Control Engineering Methods for the Design of Robust Behavioral Treatments

Korkut Bekiroglu, Constantino Lagoa, Suzan A. Murphy, and Stephanie T. Lanza

Abstract—In this paper, a robust control approach is used to address the problem of adaptive behavioral treatment design. Human behavior (e.g., smoking and exercise) and reactions to treatment are complex and depend on many unmeasurable external stimuli, some of which are unknown. Thus, it is crucial to model human behavior over many subject responses. We propose a simple (low order) uncertain affine model subject to uncertainties whose response covers the most probable behavioral responses. The proposed model contains two different types of uncertainties: uncertainty of the dynamics and external perturbations that patients face in their daily life. Once the uncertain model is defined, we demonstrate how least absolute shrinkage and selection operator (lasso) can be used as an identification tool. The lasso algorithm provides a way to directly estimate a model subject to sparse perturbations. With this estimated model, a robust control algorithm is developed, where one relies on the special structure of the uncertainty to develop efficient optimization algorithms. This paper concludes by using the proposed algorithm in a numerical experiment that simulates treatment for the urge to smoke.

Index Terms—Adaptive treatment design, adaptive-robust intervention, behavioral treatment design, min–max structured robust optimization, receding horizon control.

NOMENCLATURE

$\mathbb{N}$ Natural numbers.
$\mathbf{A}$ Matrix in $\mathbb{R}^{m \times n}$.
$\mathbf{A} \geq 0$ Positive semidefinite matrix.
$\mathbf{A}^T$ Transpose of the matrix $\mathbf{A}$.
$\mathbf{A}_{:,i}$ i^{th} column of matrix $\mathbf{A}$.
$\mathbf{I}$ Identity matrix in $\mathbb{R}^{m \times m}$.
$\mathbf{w}$ Signal in $\mathcal{L}^{\infty}$.

$\mathbf{w}_k$ kth element of signal $\mathbf{w}$, $\mathbf{w}_k \in \mathbb{R}$, and $k = [1, 2, \ldots, \infty]$.
$\mathbf{w}_{(k:k+K-1)}$ Vector $\mathbf{w}_{(k:k+K-1)} \in \mathbb{R}^K$ that is segment of signal $\mathbf{w}$: $\mathbf{w}_{(k:k+K-1)} = [\mathbf{w}_k \ \mathbf{w}_{k+1} \ \ldots \ \mathbf{w}_{k+K-1}]$.
$\dim(\mathbf{w}_{(k:k+K-1)})$ Dimension of vector $\mathbf{w}_{(k:k+K-1)}$.
$\mathrm{diag}(\mathbf{w}_{(k:k+K-1)})$ Elements of vector $\mathbf{w}_{(k:k+K-1)} \in \mathbb{R}^K$ to the diagonal of a $\mathbb{R}^{K \times K}$ matrix.
$\lfloor y \rfloor$ Largest integer less than or equal to y.
$\mathcal{N}(\mu, \sigma^2)$ Standard Gaussian distribution with mean μ and variance σ^2.
$\|\mathbf{w}_{(k:k+K-1)}\|_p$ p-norm of vector $\mathbf{w}_{(k:k+K-1)}$ that is: $\|\mathbf{w}_{(k:k+K-1)}\|_p \doteq {}^p(\sum_{i=k}^{k+K-1} \mathbf{w}_i^p)^{1/2}$.
$\|\mathbf{w}_{(k:k+K-1)}\|_0$ Number of the nonzero element of vector $\mathbf{w}_{(k:k+K-1)}$ that is $\|\mathbf{w}_{(k:k+K-1)}\|_0 \doteq \#\{i : \mathbf{w}_i \neq 0\}$.

K. Bekiroglu and C. Lagoa is with the Department of Electrical Engineering, The Methodology Center, The Pennsylvania State University, University Park, PA 16802 USA (e-mail: korkutbekiroglu@gmail.com; lagoa@engr.psu.edu).
S. A. Murphy is with the Quantitative Methodology Program, Institute for Social Research, 2068, University of Michigan, Ann Arbor, MI 48106-1248 USA (e-mail: samurphy@umich.edu).
S. T. Lanza is with the Department of Biobehavioural Health, The Methodology Center, The Pennsylvania State University, University Park, PA 16802 USA (e-mail: slanza@psu.edu).

I. INTRODUCTION

PROLIFERATION of portable devices that collect patient information very frequently (i.e., intensive longitudinal data) and enable timely treatment has opened the possibility of developing effective personalized interventions [1]. These types of interventions can be behavioral or pharmacological or a combination according to the structure of the specific behavioral problem [2]. This is currently being investigated in smoking addiction, alcohol addiction, exercise behavior, and so on. Dynamical modeling of behavior is critical in developing control algorithms [3]–[6]. However, very little attention has been paid to the fact that human behavior has a significant amount of uncertainty and that this uncertainty should be addressed systematically. This paper provides an approach for modeling patient behavior and designing a robust adaptive treatment that considers both the present state of the patient and the probable perturbations to expected behavior.

Previous researchers have applied control concepts in behavioral research [3]–[5], [7]–[9], but the use of feedback is still a novel approach. In addition to feedback, additional constraints for patients' limitations and measured disturbances can be easily addressed in controller-design-based interventions [10]. In this paper, we discuss a possible control engineering-based approach and explain how robust control tools can be used to design robust adaptive treatments. To design a robust algorithm, the first step is to define the uncertainties observed in the treatment responses. Therefore, we propose a class of uncertain affine models suitable for this task and

discuss a way to identify such models using available tools, such as lasso [11]. The second step is to show that a robust control algorithm can be developed that relies on the special structure of the uncertainty in the model to develop efficient optimization algorithms.

Only simple affine models will be considered here, because very little is known about the structure of human behavior. Nevertheless, this approach can be extended to more complex models, such as nonlinear and time-varying models if one has better knowledge about the dynamics of the human behavior.

Besides the problem of choosing the right model structure, behavioral data have other remarkable challenges. The collection of behavioral data is often done over long periods of time and one cannot usually perform repeated experiments to improve the quality of the data. As a result, unlike most physical systems, behavioral data are generally more noisy, incomplete, and inconsistent. Hence, designers of treatments typically work with an incomplete, noisy data set from multiple patients [12]–[15]. Multiple participants are used to attempt to capture a more complete data set such that the gaps in one participant's data are filled by another participant. Therefore, we also use data from several participants to determine a model that both is meaningful and can be used to design treatment for a specific group of individuals.

In this paper, we discuss a specific structure for model uncertainty that can be used for modeling human behavior. When selecting a type of uncertainty, the objective is to address: 1) differences in the behavior of different participants modeled as white Gaussian noise and 2) sparse external perturbations that can be used to model life events that affect the response to treatment. With the model structure defined earlier, lasso [11] is used for parameter identification and uncertainty quantification because it provides a way to directly estimate a model subject to sparse perturbations. Given that the models extracted from behavioral data usually have a large amount of uncertainty, there is a need for feedback and, especially, robustness. Hence, in the second part of this paper, an algorithm is proposed for a controller design that is robust with respect to the specific type of uncertainties considered.

The specific control design technique used in this paper is based on a robust model predictive controller (MPC) approach [16]. A computationally simple cost function is used, namely, a quadratic one, and the corresponding robust optimization problem is formulated. In formulating the control problem, one considers all different combinations of treatments available. The resulting robust optimization problem is then solved by relying on results from robust optimization [17]–[19] and by exploiting the specific structure of the uncertainty.

Throughout this paper, the problem of designing a treatment for smoking urge is used as our example. Although the treatment is assumed as ON/OFF (apply treatment or no treatment) in this specific example, this is not always the case. There are situations where one not only has to decide when/which treatment to apply but also the dosage (see [20] for a more detailed explanation of characteristics of behavioral problems). The presented method can be easily modified to accommodate different types (behavioral, pharmacological, or combined) and levels (dosage) of treatments. Note that in this example, the term treatment is used to refer to an intervention where text messages are sent to the subject.

A. Previous Work

Adaptive interventions are sequences of treatments that are adapted and readapted to individual circumstances and behaviors in order to achieve and maintain health behavior change [12], [13]. These interventions may be provided many times (i.e., tens, hundreds, or even thousands of treatment occasions during the entire treatment period). The interventions are often delivered via portable devices, as intensive interventions. Since these interventions adapt to individual progress, it is effective to prevent insufficient response and react immediately against the unexpected shock on behavior. Therefore, scientists in the social and behavioral sciences have been working on finding systematic ways to use portable devices such as smart phones to change health behavior or maintain healthy behavior in real time. Consequently, adaptive interventions are being developed by social and behavioral scientists for different areas, such as hypertension [14], substance abuse [15], criminal justice [21], mental health [22], and Alzheimer's disease [23].

Engineering concepts, such as dynamical modeling, have been applied to modeling and controlling behavior [6]. References [3]–[5] present special issues that arise in applying such approaches to behavioral research. These studies employ both time-invariant and time-varying models, depending on the specific problem. Some extend studies use controller design methods to design adaptive interventions for special problems in behavioral science [7]–[9]. This paper defines uncertainties in behavior more explicitly than previous studies and discusses their effects on the robustness of the algorithm. In general, according to the intervention design problem, the MPC design method is preferred. Preliminary results of this paper were presented in [24]. More detailed explanations of modeling behavior and identification are presented in this paper. The complete proofs are also provided, along with simulations.

This paper is organized as follows. In Section II, we introduce the model that will be used to approximate patient behavior and justify its use. In Section III, we discuss a way to identify the model's parameters from patient study data. In Section IV, an MPC formulation of the robust control problem is introduced. In Section V, a simulation of the application of the proposed approach is presented with a simulated smoking cessation treatment from [25]. In this section, some hypothetical individual results are also given to show how a control algorithm works. Finally, in Section VI, we present some concluding remarks.

