OPTIMAL EMG PLACEMENT FOR A ROBOTIC PROSTHESIS CONTROLLER WITH SEQUENTIAL, ADAPTIVE FUNCTIONAL ESTIMATION (SAFE)

By Jonathan Stallrich*, Md Nazmul Islam*, Ana-Maria Staicu*, Dustin Crouch†, Lizhi Pan‡ and He Huang‡

Department of Statistics, North Carolina State University*, Department of Mechanical, Aerospace, and Biomedical Engineering, University of Tennessee, Knoxville†, and Department of Biomedical Engineering, North Carolina State University‡

Robotic hand prostheses require a controller to decode muscle contraction information, such as electromyogram (EMG) signals, into the user’s desired hand movement. State-of-the-art decoders demand extensive training, require data from a large number of EMG sensors, and are prone to poor predictions. Biomechanical models of a single movement degree-of-freedom tell us that relatively few muscles, and hence fewer EMG sensors, are needed to predict movement. We propose a novel decoder based on a dynamic, functional linear model with velocity or acceleration as its response and the recent past EMG signals as functional covariates. The effect of each EMG signal varies with the recent position to account for biomechanical features of hand movement, increasing the predictive capability of a single EMG signal compared to existing decoders. The effects are estimated with a multi-stage, adaptive estimation procedure we call Sequential Adaptive Functional Estimation (SAFE). Starting with 16 potential EMG sensors, our method correctly identifies the few EMG signals that are known to be important for an able-bodied subject. Furthermore, the estimated effects are interpretable and can significantly improve understanding and development of robotic hand prostheses.

1. Introduction. Over 160,000 Americans are transradial (i.e. below-elbow) amputees, henceforth TRAs, and must learn how to perform tasks without their intact hand (Ziegler-Graham et al., 2008). Passive hand prostheses are useful but cannot fully emulate the functionality of an intact hand. State-of-the-art multifunctional robotic prostheses, such as the FDA-approved DEKA arm system (Resnik, 2011; Resnik et al., 2011), have more advanced mechanical systems that move according to a prosthesis controller.
(PC) that decodes user inputs into movement. These inputs are often electromyogram (EMG) signals because they measure activation of muscles due to contractions, which are known to cause hand movement for an able-bodied (AB) subject.

Figure (1) summarizes the movement process for an AB subject and a TRA equipped with a robotic prosthesis. For both AB subjects and TRAs, intended hand movement begins with initiation of action potentials that are conducted along motor neurons from the central nervous system. The neural signals travel along the motor neural pathway and propagate onto the muscle fibers, causing the muscle to contract. The action potential measured from the muscle fibers is called motor unit action potential (MUAP). An EMG signal taken from a surface electrode placed on an intact or residual forearm represents the sum of MUAPs across one or more forearm muscles and conveys information about the magnitude and duration of muscle contractions. For an AB subject, the contractions stretch tendons connected to bones in the hand, leading to hand movement. These physical connections no longer exist for a TRA, but TRAs can still activate their residual muscles and may sense movement in their missing limb (Mercier et al., 2006).

Two important and related questions are: (1) how many EMG sensors are needed to capture relevant muscle contraction information and (2) how should the PC decode the EMG data to produce movement? A traditional direct myoelectric PC uses two EMG sensors corresponding to antagonistic muscles for control of a single movement along a degree-of-freedom (DOF), such as wrist flexion/extension. To determine the sensor locations, prosthetists first palpate a muscle while the TRA contracts their muscles to
attempt movement along a single DOF. The EMG signal on the targeted skin surface areas is visually checked by probing different locations with a bipolar electrode. A chosen EMG sensor location should be independent of a co-activating antagonistic muscle. Unfortunately, it is difficult to use these PCs to operate multiple DOFs intuitively (Resnik et al., 2018).

EMG pattern recognition (PR) is the state-of-the-art decoding method for prosthesis control (Scheme and Englehart, 2011). PR assumes that the varying patterns of multi-channel EMG signals can be decoded to different hand/wrist motions. In both clinical practice and research settings, the placement of EMG sensors is not muscle specific because EMG cross-talks (i.e., signals detected from nearby muscles) do not significantly affect decoding performance (Resnik et al., 2018). Rather, the amount and variety of neural information recorded in EMG signals determines the accuracy for classifying motions. To ensure that sufficient neural information is captured, studies have used redundant (Huang et al., 2010) or high-density electrodes (Zhou et al., 2007) to saturate the surface of the residual limb. However, this electrode saturation comes with costs: (1) it requires frequent, lengthy calibration, (2) its implementation increases the challenges in computer hardware design for data streaming and real-time EMG processing; and (3) its prediction performance may degrade due to noise introduced by additional sensor information.

Unlike EMG PR, an EMG-driven musculoskeletal model decodes the EMG input into movement based on human physiology (Crouch and Huang, 2016). For these models, saturation of the EMG recording surface does not necessarily increase the neural information captured. In fact, there are only 20 muscles in the human forearm (O’Rahilly and Müller, 1983). If muscle-specific EMG recordings can be captured, the PC decoder needs fewer electrodes. Crouch and Huang (2017) proposed a low-dimensional PC decoder based on a planar link-segment dynamic model that includes only four virtual muscles while still incorporating relevant knowledge of the biomechanical system. Their model accurately predicted continuous wrist and finger movement with only four independent EMG signals, one for each virtual muscle. However, selection of the relevant EMG sensors remains an important and critical problem for TRAs as many muscles are inaccessible to external EMG sensors due to muscle loss or depth of the muscle in the arm.

This paper introduces a novel EMG-based decoder and functional estimation procedure that is selective of the EMG without sacrificing prediction performance. The velocity or acceleration of a given movement DOF is predicted by multiple functional covariates corresponding to the recent past behavior EMG signals, whose effects can vary with the recent DOF’s posi-
tion. EMG selection and estimation is done simultaneously through our proposed Sequential Adaptive Functional Estimation (SAFE) procedure that uses multiple, adaptive stages of a penalized fitting criterion.

The paper is organized as follows. Section 2 describes the collection and processing of the EMG and movement data and provides motivation for our proposed decoder. Section 3 briefly reviews current EMG-based PCs and highlights their deficiencies. Section 4 details the dynamic, functional linear model that makes up our decoder and Section 5 details the SAFE approach. Section 6 presents the analysis results from data collected from an AB subject, including EMG selection, effect interpretations, and prediction performance. A simulation study is performed in Section 7, showing the robustness of our method to varying covariance assumptions. Section 8 concludes the paper with a discussion of impacts and extensions.

2. Data collection and processing. EMG and movement data were collected from an AB subject’s right limb. Figure 2 shows the placement of 15 EMG surface electrodes on the subject’s limb and labels those electrodes placed near muscles known to contribute to the movement DOFs of interest: metacarpophalangeal (MCP) flexion/extension and wrist flexion/extension. For simplicity, we refer to MCP flexion/extension as finger flexion/extension, because it involves simultaneous movement of all fingers excluding the thumb. To provide a baseline for selection performance, we also generated an external signal unrelated to movement, giving a total of 16 signals, denoted by $X_1, \ldots, X_{16}$ with $X_9$ being the external signal.

