Elastic Net Regression for Prosthesis Control in Short Residual Limb Amputees: Performance and Generalizability

Estimation of 1-DOF Isometric Forces using Intramuscular Electromyography Data

Oskar Berg

2023

Master's Thesis in
Biomedical Engineering

Supervisor: Alexander Olsson

LUND
UNIVERSITY

Faculty of Engineering LTH
Department of Biomedical Engineering

Abstract

This Master's thesis in Biomedical Engineering investigates the performance and generalizability of linear regression models in context of prosthesis control for short residual limb amputees. This thesis uses intramuscular electromyography data, and a regression and emplys a regression technique called Elastic Net Regression - a technique that combines L1 and L2-regularization - to predict 1-DOF isometric forces outputs from fingers and the wrist. The elastic net not only functions as a regression model but also as a feature selector, which is especially useful with higher-order interaction terms. The aim of the thesis was not merely to create a working model with high performance metrics but also to possibly train a multi-channel model that can be readily used on a new amputee without need for recalibration. Another goal was to ensure the model remains transparent and easily interpretable.

The results however, indicate that while the elastic net regression offers improved performance over standard single-channel models for the same subject, it struggled to generalize across different subjects, likely due to overfitting to individual subjects distinct characteristics. The elastic net regression model generally performed worse with lower R2-scores than the bare bones single-channel model when applying the model to new subjects.

Acknowledgements

I would like to acknowledge my supervisor, Alexander Olsson, for his guidance and expertise. Our discussions were constructive in addressing the ideas and challenges encountered in this thesis work. I appreciate his time and support throughout this project.

Contents

1 Introduction

1.1 Purpose and motivation

This thesis is exploratory work into the idea of using multi-channel intramuscular electromyography (iEMG) to estimate 1-DOF, isometric force outputs from wrist and finger actions. In this thesis a more transparent and interpretable multi-channel model will be tried and tested, compared to the more black box, neural network approach used in previous papers [1]. A model hopefully providing a clearer insight into the relationship dynamics between the inputs and outputs. Specifically, the model under investigation will be elastic net regression model, a model which combines Ridge and Lasso regularization. Furthermore, polynomial interaction terms will be integrated into the model to discern and capture the non-linear components between the iEMG and force dynamics. Aside from just obtaining robust performance on validation sets, another aspiration of this thesis is to create a model with high generalizability to new subjects. A multi-channel model that would seamlessly accommodate new subjects, and not need re-calibration.

In the work by [1], it was shown that using multiple iEMG channels (as input to an ANN) of nearby muscles to estimate finger and wrist forces outperforms One-to-One strategies, where only the matching iEMG channel for a specific hand movement is used. This finding suggests that, in spite of the spatially local nature of signals, iEMG from muscles not directly actuating the relevant DOF can provide contextual information that aid in decoding motor intent [1]. Although the degree to which activity in nearby muscles impacts the resultant force exerted by the hand remains unknown.

The concept of muscle synergies states that the nervous system simplifies the control of muscles by activating groups of muscles together as a unit, rather than individually. This allows for efficient and coordinated movement. It is suggested that a limited number of these muscle synergies are sufficient to explain a wide range of movements ([2], [3]). The concept of muscle synergies could therefore justify the use of multiple iEMG muscle channel recordings when estimating single degree-of-freedom (DOF) finger and wrist forces. Measurements of the activity of multiple muscles could potentially provide more accurate estimates of the isometric forces produced by fingers and the wrist.

In the literature there are many findings suggesting that there is high linear correlation between iEMG features and isometric finger and wrist forces [4]. Also, the Pearson correlation coefficient between features extracted from iEMG and grasping force was found to be close to 0.9 by [5]. It is described in [6] that, for small muscles with narrow motor unit recruitment force ranges, such as the first dorsal interosseous (FDI) muscle, the observed relation between force and the average rectified value (ARV) of surface EMG is reported as being approximately linear. This would support the idea that a linear model for iEMG might also be suitable for different type of 1-DOF hand forces. However, a study investigating the relationship between integrated sEMG and extensor carpi radialis (ECR) force found the relationship to be more nonlinear. This study suggested that a double exponential function could best describe the two main mechanisms behind voluntary contraction, namely recruitment and firing rate ([7]). Regarding grasping force estimation using iEMG, [8] showed that an exponential fit between grasping force and features of the iEMG is superior to a linear fit. The authors also showed that an ANN performed better than a linear model, indicating that some type of non-linear regression model could work to improve finger and wrist force estimation as well.

1.2 Dataset Overview and Delimitations

The iEMG-force-dataset is collected by the neuroengineering group at department of Biomedical Engineering at LTH [9]. It has recorded intramuscular electromyography (iEMG) and synchronous forces as a subject is performing a number of ”isometric movement” tasks. The dataset includes simpler 1-DOF tasks such as, flexion and extension of the wrist and fingers, as well as more complex multi-DOF movements such as pinching or grasping.

There is a total of fourteen able-bodied, male subjects recorded in the dataset, where half are differentiated with two different recording protocols, the short residual limb (SRL) and the long residual limb (LRL) [9]. The SRL protocol targeted the following muscles:

- Flexor Carpi Radialis (FCR) - responsible for wrist flexion
- Extensor Carpi Radialis longus (ECR) - responsible for wrist extension
- Pronator Teres (PT) - responsible for forearm pronation
- Flexor Digitorum Profundus (FDP) - responsible for flexion of fingers D2-D5
- Extensor Digitorum Communis (EDC) - responsible for extension of fingers D2-D5
- Abductor Pollicis Longus (APL) - responsible for thumb abduction

This master thesis is limited the analysis to 1-DOF movements of the short residual limb protocol, using data from four subjects: subjects 2, 3, 7, and 15. These were specifically chosen for having the best quality iEMG recordings according to [4]. Furthermore, only 5 forces out of 18 in total were chosen: Middle finger extension, little finger flexion, wrist extension, wrist flexion, and thumb flexion. The rationale behind selecting these five forces is that they all have direct correspondence to specific iEMG channels in the SRL protocol ([4]). This allows for a comparison between multi-channel models and single-channel models, aswell as making the results more concise and comprehensible.

1.3 Problem Statements

In essence, this is a sequential or time series dataset. Standard procures when working with a timeseries, e.g. a series daily temperatures, is to use previous values or prediction errors in the time series to predict future values, i.e. using a standard AutoRegressive (AR) or more complex models such as Seasonal AutoRegressive Integrated Moving Average (SARIMA) models [10]. In time series modeling one can also add exogenous variables to predict new values in a time series. An example of such a model would be an ARMAX (AutoRegressive Moving Average with eXogenous) model [10]. However, in the context of an amputee, one would not have access to preceding force values when estimating current force values. The goal is ultimately prosthesis control for amputees, meaning that you would not be able to measure any type of finger or wrist forces to feed back into the model. Moreover, in real scenario there is not really any discernible time dependence that can be used. The force exerted by a finger at t = 0s provides little to no information about the force exertion at t = 5s, which could range anywhere from 0 to maximum voluntary contraction (MVC). In this data set it is known how the force signal will move, so to use too much temporal information may perhaps even be detrimental to real life applications, where the amputee should be able to manipulate their hand prosthesis however they please. It could create a model which assumes a certain temporal structure, which might work well within the dataset, but in reality where the temporal structure is not pre-determined, the model could be completely erroneous.

Another tricky thing in regards to using simple or multiple linear regression for non-stationary time series applications, is the issue of misinterpretation and spurious relationships. One might mistakenly interpret that one time series is is highly predictive or causative of another time series, when in fact it might just be the result of coincidental pattern or a 'lurking' variable which drive both processes. However, in the context of this thesis, it is physiologically known that muscle activity in the forearm cause movements of the hand.

2 Theory

2.1 Electromyography

Electromyography (EMG) is a technique used to measure the electrical activity of muscles. EMG signals are generated by the activation of muscle fibers, which produce electrical potentials that can be detected by electrodes placed on the skin or within the muscle tissue [8]. EMG signals can provide information about the intensity and timing of muscle activity, which can be used to control prosthetic devices or to study muscle function [8].

There are two main types of EMG: surface electromyography (sEMG) and intramuscular electromyography (iEMG). sEMG involves placing electrodes on the surface of the skin to measure the electrical activity of muscles, while iEMG involves inserting fine-wire electrodes directly into the muscle tissue to measure the electrical activity of individual muscle fibers [9].

sEMG is non-invasive and provides information about the overall muscle activity, making it a popular choice for many applications. However, the use of sEMG as a control signal for prostheses has some limitations such as being able to measure only from superficial muscles, being sensitive to crosstalk, and causing skin irritation with repeated use [11].

iEMG, on the other hand, is an invasive method that provides more detailed and local information about the electrical activity of individual muscle fibers. It is less affected by noise and crosstalk from adjacent muscles and it can be recorded from deep muscles. iEMG has been proposed for use in controlling prosthetic devices as it offers possibilities for chronic implants and can overcome some limitations of sEMG. However, it should be noted that due to its high selectivity, iEMG signals may offer limited representation of the global muscle activity and force produced by the muscle [5].

