Safety Embedded Stochastic Optimal Control of Networked Multi-Agent Systems via Barrier States

Lin Song¹, Pan Zhao¹, Neng Wan¹, and Naira Hovakimyan¹

Abstract—This paper presents a novel approach for achieving safe stochastic optimal control in networked multi-agent systems (MASs). The proposed method incorporates barrier states (BaSs) into the system dynamics to embed safety constraints. To accomplish this, the networked MAS is factorized into multiple subsystems, and each one is augmented with BaSs for the central agent. The optimal control law is obtained by solving the joint Hamilton-Jacobi-Bellman (HJB) equation on the augmented subsystem, which guarantees safety via the boundedness of the BaSs. The BaS-based optimal control technique yields safe control actions while maintaining optimality. The safe optimal control solution is approximated using path integrals. To validate the effectiveness of the proposed approach, numerical simulations are conducted on a cooperative UAV team in two different scenarios.

I. INTRODUCTION

Optimal control has achieved remarkable success in both theory and applications [2], [3]. Obtaining optimal control usually requires solving a nonlinear, second-order partial differential equation (PDE), known as Hamilton-Jacobi-Bellman (HJB) equation. Stochastic optimal control (SOC) problems involve solving the control problem by minimizing expected costs [4]. By applying an exponential transformation to the value function [5], a linear-form HJB PDE is obtained, enabling related research including linearly-solvable optimal control (LSOC) [6] and path-integral control (PIC) [4], [7]. The benefits of LSOC problems include compositionality [8], [9] and the path-integral representation of the optimal control solution. However, solving SOC problems in large-scale systems is challenging due to the curse of dimensionality [10]. To overcome computational challenges, many approximation-based approaches have been developed, such as path-integral (PI) formulation [11], value function approximation [12], and policy approximation [13]. In [14], a PI approach is used to approximate optimal control actions on multi-agent systems (MASs), and the optimal path distribution is predicted using the graphical model inference approach. A distributed PIC algorithm is proposed in [15], in which a networked MAS is partitioned into multiple subsystems, and local optimal control actions are determined using local observations. However, these approaches seldom consider safety in the problem formulation, which may limit their real-world applications.

Safety refers to ensuring that a system’s states remain within appropriate regions at all times for deterministic systems, or with a high probability for stochastic systems. Reachability analysis is a formal verification approach used to prove safety and performance guarantees for dynamical systems [16], [17]. Hamilton-Jacobi (HJ) reachability analysis identifies the initial states that the system needs to avoid as well as the associated optimal control for the sake of remaining safe [18]. However, computing the reachable set in reachability analysis is typically expensive, making it challenging to apply to multi-agent and high-dimensional systems. To enable safe optimal control, safety metrics can be incorporated into the optimal control framework, either as objectives or constraints. In [19], temporal logic specifications are used as constraints for safety enforcement in optimal control development. The control barrier function (CBF) is a potent tool that can be used to enforce system safety by solving optimal control with constraints in a minimally invasive fashion [20], [21]. CBF-based methods have also been extended to stochastic systems with high-probability guarantees [22], [23]. A multi-agent CBF framework that generates collision-free controllers is discussed in [24]. Furthermore, guaranteed safety-constraint satisfaction in the network system is achieved in [25] under a valid assume-guarantee contract, with CBFs implemented onto subsystems. However, implementing CBFs as safety filters into the optimal control inputs may hinder ultimate optimality and be typically reactive to given constraints. Additionally, the feasibility of the quadratic programming (QP) introduced by CBF-based methods was not always guaranteed until the recent work in [26]. The barrier state (BaS) method is a novel methodology studied in [27], where the stability analysis of a BaS-augmented system encodes both stabilization and safety of the original system, and thus potential conflicts between control objectives and safety enforcement are avoided. In [28], discrete BaS (DBaS) is employed with differential dynamic programming (DDP) in trajectory optimization, and it has been shown that bounded DBaS implies the generation of safe trajectories. The DBaS have also been integrated into importance sampling to improve sample efficiency in safety-constrained sampling-based control problems in [29].

*This work is supported by Air Force Office of Scientific Research (AFOSR) (award #FA9550-21-1-0411) and National Aeronautics and Space Administration (NASA) (awards #80NSSC22M0070 and #80NSSC17M0051). In this version, we have omitted the theorem proof, several remarks on the results, and one simulation example due to space limitations. Interested readers can refer to [1] for the complete details.

¹Lin Song, Pan Zhao, Neng Wan, and Naira Hovakimyan are with the Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA {linsong2, panzhao2, nengwan2, nhovakim}@illinois.edu
in the solution boundedness, which prevents potential conflicts between control performance and safety requirements. However, the methodology of addressing safety issues without sacrificing optimality in networked MASs remains an open problem. In this paper, we propose a safety-embedded SOC framework for networked MASs using BaSs proposed in [27]. We adopt the MAS framework considered in [30], [31], where each agent computes optimal control based on local observations. However, [31] does not consider system safety, while [30] formulates safety concerns in the CBF framework and is potentially subject to the aforementioned issues. To address the safety-guarantee deficiency in optimal controls, we augment the dynamics of the central agent in each subsystem with BaSs that embed safety constraints and formulate the optimal control problem using the augmented dynamics. Bounded solutions to the revised optimal control problem then ensure safety due to the characteristics of BaSs.

The rest of the paper is structured as follows: Section II introduces the preliminaries of formulating SOC problems and constructing BaS; Section III presents the safety-embedded SOC framework on MASs, along with the path integral formulation to approximate the solution control; and Section IV provides numerical simulations in two scenarios to validate the effectiveness of the proposed approach. Finally, Section V concludes the paper and discusses future research directions. Several notations used in this paper are defined as follows: \( |S| \) denotes the cardinality of set \( S \), \( \det(X) \) denotes the determinant of matrix \( X \), \( \text{tr}(X) \) denotes the trace of matrix \( X \), \( \nabla_v V \) and \( \nabla^2_v V \) refer to the gradient and Hessian matrix of scalar-valued function \( V \), and \( \|v\|^2_M \coloneqq v^T M v \) denotes the weighted square norm. Due to space limitations, we have omitted several remarks and one numerical example. Interested readers can refer to [1] for more details.