![Fig 2. EMG sensor labels differentiated by movement DOF contribution. EMG label 7 measures a muscle that contributes to both finger and wrist extension and so has a combined symbol. The posture used for data collection is shown in the bottom right.](image)

The EMG signals were normalized between 0 (no contraction) and 1 (maximal contraction) based on previous training. They then were high-pass filtered at 40 Hz, rectified, and low-pass filtered at 6 Hz using a 4th order Butterworth zero-phase filter (Butterworth, 1930). Movement data were collected by reflective markers placed on 9 anatomical locations on the forearm, wrist, and hand, and were recorded at 120 Hz using an infrared
motion capture system (Vicon Motion Systems Ltd., UK). The movement data were filtered at 6 Hz using a 4th order Butterworth filter and the joint angles (in degrees) were calculated from a musculoskeletal model (Holzbaur et al., 2005) in OpenSim (Delp et al., 2007).

EMG and joint angle data were collected synchronously for approximately 30 seconds while the subject performed basic hand movements in a fixed arm posture shown in Figure 2. Isolated, single DOF movements were performed following either a consistent pattern or a random pattern. The movements during a random pattern were determined by the AB subject. The consistent pattern had the subject alternate between performing maximal finger flexion and extension. Before performing the opposite movement, they allowed their hand to return to a neutral position by relaxing their muscles. There were six independent data sets for isolated finger movement, three observing consistent movement patterns and three observing random movement patterns. The same was done for isolated wrist movement.

Figure 3 is a snapshot of synchronous data for a consistent finger movement pattern. Prior to Figure 3(a), $X_7$’s signal concurrently increases as the fingers extend. In Figure 3(a), the fingers are at maximal extension, while $X_7$’s signal decreases, meaning the subject is relaxing the corresponding muscle. Figure 3(b) has no concurrent muscle contractions while the fingers flex towards a neutral state of 20 degrees. This type of movement is attributable to passive forces generated by muscle relaxation. Note the time delay between the muscle’s relaxation in Figure 3(a) and the resulting passive force movement in Figure 3(b).

**Fig 3.** Joint finger position (degrees) for extension and flexion movements and normalized EMG signals for $X_7$ and $X_{12}$. Event (a) demonstrates restricted movement due to maximal finger extension. Event (b) demonstrates movement due to passive forces, lacking concurrently active EMG signals.
Figure 3 is convincing that EMG data can predict movement, but omits all the potential EMG signals that complicate EMG selection. The left plot in Figure 4 shows all 15 EMG signals (ignoring $X_9$) for the same time period as in Figure 3 and the right plot shows the concurrent correlations among all 16 signals, calculated across the entire 30-second data window. Between 0.5 and 1.5 seconds, the dominant signal belongs to $X_7$, but many other signals are active as well, such as $X_5$. The high concurrent correlations indicate redundant EMG information and the presence of latent signals that complicate EMG selection. The next section discusses how decoders in current EMG-based PCs address these issues.

**Fig 4.** Overlay plot (left) of 15 EMG signals during one instance of finger extension and flexion observed and concurrent correlation plot (right) between all 16 signals for the entire data set.

3. Current EMG-based control strategies. The $i$-th instance ($i = 1, \ldots, N$) of data collection includes $K$ measured and processed EMG signals, $x_{ik}$ ($k = 1, \ldots, K$), and a measure for each movement DOF of interest. For example, we let $z_i$ denote the movement DOF position. It is common to divide the data into time windows and summarize the EMG and movement data within each window. In some cases, movement in each window is assigned to a movement category, such as flexion or extension. The $x_{ik}$'s are often converted into a feature set summarizing its behavior in the time window, such as the moving average or slope (Hargrove et al., 2007), and these feature sets are the decoder’s inputs instead of the original $x_{ik}$'s.

Hahne et al. (2014) simultaneously predicted multiple wrist movement DOFs with linear and nonlinear regression (e.g., mixture of linear experts and kernel ridge regression) based on feature sets from 192 EMG signals. They compared performance across different subsets of the EMG sensors and determined fewer signals produced comparable predictions to the full set, but no formal variable selection was performed. Scheme and Englehart
(2011) summarized and compared several statistical PR classifiers (e.g., \( K \)-nearest-neighbors, support vector machine, quadratic discriminant analysis) on EMG/movement data and discussed challenges faced by these classifiers. Both papers noted the importance of the amount and type of EMG information in the feature set for a decoder’s prediction performance.

Jiang et al. (2009) and Jiang et al. (2014) proposed a generative model to incorporate underlying neurophysiological processes largely ignored by PR approaches. The \( j \)-th movement DOF \((j = 1, \ldots, M)\) was assigned two opposing latent control signals representing positive and negative directions (e.g., extension/flexion), denoted by \( c^p_{ij} \) and \( c^n_{ij} \), respectively. Let \( \mathbf{x}_i = (x_{i1}, \ldots, x_{iK})^T \) be the vector of \( K \) EMG signals at instance \( i \). Equating \( \mathbf{x}_i = \mathbf{W} \mathbf{c}_i \) where \( \mathbf{W} \) is the \( K \times 2M \) synergy matrix and \( \mathbf{c}_i = (c^p_{i1}, c^n_{i1}, \ldots, c^p_{iM}, c^n_{iM})^T \), they solved for \( \mathbf{W} \) and \( \mathbf{c}_i \) requiring both to have nonnegative elements. Their approach does not require any measured movement data, but restricts the DOF of movements the subject performs to help identify the control signals during the calibration session.

Crouch and Huang (2016) developed a decoder based on a planar lumped-parameter musculoskeletal model for the two DOFs considered in this paper. The decoder reliably predicted continuous movement with only four EMG signals corresponding to muscles known to contribute to these movements for AB subjects. The results relied on their prior knowledge of the important muscles and their ability to target them with the four surface electrodes. This knowledge is unknown for a given TRA whose musculoskeletal structure has been significantly altered and it is unclear how the important EMG signals would be selected for a TRA. We next describe a new class of EMG-based decoders based on a flexible statistical model that can recover many of the biomechanical features directly incorporated in the musculoskeletal model in Crouch and Huang (2016) and lends itself to an EMG selection method.

4. Proposed decoder and penalized estimation. As demonstrated in Figure 3, movement can be influenced by concurrent and recent past changes in \( X_k \). The effect may also depend on the position, as observed in Figure 3(a) in which there is no change in position despite a decreasing EMG signal. Hence our proposed decoder is based on a functional linear model with a scalar response, \( y_i \), and \( K \) functional predictors defined on \( S \) whose effects can vary with covariate, \( z_i \in Z \),

\[
E[y_i | X_{i1}, \ldots, X_{iK}, z_i] = \sum_{k=1}^{K} \int_S X_{ik}(s) \gamma_k(s, z_i) ds .
\]
Here \( \gamma_k(\cdot, \cdot) \) is an unknown bivariate function defined on \( S \times Z \) that quantifies the effect of the \( k \)-th functional covariate, briefly written as \( X_{ik} \), on the mean of \( y_i \) conditional on \( z_i \). When possible, we shorten \( \gamma_k(\cdot, \cdot) \) as simply \( \gamma_k \).

