Entropy for Optimal Control on a Simplex with an Application to Behavioral Nudging

Ahdab, Mohamad Al; Knudsen, Torben; Stoustrup, Jakob; Leth, John-Josef

Published in:
IEEE Control Systems Letters

Publication date:
2023

Document Version
Early version, also known as pre-print

Entropy for Optimal Control on a Simplex with an Application to Behavioral Nudging

Mohamad Al Ahdab, Torben Knudsen, Jakob Stoustrup, and John Leth

***Abstract*— We study the utilization of the entropy function of inputs in solving an Optimal Control Problem (OCP) with linear dynamics and inputs constrained to a variable-sized simplex in which the size is also an input. By using the entropy function as part of the objective functional in the OCP, we are able to derive a closed-form solution. Additionally, we present an example of how the studied OCP can be applied to choose between nudging techniques to discourage a specific behavior, such as non-adherence to medication, through the lens of behavioral momentum theory.**

***Index Terms*— Emerging control applications, Optimal control.**

I. Introduction

Optimal control problems (OCPs) are concerned with finding an optimal input trajectory for a dynamical system which maximizes or minimizes an objective functional while satisfying specific constraints. A class of problems in which the input trajectory is constrained to be on a simplex arises in many applications such as portfolio optimization in finance, resource allocation in energy systems, mixing chemicals in chemical reactions, and when the control input is a discrete probability distribution. The use of entropy in the objective for continuous-time dynamics has been studied in works of [1] and [2]. In [1], the authors analyzed the use of the entropy function for stochastic linear optimal control problems. As for the work in [2], the authors derived a class of Hamilton–Jacobi–Bellman (HJB) equations for optimal control problem in which the input is a probability measure. The optimization in the mentioned papers is performed over a probability measure with the dynamics having inputs drawn from the probability measure, and the optimal control problem considers averaged dynamics with respect to the probability measure in addition to averaged objective terms with respect to the probability measure. In this paper, we consider an OCP with linear time-varying dynamics and an input vector $\boldsymbol{u}$ constrained to a simplex of size $v > 0$ with v being an input itself. In particular, we show how the use of the entropy function in the objective in addition to a linear objective in state, a linear objective in $\boldsymbol{u}$, and a quadratic objective in v will yield a closed-form solution using the necessary conditions of the maximum principle with Arrow type sufficient conditions [3]. Although setting $v = 1$ in our setup will make our problem a special case of those considered in [1], [2], these works do not explicitly address and solve the case of a discrete probability measure with continuous linear time-varying dynamics and a linear objective function, as we do in this paper. Moreover, we introduce the size of the simplex as an additional optimization input, which further expands the scope of the problem. Furthermore, we present an example on how the OCP of interest in this paper can be used to schedule different nudging techniques using behavioral momentum theory [4] to discourage an unhealthy behavior in people. Behavioral momentum theory suggests that behaviors that are reinforced more frequently are more resistant to changes in the environment. With this theory, we model the dynamics of the average rate of a target behavior by a linear model, with the average rate of a reinforcement being a parameter.

To discourage an unhealthy behavior, we introduce different nudges to the behavior as inputs to the OCP framework. In the context of our framework, we represent the probabilities of selecting the nudges as inputs belonging to a simplex, with the size of the simplex representing their overall rate. Our objective is to optimize the choice of different nudges and their average rate to minimize the average rate of the targeted behavior, while also considering the cost of each nudge and ensuring a diversity of nudges.

The use of computational and machine learning techniques have recently been investigated for the design of nudges for medical care professionals such as in [5] and to encourage patients to adhere to their prescribed medicine in [6]. Additionally, the work in [7] considers the optimal design of nudges within a Markov decision process framework derived from resource-rational analysis. In this paper, we consider the problem of choosing between nudges while minimizing their average rate within a continuous optimal control framework derived from behavioural momentum theory. Our work offers an alternative framework and perspective for the problem of behavioral nudging in healthcare. We hope that our discussion in this paper can be one of the early works towards the application of control theory concepts in behavioral nudging of people in healthcare.

The summary of the contributions of this paper is as follows

- We derive closed-form solution for an OCP with inputs constrained to a simplex in which the size of the simplex itself is also another input.
- We present examples of how the OCP of interest and behav-

Submitted on March 15, 2023. This work was funded by the IFD Grand Solution project ADAPT-T2D, project number 9068-00056B.

The authors are with Section of Automation and Control, Department of Electronic Systems, Aalborg University, Aalborg Øst, Denmark {maah,tk,jakob,jjl}@es.aau.dk

ioral momentum theory can be used to assist in the choice of different nudging techniques aimed at discouraging an undesired behaviour, such as non-adhering to medication. To our knowledge, this is the first time control theory techniques have been used in connection with behavioral momentum theory.

II. Notations

All vectors are considered as column vectors. We let $[a, b]$ denote the closed interval from a to b, and $[a\ b]$ denote the row vector with coordinates a and b. The symbols $\boldsymbol{I}_n$ and $\boldsymbol{0}_{n\times m}$ are used to denote the $n \times n$ identity and the $n \times m$ zero matrix, respectively. The symbol $\boldsymbol{1}_n$ is used to denote the n-dimensional column vector of 1s. The symbols $\geq_e, >_e$ are used for element-wise $\geq$ and $>$. For $\boldsymbol{u} \in \triangle_n^v := \left\{\boldsymbol{u} \in \mathbb{R}^n_{\geq 0} \mid \|\boldsymbol{u}\|_1 = v\right\}$, we write the entropy function as $\phi(\boldsymbol{u}) = -\sum_{i=1}^n u_i \ln(u_i)$ and we take $0\ln(0) := 0$. For $\boldsymbol{u} \in \triangle_n^v$ and $\boldsymbol{w} >_e 0$, we write the Kullback–Leibler (KL) divergence (relative entropy) as $D_{KL}(\boldsymbol{u} \mid\mid \boldsymbol{w}) = \sum_{i=1}^n u^i \ln\left(u^i/w^i\right)$. We use $\exp_e(\boldsymbol{x})$ and $\ln_e(\boldsymbol{x})$ for the element-wise exponential and logarithm of a vector $\boldsymbol{x}$, respectively.

