Detection of linear translation using inertial sensors

Samuel Juul Barker

April 29, 2024

Abstract

Human Activity Recognition (HAR) using inertial sensors is an evolving field with many potential applications in healthcare and beyond. Much of the existing literature pertains to the use of data-driven methods which suffer from difficulties in generalisation and a large demand for labelled data, showing a need for investigation into HAR by statistical measures, for example the generalised likelihood ratio test (GLRT) as described in this document. Prior research exists showing zero velocity detection by this method, as well as detection of rotations about a fixed axis, and this report details the extension of this concept for detecting linear translational motions using an accelerometer unit (such as those in smartphones) without the use of a gyroscope. The methodology shows promise in detecting linear translations in accelerometer data.

I certify that all material in this dissertation which is not my own work has been identified.

Contents

1 Introduction

This project aims to determine the feasibility of detecting periods of linear translational motion in accelerometer data (without the use of a gyroscope) using a generalised likelihood ratio test (GLRT) based optimisation approach, in an effort to perform this classification by statistical means where much of activity classification is performed with data-driven machine learning approaches [1], [2] which suffer from a lack of rigorous theoretical basis and generalisability despite the approaches' effectiveness in an academic context [3].

1.1 Background and motivation

Inertial measurement units (IMUs) have long been the subject of research in many applications, [4] most relevantly human activity recognition (HAR) where IMUs have been employed in, for example, gait recognition, [5]–[7] pedestrian tracking, [8] navigation & positioning [9], [10] and activity recognition for purposes like healthcare, e.g. fall detection, rehabilitation etc [11]–[13].

However, in applications where inertial sensors are used to estimate position, error growth in estimated position is quadratic. This is a result of said estimate being calculated by double-integrating an accelerometer signal (as well as integrating angular velocity from a gyroscope to determine the direction of gravity) as is aptly summarised by Kok et al. [14] and illustrated in Figure 1 The quadratic growth in error due to the presence of both constant bias and random noise in the integrated signal results in position estimates only remaining useful for no more than a few seconds [15].

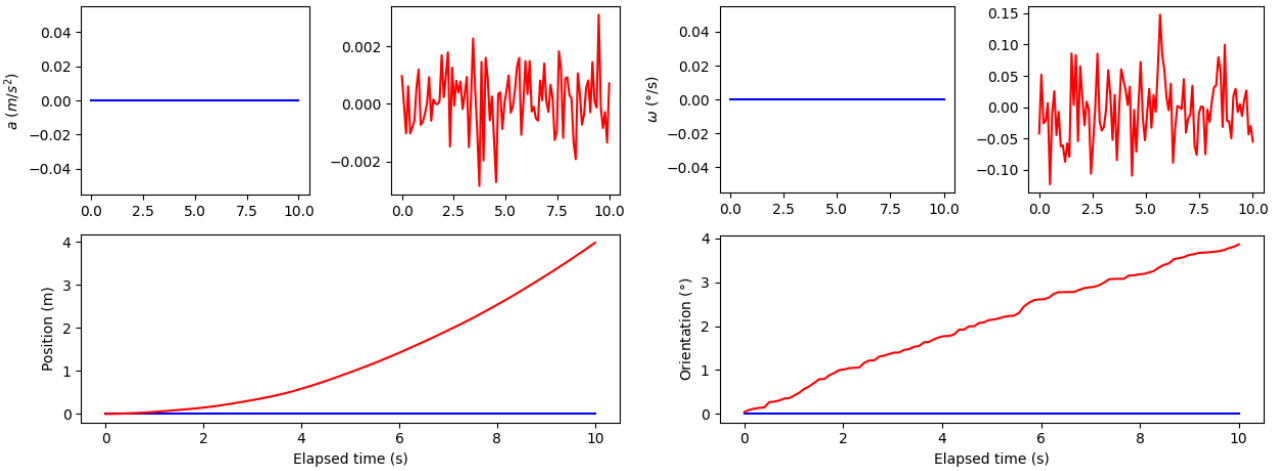

Figure 1: An illustration of quadratic error growth in estimated position (left) and linear error growth in estimated orientation (right). In both subfigures, top-left: ground truth data (zero for a stationary sensor), top-right: Gaussian noise, bottom: integration-derived estimate.

To mitigate this problem, much research has been done into methods of bounding this error growth in position estimates, chief among them being supplementing inertial measurements with external information like GPS position [8]. However, in circumstances where GPS or other such external sources of position information are absent, a commonly employed method of bounding error is zero-velocity updates (ZUPTs), wherein known, external information about the system being measured is used to identify periods where the sensor unit's velocity is assumed to be zero, thereby reducing error growth, one example being the use of ZUPTs in foot-mounted sensor pedestrian tracking by Foxlin et al [16].

Research has been done into the use of generalised-likelihood ratio tests (GLRT) for the detection of rotations about a fixed axis using inertial sensors, [17] but as of yet, this statistically motivated methodology has not been applied to the detection of linear translations – said application is explored in this project. If detection of linear translation by this method is found to be feasible, it could find use not only in correcting error growth in integration-based orientation/position estimation, but also in, for example, gesture recognition for inertial sensor based human-computer interaction (HCI), since many simpler gestures in HCI consist of translational motions.

2 Objectives and requirements

ID	Requirement
Functional requirements	
1	The program must be able to take input of sensor data in standard CSV format.
2	The creation of a functional program implementing statistical optimisation methods for detecting linear translation in Python.
3	The program must output results in an easily understood and insightful format such as tables, graphs and figures to display results in a manner conducive to interpretation.
Non-functional requirements	
4	The program must finish execution in less than 25 minutes on a standard personal computer/laptop.
5	Documentation provided must adequately explain the principles behind the methods utilised such that a non-specialist with appropriate mathematical experience could understand and reproduce the experiment.
6	Documentation provided with the source code must be readable and understandable such that a non-specialist could understand the program's design and execute the program.
7	Detailed evaluation must be done of the analysis approach and its strengths, weaknesses and aspects for potential improvement and further work.

Table 1: Specified requirements for the project.

Being a more research oriented project, the requirements specified previously were devised with the overall objective of simply determining and displaying the results of this approach to detecting linear translation, as opposed to specifying that success must be found in the endeavour. Therefore, this section details an evaluation of the project requirements and documents the extent to which objectives were fulfilled by the final program.

2.1 Evaluation of objectives

ID	Success criteria	Method of evaluation
Functional requirements		
1	The sensor data was read in CSV format by the program and processed without error.	The sections (notebook cells) of the program that read CSV data, and any sections it depends on, will be run, and this objective will be satisfied given the absence of runtime errors in the program in CSV-processing code sections.
2	The program executes without error and accuracy values for both exercises are ≥ 0.65.	The program will be run from a cold start, and this objective will be satisfied if the resultant decision accuracy obtained using the data outlined in subsection 3.2 is ≥ 0.65 for both squats and curls.
3	The program runs without error and all figures correctly render and display correct results.	The program will be run from a cold start, and all figures, tables and graphs in the Jupyter notebook will correctly render and display the results obtained by the execution of the program.
Non-functional requirements		
4	All cells are executes in less than 25 minutes total on the machine described.	The whole program will be executed from a cold start 20 times and the objective will be satisfied if the average execution time is under 25 minutes on a Lenovo Yoga 6 laptop (AMD Ryzen 7 5700U @ 4.372GHz, 16GB RAM.)
5	N/A	N/A

6	Docstrings were written to describe every function defined in the program. Code comments were written in every executable notebook cell	A static analysis program will be used to ensure all functions have a docstring.
7	N/A	N/A

Table 2: Criteria for the satisfaction of each objective, and the method by which each will be evaluated.

Upon re-evaluation of the objectives devised, it was decided that objectives 5 and 7 were too vague to be considered actionable requirements, leading to N/A in Table 2; no concrete evaluation can be done.

