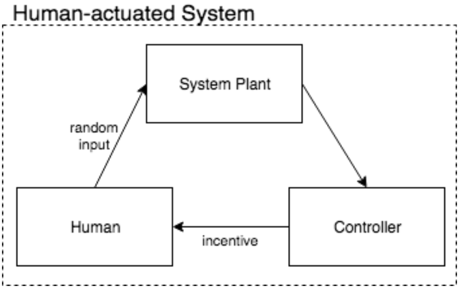
Modeling & Control of Human Acuated Systems

Sangjae Bae, Sang Min Han, Scott Moura

December 19, 2018

Human Actuated Systems (HAS)?



Dynamical systems where
the system inputs are induced by human behaviours.

In such system...

- we cannot directly command human behavior
- still, their behaviours can be "encouraged" with (price) incentives.

Human Actuated Systems (HAS)?

Motivation and Objective

Motivation:

- Large-scale infrastructure has stochastic human decision-makers in the loop (e.g. smart grids, smart cities)
- We can collect data to better understand human behavior
- Behavioral economics and control theory are disparate disciplines

Objective of this study

Construct a control-theoretic framework to optimize performance of human-actuated systems.

Specifically, we are trying to answer:

1. How to model human behaviours with dynamical systems?
2. How to incentivize human actuators to make desired behaviors for a system-wide benefit?

Overview

1. Literature
2. System Model
3. Inventory Control Problem
4. Conclusion
5. Appendix: Reference Tracking Problem
6. Appendix: formulation of SQP problem

Overview

Existing Literature

In the existing controls literature,

- Human behaviors are addressed as *noises/disturbances*, [1-3]
- Human behaviors are *to improve system performance* [4]
- Finding *optimal behaviors* that mimic or improve upon human behavior [5]

In our previous study,

- Added a new perspective: *Desired human behaviors are encouraged by incentives* [6]
- Convex optimization with convexity constraints.
  - Restricted solution domain, allowing just few alternatives.

In this study:

- Generalized model of human actuated system and control schemes, without restricting domain and number of alternatives.

Overview

System Modeling Framework

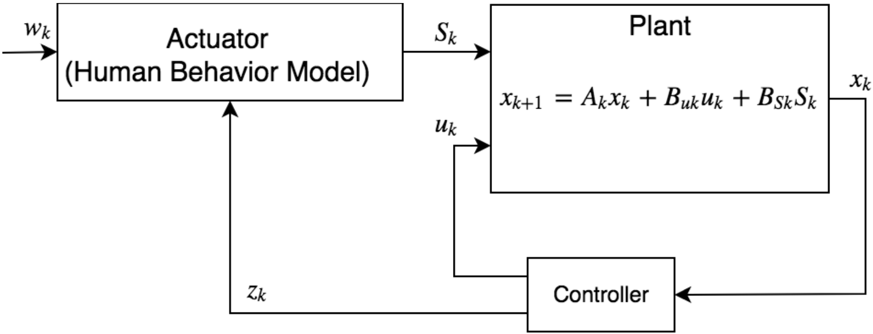
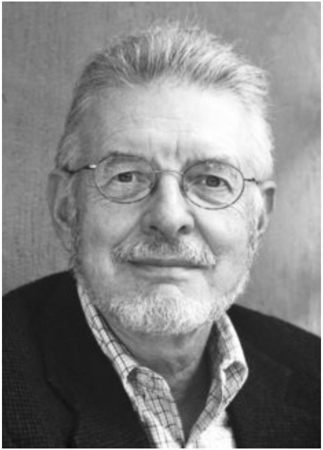


Figure: Block diagram of the Human-actuated system

Discrete Choice Model (DCM)



Professor Daniel Mcfadden

2000 Nobel Prize Winner in Economics

Berkeley, MIT, USC

Focus Area: Discrete Choice Theory

Discrete Choice Model (DCM)

Example of discrete choices:

If these were your only options, which of the following laptops would you choose to purchase?

