An Efficient Understandability Objective for Dynamic Optimal Control

D. Livingston McPherson, S. Shankar Sastry

***Abstract*— Motion optimization for legible robot intent has largely ignored the robot's dynamics, citing burdensome complexity that prevents online deployment. Even where the original task (to be communicated) could be solved on the dynamical system, the legibility problem (to communicate that task's intent) could not. This work simplifies the legibility objective to have equivalent computational complexity as the original objective to be communicated. This enables any optimal control algorithm that can solve the original task to also solve the legible version of that task.**

Along the way, we expand the definition of "intent" to include any parameter of the optimal control problem, thereby opening the door to extend communications beyond merely desired endpoints to running preferences or even, in the future, hard capabilities or safety constraints. We demonstrate how this method can replicate the properties introduced in previous communicative motion state-of-the-art (like legibility, exaggeration, and anticipation) as well as apply to non-holonomic dynamical systems.

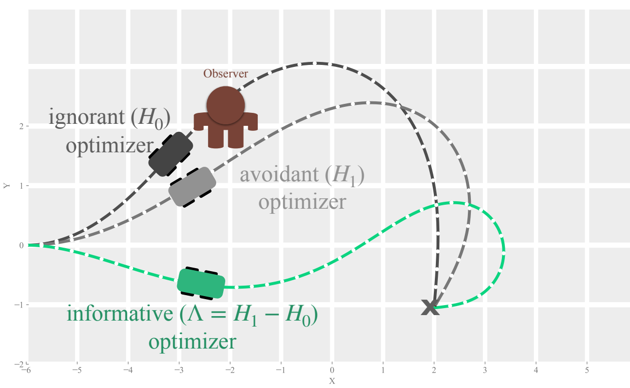

Fig. 1. *Overview*: A human observes three possible robot motions with bicycle dynamics. The black path (top) is ignorant of needing to avoid the human and its control-minimizing path is a collision course. The gray path (middle) is optimized with a collision avoidance term in addition to minimizing control effort. It's path succeeds in avoiding the human, but to minimize control cost it comes concerningly close. The light green path (bottom) is optimized to evidence its awareness of the human's needs so the human is informed. This legibility optimization metric was historically too complex to apply to non-holonomic dynamics tractably (see conclusions of [5]). This work derives a tractable equivalent metric.

1. Introduction

Every good collaboration is founded upon mutual understanding. Unfortunately, the old paradigm of factory-caged robots has no methods for informing other agents of their intent. It is essential for robots to synchronize their task understanding and communicate task intent to their human collaborators. For example, even when coordinating something as simple as handing over an object, task intent of "what", "when", and "where" must be synchronized [21]. This problem of conveying action intent is tackled by the nascent field of "Explainable Agency". The goal here is for the agent to explain their latent planning variables, such as preferences or future plans, to other relevant agents to aid collaboration [26], create transparency, or increase trust (the three most popular motivations in a recent literature review of the area [1]). Agents can generate and communicate explanations of their intent through a variety of modalities: text, graphical user interfaces, and expressive lights, amongst others [1]. Yet each of these approaches requires receivers to monitor a channel outside the task that occupies them.

Instead of investing resources and attention into a side channel, recent research has focused on using the performance of the task itself to communicate its goals. Hand-animated performances of robot characters improved test subjects' confidence and accuracy at interpreting the robot's intent [23]. Human-performed gestures were analyzed and re-mapped to robot motions by automatically detecting the salient keyframes of the gesture and emphasize them on robot kinematics [10]. Eschewing human data generation in favor of animation's first principles, Szafir et alius [22] proposed three motion primitives they hypothesized would signal future motions for their flying robots. Similarly, Zhao et alius [27] identified motion primitives and signaling features for their robot handover task. These identified features are specific to the robot morphology and task-type being conveyed (e.g. robot flying or manipulator handover), and so must be re-designed for new application contexts.

In an attempt to identify a generalizable principle, researchers have turned to models of human interpretation. Particularly amenable to robotics usage, is the "theory theory" that humans construct and test models from observed data [11][12]. Particularly, action interpretation has been described as Bayesian inferences on latent planning preferences [3][4][14][17]. Designers of explainable AIs have turned this cognitive model of action understanding into an optimization objective for motion planning [7][9][13].

Instead of identifying trajectory-space features like [27], Dragan and Srinivasa [8] derived an optimization metric they dub "legibility" based on principles of how humans observe and interpret robot motion. Optimizing trajectories with respect to this metric [7] recovered behaviors like arcing and exagerating identified in previous approaches (like Szafir et alius [22]). Furthermore, this principle promises

This work is supported by the National Science Foundation under Grant CNS-1545126 (VeHICaL)

All authors are with the Department of Electrical Engineering and Computer Science, University of California, Berkeley {david.mcpherson, sastry}@eecs.berkeley.edu

to generalize to a variety of morphologies and applications, such as self-driving vehicles [5].