For model (4.1) to be used as an EMG-based decoder, \( y_i \) is set to either movement velocity or acceleration, \( z_i \) is the movement position, and \( X_{ik} \) is a curve describing the recent past behavior of the \( k \)-th EMG signal. Each \( X_{ik} \) is defined on \( S = [-\delta, 0] \), where \( s = 0 \) refers to the concurrent time relative to the velocity \( y_i \) and \( \delta \) is a pre-determined historical length. Consistent with our application, each \( X_{ik} \) is assumed to be measured at \( \delta + 1 \) equally-spaced time points, \( X_{ik}(j) = x(i+j)k \) for \( j = -\delta, \ldots, 0 \). As in Section 3, we window the data; however, we do not reduce the EMG data into a feature set. Section 6.1 and Section A of the Supplementary Materials details how the \( y_i \) values were estimated from the observed \( z_i \) and how the \( X_{ik} \) were created from the EMG data for this application.

With \( y_i \) being the first or second derivative of \( z_i \), model (4.1) is a forced differential equation system with nonlinear forcing functions for each EMG signal (see Ramsay and Hooker (2017)). The covariate-varying effects in the forcing functions are a direct extension of the functional linear model described in Cardot et al. (2003); Ramsay and Silverman (2005); Goldsmith et al. (2011); Ferraty et al. (2012); McLean et al. (2014) and have been recently applied to functional data in Cardot and Sarda (2008); Wu et al. (2010); Davenport (2013). In their Section 1.1.6, Ramsay and Hooker (2017) analyzed handwriting movements of Chinese characters by modeling the acceleration with a harmonic oscillator, but they did not use EMG-based forcing functions. Both velocity and acceleration models are considered in Section 6, as well as forcing functions that do not depend on position.

### 4.1. Model approximation

Following Wood (2006) and Eilers and Marx (2003), we approximate \( \gamma_k \) with a tensor product of two finite dimensional univariate basis functions, \( \{\omega_l(\cdot)\} \) and \( \{\tau_m(\cdot)\} \), defined on \( S \) and \( Z \), respectively. We use the same bases for all \( \gamma_k \), so \( \gamma_k(s, z) = \sum_{l=1}^{L} \sum_{m=1}^{M} \omega_l(s) \tau_m(z) \beta_{klm} \), where the \( \beta_{klm} \)'s are basis coefficients for the \( k \)-th functional covariate’s effect that need to be estimated. In matrix form, \( \gamma_k(s, z) \approx \omega^T(s) B_k \tau(z) \) where \( \omega^T(s) = (\omega_1(s), \ldots, \omega_L(s)) \), \( \tau^T(z) = (\tau_1(z), \ldots, \tau_M(z)) \), and \( B_k = (\beta_{klm}) \) is an \( L \times M \) coefficient matrix.

Approximate the integrals in (4.1) by

\[
\int_S X_{ik}(s) \gamma_k(s, z_i) ds \approx \left\{ \sum_{j=-\delta}^{0} x(i+j)k \omega(j)^T \right\} B_k \tau(z_i) = X_{ik\omega}^T B_k \tau(z_i) .
\]
Define $X_{ik\omega T}^T = X_{ik\omega}^T \otimes \tau(z_i)$, where $\otimes$ is the Kronecker product operator. For notational simplicity, denote $X_{ik\omega T}$ by $\tilde{X}_{ik}^T$ and let $\beta_k$ denote the vectorized version of $B_k$. Then we approximate model (4.1) by

\begin{equation}
E[y_i|\tilde{X}_{i1}, \ldots, \tilde{X}_{iK}] \approx \sum_{k=1}^{K} \tilde{X}_{ik}^T \beta_k,
\end{equation}

so that estimation of $\gamma_k$ is done through estimation of $\beta_k$.

4.2. **Penalized estimation.** It is necessary to regularize estimation of $\beta_k$ to prevent overfitting and encourage interpretable $\hat{\gamma}_k$. We further desire the estimation procedure to force some estimates to zero to determine a subset of important functional covariates. Matsui and Konishi (2011); Gertheiss et al. (2013); Fan et al. (2015); Pannu and Billor (2017) have discussed functional variable selection with scalar-on-functional regression models. Our proposed approach is inspired by these ideas and accommodates more complex regression coefficients.

Following Gertheiss et al. (2013), we adopt the penalized least squares approach for estimation that simultaneously induces sparsity and controls smoothness of the $\gamma_k$'s. For simplicity, assume $\{\omega_l()\}_{l=1}^L$ and $\{\tau_m()\}_{m=1}^M$ are orthogonal B-spline bases. We allow $L$ and $M$ to be sufficiently large to capture the complexity of the regression surfaces and penalize the degree of smoothness. We measure smoothness through the total curvature in each direction with $||\gamma''_{k,s}||$ and $||\gamma''_{k,z}||$, where $||\gamma''_{k,s}||^2 = \int_S \int_Z (\gamma''_{k,s}(s,z))^2 \, dz \, ds$ and $\gamma''_{k,s} = \partial^2 \gamma_k / \partial s^2$. Sparsity results from penalizing the effect's total magnitude, measured by $||\gamma_k||$. By orthogonality of the bases,

\begin{equation}
||\gamma_k||^2 = \beta_k^T \beta_k, \quad ||\gamma''_{k,s}||^2 = \beta_k^T (\Omega_s \otimes I_M) \beta_k, \quad ||\gamma''_{k,z}||^2 = \beta_k^T (I_L \otimes \Omega_z) \beta_k,
\end{equation}

where $\Omega_s = \int \omega''(s)\omega''(s)^T \, ds$ and $\omega''()$ is the vector of second derivatives $\omega''(),$ with a similar definition for $\Omega_z$. The three norms in (4.4) are combined into the penalty term

\begin{equation}
P_\phi(\gamma_k) = (||\gamma_k||^2 + \phi_s||\gamma''_{k,s}||^2 + \phi_z||\gamma''_{k,z}||^2)^{1/2} = (\beta_k^T Q_\phi \beta_k)^{1/2},
\end{equation}

where $Q_\phi = I_{LM} + \phi_s (\Omega_s \otimes I_M) + \phi_z (I_L \otimes \Omega_z)$ is positive-definite and the tuning parameters $\phi = (\phi_s, \phi_z) > 0$ control the smoothness of $\gamma_k$ in the direction $s$ and $z$. For instance, large values of $\phi_s$ force $\hat{\gamma}_k$ to be marginally linear in $S$. We estimate $\beta_k$ by minimizing the criterion

\begin{equation}
\frac{1}{N} \sum_{i=1}^{N} \left( y_i - \sum_{k=1}^{K} \tilde{X}_{ik}^T \beta_k \right)^2 + \lambda \sum_{k=1}^{K} (\beta_k^T Q_\phi \beta_k)^{1/2},
\end{equation}
where $\lambda > 0$ controls the sparsity of the $\gamma_k$'s.