2.2 Linear regression

Linear regression is a statistical modeling technique used to establish a relationship between a dependent variable and one or more independent variables [12]. Simple linear regression focuses on understanding the relationship between a single dependent variable and one independent variable, while multiple linear regression extends the analysis to include multiple independent variables.

The general idea behind multiple linear regression is to fit a linear equation to the data points that best represents the relationship between the dependent variable and the independent variables. The equation for multiple linear regression can be written as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + \epsilon \quad (1)$$

Where Y is the dependent variable, $X_1, X_2, ..., X_p$ are the independent variables, β_0 is the intercept, $\beta_1, \beta_2, ..., \beta_n$ are the coefficients associated with each independent variable, and ϵ is the residual error. In standard linear regression the ϵ is assumed to be normally distributed. The goal is to estimate the coefficients $\beta_0, \beta_1, ..., \beta_n$ that best describe the relationship between the independent variables and the dependent variable. There are a number of assumptions associated with multiple linear regression [13],

- **Linearity**: The relationship between the dependent variable and the independent variables is linear. This assumption implies that the coefficients represent the change in the dependent variable associated with a one-unit change in the independent variables.
- **Independence**: The observations are independent of each other. In the context of sequential data, this assumption implies that there is no time dependency between observations. Autocorrelation occurs when the residual errors of the model exhibit (statistically significant) correlation or dependence over time. Violation of this assumption may lead to biased coefficient estimates, i.e. coefficient estimates tend to consistently overestimate or underestimate the true relationship.
- **Homoscedasticity**: The residual errors have constant variance across all levels of the independent variables. Homoscedasticity ensures that the variability of the errors is consistent throughout the range of the independent variables. Heteroscedasticity, where the variance of the errors varies across the independent variables, can lead to inefficient coefficient estimates, i.e. they are no longer the best estimates, in the sense of having the smallest possible variance. Techniques like weighted least squares (WLS) regression can mitigate the inefficiency caused by heteroscedasticity.
- **Normality**: The residual errors are normally distributed. This assumption allows for the use of statistical tests and confidence intervals based on the normal distribution.
- **No multicollinearity**: The independent variables are not highly correlated with each other. Multicollinearity can lead to unstable and unreliable coefficient estimates, making it challenging to isolate the individual effects of the independent variables. Slight changes in the data could lead to substantially different coefficient estimates.

2.3 Ridge (L2) regularization

Ridge Regression is used to address overfitting, a common issue where models with a high number of independent variables perform well on training data but poorly on unseen data. This problem is particularly pronounced in the presence of multicollinearity where coefficients can become highly inflated. In multiple linear regression, ridge Regression is a regularization technique that offers a solution to combat the issue of overfitting and poor generalizability of models with highly correlated independent variables [14]. The ridge regression optimization function can be written as:

$$\sum_i (Y_i - (\beta_0 + \textstyle\sum_j \beta_j X_{ij}))^2 + \lambda \textstyle\sum_j (\beta_j)^2 \quad (2)$$

Where the left hand side is the optimization function for standard multiple linear regression, and the right hand side the L2 penalty term. If the coefficients take on large values the optimization function is penalized, effectively promoting smaller coefficients. The regularization parameter, λ controls the amount of shrinkage applied to the coefficient estimates.

2.4 LASSO (L1) regularization

In addition to shrinking coefficients, LASSO Regression is also able to set some of them to exactly zero. This inherent feature selection makes Lasso regression especially useful if one is looking for a sparse model [15]. The LASSO optimization function is written as:

$$\sum_i (Y_i - (\beta_0 + \sum_j \beta_j X_{ij}))^2 + \lambda \sum_j |\beta_j| \tag{3}$$

Similarly to ridge regression, the regularization parameter λ controls the amount of regularization applied. Briefly explained, due to the geometric shape of the L1 constraint, a diamond shape in case of a 2D-model, the best fit often occurs at the points where this diamond touches an axes, which corresponds to one of the coefficients being set to zero.

The penalty term effectively selects a subset of the most important features and discarding less relevant ones, creating a parsimonious model [14]. However, this strength can also be a shortcoming in situations where features are highly correlated. LASSO tends to arbitrarily pick only one of the collinear variables and shrink the others to zero. It has an inconsistent selection and cannot perform grouped selection [15].

2.5 Elastic net regression

Elastic net regression combines the benefits of both ridge regression and LASSO regression. It utilizes a combination of L1 (LASSO) and L2 (ridge) penalties to balance between the advantages of both regularization techniques [16]. The elastic net regression optimization function can be written as:

$$\sum_i (Y_i - (\beta_0 + \sum_j \beta_j X_{ij}))^2 + \alpha\lambda \sum_j |\beta_j| + (1-\alpha)\lambda \sum_j (\beta_j)^2 \tag{4}$$

Where λ is the regularization strength and α, a variable defined in the range (0,1), specifies the mix between L1 and L2 regularization. An $\alpha = 0$ would mean the model is equivalent to ridge regression and conversely, an $\alpha = 1$ would mean it is equivalent to LASSO regression. The combination of these penalties allows elastic net regression to perform a 'grouped' feature selection [14], which might be useful in certain circumstances. When faced with multicollinearity due to e.g. interaction terms, elastic net performs this grouped feature selection. Meaning, instead of arbitrarily choosing one interaction term over another, it keeps or discards groups of correlated variables together. This is useful when comparing the subsets of chosen (independent) variables between models trained on different data, where just LASSO would make it difficult due to its arbitrary feature selection.

2.6 Performance analysis, R2-score

R-squared (R2) is a commonly used performance measure in regression analysis that represents the proportion of the variance in the dependent variable that is explained by the independent variables. The equation for calculating R-squared (R2) is as follows [17]:

$$R2 = 1 - \frac{SSR}{SST} = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \tag{5}$$

Where SSR represents the sum of squared residuals (also known as the sum of squared errors) and SST represents the total sum of squares. Y_i is the observed value of the dependent variable and $\hat{Y}_i$ is the predicted value of the dependent variable for the i^{th} observation. Lastly, $\bar{Y}$ is the mean of the observed values of the dependent variable.

The sum of squared residuals (SSR) measures the variation that is not explained by the regression model, representing the discrepancy between the observed values and the predicted values from the model. The total sum of squares (SST) measures the total variation in the dependent variable. R2 normally ranges from 0 to 1, where a value of 1 indicates that the model explains all the variation in the dependent variable, and a value of 0 suggests that the model provides no improvement over the mean.The R2 score can be negative if the model fits worse than the mean.

3 Methods

The preprocessing of the dataset was exclusively done using the Python programming language [18]. Signal processing tools from Scipy package, such as butterworth filters, filtering functions, spline interpolation were used. Figure 1 shows the basic outline of the preprocessing.

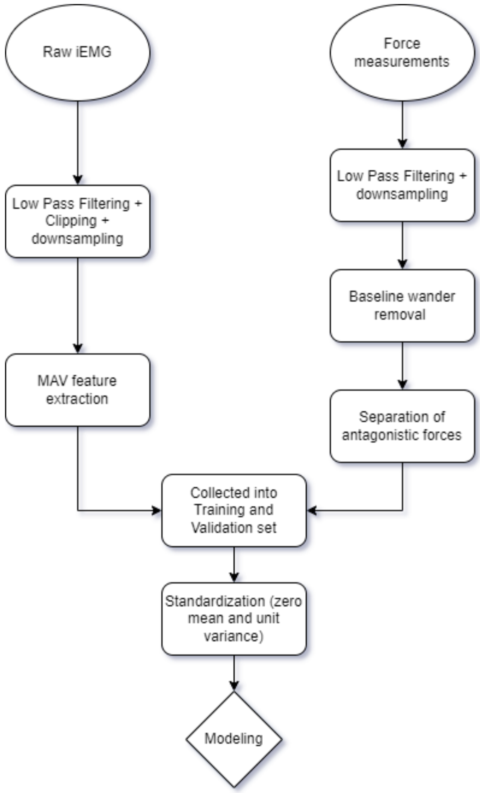


Figure 1: Preprocessing flowchart.

3.1 Data Collection and Preprocessing

The movements of interest in the dataset pertain to 1-DOF movements. There are a total of eight such movements, which include:

- Index finger: flexion-extension
- Middle finger: flexion-extension
- Ring finger: flexion-extension
- Little finger: flexion-extension
- Thumb: flexion-extension
- Thumb: adduction-abduction
- Wrist: flexion-extension
- Wrist: supination-pronation

The recording procedure for each of these 1-DOF movements consisted of two tasks:

1. **Maximal Voluntary Contraction (MVC) Stage:** In this stage, the subject exerts the maximum voluntary force in the two opposing directions specific to the 1-DOF movement.

