Robust Control Under Uncertainty via Bounded Rationality and Differential Privacy

Vincent Pacelli and Anirudha Majumdar

***Abstract*— The rapid development of affordable and compact high-fidelity sensors (e.g., cameras and LIDAR) allows robots to construct detailed estimates of their states and environments. However, the availability of such rich sensor information introduces two challenges: (i) the lack of analytic sensing models, which makes it difficult to design controllers that are robust to sensor failures, and (ii) the computational expense of processing the high-dimensional sensor information in real time. This paper addresses these challenges using the theory of *differential privacy*, which allows us to (i) design controllers with bounded sensitivity to errors in state estimates, and (ii) bound the amount of state information used for control (i.e., to impose decision-making under *bounded rationality*). The resulting framework approximates the *separation principle* and allows us to derive an upper-bound on the cost incurred with a faulty state estimator in terms of three quantities: the cost incurred using a perfect state estimator, the magnitude of state estimation errors, and the level of differential privacy. We demonstrate the efficacy of our framework numerically on different robotics problems, including nonlinear system stabilization and motion planning.**

I. Introduction

Despite the increasing availability of high-resolution sensors for robotic systems, *partial-observability* remains a challenge for controlling such systems. Sensing modalities such as vision and LIDAR are ultimately noisy and only provide partial information about the robot's state and environment. In general, solving optimal control problems with partial observability is computationally intractable. One of the most common approaches to tackling this challenge is to assume the *separation principle* [1], i.e., to *independently* design (i) a state estimator, and (ii) a controller, e.g., one based on model-predictive control (MPC), that is optimal assuming perfect state estimation. The modularity afforded by such an approach coupled with the relative tractability of tackling the estimation and control problems independently make this framework appealing. However, the separation principle does not hold in general for robotic systems due to nonlinear dynamics / measurement models. A controller that assumes perfect state estimation can thus be highly sensitive to small errors in the state estimate, leading to significant brittleness of the overall control system. This is particularly challenging in the increasingly common case where a (deep) learning model is used as part of the robot's state estimation pipeline (due to potential over-fitting). As a result, robots often behave erratically when faced with unforeseen measurement errors despite possessing high-resolution sensors.

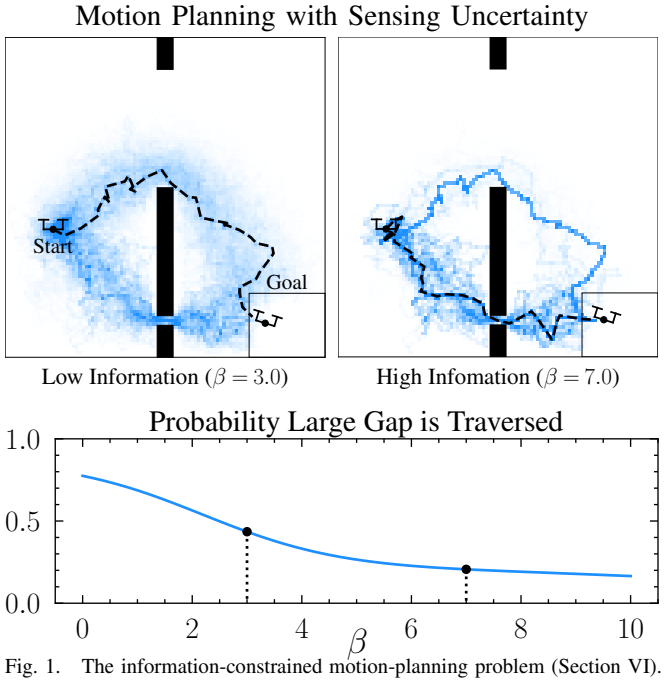

Fig. 1. The information-constrained motion-planning problem (Section VI). The robot must navigate from the start to the goal using a limited amount of state information (correlated with β). The distribution of trajectories is visualized for two values of β. At low information, the robot prefers the longer path through the wider gap, which is more robust to positional uncertainty than the direct route preferred at higher information usage.

In contrast, the cognitive science literature on *bounded rationality* — a framework for analyzing decision-making under computation or information constraints — demonstrates that humans display impressive levels of robustness and generalization in dexterous tasks such as locomotion and ball-catching without relying on highly-accurate state estimates [2, 3]. For example, *gaze heuristics* are control laws used by humans to catch freely-falling objects by adjusting their running speed based on the motion of the object in their visual field [4, 5]. Unlike MPC strategies, gaze heuristics do not require an accurate estimate of the system state or other quantities such as the wind speed or object mass. In addition to the robustness afforded by such heuristics, they are often extremely computationally efficient. Interestingly, the dual benefits of robustness and computational efficiency are related to each other in the above heuristics — boundedness of online computation (i.e., bounded rationality) ensures that control actions are only loosely coupled to sensor measurements. Such a loose coupling can prevent uncertainty in measurements from significantly impacting controller performance.

Statement of Contributions. The key conceptual contri-

The authors are affiliated with the Mechanical and Aerospace Engineering department of Princeton University, Princeton, NJ, 08544, USA {vpacelli, ani.majumdar}@princeton.edu.

This work is partially supported by the Office of Naval Research [N00014-21-1-2803] and the NSF CAREER award [2044149].

bution of this work is to use the observation that bounded rationality is linked to robust decision-making as a guide and formalize it using the theory of *differential privacy* (DP) [6–8], a framework that uses an algorithm's input-output sensitivity to quantify the information about an input gained by observing its output. Originally developed as a framework for ensuring the privacy of individuals' data (e.g., in a database with published statistics), DP allows formalization of the idea that control inputs should not depend too tightly on state estimates. In particular, our approach preserves the modular structure of the separation principle by ensuring that the mapping from state estimates to control inputs is differentially private. Moreover, it exploits the interpretation of the differentially private *exponential mechanism* as synthesizing *information-constrained* (i.e., bounded-rationality) controllers.

The primary theoretical contribution of this paper is a novel bound on the expected performance of such controllers in the presence of estimation error. The bound depends only on the cost incurred using a perfect state estimator, the magnitude of estimation errors, and the level of differential privacy. Moreover, due to the choice of the differentially-private mechanism on which our policy is based, our bound naturally applies to many popular inference-based control algorithms, such as optimal control methods based on importance sampling, Stein-variational gradient descent, and Q-learning. We are thus able to demonstrate the efficacy of the bounded-rationality approach we propose on a number of robotics problems of interest, such as stabilizing a nonlinear planar quadrotor and robust motion planning. To our knowledge, the approach presented in this paper is the first to utilize the framework of differential privacy to achieve robust and bounded-rational control of robotic systems.

