Journal Title
XX(X):1–27
©The Author(s) 2019
Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

SACBP: Belief Space Planning for Continuous-Time Dynamical Systems via Stochastic Sequential Action Control

Haruki Nishimura[1] **and Mac Schwager**[1]

Abstract

We propose a novel belief space planning technique for continuous dynamics by viewing the belief system as a hybrid dynamical system with time-driven switching. Our approach is based on the perturbation theory of differential equations and extends Sequential Action Control Ansari and Murphey (2016) to stochastic dynamics. The resulting algorithm, which we name SACBP, does not require discretization of spaces or time and synthesizes control signals in near real-time. SACBP is an anytime algorithm that can handle general parametric Bayesian filters under certain assumptions. We demonstrate the effectiveness of our approach in an active sensing scenario and a model-based Bayesian reinforcement learning problem. In these challenging problems, we show that the algorithm significantly outperforms other existing solution techniques including approximate dynamic programming and local trajectory optimization.

Keywords

Belief Space Planning, Active Sensing, Mobile Robots, Optimization and Optimal Control, Probabilistic Reasoning, Vision and Sensor-based Control

1 Introduction

Planning under uncertainty still remains as a challenge for robotic systems. Various types of uncertainty, including unmodeled dynamics, stochastic disturbances, and imperfect sensing, significantly complicate problems that are otherwise easy. For example, suppose that a robot needs to manipulate an object from some initial state to a desired goal. If the mass properties of the object are not known beforehand, the robot needs to simultaneously estimate these parameters and perform control, while taking into account the effects of their uncertainty; the exploration and exploitation trade-off needs to be resolved Slade et al. (2017). On the other hand, uncertainty is quite fundamental in motivating some problems. For instance, a noisy sensor may encourage the robot to carefully plan a trajectory so the observations taken along it are sufficiently informative. This type of problem concerns pure information gathering and is often referred to as active sensing Mihaylova et al. (2002), active perception Bajcsy (1988), or informative motion planning Hollinger and Sukhatme (2014).

A principled approach to address all those problems is to form plans in the belief space, where the planner chooses sequential control inputs based on the evolution of the belief state. This approach enables the robot to appropriately execute controls under stochasticity and partial observability since they are both incorporated into the belief state. Belief space planning is also well suited for generating information gathering actions Platt et al. (2010).

This paper proposes a novel online belief space planning algorithm. It does not require discretization of the state space or the action space, and can directly handle continuous-time system dynamics. The algorithm optimizes the expected value of a first-order cost reduction with respect to a nominal control policy at every re-planning time, proceeding in a receding-horizon fashion. We are inspired by the Sequential Action Control (SAC) algorithm recently proposed in Ansari and Murphey (2016) for model-based deterministic optimal control problems. SAC is an online method to synthesize control signals in real time for challenging (but deterministic) physical systems such as a cart pendulum and a spring-loaded inverted pendulum. Based on the concept of SAC, this paper develops an algorithmic framework to control stochastic belief systems whose dynamics are governed by parametric Bayesian filters.

1.1 Related Work in Belief Space Planning

There is a large body of literature in belief space planning. Below we briefly review relevant work in three major types of solution methods: greedy strategies, trajectory optimization methods, and belief MDP and POMDP approaches. We then change our perspective and discuss advantages and drawbacks of closed-loop and open-loop planning schemes, while also introducing other relevant work that do not necessarily fall into the three categories mentioned above.

Greedy Strategies Belief space planning is known to be challenging for a couple of reasons. First, the belief state is continuous and can be high-dimensional even if the underlying state space is small or discrete. Second, the dynamics that govern the belief state transitions are stochastic due to unknown future observations. Greedy

[1] Stanford University, USA

Corresponding author:

Haruki Nishimura, Department of Aeronautics and Astronautics, Stanford University, Stanford, CA 94305, USA.

Email: hnishimura@stanford.edu

approaches alleviate the complexity by ignoring long-term effects and solve single-shot decision making problems sequentially. Despite their suboptimality for long-term planning, these methods are often employed to find computationally tractable solutions and achieve reasonable performance in different problems Bourgault et al. (2002); Seekircher et al. (2011); Schwager et al. (2017), especially in the active sensing domain.

Trajectory Optimization Methods In contrast to the greedy approaches, trajectory optimization methods take into account multiple timesteps at once and find non-myopic solutions. In doing so, it is often assumed that the maximum likelihood observation (MLO) will always occur during planning Platt et al. (2010); Erez and Smart (2010); Patil et al. (2014). This heuristic assumption yields a deterministic optimal control problem, to which various nonlinear trajectory optimization algorithms are applicable. However, ignoring the effects of stochastic future observations can degrade the performance van den Berg et al. (2012). Other methods van den Berg et al. (2012); Rafieisakhaei et al. (2017) that do not rely on the MLO assumption are advantageous in that regard. In particular, belief iLQG van den Berg et al. (2012) performs iterative local optimization in a Gaussian belief space by quadratically approximating the value function and linearizing the dynamics to obtain a time-varying affine feedback policy. However, this method as well as many other solution techniques in this category result in multiple iterations of intensive computation and can require a significant amount of time until convergence.

Belief MDP and POMDP Approaches Belief space planning can be modeled as a Markov decision process (MDP) in the belief space, given that the belief state transition is Markovian. If the reward (or cost) is defined as an explicit function of the state and the control, the problem is equivalent to a partially observable Markov decision process (POMDP) Kaelbling et al. (1998). A key challenge in POMDPs and belief MDPs has been to address problems with large state spaces. This is particularly important in belief MDPs since the state space for a belief MDP is a continuous belief space. To handle continuous spaces, Couëtoux et al. (2011) introduce double progressive widening (DPW) for Monte Carlo Tree Search (MCTS) Browne et al. (2012). In Slade et al. (2017), this MCTS-DPW algorithm is run in a Gaussian belief space to solve the object manipulation problem mentioned earlier. We have also presented a motion-based communication algorithm in our prior work, which uses MCTS-DPW for active intent inference with monocular vision Nishimura and Schwager (2018a).

While MCTS-DPW as well as other general purpose POMDP methods Somani et al. (2013); Sunberg and Kochenderfer (2017) are capable of handling continuous state spaces, their algorithmic concepts are rooted in dynamic programming and tree search, requiring a sufficient amount of exploration in the tree. The tree search technique also implicitly assumes discrete-time transition models. In fact, most prior works reviewed in this section are intended for discrete-time systems. A notable exception is Chaudhari et al. (2013), in which a sequence of discrete-time POMDP approximations is constructed such that it converges to a true continuous-time model. However, this approach still relies on an existing POMDP solver Kurniawati et al. (2008) that is designed for discrete systems. There still remains a need for an efficient and high-performance belief space planning algorithm that is capable of directly handling systems with inherently continuous-space, continuous-time dynamics, such as maneuvering micro-aerial vehicles, or autonomous cars at freeway speeds.

Closed-loop and Open-loop Planning An important aspect of belief space planning is the stochastic belief dynamics, which essentially demand that one plan over closed-loop (i.e. feedback) policies. This is a primary assumption in belief MDP and POMDP frameworks where an optimal mapping from beliefs to control actions are sought Huber (2009); Kochenderfer (2015). However, computing exact optimal policies in general is intractable due to curse of dimensionality and curse of history Papadimitriou and Tsitsiklis (1987); Madani et al. (1999); Somani et al. (2013). Therefore, in practice we need to make certain approximations to achieve a tractable solution within a reasonable computation time, whether the solution method is online or offline. For POMDP methods, those approximations are often done by discretization and/or sampling Kurniawati et al. (2008); Chaudhari et al. (2013); Somani et al. (2013); Sunberg and Kochenderfer (2017). Belief iLQG van den Berg et al. (2012) makes an approximation by restricting the policy to a time-varying affine feedback controller and performing local optimization. Other methods that provide closed-loop policies include Feedback-based Information RoadMap (FIRM) Agha-Mohammadi et al. (2014), its extension with online local re-planning Agha-mohammadi et al. (2018), and T-LQG Rafieisakhaei et al. (2017). FIRM-based methods, which are developed primarily for motion planning problems, construct a graph in the belief space whose edges correspond to local LQG controllers. T-LQG approximately decouples belief space planning into local trajectory optimization and LQG tracking via a use of the separation principle. The local trajectory optimization part achieves one of the lowest asymptotic computational complexities among existing trajectory optimization methods, and thus can be solved efficiently with a commercial nonlinear programming (NLP) solver Rafieisakhaei et al. (2017).

Even though closed-loop methods are appealing, their computation cost can be prohibitive in some challenging planning and control problems, especially when the state space is large and yet real-time performance is required. In such cases, a practical strategy to alleviate the computational burden is open-loop planning, wherein one seeks to optimize a static sequence of control actions. Often times feedback is provided through re-planning of a fixed horizon, open-loop control sequence after making a new observation. This combination of open-loop planning and feedback through re-planning is called receding horizon control (RHC) or model predictive control (MPC) Huber (2009). Although they do not account for feedback at each planning time, RHC methods in general have successfully solved challenging planning and control problems with high efficiency where fast closed-loop policy computation can be impractical. Examples of such real-time applications include trajectory tracking with quadrotors Bangura and Mahony (2014) and agile autonomous driving on dirt Williams et al. (2018), although they are both in the optimal control literature. Within belief space planning, some methods based on the MLO assumption Platt (2013); Patil et al. (2014) are run as RHC.

1.2 Contributions

Our approach presented in this paper takes the form of RHC, which is reasonable for high-dimensional and continuous belief space planning problems that also demand highly-dynamic, real-time maneuvers with high frequency observation updates. However, the proposed method is significantly different than any of the previous approaches discussed in Section 1.1. We view the stochastic belief dynamics as a hybrid system with time-driven switching Heemels et al. (2009), where the controls are applied in continuous time and the observations are made in discrete time. A discrete-time observation creates a jump discontinuity in the belief state trajectory due to a sudden Bayesian update of the belief state. This view of belief space planning yields a continuous-time optimal control problem of a high-dimensional hybrid system. We then propose a model-based control algorithm to efficiently compute control signals in a receding-horizon fashion. The algorithm is based on Sequential Action Control (SAC) Ansari and Murphey (2016). SAC in its original form is a deterministic, model-based hybrid control algorithm, which “perturbs” a nominal control trajectory in a structured way so that the cost functional is optimally reduced up to the first order. The key to this approach is a careful use of the perturbation theory of differential equations that is often discussed in the mode scheduling literature Egerstedt et al. (2006); Wardi and Egerstedt (2012). As a result, SAC derives the optimal perturbation in closed form and synthesizes control signals at a high frequency to achieve a significant improvement over other optimal control methods that are based on local trajectory optimization Ansari and Murphey (2016).

We apply the perturbation theory to parametric Bayesian filters and derive the optimal control perturbation using the framework of SAC. Even though each control perturbation is small, high-frequency control synthesis and online re-planning yield a series of control actions that is significantly different than the nominal control, reacting to stochastic observations collected during execution or online changing conditions. Furthermore, we extend the original SAC algorithm to also account for stochasticity in the future observations during planning, by incorporating Monte Carlo sampling of nominal belief trajectories. Our key contribution is the resulting continuous belief space planning algorithm, which we name SACBP. The algorithm has the following desirable properties:

1. Although the form of control perturbation is open-loop with on-line re-planning, the perturbation computed by SACBP in near real-time is optimized for better average performance over the planning horizon than a given nominal control, whether it is open-loop or closed-loop.

2. SACBP does not require discretization of the state space, the observation space, or the control space. It also does not require discretization of time other than for numerical integration purposes.

3. General nonlinear parametric Bayesian filters can be used for state estimation as long as the system is control-affine and the control cost is quadratic.

4. Stochasticity in the future observations are fully considered.

5. SACBP is an anytime algorithm. Furthermore, the Monte Carlo sampling part of the algorithm is naturally parallelizable.

6. Even though SACBP is inherently suboptimal for the original stochastic optimal control problem, empirical results suggest that it is highly sample-efficient and outperforms other open-loop and closed-loop methods when near real-time performance is required.

There exists prior work Mavrommati et al. (2018) that uses SAC for active sensing, but its problem formulation relies on the ergodic control framework, which is significantly different from the belief space planning framework we propose here. We show that our SACBP outperforms projection-based ergodic trajectory optimization, MCTS-DPW, T-LQG, and a greedy method on an active multi-target tracking example. We also show that SACBP outperforms belief iLQG, MCTS-DPW, and T-LQG on a manipulation scenario.

This paper is an extension of the theory and results previously presented by the authors in Nishimura and Schwager (2018b). Compared to the conference version, we provide a more detailed derivation of the algorithm (Section 2) as well as a thorough mathematical analysis of the open-loop control perturbation for stochastic hybrid systems (Section 3 and Appendix A). This analysis leads to a guarantee for SACBP that, with an appropriate choice of the perturbation duration, the algorithm is expected to perform no worse than the nominal control. Since the nominal control can be arbitrary, one could even provide a discrete POMDP policy derived offline as a nominal policy to “warm-start” the planning. We also report new simulation results that involve a comparison of SACBP with T-LQG (Section 4), where we observe superior performance of SACBP for real-time applications.

In the next section we derive relevant equations and present the SACBP algorithm along with a discussion on computational complexity. Section 3 provides the key results of the mathematical analysis. Section 4 summarizes the simulation results. Conclusions and future work are presented in Section 5.

2 SACBP Algorithm

We first consider the case where some components of the state are fully observable. We begin with this mixed observability case as it is simpler to explain, yet still practically relevant. For example, this is a common assumption in various active sensing problems Schwager et al. (2017); Le Ny and Pappas (2009); Popović et al. (2017) where the state of the robot is perfectly known, but some external variable of interest (e.g. a target’s location) is stochastic. In addition, deterministic state transitions are often assumed for the robot. Therefore, in Section 2.1 we derive the SACBP control update formulae for this case. The general belief space planning where none of the state is fully observable or deterministically controlled is discussed in Section 2.2. An extension to use a closed-loop policy as the nominal control is presented in Section 2.3. The computation time complexity is discussed in Section 2.4.

2.1 Problems with Mixed Observability

Suppose that a robot can fully observe and deterministically control its own state $p(t) \in \mathbb{R}^{n_p}$. Other external states over which the robot does not have direct control are not known and are estimated with the belief vector $b(t) \in \mathbb{R}^{n_b}$. This belief vector characterizes a probability distribution that the robot uses for state estimation. If the belief is Gaussian, for example, the covariance matrix can be vectorized column-wise and stacked all together with the mean to form the belief vector. We define the augmented state as $s \triangleq (p^{\mathrm{T}}, b^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{n_s}$.

2.1.1 Dynamics Model The physical state p is described by the following ODE:

$$\dot{p}(t) = f\left(p(t), u(t)\right), \tag{1}$$

where $u(t) \in \mathbb{R}^m$ is the control signal. On the other hand, suppose that the belief state only changes in discrete time upon arrival of a new observation from the sensors. In the case of target tracking, for example, this means that the change of the location of the target is described by a discrete time model from the robot's perspective. The behavior of such a system can be estimated by a discrete-time Bayesian filter. We will discuss the more general continuous-discrete time filtering case in Section 2.2. Let t_k be the time when the k-th observation becomes available to the robot. The belief state transition is given by

$$\begin{cases} b(t_k) = g(p(t_k^-), b(t_k^-), y_k) \\ b(t) = b(t_k) \qquad\qquad\qquad \forall t \in [t_k, t_{k+1}), \end{cases} \tag{2}$$

where t_k^- is infinitesimally smaller than t_k. Nonlinear function g corresponds to a discrete-time, parametric Bayesian filter (e.g., Kalman filter, extended Kalman filter, discrete Bayesian filter, etc.) that forward-propagates the belief for prediction, takes the new observation $y_k \in \mathbb{R}^q$, and returns the updated belief state. The concrete choice of the filter depends on the instance of the problem. Note that the belief state stays constant for the most of the time in (2). This is because we are viewing a discrete time model in a continuous time framework.

Equations (1) and (2) constitute a hybrid system with time-driven switching Heemels et al. (2009). This hybrid system representation is practical since it captures the fact that the observation updates occur less frequently than the control actuation in general, due to expensive information processing of sensor readings. Furthermore, with this representation one can naturally handle agile robot dynamics as they are without coarse discretization in time.

Given the initial state $s_0 \triangleq (p(t_0)^{\mathrm{T}}, b(t_0)^{\mathrm{T}})^{\mathrm{T}}$ and some control $u(t)$ for $t \in [t_0, t_f]$, the system evolves stochastically according to the hybrid dynamics equations. The stochasticity is due to a sequence of stochastic future observations that will be taken by the end of the planning horizon t_f. In this paper we assume that the observation interval $t_{k+1} - t_k \triangleq \Delta t_o$ is fixed, and the control signals are recomputed when a new observation is incorporated in the belief, although one can also use a variable observation interval.

2.1.2 Perturbed Dynamics The control synthesis of SACBP begins with a given nominal control schedule $u(t)$ for $t \in [t_0, t_f]$. For simplicity we assume here that the nominal control schedule is open-loop, but we remind the reader that SACBP can also handle closed-loop nominal policies, which we discuss in Section 2.3. Suppose that the nominal control is applied to the system and a sequence of T observations $(y_1, \dots, y_T)$ is obtained. Conditioned on the observation sequence, the augmented state evolves deterministically. Let $s = (p^{\mathrm{T}}, b^{\mathrm{T}})^{\mathrm{T}}$ be the nominal trajectory of the augmented state induced by $(y_1, \dots, y_T)$.

Now let us consider perturbing the nominal trajectory at a fixed time $\tau < t_1$ for a short duration ϵ. The perturbed control u^ϵ is defined as

$$u^\epsilon(t) \triangleq \begin{cases} v & \text{if } t \in (\tau - \epsilon, \tau] \\ u(t) & \text{otherwise.} \end{cases} \tag{3}$$

Therefore, the control perturbation is determined by the nominal control $u(t)$, the tuple (τ, v), and ϵ. Given (τ, v), the resulting perturbed system trajectory can be written as

$$\begin{cases} p^\epsilon(t) \triangleq p(t) + \epsilon\Psi_p(t) + o(\epsilon) \\ b^\epsilon(t) \triangleq b(t) + \epsilon\Psi_b(t) + o(\epsilon), \end{cases} \tag{4}$$

where $\Psi_p(t)$ and $\Psi_b(t)$ are the state variations that are linear in the perturbation duration ϵ:

$$\Psi_p(t) = \left.\frac{\partial_+}{\partial\epsilon} p^\epsilon(t)\right|_{\epsilon=0} \triangleq \lim_{\epsilon \to 0^+} \frac{p^\epsilon(t) - p(t)}{\epsilon} \tag{5}$$

$$\Psi_b(t) = \left.\frac{\partial_+}{\partial\epsilon} b^\epsilon(t)\right|_{\epsilon=0} \triangleq \lim_{\epsilon \to 0^+} \frac{b^\epsilon(t) - b(t)}{\epsilon}. \tag{6}$$

The notation $\frac{\partial_+}{\partial\epsilon}$ represents the right derivative with respect to ϵ. The state variations at perturbation time τ satisfy

$$\begin{cases} \Psi_p(\tau) = f(p(\tau), v) - f(p(\tau), u(\tau)) \\ \Psi_b(\tau) = 0. \end{cases} \tag{7}$$

The initial belief state variation $\Psi_b(\tau)$ is zero because the control perturbation u^ϵ has no effect on the belief state until the first Bayesian update is performed at time t_1, according to the hybrid system model (1)(2). For $t \geq \tau$, the physical state variation Ψ_p evolves according to the following first-order ODE:

$$\begin{aligned} \dot{\Psi}_p(t) &= \frac{d}{dt}\left(\left.\frac{\partial_+}{\partial\epsilon} p^\epsilon(t)\right|_{\epsilon=0}\right) && (8) \\ &= \left.\frac{\partial_+}{\partial\epsilon} \dot{p}^\epsilon(t)\right|_{\epsilon=0} && (9) \\ &= \left.\frac{\partial_+}{\partial\epsilon} f(p^\epsilon(t), u(t))\right|_{\epsilon=0} && (10) \\ &= \frac{\partial}{\partial p} f\left(p(t), u(t)\right)\Psi_p(t), && (11) \end{aligned}$$

where the chain rule of differentiation and $p^\epsilon(t)|_{\epsilon=0} = p(t)$ are used in (11). For a more rigorous analysis, see Appendix A. The dynamics of the belief state variation Ψ_b in the continuous region $t \in [t_k, t_{k+1})$ satisfy $\dot{\Psi}_b(t) = 0$ since the belief vector $b(t)$ is constant according to (2). However, across the jumps the belief state variation Ψ_b changes discontinuously and satisfies

$$\Psi_b(t_k) = \frac{\partial_+}{\partial \epsilon} b^\epsilon(t_k)\bigg|_{\epsilon=0} \tag{12}$$

$$= \frac{\partial_+}{\partial \epsilon} g\left(p^\epsilon(t_k^-), b^\epsilon(t_k^-), y_k\right)\bigg|_{\epsilon=0} \tag{13}$$

$$= \frac{\partial}{\partial p} g\left(p(t_k^-), b(t_k^-), y_k\right)\Psi_p(t_k^-) + \frac{\partial}{\partial b} g\left(p(t_k^-), b(t_k^-), y_k\right)\Psi_b(t_k^-). \tag{14}$$

2.1.3 Perturbed Cost Functional

Let us consider a total cost of the form

$$\int_{t_0}^{t_f} c\left(p(t), b(t), u(t)\right) dt + h(p(t_f), b(t_f)), \tag{15}$$

where c is the running cost and h is the terminal cost. Following the discussion above on the perturbed dynamics, let J denote the total cost of the nominal trajectory conditioned on the given observation sequence $(y_1, \ldots, y_T)$. Under the fixed (τ, v), we can represent the perturbed cost J^ϵ in terms of J as

$$J^\epsilon \triangleq J + \epsilon\nu(t_f) + o(\epsilon), \tag{16}$$

where $\nu(t_f) \triangleq \frac{\partial_+}{\partial \epsilon} J^\epsilon|_{\epsilon=0}$ is the variation of the total cost with respect to the perturbation. For further analysis it is convenient to express the running cost in the Mayer form Liberzon (2011). Let $\hat{s}(t)$ be a new state variable defined by $\dot{\hat{s}}(t) = c\left(p(t), b(t), u(t)\right)$ and $\hat{s}(t_0) = 0$. Then the total cost is a function of the appended augmented state $\overline{s} \triangleq (\hat{s}, s^\mathrm{T})^\mathrm{T} \in \mathbb{R}^{1+n_s}$ at time t_f, which is given by

$$J = \hat{s}(t_f) + h\left(s(t_f)\right). \tag{17}$$

Using this form of the total cost J, the perturbed cost (16) becomes

$$J^\epsilon = J + \epsilon \begin{bmatrix} 1 \\ \frac{\partial}{\partial p} h\left(p(t_f), b(t_f)\right) \\ \frac{\partial}{\partial b} h\left(p(t_f), b(t_f)\right) \end{bmatrix}^\mathrm{T} \overline{\Psi}(t_f) + o(\epsilon), \tag{18}$$

where $\overline{\Psi}(t_f) \triangleq \left(\hat{\Psi}(t_f), \Psi_p(t_f)^\mathrm{T}, \Psi_b(t_f)^\mathrm{T}\right)^\mathrm{T}$ and $\hat{\Psi}$ is the variation of $\hat{s}$. Note that the dot product in (18) corresponds to $\nu(t_f)$ in (16). The variation $\hat{\Psi}$ follows the variational equation for $t \geq \tau$:

$$\dot{\hat{\Psi}}(t) = \frac{d}{dt}\left(\frac{\partial_+}{\partial \epsilon}\hat{s}^\epsilon(t)\bigg|_{\epsilon=0}\right) \tag{19}$$

$$= \frac{\partial_+}{\partial \epsilon}\dot{\hat{s}}^\epsilon(t)\bigg|_{\epsilon=0} \tag{20}$$

$$= \frac{\partial_+}{\partial \epsilon} c(p^\epsilon(t), b^\epsilon(t), u(t))\bigg|_{\epsilon=0} \tag{21}$$

$$= \frac{\partial}{\partial p} c(p(t), b(t), u(t))^\mathrm{T}\Psi_p(t) + \frac{\partial}{\partial b} c(p(t), b(t), u(t))^\mathrm{T}\Psi_b(t), \tag{22}$$

where the initial condition is given by $\hat{\Psi}(\tau) = c(p(\tau), b(\tau), v) - c(p(\tau), b(\tau), u(\tau))$.