However, this generalization to dynamic control is rare amongst interpretable motion optimization (both inference-model base [7][9][13] and otherwise [6][18]). This is likely due to the prohibitive computation costs required of dynamically-constrained optimizations with complex objective functions; the only prior work on legibility optimization for nonlinear dynamics [5] emphasized that the algorithm lacked the efficiency to run online.

This paper derives a simplified equivalent metric for legibility that has the identical complexity class for access as the originally intended task to be signaled. That is, optimizing the legibility problem requires no extra complexity over the non-legible version.

A. Contributions and Outline

The main contribution of this paper is the simplified legibility formulation that will be laid out in Section 4. Along the way, a more powerful hypothesis class for modeling "intent" is introduced in Section 3. This broader class opens the door to communicating more than just endpoints, as illustrated in Fig. 1. The beginnings of this will be illustrated in Section 6 while the promising future directions will be discussed in Section 7.

2. Mathematical Background

The robotic motion problem is to select actions u from the set of available choices $U = \mathbb{R}^{n_u}$ to steer the evolving state x in the state space $X = \mathbb{R}^{n_x}$. The actions influence the state through the dynamic difference equation:

$$x(t+1) = \bar{f}(x(t), u(t)), \quad \forall t \in [0, 1, \cdots, T] \tag{1}$$

for discrete-time systems or through the dynamics differential equation:

$$\dot{x}(t) = f(x(t), u(t)), \quad \forall t \in [0, T) \tag{2}$$

for continuous time systems. Here the time-indexing set $[0, 1, \cdots, T] \subset \mathbb{Z}$ or $[0, T) \subset \mathbb{R}$ is called the time horizon $\mathcal{T}$ with T being the final time. Let $\mathcal{X}$ be the set of functions $x(\cdot) : \mathcal{T} \to X$ and $\mathcal{U}$ be the set of functions $u(\cdot) : \mathcal{T} \to U$.

Aside: For the rest of this paper, we will focus on the continuous-time case but the mathematics will apply straightforwardly to discretely-indexed functions. Similarly we focus attention on continuous state and action spaces, but the contributions can be re-derived for discrete state or action spaces with some change in notation.

Actions can be chosen following a variety of paradigms. This paper follows the optimization-based paradigm for reflecting goal-driven "intelligent" behavior. Here possible action choices, along with their resultant state trajectories starting from some given initial state $x(0) = x_0$, are ranked by some objective function $J(x(\cdot), u(\cdot))$. Often these objectives are time-decomposable and so can be broken into a running cost $L(x(t), u(t))$ and terminal cost $\phi(x(T))$ as:

$$J(x(\cdot), u(\cdot)) = \phi(x(T)) + \int_0^T L(x(\tau), u(\tau)) d\tau \tag{3}$$

The goal is to choose the actions to achieve an extremum of this objective function (a minimum when J is considered a *cost* or a maximum when J is considered a *reward*). This is notated mathematically:

$$\begin{aligned} &\min_{u(\cdot)} J(x(\cdot), u(\cdot)) \\ &\text{subject to} \\ &\qquad \dot{x}(t) = f(x, u, t) \quad \forall t \in \mathcal{T} \\ &\qquad x(0) = x_0 \end{aligned} \tag{4}$$

Algorithms focused on planning the path abstract the problem by allowing the agent to directly choose states x. This configuration is equivalent to setting the action space equal to the state space $U = X$ and setting

$$f(x, u, t) = u \tag{5}$$

which is a linear, fully controllable dynamic system making it simple to optimize. Applied systems rarely have this luxury, but sacrificing dynamic feasibility is often made in order to tackle more complex objectives, like in joint task-and-motion planning or for non-convex human-factors objectives (like in [18]). This work will simplify the legibility metric enough that it can be optimized with respect to even nonlinear dynamics.

For notational concision and to focus on the actual decision variable $u(\cdot)$, we will combine the cost function $J(x(\cdot), u(\cdot))$ with the constraints $\dot{x}(t) = f(x, u, t) \quad \forall t \in \mathcal{T}$ and $x(0) = x_0$

$$\mathcal{H}_J(u(\cdot)) = J(\rho(u(\cdot), x_0), u(\cdot)), \tag{6}$$

using the unique solution map $\rho(u(\cdot), x_0)$ (that exists from the Picard-Lindelof theorem). We will narrow $\mathcal{U}$ to the set of piecewise continuous functions and require that the dynamics $f(x, u, t)$ are Lipschitz continuous in state x, continuous in u and piecewise continuous in t.