3 Design, methods and implementation

3.1 Toolset and approach

The project was carried out using Python in a Jupyter notebook format, for its ease of use and prototyping/exploratory nature of continuous re-evaluation. Essential to the project were the statistical libraries *numpy*, [18] *scipy* [19] and *pandas* [20].

3.2 Experimental accelerometer data

In order to implement and evaluate this statistically motivated approach to detecting linear translations with only accelerometer data, real-world exercise data was recorded and preprocessed. The sensor unit used was the 3-axis accelerometer on a Samsung Galaxy A10 smartphone, fastened to the underside of the subject's wrist during recording, which was performed at a rate of 100Hz. These data were saved and stored in CSV (comma separated value) format, and data recorded were as follows in Table 3.

time	**seconds_elapsed**	**x**	**y**	**z**
Unix epoch time corresponding to the instant of measurement (s)	Seconds since measurement began (s)	Accelerometer reading in x direction (m/s^2)	Accelerometer reading in y direction (m/s^2)	Accelerometer reading in z direction (m/s^2)

Table 3: Measurements taken by the sensor unit.

Exercises were performed using a 7kg dumbbell in each of the subject's hands, and were 8 repetitions squats with pauses between reps, 10 repetitions bicep curls without pauses. These particular motions were selected as typical examples of the types of motion this methodology is intended to discriminate between – translational motion in dumbbell squats, and purely rotational motion in dumbbell curls.

The data collected during these two exercise sessions was then trimmed to remove the parts of the data where the weights were being picked up at the start, and put down at the end, to restrict the domain to relevant data only.

After the measurement sequences were trimmed to the relevant periods, their durations were manually labelled with the appropriate hypothesis, represented as 0 meaning the sensor was not undergoing translational motion ($\mathcal{H}_0$) or 1 meaning the sensor was undergoing translational motion ($\mathcal{H}_1$). In the particular case of the two exercises used in this experiment, the dumbbell curl is an exercise where none of the data was labelled as translational motion since the entire motion consists of rotation about a fixed axis, the elbow, and the motion was done with no pauses. This is illustrated in Figure 3.

On the other hand, the dumbbell squat was labelled as having multiple periods of translational ($\mathcal{H}_1$) and non-translational motion ($\mathcal{H}_0$) across repetitions of the exercise, as the exercise's primary motions (rising up and squatting down) consist of purely translational motion in the arm on which the sensor unit was placed, and the periods between rising and squatting motions were mostly still with small non-translational motion as the body responded to the rapid deceleration. This is illustrated in Figure 4

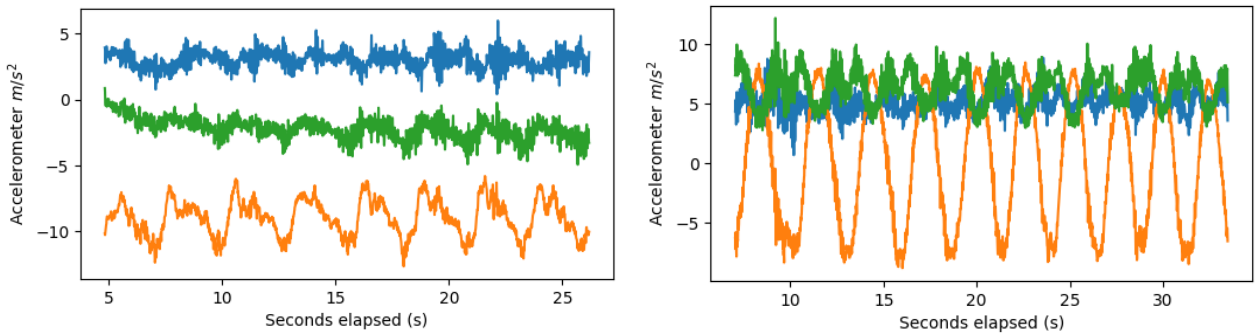

Figure 2: Accelerometer data from the sensor unit on three axes in m/s^2 plotted against seconds elapsed during the exercise. Left: dumbbell squats, right: dumbbell curls.

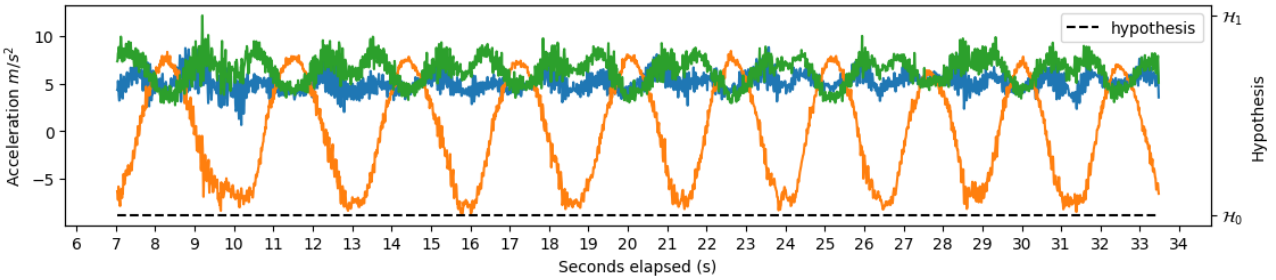

Figure 3: Accelerometer data for dumbbell curls with hypothesis plotted (black dotted line) – the ground truth for the entire measurement sequence is $\mathcal{H}_0$.

3.3 Hypothesis testing approach

As detailed in section 1, this project employs a hypothesis testing-based approach to identifying periods of linear translation in a similar way to Wahlström's work on identifying rotation about a fixed axis, [17] wherein a purely statistical and optimisation-based methodology is employed as opposed to a data-driven, machine learning based one. The task of detecting periods of linear translation is formulated as a decision between two hypotheses as described in Equation 1.

$$\begin{aligned} \mathcal{H}_0: &\quad \text{The sensor unit is not undergoing linear translation.} \\ \mathcal{H}_1: &\quad \text{The sensor unit is undergoing linear translation.} \end{aligned} \tag{1}$$

the result of the generalised-likelihood ratio test (GLRT) being given by

$$L_G(\mathbf{z}_n) = \frac{p(\mathbf{z}_n \mid \hat{\boldsymbol{\theta}}_1, \mathcal{H}_1)}{p(\mathbf{z}_n \mid \hat{\boldsymbol{\theta}}_0, \mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma \tag{2}$$

where $\mathbf{z}_n$ represents an IMU measurement sequence (here, from an accelerometer) and $\hat{\boldsymbol{\theta}}$ represents the unknown parameters. The hypothesis is decided upon by calculating a maximum-likelihood estimate (MLE) and comparing it to a threshold value γ.

Given a measurement sequence $\mathbf{y}_k^a \in \mathbb{R}^3$ in sample k, measurements are modelled as such, following notation from [15] and [17]:

$$\mathbf{y}_k^a = \mathbf{s}_k^a(\boldsymbol{\theta}) + \mathbf{v}_k \tag{3}$$

with $\mathbf{s}_k^a(\boldsymbol{\theta})$ denoting the specific force experienced by the sensor unit and $\mathbf{v}_k$ the noise (assumed to be a temporally uncorrelated Gaussian random variable with zero mean). In cases where the sensor is experiencing translational motion, i.e. under $\mathcal{H}_1$, the sensor signal is modelled as such:

$$\mathcal{H}_1 : \forall k \in \Omega_n,\ \mathbf{s}_k^a(\boldsymbol{\theta}) = g\mathbf{u}_g + c_k\mathbf{u}_c \tag{4}$$