III. Solution of the Optimal Control Problem

In this section, we first present the OCP of interest in III-A, and derive an explicit solution for it in III-B.

A. Problem Setup

To define the OCP of interest (OCPv) in this work, we begin by defining

$$L(\boldsymbol{x}, \boldsymbol{u}, v, t) := \frac{1}{\eta}\phi(\boldsymbol{u}) + \boldsymbol{c}^{\mathrm{T}}(t)\boldsymbol{u} + \boldsymbol{d}^{\mathrm{T}}\boldsymbol{x} + qv^2,$$

and $\boldsymbol{S}(\boldsymbol{x}) := \boldsymbol{e}^{\mathrm{T}}\boldsymbol{x}$, with $\eta > 0$, $\boldsymbol{c}(t) \in \mathbb{R}^{n_u}$ being continuously differentiable, $\boldsymbol{d} \in \mathbb{R}^{n_x}$, and $\boldsymbol{e} \in \mathbb{R}^{n_x}$. The OCP in this paper has the following form

$$\max_{\boldsymbol{u}, v} \int_{t_0}^{t_f} L\left(\boldsymbol{x}(t), \boldsymbol{u}(t), v(t), t\right)dt + S\left(\boldsymbol{x}(t_f)\right) \tag{1a}$$

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{B}(t)\boldsymbol{u}(t),\ \boldsymbol{x}(t_0) = \boldsymbol{x}_0, \tag{1b}$$

$$v(t) - \boldsymbol{1}^{\mathrm{T}}\boldsymbol{u}(t) = 0,\ \boldsymbol{u}(t) \geq_e 0,\ v(t) \geq 0. \tag{1c}$$

Note that for the case in which v is set to 1, the inputs $\boldsymbol{u}$ will be constrained to the unit simplex $\triangle_{n_u}^1$. Also note that $B(t)$ is assumed to be an explicit function of time t e.g., see below (11).

B. Closed-Form Solution

In order to find the solution for OCPv, we use the necessary conditions of the maximum principle. To summarize the necessary conditions, it is convenient to define the Hamiltonian function for our problem

$$H(t, \boldsymbol{x}, \boldsymbol{u}, v, \boldsymbol{\lambda}) = \tilde{L}(\boldsymbol{x}, \boldsymbol{u}, v, t) + \boldsymbol{\lambda}^{\mathrm{T}}\left(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}(t)\boldsymbol{u}\right), \tag{2}$$

for all $(t, \boldsymbol{x}, \boldsymbol{u}, v, \boldsymbol{\lambda}) \in \left[t_0, t_f\right] \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R} \times \mathbb{R}^{n_x}$, where $\boldsymbol{\lambda}$ is called the adjoint variable[1]. The necessary conditions for the tuple $\left(\boldsymbol{x}^*(t), \boldsymbol{u}^*(t), v^*(t)\right)$, $t \in \left[t_0, t_f\right]$, to be a solution of the OCP in (1) are summarized as follows

$$\left(\boldsymbol{u}^*(t), v^*(t)\right) \in \underset{\boldsymbol{u}\in\triangle_{n_u}^v, v\geq 0}{\operatorname{argmax}} H\left(\boldsymbol{x}(t), \boldsymbol{u}, v, \boldsymbol{\lambda}(t), t\right), \tag{3a}$$

$$\dot{\boldsymbol{\lambda}}(t) = -\boldsymbol{H}_{\boldsymbol{x}}\left(t, \boldsymbol{x}^*(t), \boldsymbol{u}^*(t), v^*(t), \boldsymbol{\lambda}(t)\right), \tag{3b}$$

$$\boldsymbol{\lambda}(t_f) = -\boldsymbol{S}_x(t_f) \tag{3c}$$

From the maximality condition (3a), we get (see Appendix I)

$$\left(\boldsymbol{u}^*(t), v^*(t)\right) = \underset{\boldsymbol{u}\in\triangle_{n_u}^v, v\geq 0}{\operatorname{argmax}} H\left(\boldsymbol{x}(t), \boldsymbol{u}, v, \boldsymbol{\lambda}(t), t\right), \tag{4a}$$

$$\boldsymbol{u}^*(t) = v^*(t)\frac{\exp_{\mathrm{e}}\left(\eta\boldsymbol{B}^{\mathrm{T}}(t)\boldsymbol{\lambda}(t) + \eta\boldsymbol{c}(t)\right)}{\boldsymbol{1}^{\mathrm{T}}\exp_{\mathrm{e}}\left(\eta\boldsymbol{B}^{\mathrm{T}}(t)\boldsymbol{\lambda}(t) + \eta\boldsymbol{c}(t)\right)} >_e \boldsymbol{0}, \tag{4b}$$

$$v^*(t) = \frac{-\mathrm{W}_0\left(-2q\eta\exp(-1)\,\boldsymbol{1}^{\mathrm{T}}\exp_{\mathrm{e}}\left(\eta\boldsymbol{B}^{\mathrm{T}}(t)\boldsymbol{\lambda}(t) + \eta\boldsymbol{c}(t)\right)\right)}{2q\eta}, \tag{4c}$$

