want to describe it such that it depends on a some angle θ instead.

Definition 3. *Given a set of regressors with uncertainty, $\overline{\mathbf{X_i}+\mathbf{u_i}}$, their projection on the sphere S^{n-1} would consist of all rotations of $\mathbf{X_i^*}$ of angle at most θ_i from $\mathbf{X_i}$ on the sphere, giving the uncertainty cone:*

$$\mathcal{C}(\mathbf{X_i}, \theta_i) = \{tR(\theta_i', \mathbf{n})\mathbf{X_i} | t \in \mathbb{R}, \theta_i' \leq \theta_i, \mathbf{n} \in \mathbb{R}^m\} \tag{67}$$

The set of uncertainty cones is then denoted $\mathcal{C}(\mathbf{X}, \theta)$ where row i denotes the uncertainty cone of regressor i.

The size of the angle is found by:

Theorem 2. *Nordlings uncertainty cone defined through uncertainty length u, can be redefined through uncertainty on angle θ such that*

$$\mathcal{C}(\mathbf{X_i}, u_i) = \mathcal{C}(\mathbf{X_i}, \sin^{-1}(\frac{u_i}{|x|})) \tag{68}$$

Proof. To compute θ, we first observe that the maximum angle of rotation is independent of the direction in which we rotate. Hence we can reduce our problem of finding θ to being a 2-dimensional one regardless of the number n dimensions by looking at the plane in which we are rotating. This plane would be defined by the regressor X and the direction of the rotation, but

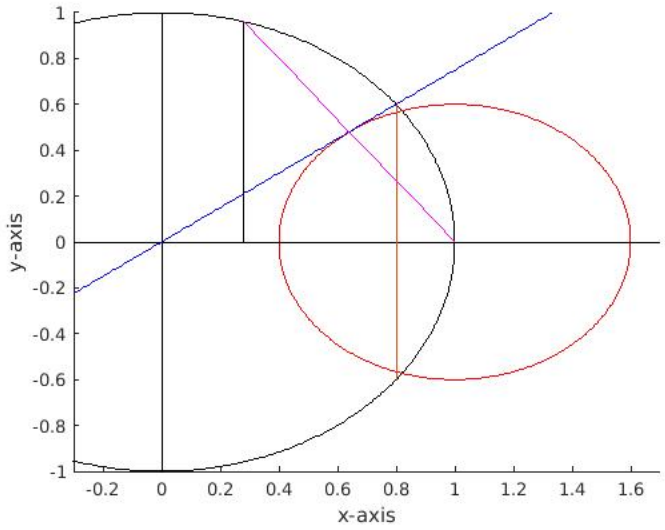


Figure 3: How to find the angle θ

Let the larger, black, circle be the unit circle. The blue line is tangent to the smaller, red, circle of radius u, and is perpendicular to the magenta line creating a chord to the unit circle of lengt $2u$. Hence the blue line is at an angle of $\theta = \sin^{-1}(u)$ from the x-axis. Note that a rotation θ of a regressor $\boldsymbol{X}$ will then rotate outside of its uncertainty set, even though it is still within the cone.

we would assume that is the x-y plane without loss of generality. We get something looking like figure 3 .

First we let x lie on the unit circle, hence scaling all lengths by $1/|x|$. Then we get the uncertainty circle of radius $u/|x|$ (red in the figure). Take one of the two lines tangent to this circle, that also goes through the origin. Using it as a mirror, we get a second circle, with origin at x^* lying on the unit circle, such that the (pink) line, called chord, between x and x^* is $2u$. We get the following computation:

$$|x|\text{chord}(2\theta) = 2u \;\Rightarrow\; \text{chord}(2\theta) = 2u/|x| \;\Rightarrow\; \theta = \sin^{-1}(u/|x|) \tag{69}$$

with the chord defined as: $\text{chord}(\theta) = 2\sin(\theta/2)$.