The perturbed cost equation (18), especially the dot product expressing $\nu(t_f)$, is consequential; it tells us how the total cost changes due to the perturbation applied at some time τ, up to the first order with respect to the perturbation duration ϵ. At this point, one could compute the value of $\nu(t_f)$ for a control perturbation with a specific value of (τ, v) by simulating the nominal dynamics and integrating the variational equations (11)(14)(22) from τ up to t_f.

2.1.4 Adjoint Equations

Unfortunately, this forward integration of $\nu(t_f)$ is not so useful by itself since we are interested in finding the value of (τ, v) that achieves the largest possible negative $\nu(t_f)$, if it exists; it would be computationally intensive to apply control perturbation at different application times τ with different values of v and re-simulate the state variation $\overline{\Psi}$. To avoid this computationally expensive search, we mirror the approach presented in Ansari and Murphey (2016) and introduce the adjoint system $\overline{\rho} \triangleq (\hat{\rho}, \rho_p^\mathrm{T}, \rho_b^\mathrm{T})^\mathrm{T}$ with which the dot product remains invariant:

$$\frac{d}{dt}\left(\overline{\rho}(t)^\mathrm{T}\overline{\Psi}(t)\right) = 0 \ \ \forall t \in [t_0, t_f]. \tag{23}$$

If we let

$$\overline{\rho}(t_f) \triangleq \left(1, \frac{\partial}{\partial p} h\left(p(t_f), b(t_f)\right)^\mathrm{T}, \frac{\partial}{\partial b} h\left(p(t_f), b(t_f)\right)^\mathrm{T}\right)^\mathrm{T} \tag{24}$$

so that its dot product with $\overline{\Psi}(t_f)$ equals $\nu(t_f)$ as in (18), the time invariance gives

$$\nu(t_f) = \overline{\rho}(t_f)^\mathrm{T}\overline{\Psi}(t_f) \tag{25}$$

$$= \overline{\rho}(\tau)^\mathrm{T}\overline{\Psi}(\tau) \tag{26}$$

$$= \overline{\rho}(\tau)^\mathrm{T}\begin{bmatrix} c\left(p(\tau), b(\tau), v\right) - c\left(p(\tau), b(\tau), u(\tau)\right) \\ f(p(\tau), v) - f(p(\tau), u(\tau)) \\ 0 \end{bmatrix}. \tag{27}$$

Therefore, we can compute the first-order cost change $\nu(t_f)$ for different values of τ once the adjoint trajectory is derived. For $t \in [t_k, t_{k+1})$ the time derivative of $\overline{\Psi}$ exists, and the invariance property leads to the following equation:

$$\dot{\overline{\rho}}(t)^\mathrm{T}\overline{\Psi}(t) + \overline{\rho}(t)^\mathrm{T}\dot{\overline{\Psi}}(t) = 0. \tag{28}$$

It can be verified that the following system satisfies (28) with $\overline{\rho}(t) = \left(\hat{\rho}(t), \rho_p(t)^\mathrm{T}, \rho_b(t)^\mathrm{T}\right)^\mathrm{T}$:

$$\begin{cases} \dot{\hat{\rho}}(t) = 0 \\ \dot{\rho}_p(t) = -\frac{\partial}{\partial p} c(p(t), b(t), u(t)) - \frac{\partial}{\partial p} f(p(t), u(t))^\mathrm{T}\rho_p(t) \\ \dot{\rho}_b(t) = -\frac{\partial}{\partial b} c(p(t), b(t), u(t)). \end{cases} \tag{29}$$

Analogously, across discrete jumps we can still enforce the invariance by setting $\overline{\rho}(t_k)^\mathrm{T}\overline{\Psi}(t_k) = \overline{\rho}(t_k^-)^\mathrm{T}\overline{\Psi}(t_k^-)$, which holds for the following adjoint equations:

$$\begin{cases} \hat{\rho}(t_k^-) = \hat{\rho}(t_k) \\ \rho_p(t_k^-) = \rho_p(t_k) + \frac{\partial}{\partial p} g\left(p(t_k^-), b(t_k^-), y_k\right)^{\mathrm{T}} \rho_b(t_k) \\ \rho_b(t_k^-) = \frac{\partial}{\partial b} g\left(p(t_k^-), b(t_k^-), y_k\right)^{\mathrm{T}} \rho_b(t_k). \end{cases} \tag{30}$$

Note that the adjoint system integrates backward in time as it has the boundary condition (24) defined at t_f . More importantly, the adjoint dynamics (29)(30) only depend on the nominal trajectory of the system (p, b) and the observation sequence $(y_1, \ldots, y_T)$. Considering that $\hat{\rho}(t) = 1$ at all times, the cost variation term $\nu(t_f)$ is finally given by

$$\nu(t_f) = c(p(\tau), b(\tau), v) - c(p(\tau), b(\tau), u(\tau)) + \rho_p(\tau)^{\mathrm{T}} \left\{ f(p(\tau), v) - f(p(\tau), u(\tau)) \right\}. \tag{31}$$

2.1.5 Control Optimization In order to efficiently optimize (31) with respect to (τ, v), we assume hereafter that the control cost is additive quadratic $\frac{1}{2}u^{\mathrm{T}}C_u u$ and the dynamics model $f(p, u)$ is control-affine with linear term $H(p)u$, where $H\colon \mathbb{R}^{n_p} \to \mathbb{R}^m$ can be any nonlinear mapping. Although the control-affine assumption may appear restrictive, many physical systems possess this property in engineering practice. As a result of these assumptions, (31) becomes

$$\nu(t_f) = \frac{1}{2}v^{\mathrm{T}}C_u v + \rho_p(\tau)^{\mathrm{T}} H(p(\tau))(v - u(\tau)) - \frac{1}{2}u(\tau)^{\mathrm{T}}C_u u(\tau). \tag{32}$$

So far we have treated the observation sequence $(y_1, \ldots, y_T)$ as given and fixed. However, in practice it is a random process that we have to take into account. Fortunately, our control optimization is all based on the nominal control schedule $u(t)$, with which we can both simulate the augmented dynamics and sample the observations. To see this, let us consider the observations as a sequence of random vectors $(Y_1, \ldots, Y_T)$ and rewrite $\nu(t_f)$ in (32) as $\nu(t_f, Y_1, \ldots, Y_T)$ to clarify the dependence on it. The expected value of the first order cost variation is given by

$$\mathbb{E}[\nu(t_f)] = \int \nu(t_f, Y_1, \ldots, Y_T) d\mathbb{P}, \tag{33}$$

where $\mathbb{P}$ is the probability measure associated with these random vectors. Although we do not know the specific values of $\mathbb{P}$, we have the generative model; we can simulate the augmented state trajectory using the nominal control and sequentially sample the stochastic observations from the belief states along the trajectory.

Using the linearity of expectation for (32), we have

$$\mathbb{E}[\nu(t_f)] = \frac{1}{2}v^{\mathrm{T}}C_u v + \mathbb{E}[\rho_p(\tau)]^{\mathrm{T}} H(p(\tau))(v - u(\tau)) - \frac{1}{2}u(\tau)^{\mathrm{T}}C_u u(\tau). \tag{34}$$

Notice that only the adjoint trajectory is stochastic. We can employ Monte Carlo sampling to sample a sufficient number of observation sequences to approximate the expected adjoint trajectory. Now (34) becomes a convex quadratic in v for a positive definite C_u. Assuming that C_u is also diagonal, analytical solutions are available to the following convex optimization problem with an input saturation constraint.

$$\begin{aligned} \underset{v}{\text{minimize}} \quad & \mathbb{E}[\nu(t_f)] \\ \text{subject to} \quad & a \preceq v \preceq b, \end{aligned} \tag{35}$$

where $a, b \in \mathbb{R}^m$ are some saturation vectors and $\preceq$ is an elementwise inequality. This optimization is solved for different values of $\tau \in (t_0 + t_{\text{calc}} + \epsilon, t_0 + t_{\text{calc}} + \Delta t_o)$, where t_{calc} is a pre-allocated computation time budget and Δt_o is the time interval between two successive observations as well as control updates. We then search over $(\tau, v^*(\tau))$ for the optimal perturbation time τ^* to globally minimize $\mathbb{E}[\nu(t_f)]$. There is only a finite number of such τ to consider since in practice we use numerical integration such as the Euler scheme with some step size Δt_c to compute the trajectories. In Ansari and Murphey (2016) the finite perturbation duration ϵ is also optimized using line search, but in this work we set ϵ as a tunable parameter to reduce the computation time. The complete algorithm is summarized in Algorithm 1. The call to the algorithm occurs every Δt_o[s] in a receding-horizon fashion, after the new observation is incorporated in the belief.

Algorithm 1 SACBP Control Update for Problems with Mixed Observability

```
INPUT: Current augmented state $s_0 = (p(t_0)^{\mathrm{T}}, b(t_0)^{\mathrm{T}})^{\mathrm{T}}$, nominal control schedule $u(t)$ for $t \in [t_0, t_f]$, perturbation duration $\epsilon$
OUTPUT: Optimally perturbed control schedule $u^{\epsilon}(t)$ for $t \in [t_0, t_f]$
for $i = 1{:}N$ do
    Forward-simulate nominal augmented state trajectory (1)(2) and sample observation sequence $(y_1^i, \ldots, y_T^i)$ along the augmented state trajectory.
    Backward-simulate nominal adjoint trajectory $\rho_p^i$, $\rho_b^i$ (29)(30) with sampled observations.
end for
Compute Monte Carlo estimate: $\mathbb{E}[\rho_p] \approx \frac{1}{N}\sum_{i=1}^{N} \rho_p^i$.
for $(\tau = t_0 + t_{\text{calc}} + \epsilon;\ \tau \leq t_0 + t_{\text{calc}} + \Delta t_o;\ \tau \leftarrow \tau + \Delta t_c)$ do
    Solve quadratic minimization (35) with (34). Store optimal value $\nu^*(\tau)$ and optimizer $v^*(\tau)$.
end for
$\tau^* \leftarrow \arg\min \nu^*(\tau),\ v^* \leftarrow v^*(\tau^*)$
$u^{\epsilon} \leftarrow PerturbControlTrajectory(u, v^*, \tau^*, \epsilon)$ (3)
return $u^{\epsilon}$
```

2.2 General Belief Space Planning Problems

If none of the state is fully observable, the same stochastic SAC framework still applies almost as is to the belief sate b. In this case we consider a continuous-discrete filter Xie et al. (2007) where the prediction step follows an ODE and the update step provides an instantaneous discrete jump. The hybrid dynamics for the belief vector are given by

$$\begin{cases} b(t_k) = g(b(t_k^-), y_k) \\ \dot{b}(t) = f(b(t), u(t)) \qquad \forall t \in [t_k, t_{k+1}). \end{cases} \tag{36}$$

Letting $\Psi(t) = \frac{\partial_+}{\partial \epsilon} b^\epsilon(t)\big|_{\epsilon=0}$, the variational equation yields

$$\begin{cases} \Psi(t_k) = \frac{\partial}{\partial b} g(b(t_k^-), y_k)\Psi(t_k^-) \\ \dot{\Psi}(t) = \frac{\partial}{\partial b} f(b(t), u(t))\Psi(t) \qquad \forall t \in [t_k, t_{k+1}) \end{cases} \tag{37}$$

with initial condition $\Psi(\tau) = f(b(\tau), v) - f(b(\tau), u(\tau))$.

Let the total cost be of the form:

$$\int_{t_0}^{t_f} c(b(t), u(t))dt + h(b(t_f)). \tag{38}$$

Under the given (τ, v) and $(y_1, \dots, y_T)$, the variation $\nu(t_f)$ of the total cost can be computed as

$$\begin{aligned} \nu(t_f) = c(b(\tau), v) &- c(b(\tau), u(\tau)) \\ &+ \int_{\tau}^{t_f} \frac{\partial}{\partial b} c(b(t), u(t))^{\mathrm{T}} \Psi(t)dt \\ &+ \frac{\partial}{\partial b} h(b(t_f))^{\mathrm{T}} \Psi(t_f). \end{aligned} \tag{39}$$

This is equivalent to

$$\begin{aligned} \nu(t_f) &= c(b(\tau), v) - c(b(\tau), u(\tau)) \\ &+ \rho(\tau)^{\mathrm{T}} \left\{ f(b(\tau), v) - f(b(\tau), u(\tau)) \right\}, \end{aligned} \tag{40}$$

where ρ is the adjoint system that follows the dynamics:

$$\begin{cases} \rho(t_k^-) = \frac{\partial}{\partial b} g(b(t_k^-), y_k)^{\mathrm{T}} \rho(t_k) \\ \dot{\rho}(t) = -\frac{\partial}{\partial b} c(b(t), u(t)) - \frac{\partial}{\partial b} f(b(t), u(t))^{\mathrm{T}} \rho(t) \end{cases} \tag{41}$$

with the boundary condition $\rho(t_f) = \frac{\partial}{\partial b} h(b(t_f))$. Under the control-affine assumption for f and the additive quadratic control cost, the expected first order cost variation (40) yields

$$\begin{aligned} \mathbb{E}[\nu(t_f)] = \frac{1}{2} v^{\mathrm{T}} C_u v &+ \mathbb{E}[\rho(\tau)]^{\mathrm{T}} H(b(\tau))(v - u(\tau)) \\ &- \frac{1}{2} u(\tau)^{\mathrm{T}} C_u u(\tau), \end{aligned} \tag{42}$$

where $H(b(\tau))$ is the control coefficient term in f.

Although it is difficult to state the general conditions under which this control-affine assumption holds, for instance one can verify that the continuous-discrete EKF Xie et al. (2007) satisfies this property if the underlying system dynamics f_{sys} is control-affine.

$$\begin{cases} \dot{\mu}(t) = f_{\text{sys}}(\mu(t), u(t)) \\ \dot{\Sigma}(t) = A\Sigma + \Sigma A^{\mathrm{T}} + Q \end{cases} \tag{43}$$

In the above continuous-time prediction equations, A is the Jacobian of the dynamics function $f_{\text{sys}}(x(t), u(t))$ evaluated at the mean $\mu(t)$ and Q is the process noise covariance. If f_{sys} is control-affine, so is A and therefore so is $\dot{\Sigma}$. Obviously $\dot{\mu}$ is control affine as well. As a result the dynamics for the belief vector $b = (\mu^{\mathrm{T}}, \text{vec}(\Sigma)^{\mathrm{T}})^{\mathrm{T}}$ satisfy the control-affine assumption.

Mirroring the approach in Section 2.1, we can use Monte Carlo sampling to estimate the expected value in (42). The resulting algorithm is presented in Algorithm 2.

Algorithm 2 SACBP Control Update for General Belief Space Planning Problems

```
INPUT: Current belief state $b_0 = b(t_0)$, nominal control
    schedule $u(t)$ for $t \in [t_0, t_f]$, perturbation duration $\epsilon$
OUTPUT: Optimally perturbed control schedule $u^\epsilon(t)$ for
    $t \in [t_0, t_f]$.
 1: for $i$ = 1:$N$ do
 2:     Forward-simulate nominal belief state trajectory (36)
        and sample observation sequence $(y_1^i, \dots, y_T^i)$ along the
        belief trajectory.
 3:     Backward-simulate nominal adjoint trajectory $\rho^i$
        (41) with sampled observations.
 4: end for
 5: Compute Monte Carlo estimate: $\mathbb{E}[\rho] \approx \frac{1}{N}\sum_{i=1}^{N} \rho^i$.
 6: for ($\tau = t_0 + t_{\text{calc}} + \epsilon$;  $\tau \leq t_0 + t_{\text{calc}} + \Delta t_o$;  $\tau \leftarrow$
    $\tau + \Delta t_c$) do
 7:     Solve quadratic minimization (35) with (42). Store
        optimal value $\nu^*(\tau)$ and optimizer $v^*(\tau)$.
 8: end for
 9: $\tau^* \leftarrow \arg\min \nu^*(\tau)$, $v^* \leftarrow v^*(\tau^*)$
10: $u^\epsilon \leftarrow PerturbControlTrajectory(u, v^*, \tau^*, \epsilon)$ (3)
11: return $u^\epsilon$
```

2.3 Closed-loop Nominal Policy

In Sections 2.1 and 2.2 we assumed that the nominal control was an open-loop control schedule. However, one can think of a scenario where a nominal control is a closed-loop policy computed offline, such as a discrete POMDP policy that maps beliefs to actions Kurniawati et al. (2008). Indeed, SACBP can also handle closed-loop nominal policies. Let π be a closed-loop nominal policy, which is a mapping from either an augmented state $s(t)$ or a belief state $b(t)$ to a control value $u(t)$. Due to the stochastic belief dynamics, the control values returned by π in the future is also stochastic for $t \geq t_1$. This is reflected when we forward-propagate the nominal dynamics. Specifically, each sampled trajectory has a different control trajectory in addition to a different observation sequence. However, the equations are still convex quadratic in v as shown below. For problems with mixed observability, we have

$$\begin{aligned} \mathbb{E}[\nu(t_f)] = \frac{1}{2} v^{\mathrm{T}} C_u v &+ \mathbb{E}[\rho_p(\tau)]^{\mathrm{T}} H(p(\tau)) \left\{ v - \pi(s(\tau)) \right\} \\ &- \frac{1}{2} \pi(s(\tau))^{\mathrm{T}} C_u \pi(s(\tau)). \end{aligned} \tag{44}$$

The general belief space planning case also yields a similar equation:

$$\begin{aligned} \mathbb{E}[\nu(t_f)] = \frac{1}{2} v^{\mathrm{T}} C_u v &+ \mathbb{E}[\rho(\tau)]^{\mathrm{T}} H(b(\tau)) \left\{ v - \pi(b(\tau)) \right\} \\ &- \frac{1}{2} \pi(b(\tau))^{\mathrm{T}} C_u \pi(b(\tau)). \end{aligned} \tag{45}$$

Note that $s(\tau)$ and $b(\tau)$ are both deterministic since the first observation y_1 is not yet taken at $\tau < t_1$. The expectations in (44) and (45) can be estimated using Monte Carlo sampling. The forward-simulation of the nominal trajectory in Line 2 of Algorithms 1 and 2 is now with the closed loop policy π, and the equations in Line 7 need to be replaced with (44) and (45), respectively. However, the rest remains unchanged.

2.4 Computation Time Complexity

Let us analyze the time complexity of the SACBP algorithm. The bottleneck of the computation is when the forward-backward simulation is performed multiple times (lines 1–4 of Algorithms 1 and 2). The asymptotic complexity of this part is given by $O(N(\frac{t_f - t_0}{\Delta t_o})(M_{\text{forward}} + M_{\text{backward}}))$, where M_{forward} and M_{backward} are the times to respectively integrate the forward and backward dynamics between two successive observations. For a more concrete analysis let us use the Gaussian belief dynamics given by EKF as an example. For simplicity we assume the same dimension n for the state, the control, and the observation. The belief state has dimension $O(n^2)$. Using the Euler scheme, the forward integration takes $M_{\text{forward}} = O((\frac{\Delta t_o}{\Delta t_c} + 1)n^3)$ since evaluating continuous and discrete EKF equations are both $O(n^3)$. Computation of the continuous part of the adjoint dynamics (41) is dominated by the evaluation of the Jacobian $\frac{\partial f}{\partial b}$, which is $O(n^5)$ because $O(n^3)$ operations to evaluate f are carried out $O(n^2)$ times. The discrete part is also $O(n^5)$. Therefore, $M_{\text{backward}} = O((\frac{\Delta t_o}{\Delta t_c} + 1)n^5)$. Overall, the time complexity is $O(N(\frac{t_f - t_0}{\Delta t_o})(\frac{\Delta t_o}{\Delta t_c} + 1)n^5)$. This is asymptotically smaller in n than belief iLQG, which is $O(n^6)$. See Rafieisakhaei et al. (2017) for a comparison of time complexity among different belief space planning algorithms. We also remind the readers that SACBP is an online method and a naive implementation already achieves near real-time performance, computing control over a 2[s] horizon in about 0.1[s] to 0.4[s]. By near real-time we mean that a naive implementation of SACBP requires approximately $0.7 \times t_{\text{calc}}$ to $3 \times t_{\text{calc}}$ time to compute an action that must be applied t_{calc}[s] in the future. We expect that parallelization in a GPU and a more efficient implementation will result in real-time computation for SACBP.

