Modeling & Control of Human Actuated Systems

Sangjae Bae * **Sang Min Han** ** **Scott Moura** *,***

* *Energy, Controls, and Applications Lab (eCAL), Department of Civil and Environmental Engineering, University of California, Berkeley, USA. (e-mail: {sangjae.bae, smoura}@berkeley.edu).*
** *Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, USA. (e-mail: smhan@berkeley.edu)*
*** *Smart Grid and Renewable Energy Lab, Tsinghua-Berkeley Shenzhen Institute, Shenzhen, P.R. China*

Abstract:
This paper investigates a Cyber-Physical & Human System (CPHS) comprised of a deterministic dynamical system plant model and a human actuator model. Namely, human decisions are stochastic inputs to the plant model. We examine a framework where human decisions cannot be directly controlled, but can be influenced via incentive control signals. Specifically, we use the framework of discrete choice models (DCMs) to capture human decision making, and then design optimal controllers for these human actuated dynamical systems. Existing literature on CPHS often treats human inputs as stochastic and exogenous inputs, and then formulates a disturbance rejection problem. Instead of treating human decision-making as an uncontrollable exogenous input, we directly incorporate human decision making into the modeling framework with DCMs. This paper thus adds two original contributions. (i) We develop a generalized human-actuated system framework based on DCM that predicts the probability of human decisions, conditioned on controllable incentives. (ii) We show that existing optimization schemes, such as Sequential Quadratic Programming (SQP) and Dynamic Programming (DP), can be applied to control the proposed human-actuated system. We conclude this paper by demonstrating the framework on a reference tracking problem and an inventory control problem.

Keywords: cyber-physical and human system, human-in-the-loop system, human behavior modeling and control, discrete choice model and control

1. INTRODUCTION

Cyber-physical systems (CPS) are engineered systems that integrate automatic control, network communication, sensors, and computation in applications such as smart grids and smart cities (see Cocchia (2014); Karnouskos (2011)). In CPS, human input can be a crucial component, but plays different roles in different contexts. From the control designers' perspective, human behavior is often regarded as a random disturbance, yielding a stochastic system that we seek to stabilize in some sense (see Pentland and Liu (1999)). Human inputs are also sometimes regarded as a correction mechanism designed to compensate errors in human-incorporated monitoring systems (see Mukhopadhyay (2015)). In practice, human beings are not necessarily irrational decision makers that are best modeled as simple stochastic disturbances. On the other hand, humans are imperfect decision makers that can be approximately represented by deterministic mathematical models. Instead, we consider humans as independent, intelligent agents who act to maximize their idiosyncratic utility functions. Their actions are ultimately influenced by both internal and external factors, e.g. price incentives and weather. Developing a mathematical framework to account for behavior from independent, intelligent agents in a control system is a technically challenging problem that Munir et al. (2013) argues should be studied more thoroughly. This paper addresses this challenge.

Different approaches for modeling human behavior have been proposed in the past, but this modeling remains a difficult task in CPHS. In automotive control, human behavior is often modeled and predicted using a probabilistic model (e.g. Markov Chains), and are regarded as random system inputs that are not subject to control (see Pentland and Liu (1999)). In economics and social sciences, discrete choice models (DCMs) have been studied to formalize the human decision making process under a finite set of alternatives and the corresponding retributions (see McFadden and Train (2000)).

In our previous study (Bae et al. (2018)), we modeled a dynamical system with human actuators based on the state-space representation and a discrete choice model. We then proposed a convex optimization scheme that determines the optimal price incentives, which induce certain human behavior. The convex optimization approach however has

⋆ This work was funded, in part, by Total S.A. and the Tsinghua-Berkeley Shenzhen Institute (TBSI).

limitations in the scope of practical applications as the proposed scheme is limited to human behavior models with only a few alternatives. Additionally, a convexity constraint in (Bae et al. (2018)) significantly restricts the solution domain of the optimization problem, and thus yields overly conservative solutions. Furthermore, the price incentive determined by the optimization problem is not alternative specific, i.e. the price incentive is calculated after collectively considering all alternatives in retrospect.

In this paper, we propose a more generalized model of human-in-the-loop systems and control schemes that circumvent the aforementioned limitations, and are applicable to a wider range of problems. The main contributions of this paper are twofold. (i) We first propose a generalized model of the human-in-the-loop system with multiple alternatives that can be controlled by various methods. (ii) We then describe Sequential Quadratic Programming (SQP) and Dynamic Programming (DP) approaches for control, which determine the optimal incentive signals that induce human actions that are preferred by the system operator.

The paper is organized in the following manner. Section 2 describes the CPHS modeling and Section 3 and 4 demonstrate the proposed framework on reference tracking and inventory control problems, respectively. Section 5 discusses potential extensions to the proposed framework. We conclude with discussions of the modeling and control framework in Section 6.

2. SYSTEM MODEL

Figure 1 illustrates the CPHS framework, which consists of three main components: (i) the human actuator, (ii) plant, and (iii) controller. A human actuator produces a random decision $\boldsymbol{S}_k$ at time step k, based on a discrete choice model (DCM). Decision $\boldsymbol{S}_k$ is then sent as an input to the plant. A controller produces two control signals: a plant control signal $\boldsymbol{u}_k$, and an incentive control signal $\boldsymbol{z}_k$ that influences the human's choices. The ultimate objective is to synthesize a controller that optimally manages the closed-loop system. We mathematically formalize the human behavior model and incorporate it into the plant's system dynamics. Specifically, we use a parametric discrete choice model for human behavior and incorporate it into the general state space-based dynamical system where existing control theoretic algorithms can be applied. The corresponding control methodology is discussed in the next section. We limit ourselves to linear system dynamics, i.e. linear state evolution equation, for a more intuitive presentation. The presented framework can be extended to nonlinear system dynamics, and is discussed in Section 5.