Let $R_\phi$ be the Cholesky factor of $Q_\phi$, i.e., $Q_\phi = R_\phi R_\phi^T$, and set $\overline{\beta}_k = R_\phi^T \beta_k$ and $W_{ik} = R_\phi^{-1} \overline{X}_{ik}$. The criterion (4.6) can be written as

$$\sum_{i=1}^{N} (y_i - \sum_{k=1}^{K} W_{ik}^T \overline{\beta}_k)^2 + \lambda \sum_{k=1}^{K} ||\overline{\beta}_k||_2,$$

which for a fixed $\phi$ is the group LASSO criterion described in Yuan and Lin (2006) and Yang and Zou (2013, 2015). Minimizing (4.7) can be done efficiently using the groupwise-majorization-descent algorithm (Yang and Zou, 2013) for given $\lambda$ and $\phi$. Rebuilding $\overline{B}_k$ from $\overline{\beta}_k = (R_\phi^T)^{-1} \overline{\beta}$, we estimate $\gamma_k$ with $\overline{\gamma}_k(s, z) = \omega(s)^T \overline{B}_k \tau(z)$.

### 4.3. Selection of $\lambda$ and $\phi$.

The optimal tuning parameters $\pi^* = (\lambda^*, \phi_s^*, \phi_z^*)$ are chosen empirically by comparing the prediction performance of the minimizer of (4.7) across multiple settings in the tuning parameter space, e.g., $K$-fold cross-validation (CV). For longitudinal data, like this data application, $K$-fold block CV is recommended, in which the data are not partitioned randomly but rather into $K$ equally-sized sequential sections (Roberts et al., 2017). CV is performed across a set of $\pi$'s, each producing a set of estimates, $\{\hat{\gamma}_{k, \pi}\}$ and an average prediction error ($APE_{\pi}$) across the folds. Let $\pi_{\text{min}}$ be the $\pi$ having $APE_{\text{min}}$, the smallest $APE_{\pi}$. The $\pi_{\text{min}}$ is one choice for the optimal $\pi$, but we recommend a version of the one-standard-error rule (Friedman et al., 2001; Krstajic et al., 2014; Yang and Zou, 2015). First, gather the considered $\pi$'s whose corresponding $APE_{\pi} \leq APE_{\text{min}} + SE_{\text{min}}$, where $SE_{\text{min}}$ is the estimated standard error of $APE_{\text{min}}$. From this subset, chose the $\pi$ whose estimates minimize $\sum_k P_\phi(\hat{\gamma}_{k, \pi})$ with $\phi = (1, 1)$, being a measure of model complexity balancing sparsity and smoothness of the $\gamma_k$'s.

### 5. Sequential, Adaptive Functional Estimation (SAFE).

The fitting criterion (4.6) does not allow for different shrinkage and smoothness for the functional predictors, which may inflate the number of false positives. Adaptive estimation is one way to improve variable selection performance (Zou, 2006; Meier, 2009; Gertheiss et al., 2013; Guo et al., 2015; Ciuperca, 2016; Ivanoff et al., 2016). This section describes a sequential, adaptive fitting procedure that is generally applicable to linear models fit using a group LASSO-type penalty.

#### 5.1. Adaptive weight penalty.

Adaptive LASSO is an iterative fitting procedure that translates preliminary estimates $\hat{\gamma}_k$ into weights that are incorporated into penalty (4.5) to improve variable selection performance.
Here, the $k$-th functional covariate is assigned the positive weight vector $w_k = (f_k, g_k, h_k)$, where $f_k$ weights the sparsity penalty $||\gamma_k||^2$ and $g_k$ and $h_k$ weight the smoothing penalties $||\gamma''_{k,s}||^2$ and $||\gamma''_{k,z}||^2$, respectively. We set the weights to $f_k = 1/||\gamma_k||^{1/2}$, $g_k = 1/||\gamma''_{k,s}||^{1/2}$, and $h_k = 1/||\gamma''_{k,z}||^{1/2}$. The adaptive penalty function is denoted

\[(5.1) \quad P_{\phi,w}(\gamma_k) = (f_k||\gamma_k||^2 + g_k\phi_s||\gamma''_{k,s}||^2 + h_k\phi_z||\gamma''_{k,z}||^2)^{1/2}.\]

The weighted penalty for $\gamma_k$ can be efficiently calculated by $P_{\phi,w}(\gamma_k) = (\beta_k^TQ_{\phi,w}\beta_k)^{1/2}$, where $Q_{\phi,w} = f_kI_{LM} + g_k\phi_s(\Omega_s \otimes I_M) + h_k\phi_z(I_L \otimes \Omega_z)$. Constructing (4.7) with $Q_{\phi,w}$ instead of $Q_{\phi}$ incorporates these weights in the estimation procedure.

Selection performance of the adaptive group LASSO procedure depends on the accuracy of the $\hat{\gamma}_k$'s (Zou, 2006). Initial estimation by minimizing (4.7) without penalizing $||\gamma_k||^2$ (Gertheiss et al., 2013) may produce poor estimates for this data application due to high correlation between the functional predictors (see Figure 4) and dependency of the $X_{ik}$ distributions on $z_i$ (see Section A of the Supplementary Materials). Instead, we recommend initial estimates by optimizing (4.7), which includes the sparsity penalty. This is essentially a two-stage, adaptive estimation procedure where the first stage sets $w_k = (1, 1, 1)$ for all covariates to generate initial estimates. The weights are updated and a second stage of adaptive estimation is performed. Next we formalize and generalize this sequential, adaptive procedure.

5.2. Sequential adaptive algorithm. The first stage of Sequential, Adaptive Functional Estimation (SAFE) performs adaptive group LASSO with $w_k^1 = (1, 1, 1)$, producing the stage 1 estimates $\{\hat{\gamma}_k^1\}_{k=1}^K$ for optimal tuning parameters $\pi^{1*}$. Let $K^1$ be the active variable subset of $\{1, \ldots, K\}$ where $\hat{\gamma}_k^1 = 0$ and let $\Gamma^1 = \{\hat{\gamma}_k^1; k \in K^1\}$. All $k \notin K^1$ are removed from consideration for the remaining stages of the estimation procedure.

The second stage of SAFE starts by calculating weights $w_k^2$ for $k \in K^1$ from their corresponding $\hat{\gamma}_k^1 \in \Gamma^1$. Adaptive group LASSO is performed with these $w_k^2$, producing the stage 2 estimates $\{\hat{\gamma}_k^2\}_{k \in K^1}$ for optimal tuning parameters $\pi^{2*}$. Note, the $\pi^{2*}$ are found from a new implementation of K-fold CV (or block CV). This gives the second stage active variable set $K^2$ and active variable estimates $\Gamma^2$. This two-stage approach is similar to the relaxed LASSO (Meinshausen, 2007) and the work by Wei and Huang (2010) and Guo et al. (2015).