3. External Observer Modeling

Traditionally, when animating a robot using optimal control, the robotic agent optimizes its behavior to concord with some agenda $J(x(\cdot), u(\cdot))$ (e.g. merge onto a highway without collisions). Yet optimizing just a task objective ignores the needs of external observers working around the robot. Robots do not operate in an informational vacuum. As robots start working in human-spaces, these other agents will observe and interpret the robot's motion to understand their behavior. In this work, we are particularly interested in how humans understand the robot's intended goals. These can be encoded as optimization parameters. So the human's interpretation is deciding between hypothesized optimization problems that might be generating the robot's behavior. Call these hypotheses H_i each corresponding to optimizing a different metric $\mathcal{H}_{J_i}(u(\cdot))$. This paper focuses on binary hypothesis testing: the human is deciding between a default case H_0 and an alternative H_1. Multiple hypothesis testing can be performed as iterated binary hypothesis testing with corrections (e.g. Bon Ferroni), but this problem is left for future work.

Following the hypothesis testing framework, we assume there is a null hypothesis that the observer will default to believing and an alternative hypothesis that describes the true characteristic of our robotic agent. These could be some essential binary (e.g. safe/unsafe) or a choice between two options (reaching for the left object or the right one). Applications like these are explored further in Sections 6 and 7.

For any binary hypotheses, the uniformly most powerful test that an observer can use to judge the robot's performance is a Likelihood Ratio Test:

$$\begin{aligned} &\text{if } \Lambda(u(\cdot)) \leq \eta \quad &&\text{, rejects } H_0 \\ &\text{if } \Lambda(u(\cdot)) > \eta \quad &&\text{, fails to reject } H_0 \end{aligned}$$

where the deciding factor is the ratio between the probability of observing the robot's choices given the different hypotheses $P(u(\cdot)|H_i)$:

$$\Lambda(u(\cdot)) \coloneqq \frac{P(u(\cdot)|H_0)}{P(u(\cdot)|H_1)} \tag{7}$$

Following cognitive scientific models such as [17][19], we formulate the hypothesized likelihood of observing a performed motion as an exponential distribution with cost as the sufficient statistic:

$$P(u(\cdot)|H_i) \propto \frac{e^{-\mathcal{H}_{J_i}(u(\cdot))}}{Z_i} \tag{8}$$

This distribution models the human observer as expecting optimal behavior (according to their hypothesized $\mathcal{H}_{J_i}$), but leaving sub-optimal behaviors as possible due to their inconfidence in $\mathcal{H}_{J_i}$ [19][28]. This distribution is reffered to as the "Boltzmann distribution" by analogy to the statistical mechanical equilibrium distribution where the negative cost function is interpreted as the energy.

This Boltzmann distribution was used in the prior state-of-the-art [7], but required calculating the normalization constant Z_i. This requires integrating over all possible time-varying controllers:

$$\begin{aligned} P(u(\cdot)|H_i) &= \frac{e^{-\mathcal{H}_{J_i}(u(\cdot))}}{Z_i} \\ &= \frac{e^{-\mathcal{H}_{J_i}(u(\cdot))}}{\int_{u(\cdot)} e^{-\mathcal{H}_{J_i}(u(\cdot))} du(\cdot)} \end{aligned} \tag{9}$$

Unfortunately, this is an infinite dimensional space. Dragan and Srinivasa sidestepped this Sisyphean task by approximating it instead by fully solving for the optimal cost-to-go in the control problem. For their focus on unconstrained planning problems with quadratic costs, this could be done in closed form. Unfortunately, when working with nonlinear dynamics, the problem can no longer be solved analytically.

In the next section we will introduce an equivalent optimization problem that *does not require calculating the partition function Z_i at all.*

4. Optimizing Controls for Informativeness

The model of human action interpretation in Section 3 classifies some actions as evidence for the null hypothesis and others as evidence for the alternative. Our robot can optimize its chosen actions to ensure it evidences the correct alternative:

Lemma 4.1. *For every likelihood ratio testing observer, the control that optimizes:*

$$\min_{u(\cdot)} \Lambda(u(\cdot)) = \min_{u(\cdot)} \frac{P(u(\cdot)|H_0)}{P(u(\cdot)|H_1)} \tag{10}$$

is guaranteed to be evidence to reject the null hypothesis for every obsever with non-empty rejection region:

$$R_{NP} = \left\{ x : \Lambda(u(\cdot)) = \frac{P(u(\cdot)|H_0)}{P(u(\cdot)|H_1)} \leq \eta \right\} \neq \varnothing \tag{11}$$

Proof of Lemma 4.1. From the assumed property (11) there exists an $u_r(\cdot) \in R_{NP}$ that, by definition, satisfies:

$$\Lambda(u_r(\cdot)) \leq \eta$$

Yet this $u_r(\cdot)$ cannot have smaller $\Lambda(u(\cdot))$ than $u^*(\cdot)$ since $u^*(\cdot)$ optimizes Λ. Indeed,

$$\Lambda(u^*(\cdot)) \leq \Lambda(u_r(\cdot)) \leq \eta$$

Therefore the optimizer is in the rejection region for any testing value of η.

