Automatica 171 (2025) 111978

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

Approximate constrained stochastic optimal control via parameterized input inference ☆

Shahbaz P. Qadri Syed, He Bai *

Mechanical and Aerospace Engineering, Oklahoma State University, Stillwater, OK, 74078, United States of America

ARTICLE INFO

Article history:
Received 28 July 2023
Received in revised form 1 June 2024
Accepted 23 August 2024
Available online xxxx

Keywords:
Inference-based control
Structured control
Parametric optimization
Multi-agent systems
Stochastic control

ABSTRACT

Approximate methods to solve stochastic optimal control (SOC) problems have received significant interest from researchers in the past decade. Probabilistic inference approaches to SOC have been developed to solve nonlinear quadratic Gaussian problems. In this work, we propose an Expectation–Maximization (EM) based inference procedure to generate state-feedback controls for constrained SOC problems. We consider the inequality constraints for the state and controls and also the structural constraints for the controls. We employ barrier functions to address state and control constraints. We show that the expectation step leads to smoothing of the state-control pair while the maximization step on the non-zero subsets of the control parameters allows inference of structured stochastic optimal controllers. We demonstrate the effectiveness of the algorithm on unicycle obstacle avoidance and four-unicycle formation control examples. In these examples, we perform an empirical study on the parametric effect of barrier functions on the state constraint satisfaction. We also present a comparative study of smoothing algorithms on the performance of the proposed approach.

1. Introduction

Stochastic optimal control (SOC) is defined as the problem of finding a controller that minimizes an expected cost in the presence of uncertainty and dynamics constraint. The uncertainty is either in the form of noisy observations or process noise that approximates model uncertainties in the system. A solution to the SOC problem can be found by solving the nonlinear stochastic Hamilton–Jacobi–Bellman (HJB) equation (Stengel, 1994). In general, its numerical solution is computationally intractable due to the curse of dimensionality resulting from the discretization of the space and time (Todorov, 2006). A fast and locally approximate solution to the SOC problem is the Linear Quadratic Gaussian (LQG) case where the SOC problem is solved for the noise-free optimal trajectory and a local LQG model is constructed as perturbation around this trajectory. The local linear quadratic regulator computes a reasonable approximate solution to the original SOC problem if the model is close to the optimal noise-free trajectory.

The general duality between control and estimation (Todorov, 2008) and the notion of relating the cost and log-likelihood have motivated a new class of methods to approximately solve the SOC problem in a non-LQG setting. These methods are often referred to as *control-as-inference* methods in literature which solve the SOC problem as an inference problem on a probabilistic graphical model (PGM). A PGM is a graphical model encoding complex relationships between random variables in the form of a graph. It is widely used in statistics and machine learning to model joint probability distributions of random variables. This graphical representation of probability distribution is advantageous as it allows the decomposition of the joint probability distribution as a product of factors by exploiting the structure of the model. Moreover, algorithms developed in this framework have shown propitious results in real-world applications (see e.g. Itoh et al., 2017; Rawlik, Toussaint, & Vijayakumar, 2010; Rückert & Neumann, 2013; Toussaint, 2009; Watson, Abdulsamad, Findeisen, & Peters, 2021; Watson, Abdulsamad, & Peters, 2020). A common limitation of the above inference-based control approaches is the restriction to linear feedback controllers to achieve closed-form updates in a Gaussian setting. It is well known that nonlinear systems typically admit nonlinear optimal controllers, and hence the use of the existing linear controllers will yield sub-optimal performance in a nonlinear setting. In our prior work (Syed & Bai, 2023), we propose the Parameterized Input Inference for Control (*PIIC*) algorithm where the controller is parameterized by a (possibly) nonlinear basis function of the state which allows

☆ The work was supported by the U.S. National Science Foundation (NSF) under Grant Nos. 1925147, 2212582, and 2241585. The material in this paper was partially presented at the 2023 American Control Conference (ACC), May 31–June 2, 2023, San Diego, California, USA. This paper was recommended for publication in revised form by Associate Editor Simone Formentin under the direction of Editor Alessandro Chiuso.

* Corresponding author.
E-mail addresses: shahbaz_qadri.syed@okstate.edu (S.P.Q. Syed), he.bai@okstate.edu (H. Bai).

https://doi.org/10.1016/j.automatica.2024.111978
0005-1098/

formulating the unconstrained SOC problem as a parameter inference problem. Hence, one of the contributions of this paper is that we employ a barrier function approach to solve constrained SOC problems using the PIIC algorithm.

In recent years, the design of structured controllers has received a lot of attention for applications in large-scale systems and multi-agent systems. A structured controller reduces the computational load by translating the topology of networked systems to the sparsity of the controller, facilitating distributed controls at subsystems. An example of structured control is distributed optimal control for multi-agent systems, where the control of each agent contains information only from a subset of the agents. However, to the best of our knowledge, none of the existing inference-based control approaches have been developed in the structured control domain owing to the challenge of encoding and preserving the structure imposed on the control gain. Hence, the main contribution of this work is that we propose a *structured*-PIIC algorithm to solve structured SOC problems in an inference-based control framework.

The main contributions of this work are as follows: (1) We enhance the formulation of the PIIC algorithm (Syed & Bai, 2023) to address constrained SOC problems, where the constraints include both state, control constraints and structural constraints on the state-feedback controllers. Although structured optimal control has been investigated for deterministic systems (see e.g., Fardad & Jovanović, 2014; Jovanović & Dhingra, 2016; Lin, Fardad, & Jovanović, 2011), our approach provides an effective structured control solution for stochastic systems. The resulting algorithm is an instance of the EM procedure which has a guaranteed convergence to local optima. (2) We empirically demonstrate the effectiveness of the proposed algorithm with respect to constraint satisfaction and structured control using unicycle control problems. The algorithm outperforms the commonly-used Iterative Linear Quadratic Gaussian (ILQG) approach (Todorov & Li, 2005) with reduced mean cost and cost variance.

The rest of the paper is organized as follows. Section 2 reviews the formulation of the SOC problem in an inference-based control framework. Section 3 presents our algorithm to address constrained SOC problems. Section 4 demonstrates the efficacy of our approach on a unicycle model in constrained control and structured control scenarios. Section 5 concludes the paper.

Notation: Let $\mathcal{N}(y|a, A)$ represent a random variable y satisfying a Gaussian distribution in the normal form with mean $a \in \mathbb{R}^d$ and covariance $A \in \mathbb{R}^{d\times d}$ given by $\mathcal{N}(y|a, A) = \frac{1}{(2\pi)^{\frac{d}{2}}|A|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(y-a)^\intercal A^{-1}(y-a)\right)$, where $|A|$ represents the determinant of A. We use blkdiag$(A_1, A_2, \ldots, A_n)$ to denote a block diagonal matrix with matrices $A_1, A_2, \ldots, A_n$ on its principal diagonal. $\mathbb{I}_n$ denotes the identity matrix of size n. $\otimes$ denotes the Kronecker product. Tr$(\cdot)$ denotes the trace operator, and $\mathbb{E}(\cdot)$ denotes the expectation operator. $\mathbf{1}_{m\times n}$, $\mathbf{0}_{m\times n}$ denote the $m \times n$ matrices with entries 1 and 0, respectively.