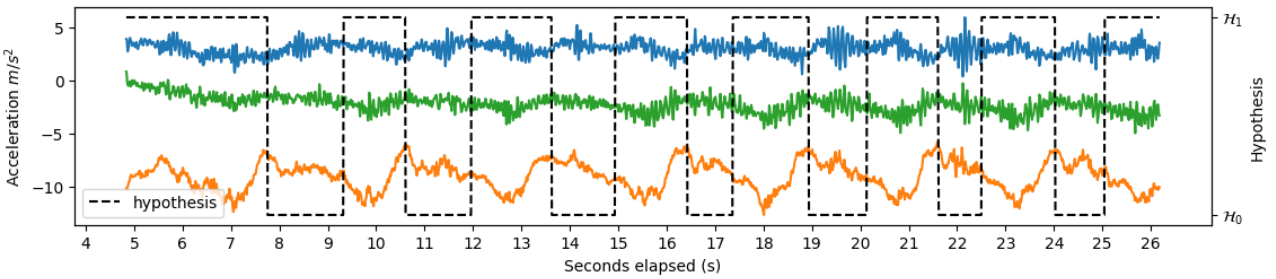

Figure 4: Accelerometer data for dumbbell squats with hypothesis plotted (black dotted line) – the ground truth for the entire measurement sequence consists of periods of translational motion ($\mathcal{H}_1$) and periods of non-translational motion ($\mathcal{H}_0$).

where g and c_k are respectively defined as the average magnitude of Earth's gravity (here 9.81 m/s^2) and the magnitude of residual acceleration over $\Omega_n = \{\ell \in \mathbb{N} : n \leq \ell \leq n+N-1\}$ [15] experienced by the sensor unit, and where $\mathbf{u}_g$ is defined as the unit vector representing the direction of gravity and $\mathbf{u}_c$ is the unit vector representing the direction of acceleration.

With these definitions, the problem of detecting linear translation can be defined as an optimisation problem as follows [21].

$$\begin{aligned}
&\min_{\mathbf{u}_g,\mathbf{u}_c,\{c_k\}_{k+n}^{n=N-1}} \sum_{k\in\Omega_n} ||\mathbf{y}_k^a - (g\mathbf{u}_g + c_k\mathbf{u}_c)|\,|^2 \\
&= \min_{\mathbf{u}_g,\mathbf{u}_c} \sum_{k\in\Omega_n} \frac{\|(\mathbf{y}_k^a - g\mathbf{u}_g)\times\mathbf{u}_c\|^2}{\|\mathbf{u}_c\|^2} \\
&= \min_{\mathbf{u}_g,\mathbf{u}_c} \sum_{k\in\Omega_n} \frac{\mathbf{u}_c^\intercal[(\mathbf{y}_k^a - g\mathbf{u}_g)]_\times^\intercal[(\mathbf{y}_k^a - g\mathbf{u}_g)]_\times\mathbf{u}_c}{\mathbf{u}_c^\intercal\mathbf{u}_c} \\
&= \min_{\mathbf{u}_c,\mathbf{u}_g} \mathcal{R}(\mathbf{M}_n^a(\mathbf{u}_g),\mathbf{u}_c) \\
&= \min_{\mathbf{u}_g} \lambda_{\min}
\end{aligned} \tag{5}$$

Here the problem is formulated as minimising the sum over all measurement windows $k \in \Omega_n$ of the squared differences between the observed measurements $\mathbf{y}_k^a$ and $g\mathbf{u}_g + c_k\mathbf{u}_c$. This minimisation is done by optimising gravity's direction, the direction of acceleration and the magnitude of acceleration. The problem is then reduced mathematically to the Rayleigh quotient, i.e. the minimum eigenvalue, of the matrix $\mathbf{M}_n^a$ and $\mathbf{u}_c$, where, extending notation from [17]

$$\mathbf{M}_n^a = \sum_{k\in\Omega_n} [(\mathbf{y}_k^a - g\mathbf{u}_g)]_\times^\intercal[(\mathbf{y}_k^a - g\mathbf{u}_g)]_\times \tag{6}$$

and $[\mathbf{a}]_\times$ is the skew-symmetric matrix such that $[\mathbf{a}]_\times \times \mathbf{b}$ is the cross product $\mathbf{a}\times\mathbf{b}$, and $\mathcal{R}(\mathbf{M},\mathbf{a}) = \mathbf{a}^\top\mathbf{M}\mathbf{a}/(\mathbf{a}^\top\mathbf{a})$ is equal to the Rayleigh quotient of $\mathbf{M}$, a matrix, and $\mathbf{a}$, a non-zero vector [17].

By this method the optimisation problem is reduced to a 3 dimensional one, the 3 variables being the x, y and z directions of $\mathbf{u}_g$, which is constrained in the optimisation to remain a unit vector since it represents a direction. The minimum minimum eigenvalue of $\mathbf{M}_n^a$, $\lambda_{\min}$, serves as the test statistic (the MLE) for the hypothesis testing process, and will be compared against some threshold value $\gamma \in \Gamma$ to reject or maintain $\mathcal{H}_0$, thereby categorising the motion as translational or non-translational, such that decisions best reflect the manual labels in the data. Again extending notation from [17], [21] the GLRT can be represented as in Equation 7.

$$L_G(\mathbf{y}_k^a,\ldots,\mathbf{y}_{k+n}^a) = \lambda_{\min} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma. \tag{7}$$

4 Development and results

This section encompasses both the process of developing the program and a showcase of the program's preliminary results, since the iterative nature of development using Jupyter notebook led to the continuous refinement of the program, and gradual accumulation of results stage by stage.

4.1 Preparation of experimental data

As described in subsection 3.2, the accelerometer data was recorded using a Samsung Galaxy A10's inbuilt accelerometer unit at a rate of 100Hz, and the data gathered from performing 8 repetitions of dumbbell squats, and 10 repetitions of dumbbell curls. The data were then saved in CSV format, trimmed to relevant periods, loaded into the Jupyter notebook, and manually labelled as seen in Figure 3 and Figure 4.

4.2 Computing the min. min. eigenvalue MLEs

After the data was loaded and prepared, it was initially separated into sliding windows each 100 measurements in size (equivalent to one whole second at 100Hz) and 10 measurements apart (henceforth "with a stride of 10.") These parameters were chosen as a starting point to discern the robustness of the process as a whole, as well as hopefully provide some insight into the objective of the project, that is, determining the feasibility of detecting linear translation by the methodology described in section 3. Later, the process was repeated with different window size and strides, and categorisation performance was compared between them.

To begin, functions were defined for creating the skew-symmetric matrices whose sum was $\mathbf{M}_n^a$ as described in Equation 6, calculating the minimum eigenvalue of $\mathbf{M}_n^a$, and optimising on $\mathbf{u}_g$ as described in Equation 5. This optimisation was done using *scipy* [19] using the constrained trust region algorithm (*"trust-constr"*) provided by *scipy*, where the parameters (the three elements of the unit vector $\mathbf{u}_g$) are constrained to ensure $\mathbf{u}_g$ remained a unit vector. This method was used as it was the only method in *scipy*'s optimisation library that could utilise the experimental data while constraining $\mathbf{u}_g$.

Once the test statistics were calculated for the 100 measurement, 10 stride window sets for both exercises, they were plotted as a line graph, as seen in Figure 5.

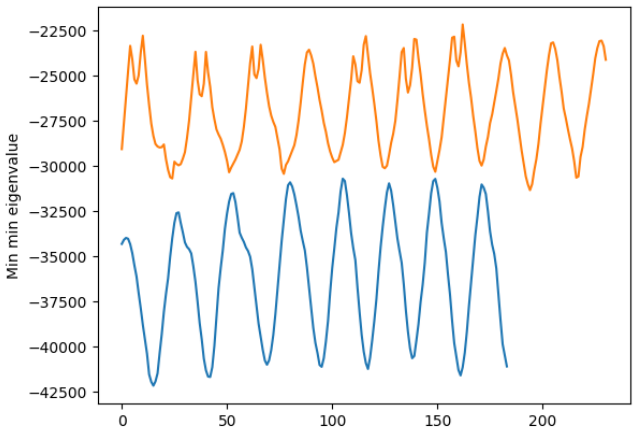

Figure 5: A line plot of the test statistics (minimum minimum eigenvalues of $\mathbf{M}_n^a$) for both exercises; curls in orange (above), squats in blue (below.)