■

This optimization turns out to have a simple form in terms of the hypotheses' costs:

Theorem 4.2 (Simplified Equivalent of Maximizing Self-Evidence)**.** *The problem of maximizing the observer's likelihood ratio in favor of the alternative hypothesis:*

$$\min_{u(\cdot)} \Lambda(u(\cdot)) = \min_{u(\cdot)} \frac{P(u(\cdot)|H_0)}{P(u(\cdot)|H_1)} \tag{12}$$

has the same optima as:

$$\min_{u(\cdot)} \mathcal{H}_{J_1}(u(\cdot)) - \mathcal{H}_{J_0}(u(\cdot)) \tag{13}$$

Call this equivalent objective:

$$L(u(\cdot)) \coloneqq \mathcal{H}_{J_1}(u(\cdot)) - \mathcal{H}_{J_0}(u(\cdot))$$

Proof of Theorem 4.2.

$$\begin{aligned}\min_{u(\cdot)} \Lambda(u(\cdot)) &= \min_{u(\cdot)} \frac{P(u(\cdot)|H_0)}{P(u(\cdot)|H_1)} \\ &= \min_{u(\cdot)} \frac{\frac{e^{-\mathcal{H}_{J_0}(u(\cdot))}}{Z_0}}{\frac{e^{-\mathcal{H}_{J_1}(u(\cdot))}}{Z_1}} \\ &= \min_{u(\cdot)} \frac{Z_1}{Z_0} \frac{e^{-\mathcal{H}_{J_0}(u(\cdot))}}{e^{-\mathcal{H}_{J_1}(u(\cdot))}} \\ &= \min_{u(\cdot)} \frac{Z_1}{Z_0} e^{-\mathcal{H}_{J_0}(u(\cdot))+\mathcal{H}_{J_1}(u(\cdot))}\end{aligned}$$

An objective can be composed with any non-decreasing function and have equivalent optima. Since the partitions Z_i are non-negative and the logarithm is non-decreasing we can compose an equivalent objective as:

$$\begin{aligned}L(u(\cdot)) &\coloneqq \log\left(\frac{Z_0}{Z_1}\Lambda(u(\cdot))\right) \\ &= \log\left(\frac{Z_0}{Z_1}\frac{Z_1}{Z_0} e^{-\mathcal{H}_{J_0}(u(\cdot))+\mathcal{H}_{J_1}(u(\cdot))}\right) \\ &= \log\left(e^{-\mathcal{H}_{J_0}(u(\cdot))+\mathcal{H}_{J_1}(u(\cdot))}\right) \\ &= -\mathcal{H}_{J_0}(u(\cdot)) + \mathcal{H}_{J_1}(u(\cdot)) \\ &= \mathcal{H}_{J_1}(u(\cdot)) - \mathcal{H}_{J_0}(u(\cdot)) \qquad (14)\end{aligned}$$

■

Therefore the equivalent objective $L(u(\cdot))$ defined in Equation 14 can be used instead of the likelihood ratio, thereby avoiding calculating partition functions Z_i. This goal can be understood as aiming to improve the alternative hypothesis' cost while performing worse at the null hypothesis' cost. The simple linearity of Equation 14 makes informative control as tractable as the original optimization:

Corollary 4.2.1 (Time Complexity of Theorem 4.2's objective)**.** *For any gradient-based control optimization method, finding the optima of the likelihood ratio inherits the order of time-complexity from the original (non-communicative) optimizations in equation (4) (whichever has higher time-complexity) .*

Proof of Corollary 4.2.1. The optima of the likelihood ratio can be found as the optima of Equation 14. Because of the linear form of Equation 14, all queries to the objective function and its derivatives must only query the two hypotheses' respective lookups and combine them. Therefore the computational complexity of each iteration will be at most the sum of the two hypotheses' complexities. Thus they will share the same growth-rate/order of time-complexity.

■

The hypotheses' rewards are combined linearly with equal weights in Theorem 4.2. However, if the designer desires the robot to not only optimally communicate but also optimize the original reward, more weight on the H_1 term can be added in:

$$\begin{aligned}&\min_{u(\cdot)} L(u(\cdot)) + \alpha\mathcal{H}_{J_1}(u(\cdot)) \\ &= \min_{u(\cdot)} (1+\alpha)\mathcal{H}_{J_1}(u(\cdot)) - \mathcal{H}_{J_0}(u(\cdot)) \qquad (15)\end{aligned}$$

This weight α could be interpreted formally as a Lagrange multiplier on a sub-optimality bound as suggested in Section VI of [7]. This relative weighting between the original optimization and the informativeness objective can be dynamically shifted to create another desirable property: anticipativeness.