3 Analysis of Mode Insertion Gradient for Stochastic Hybrid Systems

The SACBP algorithm presented in Section 2 as well as the original SAC algorithm Ansari and Murphey (2016) both rely on the local sensitivity analysis of the cost functional with respect to the control perturbation. This first-order sensitivity term (i.e. $\nu(t_f)$ in our notation) is known as the mode insertion gradient in the mode scheduling literature Egerstedt et al. (2006); Wardi and Egerstedt (2012). In Ansari and Murphey (2016) the notion of the mode insertion gradient has been generalized to handle a broader class of hybrid systems than discussed before. What remains to be seen is a further generalization of the mode insertion gradient to stochastic hybrid systems, such as the belief dynamics discussed in this paper. Indeed, the quantity we can optimize in (35) is essentially the expected value of the first-order sensitivity of the total cost. This is not to be confused with the first-order sensitivity of the expected total cost, which would be the natural generalization of the mode insertion gradient to stochastic systems. In general, those two quantities can be different, since the order of expectation and differentiation may not be swapped arbitrarily. In this section, we provide a set of sufficient conditions under which the order can be exchanged. By doing so we show that 1) the notion of mode insertion gradient can be generalized to stochastic hybrid systems, and 2) the SACBP algorithm optimizes this generalized mode insertion gradient. Through this analysis we will see that the SACBP algorithm has a guarantee that, in expectation it performs at least as good as the nominal policy with an appropriate choice of ϵ.

3.1 Assumptions

Let us begin with a set of underlying assumptions for the system dynamics, the control, and the cost functions. Without loss of generality, we assume that the system starts at time $t = 0$ and ends at $t = T$, with a sequence of T observations $(y_1, \dots, y_T)$ made every unit time. For generality, we use notation x to represent the state variable of the system in this section, in place of b or s that respectively represented the belief state or the augmented state in Section 2. This means that the analysis presented here is not restricted to belief systems where the dynamics are governed by Bayesian filters, but rather applies to a broader class of stochastic systems.

Assumption 1. Control Model. *The controls are in $\widetilde{C}^{0,m}[0,T]$, the space of piecewise continuous functions from $[0,T]$ into $\mathbb{R}^m$. We further assume that there exists some $\rho_{\max} < \infty$ such that for all $t \in [0,T]$, we have $u(t) \in B(0, \rho_{\max})$ where $B(0, \rho_{\max})$ is the closed Euclidean ball of radius $\rho_{\max}$ centered at 0, i.e. $\|u(t)\|_2 \leq \rho_{\max}$. We denote this admissible control set by $U \triangleq \{u \in \widetilde{C}^{0,m}[0,T] \mid \forall t \in [0,T]\ u(t) \in B(0, \rho_{\max})\}$.*

Remark 1. *For the sake of analysis, the control model described above takes the form of an open-loop control schedule, where time t determines the control signal $u(t)$. However the analysis can be extended to closed-loop nominal policies in a straightforward manner, which is discussed in Appendix A. (See Remark 5.)*

Assumption 2. Dynamics Model. *Let $x_0 \in \mathbb{R}^{n_x}$ be the given initial state value at $t = 0$. Given a control $u \in U$ and a sequence of observations $(y_1, \dots, y_T) \in \mathbb{R}^{n_y} \times \dots \times \mathbb{R}^{n_y}$, the dynamics model is the following hybrid system with time-driven switching:*

$$x(t) \triangleq x_i(t)\ \forall t \in [i-1, i)\ \forall i \in \{1, 2, \dots, T\}, \tag{46}$$

where x_i is the i-th "mode" of the system state defined on $[i-1, i]$ as:

$$x_i(i-1) = g\left(x_{i-1}(i-1), y_{i-1}\right) \tag{47}$$

$$\dot{x}_i(t) = f\left(x_i(t), u(t)\right) \qquad \forall t \in [i-1, i], \tag{48}$$

with $x(0) = x_1(0) = x_0$. We also define the final state as $x(T) \triangleq g(x_T(T), y_T)$.

For the transition functions f and g we assume the following:

- *(2a) the function $f\colon \mathbb{R}^{n_x} \times \mathbb{R}^m \to \mathbb{R}^{n_x}$ is continuously differentiable;*
- *(2b) the function $g\colon \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \to \mathbb{R}^{n_x}$ is continuous. It is also differentiable in x;*
- *(2c) for function f, there exists a constant $K_1 \in [1, \infty)$ such that $\forall x', x'' \in \mathbb{R}^{n_x}$ and $\forall u', u'' \in B(0, \rho_{\max})$,*

the following condition holds:

$$\|f(x',u')-f(x'',u'')\|_2 \leq K_1\left(\|x'-x''\|_2+\|u'-u''\|_2\right) \quad (49)$$

(2d) *for function g, there exist finite non-negative constants K_2, K_3, K_4, K_5 and positive integers L_1, L_2 such that $\forall x \in \mathbb{R}^{n_x}$ and $\forall y \in \mathbb{R}^{n_y}$, the following relations hold:*

$$\|g(x,y)\|_2 \leq K_2 + K_3\|x\|_2^{L_1} + K_4\|y\|_2^{L_2} + K_5\|x\|_2^{L_1}\|y\|_2^{L_2} \quad (50)$$

$$\left\|\frac{\partial}{\partial x}g(x,y)\right\|_2 \leq K_2 + K_3\|x\|_2^{L_1} + K_4\|y\|_2^{L_2} + K_5\|x\|_2^{L_1}\|y\|_2^{L_2} \quad (51)$$

Remark 2. *Assumptions (2a) and (2c) are used to show existence and uniqueness of the solution to the differential equation (48) as well as the variational equation under control perturbation. (See Propositions 3 and 16 in Appendix A.) Note that Assumption (2c) is essentially a Lipschitz continuity condition, which is a quite common assumption in the analysis of nonlinear ODEs, as can be seen in Assumption 5.6.2 in Elijah (1997) and Theorem 2.3 in Khalil and Grizzle (2002). Assumptions (2b) and (2d) are the growth conditions on x across adjacent modes. Recall that in belief space planning where the system state x is the belief state b, the jump function g corresponds to the observation update of the Bayesian filter. The form of the bound in (50) and (51) allows a broad class of continuous functions to be considered as g, and is inspired by a few examples of the Bayesian update equations as presented below.*

Proposition 1. Bounded Jump for Univariate Gaussian Distribution. *Let $b = (\mu, s)^{\mathrm{T}} \in \mathbb{R}^2$ be the belief state, where μ is the mean parameter and $s > 0$ is the variance. Suppose that the observation y is the underlying state $x \in \mathbb{R}$ corrupted by additive Gaussian white noise $v \sim \mathcal{N}(0,1)$. Then, the Bayesian update function g for this belief system satisfies Assumption (2d).*

Proof. The Bayesian update formula for this system is given by $g(b,y) = \hat{b} \triangleq (\hat{\mu}, \hat{s})^{\mathrm{T}}$, where

$$\hat{\mu} = \mu + \frac{s}{s+1}(y-\mu) \quad (52)$$

$$\hat{s} = s - \frac{s^2}{s+1} \quad (53)$$

is the update step of the Kalman filter. Rearranging the terms, we have

$$g(b,y) = \frac{1}{s+1}\begin{pmatrix}\mu + sy \\ s\end{pmatrix} \quad (54)$$

and consequently,

$$\frac{\partial}{\partial b}g(b,y) = \frac{1}{(s+1)^2}\begin{pmatrix}s+1 & y \\ 0 & 1\end{pmatrix}. \quad (55)$$

We will show that the function g satisfies Assumption (2d). For the bound on $g(b,y)$,

$$\begin{aligned}\|g(b,y)\|_2^2 &= \frac{1}{(s+1)^2}\left\{(\mu+sy)^2+s^2\right\} && (56)\\ &\leq (\mu+sy)^2+s^2 && (57)\\ &\leq \|b\|_2^2 + \left(b^{\mathrm{T}}\begin{pmatrix}y\\1\end{pmatrix}\right)^2 && (58)\\ &\leq \|b\|_2^2(2+\|y\|_2^2), && (59)\end{aligned}$$

where we have used $(s+1)^2 \geq 1$ and the Cauchy-Schwarz inequality. Thus,

$$\begin{aligned}\|g(b,y)\|_2 &\leq \|b\|_2\sqrt{2+\|y\|_2^2} && (60)\\ &\leq \sqrt{2}\|b\|_2 + \|b\|_2\|y\|_2 && (61)\end{aligned}$$

Similarly, the bound on the Jacobian yields

$$\begin{aligned}\left\|\frac{\partial}{\partial b}g(b,y)\right\|_2^2 &\leq \left\|\frac{\partial}{\partial b}g(b,y)\right\|_{\mathrm{F}}^2 && (62)\\ &= \frac{1}{(s+1)^4}\left\{(s+1)^2+y^2+1\right\} && (63)\\ &= \frac{1}{(s+1)^2}+\frac{1}{(s+1)^4}(y^2+1) && (64)\\ &\leq 2+\|y\|_2^2. && (65)\end{aligned}$$

Therefore,

$$\left\|\frac{\partial}{\partial b}g(b,y)\right\|_2 \leq \sqrt{2}+\|y\|_2 \quad (66)$$

This shows that the jump function g for the above univariate Gaussian model satisfies Assumption (2d) with $(K_2, K_3, K_4, K_5) = (\sqrt{2}, \sqrt{2}, 1, 1)$ and $(L_1, L_2) = (1, 1)$.

Proposition 2. Bounded Jump for Categorical Distribution. *Let $b = (b_1, \ldots, b_n)^{\mathrm{T}} \in \mathbb{R}^n$ be the n-dimensional belief state representing the categorical distribution over the underlying state $x \in \{1, \ldots, n\}$. We choose the unnormalized form where the probability of $x = i$ is given by $b_i / \sum_{i=1}^n b_i$. Let the observation $y \in \{1, \ldots, m\}$ be modeled by a conditional probability mass function $p(y \mid x) \in [0,1]$. Then, the Bayesian update function g for this belief system satisfies Assumption (2d).*

Proof. The Bayes rule gives $g(b,y) = \hat{b} \triangleq (\hat{b}_1, \ldots, \hat{b}_n)$, where

$$\begin{pmatrix}\hat{b}_1\\ \hat{b}_2\\ \vdots\\ \hat{b}_n\end{pmatrix} = \begin{pmatrix}p(y \mid 1)b_1\\ p(y \mid 2)b_2\\ \vdots\\ p(y \mid n)b_n\end{pmatrix}. \quad (67)$$

Therefore, we can easily bound the norm of the posterior belief $\hat{b}$ by

$$\|g(b,y)\|_2 = \|\hat{b}\|_2 \leq \|b\|_2, \quad (68)$$

as $p(y \mid x) \leq 1$. The Jacobian is simply the diagonal matrix $\mathrm{diag}(p(y \mid 1), \ldots, p(y \mid n))$, and hence

$$\left\|\frac{\partial}{\partial b}g(b,y)\right\|_2 \leq 1. \quad (69)$$

This shows that the jump function g for the categorical belief model above satisfies Assumption (2d) with $(K_2, K_3, K_4, K_5) = (1, 1, 0, 0)$ and $(L_1, L_2) = (1, 1)$.

Assumption 3. Cost Model. *The instantaneous cost $c\colon \mathbb{R}^{n_x} \times \mathbb{R}^m \to \mathbb{R}$ is continuous. It is also continuously differentiable in x. The terminal cost $h\colon \mathbb{R}^{n_x} \to \mathbb{R}$ is differentiable. Furthermore, we assume that there exist finite non-negative constants K_6, K_7 and a positive integer L_3 such that for all $x \in \mathbb{R}^{n_x}$ and $u \in B(0, \rho_{\max})$, the following relations hold:*

$$|c(x, u)| \leq K_6 + K_7 \|x\|_2^{L_3} \tag{70}$$

$$\left\| \frac{\partial}{\partial x} c(x, u) \right\|_2 \leq K_6 + K_7 \|x\|_2^{L_3} \tag{71}$$

$$|h(x)| \leq K_6 + K_7 \|x\|_2^{L_3} \tag{72}$$

$$\left\| \frac{\partial}{\partial x} h(x) \right\|_2 \leq K_6 + K_7 \|x\|_2^{L_3}. \tag{73}$$

Remark 3. *Assumption 3 is to guarantee that the cost function is integrable with respect to stochastic observations, which are introduced in Assumption 4. Note that even though the above bound is not general enough to apply to all analytic functions, it does include all finite order polynomials of $\|x(t)\|_2$ and $\|u(t)\|_2$, for example, since $\|u(t)\|_2$ is bounded by Assumption 1.*

Assumption 4. Stochastic Observations. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $(Y_1, \ldots, Y_T)$ be a sequence of random vectors in $\mathbb{R}^{n_y}$ defined on this space, representing the sequence of observations. Assume that for each Y_i all the moments of the ℓ^2 norm is finite. That is,*

$$\forall i \in \{1, \ldots, T\}\ \forall k \in \mathbb{N}\ \ \mathbb{E}\left[\|Y_i\|_2^k\right] < \infty. \tag{74}$$

Definition 1. Perturbed Control. *Let $u \in U$ be a control. For $\tau \in (0, 1)$ and $v \in B(0, \rho_{\max})$, define the perturbed control u^ϵ by*

$$u^\epsilon(t) \triangleq \begin{cases} v & \textit{if } t \in (\tau - \epsilon, \tau] \\ u(t) & \textit{otherwise}, \end{cases} \tag{75}$$

where $\epsilon \in [0, \tau]$. By definition if $\epsilon = 0$ then u^ϵ is the same as u. We assume that the nominal control $u(t)$ is left continuous in t at $t = \tau$.

3.2 Main Results

The main result of the analysis is the following theorem.

Theorem 1. Mode Insertion Gradient. *Suppose that Assumptions 1 – 4 are satisfied. For a given (τ, v), let u^ϵ denote the perturbed control of the form (75). The perturbed control u^ϵ and the stochastic observations $(Y_1, \ldots, Y_T)$ result in the stochastic perturbed state trajectory x^ϵ. For such u^ϵ and x^ϵ, let us define the mode insertion gradient of the expected total cost as*

$$\frac{\partial_+}{\partial \epsilon} \mathbb{E}\left[\int_0^T c(x^\epsilon(t), u^\epsilon(t)) dt + h(x^\epsilon(T))\right]\Bigg|_{\epsilon=0}. \tag{76}$$

Then, this right derivative exists and we have

$$\begin{aligned} &\frac{\partial_+}{\partial \epsilon} \mathbb{E}\left[\int_0^T c(x^\epsilon(t), u^\epsilon(t)) dt + h(x^\epsilon(T))\right]\Bigg|_{\epsilon=0} \\ &\quad = c(x(\tau), v) - c(x(\tau), u(\tau)) \\ &\quad + \mathbb{E}\Bigg[\int_\tau^T \frac{\partial}{\partial x} c(x(t), u(t))^{\mathrm{T}} \Psi(t) dt \\ &\qquad + \frac{\partial}{\partial x} h(x(T))^{\mathrm{T}} \Psi(T)\Bigg], \end{aligned} \tag{77}$$

where $\Psi(t) = \frac{\partial_+}{\partial \epsilon} x^\epsilon(t)\big|_{\epsilon=0}$ is the state variation.

The proof of the theorem is deferred to Appendix A. One can see that the mode insertion gradient (76) is a natural generalization of the ones discussed in Egerstedt et al. (2006); Wardi and Egerstedt (2012); Ansari and Murphey (2016) to stochastic hybrid systems. Furthermore, by comparing (77) with (39) it is apparent that the right hand side of (77) is mathematically equivalent to $\mathbb{E}[\nu(t_f)]$, the quantity to be optimized with the SACBP algorithm in Section 2.

The fact that SACBP optimizes (76) leads to a certain performance guarantee of the algorithm. In the open-loop nominal control case, the term $\mathbb{E}[\nu(t_f)]$ as in (34) or (42) becomes 0 if the control perturbation v is equal to the nominal control $u(\tau)$. Therefore, as long as $u(\tau)$ is a feasible solution to (35) the optimal value is guaranteed to be less than or equal to zero. Furthermore, in expectation the actual value of $\mathbb{E}[\nu(t_f)]$ matches the one approximated with samples, since the Monte Carlo estimate is unbiased. In other words, the perturbation (τ^*, v^*) computed by the algorithm is expected to result in a non-positive mode insertion gradient. If the mode insertion gradient is negative, there always exists a sufficiently small $\epsilon > 0$ such that the expected total cost is decreased by the control perturbation. In the corner case that the mode insertion gradient is zero, one can set $\epsilon = 0$ to not perturb the control at all. Therefore, for an appropriate choice of ϵ the expected performance of the SACBP algorithm over the planning horizon is at least as good as that of the nominal control.

The same discussion holds for the case of closed-loop nominal control policies, when the expression for $\mathbb{E}[\nu(t_f)]$ is given by (44) or (45). This is because Theorem 1 still holds if the nominal control $u(t)$ is a closed-loop policy as stated in Remark 5 (Appendix A). Therefore, the expected worst-case performance of the algorithm is lower-bounded by that of the nominal policy. This implies that if a reasonable nominal policy is known, at run-time SACBP is expected to further improve it while synthesizing continuous-time control inputs efficiently.

4 Simulation Results

We evaluated the performance of SACBP in the following two simulation studies: (i) active multi-target tracking with range-only observations; (ii) object manipulation under model uncertainty. SACBP as well as other baseline methods were implemented in Julia*, except for T-LQG Rafieisakhaei et al. (2017) whose NLP problems were modeled by CasADi Andersson et al. (2019) in Python and then solved by Ipopt Wächter and Biegler (2006), a standard NLP solver based on interior-point methods. All the computation was performed on a desktop computer with Intel® Core™ i7-8750H CPU and 32.1GB RAM. The Monte Carlo sampling of SACBP was parallelized on multiple cores of the CPU.

4.1 Active Multi-Target Tracking with Range-only Observations

This problem focuses on pure information gathering, namely identifying where the moving targets are in the environment. In doing so, the surveillance robot modeled as a single integrator can only use relative distance observations. The robot's position p is fully observable and the transitions are deterministic. Assuming perfect data association, the observation for target i is $d_i = ||q_i - p + v_i||_2$, where q_i is the true target position and v_i is zero-mean Gaussian white noise with state-dependent covariance $R(p, q_i) = R_0 + ||q_i - p||_2 R_1$. We use $0.01 I_{2\times 2}$ for the nominal noise R_0. The range-dependent noise $R_1 = 0.001 I_{2\times 2}$ degrades the observation quality as the robot gets farther from the target. The discrete-time UKF is employed for state estimation in tracking 20 independent targets. The target dynamics are modeled by a 2D Brownian motion with covariance $Q = 0.1 I_{2\times 2}$. Similarly to Spinello and Stilwell (2010), an approximated observation covariance $R(p, \mu_i)$ is used in the filter to obtain tractable estimation results, where μ_i is the most recent mean estimate of q_i.

The SACBP algorithm produces a continuous robot trajectory over 200[s] with planning horizon $t_f - t_0 = 2[s]$, update interval $\Delta t_o = 0.2[s]$, perturbation duration $\epsilon = 0.16[s]$, and $N = 10$ Monte Carlo samples. The Euler scheme is used for integration with $\Delta t_c = 0.01[s]$. The Jacobians and the gradients are computed either analytically or using an automatic differentiation tool Revels et al. (2016) to retain both speed and precision. In this simulation $t_{\text{calc}} = 0.15[s]$ is assumed no matter how long the actual control update takes. We use $c(p, b, u) = 0.05 u^{\mathrm{T}} u$ for the running cost and $h(p, b) = \sum_{i=1}^{20} \exp(\text{entropy}(b_i))$ for the terminal cost, with an intention to reduce the worst-case uncertainty among the targets. This expression for $h(p, b)$ is equivalent to:

$$h(p, b) = \sum_{i=1}^{20} \sqrt{\det(2\pi e \Sigma_i)}, \tag{78}$$

where Σ_i is the covariance for the i-th target. The nominal control $u(t)$ is constantly zero.

We compared SACBP against four baseline methods: (i) a greedy algorithm based on the gradient descent of terminal cost h, similar to Schwager et al. (2017); (ii) MCTS-DPW Couëtoux et al. (2011); Egorov et al. (2017) in the Gaussian belief space; (iii) projection-based trajectory optimization for ergodic exploration Miller and Murphey (2013); Miller et al. (2016); Dressel and Kochenderfer (2018); (iv) T-LQG Rafieisakhaei et al. (2017). We also attempted to implement the belief iLQG van den Berg et al. (2012) algorithm, but the policy did not converge for this problem. We suspect that the non-convex terminal cost h contributed to this behavior, which in fact violates one of the underlying assumptions made in the paper van den Berg et al. (2012).

MCTS-DPW uses the same planning horizon as SACBP, however it draws $N = 25$ samples from the belief tree so the computation time of the two algorithms matches approximately.

For T-LQG, an NLP is formulated so that the optimization objective is a discrete-time equivalent of the SACBP objective, with the discrete time interval of $\Delta t_o = 0.2[s]$. Unfortunately, the Ipopt solver takes a significantly long time to solve this NLP; on average the optimization over 10 timesteps (thus 2[s]) takes 43.9[s] to converge to a local minimum, with the worst case of over 250[s], due to the high dimensionality of the joint state space. This high computation cost makes it prohibitive to derive an offline policy over the entire 200[s] simulation horizon prior to the execution. Therefore, we only test T-LQG in an online manner for this problem, with the same planning horizon as SACBP. For a fair comparison with the other methods, we adjust the max_cpu_time parameter of the solver so that the Ipopt iterations terminate within a computation time that is the same order of magnitude as SACBP, MCTS-DPW, and ergodic exploration. Note also that T-LQG is a closed-loop planning method that comes with a local LQG controller to track the optimized nominal belief trajectory. Designed to require minimal online computation, re-planning for T-LQG only happens if the end of the planning horizon is reached or the symmetric KL-divergence between the nominal belief state and the actual belief state exceeds a certain threshold d_{th}. We set d_{th} to $1e6$ for this problem, since we noticed that a smaller value would result in re-planning after almost each observation update. With our choice, re-planning almost only happens at the end of the planning horizon, which allows for efficient execution of the policy. More details about the re-planning for T-LQG can be found in Rafieisakhaei et al. (2016, 2017).

Ergodic exploration is not a belief space planning approach but has been used in the active sensing literature. Beginning with the nominal control of zero, it locally optimizes the ergodicity of the trajectory with respect to the spatial information distribution based on Fisher information. This optimization is open-loop since the effect of future observations is not considered. As a new observation becomes available, the distribution and the trajectory are recomputed.

All the controllers were saturated at the same limit. The results presented in Figure 1 clearly indicates superior performance of SACBP while achieving real-time computation. More notably, SACBP generated a trajectory that periodically revisited the two groups whereas other methods failed to do so (Figure 2). With SACBP the robot was moving into one of the four diagonal directions for most of the time. This is plausible, as SACBP solves the quadratic program with a box input constraint (35), which tends to find optimal solutions at the corners. MCTS-DPW resulted in a highly non-smooth trajectory and failed to fully explore the environment. The greedy approach improved

*The code for SACBP as well as the baseline methods is publicly available at https://github.com/StanfordMSL/SACBP.jl.