2. **Sine Cue Force Matching Stage:** After determining the MVCs, the subject then engages in a task where they attempt to match their force output to a sinusoidal cue signal, with a duration of 9 periods and a frequency of 0.1 Hz. During this task, the subject uses both the protagonist and antagonist muscle pairs to create a force curve which matches the sine cue signal.

An unfortunate consequence of combining opposing forces is that it becomes necessary to split the force signal into two distinct signals, separating the forces produced by the antagonistic muscles. This separation is necessary to be able to compare single-channel estimation with multi-channel models. This would need to be dealt with in the preprocessing stage.

Firstly, all force sensor signals and iEMG signals were low pass filtered at 10 Hz and 500 Hz respectively with a 2nd order Butterworth filter. Also, the iEMG signals were clipped at 1st and 99th percentile to mitigate the effect of voltage spikes. Both the force signals and iEMG were then downsampled 1/10th to reduce computational costs. Resulting sampling frequency was then 1024 Hz which roughly equals 1 ms between samples.

3.1.1 iEMG feature extraction

The Mean Absolute Value (MAV) is a commonly used feature in sEMG signal processing. It quantifies the average magnitude of the EMG signal and provides an estimate of the overall muscle activity. The MAV can be calculated using the following equation:

$$MAV(k) = \frac{1}{N} \sum_{l=k-N}^{k} |X(l)| \tag{6}$$

where, MAV represents the Mean Absolute Value at discrte time point, k. N is the window size and $X(l)$ represents the iEMG signal voltage amplitude at time point k. There are many other possible feature to use, such as Zero Crossings, Slope Sign Changes, Variance etc but MAV was chosen for this thesis, because the following reasons:

1. **Scope**: This thesis is more about the use of multi-channel EMG than analysis of different feature extractions. Thus, the work was limited to one type of extraction feature.

2. **Prevalence in EMG Studies**: It is a very common feature used in EMG studies. However whether it is the optimal feature for iEMG analysis is something that could be discussed. In [4] one can see many different iEMG features compared, and MAV ranks high among them.

3. **Straightforward and Robustness to noise**: MAV offers ease of interpretation due to its straightforward computation. MAV provides an average measure, making it inherently robust to transient spikes or anomalies in the data.

Mean absolute value (MAV) transformation of the rectified raw iEMG signals was performed using a convolution between an iEMG channel and a rectangular window of unit amplitude and 500 ms length (512 samples). An overlapping window approach was employed with a step size of 32. In [9] it is stated that a window size of 250 ms has the best trade-off between RMSE, correlation, and the controller delay. However, windows sizes around 500 ms had the best RMSE and correlation and was therefore chosen. Window sizes larger than 500 ms was shown in same paper to yield negligible improvements and even decreasing results as the window size increased.

3.1.2 Force sensor measurements

In many of the force signals, a noticeable DC offset was present and in some signal, there was even a noticeable baseline wander present jointly with the DC component, as seen in figure 2. Obviously this type of baseline trend would affect the models estimation implemented in future stages and was something that needed to be fixed. Two methods were considered when fixing this issue, a high pass filter and spline interpolation.

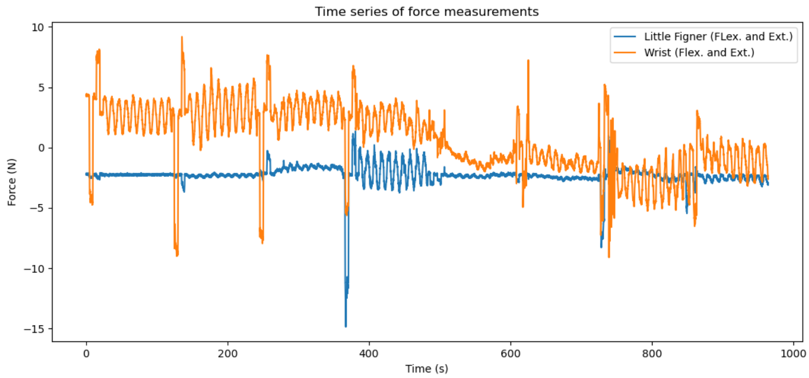


Figure 2: Two examples of force measurement sensor signals recorded from subject 7. Some form of baseline wander is present in the Wrist flexion- and extension force signal, jointly with a DC offset. However in the little finger flexion-extnesion, there is only a DC offset, but no noticable baseline wander.

For the high pass method filter method a frequency analysis had to be performed. Since the sine cue signal that the subjects were supposed to follow had a frequency of 0.1 Hz, this frequency needed to remain largely undistorted. If the cutoff frequency is too close to this frequency, the amplitude response of the filter would attenuate this particular frequency too much and ultimately distort the signal of interest. On the other hand, the cut off frequency shoyld be kept as high as possible to remove as much of the low frequency noise as possible. Figure 3 shows the amplitude response @ $f = 0.1$ Hz of a 2nd order Butterworth filter plotted against cut-off frequencies ranging from 0.1 mHz to 0.1 Hz.

Forward-backward filtering ('filtfilt' from Scipy) was used to prevent phase distortion and ensure a linear phase response [19]. By filtering data in both forward and backward directions, phase shifts from the forward pass are negated in the backward pass. A consequence of the forward-backward filtering, is that the amplitude attenuation is applied twice, i.e. the magnitude response of the filter is squared. Since the high pass filter method removed the baseline wander and DC component, one could separate the signal easily into flexion and extension by separating at zero.

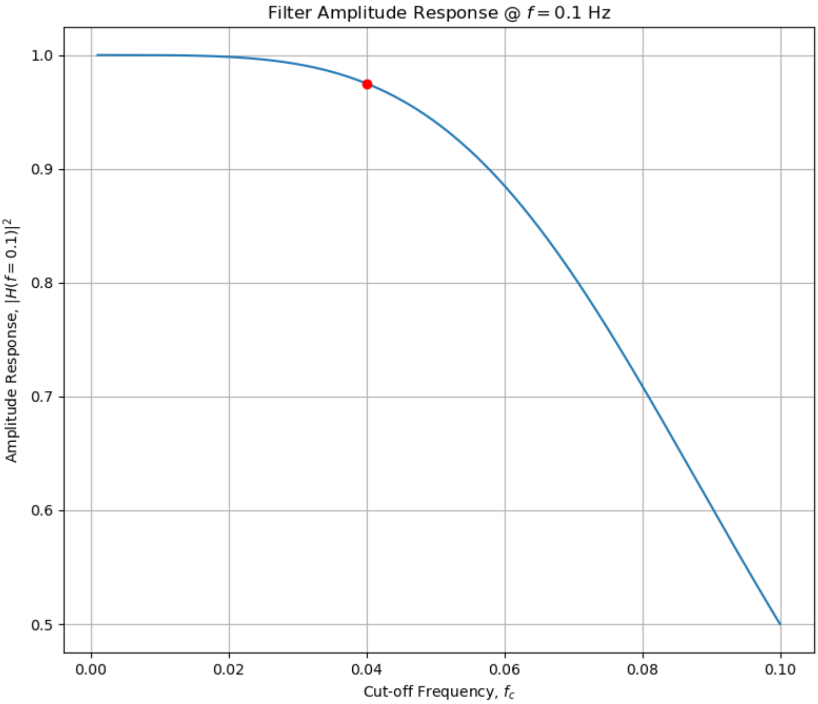


Figure 3: Squared amplitude response @ f = 0.1 Hz, using high pass filter with different cut-off frequencies ranging from 0.001 to 0.10. The red dot specifies the chosen cut-off frequency.

The spline interpolation method consisted of using information from the sine cue signal. When the sinusoidal signal switches from positive to negative, it is the same time that the subject should switch from e.g. flexion to extension, i.e. that's where the signal should be separated. Due to the fact that there can be noise at the zero-crossing point, an average of the 50 points before and after the zero-crossing point was used as the interpolation point. The sine cue signal is 9 periods per task, so in total there ends up being 20 points which are spline interpolated. The interpolated signal is then used as the separating line between the antagonistic forces.

In the end, the high pass filter method was chosen, due to being a more straight forward approach and easier to generalize. Figure 4 shows the before and after filtering using the high pass filter method. A similar type of reprocessing on force measurement signals was found to have been done in [20]. Even though the sine cue tasks (x.3) was largely undistorted, the high pass filter did unfortunately distort the data of the MVC tasks (x.1 and x.2). The MVC stages are essentially two consecutive rectangular pulses (one negative and one positive), which contain significant low-frequency components due to its sudden transitions and finite width, and is thus distorted by the high pass filter.