where W_0 is the principal branch of the Lambert W function. Letting $\left(\boldsymbol{u}^0, v^0\right) := \operatorname{argmax}_{\boldsymbol{u}\in\triangle_{n_u}^v, v\geq 0} H\left(\boldsymbol{x}, \boldsymbol{u}, v, \boldsymbol{\lambda}, t\right)$, we get that $(\boldsymbol{u}^0, v^0)$ is $(\boldsymbol{u}^*, v^*)$ for a given $\boldsymbol{\lambda} \in \mathbb{R}^{n_x}$, which is a unique solution due to the strict concavity of H in $(\boldsymbol{u}, v)$. Substituting $(\boldsymbol{u}^0, v^0)$ in the Hamiltonian we get that $H(\boldsymbol{x}, \boldsymbol{u}^0(\boldsymbol{x}, \boldsymbol{\lambda}, t), v^0(\boldsymbol{x}, \boldsymbol{\lambda}, t), \boldsymbol{\lambda}, t)$ is an affine function in $\boldsymbol{x}$ which is concave. Additionally, since $\boldsymbol{S}(\boldsymbol{x})$ is also concave in $\boldsymbol{x}$, the necessary conditions (3) are sufficient (Arrow type sufficient conditions [3]). Now using the adjoint equation (3b) together with the transversality condition (3c), we get

$$\dot{\boldsymbol{\lambda}}(t) = -\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda}(t) - \boldsymbol{d},\ \boldsymbol{\lambda}\left(t_f\right) = \boldsymbol{e}. \tag{5}$$

Referring to [8], we can obtain the solution to (5) as[2]:

$$\boldsymbol{\lambda}(t) = \boldsymbol{M}_A(t)\boldsymbol{e} + \boldsymbol{M}_d(t), \tag{6}$$

where

$$\begin{bmatrix}\boldsymbol{M}_A(t) & \boldsymbol{M}_d(t)\\ \boldsymbol{0}_{1\times n_x} & 1\end{bmatrix} = \mathrm{e}^{\boldsymbol{M}(t-t_f)}, \boldsymbol{M} = \begin{bmatrix}-\boldsymbol{A}^{\mathrm{T}} & -\boldsymbol{d}\\ \boldsymbol{0}_{1\times n_x} & 0\end{bmatrix}.$$

Substituting the adjoint solution (6) in (4), we get the solution

$$\boldsymbol{u}^*(t) = v^*(t)\frac{\exp_{\mathrm{e}}\left(\zeta(t)\right)}{\boldsymbol{1}^{\mathrm{T}}\exp_{\mathrm{e}}\left(\zeta(t)\right)} >_e \boldsymbol{0}, \tag{7a}$$

$$v^*(t) = \frac{-\mathrm{W}_0\left(-2q\exp(-1)\eta\boldsymbol{1}^{\mathrm{T}}\exp_{\mathrm{e}}\left(\zeta(t)\right)\right)}{2q\eta}, \tag{7b}$$

$$\zeta(t) := \eta\boldsymbol{B}^{\mathrm{T}}(t)\left(\boldsymbol{M}_A(t)\boldsymbol{e} + \boldsymbol{M}_d(t)\right) + \eta\boldsymbol{c}(t). \tag{7c}$$

[1]The adjoint variable λ_0 in $H(t, \boldsymbol{x}, \boldsymbol{u}, v, \boldsymbol{\lambda}, \lambda_0) = \lambda_0 L(\boldsymbol{x}, \boldsymbol{u}, v, t) + \boldsymbol{\lambda}^{\mathrm{T}}\left(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}(t)\boldsymbol{u}\right)$ is set to $\lambda_0 = 1$ since the end-point $\boldsymbol{x}(t_f)$ is free [3].

[2]Note that for a non-singular $\boldsymbol{A}$, we can write $\boldsymbol{M}_A(t) = \exp\left(\boldsymbol{A}\left(t - t_f\right)\right)$, and $\boldsymbol{M}_d(t) = \boldsymbol{A}^{-1}\left(\boldsymbol{M}_A(t) - \boldsymbol{I}\right)\boldsymbol{d}$.

Remark 3.1: For the case when v is set to 1 and it is not optimized over, the solution $\boldsymbol{u}^*$ can then be shown to be $\boldsymbol{u}^* = \frac{\exp_e(\zeta(t))}{\mathbf{1}^{\mathrm{T}}\exp_e(\zeta(t))}$ by following the same procedure to obtain (7a). Moreover, if $\boldsymbol{d} = \boldsymbol{e} = \mathbf{0}$ and $\boldsymbol{c}(t) = \boldsymbol{c}$, then the solution (7) is a constant input $\boldsymbol{u}^* = \exp_{\mathrm{e}}(\eta\boldsymbol{c})/\mathbf{1}^{\mathrm{T}}\exp_{\mathrm{e}}(\eta\boldsymbol{c})$. Additionally, if the inputs are weighted equally (i.e., $\boldsymbol{c} = c\mathbf{1}$ for some scalar $c \in \mathbb{R}$), then the solution simplifies to $\boldsymbol{u}^* = \frac{1}{n_u}\mathbf{1}$. This is the well-known solution for the maximum entropy on a simplex.