Immediately, these test statistic plots show an evident separation between the two sets of values for squats and curls, which shows promise in telling the two types of motion apart.

4.3 Hypothesis testing

With the test statistics calculated, a simple grid-search method was employed in order to decide on a threshold γ that would best identify periods of translational motion. Given the easily visibly

determined numerical ranges that the test statistics lie within, for this trial of 100 measurement, 10 stride window sets, the upper and lower bounds for γ ranges to be searched were set as the maximum and minimum of the test statistics for each exercise, as beyond either boundary all test statistics would be either greater or less than the threshold, leading to the same decision for all measurements no matter what – much like overfitting in machine learning approaches. Given the relatively low computational intensity of this statistical method as opposed to more intensive machine learning categorisation, the step value was taken to be 1, i.e. it was viable to grid search a number of γ values in the order of tens of thousands, and it was deemed not worth searching with a lower step size.

The "accuracy" of classifications (i.e. the proportion of decisions either way that were correct) was first calculated for each exercise over the threshold space. As seen in Figure 6, the accuracy for each exercise reflects the fact that squats have periods of translational and non-translational motion, whereas curls have only the latter. How this phenomenon in accuracy curves reflects the nature of the two different exercises is discussed further in section 5.

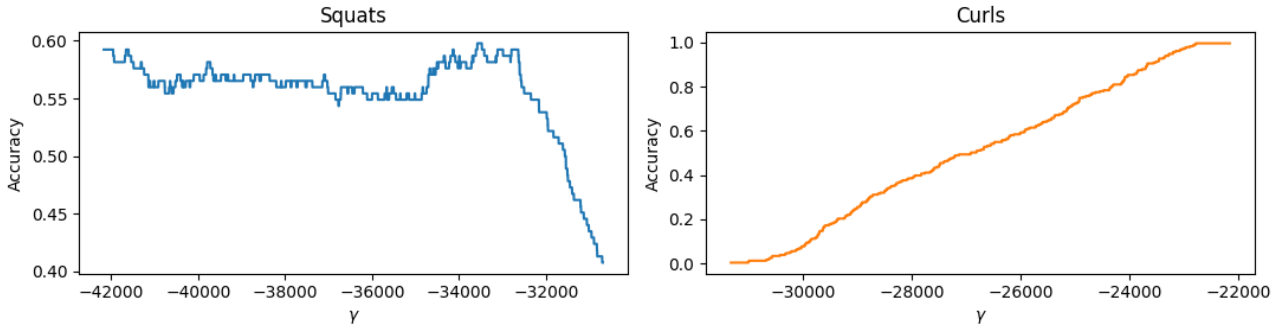

Figure 6: Line graphs of accuracy against the threshold γ for squats and curls.

With these statistics calculated and plotted for each value of γ, it is pertinent to analyse the actual classifications obtained from which the statistics were calculated. Figure 7 shows the hypotheses decisions that yielded the highest accuracy (of ∼0.598) for the dumbbell squat exercise. It can be observed in Figure 7 that the predicted hypotheses follow a pattern of repetitions reflecting the exercise itself, and periods of transitional motion (where $\mathcal{H}_1$ is true) appear to be predicted successfully with a time latency of ∼0.5-1.0 seconds on the transition to deciding on $\mathcal{H}_1$, but not the other way. This indicates that the method shows promise; see more discussion in section 5.

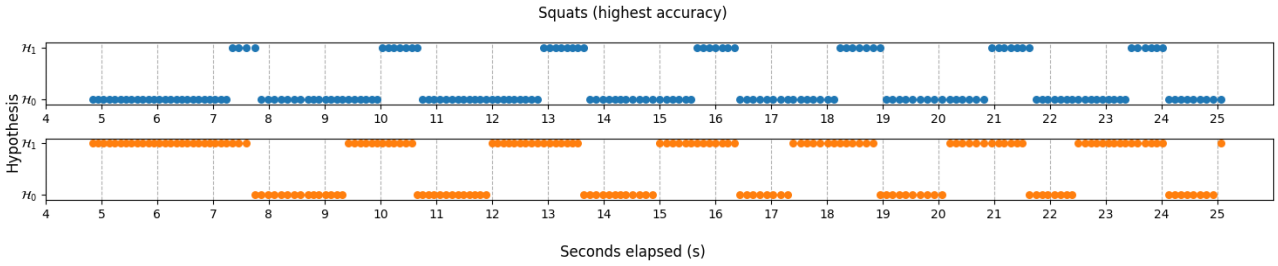

Figure 7: A plot of the set of hypothesis decisions that yielded the highest accuracy (∼0.598) for the dumbbell squat exercise. Top: predicted hypothesis, bottom: true hypothesis (manually labelled.) Hypothesis testing GLRT method (top) shows approximation of true hypotheses.

For dumbbell curls, the highest-accuracy prediction set, shown in Figure 8, was as one might predict; practically all decisions were $\mathcal{H}_0$. This is because the ideal γ for an only-$\mathcal{H}_0$ measurement set is $\leq$ the lowest test statistic (the lowest value of λ_{min}) meaning a threshold derived from motions of this kind without any periods of translational motion whatsoever may not generalise; again see discussion in section 5.

With the results of hypothesis testing calculated for all values of γ in the specified range, the binary nature of this hypothesis testing methodology lead to the application of true positive, false positive, true positive and true negative rates to this approach, where these metrics are interpreted as in Table 4.

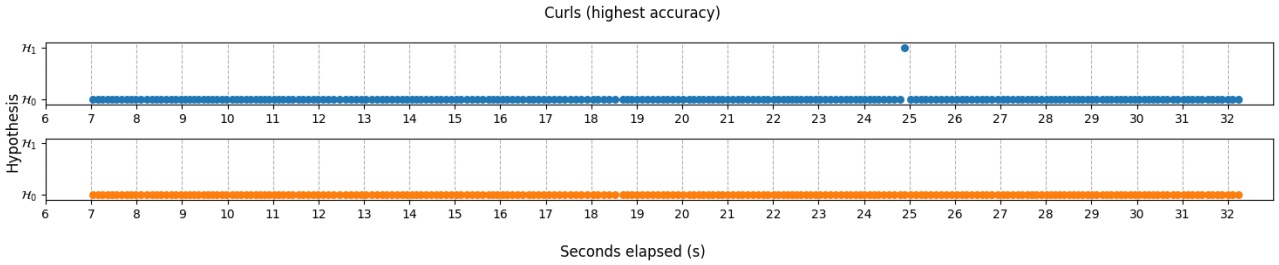

Figure 8: A plot of the set of hypothesis decisions that yielded the highest accuracy (∼0.996) for the dumbbell squat exercise. Top: predicted hypothesis, bottom: true hypothesis (manually labelled.) Expectedly, hypothesis decisions are almost always $\mathcal{H}_0$.

Metric	True hypothesis	Hypothesis decided upon
True positive	$\mathcal{H}_1$	$\mathcal{H}_1$
False positive	$\mathcal{H}_1$	$\mathcal{H}_0$
True negative	$\mathcal{H}_0$	$\mathcal{H}_0$
False negative	$\mathcal{H}_0$	$\mathcal{H}_1$

Table 4: Representation of common binary classification metrics as applied to this hypothesis testing method.

These metrics were calculated and plotted as line graphs for each exercise, as illustrated in Figure 9. In the case of dumbbell squats, metrics follow relatively linear progressions as the threshold value γ changes and measurement windows are classified differently. However, in the case of curls, wherein the true hypothesis is always $\mathcal{H}_0$, there can be no "true positives" as there is no instance where $\mathcal{H}_1$ is the correct hypothesis to predict.