In the end, only the sine cue stages were used as modeling data, due to both of the baseline wander removal methods not working properly on the MVC stages. Another method that would not affect the MVC stages could be explored, but since there was more than enough good data from the x.3 stages, this method sufficed. Additionally, the MVC stages are very short and more data of constant contraction is not of great interest.

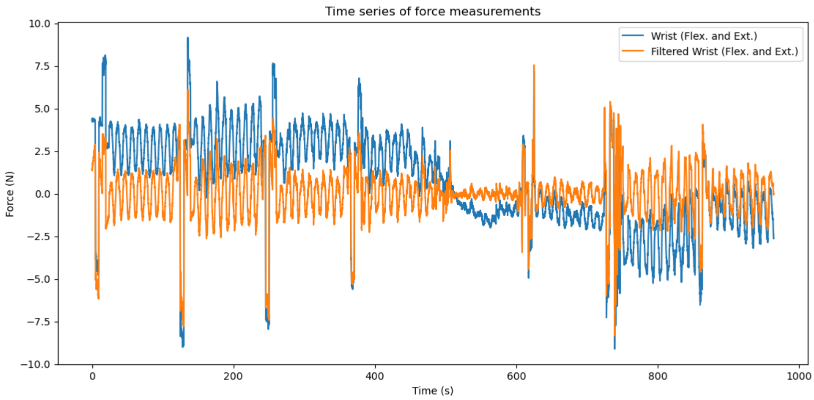


Figure 4: Unfiltered wrist force signal (blue) and filtered signal (orange). As seen in the filtered signal, the baseline wander in no longer present. The sine cue-stages are largely undistorted. However, upon more detailed inspection the MVC-stages were heavily distorted.

3.2 Modeling

The Python package, Statsmodels [21] was used for modeling. A simple linear regression model between related iEMG channel and force type was created as a standard to compare against the multi-channel model. An excel look-up sheet constructed by [9] was used to find related iEMG channel and force.

3.2.1 iEMG channel Correlation matrix analysis

Prior to modeling, a correlation analysis was performed. In the use of multiple linear regression models, one has to be careful if our independent variables are highly correlated with each other, due to the issue of multicollinearity. Multicollinearity can cause problems when applying the model to new, unseen data. The presence of multicollinearity can lead to overfitting, which may not generalize well to new data. Another issue arises is when we start dissecting the model to understand how each variable is contributing individually. This is where multicollinearity muddles the waters because it can obscure the importance of individual variables and make it hard to identify redundant variables. Unlike sEMG, iEMG does not typically present issues with crosstalk. However, the possibility for multicollinearity still needs to be investigated.

3.2.2 Specific Movement Task Modeling

In the Specific Movement Task Modeling, only data from the task, where the force of interest is active, is used in training and validation. For example, if the force of interest is middle finger extension, then only data from the middle finger flexion-extension movement task is used in the modeling. As stated in the introduction of the dataset, the sine cue signal for each stage is approximately a 9-period long sinusoidal signal. The first 4/5 of the signal is used for training and the last 1/5-part of the sine cue stage is left as a validation set. The iEMG feature signals were standardized to have a mean of 0 and variance of 1. This ensures that the estimated coefficients fall within a comparable magnitudes, allowing for meaningful comparisons. Furthermore, the dependent variables were standardized to eliminate the need for an intercept in the regression models.

An initial idea was to use standard multiple linear regression as an all-to-one model. Meaning one coefficient for each iEMG feature channel is estimated. Then use significance tests as basis for investigating which iEMG channels are relevant when estimating a specific force type. A standard statistical method is to use t-tests in order to determine whether independent variables are statistically significant. However, due to unfulfilled assumptions (autocorrelation, heteroschedasticity, non-normality) and large amount of data (high sampling frequency), the accuracy and reliability of p-values became questionable. In addition to the violated assumptions, the issue with a large amount of data points, is that even with minor relationships between the independent variables and the dependent variable can result in statistically significant p-values. While certain iEMG channels may appear significant through p-values, the coefficient magnitude determines their actual importance.

Regularization was introduced to the multiple linear regression models to more rigorously identify the relevant and irrelevant iEMG channels. LASSO seemed an apt choice, since we are expecting a sparse solution, i.e. some iEMG channels are useful and others not so much. LASSO had success in that it reduced the amount of coefficients, making a simpler underlying model. However, even in these multi-channel models there was still temporal structure left uncaptured, a notable large 'seasonal' component present at 0.1 Hz in the Auto Correlation Function (ACF). This indicated that the sinusoidal structure embedded into the force signals Was not entirely captured by the multiple linear regression LASSO-model.

An idea to combat this was to use interaction terms to capture and identify the non-linear components in the iEMG-force relationship. Elastic net was chosen as the ideal regularization technique. Not only can it deal with overfitting issues that happens with many (and likely collinear) interaction terms. It can also perform grouped feature selection of the independent variables. Allowing us to perhaps identify a set of independent variables that explains the underlying iEMG-force-relationship, and investigate whether this set is consistent across all subjects.

In elastic net regression, the hyperparameters λ and α from equation 4 need to be tuned before performing regression. A conventional method for hyperparameter tuning in machine learning is grid search [22], where we search over a 2D-space to determine the optimal pair, λ and α, that maximizes the R2-score of the validation data. An example grid search 2D-space from subject 7 is illustrated in figure 5. However, variability was encountered in the optimal parameter pair across different forces and subjects. As a result, I opted for a manual selection of parameters, settling on $\lambda = 0.2$ and $\alpha = 0.6$. This choice leans slightly towards LASSO, with a relatively high regularization strength to emphasize a sparse model. However, even with the added interaction there were still uncaptured structure left in the residuals. In future work, a study of the utility of even higher-order interaction terms could be examined. However, the focus of this thesis will stay on the results of the 2nd order interaction terms.

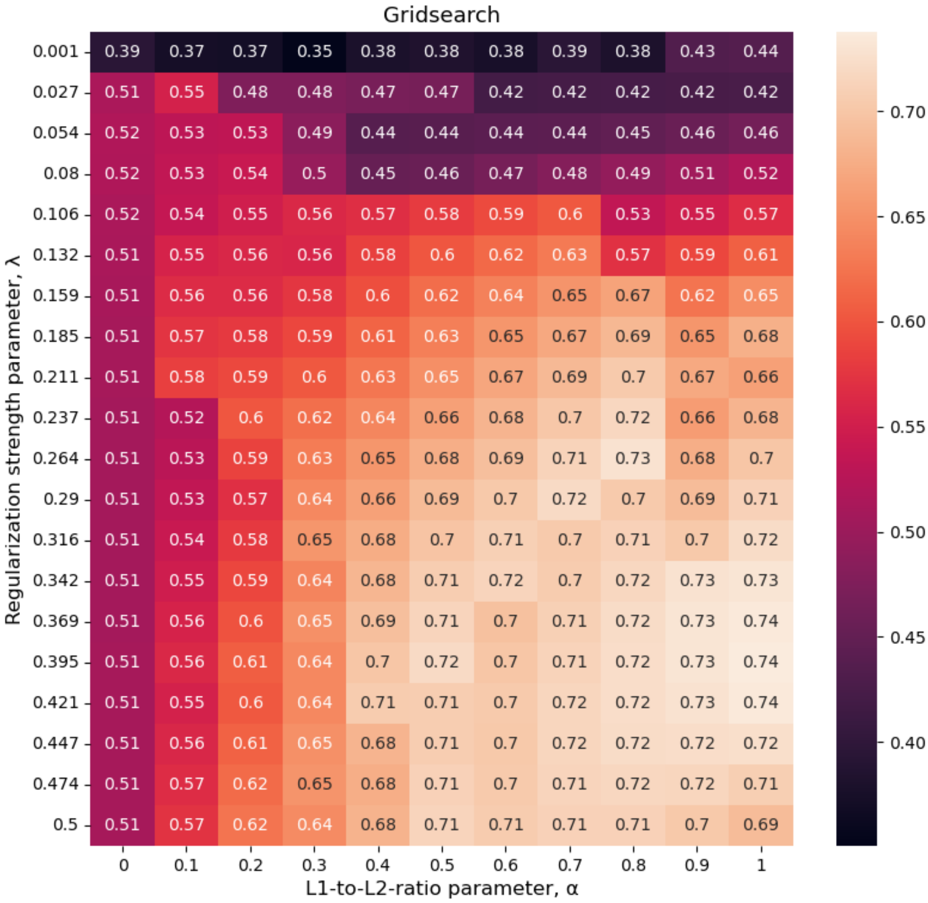


Figure 5: R2-score from validation data estimation using different hyperparameter values. Data collected from Thumb flexion task performed by subject 7.