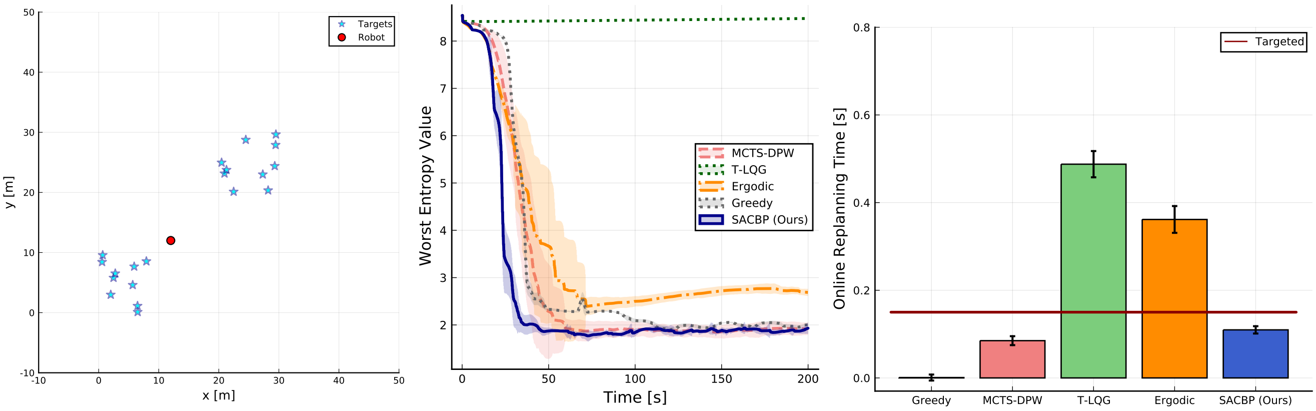

Figure 1. (Left) Simulation environment with 20 targets and a surveillance robot. (Middle) The history of the worst entropy value among the targets averaged over 20 random runs, plotted with the standard deviation. With the budget of 10 Monte Carlo samples, SACBP showed small variance for the performance curve and resulted in the fastest reduction of the worst entropy value compared to every other baseline. (Right) Computation times for Greedy, MCTS-DPW, and SACBP achieved real-time performance, taking less time than simulated $t_{\text{calc}} = 0.15[s]$.

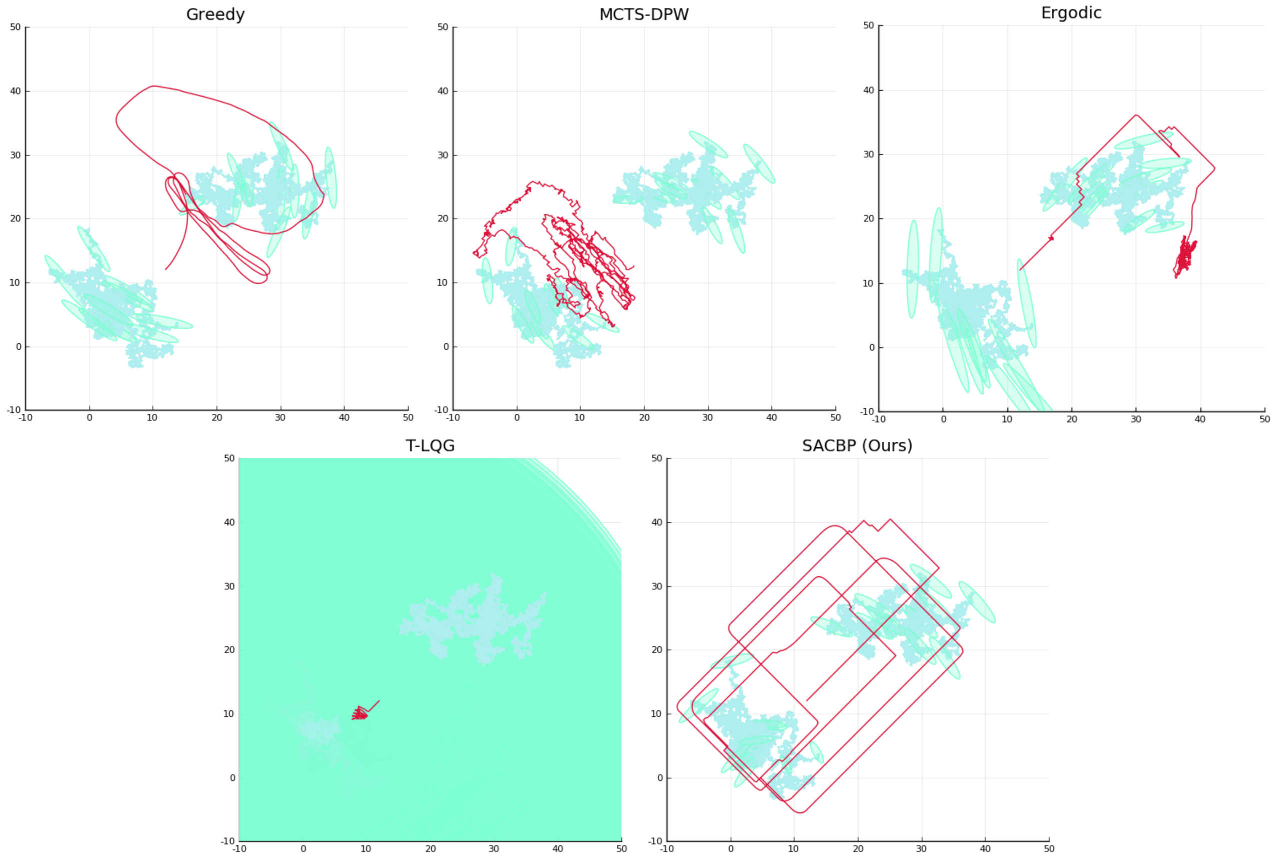

Figure 2. Sample robot trajectories (depicted in red) generated by each algorithm. Greedy, MCTS-DPW, and Ergodic did not result in a trajectory that fully covers the two groups of the targets. T-LQG failed to reduce the estimation uncertainty even after 200[s], due to insufficient time to solve the NLP with high-dimensional joint states in an online manner. SACBP successfully explored the space and periodically revisited both of the two target groups. With SACBP, the robot traveled into one of the four diagonal directions for most of the time. This is due to the fact that SACBP optimizes a convex quadratic under a box saturation constraint, which tends to find optimal solutions at the corners. In all the figures, the blue lines represent the target trajectories and the shaded ellipses are 99% error ellipses at $t = 200$[s].

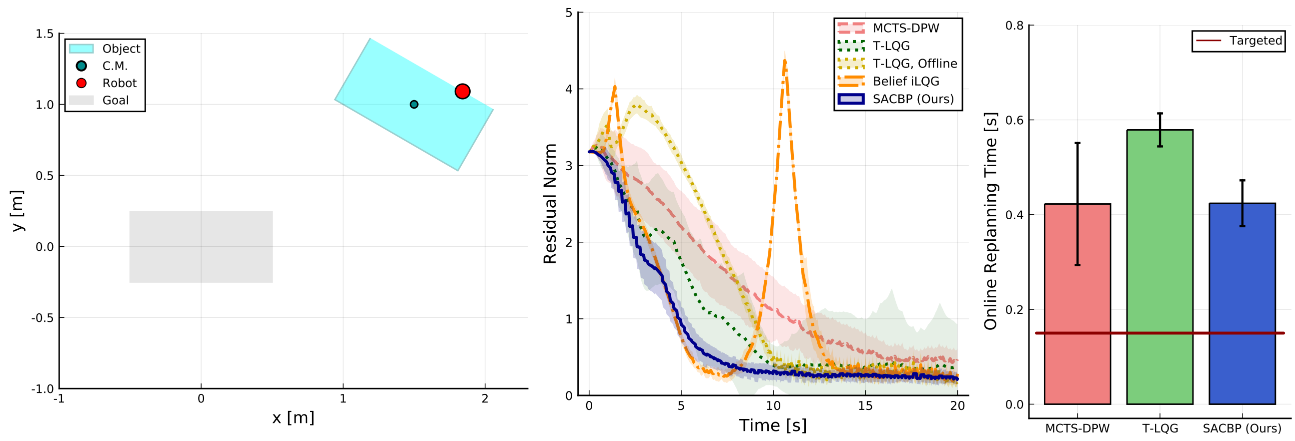

Figure 3. (Left) The robot is attached to the rectangular object. (Middle) The history of the l_2 norm of the residual between the goal state and the true object state averaged over 20 runs. SACBP with $N = 10$ samples successfully brought the object close to the goal. The reduction of the residual norm was much slower for MCTS-DPW. T-LQG was not as successful either, regardless of whether the policy was derived offline (without re-planning) or online (with re-planning), although it eventually achieved similar residual norms to SACBP. Belief iLQG resulted in large overshoots at around 2[s] and 11[s]. (Right) Computation time of SACBP was increased from the multi-target tracking problem due to increased complexity related to the continuous-discrete belief dynamics, but still achieved a reasonable value. Note that the computation times of the offline algorithms were significantly longer and are not shown in this plot.

the smoothness, but the robot eventually followed a cyclic trajectory in a small region of the environment. To our surprise, the ergodic method did not generate a trajectory that covers the two groups of the targets. This is likely due to the use of a projection-based trajectory optimization method, which has been recently found to perform rather poorly with rapid re-planning Dressel (2018). Among all the baselines implemented, T-LQG performed the worst, leaving large estimation covariances even after the full 200[s]. This is due to the insufficient time budget to solve the NLP, which indicates that T-LQG is not suited for high-dimensional belief space planning problems such as this active multi-target tracking task.

4.2 Object Manipulation under Model Uncertainty

This problem is identical to the model-based Bayesian reinforcement learning problem studied in Slade et al. (2017), therefore a detailed description of the nonlinear dynamics and the observation models are omitted. See Figure 3 for the illustration of the environment. A 2D robot attached to a rigid body object applies forces and torques to move the object to the origin. The object's mass, moment of inertia, moment arm lengths, and linear friction coefficient are unknown. These parameters as well as the object's 2D state need to be estimated using EKF, with noisy sensors which measure the robot's position, velocity, and acceleration in the global frame. The same values for $t_f - t_0$, Δt_o, Δt_c, t_{calc} as in the previous problem are assumed. Each simulation is run for 20[s]. SACBP uses $\epsilon = 0.04[s]$ and $N = 10$. The nominal control for SACBP is a position controller whose input is the mean x-y position and the rotation estimates of the object. The cost function is quadratic in the true state x and control u, given by $\frac{1}{2}x^{\mathrm{T}}C_x x + \frac{1}{2}u^{\mathrm{T}}C_u u$. Taking expectations yields the equivalent cost in the Gaussian belief space $c(b, u) = \frac{1}{2}\mu^{\mathrm{T}}C_x\mu + \frac{1}{2}\text{tr}(C_x\Sigma) + \frac{1}{2}u^{\mathrm{T}}C_u u$, where Σ is the covariance matrix. We let terminal cost h be the same as c except that it is without the control term.

We compared SACBP against (i) MCTS-DPW in the Gaussian belief space, (ii) belief iLQG, and (iii) T-LQG. MCTS-DPW uses the same planning horizon as SACBP, and is set to draw $N = 240$ samples so that the computation time is approximately equal to that of SACBP. Furthermore, MCTS-DPW uses the position controller mentioned above as the rollout policy, which is suggested in Slade et al. (2017).

The policy computation with belief iLQG is performed offline over the entire simulation horizon of 20[s]. The solver is initialized with a nominal trajectory generated by the same position controller. The entire computation of the policy takes 5945[s], a significant amount of time until convergence to a locally optimal affine feedback policy. Note however that the online policy execution can be performed instantaneously.

For T-LQG, we test the algorithm in both the online and the offline modes. The online mode is equivalent to the implementation for the active multi-target tracking problem. Re-planning happens when the end of the planning horizon is reached or the symmetric KL-divergence surpasses $d_{\text{th}} = 25$, which was rarely exceeded during the simulation for this problem. Furthermore, the max_cpu_time parameter of Ipopt is adjusted so the computation time is comparable to SACBP and MCTS-DPW. This is because the full optimization over 10 timesteps (i.e. 2[s]) takes 2.5[s] on average and 7[s] in the worst case, which is better than in the active multi-target tracking problem but still prohibitively slow for online control computation, taking about $15 \times t_{\text{calc}}$ to $45 \times t_{\text{calc}}$ with $t_{\text{calc}} = 0.15[s]$. On the other hand, the offline mode computes the closed-loop policy once for the entire simulation horizon without online re-planning or limiting max_cpu_time. This setup is identical to belief iLQG. In this mode, it takes T-LQG 1065[s] to compute the policy, which is about 5 to 6 times faster than belief iLQG. This improved efficiency is congruous with the complexity analysis provided in Rafieisakhaei et al. (2017). Note also that we use the aforementioned position controller to initialize the NLP solver in both the online and the offline modes.

Overall, the results presented in Figure 3 demonstrate that SACBP outperformed all the baselines in this task with only 10 Monte Carlo samples, bringing the object close to the goal within 10[s]. Although the computation time increased from the previous problem due to the continuous-discrete filtering, it still achieved near real-time performance with less than $3 \times t_{\mathrm{calc}}$[s] on average. Compared to SACBP, reduction of the residual norm was slower for MCTS-DPW and online T-LQG. The two offline algorithms tested, belief iLQG and offline T-LQG, both had a large overshoot at around 2[s] and 3[s], respectively. We suppose that this was caused by the offline nature of the policies, as well as a mismatch between the discrete-time model used for planning and the continuous-time model employed for dynamics simulation. Another large overshoot for belief iLQG at 11[s] was likely due to a locally optimal behavior of the iLQG solver.

5 Conclusions and Future Work

In this paper we present SACBP, a novel belief space planning algorithm for continuous-time dynamical systems. We view the stochastic belief dynamics as a hybrid system with time-driven switching and derive the optimal control perturbation based on the perturbation theory of differential equations. The resulting algorithm extends the framework of SAC to stochastic belief dynamics and is highly parallelizable to run in near real-time. The rigorous mathematical analysis shows that the notion of mode insertion gradient can be generalized to stochastic hybrid systems, which leads to the property of SACBP that the algorithm is expected to perform at least as good as the nominal policy with an appropriate choice of the perturbation duration. Through an extensive simulation study, we have confirmed that SACBP outperforms other algorithms including a greedy algorithm and non-myopic closed-loop planners that are based on approximate dynamic programming and/or local trajectory optimization. In future work, we are interested to consider a distributed multi-robot version of SACBP as well as problems with hard state constraints. We also plan to provide additional case studies for more complex belief distributions with efficient implementation.

Funding

Toyota Research Institute ("TRI") provided funds to assist the authors with their research but this article solely reflects the opinions and conclusions of its authors and not TRI or any other Toyota entity. This work was also supported in part by NSF grant CMMI1562335, ONR grant N00014-18-1-2830, a JASSO fellowship, and a Masason Foundation fellowship. The authors are grateful for this support.

References

Agha-mohammadi A, Agarwal S, Kim S, Chakravorty S and Amato NM (2018) Slap: Simultaneous localization and planning under uncertainty via dynamic replanning in belief space. *IEEE Transactions on Robotics* 34(5): 1195–1214.

Agha-Mohammadi AA, Chakravorty S and Amato NM (2014) Firm: Sampling-based feedback motion-planning under motion uncertainty and imperfect measurements. *The International Journal of Robotics Research* 33(2): 268–304.

Andersson JAE, Gillis J, Horn G, Rawlings JB and Diehl M (2019) CasADi – A software framework for nonlinear optimization and optimal control. *Mathematical Programming Computation* 11(1): 1–36. DOI:10.1007/s12532-018-0139-4.

Ansari AR and Murphey TD (2016) Sequential Action Control: Closed-Form Optimal Control for Nonlinear and Nonsmooth Systems. *IEEE Transactions on Robotics* 32(5): 1196–1214.

Bajcsy R (1988) Active perception. *Proceedings of the IEEE* 76(8): 966–1005.

Bangura M and Mahony R (2014) Real-time model predictive control for quadrotors. *IFAC Proceedings Volumes* 47(3): 11773–11780.

Bourbaki N and Spain P (2004) *Elements of Mathematics Functions of a Real Variable: Elementary Theory*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Bourgault F, Makarenko A, Williams S, Grocholsky B and Durrant-Whyte H (2002) Information based adaptive robotic exploration. In: *IEEE/RSJ International Conference on Intelligent Robots and System*, volume 1. IEEE, pp. 540–545.

Browne CB, Powley E, Whitehouse D, Lucas SM, Cowling PI, Rohlfshagen P, Tavener S, Perez D, Samothrakis S and Colton S (2012) A survey of monte carlo tree search methods. *IEEE Transactions on Computational Intelligence and AI in Games* 4(1): 1–43.

Chaudhari P, Karaman S, Hsu D and Frazzoli E (2013) Sampling-based algorithms for continuous-time pomdps. In: *2013 American Control Conference*. IEEE, pp. 4604–4610.

Couëtoux A, Hoock JB, Sokolovska N, Teytaud O and Bonnard N (2011) Continuous Upper Confidence Trees. In: *2011 International Conference on Learning and Intelligent Optimization*. Springer, Berlin, Heidelberg, pp. 433–445.

Diestel J and Uhl J (1977) *Vector Measures*. American Mathematical Society.

Dressel L and Kochenderfer MJ (2018) Tutorial on the generation of ergodic trajectories with projection-based gradient descent. *IET Cyber-Physical Systems: Theory & Applications* .

Dressel LK (2018) *Efficient and Low-cost Localization of Radio Sources with an Autonomous Drone*. PhD Thesis, Stanford University.

Egerstedt M, Wardi Y and Axelsson H (2006) Transition-time optimization for switched-mode dynamical systems. *IEEE Transactions on Automatic Control* 51(1): 110–115.

Egorov M, Sunberg ZN, Balaban E, Wheeler TA, Gupta JK and Kochenderfer MJ (2017) Pomdps. jl: A framework for sequential decision making under uncertainty. *Journal of Machine Learning Research* 18(26): 1–5.

Elijah P (1997) *Optimization: Algorithms and Consistent Approximations*. Springer Verlage Publications.

Erez T and Smart WD (2010) A scalable method for solving high-dimensional continuous pomdps using local approximation. In: *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*, UAI'10. Arlington, Virginia, United States: AUAI Press, pp. 160–167.

Gowrisankaran K (1972) Measurability of functions in product spaces. *Proceedings of the American Mathematical Society* 31(2): 485–488.

Heemels W, Lehmann D, Lunze J and Schutter BD (2009) Introduction to hybrid systems. In: Lunze J and Lamnabhi-Lagarrigue F (eds.) *Handbook of Hybrid Systems Control – Theory, Tools, Applications*, chapter 1. Cambridge University Press, pp. 3–30.

Hollinger GA and Sukhatme GS (2014) Sampling-based robotic information gathering algorithms. *The International Journal of Robotics Research* 33(9): 1271–1287.

Huber M (2009) *Probabilistic framework for sensor management*, volume 7. KIT Scientific Publishing.

Kaelbling LP, Littman ML and Cassandra AR (1998) Planning and acting in partially observable stochastic domains. *Artif. Intell.* 101(1-2): 99–134.

Khalil HK and Grizzle JW (2002) *Nonlinear systems*, volume 3. Prentice hall Upper Saddle River, NJ.

Kochenderfer MJ (2015) *Decision making under uncertainty: theory and application*. MIT press.

Kurniawati H, Hsu D and Lee WS (2008) Sarsop: Efficient point-based pomdp planning by approximating optimally reachable belief spaces. In: *Robotics: Science and Systems (RSS)*. Zurich, Switzerland.

Le Ny J and Pappas GJ (2009) On trajectory optimization for active sensing in Gaussian process models. In: *Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*. IEEE, pp. 6286–6292.

Liberzon D (2011) *Calculus of variations and optimal control theory: A concise introduction*. Princeton University Press.

Madani O, Hanks S and Condon A (1999) On the undecidability of probabilistic planning and infinite-horizon partially observable markov decision problems. In: *AAAI/IAAI*. pp. 541–548.

Mavrommati A, Tzorakoleftherakis E, Abraham I and Murphey TD (2018) Real-time area coverage and target localization using receding-horizon ergodic exploration. *IEEE Transactions on Robotics* : 62–80.

Mihaylova L, Lefebvre T, Bruyninckx H, Gadeyne K and De Schutter J (2002) Active Sensing for Robotics - A Survey. In: *Proceedings of 5th International Conference on Numerical Methods and Applications*. pp. 316–324.

Miller LM and Murphey TD (2013) Trajectory optimization for continuous ergodic exploration. In: *American Control Conference (ACC), 2013*. IEEE, pp. 4196–4201.

Miller LM, Silverman Y, MacIver MA and Murphey TD (2016) Ergodic exploration of distributed information. *IEEE Transactions on Robotics* 32(1): 36–52.

Nishimura H and Schwager M (2018a) Active motion-based communication for robots with monocular vision. In: *2018 IEEE International Conference on Robotics and Automation (ICRA)*. pp. 2948–2955.

Nishimura H and Schwager M (2018b) Sacbp: Belief space planning for continuous-time dynamical systems via stochastic sequential action control. In: *The 13th International Workshop on the Algorithmic Foundations of Robotics (WAFR)*. Mérida, México.

Papadimitriou CH and Tsitsiklis JN (1987) The complexity of markov decision processes. *Mathematics of operations research* 12(3): 441–450.

Patil S, Kahn G, Laskey M, Schulman J, Goldberg K and Abbeel P (2014) Scaling up gaussian belief space planning through covariance-free trajectory optimization and automatic differentiation. In: *WAFR*, *Springer Tracts in Advanced Robotics*, volume 107. Springer, pp. 515–533.

Platt R (2013) Convex receding horizon control in non-gaussian belief space. In: *Algorithmic Foundations of Robotics X*. Springer, pp. 443–458.

Platt R, Tedrake R, Kaelbling L and Lozano-Perez T (2010) Belief space planning assuming maximum likelihood observations. In: *Robotics: Science and Systems (RSS)*.

Popović M, Hitz G, Nieto J, Sa I, Siegwart R and Galceran E (2017) Online informative path planning for active classification using uavs. In: *2017 IEEE International Conference on Robotics and Automation (ICRA)*. pp. 5753–5758.

Rafieisakhaei M, Chakravorty S and Kumar P (2016) Belief space planning simplified: Trajectory-optimized lqg (t-lqg). *arXiv preprint arXiv:1608.03013* .

Rafieisakhaei M, Chakravorty S and Kumar PR (2017) T-lqg: Closed-loop belief space planning via trajectory-optimized lqg. In: *2017 IEEE International Conference on Robotics and Automation (ICRA)*. pp. 649–656.