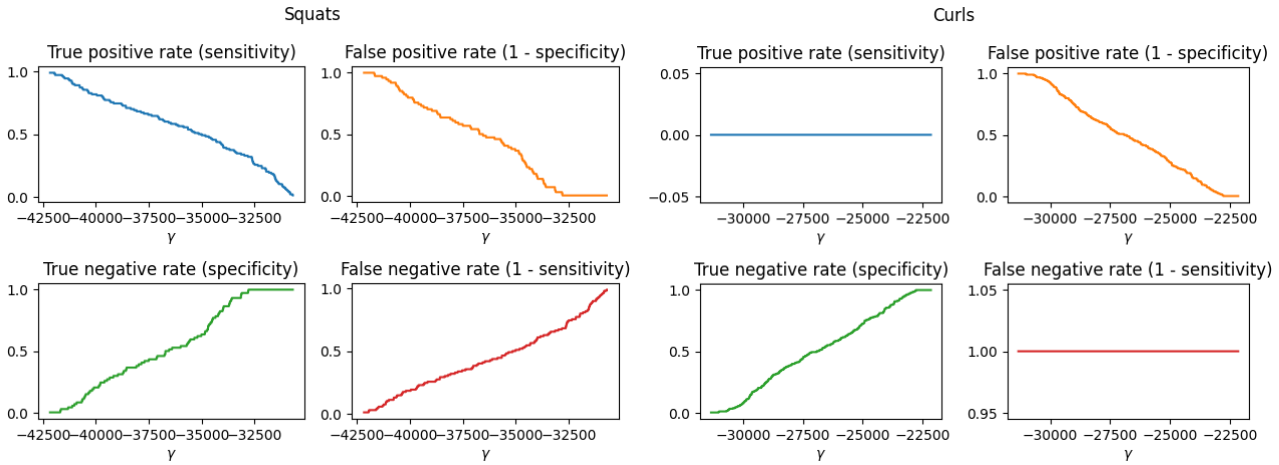

Figure 9: True positive, false positive, true negative and false negative rates obtained by grid searching γ and the hypothesis testing process at each value of γ.

To examine the effects on the hypothesis testing process of selecting for values of γ that minimise or maximise the aforementioned four statistics, the hypotheses reflecting these extrema were plotted.

Figure 10: Hypothesis plots for the highest and lowest values for the four primary classification metrics, calculated for dumbbell squats.

In the case of dumbbell squats, Figure 10 shows the decisions that lead to the highest true negative rate TNR_{max} and lowest false positive rate FPR_{min} are both identical to each other, and very similar to the decisions that lead to the highest accuracy (see Figure 7.) Also identical are the highest true positive TPR_{max}, highest false positive FPR_{max}, lowest true negative TNR_{min} and lowest false negative FNR_{min} rates, the decisions leading to these extrema being almost all $\mathcal{H}_1$, meaning the γ that lead to the extrema of these metrics was equal to the upper bound of the test statistics. Conversely, the identical decisions for the lowest true positive TPR_{min} and highest false negative FNR_{max} rates imply that γ was equal to the lower bound of the test statistics for these decisions.

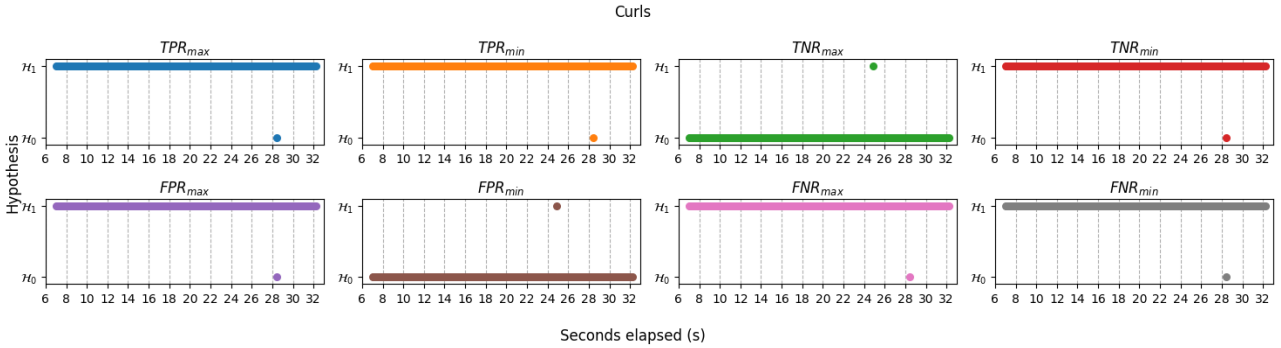

Figure 11: Hypothesis plots for the highest and lowest values for the four primary classification metrics, calculated for dumbbell curls.

An equivalent plot for curls, however, as illustrated in Figure 11, shows little variation in the decisions made that lead to extrema in each metric – there are two groups of identical decision sets, those being the group of TPR_{max}, TPR_{min}, TNR_{min}, FPR_{max}, FNR_{max} and FNR_{min}, and the pair TNR_{max} and FPR_{min}. γ was equal to the highest test statistic value in the first group and the lowest test statistic value in the second.

How these hypothesis plots and their implications for the use of these metrics in this project will be discussed in section 5, along with the extent to which these provide additional insight into the problem at hand of detecting linear translation.

4.4 Comparison between window sizes and strides

After implementing the methodology as a program and evaluating the results with a window size of 100 measurements (1 second at 100Hz) and a stride of 10, the same methodology was applied onto data separated into windows 50 measurements in size (half a second at 100Hz) with a stride of 10, again for data separated into windows 150 measurements in size (1.5 seconds at 100Hz) with a stride of 10, and the same again, with a stride of 25, to measure the effect stride had on the effectiveness of the method. This report shows only the results of classification on the squats exercise, as the results on curls showed no difference between each other, and therefore were not included.

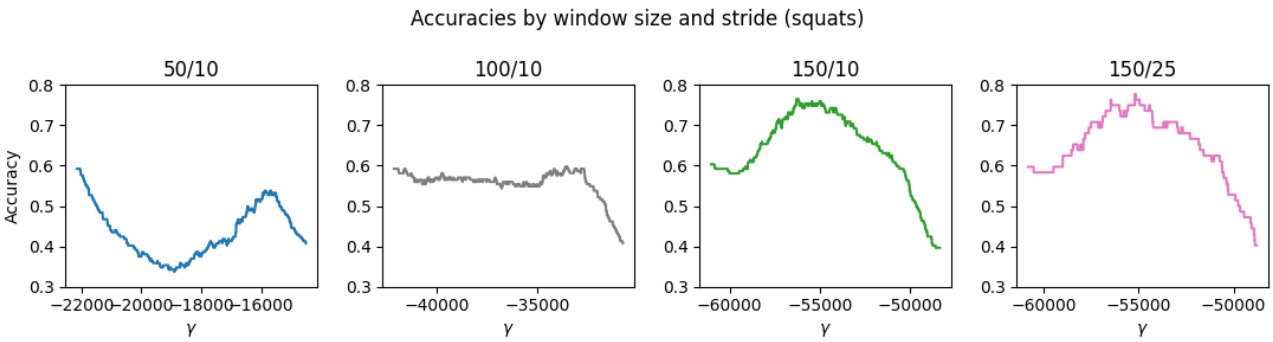

Figure 12: Line plots of accuracy against γ in the four window size and stride variations – titles indicate window size and stride separated by a forward slash.

As seen in Figure 12, the accuracy profile was found to vary rather noticeably by window size, but the minor variation in stride did not make a noticeable difference in accuracy. Due to the relatively short length of the exercise sequences recorded, higher stride values were not experimented with, and due to time constraints, lower stride values were not experimented with. See section 5 for discussion.