SAFE is easily generalized to more than two stages and, as shown in Section 6, may perform better with more stages. Let $r = 1, \ldots, R$ index the stages of SAFE where stage $r$ uses weights, $w_k^r$, calculated from $\Gamma^{r-1}$ for

\[\text{imsart-aoas ver. 2014/10/16 file: draft_AAS_20180727.tex date: September 16, 2019}\]
$r \geq 2$ and $w_k^1 = (1, 1, 1)$. This gives $R$ active variable sets, $\mathcal{K}^1, \ldots, \mathcal{K}^R$, and active variable estimates, $\Gamma^1, \ldots, \Gamma^R$. The number of stages could be preset or chosen based on a stopping criterion that assesses the consistency of the $\mathcal{K}^r$ and/or $\Gamma^r$ for increasing $r$. For example, SAFE could stop if $\mathcal{K}^r = \mathcal{K}^{r-1}$ or when changes in the effect estimates in $\Gamma^r$ compared to $\Gamma^{r-1}$ are below some threshold. In Section 6, we fixed $R = 5$ and assessed convergence by plotting the norms in (4.4) across the stages to check whether any of the non-zero $\hat{\gamma}_k$ approached 0. An example is given in Section C.4 of the Supplementary Materials.

If the tuning parameter space is not explored well, it is possible that the $\text{APE}_{\text{min}}$ could increase significantly for increasing stages. To prevent this, we retain the $\text{APE}_{\text{min}}$ and its $\text{SE}_{\text{min}}$ from the first stage and include these values in the second stage CV. Denote these retained values by $\text{APE}_{\text{min}}^\ast$ and $\text{SE}_{\text{min}}^\ast$. If the second stage $\text{APE}_{\text{min}} \leq \text{APE}_{\text{min}}^\ast$, we continue with CV as usual and updated $\text{APE}_{\text{min}}^\ast = \text{APE}_{\text{min}}$ and $\text{SE}_{\text{min}}^\ast = \text{SE}_{\text{min}}$. Otherwise, we collect those second stage $\pi$’s with $\text{APE}_{\pi} \leq \text{APE}_{\pi}^\ast + \text{SE}_{\pi}^\ast$ and select $\pi^2^\ast$ as the one that minimizes $\sum_k P_\phi(\hat{\gamma}_k, \pi)$ with $\phi = (1, 1)$. If $\text{APE}_{\pi} > \text{APE}_{\pi}^\ast + \text{SE}_{\pi}^\ast$ for all second stage $\pi$’s, we restart SAFE with a finer grid of tuning parameter candidates. The process for more than two stages is straightforward.

In a similar spirit to Leeb et al. (2015) and Zhao et al. (2017), after the last SAFE stage we may re-estimate the $\gamma_k$ for $k \in \mathcal{K}^R$ using only a smooth regularization to potentially reduce the estimation bias caused by the sparsity penalty. We recommend these estimates be compared to those in $\Gamma^R$, as the latter estimates may be more interpretable due to their penalization of $||\gamma_k||^2$.

SAFE’s approach is easily generalized to models other than (4.1). The model need not even be functional. In the next section, we implement variations of SAFE with three other functional variable selection algorithms and show how it can also improve their selection performance.

6. EMG selection and prediction for hand movement. We applied the proposed methods to the AB subject’s data described in Section 2. Based on the structure of a biological limb, at least one signal is needed to control flexion and extension of a given movement DOF. However, the AB musculoskeletal system has muscle redundancies, meaning multiple muscles can contribute to a given movement DOF (see Figure 2). The ideal decoder would need only one EMG signal for extension and another for flexion without sacrificing prediction performance.

We expected estimation of $\gamma_k$ to be challenging for this data application.
for two reasons. Figure 4 suggested the presence of latent factors among the 16 measured signals, which can be explained by the EMG sensors being indirect measures of redundant muscle contractions. Moreover, Section A of the Supplementary Materials demonstrates how the distribution of the $X_{ik}$ can vary with position, with some positions having $X_{ik} \approx 0$. The latter can cause some inconsistent behavior of the estimated effects as we will see in Section 6.4.

We considered three competitors that perform functional variable selection and estimation on a model with a smooth, position effect and covariate-invariant effects for the EMG:

$$(6.1) \quad E[y_i|X_{i1}, \ldots, X_{iK}, z_i] = \alpha(z_i) + \sum_k \int_{-\delta}^{0} X_{ik}(s)\gamma_k(s)ds .$$

Since $y_i$ is either the first or second derivative of $z_i$, model (6.1) resembles a forced harmonic oscillator (Ramsay and Hooker, 2017), $D^2z = -\beta z + u$, where $u$ is a forcing input. Model (6.1) sets $u = \sum_k \int_{-\delta}^{0} X_{ik}(s)\gamma_k(s)ds$ and replaces $-\beta z$ with a smooth effect $\alpha(z_i)$, due to the positional constraints on extension/flexion of the fingers and wrist. The three competitors represent current functional variable selection methods and represent the approaches described in Section 3. All four methods are summarized here:

- **SAFE(z):** Model (4.1) with SAFE.
- **AGL:** Model (6.1) with SAFE, essentially adaptive group LASSO from Gertheiss et al. (2013).
- **LAD:** Model (6.1) with SAFE minimizing least absolute deviation (Pannu and Billor, 2017)
- **FAR:** Model (6.1) with group smoothly clipped absolute deviation penalty, i.e., linear functional additive regression from Fan et al. (2015)

Adaptive weighting was incorporated into AGL and LAD but not FAR because it was unclear how weighting should be performed. FAR was performed for $R = 2$ stages due to its lack of weighting, while the other three methods used $R = 5$ stages. The position effect was approximated by $\alpha(z_i) = \tau(z_i)^T\beta$, and its weighted sparse/smooth measures were included in penalty (5.1).

6.1. **Computational details.** The estimation procedure is similar to gradient matching (Ramsay and Hooker (2017), Chapter 8), which uses initial estimates of the response (i.e., velocity or acceleration) from the observed position data. The velocity values, $y_i$, were estimated from a smoother of the position data, $z_i$, based on a large set of sixth-order B-spline basis functions fit with a third-order penalty using fda (Ramsay and Silverman, 2005; Ramsey et al., 2014) in R (R Core Team, 2017). The acceleration values were
similarly estimated but with a fourth-order penalty. The \( \hat{y}_i \) were centered and scaled to have mean 0 and variance 1, and thinned by selecting every 20th observation in order to reduce the high temporal correlation, leaving approximately \( N = 200 \) responses. For each \( \hat{y}_i \), we extracted the previous \( \delta + 1 \) EMG observations, ending with the \( i \)-th EMG value. We chose a past time window of approximately 300ms (\( \delta = 40 \)) based on observed passive force movement from Figure 3. All \( X_{ik} \) were previously normalized between 0 and 1, so we only centered each curve at time points \( s = \delta, \ldots, 0 \). More details may be found in Section A of the Supplementary Materials.