2.1 Human Behavior Modeling

Human behavior can be modeled using either parametric or nonparametric probabilistic models. We use parametric models here to quantitatively evaluate the impact of a specific parameter associated with a controlled variable z_k, e.g. price incentives. The actuator model also takes exogenous input variable w_k, which are not controlled, e.g. weather. We assume the human makes decisions among a finite set of actions, and chooses one action at a time, i.e. the action space for $\boldsymbol{S}_k$ is finite. We then consider the discrete choice model with multinomial logistic regression as an empirical behavior model (see Ben-Akiva and Lerman (1985)).

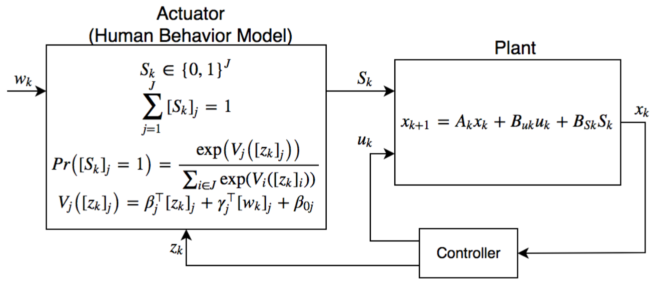

Fig. 1. Block diagram of the human-actuated system.

In the discrete choice model, each choice, also called an "alternative" in the econometrics literature, has a specific utility function. One alternative is chosen when its utility is greater than all others. For the j-th alternative, $j \in \{1, 2, \cdots, J\}$, the utility function is

$$U_j = f_j(z_j, w_j) \doteq \beta_j^\top z_j + \gamma_j^\top w_j + \beta_{0j} + \epsilon_j, \qquad (1)$$

where $z_j \in \mathbb{R}^{n_z}$ is a vector of controllable inputs; $w_j \in \mathbb{R}^{n_w}$ is a vector of uncontrollable inputs for alternative j; $\beta_j \in \mathbb{R}^{n_z}$ and $\gamma_j \in \mathbb{R}^{n_w}$ are parameter vectors [1] for the controllable inputs and uncontrollable inputs, respectively; β_{j0} is the "alternative specific" constant; and ϵ_j accounts for unspecified errors. Note that U_j need not be positive definite. In fact, with the choice of probability distribution for ϵ_j given below, having the same utility for all alternatives, i.e. $U_j = U_i \ \forall \ i \neq j$, corresponds to all alternatives being equally probable.

Denote by $[S_k]_j$ a random binary variable, $\{0, 1\}$, that indicates human selection of alternative j, i.e. $[S_k]_j = 1$ if alternative j is chosen and $[S_k]_j = 0$ otherwise. The random errors ϵ_j induce a probability distribution on S_k, which is characterized by a probability mass function (PMF) corresponding to the selection of each alternative. We assume that the errors ϵ_j are *i.i.d* and follow the Extreme Value distribution, which is generally known as the multinomial logit model in econometrics. Under these assumptions, the probability of choosing alternative j at time step k is

$$\Pr([S_k]_j = 1) = \Pr\left[\bigcap_{i \neq j} (U_j > U_i)\right] = \frac{e^{V_j}}{\sum_{i=1}^{J} e^{V_i}}, \qquad (2)$$

where $V_j \doteq \beta_j^\top z_j + \gamma_j^\top w_j + \beta_{0j}$ is the utility of alternative j without errors ϵ_j.

2.2 Human Behavior Incorporation into System Dynamics

The aforementioned behavior model describes probabilistic human decisions on the plant. We consider a dynamical system where the actuator is mathematically represented by the human behavior model in (2) to quantitatively evaluate the effect of human inputs on the plant. Human decisions are ultimately random inputs to the plant, characterized by a probability mass function that is modulated by a control input u_k. Formally, the probability of choosing

[1] Parameters can be general or specific over alternatives

alternative j is a function of $z = \{z_1, z_2, \ldots, z_J\}$ and $w = \{w_1, w_2, \ldots, w_J\}$:

$$g_j(z, w) = \frac{e^{V_j(z_j, w_j)}}{\sum_{i=1}^{J} e^{V_i(z_i, w_i)}}. \tag{3}$$

We define *human actuated system* to be a system that evolves according to both direct controlled inputs u_k and random human inputs $\boldsymbol{S}_k$. For $k \in \{0, \ldots, N-1\}$, $\mathcal{J} = \{1, \ldots, J\}$, $(N, J) \in \mathbb{N}^+ \times \mathbb{N}^+$ [2], the linear time-varying human actuated system, with a single human actuator, is

$$x_{k+1} = A_k x_k + B_{uk} u_k + B_{Sk} S_k, \tag{4}$$

where $S_k \in \{0,1\}^J$, $[S_k]_j = 1$ if alternative j is chosen and 0 otherwise, and

$$\sum_{j \in \mathcal{J}} [S_k]_j = 1. \tag{5}$$

The state is $x_k \in \mathbb{R}^n$, $A_k \in \mathbb{R}^{n \times n}$ is the system matrix, $B_{uk} \in \mathbb{R}^{n \times n_u}$ is the direct input matrix, and $B_{Sk} \in \mathbb{R}^{n \times J}$. We denote by $u_k \in \mathbb{R}^{n_u}$ and $S_k \in \mathbb{R}^J$ the direct input and the random human input to the system, respectively. Equation (5) indicates that the system accepts only one decision from the human actuator, and that B_{Sk} evaluates the system impact of only one action at a given time. Then, the probability of choosing alternative j at time step k can be written