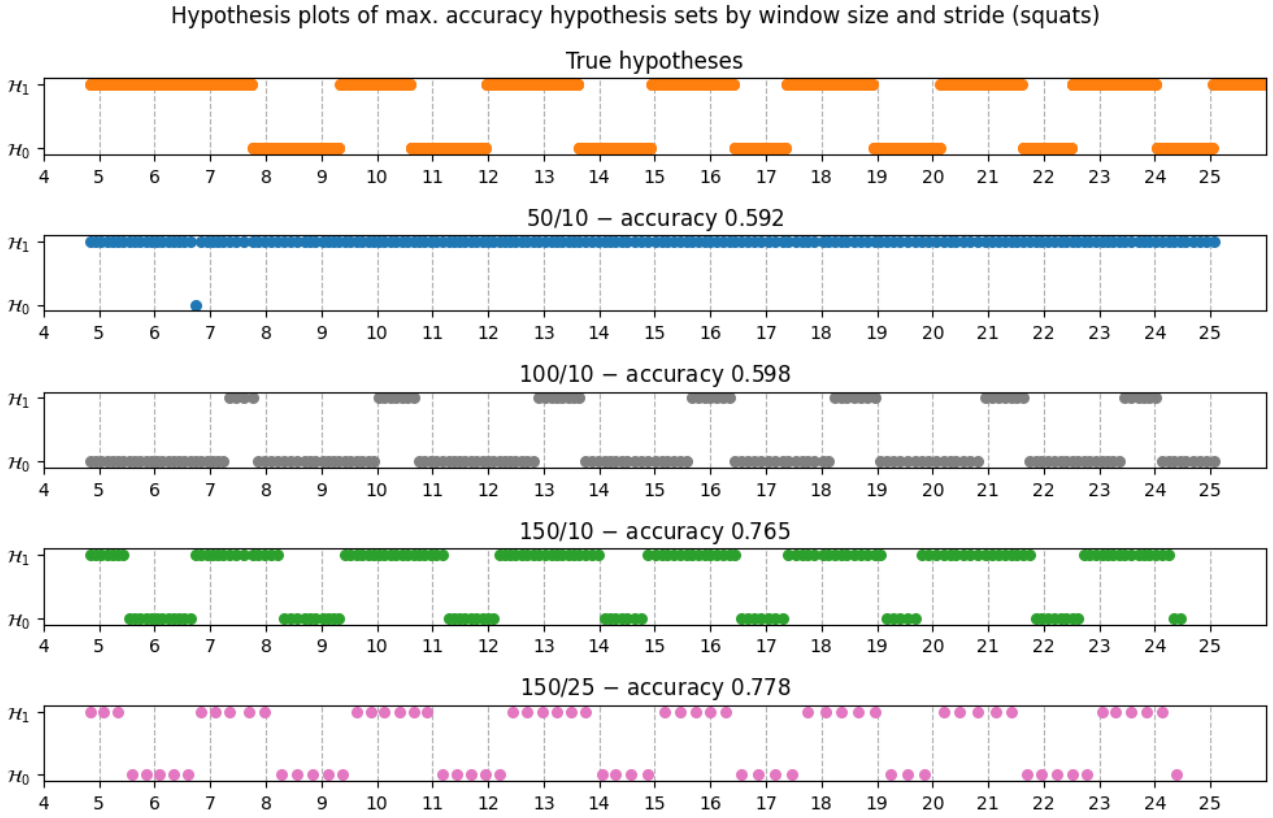

Figure 13: Hypothesis plots for the true hypotheses (manually labelled) and the maximum accuracy decision sets for each window size and stride variation.

With accuracies profiles plotted for each variation in window size and stride, it was deemed useful to plot the hypotheses that lead to the highest accuracy in each to determine the closeness to the true hypotheses achieved by the most accurate decision set for each variation. This plot is shown in Figure 13. The sparser hypothesis in the fifth plot in this figure reflect the increased stride and therefore the lower number of hypothesis data points present in that decision set.

Additionally, in order to further compare the classification performance for the different values of window size and stride, a receiver operating characteristic (ROC) curve (shown in Figure 14) was plotted showing true positive rate against false positive rate for each variation, along with a $y = x$ line

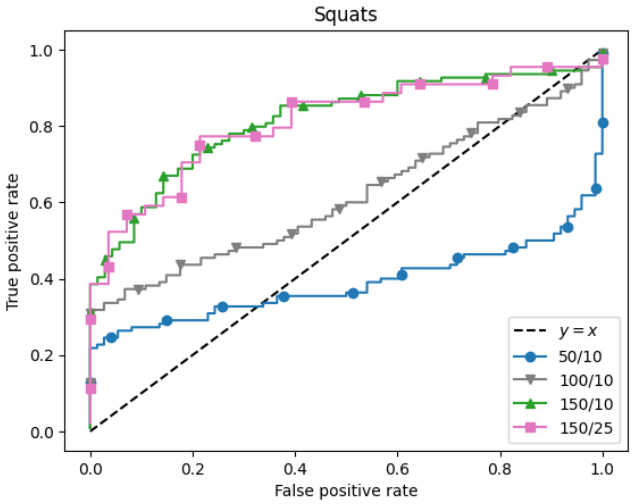

Figure 14: The receiver operating characteristic (ROC) curve plotting true positive rate against false positive rate for the results of each window size and stride variation.

representing random guessing.

5 Discussion

This section discusses the results of the project from a research standpoint, evaluating the extent to which it managed to provide successful insight into the problem at hand, the extent to which any insights gleaned or methods used were novel, a self-evaluation from a research standpoint and a discussion of potential future directions of investigation and research.

5.1 Experimental data

The widely used [22] method by which the accelerometer data was collected, that is, using a smartphone, proved suitable for the project due to the ready availability and ubiquitous ownership of smartphones, as well as the acceptable sampling rate of the unit itself at 100Hz. The fact that the particular unit, the Samsung Galaxy A10, does not have a gyroscope was not an issue as gyroscope measurements were not needed. The physical form factor of the smartphone also posed no hindrance in the actual recording of data via exercise, as it was fastened to the forearm tightly. Although the smartphone's weight at about 170 grams (excluding strap) is believed not to have a noticeable impact on the exercise and therefore the sensor data, a lighter IMU-bearing device such as a smartwatch or bespoke device may have reduced any mass-based interference with motions performed, enhancing the quality of the data further.

5.2 Detection methodology and results

Although taking a statistical approach to detecting types of motion using generalised-likelihood ratio tests (GLRTs), most notably deriving the minimum minimum eigenvalue of quantities derived from the IMU signal is not in itself novel, being in this case extended from work done in [17], [21], no existing research was found applying any such method to the detection of linear translations, and in this capacity the project is believed to be novel.

A major advantage of this method, as described in subsection 3.3, is the far lower computational power required to perform classification, since no large neural network or other machine learning model must be trained on large quantities of pre-labelled input to yield a result. This lack of a requirement for large quantities of training data also lead to a far more rapid and iterative development process for the program implementing this approach, both in terms of time saved on recording and labelling data, but also on time spent training a machine learning model, an often computationally intensive

task. On the data which was separated into 100-measurement windows, 10 measurements apart each, the calculation of test statistics as described in Equation 5 (the primary computational load in the program) took an average of 26.7s (over 10 runs) for squats, and 67 for curls, performed on the device described in Table 2. This keeps the total execution time of the program well under the target 25 minutes, and again exemplifies the relative computational efficiency of this technique.

Displayed in Figure 5 is the two exercises' test statistics plotted as line graphs, showing as being clearly separate with very little overlap in their ranges. The fact that an exercise composed of periods of both translational motion and non-translational motion is so clearly separate from an exercise without any translational motion shows promise for this methodology, however, this itself implies that any γ threshold on, in this example, squats, cannot be re-used to classify on, in this example, curls. This undermines the aim for this hypothesis-testing based method to better generalise between people and exercises than contemporary data-driven methods. One potential avenue for further investigation is normalising accelerometer data to remove variation caused by the difference in acceleration magnitude between exercises.