5. Anticipation through Receding Horizon Control

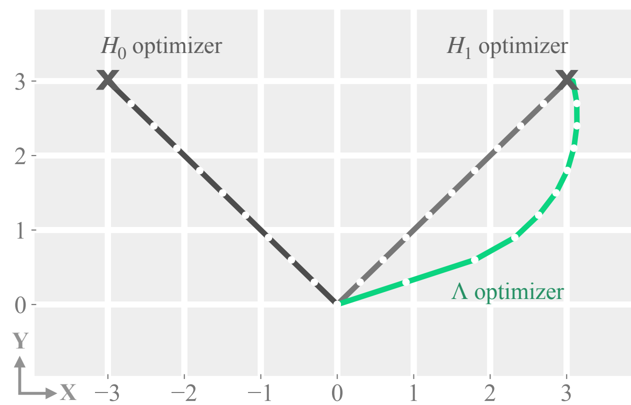

Fig. 2. Optimized paths through x-y space reaching for either the leftwards destination (H_0) or the rightwards destination (H_1). The anticipative trajectory (in green) that optimizes Λ leads rightwards early; as opposed to the non-anticipative trajectories (in gray) which indicate much more slowly. The H_1 trajectory takes three times longer to move rightwards to $x = 2$

Previous work emphasized the importance of communicating intent earlier in the motion. Gielniak and Thomaz [10] promoted the concept of "anticipativeness" and tweaked motions to express salient gestures earlier on in the time horizon.

Dragan and Srinivasa [7] incorporated this concept of anticipativeness by ensuring that unfininshed viewings of the motion plan would also push the Bayesian observer towards the correct conclusion. They formulated this early expressiveness by optimizing for all incomplete viewings on time horizons $[0, t]$ for $t \in [0, T]$ simultaneously. They chose to combine these multiple sub-objectives through a weighted linear combination:

$$\int_{t=0}^{T} (T-t) P(H_1|u(0:t)) dt$$

This sum of probabilities again requires weighting by the normalizing partition constants. Instead, an equivalent prioritization of legibility for earlier controls can be accomplished through receding horizon control. Anticipativeness can be obtained by reweighting informativeness more heavily during earlier horizons and then shifting to prioritizing pure efficiency in later horizons by increasing the α in Equation 15.

Let the problem be replanned at times $\bar{t}_0, \bar{t}_1, \cdots \bar{t}_M$. Let $\alpha(t)$ be some increasing function of time that will prioritize efficiency over communicativeness in later replanning horizons. Let $\bar{u}_i(\cdot)$ be the optimal controls from one of M optimization problems on horizons $[\bar{t}_i, T)$

$$\bar{u}_i = \arg\min_{u(\cdot)} (\alpha(\bar{t}_i) + 1)\mathcal{H}_{J_1}(u(\cdot)) - \mathcal{H}_{J_0}(u(\cdot)) \quad (16)$$

The robot will follow the controls from $\bar{u}_i(\cdot)$ for all times $\bar{t}_i \leq t < \bar{t}_{i+1}$.

Whereas the original path planning problem would have been too expensive to replan, the streamlined objective in Theorem 4.2 is tractable enough to replan online. In fact, for the problem in Dragan and Srinivasa [7] it becomes a Linear-Quadratic Regulator (LQR). This is because they use the path-planning approximation of Equation (5) for the dynamics and a quadratic penalty:

$$J_i(x(\cdot), u(\cdot)) = \|x(T) - g_i\|_2^2 + \int_0^T \|u(t)\|_2^2 dt \quad (17)$$

where g_i was one of two goals the robot could be reaching to grab.

This receding horizon anticipation optimization was used to recreate Dragan and Srinivasa's [7] exagerated arcing path plans, as seen in Figure 2. Unlike that method, the solutions could be found analytically thanks to the streamlined objectives of Equation (13) leaving the LQR structure intact.

6. Experimental Applications

Beyond replicating previous art with greater efficiency, the legible model predictive control (LMPC) algorithm can solve entirely new problems:

A. Responsiveness to Updates

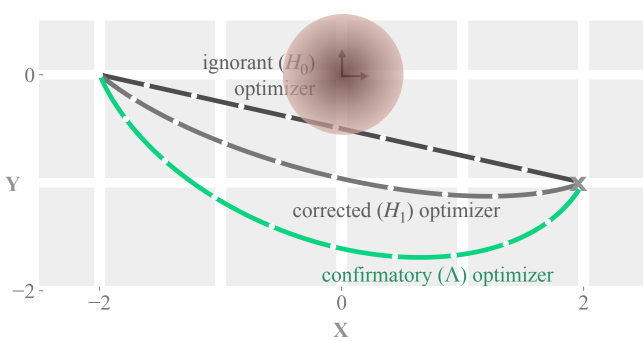

Fig. 3. Optimized paths in x-y space: After the instructor corrects the robot to avoid the red region around the origin, the robot must demonstrate its new understanding. The dark path is the optimum pre-correction H_0 (from Equation 18), the gray path is the optimum post-correction H_1 (from Equation 19), and the green path is the optimum for informing the corrector of the successful correction Λ; all here with $g = [2, -2]^T$, $a_1 = 40$, $a_2 = 25$, $a_3 = 1$.