For performance analysis, an idea of mine was to make use of having recordings from multiple subjects in the validation process. Patterns that consistently emerge across different subjects are very likely to be genuine, while randomness will not be consistently be replicated. However, every subject's body is unique, and as a result, a specific iEMG-force model that performs well for one subject may not necessarily apply universally. Therefore, a mix of performance measures for intra-subject performance and inter-subject performance is presented in the results. Annotated heatmaps were created to visually to present the R2-scores derived from validation data. Two types of R2-heatmaps were constructed:

1. **Intra- and Inter-Subject, R2-score Heatmaps:**
   Here, models are trained on only a single subject's training data. Then, tested on the validation sets from all other subjects (including its own). The purpose of these kinds of heatmaps is to get an idea of performance on its own validation data as well as generalizability to the other subjects.

2. **Leave-One-Subject-Out, R2-score Heatmaps:**
   Here, the models are trained on data from all subjects except one, which is left out for validation . This method parallels the leave-one-out cross-validation commonly seen in machine learning. The point of the leave-one-subject-out heatmap is to measure whether there is an increased generalizability when using training data from multiple subjects and testing it on validation data from a new, unseen subject.

Additionally, a heatmap consisting of jaccard indexes, a measure of overlapping features between subjects was constructed. The idea of this analysis was to explore the possibility of there being an underlying set of channel interaction terms that explains a particular force, i.e if there is a common set of independent variables chosen by the elastic net that is the same across the four subjects.

The jaccard index is defined as

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{7}$$

where A would be (in our context) the set of independent variables with non-zero coefficients estimated subject X's training data and B would be the set of non-zero independent variables with non-zero coefficients estimated from subject Y's training data.

3.2.3 All Movement Tasks Modeling

In the All Movement Tasks Modeling, data from all 1-DOF sine cue tasks are used ([1-8].3) in training and validation. This type of model and evaluation might be considered more realistic, because even during tasks where a certain force is inactive and is in a rest-state, the iEMG amplitude and output force should be matching. Similarly to the specific task modeling, the last 1/5th part from each task is extracted and collected into a validation set, while the remaining is collected into a training set. Also, same type of performance analysis and heatmaps were constructed as described in the specific movement task modeling.

4 Results

Figure 6 illustrates the estimation procedure on a validation set for the Specific Movement Task Models. Both the multi-channel and the single-channel models are illustrated with an arbitrary task and force. The All Movement Tasks Modeling approach would essentially be equivalent but with multiple tasks as opposed to the single task presented.

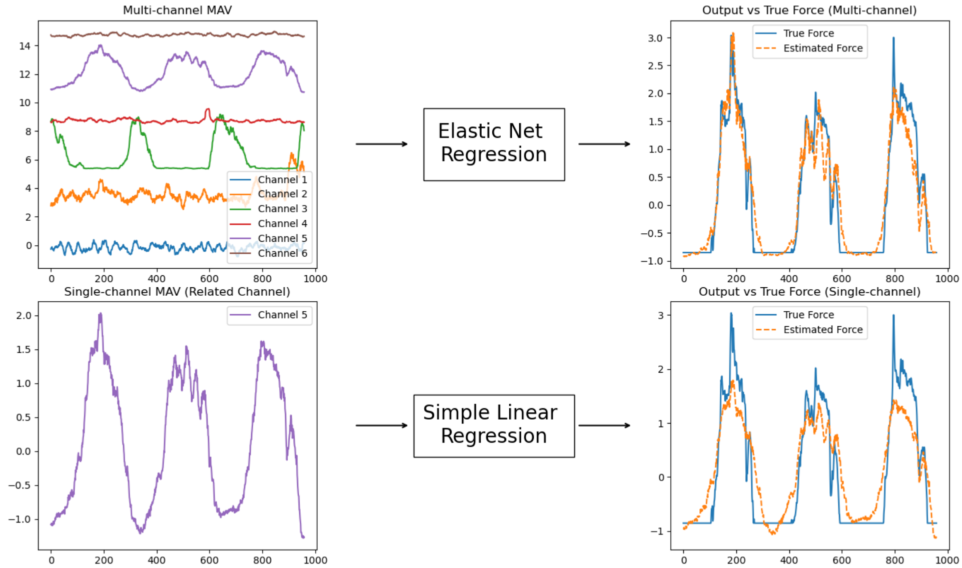


Figure 6: Illustration of the estimation procedure for Specific Movement Task Models. The elastic net (multi-channel) model uses all available MAV-channel data as well as the 2nd order interaction terms of them. The available MAV-channels are presented with an offset for increased visibility. The 2nd order interaction terms are not illustrated in the figure.

4.1 iEMG channel correlation

High correlation between some channels was found in Figure 7, notably for PT-FCR and ECR-APL (for subject 3). Knowing that crosstalk is unlikely and the performed movements where solely 1-DOF, the high correlation could suggest that there is some form of co-activated and synergistic activation between these muscles. It not unreasonable that some muscles have a secondary or stabilizing role in conjunction with the primary muscle for certain movements, especially if they share close anatomical proximity.

Additionally, a large variation between correlation scores across subjects was observed. For example, the cross-correlation between the PT- and FCR-channel were (0.56, 0.24, 0.82, 0.71) for subjects (s7, s3, s2, s15) respectively, indicating that there might be quite large differences in the co-activation patterns across subjects. One would expect some differences due to variation in environmental setting and electrode insertion, but correlation differences to this degree was unexpected, indicating that it might be difficult for multi-channel models to generalize to subject data that is not included in the training set. Especially, standard multiple linear regression which would assume that the iEMG channel-to-force coefficients have similar magnitude (and sign) across time and subject.

Figure 7: Correlation matrices constructed from data from subjects 7, 3, 2, and 15. Showing cross-correlations between MAV-transformed iEMG channels. Abbreviated names of the iEMG channels are displayed on the rows and columns.

4.2 Specific Movement Task Modeling

In this section specific movement task modeling is employed. Figure 8 uses Elastic net regression models on data from single subjects and presents R2-scores from intra- and inter-subject validation sets. For every force type, there are four models trained on each subject's training data. Each model is then tested on the validation set of all other subjects, including its own validation set. Thereby, creating a heatmap of dimension 4x4. In total, there are five of these heatmaps, one for each force type.

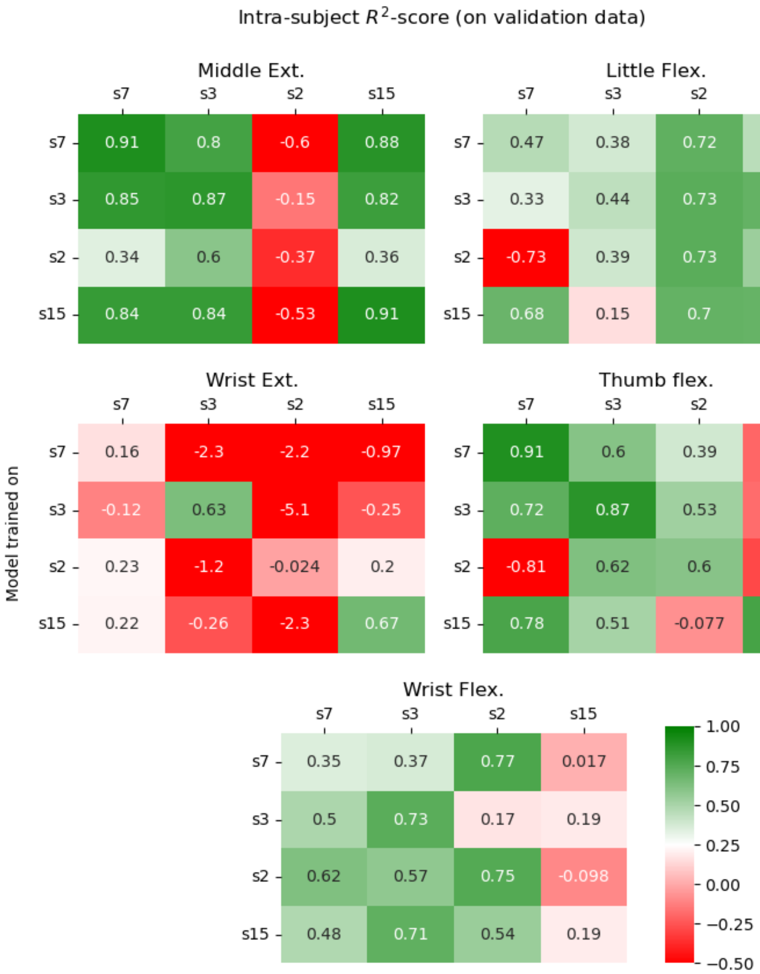


Figure 8: Multi-channel (Elastic net) models validated on other subjects using R2 as performance measure. The purpose of the heatmaps is to get an idea of performance and generalizability between subjects on validation data. Note that the subject data from which the model is trained stated on the rows of the heatmaps and the columns state from which subject the validation data comes from.