2. Inference-based stochastic optimal control

Consider a dynamical system given by

$$x_{t+1} = F(\tau_t) + \eta_t, \tag{1}$$

where $\tau_t = [x_t^\intercal, u_t^\intercal]^\intercal \in \mathbb{R}^{n_x+n_u}$ is the state-control vector at time t, $x_t \in \mathbb{R}^{n_x}$ and $u_t \in \mathbb{R}^{n_u}$ denote the state and control at time t, respectively. F: $\mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_x}$ is a nonlinear mapping of x_t, u_t, and $\eta_t \sim \mathcal{N}(\eta_t|0, \Sigma_{\eta_t})$ represents additive Gaussian noise that models the uncertainty in the dynamics. For a given finite-horizon T, and a state-control sequence $[x_T, \tau_{0:T-1}]$, define the trajectory cost as $\mathcal{C}(x_T, \tau_{0:T-1}) = c_T(x_T) + \sum_{t=0}^{T-1} c_t(\tau_t)$, where $c_t : \mathbb{R}^{n_x+n_u} \to \mathbb{R}$ is a nonlinear mapping from the state-control space to the cost space for $t < T$ and $c_T : \mathbb{R}^{n_x} \to \mathbb{R}$ is a nonlinear mapping from the state space to the cost space at the terminal time T. The considered SOC problem is given by

$$\min_{u_{0:T-1}} \mathbb{E}[\mathcal{C}(x_T, \tau_{0:T-1})] \tag{2}$$
$$\text{such that } x_{t+1} \sim \mathcal{N}(x_{t+1}|F(\tau_t), \Sigma_{\eta_t}),$$
$$\mathcal{K}(\tau_t) > 0,$$

where $\mathcal{K}(\cdot) \in \mathbb{R}^{n_{\text{in}}}$ is such that $\mathcal{K}_j(\cdot) : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \to \mathbb{R}$, $j = 1, \ldots, n_{\text{in}}$ is a nonlinear mapping that defines an inequality constraint. We assume that the feedback controller u_t at each time step is parameterized by a (possibly nonlinear) basis function of the state, $\mathcal{B}_t(x_t) \in \mathbb{R}^{n_b}$, and unknown parameters $\Theta_t \in \mathbb{R}^{n_b\times n_u}$ such that

$$p(u_t|x_t) = \mathcal{N}(u_t|\Theta_t^\intercal \mathcal{B}_t(x_t), \Sigma_{\delta_t}), \tag{3}$$

where δ_t represents a zero-mean Gaussian noise with covariance Σ_{δ_t} that models the uncertainty in control.

The PGM for the SOC problem (2) is constructed with the state-control sequence as latent variables and the sequence of binary random variables $\mathcal{O}_t \in \{0, 1\}$, $t = 0, \ldots, T$, as observed variables. The binary random variable $\mathcal{O}_t$ represents the notion of optimality or task fulfillment at each time step, i.e., $\mathcal{O}_t = 1$ when optimal state and action are observed at time t. Similar to the general duality between estimation and control (Todorov, 2008), probabilistic inference approaches relate the probabilities to cost by assuming that the negative log-likelihood of observing the optimality/task fulfillment at time t is proportional to the stage cost c_t, i.e.,

$$p(\mathcal{O}_t = 1|\tau_t) \propto \exp\{-c_t(\tau_t)\}. \tag{4}$$

Hence, the likelihood of observing optimality at each time step is high if and only if the cost incurred is low. We have shown in our prior work (Syed & Bai, 2023) that the parameterization in (3) yields nonlinear controllers for the unconstrained version of (2) using the EM procedure. The focus of this work is to extend the formulation to constrained and structured SOC problems.

3. Constrained stochastic optimal control

We consider two types of constraints in the SOC problem. Section 3.1 addresses inequality constraints on τ_t, which are particularly useful for maintaining safety of the system and creating bounded controls. Section 3.2 examines structural constraints on the control, which can be used for designing distributed controllers. Corresponding examples are demonstrated in Section 4.

3.1. State and control constraints

We present an approach to embed inequality constraints on τ_t into the inference-based control formulation in Section 2. We are motivated by the barrier function method, which is a popular approach in optimization literature to solve a constrained optimization problem as a sequence of unconstrained optimization problems by adding a high cost for approaching the boundary of feasibility region from the interior (Bertsekas, 2016, Chapter 5). It is also similar to the potential function approach commonly used for collision avoidance and motion planning (Kavraki & LaValle, 2008).

Let the safe set for constraint $j = \{1, \ldots, n_{\text{in}}\}$ be given by $\mathcal{C}_{s,j} = \{\tau_t \in \mathbb{R}^{n_x+n_u} | \mathcal{K}_j(\tau_t) > 0\}$, where $\mathcal{C}_{s,j}$ is assumed to be non-empty $\forall j$. A barrier function $B(\tau)$ is continuous in the interior of $\mathcal{C}_{s,j}$ and goes to ∞ as one of the constraints $\mathcal{K}_j$ approaches 0 from positive values. Motivated by this approach, we define a

relaxed barrier function for each constraint, denoted by $c_{\text{in},j}(\tau_t)$, that evaluates to zero if and only if $\tau_t \in \mathcal{C}_{s,j}$, and is positive otherwise, i.e.,

$$c_{\text{in},j}(\tau_t) = (\psi_j(\tau_t))^\intercal Q_j^{\text{in}} \psi_j(\tau_t) \begin{cases} = 0, & \text{if } \tau_t \in \mathcal{C}_{s,j} \\ > 0, & \text{otherwise,} \end{cases} \tag{5}$$

where $\psi(\tau_t)$ is a (possibly) nonlinear function of τ_t. The $c_{\text{in},j}(\tau_t)$ can be considered the cost for the satisfaction of constraint j. It is positive when the constraint is violated and zero otherwise. As shown later, we employ a likelihood function $\exp(-c_{\text{in},j}(\tau_t))$ to encode the satisfaction of constraint j into our inference-based control formulation. According to (5), the likelihood function evaluates to 1 in the safe set $\mathcal{C}_{s,j}$, which is the maximum of $\exp(-c_{\text{in},j}(\tau_t))$. Thus, satisfaction of constraint j is encoded with a higher likelihood of occurrence.

Let $\mathcal{O}_t^\tau$, $\mathcal{O}_t^{\text{in},j}$ denote the binary random variables corresponding to observing optimality in the cost, and in the satisfaction of constraint j, respectively. We prescribe $p(\mathcal{O}_t = 1|\tau_t) \propto p(\mathcal{O}_t^\tau = 1|\tau_t) \prod_j p(\mathcal{O}_t^{\text{in},j} = 1|\tau_t)$. Letting $p(\mathcal{O}_t^{\text{in},j} = 1|\tau_t) \propto \exp(-c_{\text{in},j}(\tau_t))$, we rewrite (4) as

$$p(\mathcal{O}_t = 1|\tau_t) \propto \exp\{-c_t(\tau_t) - \sum_{j=1}^{n_{\text{in}}} c_{\text{in},j}(\tau_t)\}. \tag{6}$$