$$\Pr([S_k]_j = 1) = g_j(z_k, w_k) \ \forall \ j = 1, \ldots, J. \tag{6}$$

Claim 1. The model can be extended to account for $M > 1$ human actuators, each with multiple alternatives, by simply extending the input matrix B_{Sk} and S_k appropriately. For every additional human actuator, B_{Sk} increases in size by J columns and S_k increases by J rows, i.e. we append the J number of alternatives of each additional human actuator to S_k. More generally, if the m-th human actuator has J_m number of alternatives, B_{Sk} will have $\sum_{m=1}^{M} J_m$ columns and S_k will have $\sum_{m=1}^{M} J_m$ elements. With this formulation, the system impact induced by M human actions is the aggregate system impact over M human actuators.

Incorporating the human actuated system into existing feedback control mechanisms is, unfortunately, mathematically difficult due to its probabilistic nature. Human input $\boldsymbol{S}_k$ is a binary random vector with a distribution that is modulated by incentive control z_k. We consider controlling the mean dynamics. By considering the expected value of the state at each time step k, each indicator function $[S_k]_j$ is replaced by the probability of the j-th human action. The human-actuated mean dynamics is a deterministic nonlinear system where the expected system input $g(z_k, w_k)$ includes probabilities of human actions as a function of the incentive control z_k and the exogenous variable w_k. This system is formulated as

$$\bar{x}_{k+1} = A_k \bar{x}_k + B_{uk} u_k + B_{Sk} g(z_k, w_k), \tag{7}$$

where $\bar{x}_k \doteq \mathbb{E}[x_k]$, $g(z_k, w_k) \doteq [\mathbb{E}\left[[S_k]_1\right] \cdots \mathbb{E}[[S_k]_J]]^\top$, and for every $j \in \{1, \ldots, J\}$,

$$\mathbb{E}[[S_k]_j] = g_j(z_k, w_k) = \frac{e^{V_{jk}([z_k]_j, [w_k]_j)}}{\sum_{i=1}^{J} e^{V_{ik}([z_k]_i, [w_k]_i)}}, \tag{8}$$

which is equivalent to $\Pr([S_k]_j = 1)$.

[2] $\mathbb{N}^+$ denotes strictly positive integers

Claim 2. Other parametric discrete choice probability models can fit this modeling framework by simply replacing the probability of every j-th alternative $\Pr\left([S_k]_j = 1\right)$ appropriately.

3. REFERENCE TRACKING PROBLEM

In this section, we apply the CPHS modeling framework to solve a reference tracking problem. We consider a sequential decision making problem from the perspective of a systems engineer. The objective of the controller is to have the plant state follow a reference trajectory x_k^{ref}. Practical applications include modulating food prices to follow a desirable agricultural output trajectory, modulating electricity prices to follow renewable generation, or a human and robot symbiotically operating a machine. The engineer must induce human behavior in a manner that minimizes the plant state's deviation from the reference trajectory. At every time step, the controller must produce two signals: (i) a direct control to steer the plant's state towards the reference trajectory; and (ii) an incentive to induce a desired human input that minimizes the plant state's deviation from the reference trajectory.

3.1 Formulation of Optimization Problem

The objective is to design a controller that minimizes the deviation of the plant's state trajectory from the reference trajectory while simultaneously balancing the state deviation cost and the control effort. For the time-invariant human actuated system $f(\bar{x}_k, u_k, z_k, w_k) = A\bar{x}_k + B_u u_k + B_S g(z_k, w_k)$, $k \in \{0, ..., N-1\}$, the optimization problem is formulated as

$$\begin{aligned}
\text{minimize}_{\bar{x}_k, u_k, z_k} \quad F = \sum_{k=0}^{N-1} \Big[& (\bar{x}_k - x_k^{\text{ref}})^\top Q_k (\bar{x}_k - x_k^{\text{ref}}) \\
& + (u_k - u_k^{\text{ref}})^\top R_{uk} (u_k - u_k^{\text{ref}}) \\
& + (z_k - z_k^{\text{ref}})^\top R_{zk} (z_k - z_k^{\text{ref}}) \Big] \\
& + (\bar{x}_N - x_N^{\text{ref}})^\top Q_N (\bar{x}_N - x_N^{\text{ref}}) \qquad (9)
\end{aligned}$$

$$\begin{aligned}
\text{subject to: } & \bar{x}_0 = \bar{x}_{\text{init}}, & (10) \\
& \bar{x}_{k+1} = f(\bar{x}_k, u_k, z_k, w_k), & (11) \\
& z_k \geq 0, & (12)
\end{aligned}$$

where Q_k, R_{uk}, R_{zk} are weight matrices for the errors in the mean state, direct control effort, and incentive control effort at time step k, respectively. Signals x_k^{ref}, u_k^{ref}, and z_k^{ref} are reference state trajectories, reference direct control, and reference incentive control, respectively. The final weight matrix is Q_N, $\bar{x}_N$ is a final mean state, and $\bar{x}_N^{\text{ref}}$ is a final reference state. The first, second, and third terms in the objective function (9) indicate the error penalties of the mean state, direct control effort, and the incentive control effort at time step k, respectively, and the last term evaluates the error penalty of the mean state at the final step. The initial condition $\bar{x}_{\text{init}}$ is given in (10). The equality constraint (11) incorporates the human actuated system model into the optimization problem and the inequality constraint (12) ensures nonnegative incentives.