II. MODEL

As mentioned in Section I, data on human behavior usually involve several subjects with many gaps in information over time. Moreover, behavior might change with time and vary from subject to subject. Therefore, it is very unlikely that one can accurately specify a complex model of behavior. One can use a simple model to approximate patient behavior or the

main dynamics of the subject and, at the same time, highlight the fact that there is a substantial amount of uncertainty in it. Hence, a possible model for identification is a set of affine difference equations. The difference model used in this paper is of the form

$$\mathbf{y}_{k+1} = \mathbf{a}_f + \sum_{i=0}^{n-1}[A_i\mathbf{y}_{k-i} + B_i\mathbf{T}_{k-i}] + D\varepsilon_k + E\mathbf{w}_k \quad (1)$$

where $\mathbf{y}_k \in \mathbb{R}^m$ is the measured output vector at time k, $\mathbf{T}_k \in \mathbb{R}$, and $\mathbf{T}_k \in \mathcal{T}$ is the control input (i.e., treatment), where $\mathcal{T}$ is a set of all available treatments. In many cases, this set has a finite number of elements. Unmeasurable sparse exogenous perturbations are denoted by $\mathbf{w}_k \in \mathbb{R}$ and $\varepsilon_k \in \mathbb{R}^m$, which represents uncertainty in the model. The affine term vector is $\mathbf{a}_f \in \mathbb{R}^m$, and $A_i \in \mathbb{R}^{m\times m}$, $B_i \in \mathbb{R}^m$, $D \in \mathbb{R}^{m\times m}$, and $E \in \mathbb{R}^m$ are coefficients matrices.

The dynamics of behavior are very complex and hard to control because of the uncertainties on behaviors. Therefore, the uncertain difference equation in (1) is used to model a large set of responses to behavioral treatment. In addition to the dynamics of behavior, which are contaminated by Gaussian noise, an unmeasured exogenous input is added to the model to represent unexpected influences.

We use as a motivating example a hypothetical study of smoking cessation. In this example, we address the problem of treatment design after quit day (i.e., day on which an individual quits smoking). Patients usually have different dynamics before and after quit day and different treatments should be designed for each of these cases. Since the objective of this paper is to develop a general design procedure, we only consider the after quit day case as an example of treatment design. The procedure developed can also be used for before quit day case.

In the smoking cessation model, there are three different variables that one can measure: smoking urge (denoted by su); negative affect, (na), which is a single scale indicating an adverse mood state; and self-efficacy, (se), which represents an individual's belief in their ability to abstain from smoking. With these variables, the following model is proposed:

$$\begin{bmatrix} su_{k+1} \\ na_{k+1} \\ se_{k+1} \end{bmatrix} = \begin{bmatrix} a_f \\ d_f \\ g_f \end{bmatrix} + \sum_{i=0}^{n-1}\left\{\begin{bmatrix} a_i & b_i & c_i \\ d_i & e_i & f_i \\ g_i & h_i & r_i \end{bmatrix}\begin{bmatrix} su_{k-i} \\ na_{k-i} \\ se_{k-i} \end{bmatrix}\right\}$$
$$+ \sum_{i=0}^{n-1}\left\{\begin{bmatrix} 0 \\ q_i \\ s_i \end{bmatrix}\mathbf{T}_{k-i}\right\} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\mathbf{w}_k$$
$$+ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} \varepsilon_k^{(su)} \\ \varepsilon_k^{(na)} \\ \varepsilon_k^{(se)} \end{bmatrix}. \quad (2)$$

Remark 1: The signs of several of these coefficients are assumed to be known in advance for this example. For example, negative affect (na) increases smoking urge (su), whereas self-efficacy (se) decreases it. More information about the dynamics of the smoking urge can be found in [26] and [27]. This information can be used in two different ways: 1) it can be

given to the identification algorithm as additional constraints or 2) it can be utilized to validate the model.

This model also shows that smoking urge cannot be directly controlled, but if the negative affect can be decreased and self-efficacy can be increased by applying treatment **T**, the desired smoking urge level can be achieved under the uncertainties.

One should first note that the model above is an uncertain affine model. One of the reasons why this structure was chosen was the fact that a model should not be obtained just from data, and it should also leverage characteristics determined by prior behavioral research. For instance, patients smoke cigarettes to regulate the smoking urge set point (see [3], [28] for more detailed explanation of smoking dynamics). Therefore, an affine term (constant term in the model) is added to the model to represent this set point.

The model in (1) is called a structural equation model (SEM) in social science [29]. Similar to SEM models, an affine model for smoking cessation is preferred, because the equilibrium point of the uncontrolled system is often not the origin. The model is obtained by using data from several patients and differences in patients are typically modeled as Gaussian random variables ε. The external perturbation $\mathbf{w} \in \mathcal{L}^\infty$ is motivated by a different kind of uncertainty in human behavior. The way the subjects behave is influenced by external sparse events that temporarily affect his/her response to treatment. Hence, one needs an uncertainty that is sparse, bounded in magnitude, and with limited cumulative effect. As a result, it is assumed that the segment of perturbation signal $\mathbf{w} \in \mathcal{L}^\infty$ is bounded in the ℓ_1-norm, ℓ_∞-norm, and ℓ_0-norm. For this situation, a possible signal set $\mathcal{W}$, which is used in this paper, is the following. Given an MPC control window of length K and k, the set $\mathcal{W}$ is

$$\mathcal{W} = \{\mathbf{w} \in \mathcal{L}^\infty : \text{for all } k, \ \|\mathbf{w}_{(k:k+K-1)}\|_\infty \leq \alpha,$$
$$\|\mathbf{w}_{(k:k+K-1)}\|_1 \leq \gamma, \ \|\mathbf{w}_{(k:k+K-1)}\|_0 = K/\varsigma\}. \quad (3)$$

$\alpha \in \mathbb{N}$ is bound on the magnitude of the perturbation, $\gamma \in \mathbb{N}$ is bound on cumulative effect, and $\varsigma \in \mathbb{N}$ enforces the sparsity constraint on vector $\mathbf{w}_{(k:k+K-1)}$.

III. IDENTIFICATION OF THE MODEL AND PERTURBATIONS

To estimate the coefficients from study data, we start by noting that the model above can be taken to represent the relations among the variables as

$$Y = H\beta + \varepsilon. \quad (4)$$

In this model, the unknown vector β contains the parameters of the model and vector w, vector Y is a function of the measurements, and matrix H is a function of the measurements and inputs (treatment). For the smoking cessation example, the structure and the dimension of the measurement vector Y, unknown vector β, and matrix H in (4) are given in Appendix A.

There are several ways to estimate vector β. In the examples provided in this paper, lasso [11] is used as an identification tool. Assuming $\mathcal{P}$ is the number of the patients, N is the length of the data from each patient, and n is the order of the system,

Algorithm 1 Lasso Iterative Identification Algorithm

1: Given μ, tolerance δ, iteration $j = 0, 1, ..., \beta_0 = 0$, and randomly chosen $\beta_1 = \beta_{initial}$ ▷ Initialization
2: **while** $|\beta_{j+1} - \beta_j| \leq \delta$ **do**
3: $\quad \hat{\sigma}^2 \leftarrow \frac{\|Y - H\beta_{j+1}\|_2^2}{n}$
4: $\quad \hat{\beta} \leftarrow \text{argmin}_{\beta,w} \frac{\|Y - H\beta\|_2^2}{2n\hat{\sigma}} + \frac{\hat{\sigma}}{2}$
    $\quad\quad$ subject to
    $\quad\quad \|w\|_1 \leq \mu$
5: $\quad \beta_{j+1} \leftarrow \hat{\beta}$
6: **end while**
7: $\sigma \leftarrow \hat{\sigma}$ and $\beta \leftarrow \hat{\beta}$
8: α, γ and $\varsigma \leftarrow$ Parameters of set $\mathcal{W}$ from statistic of vector w,

---

this algorithm provides a way to determine an estimate of the parameters that balances the size of the noise $\varepsilon \in \mathcal{N}(0, \sigma^2)$ and the sparsity of the exogenous perturbation w in the model. Note that there is also a balance between the order of the system and the size of the perturbation (magnitude, sparsity, and so on). Sparsity of signal $\mathbf{w}$ is crucial in this context, as it represents perturbations that a patient faces infrequently

$$\{\hat{\beta}, \hat{\sigma}\} = \underset{\beta, \sigma, w}{\arg\min} \frac{\|Y - H\beta\|_2^2}{2n\sigma} + \frac{\sigma}{2}$$
$$\text{s.t.} \quad \|w\|_1 \leq \mu. \quad\quad (5)$$

The optimization problem in (5) is a convex minimization of a penalized joint loss function with a regularization parameter μ given for the regression coefficient, noise level, and vector w, which is a finite dimensional signal.

The design matrix H and the response vector Y are constructed from measurements from $\mathcal{P}$ different patients (see Appendix A). Since the problem is jointly convex with respect to β and σ, if the initial estimate of β is known or given, then σ, which minimizes the objective function, can be calculated by differentiating the objective function (5) with constant β vector (step 3 in Algorithm 1). Then, the optimization problem in step 4 in Algorithm 1 is solved to estimate vector β.

The following lasso algorithm shows the identification procedure in [11], where j shows the iteration number.

The terms in the objective function (5) aim at finding the maximum likelihood estimate of the parameters of the model and the variance of the noise ε while the ℓ_1 constraint searches for a sparse exogenous input. Here, the ℓ_1-norm is used as a convex approximation of the ℓ_0-norm [30]. Thus, given the data and the order of the model n, at the end of the identification algorithm, the parameters of the model in (2), the noise levels of $\varepsilon \in \mathcal{N}(0, \sigma^2)$, and unmeasurable exogenous input $w \in \mathbb{R}^{\mathcal{P}(N-n)}$ are estimated. Then, the statistic of identified vector $w \in \mathbb{R}^{\mathcal{P}(N-n)}$ is used to define signal set $\mathcal{W}$ in (3).

In the smoking cessation example, since an uncertain model is searched for all $\mathcal{P}$ individuals (patients), all observations of $\text{su}_k^i, \text{na}_k^i$ and se_k^i ($i = 1, 2, \ldots, \mathcal{P}$) are used to construct the regression model (4). Again the mathematical details of model (1) for smoking cessation example are defined in Appendix A.
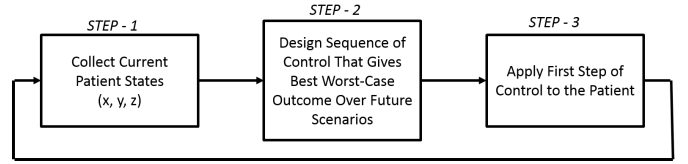

Fig. 1. Adaptive intervention algorithm.