Suppose that the trajectory cost $c_t(\tau_t)$ is quadratic. Adding the barrier function in (5) as a cost to $c_t(\tau_t)$ yields $\forall\ t = 0, \dots, T$,

$$\begin{aligned} c_t(\tau_t) + \sum_{j=1}^{n_{\text{in}}} c_{\text{in},j}(\tau_t) &= (x_t - x_t^d)^\intercal Q_t (x_t - x_t^d) \\ &+ (u_t - u_t^d)^\intercal R_t (u_t - u_t^d) + \sum_{j=1}^{n_{\text{in}}} (\psi_j(\tau_t))^\intercal Q_j^{\text{in}} (\psi_j(\tau_t)), \end{aligned} \tag{7}$$

where $Q_t \succeq 0$, $R_t \succeq 0$, and $Q_j^{\text{in}} \succeq 0$, $j = 1, \dots, n_{\text{in}}$, are the cost matrices. It then follows from (6) and (7) that

$$\begin{aligned} p(\mathcal{O}_t = 1|\tau_t) &\propto \exp\{-\alpha (z_t^* - h(\tau_t))^\intercal \Gamma_t (z_t^* - h(\tau_t))\} \\ &= \mathcal{N}(z_t = z_t^* | h(\tau_t), (\alpha \Gamma_t)^{-1}), \end{aligned} \tag{8}$$

where $\Gamma_t = \text{blkdiag}(Q_t, R_t, Q_1^{\text{in}}, \dots, Q_{n_{\text{in}}}^{\text{in}}) \in \mathbb{R}^{n^* \times n^*}$, $h(\tau_t) = \begin{bmatrix} \tau_t^\intercal & \psi(\mathcal{K}_1(\tau_t)) & \cdots & \psi(\mathcal{K}_{n_{\text{in}}}(\tau_t)) \end{bmatrix}^\intercal \in \mathbb{R}^{n^*}$, $z_t^* = \begin{bmatrix} (\tau_t^d)^\intercal & 0 & \cdots & 0 \end{bmatrix}^\intercal \in \mathbb{R}^{n^*}$ with $\tau_t^{d\intercal} = [(x_t^d)^\intercal\ (u_t^d)^\intercal]$, $n^* = (n_x + n_u + n_{in})$, and α is the scale factor (hyperparameter) introduced to optimize the covariance of $\mathcal{O}_t$ to maximize the expected log-likelihood.

An optimal trajectory is computed as the mean of the conditional or joint posterior distribution of the state-control trajectory given that the optimality is observed throughout the entire trajectory, i.e., $\mathcal{O}_{0:T} = 1$. The objective of the PIIC algorithm is to infer the parameters $\Theta_{0:T-1}$ and α that maximize the log-likelihood, i.e.,

$$\Theta_{0:T-1}^*, \alpha^* = \underset{\Theta_{0:T-1}, \alpha}{\text{argmax}} \log[p(\mathcal{O}_{0:T} = 1|\Theta_{0:T-1}, \alpha)]. \tag{9}$$

The optimization problem in (9) is generally intractable. Thus, we resort to computing the parameters using the EM algorithm. The EM algorithm is an iterative algorithm used to find maximum likelihood solutions for models with latent variables. It performs consecutive expectation (E-step) and maximization (M-step) steps in each iteration. The E-step computes the expected log-likelihood over the posterior distribution of latent variables and the consequent M-step computes the parameters that maximize this expectation. Each iteration of the EM algorithm results in a non-decreasing expected log-likelihood, thus guaranteeing convergence to a local maximum. We refer interested readers to Bishop (2006) for a detailed introduction to the EM algorithm.

Denote $\tau_{0:T-1}$ by τ, $\mathcal{O}_{0:T} = 1$ by $\mathcal{O}$, and $\Theta_{0:T-1}$ by $\boldsymbol{\Theta}$. Then the objective in (9) is rewritten as

$$\log[p(\mathcal{O}|\boldsymbol{\Theta}, \alpha)] = \log\Big[\int p(x_T, \tau, \mathcal{O}|\boldsymbol{\Theta}, \alpha) d\tau\, dx_T\Big]. \tag{10}$$

The integrand in (10) is proportional to the joint posterior distribution given by

$$\begin{aligned} p(x_T, \tau, \mathcal{O}, \boldsymbol{\Theta}, \alpha) &= p(x_0) p(\mathcal{O}_T = 1|x_T, \alpha) \\ &\prod_{t=0}^{T-1} p(x_{t+1}|\tau_t) p(\mathcal{O}_t = 1|\tau_t, \alpha) p(u_t|x_t, \Theta_t). \end{aligned} \tag{11}$$

Introducing $q(x_T, \tau)$, a known tractable distribution of x_T and τ, we obtain

$$\log[p(\mathcal{O}|\boldsymbol{\Theta}, \alpha)] = \log\Big[\mathop{\mathbb{E}}_{q(x_T,\tau)}\Big[\frac{p(x_T, \tau, \mathcal{O}|\boldsymbol{\Theta}, \alpha)}{q(x_T, \tau)}\Big]\Big]$$

. Using Jensen's inequality, we further get

$$\log[p(\mathcal{O}|\boldsymbol{\Theta}, \alpha)] \geq \mathop{\mathbb{E}}_{q(x_T,\tau)} \log\Big[\frac{p(x_T, \tau, \mathcal{O}|\boldsymbol{\Theta}, \alpha)}{q(x_T, \tau)}\Big]. \tag{12}$$

Note that (12) becomes equality for $q(x_T, \tau) = p(x_T, \tau|\mathcal{O})$. The PIIC algorithm optimizes the right-hand side of (12) based on the EM procedure. Hence, convergence to a local maximum is guaranteed (Moon, 1996).

Substituting (11) in the M-step yields

$$\begin{aligned} \underset{\boldsymbol{\Theta}, \alpha}{\text{argmax}} \mathop{\mathbb{E}}_{q(x_T,\tau)} \Big[\log p(x_0) + \sum_{t=1}^{T-1} \log p(x_{t+1}|\tau_t) + \\ \sum_{t=0}^{T} \log p(\mathcal{O}_t = 1|\tau_t, \alpha) + \sum_{t=0}^{T-1} \log p(u_t|x_t, \Theta_t)\Big]. \end{aligned} \tag{13}$$

To find Θ_t^{k+1}, we take gradient of (13) with respect to Θ_t and set it to zero, which yields

$$\Theta_t^{k+1} = \Big[\mathop{\mathbb{E}}_{q(\tau_t)} (\mathcal{B}_t(x_t)(\mathcal{B}_t(x_t))^\intercal)\Big]^{-1} \mathop{\mathbb{E}}_{q(\tau_t)} (\mathcal{B}_t(x_t) u_t^\intercal). \tag{14}$$

It is straightforward to show that if the control parameter is time-invariant i.e., $\Theta_{0:T-1} = \Theta$ then

$$\Theta^{k+1} = \Big[\mathop{\mathbb{E}}_{q(\tau)} \Big(\sum_{t=0}^{T-1} \mathcal{B}_t(x_t)(\mathcal{B}_t(x_t))^\intercal\Big)\Big]^{-1} \mathop{\mathbb{E}}_{q(\tau)} \Big(\sum_{t=0}^{T-1} \mathcal{B}_t(x_t) u_t^\intercal\Big). \tag{15}$$

Similarly, to find α^{k+1} we take gradient of (13) with respect to α and set it to zero, which yields

$$\alpha^{k+1} = \frac{(T-1)n_z + n_{z_T}}{\sum_{t=0}^{T} \text{Tr}(\Gamma_t \mathop{\mathbb{E}}_{q(x_T,\tau)} [(z_t^* - z_t)(z_t^* - z_t)^\intercal])}, \tag{16}$$

where $q(x_T, \tau) = p(x_T, \tau|\mathcal{O})$.