II. Related Work

This section briefly reviews three areas of the literature that are typically considered independently from one another. However, the unifying theme is that they exploit various properties of the Gibbs measure.

A. Differential Privacy and Applications in Control Theory

Differential privacy (DP) [6–9] is an algorithmic framework that arose to meet the conflicting needs of statisticians (who ultimately want to publish analyses of sensitive data sets) and participants (who want to keep their data private). DP offers an elegant solution based on the intuition that if the results published from a study are insensitive to the substitution of data from any given individual, then this offers privacy. The literature on DP is vast and many differentially-private *mechanisms* (i.e., stochastic algorithms) have been proposed. The most relevant for our purpose is the *exponential mechanism* [8] defined by the Gibbs measure; this mechanism provides a method for releasing the results of an optimal decision-making procedure based on data (e.g., Bayesian inference [10]). Additionally, DP enjoys several appealing theoretical properties, namely the preservation of privacy under composition of mechanisms, and the ability to rigorously analyze the trade-off between privacy and accuracy [8].

In the context of control theory, application of the DP formalism is largely limited to networked control and distributed systems [11–16]. These include privately solving distributed optimization problems, modifying control inputs to keep the system state private from an external observer, and aggregating measurements from privacy-concerned agents for filtering performed by a central entity. In contrast, we *do not* use DP for the sake of privacy; instead, DP formalizes the intuition described in Section I: limiting the sensitivity of control inputs to state estimates affords robustness to estimation error. To our knowledge, our work is the first to utilize DP for this purpose.

B. Bounded Rationality and Robust Decision-Making

Bounded rationality is a model of decision-making first introduced to address misalignment between economic theory based on rational agents and the reality of sub-optimal human decision-making [17]. In short, while agents often have an objective they are trying to optimize, their ability to optimize this objective is bounded by informational and computational constraints. The cognitive science literature has identified several *heuristics* used by humans to deal with such constraints [2, 3, 18, 19]. Empirical work in cognitive science and robotics suggests that in addition to efficiency, such heuristics can provide generalization to new environments [2–4, 20, 21].

While bounded rationality is formalized in a number of ways for artificial agents, the most relevant framework adds an *information-theoretic constraint* to an otherwise rational agent. More precisely, a bounded-rational agent is modeled as solving an *entropy-maximization* or *rate-distortion* problem [22–27]. Despite empirical evidence that such a bounded-rational agent can be robust to uncertainty or noise in sensor measurements and dynamics [20, 21, 28–30], formal analyses justifying such robustness benefits are limited. One approach is to use a variational representation of the (relative) entropy (e.g., the Donsker-Varadhan representation [31]) to derive a zero-sum game that the bounded-rational agent is implicitly playing against an adversary that chooses a cost function [25, 26, 32]. This approach is closely related to the *maximum-entropy principle* from Bayesian statistics [33] and leads to generalization results for some estimation problems. However, the payoff of the game is difficult to interpret (especially for the adversary) in the context of most control problems, which limits the usefulness of the analysis. Alternatively, an upper bound on the performance degradation of a bounded-rational agent under measurement error exists [20] using variational representations of the entropy, but its assumptions make the bound difficult to interpret practically. Instead, this paper demonstrates that the connection between DP and bounded rationality through the Gibbs measure allows for the derivation of a bound on control performance in the presence of estimation error through a large-deviation analysis. The trade-offs presented by this bound are easily understood and, importantly, the bound only depends on access to simulation or lab data without actually deploying the robot in the target environment.

C. Optimal Control as Bayesian Inference

Another relevant vein of research at the intersection of information theory, Bayesian inference, and control theory is *linearly-solvable optimal control* (LSOC) [34–43]. These techniques exploit an identified equivalence between optimal control and Bayesian inference: an approximately optimal control sequence is found by sampling from a Gibbs measure over control inputs conditioned on the current state and the event that the sequence results in an optimal trajectory [36]. The result is that the nonlinear stochastic optimal control problem is solved as a linear, albeit infinite-dimensional, differential equation using a transformed cost function, and the solution is approximately computed using inference algorithms that include importance sampling [44], Stein-variational gradient descent (SVGD) [45], and Q-learning [46]. Despite the growing popularity of these algorithms for optimal control and empirical evidence for the generalization benefits they confer [47], theoretical knowledge of their robustness properties is limited [48]. This paper aims to fill this gap by providing a new, concrete analysis that these algorithms are provably robust to estimation error.

III. Notation

Random variables are denoted by uppercase letters (e.g., X), and realized quantities are denoted by lowercase letters (e.g., x). Deterministic functions appear in either case. Finite sequences are represented as $x_{i:j} = (x_k)_{k=i}^{j}$ for $i \leq j$. The indicator function for a set $\mathbf{A}$ is denoted $1_{\mathbf{A}}(\cdot)$. Functionals are double-struck, namely decorated varieties of the expectation $\mathbb{E}[\cdot]$, the relative entropy $\mathbb{D}[\cdot||\cdot]$, and the tightest Lipschitz constant of a scalar-valued function $\mathbb{L}[\cdot]$. Expectations are taken over the uppercase random variables (and mechanisms), e.g., in $\mathbb{E}[H(x, U)]$, x is fixed and integration is over U. Sets are in boldface and $\mathbf{\Delta}(\mathbf{A})$ is the set of distributions with support $\mathbf{A}$. For brevity, it is assumed that all necessary moments of random variables exist, and spaces are measurable with their subsets coming from appropriate σ-algebras.

A *mechanism* (denoted by uppercase script) refers to a (randomized) algorithm (formally, a transition kernel) between sets $\mathbf{X}$ and $\mathbf{Y}$. A mechanism $\mathcal{M} : \mathbf{X} \to \mathbf{\Delta}(\mathbf{Y})$ defines a probability distribution on $\mathbf{Y}$ for each $x \in \mathbf{X}$. Denote by $\mathcal{M}(x)\{\cdot\}$ the density and measure of this distribution when applied to elements and measurable subsets of $\mathbf{Y}$ respectively. When clear from context, we will overload this notation by treating $\mathcal{M}(x)$ as a random variable with support $\mathbf{Y}$. Mechanisms may be composed, i.e., if $\mathcal{M}' : \mathbf{Y} \to \mathbf{\Delta}(\mathbf{Z})$ is another mechanism, then $(\mathcal{M}' \circ \mathcal{M})(\cdot) \coloneqq \mathcal{M}'(\mathcal{M}(\cdot))$.