For SAFE(\( z \)), \( \gamma_k(\cdot, \cdot) \) was modeled with orthogonal cubic B-splines of dimension \( L = 10 \) and \( M = 10 \) in the \( s \) and \( z \) direction, respectively. Both alternatives AGL and LAD set \( L = 10 \) basis functions for \( \gamma_k(\cdot) \), while FAR used \( L = 5 \) because it does not penalize smoothness. The optimal tuning parameters were found using 5-fold block CV as described in Section 4.3 with tuning parameter values \( \log(\phi_s), \log(\phi_z) \in \{-10, -5, 0, 5, 10\} \) and \( \log(\lambda) \) from \(-20\) to \(0\) in increments of \(0.25\). We required \( g_k, h_k \leq \exp(10) \) to circumvent computational issues that arose when \( \hat{\gamma}_k \) was approximately linear in either the \( s \) or \( z \) direction. We used the R package \texttt{gglasso} (Yang and Zou, 2013) for AGL and SAFE(\( z \)), \texttt{rqPen} (Sherwood et al., 2017) for LAD, and R code from the corresponding author of Fan et al. (2015) for FAR.

6.2. Performance metrics. For the AB subject, the muscles driving the movement of interest (finger or wrist extension/flexion) and the subset of EMG sensors that best measure these muscles’ contractions were known. Let \( \mathcal{K} = \mathcal{K}_F \cup \mathcal{K}_E \) denote the index set of these important EMG signals where \( \mathcal{K}_F \) and \( \mathcal{K}_E \) partition \( \mathcal{K} \) into index sets corresponding to flexion and extension, respectively. From Figure 2, these sets for finger and wrist movements were

- **Finger:** \( \mathcal{K}_F = \{X_{12}\} \quad \mathcal{K}_E = \{X_5, X_7\} \)
- **Wrist:** \( \mathcal{K}_F = \{X_8, X_{10}, X_{11}, X_{14}\} \quad \mathcal{K}_E = \{X_2, X_7, X_{13}, X_{15}\} \)

We compared the four methods with metrics focusing on identification of \( \mathcal{K}_F \) and \( \mathcal{K}_E \) as well as prediction accuracy:

- **Size** = \( |\hat{\mathcal{K}}| \), ideal size = 2 with \( |\hat{\mathcal{K}} \cap \mathcal{K}_F| = 1 \) and \( |\hat{\mathcal{K}} \cap \mathcal{K}_E| = 1 \)
- **Sparsity (SP)** = 1 - \( |\hat{\mathcal{K}}|/|\mathcal{K}| \), ideal sparsity = 14/16 = 0.875. Also, relative sparsity (RSP) = SP/0.875
- **False positive rate (FPR)** = \( FP/|\mathcal{K}| \), where \( \mathcal{K}^c \) is the complement of \( \mathcal{K} \) and \( FP = |\hat{\mathcal{K}} \cap \mathcal{K}^c| \)
- **True positive rate (TPR)** = \( TP/2 \), where \( TP = 1(|\hat{\mathcal{K}} \cap \mathcal{K}_F| \geq 1) + 1(|\hat{\mathcal{K}} \cap \mathcal{K}_E| \geq 1) \) focuses on whether we find a member of \( \mathcal{K}_F \) and \( \mathcal{K}_E \)
- **Average Prediction Error (APE)** = \( \sum_{i=1}^{N_0}(y_{0i} - \hat{y}_{0i})^2/N_0 \) for \( N_0 \)
6.3. Variable selection results. Table 1 shows the results of the variable selection performance for the six finger movement data sets with velocity as the response; acceleration and wrist movement results may be found in the Supplementary Materials. The data sets are referenced by movement DOF (F=Finger, W=Wrist), movement pattern (C=Consistent, R=Random), and replicate (1,2,3), e.g., FC1 = Finger movement, Consistent pattern, replicate 1. In Table 1, all methods tended to overselect in the initial stage. The final iteration of SAFE(z) selected signals \{X_7, X_{12}\} in all six data sets; for FR2 it also selected X_5. Both AGL and LAD benefited from adopting the SAFE approach, generally having better selection properties in the final stage than the initial stage. LAD’s final selection iteration was similar to SAFE(z) except for FC1, selecting six total EMG signals. AGL overselected for all data sets except FR2 and FR3. FAR generally performed well but overselected for FC2 and FR1-FR3.

![Table 1](https://example.com/table1.png)

Variable selection performance metrics for consistent (top three rows) and random (bottom three rows) finger movement patterns with velocity as the response. The top and bottom row in each cell represents the results of the initial and final stage, respectively.

Selection results for the acceleration-based models were generally worse than those shown in Table 1 (see Table 1 in the Supplementary Materials). SAFE(z) identified the correct signals for all data sets, but dropped X_{12} for FR3. The competing methods had more erratic results, both over and under selecting. For these reasons, we will focus our attention only on the velocity-based model.

For wrist movement, AGL, LAD, and FAR selected many EMG signals from \(K_F\) and \(K_E\) and often had nonzero FPRs (see Table 2 in the Supple-
SAFE(\(z\)) always selected one EMG sensor from each \(K_F\) and \(K_E\) except for WC1 when it selected two signals from each group. SAFE(\(z\)) also maintained FPR=0 for all data sets.

6.4. **EMG effect interpretations.** The SAFE(\(z\)) estimated effects for all finger/wrist data sets can be found in Sections B.3 and C.3 of the Supplementary Materials, respectively. Figure 5 here shows two sets of SAFE(\(z\)) estimates \(\hat{\gamma}_7\) and \(\hat{\gamma}_{12}\) for FC3, one using only a smoothness penalty (top panels) and another from the final SAFE stage (bottom panels). Fixing \(s = 0\), concurrent activation of \(X_7\) led to finger extension (negative velocity), while concurrent activation of \(X_{12}\) led to finger flexion (positive velocity). The top panels suggest the effect of \(X_{12}\) was strongest when the hand was fully extended (negative position). This is impossible for the consistent movement pattern because the flexor digitorum, measured by \(X_{12}\), never contracted during maximal finger extension, meaning \(X_{i,12}(s) \approx 0\) at this position. By not penalizing \(\|\gamma_{12}\|\), the estimated effect deviated from 0 while maintaining its smoothness and without degrading the model’s fit. When this penalty was included, the estimated effects approached 0 at these positions, providing clearer interpretations of the EMG’s movement contributions. For this reason, we use the final stage estimates for interpretation and prediction.

**Fig 5.** Estimated coefficients for \(X_7\) (left) and \(X_{12}\) (right) under a smoothness-only penalty (top) and the sparse/smooth estimates from SAFE(\(z\)) (bottom). To help interpretation, the x-axis is in terms of previous seconds from the concurrent time, indicated by 0.00. Red values indicate positive velocity (flexion); blue values indicate negative velocity (extension); gray values indicate zero velocity.