In this paper, we define *approximate inference* as the inference of the latent variables of a PGM. Approximate inference can also be defined as an approximation of the true posterior with a family of distributions that minimizes the KL divergence (Rawlik, Toussaint, & Vijayakumar, 2013). Let $q_\pi(\tau) = \prod_{t=0}^{T-1} p(u_t|x_t) p(x_{t+1}|\tau_t)$, be the state-control distribution parameterized by Θ and $q_s(\tau) = p(\tau|\mathcal{O})$ be the smoothed state-control distribution.

Proposition 1. *The minimization of the KL divergence $KL(q_s \| q_\pi)$ is equivalent to the minimization of the objective (13) with respect to the parameter Θ.*

Proof. From the definition, we have

$$KL(q_s\|q_\pi) = \int_\tau q_s \log\left(\frac{q_s}{q_\pi}\right) d\tau = \mathop{\mathbb{E}}_{q_s}\log(q_s) - \mathop{\mathbb{E}}_{q_s}\log(q_\pi). \tag{17}$$

To minimize (17) w.r.t. Θ, we take the gradient and set it to zero, resulting in

$$\Theta^* = \left[\mathop{\mathbb{E}}_{q_s}\left[\sum_{t=0}^{T-1}\mathcal{B}_t(x_t)(\mathcal{B}_t(x_t))^\intercal\right]\right]^{-1}\left[\mathop{\mathbb{E}}_{q_s}\left[\sum_{t=0}^{T-1}\mathcal{B}_t(x_t)u_t^\intercal\right]\right], \tag{18}$$

which is equivalent to (15). □

3.2. Structured control

Structured optimal control primarily deals with the design of static optimal controllers for interconnected systems with topological constraints. These topological constraints are translated as sparsity in the feedback gain. The problem of designing optimal controllers with structured feedback gains has been well studied for deterministic systems (e.g., see Fardad & Jovanović, 2014; Jovanović & Dhingra, 2016; Lin et al., 2011). However, it has not been fully explored for stochastic systems. We impose a structural constraint on the controller gain matrix Θ_t. We assume that the state x_t and the control u_t are composed of N subsystem states and M subcontrols, respectively, i.e., $x_t = [(x_t^1)^\intercal, \dots, (x_t^N)^\intercal]^\intercal$ and $u_t = [(u_t^1)^\intercal, \dots, (u_t^M)^\intercal]^\intercal$. The x_t^i and u_t^j can be multidimensional, $i = 1, \dots, N$, $j = 1, \dots, M$. Denote by $b^i(x_t^i) \in \mathbb{R}^{n_{b^i}}$ and by θ_t^{ij} the basis function corresponding to x_t^i and the submatrix of the controller gain Θ_t corresponding to u^j and $b^i(x_t^i)$, respectively. The subcontrols u_t^j, $j = 1, \dots, M$, are parameterized as

$$u_t^j = \begin{bmatrix}(\theta_t^{1j})^\intercal & (\theta_t^{2j})^\intercal & \cdots & (\theta_t^{Nj})^\intercal\end{bmatrix}\mathcal{B}_t(x_t) + \delta_t^j, \tag{19}$$

where $\mathcal{B}_t(x_t) = [(b^1(x_t^1))^\intercal\ (b^2(x_t^2))^\intercal\ \cdots\ (b^N(x_t^N))^\intercal]^\intercal$, $\delta_t^j \sim \mathcal{N}(\delta_t^j|0, (\sigma_t^j)^2)$. We assume that δ_t^j, $j = 1, \dots, M$, are i.i.d. zero mean Gaussian noise. Following the notation in (3), we have

$$\Theta_t = \begin{bmatrix}(\theta_t^{11})^\intercal & (\theta_t^{21})^\intercal & \cdots & (\theta_t^{N1})^\intercal \\ \vdots & \vdots & \vdots & \vdots \\ (\theta_t^{1M})^\intercal & (\theta_t^{2M})^\intercal & \cdots & (\theta_t^{NM})^\intercal\end{bmatrix}^\intercal.$$

Let $\mathcal{F}$ be the set of ordered pairs such that $(i,j) \in \mathcal{F}$ if the subcontrol u_t^j can receive information from the subsystem state x_t^i. Consider the structured SOC problem:

$$\min_{u_{0:T-1}} \mathbb{E}[\mathcal{C}(x_T, \tau_{0:T-1})] \tag{20}$$

such that $x_{t+1} \sim \mathcal{N}(x_{t+1}|F(\tau_t), \Sigma_{\eta_t})$,

$$\mathcal{K}(\tau_t) > 0,\ \theta_t^{ij} = \mathbf{0}_{n_{b^i}\times n_{u^j}},\ \text{if } (i,j) \notin \mathcal{F}.$$

Our key idea to solve the structured SOC problem is to decompose the problem into multiple unstructured SOC problems in a lower dimensional subspace of nonzero entries corresponding to each element of u_t. Then, the inferred parameters are mapped back to the original vector space through an inverse transformation which preserves the structure during the inference procedure.

To capture the structural constraints, we define a structural identity (under element-wise matrix multiplication) of the feedback gain Θ_t, denoted by $\Phi \in \mathbb{R}^{n_b\times n_u}$. The Φ is a block matrix whose (i,j)th block is all ones if u_t^j depends on $b^i(x_t^i)$ and otherwise all zeros, that is, $\forall\ i = 1, \dots, N, j = 1, \dots, M$,

$$\Phi_{ij} = \begin{cases}\mathbf{1}_{n_{b^i}\times n_{u^j}}, & \text{if } (i,j) \in \mathcal{F} \\ \mathbf{0}_{n_{b^i}\times n_{u^j}}, & \text{otherwise.}\end{cases} \tag{21}$$

Let u_t^p be the pth element of u_t and $\Phi_p \in \mathbb{R}^{n_b}$ be the pth column of Φ, where $p = \{1, \dots, n_u\}$. For every Φ_p, there exists an $\mathcal{S}_p : \mathbb{R}^{n_b} \to \mathbb{R}^{\tilde{n}_b}$ that maps Φ_p to its lower dimensional non-zero entries $\tilde{\Phi}_p \in \mathbb{R}^{\tilde{n}_b}$, where $\tilde{n}_b \le n_b$. Hence, $\mathcal{S}_p$ can be applied to the pth column of Θ_t, denoted by Θ_t^p, to extract its non-zero entries, denoted by $\tilde{\Theta}_t^p \in \mathbb{R}^{\tilde{n}_b}$, i.e., $\tilde{\Theta}_t^p = \mathcal{S}_p\Theta_t^p$. Similarly, we let $\tilde{\mathcal{B}}_t^p(x_t) = \mathcal{S}_p\mathcal{B}_t^p(x_t)$. Also, for every $\mathcal{S}_p$, there exists an $\mathcal{S}_p' : \mathbb{R}^{\tilde{n}_b} \to \mathbb{R}^{n_b}$ that maps $\tilde{\Theta}_t^p$ back to Θ_t^p.

