Annual Reviews in Control 57 (2024) 100948

Contents lists available at ScienceDirect

Annual Reviews in Control

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

Learning safe control for multi-robot systems: Methods, verification, and open challenges

Kunal Garg *, Songyuan Zhang, Oswin So, Charles Dawson, Chuchu Fan

Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, 02139, MA, USA

ARTICLE INFO

Keywords:
Safe multi-agent reinforcement learning
Certificate-based multi-agent control
Verification for multi-agent systems

ABSTRACT

In this survey, we review the recent advances in control design methods for robotic multi-agent systems (MAS), focusing on learning-based methods with safety considerations. We start by reviewing various notions of safety and liveness properties, and modeling frameworks used for problem formulation of MAS. Then we provide a comprehensive review of learning-based methods for safe control design for multi-robot systems. We start with various shielding-based methods, such as safety certificates, predictive filters, and reachability tools. Then, we review the current state of control barrier certificate learning in both a centralized and distributed manner, followed by a comprehensive review of multi-agent reinforcement learning with a particular focus on safety. Next, we discuss the state-of-the-art verification tools for the correctness of learning-based methods. Based on the capabilities and the limitations of the state-of-the-art methods in learning and verification for MAS, we identify various broad themes for open challenges: how to design methods that can achieve good performance along with safety guarantees; how to decompose single-agent-based centralized methods for MAS; how to account for communication-related practical issues; and how to assess transfer of theoretical guarantees to practice.

1. Introduction

1.1. Motivation and applications of MAS

Multi-agent systems (MAS) have received tremendous attention from scholars in different disciplines, including computer science and robotics, as a means to solve complex problems by subdividing them into smaller tasks (Dorri, Kanhere, & Jurdak, 2018). Some examples of MAS include smart grids (Ringler, Keles, & Fichtner, 2016), search and rescue teams (Queralta et al., 2020; Vorotnikov, Ermishin, Nazarova, & Yuschenko, 2018), edge computing (Wang, Wang et al., 2020), wireless communication networks (Cui, Liu, & Nallanathan, 2019), space systems (Huang, Zhang, Tian, & Chen, 2023; Ren, 2007; Wei, Luo, Dai, & Duan, 2018), package delivery (Salzman & Stern, 2020), power systems (Molzahn et al., 2017), and micro-grids (Espina et al., 2020). The design and analysis of MAS controllers present unique challenges, such as scalability, verification, and robustness to factors such as communication issues, adversarial or non-cooperative agents, and partial observability. While we will provide a detailed discussion on the limitations and challenges of learning-based methods for MAS, interested readers on current broad challenges in various aspects of MAS are referred to Canese et al. (2021), Du and Ding (2021), Ismail, Sariff, and Hurtado (2018) and Nweye, Liu, Stone, and Nagy (2022).

Learning-based methods have been successfully deployed on multi-robot systems demonstrating collision-free safe behaviors in physical robots, e.g., for drones in Vinod et al. (2022) and ground vehicles in Cui et al. (2022) and Everett, Chen, and How (2018). Some recent works have shown the benefits of learning-based methods compared to *classical* methods for robotic MAS-specific tasks such as cooperative exploration (Lyu, Xiao, Daley, & Amato, 2021), where learning-based methods can achieve coverage of an unknown region in half the time on a hardware robot car platform. Inspired by these recent successes of learning-based methods in safe MAS control, we provide a comprehensive summary of the current state-of-the-art learning-based methods for safe MAS control.

1.2. Scope of this survey

The four major topics of focus in this survey are

Shielding-based methods — Section 3: Methods that delegate safety to shielding function or safety filter which preserves the safety of the MAS. These include non-learning-based Control Barrier Function (CBF), Predictive Safety Filters (PSF), Hamilton–Jacobi (HJ) Reachability, Automata-based methods, and some heuristic methods.

* Corresponding author.
E-mail address: kgarg@mit.edu (K. Garg).

https://doi.org/10.1016/j.arcontrol.2024.100948
Received 21 November 2023; Accepted 5 March 2024
Available online 21 March 2024

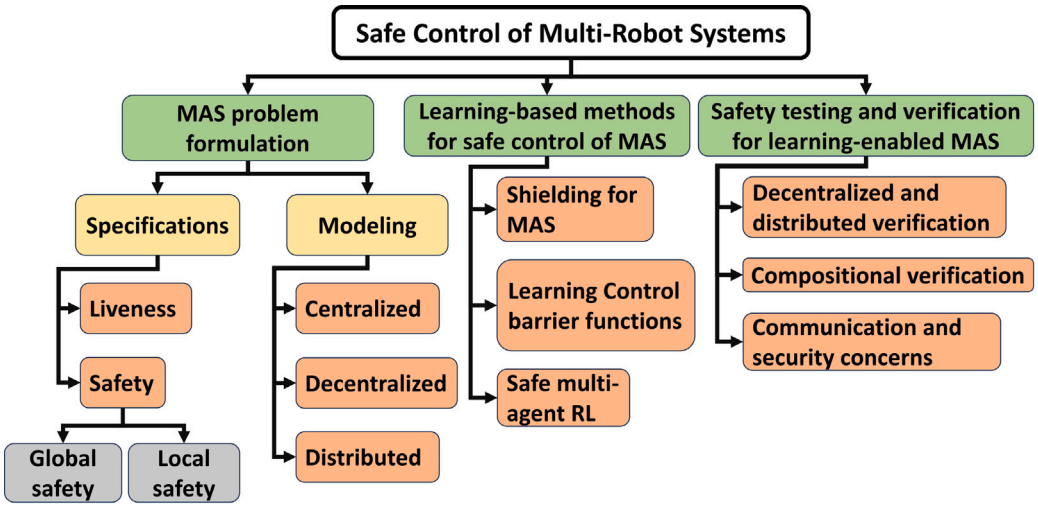

Fig. 1. Overview of the survey taxonomy on safe learning for MAS.

Methods for learning CBF — Section 4: Methods for learning centralized and distributed CBF along with a control policy for MAS.

Multi-agent reinforcement learning (MARL) — Section 5: Methods that apply reinforcement learning to MAS and directly tackle safety constraints.

Verification of learning-enabled MAS — Section 6: The challenges inherent in verifying learned controllers for MAS, how various centralized verification tools have been extended to handle MAS, and communication issues specific to MAS (see Fig. 1).

1.3. Main takeaway from this survey

- *Focus on MAS safety*: This is the first survey on robot MAS with an explicit focus on safety. The survey walks the reader through the taxonomy of the MAS control design problems, various learning-based methods for safe control synthesis, and the open problems in the field of safe MAS control.
- *Comprehensive survey of learning-based methods for multi-robot systems*: Unlike numerous existing surveys that only focus on MARL, this survey is intended to be a starting point for researchers getting started with learning-based safe control of MAS with a discussion of the capabilities and limitations of the existing learning-based tools.
- *Vision for the future of safe MAS control*: The survey identifies a range of open problems in the field of safe learning-based control, and verification thereof, for robot MAS and it informs the future research on the topic.

