TALLINN UNIVERSITY OF TECHNOLOGY
SCHOOL OF ENGINEERING
Department of Electrical Power Engineering and Mechatronics

# SINGULAR VALUE DECOMPOSITION BASED HUMAN GAIT ANALYSIS USING IMU SENSORS

## INIMESE KÕNNAKU ANALÜÜS AINU-VÄÄRTUSE DEKOMPOSITSIOONI PÕHJAL IMU ANDURITE ABIL

MASTER THESIS

Student: Dilshan Malith Priyasad

Student Code: 195280 MAHM

Supervisor: Jeffrey Andrew Tuhtan, Associate Professor
Cecilia Monoli, MSc

Co-Supervisor: Mart Tamre, Professor

Tallinn 2021

## AUTHOR'S DECLARATION

Hereby I declare, that I have written this thesis independently.
No academic degree has been applied for based on this material. All works, major viewpoints and data of the other authors used in this thesis have been referenced.

"....." ..... 20…..

Author: .....
*/signature /*

Thesis is in accordance with terms and requirements

"....." ..... 20….

Supervisor: ….....
*/signature/*

Accepted for defence

"....."..... 20… .

Chairman of theses defence commission: .....
*/name and signature/*

**Non-exclusive licence for reproduction and publication of a graduation thesis[1]**

I, Arukattu Patabendige Dilshan Malith Priyasad

1. grant Tallinn University of Technology free license (non-exclusive license) for my thesis "Singular value decomposition based human gait analysis using IMU sensors"

supervised by Prof. Jeffrey Andrew Tuhtan

1.1 to be reproduced for the purposes of preservation and electronic publication of the graduation thesis, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright.

1.2 to be published via the web of Tallinn University of Technology, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright.

2. I am aware that the author also retains the rights specified in clause 1 of the non-exclusive license.

3. I confirm that granting the non-exclusive license does not infringe other persons' intellectual property rights, the rights arising from the Personal Data Protection Act or rights arising from other legislation.

---

_____ (date)

[1] *The non-exclusive licence is not valid during the validity of access restriction indicated in the student's application for restriction on access to the graduation thesis that has been signed by the school's dean, except in case of the university's right to reproduce the thesis for preservation purposes only. If a graduation thesis is based on the joint creative activity of two or more persons and the co-author(s) has/have not granted, by the set deadline, the student defending his/her graduation thesis consent to reproduce and publish the graduation thesis in compliance with clauses 1.1 and 1.2 of the non-exclusive licence, the non-exclusive license shall not be valid for the period.*

**Department of Electrical Power Engineering and Mechatronics**

# THESIS TASK

**Student**: Arukattu Patabendige Dilshan Malith Priyasad, 195280MAHM

Study programme MAHM02/18-Mechatronics

main speciality:

Supervisor(s): Jeffrey Andrew Tuhtan, Associate Professor

Cecilia Monoli, MSc

Consultants: Mart Tamre, Professor

**Thesis topic**:

(in English) Singular value decomposition based human gait analysis using IMU sensors.

(in Estonian) Inimese kõnnaku analüüs ainu-väärtuse dekompositsiooni põhjal IMU andurite abil

**Thesis main objectives**:

1. Apply singular Value Decomposition (SVD) as a data-driven model for gait analysis based on wearable IMU sensors.
2. Evaluate the proposed model for successful approximation of underwater gait behaviour vs on land using a previously collected dataset.
3. Use a Python script to identify combinations of variables which dominate during a gait cycle.

**Thesis tasks and time schedule:**

| No | Task description | Deadline |
|---|---|---|
| 1. | Development of python GUI | 30.02.2021 |
| 2. | Integrating SVD based low rank approximation | 30.03.2021 |
| 3. | Testing and validation of results | 30.04.2021 |

**Language:** English **Deadline for submission of thesis:** 18" May 2021

**Student:** Dilshan Malith Priyasad ..... ".....".....20….a

*/signature/*

**Supervisor:** Jeffrey Andrew Tuhtan ..... ".....".....20….a

*/signature/*

**Consultant:** ..... ".....".....20….a

*/signature/*

**Head of study programme:** Mart Tamre ..... "....."..... 20…..a

*/signature/*

*Terms of thesis closed defence and/or restricted access conditions to be formulated on the reverse side*

# PREFACE

Human gait is becoming a primary field of research in biomechanics due to the demand in various applications. Each area of human gait related research can be classified based on evaluation methods and the means of data collection. Primarily, the gait research is focused on biometric identification and rehabilitation for patients with gait disorders. Inertial measurement unit sensors are used in all the above-mentioned instances as a source of data acquisition. The IMU sensor development has been rapid over the years and currently they provide smart interfaces with intelligent sensors at a relatively low price. Therefore, the popularity of using IMU sensors have been rising.

Data driven methods are especially suitable for analysing dynamic systems which are harder to be explained using governing equations or where governing equations are difficult to derive. Data driven methods utilize the obtained data from a dynamic system to find correlations between them in a way that it symbolises governing equations. Singular value decomposition is a data driven method which can be used to identify dominant patterns of a dataset and provide an accurate low rank approximation to the original dataset. Singular value decomposition is identified as a primary application to dimensionality reduction.

The main objective of this work is to propose singular value decomposition as a suitable data driven method to approximate gait data collected via IMU sensors. Various data sets collected from multiple subjects, under different conditions have been analysed in the study. The low rank approximation is a key component in this research and comparisons between rank-r approximations are discussed in detail.

The SVD algorithm is integrated to a python based graphical user interface to provide a software-based representation and help the used interpreting the results. The python GUI is a smart tool which can be used to optimize low rank approximation and identify dominant patterns in gait cycles. One of the two main outcomes of the thesis was identifying the optimum low rank approximation for terrestrial gaits and underwater gaits which can capture 99% of the energy. Identifying dominant variables of a gait cycle was the other main outcome of the study.

This thesis is written in English language and is 64 pages long including 6 chapters.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF ABBREVIATIONS

GA- Gait Analysis

FT-Fourier Transform

DFT- Discrete Fourier Transform

WT-Wavelet Transform

CWT-Continuous Wavelet Transform

SVD-Singular Value Decomposition

DMD- Dynamic Mode Decomposition

EMD- Empirical Mode Decomposition

IMU-Inertial Measurement Unit

GUI- Graphical User Interface

IMF- Intrinsic Mean Functions

GPR-Gaussian Progress Regression

RMSE- Root Mean Square Error

R -Correlation Coefficient

# 1 INTRODUCTION

Gait is primarily defined as the manner of walking. It is recognized as a distinctive feature because of its dependence on individual weight, limb length, foot wear and posture combined with motion [1]. Human gait can be further described as a locomotion pattern with distinguishable features characterized by kinematics and kinetics [2]. A gait cycle consists of multiple phases which describe the motion at each position [3]. The phases of a gait cycle are illustrated in the Figure 1.1.

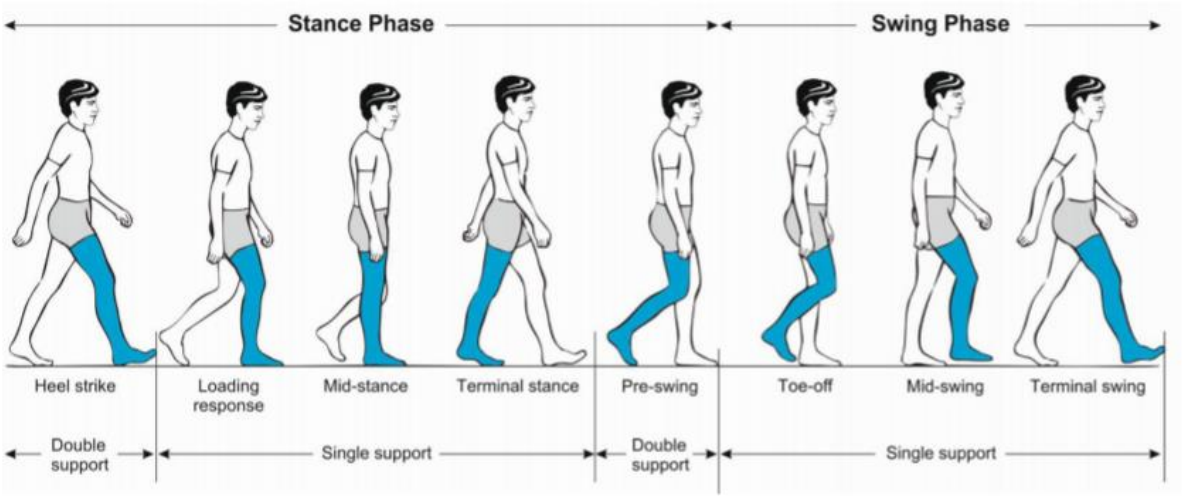


Figure 1.1: Phases of a gait cycle [3].

Due to the uniqueness of human gait, it is considered as an important biometric identifier for everyone. Though, most common human identifiers such as retinal scan, fingerprint etc. require close human contact with the sensing object, human gait can be identified from a considerable distance. Therefore, gait recognition techniques and feature extraction research have attracted growing attention in the recent years [4].

Gait analysis techniques can be classified into two main categories [5].

1) Silhouette based feature recognition.
2) Model based feature extraction.

Silhouette based feature recognition, also known as image-based recognition includes capturing images of a continuous gait cycle and extracting features based on either a temporal based technique or a template-based technique. In temporal based technique, all the silhouettes are analysed based on a chosen feature extraction method. The criteria for feature recognition are the correlation between adjacent silhouettes which is mapped over the complete sequence. Template based feature recognition is based on converting the whole

sequence of silhouettes into a single image which is known as the gait template. Gait Energy Image (GEI) is the averaged silhouette over a sequence of walking gait, and it is the most popular gait template. Various gait templates have been developed over the years namely Gait Flow Image (GFI), Gait Silhouette volume (GSV), Space time interest points (STIP) etc. with different criteria. Silhouette based feature recognition techniques involve multiple covariate factors such as camera viewpoint, surface, clothing, shoes, carrying condition etc. To minimize the effect of covariate factors, multiple classifiers are used in calculations which increases the computational cost associated with the method [6].

Model based feature extraction techniques are based on a solid mathematical model which is characterized by kinematic and kinetic parameters during a sequence of gait cycles. Model based techniques involve various types of sensors to extract a series of static and dynamic variables in a continuous time frame. As mentioned in [7], currently used sensors are accelerometer, gyroscope, magnetometer and pressure sensors. Research [8] illustrates how MEMS based Inertial Measurement Unit (IMU) sensors have been used to formulate a system called Inertial Navigation system (INS) to study the gait of patients with gait disorders. Due to the low cost and availability, IMU sensors have become so popular among researchers involved in GA. Data acquisition systems related with IMU sensors have been using smart devices and applications to filter raw sensor data and sometimes transformations for high accuracy. Reference [7] describes which gait features can be extracted from IMU sensors including Gait speed, Joint angle, Gait regularity, Gait stability, Gait complexity etc. Some temporal features like Cadence, Swing time and Double stance time can also be extracted using event detection or the hidden Markov model [7]. It is also possible to pre-process the sensor data using filtering methods and Normalizing. Research [9] shows types of data sets that can be prepared for gait analysis including standard datasets, multi view datasets, infrared gait datasets, 3d gait datasets etc.

Selection of the mathematical model depends on factors such as the objective of the research, types of datasets and behaviour of the signal. Objective of a model-based GA can vary from human recognition and classification, improving the accuracy of an existing model, proposing a model to an application and creating a new mathematical model. Types of datasets can be described by the dimensions of the input data matrix. Mathematical models can be either univariate or multivariate, and therefore the selection of the mathematical model can vary depending on the number of variables considered in the input matrix. Mostly used

mathematical models and their pros and cons are described in the next chapter. For this research we will consider Singular Value as a suitable mathematical model for gait analysis.

Our approach is to apply Singular Value Decomposition (SVD) as a comprehensive and data driven mathematical model for gait analysis based on IMU sensors. Idea is to identify low rank approximations which are optimum for dimensionality reduction. Identification of dominant variables in a gait cycle is also an objective of the study. We will also develop the SVD algorithm embedded to a python script which will include a Graphical User Interface (GUI) for parameterization.

# 2 LITERATURE REVIEW

## 2.1 Background of GA Research

Gait analysis techniques involve capturing and measuring information related to various gait parameters such as gait speed, step length, knee joint angle, ankle joint angle etc.[8],[10]. Inertial Measurement Unit (IMU) sensors have been an effective tool in Human gait analysis related research in recent years due to their capability in measuring gait parameters. Due to clinical applications and increasing demand, various types of IMU sensors are available in the market [11]. Most recent versions of wearable IMU sensors consist of accelerometers, gyroscopes and magnetic sensors which are available for use at a low cost [12]. Reference [13] utilized IMU sensors to study the underwater gait kinematics and land-based gait kinematics. The study has evaluated the applicability of the IMU sensor system in studying underwater gait kinematics in comparison with an optoelectronic system and a camera-based motion capture system. The GPR calculation steps and comparison results of knee angle estimation is shown in [13].Furthermore, the research shows that proposed IMU system is suitable to be used in land and underwater environment to study gait kinematics. When IMU sensors are utilized in a gait analysis, the focus should be the analysis method and classifiers which must be used. Mostly used GA techniques have been identified as Fourier transform, Wavelet transform, Empirical mode decomposition, Dynamic mode decomposition and Singular value decomposition. In this section, analytical techniques and their applications will be discussed in detail.

The Fourier transform is a mathematical model which is defined for continuous functions and used in digital applications and analysers [14]. Discrete Fourier transform (DFT) is the most widely used version of the Fourier transform, in which the frequency domain of physical signals can be explored. DFT has been used in [15] to detect abnormal gait based on the joint angle characteristics in frequency domain. In which, the joint angle was expressed as a complex number z(n) = ankle angle(n) + j knee angle (n) in a sample size N, where n=0,1,2…...N. The complex angle is then converted to the frequency domain using the Fourier transform.

$$Z(k) = \sum_{n=0}^{N-1} Z(n) e^{-j\,2\pi kn/N} \qquad (2.1)$$

Where Z(k) represents the frequency domain function response and k=0,1,2…. N-1. Then Z(k) is analysed by plotting the magnitude and phase angle of each gait cycle. After the

graphical representation, abnormal gait has been identified by the offset of the frequency domain response, whenever the subject attempts an abnormal gait compared to the normal gait response.

Furthermore, in a similar analysis carried out in the frequency domain, a gait recognition algorithm has been proposed using key Fourier descriptors (KFD) [16]. In which, the gait cycle is captured using a fixed camera and divided into a predetermined number of silhouettes. Then, using a boundary following algorithm, the outer contour of each silhouette has been obtained. Notably, an equal number of contour points were then identified in each silhouette. Each point on the contour was represented by a complex number $S = X_i + j.Y_i$ (i = 0,1, 2....N-1) where N is the number of contour points. If the number of silhouettes is given by T and if $Z_{ki}$ denotes the i th complex point on k th contour, the gait motion can be represented using a N x T length vector.

$$Z = [\ Z_{00}, Z_{01}, Z_{02}\ldots\ldots Z_{0N-1}, Z10, Z11\ldots\ldots\ldots\ldots Z_{(T-1)(N-1)}]^T \quad (2.2)$$

Then Fourier descriptors (FD) have been calculated using the Fourier transform equation, out of which KFDs are selected based on energy distribution vs frequency. In terms of using Fourier transform for gait analysis, it is proved that KFDs are more important for gait identification than other descriptors by obtaining F-statistic in analysis of variance (ANOVA) [14].

Fast Fourier transform (FFT) is an improved version of DFT. It takes advantage over the DFT in the means of calculating DFT coefficients efficiently by carrying out calculations iteratively. In [17], FFT has been proposed to identify Parkinson's Disease by extracting features from gait signals. The research has been carried out using a dataset obtained via an open resource named Physionet, which comprise of measurements of the vertical ground reaction force (VGRF). All the force readings were converted to frequency domain using FFT and features including maxima, minima, standard deviation, log entropy and normalized energy were calculated. Those features were then classified using various classifiers such as Naïve Bayes, Logistic regression, Random Forest etc. to identify PD subjects.

Modern digital filtering techniques and signal processing systems are based on Short Time Fourier Transform (STFT), a technique which can describe any quasi-stationary signal. Compared to DFT, Short time Fourier transform can be effectively used to describe the frequency or spectral content of a signal at each time point. Reference [18] proposed a

method to extract gait data from embedded IMU sensors in a smart phone and analysing the data using a combination of STFT and wavelet transform (WT). Objective of the research has been to identify different pedestrian motions and distinguishing between different activity levels. The signal was first divided into narrow windows using STFT, which assumes any varying signal can be considered stationary for short periods of time. The Same set of data is then transformed to the frequency domain using the Wavelet transform. Unlike the fixed duration sampling window in STFT, the analysis window in WT can be changed to match the frequency change in the signal. Therefore, it has been observed that STFT is more suitable to analyse slow changing steady motions while the WT is more ideal for unpredictable and quick motion changes due to high sampling rate at higher frequencies which improves the resolution of readings significantly. During the research, it is also noted that out of most used Wavelet functions such as Haar, Morlet, Meyer, Mexican Hat, sym8 etc. it is more suitable to use the Morlet function in gait analysis.

Wavelet transform (WT) is recognized as a time-frequency representation for continuous time signals or functions. Wavelet type signals are commonly found in electronics, underwater communication, biomedical engineering and geophysics. Reference [12] describes a method to distinguish patients with Parkinson disease from others using a feature extraction of gait characteristics based on WT. The proposed method includes recording vertical ground reaction force on both feet using sensors which are located under each foot and converting the summation of forces using Continuous wavelet transform (CWT). The resultant approximations and detail coefficients were further analysed using three classifiers including an artificial neural network, a support vector machine (SVM) and Naïve Bayes classifier. From the research, it is evident that the neural network with a resilient backpropagation algorithm has been the most efficient classifier with an accuracy of 85%. It can also be noted that the WT is used as a partial analytical tool in GA related research since it involves various classifiers for the interpretation of end results.

WT has also been used with gait template (image) based recognition. According to [6], a gait feature template can be generated by concatenating multiple sub images. Gait template is a combination of silhouettes converted to a single image which represents a sequence of walking gait such as Chrono-gait image (CGI), Gait Silhouette volume (GSV) etc. These sub images are obtained by applying 2D- Dual-tree complex Wavelet transform (DT-CWT) to silhouettes. DT-CWT has been introduced to overcome some disadvantages associated with Discreet Wavelet Transform (DWT). Unlike the CWT, the modified version can detect the change of a signal in multiple directions. The feature extraction was done by applying the 2D

Principal Component Analysis (PCA) which is an unsupervised feature extracting technique. Finally, to reduce the effect of covariate factors, Random Subspace Method (RSM) classifier was introduced. From WT based gait research it can be identified that high computational cost is a common problem associated with the mathematical model.

Empirical mode decomposition is a recently developed method that allows the decomposition of one-dimensional signals into intrinsic oscillatory modes. When applied to a seismic trace, EMD can be used to study different intrinsic oscillatory modes and instantaneous frequencies of the device. Seismic signals are mostly non stationery systems, making Fourier analysis unsatisfactory due to significant frequency changes over time. Fourier decomposition or the wavelet decomposition use fixed functions as their basis, and they do not compensate for the varying nature of the signals. The basis of Empirical mode decomposition is to adaptively decompose a signal into oscillating components.

Empirical mode decomposition has been used in [19], for de-noising the gait data. In this research, a mathematical concept named Gaussian process is integrated to EMD to optimize the signal. As the improved EMD decomposes a signal into IMFs, first IMF contains the highest frequency values and therefore high noise is expected. After selecting the threshold value, only the IMFs which have higher energy than the threshold was chosen to reconstruct the signal. EMD has also been used for gait feature extraction from Ear-worn sensors[20]. E-AR sensor is placed inside the ear to ensure the human motion is similar to what is sensed by the internal vestibular system. After decomposing the signal using EMD, first IMF is discarded to filter out the noise and low frequency components are also discarded for the Baseline removal. Then the signal is reconstructed using the IMFs which are above the frequency boundary and the peak detection is used to extract essential points in the gait cycle. It has also shown that the method of decomposition via EMD and reconstruction is effective for patients with Parkinson's disease.

Reference [21] describes a gait rhythm fluctuation analysis to identify patients with neurodegenerative diseases. The research includes decomposing of the data set based on EMD and obtaining first five IMFs for the analysis. Then, using the Kendall's co-efficient of concordance, association of ranked data are measured based on two variables named total rank R and the Kendall's W.

Dynamic Mode Decomposition was originally used in fluid mechanics, atmospheric science and nonlinear waves as a data based method and recovery [22]. DMD involves a mathematical

model that is developed in order to analyze, control and simulate complex, nonlinear systems without knowing the governing equations that drives the system. DMD has been used in analyzing data, in which the data have been collected in regular time intervals, where DMD approximated the low dimensional modes of the linear, time independent Koopman operator in order to estimate the potentially nonlinear dynamics [23].DMD can be further modified to Extended Dynamic Mode Decomposition (EDMD) at the expense of computational cost and storage cost. Reference [24] proposed a method called Output Dynamic Mode Decomposition (ODMD) which is an extension DMD based on output functional expanses.

Even though, Dynamic mode decomposition has been mostly associated with fluid dynamics it has also been used in motion related studies. Reference [25] proposed a method to simulate the response of ankle exoskeletons using DMD. The research shows the DMD generated similar or more accurate locomotor response predictions to exoskeletons than the simulated models on 'Opensim' application. A gait recognition differences under carrying or clothing based on windowed DMD has been proposed in [26]. The method includes obtaining silhouettes to create the gait energy image (GEI) and decomposing the GEI to enhance the texture which are not affected by clothing or carry condition variations.

Singular value decomposition (SVD) is a stable numerical method which can be applied to decompose matrices. Unlike FFT, SVD is a pure data driven technique which is heavily utilized to obtain low rank approximations to matrices [27]. More importantly, low dimensional approximation of SVD can be applied to high dimensional data with unknown dominant patterns. As a data driven method, SVD does not use a preselected set of functions and instead it uses a different structure to identify repeating patterns and dominating patterns by decomposing a matrix into a product of three matrices [28]. Apart from the dimensionality reduction, the SVD can also be used to calculate the pseudo inverse of non-square matrices which provides solutions to various underdetermined and overdetermined matrix equations[27]. For a given rectangular matrix X of size m x n, the SVD can be applied as follows.

$$X = U\,\Sigma\,V^T \tag{2.3}$$

Where, U is an orthogonal matrix of size m x m. V is also an orthogonal matrix but in the size of n x n. $\boldsymbol{\Sigma}$ is a diagonal matrix of singular values. Rank (r) is a term closely associated with SVD and in this example, rank of the matrix X: r=rank(X) can be identified as the maximum

number of linearly independent rows or columns in the matrix. Columns of the U matrix are called left eigen vectors or left singular vectors of vector X. The columns of matrix V are called right eigen vectors or right singular vectors of matrix X. Then, $\Sigma$ matrix will only have r number of non-zero values on the diagonal. All the other n-r values will be zeros [29].

In [28] is showed that spatio-temporal parameters associated with gait cycles and trajectory patterns are dominating characteristics in human gait analysis. Due to the quasi-periodic behaviour of these parameters, researchers have used sinusoidal functions and algorithms such as FFT for analysing over the years. But the research shows that to extract dominant patterns from the gait analysis, harmonic algorithms are inappropriate. Furthermore, [21] used signal-to-noise ratio (SNR) which is derived from the SVD result to identify the most dominant pattern named as Average Energy Component (AEC). It shows that for a data matrix with strict periodic values and no noise, the total energy of the matrix can be given as the sum of singular values. Therefore, the degree of periodicity of the relative column can be calculated based on the percentage of the total energy of the square of dominant singular value. Similarly, in a data matrix with noise the percentage of the square of singular value represents the SNR.

Reference [30] proposed a method to extract embodied knowledge or tacit knowledge which is the knowledge remembered by the human body and also reflected by the dexterity, using SVD. The proposed method has been applied to a hand gesture recognition trial and on a walking disability evaluation trial. Notably, the data for the hand gesture trial have been measured using an image processing motion analysis software with a frame-by-frame analysis using cameras. Markers attached to various points on the hand records the position in X, Y and Z vector frame. The resulting data matrix is a thin matrix with three columns. Here, the focus has been the gesture recognition based on the left singular vectors extracted from the time series-based dataset. The research has used two derived criteria based on the left singular vectors to recognize hand gestures.

The data in [30] for the walking disability trial has been gathered using three axis accelerometers. Therefore, the data matrix only included three columns of data with respect to X, Y and Z axes. The research only focused on the acceleration within the period of landing of the right foot and taking off during a gait cycle. It shows that the first singular value received from the SVD is sufficient to identify the walking disability. More precisely, it shows that the first singular value is inversely proportional to the increase of walking disability.

SVD has also been used in a gait recognition study to create a View Transformation Model (VTM) based on the Gait Energy Image (GEI) in [31]. The research shows that the SVD is a successor to the Fourier transform which was used in earlier studies to obtain VTM through a factorization process. Instead, the research uses silhouettes obtained from various cameras at different angles during a gait cycle to create the GEI feature. Then, a Linear Discriminant Analysis (LDA) is applied to the GEI feature to optimize the feature. SVD is then applied on the optimized GEI feature and then the signal is reconstructed using the low rank optimization. The newly created matrix is recognized as the VTM. As explained in the research, the low rank reconstruction improves the accuracy of the VTM as it avoids fitting problems by removing all the less important elements. It also makes the algorithm run speed faster due to the dimension reduction. The specific low rank recreation variant used in the research is Truncated Singular Value Decomposition (TSVD).

## 2.2 Time-Frequency Resolution Comparison

The frequency resolution comparison is shown in the Figure 1.1.