II. PRELIMINARIES AND PROBLEM FORMULATION

A. Stochastic optimal control problems

1) MASs and factorial subsystems: We consider a MAS with \( N \) homogeneous agents indexed by \( \{1, 2, \ldots, N\} \). To describe the networked MAS, we use a connected and undirected graph \( \mathcal{G} = (V, E) \), where vertex \( v_i \in V \) represents agent \( i \), and undirected edge \( (v_i, v_j) \in E \) indicates that agent \( i \) and \( j \) can communicate with each other. We define the index set of all agents neighboring agent \( i \) as \( \mathcal{N}_i \), and factorize the networked MAS into multiple subsystems \( \mathcal{N}_i = \mathcal{N}_i \cup \{i\} \), where each factorial subsystem consists of a central agent and all its neighboring agents. Figure 1 provides an illustrative example of the factorization scheme, where \( x_i \) and \( \bar{u}_i \) denote the joint states and joint control actions of factorial subsystem \( N_i \). The local control action \( u_j \) is determined by minimizing a joint cost function on subsystem \( N_j \), which depends on the local observation \( \bar{x}_j \). Computing optimal control actions and sampling are therefore related to the size of each subsystem, rather than the entire network, which reduces computational complexity. For more discussions on the distributed control for LSOC problems on MASs, interested readers can refer to [15].

2) Stochastic optimal control of MASs: We use the Itô diffusion process to describe the joint dynamics of subsystem \( \mathcal{N}_i \) in a networked MAS consisting of \( N \) homogeneous agents governed by mutually independent passive dynamics. The process is represented by the following equation:

\[
\begin{align*}
\dot{x}_i &= \bar{g}_i(x_i(t), t)dt + B_i(x_i)[\bar{u}_i(x_i(t), t)dt + \sigma_i d\omega_i],
\end{align*}
\]

where \( \bar{g}_i(x_i(t), t) \in \mathbb{R}^{d_i} \) is the joint state vector and \( M \) represents the state dimension of each individual agent, \( \bar{g}_i(x_i(t), t) = [g_i(x_i(t), t)^T, g_j \in \mathcal{N}_i(x_j(t), t)^T] \in \mathbb{R}^{d_i} \) is the joint control matrix, \( \bar{u}_i(x_i, t) = [u_i(x_i, t), u_j \in \mathcal{N}_i(x_j, t) \in \mathbb{R}^{d_i}] \) is the joint control vector, and \( \omega_i = [\omega_i^T, \omega_j \in \mathcal{N}_i] \in \mathbb{R}^{d_i} \) is the joint noise vector with covariance matrix \( \bar{\sigma}_i = \text{diag}(\sigma_i, \sigma_j \in \mathcal{N}_i) \in \mathbb{R}^{d_i} \). To ensure the uniqueness of the solution, we assume that \( \bar{g}_i, \bar{B}_i, \bar{\sigma}_i \) are locally Lipschitz continuous.

We use \( \bar{B}_i \) to denote the set of joint terminal states and \( \bar{\mathcal{N}}_i \) to denote the set of joint non-terminal states. The entire allowable joint state space \( \bar{S}_i \) is partitioned into \( \bar{\mathcal{N}}_i \) and \( \bar{B}_i \). For \( \bar{x}_i \in \bar{\mathcal{N}}_i \), we define the running cost function as \( c_i(\bar{x}_i, \bar{u}_i) = q_i(\bar{x}_i) + \frac{1}{2} \bar{u}_i(x_i, t)^T R_i \bar{u}_i(x_i, t), \) where \( q_i(\bar{x}_i) \in \mathbb{R}_{\geq 0} \) is a joint state cost, and \( \bar{u}_i(x_i, t)^T R_i \bar{u}_i(x_i, t) \) is a control-quadratic cost with positive definite matrix \( R_i \in \mathbb{R}^{d_i \times d_i} \). When \( \bar{x}_i \in \bar{B}_i \), the terminal cost function is denoted by \( \phi_i(\bar{x}_i(t), t) \), for \( t_f \) is the final time. We also have the terminal cost \( \phi_i(\bar{x}_i(t), t) \) defined for \( \bar{x}_i(t) \in \bar{B}_i \). In the first exit formulation, \( t_f \) is determined online as the first time a joint state \( \bar{x} \in \bar{B}_i \) is reached. The cost-to-go function \( J^{\bar{u}_i}(\bar{x}_i(t), t) \) under joint control \( \bar{u}_i \) is defined as

\[
J^{\bar{u}_i}(\bar{x}_i(t), t) = \mathbb{E}_{\bar{x}_i(t)}\left[\phi_i(\bar{x}_i(t, t) + \int_{t_f}^{t_f} c_i(\bar{x}_i(\tau), \bar{u}_i(\tau))d\tau)\right],
\]

where the expectation is taken with respect to the probability measure under which \( \bar{x}_i \) satisfies (1) under given joint control \( \bar{u}_i \) starting from the initial condition \( \bar{x}_i(t) \). The optimal cost-to-go function (or value function) is formulated as \( V_i(\bar{x}_i(t), t) = \min_{\bar{u}_i} J^{\bar{u}_i}(\bar{x}_i(t), t) \), which is the minimum expected cumulative running cost starting from joint state \( \bar{x}_i \). For the sake of brevity, we use the notation \( \bar{x}_i \) to represent \( x_i(t) \) and \( \bar{x}_i(t) \) in the following context.

