8-25-2016

Hybrid System Identification of Manual Tracking Submovements in Parkinson's Disease

Carlos Gonzalez

Carlos Gonzalez

Candidate

Electrical and Computer Engineering

Department

This thesis is approved, and it is acceptable in quality and form for publication:

Approved by the Thesis Committee:

Meeko Oishi , Chairperson

Rafael Fierro

Chaouki Abdallah

Hybrid System Identification of Manual Tracking Submovements in Parkinson's Disease

by

Carlos Gonzalez

B.S., Mechanical Engineering, University of New Mexico, 2014

THESIS

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Master of Science
in Electrical Engineering

The University of New Mexico

Albuquerque, New Mexico

July, 2016

Dedication

To my parents, my aunt and uncle, and my friends who had faith in me.

Acknowledgments

I would like to start by thanking my lab partners: Abraham Vinod, Joseph Gleason, and Steven Cutlip, for the priceless experience, support, and insights they brought up on my work. I would also like to give a special thanks to my advisor Dr. Meeko Oishi for her valuable guidance throughout the beginning of my research career and to Dr. Martin McKeown for providing the data used in this thesis work as well as for his help in interpreting results from a neuroscience perspective.

This material is based upon work supported by the National Science Foundation under Grant Number CMMI- 1254990 (CAREER, Oishi) and by Ian and Rosemary Mottershead and The Charros Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Hybrid System Identification of Manual Tracking Submovements in Parkinson's Disease

by

Carlos Gonzalez

B.S., Mechanical Engineering, University of New Mexico, 2014

M.S., Electrical Engineering, University of New Mexico, 2016

Abstract

Seemingly smooth motions in manual tracking, (e.g., following a moving target with a joystick input) are actually sequences of submovements: short, open-loop motions that have been previously learned. In Parkinson's disease, a neurodegenerative movement disorder, characterizations of motor performance can yield insight into underlying neurological mechanisms and therefore into potential treatment strategies. We focus on characterizing submovements through Hybrid System Identification, in which the dynamics of each submovement, the mode sequence and timing, and switching mechanisms are all unknown. We describe an initialization that provides a mode sequence and estimate of the dynamics of submovements, then apply hybrid optimization techniques based on embedding to solve a constrained nonlinear program. We also use the existing geometric approach for hybrid system identification to analyze our model and explain the deficits and advantages of each. These methods are applied to data gathered from subjects with Parkinson's disease (on and off L-dopa medication) and from age-matched control subjects, and the results compared

across groups demonstrating robust differences. Lastly, we develop a scheme to estimate the switching mechanism of the modeled hybrid system by using the principle of maximum margin separating hyperplane, which is a convex optimization problem, over the affine parameters describing the switching surface and provide a means o characterizing when too many or too few parameters are hypothesized to lie in the switching surface.

Contents

List of Figures

List of Tables

Glossary

$A^\top$	Transpose of matrix A
$A^\dagger$	Moore-Penrose pseudoinverse of A
$u[k]$	Input at time k
$y[k]$	Output at time k
$\mathbb{R}$	The set of real numbers
$\mathbb{Z}$	The set of integers
$\|\cdot\|$	The Euclidean norm
$\Leftrightarrow$	If and only if
$\in$	Belongs to
$\preceq$	Component-wise inequality
$\triangleq$	Equal by definition

Chapter 1

Introduction

1.1 Overview

In this thesis, we develop a method to identify multiple parameters of hybrid systems. We do this by looking at two different problems: (1)identifying the hybrid system dynamics and the active modes at all times, and (2) estimating a guard condition that meets the hybrid system switching criteria.

Our approach to identify the former problem is based on perspectives of hybrid optimal control. This approach is applied on actual data from a tracking task performed by 'normal' subjects and subject with Parkinson's disease before and after medication. We then compare our results to the solutions obtained from a current geometric approach to hybrid system identification, which is currently one of the most robust-to-noise methods. We then verify the validity of our results with current perspectives from the neuroscience community, specially those which are specific to Parkinson's disease. The latter problem is addressed by finding the separating hyperplane with maximum margin at mode transition times. In this regard, we also show some properties that indicate the we have estimated the true guard.

1.2 Parkinson Disease

Parkinson's disease is a neurodegenerative movement disorder – a progressive condition with no cure, and no definitive means to alter the progression of the disease. It is characterized by tremor, rigidity, bradykinesia (slowness of movement) and postural instability, and is the second most common neurodegenerative disease after Alzheimer's, affecting approximately 1 million people in the US. Pharmacological treatment of Parkinson's disease, predominately based on dopaminergic replacement, works well in the early disease stages, but motor complications such as dyskinesias (involuntary writhing movements) may occur.

While external non-invasive brain stimulation has the potential to provide patient-specific, time-sensitive, targeted therapy, the particular disease features or biomarkers that such a treatment should target are unclear. Clinical scales and questionnaires are often crude and subjective; imaging and brain mapping techniques (fMRI, PET, MEG) are expensive, non-portable, and preclude frequent testing. Furthermore, the relation between altered brain activation and motor performance may not be straightforward, so therapies that solely optimize brain activity patterns may not translate into clinically significant benefits. However, measures of motor performance, correlated with brain function, may provide a necessary linkage between overall motor behavior (ultimately important for any successful therapy) and neurological mechanisms.

1.3 Hybrid Systems Models

Hybrid systems are dynamical systems that couple the interaction of continuous and discrete dynamics [3]. These systems appear naturally in many areas, e.g., in air traffic control [4], biped locomotion modeling [5], control of DC-DC boost

converters [6], etc. In many of these areas, physics-based principles can often be applied to arrive to descriptive models which can be later altered through control mechanisms. In other cases, the models, though known to be hybrid, can be difficult and complicated to model. For such situations we have to resort to hybrid system identification techniques.

In this work, we will solely discuss discrete-time hybrid systems, which are often modeled in one of two forms: in state-space or in input-output form [1]. The former are described by

$$\begin{aligned} x[k+1] &= A_{q[k]}x[k] + B_{q[k]}u[k] + f_{q[k]} + w[k] \\ y[k] &= C_{q[k]}x[k] + D_{q[k]}u[k] + g_{q[k]} + v[k] \end{aligned} \tag{1.1}$$

where $x[k] \in \mathbb{R}^n$ is the continuous state vector, $u[k] \in \mathbb{R}^p$ is the continuous input vector, $y[k] \in \mathbb{R}^m$ is the continuous output vector, $w[k] \in \mathbb{R}^n$ is the process noise, and $v[k] \in \mathbb{R}^m$ is the measurement noise term, all at time $k \in \mathbb{Z}$. In this framework, the discrete state $q[k] \in \{1, 2, \cdots, s\}$ selects the affine mode that is active at time k. The matrices A_i, B_i, f_i, C_i, D_i, and g_i, of corresponding dimensions, describe the dynamics evolution. In general, $q[k]$ can be an exogenous input, or a function of $x[k]$ and $u[k]$. Systems of the form (1.1) are known as Switched Affine (SWA) models. When $q[k)]$ is described by a polyhedral partition of the state space, i.e., if

$$q[k] = i \Leftrightarrow \begin{bmatrix} x[k] \\ u[k] \end{bmatrix} \in \Omega_i \tag{1.2}$$

then the system is said to be Piece-Wise Affine (PWA), where $\{\Omega_i\}_{i=1}^s$ define a complete partition of the domain.

For systems described by (1.1), the problem is described as follows [7].

Problem 1 *Given the set of input-output pairs $\{(u[k], y[k])\}_{k=1}^N$, estimate the model order n, the number of modes s, the matrices A_i, B_i, f_i, C_i, D_i, and g_i, and, if the model is PWA, find the polyhedral regions Ω_i, $i = 1, 2, \cdots, s$.*

On the other hand, systems modeled in input-output form are described by the collection of ARX systems, given by

$$y[k] = \theta_{q[k]}^{\top}\varphi[k] + e[k] \tag{1.3}$$

with parameter vectors θ_i describing the dynamics, noise term $e[k]$, and extended regressor vector $\varphi[k] = [r[k]^{\top}\ 1]^{\top}$ containing the regressor vector

$$r[k] = \begin{bmatrix} y[k-1]^{\top} & \cdots & y[k-n_a]^{\top} & u[k]^{\top} & u[k-1]^{\top} & \cdots & u[k-n_b]^{\top} \end{bmatrix}^{\top}$$

Systems described by (1.3) are called Switched ARX (SARX) models and represent a subclass of the SWA model (1.1). If the switching rule of (1.3) is defined by

$$q[k] = i \quad \Leftrightarrow \quad r[k] \in \mathcal{R}_i, \quad i = 1, 2, \cdots, s \tag{1.4}$$

with $\mathcal{R} \subseteq \mathbb{R}^d$, where $d = m \cdot n_a + p \cdot (n_b + 1)$, it is a PieceWise Affine ARX (PWARX) model. In this case, $\{\mathbb{R}_i\}_{i=1}^{s}$ are polyhedrons forming a complete partition of the regressor domain $\mathcal{R}$, i.e., each region $\mathcal{R}_i$ are defined as

$$\mathcal{R}_i = \left\{ r \in \mathbb{R}^d \ : \ H_i \begin{bmatrix} r \\ 1 \end{bmatrix} \preceq_{[i]} 0 \right\} \tag{1.5}$$

For systems of the form (1.3), the identification problem is stated as follows [7].

Problem 2 *Given the set of input-output pairs $\{(u[k], y[k])\}_{k=1}^{N}$, estimate the model order n_a and n_b, the number of modes s, the parameter vectors $\theta_i, i = 1, 2, \cdots, s$, and the discrete state $q[k]$ for $k > \max\{n_a, n_b\}$. In addition, if the model is PWARX, estimate the regions $\mathcal{R}_i,\ i = 1, 2, \cdots, s$.*

Other relevant hybrid system models exist, such as Mixed Logical Dynamical (MLD) models, Jump-Markov Linear (JML) models, Max-Min-Plus-Scaling (MMPS), will not be discussed here but the interested reader is referred to the survey papers [1, 7], which give some mention and further references for these types of systems.

1.4 Thesis Outline

The remainder of the thesis is organized as follows: chapter 2 gives a brief exposition of the different existing methods for the identification of PWARX and SARX systems. Chapter 3 describes our modeling framework used to characterize submovements, which borrows techniques from optimal hybrid control and applies them in the area of hybrid system identification. This methodology is applied to real data obtained from 'healthy' subject, and subjects with Parkinson Disease performing a tracking experiment. The results are compared to another different hybrid system identification technique using a geometric approach to hybrid system identification. In chapter 4, we describe our current approach to obtain the switching rule for the modeled SARX system. We conclude this work with chapter 5, where we present a summary and discuss some future work.

1.5 Contributions

Hybrid systems identification has gained an increase interest and ample development over the last decades. One of the most common approaches used to test the effectiveness of these methods is through synthetic data with varying noise statistics. Few works have shown results obtained in real applications, moreover, fewer works focused on SARX systems. We developed an identification approach based on results derived from the area of optimization of hybrid systems, which uses a relaxation of the MIQP, and applied our algorithm to real human data, which is a major challenge as data is not only noisy but also nonlinear and heterogeneous. Our modeling framework is simple enough that it can provide neurological meaning through the description of a single parameter taking multiple values over different modes of the hybrid system. This decision allows us to compare results across subjects and popu-

lations (comparing a switching matrix would be not only more complex but harder to interpret from a neurological point of view).

Secondly, we work towards the problem of guard estimation of SARX systems (with unknown switching rule) and assess some properties that allow us to determine when the right parameter space has been found. This is a major challenge which, once fully developed, could aid in providing insight into the different ways in which the brain operates once certain regions have been compromised. For instance, it is known that the basal ganglia and the cerebellum are involved in motor tracking tasks and are often impaired in Parkinson's disease, hence, being able to characterize these in a model which can be enhanced through either the right dosage of medication or with external inputs, can bring advances to the biomedical and neuroscience field.

Chapter 2

Identifying Dynamics of Hybrid Systems

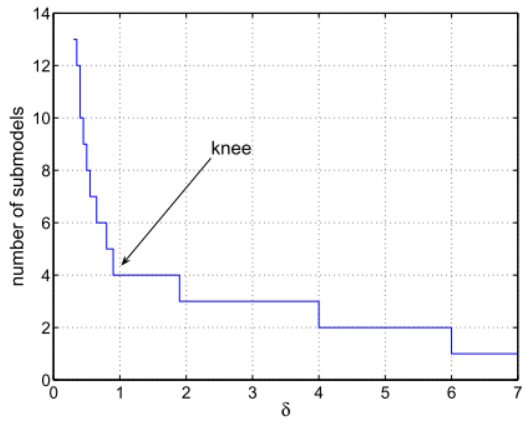
Current methods used to identify the multiple components of a hybrid system exist for models in state-space (e.g., [8, 9, 10, 11]) and input-output form (e.g., [12, 13, 14, 15, 16]). As previously stated, we focus on problems posed in the latter form for single-input single-output (SISO) systems. These methods fall into the two categories mentioned in § 1.3: SARX or PWARX.

In general, if the number of modes s of the hybrid system is known, the identification problem can be posed as the following Mixed Integer Program (MIQP)

$$\begin{aligned} \text{minimize} \quad & \sum_{k=1}^{N} \sum_{i=1}^{s} \ell(y[k] - \varphi[k]^\top \theta_i) q_{k,i} & \\ \text{subject to} \quad & \sum_{i=1}^{s} q_{k,i} = 1 & \forall\ k \\ & q_{k,i} \in \{0,1\} & \forall\ k, i \end{aligned} \tag{2.1}$$

where $\ell(\cdot)$ is a nonnegative loss function. Two common loss functions are $\ell(\cdot) = \epsilon^2$

Figure 2.1: Plot of number of modes s as the tunning parameter δ is varied. The "knee" of the plot provides a heuristic to finding the number of modes in a hybrid system. Figure from [1].



and $\ell(\cdot) = |\epsilon|$ in which case (2.1) becomes a Mixed Integer Quadratic Program (MIQP) or Mixed Integer Linear Program (MILP), respectively. However, solving (2.1) is an NP-hard problem, thus multiple researcher have looked at alternate methods to solve the hybrid system identification problem. Some of these methods are mentioned and briefly explained in § 2.1.

When the number of modes s is not readily available, one often poses the following problem [1].