Section 3 extended the communicatable payload from just end-destinations (as was the case in [7][9]) to include the dynamic cost and constraints as well. This opens the door to communicate differences in running costs as in [13] but through communicative motion instead of scenario generation. This could help improve in-task teaching (like the physical in-task value alignment in [2]) by confirming whether the lesson was learned correctly, thereby completing the communication loop proposed in [1]'s roadmap. Figure 3 takes the use case of [2] where the user corrects the robot and adds a penalty on approaching the obstacle at the origin:

$$J_0(x(\cdot), u(\cdot)) = a_1\|x(T) - g\|_2^2 + \int_0^T a_2\|u(t)\|_2^2 dt \quad (18)$$

$$\begin{aligned} J_1(x(\cdot), u(\cdot)) = {} & a_1 * \|x(T) - g\|_2^2 \\ & + \int_0^T a_2\|u(t)\|_2^2 - a_3\|x(t)\|_2^2 dt \quad (19) \end{aligned}$$

By accentuating the alternative hypothesis H_1 the robot disambiguates whether or not it's understood the correction. If the human likes the robot's understanding of J_1 they can rest easy. If they find the new behavior still unsatisfactory, they are now informed to know what else must be added.

B. Legibility for Nonlinear Systems

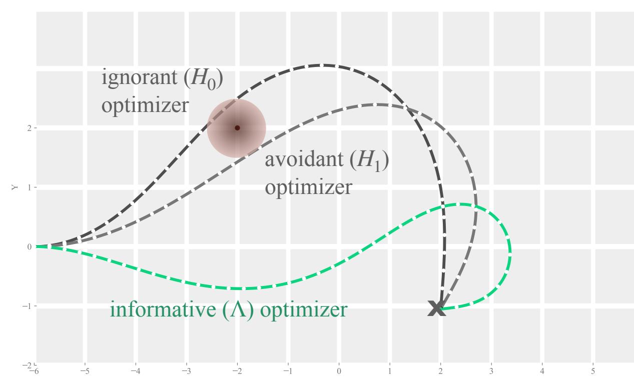

Fig. 4. The informative control optimization can even apply to nonlinear dynamics. After adding a quadratic penalty to nearing state $[2, -2]^T$, the avoidant optimum to H_1 (in gray) indeed has a farther *integral* than the ignorant optimum to H_0 (in black), but the path still looks qualitatively the same. In contrast, the informative optimizer to Λ makes its avoidance obvious. Here $g = [2, -1]^T$, $h = [-2, 2]^T$, $a_1 = 800$, $a_2 = 10$, $a_3 = 2$.

Not only does the reformulation in Section 3 allow communicating differences in running *costs*, it is also built to apply to systems with running dynamic *constraints*. And the new formulation in Section 4 is lightweight enough to be tractable for the numerical methods often necessary for nonlinear optimal control.

This capacity is demonstrated in Figure 4 for the following example nonlinear dynamical control problem. The streamlined objective in Equation 14 can be optimized for nonlinear dynamics using established nonlinear control frameworks, for example the iterative Linear Quadratic Regulator approach [25][24]. Consider the three dimensional Dubins vehicle with constant velocity $v = 3$:

$$\frac{d}{dt}x(t) = \begin{pmatrix} v\cos(x_3(t)) \\ v\sin(x_3(t)) \\ u(t) \end{pmatrix} \tag{20}$$

The null hypothesis minimizes the control effort added to the final distance from the goal $x-y$ point $g = [2, -1]$ (it is agnostic to angle), while the alternative hypothesis also quadratically penalizes proximity to an undesirable $x-y$ position at $a = [-2, 2]$:

$$J_0(x(\cdot), u(\cdot)) = a_1\|Px(T) - g\|_2^2 + \int_0^T a_2\|u(t)\|_2^2 dt \tag{21}$$

$$J_1(x(\cdot), u(\cdot)) = a_1\|Px(T) - g\|_2^2 + \int_0^T a_2\|u(t)\|_2^2 - a_3\|Px(t) - h\|_2^2 dt \tag{22}$$

where P is the projection from the three dimensional state of planar position and angle down only to planar position:

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

C. Predictability versus Legibility Discussion

Section 4 proved the interchangeability of motion communicativeness with the streamlined objective $L(u(\cdot))$ along with time-complexity and outcome guarantees. The linear structure of $L(u(\cdot))$ also clarifies a disagreement in the literature on the relation of "predictablity" and "legibility". Dragan and Srinivasa [7] emphasized the distinction between "predictablity", defined as optimizing expectedness given the task $H_1(u(\cdot))$, and "legibility", defined as optimizing making the task clear which in our binary hypothesis setting is $\Lambda(u(\cdot))$. Crucially they state in [8]:

> "Predictability and legibility are fundamentally different and often contradictory properties of motion."

Lichtenthaler and Kirsch questioned this contradictoriness and concluded that "the two factors are coherent" [16].