Facilitated by the exponential transformation of the value function, the optimal control action for the stochastic system (1) can be expressed in a linear form. The linear-form optimal control solution was initially proposed for a single-agent system in [32], and later extended to a multi-agent scenario in [31]. Here, we summarize the main results. The desirability function \( Z(\bar{x}_i, t) = \exp[-V_i(\bar{x}_i(t), t)/\lambda_i] \) is
defined over the joint state $\bar{x}_i$, and the nonlinearity cancellation condition $\bar{\sigma}_i \bar{\sigma}^\top_i = \lambda_i \bar{R}^{-1}_i$ is imposed to eliminate the nonlinear terms. Then, the linear-form joint optimal control action for the factorial subsystem $\bar{N}_i$ in the networked MAS under the discussed decentralization topology takes the form of

$$u^*_i(\bar{x}_i, t) = \bar{\sigma}_i \bar{B}_i(\bar{x}_i)^\top \nabla_{\bar{x}_i}Z(\bar{x}_i, t) / Z(\bar{x}_i, t).$$  (3)

The nonlinearity cancellation condition essentially indicates that control is costly when the noise variance in a control channel is low, and therefore large control effort is avoided.

B. Safety-embedded control via BaSs

The recently proposed safety-embedded control methodology through BaSs is a novel approach that enforces CBF constraints by augmenting the original system with BaSs [27], [28]. This methodology has demonstrated that the boundedness of the BaSs implies the safety of the original system. In this context, we provide a brief summary of the construction of BaSs for nonlinear control-affine systems, which our proposed optimal control framework on the augmented MASs embedding safety builds upon. We consider nonlinear control-affine dynamical systems modeled as

$$\dot{x} = g(x) + b(x)u,$$  (4)

where $x \in \mathcal{D} \subseteq \mathbb{R}^M$, $u \in \mathbb{R}^P$, $g : \mathbb{R}^M \to \mathbb{R}^M$, and $b : \mathbb{R}^M \to \mathbb{R}^{M \times P}$ are continuously differentiable functions. The set $\mathcal{D}$ is the domain of operation, and we denote $h : \mathcal{D} \to \mathbb{R}$ as a continuously differentiable function representing the safe set, e.g., the safe operating region $\mathcal{C} := \{x \in \mathcal{D} \mid h(x) > 0\}$. A critical property of the scalar-valued barrier function (BF) is that its value remains bounded except when $x$ approaches the boundaries of the safe operating region $\mathcal{C}$. We consider the composite barrier function (BF) in the form of $\beta(x) = B \circ h(x)$, where $h(x)$ defines a safe set. The dynamics of the BaSs $z$ are modeled as

$$\dot{z} = \phi_0(z + \beta_0 h(x)) \doteqdot \phi_0(z + \phi_1(z + \phi_0 h(x))),$$  (5)

where $\beta_0 = (\beta(0), \dot{h}(x) = L h(x) + L_\gamma h(x) u, \gamma \in \mathbb{R}_{> 0}$, $\phi_0(\cdot), \phi_1(\cdot)$ are analytic functions formulated based on the choice of BF $B(\cdot), h(x)$, and are subject to certain conditions proposed in [27]. We introduce a lemma from [27] that connects the boundedness of the BaSs with the safety of the generated trajectories.

**Lemma 1** ([27]). Suppose that $z(0) = \beta(x(0)) - \beta(0)$ and $\beta(x(0)) < \infty$, the BaSs $z(t)$ generated from (5) with the nonlinear dynamics (4) is bounded if and only if $\beta(x(t))$ is bounded for all $t > 0$.

**Remark 1.** From Lemma 1, we can observe that ensuring the boundedness of the BaSs implies boundedness of $\beta(x(t))$. This implies that $h(x(t)) > 0$ due to the properties of the BF; ensuring that the system trajectories always remain in the safe operating region $\mathcal{C} = \{x \in \mathcal{D} \mid h(x) > 0\}$.

III. BARRIER STATE AUGMENTED STOCHASTIC OPTIMAL CONTROL OF MASs

A. Safe optimal control of MASs via BaSs

Using the system factorization methodology introduced in Section II-A.1, we can now formulate a safety-embedded optimal control on MASs using BaSs. Based on the distributed control framework we build upon, we conjecture that appending only the BaSs corresponding to the central agent for each subsystem is sufficient since only agent $i$ samples the local optimal control action from the computation results on subsystem $\bar{N}_i$. By determining a bounded solution to the optimal control problem on each augmented subsystem, we can certify the safety of the included central agent. Taking certified-safe control actions for each agent from its corresponding subsystem (in which it acts as the central agent) collectively establishes the safety of the entire network.

We first consider the continuous-time dynamics of agent $i$ within $\bar{N}_i$, which can be described by the following Itô process

$$dx_i = g_i(x_i, t)dt + B_i(x_i)[u_i(\bar{x}_i, t)dt + \sigma_i d\omega_i],$$  (6)

where $x_i \in \mathbb{R}^M$ is the state vector, $g_i(x_i, t) \in \mathbb{R}^M$ is the passive dynamics vector, $B_i(x_i) \in \mathbb{R}^{M \times P}$ is the control matrix, $u_i(\bar{x}_i, t) \in \mathbb{R}^P$ is the control action vector, and $\omega_i \in \mathbb{R}^P$ is the noise vector with covariance matrix $\sigma_i \in \mathbb{R}^{P \times P}$. The construction of BaS introduced in Section II-B is based on general nonlinear control-affine dynamical systems. Here, we specifically consider the dynamics in form of (6) and construct the corresponding BaSs.