Figure 9 presents intra- and inter-subject R2-scores for single-channel models. The results here is used as a standard to compare against the elastic net model previous figure. Single-channel models, refer to the fact that only one independent variable, the corresponding iEMG channel to force type, is used in the linear regression model. In other terms, it is just plain Simple Linear Regression, the most bare bones model possible. For every force type, there are four models trained on each subject's training data. Each model is then tested on the validation set of all other subjects, including its own validation set. Thereby, creating a heatmap of dimension 4x4. In total, there are five of these heatmaps, one for each force type. A list of the related iEMG-channel and force type used to construct figure 9 is presented here for increased clarity for the reader:

- Extensor Digitorum Communis (EDC) - Middle finger extension
- Flexor Digitorum Profundus (FDP) - Little finger flexion
- Flexor Carpi Radialis (FCR) - Wrist flexion
- Extensor Carpi Radialis longus (ECR) - Wrist extension
- Abductor Pollicis Longus (APL) - Thumb flexion

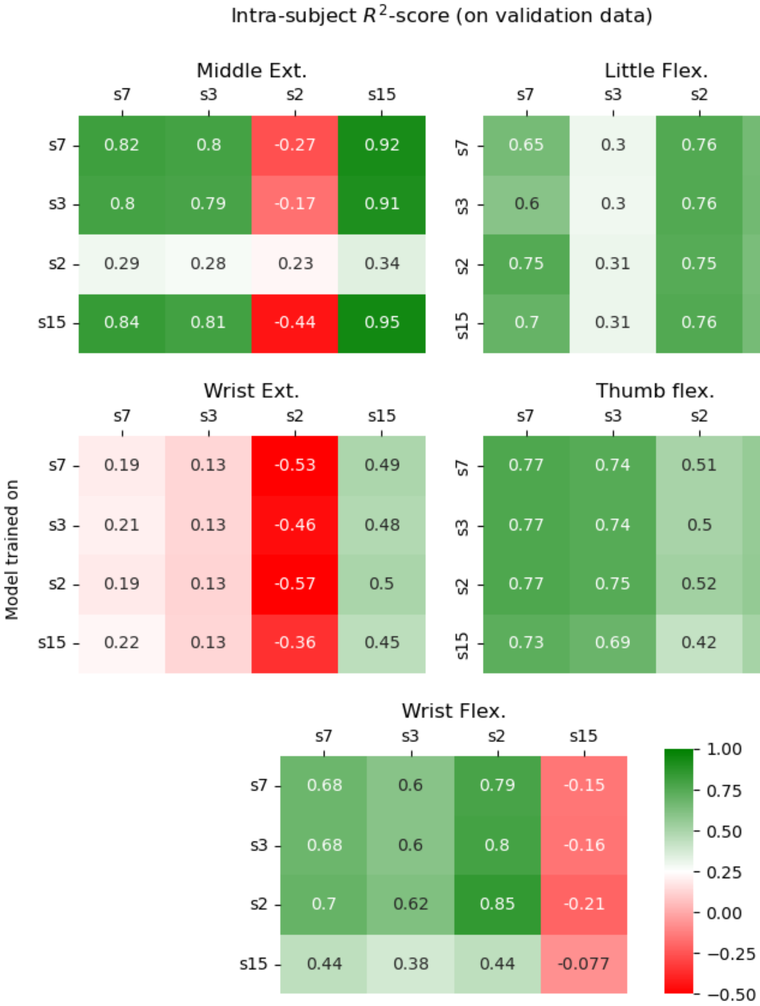


Figure 9: Single-channel models validated on other subjects using R2 as performance measure. The purpose of this figure is to see how single-channel performs and generalizes compared to multi-channel models in figure 8. Similarly to aforementioned figure, the subject data from which the model is trained stated on the rows of the heatmaps and the columns state from which subject the validation data comes from.

Figure 10 presents five jaccard index heatmaps, representing the common features chosen by the elastic net regularization technique. A Jaccard index of 1 would indicate that two models, each trained on data from different subjects, have the exact same non-zero independent variables in the elastic net model. Conversely, an index of 0 would indicate that there are no overlapping features, indicating that the chosen, underlying independent variables of the models are completely different. Note that the diagonal elements are all 1 due to showing the common features for the same subject model.

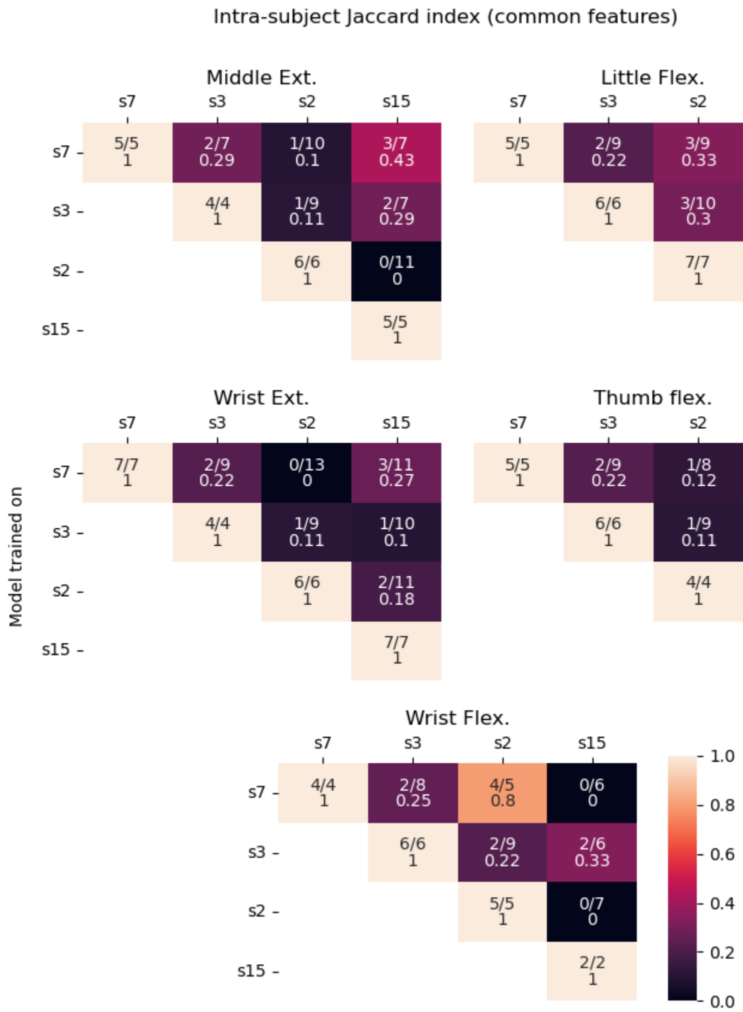
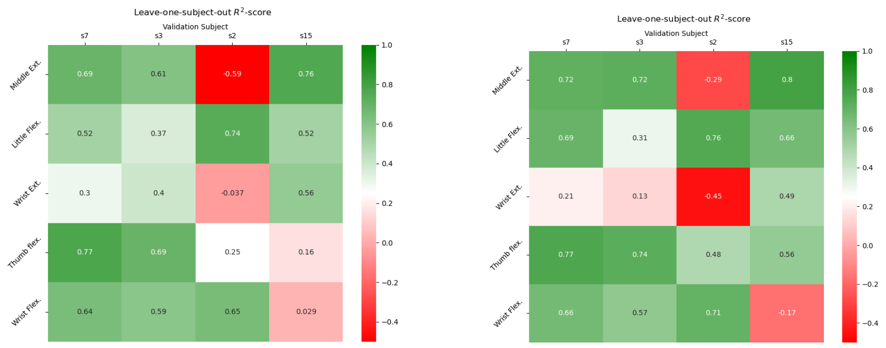


Figure 10: Jaccard index matrices for single subject models, showing common independent variables selected by the elastic net regularization between subjects.

Figure 11 presents R2-score for the leave-one-subject-out validation. It shows the elastic net and single-channel models side-by-side for easy comparison.



(a) Elastic Net models

(b) Single-channel models

Figure 11: Leave-one-subject-out validation heatmaps. Rows indicate which force type the model estimates, while columns represent the validation subject. For instance, the top-left square presents the R^2-score for Middle finger extension force estimated using validation data from subject 7. Thus, the model was trained on data from subjects s3, s2, and s15, but excluded data from subject 7.

4.3 All Movement Tasks Modeling

In this section All movement Tasks Modeling is employed, meaning data from all sine cue tasks are collected into the training and validation sets. All figures are equivalent to those in the previous section (Section 4.2), but the difference being the models are trained on data from all movement tasks.

- Intra- and inter-subject R2-scores for Elastic net models are presented in Figure 12.
- Jaccard indexes from the Elastic net regression are shown in Figure 13.
- The single-channel models, used as a comparison standard against multi-channel models, can be found in Figure 14.
- Figure 15 displays the leave-one-subject-out validation, with side-by-side comparison of both the elastic net models and single-channel models.