This means that the uncertainty cone of $\mathbf{X_i}$ can be described as rotating a degree $\theta_i \leq 2\sin(u/|\mathbf{X_i}|)$ in any direction. □

Now that we know the angle θ we can compute the 'distance' d between two uncertainty cones of $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$ by first computing the distance between $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$ on the unit sphere with our preferred metric ϕ, and then subtracting the uncertainty angles. When d is negative, the two cones intersect. Note that this means that d is not a metric.

$$d = \phi(\boldsymbol{X_1}, \boldsymbol{X_2}) - \theta_1 - \theta_2 \tag{70}$$

4.1.2 The problem with two intersected uncertainty cones

One immediate problem which arises when two cones intersect (on more than a single point) is that *pspace* in one instant becomes all of $\mathbb{R}^m$. For any point X_1 in the intersection we can find a small enough neighbourhood which is also completely within the intersection. Hence for any point $p \in \mathbb{R}^m$ we can find a point X_2 in that neighbourhood such that $X_1 + tX_1\vec{X}_2 = p$ for some $t \in \mathbb{R}$.

Although mathematically correct, the resulting *pspace* is probably not very interesting, or helpfull, at least not when t starts becoming very large. One way to avoid such a situation would be to use the variable selection technique called *clustering* (see 2.1.2) and merge two regressors together. This could be done by letting the new regressor be $u_2/(u_1+u_2)\boldsymbol{X_1}+u_1/(u_1+u_2)\boldsymbol{X_2}$, with the uncertainty being $(\phi_5(\boldsymbol{X_1}, \boldsymbol{X_2}) + \theta_1 + \theta_2)/2$.

The resulting cone $\mathcal{C}(X^*)$ would be much larger, and even though it might seem strange to use a point not covered by any of the cones $\mathcal{C}(\boldsymbol{X_i})$ which created it, we know we can always reach that point using some combination of $\boldsymbol{X_i}$.

It might actually be convenient to merge two cones even though they do not intersect, but are simply very close to each other. This is because, even though they might not cover all of $\mathbb{R}^m$ they could cover a very large part of it in such a way that it is difficult to interpret any results, since the regressand will most likely be in it.

4.2 pspace and the rotation matrix

4.2.1 3 dimensions

To better understand how the rotation matrix effects a solution, we might first want to get a better understanding of *pspace*.

It might be tempting to define the *pspace* as rotating the space created by the regressor $\mathbf{X}$ an angle θ for each regressor. However, as shown by figure 3 this will not work. One might want to argue that the image is of a flat surface and not on the unit sphere, however as we have already argued before, the sphere is locally indifferent from the plane. For example, the unit sphere could be the size of the earth. Then our *pspace* would be too small.

To get the true value we are looking for we get the following theorem for 3 dimensions:

Theorem 3. *Given a normalized regressor* $\mathbf{X_1} = (x_1, y_1, z_1)$ *with uncertainty cone of radius* u *on the unit sphere, and a normalized regressor* $\mathbf{X_2}$ *with uncertainty cone of size* $\mathbf{0}$ *(not intersecting?), we can find the angle* 2ϕ *within which all uncertainty planes must lie.*

Furthermore, let R be the rotations matrix which rotates $\mathbb{R}^3$ such that $\mathbf{X_2}$ lies on the point (1,0,0), and $\mathbf{X_1}$ on the x-z-plane, and let $R\boldsymbol{X_1} = \boldsymbol{X_1^*} = (x_1^*, y_1^* = 0, z_1^*)$, then we can compute the angle as

$$\phi = \sin^{-1}\left(\sqrt{\frac{u - (1/4z_1^* \pm \sqrt{(1/16)(z_1^*)^2 + (u-\pi/2)^2})}{1 - (x_1^* z_1^* - \sqrt{1-u^2}x_1^*)z_1^* - \sqrt{1-u^2}x_1^*)}}\right) \tag{71}$$

Proof. We can always rotate $\mathbb{R}^3$ such that the line going through $\mathbf{X_2}$ and $-\mathbf{X_2}$ is equal to the x-axis, and $\mathbf{X_1}$ lies on the $x-z$ plane, i.e. $y_1 = 0$. Hence without loss of generality we will assume this for the rest of the proof, to simplify computations. (We choose a basis such that this is true).

Outline of the proof: The idea is to describe 1) the different planes going through $\boldsymbol{X_2}$ and $\overline{\boldsymbol{X_1 + u}}$, and 2) the circle created by the edge of the uncertainty cone $t\overline{\boldsymbol{X + u}}$ and the unit sphere. We do this by describing the (Euclidean) length t between the two points $p_1 = (p_x, p_y, p_z)$ and $p_2 = (p_x, p_y, -p_z)$ and then maximize the angle such that it stay within the bounderies of the circle.