IV. Robust Single-Step Decision-Making

This section details the robustness properties that follow from applying a bounded-rationality approach to a single-step decision-making problem. Section V utilizes composition properties of DP to extend the analysis to multi-step optimal control problems.

A. Problem Statement

Let the state of the robotic system be $x \in \mathbf{X}$. The goal of the agent (robot) is to process the information contained in this state and select a control input $u \in \mathbf{U}$ that minimizes a cost $H(x, u)$. The state $X \sim \mathcal{X}$ is random and the agent must find a feedback mechanism $\mathcal{U} : \mathbf{X} \to \mathbf{\Delta}(\mathbf{X})$ that solves:

$$\min_{\mathcal{U}} \mathbb{J}^{\text{off}}[\mathcal{U}] \coloneqq \mathbb{E}[H(X, \mathcal{U}(X))]. \tag{OFF}$$

Since X is made available to the agent, this problem is fully-observable. It is referred to as the *offline problem* since it corresponds to, e.g., a lab or simulation setting where the agent makes arbitrarily fine measurements.

In practice, the primary concern is performance of the feedback controller on the *online problem*, where the agent only has access to a noisy state estimate $\hat{X} \sim \hat{\mathcal{X}}(X)$.[1] The control input $U \sim \mathcal{U}(\hat{X})$ is selected using this estimate, and the goal is to solve:

$$\min_{\mathcal{U}} \mathbb{J}^{\text{on}}[\mathcal{U}] \coloneqq \mathbb{J}^{\text{off}}[\mathcal{U} \circ \hat{\mathcal{X}}]. \tag{ON}$$

This is a *partially-observable* decision problem and its general solution requires reformulating the problem into an intractably high-dimensional (often infinite-dimensional) optimization problem [49, 50]. Instead, we will demonstrate that a *bounded-rational* controller allows for the value of (ON) to be bounded in terms of (OFF). That is, *performance on the online problem can be guaranteed using only information available offline* due to a property of the bounded rationality controller known as *differential privacy*. The difference in performance between these two problems is defined to be $\Delta\mathbb{J}[\mathcal{U}] \coloneqq \mathbb{J}^{\text{on}}[\mathcal{U}] - \mathbb{J}^{\text{off}}[\mathcal{U}]$, and the main contribution of this paper is to bound this quantity.

B. Differential Privacy

Differential privacy (DP) formalizes the observation that a mechanism does not reveal information about its input if the input may be replaced by a similar one without impacting the output distribution significantly. Specifically, *random metric DP* [9, 10, 51] encodes similarity via a metric[2] on the space of input data and is a natural fit for control applications where the input comes from a metric state space:

Definition (Differential Privacy)**.** Let $(\mathbf{X}, \rho)$ be a pseudometric space and $X, \hat{X}$ be two random variables supported on $\mathbf{X}$. A mechanism $\mathcal{U} : \mathbf{X} \to \mathbf{\Delta}(\mathbf{U})$ is said to have (ρ, γ)-*random differential privacy* ((ρ, γ)-DP) if, with probability $1 - \gamma$:

$$\forall u \in \mathbf{U}, \quad \log \mathcal{U}(X)\{u\} - \log \mathcal{U}(\hat{X})\{u\} \leq \rho(X, \hat{X}). \tag{1}$$

Intuitively, metric DP describes a kind of Lipschitz continuity[3] (in probability) and characterizes the sensitivity of mechanisms (stochastic controllers $\mathcal{U}$) to replacement of X with $\hat{X}$; specifically, the ratio of output densities is bounded

[1] Typically $\hat{\mathcal{X}}(x)$ is the composition of a state estimator with a noisy sensor, but these mappings are implementation dependent and irrelevant to the analysis. It is simpler and without loss of generality to work only with their composition.

[2] A metric suited for the control problem at hand may present itself (see Section VI for problems with quadratic costs), or it may simply be taken to be a Euclidean metric.

[3] Specifically, it is Lipschitz continuity with a specific metric on probability measures [9].

by $\exp(\rho(X, \hat{X}))$. An important aspect of DP is that it is preserved under composition of mechanisms [8]. Two such composition properties are:

Proposition 1 (Post-Processing)**.** If $\mathcal{M}_1 : \mathbf{X} \to \mathbf{\Delta}(\mathbf{U})$ is (ρ, γ)-DP, then for any $\mathcal{M}_2$, $\mathcal{M}_2 \circ \mathcal{M}_1$ is (ρ, γ)-DP.

Proposition 2 (Composition)**.** Let $\mathcal{M}_1, \mathcal{M}_2 : \mathbf{X} \to \mathbf{\Delta}(\mathbf{U})$ be (ρ_1, γ_1)-DP and (ρ_2, γ_2)-DP mechanisms respectively. Then $\mathcal{M} = (\mathcal{M}_1, \mathcal{M}_2)$ is (ρ, γ)-DP where $\rho(x, \hat{x}) \coloneqq \rho_1(x, \hat{x}) + \rho_2(x, \hat{x})$ and $\gamma \coloneqq \gamma_1 + \gamma_2$.

Proof. See the extended version of this paper [52]. □

These results allow for the recursive composition of system dynamics and private controllers to yield a mechanism that computes costs in a private manner; this will allow us to extend the robustness theorem proven for single-step decision-making in the next subsection to multi-step problems (Section V).

C. The Exponential Mechanism and Bounded Rationality

A popular mechanism for DP is the *exponential mechanism* [8], which provides privacy in problems where the solution of an optimization problem is desired as output. Thus, it is a natural starting point for designing a *differentially-private optimal control algorithm*. Specifically, the exponential mechanism $\mathcal{U}^\beta : \mathbf{X} \to \mathbf{\Delta}(\mathbf{U})$ is defined as the solution to the optimization problem:

$$\min_{\mathcal{U}} \ \mathbb{E}\left[H(X, U) + \beta^{-1}\mathbb{D}[\mathcal{U}(X)||\mathcal{U}^\perp]\right]. \qquad \text{(BR-SS)}$$

Here $\mathcal{U}^\perp \in \mathbf{\Delta}(\mathbf{U})$ is a "prior" supported on $\mathbf{U}$ independent of X (emphasized by the superscript $\perp$) and β^{-1} is interpreted as the Lagrange multiplier of a relative entropy constraint. Specifically, for each value of β^{-1}, there exists $d > 0$ such that $\mathcal{U}^\beta$ solves the constrained problem:

$$\min_{\mathcal{U}} \ \mathbb{E}\left[H(X, U)\right] \quad \text{s.t.} \quad \mathbb{E}[\mathbb{D}[\mathcal{U}(X)||\mathcal{U}^\perp]] \leq d.$$

The values of β^{-1} and d are linked in that increasing β^{-1} corresponds to an increase in d, but in general an exact relationship is elusive.