Problem 3 *Given $\delta > 0$, find the smallest number s of vectors θ_i, $i = 1, 2, \cdots, s$ and a mapping $k \mapsto q[k]$ such that*

$$|y[k] - \varphi^\top[k]\theta_{q[k]}| \leq \delta, \quad \forall\ k = \bar{n}, \cdots, N \tag{2.2}$$

where $\bar{n} = \max\{n_a, n_b\}$.

The value δ in Problem 3 is a tunning parameter accounting for the trade off between fit and accuracy. A typical approach in finding the number of modes in a system is by finding the knee of the s-curve, shown in Figure 2.1.

2.1 Methods to Identify SARX Models

SARX models are represented by the general description (1.3) without requirement (1.4), i.e., with arbitrary switching times allowed between modes.

One of the pioneering works in the identification of SARX models is the algebraic procedure [17]. This approach introduces the notion of the hybrid decoupling constraint, given by

$$\prod_{i=1}^{s}(b_i^\top z[k]) = 0 \tag{2.3}$$

with $b_i = [1\ \theta_i^\top]^\top$ and $z[k] = [-y[k]\ \varphi^\top[k]]^\top$, which decouples the identification of the parameter vectors θ_i from the switching sequence and switching rule, hence aiding in the process of identifying both SARX and PWARX models. Although initially formulated to provide a close form solution to noiseless systems, it addresses the case of noisy data by manipulating the learning rate of the proposed algorithm [18]. Moreover, an extension to the algebraic procedure is shown in [19], where the problem is reformulated into a constrained rank minimization problem, which is known to be NP-hard, this is relaxed into a convex optimization problem.

An additional method is the sparse optimization method, inspired in recent results in compressed sensing[20]. Sparse optimization is in general a non-convex problem, so it is relaxed into an ℓ_1 minimization problem. A continuous optimization framework is presented in [15], where the MIQP (2.1) is approached by instead posing the problem as the non-convex unconstrained problem

$$\underset{f_j}{\text{minimize}} \quad \sum_{i=1}^{N}\Big(\min_{j=1,\cdots,s}\ell(y_i - f_j(\boldsymbol{x}_i))\Big) \tag{2.4}$$

which manages to reduce the complexity of the MIQP problem. This work also shows how this formulation can be posed for additional loss functions, such as the Maximum-Likelihood framework and the product-of-errors estimators. A more recent approach which has been shown to be more robust to noise is the geometric

approach [21]. In here, the problem is approached by looking at the parameter space instead of the data space. They utilize the well-studied Principal Component Analysis (PCA) to separate the hyperspheres describing the parameters of the parameter vectors θ_i, which are then found using linear regression. This method is compared against other well-known methods, namely the algebraic approach [17], the continuous optimization approach [15], and two sparse optimization methods [20, 2], and the geometric approach [21] is numerically shown to be more robust to noise than these other ones. Therefore, we choose this model as a basis for comparison when we implement system identification to our data set in Chapter 3.

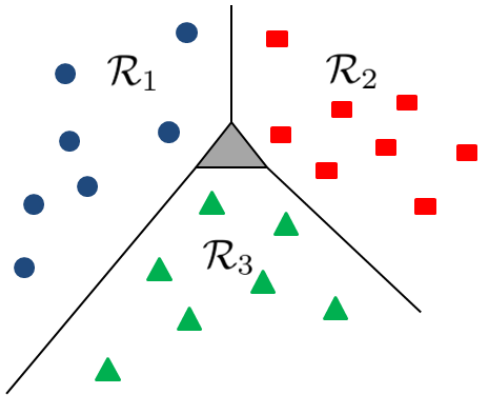
2.2 Methods to Identify PWARX Models

PWARX models are described by (1.3) with switching rule defined by a partition of the regressor space, as given by (1.4). Thus, it is not uncommon to see that methods used to identify SARX models can be used to identify PWARX models, given that the constraint (1.4) is treated afterwards. Note, however, that expressing the constraint that the regions $\{\mathcal{R}_i\}_{i=1}^{s}$ form a complete partition of the regressor domain $\mathcal{R}$ is one of the main difficulties that arises in the identification of PWARX models. This can be more clearly visualized in Figure 2.2 where for systems with more than two modes, the classification techniques used can lead to having holes. One common technique used to overcome this difficulty consists in using multi-category support vector machines (SVM) [22].

Initially, the procedure to PWARX system identification consisted on first classifying the data points and estimating the model parameters, then reconstructing the partitions of the regressor domain. This was typically achieved by using linear support vector machines (LSVM).

Alternate methods to the one mentioned above include a bounded error approach

Figure 2.2: Example of system with three modes with a classification scheme that leads to an incomplete partition of the regressor domain $\mathcal{R}$, with the gray shaded area being the not covered region.



[23] which uses set membership techniques, a statistical clustering technique [24] which uses the expectation-maximization algorithm and support vector classifiers, and a convex approach that starts with an over-parameterization of the parameter vector then uses a sum of norms regularization to cluster subsystems [25, 26].

Chapter 3

Detection of Submovements in Parkinson's Disease

3.1 Relevance of Submvements in Parkinson's Disease

Submovements are open-loop motions ("motor programs") that have been previously learned. Seemingly smooth motions (such as in manual tracking) are actually sequences of submovements [27, 28], in which motor planning is accomplished via selection between motor programs [29, 30]. Corrective submovements are thought to complementarily involve the cerebellum and basal ganglia [31, 32], with roughly decision and timing of the corrections determined by the basal ganglia and cerebellum, respectively. Deficiencies in these regions are well established in Parkinson's disease, however little has been done to characterize how these deficiencies affect submovements in Parkinson's disease. We aim to detect submovements using hybrid system identification, with the ultimate goal of characterization for personalized brain stimulation therapy.

We seek methodological characterization of submovements that are amenable to future design of subject-specific feedback control laws. Hence, we focus on optimization and system identification. Other researchers have considered related problems in characterization of tasks on human movements ([33]), albeit it at a different scale and with a focus on characterization through classification. Motions are decomposed into fundamental tasks, denoted *movemes*, via pattern recognition techniques that are used to train Gaussian classifiers for each moveme.

Previous work has demonstrated that simple second-order linear time-invariant (LTI) models can describe manual tracking tasks [34, 35, 36] at a high level in which submovements are not explicitly considered. We presume that each submovement can be characterized via a second-order LTI system, and that transitions between submovements are instantaneous, yielding a switched linear dynamical system. However, since our hybrid model of manual tracking is not based on first principles, we seek to determine a) the model parameters of the LTI dynamics that describe each submovement, and b) the timing of submovements (e.g., the mode). Hence, we are interested in methods that are agnostic to all of these elements.

3.2 Experiment Setup

The experiments are completed by subjects with clinically defined mild to moderate Parkinson's disease (measured clinically as Hoehn & Yahr stage 1-3), both off and on dopaminergic medication, as well as by "normal" age-matched subjects without Parkinson's disease or other confounding ailments. Data is gathered at the Pacific Parkinson's Research Centre at the University of British Columbia under the supervision of Dr. McKeown, in accordance with institutional review board protocols. Subjects practice the task during a training session until their tracking errors converged to a constant value. Data we consider here are gathered after completion of

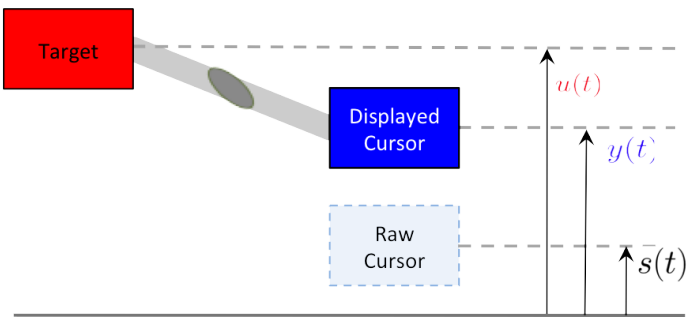


Figure 3.1: Experimental setup. The subject controls the cursor, with the goal of tracking the height of the target. The cursor position is manipulated by scaling the error between the target height and the raw cursor height, to appear to be "better", "worse", or "normal".

the training period.

This experiment is designed to assess responsiveness to error, adaptation to sudden change, and reward mechanisms, and is EEG compliant. The experimental setup is illustrated in Figure 3.1, where subjects use a joystick to control the vertical position of a cursor (blue box) which is connected to a target (red box) via a 'glass rod'. Subjects are instructed to track the vertical position of the target, which follows a smooth but seemingly unpredictable path $u(t) = A_1 \sin(\omega_1 t) + A_2 \sin(\omega_2 t)$. Frequencies ω_1, ω_2 are customized for each subject to assure that the task is not too slow and easy or too fast and hard.

Three separate tracking tasks are performed in 30 second blocks. In each block, the visual feedback of the tracking error is either unaltered, attenuated, or amplified, such that the cursor is displayed at $y(t) = u(t) + \alpha_i \left(s(t) + u(t)\right)$ with raw cursor position $s(t)$ and scaling factor α_i dependent on the mode $i \in \{\text{Normal}, \text{Better}, \text{Worse}\}$. In the Normal task ($\alpha_{\text{Normal}} = 1$), the vertical distance between the target and the displayed cursor displayed reflects the actual error generated by the subject. In the Better task ($\alpha_{\text{Better}} = 0.3$), the tracking error appears better than expected. In the Worse task ($\alpha_{\text{Worse}} = 2.0$), the tracking error is magnified, and appears worse than

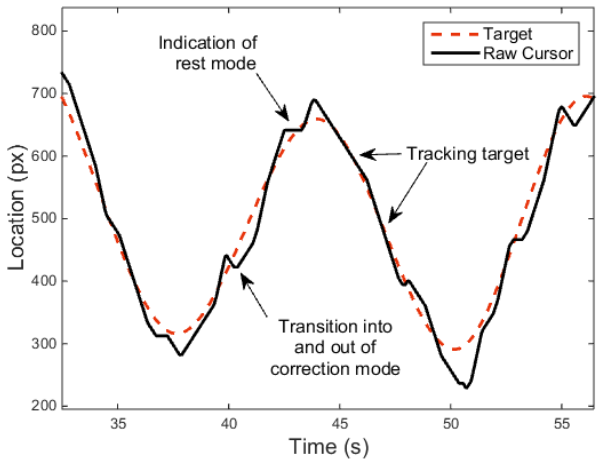


Figure 3.2: Typical data from the joystick tracking task reveals three types of submovements: one in which corrective motion occurs to redirect the cursor back towards the desired trajectory, one in which the user remains in a rest state, and one in which motion appears to follow the desired trajectory without any convergence in error.

expected. Subjects perform a total of 8×3 tasks. A set of training data was obtained for all subjects under the Normal task. Following training, 8 trials per subject were completed, each of which contained 3 tasks, alternating between Better-Worse-Better and Worse-Better-Worse.

Motor data is available at `http://www.unm.edu/~oishi/data`, courtesy of Dr. Martin J. McKeown. The available data consist of training data and experimental data for 3 prototypical normals subjects, and 3 subjects with Parkinson's disease, both on medication and off medication.

3.3 Problem Formulation

We start with the assumption that tracking a target using muscle movement involves nonconcurrent and distinct sensorimotor feedback systems. As seen in Figure 3.2, the

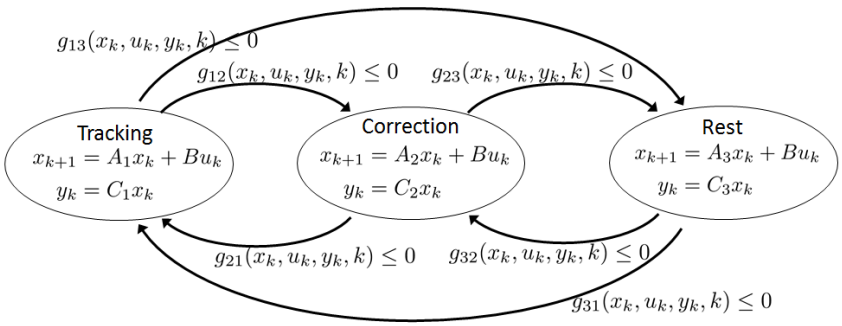


Figure 3.3: Hybrid dynamical model of submovements in manual tracking. Tracking represents a cursor-following submovement, Correction represents a corrective submovement, and Rest represents an idle state.

output generated by the system can reasonably mimic the movement of the target with the inclusion of an error deadzone, which is a known reaction ([37, 27, 38]). A difference equation that can simulate this behavior is given by

$$y[k+1] = y[k] + \beta_q \left(u[k] - u[k-1]\right) \tag{3.1}$$

where β_q is a real-valued constant that changes with mode $q \in \{1, \cdots, s\}$. For modeling purposes, we define the state $x[k] = [y[k-1], \cdots, y[k-n_a], u[k-1], \cdots, u[k-n_b]]^T$ with $n_a = 1$ and $n_b = 2$ as in (3.1) to obtain the discrete LTI system

$$\begin{aligned} x[k+1] &= \begin{bmatrix} 1 & \beta_q & -\beta_q \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} x[k] + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} u[k] \\ y[k] &= \begin{bmatrix} 1 & \beta_q & -\beta_q \end{bmatrix} x[k] \end{aligned} \tag{3.2}$$

The data collected on the tracking tasks (e.g. in Figure 3.2) shows evidence of at least three distinct modes characterizing the system dynamics. We therefore propose a three-mode hybrid model as shown in Figure 3.3 with discretized LTI dynamics

(3.2). This model arises in which transitions are described by unknown functions and the dynamics are parameterized in the state matrix by the real-valued constant β_q that changes with mode. The continuous state is $x \in \mathbb{R}^3$, the continuous input is $u \in \mathbb{R}$, and the continuous output is $y \in \mathbb{R}$. The discrete mode is $q[k] \in \{1, 2, 3\}$, in which mode 1 represents a cursor following submovement, mode 2 reflects a corrective submovement, and mode 3 indicates a rest state.