Revels J, Lubin M and Papamarkou T (2016) Forward-mode automatic differentiation in julia. *arXiv:1607.07892 [cs.MS]* .

Schwager M, Dames P, Rus D and Kumar V (2017) A multi-robot control policy for information gathering in the presence of unknown hazards. In: *Robotics Research : The 15th International Symposium ISRR*. Springer International Publishing, pp. 455–472.

Seekircher A, Laue T and Röfer T (2011) Entropy-based active vision for a humanoid soccer robot. In: *RoboCup 2010: Robot Soccer World Cup XIV*, volume 6556 LNCS. Springer Berlin Heidelberg, pp. 1–12.

Slade P, Culbertson P, Sunberg Z and Kochenderfer M (2017) Simultaneous active parameter estimation and control using sampling-based bayesian reinforcement learning. In: *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. pp. 804–810.

Somani A, Ye N, Hsu D and Lee WS (2013) Despot: Online pomdp planning with regularization. In: *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'13. USA: Curran Associates Inc., pp. 1772–1780.

Spinello D and Stilwell DJ (2010) Nonlinear estimation with state-dependent gaussian observation noise. *IEEE Transactions on Automatic Control* 55(6): 1358–1366.

Sunberg Z and Kochenderfer MJ (2017) POMCPOW: an online algorithm for pomdps with continuous state, action, and observation spaces. *CoRR* abs/1709.06196.

van den Berg J, Patil S and Alterovitz R (2012) Motion planning under uncertainty using iterative local optimization in belief space. *The International Journal of Robotics Research* 31(11): 1263–1278.

Wächter A and Biegler LT (2006) On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical programming* 106(1): 25–57.

Wardi Y and Egerstedt M (2012) Algorithm for optimal mode scheduling in switched systems. In: *2012 American Control Conference (ACC)*. pp. 4546–4551.

Williams G, Drews P, Goldfain B, Rehg JM and Theodorou EA (2018) Information-theoretic model predictive control: Theory and applications to autonomous driving. *IEEE Transactions on Robotics* 34(6): 1603–1622.

Xie L, Popa D and Lewis FL (2007) *Optimal and robust estimation: with an introduction to stochastic control theory*. CRC press.

A Detailed Analysis of Mode Insertion Gradient for Stochastic Hybrid Systems

In this appendix, we provide a thorough analysis of the stochastic hybrid systems with time-driven switching that satisfy Assumptions 1 – 4. Our goal is to prove Theorem 1.

A.1 Nominal Trajectory under Specific Observations

First, we analyze the properties of the system $x(t)$ for $t \in [0, T]$ under a given initial condition x_0, control $u \in U$, and a specific sequence of observations $(y_1, \dots, y_T)$ sampled from $(Y_1, \dots, Y_T)$.

Proposition 3. Existence and Uniqueness of Solutions. *Given a control $u \in U$ and a sequence of observations $(y_1, \dots, y_T)$, the system $x(t)$ starting at x_0 has a unique solution for $t \in [0, T]$.*

Proof. We will show that each x_i for $i \in \{1, \dots, T\}$ has a unique solution, and thus x is uniquely determined as a whole. First, by Assumption (2a), (2c) and the Picard Lemma (Lemma 5.6.3 in Elijah (1997)), the differential equation

$$\dot{x}_1(t) = f(x_1(t), u(t)) \tag{79}$$

with initial condition $x_1(0) = x_0$ has a solution for $t \in [0, 1]$. Furthermore, Proposition 5.6.5 in Elijah (1997) assures that the solution x_1 is unique under Assumption (2a) and (2c). This guarantees that the initial condition for x_2 defined by $x_2(1) = g(x_1(1), y_1)$ is unique. Therefore, proceeding by induction each $x_1, \dots, x_T$ has a unique solution, which completes the proof.

Corollary 4. Right Continuity. *Given a control $u \in U$ and a sequence of observations $(y_1, \dots, y_T)$, the system $x(t)$ starting at x_0 is right continuous in t on $[0, T]$.*

Proof. By Proposition 3 each x_i has a unique solution that follows $\dot{x}_i = f(x, u)$. Clearly each x_i is continuous on $[i-1, i]$, which proves that $x(t) \triangleq x_i(t)\ \forall t \in [i-1, i)\ \forall i \in \{1, 2, \dots, T\}$ with $x(T) \triangleq g(x_T(T), y_T)$ is right continuous on $[0, T]$.

Lemma 5. *Let ξ_i denote the initial condition for x_i. Then, there exists a constant $K_8 < \infty$ such that for all $i \in \{1, \dots, T\}$,*

$$\forall t \in [i-1, i] \quad \|x_i(t)\|_2 \le (1 + \|\xi_i\|_2)\, e^{K_8} \tag{80}$$

Proof. Using Assumption (2a) and (2c), the claim follows directly from Proposition 5.6.5 in Elijah (1997).

Proposition 6. Lipschitz Continuity. *For each $i \in \{1, \dots, T\}$, let ξ_i' and ξ_i'' be two distinct initial conditions for x_i. Furthermore, let u' and u'' be two controls from U. The pairs (ξ', u') and (ξ'', u'') respectively define two solutions x_i' and x_i'' to ODE $\dot{x}_i = f(x_i, u)$ over $[i-1, i]$. Then, there exists an $L < \infty$, independent of ξ_i', ξ_i'', u' and u'', such that*

$$\begin{aligned}\forall i \in \{1, \dots, T\}\ \forall t \in [i-1, i] \quad & \|x_i'(t) - x_i''(t)\|_2 \le \\ & L\left(\|\xi_i' - \xi_i''\|_2 + \int_{i-1}^{i} \|u'(t) - u''(t)\|_2 dt\right). \end{aligned}\tag{81}$$

Proof. The proof is similar to that of Lemma 5.6.7 in Elijah (1997). Making use of the Picard Lemma (Lemma 5.6.3 in Elijah (1997)) and Assumption (2c), we obtain

$$\begin{aligned}&\|x_i'(t) - x_i''(t)\|_2 \le \\ &e^{K_1}\left(\|\xi_i' - \xi_i''\|_2 + K_1 \int_{i-1}^{i} \|u'(t) - u''(t)\|_2 dt\right).\end{aligned}\tag{82}$$

As $K_1 \ge 1$,

$$\begin{aligned}&\|x_i'(t) - x_i''(t)\|_2 \le \\ &K_1 e^{K_1}\left(\|\xi_i' - \xi_i''\|_2 + \int_{i-1}^{i} \|u'(t) - u''(t)\|_2 dt\right).\end{aligned}\tag{83}$$

Defining $L \triangleq K_1 e^{K_1} < \infty$ completes the proof.

Corollary 7. Uniform Continuity in Initial Conditions. *Let $u \in U$ be a given control. For each $i \in \{1, \dots, T\}$, let ξ_i' and ξ_i'' be two distinct initial conditions for x_i. The pairs (ξ', u) and (ξ'', u) respectively define two solutions x_i' and x_i'' to ODE $\dot{x}_i = f(x_i, u)$ over $[i-1, i]$. Note that they share the same control but have different initial conditions, unlike Proposition 6. Then, for any $\epsilon > 0$ there exists $\delta > 0$ such that*

$$\begin{aligned}\forall i \in \{1, \dots, T\}\ \forall t \in [i-1, i] \quad & \|\xi_i' - \xi_i''\|_2 < \delta \\ & \Rightarrow \|x_i'(t) - x_i''(t)\|_2 < \epsilon.\end{aligned}\tag{84}$$

Proof. By Proposition 6, we find

$$\|x_i'(t) - x_i''(t)\|_2 \le L\|\xi_i' - \xi_i''\|_2 \tag{85}$$

for all $i \in \{1, \dots, T\}$ and $t \in [i-1, i]$, where $L < \infty$. Take any $\delta < \frac{\epsilon}{L}$ to prove the claim.

Proposition 8. Continuity in Observations. *Given a control $u \in U$, the map $(y_1, \dots, y_T) \mapsto x(t)$ is continuous for all $t \in [0, T]$, where x represents the solution to the system under Assumption 2 starting at $x(0) = x_0$.*

Proof. We will show the continuity of $(y_1, \dots, y_T) \mapsto x_i(t)$ for each $i \in \{1, \dots, T\}$ by mathematical induction. First, by Assumption 2 the value of $x_1(t)$ is solely determined by x_0 and u. Therefore, for any $t \in [0, 1]$ the function $(y_1, \dots, y_T) \mapsto x_1(t)$ is a constant map, which is continuous. Next, suppose that $\forall t \in [i-1, i]\ (y_1, \dots, y_T) \mapsto x_i(t)$ is continuous for some $i \in \{1, \dots, T-1\}$. Now consider x_{i+1}. Let $F_{i+1}(\xi_{i+1}, t)$ be the map from an initial condition $x_{i+1}(i) = \xi_{i+1}$ to the solution x_{i+1} at $t \in [i, i+1]$ under the given u. In other words, $F_{i+1}(\xi_{i+1}, t)$ is equivalent to the integral equation

$$F_{i+1}(\xi_{i+1}, t) \triangleq \xi_{i+1} + \int_i^t f(x_{i+1}(a), u(a))da \tag{86}$$

and takes initial condition ξ_{i+1} as well as time t as its arguments. Substituting $\xi_{i+1} = g(x_i(i), y_i)$, $F_{i+1}\left(g(x_i(i), y_i), t\right)$ gives the actual value of $x_{i+1}(t)$. We will prove the continuity of $F_{i+1}(g(x_i(i), y_i), t)$ as follows. First, note that the map from $(y_1, \dots, y_T)$ to $F_{i+1}\left(g(x_i(i), y_i), t\right)$ is the result of the composition:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_T \end{pmatrix} \mapsto \begin{pmatrix} x_i(i) \\ y_i \end{pmatrix} \mapsto g(x_i(i), y_i) \mapsto F_{i+1}(g(x_i(i), y_i), t). \tag{87}$$

The first map is continuous since $x_i(i)$ is continuous in $(y_1, \dots, y_T)$ by the induction hypothesis. The second map is also continuous by Assumption (2b). Lastly, Corollary 7 shows that $F_{i+1}(\xi_{i+1}, t)$ is (uniformly) continuous in ξ_{i+1} for $t \in [i, i+1]$. Therefore, $(y_1, \dots, y_T) \mapsto x_{i+1}(t)$ is continuous for all $t \in [i, i+1]$.

Proposition 9. Bounded State Trajectory. *Given a control $u \in U$ and a sequence of observations $(y_1, \dots, y_T)$, the system $x(t)$ starting at $x(0) = x_1(0) = x_0$ has the following bound:*

$$\forall i \in \{2, \dots, T\}\ \forall t \in [i-1, i]$$

$$\|x_i(t)\|_2 \leq \sum_{(j_1, \dots, j_{i-1}) \in \mathcal{K}_i} \alpha_i^{(j_1, \dots, j_{i-1})}(x_0) \prod_{m=1}^{i-1} \|y_m\|_2^{j_m}, \tag{88}$$

where $\mathcal{K}_i$ is a finite set of sequences of non-negative integers of length $i-1$, and $\alpha_i^{(j_1, \dots, j_{i-1})}(x_0)$ is a finite positive constant that depends on x_0 and $(j_1, \dots, j_{i-1})$ but not on any of the observations or the control.

For $i = 1$ the bound is given by $\forall t \in [0, 1]$ $\|x_1(t)\|_2 \leq \alpha_1(x_0)$ for some finite positive constant $\alpha_1(x_0)$.

Proof. For $i = 1$, Lemma 5 gives $\forall t \in [0, 1]$ $\|x_1(t)\|_2 \leq (1 + \|x_0\|_2)\, e^{K_8} \triangleq \alpha_1(x_0)$. For $i = 2$, by Assumption (2d) and the case for $i = 1$, we have

$$\|x_2(1)\|_2 = \|g(x_1(1), y_1)\|_2 \tag{89}$$

$$\begin{aligned} &\leq K_2 + K_3\|x_1(1)\|_2^{L_1} + K_4\|y_1\|_2^{L_2} \\ &\qquad + K_5\|x_1(1)\|_2^{L_1}\|y_1\|_2^{L_2} \end{aligned} \tag{90}$$

$$\begin{aligned} &\leq K_2 + K_3\alpha_1(x_0)^{L_1} + K_4\|y_1\|_2^{L_2} \\ &\qquad + K_5\alpha_1(x_0)^{L_1}\|y_1\|_2^{L_2}. \end{aligned} \tag{91}$$

Then, by Lemma 5, $\forall t \in [1, 2]$

$$\|x_2(t)\|_2 \leq (1 + \|x_2(1)\|_2)e^{K_8} \tag{92}$$

$$\begin{aligned} &\leq e^{K_8}\Big(1 + K_2 + K_3\alpha_1(x_0)^{L_1} + K_4\|y_1\|_2^{L_2} \\ &\qquad + K_5\alpha_1(x_0)^{L_1}\|y_1\|_2^{L_2}\Big) \end{aligned} \tag{93}$$

$$\triangleq \sum_{(j_1) \in \mathcal{K}_2} \alpha_2^{(j_1)}(x_0)\|y_1\|_2^{j_1}, \tag{94}$$

where $\mathcal{K}_2 = \{(0), (L_2)\}$, and

$$\alpha_2^{(0)}(x_0) = e^{K_8}\left(1 + K_2 + K_3\alpha_1(x_0)^{L_1}\right) \tag{95}$$

$$\alpha_2^{(L_2)}(x_0) = e^{K_8}\left(K_4 + K_5\alpha_1(x_0)^{L_1}\right) \tag{96}$$

are both finite positive constants that depend on x_0 but not on any of the observations or the control.

Next, suppose that the claim holds for some $i \geq 2$. That is,

$$\forall t \in [i-1, i]$$

$$\|x_i(t)\|_2 \leq \sum_{(j_1, \dots, j_{i-1}) \in \mathcal{K}_i} \alpha_i^{(j_1, \dots, j_{i-1})}(x_0) \prod_{m=1}^{i-1} \|y_m\|_2^{j_m}, \tag{97}$$

where $\mathcal{K}_i$ and $\alpha_i^{(j_1, \dots, j_{i-1})}(x_0)$ are as defined in the statement of the proposition. Making use of this assumption, Assumption (2d) and Lemma 5, we find that for all $t \in [i, i+1]$,

$$\|x_{i+1}(t)\|_2 \leq (1 + \|g(x_i(i), y_i)\|_2)e^{K_8} \tag{98}$$

$$\begin{aligned} &\leq e^{K_8}(1 + K_2) + e^{K_8}K_4\|y_i\|_2^{L_2} \\ &\qquad + e^{K_8}\|x_i(i)\|_2^{L_1}(K_3 + K_5\|y_i\|_2^{L_2}). \end{aligned} \tag{99}$$

The first two terms in the above sum can be rewritten as

$$e^{K_8}(1 + K_2) = e^{K_8}(1 + K_2)\|y_1\|_2^0 \times \dots \times \|y_i\|_2^0 \tag{100}$$

$$e^{K_8}K_4\|y_i\|_2^{L_2} = e^{K_8}K_4\|y_1\|_2^0 \times \dots \times \|y_{i-1}\|_2^0 \times \|y_i\|_2^{L_2} \tag{101}$$

For the last term, we can use (97) and the multinomial theorem to write

$$\|x_i(i)\|_2^{L_1}$$

$$\leq \left(\sum_{(j_1, \dots, j_{i-1}) \in \mathcal{K}_i} \alpha_i^{(j_1, \dots, j_{i-1})}(x_0) \prod_{m=1}^{i-1} \|y_m\|_2^{j_m} \right)^{L_1} \tag{102}$$

$$\begin{aligned} &= \sum_{k_1 + \dots + k_{|\mathcal{K}_i|} = L_1} \binom{L_1}{k_1, \dots, k_{|\mathcal{K}_i|}} \\ &\quad \times \left\{ \prod_{l=1}^{|\mathcal{K}_i|} \alpha_i^{(j_1^{(l)}, \dots, j_{i-1}^{(l)})}(x_0)^{k_l} \right\} \times \prod_{m=1}^{i-1} \|y_m\|_2^{\sum_{l=1}^{|\mathcal{K}_i|} k_l j_m^{(l)}}, \end{aligned} \tag{103}$$

where $(j_1^{(l)}, \dots, j_{i-1}^{(l)})$ is the l-th element in $\mathcal{K}_i$. Note that k_l is non-negative for all $l \in \{1, \dots, |\mathcal{K}_i|\}$. By the induction hypothesis (97), exponent $\sum_{l=1}^{|\mathcal{K}_i|} k_l j_m^{(l)}$ is also non-negative for all $m \in \{1, \dots, i-1\}$.

Thus, substituting (100), (101), and (103) into (99), rearranging the sums and re-labeling the sequences of integer exponents, we can write

$$\|x_{i+1}(t)\|_2 \leq \sum_{(j_1, \dots, j_i) \in \mathcal{K}_{i+1}} \alpha_{i+1}^{(j_1, \dots, j_i)}(x_0) \prod_{m=1}^{i} \|y_m\|_2^{j_m} \tag{104}$$

for all $t \in [i, i+1]$, where $\mathcal{K}_{i+1}$ is a set of sequences of non-negative integers of length i, and each $\alpha_{i+1}^{(j_1, \dots, j_i)}(x_0)$ does not depend on any of the observations or the control. Here the cardinality of $\mathcal{K}_{i+1}$ is at most finite, since

$$|\mathcal{K}_{i+1}| \leq 2 + 2\binom{L_1 + |\mathcal{K}_i| - 1}{|\mathcal{K}_i| - 1} \tag{105}$$

by (99) and the multinomial theorem.

Finally, proceeding by mathematical induction over $i \in \{2, \dots, T\}$ completes the proof.

Proposition 10. Bounded Cost Functions. *Given a control $u \in U$ and a sequence of observations $(y_1, \dots, y_T)$, the instantaneous cost $c(x(t), u(t))$ induced by the state trajectory $x(t)$ has the following bound.*

$$\forall i \in \{2, \dots, T\}\ \forall t \in [i-1, i]\ \ |c(x_i(t), u(t))| \\ \leq \sum_{(j_1,\dots,j_{i-1}) \in \mathcal{K}'_i} \alpha_i'^{(j_1,\dots,j_{i-1})}(x_0) \prod_{m=1}^{i-1} \|y_m\|_2^{j_m}, \tag{106}$$

where $\mathcal{K}'_i$ is a finite set of sequences of non-negative integers of length $i-1$, and $\alpha_i'^{(j_1,\dots,j_{i-1})}(x_0)$ is a finite positive constant that depends on x_0 and $(j_1, \dots, j_{i-1})$ but not on any of the observations or the control.

For $i = 1$ the bound is given by

$$\forall t \in [0, 1]\ \ |c(x_1(t), u(t))| \leq \alpha'_1(x_0) \tag{107}$$

for some finite positive constant $\alpha'_1(x_0)$.

Similarly, the terminal cost $h(x(T))$ is bounded by

$$|h(x(T))| \leq \sum_{(j_1,\dots,j_T) \in \mathcal{K}'_{T+1}} \alpha_{T+1}'^{(j_1,\dots,j_T)}(x_0) \prod_{m=1}^{T} \|y_m\|_2^{j_m}. \tag{108}$$

Proof. For the instantaneous cost function $c(x_i(t), u(t))$, Assumption 3 along with Proposition 9 yields the following bound:

$$\forall i \in \{2, \dots, T\}\ \forall t \in [i-1, i]\ \ |c(x_i(t), u(t))| \leq K_6 \\ + K_7 \left(\sum_{(j_1,\dots,j_{i-1}) \in \mathcal{K}_i} \alpha_i^{(j_1,\dots,j_{i-1})}(x_0) \prod_{m=1}^{i-1} \|y_m\|_2^{j_m} \right)^{L_3}. \tag{109}$$

Making use of the multinomial expansion formula in the same manner as in the proof of Proposition 9, we conclude that

$$\forall i \in \{2, \dots, T\}\ \forall t \in [i-1, i]\ \ |c(x_i(t), u(t))| \\ \leq \sum_{(j_1,\dots,j_{i-1}) \in \mathcal{K}'_i} \alpha_i'^{(j_1,\dots,j_{i-1})}(x_0) \prod_{m=1}^{i-1} \|y_m\|_2^{j_m} \tag{110}$$

for some finite set $\mathcal{K}'_i$ of sequences of non-negative integers and finite positive constants $\alpha_i'^{(j_1,\dots,j_{i-1})}(x_0)$. Similarly, for $i = 1$ we obtain

$$|c(x_1(t), u(t))| \leq K_6 + K_7 \alpha_1(x_0)^{L_3} \triangleq \alpha'_1(x_0) \tag{111}$$

for all $t \in [0, 1]$.

To bound the terminal cost, note that

$$\begin{aligned} |h(x(T))| &\leq K_6 + K_7 \|x(T)\|_2^{L_3} && (112) \\ &= K_6 + K_7 \|g(x_T(T), y_T)\|_2^{L_3} && (113) \\ &\leq K_6 + K_7 \Big\{ K_2 + K_4 \|y_T\|_2^{L_2} \\ &\quad + \|x_T(T)\|_2^{L_1} \left(K_3 + K_5 \|y_T\|_2^{L_2} \right) \Big\}^{L_3} && (114) \end{aligned}$$

by Assumptions (2d) and 3. Since $L_3 < \infty$, (114) yields a polynomial of $\|x_T(T)\|_2$ and $\|y_T\|_2$ of finite terms, for each of which we can use Proposition 9 and apply the multinomial expansion formula to show

$$|h(x(T))| \leq \sum_{(j_1,\dots,j_T) \in \mathcal{K}'_{T+1}} \alpha_{T+1}'^{(j_1,\dots,j_T)}(x_0) \prod_{m=1}^{T} \|y_m\|_2^{j_m} \tag{115}$$

for some finite set $\mathcal{K}'_{T+1}$ of sequence of non-negative integers and finite positive constants $\alpha_{T+1}'^{(j_1,\dots,j_T)}(x_0)$.

A.2 *Perturbed Trajectory under Specific Observations*

Next, we will perturb the nominal state x of the system while assuming the same initial condition x_0 and the specific observations $(y_1, \dots, y_T)$ as in Section A.1. The perturbed control is an open-loop perturbation as defined below.

Definition 1. Perturbed Control. *Let $u \in U$ be a control. For $\tau \in (0, 1)$ and $v \in B(0, \rho_{\max})$, define the perturbed control u^ϵ by*

$$u^\epsilon(t) \triangleq \begin{cases} v & \text{if } t \in (\tau - \epsilon, \tau] \\ u(t) & \text{otherwise,} \end{cases} \tag{116}$$

where $\epsilon \in [0, \tau]$. By definition if $\epsilon = 0$ then u^ϵ is the same as u. We assume that the nominal control $u(t)$ is left continuous in t at $t = \tau$.