Our work clarifies the exact relation between "predictability" $H_1(u(\cdot))$ and "legibility" $\Lambda(u(\cdot))$ through the simplified relation stated in Theorem 4.2. By reformulating the legibility problem from $\Lambda(u(\cdot))$ into the equivalent, simplified linear combination $L(u(\cdot))$ in Eq. 13, we can state the difference between predictability and legibility quite simply: legibility $L(u(\cdot))$ increases linearly with predictability $H_1(u(\cdot))$, meaning they are indeed coherent and typically correlated as [16] asserts, while also having an uncorrelated term (equal to adding in $-H_0(u(\cdot))$) that will cause the legibility optimizers to be fundamentally different from predictability optimizers.

7. Future Work

We found a formulation for legible control that is on the exact same order of complexity as the original (illegible) control problems and demonstrated some legible motions. Yet this is only a groundwork. Much work remains to be done exploring how real observers will adapt and the rich possibilities of extensions. We outline just a few below.

A. Human Modeling

This paper derived a simplified objective for choosing controls that will communicate the robot's task-intent. And every communication requires assuming how receivers will interpret the signals. We have laid out our assumed model based on the receiver testing optimally (i.e. uniformly most powerfully) with respect to binary Boltzmann hypotheses. Which in turn are optimal distributions around a known characteristic reward function with random hidden preferences [19][28]. Yet this is not the only tenable receiver model.

There is a rich space of alternative decision models for human observers. In particular, humans rarely have infinite computation resources to judge and so will likely use heuristics (e.g. truncated cost-to-gos on their time horizon judgements, or not consider alternative outcomes fully and clip the infinite support of the Boltzmann distribution Equation 8). And even with infinite computational resources, decisions may be skewed by risk-averseness (like in [20]) if their decision has tangible outcomes (e.g. whether or not to trust an oncoming autonomous vehicle with their safety).

This paper also focused on cases where there are two clear hypotheses (e.g. left or right, ignorant or corrected). The binary hypothesis testing could be extended to multiple hypothesis testing thereby inducing a multi-objective optimization. This could be handled naively as a weighted linear combination. Or if the likelihood ratio thresholds for the human can be determined, informativeness could be encoded as constraints ensuring correct decisions along each pair:

$$\begin{aligned} \min_u \; & H_1(u) \\ \text{subject to } \; & H_1 - H_{0,1} \le \eta_1 \\ & H_1 - H_{0,2} \le \eta_2 \\ & \vdots \\ & H_1 - H_{0,N} \le \eta_N \end{aligned}$$

Such an extension could explore experimentally how many hypotheses a human observer reasonably considers simultaneously and how new hypotheses emerge.

B. Communication Extensions

Casting the hypotheses as differentiating between full optimal control tasks allowed us to communicate about more than endpoint states. A clear case where demonstrating a different task understanding is crucial is when being responsive to users' value alignment requests. Future work should explore what cognitive phenomena arise in how users perceive and plan around systems that communicate reception of their inputs.

Generalizing the hypotheses from destinations to full optimal control problems also opens the possibility of communicating what *constraints* the robot is bound to through motion. This framework can provide a new perspective on communicating capability as Kwon et al. prompted in [15]. This constraint communication could also be used to inform a supervisor whether or not the robot is obeying safety rules. The hard *evidence* of performed motion may be a uniquely suited communication channel to debug system failures; as the old adage goes "actions speak louder than words". For constraint communicating to work, future work must develop what it means to substract reward-hypotheses that don't share the same support. After deriving, we can explore cognitive phenomena on safety-critical judgement (such as risk-sensitivity [20]).

8. Conclusion

This paper derived the groundwork for informative motion on the control loop level. By broadening the subject to be conveyed from endpoint goals (as in [7][9]) to full optimal control problems, we are able to convey more details about the agent's task-intent (as we showed with running costs) and apply to more complex systems (as we demonstrated with the Dubins car). By modeling the receiving observer as testing binary hypotheses, we obtained a communicative motion optimization problem on the same order of complexity as the non-commmunicative problem.

This paper was able to recreate the desirable properties identified by previous work [10][7] through a more efficient objective that is portable to a cornucopia of optimal control approaches.

References

[1] Sule Anjomshoae, Amro Najjar, Davide Calvaresi, and Kary Främling. Explainable agents and robots: Results from a systematic literature review. In *18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019*, pages 1078–1088. International Foundation for AAMAS, 2019.

[2] Andrea Bajcsy, Dylan P Losey, Marcia K OMalley, and Anca D Dragan. Learning robot objectives from physical human interaction. *Proceedings of Machine Learning Research*, 78:217–226, 2017.

[3] Chris Baker, Rebecca Saxe, and Joshua B Tenenbaum. Bayesian models of human action understanding. In *Advances in neural information processing systems*, pages 99–106, 2006.

[4] Chris L Baker, Rebecca Saxe, and Joshua B Tenenbaum. Action understanding as inverse planning. *Cognition*, 113(3):329–349, 2009.

[5] Tim Brüdigam, Kenan Ahmic, Marion Leibold, and Dirk Wollherr. Legible model predictive control for autonomous driving on highways. *IFAC-PapersOnLine*, 51(20):215–221, 2018.