For example, consider an interconnected system with four subsystem states and three subcontrols, i.e., $x_t = [x_t^1\ x_t^2\ x_t^3\ x_t^4]^\intercal \in \mathbb{R}^4$ and $u_t = [u_t^1\ u_t^2\ u_t^3]^\intercal \in \mathbb{R}^3$. Assume that u_t^j's are linear functions of the states. Consider the following structural constraints on u_t^j's: u_t^1 depends only on x_t^1 and x_t^3, u_t^2 only on x_t^1, x_t^2, and x_t^4, and u_t^3 only on x_t^3 and x_t^4. Then (19) takes the form

$$u_t = \underbrace{\begin{bmatrix}\theta^{11} & 0 & \theta^{31} & 0 \\ \theta^{12} & \theta^{22} & 0 & \theta^{42} \\ 0 & 0 & \theta^{33} & \theta^{43}\end{bmatrix}}_{\Theta_t^\intercal}\underbrace{\mathcal{B}_t(x_t)}_{x_t} + \delta_t,$$

where $\delta_t = [\delta_t^1\ \delta_t^2\ \delta_t^3]^\intercal$. By definition, $\Phi = \begin{bmatrix}1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1\end{bmatrix}^\intercal$. Then, $\Phi_1 = \begin{bmatrix}1 & 0 & 1 & 0\end{bmatrix}^\intercal$, $\mathcal{S}_1 = \begin{bmatrix}1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0\end{bmatrix}$ and $\mathcal{S}_1' = \mathcal{S}_1^\intercal$. Therefore, $\tilde{\Theta}_t^1 = \mathcal{S}_1\Theta_t^1 = [\theta^{11}\ \theta^{13}]^\intercal$ and $\Theta_t^1 = \mathcal{S}_1'\tilde{\Theta}_t^1$. Similarly, $\mathcal{S}_2$, $\mathcal{S}_3$ can be computed corresponding to Θ_t^2, Θ_t^3 respectively.

Using the notation $\tilde{\Theta}_t^p$ and $\tilde{\mathcal{B}}_t^p(x_t)$, taking the gradient of (13) against $\tilde{\Theta}_t^p$, and equating it to zero yields the update equation for $\tilde{\Theta}_t^p$ as

$$(\tilde{\Theta}_t^p)^{k+1} = \left[\mathop{\mathbb{E}}_{q(\tau_t)}[\tilde{\mathcal{B}}_t^p(x_t)(\tilde{\mathcal{B}}_t^p(x_t))^\intercal]\right]^{-1}\mathop{\mathbb{E}}_{q(\tau_t)}[\tilde{\mathcal{B}}_t^p(x_t)(u_t^p)^\intercal]. \tag{22}$$

From an implementation perspective, a time-invariant control parameter Θ may be advantageous. Following a similar approach to (22), we obtain the time-invariant parameter update as

$$(\tilde{\Theta}^p)^{k+1} = \Big[\mathop{\mathbb{E}}_{q(\tau)}\Big[\sum_{t=0}^{T-1}\tilde{\mathcal{B}}_t^p(x_t)(\tilde{\mathcal{B}}_t^p(x_t))^\intercal\Big]\Big]^{-1}\mathop{\mathbb{E}}_{q(\tau)}\Big[\sum_{t=0}^{T-1}\tilde{\mathcal{B}}_t^p(x_t)(u_t^p)^\intercal\Big]. \tag{23}$$

The covariance of the controller σ^p is updated $\forall\ p = 1, \dots, n_u$ and $t = 0, \dots, T-1$ using

$$\sigma_t^p = \mathop{\mathbb{E}}_{q(\tau)}(u_t^p - (\tilde{\Theta}_t^p)^\intercal\tilde{\mathcal{B}}_t(x_t^p))(u_t^p - (\tilde{\Theta}_t^p)^\intercal\tilde{\mathcal{B}}_t(x_t^p))^\intercal. \tag{24}$$

Algorithm 1 summarizes the structured parameterized input inference for control (structured PIIC) algorithm. It performs the E-step and the M-step iteratively until convergence. The structure imposed on the control parameter Θ is preserved by performing updates on the non-zero subsets of each subsystem using (22) and (24). In our implementation, we claim convergence of the algorithm if the infinity norm of the difference between the state trajectories in two consecutive iterations is less than a threshold. As shown in the Appendix, Algorithm 1 recovers the Gaussian I2C (Watson et al., 2021) for linear dynamics without any constraints if $\mathcal{B}_t(x_t) = [x_t^\intercal\ 1]^\intercal$, and Θ_t does not have a specific structure.

4. Simulation examples

In this section, we demonstrate the effectiveness of the PIIC algorithm for inference of constrained and structured stochastic optimal controllers. In Section 4.1, we demonstrate the effectiveness of the barrier function approach for a unicycle obstacle

Algorithm 1 Structured PIIC algorithm

repeat
E-step: Compute

$$q^{k+1} = p(x_T, \tau|\mathcal{O}, \Theta^k, \alpha^k)$$

$$Q(\Theta, \alpha|\Theta^k, \alpha^k) = \mathop{\mathbb{E}}_{(x_T,\tau)\sim q^{k+1}} \log[p(x_T, \tau, \mathcal{O}|\Theta, \alpha)]$$

M-step:
for t= 0 : $T-1$ **do**
for p = 1 : n_u **do**
Update $\tilde{\Theta}_t^p$, σ_t^p using (22), (24), respectively.
$\Theta_t^p = \mathcal{S}_p'(\tilde{\Theta}_t^p)$
end for
Update $\Theta_t = \begin{bmatrix}\Theta_t^1 & \cdots & \Theta_t^{n_u}\end{bmatrix}^\intercal$
Update $\Sigma_{\delta_t} = \text{blkdiag}(\sigma_t^1, \cdots, \sigma_t^N)$
end for
Update α using (16)
until convergence

avoidance problem. We also study the performance of the PIIC for the choice of two smoothing approaches and compare them with the ILQG baseline. In Section 4.2, we illustrate the utility of the PIIC for distributed formation control of four unicycle robots. The common simulation parameters are step size $dt = 0.05$ sec, and the state cost matrix $Q_t = \mathbb{I}_3$.

4.1. Obstacle avoidance

Consider a unicycle robot whose dynamics are given as

$$X_{t+1} = X_t + dt\, f(X_t, u_t) + \eta_t, \tag{25}$$

where at any time instant t, $X_t = [x_t\ y_t\ \theta_t]^\intercal \in \mathbb{R}^3$ denotes the 2-dimensional positions and heading of the robot, $u_t = [v_t\ \omega_t]^\intercal \in \mathbb{R}^2$ denotes the linear and angular velocities of the robot, $f_t(X_t, u_t) = [v_t \cos(\theta_t)\ v_t \sin(\theta_t)\ \omega_t]^\intercal$ denotes the nonlinear unicycle dynamics, $\eta_t \sim \mathcal{N}(\eta_t|0, \Sigma_{\eta_t})$ corresponds to the process noise, and dt denotes the step size for discretization. We consider the controller parameterization of the form (3) where $\mathcal{B}(X_t) = [x_t\ y_t\ \theta_t\ 1]^\intercal$.