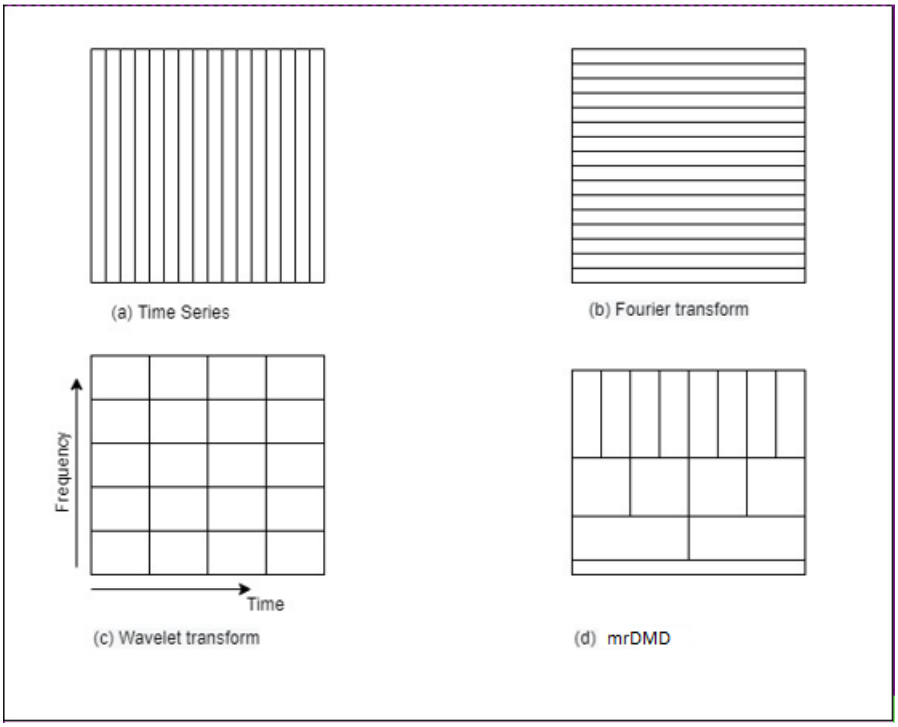


Figure 2.1:Resolution limitations in frequency analysis [25].

Graphical representation of the frequency resolution as described in [32], suggests FT is a fixed resolution time series while the Wavelet transform is a multi-resolution conversion. Wavelet transform representation in the image is the basis of spectrogram which is a powerful tool in the frequency domain. Both EMD and DMD are multi-resolution conversions and can be adjusted as fixed resolution conversions which is an advantage over FT and WT.

## 2.3 Conclusion of the Literature review

- Each mathematical model used for gait analysis have proven shortcomings. Fourier transform is more suitable for smooth and predictable signals. Though, Fourier transform is far from ideal when the signals are composed of spikes or irregularities. Whereas the wavelet transform will be more suitable to analyse a signal with irregularities and spikes. In case of a signal having both smooth (predictable) components and irregular components, application of each method would be limited.

- Empirical mode decomposition overcomes the above situation by decomposing the signal into IMFs which can reconstruct the signal accurately. However, EMD has imperfections such as effect of the end point and mode aliasing problems. Compared to those methods, DMD has been a powerful tool in fluid dynamics which has been used to predict a complex flow based on the space-time-based decompositions. Application of DMD as a data driven method in Gait Analysis using IMU sensors will be proposed in this research.

- None of the mathematical models described above, have been proven successful at predicting the gait behaviour. It is yet to be evaluated whether the proposed method is suitable enough or needs to be modified.

- As mentioned above, an IMU sensor transmits ten variables including gyroscopic readings, accelerometer readings, magnetometer readings and pressure reading. One of the project goals is to identify combinations of variables which can be used to obtain the best low rank approximation in different scenarios and conditions.

## 2.4 Project Objectives

Therefore, the main objectives of the research can be listed as,

- Apply Singular Value Decomposition (SVD) as a data-driven model for gait analysis based on wearable IMU sensors.
- Evaluate the proposed model for successful approximation of underwater gait behaviour vs on land using a previously collected dataset.
- Use a Python script to identify combinations of variables which dominate during a gait cycle.
- Create a python GUI which gives full access to the user to customize data sets with graphical representation.

# 3 MATERIALS AND METHODS

## 3.1 Investigation Procedure

All the trials and experiments related to the research have been carried out with an integrated BOSCH BNO055 waterproof IMU sensor. Its components and the structure are shown in the Figure 3.1.

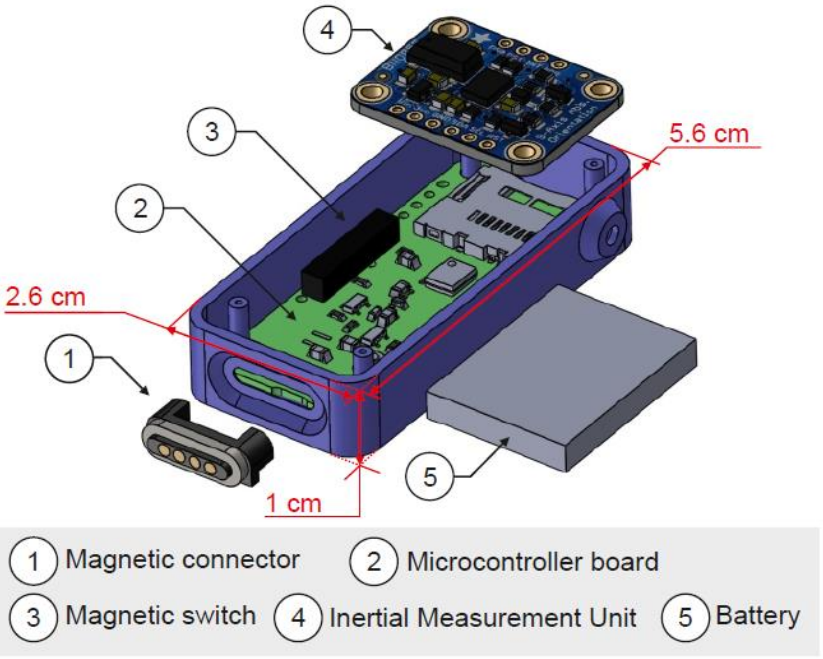


Figure 3.1: Integrated IMU sensor and components.

Technical characteristics of the IMU sensor are represented in the Table 3-1 below.

Table 3-1: Technical specifications of the IMU sensor.

| | |
|---|---|
| **Processor** | Cortex M0 |
| **IMU chip** | BNO055 |
| **Sampling rate** | 100 Hz |
| **Memory** | 16 GB |
| **Battery** | 100 mAh |
| **Data Transmission** | Serial RS-232 |
| **Outer dimensions** | 5.6 x 2.6 x 1.0 cm |
| **Dry Mass** | 22 g |

A comprehensive gait analysis based on SVD, has been performed in this study on a series of data previously collected. Gait of multiple subjects in different trials have been recorded and monitored both on land and underwater conditions, exploiting the characteristics of sensors showed previously. Dry land and underwater trials have been conducted in the Tallinn University of Technology and in a pool facility located in Keila Tervisekeskus respectively. More details on the previously collected dataset is available on [13].

A summary about the attempted trials is shown in Table 3-2.

Table 3-2:Overview of the trials.

| | |
|---|---|
| Trial classification according to the original paper [13] | Trial 2 |
| Subjects involved | 4 (1F - 3M) |
| Environment tested | - dry land<br>- underwater |
| Exercise executed | Gait (barefoot at self-selected speed) |
| Repetitions | 11 |
| Monitoring systems (tools) used | - IMU<br>- Motion capture (Camera) |
| Investigated leg | Right |
| IMU used | 2 (Thigh and Shank) |

During the trials, two identical IMU sensors were attached to each subject for data acquisition via adhesive tape. Both sensors were placed laterally on the right limb of the subject, one on the thigh and the other on the shank. Main purpose of the original investigation was the estimation of the right knee angle range of motion along the gait pattern.

Precisely, sensors were placed at the same vertical height as the centre of mass in each body segment following the outwalk protocol [33]. Reference [34] described the position of the centre of mass (COM) for both thigh and leg as indicated in the Table 3-3.

Table 3-3:Location of the centre of mass for thigh and leg.

| **Segment** | **Definition** | **Position of the COM** | |
|---|---|---|---|
| | | **Proximal** | **Distal** |
| Thigh | Greater trochanter/femoral condyles | 0.433 | 0.567 |
| Leg | Femoral condyles/medial malleolus | 0.433 | 0.567 |

Both sensors were taped to the skin on each segment, which heavily restricted the lateral motion of sensors during any movements made by the subject including the trial. Placement of sensors on two subjects is shown in the Figure 3.2.

Figure 3.2: IMU sensors attachment to the Thigh and Shank of each subject.

The default sampling rate of the sensor (100Hz) was used for the trials. Before each trial, a two-pose calibration was carried out with each subject. Each subject was requested to stand straight, which was considered as the pose 1 and then lift the right leg with knee flexed to a comfortable angle for 2-3 seconds, which was defined as the pose 2. Then, the subject was requested to walk on a straight line for five meters (5m) with self-selected comfortable speed. The only constraint was that the right leg had to complete a minimum of two gait cycles to be considered as a valid trial. Underwater trials were designed similarly.

## 3.2 Data acquisition

The IMU sensor is a combination of sensing elements which measures different physical phenomenon as mentioned in the Chapter 3.3. The embedded accelerometer and the magnetometer readings are comprised of three components associated with X, Y and Z directions, respectively. Therefore, the global coordinate system of the sensor needed to be identified before the data was analysed. The global coordinate system of the IMU sensor is showed in the Figure 3.3.


Figure 3.3: The global coordinate system of the IMU sensor.

Data is recorded on the inside memory of the sensor and then transferred to a PC using RS232-USB protocol. Each dataset could be downloaded as a '.txt' file. Each file consisted of row wise entries for each timepoint in the time frame. Each row consisted of fourteen (14) values of raw data. An example of a raw data set received from a sensor attached to Shank (Leg) is shown in the Figure 3.4.

```
31493   999.81  25.64   0       187.50  46.88   -174.44 -12.94  -10.78  -6.34   -0.03   -0.02   -0.92   -0.40
31713   999.69  25.64   0       228.88  43.19   -88.31  8.86    11.63   -6.80   -0.30   -0.27   -0.65   -0.65
31723   1000.83 25.64   0       228.56  42.25   -91.88  11.64   13.92   -6.12   -0.29   -0.28   -0.66   -0.64
31733   1000.70 25.64   0       227.69  41.75   -96.06  14.17   15.91   -4.05   -0.27   -0.30   -0.67   -0.62
31743   1000.68 25.64   0       226.12  41.75   -101.19 13.13   15.19   -1.79   -0.26   -0.31   -0.69   -0.60
31753   1000.49 25.64   0       224.00  42.06   -107.12 10.78   13.00   -0.52   -0.23   -0.33   -0.71   -0.58
31763   1000.53 25.64   0       221.25  42.38   -113.56 8.10    10.69   0.78    -0.21   -0.33   -0.73   -0.56
31863   1000.46 25.64   0       139.25  33.50   -161.81 12.35   14.06   2.32    -0.03   -0.33   -0.89   -0.32
31873   1000.77 25.64   0       141.38  31.62   -166.00 10.42   12.96   1.39    -0.01   -0.32   -0.90   -0.29
31883   1000.65 25.64   0       143.69  29.75   -169.94 8.51    12.11   1.31    0.01    -0.30   -0.92   -0.27
31893   1000.49 25.64   0       145.94  28.06   -173.44 5.81    10.53   0.50    0.02    -0.28   -0.93   -0.25
31903   1000.47 25.64   0       148.19  26.75   -176.44 4.22    9.66    0.34    0.04    -0.27   -0.94   -0.23
31913   1000.48 25.64   0       150.38  25.88   -178.94 2.99    9.44    1.22    0.05    -0.25   -0.94   -0.22
32011   1000.38 25.64   0       168.44  22.94   168.44  -4.78   7.02    0.62    0.12    -0.08   -0.97   -0.19
32020   1000.58 25.64   0       170.25  23.31   167.88  -5.20   7.25    1.56    0.12    -0.08   -0.97   -0.19
32030   1000.56 25.64   0       172.12  23.94   167.56  -5.10   7.90    2.59    0.12    -0.07   -0.97   -0.20
32041   1000.35 25.64   0       174.00  24.62   167.38  -4.92   8.00    1.70    0.11    -0.03   -0.97   -0.22
32051   1000.43 25.64   0       175.88  25.38   167.38  -4.97   7.63    0.66    0.10    -0.02   -0.97   -0.22
32061   1000.18 25.64   0       177.81  26.00   167.44  -5.03   7.28    0.64    0.10    -0.00   -0.97   -0.23
32161   1000.14 25.64   0       200.12  30.75   170.00  -11.23  -1.34   5.92    0.03    0.17    -0.95   -0.28
32168   1000.42 25.64   0       202.00  31.69   171.62  -11.23  -1.34   5.92    0.01    0.18    -0.94   -0.28
32178   1000.33 25.64   0       203.75  32.50   173.62  -11.49  -1.97   4.68    -0.01   0.20    -0.94   -0.28
32190   1000.19 25.64   0       205.50  33.25   175.75  -10.96  -4.88   0.57    -0.03   0.21    -0.93   -0.29
32200   1000.20 25.64   0       207.06  33.56   178.00  -10.96  -5.68   -0.65   -0.05   0.22    -0.93   -0.28
32210   999.95  25.64   0       208.44  33.31   -179.81 -11.95  -7.09   -0.69   -0.07   0.24    -0.93   -0.28
32310   999.90  25.64   0       207.62  29.31   -154.75 -3.43   -7.04   -11.87  -0.25   0.23    -0.92   -0.20
32318   1000.35 25.64   0       205.81  29.06   -152.06 -3.43   -7.04   -11.87  -0.27   0.21    -0.92   -0.20
32328   1000.25 25.64   0       203.94  28.56   -149.94 -3.14   -6.88   -11.41  -0.28   0.19    -0.92   -0.20
32340   1000.19 25.64   0       202.06  27.81   -148.38 0.80    -3.63   -8.87   -0.29   0.18    -0.92   -0.20
32350   1000.11 25.64   0       200.25  26.75   -147.56 1.30    -3.40   -8.44   -0.29   0.16    -0.92   -0.19
32360   1000.07 25.64   0       198.62  25.25   -147.50 2.08    -2.53   -8.26   -0.29   0.15    -0.93   -0.18
32460   999.96  25.64   0       192.06  -0.75   -164.75 1.22    4.25    -16.49  -0.13   0.10    -0.99   0.02
```

Figure 3.4: Raw data from IMU sensors as a .txt file.

The data was first analysed for integrity and continuity using a graphical approach. Each variable was plotted individually across all the steps for a subject. The basis for the test was the periodic behaviour of the signal. The signal was observed for sudden spikes which were deviated from the periodic pattern. Steps where spikes were recognized were rejected without compromising the whole data set. Behaviour of the Acceleration signal across three gait cycles (3 repetitions) for dry land and Underwater trials are shown in Figure 3.5 and Figure 3.6.

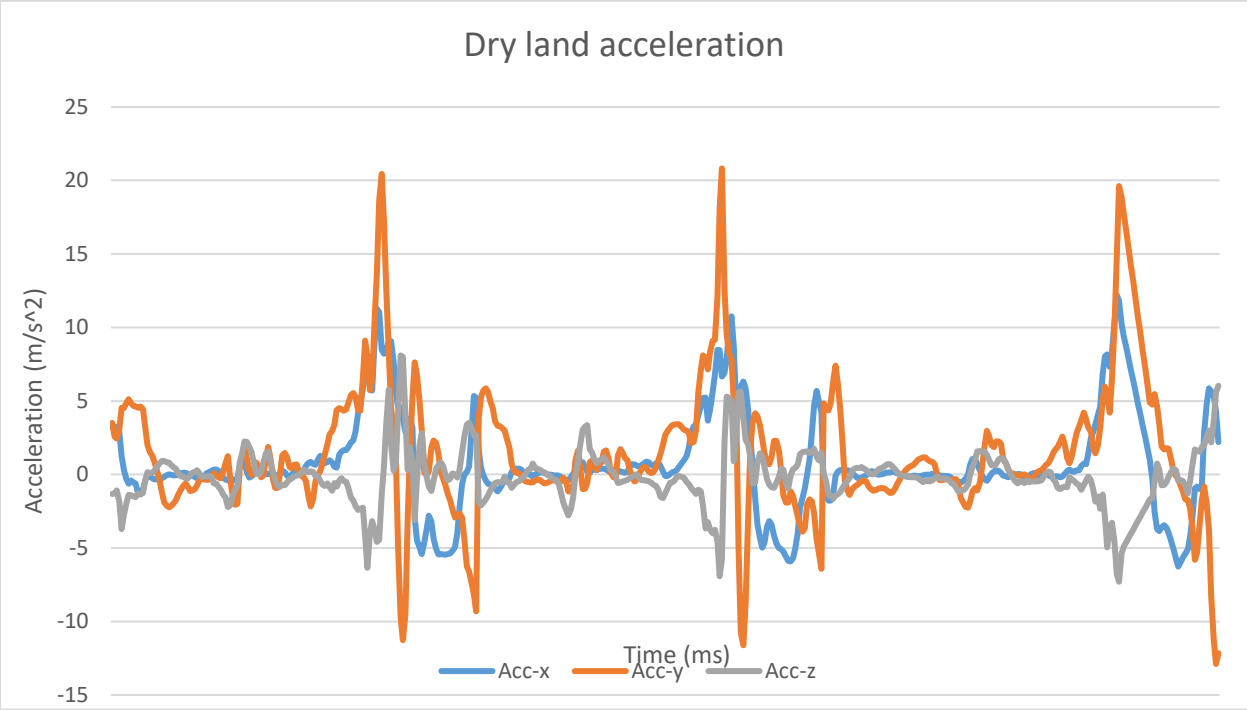

Figure 3.5: Acceleration across 3 gait cycles from shank sensor on dry trials (Subject 1).

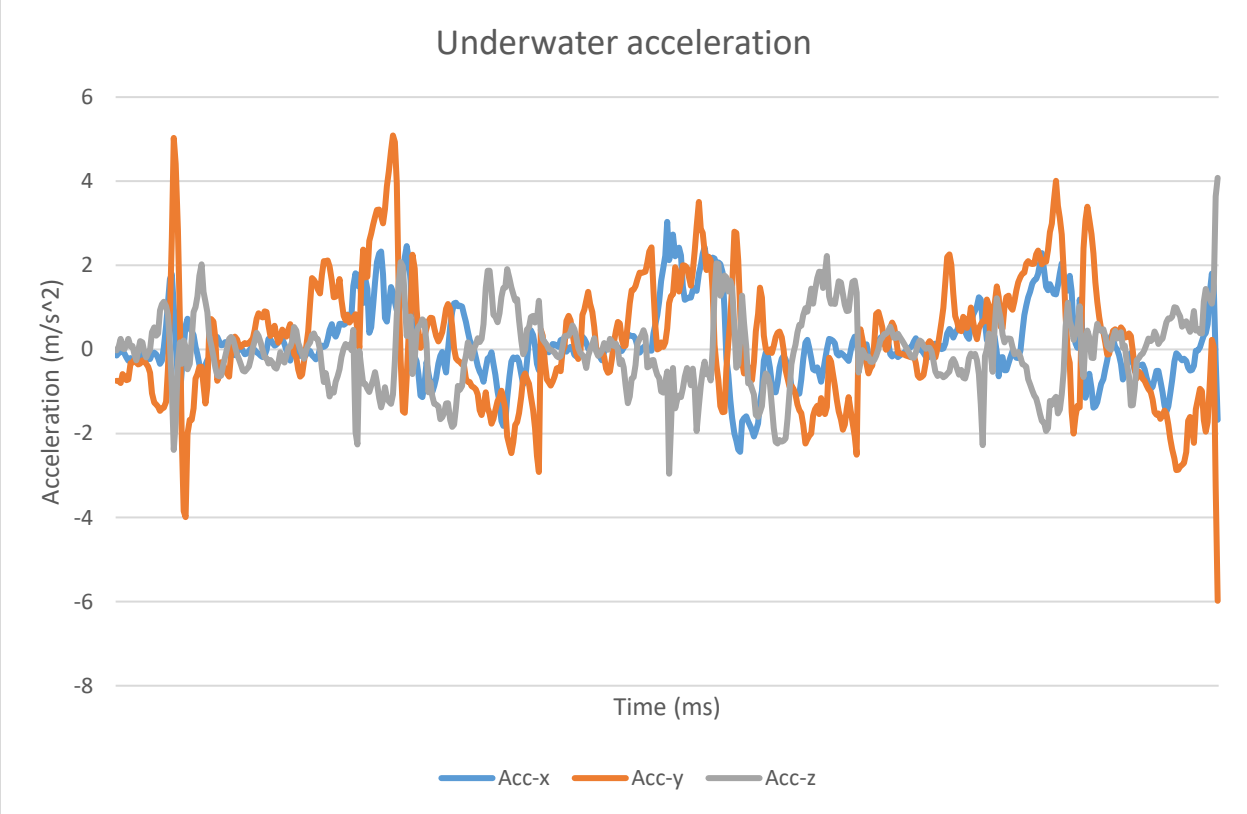

Figure 3.6 Acceleration from shank sensor in Underwater Trials (Subject 1).

## 3.3 Data Pre-Processing

During the pre-processing stage, the knee angle estimation obtained through Gaussian Process Regression (GPR) [13] was added to the IMU data set.

Knee angle is added as a separate data column by converting the text file to an Excel worksheet. Datasets are converted to '.Txt' (Text document) format after the Knee angle addition is completed. The data orientation of a text document is shown in the Table 3-4.

Table 3-4:Data orientation of the input data set.

| **Column number** | **Variable** | **Units** |
|---|---|---|
| 1 | Frame rate | Milliseconds |
| 2 | Pressure | mBar |
| 3 | Temperature | Celsius (0C) |
| 4 | Index | - |
| 5 | Acceleration X | ms-2 |
| 6 | Acceleration Y | ms-2 |
| 7 | Acceleration Z | ms-2 |
| 8 | Magnetic field X | µT |
| 9 | Magnetic field Y | µT |
| 10 | Magnetic field Z | µT |
| 11 | Quaternion | - |
| 12 | Quaternion | - |
| 13 | Quaternion | - |
| 14 | Quaternion | - |
| 15 | Knee angle | Degrees (0) |

# 4 DEVELOPMENT OF A PYTHON GUI

## 4.1 Planning and Structure of the python script

A python-based program was developed with the ability of analysing multiple data sets based on SVD. Python is a high-level object-oriented programming language which is excellent at numeric computing. Therefore, python was the most suitable language for the study. The python based Graphical User Interface (GUI) is a major component of this study. The code is available on https://github.com/dilshmalith/SVD_Utility_tool for reference. The program and the GUI components will be discussed in this chapter.

The program was structured with a combination of the waterfall model and the object-oriented approach model [35]. The waterfall model is the traditional approach for software development which is linear and sequential. The model is also designed in a way that it avoids overlap between phases. Due to the sequential behaviour of user inputs and early identification of objectives in this study, the waterfall model was a near perfect fit for the GUI development. The Object-oriented approach was only utilized in certain scripts to add some flexibility to the code. The phases of the waterfall model are requirement analysis, design, implementation, testing and maintenance [35]. Due to the calculations included in the study, design and implementation phases were merged as a single phase. Since the GUI related to the study was released as a trial version, the maintenance phase was not considered in the study and is open for future studies. Therefore, the stages of the GUI development can be given as,

1) Requirement analysis
2) Design and Implementation
3) Testing of the GUI

## 4.2 Requirement Analysis

The planning of the program was based on following factors,

1) Graphical User Interface and frames
2) User inputs and permissives
3) Data structure
4) Selection of tools and widgets

The GUI was designed as a frame-based application where each frame interacts with the user in a pre-defined sequence. This was to keep the GUI simple and easily interactable. User inputs were embedded to the code via buttons and selection menus. User inputs were properly bounded by permissives ensuring the program does not proceed to the next step if the constraint was not satisfied. Data handling within the program was done using 2D-Arrays and Data frames. The GUI was preliminary designed based on widgets and elements available on the standard Tkinter framework. The structure of the program can be illustrated using a flow chart as indicated in the Figure 4.1.

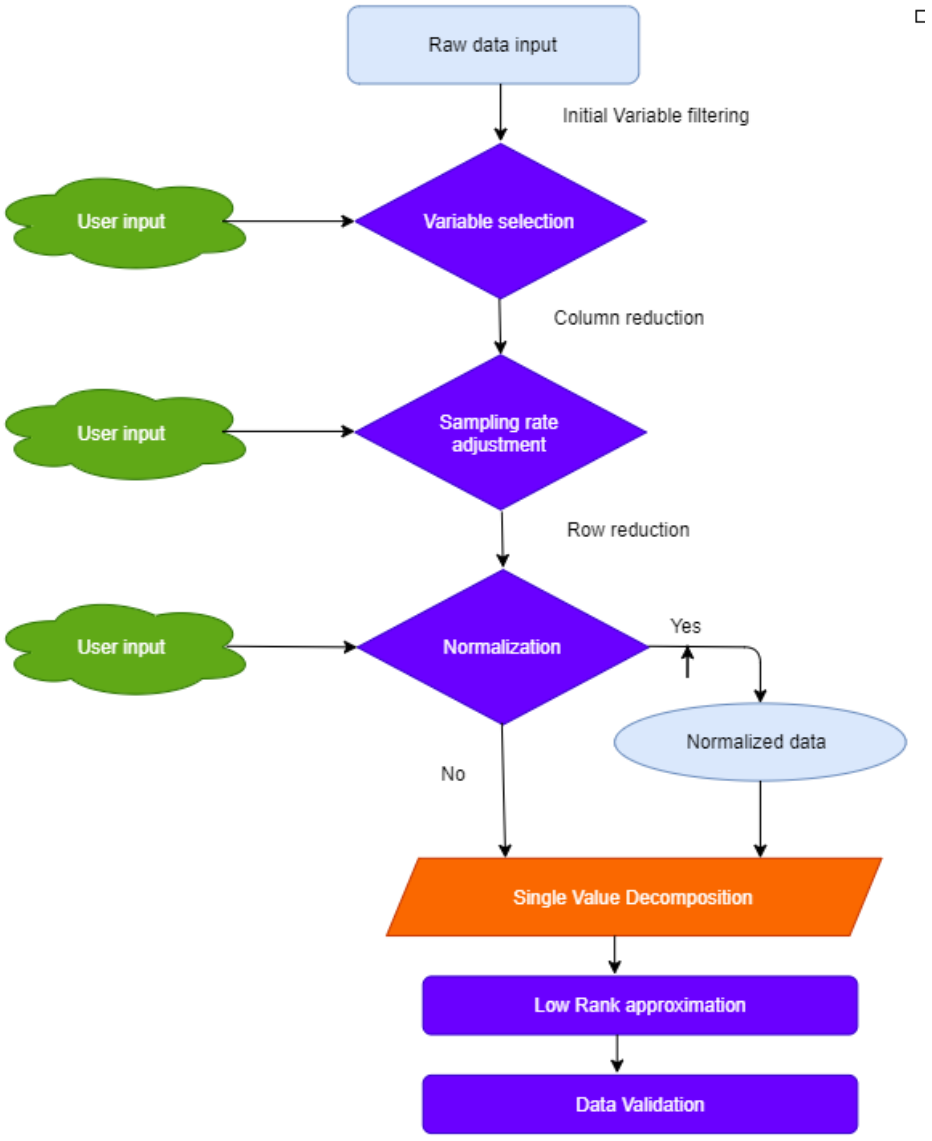


Figure 4.1: Process flow diagram of the python code.