Remark 3.2: Incorporating the entropy function into the objective of OCPv encourages the utilization of all available inputs, as the resulting solution, as shown in equation (7), is always non-zero. This encourages diversification in the inputs, which can be advantageous in certain applications where exploring diverse solutions is desirable or in situations where one or more inputs could potentially lose their effectiveness, such as in the case of a faulty actuator. Using only a linear term for the inputs in the objective of OCPv will yield a bang-bang solution in step (4). Additionally, if we use a quadratic term $-\boldsymbol{u}^{\mathrm{T}}(t)\boldsymbol{Q}\boldsymbol{u}(t), \boldsymbol{Q} \geq 0$ for the inputs in place of the entropy function, then the problem in (4) becomes a standard quadratic optimization problem (StQP). However, determining explicit solutions for StQPs is known to be NP-hard [9], even though efficient algorithms are available. In contrast, despite the need to evaluate a matrix exponential for $\boldsymbol{M}_A$ and $\boldsymbol{M}_d$, computing the explicit solution in (7) can be more efficient to implement in many scenarios (e.g., when $\boldsymbol{A}$ is diagonal). Moreover, if we intend to implement (7) recursively, as demonstrated in IV-D, the matrix exponential need only to be evaluated once. Finally, obtaining an explicit solution may prove valuable for conducting further theoretical analyses of the implemented solution's dynamics.

Remark 3.3: The solution to OCPv can be used in a receding horizon fashion by recursively estimating the dynamics parameters and solving the OCP for a fixed horizon (see IV-D for an example). To ensure that the inputs between the solutions are close to each other, we introduce a relative entropy objective $-\eta_p^{-1}\exp\left(-\rho\tilde{t}\right)D_{KL}\left(\boldsymbol{u}(t)\,||\,\boldsymbol{u}_p\right)$ with $\rho > 0$, $\eta^{-1} < \eta_p^{-1}$, and $q_p\exp\left(-\rho\tilde{t}\right)\left(v(t)-v_p\right)^2$ where $\tilde{t} := t - t_0$, $|q_p| < |q|$, with $q_p \leq 0$. The values $\boldsymbol{u}_p, v_p$ are the last inputs from the previously computed solution. In this case, the solution in (4) becomes

$$\boldsymbol{u}^*(t) = v^*(t)\frac{\exp_{\mathrm{e}}\left(\boldsymbol{\gamma}(t)\right)}{\mathbf{1}^{\mathrm{T}}\exp_{\mathrm{e}}\left(\boldsymbol{\gamma}(t)\right)}, \tag{8a}$$

$$v^*(t) = \frac{-1}{2\bar{q}(t)\bar{\eta}(t)}\mathrm{W}_0\Big(-2\bar{q}(t)\bar{\eta}(t)\mathbf{1}^{\mathrm{T}}\exp_{\mathrm{e}}\left(\boldsymbol{\gamma}(t)\right) \exp\left(-2\bar{\eta}q_p\exp\left(-\rho\tilde{t}\right)-1\right)\Big), \tag{8b}$$

$$\boldsymbol{\gamma}(t) := \bar{\eta}(t)\boldsymbol{B}^{\mathrm{T}}(t)\boldsymbol{\lambda}(t) + \bar{\eta}(t)\boldsymbol{c}(t) + \bar{\eta}(t)\eta_p^{-1}\exp\left(-\rho\tilde{t}\right)\ln_e\left(\boldsymbol{u}_p\right), \tag{8c}$$

where $\bar{\eta}(t) = \frac{1}{\eta^{-1}+\eta_p^{-1}\exp\left(-\rho\tilde{t}\right)}$, and $\bar{q}(t) = q + q_p\exp\left(-\rho\tilde{t}\right)$.

IV. Example with Behavioral Momentum Theory

In this section, we will present a simple model derived from the principles of behavioral momentum theory. The model takes the form of $\dot{x}(t) = \boldsymbol{B}(t)\boldsymbol{u}(t)$, with $\mathbf{1}^{\mathrm{T}}\boldsymbol{u}(t) = v(t)$ and will be described in detail in IV-A. We will then proceed to use this model to solve OCPv for various scenarios in IV-B, IV-C, and IV-D.

A. Behavioral Momentum Model

Behavioral momentum theory provides a quantitative basis for the idea that the rate of a behavior, which has been reinforced frequently in the past is more resistant to change with disruptions than if it has been reinforced less frequently [4], [10]. In the works of [4], [10], mathematical representations for behavioral momentum theory were introduced and validated with data obtained from different experiments. In this paper, we use a simple continuous-time version based on an averaged model from [4], [10]. Let $\beta(t) \in \mathbb{R}_{\geq 0}$ be the average rate of occurrence for a specific behavior per unit time and define $x(t) := \log_{10}\left(\beta(t)\right)$, then the change $x(t) - x(t_1)$ with $\Delta t := t - t_1 \geq 0$ is modelled with respect to disruptions and reinforcers as $x(t) - x(t_1) = \frac{-\delta(t)}{\sqrt{r(t)}}\Delta t$, where $r(t) \in \mathbb{R}_{\geq 0}$ is the average rate of a reinforcer, and $\delta(t) = b(t)v(t)$ with $v(t) \in \mathbb{R}_{\geq 0}$ being the average rate of disruption events, and $b(t) \in \mathbb{R}_{\geq 0}$ being an effect factor for the disruption events. The value $\sqrt{r}$ represents a "behavioral inertia", a higher average reinforcer rate would require a higher average rate for the effect of disruptions δ to change the behaviour. Dividing by Δt and taking the limit for $\Delta t \to 0$, we obtain

$$\dot{x}(t) = \frac{-1}{\sqrt{r(t)}}\delta(t). \tag{9}$$

Consider now that for a disruption happening with an average rate of $v(t)$, the disruption can be of n_u different types with a probability $\bar{u}_i$ of being of type i with an effect factor b_i. In that case, $\delta(t)$ in (9) becomes

$$\delta(t) = v(t)\boldsymbol{b}^{\mathrm{T}}(t)\bar{\boldsymbol{u}}(t), \tag{10}$$