Remark 4. *It is obvious that $u^\epsilon(t)$ is piecewise continuous on $[0, T]$. Therefore, for $v \in B(0, \rho_{\max})$ we have $u^\epsilon \in U$, i.e. u^ϵ is an admissible control. Thus, Proposition 3 assures that there exists a unique solution x^ϵ for the trajectory of the system under the control perturbation. In the remainder of the analysis, we assume that (τ, v) is given and fixed.*

Lemma 11. *Let $\epsilon, \epsilon' \in [0, \tau]$, and let $u^\epsilon, u^{\epsilon'} \in U$ be two perturbed controls of the form (116). Let $x_1^\epsilon, x_1^{\epsilon'}$ be the solutions of x_1 for $t \in [0, 1]$ by applying u^ϵ and $u^{\epsilon'}$ respectively to the initial condition x_0. Then, there exists an $L' < \infty$, independent of $\epsilon, \epsilon', x_1^\epsilon$ and $x_1^{\epsilon'}$, such that*

$$\forall \epsilon, \epsilon' \in [0, \tau]\ \forall t \in [0, 1]\ \ \|x_1^\epsilon(t) - x_1^{\epsilon'}(t)\|_2 \leq L' |\epsilon - \epsilon'|. \tag{117}$$

Proof. By Proposition 6, we find that

$$\forall t \in [0, 1]\ \ \|x_1^\epsilon(t) - x_1^{\epsilon'}(t)\|_2 \leq L \int_0^1 \|u^\epsilon(t) - u^{\epsilon'}(t)\|_2 dt \tag{118}$$

for some $L < \infty$. Let us derive an upper-bound on the integral on the right hand side. If $\epsilon \geq \epsilon'$,

$$\int_0^1 \|u^\epsilon(t) - u^{\epsilon'}(t)\|_2 dt = \int_{\tau-\epsilon}^{\tau-\epsilon'} \|v - u(t)\|_2 dt, \tag{119}$$

where $u(t)$ is the nominal control that both u^ϵ and $u^{\epsilon'}$ are based on. Since $u(t) \in B(0, \rho_{\max})$, we obtain

$$\int_0^1 \|u^\epsilon(t) - u^{\epsilon'}(t)\|_2 dt \leq \sup_{u \in B(0,\rho_{\max})} \|v - u\|_2 (\epsilon - \epsilon'). \tag{120}$$

Similarly, if $\epsilon < \epsilon'$ we have

$$\int_0^1 \|u^\epsilon(t) - u^{\epsilon'}(t)\|_2 dt \leq \sup_{u \in B(0,\rho_{\max})} \|v - u\|_2 (\epsilon' - \epsilon). \tag{121}$$

Put these two cases together and substitute into (118) to get

$$\forall t \in [0,1] \;\; \|x_1^\epsilon(t) - x_1^{\epsilon'}(t)\|_2 \leq L'|\epsilon - \epsilon'|, \tag{122}$$

where $L' \triangleq L \sup_{u \in B(0,\rho_{\max})} \|v - u\|_2 \leq 2L\rho_{\max} < \infty$.

Lemma 12. *Let u^ϵ and x_1^ϵ be as in Lemma 11. Then,*

$$\begin{aligned}\lim_{\epsilon \to 0^+} \frac{1}{\epsilon} \int_{\tau-\epsilon}^{\tau} \left\{ f(x_1^\epsilon(t), u^\epsilon(t)) - f(x_1(t), u(t)) \right\} dt \\ = f(x_1(\tau), v) - f(x_1(\tau), u(\tau)),\end{aligned} \tag{123}$$

where x_1 denotes the solution under the nominal control $u \in U$.

Proof. We will show that the difference norm

$$\begin{aligned}\Bigg\| \frac{1}{\epsilon} \int_{\tau-\epsilon}^{\tau} \left\{ f(x_1^\epsilon(t), u^\epsilon(t)) - f(x_1(t), u(t)) \right\} dt \\ - f(x_1(\tau), v) + f(x_1(\tau), u(\tau)) \Bigg\|_2\end{aligned} \tag{124}$$

converges to 0 as $\epsilon \to 0^+$. Indeed, (124) becomes

$$\begin{aligned}&\Bigg\| \frac{1}{\epsilon} \int_{\tau-\epsilon}^{\tau} \{ f(x_1^\epsilon(t), u^\epsilon(t)) - f(x_1(t), u(t)) \\ &\qquad - f(x_1(\tau), v) + f(x_1(\tau), u(\tau)) \} dt \Bigg\|_2 &(125)\\ &\leq \frac{1}{\epsilon} \int_{\tau-\epsilon}^{\tau} \| f(x_1^\epsilon(t), u^\epsilon(t)) - f(x_1(t), u(t)) \\ &\qquad - f(x_1(\tau), v) + f(x_1(\tau), u(\tau)) \|_2 dt &(126)\\ &\leq \frac{1}{\epsilon} \int_{\tau-\epsilon}^{\tau} \{ \| f(x_1^\epsilon(t), u^\epsilon(t)) - f(x_1(\tau), v) \|_2 \\ &\qquad + \| f(x_1(t), u(t)) - f(x_1(\tau), u(\tau)) \|_2 \} dt. &(127)\end{aligned}$$

We used the triangle inequality in (127). Now, making use of the fact that $\forall t \in (\tau - \epsilon, \tau] \;\; u^\epsilon(t) = v$, Assumption (2c) yields

$$\begin{aligned}&\| f(x_1^\epsilon(t), u^\epsilon(t)) - f(x_1(\tau), v) \|_2 \\ &\quad \leq K_1 \| x_1^\epsilon(t) - x_1(\tau) \|_2 &(128)\\ &\quad = K_1 \| x_1^\epsilon(t) - x_1(t) + x_1(t) - x_1(\tau) \|_2 &(129)\\ &\quad \leq K_1 \| x_1^\epsilon(t) - x_1(t) \|_2 + K_1 \| x_1(t) - x_1(\tau) \|_2 &(130)\\ &\quad \leq K_1 L' \epsilon + K_1 \| x_1(t) - x_1(\tau) \|_2 &(131)\end{aligned}$$

for any $t \in (\tau - \epsilon, \tau]$, where we applied Lemma 11 in (131) with $\epsilon' = 0$. Similarly,

$$\begin{aligned}&\| f(x_1(t), u(t)) - f(x_1(\tau), u(\tau)) \|_2 \\ &\quad \leq K_1 \| x_1(t) - x_1(\tau) \|_2 + K_1 \| u(t) - u(\tau) \|_2.\end{aligned} \tag{132}$$

Therefore, (124) is upper-bounded by

$$\begin{aligned}&\frac{1}{\epsilon} \int_{\tau-\epsilon}^{\tau} \{ K_1 L' \epsilon + 2K_1 \| x_1(t) - x_1(\tau) \|_2 \\ &\qquad\qquad + K_1 \| u(t) - u(\tau) \|_2 \} dt &(133)\\ &\leq K_1 L' \epsilon + 2K_1 \sup_{t \in [\tau-\epsilon, \tau]} \| x_1(t) - x_1(\tau) \|_2 \\ &\qquad + K_1 \sup_{t \in [\tau-\epsilon, \tau]} \| u(t) - u(\tau) \|_2, &(134)\end{aligned}$$

which converges to 0 as $\epsilon \to 0^+$, since

$$0 \leq \sup_{t \in [\tau-\epsilon, \tau]} \| x_1(t) - x_1(\tau) \|_2 \longrightarrow 0 \tag{135}$$

and

$$0 \leq \sup_{t \in [\tau-\epsilon, \tau]} \| u(t) - u(\tau) \|_2 \longrightarrow 0 \tag{136}$$

as $\epsilon \to 0^+$.

Lemma 13. *Let x_1^ϵ and x_1 be as in Lemma 12. Let $\Psi_1(t)$ be the right derivative of $x_1^\epsilon(t)$ with respect to ϵ evaluated at $\epsilon = 0$. That is,*

$$\Psi_1(\tau) = \left. \frac{\partial_+}{\partial \epsilon} x_1^\epsilon(\tau) \right|_{\epsilon=0} \triangleq \lim_{\epsilon \to 0^+} \frac{x_1^\epsilon(\tau) - x_1(\tau)}{\epsilon}. \tag{137}$$

Then, we have

$$\Psi_1(\tau) = f(x_1(\tau), v) - f(x_1(\tau), u(\tau)). \tag{138}$$

Proof. Let us express both $x_1^\epsilon(\tau)$ and $x_1(\tau)$ in the integral form:

$$x_1^\epsilon(\tau) = x_0 + \int_0^\tau f(x_1^\epsilon(t), u^\epsilon(t)) dt \tag{139}$$

$$x_1(\tau) = x_0 + \int_0^\tau f(x_1(t), u(t)) dt. \tag{140}$$

Note that $u^\epsilon(t) = u(t)$ and $x_1^\epsilon(t) = x_1(t)$ for $t \in [0, \tau - \epsilon]$, since no perturbation is applied to the system until $t > \tau - \epsilon$. Therefore,

$$\begin{aligned}&x_1^\epsilon(\tau) - x_1(\tau) \\ &\quad = \int_{\tau-\epsilon}^{\tau} \left\{ f(x_1^\epsilon(t), u^\epsilon(t)) - f(x_1(t), u(t)) \right\} dt.\end{aligned} \tag{141}$$

Making use of Lemma 12, we conclude that

$$\lim_{\epsilon \to 0^+} \frac{x_1^\epsilon(\tau) - x_1(\tau)}{\epsilon} = f(x_1(\tau), v) - f(x_1(\tau), u(\tau)). \tag{142}$$

Lemma 14. *Suppose that Assumption (2c) is satisfied. Then, we have the following bound on the (matrix) norm of the Jacobian of function f:*

$$\left\| \frac{\partial}{\partial x} f(x', u') \right\|_2 \leq K_1, \tag{143}$$

for any $x' \in \mathbb{R}^{n_x}$ and $u' \in B(0, \rho_{\max})$.

Proof. Let u'' be equal to u' in Assumption (2c). Then, we have

$$\|f(x',u')-f(x'',u')\|_2 \le K_1\|x'-x''\|_2 \tag{144}$$

for any $x',x''\in\mathbb{R}^{n_x}$ and any $u'\in B(0,\rho_{\max})$. Now, let us define some non-zero scalar t and some unit vector $v\in\mathbb{R}^{n_x}$ that satisfies $\|v\|_2=1$, and let

$$x'' \triangleq x'+tv. \tag{145}$$

Substituting (145) into (144), we get

$$\frac{\|f(x'+tv,u')-f(x',u')\|_2}{|t|}\le K_1. \tag{146}$$

Note that this holds for any t and v as defined above. From multivariate calculus, on the other hand, the directional derivative of f with respect to v is given by

$$\lim_{t\to 0}\frac{f(x'+tv,u')-f(x',u')}{t}=\left(\frac{\partial}{\partial x}f(x',u')\right)v. \tag{147}$$

Therefore, we have

$$\left\|\left(\frac{\partial}{\partial x}f(x',u')\right)v\right\|_2\le K_1, \tag{148}$$

which implies

$$\left\|\frac{\partial}{\partial x}f(x',u')\right\|_2 \triangleq \sup_{\|v\|_2=1}\left\|\left(\frac{\partial}{\partial x}f(x',u')\right)v\right\|_2\le K_1. \tag{149}$$

Lemma 15. *Let x_1^ϵ and x_1 be as in Lemma 12. Then, $\Psi_1(t)=\frac{\partial_+}{\partial\epsilon}x_1^\epsilon(t)\big|_{\epsilon=0}$ uniquely exists for $t\in[\tau,1]$ and follows the ODE:*

$$\dot\Psi_1(t)=\frac{\partial}{\partial x_1}f(x_1(t),u(t))\Psi_1(t), \tag{150}$$

with the initial condition $\Psi_1(\tau)$ given by Lemma 13.

Proof. Taking some $a\in(\tau,1]$, let us express $x_1^\epsilon(a)$ and $x_1(a)$ in the integral form:

$$x_1^\epsilon(a)=x_1^\epsilon(\tau)+\int_\tau^a f(x_1^\epsilon(t),u^\epsilon(t))dt \tag{151}$$

$$x_1(a)=x_1(\tau)+\int_\tau^a f(x_1(t),u(t))dt. \tag{152}$$

Thus, we have

$$\Psi_1(a)=\Psi_1(\tau) + \lim_{\epsilon\to0+}\int_\tau^a\frac{1}{\epsilon}\{f(x_1^\epsilon(t),u(t))-f(x_1(t),u(t))\}dt, \tag{153}$$

where we used $\forall t\in[\tau,a]\ \ u^\epsilon(t)=u(t)$. We will take a measure-theoretic approach to prove that the order of the limit and the integration can be switched in (153). First, think of the integral as a Lebesgue integral on the measure space $([\tau,a],\mathcal{B}([\tau,a]),\lambda)$, where $\mathcal{B}([\tau,a])$ is the Borel σ-algebra on $[\tau,a]$ and λ is the Lebesgue measure. Furthermore, consider the integrand as a function from $[\tau,a]$ into the Banach space $(\mathbb{R}^{n_x},\|\cdot\|_2)$, i.e. the Euclidean space endowed with the ℓ^2 norm. By the piecewise continuity of u^ϵ,u and the continuity of x_1^ϵ, x_1, and f, the integrand is a piecewise continuous function with respect to t, which is λ-measurable. In fact it is also Bochner-integrable, since for $t\in[\tau,a]$ we have the constant bound:

$$\begin{aligned}\frac{1}{\epsilon}&\|f(x_1^\epsilon(t),u(t))-f(x_1(t),u(t))\|_2\\ &\le\frac{1}{\epsilon}K_1\|x_1^\epsilon(t)-x_1(t)\|_2 &(154)\\ &\le K_1L' &(155)\end{aligned}$$

by Assumption (2c) and Lemma 11. Furthermore, the chain rule gives

$$\lim_{\epsilon\to0+}\frac{1}{\epsilon}(f(x_1^\epsilon(t),u(t))-f(x_1(t),u(t))) =\frac{\partial}{\partial x_1}f(x_1(t),u(t))\Psi_1(t), \tag{156}$$

assuming that $\Psi_1(t)$ exists for $t\in[\tau,a]$. Therefore, by the Bochner-integral version of the dominated convergence theorem (Theorem 3 in Diestel and Uhl (1977), Chapter II), we obtain

$$\Psi_1(a)=\Psi_1(\tau)+\int_\tau^a\frac{\partial}{\partial x_1}f(x_1(t),u(t))\Psi_1(t)dt. \tag{157}$$

This is equivalent to the ordinary differential equation:

$$\dot\Psi_1(t)=\frac{\partial}{\partial x_1}f(x_1(t),u(t))\Psi_1(t). \tag{158}$$

It remains to show that the solution $\Psi_1(t)$ that satisfies (158) does exist and is unique. First, let Ψ_1' and Ψ_1'' be two systems that follow (158) and share the same initial condition $\Psi_1(\tau)$. Then, by Lemma 14 we have

$$\begin{aligned}&\|\dot\Psi_1'(t)-\dot\Psi_1''(t)\|_2\\ &=\left\|\frac{\partial}{\partial x_1}f(x_1(t),u(t))(\Psi_1'(t)-\Psi_1''(t))\right\|_2 &(159)\\ &\le\left\|\frac{\partial}{\partial x_1}f(x_1(t),u(t))\right\|_2\cdot\|\Psi_1'(t)-\Psi_1''(t)\|_2 &(160)\\ &\le K_1\left\|\Psi_1'(t)-\Psi_1''(t)\right\|_2. &(161)\end{aligned}$$

Existence follows from this inequality in conjunction with the Picard Lemma (Lemma 5.6.3 in Elijah (1997)). To show the uniqueness, apply the Bellman-Gronwall Lemma (Lemma 5.6.4 in Elijah (1997)) to the following integral inequality:

$$\forall a\in[\tau,1]$$
$$\|\Psi_1'(a)-\Psi_1''(a)\|_2\le K_1\int_\tau^a\|\Psi_1'(t)-\Psi_1''(t)\|_2dt. \tag{162}$$

Proposition 16. Variational Equation. *Let $u\in U$ and x_1 be the nominal control and the resulting state trajectory. Let x_1^ϵ be the perturbed state induced by the perturbed control u^ϵ of the form (116). Propagating $x_1^\epsilon(t)$ through the hybrid dynamics, we get a series of modes $x_2^\epsilon,\ldots,x_T^\epsilon$ that constitutes the entire trajectory $x^\epsilon(t)$ for $t\in[\tau,T]$. Define* *the state variation $\Psi(t)$ for $t \in [\tau, T]$ by*

$$\Psi(t) = \frac{\partial_+}{\partial \epsilon} x^\epsilon(t)\bigg|_{\epsilon=0} \triangleq \lim_{\epsilon \to 0^+} \frac{x^\epsilon(t) - x(t)}{\epsilon}. \tag{163}$$

Then, $\Psi(t)$ exists for $t \in [\tau, T]$ and follows the hybrid system with time-driven switching:

$$\Psi(t) = \begin{cases} \Psi_1(t) & \forall t \in [\tau, 1) \\ \Psi_i(t) & \forall t \in [i-1, i)\ \forall i \in \{2, \dots, T\} \end{cases} \tag{164}$$

with $\Psi(T) = \frac{\partial}{\partial x_T} g(x_T(T), y_T)\Psi_T(T)$, where Ψ_1 is defined on $[\tau, 1]$ as in Lemma 15, and Ψ_i for $i \geq 2$ is defined on $[i-1, i]$ with

$$\Psi_i(i-1) = \frac{\partial}{\partial x_{i-1}} g(x_{i-1}(i-1), y_{i-1})\Psi_{i-1}(i-1) \tag{165}$$

$$\dot{\Psi}_i(t) = \frac{\partial}{\partial x_i} f(x_i(t), u(t))\Psi_i(t) \ \ \forall t \in [i-1, i]. \tag{166}$$

Proof. The case for $t \in [\tau, 1)$ follows from Lemma 15, since in this case we have

$$\Psi(t) \triangleq \lim_{\epsilon \to 0^+} \frac{x^\epsilon(t) - x(t)}{\epsilon} \tag{167}$$

$$= \lim_{\epsilon \to 0^+} \frac{x_1^\epsilon(t) - x_1(t)}{\epsilon} \tag{168}$$

$$= \Psi_1(t). \tag{169}$$

At $t = 1$, we obtain

$$\Psi(1) \triangleq \lim_{\epsilon \to 0^+} \frac{x^\epsilon(1) - x(1)}{\epsilon} \tag{170}$$

$$= \lim_{\epsilon \to 0^+} \frac{x_2^\epsilon(1) - x_2(1)}{\epsilon} \tag{171}$$

$$= \lim_{\epsilon \to 0^+} \frac{g(x_1^\epsilon(1), y_1) - g(x_1(1), y_1)}{\epsilon} \tag{172}$$

$$= \frac{\partial}{\partial x} g(x_1(1), y_1)\Psi_1(1) \tag{173}$$

by (47) and the chain rule. Let us define Ψ_2 by $\Psi_2(t) \triangleq \lim_{\epsilon \to 0^+}(x_2^\epsilon(t) - x_2(t))/\epsilon$. Similarly to the proof of Lemma 15, one can show that Ψ_2 follows the integral equation:

$$\Psi_2(a) = \Psi_2(1) + \int_1^a \frac{\partial}{\partial x_2} f(x_2(t), u(t))\Psi_2(t)dt \tag{174}$$

with $\Psi_2(1) = \Psi(1)$, and that $\Psi_2(t)$ that satisfies (165) and (166) uniquely exists. This proves the case for $t \in [1, 2)$. Proceeding by induction completes the proof.

So far we have focused entirely on the right derivative $\frac{\partial_+}{\partial \epsilon} x^\epsilon(t)$ evaluated at $\epsilon = 0$. The next proposition shows that x_1^ϵ is in fact right differentiable with respect to ϵ at all $\epsilon \in [0, \tau)$.