	Alt1	Alt2	Alt3	
Brand	Microsoft	Apple	Google	
Operating System	Windows 10	OS X	Chrome OS	
Screen Size	13.5”	13”	12”	
Battery Life	12 hours	10 hours	12 hours	
Front Camera	Yes	Yes	No	
Rear Camera	Yes	No	No	
Stylus	Yes	No	No	I would not choose any of these.
Removable Keyboard	Yes	No	No	
Price	$1,499	$1,299	$999	
	○	○	○	○

Attributes

Discrete Choice Model (DCM)

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

where

U_j : Utility of j-th alternative, $j \in \{1, ..., J\}$

β_j: Parameters of controlled attributes

z_j: Controlled attributes

γ_j: Parameters of uncontrolled attributes

w_j: Uncontrolled attributes

β_{0j}: Alternative specific constant

ϵ_j: Undefined errors

Probability of Alternatives in Discrete Choice Model

Logit model

Assuming that the undefined errors, ϵ_j, have *iid* Extreme Value distribution, the probability of choosing j-th alternative is [7]:

$$\Pr(\text{alternative } j \text{ is selected}) = \Pr\left[\bigcap_{j \neq i} (U_j > U_i)\right] = \frac{e^{V_j}}{\sum_{i=1}^{J} e^{V_i}}, \tag{2}$$

where $V_j \doteq \beta_j^\top z_k + \gamma_j^\top w_k + \beta_{0j}$.

Probability of Alternatives in Discrete Choice Model

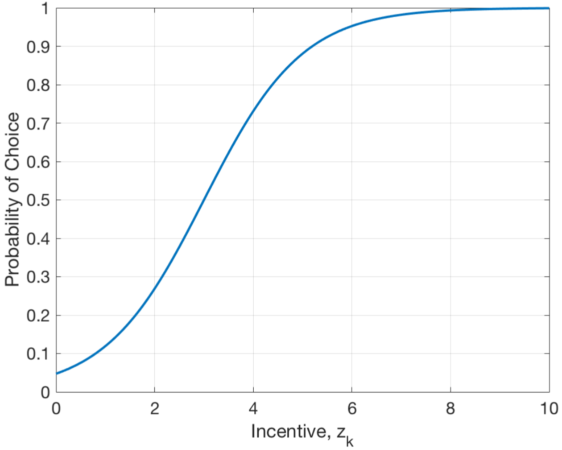


Figure: Binary Logit model example

Mathematical Formulation of Human-actuated System

With a linear system plant,

$$x_{k+1} = A_k x_k + B_{uk} u_k + B_{Sk} S_k, \tag{3}$$

where

- x_k is state
- A_k is the system matrix
- B_{uk} is the direct input matrix
- B_{Sk} is the human input matrix
- u_k is the direct input
- S_k is the random human input to the system.
  - $S_k \in \{0, 1\}^J$
  - $[S_k]_j = 1$ if alternative j is chosen and 0 otherwise
  - $\sum_{j \in \mathcal{J}} [S_k]_j = 1$

Mathematical Formulation of Human-actuated System

Consider mean dynamics,

$$\bar{x}_{k+1} = A_k \bar{x}_k + B_{uk} u_k + B_{Sk} g(z_k, w_k), \tag{4}$$

where

- $\bar{x}_k \doteq \mathbb{E}[x_k]$
- $g(z_k, w_k) \doteq \begin{bmatrix} \mathbb{E}\left[[S_k]_1\right] & \cdots & \mathbb{E}[[S_k]_J] \end{bmatrix}^\top$
- $\mathbb{E}[[S_k]_j] = g_j(z_k, w_k) = \frac{e^{V_{jk}([z_k]_j, [w_k]_j)}}{\sum_{i=1}^{J} e^{V_{ik}([z_k]_i, [w_k]_i)}}, \quad \forall j \in \{1, \ldots, J\}$
  - $\equiv \Pr([S_k]_j = 1)$

Overview

Inventory Control Problem

Objective

Maximizing revenue by giving *price discount*, z_k, and *restocking inventories*, u_k.

Inventory Control Problem

System dynamics,

$$x_{k+1} = x_k + u_k - B_S S_k, \tag{5}$$

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
- $B_S = [0, 1, ..., J-1]^\top$; number of items purchased by a customer

(Mean dynamics is developed as previously mentioned.)

Formulation of Optimization Problem

$$\text{minimize}_{u_k, z_k} \sum_{k=0}^{N-1} \left[c_u u_k - r B_S g(z_k, w_k) + \sum_{j=1}^{J} [c_z]_j [z_k]_j g_j(z_k, w_k) \right] \tag{6}$$

$$\text{subject to: } \bar{x}_{k+1} = \bar{x}_k + u_k - B_S g(z_k, w_k), \tag{7}$$

$$0 \leq \bar{x}_k \leq \bar{x}_{\max} \tag{8}$$

$$0 \leq u_k \leq u_{\max} \tag{9}$$

$$0 \leq z_k \leq z_{\max}, \tag{10}$$

where

- c_u is the cost-per-unit from the supplier
- c_z is the cost-per-incentive from the store manager
- r is the revenue-per-unit from customers

Solving Optimization with Dynamic Programming (DP)

The Bellman equation is

$$J_k(\bar{x}_k) = \min_{u_k, z_k} \left\{ c_u u_k - r B_S g(z_k, w_k) + \sum_{j=1}^{J} [c_z]_j [z_k]_j g_j(z_k, w_k) + J_{k+1}(\bar{x}_{k+1}) \right\}, \tag{11}$$

where $\bar{x}_{k+1} = \bar{x}_k + u_k - B_S g(z_k, w_k)$ for $k \in \{0, \dots, N-1\}$.