When computing and plotting the hypotheses that lead to the highest categorisation accuracy, those found to maximise accuracy for curls (at around $\sim$0.996) showed the expected result of deciding on $\mathcal{H}_0$, as illustrated in Figure 8. This showed that on a period of motion with no translational motion, the highest accuracy is gained when $\gamma \leq$ the lowest value of λ_{min}. Squats, however, create a more interesting case, as the hypothesis decisions that lead to the highest accuracy (about $\sim$0.598) showed a marked meaningful pattern of classification closely following the true hypotheses as shown in Figure 7. Restricting the threshold space Γ to between the minimum and maximum values of λ_{min} was key in achieving this, as including values at or past these extrema of λ_{min} lead to the maximum-accuracy decision set simply categorising all measurement windows as the mode hypothesis, that is, the one that was correct for the most time instants. This simple constraint method shows promise in creating meaningful classifications that avoid the optimisation for naïve guesswork analogous to overfitting in data-driven methods. The accuracy achieved for curls ($\sim$0.996) succeeded in surpassing the goal accuracy of 0.65, however, that of squats($\sim$0.598) did not.

Following on from the simpler metric of accuracy, there was also some insight to be drawn from the true and false positive and negative rates for each exercise (see section 4 for notation.) As seen in the aforementioned figure, in the case of squats, only the decision sets for TNR_{max} and FPR_{min} were particularly meaningful classifications, both being identical i.e. one value of γ lead to both the largest true negative rate and the lowest false positive rate, whereas the decision sets for the other extrema all lead to classifications of, effectively, entirely $\mathcal{H}_1$ or $\mathcal{H}_0$ for the entirety of the accelerometer data for the squats. This appears to show that, for the dumbbell squats, insight can be found from paying attention to the true negative and false positive rates. However, further investigation of other exercises/types of motion is needed to determine whether these are the most important metrics in general, or only for this particular case. Examining dumbbell curls, however, again shows little insight, since extrema for all metrics show wholly homogeneous decision sets, either $\mathcal{H}_1$ or $\mathcal{H}_0$ for the entire sequence.

Comparison of performance between classifications done on various window sizes and strides was useful in showing the effect of window size and stride, as well as the methodology in general. In the case of measurement window size 50, i.e. half a second of measurements at 100Hz with a stride of 10, i.e. one tenth of a second distance between measurement windows (henceforth 50/10) the accuracy curve (see Figure 12) was at its peak when the threshold was equal to the lowest value of the test statistic, and all other accuracy values were less than this method of simply labelling every hypothesis as $\mathcal{H}_1$, which was shown to happen in Figure 13. From the poor performance of classification with windows of this size it can be inferred that half a second of measurements in each window was insufficient to perform accurate decision making. The ROC curve (see Figure 14) shows this succinctly; the majority of the curve in the 50/10 case is below the $y = x$ line showing that the classification is actually inferior to random guessing.

In the 100/10 case, accuracy plateaued around the proportion of true hypotheses that were $\mathcal{H}_1$ ($\sim$0.598 – see Figure 12) and then decreased towards the end of the highest γ, which, together with the appropriate line on the ROC curve (Figure 14), shows that a window size of 100 measurements (one second at 100Hz) was barely superior to random guessing. The hypothesis plot (Figure 13), however,

showed that the maximum accuracy, which itself was in fact barely higher than that achieved with the smallest value of γ, produced a fairly meaningful plot that reflected the nature of the motion being performed – periods of linear translation were being classified against periods of non-translational motion.

Furthermore, increasing the window size to 150 measurements, or 1.5 seconds at 100Hz (henceforth 150/10 for a stride of 10 and 150/25 for a stride of 25) showed a further increase in effectiveness, showing accuracy curves (Figure 12) in both the 150/10 and 150/25 cases that the larger measurement window lead to superior detection of periods of linear translation, with a maximum accuracy of $\sim$0.765 for the 150/10 case and $\sim$0.778 for 150/25. The discrepancy between these is believed to be caused by the measurement window set with a stride of 25 encapsulating some portion(s) of the motion data that allowed for superior classification when processed and/or excluding information that diminished accuracy. It is also immediately visible in the hypothesis plots (Figure 13) that the hypotheses predicted with these parameters more closely match the true hypotheses. Additionally, the 150/10 and 150/25 cases also show more promising profiles in their ROC curves (Figure 14), being far closer to the ideal Γ-shaped curve than the 50/10 and 100/10 cases. This shows a general trend of larger measurement windows providing superior classification, however, the full extent to which this increase continues with window size was not investigated; nevertheless, excessively large window sizes would lead to a dramatic increase in the computational intensity of performing detection by this methodology, diminishing one of its advantages in its relative speed.

From these evaluations of these relevant and often-used metrics of classification, and especially from the hypothesis plots drawn from their extrema, evidence points to dumbbell curls being an exercise that was not especially useful or insightful to investigate except insofar as it was useful to disqualify it, despite the initial belief that including a type of motion that contained zero translational motion could provide an interesting case. One remedy to this would be to classify periods of zero motion as translational motion, however, this would lead to squats, a motion consisting of nothing but stillness and translational motion suffering the same problem as curls did in this experiment. A potential further avenue of investigation is formulating zero motion as a third nested hypothesis.

5.3 Potential future direction

Though this project has provided a first step into detecting linear translation by the statistical methods described herein, the results of this project make evident paths for future research into this area. One immediately visible avenue for further work is the application of this methodology or methodologies similar, to other types of motion, including not just other exercises as is relevant in the context of HAR, but also other types of motion, for instance that of robotic arms, to shed light on the extent to which the technique applies to motions other than those exhibited by humans.

In terms of the particular exercises done having been selected to exemplify the two different types of motion being detected, the dumbbell curls in particular were underwhelming in terms of usefulness in determining the effectiveness of the methodology, since classifications tended to form no meaningful reflection of the motion exhibited in the actual exercise itself, unlike its counterpart of dumbbell squats which showed clearly both the repetition-based nature of the exercise and varying success rates in forming a meaningful classification, something the single-hypothesis structure of dumbbell curls was unable to capture. This serves as another endorsement for further application of the herein described methodology to other forms of commonly performed exercise containing translational motion, in order to encapsulate more of the variation of natural human motion and extract insight from it.

Varying the measurement window size and stride showed a general trend of classification accuracy increasing with window size, however, the exact extent to which this trend continues and where the limits lie was not investigated comprehensively. This leaves the opportunity in future work to investigate how the performance of this methodology in detecting linear translations varies with window size and stride.

Also found slightly lacking is demonstration of the desired effect of this methodology to generalise between people, since only one subject was used in the experiment. The potential for generalisation in groups of people could therefore provide an avenue of future investigation and experimentation.

6 Project evaluation

The primary goal of this project was to determine the feasibility of using a GLRT-based hypothesis testing method to detect periods of linear translation using accelerometer data, and the project is believed to have taken a definite first step towards doing so, but was not able to provide an exhaustive investigation or certain verdict, since doing so was too large a task in scope for the timeframe given.

ID	Success criteria	*	Notes
Functional requirements			
1	The sensor data was read in CSV format by the program and processed without error.	**Y**	The program read in CSV data without fault.
2	The program executes without error and accuracy values for both exercises are ≥ 0.65.	**Y**	Accuracy for squats was $\sim$0.778 with window size 150, stride 25, accuracy for curls was $\sim$0.996 with all window sizes and strides.
3	The program runs without error and all figures correctly render and display correct results.	**Y**	No errors are encountered and the program executes without error.
Non-functional requirements			
4	All cells are executes in less than 25 minutes total on the machine described.	**Y**	After 10 runs, the average execution time for the entire notebook was 5 minutes and 20 seconds – far below the target time of 25 minutes.
5	N/A	N/A	N/A
6	Docstrings were written to describe every function defined in the program. Code comments were written in every executable notebook cell	**Y**	Each cell and function was documented.
7	N/A	N/A	N/A

Table 5: A table describing whether each objective was fulfilled, and notes relating to each. ***Whether the objective was satisfied; **Y** – yes, **N** – no.

In summary, the program ran without error, and all objectives were fulfilled apart from those which were deemed too vague and thus not applicable.