We use $N_s$ independent BaSs to describe all $N_s$ safety constraints of interest for each agent. For agent $i$, we denote the function modeling the $j$-th constraint as $h_j(x_i)$ and the corresponding BaS $\text{BaS} = \text{BaS}_{h_j(x_i)}$ associated with it. Suppose we use the inverse BF, i.e., $\beta(x_i) = 1/h(x_i)$, and we choose $\phi_0(\xi) = -\xi^2$ and $\phi_1(\xi, \eta) = \eta \xi^2 - \xi$ as in [27]. Then, the dynamics for each single BaS are reduced to:

$$d\zeta_{i(j)} = \left[-(\zeta_{i(j)} + \beta_0(\zeta_{i(j)}))^2 \frac{\partial h_j}{\partial x_i} g_i(x_i, t) + \gamma h_j(x_i) + \gamma (\zeta_{i(j)} + \beta_0(\zeta_{i(j)}))\right]dt$$

$$+ (\zeta_{i(j)} + \beta_0(\zeta_{i(j)}))^2 \frac{\partial h_j}{\partial x_i} B_i(x_i)[u_i(\bar{x}_i, t)dt + \sigma_i d\omega_i],$$

with $\beta_0(\zeta) = 1/h(\zeta)$. $\zeta_{i(j)} \in \mathbb{R},$ $g_{i(j)}(x_i, z_{i(j)}), t) \in \mathbb{R}, B_{i(j)}(x_i, z_{i(j)}) \in \mathbb{R}^{N_s \times P}$, and $u_{i(j)}(\bar{x}_i, t) \in \mathbb{R}^P$. Since each agent is subject to $N_s$ constraints, which are modeled by the BaS dynamics in (7), we can represent the reorganized BaS dynamics for agent $i$ as:

$$dx_i = g_d(x_i, z_{i(j)}, t)dt + B_{i(d)}(x_i)[u_i(\bar{x}_i, t)dt + \sigma_i d\omega_i],$$  (8)

where $z_i = [z_i(1), \ldots, z_i(N_s)]^\top \in \mathbb{R}^{N_s},$ $g_d(x_i, z_{i(j)}) = [g_{i(1)}(x_i, z_{i(1)}), \ldots,$ $g_{i(N_s)}(x_i, z_{i(N_s)})]^\top \in \mathbb{R}^{N_s},$ $B_{i(d)}(x_i, z_{i(j)}) = [B_{i(1)}(x_i, z_{i(1)})]^\top, \ldots,$ $B_{i(N_s)}(x_i, z_{i(N_s)})]^\top \in \mathbb{R}^{N_s \times P}$, and the individual elements $\beta_0(\zeta), g_{i(j)}(x_i, z_{i(j)}),$ and $B_{i(j)}(x_i, z_{i(j)})$ are defined by (7). Next, we integrate the BaS dynamics (8) corresponding to all $N_s$ constraints that the central agent $i$ is subject to. As a result, the augmented joint dynamics take the form of

$$d\bar{Y}_i = \bar{g}_i^\top(\bar{Y}_i, t)dt + B_i^\top(\bar{Y}_i)[u_i(\bar{Y}_i, t)dt + \bar{\sigma}_d d\bar{\omega}_i],$$  (9)

where $\bar{Y}_i = \left[\bar{x}_i^\top, z_{i(1)}^\top, \ldots, z_{i(N_s)}^\top\right]^\top \in \mathbb{R}^{M + N_s}$ is the augmented joint state vector, $\bar{g}_i(\bar{Y}_i, t) = \left[g_i(x_i, t)^\top, g_{i(1)}(x_i, z_{i(1)})^\top, \ldots,$ $g_{i(N_s)}(x_i, z_{i(N_s)})^\top\right]^\top \in \mathbb{R}^{M + N_s}$ is the
amplified joint passive dynamics vector, \( \bar{B}^T_i(\bar{Y}_i) = \text{diag}\{B^T_i(Y_i), B^T_j \in N_i(x_j)\} \in \mathbb{R}^{(M \cdot |\bar{N}_i| + N_S) \times (P \cdot |\bar{N}_i|)} \) with \( B^T_i(Y) = [B_{ai}(x_i), B_{bi}(x_i, z_i)]^T \in \mathbb{R}^{(M \cdot |\bar{N}_i| + N_S) \times P} \) is the augmented joint control matrix, \( \bar{u}_i(\bar{Y}_i, t) = [u_i(\bar{x}_i, t)^T, u_j \in N_i(\bar{x}_j, t)]^T \in \mathbb{R}^{P \cdot |\bar{N}_i|} \) is the joint control action for the augmented system. \( \bar{\omega}_i = [\omega^T_i, \omega^T_j \in \mathbb{R}^{N_i}] \in \mathbb{R}^{P \cdot |\bar{N}_i|} \) is the joint noise vector, and \( \bar{\sigma}_i = \text{diag}\{\sigma_i, \sigma_j \in N_i\} \in \mathbb{R}^{(P \cdot |\bar{N}_i|) \times (P \cdot |\bar{N}_i|)} \) is the covariance matrix of noise vector \( \bar{\omega}_i \). In particular, the BaS vector for agent \( i \) is denoted by \( z_i = [z_i(1), z_i(2), \ldots , z_i(N_i)]^T \in \mathbb{R}^{N_i} \), where each \( z_i(j) \) is subject to the BaS dynamics in (7). Since only some states are directly controlled, we can partition the augmented joint state \( \bar{Y}_i \) into directly actuated states \( \bar{Y}_{d}(t) \in \mathbb{R}^{D \cdot |\bar{N}_i| + N_{SD}} \) and non-directly actuated states \( \bar{Y}_{n}(t) \in \mathbb{R}^{U \cdot |\bar{N}_i| + N_{SU}} \). Then, we can express \( \bar{Y}_i = [\bar{Y}_{d}(t), \bar{Y}_{n}(t)]^T \), where \( U \) and \( D \) denote the dimensions of non-directly and directly actuated states for one agent, and \( N_{SU}, N_{SD} \) denote the dimensions of non-directly and directly actuated BaSs for one agent. Using this notation, we can rewrite the augmented joint dynamics in (9) in the following partitioned vector form