1.4. Previous surveys on multi-agent systems

Many survey and review articles appeared in the past few years on the topic of MAS. However, the key elements that differentiate this survey from the prior work are: (1) its focus on the safety of robotic MAS, (2) general learning methods as the central theme, and (3) its identification of open problems and challenges in safe learning-based MAS research. The article (Chen, Ren, et al., 2019) provides a detailed introduction to MAS taxonomy and control and Tahir, Böling, Haghbayan, Toivonen, and Plosila (2019) on Unmanned Aerial Vehicle swarms, but the discussion in these articles is restricted to non-learning-based methods. Given the popularity of MAS and safety in the context of RL, there have been many surveys that cover RL for MAS (Gronauer & Diepold, 2022; Nguyen, Nguyen, & Nahavandi, 2020; Oroojlooy & Hajinezhad, 2023; Rizk, Awad, & Tunstel, 2019; Zhang, Yang, & Başar, 2021b, 2021c; Zhou, Liu, & Tang, 2023) or for single-agent safety (Liu, Halev and Liu, 2021). Of these, only a few surveys (Gu et al., 2022; Zhou et al., 2023) cover the intersection of both topics. Despite this, many MARL works acknowledge safe MARL as a new area that has not been explored much but is a promising future direction (Gu et al., 2022; Oroojlooy & Hajinezhad, 2023; Zhang et al., 2021c).

The topic of safety has been reviewed extensively in the robotics community (Brunke et al., 2022). However, these works focus on the single-agent case, ignoring a broad category of disturbances and safety issues that are particular to MAS (e.g., communication delay and errors). While Liu, Halev et al. (2021) discusses safety but for single-agent systems. Quite a few surveys are looking at distributed optimization (Espina et al., 2020; Molzahn et al., 2017; Yang et al., 2019), power systems (Molzahn et al., 2017) and micro-grids (Espina et al., 2020). While safety is not explicitly discussed in these works, some of the methods reviewed in these surveys can incorporate it via the inclusion of safety constraints. Finally, many surveys are focusing on applications of MAS (Espina et al., 2020; Queralta et al., 2020; Xie & Liu, 2017; Yang et al., 2019).

1.5. Topics not covered by this article

Given the main focus of the survey being safe learning-based methods for MAS, various topics about robotic MAS are out of the scope of this paper. A non-comprehensive list of such topics along with recent surveys on those topics is Vector field-based methods (Gao, Bai, Fu, & Quan, 2023; Salman, Ayvali, & Choset, 2017); Model predictive control (Wang, Duan, Lv, Wang and Chen, 2020); Consensus control (Nowzari, Garcia, & Cortés, 2019); Distributed optimization; Nedić and Liu (2018) and Yang et al. (2019); and Multi-agent games (Wang et al., 2022).

1.6. Organization

We start with a general problem formulation, notations, and common definitions for MAS in Section 2. Section 3 introduces shielding methods for the safety of MAS, then Section 4 discusses learned certificates more specifically. Section 5 discusses various MARL-based methods. Section 6 covers verification techniques for MAS. Sections 7 and 8 conclude with a discussion of the challenges and open problems in the field of safe MAS control.

2. MAS problem formulation

This section defines common notions used in the context of multi-agent systems (MAS); namely, agent models, specifications, and modeling frameworks.

2.1. Definitions and notations

In this work, we focus on a general class of MAS consisting of N agents where each agent is a dynamic system modeled as

$$\dot{x}_i = F(x_i, u_i, d_i, v_{ij}(x_j)), \quad x_i \in \mathcal{X}_i, \quad u_i \in \mathcal{U}_i, \tag{1}$$

where $x_i \in \mathbb{R}^{n_i}, u_i \in \mathbb{R}^{m_i}$ denote the state and the input of agent i. The sets $\mathcal{X}_i \subseteq \mathbb{R}^{n_i}, \mathcal{U}_i \subseteq \mathbb{R}^{m_i}$ denote the operational workspace and the set of inputs for the ith agent. The term $d_i \in \mathcal{D}_i$ denotes the disturbances, uncertainties, and unmodeled dynamics for agent i, while the map $v_{ij} : \mathbb{R}^{n_j} \to \mathbb{R}^{p_i}$ denotes the influence of the other agents $j \neq i$ or in other words, the inter-agent coupling of the MAS dynamics. The joint state vector and the input vector are denoted as $\mathbf{x} = [x_1^\top, x_2^\top, \dots, x_N^\top]^\top \in \mathcal{X} \subset \mathbb{R}^{\sum n_i}$ and $\mathbf{u} = [u_1^\top, u_2^\top, \dots, u_N^\top]^\top \in \mathcal{U} \subset \mathbb{R}^{\sum m_i}$, respectively. The combined dynamics of the MAS can be written as

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}, \mathbf{u}, \mathbf{d}). \tag{2}$$

To enable theoretical analysis, it is typically assumed that the function $\mathbf{F}$ is locally Lipschitz continuous (Ames, Xu, Grizzle, & Tabuada, 2016).

For robotic systems, let $x_i \supset p_i \in \mathbb{R}^3$ denote the physical location of the ith agent in the 3D space. The state trajectory of agent i under a control policy π_i starting at an initial condition $x_i(0) \in \mathcal{X}_i$ is denoted as $\phi_i(\cdot, \pi_i; x_i(0)) : \mathbb{R}_+ \to \mathbb{R}^{n_i}$. Correspondingly, the state trajectory of the MAS is denoted as $\Phi(\cdot, \boldsymbol{\pi}; \mathbf{x}(0))$ with $\boldsymbol{\pi} : \mathcal{X} \to \mathcal{U}$ being the joint policy for the MAS. When an explicit emphasis on the underlying policy is not required, we denote the trajectories of the agents with $x_i(\cdot)$ and that of the MAS with $\mathbf{x}(\cdot)$ for the sake of brevity.

A network can be defined for the MAS with agents denoting the nodes and their communication links denoting edges. Let $R_i > 0$ be the sensing/communication radius of the ith agent and define $\mathcal{N}_i(t) = \{j \mid \|p_i - p_j(t)\| \leq R\}$ as the set of *neighbors* of the ith agent, i.e., the set of agents from (respectively, to) which the agent i can receive (respectively, send) information (Zavlanos & Pappas, 2008). Using this notion of neighbors, a graph topology can be defined for the MAS as $\mathcal{G}(t) = (\mathcal{V}, \mathcal{E}(t))$ where $\mathcal{V} = \{1, 2, \dots, N\}$ is the set of vertices denoting the agents and $\mathcal{E}(t)$ the time-varying set of connections given as $\mathcal{E}(t) = \{(i, j) : j \in \mathcal{N}_i, i \in \mathcal{V}\}$, that is, there is an edge $\mathcal{E}_{ij}(t)$ from agent j to agent i at time t if the agent i is able to receive information from agent j. A time-varying adjacency matrix $\mathcal{A}$ for this graph is defined as

$$\mathcal{A}_{ij}(t) = \begin{cases} 1 & j \in \mathcal{N}_i(t), \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

The Laplacian matrix $\mathcal{L}$ corresponding to the adjacency matrix $\mathcal{A}(t)$ is given as

$$\mathcal{L}_{ij}(\mathcal{A}(t)) = \begin{cases} \sum_{k \neq i} \mathcal{A}_{ik}(t), & j = i, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

From Mesbahi and Egerstedt (2010, Theorem 2.8), we know that the graph topology $\mathcal{G}(t)$ is connected at time t if and only if the second smallest eigenvalue of the Laplacian matrix is positive, i.e., $\lambda_2(\mathcal{L}(\mathcal{A}(t))) > 0$.