with $\bar{\boldsymbol{u}}(t) \in \triangle^1_{n_u}$ and $\boldsymbol{b}(t) \in \mathbb{R}^{n_u}_{\geq 0}$. Here, the i_{th} component $\bar{u}_i(t) \geq 0$ of $\bar{\boldsymbol{u}}(t)$ can also be understood as the average rate of a type of disruption with respect to the other types in $\delta(t)$ (average rate ratio). Note that the sum $\sum_{i=1}^{n_u} v(t)\bar{u}_i(t) = v(t)$. The value $v(t)$ is usually desired to be small enough to avoid what is known as alert fatigue [11]. Alert fatigue is when the rate of disruptions is high enough that the disruptions will lose their effect. Note that the model (9) with (10) can be written in the form of the model of OCPv (1b) by introducing $v(t)$ in the constraint (1c):

$$\dot{x} = \frac{-1}{\sqrt{r(t)}}\boldsymbol{b}^{\mathrm{T}}(t)\boldsymbol{u}(t), \mathbf{1}^{\mathrm{T}}\boldsymbol{u}(t) = v(t), \tag{11}$$

where $\boldsymbol{B}(t) = \frac{-1}{\sqrt{r(t)}}\boldsymbol{b}^{\mathrm{T}}(t)$.

Remark 4.1: For a better understanding of the averaged representation and how to obtain (10), consider the Poisson Compound Process $\Pi(t)$ defined as

$$\Pi(t) = \sum_{k=1}^{P(t)} \frac{-1}{\sqrt{r(T_k^-)}}\boldsymbol{b}^T(T_k^-)\boldsymbol{W}_k, \tag{12}$$

where $P(t)$ is a Poisson process representing the number of disruptions (jumps) up until time t with rate $v(t)$, T_k^- is the pre-disruption time value of the k_{th} random disruption type, $\boldsymbol{W} = \boldsymbol{W}_k$ is an IID stochastic process where $\boldsymbol{W}_k$ represents the type of the k_{th} disruption such that $\boldsymbol{W}_k \in \mathcal{W} = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_{n_u}\}$ with $\boldsymbol{w}_i$ being a vector of zeros except for the i_{th} element being 1 and $\mathbb{P}(\boldsymbol{W}_k = \boldsymbol{w}_i) = \bar{u}_i$. Taking the expectation of $\Pi(t)$ (Chapter 5 in [12]) will give us

$$\mathbb{E}\left[\Pi(t)\right] = \int_0^t v(s) \frac{-1}{\sqrt{r(s)}} \boldsymbol{b}^T(s) \bar{\boldsymbol{u}}(s) ds, \tag{13}$$

which is equivalent to $x(t)$ in (9) with $\delta(t)$ being chosen as in (10). This interpretation also gives us a method to apply the disruptions in real life by simulating (12).

In this paper, we will consider nudges as intentional disruptions that can change the reinforcement contingencies associated with a behavior and we will examine three different examples. The first one in section IV-B deals with a case when $v(t)$ is fixed to be 1 (see Remark 3.1) and $\boldsymbol{b}$ is constant. The second case in section IV-C is when $v(t)$ is optimized over, and the third case in section IV-D is when $\boldsymbol{b}(t)$ is time-varying compared with a receding horizon setting. It is important to note that the examples discussed are simplified abstractions. The intention of presenting the examples is to show how the solution of OCPv in this paper can potentially be used for behavior nudging with elements from behavioral momentum theory. In all of the figures, we will report $\beta(t) = 10^{x(t)}$ and $\bar{\boldsymbol{u}}(t) = \frac{1}{v(t)}\boldsymbol{u}(t)$. The code for generating the results can be found on `https://gitlab.com/aau-adapt-t2d/nudging_entropyocp`.

B. Case with a Constant Rate of Nudges

Consider a case in which a diabetic subject is not following their prescribed medication regimen, such as failing to administer the correct dose of insulin or taking a lower or higher dose than what was prescribed due to some constant average rate of a reinforcer $r = 7$ [1/Week]. Here the reinforcer could be inconveniences of administering the dose and/or economical burden. Assume that we have three different types of disruptions: $\bar{u}_1$ being the probability of sending dose reminder text messages to the subject with an effect of $b_1 = 0.2$, $\bar{u}_2$ being the probability of sending personalized encouraging text messages to the subjects (e.g., reminding them about the importance of their health to their family) with an effect of $b_2 = 0.3$, and $\bar{u}_3$ being the probability of a phone call from a medical staff reminding them about the importance of their health with an effect of $b_3 = 0.4$. Our case study assumes that having a call from a medical staff is the most effective method while sending unpersonlized reminders is the least effective. Additionally, consider that we desire to fix the rate of nudges to a constant $v = 1$ [1/Week]. Phone calls from medical staff can be costly and labor intensive. To account for this, we define a linear cost for the different options $\boldsymbol{c} = -[0.1\ 0.5\ 1]^{\mathrm{T}}$ giving a higher cost for $\bar{u}_3$ and a lower cost for $\bar{u}_1$. A higher cost for $\bar{u}_2$ than the cost for $\bar{u}_1$ is used since the second type of nudges requires obtaining personal information regarding the subjects and formulating specific text messages for them. This cannot be easily automated when compared to just sending dose reminders with $\bar{u}_1$. Additionally, we choose $d = e = -2$ to lower $x(t)$ within a time horizon $t_f = 24$ [Week]. Finally, we select $\eta = 1$ for the entropy function. Figure 1 shows the results of applying the solution in (7) with $v = 1$ compared to a solution to the problem obtained numerically by using forward-Euler with a discretization step $T_d = 0.01$ to discretize the dynamics, lift the problem, and then solve it using SDP3 [13] with CVX [14]. The numerical solution matches the closed-form solution which further validates it. We can see from the solution that the reliance on medical staff and personalized reminders is higher at the beginning than text reminders but slowly decreases with time to reduce the burden on the medical staff. Additionally, none of the nudging techniques have a zero contribution at any point of time and there is always a mix between all of them ((4b) will always be strictly positive). In figure 2, we compare our closed-form