\[
\begin{bmatrix}
\bar{d}Y_{d}(t) \\
\bar{d}Y_{n}(t)
\end{bmatrix} =
\begin{bmatrix}
\bar{g}_{d}(\bar{Y}_i, t) \\
\bar{g}_{n}(\bar{Y}_i, t)
\end{bmatrix} dt + \begin{bmatrix}
\bar{p}_{d}(\bar{Y}_i) \\
\bar{p}_{n}(\bar{Y}_i)
\end{bmatrix} d\bar{\omega}_i + \bar{\sigma}_i d\bar{\omega}_i,
\]

where \( 0 \) denotes a zero matrix with the appropriate dimensions. We define the joint running cost function involving the augmented joint state \( \bar{Y}_i \) of the network \( \bar{N}_i \) as

\[
c_i(\bar{Y}_i, \bar{u}_i) = q_i(\bar{Y}_i) + \frac{1}{2} \bar{u}_i(\bar{Y}_i, t)^T \bar{R}_i \bar{u}_i(\bar{Y}_i, t),
\]

where \( \bar{R}_i \in \mathbb{R}^{P \cdot |\bar{N}_i| \times P \cdot |\bar{N}_i|} \) is positive definite. Here, we assume that the control weights of each agent are decoupled, i.e., \( \bar{R}_i = \text{diag}\{R_i, R_j \in \mathbb{R}^{N_i}\} \), and \( \frac{1}{2} \bar{u}_i(\bar{Y}_i, t)^T \bar{R}_i \bar{u}_i = \sum_{j \in \bar{N}_i} \frac{1}{2} u_i R_i u_i \). The terminal cost function for the augmented joint state \( \bar{Y}_i \) is defined as \( \bar{\phi}_i(\bar{Y}_i) = \sum_{j \in \bar{N}_i} \omega_i^T \phi_j(\bar{Y}_i) + \omega_i^T \phi_j(\bar{Y}_i) \), with \( \omega^T_i, \omega^T_j > 0 \) denoting weights reflecting the importance of each agent. The joint cost-to-go function, subject to the control \( \bar{u}_i \) in the first-exit formulation for the augmented subsystem \( \bar{N}_i \), is defined as

\[
J^*_{\bar{u}_i}(\bar{Y}^T_i, t) = \mathbb{E}_{\bar{Y}_i, t} \left[ \bar{\phi}_i(\bar{Y}^T_i) + \int_{t}^{T} c_i(\bar{Y}_i(\tau), \bar{u}_i(\tau)) d\tau \right],
\]

and the joint value function is defined as

\[
V_i(\bar{Y}_i, t) = \min_{\bar{u}_i} \mathbb{E}_{\bar{Y}_i, t} \left[ \bar{\phi}_i(\bar{Y}^T_i) + \int_{t}^{T} c_i(\bar{Y}_i(\tau), \bar{u}_i(\tau)) d\tau \right].
\]

Next, we introduce a theorem that summarizes the solution to the optimal control problem for MASs that guarantees safety, which is achieved by solving a linear-form HJB equation with BaS augmentation.

**Theorem 1.** Consider a MAS consisting of \( N \) homogeneous agents with joint dynamics given by (1). To incorporate safety, we augment the joint dynamics for subsystem \( \bar{N}_i \) using the central agent BaS, and the augmented joint dynamics are given by (9). The augmented MAS is subject to the joint immediate cost (11) and the joint value function (12). Then, the joint optimal control action \( \bar{u}^*_i \) of subsystem \( \bar{N}_i \) ensuring safety is given by

\[
\bar{u}^*_i(\bar{Y}_i, t) = -\bar{R}_i^{-1} \bar{B}^T_i(\bar{Y}_i) \nabla_{\bar{Y}_i} V_i(\bar{Y}_i, t).
\]

We define the desirability function \( Z_i(\bar{Y}_i, t) = \exp[-V_i(\bar{Y}_i, t)/\lambda_i] \), where \( \lambda_i \in \mathbb{R} \). Under the safe optimal control action (13) and the nonlinearity cancellation condition \( \bar{R}_i = (\bar{\sigma}_i^T / \lambda_i)^{-1} \), the joint stochastic HJB equation reduces to a linear form given by

\[
\begin{aligned}
\frac{\partial}{\partial t} Z_i(\bar{Y}_i, t) &= \left[ g_i(\bar{Y}_i) Z_i(\bar{Y}_i, t)/\lambda_i - \bar{g}_i(\bar{Y}_i, t)^T \nabla_{\bar{Y}_i} Z_i(\bar{Y}_i, t) \right] \\
&- \frac{1}{2} tr(\bar{B}^T_i(\bar{Y}_i) \bar{\sigma}_i \bar{\sigma}_i^T \bar{B}_i(\bar{Y}_i)^T \nabla_{\bar{Y}_i} Z_i(\bar{Y}_i, t)),
\end{aligned}
\]

with the boundary condition \( Z_i(\bar{Y}_i, t) = \exp[-\phi_i(\bar{Y}_i)/\lambda_i] \).

The above equation can be solved in closed form as

\[
Z_i(\bar{Y}_i, t) = \mathbb{E}_{\bar{Y}_i, t} \left[ \exp(-\phi_i(\bar{Y}_i)/\lambda_i - \int_{t}^{T} \bar{q}_i(\bar{Y}_i)/\lambda_i d\tau) \right].
\]

Here, the diffusion process \( \bar{p}_i(t) \) is subject to the uncontrolled dynamics \( d\bar{p}_i(t) = \bar{g}_i(\bar{p}_i(t), \tau) d\tau + \bar{B}_i(\bar{p}_i) d\omega_i \) with initial condition \( \bar{p}_i(t) = \bar{Y}_i(t) \).

**Remark 2.** The proof of Theorem 1 is inspired by the proof of Theorem 2 in [15], with a few modifications and observations. First, since the central agent states are augmented with the BaS while other agent states remain unchanged, we need to differentiate the state-dependent terms in the value function, between the central and non-central agents in each subsystem. Second, the optimal control actions now take different forms for central and non-central agents, and therefore the nonlinear term cancellation in the HJB equation needs to be considered separately, with associated optimal control actions. Finally, the safety property of the obtained optimal controls is established by the boundedness of the BaS, which is part of the augmented joint state. A feasible solution to the optimal control problem implies a bounded cost function, and thus bounded BaSs.