3.2 Solving Optimization Problem with SQP

This optimization problem is non-convex because the plant dynamics (11) are nonlinear with respect to the incentive control z_k. Due to the non-affine nonlinearity in the control signal z_k, feedback linearization methods, such as those described in Khalil (1996), may not apply. We utilize a Sequential Quadratic Programming (SQP) approach [3] (see Nocedal and Wright (2006)). SQP iteratively finds sub-optimal control polices for nonlinearly constrained optimization problems by solving quadratic approximations to the original problem at every iteration.

Linearization

SQP essentially replaces the nonlinear constraint (11) with a linear approximation. Using Taylor's theorem, we can approximate the system as [4]

$$\tilde{x}_{k+1} = \tilde{A}_k\tilde{x}_k + \tilde{B}_{uk}\tilde{u}_k + \tilde{B}_{zk}\tilde{z}_k, \tag{13}$$

where $\tilde{x}_k \doteq \bar{x}_k - \bar{x}_k^{\text{ref}}$, $\tilde{u}_k \doteq u_k - u_k^{\text{ref}}$, $\tilde{z}_k \doteq z_k - z_k^{\text{ref}}$, and

$$\tilde{A}_k \doteq \nabla_{\bar{x}}\, f_k\left(\bar{x}_k^{\text{ref}}, u_k^{\text{ref}}, z_k^{\text{ref}}, w_k\right)^\top \tag{14}$$

$$\tilde{B}_{uk} \doteq \nabla_{u}\, f_k\left(\bar{x}_k^{\text{ref}}, u_k^{\text{ref}}, z_k^{\text{ref}}, w_k\right)^\top \tag{15}$$

$$\tilde{B}_{zk} \doteq \nabla_{z}\, f_k\left(\bar{x}_k^{\text{ref}}, u_k^{\text{ref}}, z_k^{\text{ref}}, w_k\right)^\top. \tag{16}$$

Note that $\tilde{A}_k = A$ and $\tilde{B}_{uk} = B_u$ since $f_k(\bar{x}_k, u_k, z_k, w_k)$ is linear with respect to $\bar{x}_k$ and u_k. Denote by $\tilde{B}_{zk,m}^\top$ the m-th row of $\tilde{B}_{zk}$ for $m = 1, \dots, n$. Then,

$$\tilde{B}_{zk,m}^\top = \begin{bmatrix} \frac{\beta_{k1}\exp\{V_{k1}([z_k^{\text{ref}}]_1,[w_k]_1)\}}{\left(\sum_{i=1}^J \exp\{V_{ki}([z_k^{\text{ref}}]_i,[w_k]_i)\}\right)^2}\psi_{k,m}(z_k^{\text{ref}}, w_k, 1) \\ \vdots \\ \frac{\beta_{jk}\exp\{V_{kJ}([z_k^{\text{ref}}]_J,[w_k]_J)\}}{\left(\sum_{i=1}^J \exp\{V_{ki}([z_k^{\text{ref}}]_i,[w_k]_i)\}\right)^2}\psi_{k,m}(z_k^{\text{ref}}, w_k, J) \end{bmatrix}^\top \tag{17}$$

where for $j = 1, \dots, J$, the function $\psi_{k,m}$ is defined

$$\psi_{k,m}(z_k^{\text{ref}}, w_k, j) \doteq \sum_{i=1}^{J}\left([B_{zk,m}^\top]_j - [B_{zk,m}^\top]_i\right)\exp\{V_{ik}([z_k^{\text{ref}}]_i, [w_k]_i)\}. \tag{18}$$

Reformulation of Optimization Problem

To apply the SQP framework to the aforementioned optimization problem, we first rewrite the optimization problem with respect to stacked variables v and v^{ref}

$$\text{minimize}_v \;\; (v - v^{\text{ref}})^\top H(v - v^{\text{ref}}) \tag{19}$$

subject to:

$$\begin{bmatrix} \bar{x}_{\text{init}} - \{v\}_{\bar{x}_0} \\ A\{v\}_{\bar{x}_0} + B_u\{v\}_{u_0} + B_S\{v\}_{z_0} - \{v\}_{\bar{x}_1} \\ \vdots \\ A\{v\}_{\bar{x}_{N-1}} + B_u\{v\}_{u_{N-1}} + B_S\{v\}_{z_{N-1}} - \{v\}_{\bar{x}_N} \end{bmatrix} = 0, \tag{20}$$

$$[\{v\}_{z_0}, \dots, \{v\}_{z_{N-1}}] \geq 0, \tag{21}$$

where

$$v = [\bar{x}_0, u_0, z_0, \dots, \bar{x}_{N-1}, u_{N-1}, z_{N-1}, \bar{x}_N], \tag{22}$$

$$v^{\text{ref}} = [\bar{x}_0^{\text{ref}}, u_0^{\text{ref}}, z_0^{\text{ref}}, \dots, \bar{x}_{N-1}^{\text{ref}}, u_{N-1}^{\text{ref}}, z_{N-1}^{\text{ref}}, \bar{x}_N^{\text{ref}}], \tag{23}$$

$$H = \text{diag}\{Q_0, R_{u0}, R_{z0}, \dots, Q_{N-1}, R_{u(N-1)}, R_{z(N-1)}, Q_N\}, \tag{24}$$

and $\{v\}_{(*)}$ denotes the elements of the stacked variable v that correspond to $(*)$. We then take a second-order approximation of the Lagrangian function $\mathcal{L}$ of (19)-(21) with respect to an optimal solution $\hat{v}^{(i)}$ obtained at iteration i. We also linearize the equality constraints (20) with respect to $\hat{v}^{(i)}$. We eventually formulate the SQP optimization problem