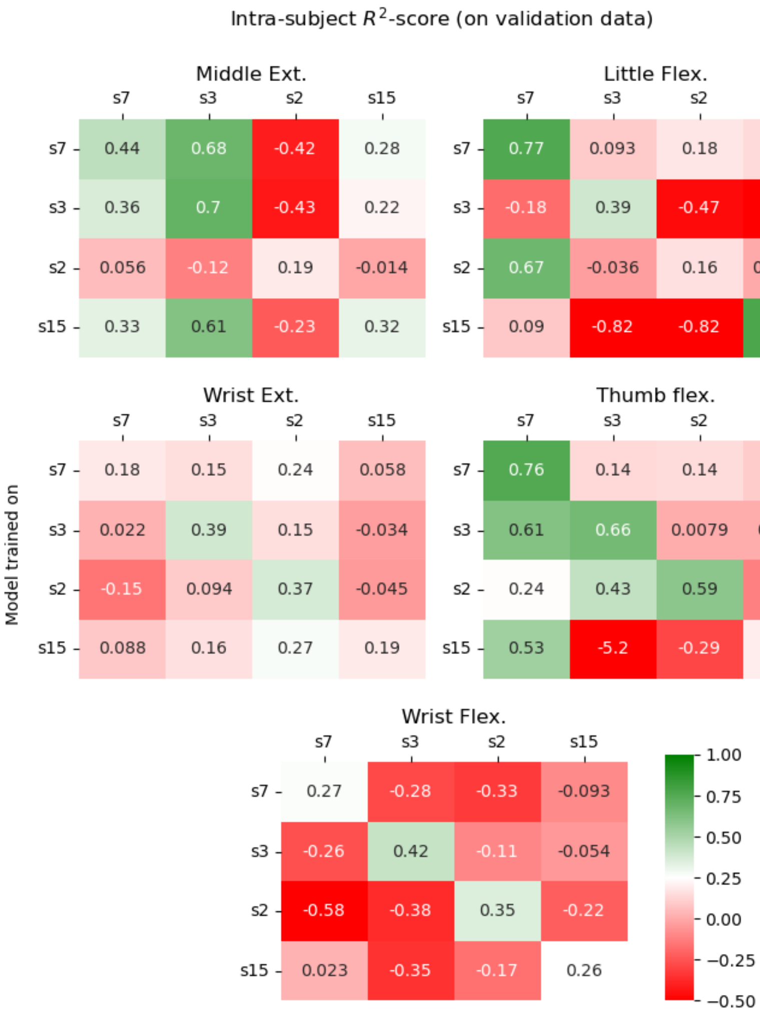


Figure 12: All Movement Tasks, Multi-channel (Elastic net) models validated on other subjects using R2 as performance measure. The purpose of the heatmaps is to get an idea of performance and generalizability between subjects on validation data. Note that the subject data from which the model is trained stated on the rows of the heatmaps and the columns state from which subject the validation data comes from.

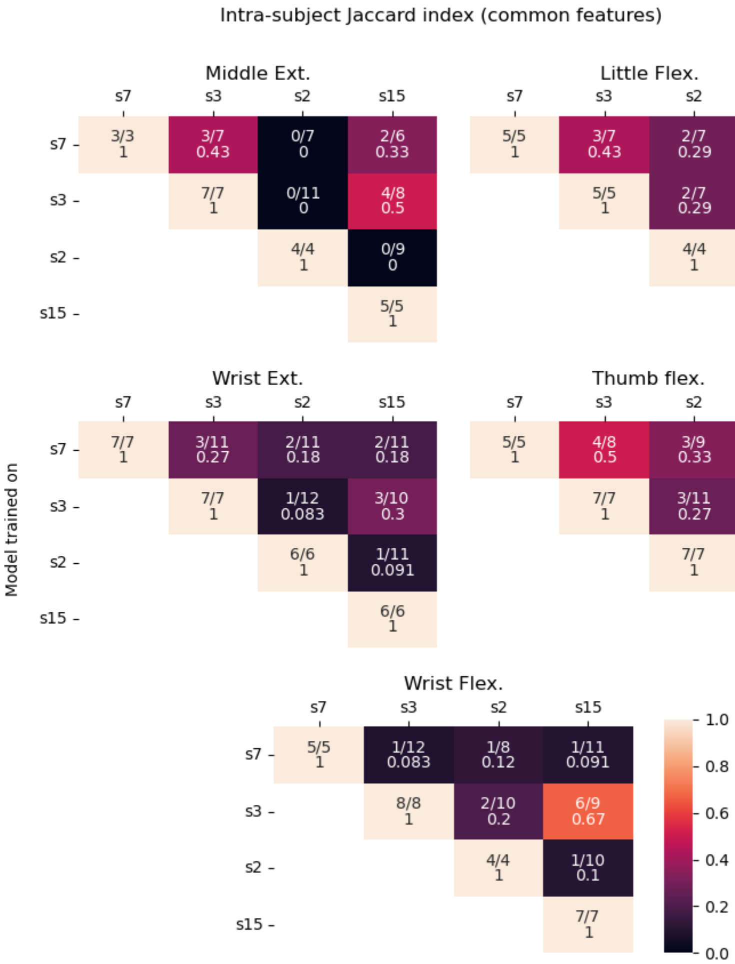
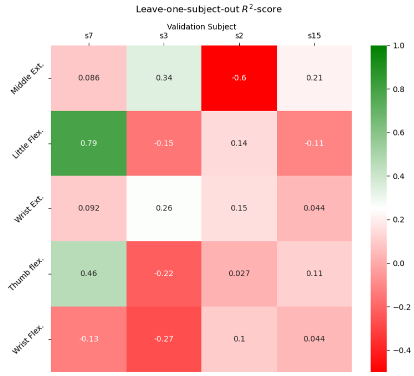
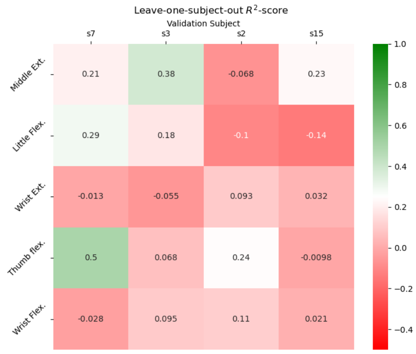


Figure 13: All Movement Tasks, Jaccard index matrices for single subject models, showing common independent variables selected by the elastic net regularization between subjects.

Figure 14: All Movement Tasks, Single-channel models validated on other subjects using R2 as performance measure. The purpose of this figure is to see how single-channel performs and generalizes compared to multi-channel models in figure 8. Similarly to aforementioned figure, the subject data from which the model is trained stated on the rows of the heatmaps and the columns state from which subject the validation data comes from.



(a) Elastic Net models



(b) Single-channel models

Figure 15: All Movement Tasks, Leave-one-subject-out validation heatmaps. Rows indicate which force type the model estimates, while columns represent the validation subject. For instance, the top-left square presents the R2-score for Middle finger extension force estimated using validation data from subject 7. Thus, the model was trained on data from subjects s3, s2, and s15, but excluded data from subject 7.

5 Discussion

5.1 Reflections on Methods

In order to delimit the thesis, only one feature extraction type, MAV, was used. However, it's debatable whether this is the optimal feature for iEMG. Further analysis could explore if there's a more physiologically accurate feature that better represents the iEMG-force dynamics. Additionally, employing multiple features or a fusion of several features might extract more relevant information from the raw iEMG signal for force estimation. For instance, combining a time domain feature with a frequency domain feature could be beneficial.

Another delimitation was the exclusive usage of R2 as performance measure in the results. Using an additional measure such as Variance, would provide a more comprehensive understanding. While R2 provides good overview of how well the model performs compared to the ground truth, a combined R2 and Variance measure could provide more information about the consistency of the model's estimations. However, the exclusive use R2 was in part a conscious choice to not overload the reader with too much information.

Additionally, in the method used for baseline wander removal, specifically the high-pass filter approach, an ideal filter of a higher order could have been constructed. After all, there are no time constraints on the preprocessing of the force signals. A high-order high pass filter with steep cut-off could potentially leave the frequencies of interest entirely undistorted.

5.2 Specific Movement Task Modeling

The elastic net regression model based on polynomial iEMG MAV features used in this project shows mixed results for estimating a concurrent 1-DOF, isometric force produced by a finger or wrist, as seen in figure 8. If a complex multichannel model like this was to be used, the model would need to be trained on same subject, due to the wildly different result when generalizing to new subjects. Perhaps contralateral (opposite side arm) training on an amputee subject is something that could be explored to calibrate model for an amputee subject. This however assumes that the opposite side arm have very similar iEMG-force-dynamics to the amputated side.