**Proof.** See the extended version [1].

### B. Path integral approximation formulation

Although Theorem 1 of Section III-A provides a closed-form solution of the optimal control action, computing the expectation term over all trajectories initiated at \( (\bar{Y}_i^T, t) \) in the solution of the desirability function (15) is usually intractable. To address this, we reformulate the optimal control solution for the augmented system using path integrals and propose a revised proposition based on Proposition 3 in [15]. However, due to space limitations, we have omitted it here. Interested readers can refer to [1] for the complete details.

## IV. SIMULATION RESULTS

In this section, we conduct numerical simulations on cooperative MASs consisting of UAVs in environments with obstacles. The objective is to reach targets, avoid obstacles, and cooperate with other agents. The continuous-time dynamics of each UAV are given by:

\[
\begin{bmatrix}
\frac{dx_i}{dt} \\
\frac{dy_i}{dt} \\
\frac{d\phi_i}{dt}
\end{bmatrix} = \begin{bmatrix}
\nu_i \cos \phi_i \\
\nu_i \sin \phi_i \\
0
\end{bmatrix} dt + \begin{bmatrix}
0 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix} \begin{bmatrix}
u_i \\
\sigma_i \\
\nu_i
\end{bmatrix} d\omega_i,
\]

(16)
where \((x_i, y_i), v_i, \phi_i\) represent the position coordinate, forward velocity, and heading angle of UAV \(i\). We use \(x_i := (x_i, y_i, v_i, \phi_i)^T\) to denote the state vector. The forward acceleration \(u_i\) and angular velocity \(w_i\) are the control inputs, and \(\omega_i\) is the standard Brownian motion disturbance. We set the noise level to \(\sigma_i = 0.1\) and \(\nu_i = 0.05\), and specify the exit time as \(t_f = 20\) seconds.

This behavior is essentially due to the fact that CBF-based methods are reactive to given safety constraints and require

where \(\beta_j(\cdot)\) is the inverse BF. The \(\gamma\)-parameter in (7) is set to 0.5, which determines the BaS’s rate of returning to \(\beta_j(x_i) - \beta_0\), as discussed in [27]. The running cost functions for the three subsystems are designed as follows:

\[
g(q(\bar{x}_i), z_j) = 2\|([x_{ij}, y_{ij}])_l - d_{ij}^{\max}\|_2 + 0.5\|([x_{ij}, y_{ij}])_l - (x_{2j}, y_{2j})\|_2 - d_{2j}^{\max} + 0.8 \leq ([x_{1j}, y_{1j}])_l - (x_{1j}, y_{1j})\|_2 - d_{1j}^{\max} + 0.8\|([x_{2j}, y_{2j}])_l - (x_{2j}, y_{2j})\|_2 - d_{2j}^{\max} + 0.8\|([x_{1j}, y_{1j}])_l - (x_{1j}, y_{1j})\|_2 - d_{1j}^{\max} + 0.8\|([x_{2j}, y_{2j}])_l - (x_{2j}, y_{2j})\|_2 - d_{2j}^{\max} + 0.8\|([x_{1j}, y_{1j}])_l - (x_{1j}, y_{1j})\|_2 - d_{1j}^{\max} + 0.8\|([x_{2j}, y_{2j}])_l - (x_{2j}, y_{2j})\|_2 - d_{2j}^{\max} + 0.8\|([x_{1j}, y_{1j}])_l - (x_{1j}, y_{1j})\|_2 - d_{1j}^{\max} + 0.8\|([x_{2j}, y_{2j}])_l - (x_{2j}, y_{2j})\|_2 - d_{2j}^{\max} + 0.8\|([x_{1j}, y_{1j}])_l - (x_{1j}, y_{1j})\|_2 - d_{1j}^{\max} + 0.8\|([x_{2j}, y_{2j}])_l - (x_{2j}, y_{2j})\|_2 - d_{2j}^{\max} + 0.8\|([x_{1j}, y_{1j}])_l - (x_{1j}, y_{1j})\|_2 - d_{1j}^{\max} + 0.8\|([x_{2j}, y_{2j}])_l - (x_{2j}, y_{2j})\|_2 - d_{2j}^{\max} + 0.8\|([x_{1j}, y_{1j}])_l - (x_{1j}, y_{1j})\|_2 - d_{1j}^{\max} + 0.8\|([x_{2j}, y_{2j}])_l - (x_{2j}, y_{2j})\|_2 - d_{2j}^{\max} + 0.8\|([x_{1j}, y_{1j}])_l - (x_{1j}, y_{1j})\|_2 - d_{1j}^{\max} + 0.8\|([x_{2j}, y_{2j}])_l - (x_{2j}, y_{2j})\|_2 - d_{2j}^{\max} + 0.8\|([x_{1j}, y_{1j}])_l - (x_{1j}, y_{1j})\|_2 - d_{1j}^{\max} + 0.8\|([x_{2j}, y_{2j}])_l - (x_{2j}, y_{2j})\|_2 - d_{2j}^{\max} + 0.8\|([x_{1j}, y_{1j}])_l - (x_{1j}, y_{1j})\|_2 - d_{1j}^{\max} + 0.8\|([x_{2j}, y_{2j}])_l - (x_{2j}, y_{2j})\|_2 - d_{2j}^{\max} + 0.8\|([x_{1j}, y_{1j}])_l - (x_{1j}, y_{1j})\|_2 - d_{1j}^{\max} + 0.8\|([x_{2j}, y_{2j}])_l - (x_{2j}, y_{2j})\|_2 - d_{2j}^{\max} + 0.8\|([x_{1j}, y_{1j}])_l - (x_{1j}, y_{1j})\|_2 - d_{1j}^{\max} + 0.8\|([x_{2j}, y_{2j}])_l - (x_{2j}, y_{2j})\|_2 - d_{2j}^{\max} + 0.8\|([x_{1j}, y_{1j}])_l - (x_{1j}, y_{1j})\|_2 - d_{1j}^{\max} + 0.8\|([x_{2j}, y_{2j}])_l - (x_{2j}, y_{2j})\|_2 - d_{2j}^{\max} + 0.8\|([x_{1j}, y_{1j}])_l - (x_{1j}, y_{1j})\|_2 - d_{1j}^{\max} + 0.8\|([x_{2j}, y_{2j}])_l - (x_{2j}, y_{2j})\|_2 - d_{2j}^{\max} + 0.8\|([x_{1j}, y_{1j}])_l - (x_{1j}, y_{1j})\|_2 - d_{1j}^{\max} + 0.8\|([x_{2j}, y_{2j}])_l - (x_{2j}, y_{2j})\|_2 - d_{2j}^{\max} + 0.8\|([x_{1j}, y_{1j}])_l - (x_{1j}, y_{1j})\|_2 - d_{1j}^{\max} + 0.8\|([x_{2j}, y_{2j}])_l - (x_{2j}, y_{2j})\|_2 - d_{2j}^{\max} + 0.8\|([x_{1j}, y_{1j}])_l - (x_{1j}, y_{1j})\|_2 - d_{1j}^{\max} + 0.8\|([x_{2j}, y_{2j}])_l - (x_{2j}, y_{2j})\|_2 - d_{2j}^{\max} + 0.8\|([x_{1j}, y_{1j}])_l - (x_{1j}, y_{1j})\|_2 - d_{1j}^{\max} + 0.8\|([x_{2j}, y_{2j}])_l - (x_{2j}, y_{2j})\|_2 - d_{2j}^{\max} + 0.8\|([x_{1j}, y_{1j}])_l - (x_{1j}, y_{1j})\|_2 - d_{1j}^{\max} + 0.8\|([x_{2j}, y_{2j}])_l - (x_{2j}, y_{2j})\|_2 - d_{2j}^{\max}
\]