The goal of the SOC problem is for the unicycle to navigate to a desired position without collision with obstacles. Let $\mathcal{A}$ be the set of obstacles. For $j \in \mathcal{A}$, we define

$$\mathcal{K}_j(\tau_t) = [(x_t - x_{obs,j})^2 + (y_t - y_{obs,j})^2 - (r_{obs,j} + r_s)^2],$$

where $(x_{obs,j}, y_{obs,j})$ and $r_{obs,j}$ are the center and the radius of the jth obstacle, respectively, and r_s denotes its safety radius. Let the unsafe set for the robot be $\mathcal{C}_u = \{(x, y) \in \mathbb{R}^2 | \mathcal{K}_j(x, y) < 0,\ \forall\, j \in \mathcal{A}\}$. Then, the safe set for the collision avoidance constraint $\mathcal{C}_s = \mathbb{R}^2 \backslash \mathcal{C}_u$. In our simulations, we choose $\psi_j(\tau_t)$ in (5) as

$$\psi_j(\tau_t) = \begin{cases} 0, & \text{if } \tau_t \in \mathcal{C}_s \\ \gamma(1 - \tanh(\epsilon \mathcal{K}_j(\tau_t))), & \text{otherwise,} \end{cases} \tag{26}$$

where $\gamma, \epsilon \in \mathbb{R}_+$ are tunable parameters to vary the tightness of the constraint and smoothness of $\psi_j(\cdot)$, respectively. For simulations, we choose $\gamma = 1$ and $\epsilon = 1$ in the barrier function (26). Other simulation parameters are given in Table 1. Fig. 1 shows the variation of $p(\mathcal{O}_t^{\text{in},j} = 1|\tau_t)$ with respect to γ in (26). We see that as γ increases, the constraint becomes more conservative, resulting in a lower likelihood of constraint violation.

We investigate the effect of the choice of smoothing algorithm on the overall performance of the PIIC algorithm. We employ unscented smoothing (UPIIC), and *factor graph optimization* (FGPIIC)

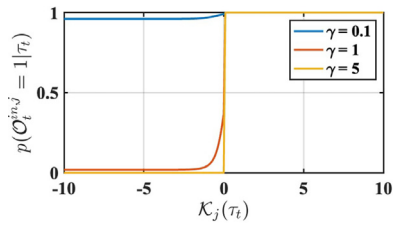

Fig. 1. Variation of $p(\mathcal{O}_t^{\text{in},j} = 1|\tau_t)$ for $Q_j^{in} = 1$ and different values of γ.

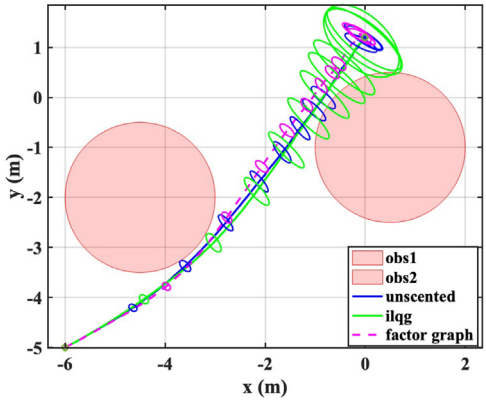

Fig. 2. Comparison of the trajectories with the feedback controllers inferred using ILQG, PIIC with unscented smoothing, and factor graph optimization.

Table 1
Simulation parameters for the unicycle example.

Simulation parameters	Value
Process noise covariance, Σ_{η_t}	$\text{diag}(10^{-3}, 10^{-3}, 10^{-3})$
Cost matrices, $\{Q_{t,\text{obs}}, Q_T, R_t\}$	$\{20,\ 10\ \mathbb{I}_3,\ 0.5\ \mathbb{I}_2\}$

in the E-step of Algorithm 1. The *unscented smoothing* is analogous to the unscented Kalman smoothing (Särkkä, 2008) except that we utilize it to compute the smoothed *state-control* distribution rather than just the state distribution. Factor graph optimization solves the smoothing problem as a nonlinear least squares problem. This is possible due to the fact that the maximum-a-posteriori (MAP) inference on a nonlinear factor graph with Gaussian noise models is equivalent to nonlinear least squares problem (Dellaert & Kaess, 2017). We interface with the GTSAM library (Dellaert & Contributors, 2022) to implement the factor graph generation and optimization. This approach is well known to be computationally efficient.

We compare the performance of the UPIIC, FGPIIC with the ILQG algorithm (Todorov & Li, 2005). The ILQG algorithm does not accommodate state constraints. Hence, we use the modified cost (7) with the barrier function candidate (26) to impose the obstacle avoidance constraint for a fair comparison. Fig. 2 shows the trajectories for 50 MC simulations with the corresponding covariance ellipses for $T = 200$ steps.

The mean and standard deviation of the incurred trajectory cost are shown in Table 2. We observe that the FGPIIC has the superior performance followed by UPIIC and ILQG. This can be attributed to the fact that each iteration of the FGPIIC performs multiple iterations of factor graph optimization until a level of convergence is reached whereas the UPIIC performs only one pass of the smoothing step per iteration, yielding in sub-optimal trajectories compared to FGPIIC. We also observe that the ILQG approach suffers from poor convergence, leading to higher variance in the trajectories and a greater number of constraint violations. We have repeated the same comparison for a target reaching problem without obstacles and the resulting trend was similar.

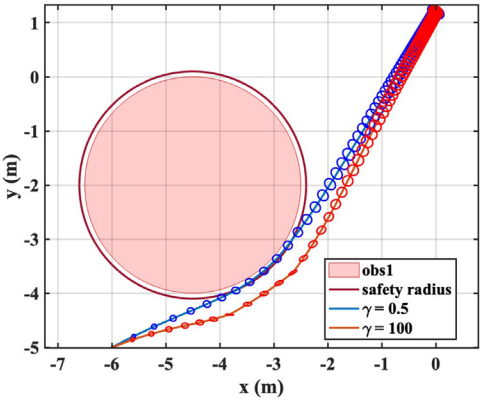

Fig. 3. Comparison of the trajectories of unicycle model with covariance ellipses for $\gamma = 0.5$ and $\gamma = 100$.

Table 2
Comparison of the average cost and standard deviation for 50 MC simulations with the feedback controllers inferred using ILQG, UPIIC, and FGPIIC for unicycle target reaching example with and without obstacles.

	Without obstacles	With obstacles
UPIIC	64.12 ± 4.26	92.26 ± 31.39
FGPIIC	**58.69 ± 3.18**	**61.89 ± 4.61**
ILQG	79.16 ± 19.02	189.02 ± 196.68

Table 3
Comparison of the number of constraint violations occurred in 50 MC simulations for various values of γ.

γ	1	2	5	≥10
# of constraint violations	38	56	11	0

We also perform an empirical study on the effect of γ in the barrier function (26) on the inferred controller and the trajectory of the unicycle robot. We restrict to a single obstacle to visualize a more pronounced effect. Fig. 3 shows the two trajectories resulting from controllers inferred using different values of γ. We observe that for higher values of γ, the minimum distance of the trajectory from the obstacle increases, i.e., the controller becomes more conservative. Due to the presence of process noise there is finite probability of constraint violation. However, for higher values of γ, the conservatism of the controller yields less constraint violations, i.e., the deviations from the inferred trajectory exist but remain in the safe set $\mathcal{C}_s$, resulting in satisfaction of the actual constraint with a higher probability. We corroborate the claim in Table 3, which shows that the number of constraint violations decreases as γ increases.