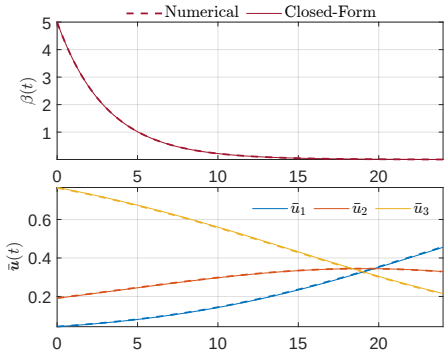

Fig. 1. Numerical solution (dashed) against the closed-form solution (solid) for OCPv with constant $\boldsymbol{v} = \mathbf{1}$.

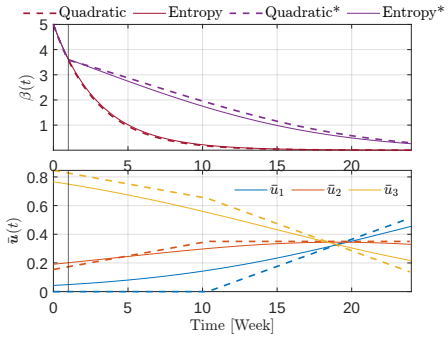

Fig. 2. Comparison between the quadratic objective case (dashed) and the entropy objective case (solid) in OCPv.

solution with a numerical solution obtained using a quadratic cost $-\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}$ instead of the entropy in OCPv. Additionally, we simulate the response $\beta(t)$ in a case where the medical staff become unavailable after the first week rendering $b_3 = 0$ in simulation only and not in the calculation of the input nudges. We can see from the figure how the input nudges with the entropy objective are smoother than the ones calculated with a quadratic cost. Additionally, we see that for the quadratic cost case, the text reminders were not used at all until almost 10 weeks from the beginning of the scheduling of nudges. This

is not preferable since it is desired for the subject to be more acquainted with the different nudging techniques as early as possible to handle technical and personal difficulties from the beginning. The average rate of the behaviour $\beta(t)$ for both the quadratic objective case and the entropy objective case are very similar. For the case when $b_3 = 0$, the entropy objective case has a lower $\beta(t)$ curve over time than the curve obtained with the quadratic objective. This is expected since maximizing the entropy encourage the use of all the available resources which offers robustness in case of the sudden absence of one resource or more.

C. Case with a Time-Varying Nudge Rate

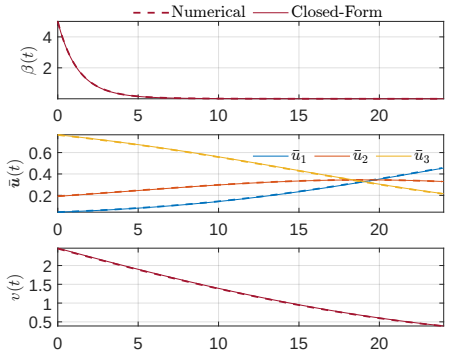

Fig. 3. Numerical Solution (dashed) against the closed-form solution (solid) for OCPv.

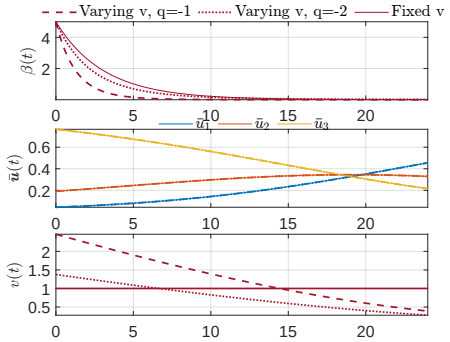

Fig. 4. Comparison between the solutions with varying $\boldsymbol{v}$ with $\boldsymbol{q = -1}$ (dashed), $\boldsymbol{q = -2}$ (dotted), and with a fixed rate $\boldsymbol{v = 1}$ (solid).

We consider in this section the same case in the previous section IV-B but when we desire to optimize over $v(t)$. Figure 3 shows the results when we choose $q = -1$ against a numerical solution obtained using CVX and SDP3. We observe from figure 3 that the numerical solution matches the closed-form solution which further validates it. We notice from the solution that the rate of nudges v at the beginning has a value greater than 2 $[1/\text{Week}]$, and phone calls from medical staff have the highest share of the different types of nudges. Afterwards, the nudge rate v decreases to be below 1 $[1/\text{Week}]$ throughout the solution while the reliance on text reminders is increasing to finally be the nudge with the highest contribution. Allowing v to vary gives the opportunity to lower it while the average behavioral rate $\beta(t)$ is decreasing, which prevents overburdening the subject with nudges that could lead to alert fatigue. In figure 4, we compare the solutions when $v = 1$ with two cases of varying v with $q = -1$ and $q = -2$. We can see from the figure that for both of the cases of varying v, the average rate $\beta(t)$ decreases faster than the case of a fixed rate due to v starting with a value greater than 1. Additionally, we observe that the inputs $\bar{\boldsymbol{u}}$ are identical for all the cases with $\bar{u}_3$ being the highest at the beginning and the lowest towards the end. Notice how increasing $|q|$ will make v starts at a lower value which helps to reduce the risk of alert fatigue from the beginning.