First we compute the angle between the two regressors $\mathbf{X_1}$ and $\mathbf{X_2}$, and get $\gamma = \sin^{-1}(x_1)$.

1) Consider the two planes P_ϕ and $P_{-\phi}$ which are the x-z-plane P rotated by ϕ respectively $-\phi$ around the x-axis. We can make the following observations:

- P, P_ϕ and $P_{-\phi}$ intersect each other on the x-axis, and hence on S^3 at the points $(-1, 0, 0)$ and $(1, 0, 0)$.
- The plane-sphere intersection for P are the points in the set $PS = \{(x, 0, z)|x^2 + z^2 = 1\} = \{(x, 0, \sqrt{1-x^2})\}$, for P_ϕ we have the set $PS_\phi = \{(x, \sqrt{1-x^2})\sin\phi, \sqrt{1-x^2}\cos\phi)|x \in [-1, 1]\}$ and for $P_{-\phi}$, using that $\sin(-\phi) = -\sin\phi$ and $\cos(-\phi) = \cos\phi$, we have the set $PS_{-\phi} = \{(x, -\sqrt{1-x^2}\sin\phi, \sqrt{1-x^2}\cos\phi)|x \in [-1, 1]\}$.
- For any $x \in [-1, 1]$ the distance between the two great circles C_ϕ and $C_{-\phi}$ is the distance between the y coordinates; $|2\sqrt{1-x^2}\sin\phi|$. Note: the greatest possible distance is $|2\sin\phi|$.

This gives:

$$t = 2|\sqrt{1-p_x}\sin\phi| \tag{72}$$

2) Imagine we have a cone which is cut of where it intersects the unit sphere, i.e. when the sides are of length 1. The cone-sphere-intersection is then described by rotating a circle $c = (x_c, \sqrt{u - x_c^2}, \sqrt{1-u^2})$ of radius u around the y-axis at an angle $\gamma - \pi/2$, i.e. the angle between $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$ and compensate since we start witht the origin of the circle at the y-axis

instead of the x-axis. The z_c-value is obtainted by using the pythagorian theorem, $z_c^2 + u^2 = 1$, since the points on the circle lie on the unit sphere.

$$\begin{pmatrix} \cos\gamma & 0 & -\sin\gamma \\ 0 & 1 & 0 \\ \sin\gamma & 0 & \cos\gamma \end{pmatrix} \begin{pmatrix} x_c \\ \sqrt{u - x_c^2} \\ \sqrt{1-u^2} \end{pmatrix} = \begin{pmatrix} x_c\cos\gamma - \sqrt{1-u^2}\sin\gamma \\ \sqrt{u-x_c^2} \\ x_c\sin\gamma + \sqrt{1-u^2}\cos\gamma \end{pmatrix} \quad (73)$$

Not surprisingly the length between the y-coordinates remains unchanged, being

$$t = 2|\sqrt{u - x_c^2}| \quad (74)$$

But from (72) we also have that $t = 2\sqrt{1-p_x}\sin\phi$ which gives the equation

$$\phi = \sin^{-1}\left(\frac{2|\sqrt{u-x^2}|}{2|\sqrt{1-p_x}|}\right) \quad (75)$$

We want to find the point $p = (p_x, p_y, p_z)$ which is tangent to the cone. We look at the point $p_x = x_c\cos\gamma - \sqrt{1-u^2}\sin\gamma$. Plugging this into (75) gives

$$\phi = \sin^{-1}\sqrt{\frac{u - x_c^2}{1 - x_c\cos\gamma - \sqrt{1-u^2}\sin\gamma}} \quad (76)$$

We now have the different angles ϕ of planes going through any point on the circle. However we would like to find the largest such ϕ. To maximise ϕ we need to maximise 76, i.e. find x_c such that

$$\sqrt{\frac{u - x_c^2}{1 - x_c\cos\gamma - \sqrt{1-u^2}\sin\gamma}} = \pi/2 \quad (77)$$

which gives

$$\begin{aligned} x_c^2 - u &= (\frac{\pi}{2})^2(x_c\cos\gamma + \sqrt{1-u^2}\sin\gamma - 1) \\ x_c &= \pi/4\cos\gamma + \sqrt{(\pi/4)^2\cos^2\gamma + (u - \pi/2(\sqrt{1-u^2}\sin\gamma - 1))} \end{aligned} \quad (78)$$