Finally, given a time interval of length K, output vector $\mathbf{y}$ is a function of uncertainties $\mathbf{w} \in \mathcal{L}^\infty$, $\varepsilon \in \mathcal{N}(0, \sigma^2)$, and the state of the system. As mentioned earlier, $\mathbf{w}$ is a bounded sparse uncertainty and ε is Gaussian noise that models differences between different subjects. Now, to be able to design robust controllers, we need to have a bounded support set for ε.

Remark 2: Since Gaussian distributions have unbounded support, we choose a set of high probability. More precisely, given a window of size K, the density of $\varepsilon_{k:k+K-1}$ has hyperspherical contours and we take a set of high probability of the form

$$\|\varepsilon_{k:k+K-1}\|_2 \leq \rho \quad\quad (6)$$

where ρ is chosen based on the identified standard deviation of the noise. This has the added advantage of leading to a formulation of the robust treatment design that is computationally tractable. We refer the reader to [31] for a more detailed discussion of the advantages of using this approach to describe this type of noise.

IV. ROBUST MODEL PREDICTIVE CONTROLLER

The method for controller design used in this paper is based on MPC [16]. Here, the usual approach in robust MPC is taken: one estimates the present value of the state variable and determines the value of the control variables over the horizon that minimizes a given objective function. The first of these control signals is applied, and the process is then repeated. The control algorithm is summarized in Fig. 1. In step 1, initial conditions or the information $\mathbf{y}_{k-1} \ldots \mathbf{y}_{k-n}$ that the control algorithm will use is collected. In the smoking cessation study, this information is $\text{su}_{k-1} \ldots \text{su}_{k-n}$, $\text{na}_{k-1} \ldots \text{na}_{k-n}$, $\text{se}_{k-1} \ldots \text{se}_{k-n}$. In addition, after consultation with a practitioner in the field, the total treatment to be given to the patient is limited to an allowable set [32]–[34]. Such limitations might require the knowledge of how much treatment was given in the recent past. As a result, recent control input information $\mathbf{T}_{k-1} \ldots \mathbf{T}_{k-l}$ ($l \geq n$) is used in order to enforce the constraints on the total treatment provided.

Consistent with the usual MPC approach, in step 2, one minimizes a given cost function subject to constraints on the total applied control in a certain range. Then, in step 3, the result of the control algorithm is applied to the patient. Thus, in this section, a decision rule is developed that dictates whether treatment is applied to each individual at each time point under the uncertainties.

The difference equation (1) is used to determine a closed-form objective function for MPC formulation with

a given receding horizon K. In the objective function, we represent $\mathcal{Y}_{k+1} \in \mathbb{R}^{Km}$ as

$$\mathcal{Y}_{k+1} = \mathcal{A}_f + \tilde{A}\mathcal{Y}_0 + \tilde{B}\tilde{\mathbf{T}}_k + \tilde{D}\tilde{\varepsilon}_k + \tilde{E}\tilde{\mathbf{w}}_k \qquad (7)$$

where $\mathcal{Y}_0 \in \mathbb{R}^{nm}$ is the vector containing the state of the model. $\tilde{\mathbf{T}}_k \in \mathbb{R}^{K+n-1}$ is the treatment. $\tilde{\mathbf{w}}_k \in \mathbb{R}^K$ is sparse disturbance and $\tilde{\varepsilon}_k \in \mathbb{R}^K$ is noise. A_f, $\tilde{A}$, $\tilde{B}$, $\tilde{D}$, and $\tilde{E}$ are matrices, which are calculated recursively from difference equation (1)

$$\mathcal{Y}_{k+1} = \begin{bmatrix} \mathbf{y}_k \\ \mathbf{y}_{k+1} \\ \vdots \\ \mathbf{y}_{k+K-1} \end{bmatrix}, \quad \mathcal{Y}_0 = \begin{bmatrix} \mathbf{y}_{k-n} \\ \mathbf{y}_{k+1-n} \\ \vdots \\ \mathbf{y}_{k-1} \end{bmatrix}, \quad \tilde{\mathbf{T}}_k = \begin{bmatrix} \mathbf{T}_{k-n} \\ \mathbf{T}_{k+1-n} \\ \vdots \\ \mathbf{T}_{k+K-2} \end{bmatrix}$$

$$\tilde{\mathbf{w}}_k = \begin{bmatrix} \mathbf{w}_{k-1} \\ \mathbf{w}_k \\ \vdots \\ \mathbf{w}_{k+K-2} \end{bmatrix}, \quad \tilde{\varepsilon}_k = \begin{bmatrix} \varepsilon_{k-1} \\ \varepsilon_k \\ \vdots \\ \varepsilon_{k+K-2} \end{bmatrix}.$$

Then, in this paper, the objective function of the following form is considered:

$$(\mathcal{Y}_{k+1} - \theta)X(\mathcal{Y}_{k+1} - \theta)^T$$

although more complex convex functions can be addressed by the framework presented here. For the smoking cessation example, since one minimizes smoking urge, matrix X is defined to choose smoking urge measure in vector $\mathcal{Y}_{k+1}$ as

$$X = \begin{cases} 1, & \text{if } x_{i,i} = k, k+m, k+2m, \ldots, k+(K-1)m \\ 0, & \text{otherwise.} \end{cases}$$

Given the tuning parameter $\theta \in \mathbb{R}$, approximation of desired average level of smoking urge, we aim to solve the following robust optimization problem at each step of the algorithm:

$$\min_{\mathbf{T} \in \mathcal{T}} \max_{\substack{\|\tilde{\varepsilon}_k\|_2 \leq \rho \\ \tilde{\mathbf{w}}_k \in \mathcal{W}_k}} (\mathcal{Y}_{k+1} - \theta)X(\mathcal{Y}_{k+1} - \theta)^T \qquad (8)$$

which is subject to the system dynamics described in (7). We also define $\mathcal{W}_k$ from the set $\mathcal{W}$ as

$$\mathcal{W}_k = \{\text{vector } \tilde{\mathbf{w}}_k \in \mathbb{R}^K \text{ satisfies} : \|\tilde{\mathbf{w}}_k\|_\infty \leq \alpha,$$
$$\|\tilde{\mathbf{w}}_k\|_1 \leq \gamma, \ \|\tilde{\mathbf{w}}_k\|_0 = K/\varsigma\}. \qquad (9)$$

The intersection of the norms in the set $\mathcal{W}_k$ forms convex polytopes. It should be noted that set $\mathcal{W}_k$ is a union of these convex polytopes; this fact is explored later when solving the robust optimization problem resulting from the MPC formulation of the control problem.

Smoking Cessation Example: We now discuss the set $\mathcal{T}$ of allowable treatments used for the smoking cessation example. In this case

$$\mathcal{T} = \left\{ \mathbf{T} : \mathbf{T}_\eta \in \{0, 1\} \text{ for } \eta = k, \ldots, k+K-1 \text{ and} \right.$$
$$\left. \sum_{\eta=k-l}^{k+K-1} \mathbf{T}_\eta \leq T_{\text{total}} \right\} \qquad (10)$$

represent the fact that one can either apply or not apply treatment at each time. In addition, the number of treatments applied in the last $K + l$ instances is constrained to be less than or equal to T_{total}. This constraint in the number of treatments is aimed at addressing the problem of treatment burden [32]–[34], which is decreasing effectiveness as the number of applied treatments increases. In this specific example, treatment $\mathbf{T}$ is binary; but in reality, it can have a different level and type of treatment.

One should note that the optimization problem described earlier is complex. Therefore, a problem equivalent to the original problem (8) is now presented that is suitable for implementation in Theorem 1. Using the results in [17]–[19] (provided in Appendix B for completeness), one can prove the following result.

Theorem 1: Consider the following semidefinite problem:

$$\min_{\tau, \tilde{\mathbf{T}}_k \in \mathcal{T}, \lambda} \tau$$

subject to

$$\begin{bmatrix} \tau - \lambda\rho^2 & 0 & * \\ 0 & \lambda I & (\tilde{D}X)^T \\ (\mathcal{A}_f - \theta + \tilde{A}\mathcal{Y}_0 + \tilde{B}\tilde{\mathbf{T}}_k + \tilde{E}\tilde{\mathbf{w}}_k)X & \tilde{D}X & I \end{bmatrix} \geq 0 \qquad (11)$$

for all $\tilde{\mathbf{w}}_k \in \mathcal{W}_{\text{ext}}$, where $\mathcal{W}_{\text{ext}}$ values are the extremes of the polytopes whose union is $\mathcal{W}_k$.

The optimum of the problem above is an upper bound on the optimum of the original optimization problem (8) and is equal if only one of the elements in the set $\mathcal{W}_{\text{ext}}$ leads to an active constraint at the optimum.

In the LMI in (11), $*$ is $((\mathcal{A}_f - \theta + \tilde{A}\mathcal{Y}_0 + \tilde{B}\tilde{\mathbf{T}}_k + \tilde{E}\tilde{\mathbf{w}}_k)X)^T$.

Proof: See Appendix C. $\qquad\square$

We now characterize the set of extremes, $\mathcal{W}_{\text{ext}}$.

Theorem 2: Consider set $\mathcal{W}_k$ defined in (9), where α, γ, and ς define the bounds on the ℓ_∞-norm, ℓ_1-norm, and sparsity, respectively. Let $n_{nz} = \dim(\tilde{\mathbf{w}}_k)/\varsigma$.