The selection of model (3.1) allows for computational advantages and agrees with our automaton selection. In particular, we benefit by the reduced complexity in the identification problem, i.e., instead of solving for $(n_a+n_b)\times s$ parameters in the system dynamics described in input-output form (1.3), we solve solely for s parameters. In addition, we can assign limits on β_q that concur with our mode assignments. This can be verified by (3.1) and a comparison with the target and cursor motions, namely: $\dot{u} > 0 \rightarrow u[k-1] - u[k-2] > 0$, and thus, $\beta_1 > 0 \rightarrow \mathsf{tracking}$, $\beta_2 < 0 \rightarrow \mathsf{correction}$, and $\beta_3 = 0 \rightarrow \mathsf{idle}$. Similarly, we find that when $\dot{u} < 0 \rightarrow u[k-1] - u[k-2] < 0$, and thus, $\beta_1 > 0 \rightarrow \mathsf{tracking}$, $\beta_2 < 0 \rightarrow \mathsf{correction}$, and $\beta_3 = 0 \rightarrow \mathsf{idle}$.

Problem 4 *Determine a) the dynamics that characterize submovements associated with cursor following and with corrections, and b) when submovements occur. That is, identify the model parameters $\beta_1, \beta_2 \in \mathbb{R}$, and the hybrid trajectory $(q[k], x[k])$ given a history $y[k]$ and the underlying model (Figure 3.3).*

3.4 System Identification

3.4.1 Embedding Approach

Mode sequence initialization

We initialize the mode sequence by first applying standard ARX identification for linear systems ([39])

$$\tilde{\theta}_k = \arg\min_{\theta} \left(y[k] - r[k]^T\theta\right)^2,\ r[k] = \begin{bmatrix} y[k-1] \\ u[k-1] \\ u[k-2] \end{bmatrix} \tag{3.3}$$

with regressor vector $r[k]$ and parameter vector $\theta = \left[1, \tilde{\beta}_k, -\tilde{\beta}_k\right]^T$ over small, moving time horizons $[k-\Delta, k+\Delta]$, $\Delta = 2$. Hence for each window indexed by time k, we obtain a single estimate for the gain coefficient $\tilde{\beta}_k$.

We use the resulting $\tilde{\theta}_k$, $k \in [0, T]$ to develop an initial mode trajectory $\hat{q}_k$

$$\hat{q}_k = \begin{cases} 1 & \text{for} \quad \tilde{\beta}_k > \delta \\ 2 & \text{for} \quad \tilde{\beta}_k < -\delta \\ 3 & \text{for} \quad \tilde{\beta}_k < |\delta| \end{cases} \tag{3.4}$$

with δ chosen close to zero to compensate for computational limitations. We initialize values for $\beta_1 = \frac{\sum_{k:\hat{q}_k=1} \tilde{\beta}_k}{\sum_{k:\hat{q}_k=1} k}$ and $\beta_2 = \frac{\sum_{k:\hat{q}_k=2} \tilde{\beta}_k}{\sum_{k:\hat{q}_k=2} k}$.

This, generates an initial switching sequence defined (with abuse of notation, indexing time by t_i, instead) as

$$\Sigma_0 = \{(t_0, \hat{q}_0(t_0)), (t_1, \hat{q}_0(t_1)), \cdots, (t_N, \hat{q}_0(t_N))\} \tag{3.5}$$

with switching instants $t_i, i \in \{0, \cdots, N\}$ between submovements.

Hybrid optimization

We form an optimization problem for switched systems with initialized (but not fixed) mode sequence and unknown switching times

$$\begin{aligned} \min_{\beta_1,\beta_2,q_{1,k},q_{2,k}} \quad & \sum_{k=1}^{N}\sum_{i=1}^{3}(y[k]-\hat{y}_{i,k})^2\sigma_{i,k} \\ \text{subject to} \quad & \sum_{i=1}^{3}\sigma_{i,k}=1, \quad \sigma_{i,k}\in\{0,1\} \end{aligned} \tag{3.6}$$

with a cost that represents the identification priority: to minimize the estimation error via a least-squares sense with respect to observed data $u[k]$, $y[k]$, with predicted output

$$\hat{y}_{q,k} = y[k-1] + \beta_q u[k-1] - \beta_q u[k-2] \tag{3.7}$$

Notice (3.6) is the same as the optimization derived by [15] for the *Minimum-of-Errors* (ME) estimator framework with $l(\cdot)=\varepsilon^2$.

With $N = 1650$, this results in a mixed-integer quadratic program [40, 41] with $3N$ integer variables $\sigma_{1,k}, \sigma_{2,k}, \sigma_{3,k}$ and 3 real-valued variables $\beta_1, \beta_2, \beta_3$. However, since our cost function and dynamics are so simple, we elect to use an embedded method [42] which relaxes the problem by transforming it into a constrained nonlinear program. This approach relaxes the discrete mode based on the fact that trajectories of the switched system are dense in the formulation of the embedded system. Thus, the relaxed nonlinear program becomes

$$\min_{\beta_q,\nu_{1,k},\nu_{2,k}} \sum_{k=1}^{N}(y[k]-\hat{y}[k])^2\nu_{1,k}+(y[k]-\hat{y}[k])^2\nu_{2,k}+(y[k]-\hat{y}[k])^2(1-\nu_{1,k}-\nu_{2,k}) \tag{3.8}$$

$$s.t. \begin{cases} 0\le\nu_{q,k}\le 1 & \forall q\in\{1,2\} \\ \nu_{1,k}+\nu_{2,k}\le 1 & \\ 0.5<\beta_1<6 & \\ -6<\beta_2<-0.5 & \\ \beta_3=0 & \end{cases}$$

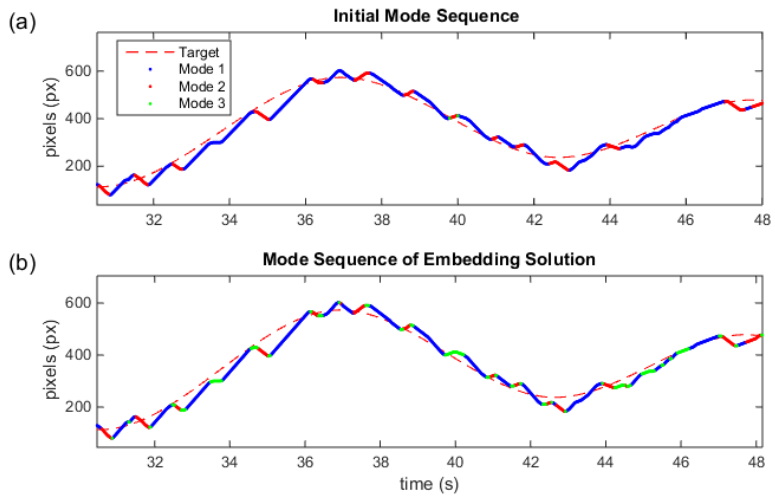


Figure 3.4: (a) Initialization of switching sequence via moving-horizon ARX identification. (b) Locally optimal switching sequence determined by embedded optimal solution.

with $\hat{y}[k]$ defined by (3.7), mode $\nu_{3,k} \triangleq 1-\nu_{1,k}-\nu_{2,k}$ and $\nu_{q,k} \in [0,1]\ \forall q \in \{1,2,3\}$, constraints on β_q obtained according to the formulation shown on Section 3.3. Moreover, the additional constraints of $\beta_1 \not< 0.5$ and $\beta_2 \not> -0.5$ are imposed to avoid over-penalizing $\beta_2 \approx \beta_3 \approx 0$ and instead look for alternate values of β_3 away from zero. The initial conditions are obtained from (3.4) and from β_1, β_2 determined by the mean values of $\tilde{\beta}_k$ in each mode.

By solving (3.8) we can obtain the solution to (3.6) with the projection $\sigma_{q,k} = 1$ for $q = \max_q \nu_{q,k}$. A typical initialization and optimization result is shown in Figure 3.4 and an academic example showing its effectiveness is shown in Appendix A.

3.4.2 Geometric Approach

In lieu of the simple dynamics (3.7) for the third mode, i.e., when $\hat{y}[k] = y[k-1]$, we are able to adapt our problem to the two mode system identification proposed

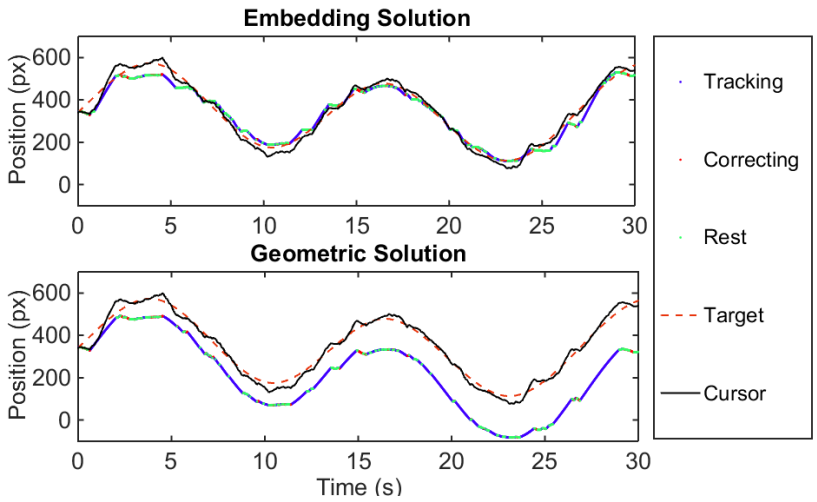


Figure 3.5: Comparison of the experimental data and the solution to the hybrid model obtained (a) via the embedding and (b) the geometric approach.

in [21]. The identification procedure goes as follows: First, we construct the set $\mathcal{S} = \{(\varphi[k], y[k])\}_{k=1}^{N}$ with the regressor $\varphi[k]$ and output $y[k]$ at all instances $k \in K = \{1, 2, \cdots, N\}$. Then, we extract the set corresponding to the rest submovements, $\mathcal{S}_r = \{(\varphi[i], y[i])\}_{i \in \mathcal{I}}$ where $\mathcal{I} = \{k \in K : y[k] = y[k-1]\}$. Lastly, we identify the parameters β_q and the switching sequence $q[k]$ for the remaining two submovements by using a geometric approach [21] for two-mode switched system identification, a method that is particularly robust to noise as compared to other methods [17, 20, 2, 15]. To identify the parameter vector in this case, we incorporate the identification of our model by using $\Theta \triangleq [1 \ \beta^* \ -\beta^*]$ instead of the generic parameter vector θ. In addition, we impose the initialization point boundaries, $0 < \beta^* < 10$, and after the Principal Component Analysis (PCA), we impose the more strict conditions $\beta_1 \in [0.5, 6]$, $\beta_2 \in [-6, -0.5]$ as before.

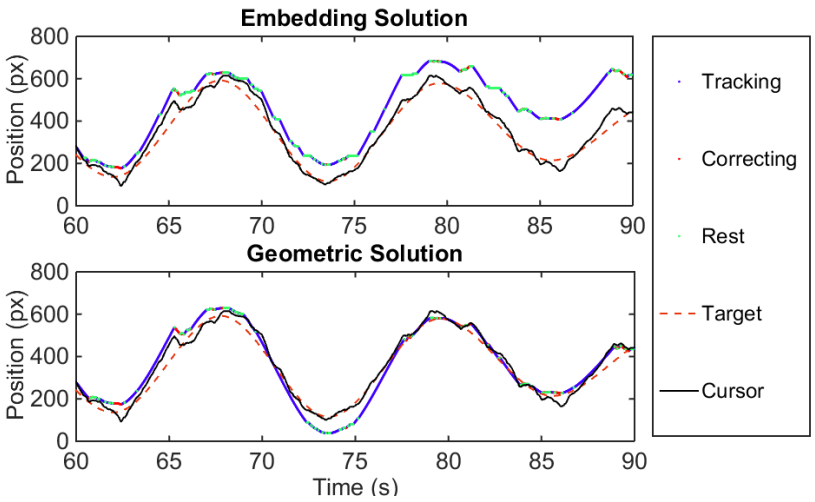


Figure 3.6: Comparison of the experimental data and the solution to the hybrid model obtained (a) via the embedding and (b) the geometric approach.

3.4.3 Results

The embedding approach discussed in subsection 3.4.1 and the geometric approach as described in subsection 3.4.2 were solved using MATLAB's constrained nonlinear minimization solver `fmincon`. In the case of the embedded nonlinear constrained problem (3.8), it took approximately 10 hours to run for all populations on a desktop with a 64-bit architecture with Intel Core i7-2600 3.40 GHz processor and 8.00 GB of RAM. Because the optimization inevitably falls in a local minima, we rely heavily on the initialization (also shown in Figure 3.4), which varies with parameter δ in (3.4). On the other hand, solving via the geometric approach (on the same computer) took roughly 3 minutes. This computational difference is expected since the minimization problem in the geometric approach is only done over the parameter vectors (i.e., β_1 and β_2) whereas the embedding approach minimizes over the switching sequence as well. By contrast, the geometric approach avoids the combinatorial problem by applying PCA on the mapping of the data points to parameter space to find the

switching sequence.

One issue that arises across all segments is the inconsistency of the 'supremacy' of one method over the other. We show two sample solutions showing the experimental data and its predicted output based on the proposed hybrid model using the embedding and the geometric approaches for segments 1 and 3 performed by the same subject on the same trial (Figure 3.5 and Figure 3.6, respectively). While in segment 1 the embedding solution seems to diverge less from the experimental data, segment 3 is better approximated by the geometric solution. Thus, we keep both approaches and study their results separately for different purposes.

In the geometric approach, we had separated the exact instances when resting occurs by creating the set $\mathcal{S}_r$ containing all instances with $\beta_3 = 0$. Then, by the classification scheme used by the geometric approach (i.e., PCA and optimization), we know we have classified (in the noiseless case) the remaining data accordingly (i.e., $\beta_1 > 0 \rightarrow$ tracking and $\beta_2 < 0 \rightarrow$ correction).

Using this scheme, then, we perform a paired-sample t-test on all Parkinson's subjects off and on medication based on the following parameters: (1) dwell time in each mode, (2) root-mean-square (RMS) tracking error at the instant prior to switching between modes, (3) RMS error accumulated during each mode, and (4) cursor speed relative to target speed at the instant prior to switching between modes. We find statistically significant trends in three parameters: in the dwell time in correction mode, $t(9) = 2.271$, $p < 0.0493$, in the dwell time in rest mode, $t(9) = 2.847$, $p < 0.0192$, and in the RMS tracking error when switching from tracking to rest, $t(9) = 2.342$, $p < 0.0439$. Note that these parameters should be interpreted under the assumptions of no dwell time constraints.