Plugging in $\gamma = \cos^{-1}(z_1) = \sin^{-1}(x_1)$ and 78 in 76 we get

$$\phi = \sin^{-1}\left(\sqrt{\frac{u - (1/4z_1 + \sqrt{(1/16)z_1^2 + (u - \pi/2(\sqrt{1-u^2}x_1 - 1)))^2}}{1 - (x_1z_1 - \sqrt{1-u^2}x_1)z_1 - \sqrt{1-u^2}x_1}}\right) \quad (79)$$

Which is what we wanted. □

This means that the angle ϕ does not only depend on the uncertainty u of regressor $\boldsymbol{X}$, but also on where the regressor $\boldsymbol{X}$ is in relation to the axis we are rotating around. However, since we know only SO(3) has an axis of rotation, we could need to change it for higher dimensions.

5 Results and discussion for further studies

When adding the arbitrary rotation matrix to Nordlings variable selection system, most of the properties can not be obtained. Only those properties which are independent of the regressand $\boldsymbol{Y}$, such as independence or colinearity of regressor $\boldsymbol{X}$, can be definitely obtained.

This is why one would need some constraints, e.g. an optimisation, of the rotation matrix. However, such optimisation could be difficult to find, since the space spanned by the regressor matrix $\boldsymbol{X}$ is difficult to explore.

The tools for exploring the space, and how the rotation matrix would effect it. Different representation methods and their respective metrics, have been suggested. Such as Euler angles, quaternions and Clifford algebra.

For each representation method it is clear that progress on rotation in dimensions higher than 4 is mostely more scarce because most applications on rotations occur in e.g. robotics (i.e. 3 dimensional space) or in space-time in physics applications.

More general work is done on the metrics of Euler angles, with regards to their Lie algebra and Haar measure. Metrics which apply not only to spheres, but to many other surfaces. However, it could still be preferred to use Clifford algebra, since that might require less swithching between representations, especially if it is possible to encorporate the rest of Nordlings variable selection system into Clifford algebra.

Due to lack of knowledge and time, I did not manage to accieve what I set out to do: To find some good optimisation method for the rotation matrix applied to Nordlings variable selection system, and then see whether something can be said about e.g. if some parameter could be selectable or positive etc.

I found that there are many possible paths to follow when trying to find such optimisation. One would want to come up with an optimisation technique for each representation technique, to be able to evaluate which technique that would compute fastest.

5.1 Some of the questions which remained unanswered

Since the topic turned out to be much larger than I had expected, many questions arose over time to which I have not yet been able to find an answer. This has partially been due to the fact that this report has been rather broad, looking in to many potential solutions. But also because the general understanding of rotations in higher dimensions seems quite limited.

I would have liked to have a better understanding of rotation metrics which would have worked in Clifford algebra, more precisely for Cl_n^+. With the constraint of being on the unit sphere i.e. all vectors have a length of 1, and some (generalised) results from Cole [Col90] perhaps this would have been possible. One difficulty might be that metric in Clifford algebra are

(understandably) limited to scalars and vectors.

In [PR97] Park and Ravani raise the question of how to measure computational error with regard to rotations, since most error measures, or results from function approximation theory, are in Hilbert space, while the math around rotations might not be. When trying to implement the rotation into Nordlings system, this could be relevant to look into, regardless of whether one uses Lie algebra or Clifford algebra to solve the problem.

It would be interesting to see how one can compute when one would only regard the projections onto the unit sphere S^{m-1} for m dimensions. Where the projection goes to/from the origin, in other words, all vectors are of length 1. This way one might be able to represent the space spanned by the regressor $\boldsymbol{X}$, the *pspace*, as some area on the unit sphere.

Whether this is actually desirable is of course up to the user, since one might sacrifice other visualization properties. The length of the seperate regressors can still be expressed as simple lines pointing through the unit sphere. However, it might still be much less intuitive if one is used to uncertainty cones in lesser dimensions, or in all other technology one interacts with.