Generally, the elastic net model performs better than single channel model in figure 9, when trained and validated on the same subject, but fails to reliably deliver when generalizing the model to new subjects and can at times produce extremely poor results with very large negative R2-scores. On average, the single-channel models seem to generalize better to new subjects, indicating that there is a pattern of overfitting to a particular subject with the elastic net models.

Even in the correlation matrix analysis in figure 7, it is apparent that there is large differences between subjects, and generalizability is probably difficult to achieve. There was a slight, perhaps naive, chance that the interaction terms would be able to correct for this, but results indicate that it was not the case.

The jaccard index in figure 10, measuring overlapping features was very small, indicating that there is no consistent polynomial interaction feature set that describes the underlying relationship between force and iEMG features, that generalizes across subjects. While there often is only 1-2 overlapping variables among the models, these few shared variables are likely the largest coefficients and represent the primary related iEMG-to-force channel. But overall, it seems that there are a lot of selected variables that are specific to the individual subject. In future studies or applications, a polynomial model with interaction terms could be used to improve intra-subject estimates, but if the goal is merely improved prediction score, a more advanced method such as an artificial neural network would be more suitable.

The leave-one-subject-out R2-scores of the multichannel models are generally worse than single channel models, as seen in figure 11, It is not surprising due to the poor generalizability seen previously. In figure 8 the results are very varied, it can perform great on some subjects, but extremely poorly on others. The pooled model of using data from multiple subjects seem to perform in the middle ground of these two extremities, selecting a set of variables that does not improve the result on the good subject pairings in figure 8, but does not overfit to specific subjects as much.

5.3 All Movement Tasks Modeling

All the models in general show poor results when all movement tasks are included. When rest stages are included in the training data, the iEMG feature and the related finger and wrist force can become widely uncorrelated, which shows in the results. For both the single-channel and multi-channel models with interaction terms, the results are poor. For some subjects, e.g subject 7, the R2-score can perform well on it's own validation data, but it is the exception rather than the rule. The leave-one-subject-out R2-scores does not improve upon much either. Additionally, considering the poor R2 performances, it is not surprising that the jaccard indexes show high dissimilarity as well.

Upon deeper analysis, the poor R2-scores may not be completely unexpected results. If

FDP and EDC are respectively responsible for flexion and extension of fingers D2-D5 as said in [9], then it is actually expected for them to not correlate over all tasks performed. Also, if the idea of muscle synergies is correct and some muscles acts as secondary or supporting for unrelated 1-DOF forces, then one would also expect them not to remain completely inactive when other tasks are performed. This is actually supported by relatively good results in Section 4.2, in comparison to Section 4.3.

A Future model that could be made to improve performance in this type of setting, where many 1-DOF tasks performed sequentially, would be a model could simultaneously classify the type of task performed and then use the Specific Task model to estimate the force produced by the task.

5.4 Model Improvements

There are several potential improvements that could be explored to further enhance performance and address certain limitations of the current elastic net model.

- Threshold Models: An improvement worth considering is the use of threshold models. These models incorporate a non-linear output function, such as the activation function, i.e. max(0, output), to account for any thresholds or limits in the dependent variable. In the case of the force data, where there is a lower limit, e.g. extension force measurements can only take on positive values, a threshold model using a non-linear output function can capture this characteristic and potentially improve the model's performance. However, using a non-linear function generally make the optimization and interpretability of coefficients more difficult.

  The simplest solution is to just threshold the outputs from the model post-hoc, however, ideally one would want a model that "understands" the dynamics of the variables and adjusts the coefficients accordingly. This simple post-hoc method would likely yield a slight increase in performance.

- HAC-Adjusted Standard Errors [23]: To obtain more accurate coefficient uncertainties, utilizing Heteroscedasticity and Autocorrelation Consistent (HAC) standard errors can be beneficial. HAC-adjusted standard errors account for heteroscedasticity and autocorrelation in the residuals and provide more reliable estimates of the coefficient uncertainties, by considering the temporal dependence and heteroscedasticity in the model errors.

- Generalized Linear Models (GLMs) [24]: GLMs allow for the specification of different error distributions and link functions to better capture the underlying characteristics of the data. This would require more rigorous study in what link function could more accurately describe the iEMG-force-relationships or is more physiologically sound. In my tangential research I was not able to find one which performs better than the linear.

- Time-dynamic models : While models utilizing preceding force values to predict future values, might not be possible due to the physical limitation of feedback mechanisms in amputees, compared to other things like temperatures or stocks indexes. If we were hypothetically given a continuous stream of force data with high sampling frequency, time-dynamic models could prove effective, since force changes tend to be smooth and not super sudden and that subsequent force values tend to remain close to preceding ones.

References

[1] Alexander E. Olsson, Nebojša Malešević, Anders Björkman, and Christian Antfolk. End-to-end estimation of hand- and wrist forces from raw intramuscular emg signals using lstm networks. *Frontiers in Neuroscience*, 15, 2021.

[2] Denise Berger and Andrea D'Avella. Effective force control by muscle synergies. *Frontiers in computational neuroscience*, 8, 2014.

[3] Alessandro Scano, Andrea Chiavenna, Lorenzo Molinari Tosatti, Henning Müller, and Manfredo Atzori. Muscle synergy analysis of a hand-grasp dataset: A limited subset of motor modules may underlie a large variety of grasps. *Frontiers in Neurorobotics*, 12, 2018.

[4] Nebojša Malešević, Anders Björkman, Gert Andersson, Christian Cipriani, and Christian Antfolk. Evaluation of simple algorithms for proportional control of prosthetic hands using intramuscular electromyography. *Sensors (Basel, Switzerland)*, 22, 07 2022.

[5] Ernest Nlandu Kamavuako, Dario Farina, Ken Yoshida, and Winnie Jensen. Relationship between grasping force and features of single-channel intramuscular emg signals. *Journal of Neuroscience Methods*, 185:143–150, 2009.

[6] Ping Zhou and William Rymer. Factors governing the form of the relation between muscle force and the emg: A simulation study. *Journal of neurophysiology*, 92:2878–86, 2004.

[7] S. Metral and G. Cassar. Relationship between force and integrated emg activity during voluntary isometric anisotonic contraction, 1981.

[8] Ernest Kamavuako, Dario Farina, Ken Yoshida, and Winnie Jensen. Estimation of grasping force from features of intramuscular emg signals with mirrored bilateral training. *Annals of biomedical engineering*, 40:648–56, 03 2012.

[9] Nebojsa Malesevic, Anders Björkman, Gert S. Andersson, Ana Matran-Fernandez, Luca Citi, Christian Cipriani, and Christian Antfolk. A database of multi-channel intramuscular electromyogram signals during isometric hand muscles contractions. *Scientific Data*, 7(10), 2020.

[10] Andreas Jakobsson. *An introduction to Time series modeling*. Studentlitteratur, 3rd edition, 2019.

[11] Ernest Kamavuako, Jakob Rosenvang, Mette Bøg, Anne Smidstrup, Ema Petersen, Marko Niemeier, Winnie Jensen, and Dario Farina. Influence of the feature space on the estimation of hand grasping force from intramuscular emg. *Biomedical Signal Processing and Control*, 8:1–5, 2013.

[12] Adi Bronshtein. Simple and multiple linear regression in python, 2018.

[13] UCLA Institute for Digital Research and Education. Regression with stata chapter 2 – regression diagnostics, 2023.

[14] Deanna Schreiber-Gregory and Henry M Jackson. Regulation techniques for multicollinearity: Lasso, ridge, and elastic nets, 2018.

[15] Aarshay Jain. A complete tutorial on ridge and lasso regression in python, 2016.

[16] Statsmodels Developers. Linear regression regularization, 2022.

[17] Abhigyan. R-squared and adjusted r-squared, 2023.

[18] Python Software Foundation. Python, 2021.

[19] SciPy developers. scipy.signal.filtfilt, 2023. Online documentation for the filtfilt function in the SciPy library.

[20] Daniele Borzelli, Sergio Gurgone, Paolo De Pasquale, Nicola Lotti, Andrea d'Avella, and Laura Gastaldi. Use of surface electromyography to estimate end-point force in redundant systems: Comparison between linear approaches. *Bioengineer*, 10(5):234–258, 2023.

[21] Skipper Seabold and Josef Perktold. *Statsmodels: Econometric and Statistical Modeling with Python*, 2010. Proceedings of the 9th Python in Science Conference.

[22] Scikit. Gridsearchcv: Exhaustive search over specified parameter values for an estimator, 2021.

[23] Christoph Hanck, Martin Arnold, Alexander Gerber, and Martin Schmelzer. *Introduction To Econometrics with R*. 2023. Online Book.

[24] Yuho Kida. Generalized linear models, 2019.