1) If $\alpha \geq \gamma$, then $\mathcal{W}_{\text{ext}}$ is the set of all signed permutations of

$$[\gamma \underbrace{0 \cdots 0}_{K-1}].$$

2) If $n_{nz}\alpha \leq \gamma$, then $\mathcal{W}_{\text{ext}}$ is the set of all signed permutations of

$$[\underbrace{\alpha \cdots \alpha}_{n_{nz}} \underbrace{0 \cdots 0}_{K-n_{nz}}].$$

3) If $\gamma < n_{nz}\alpha < n_{nz}\gamma$, then $\mathcal{W}_{\text{ext}}$ is the set of all signed permutations of

$$[\underbrace{\alpha \cdots \alpha}_{n_+} h \underbrace{0 \cdots 0}_{K-n_+-1}]$$

where

$$n_+ = \left\lfloor \frac{\gamma}{\alpha} \right\rfloor \text{ and } h = \gamma - n_+\alpha.$$

Proof: See Appendix D. $\qquad\square$

After finding the extreme points of the set $\mathcal{W}_k$, a semidefinite programming or mixed integer semidefinite programming solver could be used to solve problem (11); but for a large

receding horizon K, the number of extremes and, thus, the number of LMI constraints might be fairly large.

Section V presents a simulated application of the algorithm using a model that mimics data that might have been collected in smoking cessation studies, such as [25].

V. SIMULATION RESULTS

A. Implementation

As an implementation of the adaptive treatment algorithm that is proposed for an intervention using portable devices, we consider individuals who wish to quit smoking, and we present a hypothetical algorithm for treatment via a smart phone. In addition, the quit day (i.e., day on which an individual quits smoking) and the time immediately following the quit day are extremely important on smoking cessation interventions and in the prevention of relapse. The simulations in this example address smoking dynamics after the quit day.

In the simulation, the participant provides information on his/her level of smoking urge, negative affect, and self-efficacy on the morning of day 1; the smart phone then provides the behavioral treatment. Robust adaptive treatment begins at midday after the participant again provides his/her level of smoking urge, negative affect, and self-efficacy. Right now, su_0^i, na_0^i, se_0^i, $\mathbf{T}_0^i$ and su_1^i, na_1^i, se_1^i are known, where $\mathbf{T}_1^i$ is treatment provided in the morning. The proposed adaptive treatment algorithm is run and decides $\mathbf{T}^{i\ \text{optimal}}$ and set $\mathbf{T}_2^i = \mathbf{T}^{i\ \text{optimal}}$. That is, if $\mathbf{T}^{i\ \text{optimal}} = 1$, treatment is provided at midday; otherwise, it is not. Afterward, new su_3^i, na_3^i, se_3^i are collected; then by using these new values and su_2^i, na_2^i, se_2^i, the process $\mathbf{T}_2^i$ is repeated to obtain $\mathbf{T}_3^i = \mathbf{T}^{i\ \text{optimal}}$. This process is repeated throughout the entire course of treatment.

B. Identification Results

In order to test the performance of the approach presented in this paper, we emulate a real application. We develop a second-order ($n = 2$) model for smoking urge that approximates the behavior observed in empirical studies, such as [25]. Note that, in this paper, subjects smoke heavily. Therefore, the intervention is not expected to drive cravings to zero in this short-term program. Moreover, during the initial phase of treatment, the proposed algorithm aims to decrease smoking urge of patients as much as possible, while simultaneously reducing the amount of treatment. We aim to increase initial impact of treatment.

Here is the true model of smoking urge

$$
\begin{bmatrix} a_f \\ a_0 \\ a_1 \\ b_0 \\ b_1 \\ c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} 1.06 \\ .50 \\ .35 \\ .01 \\ 0 \\ -.12 \\ -.09 \end{bmatrix}, \begin{bmatrix} d_f \\ d_0 \\ d_1 \\ e_0 \\ e_1 \\ f_0 \\ f_1 \end{bmatrix} = \begin{bmatrix} .15 \\ .01 \\ 0 \\ .49 \\ .25 \\ -.05 \\ 0 \end{bmatrix}, \begin{bmatrix} g_f \\ g_0 \\ g_1 \\ h_0 \\ h_1 \\ r_0 \\ r_1 \end{bmatrix} = \begin{bmatrix} .52 \\ -.01 \\ 0 \\ -.02 \\ 0 \\ .49 \\ .36 \end{bmatrix}.
$$

The studies mentioned earlier do not contain treatment. Hence, we augmented the model with a nonlinear effect of treatment in order to represent treatment burden. In the literature, the treatment burden is defined as increasing the

level of treatment to the extent that it may cause suboptimal adherence and even negative outcomes. More information about treatment burden in social and medical sciences can be found in [32]–[34]. More precisely, the following terms were introduced:

$$
q_0 \mathbf{T}_{k-1} + q_1 \text{sig} \left[\sum_{u=k-44}^{u=k-2} \mathbf{T}_u \right] \mathbf{T}_{k-2} \tag{12}
$$

$$
s_0 \mathbf{T}_{k-1} + s_1 \text{sig} \left[\sum_{u=k-44}^{u=k-2} \mathbf{T}_u \right] \mathbf{T}_{k-2}
$$

where

$$
\text{sig}(x) = \frac{1}{1 + e^{-(x-17)}}.
$$

$\text{sig}(x)$ represents the sigmoid function or special case of logistic function. Then, the following parameter values are introduced:

$$
[q_0 \quad q_1 \quad s_0 \quad s_1] = [-.2 \quad .2 \quad .2 \quad -.2].
$$

This model for the effect of treatment is designed to emulate the case where the treatment has a positive impact, but its effectiveness decreases when the treatment is applied too frequently.

For this highly uncertain model, 500 different trajectories of the system are generated, with random initial conditions and random realizations of uncertainty. This is done to simulate the behavior of 500 subjects in a study where the treatment was provided at random times. Note that this is a reasonable surrogate for a study, since the model is designed to approximate real data, except for the influence of treatment. The model identification algorithm described in Section III is then applied, and a model to be used by the robust MPC algorithm was obtained. To validate the model, we can use the information in [26] and [27], showing that negative affect (na) increases smoking urge (se), where the sign of the coefficient of negative affect (na) is positive; self-efficacy (se) decreases smoking urge (su), where the sign of the coefficient of self-efficacy (se) is negative. These effects also can be seen in Fig. 3 and Table II. Moreover, the dynamical equations of negative affect and self-efficacy indicate that the effect of treatment is very small. This results in small improvement in smoking urge. However, this improvement is very important for heavy smokers at the beginning of the treatment process.

To do this, a model of the form (2) is used with order $n = 2$. The model coefficients obtained for a randomly chosen virtual patient are

$$
\begin{bmatrix} a_f \\ a_0 \\ a_1 \\ b_0 \\ b_1 \\ c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} 1.11 \\ .47 \\ .34 \\ .01 \\ 0 \\ -.10 \\ -.08 \end{bmatrix}, \begin{bmatrix} d_f \\ d_0 \\ d_1 \\ e_0 \\ e_1 \\ f_0 \\ f_1 \\ q_0 \\ q_1 \end{bmatrix} = \begin{bmatrix} .13 \\ .005 \\ 0 \\ .5 \\ .24 \\ -.05 \\ 0 \\ -0.207 \\ 0.15 \end{bmatrix}, \begin{bmatrix} g_f \\ g_0 \\ g_1 \\ h_0 \\ h_1 \\ r_0 \\ r_1 \\ s_0 \\ s_1 \end{bmatrix} = \begin{bmatrix} .52 \\ -.011 \\ 0 \\ -.02 \\ 0 \\ .49 \\ .35 \\ 0.202 \\ -0.142 \end{bmatrix}
$$

TABLE I
INITIAL CONDITION OF STATES AND INPUT

$su(0)$	$su(1)$	$se(0)$	$se(1)$	$na(0)$	$na(1)$	$T(0)$	$T(1)$
3.521	0.346	4.428	4.324	-0.354	1.082	0	0

TABLE II
INTERVENTION PERFORMANCE OF SDP

	Adaptive Treatment	No Robustness Treatment	Full Treatment	No Treatment
Mean(su)	1.31	1.36	1.46	1.51
Mean(se)	3.8412	3.76	3.6424	3.4863
Mean(na)	0.3096	0.3662	0.4571	0.5698
$\|su\|_2$	22.55	23.32	24.48	25.15
$\sum_{k=1}^{k=150} \mathbf{T}_k$	43	34	150	0

and the description of the uncertainty is, for robust MPC window of $K = 8$, the following: $\tilde{\mathbf{w}}_k$ has 25% nonzero terms with $\|\tilde{\mathbf{w}}_k\|_1 \leq 5$ and $\|\tilde{\mathbf{w}}_k\|_\infty \leq 2.5$. In order to obtain these parameters, the statistic of the identified w in Section III is used. As for Gaussian uncertainty, $\varepsilon^{\text{su}} \sim \mathcal{N}(0, 1.3)$, $\varepsilon^{\text{na}} \sim \mathcal{N}(0, 0.78)$, and $\varepsilon^{\text{se}} \sim \mathcal{N}(0, 0.55)$ are identified. The sampling period here is 8 h and, hence, data are collected three times per day. The simulation is run for 50 days, and 150 data points are collected.

C. Experiments

We then applied the proposed controller in this paper to a nonlinear model mentioned earlier, again with randomly generated uncertainty/noise and the initial conditions given at Table I. In other words, we simulate the application of our robust MPC algorithm to a patient. The parameters used are specified as follows.

1) MPC window size: $K = 8$.
2) Constraints in control: $\mathbf{T}_k \in \{0, 1\}$ and

$$T_{\text{total}} = \sum_{\eta=k-3}^{k+K-1} \mathbf{T}_\eta \leq T_{\text{total}} = 10/3.$$

In other words, one either applies or does not apply treatment at each time point, and there is a constraint of a maximum of approximately one treatment per three sample times. Full treatment is defined as $\mathbf{T}_k = 1$ for all values of k.

3) Objective function to be minimized

$$\sum_{\eta=k+1}^{k+K} (su(\eta) - 2.2)^2$$

where the target value 2.2 was chosen, so that one has a significant decrease in the mean smoking value.

4) Bound on the two-norm of the Gaussian noise: $\rho = 30$.
5) Initial conditions of the states and input in Table I.