Simulation Overview

Consider:

- 30 customers
- Three alternatives:
  - buy nothing
  - buy one
  - buy two
- customers buy less items without discount
- customers are less sensitive to price discount for two items

Simulation Result

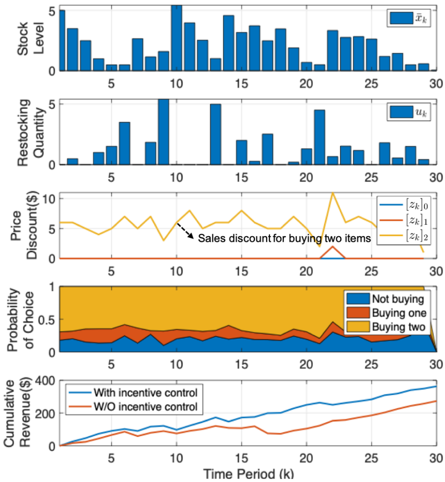


Figure: A simulation result for an inventory control problem.

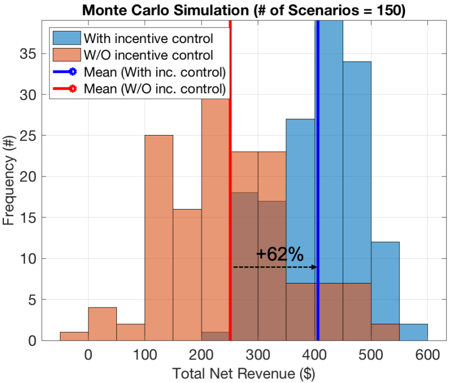


Figure: Monte Carlo simulation results for over 150 randomized scenarios.

Overview

Conclusion

Contributions:

- Proposed mathematical framework to model human-actuated dynamical systems
- Bridged decision-making in behavioral economics with control theory
- Various applications
  - e.g., Energy system operation; demand response program, EV charging station management, etcs.

On-Going work:

- Demand Response incentive control, with Smart-Grid-Smart-City dataset from Australia.
- Controller design with higher order moments, replacing simplified mean dynamics.

Reference

[1] Arnold, Ludwig. Random dynamical systems. Springer Science & Business Media, 2013.
[2] Gray, Robert M., and Lee D. Davisson. An introduction to statistical signal processing. Cambridge University Press, 2004.
[3] Maruyama, Gisiro. "Continuous Markov processes and stochastic equations." Rendiconti del Circolo Matematico di Palermo 4.1 (1955): 48.
[4] Leeper, Adam Eric, et al. "Strategies for human-in-the-loop robotic grasping." Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction. ACM, 2012.
[5] Mnih, Volodymyr, et al. "Human-level control through deep reinforcement learning." Nature 518.7540 (2015): 529.
[6] Bae, Sangjae, Sang Min Han, and Scott Moura. "System Analysis and Optimization of Human-Actuated Dynamical Systems." 2018 Annual American Control Conference (ACC). IEEE, 2018.
[7] Train, Kenneth E. Discrete choice methods with simulation. Cambridge university press, 2009.

Overview

Objective

Minimizing a deviation of the system plant's expected state trajectory $\bar{x}$ from a reference trajectory x^{ref}.

with

$$f(\bar{x}_k, u_k, z_k, w_k) = A\bar{x}_k + B_u u_k + B_S g(z_k, w_k), \quad k \in \{0, \dots, N-1\}. \tag{12}$$

The optimization problem

$$\min_{\bar{x}_k, u_k, z_k} \quad F = \sum_{k=0}^{N-1} \Big[(\bar{x}_k - x_k^{\mathrm{ref}})^\top Q_k (\bar{x}_k - x_k^{\mathrm{ref}}) + (u_k - u_k^{\mathrm{ref}})^\top R_{uk} (u_k - u_k^{\mathrm{ref}}) + (z_k - z_k^{\mathrm{ref}})^\top R_{zk} (z_k - z_k^{\mathrm{ref}}) \Big] + (\bar{x}_N - x_N^{\mathrm{ref}})^\top Q_N (\bar{x}_N - x_N^{\mathrm{ref}}) \tag{13}$$

subject to: (14)

$$\bar{x}_0 = \bar{x}_{\mathrm{init}}, \tag{15}$$

$$\bar{x}_{k+1} = f(\bar{x}_k, u_k, z_k, w_k), \tag{16}$$

$$z_k \geq 0, \tag{17}$$

Solving with Sequential Quadratic Program (SQP)

Sequential Quadratic Programming (SQP) iteratively finds sub-optimal control polices by solving quadratic approximations to the original problem at every iteration.