## 4.3 Design and Implementation

The name of the program is 'SVD utility tool' and it can be initialized by launching the program faceplate as shown in the Figure 4.2.


Figure 4.2: SVD Utility tool faceplate.

The program will launch the algorithm once the start button is pressed.

The algorithm launches the file selection window at start. The user can select whether to open a single data file or multiple data files by clicking the appropriate button. When a button is pressed, another popup window will appear where the user can navigate to the location of file(s) and make the selection. The program only allows the user to select from '.txt' files. This is to refrain the user from opening wrong formats which will result in a program error. Appearance of the file selection window is shown in the Figure 4.3.

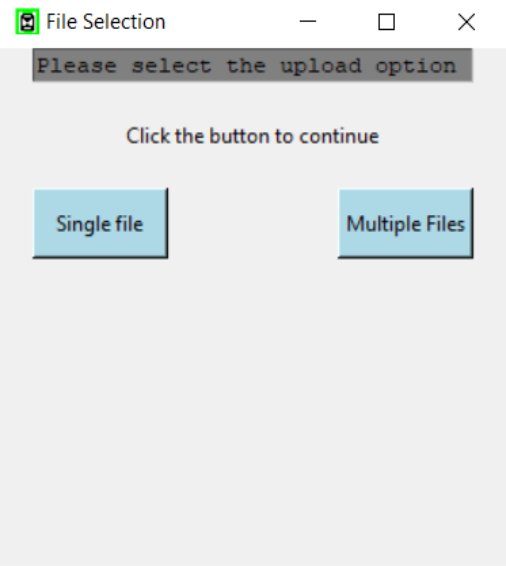


Figure 4.3:File selection window.

When the file selection is completed, the program starts reading raw data from each file, row by row, starting from top to the bottom as a giant list. During the execution of reading multiples files, each file gets stored initially as an element in a 1D array where the number of elements is equal to the number of files read. In both cases, the raw data is only available as a list of strings. To separate values from the string, the string must be split using delimiters. Once the string is separated by delimiters, an input matrix is created as a python 2D array consisting of fifteen columns and multiple rows (tall-thin matrix) where each row represents a different time entry. Data obtained from multiples files are stacked on top of each other to create the 2D array. The stacking order of files has no effect on the outcome as both SVD and DVD are data driven methods. The initial stage of the analysis is focused on the SVD and the interpretation of results. Row data had to be reshaped before the SVD algorithm was introduced in the program. The reshaping of the raw data requires some parameters in the consideration. Therefore, the program consists of several steps with user interactions and sequences with calculations.

The next step of the program is to separate independent variables from the input matrix. Column 1 (Time frame), Column 3 (Temperature) and Column 4 (Index) are independent variables of this study and are used to measure the data integrity of the program. Following the integrity test, those variables are omitted from the input matrix. This is achieved by a column reduction method after converting the input matrix to a data frame. When the rest of the columns are re-arranged, the input matrix is shaped as a tall thin matrix with 12 columns. The reduction of these three columns in the input matrix is a predefined step in the program

with no user interaction. In the next step, the user will be prompted to select from the available variables.

### 4.3.1 Variable selection

Variable selection window provides six options to choose from, as shown in the Figure 4.4. Multiple selection feature is also enabled to include different combinations of variables for the analysis. The selection of option 'All' will make sure that the input matrix will consist of all the columns. Any other selection will reshape the input matrix by applying the column reduction to remove columns of respective variables which were not selected. The layout of the variable selection window is shown in Figure 4.4.

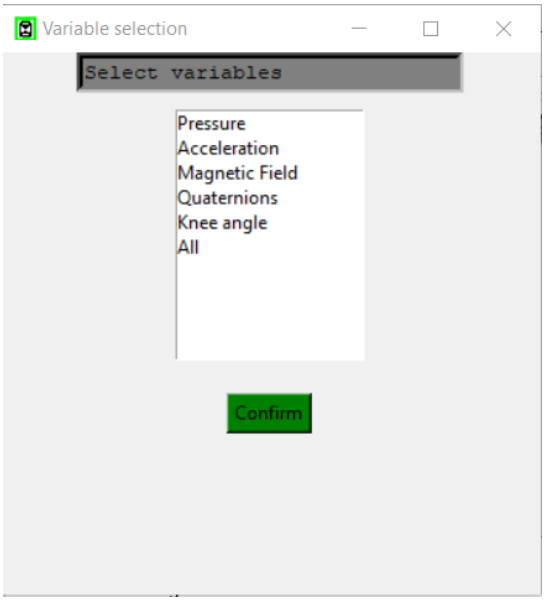


Figure 4.4:Variable selection window.

### 4.3.2 Sub Sampling rate adjustment and data normalization

Next step of the sequence is latching the sub sampling rate. The original raw data from IMU sensors have been sampled at 100 Hz (10ms intervals). The sub sampling rate is a program parameter which is introduced as a user input to adjust the sampling rate of the input data. As shown in the Sample rate window (Figure 4.5), the user is requested to enter a value between 1-10. The entered value will then be multiplied by 10ms to fixate the sub sampling interval. First row of data is always included in the reshaped input matrix as a point of reference in the time domain and from there onwards only the rows corresponding to multiples of sub sampling rate are included in the input data matrix. If the value is set as 1, no rows

will be deleted, and the sub sampling rate will match the original value of 100 Hz. If The value is greater than 1, reshaping is carried out as a row reduction procedure. The layout of the Sample rate window is shown in the Figure 4.5.

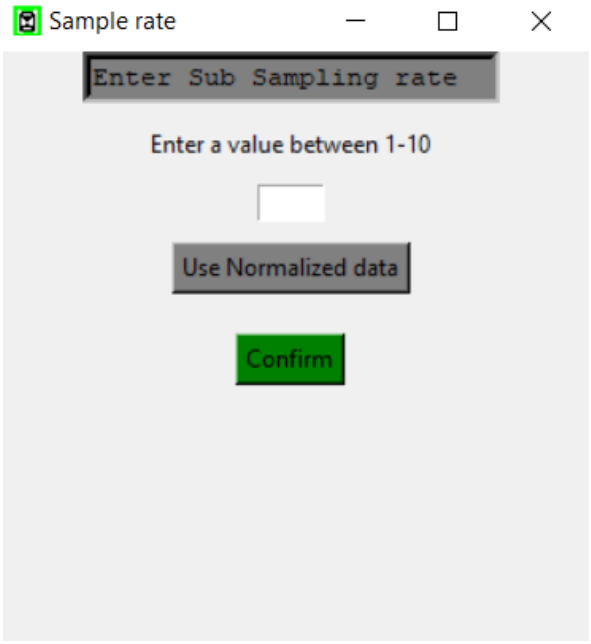


Figure 4.5:Sample rate window.

The sample rate window also includes the option to use normalized data for the analysis instead of raw data. If the option is selected, each column of data will be independently normalized based on the Z-Score equation below [36]

$$Z = \frac{X-\mu}{\sigma} \tag{4.1}$$

Where Z= Normalized value, X= Raw value, $\mu$ = Mean value and $\sigma$ = Standard deviation

The program is designed to create an instance of the normalized dataset irrespective of the selection made by the user. The instance will only be released to the next stage if the user has acknowledged the selection (By confirming the dialog box upon clicking the "Use Normalized data" button). Even though, the acknowledgement of the sub sampling rate is followed by the acknowledgement of normalization, latching the sub sampling rate has high priority in the program and will be executed prior to normalization. The normalization will be carried out when the row reduction process is complete. The program will then select which dataset (Non normalized or Normalized) to be released to the next stage as mentioned above.

The next stage of the program is applying SVD to the dataset. The SVD algorithm decomposes the input matrix into three matrices U, Sigma and $V^T$ as shown in the equation above (Eq 4.1). Instances of the three matrices are moved to another script for low rank approximation.

### 4.3.3 Low rank Approximation

Sample rate window is the next scheduled frame to appear in the program. Top portion of the window is reserved for a user input of the rank (The rank is described in the Chapter 2.1). The permitted range for the input has already been calculated based on the minimum and the maximum rank and being displayed on the window. Bottom half of the window is reserved for a set of basic graphs from the preliminary SVD results. The rank selection window is shown in Figure 4.6.

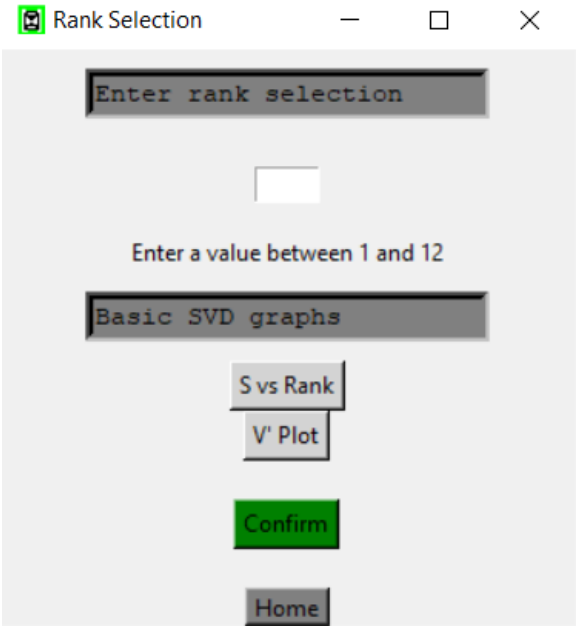


Figure 4.6:Rank selection window.

As shown in Figure 4.6, two graphical references are provided to the user to analyse the behaviour of the data and enter a suitable value for low rank approximation. The program will proceed to the next step if the user has confirmed the input and the value is within the limits. The available graphical references are,

1) S Vs Rank graph
2) V′ modes graph

#### 4.3.3.1 S vs Rank

The graphical representation of Rank Vs S graph is shown in the Figure 4.7.

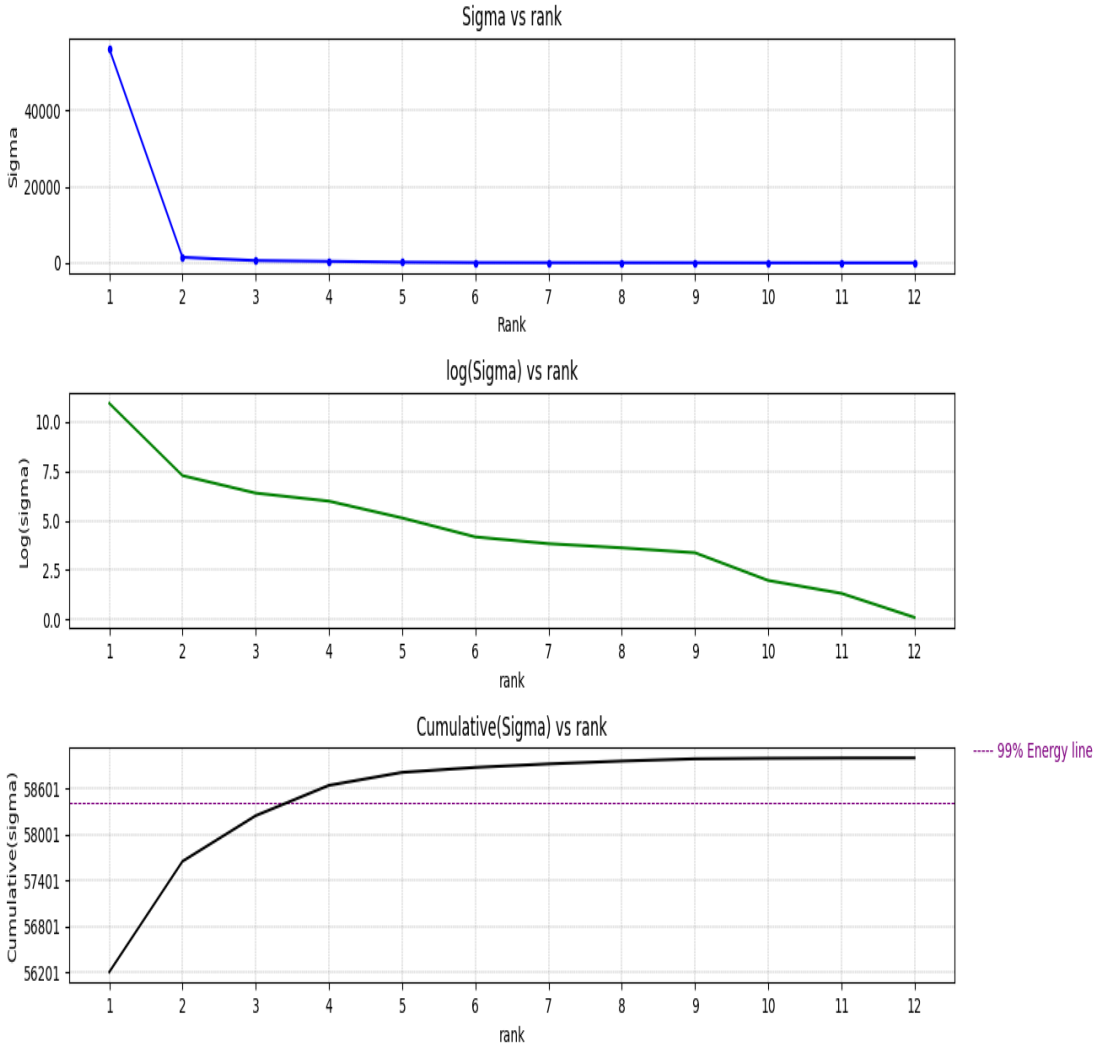


Figure 4.7:S vs Rank graphical representation.

The graph is a combination of three sub plots,

a) Sigma vs rank

The singular value (sigma) defines the strength of each eigenvariable concept corresponding to respective columns in the U matrix. First singular value ($\sigma_1$) is by far the most significant value in the series, and it indicates the domination of the first concept over all the variables. Moving along the curve, sigma values tend to reach zero as the concepts are less influential

and effective on the variables. The corresponding rank of which the sigma values begin to reach zero is a quantitative indication of the number of applicable eigenvariable concepts.

b) Log (Sigma) vs rank

The graph demonstrates the natural logarithm value opposed to the rank as an indication of the distribution of singular values. This graph can be used to fit an exponential curve to the sigma vs rank graph, considering the partial-linear behaviour of the curve. Even though the curve is not entirely linear, a line can be drawn between two suitable points on the curve that will resemble a linear plot. For the example shown above, a fitting exponential curve can be approximated as shown in the Figure 4.8.

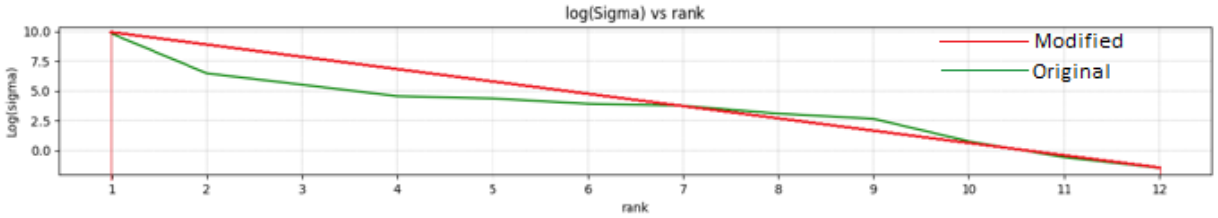


Figure 4.8:Fitting an exponential curve.

$$Y - (10) = [\frac{10-(-0.5)}{1-12}]\, r - 1 \rightarrow Y = (-0.955)\, r + 10.09 \quad (4.2)$$

Substituting for Y= Log (S)

$$\log_e (S) = (-0.955)\, r + 10.09 \rightarrow S = e^{-0.955r + 10.09} \quad (4.3)$$

Where S= sigma, r=rank.

c) Cumulative (Sigma) vs rank

Cumulative sigma value is a perfect representation of the energy distribution of each concept across r-ranks. The Y-axis is often referred to as the cumulative energy of the singular values. The cumulative graph is used to determine the cut-off point for the low rank approximation. The cut-off point has been a controversial discussion over the years. As indicated in [32], the optimal value for the cut-off is 90% of cumulative energy. Though, in this study we are interested in capturing 99% of energy through the low rank approximation. Therefore, a purple-coloured horizontal line is drawn across the grid to indicate the 99% energy limit.

### 4.3.3.2 V’ Modes

V’ modes is an additional feature provided in the program which is useful in finding the dominating correlations between the variables. The layout of the V’ modes graph is shown in the Figure 4.9.

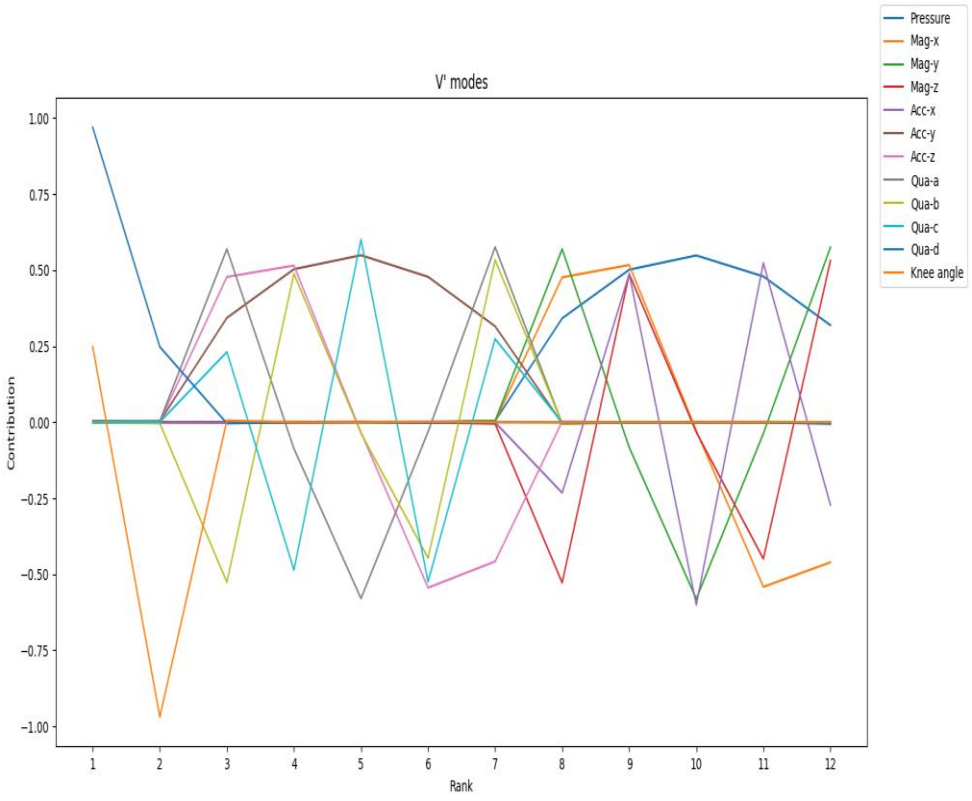


Figure 4.9: Graphical representation of V’ modes.

As described in the (Chapter 4.4), the graph is a graphical representation of correlations between the gaits and eigen gaits. Each row of the V’ matrix is recognized as a mode and all the modes are plotted on the same axes as shown in the Figure 4.9. Specifically, a mode is an indication of the link between a particular variable to each eigenvariable concept.

Each element within the mode is a floating-point number between -1 and 1. The absolute value of an element is an indication of the contribution from each concept to a particular variable. The contribution can also be recognized as a percentage of the concept to understand the mathematical principle behind the algorithm. Let, $V^*_{ij}$ be the element located on the $i^{th}$ row and the $j^{th}$ column in the V’ matrix. Then, the absolute value $| V^*_{ij} |$ determines how influential $i^{th}$ concept to the creation of the $j^{th}$ variable (column) in the parent matrix. However, the weight of this contribution is determined by the $i^{th}$ singular value ($\sigma_i$) of the sequence.

The sign of the element is to determine whether the contribution is positively influential or negatively influential. If the parent matrix is composed of vector fields, sign of the element would mean the influence on the direction. As the non-zero singular values are limited and singular values are arranged hierarchically, only the first r-modes (r=rank) are meaningful in the study.

### 4.3.4 Low rank Approximation graphs

Low rank approximation is carried out as a computational step within the program. When the rank has been entered and confirmed by the user, the program executes the low rank approximation. Rank-r approximation is achieved by keeping the leading r singular values in the matrix S and discarding the rest. The method used for the low rank approximation is derived from the truncated SVD. The truncated notation of the rank-r approximation is shown below.

$$\widetilde{X} = \sum_{k=1}^{r} \sigma_k U_k V_K^* = \sigma_1 U_1 V_1^* + \sigma_2 U_2 V_2^* + \sigma_3 U_3 V_3^* + \ldots\ldots + \sigma_r U_r V_r^* \quad (4.2)$$

Where $\tilde{X}$ is the resulting truncated matrix ($\tilde{X} \approx X$).

The next window to appear in the program is the 'Rank approximation graphs' frame. The window is a tool to compare the original matrix against the approximated matrix for each variable. The window and its composition are shown in the Figure 4.10. The purpose of providing the graphical representation is to evaluate the outcome of rank-r approximation.

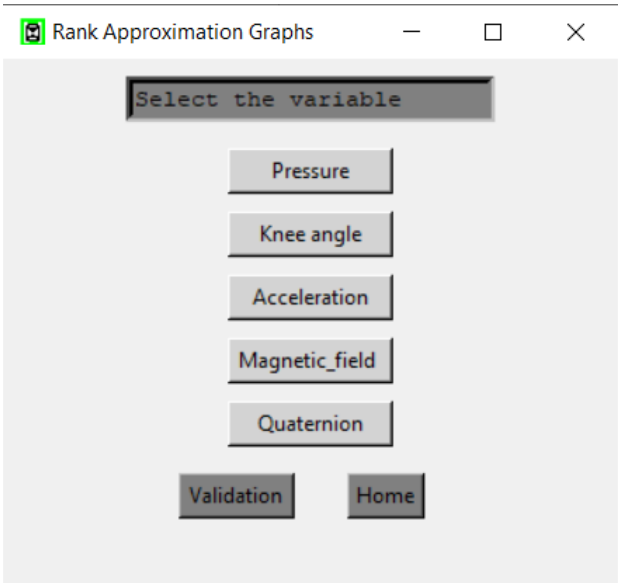


Figure 4.10: Rank Approximation Graphs window.

The evaluation of the results is entirely left to the user and the program assumes that the user is familiar with the algorithm and calculations. The program is designed to accept user commands, execute the logic, and provide the feedback without the interpretation so that the user can interpret it in an appropriate manner. An example of the obtained graphs for a rank-3 approximation of magnetometer readings are shown in Figure 4.11.

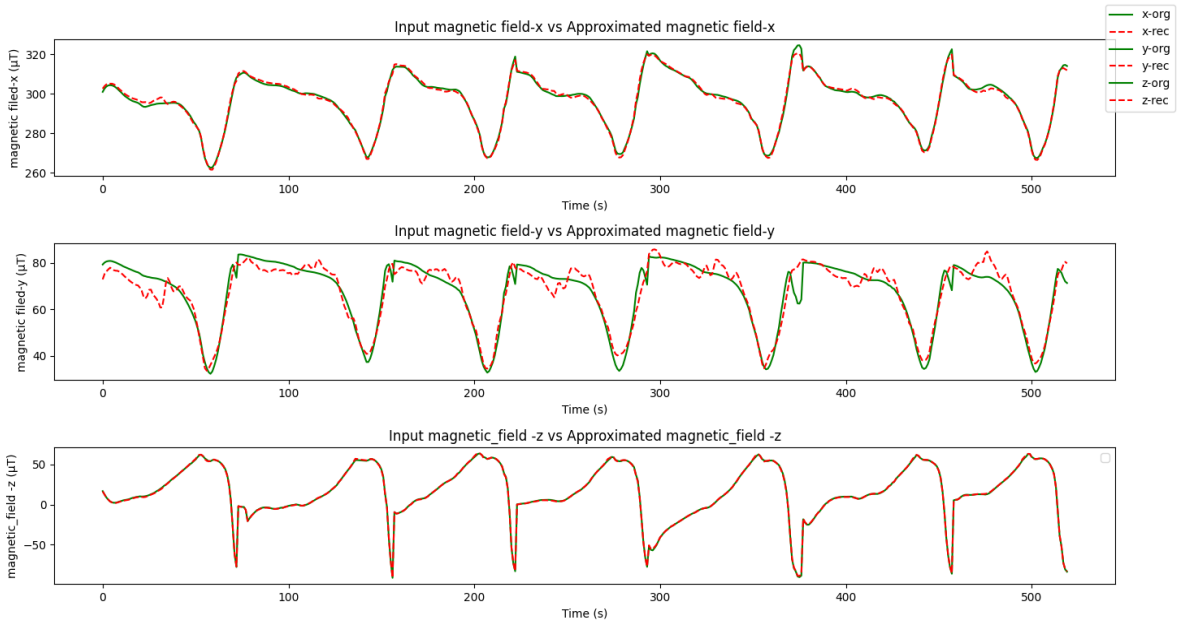


Figure 4.11: Rank-3 approximation of magnetic-field measurement.

## 4.3.5 Data validation and accuracy calculation

At the bottom of the Rank approximation graphs layout (refer the Figure 4.10) is a grey button labelled as validation. Upon clicking the validation button, the program will open a sub-window themed 'Data validation'. The layout is organized in a manner that the user can execute the Root Mean Square Error (RMSE) and Correlation Coefficient (R) calculations with button clicks. Each calculation can be graphically presented with a button click as shown in the Figure 4.12. Graphs can be individually saved by pressing the save button after browsing to the file destination. The RMSE is calculated for each column in the approximated rank-r matrix compared to the input matrix (described in the section 4.3.2). RMSE is a better indicator to compare how rank-r approximations minimize the error as discussed in the chapter 5.3. The equation used for the RMSE calculation is given below.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \tilde{x}_i)^2}{N}} \tag{4.3}$$

Where $x_i$=original value, $\tilde{x}_i$= approximated value, N= number of data points and i=$i^{th}$ iteration.

The R value is an indication of the linearity between the approximated time series vs original time series. R close to 1 suggests a strong linear relationship while the value close to zero indicates a non-linear relationship or in this case a failed attempt at approximation. The equation used for the calculation is given below.

$$R = \frac{\sum(x_i - \bar{x_i})(y_i - \bar{y_i})}{N\sigma_x\sigma_y} \quad (4.4)$$

Where R=correlation coefficient,

$x_i$ = value in the original series,

$\bar{x_i}$ = Mean of the original series,

$y_i$=approximated value,

$\bar{y_i}$ = Mean of the approximated series,

N= Number of data points

$\sigma_x$ = Standard deviation of the original series,

$\sigma_y$ = Standard deviation of the approximated series,

Data validation window is shown in the Figure 4.12.