D. Simulations Results for a Patient

The optimization problem to be solved here is a mixed-integer, semidefinite convex problem. There are many ways to solve it, but in the simulations performed here, the requirement $\mathcal{T} = \mathbf{T}_k \in \{0, 1\}$ is relaxed to $\mathcal{T} = \mathbf{T}_k \in [0, 1]$ to
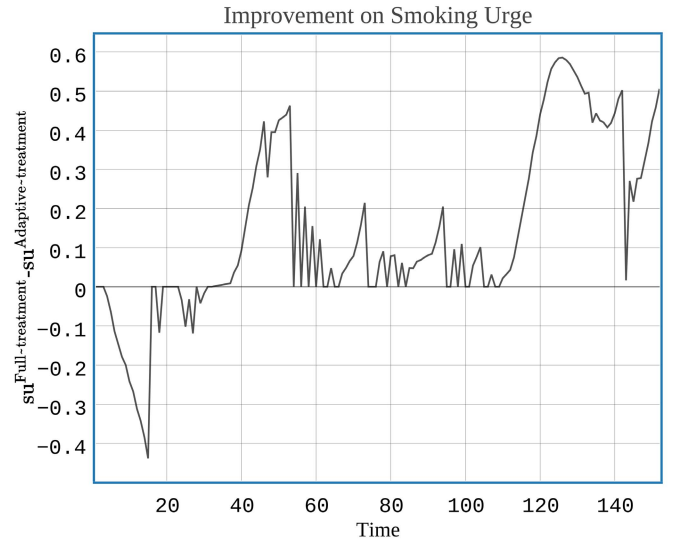

Fig. 2. Performance of adaptive intensive intervention.

use a general, semidefinite programming solver. The control applied is

$$\mathbf{T}_k^{\text{applied}} = \text{round}(\mathbf{T}_k).$$

The simulation was run for 150 time instances corresponding to a real time interval of 50 days. CVX is used with SeDuMi as the semidefinite programming solver to problem (11) [35]. The developed control law systematically improved smoking urge, since the treatment is only applied when needed. At the same time, our control law increased self-efficacy while reducing negative affect in order to reduce smoking urge. Table II represents a typical example of the results obtained.

Table II shows the performance of the adaptive treatment design in terms of objective function, average values of smoking urge, self-efficacy, negative affect, and total amount of treatment that the adaptive intervention applied. Moreover, to show the benefit of our robust adaptive intervention, the simulation is also done where controller design did not consider uncertainty (no robustness in Table II). Results indicated that the robust adaptive intervention applies the optimum amount of treatment to improve smoking urge, self-efficacy, and negative affect. The improvement in smoking urge is small, likely because this is a population of heavy smokers [25]. However, this improvement is extremely important at the beginning of the treatment process and leads to significant benefit in long term period. Note that this improvement was obtained with a minimized treatment effort. Therefore, one can see that there is a significant improvement in smoking urge not only in terms of the average but also in the fact that a consistent decrease in urge is obtained. If the treatment was more diverse than a binary signal [see the treatment set $\mathcal{T}$ in (10)] (e.g., varying dosages), the controller benefits from adaptation would be greater.

Fig. 2 shows decrease in smoking urge. In Fig. 2, $su^{\text{Full Treatment}} - su^{\text{Adaptive Treatment}}$ is depicted, where $su^{\text{Adaptive Treatment}}$ is smoking urge measured under the adaptive treatment and $su^{\text{Full Treatment}}$ is smoking urge
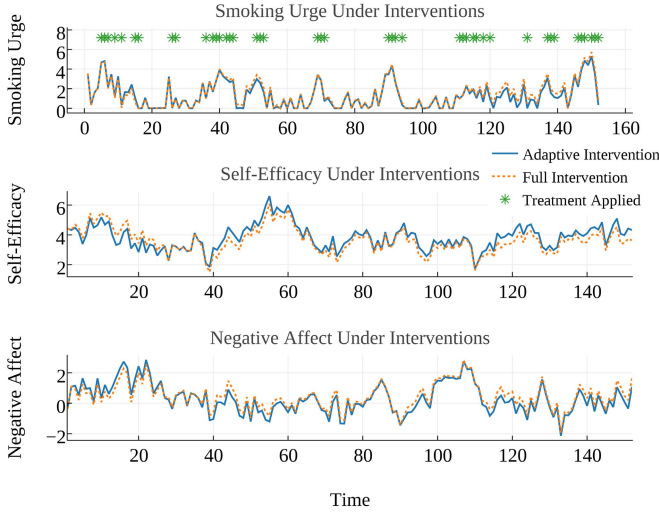
Fig. 3.   Smoking urge, negative affect, and self-efficacy under adaptive intervention.
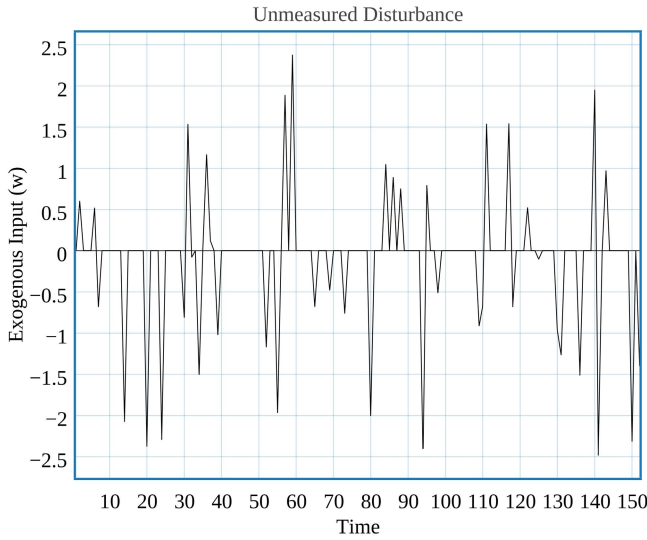

Fig. 4.   Sparse disturbance.

measured under the full treatment. Although the full treatment performs better than the adaptive treatment in the beginning of the treatment process, in the long term, the adaptive treatment works better, because the algorithm applies the treatment only when it is really needed, thus reducing treatment burden.

As expected, the algorithm carefully chooses when to apply treatment. It is mainly applied when external perturbations lead to a significant increasing trend in smoking urge. Finally, Fig. 3 shows the smoking urge, self-efficacy, and negative-affect measure of a virtual patient under both full treatment and adaptive intervention. Note that the responses shown in Fig. 3 are reflective of long-time, post-quit behavior. While important, this only represents a subset of the behavior stages seen in smoking cessation. Asterisks in Fig. 3 indicate when treatment is provided by the adaptive treatment algorithm. Fig. 4 shows the exogenous sparse disturbance that is applied to this particular patient for this simulation.

Remark 3: To be able to formulate the treatment design problem (11) as a convex optimization problem, in the
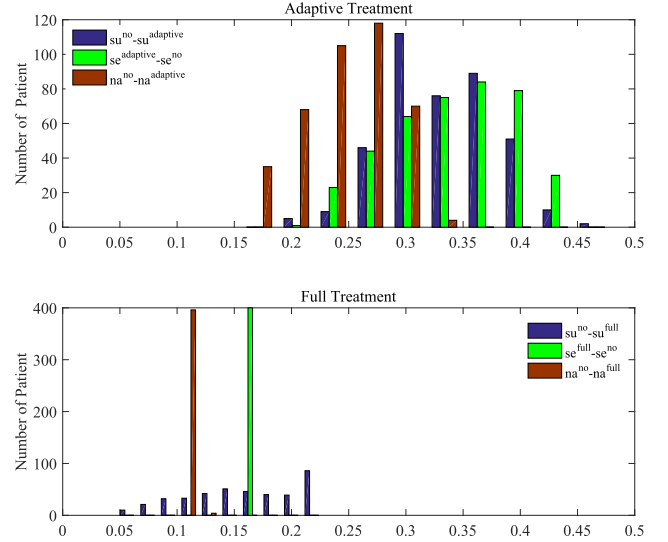

Fig. 5.   Adaptive and full treatment intervention results.

simulations above, we replaced the constraint $\mathbf{T}_k \in \{0, 1\}$ by the relaxation $\mathbf{T}_k \in [0, 1]$. Then, the treatment applied was round $(\mathbf{T}_k)$. To test the effectiveness of such relaxation, branch and bound techniques were used to solve the optimization problem under the true constraint $\mathbf{T}_k \in \{0, 1\}$. The results obtained were similar and, hence, in this example, there is no advantage in solving the more complex mixed integer optimization problem.

E. SDP Results of a Population

Since some of the parameters are as random in the simulation, the algorithm is run 400 times to estimate the average improvement in smoking urge, self-efficacy, and negative affect. The same parameters in Section V-C except that $l = 6$ in (10) are used to run the algorithm. Moreover, for the random parameters, the disturbances $\tilde{\mathbf{w}}_k$ are generated uniformly, such that $\tilde{\mathbf{w}}_k \in \mathcal{W}_k$. The Gaussian disturbances are generated under a normal distribution, such that $\varepsilon^{su} \sim \mathcal{N}(0, 1.3)$, $\varepsilon^{na} \sim \mathcal{N}(0, 0.78)$, and $\varepsilon^{se} \sim \mathcal{N}(0, 0.55)$. For the initial conditions, $su_0 \in \mathcal{U}(3.6, 0)$, $se_0 \in \mathcal{U}(4, 0)$, and $na_0 \in \mathcal{U}(2, 0)$ are generated. Fig. 5 shows average improvement in smoking urge ($su^{\text{No Treatment}} - su^{\text{Adaptive or Full Treatment}}$), self-efficacy ($se^{\text{Adaptive or Full Treatment}} - se^{\text{No Treatment}}$), and negative affect ($na^{\text{No Treatment}} - na^{\text{Adaptive or Full Treatment}}$). It is shown that the adaptive treatment increases self-efficacy while it decreases smoking urge and negative affect, even with less treatment. The average number of treatments provided is 54.6.

1) su^{no}, se^{no}, na^{no}: Smoking urge, self-efficiency, and negative affect without treatment.
2) su^{full}, se^{full}, na^{full}: Smoking urge, self-efficiency, and negative affect under full treatment.
3) su^{adaptive}, se^{adaptive}, na^{adaptive}: Smoking urge, self-efficiency, and negative affect with adaptive treatment.