D. Receding Horizon Case

In this section, we demonstrate how the solution of OCPv can be used in a receding horizon fashion to adapt to changes in the parameters of the model. We introduce "feedback" by utilizing recursively estimated values of the model's parameters for the computation of a new scheduling scheme. We choose $\rho = 5$, $q_p = 0.5q$, and $\eta_p = 10\eta$ in (8) for the receding horizon solution. For the simulation, we consider a case in which the effect b_3 of a phone call from the medical staff vanishes for a while during treatment according to $b_3(t) = 0.4 - 0.4\sigma(10(t-10)) + 0.4\sigma(10(t-18))$. Additionally, we consider a case in which the subject pays less attention to text messages on their phone over time captured by modifying the effects b_1 and b_2 according to $b_1(t) = 0.2\left(\frac{1}{2} + \frac{1}{2}\mathrm{e}^{-0.2t}\right)$ and $b_2(t) = 0.3\left(\frac{1}{2} + \frac{1}{2}\mathrm{e}^{-0.2t}\right)$. We simulate a case in which we have a perfect knowledge about $\boldsymbol{b}(t)$ (Nominal), and for a case in which we have an estimate of the effect of nudges. For the receding horizon case, we consider that every week t_j such that $t_j - t_{j-1} = 1$ [Week], $j \in \mathbb{Z}$, we obtain an estimate $\hat{\boldsymbol{b}}(t_j) = \boldsymbol{b}(t_j - 2/7) + 0.25\|\boldsymbol{b}(t_j - 2/7)\|_2\boldsymbol{\xi}(t_j)$ with $\boldsymbol{\xi}(t_j) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$[3]. The solution of the receding horizon for each week then uses a constant $\hat{\boldsymbol{b}}(t_j)$ for the entire week with $t_f = 24$ [Week]. The figure in 5 shows the results. We can see from the results that the open loop response of β with the perfect knowledge of $\boldsymbol{b}(t)$ compared to the one with the receding horizon and imperfect knowledge of $\boldsymbol{b}$ are very similar. As for the inputs, we can see how they are affected by the noise and the delay during the simulation. Despite the presence of noise and delay, the receding horizon solution is able to follow the trend of the optimal open-loop solution for the case of a perfect knowledge of $\boldsymbol{b}(t)$.

V. Conclusion and Future Work

We presented an OCP in which the inputs are constrained to a variable-sized simplex, with the size being another input to optimize over. We showed that with the inclusion of the entropy function in the objective, it is possible to derive closed-form solutions when the dynamics are linear, and the objectives are linear on the states and the simplex inputs, and

[3]Several techniques could be used to obtain estimates of the model parameters with data. The data can contain the frequency of the undesired behaviour, feedback from the subject regarding the effectiveness of the different nudges, and data on how the subject responds to nudges such as the number of times they answer phone calls or read text messages (see [6]). Since the use of these techniques is out of the scope for this paper and the goal of this section is to demonstrate how receding horizon could work, we used a random additive error with a delay of two days to simulate estimation errors.

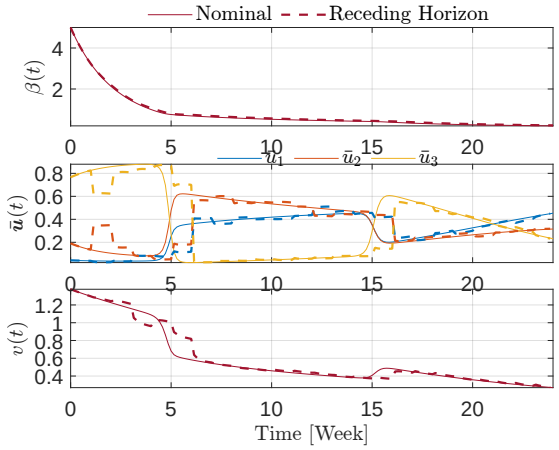

Fig. 5. Solution for the Nominal case with a perfect knowledge of $\boldsymbol{b}(\boldsymbol{t})$ (solid) and a receding horizon case (dashed).

quadratic on the size of the simplex. A possible future research direction is to study a more general class of OCPs with entropy and simplex constraints. We also demonstrated how the formulated OCP can potentially be used in conjunction with behavioral momentum theory in the help of scheduling nudges to discourage unhealthy behaviors, such as non-adherence to medication. This work is a starting point for utilizing control theory methods with the behavioral momentum theory for nudging design. Future work will focus on incorporating more complex behavioural momentum models, comparing this framework with different frameworks such as the one in [7], performing and developing system identification for behavioural momentum models, and applying the solutions in a real-life setting using a receding horizon approach.

Appendix I
Maximization with Entropy

Consider $\boldsymbol{u} \in \triangle_{n_u}^v$ with $v \geq 0$. We will derive the solution for the following problem

$$(\boldsymbol{u}^*, v^*) = \underset{\boldsymbol{u}\in\triangle_{n_u}^v, v\geq 0}{\operatorname{argmax}} \frac{1}{\eta}\phi(\boldsymbol{u}) + \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{u} + qv^2, \tag{14}$$

where $\eta > 0, q < 0$, and $\boldsymbol{\alpha} \in \mathbb{R}^{n_u}$. Define first the Lagrangian on $\operatorname{int}\left(\triangle_{n_u}^v\right) \times \mathbb{R}_{\geq 0}$ as

$$L = \frac{1}{\eta}\phi(\boldsymbol{u}) + \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{u} + qv^2 + \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{u} + \mu v + \zeta\left(v - \mathbf{1}^{\mathrm{T}}\boldsymbol{u}\right), \tag{15}$$