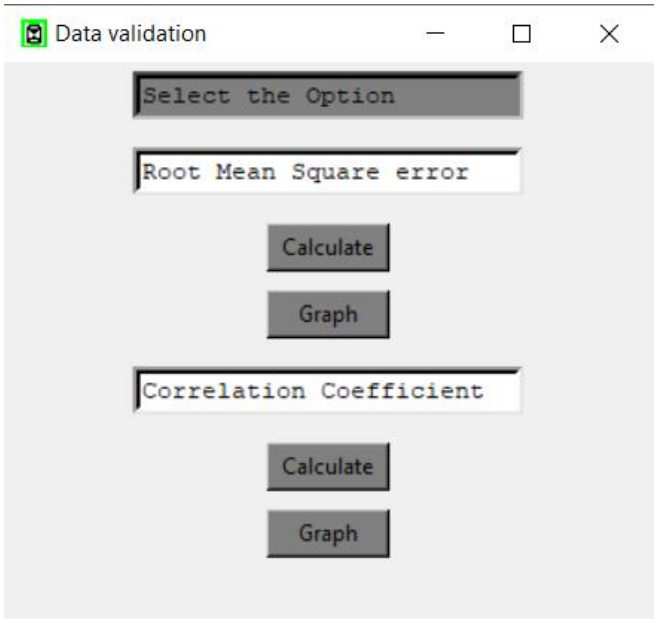


Figure 4.12:: Data validation window.

## 4.4 Interpretation of SVD results

By analysing the Singular value decomposition of a parent matrix X, matrix U can be identified as a series of column vectors arranged in a hierarchical order where, each column is an eigenvector of the parent matrix X (equation 2.3). Hierarchical order of the columns depends on the correlation between eigen vectors and the columns of the parent matrix X [32]. For example, a well-known facial recognition algorithm is based on SVD, and in [32] the authors applied SVD on a large library of facial images. From the decomposed matrices, the columns of U matrix were reshaped and consequently, they identified the images to be "eigenfaces". Therefore, it is theoretically possible to define each human face as a linear combination of eigenfaces. Furthermore, weights of the linear combinations determine the identity of each person. The weights are described by Sigma and V matrices.

Based on the interpretations discussed above, the input matrix of this study and its decomposition can be interpreted as follows. Let X be the notation for the input matrix, and $X_i$ ($X_i \in X$) be any column in the matrix X. Then $X_i$ represents the behaviour of a gait cycle or multiple gait cycles from the perspective of a variable across pre-defined phases. Number of phases can be adjusted by defining the sub-sampling rate (Chapter 4.3.2) in the program.

Let U be the notation for the left singular vectors, and $U_i$ ($U_i \in U$) be any column in the matrix U. Then $U_i$ represents a concept of an eigen gait from the perspective of each variable in the hierarchical order. An eigen gait is an intermediate state between an actual movement and no movement. Therefore, it is better described as a concept rather than a physical element. Let $\sigma_i$ be any element in the matrix $\sigma$. Then, $\sigma_i$ corresponds to the importance of concept $U_i$ where $\sigma_1$ has the highest weight.

V' matrix is an indication of correlations between eigen gaits and gaits. Moving along the first row of V' matrix will reveal how much influence the first eigen gait has on each gait, respectively. If the element is a negative floating-point number, it indicates a negative influence towards the corresponding gait. The graphical representation of the interpretation is shown in the Figure 4.13.

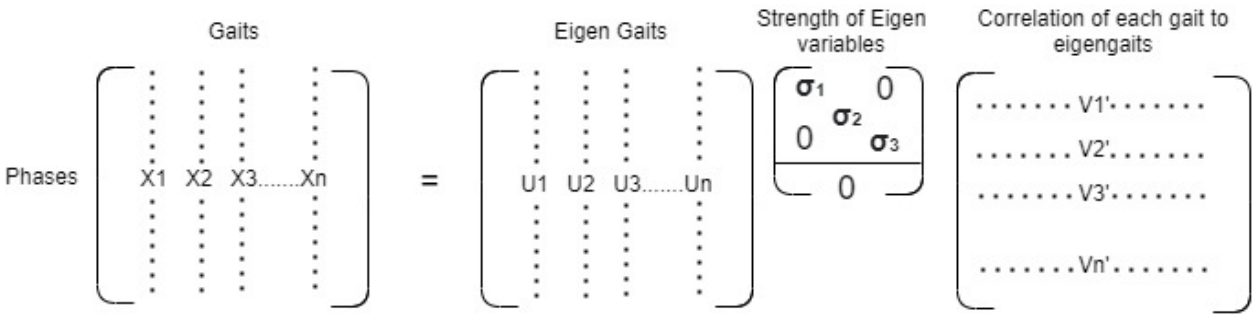


Figure 4.13: Graphical representation of SVD.

# 5 RESULTS

This Chapter is comprised of the results from the analysis conducted on the available dataset in a statistical way. The available dataset and the data acquisition procedure is described in the chapter 3.1 (Investigation procedure) and 3.2 (Data acquisition) respectively. The chapter includes a detailed analysis between two subjects in land conditions followed by an analysis on same subjects in underwater conditions. Accuracy of the approximated data will be discussed later in the chapter. The accuracy and validation analysis includes comparisons between different variables in low rank approximation.

Though, all the variables were separately analysed to study the dynamics of the approximated signal, knee angle approximation has been graphically represented throughout the chapter considering the compactness of the image compared to other variables. Comparison of the knee angle approximation is a better way to illustrate the behaviour of the approximated signal in gradual responses, step responses, peaks, and valleys across trial repetitions. Improved signal pattern over increasing r in rank-r approximations is also a basis for deciding an optimum low rank approximation.

Datasets of Subjects 1 and 2 were chosen for the in-depth analysis in this chapter. Initially, each dataset was given an identifier for analytical purposes. Dataset identifiers are shown in the Table 5-1.

Table 5-1: Dataset identifiers.

| **Subject** | **Trial Condition** | **Sensor Position** | **Identifier** |
|---|---|---|---|
| Subject 1 | dry land | Shank | A |
| | dry land | Thigh | B |
| | Underwater | Shank | C |
| | Underwater | Thigh | D |
| Subject 2 | dry land | Shank | E |
| | dry land | Thigh | F |
| | Underwater | Shank | G |
| | Underwater | Thigh | H |

## 5.1 SVD of two subjects in dry land trials

### 5.1.1 Raw data analysis for dry land trials

This section illustrates the low rank approximation of the data obtained from the shank sensor on two subjects (Dataset A & E). Dataset A was loaded to the SVD analysis tool with the combination of settings shown in the Table 5-2.

Table 5-2: Configuration of dataset A.

| **Dataset** | **Variables selected** | **Sub-Sampling rate** | **Normalized** |
|---|---|---|---|
| A | All | 1 | No |

The obtained S Vs Rank graph is shown in the Figure 5.1.

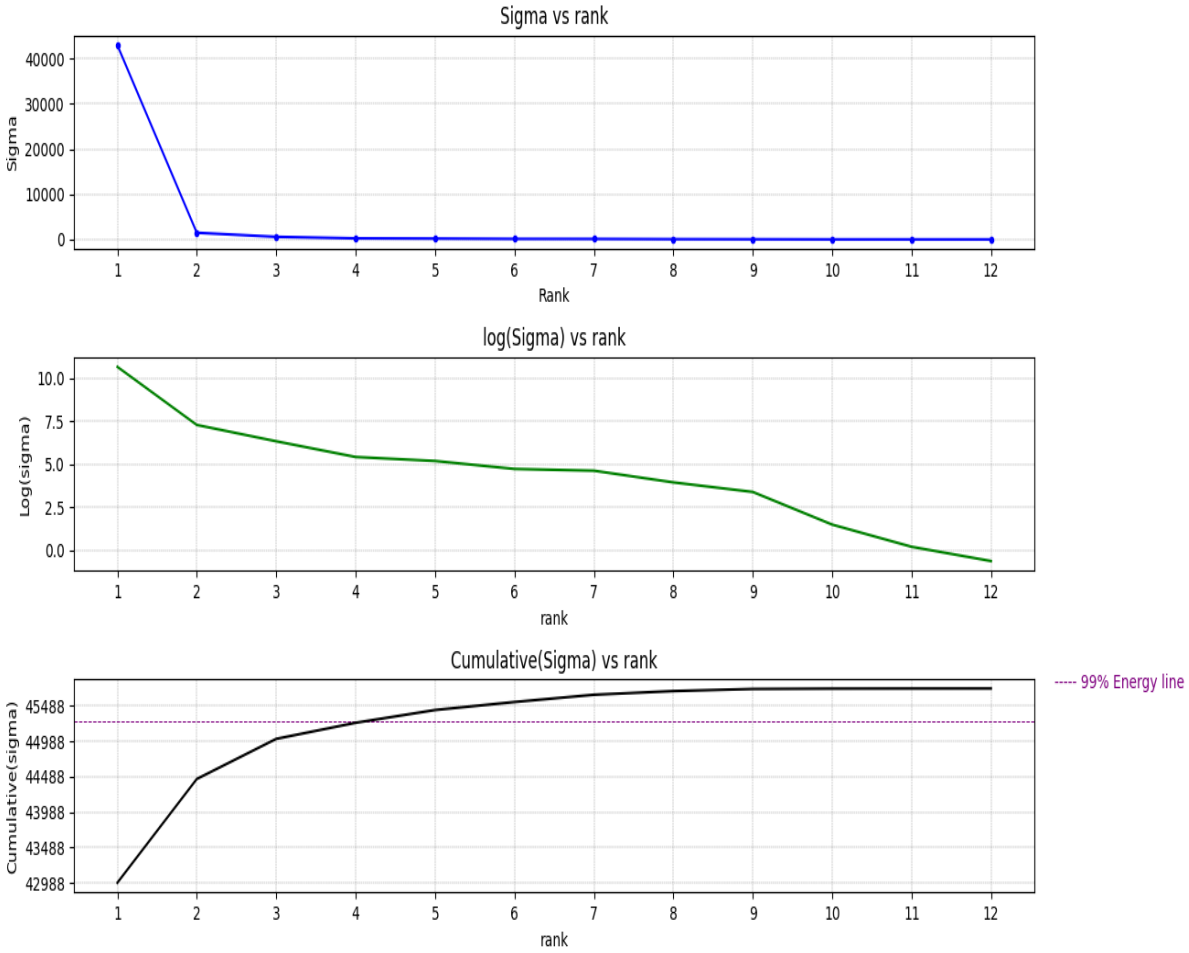


Figure 5.1: S Vs Rank graphical representation for dataset A.

From the above graph, rank-4 approximation is most likely to capture 99% of the energy. Therefore rank-1, rank-2, rank-3, and rank-4 approximations were obtained to study the difference of approximated datasets in each phase. A variable wise comparison was carried out on each approximation separately using the graphs obtained. Results for the Knee angle approximation is shown in the Figure 5.2.

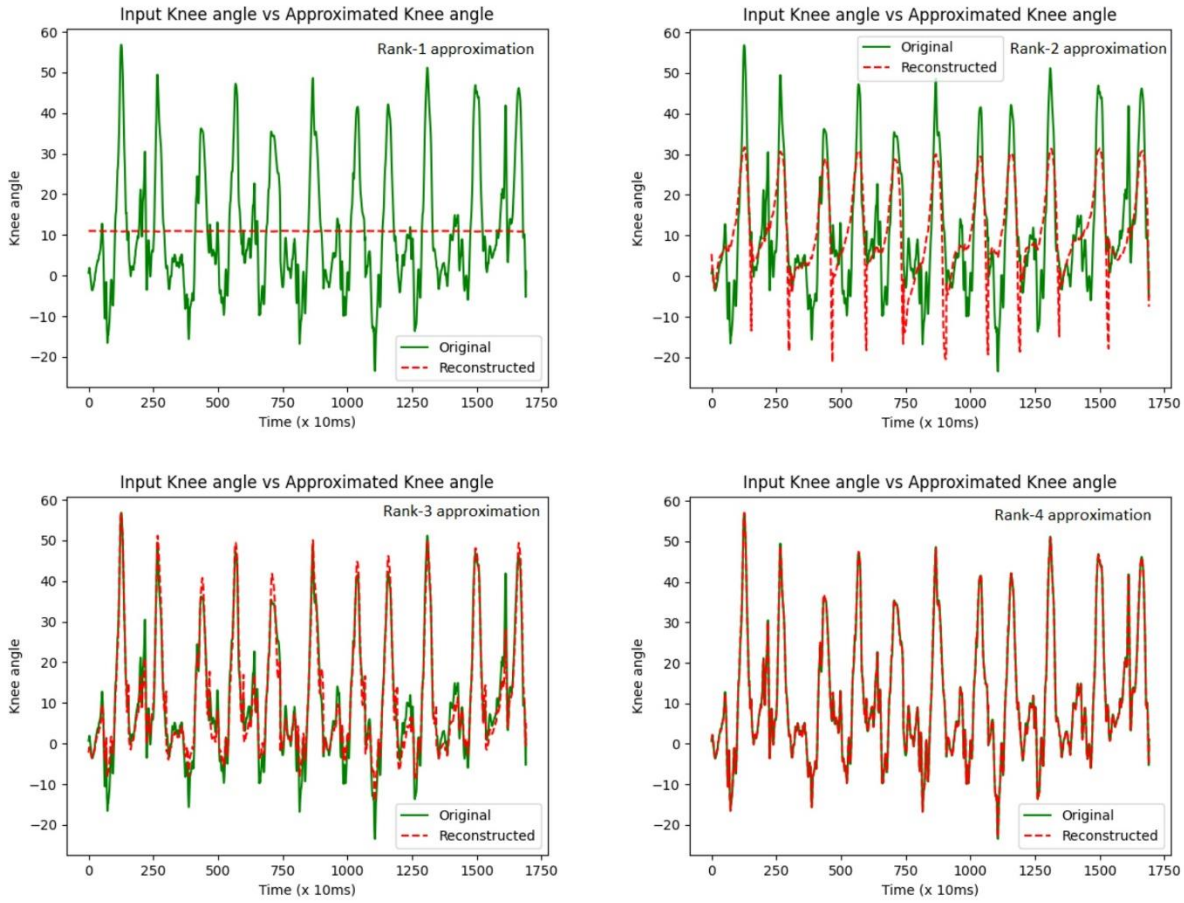


Figure 5.2: Knee Angle approximation for dataset A.

Then, Dataset E was loaded to the SVD analysis tool with the following upload configuration (Table 5-3) which is identical to the selections made for dataset A.

Table 5-3:Configuration of dataset E.

| **Dataset** | **Variables selected** | **Sub-Sampling rate** | **Normalized** |
|---|---|---|---|
| E | All | 1 | No |

The obtained S vs Rank graph for dataset E is shown in the Figure 5.3.

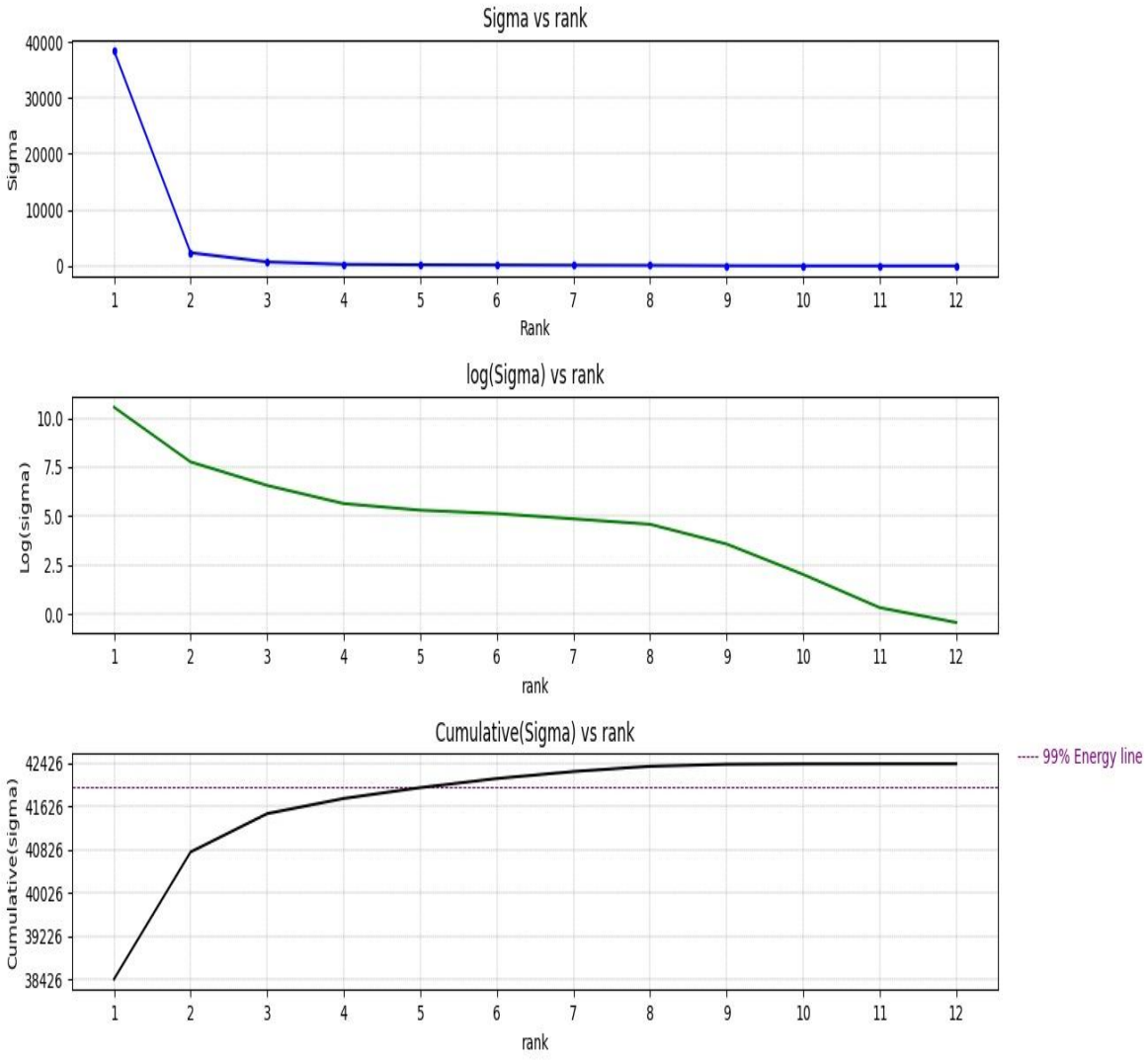


Figure 5.3: S Vs Rank graphical representation for dataset E.

Unlike dataset A, the 99% energy can only be captured with a rank-5 approximation in the dataset E. To study the behaviour on each phase, rank-1, rank-2, rank-3, and rank-5 approximations were carried out. Rank-4 was not considered since no significant energy difference between rank-4 and rank-5 was identified. Results of the Knee angle approximation from four rank-r instances are shown in the Figure 5.4.

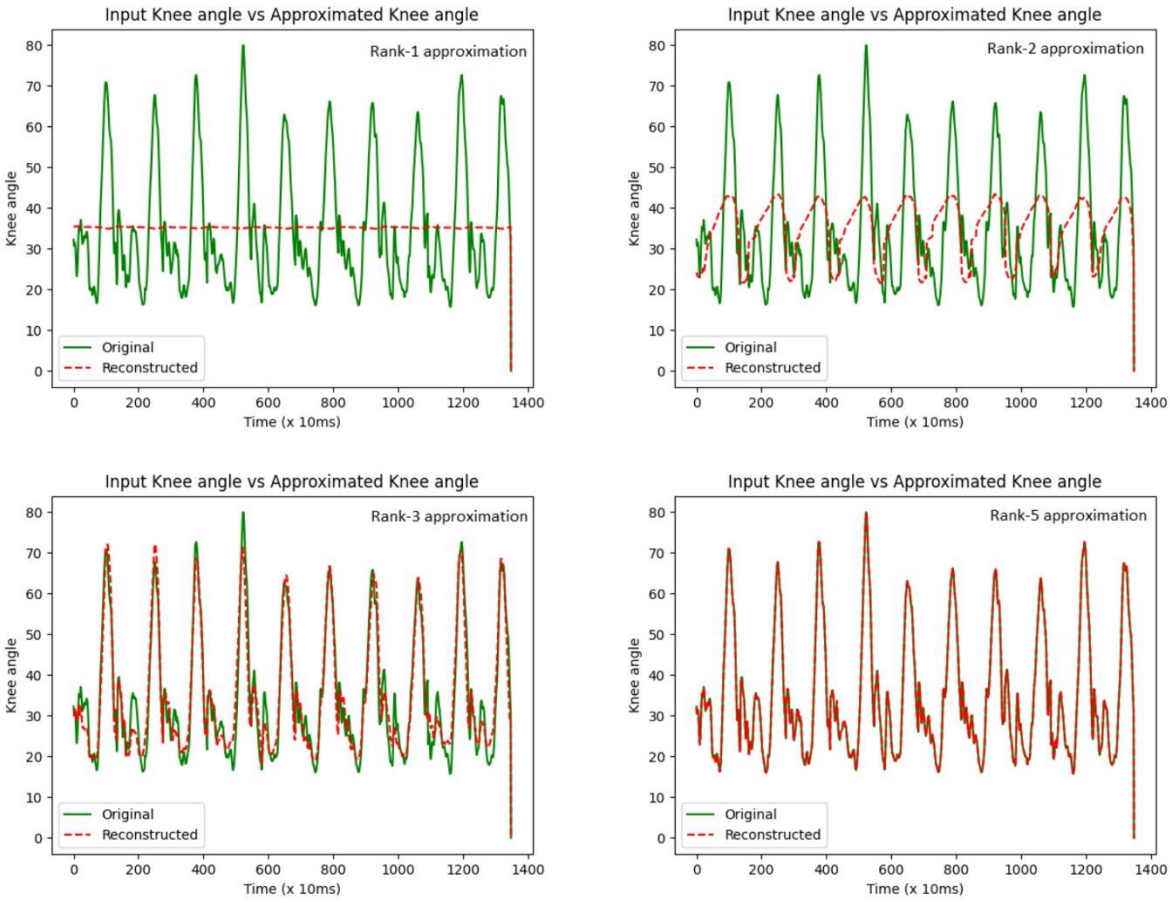


Figure 5.4: Knee angle approximation for dataset E.

## 5.1.2 Normalized data analysis for dry land trials

This section is a narrative for the use of normalized IMU sensor data in the SVD analysis. Datasets A and E were the basis for the normalized SVD analysis in dry land trials. Data normalization procedure is described in sub chapter 4.3.2 (Sub sampling rate adjustment and data normalization).

Dataset A was loaded on the analysis tool with the upload configuration shown in the Table 5-4.

Table 5-4: Configuration of dataset A (Normalized).

| **Dataset** | **Variables selected** | **Sub-Sampling rate** | **Normalized** |
|---|---|---|---|
| A | All | 1 | Yes |

The resulting S vs Rank graph is shown in the Figure 5.5.

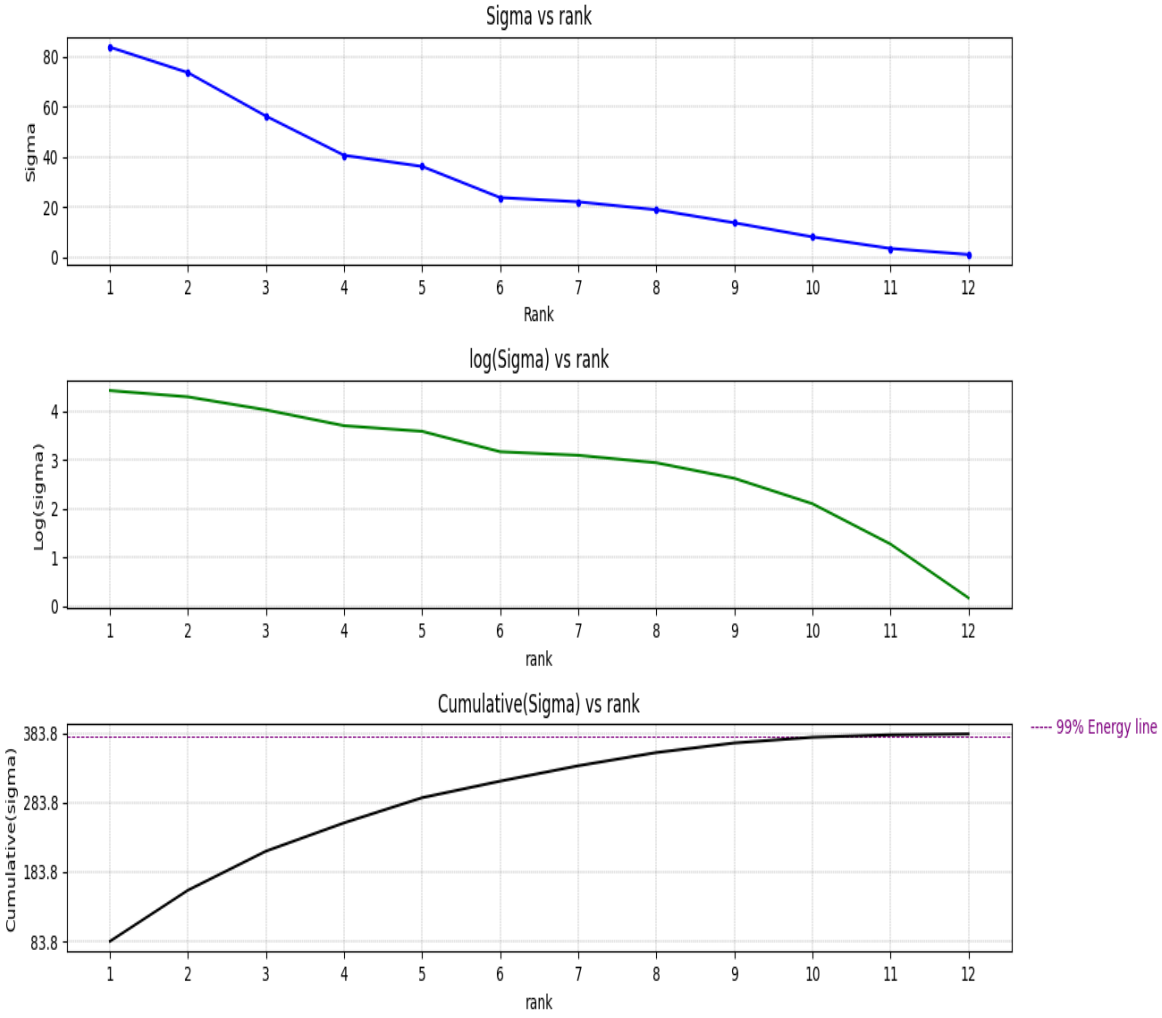


Figure 5.5: S vs rank graphical representation for normalized A.