Remark 1. This is an instance of a *maximum-entropy problem*[4] from information theory [54–56], and corresponds to finding a mechanism that minimizes the expected cost without significant deviation from the prior (measured by the relative entropy). As $\beta \to \infty$, $\mathcal{U}^\beta(x)$ simply minimizes the expected cost. As $\beta \to 0$, the solution becomes the prior. Therefore, the mechanism can be viewed as an agent employing a *bounded-rational controller*, where it only uses a finite amount of information (due to computational or sensing constraints) about the state X to deviate from its default behavior specified by $\mathcal{U}^\perp$. Information usage (rationality) is directly controlled by β (inversely related to d).

Moreover, a prior that distributes weight evenly across its support may require a large value of β for $\mathcal{U}^\beta(x)$ to concentrate about the optimal control input. Similarly, a prior with a high concentration may require a large value of β and significant computational effort to produce solutions near the optimal control input. This latter case is a kind of exploration-exploitation trade-off due to the sampling schemes used by algorithmic implementations of the exponential mechanism.

The first-order optimality conditions for the problem (BR-SS) imply that the solution is a Gibbs measure [54–56]:

$$\mathcal{U}^\beta(x)\{u\} = \mathcal{U}^\perp\{u\}\exp(-\beta H(x, u)) \ / \ Z^\beta(x), \quad (2)$$
$$Z^\beta(x) \coloneqq \mathbb{E}^\perp[\exp(-\beta H(x, U))].$$

Here, $\mathbb{E}^\perp[\cdot]$ is the expectation computed using $\mathcal{U}^\perp$ as the controller. In the language from statistical mechanics, β is the *inverse temperature*, $H(x, u)$ is the *Hamiltonian* of the system, $Z^\beta(x)$ is the *partition function*, and,

$$F^\beta(x) \coloneqq -\beta^{-1}\log Z^\beta(x), \quad (3)$$

is known as the *free energy*. The latter is notably important for a number of reasons including it being a cumulant-generating function and its equality to the Lagrangian of (BR-SS) conditioned on $X = x$ (see the appendix of [52]):

$$F^\beta(x) = \mathbb{E}[H(x, U)] + \beta^{-1}\mathbb{D}[\mathcal{U}^\beta(x)||\mathcal{U}^\perp]. \quad (4)$$

Due to its popularity, the privacy properties of the exponential mechanism under different assumptions on $H(x, u)$ are well-studied. Only a simple assumption of Lipschitz continuity in x for each u is required [10]. However, this assumption will not hold for the applications of interest in Section VI due to $\mathbf{U}$ being non-compact. In these cases, random DP may still be achieved and is suitable for robust control:

Proposition 3. Consider the set of all $u \in \mathbf{U}$ for which $H(x, u)$ is at most l-Lipschitz in x, i.e. $\mathbf{U}(l) \coloneqq \{u \in \mathbf{U} |\ \mathbb{L}[x \mapsto H(x, u)] < l\}$. Then, $\mathcal{U}^\beta(x)$ is $(2\beta l\rho, \gamma(l))$-DP, where:

$$\gamma(l) \coloneqq 1 - \mathbb{E}\left[1_{\mathbf{U}(l)}(U)\exp\left(-2\beta l\rho(X, \hat{\mathfrak{X}}(X))\right)\right].$$

Proof. See the appendix of [52]. □

The proposition characterizes the trade-offs in selecting the free parameter l and prior $\mathcal{U}^\perp$. A larger value of l implies a larger set $\mathbf{U}(l)$ but a smaller region of integration in the definition of $\gamma(l)$. For very large values of l, the latter will dominate and the probability that privacy fails, which is $\gamma(l)$, becomes almost certain. Note that larger l implies a looser privacy constraint, since $2\beta l\rho(x, \hat{x})$ will grow for any fixed pair $x, \hat{x} \in \mathbf{X}$. The choice of prior $\mathcal{U}^\perp$ may also bias U toward regions that yield a smaller Lipschitz constant.

D. Quantifying the Robustness of Bounded Rationality

The key theoretical contribution of this paper is realizing that the proposed definition of random DP can quantify the performance of the bounded-rational agent (i.e., the private controller $\mathcal{U}^\beta$) when only a noisy estimate $\hat{X}$ of the state X is available — thereby approximating the separation principle. This idea is combined with a large deviations argument that exploits the similarity between the large-deviation rate function [57] and the free energy (3) to derive the theorem:

[4]The maximum-entropy problem (BR-SS) related to the *rate-distortion problems* common in bounded rationality models [20–22, 24] via a well-known variational relationship between the relative entropy and mutual information [53]. All subsequent analysis applies to these problems as well.

Theorem 1. Define $\rho_\beta(x, \hat{x}) \coloneqq 2\beta l \rho(x, \hat{x})$. With probability at least $1 - \gamma(l)$:[5]

$$\Delta\mathbb{J}[\mathcal{U}^\beta] \leq \beta^{-1}\mathbb{E}\left[\exp\left(\rho_\beta(X, \hat{X}) + \mathbb{D}[\mathcal{U}^\beta(X)||\mathcal{U}^\perp]\right)\right].$$

Proof. See the appendix of [52] for proof. □

This theorem bounds the gap $\Delta\mathbb{J}[\mathcal{U}^\beta] \coloneqq \mathbb{J}^{\text{on}}[\mathcal{U}^\beta] - \mathbb{J}^{\text{off}}[\mathcal{U}^\beta]$ in performance between the offline and online problems in terms of β, which describes the information usage of the controller. Importantly, it relates three separate quantities: the offline expected cost (which appears in $\Delta\mathbb{J}[\mathcal{U}^\beta]$), information usage (set by β), and the quality of the state estimator as measured by $\rho(x, \hat{x})$. Critically, all terms depend *only on offline information* — that is, the state estimator is not used for feedback in any of these terms but only to measure its error. Therefore, both elements of the feedback system may be designed and evaluated independently through this bound (similar to methodologies that adopt the separation principle). Moreover, the bound can be optimized to find $\beta^\star$ that yields the tightest bound on $\mathbb{J}^{\text{on}}[\mathcal{U}^\beta]$.