2.2. MAS specifications

In the temporal logic language (Baier & Katoen, 2008, Chapter 3), the control objective for the MAS can be characterized as:

1. **Safety property**: *Something bad never happens*. Agents remain in the safe region $\mathcal{S}(t) \subset \mathbb{R}^{\sum n_i}$ at all times, i.e., $\mathbf{x}(t) \in \mathcal{S}(t)$ for all $t \geq 0$ (Wang, Ames, & Egerstedt, 2016; Zhang, Bastani, & Kumar, 2019). Generally, the safe region is defined as the complement of the occupancy set of other agents, obstacles, and restricted regions in the workspace.
2. **Liveness property**: *Something good will eventually happen*. Agents move towards minimizing[1] a (possibly joint) objective function $\Psi : \mathbb{R}^{\sum n_i} \to \mathbb{R}$, i.e., $\mathbf{x}(t) \to \operatorname{argmin}_{\mathbf{x}} \Psi(\mathbf{x})$ (Atınç, Stipanović, & Voulgaris, 2020; Chen, Li, Fan, & Williams, 2021; Prajapat, Turchetta, Zeilinger, & Krause, 2022; Sun, Chen, Mitra, & Fan, 2022; Zhang, Yang, Liu, Zhang, & Basar, 2018). The most common example of a liveness property is each agent required to reach a goal location.

Here, we use the term dynamic obstacles for agents or entities that do not follow the designed control policy. Some safety properties can be decomposed at the agent level, while some safety properties need to be stated for the MAS as a whole. For example, the inter-agent safety property needs to be specified for the MAS using a *global* safe set $\mathcal{S}$ while obstacle avoidance property can be expressed individually for each agent with a *local* safety set $\mathcal{S}_i$. Another important requirement that can be posed as a safety property in MAS is *connectivity* maintenance, which becomes an important property for various applications such as coverage (Cortes, Martinez, Karatas, & Bullo, 2004) and formation control (Mehdifar, Bechlioulis, Hashemzadeh, & Baradarannia, 2020). For the sake of performance criteria as well as maintaining safety, the agents in MAS need to sense each other and might also need to actively communicate certain information. A detailed discussion on various safety properties is provided next.

Note that there are various notions of safety used in the literature, however, since the focus of this survey is robotic MAS, we restrict our discussions to safety as it pertains to *physical* safety of the agents.

2.2.1. Safety property

For a MAS (1), the notion of safety is defined as follows.

Definition 1 (*Safety*)**.** Given a (potentially time-varying) safety constraint set $\mathcal{S}(t)$, the MAS (1) is safe with respect to $\mathcal{S}(t)$ if for all $\mathbf{x}(0) \in \mathcal{S}(0)$, the trajectories satisfy $\mathbf{x}(t) \in \mathcal{S}(t)$ for all $t \geq 0$.

Below, we give examples of some of the possible safety constraints for robotic MAS. We draw a distinction between properties that require considering the joint state of the MAS (global properties) and those that can be checked using only individual agent states (local properties).

1. Global MAS safety properties
   - *Inter-agent collision avoidance*: Given a safe distance $0 < r_s < R$, the inter-agent collision avoidance can be formulated through $\mathcal{S} = \{\mathbf{x} \mid \|p_i - p_j\| > r_s, j \neq i\}$ (Panagou, 2016; Zhang, Garg and Fan, 2023).
   - *Connectivity maintenance*: Given a communication radius $R > 0$, the MAS network connectivity maintenance can be formulated as $\mathcal{S} = \{\mathbf{x} \mid \lambda_2(\mathcal{L}(\mathcal{A}(\mathbf{x}))) > 0\}$ (Sabattini, Secchi, Chopra, & Gasparri, 2013; Zavlanos & Pappas, 2008), where

$$\mathcal{A}_{ij}(\mathbf{x}) = \begin{cases} 1, & \|p_i - p_j\| \leq R, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

[1] In some works, the objective is given as maximization instead of minimization.

2. Local safety properties

   - *Obstacle avoidance*: Given a safe distance $0 < r_s < R$ and a set of (potentially moving) obstacles $\mathcal{O}_j(t) \subset \mathbb{R}^3$, obstacle avoidance can be formulated through the set $S_i(t) = \{p_i \mid \text{argmin}_{p \in \mathcal{O}_j(t)} \|p_i - p\| > r_s, \forall j\}$ (Chen, Singletary, & Ames, 2020; Zhang, Garg et al., 2023).
   - *State limits*: Given state limits (e.g., positions and velocities) of the form $x_m^j \leq x_i^j \leq x_M^j$ where x_i^j denotes the j−th component of x_i, the safety can be formulated with the set $S_i = \{x_i \mid x_m^j \leq x_i^j \leq x_M^j, j \in \{1, 2, \dots, n_i\}\}$[2] (Borrmann, Wang, Ames, & Egerstedt, 2015; Xian, Lertkultanon, & Pham, 2017).

2.2.2. Liveness properties

Here, we discuss commonly studied liveness properties for a MAS.[3]:

1. *Consensus*: Given a consensus point x_c, the trajectories of each of the agents converge in the following sense: $\lim_{t\to\infty} x_i(t) = x_c$ (Li, Tang, & Karimi, 2020; Ren, Beard, & Atkins, 2005).[4]
2. *Formation*: Given a set of off-set vectors x_{ij} for each (i, j), $i \neq j$, the trajectories of the MAS satisfy $x_i(t) - x_j(t) \to x_{ij}$ as $t \to \infty$ (Oh, Park, & Ahn, 2015; Xue & Cao, 2019).
3. *Coverage*: Given an exploration region $\mathcal{X}_e \subset \mathcal{X}$ and a distribution function $\varphi : \mathcal{X}_e \to \mathbb{R}_+$ and a smooth increasing function $f : \mathbb{R} \to \mathbb{R}$ that measure the degradation of sensing performance, the coverage objective is to minimize the function $\Psi(x) = \sum_{i=1}^{N} \int_{V_i} f(\|x_i - z\|)\phi(z)dz$, where $V_i \subset \mathcal{X}_e$ denotes the region where agent i is responsible for coverage (Cortes et al., 2004; Santos & Egerstedt, 2018).
4. *Goal reaching*: Given a set of goal states $x_{gi} \in \mathbb{R}^{n_i}$, the trajectories of each agent reach the goal, i.e., $\lim_{t\to\infty} x_i(t) = x_{gi}$ for each i (Garg & Panagou, 2019b; Majumdar, Mallik, Salamati, Soudjani, & Zareian, 2021; Panagou, 2016).
5. *Reference tracking*: Given a reference trajectory $x_{i,ref}(\cdot)$, the trajectories of each of the agent satisfy $x_i(t) - x_{i,ref}(t) \to 0$ as $t \to \infty$ (Adaldo, Liuzza, Dimarogonas, & Johansson, 2016; Saim, Ghapani, Ren, Munawar, & Al-Saggaf, 2017).

The liveness properties are important for capturing both performance criteria (e.g. goal reaching or trajectory tracking) and certain safety properties (particularly global safety properties). For example, MAS might be required to reach a consensus or maintain a formation to maintain a connectivity property. While the focus of this survey is on safety properties, we provide brief comments on the ability of the reviewed methods to accomplish tasks that often require both safety and liveness properties (particularly, goal-reaching).

2.3. MAS modeling framework

2.3.1. Centralized, decentralized, and distributed MAS

In this section, we present various modeling frameworks for MAS. Generally, MAS is modeled under the following three paradigms (see Fig. 2):

1. *Centralized*: An MAS is termed centralized if there is a *central* node where all the information/sensor data from all the agents is collected and decisions for each of the agents are made.
2. *Decentralized*: An MAS is termed decentralized if each agent makes its own decision based on its local information/sensor data *without* communicating with other agents.
3. *Distributed*: An MAS is termed distributed if each agent makes its own decisions based on its local information/sensor data along with information received by active communication with other agents.