Unlike the raw data which is suitable for a low rank approximation (As shown in the section 5.1.1) with 99% energy capture, normalized data set behaved differently when the SVD is applied. As it can be seen from the image above, rank-10 approximation is most likely to capture 99% of energy in the distribution. Notably, singular values showed a linear behaviour vs rank compared to the exponential behaviour shown by the raw data. A detailed rank-3 vs rank-10 approximation was carried out for all the data sets in a similar manner described in the section above. The basis for selecting rank-3 approximation was to evaluate the feasibility and the accuracy of low rank approximation in a normalized IMU sensor dataset.

The result of the rank-3 vs rank-10 Knee angle approximation is illustrated in Figure 5.6.

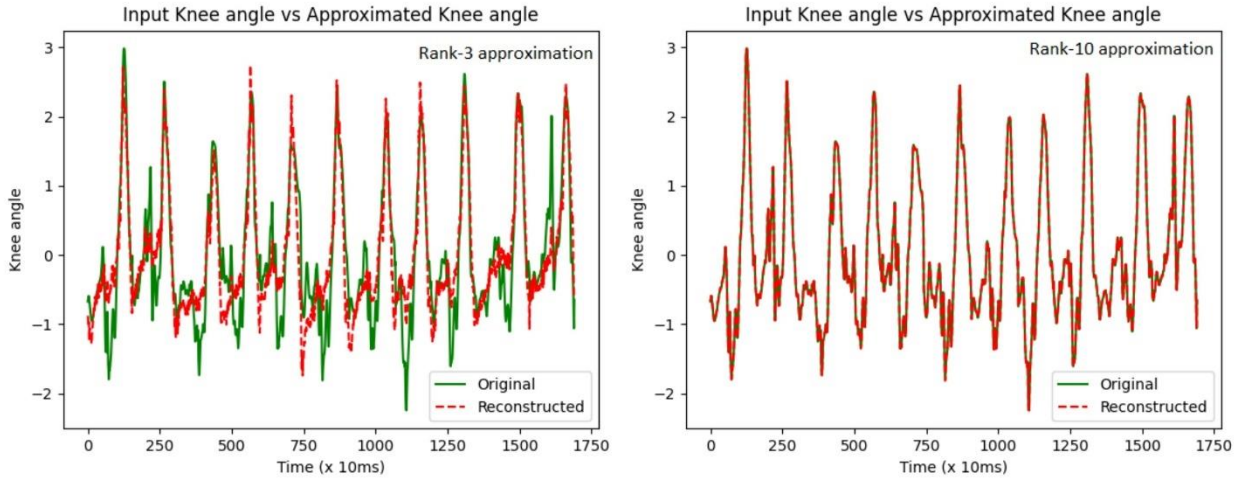


Figure 5.6: Knee angle approximation for normalized dataset A.

Similarly, data set E was analysed with the same combination of selections as indicated in the Figure 5.6 .

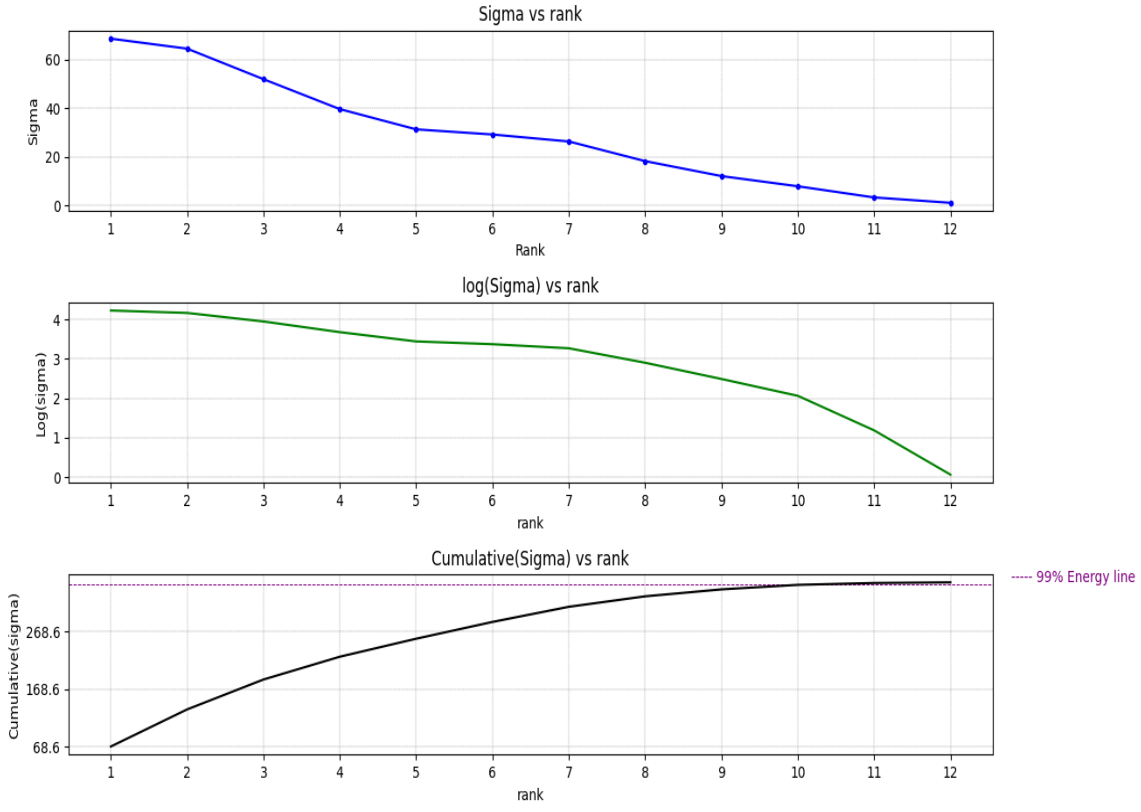


Figure 5.7: S vs Rank graphical representation for normalized dataset E.

As illustrated in the Figure 5.7, the result obtained from the normalized dataset-E is similar to the normalized dataset-A where, 99% of energy can be captured with a rank-10 approximation. A detailed rank-3 vs rank-10 approximation was analysed variable wise, and the obtained graphical Knee angle approximation comparison is shown in the Figure 5.8.

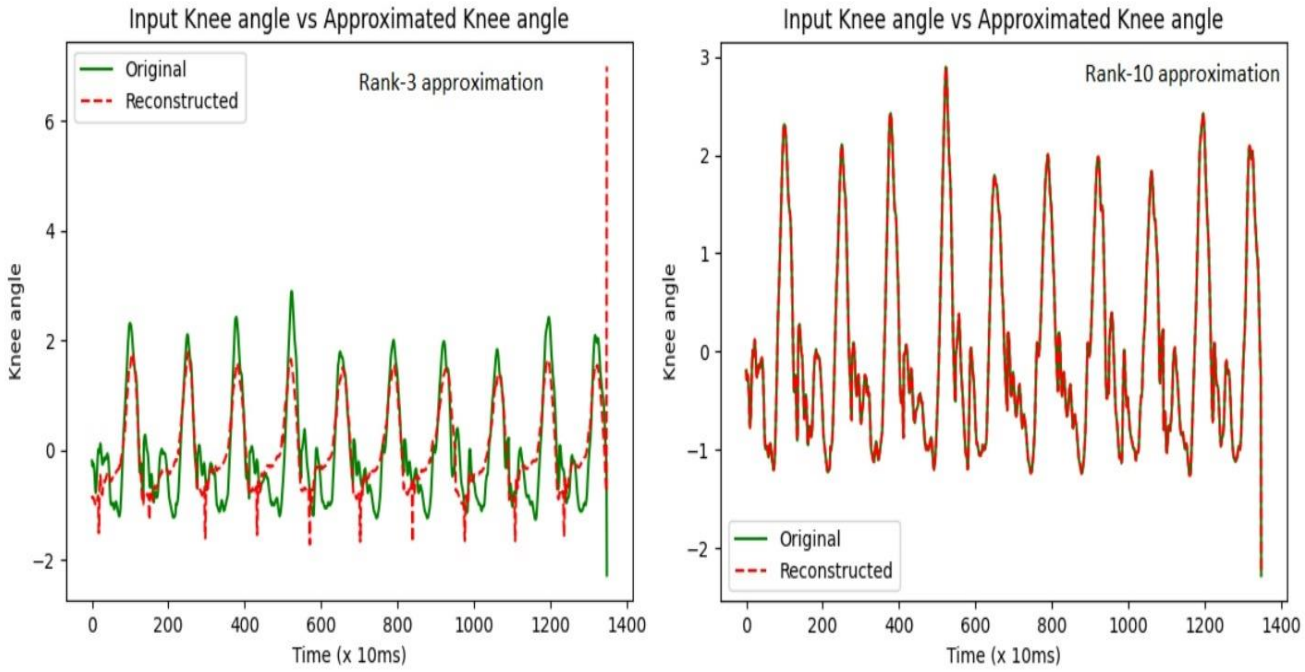


Figure 5.8: Knee angle approximation for the normalized dataset E.

In both comparisons of the normalized dataset, a rank-10 approximation perfectly captures the dynamics of the signal while a rank-3 approximation can capture the periodicity but unable to record the peaks and valleys as illustrated above.

## 5.2 SVD on two subjects on underwater trials

### 5.2.1 Raw data analysis for underwater trials

This section is a narrative of the analysis carried out on data sets obtained from underwater trials. It is comprised of an analysis between shank datasets C and G followed by a comparison between thigh datasets D and H. The basis for including shank sensor datasets is to compare the shank vs thigh approximation behaviour in underwater conditions.

The selective combination for the dataset-C is given in the Table 5-5.

Table 5-5: Configuration of the dataset C.

| **Dataset** | **Variables selected** | **Sub-Sampling rate** | **Normalized** |
|---|---|---|---|
| C | All | 1 | No |

Figure 5.9 illustrates the S vs Rank behaviour for the dataset-C.

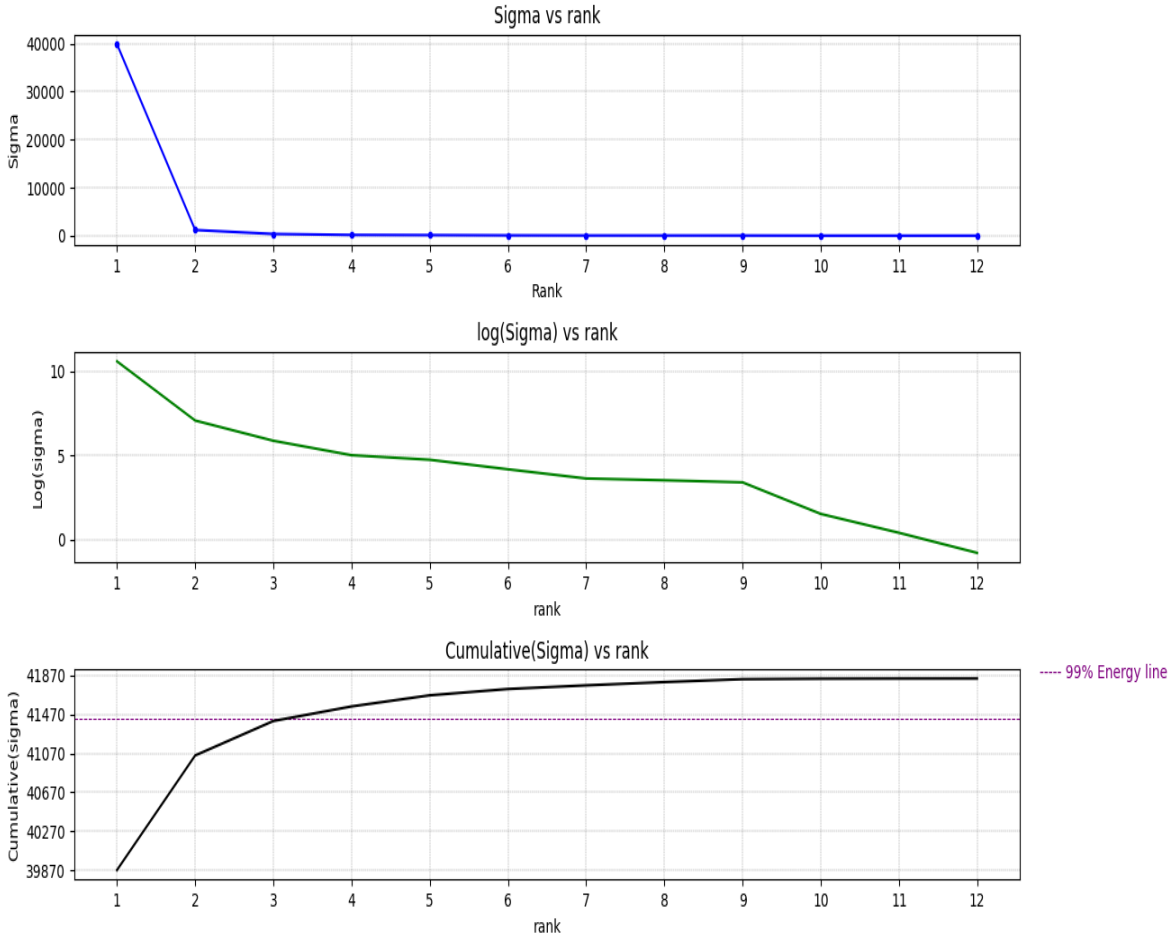


Figure 5.9: S vs Rank graphical representation for dataset C.

From the Figure 5.9 it is conclusive that 99% of energy can be captured from a rank-4 approximation. Therefore, rank-1, rank-2- rank-3 and rank-4 approximations were carried out to investigate how each approximation reconstructs the signal compared to the raw data. The combined image of the approximations is shown in the Figure 5.10

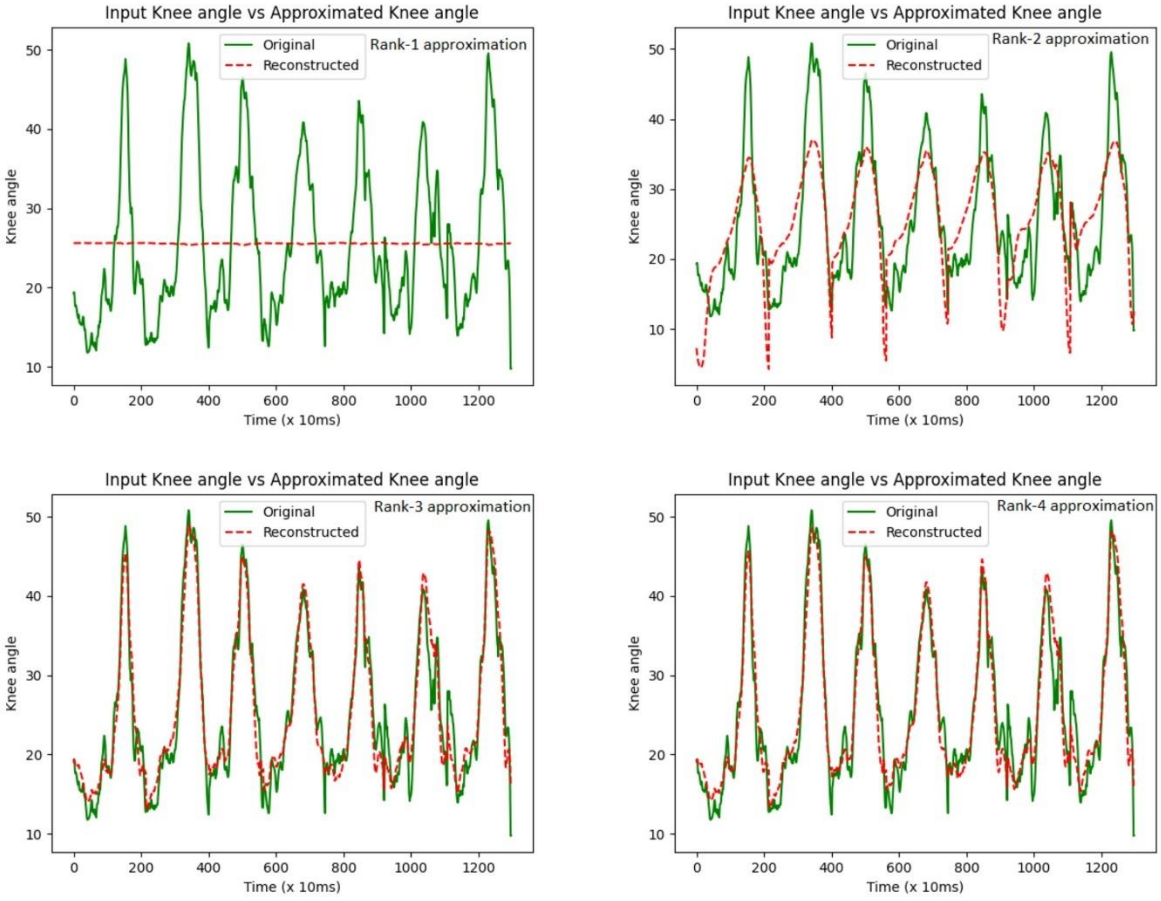


Figure 5.10: Knee angle approximation for the dataset C.

Then, dataset G was loaded with the configuration given in the Table 5-6.

Table 5-6: Upload configuration for dataset G.

| **Dataset** | **Variables selected** | **Sub-Sampling rate** | **Normalized** |
|---|---|---|---|
| G | All | 1 | No |

Figure 5.11 illustrates the S vs Rank behaviour for the dataset-G.

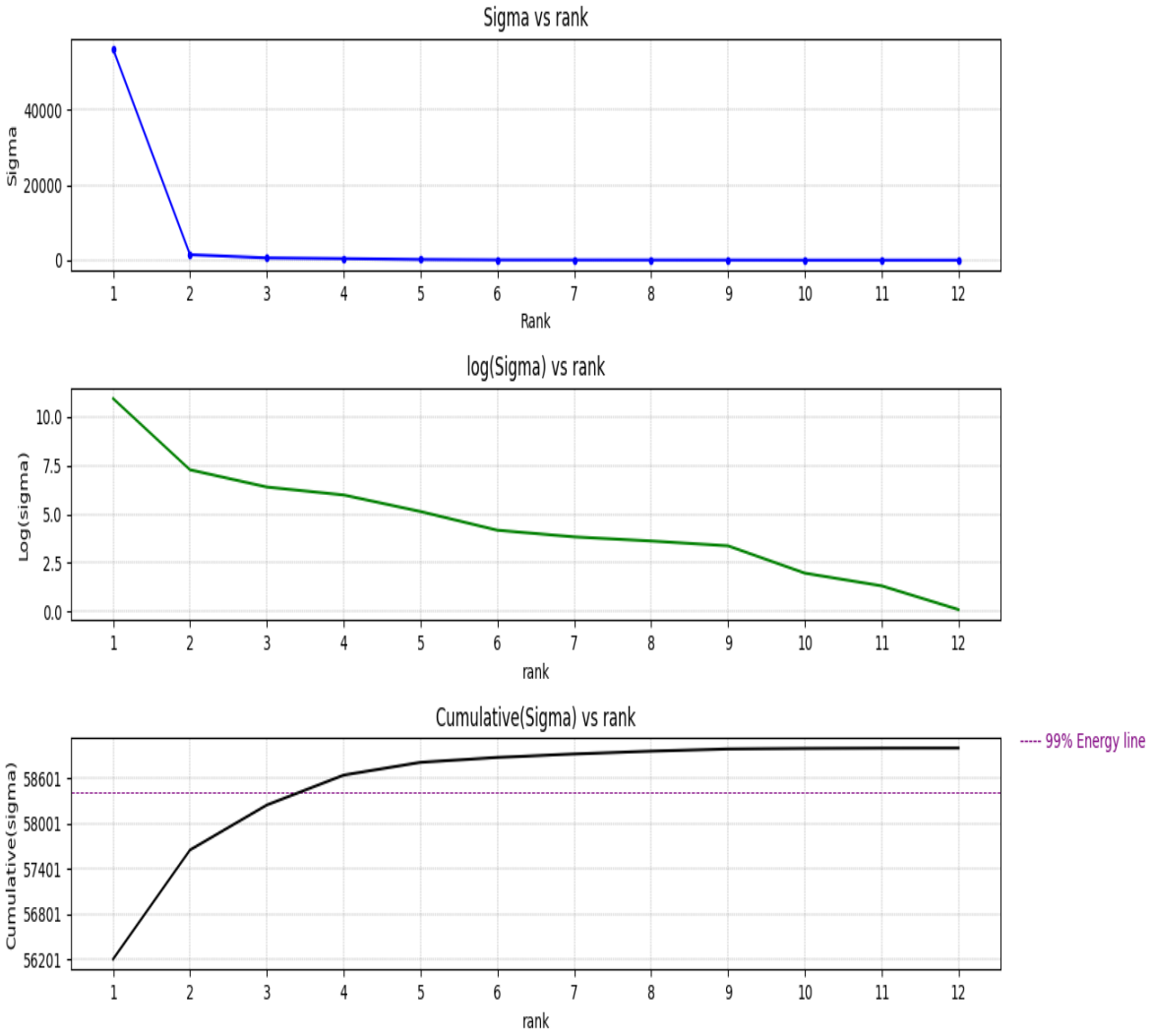


Figure 5.11: S vs Rank graphical representation for dataset G.

Based on the obtained S vs rank graph illustrated above (Figure 5.11), 99% of the energy can be captured from a rank-4 approximation. The comparison between a reconstructed rank-1, rank-2, rank-3, and rank-4 knee angle signal is illustrated in the Figure 5.12.

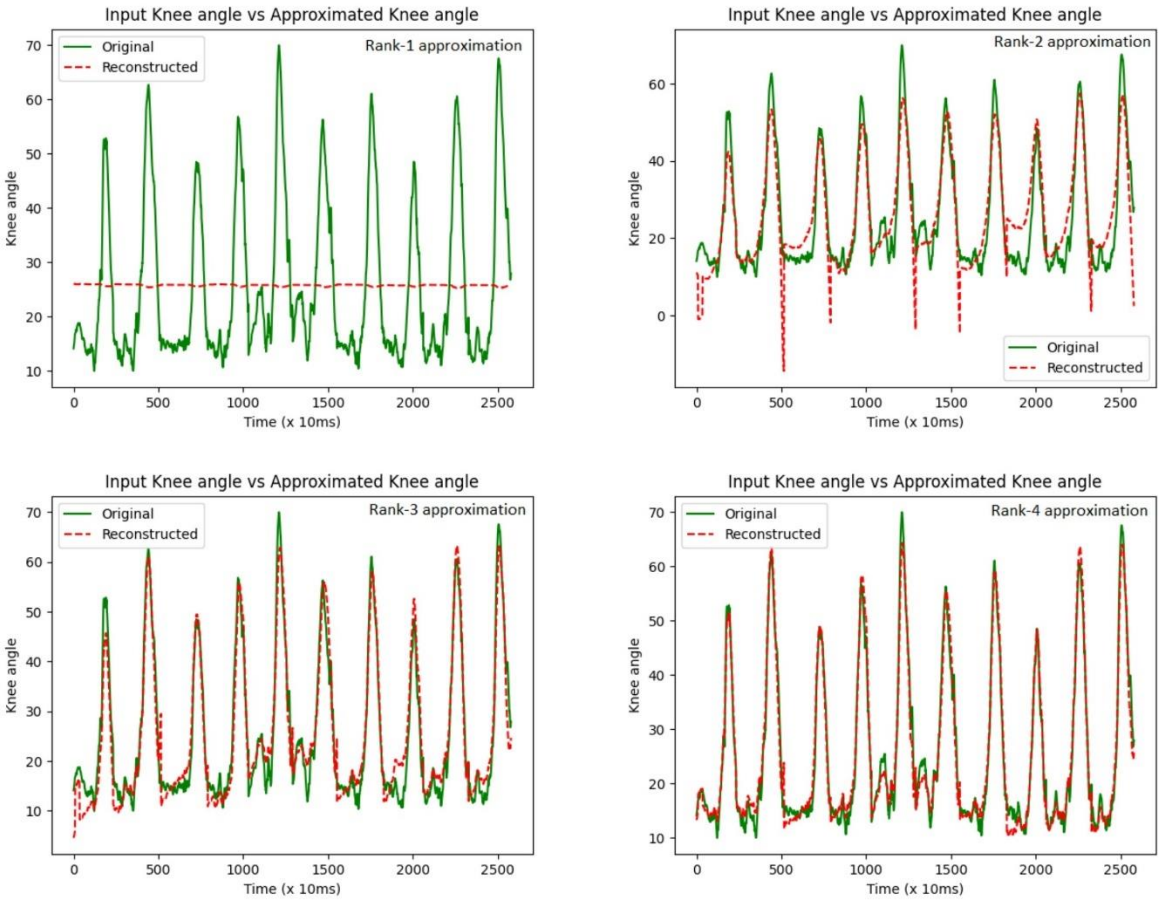


Figure 5.12: Rank-r knee angle approximation of dataset G.

Results from the thigh datasets D and H will be discussed from here onwards.

The selective combination for the dataset D is shown in the Table 5-7.

Table 5-7: Upload configuration of normalized dataset D.

| **Dataset** | **Variables selected** | **Sub-Sampling rate** | **Normalized** |
|---|---|---|---|
| D | All | 1 | No |

Figure 5.13 illustrates the S vs Rank behaviour for the dataset-D.