4.2. Formation control

We consider formation control of four unicycle robots modeled as (25). The objective is to find a stochastic optimal controller that navigates to desired goal positions with minimal control energy applied by each agent while closely maintaining a desired square formation and avoiding collision with obstacles. We define the individual cost of robot i as $\mathcal{C}^i_{x,al}(\tau^i_t) = (X^i_t - X^i_d)^\intercal Q^i_t (X^i_t - X^i_d) + \sum_{j=1}^{n_{\text{in}}} \psi_j^\intercal(\tau^i_t)\, Q_j^{i,\text{in}}\, \psi_j(\tau^i_t)$, $\mathcal{C}^i_{u,al}(\tau^i_t) = (u^i_t)^\intercal R^i_t (u^i_t)$, where at time t, X^i_t and X^i_d denote the state and the desired state of agent i, respectively, u^i_t denotes the control input of agent i, and $\psi_j(\cdot)$ is the barrier function as in (26). Let $X_t = [(X^1_t)^\intercal\ (X^2_t)^\intercal\ \cdots\ (X^N_t)^\intercal]^\intercal \in \mathbb{R}^{Nn_x}$ be the state of all the agents $i \in \mathcal{V}$ in the formation. We assume a linear controller for each agent of the form $\mathbb{E}(u^i_t) = K^i_t X_t + k^i_t$. Let $B \in \mathbb{R}^{N\times M}$ denote the incidence matrix of an undirected graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ corresponding to the formation, where M is the cardinality of the edge set $\mathcal{E}$. Define the formation cost as $\mathcal{C}_{nl}(\tau_t) = ((B \otimes \mathbb{I}_{n_x})^\intercal X_t - \delta_*)^\intercal Q_f ((B \otimes \mathbb{I}_{n_x})^\intercal X_t - \delta_*)$, where $\delta_* = [\delta_*^{1\intercal}\ \delta_*^{2\intercal}\ \cdots\ \delta_*^{M\intercal}]^\intercal \in \mathbb{R}^{Mn_x}$ represents the vector of formation targets along each edge $e \in \mathcal{E}$, and Q_f is a positive semi-definite block diagonal cost matrix. For the unit square formation in the simulation, $\mathcal{V} = \{1, 2, 3, 4\}$, $\mathcal{E} = \{(2, 1), (4, 1), (2, 3), (4, 3)\}$, and $\delta_* = [0\ 1\ 0\ \ -1\ 0\ 0\ 0\ 1\ 0\ 1\ 0\ 0]^\intercal$. The total trajectory cost for the optimal formation control problem is given by $\mathcal{C}(x_{0:T}, u_{0:T-1}) = \sum_{t=0}^{T} \mathcal{C}_{nl}(\tau_t) + \sum_{i=1}^{4} [\sum_{t=0}^{T} \mathcal{C}^i_{x,al}(\tau^i_t) + \sum_{t=0}^{T-1} \mathcal{C}^i_{u,al}(\tau^i_t)]$, where T is the time horizon set to 100 steps. Additional simulation parameters are given in Table 4.

Table 4
Simulation parameters for the formation control example.

Simulation parameters	Value
Process noise covariance, $\Sigma^i_{\eta^i_t}$	diag(10^{-3}, 10^{-3}, 10^{-4})
Linear velocity limits, v^i_t	[0, 8] m/s
Angular velocity limits, ω^i_t	[−1.5, 1.5] rad/s
$\{Q^i_t, Q^i_{t,\text{obs}}, Q^i_{t,\text{lim}}, Q^i_T, R^i_t, Q_{t,f}\}$	$\{\mathbb{I}_3, 50\ \mathbb{I}_4, 50\ \mathbb{I}_4, 50\ \mathbb{I}_3, \mathbb{I}_2, 50\ \mathbb{I}_{12}\}$

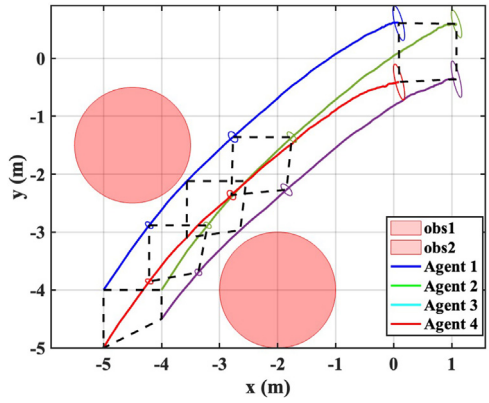

Fig. 4. Snapshots of X-Y trajectory of the unicycle formation with corresponding covariance ellipses.

We impose a *3-agent partially decentralized* structure on the controller wherein each agent has access to the state information of itself and two other agents in the formation. Fig. 4 shows the formation trajectory of the robots with covariance ellipses using Algorithm 1. We observe that the agents reach close to their target positions while avoiding the obstacles and respecting the square formation as closely as possible.

We next investigate this problem under three additional controller structures. A *centralized* structure is where each agent has access to the state information of all the agents in the formation, a *2-agent partially decentralized* structure is where each agent has access to the state information of itself and the agent diagonally opposite to it in the formation, and a *decentralized* structure is where each agent has access to only its own state information. Table 5 shows the average cost and standard deviation for 50 MC simulations. The centralized structure incurs the least average cost owing to its full information of the global state of the agents. It is followed by the 3-agent and 2-agent partially decentralized structures, respectively. The decentralized structure incurs the highest average cost. The increase of the cost is correlated to the decrease in the information available to each agent, yielding controllers with degrading performance.

5. Conclusions and future work

We present a parameterized inference-based approach to approximate constrained SOC. Our approach employs a barrier function to impose inequality constraints on the states and controls,

Table 5
Comparison of the average cost and standard deviation for 50 MC simulations with different controller structures for the 4-unicycle formation control example.

Controller structure	Average cost
Centralized	**664.67 ± 120.61**
Partially decentralized (3-agent)	**713.21 ± 116.82**
Partially decentralized (2-agent)	728.88 ± 128.41
Decentralized	749.22 ± 135.88

and creates controllers satisfying given structural constraints. We establish that our approach encompasses existing algorithms as special cases, such as the LQR and the I2C algorithm. The numerical simulations demonstrate that our approach outperforms the ILQG for constrained control and that using factor graph optimization incurs lower average cost than unscented smoothing. Our approach can also optimize control performance while satisfying structural constraints. Future work includes investigating structured control in a model-free setting for multi-agent systems.

Appendix. Equivalence of PIIC and I2C

We suppress the notation $q(\tau)$ under the expectation for brevity. Using the block matrix inversion identity in Petersen and Pedersen (2008), the inverse term in (14) yields

$$\mathbb{E}[\mathcal{B}_t(x_t)\mathcal{B}_t(x_t)^\intercal]^{-1} = \begin{bmatrix} \Sigma_{x_t}^{-1} & -\Sigma_{x_t}^{-1}\mu_{x_t} \\ -\mu_{x_t}^\intercal \Sigma_{x_t}^{-1} & 1 + \mu_{x_t}^\intercal \Sigma_{x_t}^{-1}\mu_{x_t} \end{bmatrix}, \tag{27}$$

where μ_{x_t} and Σ_{x_t} represent the mean and covariance of x_t in the smoothed state-control distribution, respectively. The second term in (14) can be expressed as

$$\mathbb{E}\left[\mathcal{B}_t(x_t)u_t^\intercal\right] = \begin{bmatrix} \Sigma_{x_t u_t} + \mu_{x_t}\mu_{u_t}^\intercal \\ \mu_{u_t}^\intercal \end{bmatrix}, \tag{28}$$

where $\Sigma_{x_t u_t} \in \mathbb{R}^{n_x \times n_u}$ is the cross-covariance between x_t and u_t, and μ_{u_t} is the mean of u_t in the smoothed state-control distribution. Substituting (27) and (28) in (14) yields

$$\Theta_t^{k+1} = \begin{bmatrix} \Sigma_{x_t}^{-1}\Sigma_{x_t u_t} \\ -\mu_{x_t}^\intercal \Sigma_{x_t}^{-1}\Sigma_{x_t u_t} + \mu_{u_t}^\intercal \end{bmatrix}. \tag{29}$$