We note there is currently no consensus in the literature on the definition of the decentralized and distributed paradigms. For example, Xuan and Lesser (2002) defines decentralized MAS where the agents communicate with their local neighbors. The authors in Roth, Simmons, and Veloso (2005) use a framework where agents communicate only when *needed*, calling this approach decentralized communication. In more recent multi-agent reinforcement learning (MARL) work (Zhang et al., 2018), the authors allow the agents to communicate over a time-varying connectivity graph and call the formulation *fully decentralized*. There are many examples of this interchangeable usage of the terms decentralized and distributed in the literature. For the sake of consistency, in this survey, we will stick with the notions as defined above (see Frampton, Baumann, & Gardonio, 2010; Molzahn et al., 2017).

Some examples of methods using centralized learning frameworks for safe control of MAS are (Dawson, Qin, Gao, & Fan, 2022; ElSayed-Aly et al., 2021; Gu et al., 2021; Khan et al., 2019; Zhang et al., 2019), the works in Cai, Cao, Lu, Zhang, and Xiong (2021) and Melcer, Amato, and Tripakis (2022) use a decentralized learning framework, and Lu, Zhang, Chen, Başar, and Horesh (2021) and Pereira, Saravanos, So, and Theodorou (2022) use a distributed learning framework. Centralized Training Decentralized Execution (CTDE) is a related paradigm, where the joint state and other global information are used to train a decentralized policy for each agent that only has access to local information (Gronauer & Diepold, 2022; Zhang, Garg et al., 2023).

2.4. Properties of algorithms for MAS safety

Finally, we discuss some desirable properties of learning-based algorithms for the safe control of MAS and categorize the main themes reviewed in this paper based on these properties (see Table 1):

Safety Guarantees - Theory: Here, we categorize the methods based on the fact that whether, under some suitable assumptions, it results in a safe policy. For instance, unconstrained MARL that relies on penalty-based mechanisms for encoding safety do not provide theoretical guarantees of safety.

Safety Guarantees - Practice: While theoretical guarantees are important, it is more useful to ask whether the assumptions needed to provide those safety guarantees from theory also hold in practice. For example, while certificate learning methods can provide safety guarantees if the certificate can be perfectly learned by the neural network (Dawson, Gao, & Fan, 2023), there is no guarantee that this will be the case. Similarly, while some constrained MARL methods guarantee that the policy at each iterate will be safe (Gu et al., 2023), this relies on the assumption that the value function is known exactly and that a trust-region optimization can be solved exactly, neither of which is true in practice.

Requirements - Domain Expertise: An important aspect that dictates the ease of usage and wide applicability of a method is whether domain expertise is needed about the specific dynamics or safety constraints to construct supporting tools and methods needed to apply the method. For instance, hand-crafting a CBF as a shield often requires domain expertise, especially under input constraints.

[2] More general constraints of the form $r(x_i) \leq 0$ for some constraint function r can also be considered.

[3] Some of these properties require compatible workspaces for the agents, i.e., $n_i = n_j = n$ for all i.

[4] More generally, it is possible to define consensus to $\rho(x_c)$ for some function ρ.

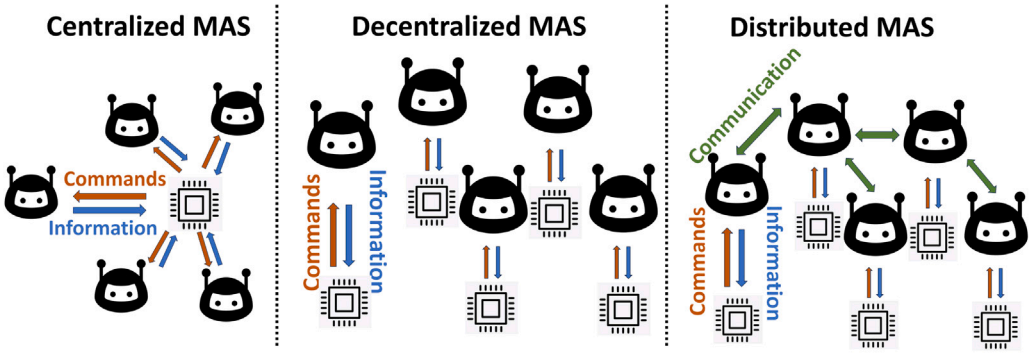

Fig. 2. MAS modeling: centralized, decentralized and distributed frameworks.

Table 1
Overview of different methods of handling safety for MAS. The tick mark denotes available features or requirements, while the cross mark denotes missing features or non-requirements. Blue denotes desirable properties, while red denotes undesirable properties.

Method	Safety guarantees		Requirements		Distributed policy
	Theory	Practice	Domain expertise	Known dynamics	
Shielding	✓	✓	✗/✓[a]	✓	✓/✗[b]
Certificate learning	✓	✗	✗	✓	✓
Unconstrained MARL	✗	✗	✗	✗	✓
Constrained MARL	✓	✗	✗	✗	✓

[a] HJ and Automata-based methods can be applied directly from safety specifications, while PSF (needs valid control-invariant set) and CBF-based (needs a valid CBF) shielding require domain expertise to use.

[b] CBF (Cai et al., 2021), PSF (Muntwiler, Wabersich, Carron, & Zeilinger, 2020) and automata-based (Melcer et al., 2022) shielding have distributed versions that maintain their safety guarantees in practice. While HJ-based shielding does have a distributed version by considering pairwise interactions (Chen, Hu, Mackin, Fisac, & Tomlin, 2015), the safety guarantees do not hold for the full MAS.

Requirement - Known Dynamics: Another factor that plays an important role in the generalizability and wide applicability of a method is whether the exact form of the dynamics is needed (e.g., for computing jacobians/querying at arbitrary states) to apply the method, or it is sufficient to have black-box evaluations of the dynamics along a set of trajectories. For instance, hand-crafted shielding methods and certificate learning methods require knowledge of the dynamics and its structure (i.e., control-affine), while RL-based methods are model-free and only require black-box evaluations.

Distributed Policy: Finally, and very importantly for large-scale MAS, one needs to ask whether the policy can be deployed in a distributed manner without a central computation/aggregation node. For instance, HJ-based shielding methods require a centralized framework for safety guarantees.

3. Shielding-based learning for MAS

One popular method of providing safety to learning-based methods is via the use of *shielding* or *safety filter*s, where an *unconstrained* learning method is paired with a *shield* or *safety filter*. Such shields are often constructed without learning with the objective of either modifying the input or the output of the learning method to maintain safety. One benefit of shielding-based methods is that safety can be guaranteed during both training and deployment since the shield is constructed prior to training. However, a drawback of some of these methods is that they require domain expertise for the construction of a valid shield, which can be challenging in the single-agent setting and becomes even more difficult for MAS. Other methods can automatically synthesize shields, but face scalability challenges.

Let $\boldsymbol{\pi} : \mathcal{X} \to \mathcal{U}$ denote the *joint* policy for MAS (2) from a learning-based controller without safety considerations. Shielding-based methods define a *shielding function* $\boldsymbol{\Pi} : \mathcal{X} \times \mathcal{U} \to \mathcal{U}$ that takes the output of $\boldsymbol{\pi}$ and returns a shielded output (Zhang et al., 2019). The level of safety and the type of safety guarantees that can be obtained depends on how the shielding function $\boldsymbol{\Pi}$ is constructed. We provide an overview of different shielding-based methods used to ensure the safety of learning for MAS in Fig. 3.

3.1. Control barrier function-based shielding

One method of constructing a shield $\boldsymbol{\Pi}$ is via a Control Barrier Function (CBF). We start by reviewing the notion of CBF.