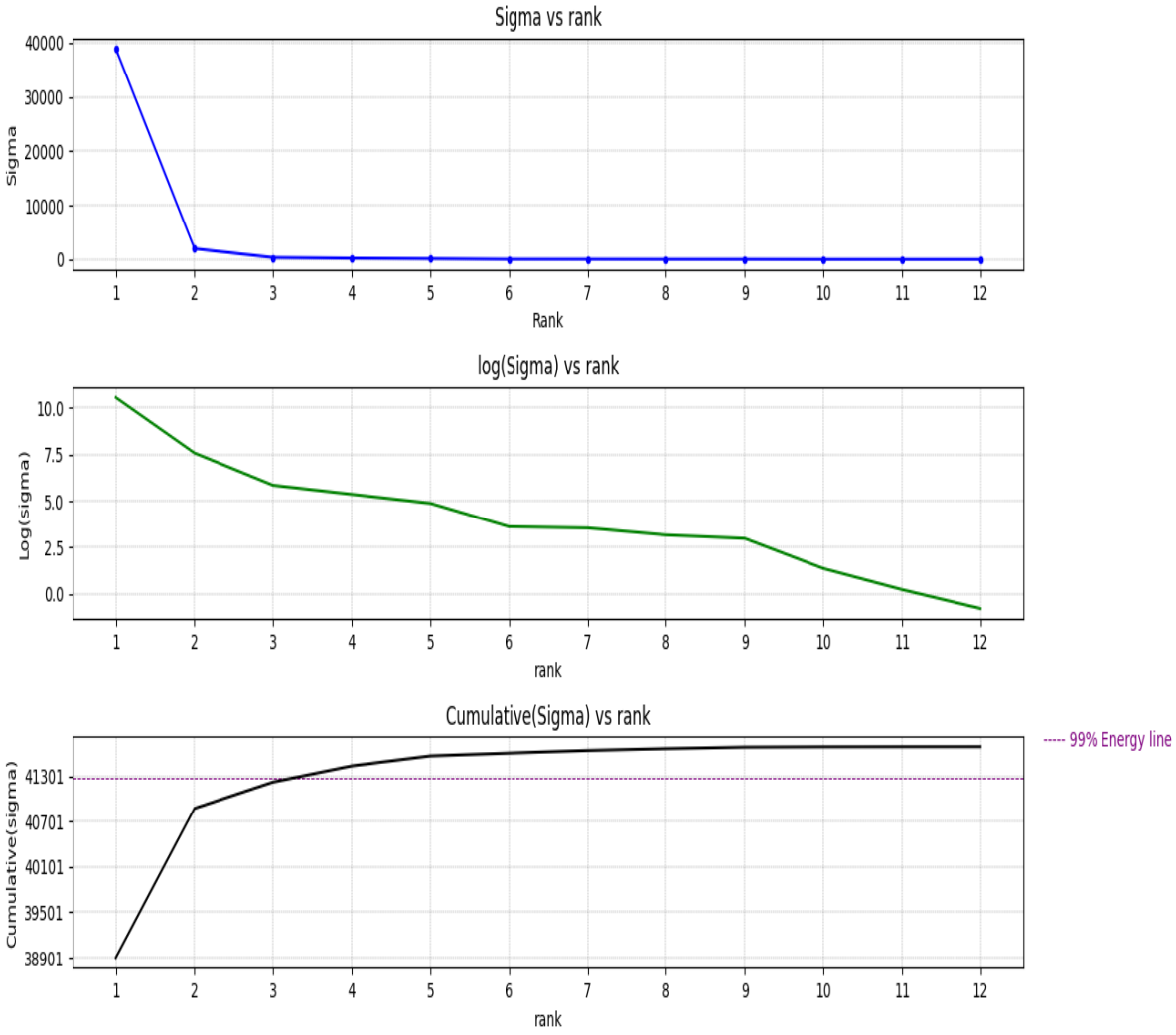


Figure 5.13: S vs Rank behaviour for the dataset D.

After evaluating the energy distribution behaviour, rank-1, rank-2, rank-3, and rank-4 approximations were carried out on the dataset-D.

The graphical representation of the Knee angle approximation across the four instances are shown in the Figure 5.14.

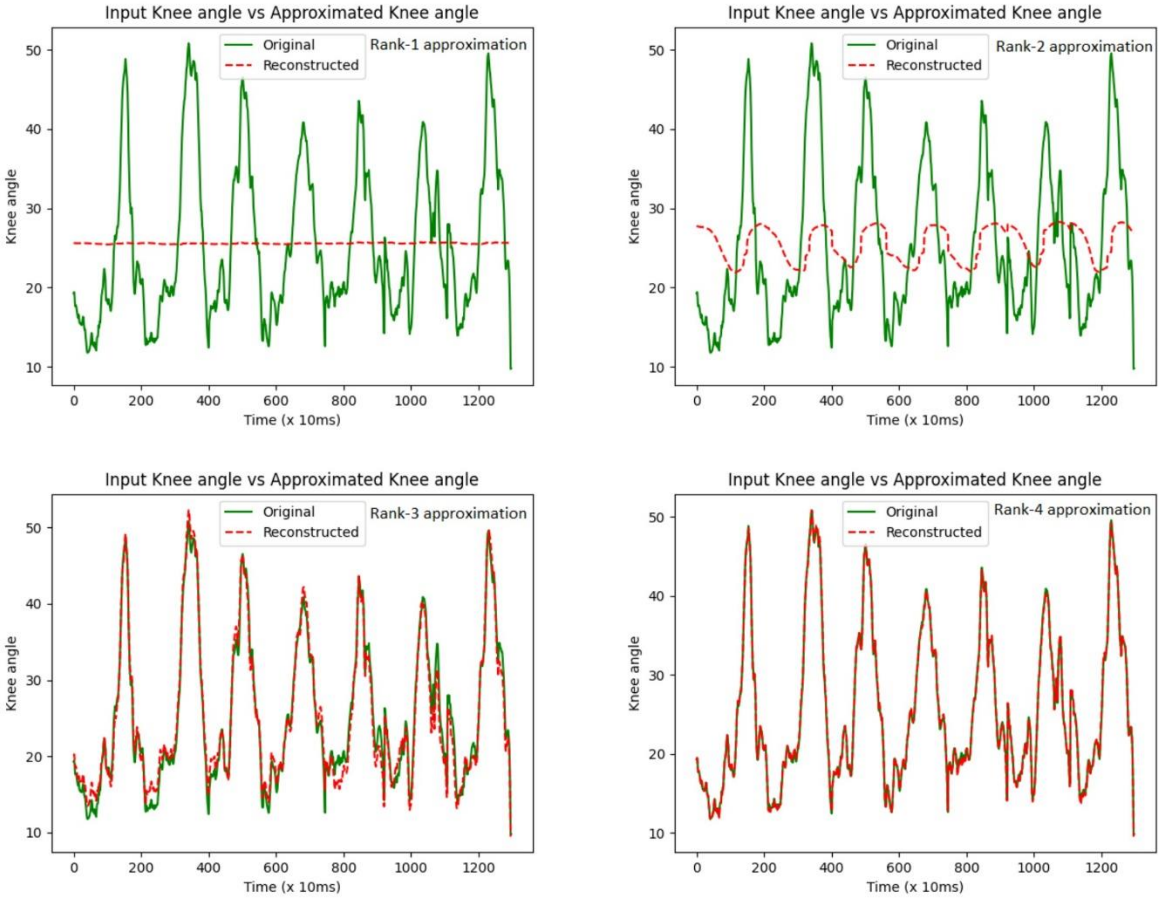


Figure 5.14: Low rank Knee angle approximation on dataset D.

Dataset H was then uploaded to the SVD utility tool with the configuration shown in the Table 5-8.

Table 5-8: Upload configuration of the dataset H.

| **Dataset** | **Variables selected** | **Sub-Sampling rate** | **Normalized** |
|---|---|---|---|
| H | All | 1 | No |

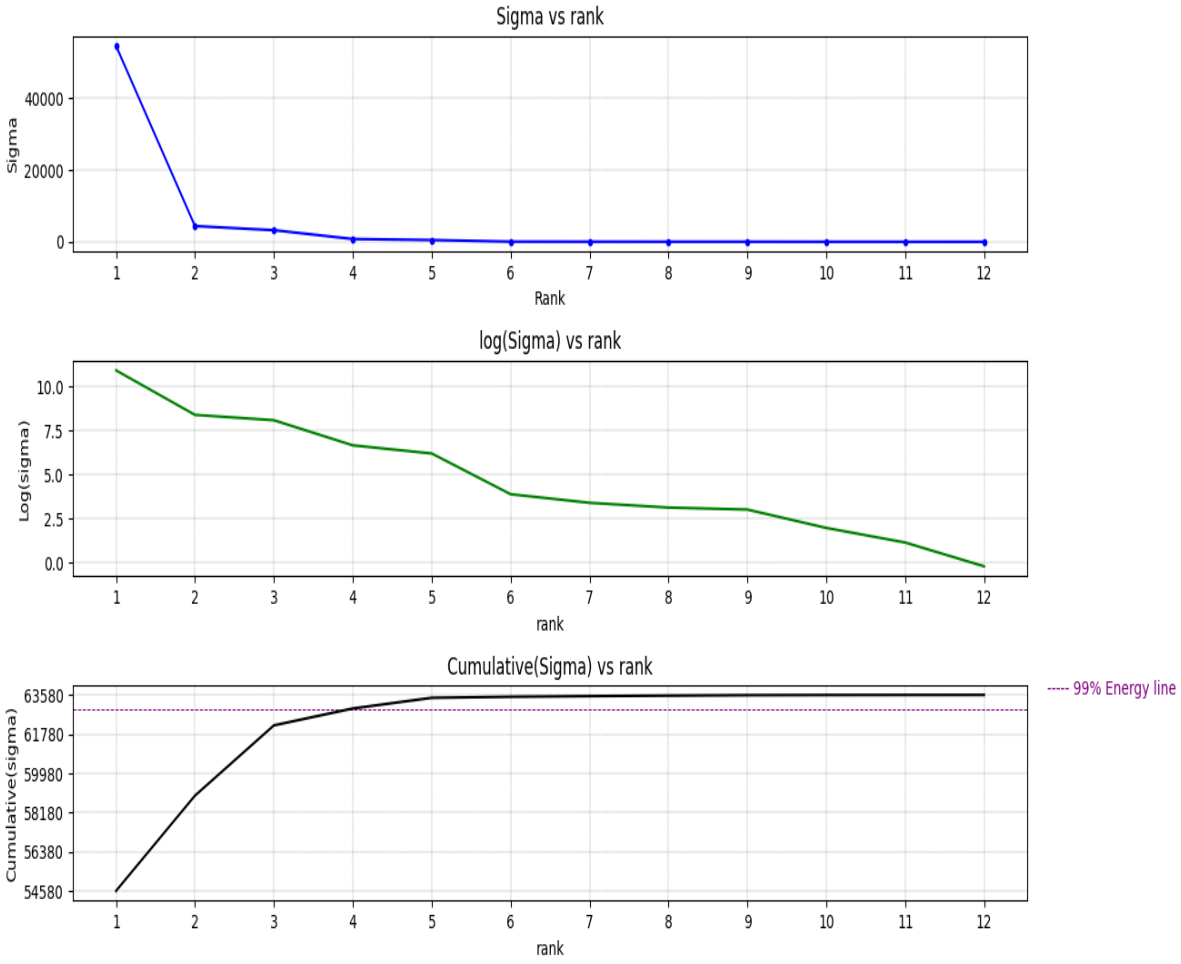


Figure 5.15: S vs Rank graphical representation of the dataset H.

From the obtained cumulative energy graph shown in the Figure 5.15, 99% of energy can be captured from a rank-5 approximation.

To evaluate the approximation accuracy, rank-1, rank-3, and rank-4 approximations were carried out alongside rank-5 approximation (expected best fit). The graphical representation of the Knee angle approximation across the four instances is shown in the Figure 5.16. Rank-2 approximation was not included in the figure below, as the pattern was much similar to that of the rank-3 approximation.

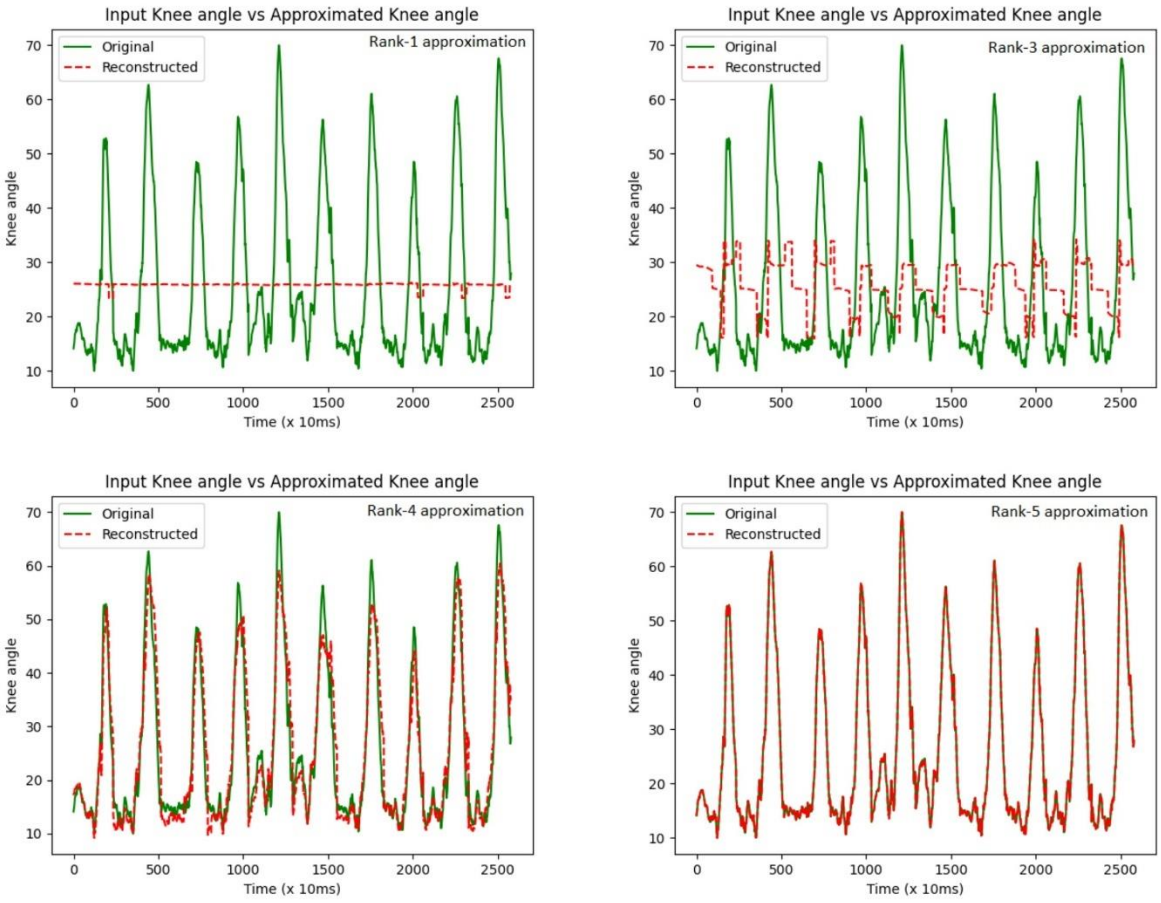


Figure 5.16: Rank-r knee angle approximation of the dataset H.

### 5.2.2 Normalized data analysis for underwater trials

This section will include a comparison of knee angle approximations on normalized datasets between shank and thigh sensor readings from trial 2. Similar to the section 5.2.1, results from set C and G will be initially indicated followed by the results from set D and H. Dataset C was first uploaded with the combination of user inputs given in the Table 5-9.

Table 5-9: Upload configuration of normalized dataset C.

| **Dataset** | **Variables selected** | **Sub-Sampling rate** | **Normalized** |
|---|---|---|---|
| C | All | 1 | Yes |

Resulting S vs Rank graph and the rank-3 vs rank-10 knee angle approximation is illustrated in the Figure 5.17 and Figure 5.18 respectively.

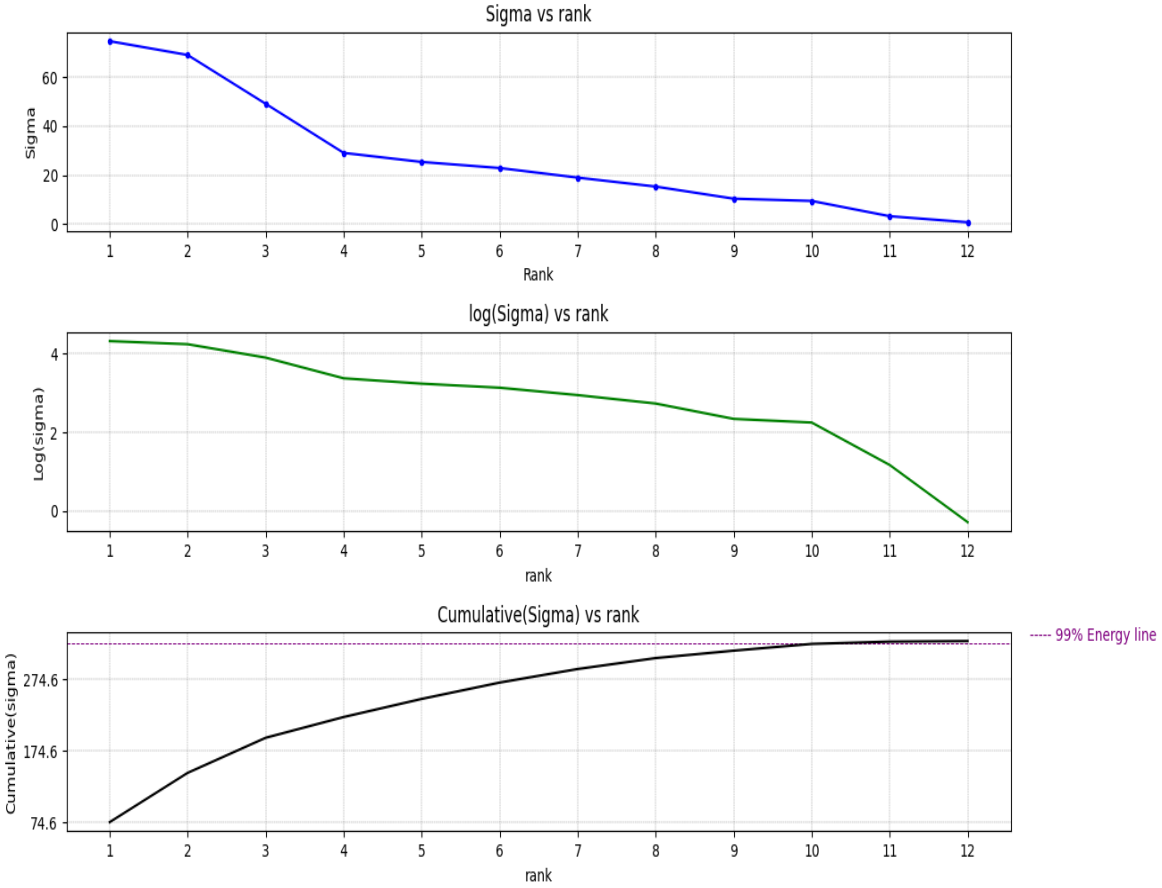


Figure 5.17: S vs Rank results from the normalized dataset C.

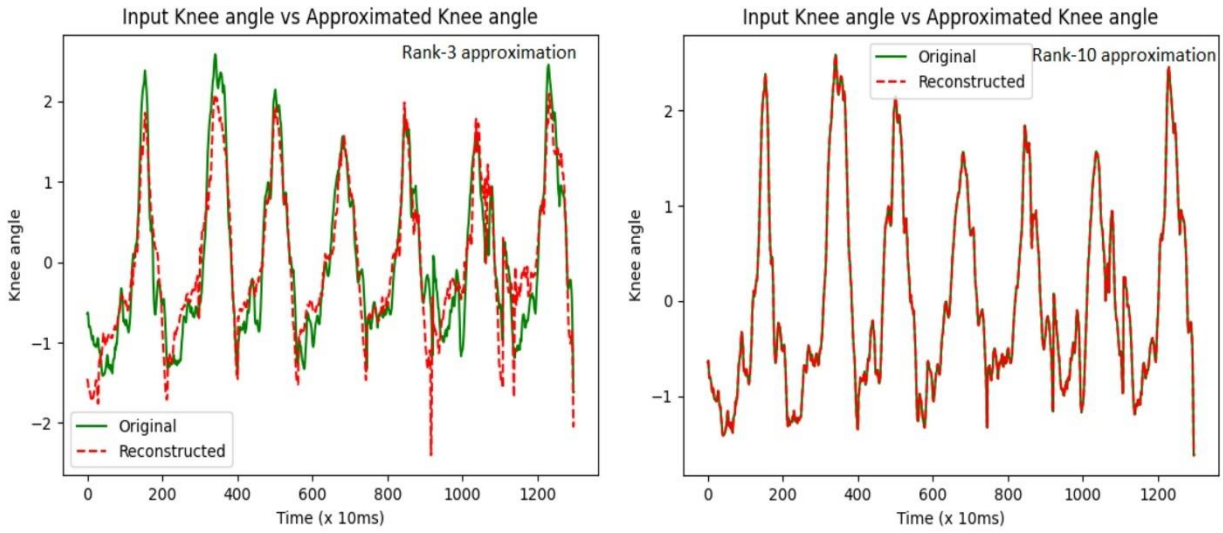


Figure 5.18: Knee angle approximation on normalized set C.

As indicated in the in the Figure 5.17, 99% energy of the signal can be captured by a rank 10 approximation. But a rank-3 approximation is also carried out to evaluate its feasibility as a possible estimator. Figure 5.18 shows a comparison between a rank-3 approximation and a rank-10 approximation. Then the dataset G was uploaded to the SVD tool with the configuration given in the Table 5-10.

Table 5-10: Upload configuration of the normalized dataset G.

| **Dataset** | **Variables selected** | **Sub-Sampling rate** | **Normalized** |
|---|---|---|---|
| G | All | 1 | Yes |

The resulting S vs rank graphs are illustrated in the Figure 5.19.

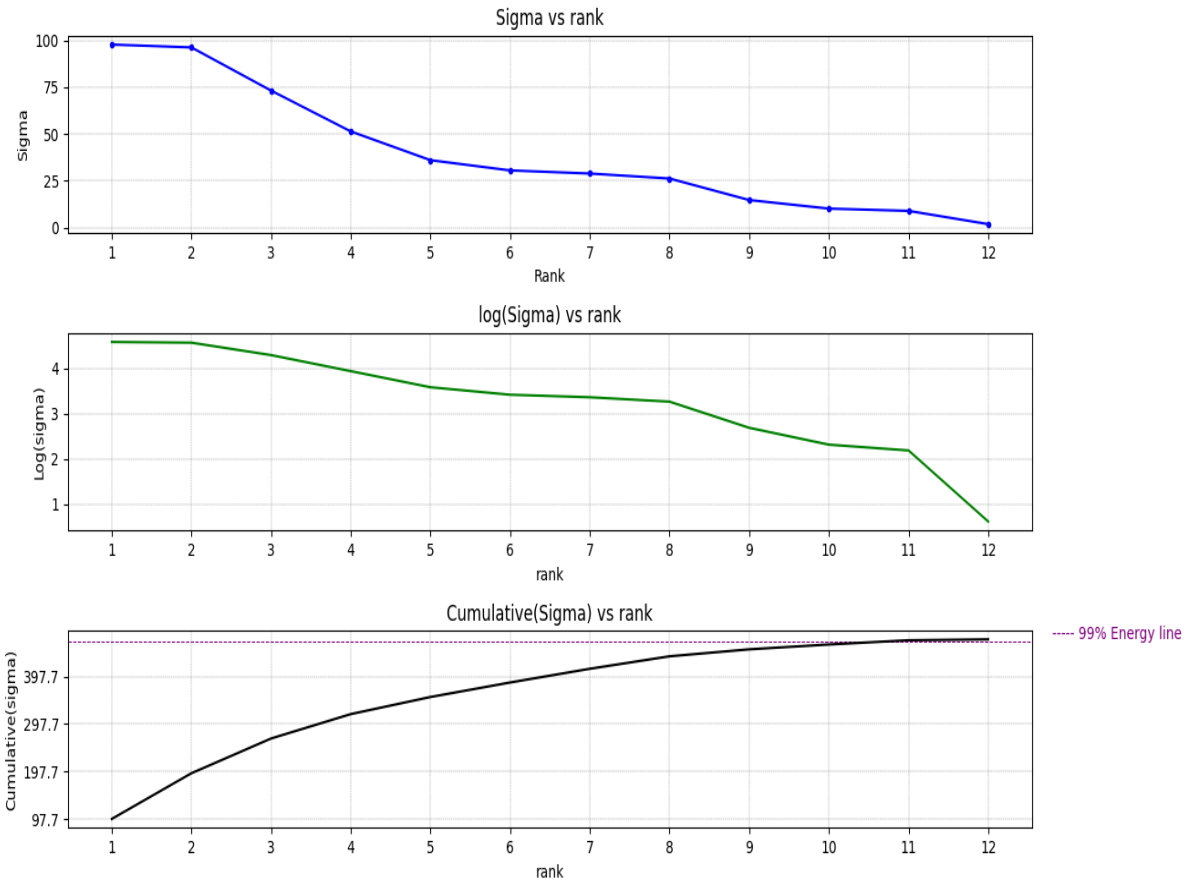


Figure 5.19:S vs Rank representation for normalized dataset G.

It is evident that a rank-11 approximation will capture 99% of energy while the feasibility of a rank-3 approximation is also evaluated as shown in the Figure 5.20.

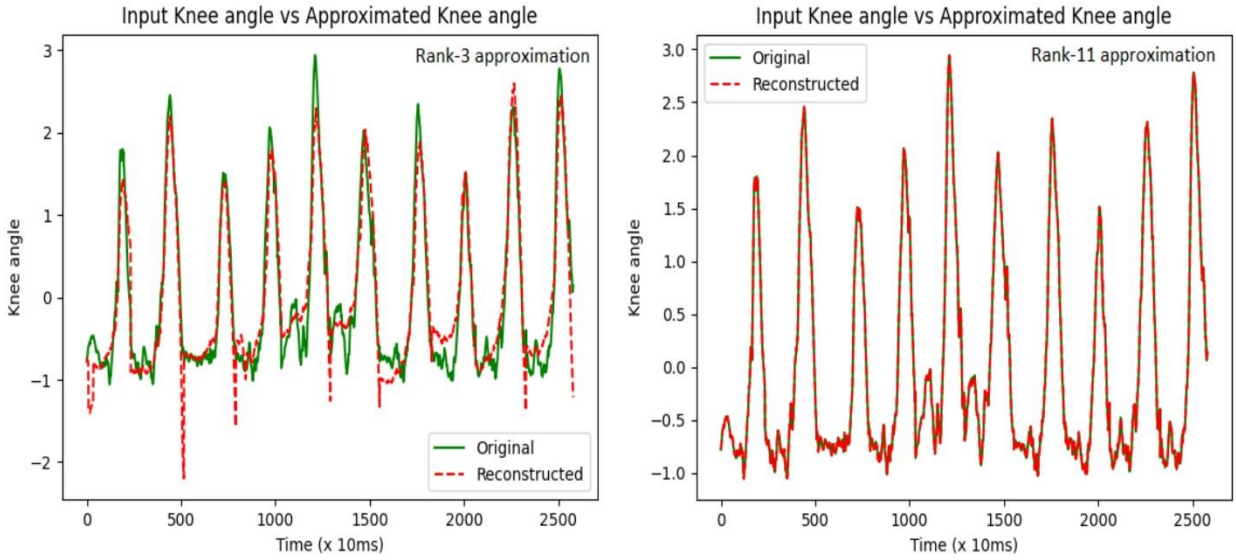


Figure 5.20: Rank-3 vs rank-11 approximation of the normalized dataset G.