(Details of reformulating it to a SQP problem is in Appendix.)

Simulation Overview

Consider a third order system

$$A = \begin{bmatrix} 0.1 & 1 & -1 \\ 1 & 0.1 & 1 \\ 1 & 0 & 0.5 \end{bmatrix}, \tag{18}$$

- the open-loop system is *unstable*, yet *controllable*
- the human's choice $\in \{-5, 0, +5\}$
- the same probability over alternatives without incentive
- the human is equally sensitive to all incentives for choosing each alternative.

Simulation Result

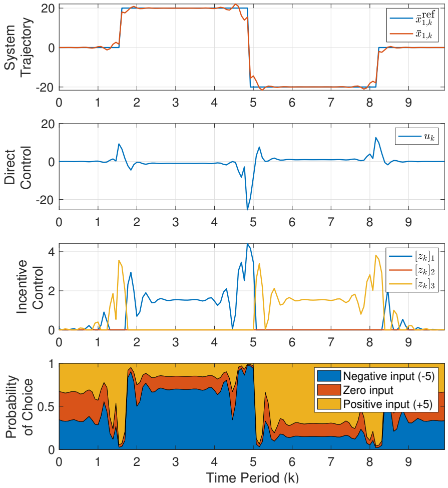


Figure: Simulated reference tracking results solved via SQP.

Findings:

(i) For $k \in [2, 4]$ and $k \in [5.5, 7.5]$, $u_k \approx 0$, yet $([z_k]_1 \neq 0, [z_k]_3 \neq 0)$

$\rightarrow$ *Optimal strategy: is to leverage human actuation z_k instead of direct control u_k.*

(ii) the control effort cost with incentive control (349.1) is lower than that without incentive control (472.13).

$\rightarrow$ *Strong potential for cost savings in practical CPHS control problems*

Linearization

SQP essentially replaces the nonlinear constraint (16) with a linear approximation. Using Taylor's theorem, we can approximate the system as[1]

$$\tilde{x}_{k+1} = \tilde{A}_k \tilde{x}_k + \tilde{B}_{uk} \tilde{u}_k + \tilde{B}_{zk} \tilde{z}_k, \tag{19}$$

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

$$\tilde{A}_k \doteq \nabla_{\bar{x}} f_k \left(\bar{x}_k^{\mathrm{ref}}, u_k^{\mathrm{ref}}, z_k^{\mathrm{ref}}, w_k \right)^\top \tag{20}$$

$$\tilde{B}_{uk} \doteq \nabla_{u} f_k \left(\bar{x}_k^{\mathrm{ref}}, u_k^{\mathrm{ref}}, z_k^{\mathrm{ref}}, w_k \right)^\top \tag{21}$$

$$\tilde{B}_{zk} \doteq \nabla_{z} f_k \left(\bar{x}_k^{\mathrm{ref}}, u_k^{\mathrm{ref}}, z_k^{\mathrm{ref}}, w_k \right)^\top . \tag{22}$$

[1] 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.

Linearization (Cont'd)

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 $m = 1, \ldots, n$. Then,

$$\tilde{B}_{zk,m}^\top = \begin{bmatrix} \frac{\beta_{k1} \exp\{V_{k1}([z_k^{\mathrm{ref}}]_1, [w_k]_1)\}}{\left(\sum_{i=1}^J \exp\{V_{ki}([z_k^{\mathrm{ref}}]_i, [w_k]_i)\}\right)^2} \psi_{k,m}(z_k^{\mathrm{ref}}, w_k, 1) \\ \vdots \\ \frac{\beta_{jk} \exp\{V_{kJ}([z_k^{\mathrm{ref}}]_J, [w_k]_J)\}}{\left(\sum_{i=1}^J \exp\{V_{ki}([z_k^{\mathrm{ref}}]_i, [w_k]_i)\}\right)^2} \psi_{k,m}(z_k^{\mathrm{ref}}, w_k, J) \end{bmatrix}^\top \tag{23}$$