The reported p-values fall below the commonly accepted threshold for significance, i.e., $p < 0.05$, which assures that the likelihood of the null hypothesis (i.e.,

Table 3.1: Mean values and p-values of the parameters tested that resulted in significant using either the embedding or the geometric approaches. The mean and p-values are shown for both methods to show that the increasing and decreasing effects are similar although it might have not resulted significant in one of the methods.

Variable tested	Embedding			Geometric		
	PD off	PD on	p-value	PD off	PD on	p-value
Time in correction(s)	2.44	3.04	0.0510	2.72	3.79	0.0493
Time in rest(s)	7.28	6.65	0.1048	4.37	3.72	0.0192
RMSE at $g_{1,3}$(px)	40.39	45.24	0.1123	40.34	45.32	0.0439
Error speed at $g_{2,3}$(px/s)	-20.42	11.63	0.0018	-22.53	-8.68	0.0428
Error speed at $g_{3,2}$(px/s)	-13.99	1.55	0.0179	-17.53	-8.80	0.1467

the respective parameter off and on medication being from the same distribution) being true is less than significant. The mean values of dwell time in correction mode increased with medication (from 2.72 s to 3.79 s), as shown in Figure 3.8(a), those in rest mode decreased with medication (from 4.37 s to 3.72 s), as shown in Figure 3.8(b), while the mean values of RMS tracking error when switching from tracking to rest increased with medication (from 40.34 pixels to 45.32 pixels), as shown in Figure 3.8(c). While these may seem non-intuitive for a medicated state, it is well established that even appropriately prescribed levels of L-dopa can generate excessive motions (e.g., "overreaching" that is seen clinically).

Doing a paired t-test on the results obtained using the embedded solution, we see significance in only two parameters: error speed (in px/s) at the correction to rest transition and error peed at the rest to correction transition. The corresponding means and p-values obtained using the embedding and geometric methods for all of the paramteres showing significance are shown in Table 3.1

Following, we obtain the mean values for β_1 and β_2 per subject across all populations as shown in Figure 3.9(a) using the embedding and Figure 3.9(b) using the geometric approach. A resemblance among the β_1 values between both approaches can be seen: the geometric approach always obtained average solutions of β_1 that

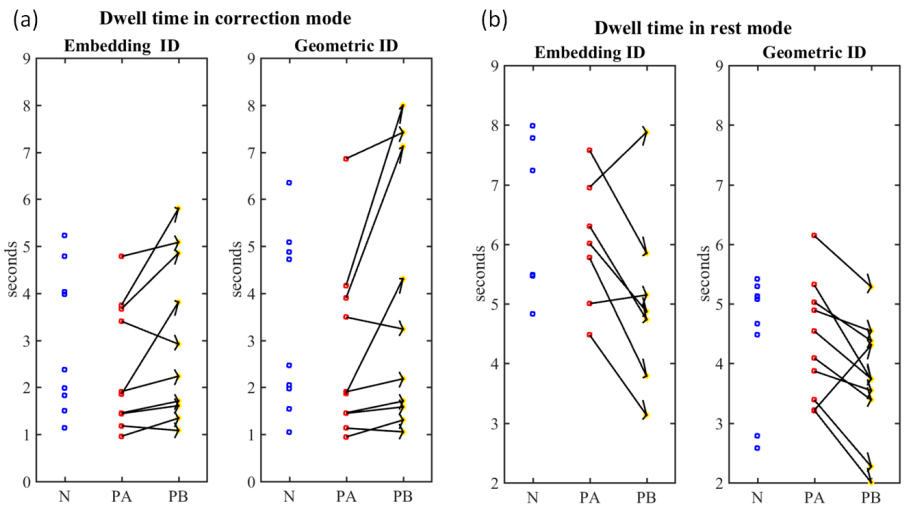


Figure 3.7: Statistical significance in a paired t-test between Parkinson's subjects off and on medication considering (a) dwell time in correction mode and (b) dwell time in rest mode. N = Normal, PA = Parkinson's off medication, and PB = Parkinson's on medication

were smaller than those from the embedding. This can be seen from the methodology that each approach was based on. Namely, by removing the set $\mathcal{S}_r$, we are leaving sampled trajectories (3.1) corresponding to $\beta_{1,k} \in (0, 0.5)$ add weight in the overall value of β_1, without allowing the solutions to enter the set of solutions in $\mathcal{S}_r$ thus decreasing its mean value. On the other hand, in the embedding approach we allow the optimization to classify a solution $\beta_{1,k} \in (0, 0.5)$ as reducing the cost by being in either $\beta_1 > 0.5$ or $\beta_3 = 0$, and hence the slight variability in the average β_1 values. In the case of β_2 (which doesn't follow this same trend), we presume that segments with insufficient number of instances in tracking mode could have not excited 'enough' the subsystem to allow its observation and lead to inaccurate values of β_2. This is supported by noticing that values of $\beta_1 \approx 1$ imply a tracking dynamic that is almost perfectly following the target, which will incur less correcting action.

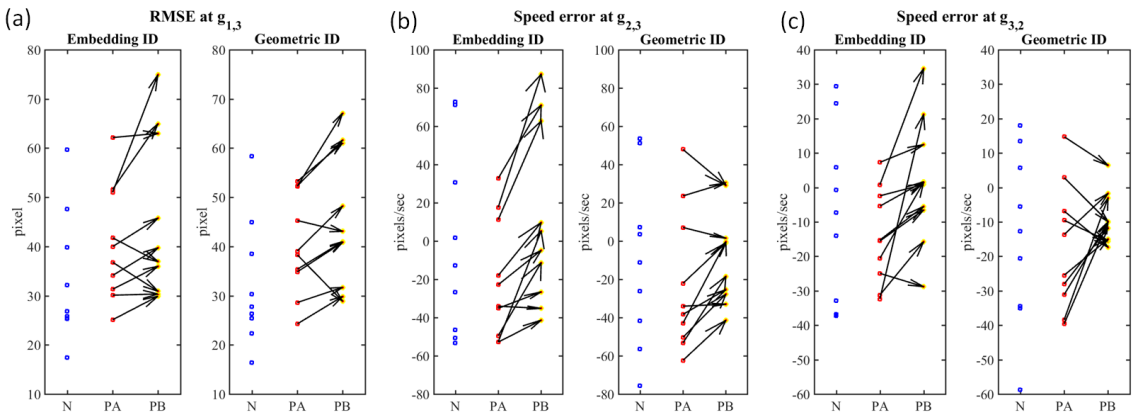


Figure 3.8: Statistical significance in a paired t-test between Parkinson's subjects off and on medication considering (a) RMS error when switching from tracking mode to rest mode, (b) error speed at the crossing of edge $E(2,3)$, and (c) error speed at the crossing of edge $E(3,2)$. N = Normal, PA = Parkinson's off medication, and PB = Parkinson's on medication

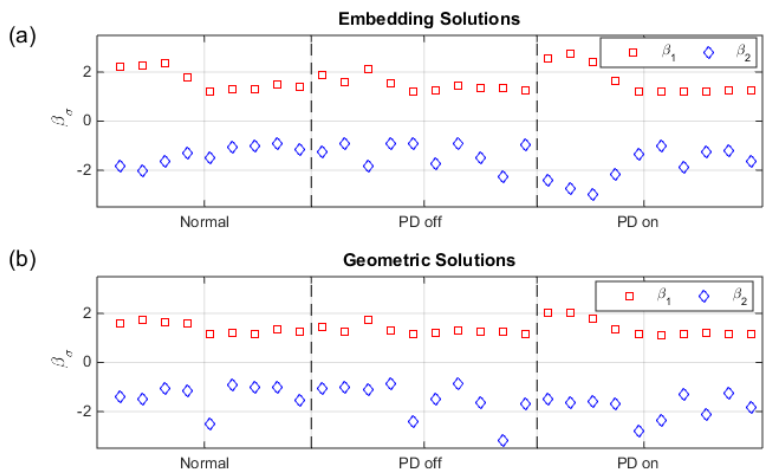


Figure 3.9: Mean values per subject obtained for β_1 and β_2 using (a) the embedding approach and (b) the geometric approaches.

3.5 Implications

Form the results shown in table 3.1, our results based on hybrid system modeling indicate that L-Dopa increases dwell time in 'correction' mode (from 2.72s to 3.79s, $p < 0.0493$) and decreases dwell time in 'rest' mode (from 4.37s to 3.72s, $p < 0.0192$). These results can be interpreted in the context of *cognitive inflexibility*, a characteristic cognitive deficit seen in PD[43]. The brain has a tradeoff between stability of representation ("continue what you are doing" whose extreme example is Obsessive Compulsive Disorder) vs flexibility ("try something new" which in excessive amounts results in Attention Deficit Disorder). Previous studies have suggested an "L-dopa overdose hypothesis" whereby L-dopa given to improve motor performance, may paradoxically worsen performance on switching tasks by promoting impulsiveness [44]. There may also be a classic "inverted-U-shaped" relationship between dopaminergic levels and performance. Our results indicating a higher RMS error attained when moving from 'tracking' mode to 'idle' mode suggest that in addition to enhanced impulsiveness, L-dopa results in a higher error tolerance in the triggering based on a prospective error-based switching logic (we get mean values for RMSE at switching instants from switching at 40.34px to 45.34px, $p < 0.0439$).

Chapter 4

Estimation of the Guard Conditions in Parkinson's Disease

We seek to assess whether transitions both within and between modes are dependent on time, error, a certain degree of stochasticity, or other factors. Optimization methods simply detect times that transition occur – they do not hypothesize possible mechanisms for the transitions. Since there is no clear neurological theory to identify mechanism underlying motor program or submovement selection, we evaluate a variety of possible scenarios. We first consider switching surfaces that are affine hyperplanes, based on visual inspection of the data.

4.1 Problem Formulation

Presume the hybrid dynamics of submovements are linear in each mode,

$$y[k] = y[k-1] + \beta_{q[k-1]}(u[k-1] - u[k-2]) + e[k] \tag{4.1}$$

with parameter $\beta_q \in \mathbb{R}$ a constant value in each of the modes $q \in Q = \{\textbf{Tracking}, \textbf{Rest}, \textbf{Correction}\}$. Further presume that the optimal switching sequence and optimal values of β_q have already been identified and hence are known.

We can rewrite the dynamics (4.1)

$$y[k] = \begin{bmatrix} 1 & 0 & \beta_q & -\beta_q \end{bmatrix} \boldsymbol{r}[k] + e[k] \tag{4.2}$$

in terms of the standard regression vector

$$\boldsymbol{r}[k] = \begin{bmatrix} y[k-1] & u[k] & u[k-1] & u[k-2] \end{bmatrix}^\top \tag{4.3}$$

We presume that the switching surface for any mode pair q, q' is an affine function of the output and regressor, that is,

$$g_{q,q'}(y[k], r[k]) = y[k] - \boldsymbol{\alpha}_{q,q'}^\top \boldsymbol{r}[k] - \gamma_{q,q'} \tag{4.4}$$

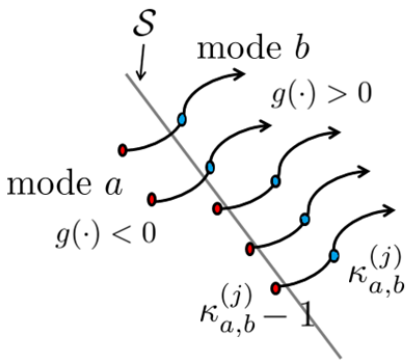
with $\alpha_{q,q'} \in \mathbb{R}^4$, $\gamma_{q,q'} \in \mathbb{R}$, such that the transition from mode q to mode q' occurs when $g_{q,q'}(\cdot,\cdot) \geq 0$. Or equivalently, using standard notation for hybrid system [3], for the edge $E(q, q')$, mode q' becomes active when $g_{q,q'}(\cdot,\cdot) \geq 0$.

Using this knowledge, we pose the following problem.

Problem 5 *Given the input-output pair* $\{(y[k], u[k])\}_{k=1}^N$*, the corresponding active modes* q_k*, and the affine component* $\boldsymbol{r}[k]$ *from* (4.4) *on which the guard lies, find the guard parameters* $\alpha_{q,q'}$ *and* $\gamma_{q,q'}$ *for all of the edges of the hybrid system.*

Moreover, we would like to determine what is the sufficient amount of information needed to compute a guard for the switched system. Hence, we analyze the true coefficients $\alpha_{q,q'}$ and $\gamma_{q,q'}$ of the switching surface (4.4) in lieu of alternate solutions $\tilde{\alpha}_{q,q'}, \tilde{\gamma}_{q,q'}$, formally phrased as follows.

Figure 4.1: Graphical representation of the active mode and role of switching surface in a generic hybrid system. In this case, mode q is described by $g_{q,q'}(\cdot,\cdot) < 0$. The blue dots correspond to the time instant k at which transition i occurs, and the red dots correspond to the time instances k just before the jump i occurs.



Problem 6 *If we know the true solution to the problem of finding the (well-posed) guard parameter $w^* \in \mathbb{R}^n$ in* (4.20)*, how does the true solution w^* compare to solving the optimization problem* (4.20) *in a higher dimensional parameter space and in a lower dimensional space than the true parameter space?*

4.2 Estimation of the guard parameter

By evaluating the switching equation at the first instant after each transition, we obtain

$$\begin{bmatrix} y[\tau_1] \\ y[\tau_2] \\ \vdots \\ y[\tau_N] \end{bmatrix} = \begin{bmatrix} \boldsymbol{r}[\tau_1]^\top & 1 \\ \boldsymbol{r}[\tau_2]^\top & 1 \\ \vdots & \vdots \\ \boldsymbol{r}[\tau_N]^\top & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \gamma \end{bmatrix} + \begin{bmatrix} \epsilon[\tau_1] \\ \epsilon[\tau_2] \\ \vdots \\ \epsilon[\tau_N] \end{bmatrix} \tag{4.5}$$

where we define $\tau_i, i \in \{1, \cdots, N\}$ as the first instant in mode q', as shown in Figure 4.1. That is, at time $k = \tau_i - 1$, the mode is q, and at time $k = \tau_i$, the mode is q'.