$$\text{minimize}_p \;\; F^{(i)} + \nabla F^{(i)\top} p + \frac{1}{2}p^\top \nabla_{vv}^2 \mathcal{L}^{(i)} p \tag{25}$$

subject to:

$$(\text{diag}\{\left[A \;\; B_u \;\; \tilde{B}_{zk}\right]\}_{k=0}^{N-1})^\top p + \begin{bmatrix} \bar{x}_{\text{init}} - \{\hat{v}^{(i)}\}_{\bar{x}_0} \\ A\{\hat{v}^{(i)}\}_{\bar{x}_0} + B_u\{\hat{v}^{(i)}\}_{u_0} + B_S\{\hat{v}^{(i)}\}_{z_0} - \{\hat{v}^{(i)}\}_{\bar{x}_1} \\ \vdots \\ \Big(A\{\hat{v}^{(i)}\}_{\bar{x}_{N-1}} + B_u\{\hat{v}^{(i)}\}_{u_{N-1}} \\ + B_S\{\hat{v}^{(i)}\}_{z_{N-1}} - \{\hat{v}^{(i)}\}_{\bar{x}_N}\Big) \end{bmatrix} = 0, \tag{26}$$

$$\nabla c(\hat{v}^{(i)})^\top p + c(\hat{v}^{(i)}) \geq 0, \tag{27}$$

where p is the SQP decision variable, and $\tilde{B}_k$ is defined in (16). Gradient $\nabla c(\hat{v}^{(i)})$ has a value of 1 in the place of elements that correspond to the incentive control z_k. It is a convex quadratic program with linear constraints and thus we can obtain p^* using canonical convex optimization methods (see Boyd and Vandenberghe (2004)).

Update algorithm

Once we obtain p^* from the aforementioned SQP, the optimal control $\hat{v}^{(i)}$ at iteration i is updated according to

$$\hat{v}^{(i+1)} = \hat{v}^{(i)} + \alpha^{(i)} p^*. \tag{28}$$

Then, the algorithm iterates until a set of convergence criteria are satisfied. SQP inherits the convergence rate of the quasi-Newton method, and convergence can be guaranteed by an appropriate choice of $\alpha^{(i)}$ (see, e.g. backtracking line-search method Nocedal and Yuan (1998); Li and Todorov (2004)) assuming appropriate regularity conditions hold and unlikely edge cases (e.g. infeasibility of the problem) do not arise.

3.3 Simulation

We illustrate a simple instance of the reference tracking problem. We consider a third order system with

[3] One drawback of SQP is that it requires the objective function to be twice differentiable, and the Hessian of the Lagrangian must be positive definite. These requirements limit some applications.

[4] The derivation takes only one vector of controlled incentive $z_k \in \mathbb{R}^J$, but can be easily extended to multiple incentive controls by extending the dimension of z_k, i.e. $z_k \in \mathbb{R}^{J\times s}$, where $s > 1$ is the number of incentive controls. This results in $s - 1$ additional columns to $\tilde{B}_{zk}$, which corresponds to the partial derivatives of $f_k(\bar{x}_k, u_k, z_k, w_k)$ with respect to each additional incentive control.

$$A = \begin{bmatrix} 0.1 & 1 & -1 \\ 1 & 0.1 & 1 \\ 1 & 0 & 0.5 \end{bmatrix}, B_u = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, B_S = \begin{bmatrix} -5 & 0 & 5 \\ -5 & 0 & 5 \\ -5 & 0 & 5 \end{bmatrix}, \tag{29}$$

Note the open-loop system is unstable, yet controllable, and each alternative, when chosen, sends a corresponding input to the system. Namely, the human's choice either (i) adds negative input (-5); (ii) adds zero input; or (iii) adds positive input $(+5)$. The DCM parameters are set to $\beta_{j0k} \equiv 1 \ \forall \ j,k$ so that the human is equally likely to choose each alternative with zero incentives and zero exogenous factors. Similarly, $\beta_{jk} \equiv 1 \ \forall \ j,k$ so that the human is equally sensitive to all incentives, for choosing each alternative. We also assume for simplicity that the exogenous variables do not influence the human decision process, i.e. $w_k \equiv 0$. We consider a scenario where the first state must follow a series of step functions. We consider reference control signals of zero, i.e. $u_k^{\text{ref}} \equiv z_k^{\text{ref}} \equiv 0$, for simplicity. The error penalty is set to $Q_k \equiv \text{diag}\{5,0,0\}$, which is higher than the control effort costs set to $R_{uk} \equiv 1$ and $R_{zk} \equiv \mathbf{I}_{3\times3}$. We apply SQP [5] to find the optimal direct control u_k and incentive control z_k for tracking the reference trajectory.

Figure 2 shows the first state $\bar{x}_{1,k}$, direct control u_k, and incentive control z_k trajectories. The SQP-derived control signals are successful in tracking the desired state trajectory. The third and fourth subplots in Figure 2 illustrate how the probability of choosing each alternative varies with the incentive control signal. We highlight two interesting observations from this simple example. First, consider the steady-state behaviors from about $k \in [2,4]$ and $k \in [5.5, 7.5]$. During these periods, u_k is essentially zero, yet the incentives for non-zero human actuation $([z_k]_1, [z_k]_3)$ are non-zero. Recall that the plant is open-loop unstable, and therefore control effort is required to maintain a constant value for $\bar{x}_1$. In this case, the optimal strategy is to leverage human actuation z_k instead of direct control u_k to stabilize the plant around constant $\bar{x}_{1,k}$ values. Second, the control effort cost with incentive control (349.1) is lower than that without incentive control (472.13). In this simple example, the cost savings increase as the role of the human input to the plant increases. This finding suggests that incentive control has strong potential for cost savings in practical CPHS control problems, where human decisions significantly impact the plant's dynamics.