where $\boldsymbol{\lambda} \geq_e 0, \mu \geq 0$, and ζ are Lagrange multipliers. We proceed by writing the Karush–Kuhn–Tucker (KKT) conditions [15] which are sufficient since the problem is concave

$$\frac{-1}{\eta}\ln_e(\boldsymbol{u}) - \frac{1}{\eta}\mathbf{1} + \boldsymbol{\alpha} + \boldsymbol{\lambda} - \zeta\mathbf{1} = 0, \tag{16a}$$

$$2qv + \mu + \zeta = 0, \tag{16b}$$

$$\boldsymbol{u} \geq_e 0,\ u_i\lambda_i = 0,\ \forall i \in \{1, \ldots, n_u\}, \tag{16c}$$

$$v \geq 0,\ \mu v = 0, \tag{16d}$$

$$\mathbf{1}^{\mathrm{T}}\boldsymbol{u} = v. \tag{16e}$$

Since we are considering $\operatorname{int}\left(\triangle_{n_u}^v\right) \times \mathbb{R}_{\geq 0}$, we get $\boldsymbol{\lambda} = \mathbf{0}$ and $\mu = 0$ from (16c), (16e), and (16d). From (16a) we get $\boldsymbol{u} = \exp_e(\eta\boldsymbol{\alpha} - \mathbf{1} - \eta\zeta\mathbf{1}) >_e 0$. From (16b), we have $\zeta = -2qv$ which we substitute back in $\exp_e(\eta\boldsymbol{\alpha} - \mathbf{1} - \eta\zeta\mathbf{1})$ and use (16e) to get

$$\begin{aligned} &\mathbf{1}^{\mathrm{T}}\exp_e(\eta\boldsymbol{\alpha})\exp(-1)\exp(2q\eta v) = v, \\ &\Rightarrow -2q\eta v\exp(-2q\eta v) = -2q\eta\exp(-1)\mathbf{1}^{\mathrm{T}}\exp_e(\eta\boldsymbol{\alpha}). \end{aligned} \tag{17}$$

Equation (17) is in the form of $y\exp(y) = x$ with $x > 0$ ($q < 0$). The solution of this equation is known to be the principle branch of the Lambert W function $y = \mathrm{W}_0(x)$. With the Lambert W function, we solve (17) for v to obtain

$$v^* = \frac{-\mathrm{W}_0\left(-2q\eta\exp(-1)\mathbf{1}^{\mathrm{T}}\exp_e(\eta\boldsymbol{\alpha})\right)}{2q\eta} \tag{18}$$

$$\boldsymbol{u}^* = \frac{v^*\exp_e(\eta\boldsymbol{\alpha})}{\mathbf{1}^{\mathrm{T}}\exp_e(\eta\boldsymbol{\alpha})}. \tag{19}$$

Since the objective function in (14) is strictly concave on $\triangle_{n_u}^v \times \mathbb{R}_{\geq 0}$, then (18) is the unique maximizer on $\triangle_{n_u}^v \times \mathbb{R}_{\geq 0}$.

References

[1] H. Wang, T. Zariphopoulou, and X. Y. Zhou, "Exploration versus exploitation in reinforcement learning: A stochastic control approach," *Available at SSRN 3316387*, 2019.

[2] J. Kim and I. Yang, "Maximum entropy optimal control of continuous-time dynamical systems," *IEEE Transactions on Automatic Control*, 2022.

[3] A. Seierstad and K. Sydsaeter, *Optimal control theory with economic applications*. Elsevier North-Holland, Inc., 1986.

[4] B. D. Greer, W. W. Fisher, P. W. Romani *et al.*, "Behavioral momentum theory: A tutorial on response persistence," *The Behavior Analyst*, vol. 39, pp. 269–291, 2016.

[5] Y. Chen, S. Harris, Y. Rogers *et al.*, "Nudging within learning health systems: next generation decision support to improve cardiovascular care," *European Heart Journal*, vol. 43, no. 13, pp. 1296–1306, 2022.

[6] B. D. Horne, J. B. Muhlestein *et al.*, "Behavioral nudges as patient decision support for medication adherence: the encourage randomized controlled trial," *American Heart Journal*, vol. 244, pp. 125–134, 2022.

[7] F. Callaway, M. Hardy, and T. Griffiths, "Optimal nudging for cognitively bounded agents: A framework for modeling, predicting, and controlling the effects of choice architectures," Jan 2022. [Online]. Available: psyarxiv.com/7ahdc.

[8] C.-T. Chen, *Linear system theory and design*. Saunders college publishing, 1984.

[9] K. G. Murty and S. N. Kabadi, "Some np-complete problems in quadratic and nonlinear programming," Tech. Rep., 1985.

[10] J. A. Nevin and T. A. Shahan, "Behavioral momentum theory: Equations and applications," *Journal of Applied Behavior Analysis*, vol. 44, no. 4, pp. 877–895, 2011.

[11] B. S. Last, A. M. Buttenheim *et al.*, "Systematic review of clinician-directed nudges in healthcare contexts," *BMJ open*, vol. 11, no. 7, p. e048801, 2021.

[12] F. B. Hanson, *Applied stochastic processes and control for jump-diffusions: modeling, analysis and computation*. SIAM, 2007.

[13] K.-C. Toh, M. J. Todd, and R. H. Tütüncü, "Sdpt3—a matlab software package for semidefinite programming, version 1.3," *Optimization methods and software*, vol. 11, no. 1-4, pp. 545–581, 1999.

[14] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.2," http://cvxr.com/cvx, Jan. 2020.

[15] W. Karush, "Minima of functions of several variables with inequalities as side conditions," in *Traces and Emergence of Nonlinear Programming*. Springer, 2014, pp. 217–245.