[6] Baptiste Busch, Jonathan Grizou, Manuel Lopes, and Freek Stulp. Learning legible motion from human–robot interactions. *International Journal of Social Robotics*, 9(5):765–779, 2017.

[7] Anca Dragan and Siddhartha Srinivasa. Generating legible motion. In *Proceedings of Robotics: Science and Systems*, Berlin, Germany, June 2013.

[8] Anca D Dragan, Kenton CT Lee, and Siddhartha S Srinivasa. Legibility and predictability of robot motion. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 301–308. IEEE, 2013.

[9] Jaime F Fisac, Chang Liu, Jessica B Hamrick, Shankar Sastry, J Karl Hedrick, Thomas L Griffiths, and Anca D Dragan. Generating plans that predict themselves. In *Algorithmic Foundations of Robotics XII*, pages 144–159. Springer, 2020.

[10] Michael J Gielniak and Andrea L Thomaz. Generating anticipation in robot motion. In *RO-MAN*, pages 449–454. IEEE, 2011.

[11] Noah D Goodman, Joshua B Tenenbaum, and Tobias Gerstenberg. Concepts in a probabilistic language of thought. Technical report, Center for Brains, Minds and Machines (CBMM), 2014.

[12] Alison Gopnik and Henry M Wellman. Reconstructing constructivism: Causal models, bayesian learning mechanisms, and the theory theory. *Psychological bulletin*, 138(6):1085, 2012.

[13] Sandy H Huang, David Held, Pieter Abbeel, and Anca D Dragan. Enabling robots to communicate their objectives. *Autonomous Robots*, 43(2):309–326, 2019.

[14] Julian Jara-Ettinger, Hyowon Gweon, Laura E Schulz, and Joshua B Tenenbaum. The naïve utility calculus: Computational principles underlying commonsense psychology. *Trends in cognitive sciences*, 20(8):589–604, 2016.

[15] Minae Kwon, Sandy H Huang, and Anca D Dragan. Expressing robot incapability. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 87–95, 2018.

[16] Christina Lichtenthäler and Alexandra Kirsch. Goal-predictability vs. trajectory-predictability: which legibility factor counts. In *Proceedings of the 2014 acm/ieee international conference on human-robot interaction*, pages 228–229, 2014.

[17] Christopher G Lucas, Thomas L Griffiths, Fei Xu, Christine Fawcett, Alison Gopnik, Tamar Kushnir, Lori Markson, and Jane Hu. The child as econometrician: A rational model of preference understanding in children. *PloS one*, 9(3):e92160, 2014.

[18] Jim Mainprice, Emrah Akin Sisbot, Thierry Siméon, and Rachid Alami. Planning safe and legible hand-over motions for human-robot interaction. In *IARP/IEEE-RAS/EURON workshop on technical challenges for dependable robots in human environments*, 2010.

[19] Daniel McFadden. Conditional logit analysis of qualitative choice behavior. In P. Zarembka, editor, *Frontiers of Econometrics*. New York: Academic Press, 1973.

[20] Lillian J Ratliff and Eric Mazumdar. Inverse risk-sensitive reinforcement learning. *IEEE Transactions on Automatic Control*, 65(3):1256–1263, 2019.

[21] Kyle Strabala, Min Kyung Lee, Anca Dragan, Jodi Forlizzi, Siddhartha S Srinivasa, Maya Cakmak, and Vincenzo Micelli. Toward seamless human-robot handovers. *Journal of Human-Robot Interaction*, 2(1):112–132, 2013.

[22] Daniel Szafir, Bilge Mutlu, and Terrence Fong. Communication of intent in assistive free flyers. In *Proceedings of the ACM/IEEE international conference on Human-Robot Interaction*, pages 358–365, 2014.

[23] Leila Takayama, Doug Dooley, and Wendy Ju. Expressing thought: improving robot readability with animation principles. In *Proceedings of the 6th international conference on Human-robot interaction*, pages 69–76, 2011.

[24] Yuval Tassa, Nicolas Mansard, and Emo Todorov. Control-limited differential dynamic programming. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1168–1175. IEEE, 2014.

[25] Emanuel Todorov and Weiwei Li. A generalized iterative lqg method for locally-optimal feedback control of constrained nonlinear stochastic systems. In *Proceedings of the 2005, American Control Conference, 2005.*, pages 300–306. IEEE, 2005.

[26] Michael Tomasello, Malinda Carpenter, Josep Call, Tanya Behne, and Henrike Moll. Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and brain sciences*, 28(5):675–691, 2005.

[27] Min Zhao, Rahul Shome, Isaac Yochelson, Kostas Bekris, and Eileen Kowler. An experimental study for identifying features of legible manipulator paths. In *Experimental robotics*, pages 639–653. Springer, 2016.

[28] Brian D Ziebart. *Modeling purposeful adaptive behavior with the principle of maximum causal entropy*. PhD thesis, Carnegie Mellon University, 2010.