3.1.1. Definition of CBF

The notion of CBF was introduced to satisfy the conditions of set invariance, where a set is termed as *forward invariant* if starting in the set, the system trajectories do not leave it (Blanchini, 1999; Brezis, 1970; Nagumo, 1942). It is also related to the notion of Control Lyapunov Function (CLF) (Ames, Galloway, Sreenath and Grizzle, 2014; Sontag, 1983) which is commonly used for liveness properties, and extends the definition of barrier certificates (Prajna, 2006; Prajna & Jadbabaie, 2004) to control systems. A comprehensive review of CBFs as a tool for safety can be found in Ames et al. (2019).

While there exists various definitions of CBF with slight variations (Ames, Grizzle and Tabuada, 2014; Ames et al., 2016; Dawson et al., 2023; Wieland & Allgöwer, 2007), we use the following definition in this survey (Ames et al., 2019):

Definition 2 (*CBF*)**.** Consider the MAS dynamics (2) with no disturbances, i.e., $\mathbf{d} = 0$. Let $\mathcal{C} \subset \mathcal{X}$ be the 0-superlevel set of a continuously differentiable function $B : \mathcal{X} \to \mathbb{R}$, i.e., $\mathcal{C} = \{\mathbf{x} \in \mathcal{X} : B(x) \geq 0\}$. Then, the function B is a CBF if there exists an extended class-$\mathcal{K}$ function[5] $\alpha : \mathbb{R} \to \mathbb{R}$ such that:

$$\sup_{\mathbf{u} \in \mathcal{U}} \left[\frac{\partial B}{\partial \mathbf{x}} \mathbf{F}(\mathbf{x}, \mathbf{u}) + \alpha\left(B(\mathbf{x})\right) \right] \geq 0, \quad \forall \mathbf{x} \in \mathcal{X}. \tag{6}$$

[5] A continuous function $\alpha : \mathbb{R} \to \mathbb{R}$ is said to be an extended class-$\mathcal{K}$ function if it is strictly increasing with $\alpha(0) = 0$.

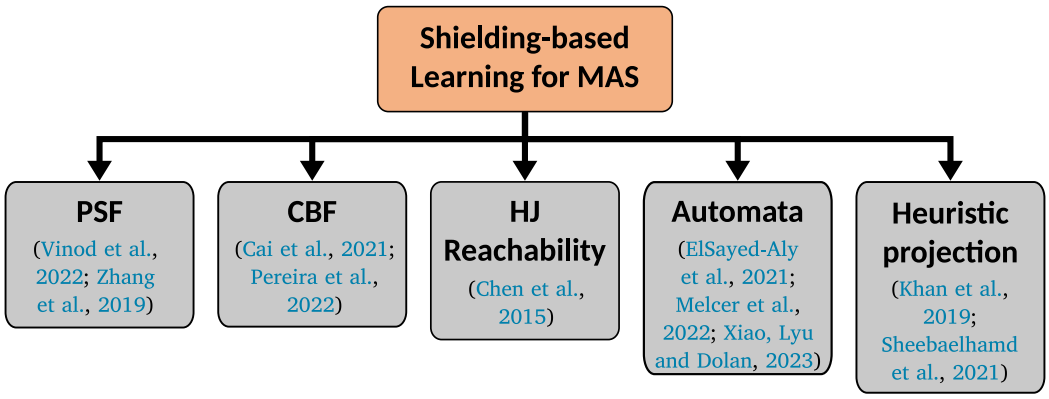

Fig. 3. Overview of Shielding-based Learning for MAS.

Safe Control Set: Based on (6), we can define a set of *safe* control input

$$K_{\text{CBF}}(\mathbf{x}) = \left\{ \mathbf{u} \in \mathcal{U} : \frac{\partial B}{\partial \mathbf{x}} \mathbf{F}(\mathbf{x}, \mathbf{u}) + \alpha\left(B(\mathbf{x})\right) \geq 0 \right\}. \tag{7}$$

The authors in Ames et al. (2019) proved the following result on forward invariance of the set $\mathcal{C}$ using the notion of CBF:

Theorem 1. *Let $\mathcal{C}$ be the 0-superlevel set of a continuous differentiable function $B : \mathcal{X} \to \mathbb{R}$, i.e., $\mathcal{C} = \{\mathbf{x} \in \mathcal{X} : B(\mathbf{x}) \geq 0\}$. If B is a CBF, and $\frac{\partial B}{\partial \mathbf{x}} \neq 0$ for all $\mathbf{x} \in \partial\mathcal{C}$, then any Lipschitz continuous policy $\boldsymbol{\pi} : \mathcal{X} \to \mathcal{U}$ with $\boldsymbol{\pi}(\mathbf{x}) \in K_{\text{CBF}}(\mathbf{x})$ renders the set $\mathcal{C}$ forward invariant.*

Remark 1. The forward invariance of the set $\mathcal{C}$ can be used for guaranteeing safety for a set $\mathcal{S}$ as follows. Based on Theorem 1, if a CBF and a controller are found on $\mathcal{X}$ that satisfy the conditions of Theorem 1 and $\mathcal{C} \subset \mathcal{S}$, then starting from any initial condition in $\mathcal{C}$, the system remains safe.

There are variations of CBF that can also render the safety of autonomous systems. For example, an appropriate choice of Lyapunov function can be used for safety since the sublevel sets of a Lyapunov function are forward invariant (Tee, Ge, & Tay, 2009). Based on this idea, Dawson et al. (2022) combined CLF and CBF, and introduced a framework for learning a Control Lyapunov Barrier Function (CLBF) that guarantees both safety and stability. The approach in Yu, Yu and Gao (2023) proposes another variation of CBF called the Control Admissibility Model (CAM), which uses a notion of CBF where the function B depends on both the state $\mathbf{x}$ and the control input $\mathbf{u}$. The central idea of most of the variations is still forward invariance as per Theorem 1.

3.1.2. CBF-based shielding synthesis

For control-affine systems $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}) + \mathbf{G}(\mathbf{x})\mathbf{u}$, condition (6) becomes a linear constraint in the control input $\mathbf{u}$. When the input constraint set $\mathcal{U}$ is a convex polytope, a centralized method to synthesize a safe control input through CBF-based shielding is using Quadratic Programming (QP):

$$\begin{aligned} \min_{\mathbf{u} \in \mathcal{U}} \quad & \|\mathbf{u} - \boldsymbol{\pi}_{\text{nom}}(\mathbf{x})\|^2, \\ \text{subject to} \quad & L_{\mathbf{F}} B(\mathbf{x}) + L_{\mathbf{G}} B(\mathbf{x})\mathbf{u} + \alpha\left(B(\mathbf{x})\right) \geq 0, \end{aligned} \tag{8}$$

where $L_{\mathbf{F}} B(\mathbf{x}) = \nabla B(\mathbf{x})^{\top} \mathbf{F}(\mathbf{x})$ and $L_{\mathbf{G}} B(\mathbf{x}) = \nabla B(\mathbf{x})^{\top} \mathbf{G}(\mathbf{x})$ and $\boldsymbol{\pi}_{\text{nom}} : \mathcal{X} \to \mathcal{U}$ is a nominal policy (from some unconstrained learning methods) that does not necessarily consider safety with respect to the set $\mathcal{C}$. Given a state $\mathbf{x}$ and a control policy $\boldsymbol{\pi}_{\text{nom}}$, the CBF-based shield $\boldsymbol{\Pi}$ outputs the solution $\mathbf{u}^*$ to the QP (8) that minimally modifies the nominal control input while guaranteeing the safety of the system. Here, the nominal policy may come from a learning-based approach without safety considerations, and the QP (8) provides a shielding mechanism for enforcing safety with this learned control policy.