We denote (4.5) in short form as

$$\boldsymbol{y}_\tau = A_\tau \boldsymbol{x} + \boldsymbol{\epsilon}_\tau \tag{4.6}$$

One common approach to finding $\alpha_{q,q'}, \gamma_{q,q'}$ that describes the switching surface would be via the least squares problem

$$\begin{aligned} \text{minimize} \quad & \|\boldsymbol{y}_\tau - A_\tau \boldsymbol{x}\| \\ \text{subject to} \quad & \boldsymbol{y}_\tau - A_\tau \boldsymbol{x} \geq \boldsymbol{0} \end{aligned} \tag{4.7}$$

with $\boldsymbol{x} = \begin{bmatrix} \alpha \\ \gamma \end{bmatrix} \in \mathbb{R}^5$, $\boldsymbol{y} = \begin{bmatrix} y[\tau_1] & y[\tau_2] & \cdots & y[\tau_N] \end{bmatrix}^T \in \mathbb{R}^N$, and $A_\tau \in \mathbb{R}^{N\times 5}$ as described in (4.5) with matrix A_τ being full rank, presuming the elements of the regressor are linearly independent. Note that despite the assumption of linear independence of columns in A_τ, we have to address the fact that since matrix A_τ contains the regressor vectors, whose elements can be very close to each other, we are prone to having an *ill-conditioned* matrix A_τ. Hence, any formulation or algorithm requiring the inversion of matrix A_τ will result in solutions which will be sensitive to small variations in the elements of matrix A_τ and elements in vector y_τ.

On the other hand, note that solving this problem via (4.7) will result in a solution resembling the dynamics. This can be seen by noticing that for small noise terms $e[k]$, the dynamics description (4.1) can be described by (4.5) with γ being a small value compensating for the small noise term $e[k]$. In the next section, we show how solving the least squares problem can easily lead to erroneous solutions.

4.2.1 Relationship to Least Squares Estimation

In here, we show the numerical issue that arises in trying to solving problem (4.7) and which prevents us from recovering the guard coefficients by this approach. We

show this by expanding the arguments in (4.7). Each $i^{\texttt{th}}$ row in (4.5) can be rewritten using (4.1), in the noise-free case, as

$$\begin{aligned} y_{\tau_i} - A_{\tau_i}x &= \big(y[\tau_i - 1] + \beta_q(u[\tau_i - 1] - u[\tau_i - 2])\big) - \big(\alpha_{y_1}y[\tau_i - 1] \\ &\qquad + \alpha_{u_0}u[\tau_i] + \alpha_{u_1}u[\tau_i - 1] + \gamma\big) \qquad (4.8) \\ &= (1 - \alpha_{y_1})y[\tau_i - 1] - \alpha_{u_0}u[\tau_i] + (\beta_q - \alpha_{u_1})u[\tau_i - 1] \\ &\qquad - (\beta_q u[\tau_i - 2] + \gamma) \qquad (4.9) \end{aligned}$$

which substituting in (4.7) results in

$$\begin{aligned} &\min. \sqrt{\sum_{i=1}^{N} \Big((1 - \alpha_{y_1})y[\tau_i - 1] - \alpha_{u_0}u[\tau_i] + (\beta_q - \alpha_{u_1})u[\tau_i - 1] - (\beta_q u[\tau_i - 2] + \gamma) \Big)^2} \\ &\text{s.t}(1 - \alpha_{y_1})y[\tau_i - 1] - \alpha_{u_0}u[\tau_i] + (\beta_q - \alpha_{u_1})u[\tau_i - 1] - (\beta_q u[\tau_i - 2] + \gamma) \geq 0, \\ &\quad \forall\ i = 1, 2, \cdots, N \end{aligned} \qquad (4.10)$$

Let the argument in parenthesis in (4.10) be $J_i(x) \triangleq \big(\ \cdot\ \big)$ and let a feasible point resembling the dynamics be of the form $x_{\text{dyn}} = [1\ \beta_q\ -\beta_q\ \gamma]$. Then we get

$$\begin{aligned} J_i(x_{\text{dyn}}) &= (1 - \alpha_{y_1})y[\tau_i - 1] - \alpha_{u_0}u[\tau_i] + (\beta_q - \alpha_{u_1})u[\tau_i - 1] \\ &\qquad - (\beta_q u[\tau_i - 2] + \gamma)\Big|_{\substack{\alpha_{y_1} = 1 \\ \alpha_{u_0} = \beta_q \\ \alpha_{u_1} = -\beta_q}} \\ &= -\beta_q u[\tau_i] + 2\beta_q u[\tau_i - 1] - \beta_q u[\tau_i - 2] - \gamma \qquad (4.11) \end{aligned}$$

Next, let us consider the cost associated with the solution of the actual guard condition $x_{\text{guard}} = [\alpha_{y_1}\ \alpha_{u_0}\ \alpha_{u_1}\ \gamma]$. We do this by procedures (4.12) thru (4.15),

explained next. Let us add and subtract the same amounts from $J_i(x_{\text{guard}})$ (as shown in equation (4.12)), then expand $J_i(x_{\text{guard}})$ in (4.12) (as shown in equation (4.13)), and collect terms as in equation (4.14).

$$J_i(x_{\text{guard}}) = J_i(x_{\text{guard}}) + \beta_q u[\tau_i] - \beta_q u[\tau_i] + \beta_q u[\tau_i - 1] - \beta_q u[\tau_i - 1] \tag{4.12}$$

$$= (1 - \alpha_{y_1}) y[\tau_i - 1] - \alpha_{u_0} u[\tau_i] + (\beta_q - \alpha_{u_1}) u[\tau_i - 1] - \beta_q u[\tau_i - 2] - \gamma + \beta_q u[\tau_i] - \beta_q u[\tau_i] + \beta_q u[\tau_i - 1] - \beta_q u[\tau_i - 1] \tag{4.13}$$

$$= (1 - \alpha_{y_1}) y[\tau_i - 1] - \alpha_{u_0} u[\tau_i] - \alpha_{u_1} u[\tau_i - 1] + \beta_q u[\tau_i] - \beta_q u[\tau_i - 1] + \left(-\beta_q u[\tau_i] + 2\beta_q u[\tau_i - 1] - \beta_q u[\tau_i - 2] - \gamma \right) \tag{4.14}$$

Lastly, we note that the last term in parenthesis in (4.14) is nothing but $J_i(x_{\text{dyn}})$ as written in (4.11), which means that

$$\begin{aligned} J_i(x_{\text{guard}}) &= (1 - \alpha_{y_1}) y[\tau_i - 1] + (\beta_q - \alpha_{u_0}) u[\tau_i] - (\beta_q + \alpha_{u_1}) u[\tau_i - 1] + J_i(x_{\text{dyn}}) \\ &= \psi_{i(\boldsymbol{\alpha}, \beta_q)} + J_i(x_{\text{dyn}}) \end{aligned} \tag{4.15}$$

Hence, since it suffices to have $J_i(x_{\text{guard}}) > J_i(x_{\text{dyn}}) \; \forall \; i = 1, 2, \cdots, N$ to get the solution x_{dyn} and not x_{guard} under the problem formulation (4.7), we can say that it suffices to show that

1. for $J_i(x_{\text{guard}}) > 0$ and $J_i(x_{\text{dyn}}) > 0$, then $\psi_{i(\boldsymbol{\alpha}, \beta_q)} > 0 \implies J_i(x_{\text{dyn}}) < J_i(x_{\text{guard}})$, and

2. for $J_i(x_{\text{guard}}) < 0$ and $J_i(x_{\text{dyn}}) < 0$, then $\psi_{i(\boldsymbol{\alpha}, \beta_q)} < 0 \implies J_i(x_{\text{dyn}}) < J_i(x_{\text{guard}})$

The above conditions are very limiting as to which guard coefficient we are properly able to recover in our minimization formulation, and without *a priori* knowledge

of what form the guard takes (i.e., which parameters it depends on), it becomes less reliable. This poses a big issue in our capacity to recover the guard of the system since solving the convex minimization problem (4.7) can easily result in a solution that we are not interested in.

Therefore, we consider the data before transition, i.e., at times $\tau_i - 1$, in conjunction with times after transitioning, i.e., times τ_i and solve the problem from a classification perspective.

4.2.2 Solution via Classification

The transition relation in (4.4) can be extended to account for two sets of data which lie in the half-spaces described as a function of each transition mode pair q, q' as

$$w_{q,q'}^\top x_i - \gamma_{q,q'} \geq 0, \quad i = 1, \cdots, N \tag{4.16a}$$

$$w_{q,q'}^\top z_i - \gamma_{q,q'} \leq 0, \quad i = 1, \cdots, N \tag{4.16b}$$

with $w_{q,q'} = [1 \text{ -}\alpha_{q,q'}]^\top$, and data features x_i and z_i corresponding to the i^{th} rows in

$$X = \begin{bmatrix} y[\tau_1] & \boldsymbol{r}^\top[\tau_1] \\ \vdots & \vdots \\ y[\tau_N] & \boldsymbol{r}^\top[\tau_N] \end{bmatrix} \quad \text{and} \quad Z = \begin{bmatrix} y[\tau_1 - 1] & \boldsymbol{r}^\top[\tau_1 - 1] \\ \vdots & \vdots \\ y[\tau_N - 1] & \boldsymbol{r}^\top[\tau_N - 1] \end{bmatrix} \tag{4.17}$$

containing the output y and regressor vector $\boldsymbol{r}$ similar to (4.3) on which the linear guard exists.

We presume our data to be noisy, so it will be convenient to look for the hyperplane that separates the data points with the largest margin. That is, we look for the hyperplane that separates the data according to (4.16) *and* maximizes the distance to the closest points to the hyperplane. This is the same problem as that solved by Support Vector Machines [45], which is summarized next.

For a given edge $E(q, q')$, we have a hyperplane described by $w^\top x - \gamma = 0$, with the closest point given by

$$d_{\min} = \min_i \left\{ \frac{w^\top x_i - \gamma}{\|w\|}, \frac{-w^\top z_i + \gamma}{\|w\|} \right\} \tag{4.18}$$

where each of the arguments denotes the distance to each data point x_i and z_i, as shown in Figure 4.2. Then, the hypeprlane with the largest margin is found by solving the problem

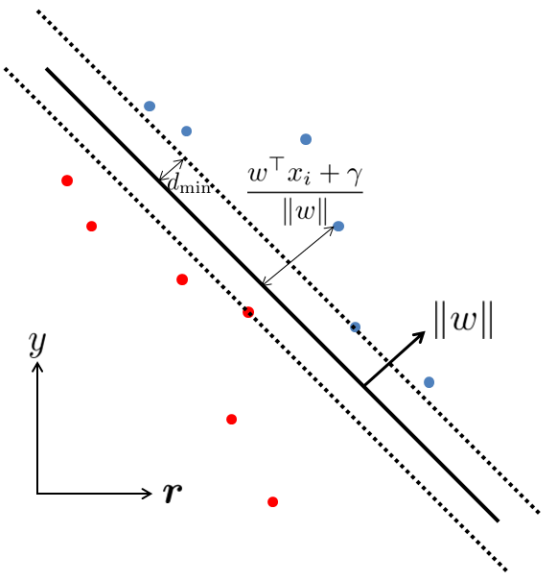
$$\begin{aligned} \text{maximize} \quad & d_{\min} \\ \text{subject to} \quad & w^\top x_i - \gamma \geq d_{\min} \\ & w^\top z_i - \gamma \leq -d_{\min} \\ & \|w\|_2 = 1 \end{aligned} \tag{4.19}$$

where we have used the fact [46] that at the the optimal solution, $\|w^*\|_2 = 1$. In addition, since w and γ are homogeneous, scaling the cost and constraints in (4.19) will not affect the optimal solution, we choose then $d_{\min}\|w\|_2 = 1$ and use the fact that maximizing $\frac{1}{\|w\|}$ is the same as minimizing $\|w\|$, to solve problem (4.19) via the following equivalent quadratic programming problem

$$\begin{aligned} \underset{w,\gamma}{\text{minimize}} \quad & \frac{1}{2}\|w\|_2^2 \\ \text{subject to} \quad & w^\top x_i - \gamma \geq 1, \quad i = 1, 2, \cdots, N \\ & w^\top z_i - \gamma \leq -1, \quad i = 1, 2, \cdots, N \end{aligned} \tag{4.20}$$

Often times, specially when the data does not seem to be separable by a hyperplane, one may seek to perform a nonlinear transform to the data (a.k.a. feature mapping) in order to find the correct classification boundaries. In these cases, the dual problem of (4.20) is studied for the reasons discussed next. After computing the Lagrangian, and computing the appropriate derivatives, we arrive at the following

Figure 4.2: Plot of output values, y, vs the regressor which the guard depends on, $\boldsymbol{r}$, at transition times τ (in blue) and $\tau - 1$ (in red).



dual problem.

$$\begin{aligned} \underset{\lambda}{\text{maximize}} \quad & \sum_{i=1}^{2N} \lambda_i - \frac{1}{2} \sum_{i=1}^{2N} \sum_{j=1}^{2N} \xi_i \xi_j \lambda_i \lambda_j \tilde{x}_i^\top \tilde{x}_j \\ \text{subject to} \quad & \lambda_i \geq 0, \quad i = 1, \cdots, 2N \\ & \sum_{i=1}^{2N} \lambda_i \xi_i = 0 \end{aligned} \tag{4.21}$$

with Lagrange multipliers λ_i, labels $\xi_i \in \{-1, +1\}$, and where $\tilde{x}_i$ corresponds to the i^{th} row of the entire data points now collected in

$$\tilde{X} = \begin{bmatrix} X \\ Z \end{bmatrix} \tag{4.22}$$

Also, from the Karush-Kuhn-Tucker (KKT) conditions [46], we get the following relationship between w and the Lagrange multipliers λ_i,

$$w = \sum_{i=1}^{2N} \lambda_i \xi_i \tilde{x}_i \tag{4.23}$$

The dual problem (4.21) is convenient in that the dot product $< x_i, x_j >$, seen in classification as the *linear Kernel*, can be often replaced with different Kernels providing a nonlinear transformation of the data, known as feature mapping. Some of the most common Kernels include the polynomial Kernel, radial basis, and neural networks [47]. Problem (4.21) is also convenient in that it is a quadratic programming problem which can be solved via Matlab's `quadprog` solver. Note, however, that solving the classification problem via (4.21) involves the computation of $\tilde{X}^\top \tilde{X}$, which, for an ill-conditioned matrix X results in a matrix $\tilde{X}^\top \tilde{X}$ with an even higher condition number, hence we solve the primal problem (4.20) as opposed to the dual (4.21) with the linear Kernel and consider a variation later in this work.