The effect of varying \(X_{12}\) on finger movement was concentrated between
0 < z_i < 55, during which the fingers were in a neutral or flexed position. For these z_i, concurrent activation of X_{12} produced finger flexion, while past activation of X_{12} contributed to finger extension. The impact of X_7 was most relevant between angles −30 < z_i < 20, a range of extended and neutral finger positions. For these z_i, concurrent activation of X_7 led to finger extension, while past activation led to finger flexion. This indicates our model successfully captured the passive force effects from past EMG behavior. Section B.4 of the Supplementary Materials gives the interested reader additional interpretations of these effects.

6.5. Prediction performance. Each model for a given data set was used to predict movement for the remaining five data sets. We calculated the mean and standard error of APEs across these five holdout samples for each data set and refer to this as out-sample prediction. Prediction was challenging across data sets, particularly when the consistent-movement models were asked to predict random movement. In-sample predictions was also considered, referring to the mean and standard error of the CV APEs across the 5 in-sample folds under the optimal π*. Figure 6 shows that across nearly all scenarios, SAFE(z) outperforms the competitors in mean APE using fewer EMG signals. As expected, finger out-sample APE for the consistent movement data sets had higher standard errors, likely due to the prediction of random movements. SAFE(z) APEs for the wrist movement data were consistently small across the two movement patterns.

7. Simulation study. Recall that the original data were thinned to circumvent potential issues resulting from temporal correlations between the y_i’s. A simulation study was performed to investigate the impact of covariance misspecification on variable selection performance. The FC3 data and its estimated coefficients (Figure 5) were used to generate responses under various error structures. The data-generating model was $y_i = \int_S X_{17}(s) \hat{\gamma}_{17}(s, z_i) ds + \int_S X_{i,12}(s) \hat{\gamma}_{12}(s, z_i) ds + \epsilon_i$ where $\epsilon_i$ was a zero-mean error process with isotropic covariance function

$$\text{cov}(\epsilon_i, \epsilon_{i'}) = \sigma_h^2 \left[ \mathbf{1}(i = i') + \theta \exp \left\{ - \left( |i - i'| / \eta \right)^2 \right\} \right].$$

Here $\theta$ is related to the dominant sources of dependence; $\theta = 0$ means that the responses are uncorrelated with $\text{var}(\epsilon_i) = \sigma_h^2$ while large $\theta$ reflects a higher degree of dependency and, for fixed $\sigma_h^2$, increases variance to $\text{var}(\epsilon_i) = \sigma_h^2 (1 + \theta)$. In addition, $\eta > 0$ controls the correlation decay where larger values imply slower correlation decay. The simulation study varied over two factors:

- $\theta = 0.25, 10, 100$

- $\eta = 1, 10, 100$
Fig 6. In-sample CV MSE means and standard errors for the optimal tuning parameters of the last selection stage (top panel) and out-sample prediction MSE means and standard errors (bottom panel) using the remaining five data sets.

- $\eta = 10$ (Fast Decay) and 100 (Slow Decay)

and sets $\sigma_h^2$ so that $\text{var}(\epsilon_i) = \sigma_h^2(1 + \theta)$ equals the in-sample MSE from the FC3 data application. For each setting, we simulated 100 independent samples with $N = 200$ observations and analyzed the data with all four competitors from Section 6 with $S = 2$ stages.

Table 2 shows the average variable selection for the six covariance scenarios. All methods had a TPR = 1 and a negligible FPR. For example, SAFE(z)'s FPR never exceeded 0.009. We also calculated the proportion of all 100 second-stage models with size greater than 2. For SAFE(z), the largest proportion was 10% while the other methods had proportions ranging from 15% (AGL, $\theta = 0.25$, $\eta = 10$) to 72% (FAR, $\theta = 100$, $\eta = 100$).

Figure 7 shows the in-sample MSE performance for the six covariance scenarios. As expected, the MSEs were larger for $\eta = 100$ than $\eta = 10$, but they only slightly increased as $\theta$ increased. These results provide evidence of SAFE’s robustness to covariance misspecification. Section D of the Supplementary Materials has results from another numerical experiment that investigated the role of model misspecification on variable selection and prediction.
Table 2
Average performance metrics across 100 data sets for six covariance settings \(\theta = 0.25, 10, 100\) and correlation decay that is slow (\(\eta = 100\)) and fast (\(\eta = 10\)) with model based on fits of data set FC3. The top and bottom row in each cell represents the results of the first and second stage, respectively. TPR=1 for all scenarios.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(\theta) Decay</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25 Slow</td>
<td>0.863 [3.92] 0.134 [1.74]</td>
<td>0.759 [5.37] 0.231 [3]</td>
<td>0.933 [2.94] 0.042 [0.55]</td>
<td>0.881 [3.66] 0.125 [1.62]</td>
</tr>
<tr>
<td></td>
<td>0.953 [2.66] 0.048 [0.62]</td>
<td>0.903 [3.36] 0.097 [1.26]</td>
<td>0.937 [2.88] 0.038 [0.5]</td>
<td>0.996 [2.06] 0.005 [0.06]</td>
</tr>
<tr>
<td>0.25 Fast</td>
<td>0.834 [4.32] 0.165 [2.14]</td>
<td>0.725 [5.85] 0.265 [3.44]</td>
<td>0.936 [2.9] 0.040 [0.52]</td>
<td>0.862 [3.93] 0.146 [1.9]</td>
</tr>
<tr>
<td></td>
<td>0.949 [2.71] 0.049 [0.64]</td>
<td>0.887 [3.58] 0.108 [1.41]</td>
<td>0.940 [2.84] 0.038 [0.5]</td>
<td>0.991 [2.12] 0.009 [0.12]</td>
</tr>
<tr>
<td>10 Slow</td>
<td>0.884 [3.62] 0.115 [1.49]</td>
<td>0.771 [5.2] 0.223 [2.9]</td>
<td>0.946 [2.75] 0.034 [0.41]</td>
<td>0.926 [3.04] 0.078 [1.02]</td>
</tr>
<tr>
<td></td>
<td>0.942 [2.81] 0.056 [0.73]</td>
<td>0.930 [2.98] 0.060 [0.9]</td>
<td>0.949 [2.7] 0.029 [0.38]</td>
<td>0.999 [2.02] 0.002 [0.02]</td>
</tr>
<tr>
<td>10 Fast</td>
<td>0.752 [5.47] 0.244 [3.17]</td>
<td>0.638 [7.07] 0.356 [4.63]</td>
<td>0.904 [3.35] 0.065 [0.84]</td>
<td>0.840 [4.24] 0.168 [2.19]</td>
</tr>
<tr>
<td></td>
<td>0.918 [3.15] 0.082 [1.07]</td>
<td>0.899 [3.41] 0.103 [1.34]</td>
<td>0.914 [3.2] 0.037 [0.74]</td>
<td>0.995 [2.07] 0.005 [0.07]</td>
</tr>
<tr>
<td>100 Slow</td>
<td>0.871 [3.81] 0.128 [1.66]</td>
<td>0.759 [5.38] 0.235 [3.05]</td>
<td>0.946 [2.76] 0.035 [0.46]</td>
<td>0.923 [3.08] 0.082 [1.06]</td>
</tr>
<tr>
<td></td>
<td>0.941 [2.82] 0.058 [0.75]</td>
<td>0.913 [3.22] 0.087 [1.13]</td>
<td>0.951 [2.69] 0.032 [0.42]</td>
<td>0.996 [2.05] 0.004 [0.05]</td>
</tr>
<tr>
<td>100 Fast</td>
<td>0.758 [5.39] 0.238 [3.09]</td>
<td>0.609 [7.48] 0.378 [4.91]</td>
<td>0.901 [3.39] 0.062 [0.8]</td>
<td>0.828 [4.41] 0.178 [2.31]</td>
</tr>
<tr>
<td></td>
<td>0.905 [2.91] 0.067 [0.87]</td>
<td>0.865 [3.89] 0.137 [1.78]</td>
<td>0.904 [3.34] 0.059 [0.77]</td>
<td>0.994 [2.08] 0.006 [0.08]</td>
</tr>
</tbody>
</table>