4. INVENTORY CONTROL PROBLEM

Next, we apply the CPHS modeling framework to solve an inventory control problem. We consider a sequential decision making problem from the perspective of a store manager, who is selling J distinct items from their store's inventory. The store manager must make two decisions at every time step: (i) how many of each item to order from their supplier to restore their inventory, and (ii) how much price discount they should apply to each item to encourage sales. The store manager must take into account the type and quantity of items that will be sold over the time horizon of interest, and set appropriate price discounts to maximize their expected net revenue. We assume orders from the supplier arrive immediately. This application fits perfectly into the proposed framework. Specifically, the

[5] We use Matlab SQP fmincon.

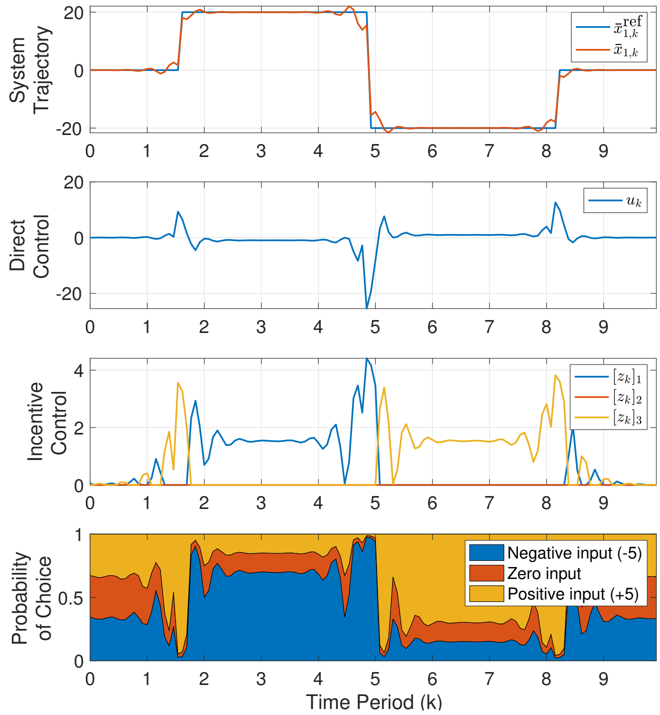

Fig. 2. Simulated reference tracking results solved via SQP, for an open-loop unstable plant model. In the third plot, $[z_k]_1$ is an incentive associated to the negative input, $[z_k]_2$ for the zero input, and $[z_k]_3$ for the positive input.

inventory dynamics represent the plant dynamics, orders from the supplier represent the direct control, and price discounts represent the human incentive.

4.1 Human Actuated System Modeling

For simplicity, we consider just one item type and have each alternative corresponding to the number of items purchased over a time period, i.e., the set of alternatives is $\{0, 1, \ldots, J-1\}$. Given a total of J alternatives, the utility function of alternative j over the time period $k \in \{0, \ldots, N-1\}$ is defined as:

$$V_{jk}([z_k]_j, [w_k]_j) \doteq \beta_{jk}[z_k]_j + \gamma_k[w_k]_j + \beta_{0jk}, \tag{30}$$

with the incentive parameter β_{jk}, incentive control $[z_k]_j$, exogenous variable parameter γ_{jk}, exogenous variable $[w_k]_j$, and alternative specific constant β_{k0j}. The system dynamics can be written as

$$x_{k+1} = x_k + u_k - B_S S_k, \tag{31}$$

where x_k is the stock level of the item, u_k is the number of items ordered from supplier, $S_k \in \{0,1\}^J$ is the vector of indicators, and $B_S = [0, 1, ..., J-1]^\top$. The mean dynamics follow the development in Section 2.2.

4.2 Formulation of Optimization problem

The objective is to minimize the overall net cost (i.e. maximize net revenue) by finding a control policy for u_k (orders from supplier) and z_k (price discount). The store manager sets different price discounts based on the quantity of items purchased. The optimal control problem can then be written as

$$\text{minimize}_{u_k,z_k} \sum_{k=0}^{N-1} [c_u u_k - rB_S g(z_k, w_k) + \sum_{j=1}^{J} [c_z]_j [z_k]_j g_j(z_k, w_k)] \quad (32)$$

$$\text{subject to: } \bar{x}_{k+1} = \bar{x}_k + u_k - B_S g(z_k, w_k), \quad (33)$$

$$0 \le \bar{x}_k \le \bar{x}_{\max} \quad (34)$$

$$0 \le u_k \le u_{\max} \quad (35)$$

$$0 \le z_k \le z_{\max}, \quad (36)$$

where c_u is the cost-per-unit from the supplier, c_z is the cost-per-incentive from the store manager, and r is the revenue-per-unit from customers. The first term in (32) is the cost of restocking inventory and the second and third terms indicate the expected revenue from customers and the expected price incentives to expend, respectively. The stock level, the number of items ordered from the supplier, and the price incentive are upper bounded by $\bar{x}_{\max}$, $u_{\max}$, and $z_{\max}$, respectively. Again, the key challenge is that (33) is nonlinear with respect to z_k.