The hyper-parameter \(d_{ij}^{\max}\) represents the maximum difference between the initial position of UAV \(i\) and its target. Similarly, \(d_{ij}^{\min}\) represents the initial distance between UAVs \(i\) and \(j\). The simulation is performed using a step size of \(\Delta t = 0.2\) s.

We employ four different approaches to compute optimal control solutions for the goal-reaching, agent-coordinating, and obstacle-avoidance task. In the conventional optimal control framework, any violation of safety constraints is penalized in the cost function. The CBF-based safe optimal controller, discussed in [23], [30], first computes a baseline optimal control and then uses CBF as a safety filter to modify the baseline control input if necessary to ensure that the safety constraints are satisfied. Here, we categorize the CBF-based safe optimal controller based on whether the obstacle-collision penalty is included in the baseline control cost function design. We compare these three methods with the BaS-based safe optimal control method proposed in this paper, which solves optimal control action on the joint dynamics with augmentation of the central agent BaSs. We conduct five simulations with each method under identical initial conditions, and the results are compared in Figure 3.

From Figure 3, it is observed that the original optimal control solution obtained by solely penalizing constraint violation is insufficient to avoid obstacles in a cluttered environment. Additionally, the CBF-based method, which involves filtering the baseline optimal control by solving a quadratic programming (QP) problem, got trapped in obstacles and failed to reach the target in some scenarios. This behavior is essentially due to the fact that CBF-based methods are reactive to given safety constraints and require
good tuning to ensure feasibility and task completion. In contrast, the proposed safe optimal control method using BaS augmentation ensures safety while successfully reaching the target and achieving desired coordination in all trails.

V. CONCLUSION AND FUTURE WORK

In this paper, we introduce a novel safety-embedded stochastic optimal control framework for networked multi-agent systems (MASs) using barrier states (BaSs) augmentation. Our approach involves augmenting the joint dynamics of each factorial subsystem by introducing BaSs that embed safety constraints for the central agent and solving the optimal control action on the augmented subsystem. The proposed method simultaneously achieves safety and other control objectives, where safety is guaranteed by the boundedness of BaSs and achieved by the feasible solutions to the reformulated optimal control problem. The safe optimal control law is obtained by solving a linear stochastic Hamilton-Jacobi-Bellman (HJB) equation, and the solution is approximated using the path-integral formulation. We validate our approach through numerical simulations on a networked team of UAV. Future work includes representing the optimal control solution in policy improvement with path integrals (PI) framework and exploring other approximation formulations, such as Relative Entropy Policy Search (REPS). Moreover, we aim to extend our safety-embedded stochastic optimal control framework to MASs with incomplete state information.

VI. ACKNOWLEDGMENT

The authors would like to appreciate constructive discussions with Hassan Almubarak.

REFERENCES

[1] L. Song, P. Zhao, N. Wan, and N. Hovakimyan, “Safety embedded stochastic optimal control of networked multi-agent systems via barrier states,” arXiv preprint arXiv:2210.03855, 2022.

[2] D. Liberzon, Calculus of Variations and Optimal Control Theory—A Concise Introduction. NJ, USA: Princeton university press, 2011.

[3] A. E. Bryson and Y.-C. Ho, Applied Optimal Control: Optimization, Estimation, and Control. Routledge, 2018.

[4] H. J. Kappen, “Linear theory for control of nonlinear stochastic systems,” Physical review letters, vol. 95, no. 20, p. 200201, 2005.

[5] E. Todorov, “Efficient computation of optimal actions,” Proceedings of the national academy of sciences, vol. 106, no. 28, pp. 11 478–11 483, 2009.

[6] K. Djidjoham and E. Todorov, “Linearly solvable optimal control,” in Reinforcement Learning and Approximate Dynamic Programming for Feedback Control. CiteSeer, 2012, vol. 17, pp. 119–141.