VI. CONCLUSION

Behavioral and social scientists have been considering the advantages of adaptive treatments. Research has indicated that adaptive treatment strategies might give better results than

fixed treatment (i.e., all patients get the same type and level of treatment) [12]. Since adaptive treatments usually provide better results than the usual methods, in this paper, we argue that control engineering methods, such as feedback or adaptation and robust optimization, can provide a systematic way to design a personalized treatment algorithm. Control engineering methods can be used to design personalized behavioral interventions while reducing treatment burden. Existing theory for adaptive treatment design is primarily qualitative and thus does not provide precise guidance regarding how much of which treatment treatment to provide to which individuals at what times. With this adaptive intervention design algorithm, treatments can be adapted and readapted in response to individuals' progress over a long period of time. These methods hold promise for maintaining desired behavior in situations where controlling behavior is challenging due to complex dynamics. In terms of future work, effort is being put into developing more efficient numerical algorithms for solving the optimization problem for MPC. This will allow for the consideration of a larger receding horizon and, hence, better performance.

APPENDIX A
FORMULATION OF THE IDENTIFICATION PROBLEM

In this section, we show how general equation (4) can be reorganized for the lasso algorithm. Start with the data set (su, na, se) of $\mathcal{P}$ patients. Assume that for each ith patient, we have collected data set su_k^i, na_k^i, se_k^i for $k = 1, 2, \ldots, N$ and $i = 1, 2, \ldots, \mathcal{P}$. Also assume that the order of the model n is given. Then, for each patient i, build the matrices in (13) and (14), shown at the bottom of this page.

Also, define

$$su^i = \begin{bmatrix} su_{n+1}^i & su_{n+2}^i \cdots su_N^i \end{bmatrix}^T \text{ where } su^i \in \mathbb{R}^{N-n}$$
$$na^i = \begin{bmatrix} na_{n+1}^i & na_{n+2}^i \cdots na_N^i \end{bmatrix}^T \text{ where } na^i \in \mathbb{R}^{N-n}$$
$$se^i = \begin{bmatrix} se_{n+1}^i & se_{n+2}^i \cdots se_N^i \end{bmatrix}^T \text{ where } se^i \in \mathbb{R}^{N-n}$$

and

$$SU = \begin{bmatrix} su^1 \\ su^2 \\ \vdots \\ su^{\mathcal{P}} \end{bmatrix} \quad NA = \begin{bmatrix} na^1 \\ na^2 \\ \vdots \\ na^{\mathcal{P}} \end{bmatrix} \quad SE = \begin{bmatrix} se^1 \\ se^2 \\ \vdots \\ se^{\mathcal{P}} \end{bmatrix}$$

where $SU \in \mathbb{R}^{\mathcal{P}(N-n)}$, $NA \in \mathbb{R}^{\mathcal{P}(N-n)}$, and $SE \in \mathbb{R}^{\mathcal{P}(N-n)}$

$$X_1 = \begin{bmatrix} X_{1,1} \\ X_{1,2} \\ \vdots \\ X_{1,\mathcal{P}} \end{bmatrix} \quad X_2 = \begin{bmatrix} X_{2,1} \\ X_{2,2} \\ \vdots \\ X_{2,\mathcal{P}} \end{bmatrix}.$$

Finally, let

$$\theta_{su} = \begin{bmatrix} a_f & a_0 & a_1 & \cdots & a_n & b_0 & \cdots & b_n & c_0 & c_1 & \cdots & c_n \end{bmatrix}$$
$$\theta_{na} = \begin{bmatrix} d_f & d_0 & d_1 & \cdots & a_n & e_0 & e_1 & \cdots & b_n & f_0 & f_1 & \cdots & f_n \\ & q_0 & q_1 & \cdots & q_n \end{bmatrix}$$
$$\theta_{se} = \begin{bmatrix} g_f & g_0 & g_1 & \cdots & a_n & h_0 & h_1 & \cdots & b_n & r_0 & r_1 & \cdots & r_n \\ & s_0 & s_1 & \cdots & s_n \end{bmatrix}$$
$$\beta = \begin{bmatrix} \theta_{su} & \theta_{na} & \theta_{se} & w \end{bmatrix}^T$$

where w is a $w \in \mathbb{R}^{\mathcal{P}(N-n)}$ dimensional vector. Then, if all the patients satisfy the model provided in (2), we have

$$Y = H\beta \tag{15}$$

where $Y \in \mathbb{R}^{3\mathcal{P}(N-n)}$, $H \in \mathbb{R}^{3\mathcal{P}(N-n) \times 3(3n+1)+N}$, and coefficient vector $\beta \in \mathbb{R}^{(11n+3+\mathcal{P}(N-n))}$

$$Y = \begin{bmatrix} SU \\ NA \\ SE \end{bmatrix} \quad H = \begin{bmatrix} X_1 & 0 & 0 & I \\ \hline 0 & X_2 & 0 & 0 \\ \hline 0 & 0 & X_2 & 0 \end{bmatrix}$$

and ε is a vector containing all noise.

APPENDIX B
THEOREM IN [17]–[19]

Theorem 3 [17]–[19]: Assume that the following problem is defined:

$$\phi(A, b, x) \triangleq \min_x \max_{\|\varepsilon\| \le \rho} \|A(\varepsilon)x - b(\varepsilon)\|_2^2. \tag{16}$$

Given $A_0, \ldots, A_p \in \mathbf{R}^{n \times m}$, $b_0, \ldots, b_p \in \mathbf{R}^n$, and $\varepsilon \in \mathbf{R}^p$, the following uncertainty structure can be defined:

$$A(\varepsilon) \triangleq A_0 + \sum_{i=1}^p \varepsilon_i A_i, \quad b(\varepsilon) \triangleq b_0 + \sum_{i=1}^p \varepsilon_i b_i. \tag{17}$$

Then, for $\rho \ge 0$, define

$$F(x) \triangleq [A_1 x - b_1 \ldots A_p x - b_p]. \tag{18}$$

$$X_{1,i} = \begin{bmatrix} 1 & su_n^i & .. & su_1^i & na_n^i & .. & na_1^i & se_n^i & .. & se_1^i \\ 1 & su_{n+1}^i & .. & su_2^i & na_{n+1}^i & .. & na_2^i & se_{n+1}^i & .. & se_2^i \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & su_{N-1}^i & .. & su_{N-n}^i & na_{N-1}^i & .. & na_{N-n}^i & se_{N-1}^i & .. & se_{N-n}^i \end{bmatrix} \tag{13}$$

$$X_{2,i} = \begin{bmatrix} 1 & su_n^i & .. & su_1^i & na_n^i & .. & na_1^i & se_n^i & .. & se_1^i & T_n^i & .. & T_1^i \\ 1 & su_{n+1}^i & .. & su_2^i & na_{n+1}^i & .. & na_2^i & se_{n+1}^i & .. & se_2^i & T_{n+1}^i & .. & T_2^i \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & su_{N-1}^i & .. & su_{N-n}^i & na_{N-1}^i & .. & na_{N-n}^i & se_{N-1}^i & .. & se_{N-n}^i & T_{N-1}^i & .. & T_{N-n}^i \end{bmatrix} \tag{14}$$

The solution of the structured robust least square problem (16) can be calculated by solving the following semidefinite problem:

$$
\min_{\tau, x, \lambda} \tau
$$
$$
\text{s.t.} \begin{bmatrix} \tau - \lambda \rho^2 & 0 & (A_0 x - b_0)^T \\ 0 & \lambda I & F(x)^T \\ (A_0 x - b_0) & F(x) & I \end{bmatrix} \geq 0. \quad (19)
$$

APPENDIX C
PROOF OF THEOREM 1

Proof: We use the structure of $\mathcal{Y}_{k+1}$ in (7) to analyze the robustness of the optimization problem. Thus, $\mathcal{Y}_{k+1}$ in the objective function in (8) is replaced by $\mathcal{Y}_{k+1}$ in (7) as

$$
\min_{\mathbf{T} \in \mathcal{T}} \max_{\substack{\|\tilde{\varepsilon}_k\|_2 \leq \rho \\ \tilde{\mathbf{w}}_k \in \mathcal{W}_k}} \|(\mathcal{A}_f - \theta + \tilde{A}\mathcal{Y}_0 + \tilde{B}\tilde{\mathbf{T}}_k + \tilde{D}\tilde{\varepsilon}_k + \tilde{E}\tilde{\mathbf{w}}_k)X\|_2^2.
$$
$$(20)$$

Given objective function in (20), we can define vectors and matrix in (17) as

$$
A_0 = \tilde{B}, \quad A_i = 0, \quad b_0(\tilde{\mathbf{w}}_k) = \mathcal{A}_f - \theta + \tilde{A}\mathcal{Y}_0 + \tilde{E}\tilde{\mathbf{w}}_k
$$

and $b_i = \tilde{D}_{:,i}$.