I also did not have the time to explore the stochastic case to any great extent, although that would have been interesting, since the uncertainty of one regressor $\boldsymbol{X_i}$ can vary depending on the uncertainty of the other regressors $\boldsymbol{X_j}$. This means that there will be more variables to take into account when checking when a regressor $\boldsymbol{Y}$ is in the *pspace*. If one wishes to optimise in certain ways, that would mean one would need to make sure one uses the uncertainty of the right regressors.

These could be some interesting topics for further investigation.

6 Acknowledgement

I want to thank my mentors Torbjörn Nordling and Yishao Zhou for their patience, help and guidance, and for introducing me to many interesting subjects.

I would also like to thank my parents who, even though they might not know much about the maths of rotations and spheres, make my world go round.

A Appendix

A.1 Rodrigues' rotation formula

[wol](Rodrigues' rotation formula): Given a vector $a = (a_1, a_2, a_3)$ around which we rotate in an angle θ, the rotation matrix is given as:

$$I + \tilde{a}\sin\theta + \tilde{a}^2(1-\cos\theta) = \begin{bmatrix} \cos\theta + a_1^2(1-\cos\theta) & -a_3\sin\theta + a_1a_2(1-\cos\theta) & a_2\sin\theta + a_1a_3(1-\cos\theta) \\ a_3\sin\theta + a_1a_2(1-\cos\theta) & \cos\theta + a_2^2(1-\cos\theta) & -a_1\sin\theta + a_2a_3(1-\cos\theta) \\ -a_2\sin\theta + a_1a_3(1-\cos\theta) & a_1\sin\theta + a_2a_3(1-\cos\theta) & \cos\theta + a_3^2(1-\cos\theta) \end{bmatrix} \quad (80)$$

Where I denotes the identity matrix and $\tilde{a}$ is the skew symmetric matrix

$$\tilde{a} = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix} \quad (81)$$

Rodrigues' rotation formula is frequently used in context with rotations in 3 dimensions.

A.2 Clifford algebra, multiplication operations

There 3 different multiplication operations for a Clifford algebra which will be explained here, even though they might not be so relevant when looking at rotations, but will still be mentioned.

There are various different *multiplication* rules defined on Clifford algebra. First there is the *geometric product* of say A and B being object with parts of any grade, written AB. This product is associative $(AB)C = A(BC)$, and distributes over addition $(A+B)C = AC + BC$. It can be shown that this multiplication is usually not commutative, however, for scalar s we have that $sC = Cs$ for any clif C consisting of parts of any grade.

The *wedge product* is defined seperately for the scalars and the vectors, and for higher grade one essentially split the blades into the vectors that compose it, and then apply the wedge product on those vectors. (Imagine having two lego houses, just putting one house on top of the other does not make a bigger house, but if one breaks the houses down into individual lego parts, one can use those parts to build a bigger house.)

For vectors P and Q we get the wedge product

$$P \wedge Q := \frac{PQ - QP}{2} \ , \ \langle P\rangle, \langle Q\rangle \leq 1 \quad (82)$$

The wedge product for vectors is antisymmetric, i.e. $P \wedge Q = -Q \wedge P$, which also means that $P \wedge P = 0$.

The definition of a blade wedge blade is now given by

$$P \wedge (Q \wedge R) := P \wedge Q \wedge R \ , \ (P \wedge Q) \wedge R := P \wedge Q \wedge R \quad (83)$$

If the first blade has grade $\langle P\rangle = p$, and the second $\langle Q\rangle = q$ then we get the new blade of grade $\langle P\wedge Q\rangle = p+q$, unless $PQ = 0$ in which case $\langle P\wedge Q\rangle = 0$. We can further extend the wedge product to any clif with the rule:

$$V \wedge (A + B) = V \wedge A + V \wedge B \tag{84}$$

Also for $A' = A + \lambda B$ we get $A' \wedge B = A \wedge B + \lambda B \wedge B = A \wedge B$.

One could think of the wedge product as 'painting' [Den], such that one brushes with one blade along the side of another. For example, brushing with vector A along vector B will create a paralellogram, this way one could also imagine brushing a bivector along a vector. The analogy fails, however, when looking at scalars, though [Den] proposes to look at $s \wedge V$ as $1 \wedge sV$, draging a point along a line of length $|sV|$, which will simply be a copy of sV.