Proposition 17. Right Differentiability of State Perturbation. *Let $x^\epsilon(t)$ be the perturbed state trajectory under the perturbed control u^ϵ defined by (116). Let $\Psi^\epsilon(t)$ denote the right derivative $\frac{\partial_+}{\partial \epsilon} x^\epsilon(t)$ evaluated at a particular $\epsilon \in [0, \tau)$. (Note that when $\epsilon = 0$ we have $\Psi^\epsilon = \Psi$.) Then, $\Psi^\epsilon(t)$ exists for $t \in [\tau - \epsilon, T]$ and follows the hybrid system with time-driven switching:*

$$\Psi^\epsilon(t) = \begin{cases} \Psi_1^\epsilon(t) & \forall t \in [\tau - \epsilon, 1) \\ \Psi_i^\epsilon(t) & \forall t \in [i-1, i)\ \forall i \in \{2, \dots, T\} \end{cases} \tag{175}$$

with $\Psi^\epsilon(T) = \frac{\partial}{\partial x_T^\epsilon} g(x_T^\epsilon(T), y_T)\Psi_T^\epsilon(T)$. Ψ_1^ϵ is defined on $[\tau - \epsilon, 1]$, where

$$\Psi_1^\epsilon(\tau - \epsilon) = f(x_1^\epsilon(\tau - \epsilon), v) - f(x_1^\epsilon(\tau - \epsilon), u^\epsilon(\tau - \epsilon)) \tag{176}$$

and

$$\dot{\Psi}_1^\epsilon(t) = \frac{\partial}{\partial x_1^\epsilon} f(x_1^\epsilon(t), u^\epsilon(t))\Psi_1^\epsilon(t) \ \ \forall t \in [\tau - \epsilon, 1]. \tag{177}$$

Ψ_i^ϵ for $i \geq 2$ is defined on $[i-1, i]$ with

$$\Psi_i^\epsilon(i-1) = \frac{\partial}{\partial x_{i-1}^\epsilon} g(x_{i-1}^\epsilon(i-1), y_{i-1})\Psi_{i-1}^\epsilon(i-1) \tag{178}$$

$$\dot{\Psi}_i^\epsilon(t) = \frac{\partial}{\partial x_i^\epsilon} f(x_i^\epsilon(t), u^\epsilon(t))\Psi_i^\epsilon(t) \ \ \forall t \in [i-1, i]. \tag{179}$$

Proof. The proposition can be proven by considering the perturbed control u^ϵ as the new nominal control, and defining a further perturbation based on this nominal control. Formally, define $\tilde{u}^{\epsilon'}$ by

$$\tilde{u}^{\epsilon'}(t) \triangleq \begin{cases} v & \text{if } \ t \in (\tau - \epsilon - \epsilon', \tau - \epsilon] \\ u^\epsilon(t) & \text{otherwise} \end{cases} \tag{180}$$

with $\epsilon' \in [0, \tau - \epsilon)$, where (τ, v) is the same pair of values as for $u^\epsilon(t)$. Since $u^\epsilon(t)$ is left continuous in t at $t = \tau - \epsilon$, $\tilde{u}^{\epsilon'}$ is a valid perturbed control of the form (116) with u^ϵ being the nominal control. Here ϵ is considered as fixed and the parameters defining this new perturbation are v, $\tau - \epsilon$, and ϵ'. This perturbation yields the new perturbed state trajectory $\tilde{x}^{\epsilon'}$ based on the nominal trajectory x^ϵ. Note that when $\epsilon' = 0$ we have $\tilde{u}^{\epsilon'} = u^\epsilon$ and $\tilde{x}^{\epsilon'} = x^\epsilon$. We can define the new state variation:

$$\widetilde{\Psi}(t) = \frac{\partial_+}{\partial \epsilon'} \tilde{x}^{\epsilon'}(t)\bigg|_{\epsilon'=0} \triangleq \lim_{\epsilon' \to 0^+} \frac{\tilde{x}^{\epsilon'}(t) - x^\epsilon(t)}{\epsilon'}. \tag{181}$$

Applying Proposition 16 to this new setting, we find that $\widetilde{\Psi}(t)$ exists for $t \in [\tau - \epsilon, T]$ and follows the hybrid system with time-driven switching:

$$\widetilde{\Psi}(t) = \begin{cases} \widetilde{\Psi}_1(t) & \forall t \in [\tau - \epsilon, 1) \\ \widetilde{\Psi}_i(t) & \forall t \in [i-1, i)\ \forall i \in \{2, \dots, T\} \end{cases} \tag{182}$$

with $\widetilde{\Psi}(T) = \frac{\partial}{\partial x_T^\epsilon} g(x_T^\epsilon(T), y_T)\widetilde{\Psi}_T(T)$. $\widetilde{\Psi}_1$ is defined on $[\tau - \epsilon, 1]$, where

$$\widetilde{\Psi}_1(\tau - \epsilon) = f(x_1^\epsilon(\tau - \epsilon), v) - f(x_1^\epsilon(\tau - \epsilon), u^\epsilon(\tau - \epsilon)) \tag{183}$$

and

$$\dot{\widetilde{\Psi}}_1(t) = \frac{\partial}{\partial x_1^\epsilon} f(x_1^\epsilon(t), u^\epsilon(t))\widetilde{\Psi}_1(t) \ \ \forall t \in [\tau - \epsilon, 1]. \tag{184}$$

$\widetilde{\Psi}_i$ for $i \geq 2$ is defined on $[i-1, i]$ with

$$\widetilde{\Psi}_i(i-1) = \frac{\partial}{\partial x^\epsilon_{i-1}} g(x^\epsilon_{i-1}(i-1), y_{i-1})\widetilde{\Psi}_{i-1}(i-1) \tag{185}$$

$$\dot{\widetilde{\Psi}}_i(t) = \frac{\partial}{\partial x^\epsilon_i} f(x^\epsilon_i(t), u^\epsilon(t))\widetilde{\Psi}^\epsilon_i(t) \ \ \forall t \in [i-1, i]. \tag{186}$$

On the other hand, notice that the new perturbed control $\tilde{u}^{\epsilon'}$ is actually equivalent to the perturbed control $u^{\epsilon+\epsilon'}$ that is based on the original nominal control u. Namely,

$$\tilde{u}^{\epsilon'}(t) = u^{\epsilon+\epsilon'}(t) \triangleq \begin{cases} v & \text{if } \ t \in (\tau-(\epsilon+\epsilon'), \tau] \\ u(t) & \text{otherwise.} \end{cases} \tag{187}$$

Consequently, the new perturbed state $\tilde{x}^{\epsilon'}$ is equal to $x^{\epsilon+\epsilon'}$, and thus

$$\begin{aligned} \widetilde{\Psi}(t) &= \lim_{\epsilon' \to 0+} \frac{\tilde{x}^{\epsilon'}(t) - x^\epsilon(t)}{\epsilon'} && (188) \\ &= \lim_{\epsilon' \to 0+} \frac{x^{\epsilon+\epsilon'}(t) - x^\epsilon(t)}{\epsilon'} && (189) \\ &= \Psi^\epsilon(t). && (190) \end{aligned}$$

This completes the proof.

Lemma 18. *Let $\Psi^\epsilon_1, \dots, \Psi^\epsilon_T$ be as given by Proposition 17. Then, for Ψ^ϵ_1 the following holds.*

$$\|\Psi^\epsilon_1(t)\|_2 \leq 2K_1\rho_{\max}e^{K_1} \qquad \forall t \in [\tau-\epsilon, 1] \tag{191}$$

Similarly, for all $i \in \{2, \dots, T\}$ we have

$$\|\Psi^\epsilon_i(t)\|_2 \leq \|\Psi^\epsilon_i(i-1)\|_2 e^{K_1} \qquad \forall t \in [i-1, i]. \tag{192}$$

Proof. We begin with the integral equation:

$$\begin{aligned} \Psi^\epsilon_1(a) &= \Psi^\epsilon_1(\tau-\epsilon) \\ &\quad + \int_{\tau-\epsilon}^{a} \frac{\partial}{\partial x^\epsilon_1} f(x^\epsilon_1(t), u^\epsilon(t))\Psi^\epsilon_1(t)dt. \end{aligned} \tag{193}$$

Therefore,

$$\begin{aligned} &\|\Psi^\epsilon_1(a)\|_2 \leq \|\Psi^\epsilon_1(\tau-\epsilon)\|_2 \\ &\quad + \int_{\tau-\epsilon}^{a} \left\| \frac{\partial}{\partial x^\epsilon_1} f(x^\epsilon_1(t), u^\epsilon(t)) \right\|_2 \cdot \|\Psi^\epsilon_1(t)\|_2 dt. \end{aligned} \tag{194}$$

Using

$$\Psi^\epsilon_1(\tau-\epsilon) = f(x^\epsilon_1(\tau-\epsilon), v) - f(x^\epsilon_1(\tau-\epsilon), u^\epsilon(\tau-\epsilon)), \tag{195}$$

Assumption (2c), and Lemma 14, we get

$$\|\Psi^\epsilon_1(a)\|_2 \leq K_1\|v - u^\epsilon(\tau-\epsilon)\|_2 + K_1\int_{\tau-\epsilon}^{a} \|\Psi^\epsilon_1(t)\|_2 dt \tag{196}$$

$$\begin{aligned} &\leq K_1 \sup_{u \in B(0, \rho_{\max})} \|v - u\|_2 \\ &\quad + K_1\int_{\tau-\epsilon}^{a} \|\Psi^\epsilon_1(t)\|_2 dt \end{aligned} \tag{197}$$

$$\leq 2K_1\rho_{\max} + K_1\int_{\tau-\epsilon}^{a} \|\Psi^\epsilon_1(t)\|_2 dt \tag{198}$$

for all $a \in [\tau-\epsilon, 1]$. Thus, by the Bellman-Gronwall Lemma (Lemma 5.6.4 in Elijah (1997)) it follows that

$$\|\Psi^\epsilon_1(t)\|_2 \leq 2K_1\rho_{\max}e^{K_1} \qquad \forall t \in [\tau-\epsilon, 1]. \tag{199}$$

For general $i \geq 2$, apply the Bellman-Gronwall Lemma to the similar integral inequality:

$$\begin{aligned} &\forall a \in [i-1, i] \\ &\|\Psi^\epsilon_i(a)\|_2 \leq \|\Psi^\epsilon_i(i-1)\|_2 + K_1\int_{i-1}^{a} \|\Psi^\epsilon_i(t)\|_2 dt \end{aligned} \tag{200}$$

to get the result.

Proposition 19. Bounded State Variation. *Given u^ϵ and $(y_1, \dots, y_T)$, Ψ^ϵ defined in Proposition 17 has the following bound:*

$$\begin{aligned} &\forall \epsilon \in [0, \tau) \ \forall i \in \{2, \dots, T\} \ \forall t \in [i-1, i] \\ &\|\Psi^\epsilon_i(t)\|_2 \leq \sum_{(j_1, \dots, j_{i-1}) \in \mathcal{L}_i} \beta^{(j_1, \dots, j_{i-1})}_i(x_0) \prod_{m=1}^{i-1} \|y_m\|_2^{j_m}, \end{aligned} \tag{201}$$

where $\mathcal{L}_i$ is a finite set of sequences of non-negative integers of length $i-1$, and $\beta^{(j_1, \dots, j_{i-1})}_i(x_0)$ is a finite positive constant that depends on x_0 and $(j_1, \dots, j_{i-1})$ but not on ϵ, u^ϵ, or $(y_1, \dots, y_T)$.

Proof. The proof of this proposition is similar to that of Proposition 9. Take any $\epsilon \in [0, \tau)$. For $i = 2$, we have $\forall t \in [1, 2]$

$$\begin{aligned} \|\Psi^\epsilon_2(t)\|_2 &\leq \|\Psi^\epsilon_2(1)\|_2 e^{K_1} && (202) \\ &\leq \left\| \frac{\partial}{\partial x^\epsilon_1} g(x^\epsilon_1(1), y_1)\Psi^\epsilon_1(1) \right\|_2 e^{K_1} && (203) \\ &\leq \left\| \frac{\partial}{\partial x^\epsilon_1} g(x^\epsilon_1(1), y_1) \right\|_2 \cdot \|\Psi^\epsilon_1(1)\|_2 e^{K_1} && (204) \\ &\leq 2K_1\rho_{\max}e^{2K_1}\Big\{ K_2 + K_4\|y_1\|_2^{L_2} + \\ &\qquad \left(K_3 + K_5\|y_1\|_2^{L_2}\right)\|x^\epsilon_1(1)\|_2^{L_1} \Big\} && (205) \end{aligned}$$

by Assumption (2d), Proposition 17, and Lemma 18. Using Proposition 9, we can bound $x^\epsilon_1(1)$ by

$$\|x^\epsilon_1(1)\|_2 \leq \sum_{(j_1) \in \mathcal{K}_2} \alpha^{(j_1)}_2(x_0)\|y_1\|_2^{j_1}. \tag{206}$$

Substituting (206) into (205) and using the multinomial theorem, one can verify that

$$\forall t \in [1, 2] \ \ \|\Psi^\epsilon_2(t)\|_2 \leq \sum_{(j_1) \in \mathcal{L}_1} \beta^{(j_1)}_1(x_0)\|y_1\|_2^{j_1} \tag{207}$$

for some finite set $\mathcal{L}_1$ and finite $\beta^{(j_1)}_1(x_0)$.

Next, suppose that the claim holds for some $i \leq 2$. That is,

$$\begin{aligned} &\forall t \in [i-1, i] \\ &\|\Psi^\epsilon_i(t)\|_2 \leq \sum_{(j_1, \dots, j_{i-1}) \in \mathcal{L}_i} \beta^{(j_1, \dots, j_{i-1})}_i(x_0) \prod_{m=1}^{i-1} \|y_m\|_2^{j_m}, \end{aligned} \tag{208}$$

where $\mathcal{L}_i$ and $\beta_i^{(j_1,\dots,j_{i-1})}(x_0)$ are as defined in the statement of the proposition. Similar to the case for $i = 2$, we have $\forall t \in [i, i+1]$

$$
\begin{aligned}
&\|\Psi_{i+1}^{\epsilon}(t)\|_2 \\
&\leq e^{K_1} \left(\sum_{(j_1,\dots,j_{i-1}) \in \mathcal{L}_i} \beta_i^{(j_1,\dots,j_{i-1})}(x_0) \prod_{m=1}^{i-1} \|y_m\|_2^{j_m} \right) \\
&\times \left\{ K_2 + K_4 \|y_i\|_2^{L_2} + \left(K_3 + K_5 \|y_i\|_2^{L_2} \right) \|x_i^{\epsilon}(i)\|_2^{L_1} \right\}
\end{aligned} \tag{209}
$$

Proposition 9 gives the following bound:

$$
\|x_i^{\epsilon}(i)\|_2 \leq \sum_{(j_1,\dots,j_{i-1}) \in \mathcal{K}_i} \alpha_i^{(j_1,\dots,j_{i-1})}(x_0) \prod_{m=1}^{i-1} \|y_m\|_2^{j_m}. \tag{210}
$$

Substituting (210) into (209) and using the multinomial theorem, we conclude that

$$
\begin{aligned}
&\forall t \in [i, i+1] \\
&\|\Psi_{i+1}^{\epsilon}(t)\|_2 \leq \sum_{(j_1,\dots,j_i) \in \mathcal{L}_{i+1}} \beta_{i+1}^{(j_1,\dots,j_i)}(x_0) \prod_{m=1}^{i} \|y_m\|_2^{j_m}
\end{aligned} \tag{211}
$$

for some finite set $\mathcal{L}_{i+1}$ and finite $\beta_{i+1}^{(j_1,\dots,j_i)}(x_0)$.

Finally, proceeding by mathematical induction over $i \in \{2, \dots, T\}$ completes the proof.

Lemma 20. *Let u^{ϵ} and x_1^{ϵ} be as in Lemma 11. Then, similarly to Lemma 12 we have*

$$
\begin{aligned}
\lim_{\epsilon \to 0^+} \frac{1}{\epsilon} \int_{\tau-\epsilon}^{\tau} \{c(x_1^{\epsilon}(t), u^{\epsilon}(t)) - c(x_1(t), u(t))\}\, dt \\
= c(x_1(\tau), v) - c(x_1(\tau), u(\tau)).
\end{aligned} \tag{212}
$$

Proof. As $u^{\epsilon}(t) = v \ \ \forall t \in (\tau - \epsilon, \tau]$, we will show that

$$
\begin{aligned}
\lim_{\epsilon \to 0^+} \frac{1}{\epsilon} \int_{\tau-\epsilon}^{\tau} \{c(x_1^{\epsilon}(t), v) - c(x_1(t), u(t))\}\, dt \\
= c(x_1(\tau), v) - c(x_1(\tau), u(\tau)).
\end{aligned} \tag{213}
$$

By Assumption 3 and the continuity of $x_1^{\epsilon}(t)$, $c(x_1^{\epsilon}(t), v)$ is continuous with respect to t on $[\tau - \epsilon, \tau]$. Thus, the mean value theorem yields

$$
\frac{1}{\epsilon} \int_{\tau-\epsilon}^{\tau} c(x_1^{\epsilon}(t), v) dt = c(x_1^{\epsilon}(\tilde{t}), v) \tag{214}
$$

for some $\tilde{t} \in [\tau - \epsilon, \tau]$. From the triangle inequality and Lemma 11 it follows that

$$
\begin{aligned}
\|x_1^{\epsilon}(\tilde{t}) - x_1(\tau)\|_2 &\leq \|x_1^{\epsilon}(\tilde{t}) - x_1(\tilde{t})\|_2 + \|x_1(\tilde{t}) - x_1(\tau)\|_2 \quad (215) \\
&\leq L'\epsilon + \|x_1(\tilde{t}) - x_1(\tau)\|_2. \quad (216)
\end{aligned}
$$

Therefore, $\lim_{\epsilon \to 0^+} x_1^{\epsilon}(\tilde{t}) = x_1(\tau)$ and

$$
\lim_{\epsilon \to 0^+} \frac{1}{\epsilon} \int_{\tau-\epsilon}^{\tau} c(x_1^{\epsilon}(t), v) dt = c(x_1(\tau), v). \tag{217}
$$

On the other hand, $u(t)$ is continuous on $[\tau - \epsilon, \tau]$ for all sufficiently small ϵ, since u is left continuous at τ by Definition 1. Therefore, $c(x_1(t), u(t))$ is continuous with respect to t on $[\tau - \epsilon, \tau]$ for ϵ small. The mean value theorem gives

$$
\frac{1}{\epsilon} \int_{\tau-\epsilon}^{\tau} c(x_1(t), u(t)) dt = c(x_1(\tilde{t}), u(\tilde{t})) \tag{218}
$$

for some $\tilde{t} \in [\tau - \epsilon, \tau]$. Taking the limit $\epsilon \to 0^+$, the right hand side converges to $c(x_1(\tau), u(\tau))$. Combining this result with (217), we conclude that

$$
\begin{aligned}
\lim_{\epsilon \to 0^+} \frac{1}{\epsilon} \int_{\tau-\epsilon}^{\tau} \{c(x_1^{\epsilon}(t), v) - c(x_1(t), u(t))\}\, dt \\
= c(x_1(\tau), v) - c(x_1(\tau), u(\tau)).
\end{aligned} \tag{219}
$$

Lemma 21. *Given $\epsilon \in [0, \tau)$, u^{ϵ} and $(y_1, \dots, y_T)$, the right derivative of the instantaneous cost function with respect to ϵ is given by*

$$
\begin{aligned}
&\forall t \in [i-1, i] \\
&\frac{\partial_+}{\partial \epsilon} c(x_i^{\epsilon}(t), u^{\epsilon}(t)) = \frac{\partial}{\partial x_i^{\epsilon}} c(x_i^{\epsilon}(t), u(t))^{\mathrm{T}} \Psi_i^{\epsilon}(t).
\end{aligned} \tag{220}
$$

for each $i \in \{2, \dots, T\}$.

For $i = 1$ we have

$$
\begin{aligned}
&\forall t \in (\tau, 1] \\
&\frac{\partial_+}{\partial \epsilon} c(x_1^{\epsilon}(t), u^{\epsilon}(t)) = \frac{\partial}{\partial x_1^{\epsilon}} c(x_1^{\epsilon}(t), u(t))^{\mathrm{T}} \Psi_1^{\epsilon}(t),
\end{aligned} \tag{221}
$$

Similarly, the right derivative of the terminal cost function with respect to ϵ is given by

$$
\frac{\partial_+}{\partial \epsilon} h(x^{\epsilon}(T)) = \frac{\partial}{\partial x^{\epsilon}} h(x^{\epsilon}(T))^{\mathrm{T}} \Psi^{\epsilon}(T). \tag{222}
$$

Proof. To prove the claim for the instantaneous cost, note that $u^{\epsilon}(t) = u(t)$ for all $t \in (\tau, T]$ and use the chain rule. The case for the terminal cost also follows from the chain rule.

Proposition 22. Bounded Cost Variations. *Given $\epsilon \in [0, \tau)$, u^{ϵ} and $(y_1, \dots, y_T)$, the right derivative of the instantaneous cost function with respect to ϵ has the following uniform bound:*

$$
\begin{aligned}
&\forall t \in [i-1, i] \\
&\left| \frac{\partial_+}{\partial \epsilon} c(x_i^{\epsilon}(t), u^{\epsilon}(t)) \right| \\
&\quad \leq \sum_{(j_1,\dots,j_{i-1}) \in \mathcal{L}'_i} \beta_i'^{(j_1,\dots,j_{i-1})}(x_0) \prod_{m=1}^{i-1} \|y_m\|_2^{j_m}
\end{aligned} \tag{223}
$$

for each $i \in \{2, \dots, T\}$, where $\mathcal{L}'_i$ is a finite set of sequences of nen-negative integers of length $i - 1$, and $\beta_i'^{(j_1,\dots,j_{i-1})}(x_0)$ is a finite positive constant that depends on x_0 and $(j_1, \dots, j_{i-1})$ but not on ϵ, u^{ϵ}, or $(y_1, \dots, y_T)$.

For $i = 1$ the bound is given by

$$
\forall t \in (\tau, 1] \quad \left| \frac{\partial_+}{\partial \epsilon} c(x_1^{\epsilon}(t), u^{\epsilon}(t)) \right| \leq \beta_1'(x_0) \tag{224}
$$

for some finite positive constant $\beta'_1(x_0)$.

Similarly, the right derivative of the terminal cost function with respect to ϵ has the following bound:

$$\left|\frac{\partial_+}{\partial\epsilon}h(x^\epsilon(T))\right| \le \sum_{(j_1,\dots,j_T)\in\mathcal{L}'_{T+1}} \beta'^{(j_1,\dots,j_T)}_{T+1}(x_0)\prod_{m=1}^{T}\|y_m\|_2^{j_m} \tag{225}$$

for some finite set $\mathcal{L}'_{T+1}$ of sequence of non-negative integers and finite positive constants $\beta'^{(j_1,\dots,j_T)}_{T+1}(x_0)$.

Proof. The proof of this proposition is similar to that of Proposition 10. Take any $\epsilon\in[0,\tau)$. For $i\in\{2,\dots,T\}$, Assumption 3 along with Lemma 21 yields

$$\left|\frac{\partial_+}{\partial\epsilon}c(x_i^\epsilon(t),u^\epsilon(t))\right| = \left|\frac{\partial}{\partial x_i^\epsilon}c(x_i^\epsilon(t),u(t))^{\mathrm{T}}\Psi_i^\epsilon(t)\right| \tag{226}$$

$$\le \left\|\frac{\partial}{\partial x_i^\epsilon}c(x_i^\epsilon(t),u(t))\right\|_2\cdot\|\Psi_i^\epsilon(t)\|_2 \tag{227}$$

$$\le (K_6+K_7\|x_i^\epsilon(t)\|_2^{L_3})\,\|\Psi_i^\epsilon(t)\|_2\,. \tag{228}$$

By Propositions 9 and 19, we have

$$\|x_i^\epsilon(t)\|_2 \le \sum_{(j_1,\dots,j_{i-1})\in\mathcal{K}_i}\alpha_i^{(j_1,\dots,j_{i-1})}(x_0)\prod_{m=1}^{i-1}\|y_m\|_2^{j_m} \tag{229}$$

$$\|\Psi_i^\epsilon(t)\|_2 \le \sum_{(j_1,\dots,j_{i-1})\in\mathcal{L}_i}\beta_i^{(j_1,\dots,j_{i-1})}(x_0)\prod_{m=1}^{i-1}\|y_m\|_2^{j_m} \tag{230}$$

for all $t\in[i-1,i]$. Substituting these into (228) and using the multinomial expansion formula, we conclude that

$$\forall i\in\{2,\dots,T\}\ \forall t\in[i-1,i]\quad \left|\frac{\partial_+}{\partial\epsilon}c(x_i^\epsilon(t),u^\epsilon(t))\right| \le \sum_{(j_1,\dots,j_{i-1})\in\mathcal{L}'_i}\beta_i'^{(j_1,\dots,j_{i-1})}(x_0)\prod_{m=1}^{i-1}\|y_m\|_2^{j_m} \tag{231}$$

for some finite set $\mathcal{L}'_i$ of sequences of non-negative integers and finite positive constants $\beta_i'^{(j_1,\dots,j_{i-1})}(x_0)$. Similarly, for $i=1$ we have $\forall t\in(\tau,1]$

$$\left|\frac{\partial_+}{\partial\epsilon}c(x_1^\epsilon(t),u^\epsilon(t))\right| \le (K_6+K_7\|x_1^\epsilon(t)\|_2^{L_3})\|\Psi_1^\epsilon(t)\|_2 \tag{232}$$

$$\le 2(K_6+K_7\alpha_1(x_0)^{L_3})K_1\rho_{\max}e^{K_1} \tag{233}$$

$$\triangleq \beta'_1(x_0), \tag{234}$$

by Proposition 9 and Lemma 18.