3.1.3. Constructing CBF

Once a CBF is given for control-affine systems, we can solve problem (8) for shielding. However, finding a CBF for MAS is not trivial, and there is no generalized framework that can find a CBF for any MAS. Here, we review approaches that can efficiently compute CBFs for certain specific types of systems.

For systems with relatively simple dynamics, such as single integrator, double integrator, and unicycle dynamics, it is possible to use a distance-based CBF (Ames et al., 2019; Ames, Grizzle et al., 2014; Ames et al., 2016; Garg, Arabi, & Panagou, 2022; Garg & Panagou, 2019a, 2021; Tong, Dawson, & Fan, 2023; Wu & Sreenath, 2016; Xu et al., 2017; Yin, Dawson, Fan, & Tsiotras, 2023). Some works also explore provable safety along with liveness by guaranteeing the feasibility of the underlying CBF-QP (Garg et al., 2022; Garg, Cosner, Rosolia, Ames, & Panagou, 2021; Garg & Panagou, 2021). For systems with multiple safety constraints, e.g., velocity constraints, and joint angle constraints, it is possible to design a CBF for each constraint and then combine them (Glotfelter, Cortés, & Egerstedt, 2017; Hsu, Xu, & Ames, 2015; Usevitch, Garg, & Panagou, 2020). However, one needs domain expertise to handcraft each of these CBFs. It is also difficult to encode input constraints when handcrafting the CBF, and therefore, the CBF-QP (8) can be infeasible.

For systems with polynomial dynamics, it is possible to use the Sum-of-Squares (SoS) method to compute a CBF. The key idea of SoS is that the CBF conditions (6) consist of a set of inequalities, which can be equivalently expressed as checking whether a polynomial function is SoS. In this manner, a CBF can be computed through convex optimization (Ahmadi & Majumdar, 2016; Clark, 2021; Srinivasan, Abate, Nilsson, & Coogan, 2021). However, these methods are limited to polynomial dynamics. Moreover, the SoS-based approaches suffer from the curse of dimensionality (i.e., the computational complexity grows exponentially with respect to the degree of polynomials involved) (Ahmadi & Majumdar, 2016).

3.1.4. Distributed CBF

While centralized CBF is an effective shield for small-scale MAS, due to its poor scalability, it is not easy to use it for large-scale MAS. To address the scalability problem, the notion of distributed CBF can be used (Cai et al., 2021; Lindemann & Dimarogonas, 2019; Panagou, Stipanović, & Voulgaris, 2013; Zhang, Garg et al., 2023). In contrast to centralized CBF where the state $\mathbf{x}$ of the MAS is used, for a distributed CBF, only the local observations and information available from communication with neighbors are used, reducing the problem dimension significantly.

Similar to a variety of notions of centralized CBF, there exists a variety of definitions of distributed CBF in the literature (Borrmann et al., 2015; Glotfelter et al., 2017; Lindemann & Dimarogonas, 2019; Qin, Zhang, Chen, Chen, & Fan, 2020; Wang, Ames, & Egerstedt, 2017; Zhang, Garg et al., 2023). Following the graph notions of MAS

introduced in Section 2.1, we review a slightly modified definition of distributed CBF from Zhang, Garg et al. (2023). Let $o_i \in \mathcal{O}_i \subset \mathbb{R}^{r_i}$ be the observation vector of agent i, and let $z_i \in \mathcal{Z}_i \subset \mathbb{R}^{q_i}$ be the encoding[6] of the information accepted by agent i from its neighbors $\mathcal{N}_i$. Note that z_i depends on the states of the neighbors of agent i.

Definition 3 (*Distributed CBF*)**.** Consider the MAS agent dynamics (1) with no disturbances and no inter-agent influences, i.e., $d_i = 0$ and $\nu_{ij}(x_j) = 0$, for all $i, j \in \mathcal{V}$. Let $\mathcal{C}_i \subset \mathcal{X}_i$ be the 0-superlevel set of a continuously differentiable function $B_i : \mathcal{X}_i \times \mathcal{O}_i \times \mathcal{Z}_i \to \mathbb{R}$. Then, the function B_i is a distributed CBF if there exists an extended class-$\mathcal{K}$ function $\alpha_i : \mathbb{R} \to \mathbb{R}$ and for each $\mathbf{x} \in \mathcal{X}$, there exists a control input $\mathbf{u} \in \mathcal{U}$, such that the following holds:

$$\frac{\partial B_i}{\partial x_i} F_i(x_i, u_i) + \frac{\partial B_i}{\partial o_i} \dot{o}_i + \sum_{j \in \mathcal{N}_i \cup \{i\}} \frac{\partial B_i}{\partial z_i} \frac{\partial z_i}{\partial x_j} F_j(x_j, u_j) + \alpha_i(B_i(x_i)) \geq 0, \quad \forall i. \tag{9}$$

Similar to the centralized CBF, under certain conditions, the distributed CBF can also guarantee the safety of the MAS (Zhang, Garg et al., 2023). According to Definition 3, the MAS is assumed to be cooperative, i.e., all the agents coordinate to satisfy CBF condition (9) for MAS (Machida & Ichien, 2021; Qin et al., 2020; Wang et al., 2016, 2017; Zhang, Garg et al., 2023). Other works consider the worst-case scenario that the agents are non-cooperative (Borrmann et al., 2015), in which case, the agents do not have any communication, resulting in a decentralized CBF formulation. In addition, Zhang, So, Garg, and Fan (2024) introduces graph control barrier functions (GCBFs), which not only can certify safety in MAS but also can generalize to an arbitrary number of agents.

3.1.5. Constructing distributed CBF

One way to construct distributed CBF is by *decomposing* centralized CBF. There are many ways to decompose centralized CBF. For example, assuming that other agents keep constant velocities, actively chasing the *ego* agent, or actively avoiding collision with the ego agent (Borrmann et al., 2015). Other works (Cosner, Chen, Leung, & Pavone, 2023; Lindemann & Dimarogonas, 2020; Wang et al., 2016, 2017) consider risk allocation among agents while decomposing the centralized CBF. In addition, decomposing the centralized CBF allows each agent to solve the optimization problem individually based on their local information in a distributed fashion, and therefore reduces computation costs (Borrmann et al., 2015; Pereira et al., 2022; Wang et al., 2017).

Another way to construct a distributed CBF is through a bottom-up approach, e.g., by composing pair-wise CBF. These approaches often encode the constraints of all the pair-wise CBF conditions in a QP to find a feasible control input that can maintain safety with respect to each of the pair-wise CBF (Chen et al., 2020; Funada et al., 2019; Glotfelter et al., 2017; Glotfelter, Cortés, & Egerstedt, 2018; Hu, Fu, & Wen, 2023; Jankovic & Santillo, 2021; Jiang & Guo, 2023; Luo, Sun, & Kapoor, 2020; Mali, Harikumar, Singh, Krishna, & Sujit, 2021; Yu, Yu et al., 2023).

3.2. Predictive safety filter-based shielding

Predictive safety filter (PSF) is another common method of constructing a shield (Wabersich & Zeilinger, 2018), which is also closely related to model predictive shielding (MPS) (Bastani, 2021; Li & Bastani, 2020).