The *dot product* for vectors P and Q is defined as

$$P \cdot Q := \frac{PQ + QP}{2} \ , \ \langle P\rangle, \langle Q\rangle = 1 \tag{85}$$

For vectors the geometric product can also be defined as $PQ = P\cdot Q+P\wedge Q$, though this will not hold for any other grade. The cross-product which exists in normal vector algebra, can be expressed with Clifford algebra, with the help of the wedge product. However, the wedge product does not require a metric, which the cross product does.

A.3 Lie groups and Lie algebras

A Lie group is a way to connect geometry and polynomial equations. As a group it can solve polynomial equations, and algebraic problems. While its smooth manifold property gives it the rigid geometric structure as every element of the group can be identified by a point in space. Because of the smooth differential manifolds, a Lie group will have Euclidian properties on a local scale.

A.3.1 Lie groups

Lie groups are groups that are differentiable manifolds, the group operation has smooth structure i.e. infinite differentiable always exist [Tay]. Gilmore gives the following definition of Lie groups:

Definition: 3. [Gil08] *A* Lie group *consists of a manifold M^n that parametrizes the group operations $(g(x), x \in M^n)$ and a combinatorial operation defined by $g(x)\circ g(y) = g(z)$ where the coordinate $z \in M^n$ depends on the coordinates $x \in M^n$ and $y \in M^n$ through a function $z = \phi(x, y)$.*

There are two topological axioms for a Lie group:

(i) ***Smoothness of the group composition map:*** *The group composition map $z = \phi(x, y)$, defined by $g(x) \circ g(y) = g(z)$, is differentiable up to any order.*

(ii) ***Smoothness of the group inversion map:*** *The group inversion map $y = \psi(x)$, degined by $g(x)^{-1} = g(y)$, is differentiable up to any order.*

Consider the 3-sphere which we know is not very flat. Locally however, in the small enough neighbourhood of a point on that sphere, the space will be indistinguishable from the plane in $\mathbb{R}^2$.

It then follows that it will be a smooth manifold. Essentially since there will be points where it is smooth, but the group structure shows that it looks the sama around any point.

The smoothness of the group composition map, matrix multiplication, is inherited from the smoothness of matrix multiplication. The inversion map is smooth, since $A^{-1} = A^T$.

It is not very surprising that, at any point $p \in S^n$, it is locally differentiable, as a Lie group should be. Since it locally looks like the normal Euclidean space.

As for the group properties, we have

Closure $(AB)^{-1} = B^{-1}A^{-1} = B^T A^T = (AB)^T$.

Associativity Inherited from matrix multiplication.

Identity The matrix identity $I \in \mathbb{R}^{n \times n}$.

Inverse $A^{-1} = A^T$ by definition of matrix representation of $\mathrm{SO}(n)$.

Hence, the matrix representation of $\mathrm{SO}(n)$ is a Lie group. This means that we can create a Lie algebra at the identity element and define a differential function there.

A.3.2 Lie algebra

A Lie algebra has a non-associative multiplication, i.e. $(ab)(cd)$ is not necessarily equal to $(a(bc)d)$ or $a(b(cd))$ etc. It is defined by its **Bracket operation**:

- Bilinear: $[ax + by, z] = a[x, z] + b[y, z]$
- Alternativity: $[x, x] = 0$
- Jacobi identity: $[x, [y, z]] + [z, [x, y]] + [y, [z, x]] = 0$

Using that we can show that

$$0 = [x+y, x+y] = [x, x+y] + [y, x+y] = \\ [x,x] + [x,y] + [y,x] + [y,y] = 0 + [x,y] + [y,x] + 0 \quad (86)$$

which means that anti-commutativity: $[x, y] = -[y, x]$.

Theorem: 16. [Tay] *The Lie algebra* $\mathfrak{so}(n) = \{A | A^T = -A, A \in GL(n, \mathbb{R}), det(A) = 1\}$ *with the bracket* $[A, B] = AB - BA$ *as its Lie bracket is related to SO(n) such that if* $A \in \mathfrak{so}(n)$ *then* $e^A \in SO(n)$.

Here $GL(n, \mathbb{R})$ stands for the linear group, in n dimensions, with real entries.

Proof. First we show that $\mathfrak{so}(n)$ is indeed a Lie algebra, in fact it is the group of skew symmetric matrices. Then we show its relation to SO(n).