Remark 2. In cases where $\mathbb{D}[\mathcal{U}^\beta(X)||\mathcal{U}^\perp]$ cannot be evaluated, tractable bounds may be available depending on the context (e.g. Gibbs inequality [54], log-Sobolev inequalities [55, 56], and variational methods [58]).

V. Extension to Multi-Step Problems

In the multi-step optimal control problem, the agent attempts to minimize a sequence of non-negative cost functions $c_0, \ldots, c_{t_f-1} : \mathbf{X} \times \mathbf{U} \to \mathbf{R}_+, c_{t_f} : \mathbf{X} \to \mathbf{R}_+$ over a time horizon t_f by selecting a feedback controller $U_t \sim \mathcal{U}_t(X_t)$ that accounts for the system dynamics, $X_{t+1} \sim \mathcal{F}_t(x_t, u_t)$. The state-input trajectories of the system generated by the choice of controller $\mathcal{U}_{0:t_f}$ is denoted $\mathcal{T}[\mathcal{U}_{0:t_f}]$ with $\mathcal{T}^\perp \coloneqq \mathcal{T}[\mathcal{U}^\perp_{0:t_f}]$ being the prior trajectory distribution.

Introduce the shorthand for trajectory cost:

$$c_{0:t_f}(x_{0:t_f}, u_{0:t_f}) \coloneqq c_0(x_0, u_0) + \cdots + c_{t_f}(x_{t_f}).$$

Adapting the single-step problem notation, the offline and online problems are written,

$$\min_{\mathcal{U}_{0:t_f}} \mathbb{J}^{\text{off}}[\mathcal{U}_{0:t_f}] \coloneqq \mathbb{E}[c_{0:t_f}(X_{0:t_f}, U_{0:t_f})],$$
$$\min_{\mathcal{U}_{0:t_f}} \mathbb{J}^{\text{on}}[\mathcal{U}_{0:t_f}] \coloneqq \mathbb{J}^{\text{off}}[(\mathcal{U} \circ \hat{\mathfrak{X}})_{0:t_f}],$$

where $\hat{\mathfrak{X}}_{0:t_f}$ are mechanisms that introduce estimation error.

Bounded rationality is achieved in a similar manner to (BR-SS) by constraining the relative entropy between the sequence of closed-loop and open-loop controllers:

$$\min_{\mathcal{U}_{0:t_f}} \mathbb{J}^{\text{off}}[\mathcal{U}_{0:t_f}] \quad \text{s.t.} \quad \mathbb{D}[\mathcal{T}[\mathcal{U}_{0:t_f}]||\mathcal{T}^\perp] \leq d_{\text{traj}}. \quad \text{(BR-MS)}$$

This optimal control problem admits a recursive solution,[6]

$$\mathcal{U}^\beta_t(x) \coloneqq \arg\min_{\mathcal{U}_t} \ \mathbb{E}\left[H_t(X, U)\right] + \beta^{-1}\mathbb{D}[\mathcal{U}(x)||\mathcal{U}^\perp_t],$$

where β is the Lagrange multiplier corresponding to the entropy constraint. Then, $\mathcal{U}^\beta_t$ is an exponential mechanism and both the Hamiltonian and value function $V_t(x)$ are given by the recursive equations,

$$H_t(x, u) \coloneqq c_t(x, u) + \mathbb{E}[V_{t+1}(\mathcal{F}_t(x, u))],$$
$$V_t(x) \coloneqq \mathbb{E}[H_t(x, U_t)],$$

where $V_{t_f}(x) \coloneqq c_{t_f}(x)$. Notably, the free energy is equivalent to the value function: $F^\beta_t(x) = V_t(x)$. The aforementioned DP composition properties allow for extension of Proposition 3, and subsequently Theorem 1, to the multi-step problem with changes made *mutatis mutandis*.

Proposition 4. Let $\mathbf{U}_t(l_t) \coloneqq \{u \mid \mathbb{L}[x \mapsto H_t(x, u)] < l_t\}$. The mechanism,

$$\mathcal{M}(x_{0:t_f}) \coloneqq (\mathcal{U}^\beta_1(x_1), \ldots, \mathcal{U}^\beta_{t_f-1}(x_{t_f-1})),$$

is (ρ_β, γ)-DP where,

$$\rho_\beta(x_{0:t_f}, \hat{x}_{0:t_f}) \coloneqq \sum_{t=0}^{t_f-1} 2\beta l_t \rho(x_t, \hat{x}_t), \qquad \gamma \coloneqq \sum_{t=0}^{t_f-1} \gamma_t,$$

and,

$$\gamma_t \coloneqq 1 - \mathbb{E}\left[1_{\mathbf{U}_t(l_t)}(U_t) \exp\left(-2\beta l_t \rho(X_t, \hat{\mathfrak{X}}_t(X_t))\right)\right].$$

Theorem 2. Let $\rho_\beta(x, \hat{x})$ and γ be as in Proposition 4 and $\mathcal{T}^\beta \coloneqq [\mathcal{U}_{0:t_f}]$. With probability at least $1 - \gamma(l_t)$,

$$\Delta\mathbb{J}[\mathcal{U}^\beta_{0:t_f}] \leq \frac{1}{\beta}\ \mathbb{E}\left[\exp\left(\rho_\beta(X_{0:t_f}, \hat{X}_{0:t_f}) + \mathbb{D}[\mathcal{T}^\beta||\mathcal{T}^\perp]\right)\right].$$

Proof. See the appendix of [52] for proofs. □

This result bounds the cumulative online cost in terms of the offline cost, level of information usage (set by β), and estimation error (measured by ρ) similarly to Theorem 1.

VI. Numerical Examples

To demonstrate the efficacy of these robustness results, three numerical examples are presented in this section: a motion planning "double-slit" experiment and optimal control of both the linearized and nonlinear planar quadrotors.[7] In each case, the dynamics of the robot are chosen to be deterministic. While the developed theory permits stochastic dynamics, deterministic dynamics emphasizes the sources of randomness with which this article is focused: measurement error and bounded-rational control policies.