For PSF, let $\mathcal{X}_{\text{CI}} \subseteq \mathcal{S}$ denote a subset of the safe set that is *control invariant*, i.e., there exists some controller under which this set is forward invariant. During deployment, at each time step, a constrained optimization problem is solved that constrains the terminal state within $\mathcal{X}_{\text{CI}}$.

$$\min_{u_k} \quad \|\boldsymbol{\pi}(\mathbf{x}_0) - \mathbf{u}_0\|^2 \tag{10a}$$

$$\text{s.t.} \quad \mathbf{x}_{k+1} = f(\mathbf{x}_k, \mathbf{u}_k), \quad \forall k = 0, 1, \ldots, N-1 \tag{10b}$$

$$\mathbf{x}_k \in \mathcal{S}, \quad \forall k = 0, 1, \ldots, N-1 \tag{10c}$$

$$\mathbf{x}_N \in \mathcal{X}_{\text{CI}}. \tag{10d}$$

The output of the safety filter $\boldsymbol{\Pi}$ is then taken as the solution $\mathbf{u}_0$ of (10). The terminal state constraint (10d) helps guarantee recursive feasibility of (10) and, as a result, guarantees infinite-horizon constraint satisfaction.

In MPS, suboptimality of the underlying constrained optimization problem is traded for substantially lowered computational costs. Instead of solving the potentially nonlinear optimization problem (10) at each time-step, a merely feasible solution is found. Let $\boldsymbol{\pi}_{\text{backup}}$ denote a *backup* policy designed to bring the system inside a set $\mathcal{X}_{\text{FI}} \subseteq \mathcal{S}$ that is forward invariant under $\boldsymbol{\pi}_{\text{backup}}$. Then, MPS chooses $\mathbf{u}_0 \in \{\boldsymbol{\pi}(\mathbf{x}_0), \boldsymbol{\pi}_{\text{backup}}(\mathbf{x}_0)\}$ and assigns $\mathbf{u}_k = \boldsymbol{\pi}(\mathbf{x}_k)$ for $k > 0$ (Bastani, 2021; Li & Bastani, 2020). If taking $\mathbf{u}_0 = \boldsymbol{\pi}(\mathbf{x}_0)$ does not lead to a feasible solution (i.e., $\mathbf{x}_N \in \mathcal{X}_{\text{FI}}$), then $\mathbf{u}_0$ is taken to be $\boldsymbol{\pi}_{\text{backup}}(\mathbf{x}_0)$.

MPS is extended to the multi-agent case in Zhang et al. (2019), where safety under $\boldsymbol{\pi}_{\text{backup}}$ is considered agent-wise as opposed to for the entire MAS. This helps avoid suboptimal cases where all agents are forced to use $\boldsymbol{\pi}_{\text{backup}}$ even when only a single agent is not safe under $\boldsymbol{\pi}$. However, Zhang et al. (2019) requires a centralized node to compute the MPS, and hence may have challenges scaling to a larger number of agents.

In Vinod et al. (2022), a PSF-like shield is used without the terminal state constraint (10d) for the *joint* MAS. Linear dynamics and linear constraints are considered, which, along with no terminal constraints, allows for (10) to be solved efficiently as a QP. Although removing the terminal state constraint also removes recursive feasibility guarantees, infeasibility was not reported in Vinod et al. (2022) during hardware experiments.

Finally, in Muntwiler et al. (2020), a distributed method of PSF for linear systems with linear safety constraints and bounded disturbances is introduced. The disturbances are handled using a robust distributed MPC technique that uses tube MPC (Conte, Zeilinger, Morari, & Jones, 2013), and a distributed negotiation procedure is introduced to allow agents to trade safety margins with neighbors.

3.3. Hamilton–Jacobi-based shielding

The HJ methods are a class of optimal control tools for finding the reachable set of a dynamical system under worst-case disturbances; Bansal, Chen, Herbert, and Tomlin (2017) provides a good introduction to this field. Most HJ methods proceed by solving a partial differential inequation for a scalar field over the joint state. This scalar field is known as the *HJ value function* and its zero superlevel set is the control invariant set. The HJ value function also defines a controller that renders the system safe. A common HJ shielding strategy is to use this controller as a shield that only activates if the value function gets too close to zero (Bansal et al., 2017). This shielding method can lead to undesirable bang–bang control behavior. Other formulations produce smoother HJ shielding controllers (Choi, Lee, Sreenath, Tomlin, & Herbert, 2021). The HJ-based methods have also been used to guide the learning of CBF controllers (Tonkens & Herbert, 2022).

A classic weakness of HJ methods is that they require solving a partial differential inequation over the state space of the system, and so the computational and memory requirements scale exponentially with the dimension of the state of the system (Bansal et al., 2017). The memory requirements can be reduced by using function approximations, such as neural networks, to represent the HJ value function (Bansal & Tomlin, 2021; Fisac, Lugovoy, Rubies-Royo, Ghosh, & Tomlin, 2019).

[6] There are many ways of encoding information, such as concatenation (Borrmann et al., 2015), summation (Qin et al., 2020), and applying attention (Zhang, Garg et al., 2023).

However, this dependence on dimensionality nevertheless prevents HJ methods from being applied to the joint state of MAS. Instead, these methods factor the MAS into pairs of agents and then solve a pairwise HJ problem (Chen et al., 2015).

3.4. Automata-based shielding

Shielding has also been applied using tools from the field of *formal methods*, where safety requirements are defined using linear temporal logic (Pnueli, 1977). Shielding for safe learning-based control using automata was introduced in a single agent case in Alshiekh et al. (2018), borrowing ideas from Bloem, Könighofer, Könighofer, and Wang (2015), where a safety game (Mazala, 2002) is solved to compute a set of states from where safety can be preserved.

In ElSayed-Aly et al. (2021), this is extended to the multi-agent case. Instead of constructing a single shield that monitors all agents, the state–space is decomposed into multiple pieces, and a shield is constructed for each piece. The authors show that this allows the algorithm to scale from two to four agents on a grid world. This work is extended in Xiao, Lyu and Dolan (2023), where this decomposition occurs dynamically. In Melcer et al. (2022), a decentralized shield is constructed, improving scalability.

However, the shield synthesis tools used in these works require a finite abstraction of the state and control spaces (Alshiekh et al., 2018; Bloem et al., 2015) and scale exponentially with the abstraction size (Könighofer et al., 2017). This can lead to conservative behavior when coarse abstractions are used (Alshiekh et al., 2018).

3.5. Heuristic projection

Finally, there are shielding methods that do not provide formal safety guarantees but rather act as heuristics. In Sheebaelhamd et al. (2021), the heuristic approach from Dalal et al. (2018) is used as a shield in a MARL framework. Here, the safety constraints are linearized, and it is assumed that only a single constraint is active at each time, giving rise to a closed-form solution that can be computed easily. In Khan et al. (2019), the velocity obstacle approach (Fiorini & Shiller, 1998) is used as a shield. Here, agents' velocities inside the velocity obstacle are projected back to the safe set. For a dynamic obstacle moving with constant velocity, a velocity obstacle defines the set of velocities of the agent that results in a collision with the dynamic obstacle (Fiorini & Shiller, 1998). When other agents are modeled as dynamic obstacles, the constant velocity assumption may not hold, but variants that make similar assumptions have been used successfully in practice (Snape, Van Den Berg, Guy, & Manocha, 2011).

The tradeoff between improved ease of use and computational cost is that these methods do not have the same safety guarantees as the previous categories of shields.

4. Learning control barrier functions for MAS

Shielding is a powerful technique to guarantee the safety of the MAS. However, hand-crafting a shield is generally difficult, requires domain expertise, and can be done for a relatively small class of problems. Furthermore, it is computationally heavy to find a shield through optimization, especially for large-scale MAS with complex dynamics. To address this problem, there has been a lot of development on using machine learning to find a shield and use it as a guide for controller synthesis. Most of these works focus on a specific kind of shield, namely, CBFs (Dawson et al., 2023), and tend to compute a CBF and a safe control policy simultaneously. In this manner, the computed control policy is encouraged to satisfy the CBF constraints and can be used for satisfying the safety requirements. In this section, we review approaches that learn a CBF and a control policy, for both centralized and distributed settings (see Fig. 4 for an overview of learning CBF methods).