8. Discussion and Extensions. In this paper, we proposed a new PC decoder based on a covariate-varying, scalar-on-function linear model to account for the biomechanical characteristics involved in hand movement. The functional predictors were the recent past behavior of EMG signals measured across the subject’s limb and the responses were finger and wrist velocity during flexion/extension. The effects for each EMG signal were allowed to vary with the current finger or wrist position. The bivariate effects were approximated using a tensor product of basis expansions whose coefficients were estimated with a group LASSO penalty that combined smoothing and sparseness penalization. We developed an adaptive, multi-stage estimation procedure, called Sequential, Adaptive Functional Estimation (SAFE), motivated by Meinshausen (2007), Wei and Huang (2010), and Guo et al. (2015). Section 6 showed our proposed decoder estimated with SAFE was able to recover a sparse set of important EMG signals for finger and wrist movement for an AB subject while competitors based on a more conventional model overselected and had worse prediction. A numerical experiment based on the data application also showed SAFE’s robustness to covariance misspecification for data with high signal-to-noise ratios.

Our methodology provides biomedical engineers new tools for developing a TRA prosthesis controller. For example, if it is difficult to identify the function and location of residual muscles in the residual limb, engineers can first place high-density EMG electrodes over the skin surface. If the TRA performs mirrored hand movements with both their intact and residual limb, movement data can be captured from their intact limb that corresponds to...
Fig 7. Average and standard deviation of in-sample MSEs across 100 data sets for six covariance settings $\theta = 0.25, 10, 100$ and Slow($\eta = 100$)/Fast($\eta = 10$) correlation decay with model based on fits of data set FC3. The bottom reference line is the in-sample MSE from the analysis of the FC3 data.

the EMG patterns collected from the residual limb. Using our decoder and SAFE, engineers can remove redundant EMG signals and examine the estimated effects to interpret the role each EMG signal has on movement. This reduces the computational burden associated with streaming and processing large amounts of EMG data that hinder state-of-the-art PR methods and provides a unique way to determine EMG electrode location for musculoskeletal model-based PCs.

The proposed SAFE procedure opens several research questions. This paper focuses on approximating the bivariate coefficients with tensor products of finite dimensional basis functions, which can cause the number of parameters to increase exponentially. Other economical basis functions should be explored, such as splines over a triangulation or radial basis functions, although these involve computational challenges for calculating the penalty. Another avenue is to extend it to functional linear models where the functional measurements are perturbed by error or are sparsely measured. In this case, one could smooth the functional covariates using existing approaches (Yao et al., 2005; Xiao et al., 2016) and employ them in a SAFE selection algorithm. We have also investigated constructing prediction intervals based on the distribution-free split-conformal prediction sets (see Section E of the Supplementary Materials for a short demonstration), which as far as the authors are aware has not been applied to functional regression.

We conjecture that SAFE requires more stages when the true model is driven by few latent variables. To see this, let $L_1(s)$ and $L_2(s)$ be two latent factors where $E(y_i|L_1(s), L_2(s)) = \int_S L_1(s)\beta_1(s)ds + \int_S L_2(s)\beta_2(s)ds$. 

\[
\]
Suppose we perform variable selection across $K$ observed covariates where, for simplicity, $X_k(s) = \alpha_{k1}L_1(s) + \alpha_{k2}L_2(s)$. If we attempt to fit a model with these covariates, we get the following model equivalency

$$
\sum_{k=1}^{K} \int_{S} X_k(s)\gamma_k(s)ds = \sum_{k=1}^{K} \int_{S} (\alpha_{k1}L_1(s) + \alpha_{k2}L_2(s))\gamma_k(s)ds \\
= \int_{S} L_1(s) \sum_{k} \alpha_{k1}\gamma_k(s)ds + \int_{S} L_2(s) \sum_{k} \alpha_{k2}\gamma_k(s)ds .
$$

For the fitted model to approximate the true model, the estimates, $\hat{\gamma}_k(s)$, should satisfy $\sum_k \alpha_{k1}\hat{\gamma}_k(s) \approx \beta_1(s)$ and $\sum_k \alpha_{k2}\hat{\gamma}_k(s) \approx \beta_2(s)$. That is, the true effects $\beta_1(s)$ and $\beta_2(s)$ will be partitioned across the $\gamma_k(s)$ and variable selection across the $X_k(s)$ implies we want a sparse partition. Across the multiple stages of SAFE, these partitioned effects will be reflected in the updated adaptive weights and performing an additional fitting stage will eventually cause the predictors with larger weight to be removed from consideration. The number of required stages to arrive at a sparse model is then likely driven by the $\alpha_{k1}$ and $\alpha_{k2}$. Paul et al. (2008) investigated variable selection for predictors with this latent factor structure and many of their ideas could be extended to the functional variable problem. Their technique may be combined with SAFE selection and would allow us to efficiently screen a large number of EMG signals.

Acknowledgments. The project described was supported by NSF grant number DIIS1527202. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NSF. A.-M. Staicu’s contributions were supported by NSF grant number DMS1454942 and NIH grant number P01CA142538. We also thank Rebecca North at NC State University for her help with the simulation studies.

SUPPLEMENTARY MATERIAL

Supplement: Algorithms and Additional Numerical Results

(doi: COMPLETED BY THE TYPESETTER; .pdf). The Supplementary Material contains details for data re-construction, additional results for finger and wrist movement analysis, simulation results, and demonstration of constructing distribution-free prediction intervals using split conformal inference.

References.


proportional neural control information for multiple-DOF prostheses from the surface electromyographic signal,” *IEEE Transactions on Biomedical Engineering*, 56, 1070–1080.


Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauen-

Department of Statistics,
North Carolina State University,
Raleigh, NC.
E-mail: jwstalli@ncsu.edu