4.3 Solving Optimization Problem with DP

The second derivative of the objective function (32) is not positive definite with respect to $\{u, z\}$, and thus SQP would not be a proper fit to solve this problem. Instead, we apply dynamic programming (DP) [6] to numerically obtain the optimal control policy. Let $J_k(\bar{x}_k)$ denote the minimum net cost from time step k to N, given that the stock level is $\bar{x}_k$ at time step k. The Bellman equation is written as

$$J_k(\bar{x}_k) = \min_{u_k,z_k} \{c_u u_k - rB_S g(z_k, w_k) + \sum_{j=1}^{J} [c_z]_j [z_k]_j g_j(z_k, w_k) + J_{k+1}(\bar{x}_{k+1})\}, \quad (37)$$

where $\bar{x}_{k+1} = \bar{x}_k + u_k - B_S g(z_k, w_k)$ for $k \in \{0, \ldots, N-1\}$. The terminal cost is $J_N(\bar{x}_N) = 0$. In other words, the revenue/cost of any remaining stock at the terminal time step is zero.

4.4 Simulation

We simulate a simple instance of the above case study, with the number of alternatives set to $J = 3$ and the number of choice situations of customers set to $N = 30$. Under this setting, a total of 30 customers enter the store and each customer either (i) does not purchase anything; (ii) buys one item; or (iii) buys two items. We set the initial stock level to $\bar{x}_0 = 5$, cost-per-unit to $c_u = 10(\$)$, cost-per-incentive to $c_z = 1(\$)$, and revenue-per-unit to $r = 20(\$)$. The DCM parameters are set to $\beta_{0jk} \equiv \{1, 0.5, 0.3\}$ so that customers are more likely to buy less items when they do not have a price discount, $\beta_{jk} \equiv \{0, 0.5, 0.3\}$ so that customers have less sensitivity to price incentives when buying two items, and $\gamma_k \equiv 0.5$ so that customers are equally sensitive to an exogenous factor as they are to price incentives. We further assume that the exogenous variables $[w_k]_j$ are sampled from the standard normal distribution, fixed, and known *a priori*. Using DP, we compute an optimal control policy $[u_k, z_k] = \pi_k(\bar{x}_k)$ for the supplier order and sales discount.

[6] DP numerically computes the optimal control inputs via Bellman's principle of optimality, without requiring linearization or conditions on the Lagrangian (see Bellman (2013)). However, it also suffers from exponential computational complexity, limiting its use to systems with relatively few states and controls.

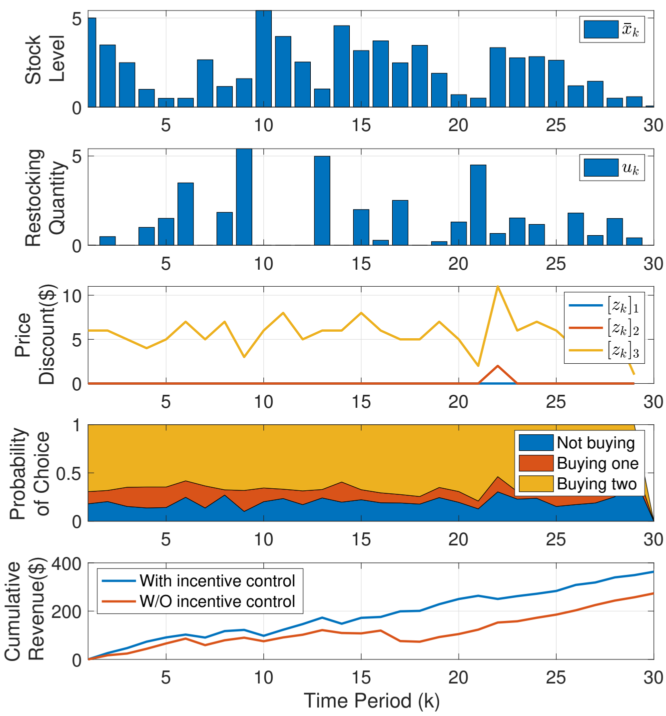

Fig. 3. A simulation result for an inventory control problem. In the third plot, $[z_k]_1$ is the incentive associated with buying zero items, $[z_k]_2$ for buying one item, and $[z_k]_3$ for buying two items.

Figure 3 presents a simulation result of the aforementioned decision problem. Note the stock $\bar{x}_k$ is depleted by the terminal time step, since remaining stock represents missed revenue. The discount for selling zero items $[z_k]_1$ is zero, since no revenue can be recovered from this action. The discount for selling two items $[z_k]_3$ is often non-zero, to incentivize sales. The fifth plot in the figure indicates that the expected cumulative net revenue is higher with optimal price incentives compared to that without incentives. The increase in the expected cumulative net revenue is further explored via Monte Carlo simulation, shown in Figure 4. The Monte Carlo simulation indicates that the expected net revenue increases significantly (+\$154.63) as a consequence of using price incentives. It is also shown that the standard deviation of the net revenue with the incentives is lower (\$76.38) than that without the incentives (\$107.58). Future work should validate these results on behavioral models parameterized on real-world data, and with real exogenous variables.