A. Motion Planning

Scenario. In this problem, the objective of the robot is to find the shortest possible path to some goal region $\mathbf{X}_g \subset \mathbf{X}$ while avoiding a set of obstacle configurations $\mathbf{X}_o \subset \mathbf{X}$ that is disjoint from $\mathbf{X}_g$. Both $\mathbf{X}_g$ and $\mathbf{X}_o$ are absorbing sets. For simplicity, the dynamics are the single integrator in the plane: $x_{t+1} = x_t + u_t$ with $\mathbf{X}, \mathbf{U} = [0, 1]^2$. The problem cost is the path length if the robot reaches the goal before intersecting with an obstacle and ∞ otherwise. The prior $\mathcal{U}^\perp_{0:t_f}$ is chosen to be a zero-mean Gaussian distribution.

The specific work space navigated by the robot is shown in Fig. 1. The region $\mathbf{X}_o$ consists of the boundary of the shown

[5] Expectations in the theorem are conditioned on the event $U \in \mathbf{U}(l)$, which is why the consequent occurs with probability $\gamma(l)$.

[6] See, e.g. [20, 24, 27, 41, 48], for detailed solutions to similar problems.

[7] The source code implementing these examples using JAX [59] and all experimental parameters are listed in a publicly available repository: `https://github.com/irom-lab/br-dp-robust`.

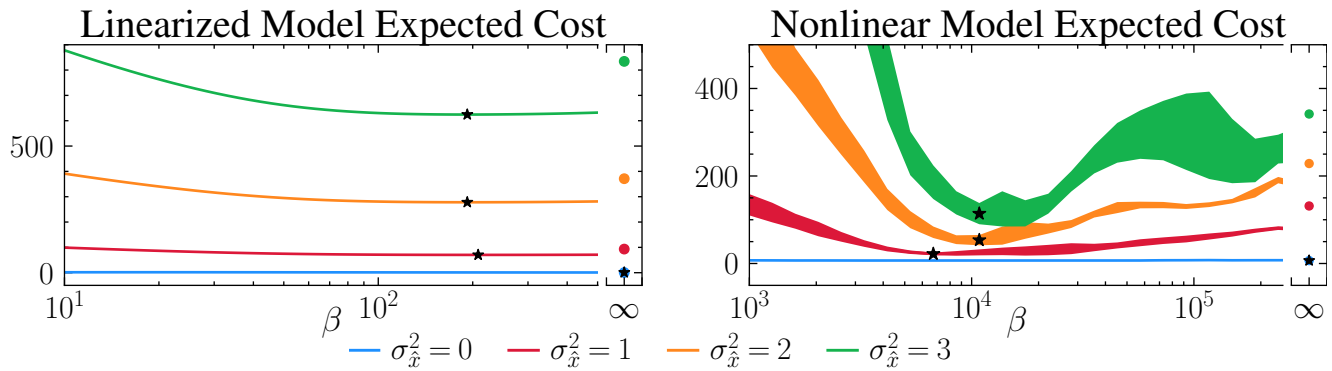

Fig. 2. Results of the numerical experiments described in Section VI. Lines indicate the expected cost of stabilizing a linearized / nonlinear quadrotor model for a range of β and $\beta = \infty$. The latter case corresponds to the performance of the LQR / MPC controllers, which are performant given perfect state information. The $\star$ marker indicates $\beta^\star$ found for each value of $\sigma^2_{\hat{x}}$. For the nonlinear example, shading indicates standard deviations for 100 trials of SV-MPC. The bounded-rationality controllers outperform LQR / MPC in the presence of estimation error.

space and a divider punctured with a large and small slit. A path through the large slit is more robust to measurement error while a path through the smaller one is less forgiving. The goal region is in the lower right. Similar experiments are present in the literature [30, 34, 35, 43].

Results. Importance sampling [44] is used to implement the bounded rationality controller. As shown in Fig. 1, as the information constraint tightens ($\beta \to 0$), the agent shifts from navigating the direct-but-treacherous gap to the robust (indirect) route. This demonstrates the robustness benefits of our differentially private control scheme.

B. Planar Quadrotor Stabilization Problems

Scenario. These examples focus on the stabilization of a planar quadrotor (constrained to the Y-Z plane and rotating about the X axis). The system state $x \in \mathbf{R}^6$ consists of the position and rotation of the robot and the time derivatives of these variables, i.e. $x = (y, z, \theta, \dot{y}, \dot{z}, \dot{\theta})$, while the control input $u \in \mathbf{R}^2$ is the thrust exerted by two opposing pairs of rotors (see [60, Section 6.6] for dynamics). The cost functions are the linear-quadratic regulator (LQR) cost functions [1]. By choosing the metric,

$$\rho(x, \hat{x}) = \frac{1}{2} \left\| xx^{\mathrm{T}} - \hat{x}\hat{x}^{\mathrm{T}} \right\|_{\mathrm{F}} + \|x - \hat{x}\|_2 ,$$

the Hamiltonian $H_t(x, u)$ satisfies the conditions for the theory from Section V to apply to both systems.

The system parameters are chosen to align with the Crazyflie 2.0 quadrotors: the mass is $0.03\,\mathrm{kg}$ and the moment of inertia is $1.43 \times 10^{-5}\,\mathrm{kg\,m^2}$ [61]. The system is temporally discretized with a time step of $\Delta t = 0.3\,\mathrm{s}$ and $t_f = 13$. The initial condition is sampled from small Gaussian perturbations about the hover state given according to $\mathcal{N}(\bar{x}_0, \sigma^2_x)$ where $\bar{y}_0 = 1\,\mathrm{m}, \bar{z}_0 = -1\,\mathrm{m}$, the remaining mean entries are zero, and $\sigma^2_{x_0} = (10^{-2}, 10^{-2}, 10^{-6}, 10^{-4}, 10^{-4}, 10^{-8})$. The open-loop prior is chosen by projecting the initial distribution through the LQR solution to find the marginal input. The estimation error is drawn from a stationary Gaussian distribution. Specifically, $\hat{\mathcal{X}}(x) = \mathcal{N}(x, \sigma^2_{\hat{x}} v)$, where $\sigma^2_{\hat{x}} \in \mathbf{R}_+$ is a scale parameter and $v = \mathrm{diag}(0.25, 0.25, 0.1, 0.25, 0.25, 0.1)$.