[7] Y. Pan, E. Theodorou, and M. Kontitsis, “Sample efficient path integral control under uncertainty,” Advances in Neural Information Processing Systems, vol. 28, 2015.

[8] E. Todorov, “Compositionality of optimal control laws,” Advances in Neural Information Processing Systems, vol. 22, 2009.

[9] L. Song, N. Wan, A. Gahlawat, N. Hovakimyan, and E. A. Theodorou, “Compositionality of linearly solvable optimal control in networked multi-agent systems,” in 2021 American Control Conference (ACC). IEEE, 2021, pp. 1334–1339.

[10] Y. D. Blondel and J. N. Tsitsiklis, “A survey of computational complexity results in systems and control,” Automatica, vol. 36, no. 9, pp. 1249–1274, 2000.

[11] E. Theodorou, J. Buchli, and S. Schaal, “A generalized path integral control approach to reinforcement learning,” The Journal of Machine Learning Research, vol. 11, pp. 3137–3181, 2010.

[12] W. B. Powell and J. Ma, “A review of stochastic algorithms with continuous value function approximation and some new approximate policy iteration algorithms for multidimensional continuous applications,” Journal of Control Theory and Applications, vol. 9, no. 3, pp. 336–352, 2011.

[13] R. S. Sutton, D. McAllester, S. Singh, and Y. Mansour, “Policy gradient methods for reinforcement learning with function approximation,” Advances in Neural Information Processing Systems, vol. 12, 1999.

[14] B. Van Den Broek, W. Wiererinck, and B. Kappen, “Graphical model inference in optimal control of stochastic multi-agent systems,” Journal of Artificial Intelligence Research, vol. 32, pp. 95–122, 2008.

[15] N. Wan, A. Gahlawat, N. Hovakimyan, E. A. Theodorou, and P. G. Voulgaris, “Allocated algorithms for linearly-solvable optimal control in networked multi-agent systems,” arXiv preprint arXiv:2102.09104, 2021.

[16] S. Summers and J. Lygeros, “Verification of discrete time stochastic hybrid systems: A stochastic reach-avoid decision problem,” Automatica, vol. 46, no. 12, pp. 1951–1961, 2010.

[17] M. P. Chapman, J. Lacotte, A. Tamar, D. Lee, K. M. Smith, V. Cheng, J. F. Fisac, S. Jha, M. Pavone, and C. J. Tomlin, “A risk-sensitive finite-time reachability approach for safety of stochastic dynamic systems,” in 2019 American Control Conference (ACC). IEEE, 2019, pp. 2958–2963.

[18] S. Bansal, M. Chen, S. Herbert, and C. J. Tomlin, “Hamilton-jacobi reachability: A brief overview and recent advances,” in 2017 IEEE 56th Annual Conference on Decision and Control (CDC). IEEE, 2017, pp. 2242–2253.

[19] M. B. Horowitz, E. M. Wolff, and R. M. Murray, “A compositional approach to stochastic optimal control with co-safe temporal logic specifications,” in 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2014, pp. 1466–1473.

[20] A. D. Ames, X. Xu, J. W. Grizzle, and P. Tabuada, “Control barrier function based quadratic programs for safety critical systems,” IEEE Transactions on Automatic Control, vol. 62, no. 8, pp. 3861–3876, 2016.

[21] C. Tao, H.-J. Yoon, H. Kim, N. Hovakimyan, and P. Voulgaris, “Path integral methods with stochastic control barrier functions,” in 2022 IEEE 61st Conference on Decision and Control (CDC), 2022, pp. 1654–1659.

[22] A. Clark, “Control barrier functions for stochastic systems,” Automatica, vol. 130, p. 109688, 2021.

[23] ——, “Control barrier functions for complete and incomplete information stochastic systems,” in 2019 American Control Conference (ACC). IEEE, 2019, pp. 2928–2935.

[24] L. Wang, A. D. Ames, and M. Egerstedt, “Safety barrier certificates for collisions-free multirobot systems,” IEEE Transactions on Robotics, vol. 33, no. 3, pp. 661–674, 2017.

[25] X. Chen, J. Anderson, K. Kalsi, A. D. Ames, and S. H. Low, “Safety-critical control synthesis for network systems with control barrier functions and assume-guarantee contracts,” IEEE Transactions on Control of Network Systems, vol. 8, no. 1, pp. 487–499, 2020.

[26] W. Xiao, C. A. Belta, and C. G. Cassandras, “Sufficient conditions for feasibility of optimal control problems using control barrier functions,” Automatica, vol. 135, p. 109960, 2022.

[27] H. Almubarak, N. Sadegh, and E. A. Theodorou, “Safety embedded control of nonlinear systems via barrier states,” IEEE Control Systems Letters, vol. 6, pp. 1328–1333, 2021.

[28] H. Almubarak, K. Stachowicz, N. Sadegh, and E. A. Theodorou, “Safety embedded differential dynamic programming using discrete barrier states,” IEEE Robotics and Automation Letters, vol. 7, no. 2, pp. 2755–2762, 2022.

[29] M. Gandhi, H. Almubarak, Y. Aoyama, and E. Theodorou, “Safety in augmented importance sampling: Performance bounds for robust mppf,” arXiv preprint arXiv:2204.05963, 2022.

[30] L. Song, N. Wan, A. Gahlawat, C. Tao, N. Hovakimyan, and E. A. Theodorou, “Generalization of safe optimal control actions on networked multi-agent systems,” IEEE Transactions on Control of Network Systems, 2022.

[31] N. Wan, A. Gahlawat, N. Hovakimyan, E. A. Theodorou, and P. G. Voulgaris, “Cooperative path integral control for stochastic multi-agent systems,” in 2021 American Control Conference (ACC). IEEE, 2021, pp. 1262–1267.

[32] W. H. Fleming, “Exit probabilities and optimal stochastic control,” Applied Mathematics and Optimization, vol. 4, no. 1, pp. 329–346, 1977.