Similarly, dataset D and H are analysed with the upload configuration indicated in the Table 5-11.

Table 5-11: Upload configuration for normalized datasets D and H.

| **Dataset** | **Variables selected** | **Sub-Sampling rate** | **Normalized** |
|---|---|---|---|
| D | All | 1 | Yes |
| H | All | 1 | Yes |

S vs Rank illustration and the Knee angle approximation comparison of the normalized dataset D is shown in Figure 5.21 and Figure 5.22 respectively. From the Figure 5.21, it is conclusive that a rank-10approximation is more likely to capture 99% of the energy in the distribution. A rank-3 approximation is also captured for comparison purposes.

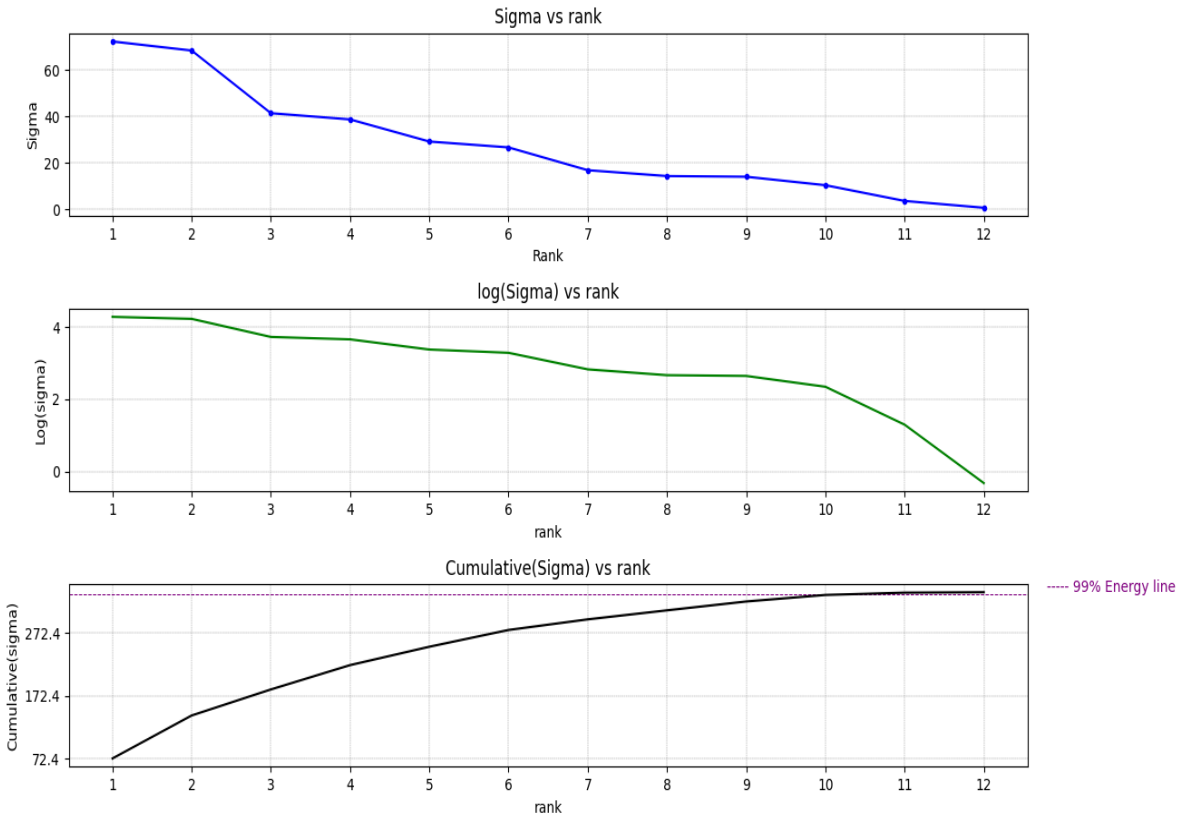


Figure 5.21: S vs Rank graphical representation for normalized D.

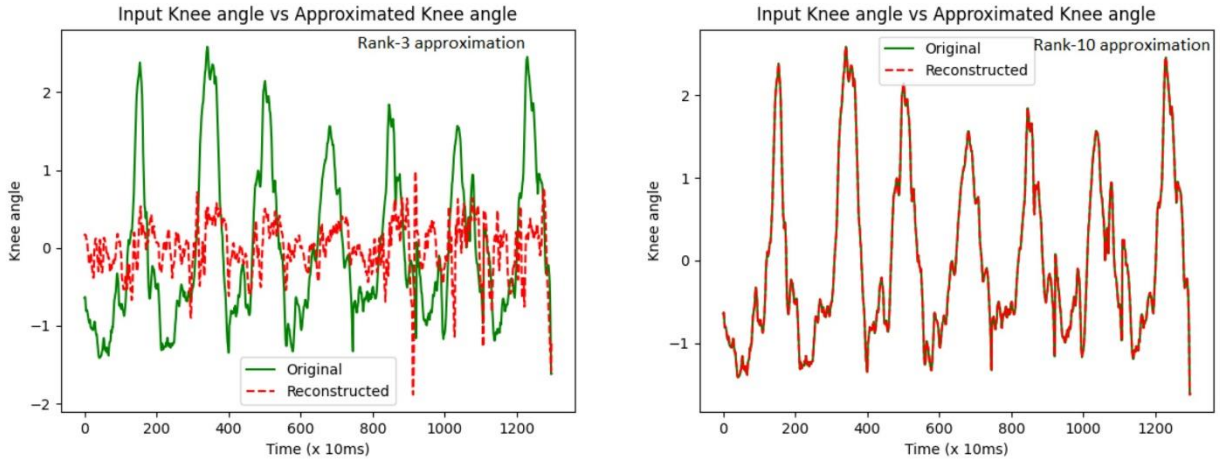


Figure 5.22: Knee angle approximation on normalized dataset D.

Then dataset H was uploaded with the configuration shown in the Table 5-11. The obtained S vs Rank graph is given in the Figure 5.23.

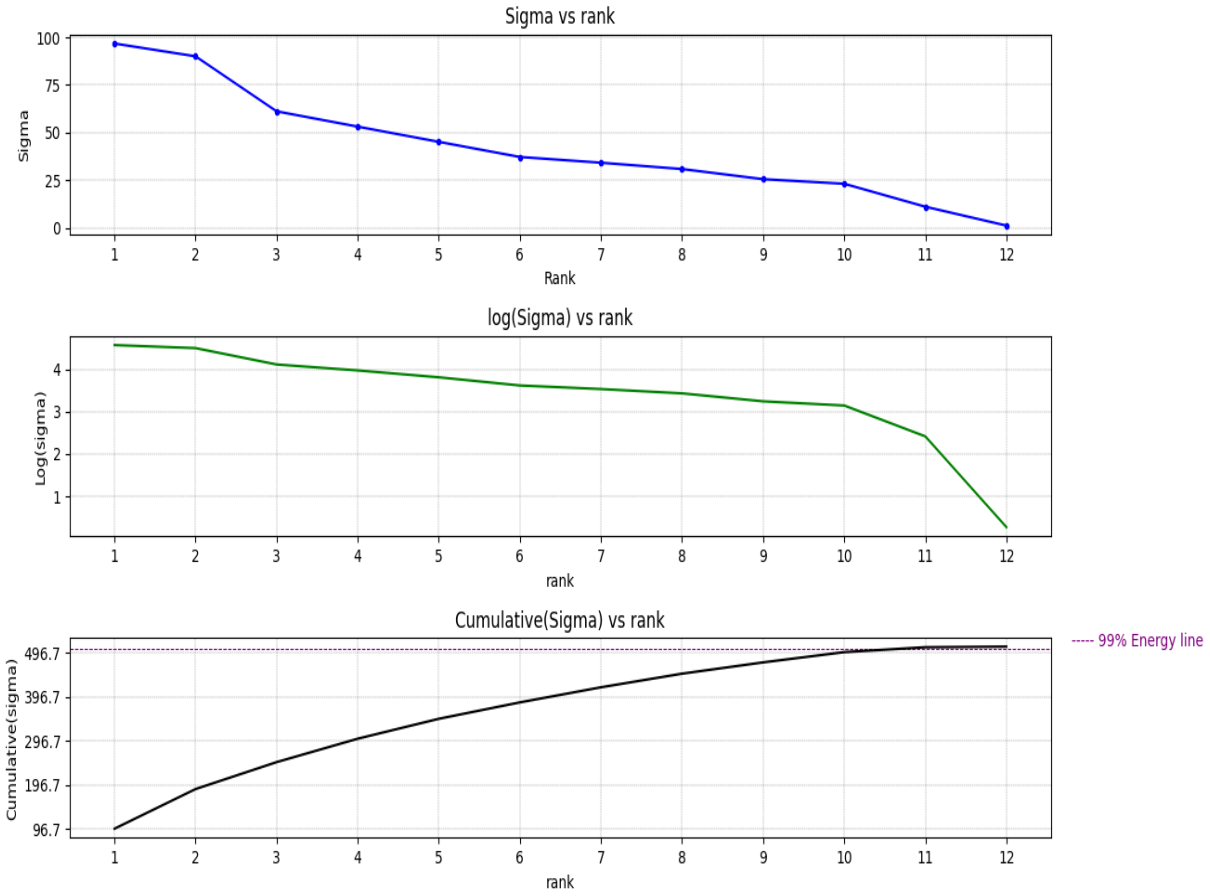


Figure 5.23: S vs Rank representation of normalized H.

From the obtained S vs Rank graph, 99% of energy can be captured with a Rank-10 approximation. A Rank-3 vs Rank-10 approximation was carried out on the normalized dataset-H to evaluate the accuracy component in each approximation. Basis for executing a rank-3 approximation is to evaluate the feasibility of fitting a rank-3 approximation to a normalized IMU sensor data set.

The obtained knee angle approximation pattern is shown in the Figure 5.24.

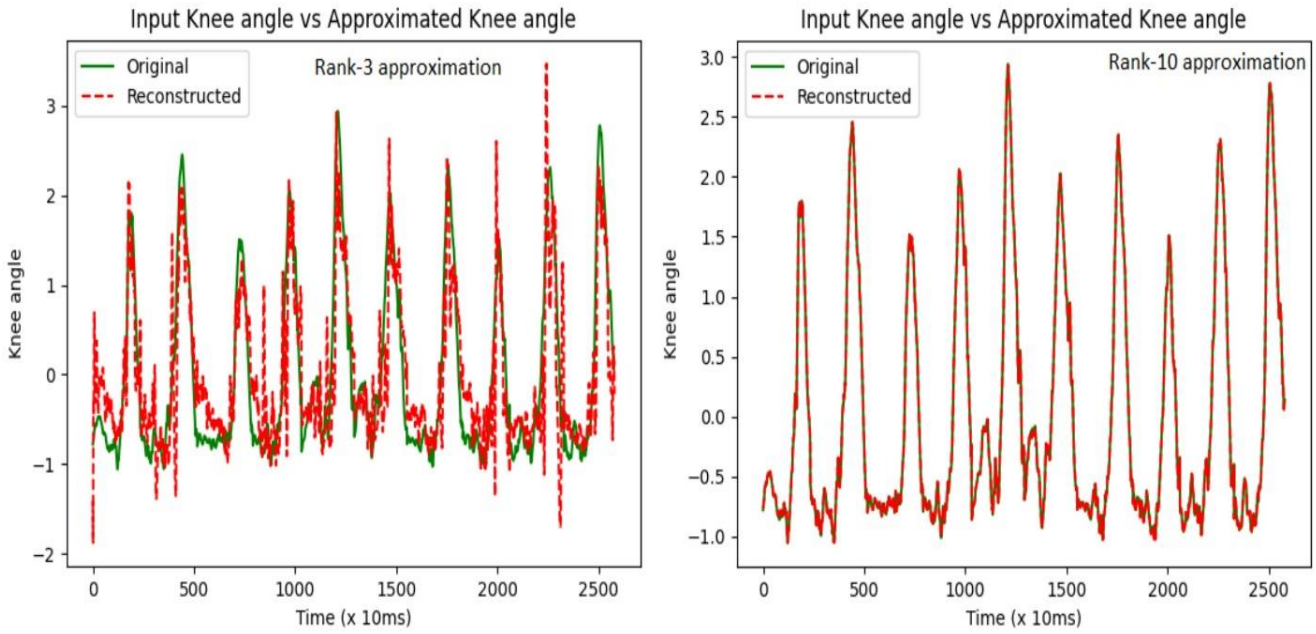


Figure 5.24:Knee angle approximation on normalized dataset H.

## 5.3 Identification of faulty data

Even though the data integrity is ensured during the data acquisition phase, some trial repetitions contain faulty data due to occasional sensor malfunction or noise. Identification of faulty data during data acquisition phase is described in the section 3.2 (Data acquisition). Due to the massive amount of data files acquired during multiple trials, some data were not comprehensively checked during the data acquisition phase. Though, the faulty data will get identified after a rank-r approximation.

The Figure 5.25 is an example of identifying three faulty repetitions (Trial 2) through a rank-3 magnetic field approximation in the dataset C. After discovering faulty datasets in a similar manner, those repetitions were removed immediately from the database and the analysis was carried out with the remaining repetitions of the dataset.

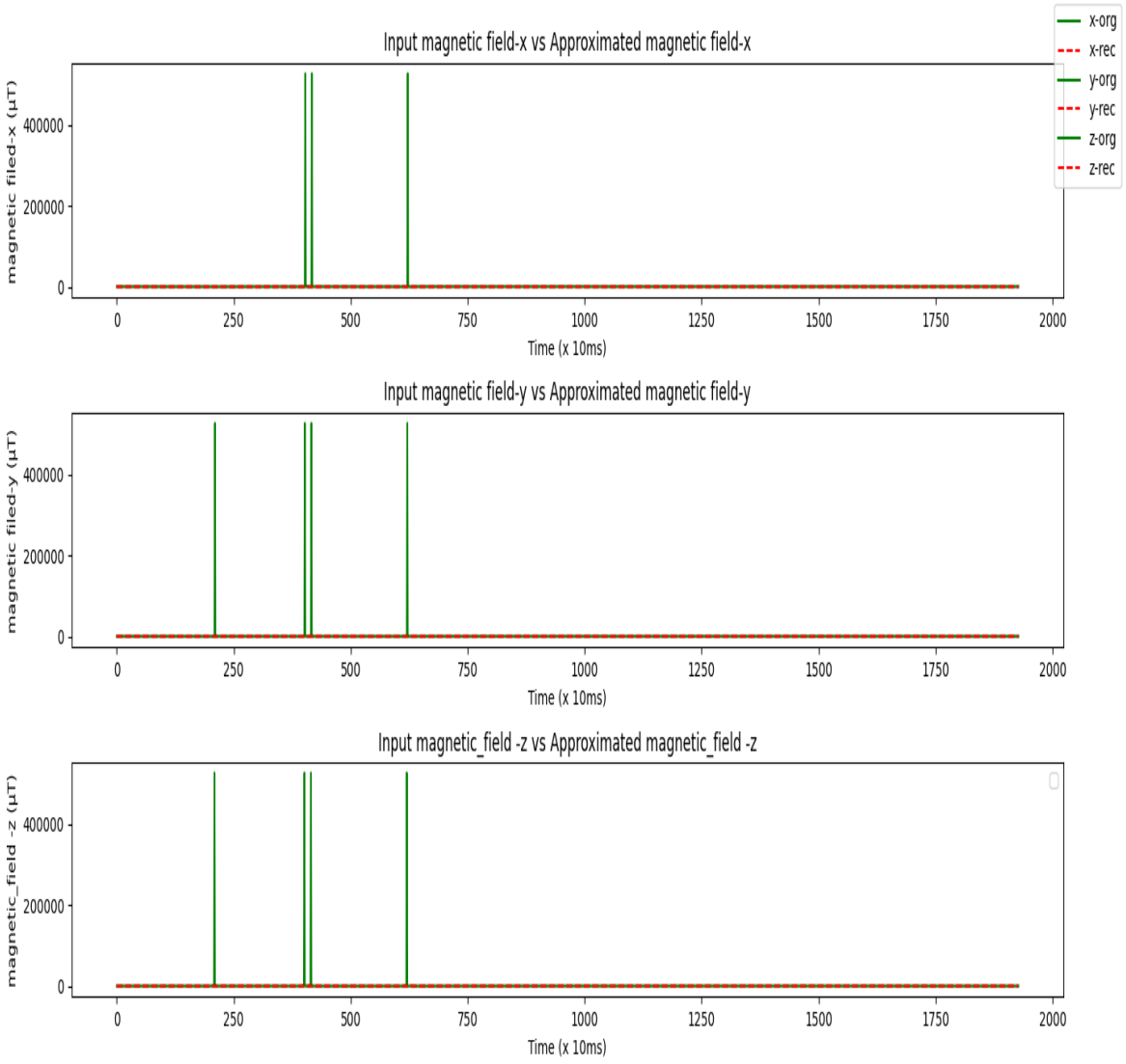


Figure 5.25: Identification of faulty data from rank-r approximation.

## 5.4 Results validation and Accuracy

This chapter describes how the approximated data were validated based on the RMSE and R calculations. Error calculation and the involved steps are described in the chapter 4.3.5.

### 5.4.1 Accuracy comparison between rank-r approximations on a single dataset

A comparison between rank-1, rank-3 and rank-4 approximations on the raw dataset-A is indicated in the Table 5-12. The best performance is marked in green, and the worst in red for each experiment.

Table 5-12: RMSE and R comparison between rank-r approximations on dataset A.

| **ID** | **Dataset =A** | **Rank 1 approximation** | | **Rank 3 approximation** | | **Rank 4 approximation** | |
|---|---|---|---|---|---|---|---|
| | **Variable** | **RMSE** | **R** | **RMSE** | **R** | **RMSE** | **R** |
| 1 | Pressure | 3.774 | 0.135 | 1.071 | 0.141 | 0.551 | 0.236 |
| 2 | Magnetic-X | 11.814 | 0.992 | 3.074 | 0.971 | 2.207 | 0.985 |
| 3 | Magnetic-Y | 13.949 | 0.958 | 3.868 | 0.962 | 1.891 | 0.991 |
| 4 | Magnetic-Z | 30.191 | -0.804 | 1.098 | 0.999 | 0.553 | 0.999 |
| 5 | Acceleration-X | 3.226 | -0.143 | 3.156 | 0.207 | 3.131 | 0.24 |
| 6 | Acceleration-Y | 4.4 | -0.206 | 3.811 | 0.499 | 3.643 | 0.56 |
| 7 | Acceleration-Z | 1.772 | 0.009 | 1.584 | 0.448 | 1.533 | 0.501 |
| 8 | Quaternion a | 0.503 | 0.126 | 0.5 | 0.113 | 0.499 | 0.116 |
| 9 | Quaternion b | 0.415 | 0.107 | 0.413 | 0.089 | 0.413 | 0.089 |
| 10 | Quaternion c | 0.198 | 0.448 | 0.175 | 0.469 | 0.174 | 0.474 |
| 11 | Quaternion d | 0.334 | -0.091 | 0.331 | 0.123 | 0.331 | 0.127 |
| 12 | Knee angle | 15.392 | -0.805 | 3.4 | 0.975 | 0.45 | 0.999 |

The data were plotted to detail the dynamic behaviour of RMSE and R across low ranks as indicated in the Figure 5.26 and Figure 5.27 respectively.

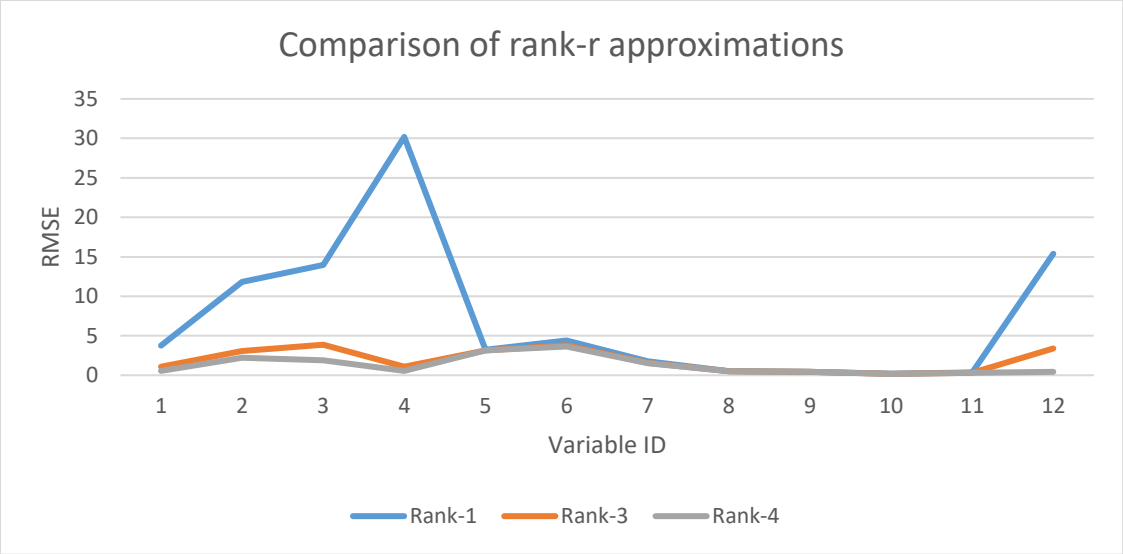


Figure 5.26 Comparison of RMSE in dataset A across rank-r approximations.

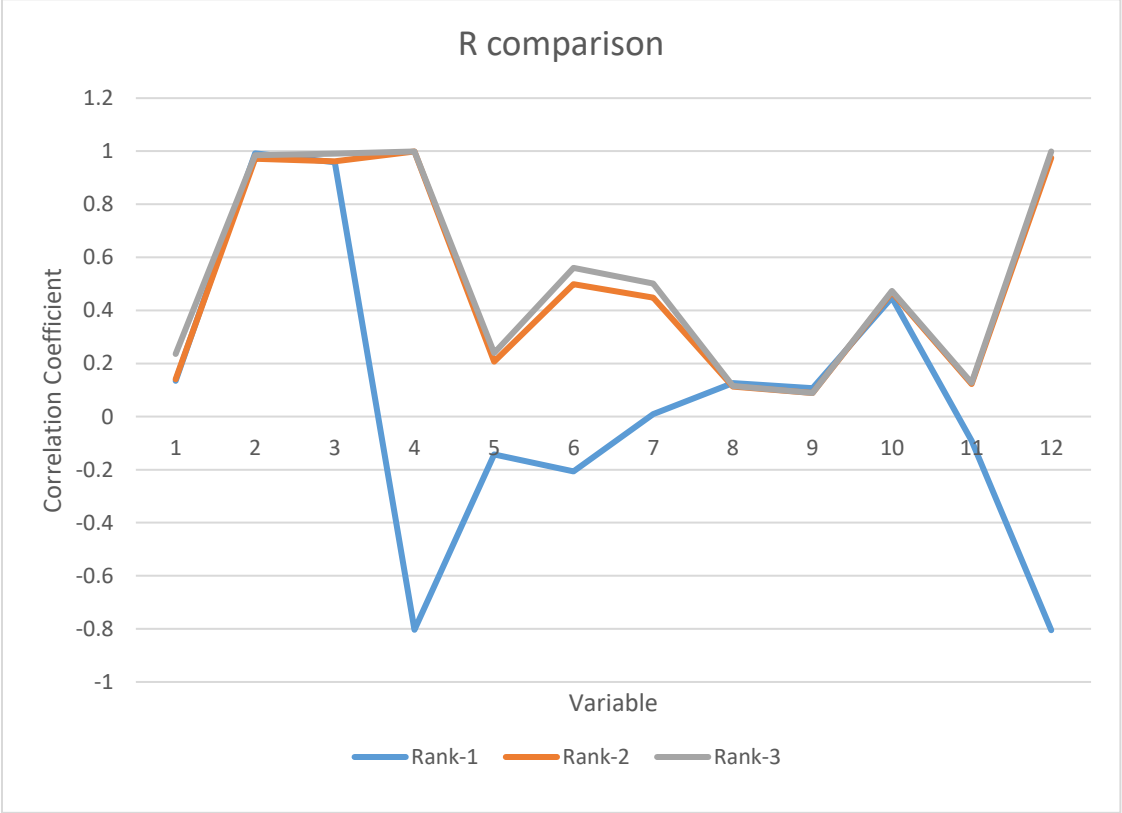


Figure 5.27 Comparison of R in dataset A across rank-r approximations.

A comparison between the rank-3 and rank-10 approximations on normalized dataset-A is shown in the Table 5-13. The best performance is marked in green, and the worst in red for

each experiment. The average RMSE and R were also calculated as the data distribution was of similar shape.

Table 5-13: RMSE and R comparison on normalized dataset A.

| **Dataset =A (Normalized)** | **Rank 3 approximation** | | **Rank10 approximation** | |
|---|---|---|---|---|
| **Variable** | **RMSE** | **R** | **RMSE** | **R** |
| Pressure | 0.845 | 0.534 | 0 | 0.999 |
| Magnetic-X | 0.264 | 0.964 | 0.017 | 0.999 |
| Magnetic-Y | 0.276 | 0.96 | 0.007 | 0.999 |
| Magnetic-Z | 0.458 | 0.888 | 0.007 | 0.999 |
| Acceleration-X | 0.884 | 0.467 | 0.001 | 0.999 |
| Acceleration-Y | 0.449 | 0.893 | 0.002 | 0.999 |
| Acceleration-Z | 0.53 | 0.847 | 0.001 | 0.999 |
| Quaternion a | 0.048 | 0.998 | 0.035 | 0.999 |
| Quaternion b | 0.085 | 0.996 | 0.053 | 0.998 |
| Quaternion c | 0.387 | 0.922 | 0.026 | 0.999 |
| Quaternion d | 0.236 | 0.971 | 0.058 | 0.998 |
| Knee angle | 0.459 | 0.888 | 0.003 | 0.999 |
| **Average** | **0.41** | **0.86** | **0.02** | **0.999** |

### 5.4.2 Accuracy comparison between approximations on two datasets

The accuracy and validation of the on-land data sets observed from both shank and thigh sensors on subject 1 and 2 will be discussed in this chapter. For this comparison, the best fit approximation (with 99% of captured energy) has been used. The results from RMSE and R calculations are given in the Table 5-14. The best performance is marked in green, and the worst in red for each experiment.