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

$$\begin{aligned} &\psi_{k,m}(z_k^{\mathrm{ref}}, w_k, j) \\ &\doteq \sum_{i=1}^{J} \left([B_{zk,m}^\top]_j - [B_{zk,m}^\top]_i \right) \exp\{V_{ik}([z_k^{\mathrm{ref}}]_i, [w_k]_i)\}. \end{aligned} \tag{24}$$

Reformulation with stacked variables

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}

$$\text{minimize}_v \quad (v - v^{\mathrm{ref}})^\top H (v - v^{\mathrm{ref}}) \tag{25}$$

subject to:

$$\begin{bmatrix} \bar{x}_{\mathrm{init}} - \{v\}_{\bar{x}_0} \\ A\{v\}_{\bar{x}_0} + B_u\{v\}_{u_0} + B_S\{v\}_{z_0} - \{v\}_{\bar{x}_1} \\ \vdots \\ A\{v\}_{\bar{x}_{N-1}} + B_u\{v\}_{u_{N-1}} + B_S\{v\}_{z_{N-1}} - \{v\}_{\bar{x}_N} \end{bmatrix} = 0, \tag{26}$$

$$[\{v\}_{z_0}, \dots, \{v\}_{z_{N-1}}] \geq 0, \tag{27}$$

Reformulation with stacked variables (Cont'd)

where

$$v = [\bar{x}_0, u_0, z_0, ..., \bar{x}_{N-1}, u_{N-1}, z_{N-1}, \bar{x}_N], \tag{28}$$

$$v^{\mathrm{ref}} = [\bar{x}_0^{\mathrm{ref}}, u_0^{\mathrm{ref}}, z_0^{\mathrm{ref}}, ..., \bar{x}_{N-1}^{\mathrm{ref}}, u_{N-1}^{\mathrm{ref}}, z_{N-1}^{\mathrm{ref}}, \bar{x}_N^{\mathrm{ref}}], \tag{29}$$

$$H = \mathrm{diag}\{Q_0, R_{u0}, R_{z0}, ..., Q_{N-1}, R_{u(N-1)}, R_{z(N-1)}, Q_N\}, \tag{30}$$

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 (25)-(27) with respect to an optimal solution $\hat{v}^{(i)}$ obtained at iteration i. We also linearize the equality constraints (26) with respect to $\hat{v}^{(i)}$

Formulation of SQP optimization problem

We eventually formulate the SQP optimization problem

$$\text{minimize}_p \;\; F^{(i)} + \nabla F^{(i)\top} p + \frac{1}{2} p^\top \nabla^2_{vv} \mathcal{L}^{(i)} p \tag{31}$$

subject to:

$$(\mathrm{diag}\{\begin{bmatrix} A & B_u & \tilde{B}_{zk} \end{bmatrix}\}_{k=0}^{N-1})^\top p + \begin{bmatrix} \bar{x}_{\text{init}} - \{\hat{v}^{(i)}\}_{\bar{x}_0} \\ A\{\hat{v}^{(i)}\}_{\bar{x}_0} + B_u\{\hat{v}^{(i)}\}_{u_0} + B_S\{\hat{v}^{(i)}\}_{z_0} - \{\hat{v}^{(i)}\}_{\bar{x}_1} \\ \vdots \\ \Big(A\{\hat{v}^{(i)}\}_{\bar{x}_{N-1}} + B_u\{\hat{v}^{(i)}\}_{u_{N-1}} \\ + B_S\{\hat{v}^{(i)}\}_{z_{N-1}} - \{\hat{v}^{(i)}\}_{\bar{x}_N}\Big) \end{bmatrix} = 0, \tag{32}$$

$$\nabla c(\hat{v}^{(i)})^\top p + c(\hat{v}^{(i)}) \geq 0, \tag{33}$$

where p is the SQP decision variable, and $\tilde{B}_k$ is defined in (22). Gradient $\nabla c(\hat{v}^{(i)})$ has a value of 1 in the place of elements that correspond to the incentive control z_k.

Update algorithm and convergence

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

$$\hat{v}^{(i+1)} = \hat{v}^{(i)} + \alpha^{(i)} p^*. \tag{34}$$

Then, the algorithm iterates until a set of convergence criteria are satisfied, e.g. $p^* < \epsilon$ for some $\epsilon > 0$.

SQP inherits the convergence rate of the quasi-Newton method, and convergence can be guaranteed by an appropriate choice of $\alpha^{(i)}$ assuming appropriate regularity conditions hold and unlikely edge cases (e.g. infeasibility of the problem) do not arise.