7 Conclusion

The aim of this project has been to determine the feasibility of detecting linear translation in accelerometer data using a GLRT-based hypothesis testing method as an alternative to traditional data-driven methods which suffer from a lack of generalisability, and a larger computational intensity. To this end, experimental data was recorded by a subject in the form of dumbbell squats and dumbbell curls, and a solution was implemented in Python formulating the problem as a 3-dimensional optimisation problem followed by a threshold selection for hypothesis testing, which returned sets of hypotheses calculated and compared against the manually labelled true hypotheses of the experimental data.

Using this method, it was determined that this methodology shows promise, but time allotted was insufficient to perform an exhaustive investigation. However, this project is believed to have succeeded in providing a suitable first step in this area of research. It was found in the case of squats that the afore described methodology lead to meaningful detection of periods of linear translation, and that classification accuracy was a useful metric in determining the performance of said classification in at least the case where the motion contains periods of both translational and non-translational motion.

Conversely, in the case of curls, little insight could be gleaned despite the high classification accuracy due to the fact that curls were a motion composed of only non-translational motion – additionally, metrics such as accuracy and true & false positive & negative rates proved unhelpful in interpreting the result in the case of curls.

Potential avenues for future research, however, presented themselves during the project. For example, carrying out this methodology for more exercise sequences, especially using different subjects, would provide much-needed information about how metrics about the classification itself indicate properties of the technique's translation detection capabilities, especially given the information-sparse results of using curls in the experiment, as well as serving to clarify the extent to which the technique generalises.

References

[1] M. H. Arshad, M. Bilal, and A. Gani, "Human activity recognition: Review, taxonomy and open challenges," *Sensors*, vol. 22, no. 17, p. 6463, Jan. 2022, Number: 17 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 1424-8220. DOI: 10.3390/s22176463.

[2] P. P. Ariza-Colpas, E. Vicario, A. I. Oviedo-Carrascal, *et al.*, "Human activity recognition data analysis: History, evolutions, and new trends," *Sensors*, vol. 22, no. 9, p. 3401, Jan. 2022, Number: 9 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 1424-8220. DOI: 10.3390/s22093401.

[3] A. Ferrari, D. Micucci, M. Mobilio, and P. Napoletano, "Deep learning and model personalization in sensor-based human activity recognition," *Journal of Reliable Intelligent Environments*, vol. 9, no. 1, pp. 27–39, Mar. 1, 2023, ISSN: 2199-4676. DOI: 10.1007/s40860-021-00167-w.

[4] D. Tazartes, "An historical perspective on inertial navigation systems," in *2014 International Symposium on Inertial Sensors and Systems (ISISS)*, Feb. 2014, pp. 1–5. DOI: 10.1109/ISISS.2014.6782505.

[5] A. Mannini and A. M. Sabatini, "A hidden markov model-based technique for gait segmentation using a foot-mounted gyroscope," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, ISSN: 1558-4615, Aug. 2011, pp. 4369–4373. DOI: 10.1109/IEMBS.2011.6091084.

[6] Q. Zou, Y. Wang, Q. Wang, Y. Zhao, and Q. Li, "Deep learning-based gait recognition using smartphones in the wild," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3197–3212, 2020, Conference Name: IEEE Transactions on Information Forensics and Security, ISSN: 1556-6021. DOI: 10.1109/TIFS.2020.2985628.

[7] P. Fernandez-Lopez, J. Liu-Jimenez, C. Sanchez-Redondo, and R. Sanchez-Reillo, "Gait recognition using smartphone," in *2016 IEEE International Carnahan Conference on Security Technology (ICCST)*, ISSN: 2153-0742, Oct. 2016, pp. 1–7. DOI: 10.1109/CCST.2016.7815698.

[8] L. Chen and H. Hu, "IMU/GPS based pedestrian localization," in *2012 4th Computer Science and Electronic Engineering Conference (CEEC)*, Sep. 2012, pp. 23–28. DOI: 10.1109/CEEC.2012.6375373.

[9] R. Ashkar, M. Romanovas, V. Goridko, M. Schwaab, M. Traechtler, and Y. Manoli, "A low-cost shoe-mounted inertial navigation system with magnetic disturbance compensation," in *International Conference on Indoor Positioning and Indoor Navigation*, Oct. 2013, pp. 1–10. DOI: 10.1109/IPIN.2013.6817896.

[10] M. Luna, G. Meifeng, Z. Xinxi, Z. Yongjian, and S. Mingliang, "An indoor pedestrian positioning system based on inertial measurement unit and wireless local area network," in *2015 34th Chinese Control Conference (CCC)*, ISSN: 1934-1768, Jul. 2015, pp. 5419–5424. DOI: 10.1109/ChiCC.2015.7260487.

[11] A. Chelli and M. Pätzold, "A machine learning approach for fall detection and daily living activity recognition," *IEEE Access*, vol. 7, pp. 38 670–38 687, 2019, Conference Name: IEEE Access, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2019.2906693.

[12] A. Avci, S. Bosch, M. Marin-Perianu, R. Marin-Perianu, and P. Havinga, "Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: A survey," in *23th International Conference on Architecture of Computing Systems 2010*, Feb. 2010, pp. 1–10.

[13] M. Zhang and A. A. Sawchuk, "Human daily activity recognition with sparse representation using wearable sensors," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 3, pp. 553–560, May 2013, Conference Name: IEEE Journal of Biomedical and Health Informatics, ISSN: 2168-2208. DOI: 10.1109/JBHI.2013.2253613.

[14] M. Kok, J. D. Hol, and T. B. Schön, "Using inertial sensors for position and orientation estimation," *Foundations and Trends® in Signal Processing*, vol. 11, no. 1, pp. 1–153, Nov. 14, 2017, Publisher: Now Publishers, Inc., ISSN: 1932-8346, 1932-8354. DOI: 10.1561/2000000094.

[15] I. Skog, P. Handel, J.-O. Nilsson, and J. Rantakokko, "Zero-velocity detection—an algorithm evaluation," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 11, pp. 2657–2666, Nov. 2010, Conference Name: IEEE Transactions on Biomedical Engineering, ISSN: 1558-2531. DOI: 10.1109/TBME.2010.2060723.

[16] E. Foxlin, "Pedestrian tracking with shoe-mounted inertial sensors," *IEEE Computer Graphics and Applications*, vol. 25, no. 6, pp. 38–46, Nov. 2005, Conference Name: IEEE Computer Graphics and Applications, ISSN: 1558-1756. DOI: 10.1109/MCG.2005.140.

[17] J. Wahlström, "Detection of rotations around a fixed axis using gyroscopes," *IEEE Sensors Letters*, vol. 7, no. 8, pp. 1–4, Aug. 2023, Conference Name: IEEE Sensors Letters, ISSN: 2475-1472. DOI: 10.1109/LSENS.2023.3296342.

[18] C. R. Harris, K. J. Millman, S. J. van der Walt, *et al.*, "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357–362, Sep. 2020, Number: 7825 Publisher: Nature Publishing Group, ISSN: 1476-4687. DOI: 10.1038/s41586-020-2649-2.

[19] P. Virtanen, R. Gommers, T. E. Oliphant, *et al.*, "SciPy 1.0: Fundamental algorithms for scientific computing in python," *Nature Methods*, vol. 17, no. 3, pp. 261–272, Mar. 2020, Number: 3 Publisher: Nature Publishing Group, ISSN: 1548-7105. DOI: 10.1038/s41592-019-0686-2.

[20] T. pandas development team, *Pandas-dev/pandas: Pandas*, version latest, Feb. 2020. DOI: 10.5281/zenodo.3509134.

[21] J. Wahlstrom, "Detection of pure translations using inertial sensors," Unpublished notes, Unpublished notes, Aug. 2015.

[22] W. Sousa Lima, E. Souto, K. El-Khatib, R. Jalali, and J. Gama, "Human activity recognition using inertial sensors in a smartphone: An overview," *Sensors*, vol. 19, no. 14, p. 3213, Jan. 2019, Number: 14 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 1424-8220. DOI: 10.3390/s19143213.