For data that is not linearly separable or is expected to have outliers, it is common to consider heuristic approaches, such as the ℓ_1 regularization, i.e., to consider the following problem

$$\begin{aligned} \underset{w,\gamma,\xi}{\text{minimize}} \quad & \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{N}\xi_i \\ \text{subject to} \quad & w^\top x_i - \gamma \geq 1 - \xi_i, \qquad i = 1, \cdots, N \\ & w^\top z_i - \gamma \leq -(1 - \xi_i), \quad i = 1, \cdots, N \\ & \xi_i \geq 0, \qquad\qquad\qquad i = 1, \cdots, N \end{aligned} \tag{4.24}$$

which includes in the cost the trade-off between the width of the gap $d_{\min}$ (first term in cost function) and the number of misclassified data points (second term in cost function). In this framework, a large coefficient C results in a more strict separating hyperplane allowing less missclassified data points, while a small C value has a lower penalty and thus results in a wider gap $d_{\min}$. We now look at the dual of the SVM

problem (4.24), which is

$$\begin{aligned} \underset{\lambda}{\text{maximize}} \quad & \sum_{i=1}^{2N} \lambda_i - \frac{1}{2} \sum_{i=1}^{2N} \sum_{j=1}^{2N} \xi_i \xi_j \lambda_i \lambda_j \tilde{x}_i^\top \tilde{x}_j \\ \text{subject to} \quad & 0 \leq \lambda_i \leq C, \quad i = 1, \cdots, 2N \\ & \sum_{i=1}^{2N} \lambda_i \xi_i = 0 \end{aligned} \tag{4.25}$$

which is of the same form as (4.21) except that the constraint on λ_i has the new upper bound C. The cost in (4.25) can be rewritten in matrix form as

$$L_D = \mathbf{1}^\top \boldsymbol{\lambda} - \frac{1}{2} \boldsymbol{\lambda}^\top \boldsymbol{\xi} R \boldsymbol{\xi} \boldsymbol{\lambda} \tag{4.26}$$

with $\boldsymbol{\lambda} \in \mathbb{R}^{2N}$, $\boldsymbol{\xi} = \text{diag}(\xi_1, \cdots, \xi_{2N})$, and $R_{ij} =< x_i^\top x_j >$. The ill-conditioning of our data, will appear then in R. This can be overcome by adding a small coefficient to the diagonal elements in R [48], i.e., by using the following Lagrangian, instead of (4.26),

$$L_D = \mathbf{1}^\top \boldsymbol{\lambda} - \frac{1}{2} \boldsymbol{\lambda}^\top \boldsymbol{\xi} (R + \delta I) \boldsymbol{\xi} \boldsymbol{\lambda} \tag{4.27}$$

which will increase the smallest singular values of R, and hence reduce the condition number. It was shown in [48] that adding δ to the diagonal elements in R has an equivalent effect as modifying C in (4.24) in the following manner: it is equivalent to applying a quadratic cost function to the slacks ξ_i whose values are $\xi_i < \delta C$ and a linear cost to the others. In this regard, solving the regularized problem (4.24) aids dealing with our ill-conditioned matrices X and Z.

In the following subsection, we show some illustrative examples of this approach in spite of recovering the guard parameters $\alpha_{q,q'}$ and $\gamma_{q,q'}$. In both cases, we use data that is linearly separable and show results based on two different cases: (1) when solving problem (4.20) and (2) when using the heuristic approach with ℓ_1 regularization, hence solving problem (4.24).

4.2.3 Examples

Example 1: 3-D toy data

Consider a system whose input is $u[k] = (f_s k)^{1.8}$ with $f_s = 0.1$, $k = 1, 2, \cdots, N$, with $N = 500$. The output is given by

$$y[k] = \theta_q^\top \varphi[k] + e[k] \tag{4.28}$$

with $q \in \{1, 2\}$, parameter vectors $\theta_1 = [1\ 1.5\ \text{-}1.5]^\top$, $\theta_2 = [1\ \text{-}0.5\ 0.5]^\top$, regressor vector $\varphi[k] = [y[k-1]\ u[k-1]\ u[k-2]]^\top$, and error term $e \sim \mathcal{N}(0, 1)$. The guard condition is given by

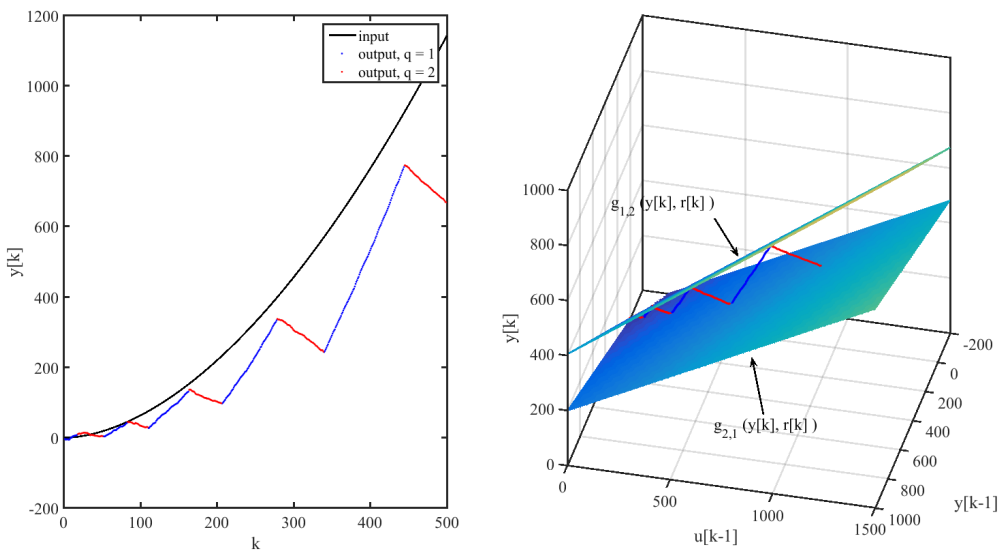
$$g_{q,q'}(y[k], \boldsymbol{r}[k]) = y_k - \boldsymbol{\alpha}_{q,q'}^\top \begin{bmatrix} y[k-1] \\ u[k-1] \end{bmatrix} - \gamma_{q,q'} \tag{4.29}$$

where we presume we readily know that the guard condition depends on the parameters $\boldsymbol{r}[k] = [y[k-1]\ u[k-1]]^\top$. In this example, we have set $\boldsymbol{\alpha}_{1,2} = [0.4\ 0.5]^\top$, $\gamma_{1,2} = 5$ and for the other edge, the coefficients vector is set to $\boldsymbol{\alpha}_{2,1} = [0.2\ 0.35]^\top$, $\gamma_{2,1} = -3$.

Dynamics (4.28) with switching rule (4.29) specify one set of input-output data, as shown in Figure (4.3). Given the modes at each instant, we find the transition times τ_i and solve the quadratic program corresponding solving the largest separating hyperplane problem (4.20) with Matlab's `quadprog` solver to obtain a candidate coefficient vector $\alpha_{q,q'}$ for edges $E(1, 2)$ and $E(2, 1)$ describing a hyperplane that separates the data by the largest margin.

Since the input-output data generated from a single initial condition spans a very reduced space in $\mathbb{R}^3$, as seen from the right plot in Figure 4.3, we are prone to not being able to capture a representative guard/hyperplane solution. To show this, let us consider the alternate hyperplanes in Figure 4.4, which are also a hyperplane satisfying the feasible solution (4.16) separating the two data sets X and Z.

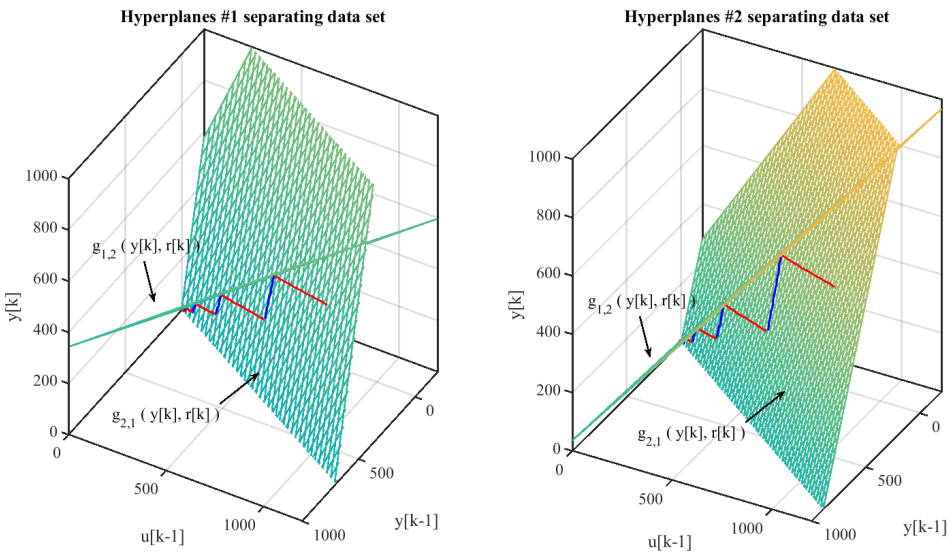
Figure 4.3: Single set of input-output data generated by system (4.28) with switching rule (4.29). Although the output seems to switch in a nonlinear fashion with respect to time (left), it can be seen that the switching rule is actually an affine function of the regressor vectors $y[k-1]$ and $u[k-1]$ (right).



In order to obtain a more representative hyperplane, we propose using multiple (and different) sets of input-output data with dynamics (4.28) and same switching rule (4.29) to see how the toy system dynamics (4.28) evolve in other areas of the space $\mathbb{R}^3$. Then, by using different sets of initial conditions to the toy system dynamics (4.28) with switching rule (4.29), we solve the non-regularized quadratic program (4.20) using the information from all transition times from all input-output sets of data. We get the resulting coefficient vectors shown in Figure 4.5 for multiple numbers of data sets. As seen in the figure, the solution converges to the true solution as the amount of data increases.

The solutions obtained using the regularized problem formulation are as follows. When a small value of C is used in the regularized solution, e.g., $C = 1$, we obtain the (estimated) guard parameters shown in Table 4.2.3. The (estimated) guard pa-

Figure 4.4: Multiple candidate hyperplanes that separate the transition times according to (4.16) but that are different from the known true hyperplane with coefficients $\boldsymbol{\alpha}$.

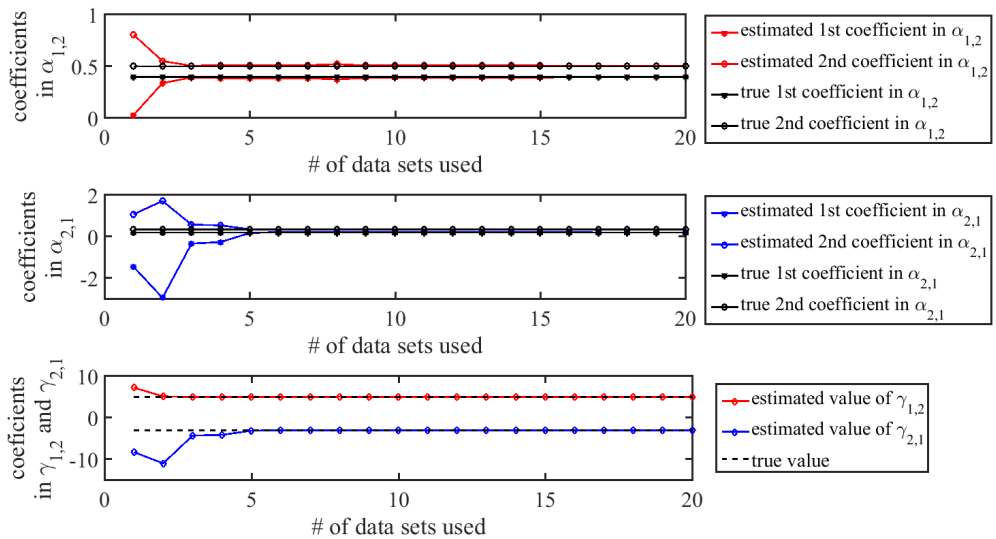


rameters obtained via the regularized solution result in the violation of some of the constraints of (4.20), but are, of course, still feasible in the regularized formulation (4.24). As shown in Table 4.2.3, the solutions of the non-regularized and the rgularized solution, do not diverge from the true solution from each other (nor from the true solution), hence we know that the solution to the quadratic program (4.20) which strictly separates all data points is a valid guard parameter solution despite

Table 4.1: True guard parameters used in example 1, along with the estimated guard coefficients obtained when solving the non-regularized problem (4.20) and the regularized problem (4.24).

		Solution	
Parameters	True Coefficients	Non-regularized	Regularized
$\alpha_{1,2}$	$[0.4\ 0.5]^\top$	$[0.393\ 0.506]^\top$	$[0.407\ 0.495]^\top$
$\gamma_{1,2}$	5	5.014	4.437
$\alpha_{2,1}$	$[0.2\ 0.35]^\top$	$[0.213\ 0.345]^\top$	$[0.213\ 0.345]^\top$
$\gamma_{2,1}$	-3	-2.960	-2.960

Figure 4.5: Estimated solution from solving QP (4.20) using multiple sets of initial conditions showing that the estimated solutions converge to the true solution as the amount of data increases. We see that the solutions converge after 10 iterations for both the coefficient vectors $\alpha_{1,2}, \alpha_{2,1}$, and the guard affine constant terms $\gamma_{1,2}$ and $\gamma_{2,1}$.



the ill-condition matrices X and Z.