Then, for fixed $\tilde{\mathbf{T}}_k$ and $\tilde{\mathbf{w}}_k$, worst case residual is defined by using the methods in [18, Sec. 4] as

$$
r_s(A, b, \tilde{\mathbf{T}}_k, \tilde{\mathbf{w}}_k)^2 = \max_{\|\tilde{\varepsilon}_k\|_2 \leq \rho} \|\tilde{B}\tilde{\mathbf{T}}_k - b(\varepsilon, \tilde{\mathbf{w}}_k)\|_2^2
$$

and from (18), define

$$
M \doteq F^T F, \quad g(\tilde{\mathbf{T}}_k, \tilde{\mathbf{w}}_k) \doteq F^T(\tilde{B}\tilde{\mathbf{T}}_k - b_0)
$$
$$
h(\tilde{\mathbf{T}}_k, \tilde{\mathbf{w}}_k) \doteq \| \tilde{B}\tilde{\mathbf{T}}_k - b_0\|_2^2.
$$

Then, without loss of generality, assuming $\rho = 1$ yields

$$
r_s(A, b, \tilde{\mathbf{T}}_k, \tilde{\mathbf{w}}_k)^2 = \max_{\|\tilde{\varepsilon}_k\|_2 \leq 1} \begin{bmatrix} 1 \\ \varepsilon \end{bmatrix}^T
$$
$$
\times \begin{bmatrix} h(\tilde{\mathbf{T}}_k, \tilde{\mathbf{w}}_k) & g(\tilde{\mathbf{T}}_k, \tilde{\mathbf{w}}_k)^T \\ g(\tilde{\mathbf{T}}_k, \tilde{\mathbf{w}}_k) & M \end{bmatrix} \begin{bmatrix} 1 \\ \varepsilon \end{bmatrix}. \quad (21)
$$

Given $\tau \geq 0$, using the $\mathcal{S}$-procedure [18, Lemma 2.1], (21) can be converted as

$$
\begin{bmatrix} 1 \\ \varepsilon \end{bmatrix}^T \begin{bmatrix} h(\tilde{\mathbf{T}}_k, \tilde{\mathbf{w}}_k) & g(\tilde{\mathbf{T}}_k, \tilde{\mathbf{w}}_k)^T \\ g(\tilde{\mathbf{T}}_k, \tilde{\mathbf{w}}_k) & M \end{bmatrix} \begin{bmatrix} 1 \\ \varepsilon \end{bmatrix} \leq \tau \quad (22)
$$

for every ε, $\varepsilon^T \varepsilon \leq 1$ if and only if there exists a scalar $\lambda \geq 0$, such that

$$
\begin{bmatrix} 1 \\ \varepsilon \end{bmatrix}^T \begin{bmatrix} \tau - \lambda - h(\tilde{\mathbf{T}}_k, \tilde{\mathbf{w}}_k) & -g(\tilde{\mathbf{T}}_k, \tilde{\mathbf{w}}_k)^T \\ -g(\tilde{\mathbf{T}}_k, \tilde{\mathbf{w}}_k) & \lambda I - M \end{bmatrix} \begin{bmatrix} 1 \\ \varepsilon \end{bmatrix} \geq 0 \quad (23)
$$

for every $\varepsilon \in \mathbb{R}^p$ and fixed $\tilde{\mathbf{T}}_k$ and $\tilde{\mathbf{w}}_k$.

Using the fact that $\lambda \geq 0$ is implied by $\lambda I \geq M$, then the above condition can be rewritten as

$$
\mathcal{F}(\tau, \lambda, \tilde{\mathbf{T}}_k, \tilde{\mathbf{w}}_k) \doteq \begin{bmatrix} \tau - \lambda - h(\tilde{\mathbf{T}}_k, \tilde{\mathbf{w}}_k) & -g(\tilde{\mathbf{T}}_k, \tilde{\mathbf{w}}_k)^T \\ -g(\tilde{\mathbf{T}}_k, \tilde{\mathbf{w}}_k) & \lambda I - M \end{bmatrix} \geq 0.
$$

Then, the upper bound on the residual such that $r_s(A, b, \tilde{\mathbf{T}}_k, \tilde{\mathbf{w}}_k)^2 \leq \tau$ can be obtained by solving the following optimization problem [18, Th. 4.1]:

$$
\tau^* = \min_{\tau, \lambda} \tau
$$
$$
\text{s.t.} \begin{bmatrix} \tau - \lambda - h(\tilde{\mathbf{T}}_k, \tilde{\mathbf{w}}_k) & -g(\tilde{\mathbf{T}}_k, \tilde{\mathbf{w}}_k)^T \\ -g(\tilde{\mathbf{T}}_k, \tilde{\mathbf{w}}_k) & \lambda I - M \end{bmatrix} \geq 0 \quad (24)
$$

for all $\tilde{\mathbf{w}}_k \in \mathcal{W}_k$.

In addition, instead of checking all $\tilde{\mathbf{w}}_k \in \mathcal{W}_k$ in LMI constraint (24), one can use the result in [36], which states that all the members of a polytope of matrices are positive semidefinite if and only if all extremes of the polytope are positive semidefinite. Furthermore, the set $\mathcal{W}_k$ is a union of sets bounded by polytopes. Using the same reasoning as in [36], it suffices for the LMI to be satisfied at the extremes of each of the polytopes. Therefore, we need to check all extremes of these polytopes. The LMI in (24) is satisfied in the convex hull of these extremes. It is just a consequence of the fact that an LMI is satisfied in a set of points; it is also satisfied in the convex hull of these points. Then, we have the following result:

$$
\tau^* = \min_{\tau, \lambda} \tau
$$
$$
\text{s.t.} \begin{bmatrix} \tau - \lambda - h(\tilde{\mathbf{T}}_k, \tilde{\mathbf{w}}_k) & -g(\tilde{\mathbf{T}}_k, \tilde{\mathbf{w}}_k)^T \\ -g(\tilde{\mathbf{T}}_k, \tilde{\mathbf{w}}_k) & \lambda I - M \end{bmatrix} \geq 0 \quad (25)
$$

for all $\tilde{\mathbf{w}}_k \in \mathcal{W}_{\text{ext}}$ and forfixed $\tilde{\mathbf{T}}_k$, where $\mathcal{W}_{\text{ext}}$ values are the extremes of the polytopes whose union is $\mathcal{W}_k$. Hence, τ^* is an upper bound on the robust performance for fixed $\tilde{\mathbf{T}}_k$.

For the optimality, assume that there is only one extreme $\tilde{\mathbf{w}}^* \in \mathcal{W}_{\text{ext}}$, such that one of the eigenvalues of LMI in (25) is equal to zero for some τ^* and λ^*, and LMI is positive definite for the rest of the extremes. Then, if the τ^* is not optimum, we can perturb τ^* or decrease it for this specific $\tilde{\mathbf{w}}^* \in \mathcal{W}_{\text{ext}}$ while keeping the LMI still positive definite for the rest of the extreme points. This contradicts the assumption that τ^* is optimum. As a result, this optimization problem gives the optimum τ^* if only one of the extremes is an active constraint.

Thus, for every fixed $\tilde{\mathbf{T}}_k$ and for all $\tilde{\mathbf{w}}_k \in \mathcal{W}_{\text{ext}}$, the results in [18, Th. 4.2] or Theorem 3 can be used to convert problem (25) to (26) for ellipsoid uncertainty $\|\tilde{\varepsilon}_k\|_2 \leq \rho$ and $\tilde{\mathbf{w}}_k \in \mathcal{W}_k$. Finally, problem in (25) is modified as follows:

$$
\min_{\tau, \tilde{\mathbf{T}}_k \in \mathcal{T}, \lambda} \tau
$$
$$
\text{s.t.}
$$
$$
\begin{bmatrix} \tau - \lambda \rho^2 & 0 & * \\ 0 & \lambda I & (\tilde{D}X)^T \\ (\mathcal{A}_f - \theta + \tilde{A}\mathcal{Y}_0 + \tilde{B}\tilde{\mathbf{T}}_k + \tilde{D}\tilde{\varepsilon}_k + \tilde{\mathbf{w}}_k)X & \tilde{D}X & I \end{bmatrix} \geq 0
$$
$$(26)$$

for all $\tilde{\mathbf{w}}_k \in \mathcal{W}_{\text{ext}}$ where again $\mathcal{W}_{\text{ext}}$ values are the extremes of the polytopes, whose union is $\mathcal{W}_k$.

In this LMI, $*$ is $((\mathcal{A}_f - \theta + \tilde{A}\mathcal{Y}_0 + \tilde{B}\tilde{\mathbf{T}}_k + \tilde{E}\tilde{\mathbf{w}}_k)X)^T$. $\square$

APPENDIX D
PROOF OF THEOREM 2

The first two cases ($\alpha \geq \gamma$ and $n_{nz}\alpha \leq \gamma$) are immediate, since they correspond to the cases where either ℓ_1 or ℓ_∞ is the

only binding constraints. Hence, in this proof, we concentrate on the third case ($\gamma < n_{nz}\alpha < n_{nz}\gamma$).

Note that, given the symmetry of the problem, we can concentrate on the subset of the elements of $\mathcal{W}_k$ that are positive and satisfy

$$\mathbf{w}_{i+1} \le \mathbf{w}_i \text{ for all } i = k : k + K - 1, \text{ and } \mathbf{w}_i \in \tilde{\mathbf{w}}_k.$$

All other extremes will be obtained from this by permutations and sign changes of the entries. Recall that, in this case, we are considering

$$\gamma < n_{nz}\alpha < n_{nz}\gamma$$

and the elements of vector $\tilde{\mathbf{w}}_k$ are in a nonincreasing order. Therefore, the extreme of the set is given by pushing as many of the first few elements as possible to their maximum value. Given that $\|\tilde{\mathbf{w}}_k\|_\infty \le \alpha$ and $\|\tilde{\mathbf{w}}_k\|_1 \le \gamma$, one can only have the first

$$n_+ = \left\lfloor \frac{\gamma}{\alpha} \right\rfloor$$

elements of vector $\tilde{\mathbf{w}}_k$ equal to α. To reach the extreme and recalling that the entries of the vector $\mathbf{w}$ are in a nonincreasing order, the $(n_+ + 1)$th element of the vector must be at its maximum value. Given that $\|\tilde{\mathbf{w}}_k\|_1 \le \gamma$ and the first N_+ terms are equal to α

$$\mathbf{w}_{n_++1} \le \gamma - n_+\alpha.$$

Hence, for vectors with entries in a decreasing order, the extreme of the set is attained by a vector of the form

$$[\underbrace{\alpha \cdots \alpha}_{n_+} \ h \ \underbrace{0 \cdots 0}_{K-n_+-1}].$$

ACKNOWLEDGMENT

REFERENCES

[1] K. L. Bierman, R. L. Nix, J. J. Maples, and S. A. Murphy, "Examining clinical judgment in an adaptive intervention design: The fast track program," *J. Consulting Clin. Psychol.*, vol. 74, no. 3, pp. 468–481, 2006.

[2] D. E. Rivera, "Optimized behavioral interventions: What does system identification and control engineering have to offer?" in *Proc. 16th IFAC Symp. Syst. Identificat.*, 2012, pp. 882–893.

[3] K. P. Timms, D. E. Rivera, L. M. Collins, and M. E. Piper, "Continuous-time system identification of a smoking cessation intervention," *Int. J. Control*, vol. 87, no. 7, pp. 1423–1437, 2014.