It is easy to check that the conditions for the bracket operation hold. Using that $[A, A] = 0$ we will show that the Jacobi identity holds:

$$[x, [y, z]] + [z, [x, y]] + [y, [z, x]] = \\ (x(yz - zy) - (yz - zy)x) + ((zxy - zyx) - (xyz - yxz)) + \\ + ((yzx - yxz) - (zxy - xzy)) = 0 \quad (87)$$

Now suppose $A, B \in \mathfrak{so}(n)$ such that $[A, B] = AB - BA$, then $[A, B]^T = (AB - BA)^T = B^T A^T - A^T B^T = -(AB - BA) = -[A, B]$, hence the Lie algebra $\mathfrak{so}(n)$ is closed under the bracket operation.

Next we wish to show that $exp : \mathfrak{so}(n) \to SO(n)$ holds, e.i. we need to show that if $A \in \mathfrak{so}(n)$ then $e^A \in$SO(n).

$iA^T = -iA$ is hermitian, hence $e^A = e^{i(-iA)}$ is unitary.

$Tr(A) = 0$, hence det(e^A) =det(e^0) = 1

when A is real, so is e^A.

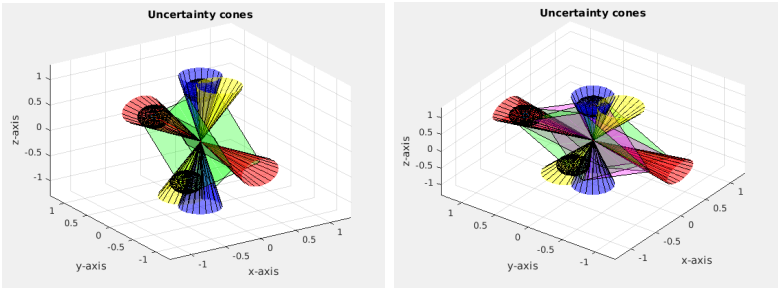
Using the logarithm, we can get the Lie algebra from SO(3). □

The Lie algebra requires we can always find an inverse.

If G is simply connected, then, for any Lie algebra homomorphism σ : $g \to h$, there is a unique Lie group homomorphism $\rho : G \to H$ such that $d\rho = \sigma$.

A.4 Pspace

Here follow some figures of *pspace* and some thoughts of how to improve the visualisation. The first figure (5a) shows part of the *pspace* of example 1 in section 2, and an expansion of figure 2. The two green planes represent part of the outline of the *pspace* of $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$.



(a) Part of pspace (b) All of pspace

Figure 4: Example representation of pspace
In a) the two green planes represent part of the outline of half of the pspace of $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$, and one can see that the yellow uncertainty cone of $\boldsymbol{X_3}$ lies within the pspace. In b) part of all of pspace, however it is clear it soon becomes quite difficult to interpret the pspace visualised like this.

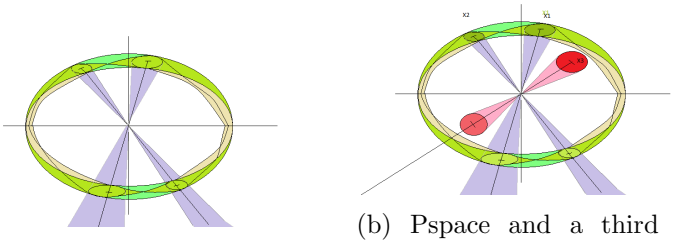
The visualisation of pspace becomes quite messy this way, and it is hard to interpret the picture even though one can move it around and look at it from different angles in matlab.

One suggested other way to visualise this in 3 dimensions would be to visualise the projections on the unit sphere. That way it might also be simpler to visualise stochastic uncertainty. It might also be possible to represent the *pspace* between multiple pairs of regressors and hence make it easier to compare those to each other.

However, when using the new visualization method one might lose vital information that would have been (more) present in the original way to visualize the data.

Another downside of this visualisation might be that it is less intuitive to interpret if one is used to other methods. Having many different visualization methods for different systems can be confusing and it can be easier to appreciate a visualisation method one is already familliar with.

Though in the end it might all come down to which method is the most computationally efficient. Unfortunately I did not have the time to find this out.