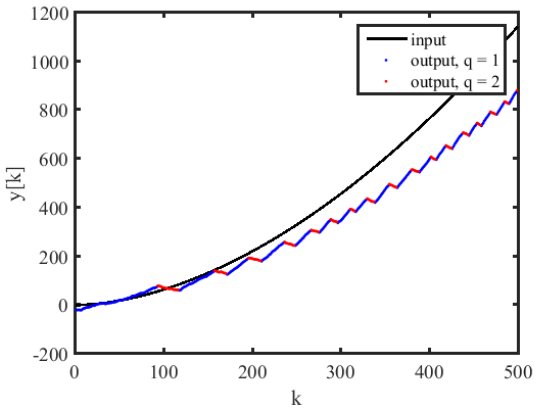
Example 2: 4-D toy data

In this case, we consider the system with same dynamics (4.28) and guard condition

$$g_{q,q'}(y[k], \boldsymbol{r}[k]) = y_k - \boldsymbol{\alpha}_{q,q'}^\top \begin{bmatrix} y[k-1] \\ u[k] \\ u[k-1] \end{bmatrix} - \gamma_{q,q'} \tag{4.30}$$

with coefficient vector $\boldsymbol{\alpha}_{1,2} = [0.8\ 0.75\ -0.6]^\top$, $\boldsymbol{\alpha}_{2,1} = [0.8\ 0.2\ -0.05]^\top$, $\gamma_{1,2} = 8$, and $\gamma_{2,1} = -1$. Again, we presume we know that the guard is a function of $\boldsymbol{r}[k] =$

Figure 4.6: Single set of input-output data generated by system (4.28) with switching rule (4.30) plotted vs time resulting in 17 transition on edge $E(1,2)$ and 16 transitions on edge $E(2,1)$.



$[y[k-1]\ u[k]\ u[k-1]]^\top$. Figure 4.6 shows a sample set of input-output data for this system.

Figure 4.7: Plots showing the convergence results on the guard coefficients vectors $\alpha_{1,2}$, $\alpha_{2,1}$, and affine constants $\gamma_{1,2}$ and $\gamma_{2,1}$.

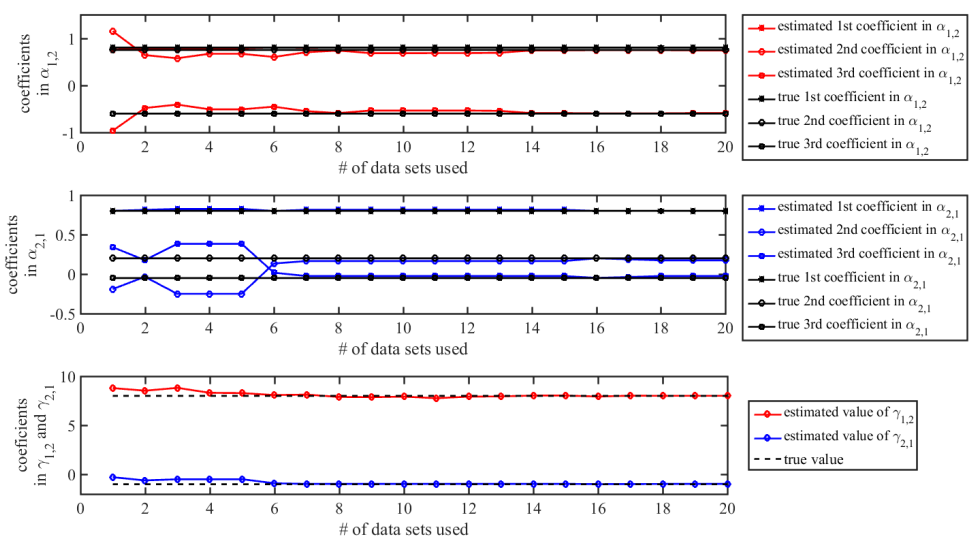


Table 4.2: True guard parameters used in example 2, along with the estimated guard coefficients obtained when solving the non-regularized problem (4.20) and the regularized problem (4.24).

		Solution	
Parameters	True Coefficients	Non-regularized	Regularized
$\alpha_{1,2}$	$[0.8\ 0.75\ -0.6]^\top$	$[0.805\ 0.821\ -0.675]^\top$	$[0.807\ 0.428\ -0.281]^\top$
$\gamma_{1,2}$	8	7.769	7.994
$\alpha_{2,1}$	$[0.8\ 0.2\ -0.05]^\top$	$[0.805\ 0.098\ 0.049]^\top$	$[0.805\ 0.204\ -0.057]^\top$
$\gamma_{2,1}$	-1	-0.906	-1.043

Applying the same methodology as in the previous example, we get convergence in all the guard coefficient vectors after 15 data sets, as shown in Figure 4.7 with non-regularized and regularized solutions shown in Table 4.2.3.

The above examples are meant to provide a base line on the efficiency of our approach, even when the values of the guard parameters are different for each edge. Next, we will study some of the additional advantages that our approach provides in terms of assessment of the computed solution.

4.3 Assessment of Switching Surfaces

In this section, we address the question posed in Problem (6) by studying the relation between the true solution $w^* \in \mathbb{R}^p$ with p being the order of the true parameter space and alternate solutions $\tilde{w} \in \mathbb{R}^q$ with $q \neq p$ being simply a different parameter space.

4.3.1 Exploring Higher Parameter Space

Let the true solution w^* be the solution to the following problem

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\|w\|_2^2 \\ \text{subject to} \quad & w_1 x_1^{(i)} + w_2 x_2^{(i)} + w_3 x_3^{(i)} - \gamma \geq 1, \quad i = 1, \cdots, N \\ & w_1 z_1^{(i)} + w_2 z_2^{(i)} + w_3 z_3^{(i)} - \gamma \leq -1, \quad i = 1, \cdots, N \end{aligned} \tag{4.31}$$

with the associated cost $J(w^*) = J^*$.

Then, we look at a similar problem in a higher parameter space $\mathbb{R}^p,\ p > n = 3$, i.e., we solve the following problem

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\|w\|_2^2 \\ \text{subject to} \quad & w_1 x_1^{(i)} + w_2 x_2^{(i)} + w_3 x_3^{(i)} + \cdots + w_p x_p^{(i)} - \gamma \geq 1, \quad i = 1, \cdots, N \\ & w_1 z_1^{(i)} + w_2 z_2^{(i)} + w_3 z_3^{(i)} + \cdots + w_p z_p^{(i)} - \gamma \leq -1, \quad i = 1, \cdots, N \end{aligned} \tag{4.32}$$

Note that since the cost is quadratic on each w_j, any $w_j \neq 0\ \forall\ j = 4, \cdots, p$ will incur a higher cost than $J^* = w_1^{*2} + w_2^{*2} + w_3^{*2}$. Moreover, since no additional constraints are added to problem (4.32), the solution w^* still remains feasible, and since it is also optimal, it follows that the optimal cost remains J^* with solution given by $\tilde{w}^* = [w_1^*\ w_2^*\ w_3^*\ 0\ \cdots\ 0]^\top$.

Evidently, the information added in $x_j,\ j = 4, \cdots, p$ was either redundant or not important. By redundant, we mean that the information provided by x_j was a linearly combination of x_1, x_2, and x_3. In the case of redundant information, the optimal solution w^* must already include the information provided by $\tilde{w}_j^* = a_1 x_1 + a_2 x_2 + a_3 x_3$ for some $a \in \mathbb{R}$.

In conclusion, if the true solution to finding the guard parameter $w \in \mathbb{R}^n$ is sought in an optimization problem where $\tilde{w} \in \mathbb{R}^p, p > n$, then the optimal solution $\tilde{w}^*$ is the zero-padded vector $\tilde{w}^* = [w_1^*\ w_2^*\ w_3^*\ 0\ \cdots 0]^\top$.

As an example, let us consider the true system given in Example 1, where we solved the problem

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\|w\|_2^2 \\ \text{subject to} \quad & w_1 y^{(i)}[k] + w_2 y^{(i)}[k-1] + w_3 u^{(i)}[k-1] - \gamma \geq 1, \qquad i = 1, \cdots, N \\ & w_1 y^{(i)}[k-1] + w_2 y^{(i)}[k-2] + w_3 u^{(i)}[k-2] - \gamma \leq -1, \quad i = 1, \cdots, N \end{aligned} \tag{4.33}$$

with true solution given by $w_{1,2} = [0.4\ 0.5]^\top$ and $w_{2,1} = [0.2\ 0.35]^\top$. Let is now pose the problem

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\|w\|_2^2 \\ \text{subject to} \quad & w_1 y^{(i)}[k] + w_2 y^{(i)}[k-1] + w_3 u^{(i)}[k-1] + w_4 u^{(i)}[k-2] - \gamma \geq 1 \\ & w_1 y^{(i)}[k-1] + w_2 y^{(i)}[k-2] + w_3 u^{(i)}[k-2] + w_4 u^{(i)}[k-3] - \gamma \leq -1 \end{aligned} \tag{4.34}$$

for all $i = 1, \cdots, N$, and where we have added the variable $u[k-2]$, which is actually known a priori from the dynamics equation (4.1) as

$$u[k-2] = u[k-1] - \frac{y[k] - y[k-1]}{\beta_q} \tag{4.35}$$

which is a linear combination of the true variables $y[k], y[k-1], u[k-1]$. Thus, adding the variable $u[k-2]$ will not alter the true solution $w_{1,2}^* = [0.4\ 0.5]^\top, w_{2,1}^* = [0.2\ 0.35]^\top$.

4.3.2 Exploring Lower Parameter Space

Let us consider now the case where the true solution is $w^* = [w_1^* \; w_2^* \; w_3^*]^\top$ but we are solving the lower dimension problem

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\|w\|_2^2 \\ \text{subject to} \quad & w_1 x_1^{(i)} + w_2 x_2^{(i)} - \gamma \geq 1, \quad i = 1, \cdots, N \\ & w_1 z_1^{(i)} + w_2 z_2^{(i)} - \gamma \leq -1, \quad i = 1, \cdots, N \end{aligned} \tag{4.36}$$

Note that the set of strict (feasibility) inequalities

$$w_1 x_1^{(i)} + w_2 x_2^{(i)} - \gamma > 0 \quad i = 1, \cdots, N \tag{4.37}$$

$$w_1 z_1^{(i)} + w_2 z_2^{(i)} - \gamma < 0 \quad i = 1, \cdots, N \tag{4.38}$$

are feasible if and only if the nonstrict inequalities in (4.36) is feasible. Hence, at the optimal solution, when the inequality becomes active, we get that either

$$X^\top w = (1 + \gamma) \quad \text{or} \quad X^\top w = -1 + \gamma \tag{4.39}$$

where $X \in \mathbb{R}^{2\times N}$ contains the data vectors x_1 and x_2, as in (4.17). However, since $w \in \mathbb{R}^{n-1}$, where n corresponds to the true order of the coefficient vector w^*, then $(1-\gamma) \notin \mathcal{R}(X^\top)$ at any point i unless the value of the extra dimension x_3 is zero, but since we know that the true solution for w_3 is $w_3^* \neq 0$, then the problem is infeasible.

In conclusion, we have shown a method to estimate the guard parameters $\alpha_{q,q'}$ and $\gamma_{q,q'}$ for all edges $E(q, q')$ in a hybrid system where the switching rule is not ruled by a complete partition of the regressor space, but rather a mapping of the form (4.4). Indeed, the variables involved in the guard condition are not always available, but if knowledge of the systems is able to provide an estimated set of variables on which the switching surface can lie, then using the proposed convex optimization

framework warrants a means to assess when we have hypothesized too many variables (resulting in a non-changing cost if we remove one of the excess parameters) and when we have hypothesized too few variables (resulting in the problem being infeasible). Moreover, some open issues remain, such as defining under which cases the high condition number of matrices X and Z will affect the solution of the guard parameters $\alpha_{q,q'}, \gamma_{q,q'}$.

Chapter 5

Summary and Future Work

5.1 Summary

We have described a hybrid model for submovements in manual tracking tasks, with application to Parkinson's disease. We use hybrid optimization techniques based in embedded solutions to solve for switching sequences and dynamics that describe when submovements occur, and how they are characterized. We also consider solutions obtained via an alternate method, in particular we use the geometric approach from [21]. This model raises interesting questions in the identification of generic hybrid automata parameters, such as guard conditions that are possibly combinatorial, multivariate, or nonlinear and that cannot be simply described by a partition of the state (or regressor) space, and the identification of systems under stochastic reset maps.

This latter one is the question that we initiate to address in this work, i.e., how to estimate the guard condition given the least amount of information about the switching rule. We derive a methodology that allows us to estimate the guard coefficients of the hybrid system in which we know the form of the guard, i.e., in which we know

the parameters on which the (linear) switching surface lies. In addition, we present some of the properties that this guard coefficient should posses, such as resulting in an optimization problem whose cost plateaus when the estimated parameter space was higher than the true parameters space, as well as an optimization problem which results unfeasible if too few parameters are used to characterize the true switching surface.

5.2 Future Work

We noted 2 issues in the identification of the parameters and active modes of the hybrid system model that may warrant additional investigation. First, we evaluate the assumption that β_1, β_2 are constant values. A histogram of initialization values $\tilde{\beta}_k$ (Figure 5.1) reveals that while our assumption of β_1, β_2 constant is reasonable, more accurate modeling might be had by showing β_1, β_2 to take on a new value every time the tracking or correction mode, respectively, is re-entered. In this case, the hybrid model in Figure 3.3 would need to be extended to accommodate a stochastic reset map with a distribution as shown in Figure 5.1. This would significantly complicate the optimization (3.6) by creating $3N$ variables $\beta_{1,k}$, $\beta_{2,k}$ and would fix the number of mode transitions based on the initialization, distinctly disadvantageous for the problem of detecting mode transitions.

Second, we ultimately seek a method to both construct the dynamics of the switched system as well as the switching conditions that dictate transitions between modes, eventually without prior knowledge of which are the linear variables on which the switching surface lies on. We have essentially decoupled these two problems, since ready solutions exist for the former, but none that we are aware of exist regarding the latter for generic hybrid systems where the switching law is not necessarily ruled by a space partition with non-overlapping planes. In chapter 4 we proposed an approach

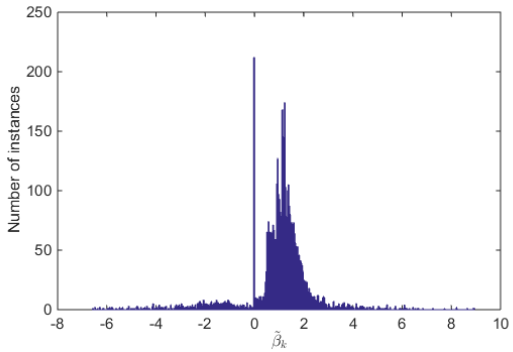


Figure 5.1: Histogram of the values of $\tilde{\beta}_k$ in the training data showing the distributions taken throughout the tracking task.

that can lead to identifying the switching surface when the parameters involved in the switching surface are known. Although a reasonable starting point, in order to apply this methodology to the real data, we must characterize the feasible parameters that could be potentially involved in the switching surface before running the algorithm on the data since numerical issues can arise, specially when running it over large data sets. Since the question of what neurological mechanisms trigger the switching between submovements is an unanswered topic at the moment, it is not uncommon to test large sets of data as candidate parameters involved in orchestrating the switching rule, hence, exploring techniques such as feature selection [49] and Kernel mappings could potentially bring some rich insight as to what these mechanisms might be as well as in determining which might be the most prominent components.