[4] J.-E. Navarro-Barrientos, D. E. Rivera, and L. M. Collins, "A dynamical model for describing behavioural interventions for weight loss and body composition change," *Math. Comput. Model. Dyn. Syst.*, vol. 17, pp. 183–203, Jan. 2011.

[5] Y. Dong *et al.*, "A dynamical systems model for improving gestational weight gain behavioral interventions," in *Proc. Amer. Control Conf. (ACC)*, 2012, pp. 4059–4064.

[6] D. E. Rivera, M. D. Pew, and L. M. Collins, "Using engineering control principles to inform the design of adaptive interventions: A conceptual introduction," *Drug Alcohol Dependence*, vol. 88, no. 2, pp. S31–S40, May 2007.

[7] N. N. Nandola and D. E. Rivera, "Model-on-demand predictive control for nonlinear hybrid systems with application to adaptive behavioral interventions," in *Proc. 19th IEEE Conf. Decision Control*, Dec. 2010, pp. 6113–6118.

[8] A. Zafra-Cabeza, D. E. Rivera, L. M. Collins, M. A. Ridao, and E. F. Camacho, "A risk-based model predictive control approach to adaptive interventions in behavioral health," *IEEE Trans. Control Syst. Technol.*, vol. 19, no. 4, pp. 891–901, Jul. 2011.

[9] N. N. Nandola and D. E. Rivera, "An improved formulation of hybrid model predictive control with application to production-inventory systems," *IEEE Trans. Control Syst. Technol.*, vol. 21, no. 1, pp. 121–135, Jan. 2013, doi: 10.1109/TCST.2011.2177525.

[10] S. Deshpande, N. N. Nandola, D. E. Rivera, and J. W. Younger, "Optimized treatment of fibromyalgia using system identification and hybrid model predictive control," *Control Eng. Pract.*, vol. 33, pp. 161–173, Dec. 2014.

[11] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc. B, Methodol.*, vol. 58, no. 1, pp. 267–288, 1996.

[12] L. M. Collins, S. A. Murphy, and K. L. Bierman, "A conceptual framework for adaptive preventive interventions," *Prevention Sci.*, vol. 5, no. 3, pp. 185–196, Sep. 2004.

[13] S. A. Murphy and L. M. Collins, "Customizing treatment to the patient: Adaptive treatment strategies," *Drug Alcohol Dependence*, vol. 88, no. 2, pp. S1–S3, May 2007.

[14] M. S. Glasgow, B. T. Engel, and B. C. D'Lugoff, "A controlled study of a standardized behavioral stepped treatment for hypertension," *Psychosomatic Med.*, vol. 51, no. 1, pp. 10–26, 1989.

[15] W. F. Velicer and J. O. Prochaska, "An expert system intervention for smoking cessation," *Patient Edu. Counseling*, vol. 36, pp. 119–129, Feb. 1999.

[16] A. Bemporad and M. Morari, "Robust model predictive control: A survey," in *Robustness in Identification and Control* (Lecture Notes in Control and Information Sciences), vol. 245. London, U.K.: Springer, 1999, pp. 207–226.

[17] H. A. Hindi and S. P. Boyd, "Robust solutions to l_1, l_2, and l_8 uncertain linear approximation problems using convex optimization," in *Proc. Amer. Control Conf. (ACC)*, vol. 6. 1998, pp. 3487–3491.

[18] L. El Ghaoui and H. Lebret, "Robust solutions to least-squares problems with uncertain data," *SIAM J. Matrix Anal. Appl.*, vol. 18, pp. 1035–1064, Oct. 1997.

[19] M. Salhov, "A subgradient solution to structured robust least squares problems," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4761–4770, Sep. 2010.

[20] K. P. Timms, D. E. Rivera, M. E. Piper, and L. M. Collins, "A hybrid model predictive control strategy for optimizing a smoking cessation intervention," in *Proc. Amer. Control Conf.*, Jun. 2014, pp. 2389–2394.

[21] D. B. Marlowe, D. S. Festinger, K. L. Dugosh, K. M. Benasutti, G. Fox, and J. R. Croft, "Adaptive programming improves outcomes in drug court: An experimental trial," *Criminal Justice Behavior*, vol. 39, pp. 514–532, Apr. 2012.

[22] A. J. Rush *et al.*, "Sequenced treatment alternatives to relieve depression (STAR*D): Rationale and design," *Controlled Clin. Trials*, vol. 25, pp. 119–142, Feb. 2004.

[23] L. S. Schneider *et al.*, "National Institute of Mental Health clinical antipsychotic trials of intervention effectiveness (CATIE): Alzheimer disease trial methodology," *Amer. J. Geriatric Psychiatry*, vol. 9, no. 4, pp. 346–360, Jan. 2001.

[24] K. Bekiroglu, C. Lagoa, S. A. Murphy, and S. T. Lanza, "A robust MPC approach to the design of behavioural treatments," in *Proc. IEEE 52nd Annu. Conf. Decision Control*, Dec. 2013, pp. 3505–3510.

[25] S. Shiffman *et al.*, "Dynamic effects of self-efficacy on smoking lapse and relapse," *Health Psychol.*, vol. 19, no. 4, pp. 315–323, Jul. 2000.

[26] M. P. Shiyko, S. T. Lanza, X. Tan, R. Li, and S. Shiffman, "Using the time-varying effect model (TVEM) to examine dynamic associations between negative affect and self confidence on smoking urges: Differences between successful quitters and relapsers," *Prevention Sci.*, vol. 13, pp. 288–299, Jul. 2012.

[27] R. M. P. Van Zundert, S. G. Ferguson, S. Shiffman, and R. C. M. E. Engels, "Dynamic effects of self-efficacy on smoking lapses and relapse among adolescents," *Health Psychol.*, vol. 29, pp. 246–254, May 2010.

[28] K. P. Timms, D. E. Rivera, L. M. Collins, and M. E. Piper, "A dynamical systems approach to understand self-regulation in smoking cessation behavior change," *Nicotine Tobacco Res.*, vol. 16, pp. S159–S168, Sep. 2013.

[29] K. A. Bollen and M. D. Noble, "Structural equation models and the quantification of behavior," *Proc. Nat. Acad. Sci. USA*, vol. 108, pp. 15639–15646, Oct. 2011.

[30] C. Ramirez, V. Kreinovich, and M. Argaez, "Why ℓ_1 is a good approximation to ℓ_0: A geometric explanation," *J. Uncertain Syst.*, vol. 7, no. 3, pp. 203–207, 2013.

[31] F. Paganini, "A set-based approach for white noise modeling," *IEEE Trans. Autom. Control*, vol. 41, no. 10, pp. 1453–1465, Oct. 1996.

[32] K. Gallacher *et al.*, "Qualitative systematic reviews of treatment burden in stroke, heart failure and diabetes—Methodological challenges and solutions," *BMC Med. Res. Methodol.*, vol. 13, p. 10, Jan. 2013.

[33] A. Sav *et al.*, "Burden of treatment for chronic illness: A concept analysis and review of the literature," *Health Expectations*, vol. 18, no. 3, pp. 312–324, Jan. 2013.

[34] B. Spring, S. Pagoto, R. Pingitore, N. Doran, K. Schneider, and D. Hedeker, "Randomized controlled trial for behavioral smoking and weight control treatment: Effect of concurrent versus sequential intervention," *J. Consulting Clin. Psychol.*, vol. 72, no. 5, pp. 785–796, Oct. 2004.

[35] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control* (Lecture Notes in Control and Information Sciences), V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag, pp. 95–110, 2008.

[36] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, vol. 15. Philadelphia, PA, USA: SIAM, 1994.

Constantino Lagoa received the B.S. and M.S. degrees from the Instituto Superior Técnico, Technical University of Lisbon, Lisbon, Portugal, in 1991 and 1994, respectively, and the Ph.D. degree from the University of Wisconsin-Madison, Madison, WI, USA, in 1998.

He joined the Electrical Engineering Department, The Pennsylvania State University, University Park, PA, USA, in 1998, where he is currently a Professor. His current research interests include robust optimization and control, chance constrained optimization, controller design under risk specifications, system identification, control of computer networks, and application of control approaches to the analysis and design of adaptive behavioral interventions.

Dr. Lagoa is an Associate Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL and *Automatica*. He is also a member of the IFAC Technical Committee on Robust Control.

Suzan A. Murphy is currently the H.E. Robbins Distinguished University Professor of Statistics, a Professor of Psychiatry, and a Research Professor with the Institute for Social Research, Oslo, Norway. Her current research interests include improving sequential, individualized decision making in health, in particular on clinical trial design and data analysis to inform the development of mobile health treatment policies.

Ms. Susan is a fellow of the Institute of Mathematical Statistics, a fellow of The College on Problems in Drug Dependence, a Former Editor of the *Annals of Statistics*, a member of the U.S. National Academy of Sciences and the U.S. National Academy of Medicine, and a 2013 MacArthur Fellow.

Korkut Bekiroglu received the B.S. degree from the Department of Electrical and Electronics Engineering, Karadeniz Technical University, Trabzon, Turkey, in 2005, the M.S. degree from the Department of Electrical Engineering, Northeastern University, Boston, MA, USA, in 2010, and the Ph.D. degree from the Department of Electrical Engineering, The Pennsylvania State University, University Park, PA, USA, in 2010.

He is currently a Post-Doctoral Scholar with the Methodology Center, The Pennsylvania State University. His current research interests include the interface of system identification, longitudinal data analysis, robust control, robust model predictive controller design, robust interventions, adaptive treatment, discrete event dynamical systems, and robust optimization.

Stephanie T. Lanza is currently a Professor of Biobehavioral Health and the Scientific Director of The Methodology Center with The Pennsylvania State University, University Park, PA, USA. She is currently involved in the areas of finite mixture modeling and time-varying effect modeling, and is devoted to disseminating cutting-edge statistical methods to applied researchers. Her current research interests include advancing analytic methods relevant for studying multidimensional and dynamic health behaviors.

Ms. Lanza is a Board Member of the Society for Prevention Research and an Associate Editor of the *Prevention Science*.