Linearized Results. The system is linearized and the dynamic programming equations specifying the control policy are solved exactly for a feedback policy that is linear in the state with additive Gaussian noise. Details are included in the appendix of [52]. The performance of the bounded-rationality controller is evaluated and compared with a controller that is optimal assuming perfect state estimation (LQR) paired with a sub-optimal state estimator. As shown in Fig. 2, the bounded-rationality controller is more robust to estimation error than the LQR controller. The optimal information usage $\beta^\star$ decreases with the uncertainty $\sigma^2_{\hat{x}}$.

Nonlinear Results. The experiments were then repeated using the nonlinear planar quadrotor dynamics. There is no closed-form solution to (BR-MS), but numerical methods for sampling from $\mathcal{U}^\beta_t(x)$ are available (see Section II-C). The SV-MPC algorithm [41] is chosen due to its increased efficiency compared to importance sampling and the fact that it reduces to gradient-based optimization of the trajectory with multiple random initializations in the $\beta \to \infty$ limit, which is a common MPC algorithm. The results are similar to the linear case: there is a prominent local minimum for β that outperforms MPC (which is optimal assuming perfect estimation) in the presence of estimation error when averaged over 100 trials. In this case, the monotonic relationship between $\beta^\star$ and $\sigma^2_{\hat{x}}$ is not seen — possibly due to SV-MPC only approximating the controller's distribution.

VII. Conclusion

This paper proposes new theoretical justifications for the robustness of bounded-rational control policies using differential privacy. The stated performance guarantee for such policies has a modular structure reminiscent of the separation principle. Multiple numerical simulations demonstrate that using DP to create controllers provides robustness to estimation error.

Future Work. There remain a number of useful properties that need to be determined about the stated bounds that extend beyond the scope of this paper, e.g., the tightness of the bounds and whether the relationship between $\beta^\star$ and $\sigma^2_{\hat{x}}$ is monotonic as suggested by the results for the linearized system in Fig. 2. Developing tractable method to evaluate the robustness bounds is also of great interest. Opportunities to extend the experimental results of the article to new applications, such as sim-to-real transfer, and adapting other mechanisms from the DP literature to robust controller design may be considered.

REFERENCES

[1] B. Anderson and J. Moore, *Optimal Control: Linear Quadratic Methods*. Courier Corporation, 2007.

[2] G. Gigerenzer and H. Brighton, "Homo heuristicus: Why biased minds make better inferences," *Topics in Cognitive Science*, vol. 1, no. 1, pp. 107–143, 2009.

[3] G. Gigerenzer and W. Gaissmaier, "Heuristic decision making," *Annual Review of Psychology*, vol. 62, pp. 451–482, 2011.

[4] S. Höfer, J. Raisch, *et al.*, "No free lunch in ball catching: A comparison of cartesian and angular representations for control," *PloS One*, vol. 13, no. 6, 2018.

[5] B. Belousov, G. Neumann, *et al.*, "Catching heuristics are optimal control policies," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2016, pp. 1426–1434.

[6] C. Dwork, F. McSherry, *et al.*, "Calibrating noise to sensitivity in private data analysis," in *Proceedings of the Theory of Cryptography Conference*, Springer, 2006, pp. 265–284.

[7] C. Dwork, "Differential privacy," *Encyclopedia of Cryptography and Security*, pp. 338–340, 2011.

[8] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, no. 3-4, pp. 211–407, 2014.

[9] K. Chatzikokolakis, M. E. Andrés, *et al.*, "Broadening the scope of differential privacy using metrics," in *Proceedings of the International Symposium on Privacy Enhancing Technologies*, Springer, 2013, pp. 82–102.

[10] C. Dimitrakakis, B. Nelson, *et al.*, "Differential privacy for Bayesian inference through posterior sampling," *Journal of Machine Learning Research*, vol. 18, no. 11, pp. 1–39, 2017.

[11] J. Le Ny and G. J. Pappas, "Differentially private filtering," *IEEE Transactions on Automatic Control*, vol. 59, no. 2, pp. 341–354, 2013.

[12] J. Cortés, G. E. Dullerud, *et al.*, "Differential privacy in control and network systems," in *Proceedings of the Conference on Decision and Control (CDC)*, IEEE, 2016, pp. 4252–4272.

[13] E. Nozari, P. Tallapragada, *et al.*, "Differentially private distributed convex optimization via objective perturbation," in *Proceedings of the American Control Conference (ACC)*, IEEE, 2016, pp. 2061–2066.

[14] S. Han, U. Topcu, *et al.*, "Differentially private distributed constrained optimization," *IEEE Transactions on Automatic Control*, vol. 62, no. 1, pp. 50–64, 2016.

[15] F. Koufogiannis and G. J. Pappas, "Differential privacy for dynamical sensitive data," in *Proceedings of the Conference on Decision and Control (CDC)*, IEEE, 2017, pp. 1118–1125.

[16] S. Han and G. J. Pappas, "Privacy in control and dynamical systems," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, pp. 309–332, 2018.

[17] H. A. Simon, "A behavioral model of rational choice," *The Quarterly Journal of Economics*, vol. 69, no. 1, pp. 99–118, 1955.

[18] W. H. Warren Jr., D. S. Young, *et al.*, "Visual control of step length during running over irregular terrain," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 12, no. 3, p. 259, 1986.

[19] D. N. Lee, "A theory of visual control of braking based on information about time-to-collision," *Perception*, vol. 5, no. 4, pp. 437–459, 1976.

[20] V. Pacelli and A. Majumdar, "Task-driven estimation and control via information bottlenecks," in *Proceedings of the International Conference on Robotics and Automation (ICRA)*, IEEE, 2019, pp. 2061–2067.

[21] ——, "Learning task-driven control policies via information bottlenecks," in *Proceedings of Robotics: System and Science (RSS)*, 2020.

[22] N. Tishby and D. Polani, "Information theory of decisions and actions," in *Perception-Action Cycle: Models, Architectures, and Hardware*, Springer, 2011, pp. 601–636.

[23] R. Fox and N. Tishby, "Minimum-information lqg control part I: Memoryless controllers," in *Proceedings of the Conference on Decision and Control (CDC)*, IEEE, 2016, pp. 5610–5616.

[24] ——, "Bounded planning in passive POMDPs," in *Proceedings of the International Conference on Machine Learning (ICML)*, 2012.

[25] P. A. Ortega and D. D. Lee, "An adversarial interpretation of information-theoretic bounded rationality," in *Proceedings of the Conference on Artificial Intelligence*, 2014.