Table 5-14: RMSE and R comparison between on land trials.

| **Dataset (rank-r)** | **A (rank 4)** | | **B (rank 4)** | | **E (rank 5)** | | **F (rank 5)** | |
|---|---|---|---|---|---|---|---|---|
| **Variable** | **RMSE** | **R** | **RMSE** | **R** | **RMSE** | **R** | **RMSE** | **R** |
| Pressure | 0.551 | 0.236 | 1.195 | 0.014 | 1.112 | 0.998 | 0.941 | 0.18 |
| Magnetic-X | 2.207 | 0.985 | 2.926 | 0.983 | 3.971 | 0.974 | 2.538 | 0.975 |
| Magnetic-Y | 1.891 | 0.991 | 4.194 | 0.924 | 1.479 | 0.995 | 1.531 | 0.982 |
| Magnetic-Z | 0.553 | 0.999 | 0.279 | 0.999 | 0.713 | 0.999 | 0.293 | 0.999 |
| Acceleration-X | 3.131 | 0.24 | 2.765 | 0.407 | 2.606 | 0.852 | 3.197 | 0.485 |
| Acceleration-Y | 3.643 | 0.56 | 3.098 | 0.478 | 1.922 | 0.948 | 1.973 | 0.94 |
| Acceleration-Z | 1.533 | 0.501 | 2.187 | 0.352 | 3.234 | 0.665 | 3.021 | 0.429 |
| Quaternion a | 0.499 | 0.116 | 0.473 | 0.127 | 0.638 | 0.05 | 0.69 | 0.071 |
| Quaternion b | 0.413 | 0.089 | 0.406 | 0.119 | 0.609 | 0.048 | 0.603 | 0.074 |
| Quaternion c | 0.174 | 0.474 | 0.325 | 0.202 | 0.262 | 0.06 | 0.322 | 0.088 |
| Quaternion d | 0.331 | 0.127 | 0.255 | 0.31 | 0.387 | 0.038 | 0.197 | 0.083 |
| Knee angle | 0.45 | 0.999 | 0.651 | 0.999 | 0.223 | 0.999 | 0.441 | 0.999 |

Similarly, the accuracy of the results from underwater trials can also be summarized as indicated in the Table 5-15. The best performance is marked in green, and the worst in red

for each experiment. The best fit approximation has been used for the comparison as indicated above.

Table 5-15: RMSE and R comparison between underwater trials.

| **Dataset (rank-r)** | **C (rank-4)** | | **D (rank-4)** | | **G (rank-4)** | | **H (rank-4)** | |
|---|---|---|---|---|---|---|---|---|
| **Variable** | **RMSE** | **R** | **RMSE** | **R** | **RMSE** | **R** | **RMSE** | **R** |
| Pressure | 0.334 | 0.982 | 0.863 | 0.955 | 0.112 | 0.999 | 0.621 | 0.985 |
| Magnetic-X | 0.552 | 0.998 | 2.042 | 0.981 | 0.411 | 0.999 | 0.45 | 1 |
| Magnetic-Y | 1.865 | 0.989 | 2.908 | 0.919 | 2.51 | 0.988 | 8.66 | 0.737 |
| Magnetic-Z | 0.327 | 0.999 | 0.085 | 0.999 | 0.347 | 0.999 | 0.308 | 1 |
| Acceleration-X | 1.14 | 0.128 | 0.688 | 0.218 | 1.04 | 0.284 | 0.447 | 0.412 |
| Acceleration-Y | 1.646 | 0.434 | 0.992 | 0.558 | 1.02 | 0.416 | 0.587 | 0.718 |
| Acceleration-Z | 1.058 | 0.374 | 0.632 | 0.156 | 0.655 | 0.111 | 0.45 | 0.291 |
| Quaternion a | 0.599 | 0.295 | 0.689 | 0.083 | 0.572 | 0.314 | 0.66 | 0.093 |
| Quaternion b | 0.534 | 0.291 | 0.594 | 0.09 | 0.565 | 0.323 | 0.628 | 0.103 |
| Quaternion c | 0.236 | 0.346 | 0.312 | 0.068 | 0.173 | 0.383 | 0.273 | 0.132 |
| Quaternion d | 0.445 | 0.287 | 0.227 | 0.08 | 0.423 | 0.287 | 0.188 | 0.138 |
| Knee angle | 2.435 | 0.968 | 0.223 | 0.999 | 2.115 | 0.99 | 4.166 | 0.96 |

## 5.5 Dominant pattern recognition

Dominant patterns in gait cycle(s) can be identified from a truncated V’ modes graph. Based on the correlations between dominating eigen gaits and the dominating singular values, dominating variables can be identified. A dominant pattern recognition based on the dataset B is shown in the Figure 5.28.

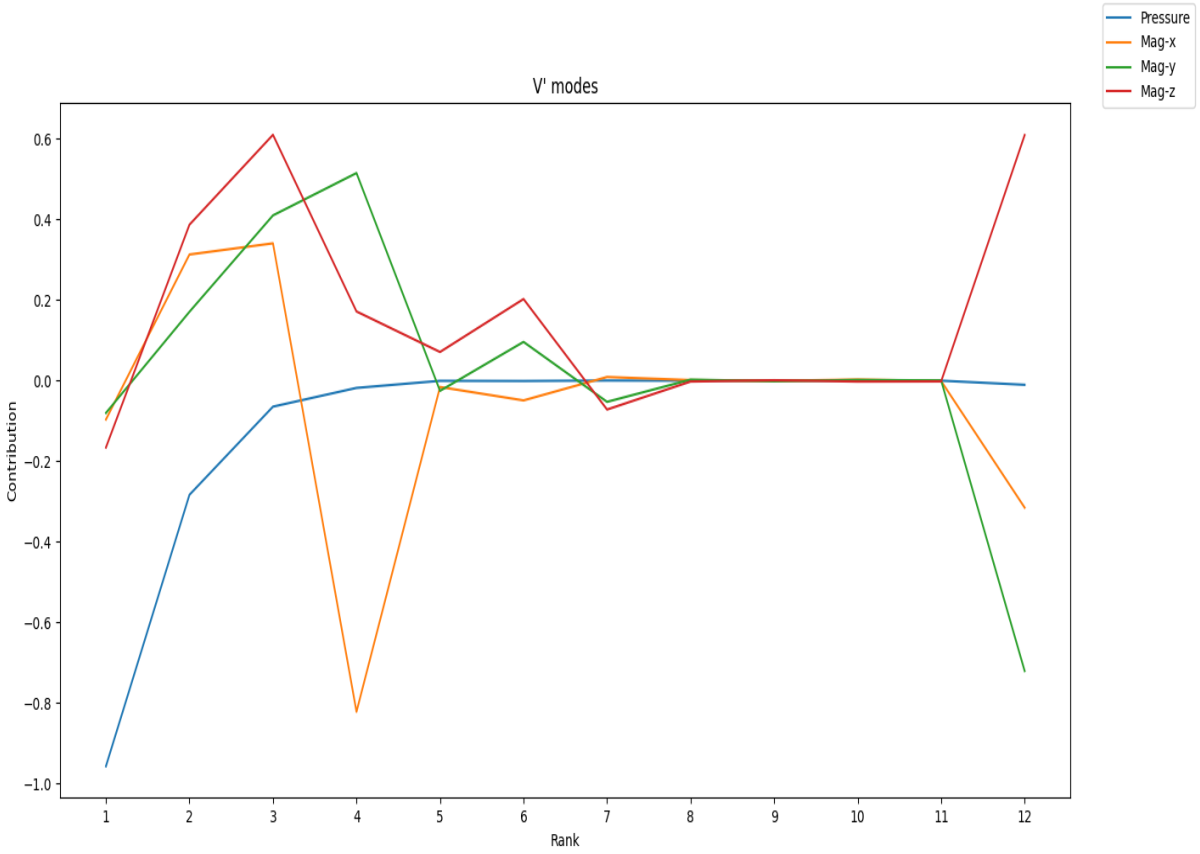


Figure 5.28: Truncated V' modes graph on dataset B.

# SUMMARY

A singular value decomposition based low rank approximation model has been proposed in this study to analyse gait data. The study was aimed to identify the dominant patterns in high-dimensional IMU sensor data with a low-dimensional approximation. The model is integrated to a python-based GUI with various features described in chapter 4. The GUI runs on Tkinter framework which provides flexibility and user-friendly environment. The dataset collected for this work was unique and adequate to evaluate the feasibility of the model.

As mentioned in the project objectives (Chapter 2.4), first objective was to apply SVD as a data driven model in the analysis. SVD was successfully applied on various datasets from various conditions and the interpretation of the decomposed matrices was identified as shown in the chapter 4.4. The concept of eigen gaits was recognized as the basis of the U matrix and the modes in V' matrix was identified as weights for each eigen gait. A detailed analysis on V' modes was identified as future work.

Second objective was to evaluate the model for successful approximation of gait data under various conditions. As shown in the chapter 5, gait data obtained from land trials are best approximated with a rank-4 approximation which captures more than 99% of the energy. Although, some on-land trial results demanded a rank-5 approximation to capture 99% of the energy in the data distribution. Even then, a rank-4 approximation still captures more than 90% of the energy which is ideal for an approximation. Results from the underwater trials suggest that a rank-4 approximation is the best dimensionality reduction that can be achieved with the objective of capturing more than 99% of the energy. Observed approximated patterns also indicate that low rank approximation is not only effective on gradually increasing or decreasing signals but also on step responses. Based on the calculated RMSE and R data shown in the chapter 5.4, it can be clearly identified that magnetic flux-Z shows the best linear behaviour across both land trials and underwater trials with a 0.999 correlation coefficient in the approximated series. Land trials also show that the knee angle approximation is also perfectly linear with a R value of 0.999. The most accurate approximation to on land gait was observed from quaternion data based on the minimum error. Major difference between land trials and underwater trials is the approximation of pressure. Underwater trials approximate the pressure with high accuracy and linearity while the land trials show non-linearity in most low rank approximations.

Both on-land data and underwater data were also normalized to study the feasibility of low rank approximation. It was found that normalized gait data requires a rank-10 approximation to capture 99% of the energy. Unlike the raw data, a rank-3 approximation leaves an average root mean square error around 0.4 on all the datasets compared to 0.02 in a rank-10 approximation. Correlation coefficient indicates the approximation is more than 80% linear on average. The SVD results on the normalized datasets also shows near equal distribution of singular values in the decomposition which is an indication of equally dominating eigen gaits. This suggests that the dominant patterns in a normalized gait distribution is hard to be recognized due to the data compression. Further SVD analysis of normalized data is open for the future exploration.

The third and fourth objectives of this study was to use a python script to identify the dominating combinations using a python script and to develop a GUI which provides full access to the user. Dominating variables were identified by observing the truncated V' modes graph (Chapter 5.5) which was generated by the python code. The basis for the identification was the number of correlations between the eigen gaits and the gaits (variables). Both on-land and underwater gait data were compared for the validation of the result. Magnetic field (Mag-x, Mag-y & Mag-z) and the knee angle were identified as dominating variables in a gait cycle.

As described in the chapter 4, the python GUI was designed with the intention of giving full access to the user to make choices and decisions within a permittable range. The object-oriented approach and the user defined functions used in the code allows a developer to integrate modifications to the functionality in the future. Also, while the computing is done in python the framework can be changed to '.Net' or a suitable to integrate the software to windows platform as a future addition.

The future development of the study is to apply the dynamic mode decomposition to the approximated dataset and create an interpretation based on the results.

# KOKKUVÕTE

Kõnnaku andmete analüüsimiseks on selles uuringus välja pakutud singulaarväärtuse dekompositsiooni madala astme lähendamismudel. Uuringu eesmärk oli tuvastada IMU kõrgemõõtmeliste andurite andmete domineerivad mustrid madala mõõtmega lähendusega. Mudel on integreeritud Python-põhisesse GUI-sse, millel on erinevad funktsioonid, mida on kirjeldatud 4. peatükis. GUI töötab Tkinteri raamistikul, mis pakub paindlikkust ja kasutajasõbralikku keskkonda. Selle töö jaoks kogutud andmekogum oli ainulaadne ja adekvaatne mudeli teostatavuse hindamiseks.

Nagu projekti eesmärkides (peatükk 2.4) mainitud, oli esimene eesmärk rakendada SVD-d analüüsis andmepõhise mudelina. SVD-d rakendati edukalt erinevatest tingimustest koosnevatele erinevatele andmekogudele ja tuvastati lagunenud maatriksite tõlgendamine, nagu on näidatud peatükis 4.4. Eigeni kõnnaku kontseptsioon tunnistati U-maatriksi aluseks ja V ’-maatriksis olevad režiimid tuvastati iga eigeni kõnnaku kaaluna. Tulevase tööna määratleti V ’režiimide üksikasjalik analüüs.

Teine eesmärk oli hinnata kõnnaku andmete edukaks lähendamiseks mõeldud mudelit erinevates tingimustes. Nagu on näidatud 5. peatükis, on maakatsetel saadud kõnnaku andmed kõige paremini kirjeldatud 4. järgu lähendusega, mis haarab rohkem kui 99% energiast. Kuigi mõned maapealsed tulemused vajasid 99% energia hõivamiseks 5. järgu lähendust. Isegi siis haarab 4. järgu lähendus rohkem kui 90% energiast, mis on lähenduse kohta ideaalne. Veealuste katsete tulemused näitavad, et 4. järgu lähendus on parim mõõtmete vähendamine, mida on võimalik saavutada eesmärgiga haarata üle 99% energiast. Vaadeldud ligikaudsed mustrid näitavad ka seda, et madala astme lähendamine efektiivne mitte ainult järk-järgult suurenevate või vähenevate signaalide korral, vaid ka hüppekaja korral. Peatükis 5.4 esitatud arvutatud RMSE ja R andmete põhjal saab selgelt kindlaks teha, et magnetvoog-Z näitab parimat lineaarset käitumist nii maismaakatsete kui ka veealuste katsete korral, ligikaudse rea korral on korrelatsioonikordaja 0,999. Maakatsed näitavad ka seda, et põlvenurga lähendamine on samuti täiesti lineaarne R-väärtusega 0,999. Kõige täpsemat lähendamist maismaakõnnakule täheldati kvaternioni andmetest minimaalse vea põhjal. Suurim erinevus maapealsete ja veealuste katsete vahel on rõhu lähendamine. Veealused uuringud annavad rõhu suure täpsuse ja lineaarsusega, samal ajal kui maapealsed katsed näitavad mittelineaarsust enamikus madala astme lähendustes.

Madalama astme hinnangu teostatavuse uurimiseks normaliseeriti nii pinna- kui ka veealused andmed. Leiti, et normaliseeritud kõnnakuandmed vajavad 99-protsendilise energia hõivamiseks 10. järgu lähendamist. Erinevalt toorandmetest jätab 3. järgu lähendus ruutkeskmise vea keskmiselt 0,4 kõigis andmekogumites, võrreldes 0.02 10. järgu lähenduses. Korrelatsioonikordaja näitab, et lähendatud väärtus on keskmiselt üle 80% lineaarne. Normaliseeritud andmekogumite SVD tulemused näitavad ka singulaarväärtuste peaaegu võrdset jaotust lagunemises, mis näitab võrdselt domineerivaid eigeni kõnnakuid.See viitab sellele, et andmete tihendamise tõttu on normaliseeritud kõnnakujaotuses domineerivaid mustreid raske ära tunda. Edasiseks uurimiseks on normaliseeritud andmete SVD edasine analüüs avatud.

Selle uuringu kolmas ja neljas eesmärk oli kasutada Pythoni skripti, et tuvastada domineerivad kombinatsioonid Pythoni skripti abil ja töötada välja kasutajale täielik juurdepääs võimaldav GUI. Domineerivad muutujad tuvastati pütooni koodi genereeritud kärbitud V ’režiimide graafiku (peatükk 5.5) jälgimise teel. Identifitseerimine põhines iga eigeni kõnnaku ja liikumise (muutujate) vaheliste korrelatsioonide arvul. Tulemuse kinnitamiseks võrreldi nii pinna- kui ka veealuse kõndimise andmeid. Magnetväli (Mag-x, Mag-y & Mag-z) ja põlvenurk tuvastati kõnnakutsükli domineerivate muutujatena.

Nagu on kirjeldatud 4. peatükis, kujundati pythoni graafiline kasutajaliides eesmärgiga anda kasutajale täielik juurdepääs valikute ja otsuste tegemiseks lubatavas vahemikus. Objektorienteeritud lähenemine ja koodis kasutatud kasutaja määratletud funktsioonid võimaldavad arendajal integreerida tulevikus funktsionaalsuse muudatusi. Kuigi arvutamine toimub pythonis, saab raamistiku muuta. Net-iks või sobivaks tarkvara integreerimiseks Windowsi platvormi tulevase täiendusena.

Uuringu edasine areng on dünaamilise režiimi lagundamise rakendamine ligikaudsele andmekogumile ja tulemuste põhjal tõlgenduse loomine.

# REFERENCES


[1] L. Havasi, Z. Szlávik, and T. Szirányi, "Detection of gait characteristics for scene registration in video surveillance system," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 503–510, 2007, doi: 10.1109/TIP.2006.888339.

[2] L. Lee and W. E. L. Grimson, "Gait analysis for recognition and classification," *Proc. - 5th IEEE Int. Conf. Autom. Face Gesture Recognition, FGR 2002*, no. Mld, pp. 155–162, 2002, doi: 10.1109/AFGR.2002.1004148.

[3] W. Pirker and R. Katzenschlager, "Gait disorders in adults and the elderly: A clinical guide," *Wien. Klin. Wochenschr.*, vol. 129, no. 3–4, pp. 81–95, 2017, doi: 10.1007/s00508-016-1096-4.

[4] D. Xu, S. Yan, D. Tao, L. Zhang, X. Li, and H. J. Zhang, "Human gait recognition with matrix representation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 7, pp. 896–903, 2006, doi: 10.1109/TCSVT.2006.877418.

[5] T. Liu, B. Sun, M. Chi, and X. Zeng, "Gender recognition using dynamic gait energy image," *Proc. 2017 IEEE 2nd Inf. Technol. Networking, Electron. Autom. Control Conf. ITNEC 2017*, vol. 2018-Janua, pp. 1078–1081, 2018, doi: 10.1109/ITNEC.2017.8284905.

[6] A. Badiezadeh, F. Ayatollahi, M. H. Ghaeminia, and S. B. Shokouhi, "Human gait recognition using Dual-Tree Complex Wavelet Transform," *2017 25th Iran. Conf. Electr. Eng. ICEE 2017*, no. lCEE20 17, pp. 461–466, 2017, doi: 10.1109/IranianCEE.2017.7985495.

[7] S. Chen, J. Lach, B. Lo, and G. Z. Yang, "Toward Pervasive Gait Analysis With Wearable Sensors: A Systematic Review," *IEEE J. Biomed. Heal. Informatics*, vol. 20, no. 6, pp. 1521–1537, 2016, doi: 10.1109/JBHI.2016.2608720.

[8] H. Zhao, Z. Wang, S. Qiu, Y. Shen, and J. Wang, "IMU-based gait analysis for rehabilitation assessment of patients with gait disorders," *2017 4th Int. Conf. Syst. Informatics, ICSAI 2017*, vol. 2018-Janua, no. 61473058, pp. 622–626, 2017, doi: 10.1109/ICSAI.2017.8248364.

[9] R. G. Birdal, A. Sertbas, and B. Mihendisligi, "Human Identification Based on Gait

Analysis: A survey," *UBMK 2018 - 3rd Int. Conf. Comput. Sci. Eng.*, pp. 489–493, 2018, doi: 10.1109/UBMK.2018.8566368.

[10] D. Tarniţă, "Wearable sensors used for human gait analysis," *Rom. J. Morphol. Embryol.*, vol. 57, no. 2, pp. 373–382, 2016.

[11] D. T. P. Fong and Y. Y. Chan, "The use of wearable inertial motion sensors in human lower limb biomechanics studies: A systematic review," *Sensors (Switzerland)*, vol. 10, no. 12, pp. 11556–11565, 2010, doi: 10.3390/s101211556.

[12] M. S. Baby, A. J. Saji, and C. S. Kumar, "Parkinsons disease classification using wavelet transform based feature extraction of gait data," *Proc. IEEE Int. Conf. Circuit, Power Comput. Technol. ICCPCT 2017*, pp. 1–6, 2017, doi: 10.1109/ICCPCT.2017.8074230.

[13] C. Monoli, J. F. Fuentez-Perez, N. Cau, P. Capodaglio, M. Galli, and J. A. Tuhtan, "Land and Underwater Gait Analysis Using Wearable IMU," *IEEE Sens. J.*, vol. XX, no. Xx, pp. 1–11, 2021, doi: 10.1109/JSEN.2021.3061623.

[14] M. H. Richardson, "Fundamentals of the discrete Fourier transform," *S V Sound Vib.*, vol. 12, no. 3, pp. 40–46, 1978.

[15] A. Mostayed, M. Mynuddin, G. Mazumder, S. Kim, and S. J. Park, "Abnormal gait detection using discrete fourier transform," *Proc. - 2008 Int. Conf. Multimed. Ubiquitous Eng. MUE 2008*, pp. 36–40, 2008, doi: 10.1109/MUE.2008.59.

[16] S. Yu, L. Wang, W. Hu, and T. Tan, "Gait analysis for human identification in frequency domain," *Proc. - Third Int. Conf. Image Graph.*, pp. 282–285, 2004, doi: 10.1109/icig.2004.72.

[17] S. Jeba Priya, A. J. Rani, and S. Su Ma, "Diagnosis of Parkinson's Disease using Fast Fourier Transform," *ICDCS 2020 - 2020 5th Int. Conf. Devices, Circuits Syst.*, pp. 198–202, 2020, doi: 10.1109/ICDCS48716.2020.243580.

[18] D. Han, V. Renaudin, and M. Ortiz, "Smartphone based gait analysis using STFT and wavelet transform for indoor navigation," *IPIN 2014 - 2014 Int. Conf. Indoor Position. Indoor Navig.*, pp. 157–166, 2014, doi: 10.1109/IPIN.2014.7275480.

[19] S. Wen, F. Wang, C. Wu, and Y. Zhang, "Gait data de-noising based on improved

EMD," *2010 Chinese Control Decis. Conf. CCDC 2010*, pp. 2766–2770, 2010, doi: 10.1109/CCDC.2010.5498727.

[20] L. Li, L. Atallah, B. Lo, and G. Z. Yang, "Feature extraction from ear-worn sensor data for gait analysis," *2014 IEEE-EMBS Int. Conf. Biomed. Heal. Informatics, BHI 2014*, pp. 560–563, 2014, doi: 10.1109/BHI.2014.6864426.

[21] P. Ren *et al.*, "Gait Rhythm Fluctuation Analysis for Neurodegenerative Diseases by Empirical Mode Decomposition," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 1, pp. 52–60, 2017, doi: 10.1109/TBME.2016.2536438.

[22] N. B. Erichson and C. Donovan, "Randomized low-rank Dynamic Mode Decomposition for motion detection," *Comput. Vis. Image Underst.*, vol. 146, pp. 40–50, 2016, doi: 10.1016/j.cviu.2016.02.005.

[23] Z. Dang, Y. Lv, Y. Li, and G. Wei, "Improved dynamic mode decomposition and its application to fault diagnosis of rolling bearing," *Sensors (Switzerland)*, vol. 18, no. 6, 2018, doi: 10.3390/s18061972.

[24] T. Runolfsson, "Output Dynamic Mode Decomposition: An Extension of Dynamic Mode Decomposition Based on Output Functional Expansions," *Proc. IEEE Conf. Decis. Control*, vol. 2018-Decem, no. Cdc, pp. 7148–7152, 2019, doi: 10.1109/CDC.2018.8619601.

[25] U. Rl, "I n e r ti al s e n s o r-b a s e d k n e e flexio n / ex t e n sio n a n gl e e s ti m a tio n," 2009.

[26] J. Wang, E. B. Garcia, S. Yu, and D. Zhang, "Windowed DMD for Gait Recognition Under Clothing and Carrying Condition Variations," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10568 LNCS, no. April 2020, pp. 484–492, 2017, doi: 10.1007/978-3-319-69923-3_52.

[27] M. Bertero and P. Boccacci, "Singular value decomposition (SVD)," *Introd. to Inverse Probl. Imaging*, no. Dmd, pp. 3–46, 2004, doi: 10.1887/0750304359/b461c9.

[28] V. P. Stokes and A. Thorstensson, "from 3-D Kinematic Data," vol. 46, no. 1, pp. 100–106, 1999.

[29] O. Uzhga-Rebrov and G. Kuleshova, "Using Singular Value Decomposition to Reduce

Dimensionality of Initial Data Set," *2020 61st Int. Sci. Conf. Inf. Technol. Manag. Sci. Riga Tech. Univ. ITMS 2020 - Proc.*, pp. 20–23, 2020, doi: 10.1109/ITMS51158.2020.9259304.

[30] Y. Jiang, I. Hayashi, and S. Wang, "Knowledge acquisition method based on singular value decomposition for human motion analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 12, pp. 3038–3050, 2014, doi: 10.1109/TKDE.2014.2316521.

[31] W. Kusakunniran, Q. Wu, H. Li, and J. Zhang, "Multiple views gait recognition using view transformation model based on optimized gait energy image," *2009 IEEE 12th Int. Conf. Comput. Vis. Work. ICCV Work. 2009*, pp. 1058–1064, 2009, doi: 10.1109/ICCVW.2009.5457587.

[32] S. L. Brunton and J. N. Kutz, "Data-Driven Science and Engineering."

[33] A. G. Cutti, A. Ferrari, P. Garofalo, M. Raggi, A. Cappello, and A. Ferrari, "'Outwalk': A protocol for clinical gait analysis based on inertial and magnetic sensors," *Med. Biol. Eng. Comput.*, vol. 48, no. 1, pp. 17–25, 2010, doi: 10.1007/s11517-009-0545-x.

[34] D. A. Winter, *Biomechanics and Motor Control of Human Movement: Fourth Edition*. 2009.

[35] A. Adetokunbo and A. Basirat, "Software Engineering Methodologies: A Review of the Waterfall Model and Object- Oriented Approach," *Int. J. Sci. Eng. Res.*, vol. 4, no. 7, pp. 427–434, 2014.

[36] V. Aggarwal, V. Gupta, P. Singh, K. Sharma, and N. Sharma, "Detection of spatial outlier by using improved Z-score test," *Proc. Int. Conf. Trends Electron. Informatics, ICOEI 2019*, vol. 2019-April, no. Icoei, pp. 788–790, 2019, doi: 10.1109/icoei.2019.8862582.