(a) Pspace

(b) Pspace and a third (non-colinear) regressor

Figure 5: Another example of representation of pspace
In a) the green/beige area would represent the pspace of the two incertainty cones. Then hopefully it would be easier to see if another regressor is colinear by checking if it is covered by the pspace or not, as in b).

References

[Bae01] John Baez. Clifford algebras. http://math.ucr.edu/home/baez/octonions/node6.html, 2001. Online page.

[Bak16] Martin John Baker. Euclidean space, 1998-2016. http://www.euclideanspace.com/maths/algebra/clifford/theory/extendXProduct/index.htm.

[Bel] Jim Belk. Haar measure on so(n). http://math.stackexchange.com/questions/1364495/haar-measure-on-son. Online answer to question.

[Can96] Di Anselmo Canfora. Parametrizzazione del gruppo ortogonale speciale, 1996. http://ansi.altervista.org/.

[Col90] F. N. Cole. On rotations in space of four dimensions. *American J. Math.*, 1890. https://jemebius.home.xs4all.nl/Papers/Cole-1890/COLE001.htm.

[Den] John Denker. Physics. http://www.av8n.com/physics/clifford-intro.htm. Online litterature.

[Dor01] Leo Dorst. Honing geometry algebra for its use in the computer sciences. *Springer verlag*, 2001.

[EG03] André Elisseeff and Isabelle Guyon. An introduction to feature and variable selection. *Journal of Machine Learning Research*, 3, 2003.

[EJ01] Bayro-Corrochano EJ. Geometric neural computing. *IEEE Trans Neural Netw.*, 2001.

[EN10] Bayro-Corrochano EJ and Arana-Daniel N. Clifford support vector machines for classification, regression, and recurrence. *IEEE Trans Neural Netw.*, 2010.

[Gil08] Robert Gilmore. *Lie Groups, Physics and Geometry.* Cambridge University press, 2008.

[Han] Andrew J. Hanson. Visualizing quaternions, part iv: Clifford algebras. Lecture notes.

[Huy09] Du Q. Huynh. Metrics for 3d rotations: Comparison and analysis. *Springer science+business media*, Online, 2009.

[LaV06] Steven M. LaValle. *Planning Algorithms.* Cambridge University Press, 2006. http://planning.cs.uiuc.edu/.

[LH14] J Li and W Hong. Feature extraction for breast cancer data based on geometric algebra theory and feature selection using differential evolution. *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi*, 2014.

[LMV87] P. Lounesto, R. Mikkola, and V. Vierros. *CLICAL User Manual: Complex Number, Vector Space and Clifforn Algebra Calculator for MS-DOS Personal Computers*. Institute of mathematics, Helsinki University of technology, 1987.

[Lou01] Pertti Lounesto. *Clifford Algebra and Spinors*. Cambridge University Press, 2nd edition, 2001.

[Nor13] Torbjörn Nordling. *Robust inference of gene regulatory network*. PhD thesis, KTH, 2013.

[Not] Lecture Notes. Compact groups. http://www.maths.tcd.ie/pub/coursework/424/GpReps-II.pdf.

[PR97] F. C. Park and Bahram Ravini. Smooth invariant interpolation of rotations. *ACM Transactions on Graphics*, 16(3):277–295, 1997.

[Tay] Michael Taylor. Lectures on lie groups.

[TPG15] Frederico Thomas and Alba Pérez-Gracia. On cayley's factorization of 4d rotations and applications. ., 2015. http://www.iri.upc.edu/people/thomas/papers/AGACSE2015a.pdf.

[Tri09] Christopher Triola. *Special Orthogonal Groups and Rotations*. University of Mary Washington?, 2009.

[vE97] L. van Elfrinkhof. Eene eigenschap van de othogonale substitutie van de vierde orde. *Handelingen van het Nederlandse natuur- en geneeskundig congres*, 1897. https://jemebius.home.xs4all.nl/Elfrinkhof.htm.

[Wil05] Andrew Jowan Wilson. A glance through the gate, clifford's geometrical algebra, 2005.

[wol] Wolfram mathworld. http://mathworld.wolfram.com/. Authors; [Clifford Algebra: Todd Rowland],[Rodrigues' Rotation Formula: Serge Belongie],[Spherical Coordinates: Eric W. Weisstein].

[WW05] Joel L. Weiner and George R. Wilkens. Quaternions and rotations in e4. *The Mathematical Association of America*, 112, 2005.