[26] P. A. Ortega, D. A. Braun, *et al.*, "Information-theoretic bounded rationality," *arXiv preprint arXiv:1512.06789*, 2015.

[27] E. Shafieepoorfard, M. Raginsky, *et al.*, "Rationally inattentive control of Markov processes," *SIAM Journal on Control and Optimization*, vol. 54, no. 2, pp. 987–1016, 2016.

[28] M. Igl, K. Ciosek, *et al.*, "Generalization in reinforcement learning with selective noise injection and information bottleneck," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 32, pp. 13 978–13 990, 2019.

[29] X. Lu, K. Lee, *et al.*, "Dynamics generalization via information bottleneck in deep reinforcement learning," *arXiv preprint arXiv:2008.00614*, 2020.

[30] A. R. Pedram, J. Stefarr, *et al.*, "Rationally inattentive path-planning via RRT," in *Proceedings of the American Control Conference (ACC)*, IEEE, 2021, pp. 3440–3446.

[31] M. D. Donsker and S. R. S. Varadhan, "Asymptotic evaluation of certain Markov process expectations for large time. IV," *Communications on Pure and Applied Mathematics*, vol. 36, no. 2, pp. 183–212, 1983.

[32] B. Eysenbach and S. Levine, "Maximum entropy RL (provably) solves some robust RL problems," *arXiv preprint arXiv:2103.06257*, 2021.

[33] P. D. Grünwald and A. P. Dawid, "Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory," *Annals of Statistics*, vol. 32, no. 4, pp. 1367–1433, 2004.

[34] H. J. Kappen, "Linear theory for control of nonlinear stochastic systems," *Physical Review Letters*, vol. 95, no. 20, p. 200 201, 2005.

[35] ——, "Path integrals and symmetry breaking for optimal control theory," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2005, no. 11, 2005.

[36] E. Todorov, "General duality between optimal control and estimation," in *Proceedings of the Conference on Decision and Control (CDC)*, IEEE, 2008, pp. 4286–4292.

[37] ——, "Efficient computation of optimal actions," *Proceedings of the National Academy of Sciences*, vol. 106, no. 28, pp. 11 478–11 483, 2009.

[38] G. Williams, N. Wagener, *et al.*, "Information theoretic MPC for model-based reinforcement learning," in *Proceedings of the International Conference on Robotics and Automation (ICRA)*, IEEE, 2017, pp. 1714–1721.

[39] E. A. Theodorou and E. Todorov, "Relative entropy and free energy dualities: Connections to path integral and KL control," in *Proceedings of the Conference on Decision and Control (CDC)*, IEEE, 2012, pp. 1466–1473.

[40] D. A. Braun, P. A. Ortega, *et al.*, "Path integral control and bounded rationality," in *Proceedings of the Symposium on Adaptive Dynamic Programming and Reinforcement Learning*, IEEE, 2011, pp. 202–209.

[41] A. Lambert, A. Fishman, *et al.*, "Stein variational model predictive control," in *Proceedings of the Conference on Robot Learning*, 2020.

[42] L. Barcelos, A. Lambert, *et al.*, "Dual online Stein variational inference for control and dynamics," in *Proceedings of Robotics: System and Science (RSS)*, 2021.

[43] M. B. Horowitz and J. W. Burdick, "Optimal navigation functions for nonlinear stochastic systems," in *Proceedings of the International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2014, pp. 224–231.

[44] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.

[45] Q. Liu and D. Wang, "Stein variational gradient descent: A general purpose Bayesian inference algorithm," in *Advances in Neural Information Processing Systems*, 2016.

[46] T. Haarnoja, H. Tang, *et al.*, "Reinforcement learning with deep energy-based policies," in *Proceedings of the International Conference on Machine Learning (ICML)*, 2017, pp. 1352–1361.

[47] T. Haarnoja, V. Pong, *et al.*, "Composable deep reinforcement learning for robotic manipulation," in *Proceedings of the International Conference on Robotics and Automation (ICRA)*, IEEE, 2018, pp. 6244–6251.

[48] S. Levine, "Reinforcement learning and control as probabilistic inference: Tutorial and review," *arXiv preprint arXiv:1805.00909*, 2018.

[49] D. P. Bertsekas, *Dynamic Programming and Optimal Control*. Athena Scientific, 2017.

[50] M. R. James and J. S. Baras, "Partially observed differential games, infinite-dimensional Hamilton-Jacobi-Isaacs equations, and nonlinear H_∞ control," *SIAM Journal on Control and Optimization*, vol. 34, no. 4, pp. 1342–1364, 1996.

[51] R. Hall, A. Rinaldo, *et al.*, "Random differential privacy," *Journal of Privacy and Confidentiality*, vol. 4, no. 2, pp. 43–59, 2012.

[52] V. Pacelli and A. Majumdar, "Robust control under uncertainty via bounded rationality and differential privacy," *arXiv preprint arXiv:2109.08262*, 2021.

[53] X. Chen, A. Guntuboyina, *et al.*, "On Bayes risk lower bounds," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 7687–7744, 2016.

[54] T. M. Cover, *Elements of Information Theory*. John Wiley & Sons, 1999.

[55] M. Raginsky and I. Sason, "Concentration of measure inequalities in information theory, communications, and coding," *Foundations and Trends in Communications and Information Theory*, vol. 10, no. 1-2, pp. 1–247, 2013.

[56] A. Maurer, "Thermodynamics and concentration," *Bernoulli*, vol. 18, no. 2, pp. 434–454, 2012.

[57] H. Touchette, "A basic introduction to large deviations: Theory, applications, simulations," *arXiv preprint arXiv:1106.4146*, 2011.

[58] B. Poole, S. Ozair, *et al.*, "On variational bounds of mutual information," in *Proceedings of the International Conference on Machine Learning (ICML)*, ser. Proceedings of Machine Learning Research, vol. 97, PMLR, 2019, pp. 5171–5180.

[59] J. Bradbury, R. Frostig, *et al.*, *JAX: Composable transformations of Python+NumPy programs*, 2018.

[60] J. Steinhardt and R. Tedrake, "Finite-time regional verification of stochastic non-linear systems," *International Journal of Robotics Research*, vol. 31, no. 7, pp. 901–923, 2012.

[61] J. Förster, "System identification of the Crazyflie 2.0 nano quadrocopter," B.S. Thesis, ETH Zurich, 2015.