To bound the right derivative of the terminal cost, note that

$$\left|\frac{\partial_+}{\partial\epsilon}h(x^\epsilon(T))\right| = \left|\frac{\partial}{\partial x^\epsilon}h(x^\epsilon(T))^{\mathrm{T}}\Psi^\epsilon(T)\right| \tag{235}$$

$$\le (K_6+K_7\|x^\epsilon(T)\|_2^{L_3})\|\Psi^\epsilon(T)\|_2 \tag{236}$$

$$= \left(K_6+K_7\|g(x_T^\epsilon(T),y_T)\|_2^{L_3}\right)\|\Psi^\epsilon(T)\|_2 \tag{237}$$

$$\le \left(K_6+K_7\|g(x_T^\epsilon(T),y_T)\|_2^{L_3}\right) \times\left\|\frac{\partial}{\partial x_T^\epsilon}g(x_T^\epsilon(T),y_T)\right\|_2\cdot\|\Psi_T^\epsilon(T)\|_2 \tag{238}$$

by Assumption 3, Proposition 17, and Lemma 21. One can apply Assumption (2d) to bound the norms of g and its Jacobian in terms of $\|x_T^\epsilon(T)\|_2$ and $\|y_T\|$. Then, (238) becomes a polynomial of $\|x_T^\epsilon(T)\|_2$ and $\|y_T\|$, multiplied by $\|\Psi_T^\epsilon(T)\|_2$. Finally, using (229) and (230) with $i=T$ to replace $\|x_T^\epsilon(T)\|_2$ and $\|\Psi_T^\epsilon(T)\|_2$, one can verify that

$$\left|\frac{\partial_+}{\partial\epsilon}h(x^\epsilon(T))\right| \le \sum_{(j_1,\dots,j_T)\in\mathcal{L}'_{T+1}}\beta'^{(j_1,\dots,j_T)}_{T+1}(x_0)\prod_{m=1}^{T}\|y_m\|_2^{j_m} \tag{239}$$

for some finite set $\mathcal{L}'_{T+1}$ of sequence of non-negative integers and finite positive constants $\beta'^{(j_1,\dots,j_T)}_{T+1}(x_0)$.

Lemma 23. *Let x^ϵ be the perturbed state induced by the perturbed control u^ϵ, and let $(y_1,\dots,y_T)$ be the given observations. Then, the function $\epsilon\mapsto c(x^\epsilon(t),u^\epsilon(t))$ is continuous with respect to $\epsilon\in[0,\tau]$ for all $t\in(\tau,T]$.*

Proof. Note that for $t\in(\tau,T]$ we have $u^\epsilon(t)=u(t)$. Thus, $c(x^\epsilon(t),u^\epsilon(t))=c(x^\epsilon(t),u(t))$. The continuity of $x_1^\epsilon(t)$ with respect to $\epsilon\in[0,\tau]$ follows from Lemma 11. In particular, $x_1^\epsilon(1)$ is continuous with respect to ϵ. Next, suppose that $\epsilon\mapsto x_i^\epsilon(i)$ is continuous for some $i\in\{1,\dots,T\}$. Then, by Assumption (2b) and Corollary 7 it follows that $\epsilon\mapsto x_{i+1}^\epsilon(t)$ is continuous for all $t\in[i,i+1]$. Proceeding by mathematical induction, we conclude that $x^\epsilon(t)$ is continuous with respect to $\epsilon\in[0,\tau]$ for all $t\in(\tau,T]$. Therefore, $\epsilon\mapsto c(x^\epsilon(t),u(t))$ is continuous by Assumption 3.

Proposition 24. *Let $u\in U$ be a control, which yields the nominal state x. Let x^ϵ be the perturbed state induced by the perturbed control u^ϵ, and let $(y_1,\dots,y_T)$ be the given observations. Then, the following bounds hold for all $\epsilon\in[0,\tau]$:*

$$\forall t\in(\tau,1]\quad |c(x_1^\epsilon(t),u^\epsilon(t))-c(x_1(t),u(t))|\le\epsilon\beta'_1(x_0) \tag{240}$$

and

$$\forall i\in\{2,\dots,T\}\ \ \forall t\in[i-1,i]\quad |c(x_i^\epsilon(t),u^\epsilon(t))-c(x_i(t),u(t))| \le\epsilon\sum_{(j_1,\dots,j_{i-1})\in\mathcal{L}'_i}\beta_i'^{(j_1,\dots,j_{i-1})}(x_0)\prod_{m=1}^{i-1}\|y_m\|_2^{j_m}, \tag{241}$$

where $\beta'_1(x_0)$, $\beta_i'^{(j_1,\dots,j_{i-1})}(x_0)$ and $\mathcal{L}'$ are as defined in Proposition 22.

Proof. For $\epsilon \in [0, \tau]$ and $t \in (\tau, T]$, the function $\epsilon \mapsto c(x^\epsilon(t), u^\epsilon(t))$ is continuous by Lemma 23 and finite by Proposition 10. It is also right differentiable with respect to ϵ for $\epsilon \in [0, \tau)$ and $t \in (\tau, T]$ by Lemma 21. Therefore, the mean value theorem (Corollary in Bourbaki and Spain (2004), p.15) along with Proposition 22 proves the claim.

Lemma 25. *Let x^ϵ be the perturbed state induced by the perturbed control u^ϵ, and let $(y_1, \ldots, y_T)$ be the given observations. Then, the function $\epsilon \mapsto h(x^\epsilon(T))$ is continuous with respect to $\epsilon \in [0, \tau]$.*

Proof. By the proof of Lemma 23 it follows that the function $\epsilon \mapsto x^\epsilon(T)$ is continuous with respect to $\epsilon \in [0, \tau]$. The continuity of h by Assumption 3 completes the proof.

Proposition 26. *Let $u \in U$ be a control, which yields the nominal state x. Let x^ϵ be the perturbed state induced by the perturbed control u^ϵ, and let $(y_1, \ldots, y_T)$ be the given observations. Then, the following bound holds for all $\epsilon \in [0, \tau]$:*

$$|h(x^\epsilon(T)) - h(x(T))| \\ \leq \epsilon \sum_{(j_1,\ldots,j_T) \in \mathcal{L}'_{T+1}} \beta'^{(j_1,\ldots,j_T)}_{T+1}(x_0) \prod_{m=1}^{T} \|y_m\|_2^{j_m}, \tag{242}$$

where $\beta'^{(j_1,\ldots,j_T)}_{T+1}(x_0)$ and $\mathcal{L}'_{T+1}$ are as defined in Proposition 22.

Proof. The proof is very similar to that of Proposition 24. Use Proposition 10, Lemma 25, and Lemma 21 to show finiteness, continuity, and right differentiability of $\epsilon \mapsto h(x^\epsilon(T))$. Then use the same mean value theorem with Proposition 22 to prove the claim.

A.3 Expected Total Cost under Stochastic Observations

In this last part of the analysis, we finally let the observations $(y_1, \ldots, y_T)$ take random values; more formally, we treat them as a sequence of random variables $(Y_1(\omega), \ldots, Y_T(\omega))$ for $\omega \in \Omega$, where $(\Omega, \mathcal{F}, \mathbb{P})$ is the probability space and each Y_i satisfies Assumption 4. With (τ, v) given and fixed, (ω, t) and ϵ uniquely determine the perturbed control u^ϵ and the observations, hence the resulting state trajectory x^ϵ and the costs $c(x^\epsilon(t), u^\epsilon(t)), h(x^\epsilon(T))$.

Lemma 27. *Let $([\tau, T], \mathcal{B}([\tau, T]), \lambda)$ be a measure space, where $\mathcal{B}([\tau, T])$ is the Borel σ-algebra on $[\tau, T]$ and λ is the Lebesgue measure. Let $\mu \triangleq \lambda \times \mathbb{P}$ be the product measure defined on the product space $(\Omega \times [\tau, T], \mathcal{F} \otimes \mathcal{B}([\tau, T]))$, where $\mathcal{F} \otimes \mathcal{B}([\tau, T])$ is the product σ-algebra. Then, the function $(\omega, t) \mapsto c(x^\epsilon(t), u^\epsilon(t))$ is $\mathcal{F} \otimes \mathcal{B}([\tau, T])$-measurable for every $\epsilon \in [0, \tau]$.*

Proof. Take any $\epsilon \in [0, \tau]$. Then, u^ϵ is in U and thus the function $(Y_1(\omega), \ldots, Y_T(\omega)) \mapsto x^\epsilon(t)$ is continuous for every $t \in [\tau, T]$ by Proposition 8. Therefore, the map $(\omega, t) \mapsto x^\epsilon(t)$ as a function of ω is $\mathcal{F}$-measurable for every $t \in [\tau, T]$. By Corollary 4, $x^\epsilon(t)$ is also right continuous with respect to t for every $\omega \in \Omega$. Therefore, from Theorem 3 in Gowrisankaran (1972) it follows that $(\omega, t) \mapsto x^\epsilon(t)$ is measurable with respect to the product σ-algebra $\mathcal{F} \otimes \mathcal{B}([\tau, T])$.

On the other hand, $u^\epsilon(t)$ is piecewise continuous in t and is constant with respect to $(Y_1(\omega), \ldots, Y_T(\omega))$. Therefore, $(\omega, t) \mapsto u^\epsilon(t)$ is also measurable with respect to $\mathcal{F} \otimes \mathcal{B}([\tau, T])$.

Finally, the continuity of the instantaneous cost c by Assumption 3 proves the claim.

Proposition 28. *For the perturbed control u^ϵ and the perturbed state x^ϵ, we have*

$$\frac{\partial_+}{\partial \epsilon} \mathbb{E}\left[\int_\tau^T c(x^\epsilon(t), u^\epsilon(t)) dt\right]\Bigg|_{\epsilon=0} \\ = \mathbb{E}\left[\int_\tau^T \frac{\partial}{\partial x} c(x(t), u(t))^{\mathrm{T}} \Psi(t) dt\right], \tag{243}$$

where Ψ is the state variation defined in Proposition 16.

Proof. By definition, the left hand side of (243) is

$$\frac{\partial_+}{\partial \epsilon} \mathbb{E}\left[\int_\tau^T c(x^\epsilon(t), u^\epsilon(t)) dt\right]\Bigg|_{\epsilon=0} \\ = \lim_{\epsilon \to 0^+} \mathbb{E}\left[\int_\tau^T \frac{1}{\epsilon}\left\{c(x^\epsilon(t), u^\epsilon(t)) - c(x(t), u(t))\right\} dt\right]. \tag{244}$$

Consider the expected value above as the equivalent Lebesgue integral:

$$\int_\Omega \left(\int_{[\tau, T]} \frac{1}{\epsilon}\left\{c(x^\epsilon(t), u^\epsilon(t)) - c(x(t), u(t))\right\} d\lambda(t)\right) \\ \times d\mathbb{P}(\omega). \tag{245}$$

By Lemma 27 the integrand is measurable in the product space. In addition, Proposition 24 shows that the absolute value:

$$\frac{1}{\epsilon}\left|c(x^\epsilon(t), u^\epsilon(t)) - c(x(t), u(t))\right| \tag{246}$$

is bounded by an integrable function for μ-a.e. (ω, t). Indeed, if we let $\hat{c}(\omega, t)$ to be a function defined by

$$\hat{c}(\omega, t) = \begin{cases} \beta'_1(x_0) \ \ \forall t \in [\tau, 1) \\ \sum_{(j_1,\ldots,j_{i-1}) \in \mathcal{L}'_i} \beta'^{(j_1,\ldots,j_{i-1})}_i(x_0) \prod_{m=1}^{i-1} \|Y_m\|_2^{j_m} \\ \qquad \forall t \in [i-1, i) \ \forall i \in \{1, \ldots, T\}, \end{cases} \tag{247}$$

then we have

$$\frac{1}{\epsilon}\left|c(x^\epsilon(t), u^\epsilon(t)) - c(x(t), u(t))\right| \leq \hat{c}(\omega, t) \tag{248}$$

for every non-zero ϵ and μ-a.e. (ω, t), and

$$\int_\Omega \left(\int_{[\tau, T]} \hat{c}(\omega, t) d\lambda(t)\right) d\mathbb{P}(\omega) = \beta'_1(x_0)(1 - \tau) \\ + \sum_{i=2}^{T} \sum_{(j_1,\ldots,j_{i-1}) \in \mathcal{L}'_i} \beta'^{(j_1,\ldots,j_{i-1})}_i(x_0) \\ \times \mathbb{E}\left[\prod_{m=1}^{i-1} \|Y_m\|_2^{j_m}\right], \tag{249}$$

where

$$\mathbb{E}\left[\prod_{m=1}^{i-1}\|Y_m\|_2^{j_m}\right] \leq \sqrt{\mathbb{E}\left[\|Y_1\|_2^{j_1}\right]}\sqrt{\mathbb{E}\left[\prod_{m=2}^{i-1}\|Y_m\|_2^{j_m}\right]} \tag{250}$$

$$\leq$$

$$\vdots$$

$$\leq \prod_{m=1}^{i-1}\left(\mathbb{E}\left[\|Y_m\|_2^{j_m}\right]\right)^{\frac{1}{2^m}} < \infty \tag{251}$$

by the Cauchy-Schwarz inequality and Assumption 4.

Furthermore, Lemma 21 proves that

$$\lim_{\epsilon\to 0^+}\frac{1}{\epsilon}\left\{c(x^\epsilon(t),u^\epsilon(t)) - c(x(t),u(t))\right\} = \frac{\partial}{\partial x}c(x(t),u(t))^{\mathrm{T}}\Psi(t) \tag{252}$$

for μ-a.e. (ω,t). Therefore, the dominated convergence theorem yields

$$\frac{\partial_+}{\partial\epsilon}\int_\Omega\left(\int_\tau^T c(x^\epsilon(t),u^\epsilon(t))d\lambda(t)\right)d\mathbb{P}(\omega)\Bigg|_{\epsilon=0} = \int_\Omega\left(\int_\tau^T \frac{\partial}{\partial x}c(x(t),u(t))^{\mathrm{T}}\Psi(t)d\lambda(t)\right)d\mathbb{P}(\omega). \tag{253}$$

Proposition 29. *For the perturbed control u^ϵ and the perturbed state x^ϵ, we have*

$$\frac{\partial_+}{\partial\epsilon}\mathbb{E}\left[h(x^\epsilon(T))\right]\Big|_{\epsilon=0} = \mathbb{E}\left[\frac{\partial}{\partial x}h(x(T))^{\mathrm{T}}\Psi(T)\right], \tag{254}$$

where Ψ is the state variation defined in Proposition 16.

Proof. The proof is similar to that of Proposition 28. By Assumption 3 and Proposition 8, $(Y_1(\omega),\dots,Y_T(\omega)) \mapsto h(x^\epsilon(T))$ is a continuous map for every $\epsilon\in[0,\tau]$. Therefore, the function $\omega\mapsto h(x^\epsilon(T))$ is $\mathcal{F}$-measurable.

By definition, the left hand side of (254) is

$$\frac{\partial_+}{\partial\epsilon}\mathbb{E}\left[h(x^\epsilon(T))\right]\Big|_{\epsilon=0} = \lim_{\epsilon\to 0^+}\int_\Omega\frac{1}{\epsilon}\left\{h(x^\epsilon(T)) - h(x(T))\right\}d\mathbb{P}(\omega) \tag{255}$$

The analysis above implies that the integrand is $\mathcal{F}$-measurable. In addition, Proposition 26 shows that the absolute value:

$$\frac{1}{\epsilon}\left|h(x^\epsilon(T)) - h(x(T))\right| \tag{256}$$

is bounded by an integrable function for every non-zero ϵ and every $\omega\in\Omega$. Furthermore, Lemma 21 proves that

$$\lim_{\epsilon\to 0^+}\frac{1}{\epsilon}\left\{h(x^\epsilon(T)) - h(x(T))\right\} = \frac{\partial}{\partial x}h(x(T))^{\mathrm{T}}\Psi(T) \tag{257}$$

for every $\omega\in\Omega$. Therefore, the dominated convergence theorem yields

$$\frac{\partial_+}{\partial\epsilon}\int_\Omega h(x^\epsilon(T))d\mathbb{P}(\omega)\Big|_{\epsilon=0} = \int_\Omega\frac{\partial}{\partial x}h(x(T))^{\mathrm{T}}\Psi(T)d\mathbb{P}(\omega). \tag{258}$$

Theorem 1. Mode Insertion Gradient. *Suppose that Assumptions 1 – 4 are satisfied. For a given (τ,v), let u^ϵ denote the perturbed control of the form (116). The perturbed control u^ϵ and the stochastic observations $(Y_1,\dots,Y_T)$ result in the stochastic perturbed state trajectory x^ϵ. For such u^ϵ and x^ϵ, let us define the mode insertion gradient of the expected total cost as*

$$\frac{\partial_+}{\partial\epsilon}\mathbb{E}\left[\int_0^T c(x^\epsilon(t),u^\epsilon(t))dt + h(x^\epsilon(T))\right]\Bigg|_{\epsilon=0}. \tag{259}$$

Then, this right derivative exists and we have

$$\begin{aligned}\frac{\partial_+}{\partial\epsilon}\mathbb{E}\left[\int_0^T c(x^\epsilon(t),u^\epsilon(t))dt + h(x^\epsilon(T))\right]\Bigg|_{\epsilon=0} &= c(x(\tau),v) - c(x(\tau),u(\tau)) \\ &+ \mathbb{E}\Bigg[\int_\tau^T\frac{\partial}{\partial x}c(x(t),u(t))^{\mathrm{T}}\Psi(t)dt \\ &+ \frac{\partial}{\partial x}h(x(T))^{\mathrm{T}}\Psi(T)\Bigg],\end{aligned} \tag{260}$$

where Ψ is the state variation defined in Proposition 16.

Proof. We first consider the instantaneous cost c. Split the integration interval to get

$$\begin{aligned}\mathbb{E}\left[\int_0^T c(x^\epsilon(t),u^\epsilon(t))dt\right] &= \mathbb{E}\left[\int_0^{\tau-\epsilon} c(x^\epsilon(t),u^\epsilon(t))dt\right] \\ &+ \mathbb{E}\left[\int_{\tau-\epsilon}^{\tau} c(x^\epsilon(t),u^\epsilon(t))dt\right] \\ &+ \mathbb{E}\left[\int_\tau^T c(x^\epsilon(t),u^\epsilon(t))dt\right]\end{aligned} \tag{261}$$

For the first two terms in the sum, recall that the evolution of the state $x^\epsilon(t)$ is not affected by any observations for all $t\in[0,\tau]$. Thus,

$$\mathbb{E}\left[\int_0^{\tau-\epsilon} c(x^\epsilon(t),u^\epsilon(t))dt\right] = \int_0^{\tau-\epsilon} c(x^\epsilon(t),u^\epsilon(t))dt \tag{262}$$

$$\mathbb{E}\left[\int_{\tau-\epsilon}^{\tau} c(x^\epsilon(t),u^\epsilon(t))dt\right] = \int_{\tau-\epsilon}^{\tau} c(x^\epsilon(t),u^\epsilon(t))dt. \tag{263}$$

Note that (262) is constant with respect to ϵ, since for all $t\in[0,\tau-\epsilon]$ we have $u^\epsilon(t)=u(t)$ and $x^\epsilon(t)=x(t)$. On the other hand, for (263) we can apply Lemma 20 to obtain

$$\frac{\partial_+}{\partial\epsilon}\mathbb{E}\left[\int_{\tau-\epsilon}^{\tau} c(x^\epsilon(t),u^\epsilon(t))dt\right]\Bigg|_{\epsilon=0} = c(x(\tau),v) - c(x(\tau),u(\tau)) \tag{264}$$

For the last term, Proposition 28 gives

$$\frac{\partial_+}{\partial \epsilon}\mathbb{E}\left[\int_{\tau}^{T} c(x^\epsilon(t), u^\epsilon(t))dt\right]\Bigg|_{\epsilon=0} \\ = \mathbb{E}\left[\int_{\tau}^{T} \frac{\partial}{\partial x} c(x(t), u(t))^{\mathrm{T}} \Psi(t) dt\right]. \quad (265)$$

Finally, for the terminal cost h we have

$$\frac{\partial_+}{\partial \epsilon}\mathbb{E}\left[h(x^\epsilon(T))\right]\Big|_{\epsilon=0} = \mathbb{E}\left[\frac{\partial}{\partial x} h(x(T))^{\mathrm{T}} \Psi(T)\right] \quad (266)$$

by Proposition 29.

Remark 5. Closed-loop Nominal Policy. *As far as the control is concerned, the analysis above only requires that the nominal control u is in U (as in Assumption 1) and that the perturbed control u^ϵ is measurable with respect to $\mathcal{F} \otimes \mathcal{B}([\tau, T])$ (as in Lemma 27). To satisfy these requirements with a closed-loop nominal policy $\pi : \mathbb{R}^{n_x} \to \mathbb{R}^m$, it is sufficient that π is a measurable map and that the induced nominal control trajectory $u(t) = \pi(x(t))$ for $t \in [0, T]$ belongs to U for any given observations $(y_1, \ldots, y_T)$.*

Notice that even with state feedback, both the nominal control trajectory $u \in U$ and the nominal state trajectory x are uniquely determined under a fixed sequence of observations. Note also that the type of the control perturbation considered in this sensitivity analysis is still open-loop:

$$u^\epsilon(t) = \begin{cases} v & \textit{if } t \in (\tau - \epsilon, \tau] \\ \pi(x(t)) & \textit{otherwise}, \end{cases} \quad (267)$$

because we still follow Definition 1 for the control perturbation model. That is, the nominal state trajectory x is used in the control feedback. This is not to be confused with the closed-loop perturbation:

$$u^\epsilon_{\textit{closed}}(t) = \begin{cases} v & \textit{if } t \in (\tau - \epsilon, \tau] \\ \pi(x^\epsilon(t)) & \textit{otherwise}, \end{cases} \quad (268)$$

where the perturbed state trajectory x^ϵ is fed back to the perturbed control.