Appendices

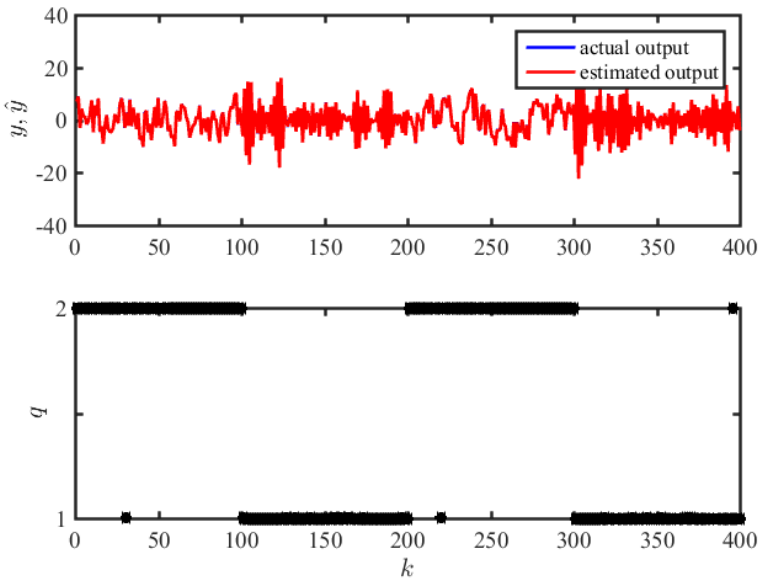
Appendix A

Hybrid System Identification Example Using Embedding Method

In here we compare our proposed hybrid system identification method described in subsection 3.4.1 to existing methods. Although our methodology is highly dependent on initial conditions, we exemplify the case of a two modes hybrid system where we initialize the optimization with values $\nu_k = 0.5, \forall\ k = 1, \cdots, N$, meaning we are unbiased (at all time instances) towards any of the modes being active. Notice that although this is not guaranteed to be always the case, it exemplifies that it can still identify more accurately the hybrid system with no "smart" initialization using a very simple computational method.

The following example is borrowed from Example 1 in [21], which is modified from [2]. We used the modified version because our algorithm shows no errors in the example as given in [2], but does show some small errors in the modified version shown in [21]. Consider the two-mode system with random input generated from

Figure A.1: Academic example originally from [2] used to exemplify the effectiveness of our approach on a two-mode hybrid system. We show the actual and estimated output(top) and the estimated active mode sequence (bottom) showing only three misclassified data points, which do not result in major changes in the estimated output.



a uniform distribution $\mathcal{U} \in [-4, 4]$and output data generated by the SARX hybrid system

$$y[k] = \begin{cases} 0.2y[k-1] + 0.24y[k-2] + 2u[k-1] + e[k], & q_k = 1 \\ -1.4y[k] - 0.53y[k-2] + u[k-1] + e[k], & q_k = 2 \end{cases} \tag{A.1}$$

with mode $q = 1$ being active during times $k \in [1, 100] \cup [201, 300]$ and mode $q = 2$ being active during times $k \in [101, 200] \cup [301, 400]$. In this case, the signal is perfectly reconstructed when there is no noise, so we further consider AWGN with standard deviation $\sigma_e = 0.2$. The actual and estimated output, along with the resulting active modes (without mode projection used) are shown in Figure A.1. With the inclusion of noise, only three data points were misclassified, and the estimated parameter vectors are $\theta_1 = [0.20\ 0.24\ 1.99]$ and $\theta_2 = [-1.41\ \ -0.53\ 1]$.

References

[1] S. Paoletti, A. L. Juloski, G. Ferrari-Trecate, and R. Vidal, "Identification of Hybrid Systems A Tutorial," *European Journal of Control*, vol. 13, no. 2-3, pp. 242–260, 2007.

[2] N. Ozay, M. Sznaier, C. M. Lagoa, and O. I. Camps, "A sparsification approach to set membership identification of switched affine systems," *IEEE Transactions on Automatic Control*, vol. 57, no. 3, pp. 634–648, 2012.

[3] J. Lygeros, K. Johansson, S. Simic, Jun Zhang, and S. Sastry, "Dynamical properties of hybrid automata," *IEEE Transactions on Automatic Control*, vol. 48, no. 1, pp. 2–17, 2003.

[4] I. Hwang, H. Balakrishnan, and C. Tomlin, "State estimation for hybrid systems: applications to aircraft tracking," *IEE Proceedings of Control Theory Applications*, vol. 153, no. 5, pp. 556–566, 2006.

[5] A. D. Ames, R. Vasudevan, and R. Bajcsy, "Human-data based cost of bipedal robotic walking," *Proceedings of the 14th international conference on Hybrid systems: computation and control - HSCC '11*, pp. 153–162, 2011. [Online]. Available: http://dl.acm.org/citation.cfm?id=1967701.1967725

[6] S. Mariethoz, S. Almer, M. Baja, a.G. Beccuti, D. Patino, a. Wernrud, J. Buisson, H. Cormerais, T. Geyer, H. Fujioka, U. Jonsson, C.-Y. K. C.-Y. Kao, M. Morari, G. Papafotiou, a. Rantzer, and P. Riedinger, "Comparison of Hybrid Control Techniques for Buck and Boost DC-DC Converters," *IEEE Transactions on Control Systems Technology*, vol. 18, no. 5, pp. 1126–1145, 2010.

[7] A. Garulli, S. Paoletti, and A. Vicino, "A survey on switched and piecewise affine system identification," *IFAC Proceedings Volumes*, vol. 16, pp. 344–355, 2012.

[8] L. Bako, G. Mercère, and S. Lecoeuche, "On-line structured subspace identification with application to switched linear systems," *International Journal of Control*, vol. 82, no. 8, pp. 1496–1515, 2009.

[9] L. Bako, G. Mercère, R. Vidal, and S. Lecoeuche, "Identification of Switched Linear State Space Models without Dwell Time," *15th IFAC Symposium on System Identification*, pp. 569–574, 2009.

[10] S. Gil and B. Williams, "Beyond local optimality: An improved approach to hybrid model learning," *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*, pp. 3938–3945, 2009.

[11] V. Verdult and M. Verhaegen, "Subspace identification of piecewise linear systems," *2004 43rd IEEE Conference on Decision and Control (CDC) (IEEE Cat. No.04CH37601)*, vol. 4, pp. 3838–3843, 2004.

[12] D. Liberati, "Biomedical applications of piece-wise affine identification for hybrid systems," *Annals of Biomedical Engineering*, vol. 37, no. 9, pp. 1871–1876, 2009.

[13] K. Boukharouba, L. Bako, and S. Lecoeuche, "Temporal video segmentation using a switched affine models identification technique," *2010 2nd International Conference on Image Processing Theory, Tools and Applications, IPTA 2010*, no. 2, pp. 157–160, 2010.

[14] G. Ferrari-Trecate, M. Muselli, D. Liberati, and M. Morari, "A clustering technique for the identification of piecewise affine systems," *Automatica*, vol. 39, pp. 205–217, 2003.

[15] F. Lauer, G. Bloch, and R. Vidal, "A continuous optimization framework for hybrid system identification," *Automatica*, vol. 47, no. 3, pp. 608–613, 2011.

[16] F. Lauer, "Estimating the probability of success of a simple algorithm for switched linear regression," *Nonlinear Analysis: Hybrid Systems*, vol. 8, no. 1, pp. 31–47, 2013. [Online]. Available: http://dx.doi.org/10.1016/j.nahs.2012.10.001

[17] R. Vidal, S. Soatto, and S. Sastry, "An algebraic geometric approach to the identification of a class of linear hybrid systems," *Conference on Decision and Control*, vol. 1, no. December, pp. 167–172, 2003.

[18] R. Vidal, "Recursive identification of switched ARX systmems," *Automatica*, vol. 44, pp. 2274–2287, 2008.

[19] N. Ozay, C. Lagoa, and M. Sznaier, "Robust Identification of Switched Affine Systems via Moments Based Convex Optimization," *Joint 48th IEEE Conference on Decision and Control*, pp. 4686–4691, 2009.

[20] L. Bako, "Identification of switched linear systems via sparse optimization," *Automatica*, vol. 47, no. 4, pp. 668–677, 2011.

[21] V. L. Le, F. Lauer, and G. Block, "Identification of linear hybrid systems : a geometric approach," *American Control Conference*, pp. 830–835, 2013.

[22] E. Bredensteiner and K. Bennett, "Multicategory classification by support vector machines," *Computational Optimization and Applications*, vol. 12, pp. 53–79, 1999.

[23] A. Bemporad, A. Garulli, S. Paoletti, and A. Vicino, "A bounded-error approach to piecewise affine system identification," *IEEE Transactions on Automatic Control*, vol. 50, no. 10, pp. 1567–1580, 2005.

[24] H. Nakada, K. Takaba, and T. Katayama, "Identification of piecewise affine systems based on statistical clustering technique," *Automatica*, vol. 41, no. 5, pp. 905–913, 2005.

[25] H. Ohlsson and L. Ljung, "Identification of Piecewise Affine Systems Using Sum-of-Norms Regularization," *IFAC World Congress, 18th*, pp. 6640–6645, 2011.

[26] ——, "Identification of switched linear regression models using sum-of-norms regularization," *Automatica*, vol. 49, no. 4, pp. 1045–1050, 2013.

[27] R. C. Miall, D. J. Weir, and J. F. Stein, "Intermittency in human manual tracking tasks," *Journal of Motor Behavior*, vol. 25, no. 1, pp. 53–63, 1993.

[28] H. I. Krebs, M. L. Aisen, B. T. Volpe, and N. Hogan, "Quantization of continuous arm movements in humans with brain injury," *Proceedings of the National Academies of Science*, vol. 96, pp. 4645–4649, April 1999.

[29] A. Karniel and F. Mussa-Ivaldi, "Does the motor control system use multiple models and context switching to cope with a variable environment?" *Experimental Brain Research*, vol. 143, pp. 520–524, 2002.

[30] P. Gawthrop, I. Loram, M. Lakie, and H. Gollee, "Intermittent control: a computational theory of human control," *Biological Cybernetics*, vol. 104, pp. 31–51, 2011.

[31] S. T. Grafton and E. Tunik, "Human basal ganglia and the dynamic control of force during on-line corrections," *The Journal of Neuroscience*, vol. 31, no. 5, pp. 1600–1605, 2011.

[32] E. Tunik, J. C. Houk, and S. T. Grafton, "Basal ganglia contribution to the initiation of corrective submovements," *NeuroImage*, vol. 47, no. 4, pp. 1757 – 1766, 2009.

[33] D. D. Vecchio, R. Murray, and P. Perona, "Decomposition of human motion into dynamics based primitives with application to drawing tasks," *Proceedings of the 2003 American Control Conference, 2003.*, vol. 2, no. July, 2003.

[34] W. Au, N. Lei, M. Oishi, and M. McKeown, "L-Dopa produces under-damped visually guided motor responses in Parkinson's disease," *Experimental Brain Research*, vol. 202, no. 3, pp. 553–559, May 2010.

[35] M. Oishi, P. TalebiFard, and M. J. McKeown, "Assessing manual pursuit tracking in parkinson's disease via linear dynamical systems," *Annals of Biomedical Engineering*, vol. 39, no. 8, pp. 2263–2273, August 2011.

[36] A. Abdel-Malek and V. Z. Marmarelis, "Modeling of task-dependent characteristics of human operator dynamics pursuit manual tracking," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 18, no. 1, pp. 163–172, January/February 1988.

[37] K. J. W. Craik, "Theory of the human operator in control systems," *British Journal of Psychology*, vol. 38, pp. 56–61, 1947.

[38] D. M. Wolpert, R. C. Miall, J. L. Winter, and J. F. Stein, "Evidence for an error deadzone in compensatory tracking," *Journal of motor behavior*, vol. 24, no. 4, pp. 299–308, 1992.

[39] L. Ljung, *System Identification: Theory for the User*. Prentice Hall PTR, 1999.

[40] X. Xu and P. Antsaklis, "Optimal control of switched systems based on parameterization of the switching instants," *IEEE Transactions on Automatic Control*, vol. 49, no. 1, pp. 2–16, Jan 2004.

[41] Y. Wardi, M. Egerstedt, and M. Hale, "Switched-mode systems: gradient-descent algorithms with armijo step sizes," *Discrete Event Dynamic Systems*, pp. 1–29, 2014.

[42] S. C. Bengea and R. A. DeCarlo, "Optimal control of switching systems," *Automatica*, vol. 41, pp. 11–27, 2005.

[43] R. Cools, "Dopaminergic modulation of cognitive function-implications for l-DOPA treatment in Parkinson's disease," *Neuroscience & Biobehavioral Reviews*, vol. 30, no. 1, pp. 1–23, 2006. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0149763405000540

[44] R. Cools, R. Barker, B. Sahakian, and T. Robbins, "L-Dopa medication remediates cognitive inflexibility, but increases impulsivity in patients with parkinson's disease," *Neuropsychologia*, vol. 41, pp. 1431–1441, 2003.

[45] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A Training Algorithm for Optimal Margin Classifiers," *Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory*, pp. 144–152, 1992.

[46] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York: Cambridge University Press, 2004.

[47] T. Hastie, R. Tibshiranu, and J. Friedman, *The Elements of Statistical Learning*. New York: Springer, 2008.

[48] M. Martinez-Ramon and C. Christodolou, *Support Vector Machines for Antenna Array Processing and Electromagnetics*. Morgan and Claypool, 2006.

[49] I. Guyon and a. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.