5. EXTENSIONS AND LIMITATIONS

The proposed system modeling framework can be extended to dynamical systems that are nonlinear in the states, without loss of generality. In this case, the incorporated system of mean dynamics is generalized as $\bar{x}_{k+1} = f(\bar{x}_k, u_k, z_k, w_k)$, $k \in \{0, \ldots, N-1\}$, with state mean $\bar{x}_k$, direct control u_k, and incentive control z_k. Then optimal solutions are found by SQP or DP, as described

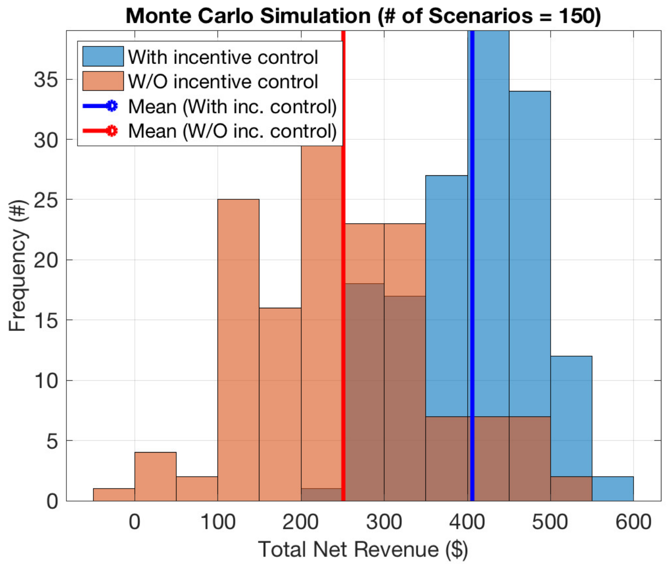

Fig. 4. Monte Carlo simulation results for over 150 randomized scenarios. In each scenario, human choices are sampled from probabilistic distributions generated by the DCM with multinomial logit regression. The net revenue with price incentive control is calculated as the total profit minus the total price incentive and restocking cost.

in Section 3 and Section 4. In addition, the framework with multinomial logistic (ML) regression can be further generalized by mixed multinomial logistic (MML) regression (see Bhat (2001)), which addresses partially observed behaviors, i.e. DCM parameters are random variables.

One limitation of the framework is that increasing the dimension of the incentive control z_k is followed by a linear increase in the total number of alternatives within the system. Nonetheless, resolving the curse of dimensionality is an open problem, and approximation methods described by Fodor (2002) can be incorporated. Another limitation is that the SQP and DP-based control strategies discussed in this paper are based upon the deterministic mean dynamics. Consequently, they fail to capture heterogeneous individual behavior, and emergent dynamics. However, the control strategies in this paper can be extended to account for stochastic dynamics, via stochastic dynamic programming (see Bertsekas (1995)).

6. CONCLUSION

This paper provides a formalized mathematical modeling and control framework for human-actuated systems, in the field of cyber-physical & human systems (CPHS). A human-actuated system quantitatively evaluates the impact of human behaviors on system dynamics. Human behaviors are random, and the probabilities associated with specified decisions are assessed by Discrete Choice Models (DCMs) developed in behavioral economics McFadden and Train (2000). We mathematically incorporate the human-actuated system into a nonlinear optimal control problem, solved via Sequential Quadratic Programming (SQP) and Dynamic Programming (DP). Finally, we applied the proposed modeling and control framework on two problems: (i) A reference tracking problem where human inputs help the system follow a reference state trajectory; (ii) an inventory control problem where the objective is to find an optimal control policy for sales price discounts to maximize net revenue. A simple empirical validation is performed for the inventory control problem using Monte Carlo simulation. Future work includes developing an active learning algorithm for identifying the human behavioral model parameters as observations are gathered.

REFERENCES

Bae, S., Han, S.M., and Moura, S. (2018). System analysis and optimization of human-actuated dynamical systems. In *2018 Annual American Control Conference (ACC)*, 4539–4545. IEEE.

Bellman, R. (2013). *Dynamic programming.* Courier Corporation.

Ben-Akiva, M.E. and Lerman, S.R. (1985). *Discrete choice analysis: theory and application to travel demand*, volume 9. MIT press.

Bertsekas, D.P. (1995). *Dynamic programming and optimal control*, volume 1. Athena Scientific Belmont, MA.

Bhat, C.R. (2001). Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B: Methodological*, 35(7), 677–693.

Boyd, S. and Vandenberghe, L. (2004). *Convex optimization.* Cambridge university press.

Cocchia, A. (2014). Smart and digital city: A systematic literature review. In *Smart city*, 13–43. Springer.

Fodor, I.K. (2002). A survey of dimension reduction techniques. Technical report, Lawrence Livermore National Lab., CA (US).

Karnouskos, S. (2011). Cyber-physical systems in the smartgrid. In *Industrial Informatics (INDIN), 2011 9th IEEE International Conference on*, 20–23. IEEE.

Khalil, H.K. (1996). Noninear systems. *Prentice-Hall, New Jersey*, 2(5), 5–1.

Li, W. and Todorov, E. (2004). Iterative linear quadratic regulator design for nonlinear biological movement systems. In *ICINCO (1)*, 222–229.

McFadden, D. and Train, K. (2000). Mixed mnl models for discrete response. *Journal of applied Econometrics*, 447–470.

Mukhopadhyay, S.C. (2015). Wearable sensors for human activity monitoring: A review. *IEEE sensors journal*, 15(3), 1321–1330.

Munir, S., Stankovic, J.A., Liang, C.J.M., and Lin, S. (2013). Cyber physical system challenges for human-in-the-loop control. In *Feedback Computing.*

Nocedal, J. and Wright, S.J. (2006). *Sequential quadratic programming.* Springer.

Nocedal, J. and Yuan, Y.x. (1998). Combining trust region and line search techniques. In *Advances in nonlinear programming*, 153–175. Springer.

Pentland, A. and Liu, A. (1999). Modeling and prediction of human behavior. *Neural computation*, 11(1), 229–242.