Comparing (29) and $\Theta_t = \begin{bmatrix} K_t & k_t \end{bmatrix}^\intercal$ yields

$$\begin{aligned} K_t &= \Sigma_{x_t u_t}^\intercal \Sigma_{x_t}^{-T} = \Sigma_{x_t u_t}^\intercal \Sigma_{x_t}^{-1}, \\ k_t &= \mu_{u_t} - \Sigma_{x_t u_t}^\intercal \Sigma_{x_t}^{-T}\mu_{x_t} = \mu_{u_t} - K_t\mu_{x_t}. \end{aligned} \tag{30}$$

The covariance Σ_{δ_t} can be written as

$$\Sigma_{\delta_t} = \mathbb{E}[(u_t - K_t x_t - k_t)(u_t - K_t x_t - k_t)^\intercal]. \tag{31}$$

Substituting (30) in (31) and rearranging yields

$$\Sigma_{\delta_t} = \Sigma_{u_t} - \Sigma_{x_t u_t}^\intercal \Sigma_{x_t}^{-1}\Sigma_{x_t u_t}. \tag{32}$$

Note that (30) and (32) correspond to the parameter update equations for the conditional control distribution in Watson et al. (2021). Hence, for the given assumptions on $\mathcal{B}_t(x_t)$ and Θ_t, the PIIC and the Gaussian-I2C formulations are equivalent. Since Watson et al. (2020) guarantees the equivalence of I2C to LQR for linear deterministic dynamics with infinitely broad priors (i.e., $\Sigma_{\eta_t} \to 0$, $\Sigma_{\delta_t}^{-1} \to 0$), we omit details of the derivation and extend the claim to the PIIC.

References

Bertsekas, Dimitri P. (2016). *Nonlinear programming 3rd edition*. Athena Scientific.

Bishop, Christopher M. (2006). *Pattern recognition and machine learning*. Berlin, Heidelberg: Springer-Verlag.

Dellaert, Frank, & Contributors, GTSAM (2022). Borglab/gtsam.

Dellaert, Frank, & Kaess, Michael (2017). *vol. 6, Factor graphs for robot perception*. Foundations and Trends in Robotics.

Fardad, Makan, & Jovanović, Mihailo R. (2014). On the design of optimal structured and sparse feedback gains via sequential convex programming. In *American control conference* (pp. 2426–2431). IEEE.

Itoh, Hideaki, Sakai, Yoshitaka, Kadoya, Toru, Fukumoto, Hisao, Wakuya, Hiroshi, & Furukawa, Tatsuya (2017). Using model uncertainty for robust optimization in approximate inference control. *Artificial Life and Robotics*, *22*(3), 327–335.

Jovanović, Mihailo R., & Dhingra, Neil K. (2016). Controller architectures: Tradeoffs between performance and structure. *European Journal of Control*, *30*, 76–91.

Kavraki, Lydia E., & LaValle, Steven M. (2008). Motion planning. In *Springer handbook of robotics* (pp. 109–131). Berlin, Heidelberg: Springer Berlin Heidelberg.

Lin, Fu, Fardad, Makan, & Jovanović, Mihailo R. (2011). Augmented Lagrangian approach to design of structured optimal state feedback gains. *IEEE Transactions on Automatic Control*, *56*(12), 2923–2929.

Moon, Todd K. (1996). The expectation-maximization algorithm. *IEEE Signal Processing Magazine*, *13*(6), 47–60.

Petersen, Kaare Brandt, & Pedersen, Michael Syskind (2008). The matrix cookbook. *Technical University of Denmark*, *7*(15), 510.

Rawlik, Konrad, Toussaint, Marc, & Vijayakumar, Sethu (2010). An approximate inference approach to temporal optimization in optimal control. *Advances In Neural Information Processing Systems*, *23*.

Rawlik, Konrad, Toussaint, Marc, & Vijayakumar, Sethu (2013). On stochastic optimal control and reinforcement learning by approximate inference. In *Proceedings of the twenty-third international joint conference on artificial intelligence* (pp. 3052–3056). AAAI Press.

Rückert, Elmar A., & Neumann, Gerhard (2013). Stochastic optimal control methods for investigating the power of morphological computation. *Artificial Life*, *19*(1), 115–131.

Särkkä, Simo (2008). Unscented Rauch-Tung-Striebel smoother. *IEEE Transactions on Automatic Control*, *53*(3), 845–849.

Stengel, Robert F. (1994). *Optimal control and estimation*. Courier Corporation.

Syed, Shahbaz P. Qadri, & Bai, He (2023). Parameterized input inference for approximate stochastic optimal control. In *American control conference* (pp. 2574–2579).

Todorov, Emanuel (2006). Optimal control theory. *Bayesian Brain: Probabilistic Approaches to Neural Coding*, 268–298.

Todorov, Emanuel (2008). General duality between optimal control and estimation. In *47th IEEE conference on decision and control* (pp. 4286–4292). IEEE.

Todorov, Emanuel, & Li, Weiwei (2005). A generalized iterative LQG method for locally-optimal feedback control of constrained nonlinear stochastic systems. In *2005 American control conference* (pp. 300–306). IEEE.

Toussaint, Marc (2009). Robot trajectory optimization using approximate inference. In *Proceedings of the 26th annual international conference on machine learning* (pp. 1049–1056).

Watson, Joe, Abdulsamad, Hany, Findeisen, Rolf, & Peters, Jan (2021). Efficient stochastic optimal control through approximate Bayesian input inference. arXiv preprint, arXiv:2105.07693.

Watson, Joe, Abdulsamad, Hany, & Peters, Jan (2020). Stochastic optimal control as approximate input inference. In *Proceedings of conference on robot learning* (pp. 697–716). PMLR.

Shahbaz P Qadri Syed is currently pursuing his Ph.D. in the department of Mechanical and Aerospace Engineering at Oklahoma state university (USA). Prior to this, he received his M.S degree in Mechanical and Aerospace Engineering from the Oklahoma state university (USA) in 2022 and B.E degree in Mechanical Engineering from the Osmania University (India) in 2019. His research interests include stochastic optimal control, Bayesian networks, multi-agent learning and control.

He Bai received his B.S. degree from the Department of Automation at the University of Science and Technology of China, Hefei, China, in 2005, and the M.S. and Ph.D. degrees in Electrical Engineering from Rensselaer Polytechnic Institute in 2007 and 2009, respectively. From 2009 to 2010, he was a Post-doctoral Researcher at Northwestern University, Evanston, IL. From 2010 to 2015, he was a Senior Research and Development Scientist at UtopiaCompression Corporation. In 2015, he joined the School of Mechanical and Aerospace Engineering at Oklahoma State University, where he is currently an associate professor. His research interests include multi-agent learning and control, nonlinear estimation, robotics, and autonomous systems.