4.1. Centralized CBF

There is a plethora of work on learning centralized CBF for safety (Jin, Wang, Yang, & Mou, 2020; Peruffo, Ahmed, & Abate, 2021; Qin, Sun, & Fan, 2022; Robey et al., 2020; Saveriano & Lee, 2019; So et al., 2024; Srinivasan, Dabholkar, Coogan, & Vela, 2020; Wang et al., 2023a, 2023b; Xiao, Wang et al., 2023). The general idea of these approaches is to use *self-supervised learning* for learning a common CBF for the entire MAS (Dawson et al., 2023). To this end, first, a centralized CBF $B_\theta : \mathcal{X} \to \mathbb{R}$ and a centralized control policy $\boldsymbol{\pi}_\phi : \mathcal{X} \to \mathcal{U}$ are parameterized using Neural Networks (NNs) with parameter θ and ϕ, respectively. Next, a loss function is designed to map the CBF constraint (6) to a penalty term:

$$\mathcal{L}_{\text{deriv}}(\theta,\phi) = \frac{1}{N_{\text{sample}}} \sum_{\mathbf{x}\in\mathcal{X}} \left[-\frac{\partial B_\theta}{\partial \mathbf{x}} \mathbf{F}(\mathbf{x}, \boldsymbol{\pi}_\phi(\mathbf{x})) - \alpha\left(B_\theta(\mathbf{x})\right) \right]^+, \tag{11}$$

where $[\cdot]^+ = \max(\cdot, 0)$ denotes the ReLU function, N_{sample} is the total number of state samples collected, and α is an extended class-$\mathcal{K}$ function.[7] To render the safety of the system, following Remark 1, it is essential that the 0-superlevel set of B_θ, denoted as C_θ, is a subset of the *known* safe set S, i.e., $C_\theta \subset S$. However, ensuring such a condition is not straightforward during the learning process.

Therefore, one cannot naively sample from the safe space and constrain the value of B_θ on these samples to be non-negative. To design self-supervised learning losses, most works consider an alternative way: making B_θ negative everywhere in the unsafe space $\mathcal{X} \setminus S$ (Jin et al., 2020; Srinivasan et al., 2020). Such a loss function can be readily defined as

$$\mathcal{L}_{\text{unsafe}}(\theta) = \frac{1}{N_{\text{unsafe}}} \sum_{\mathbf{x}\in\mathcal{X}\setminus S} \left[B_\theta(\mathbf{x})\right]^+, \tag{12}$$

where N_{unsafe} is the number of state samples in the unsafe space. However, if a loss of the form $\mathcal{L}_{\text{unsafe}}(\theta) + \mathcal{L}_{\text{deriv}}(\theta,\phi)$ is used, since there are only *negative* samples, the learned policy can easily converge to a sub-optimal solution where the forward-invariant set C_θ is relatively very small, if not empty. Hence, it is essential that *positive* samples from the safe set are also used in learning the barrier function. Naively using *positive* samples from the safe set S and *negative samples* from the unsafe set $\mathcal{X} \setminus S$ does not work since a point being safe (i.e., $x \in S$) does not imply that the dynamical system can remain safe in the future if it passes through this point. Hence, the *positive* samples must be collected from a control invariant set $C \subset S$ so that the system can be guaranteed to remain safe at all future times if it passes through these samples. Based on this, most work on CBF learning e.g., Dai, Krishnamurthy, and Khorrami (2022), Jin et al. (2020), Srinivasan et al. (2020) and Yu, Hirayama, Yu, Herbert and Gao (2023) consider an additional loss term:

$$\mathcal{L}_{\text{safe}}(\theta) = \frac{1}{N_{\mathcal{X}_C}} \sum_{\mathbf{x}\in\mathcal{X}_C} \left[-B_\theta(\mathbf{x})\right]^+, \tag{13}$$

where $\mathcal{X}_C \subset S$ is an approximation of the *unknown* ground truth forward invariant set C, and $N_{\mathcal{X}_C}$ is the number of state samples in this set, which act as *positive* samples. Since the ground truth forward invariant set C is computationally expensive to find, researchers generally approximate it with the set $\mathcal{X}_C$ through various methods, such as using the set of initial conditions if it is known to be part of the forward-invariant set (Jin et al., 2020), using distance-based heuristic methods (Srinivasan et al., 2020; Yu, Hirayama et al., 2023), or approximating it with a look-forward mechanism (Zhang et al., 2024).

Furthermore, to avoid learning a *flat* CBF whose value is close to 0 over the entire state space, many works also add a small margin $\nu > 0$ in the loss terms (Dawson et al., 2022). Some works also consider

[7] In practice, a linear function $\alpha(y) := \alpha y$ is chosen, for some $\alpha > 0$.

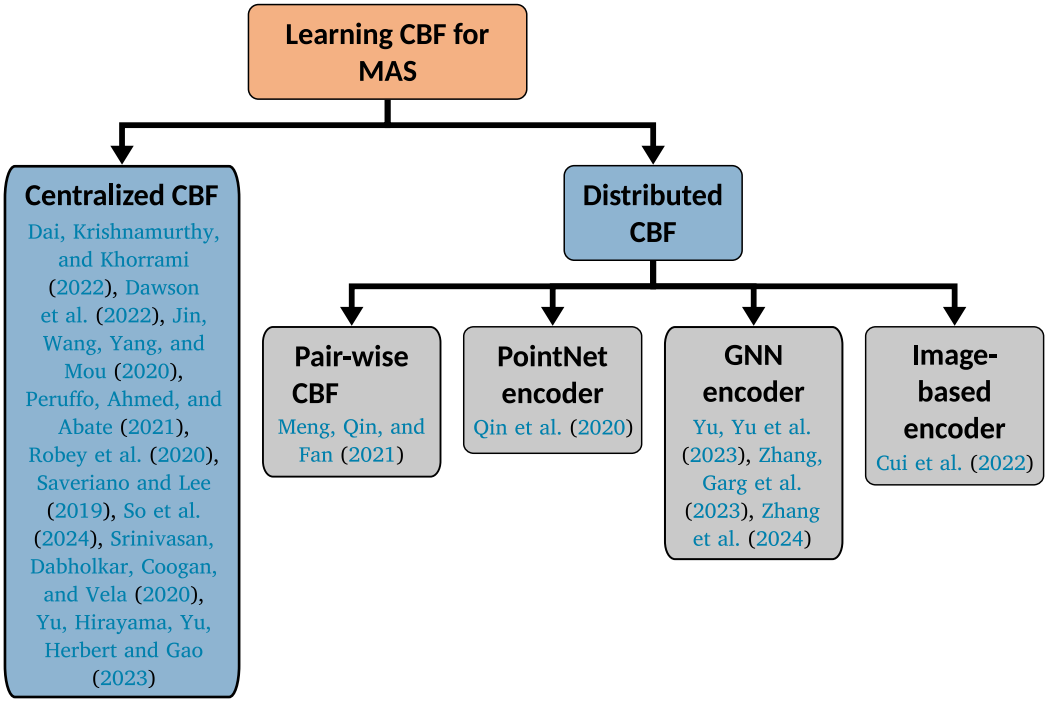

Fig. 4. Overview of learning CBF for MAS.

a reference behavior cloning controller $\boldsymbol{\pi}_{\text{BC}}$ for the liveness property, such as reaching a goal location. As a result, another loss term

$$\mathcal{L}_{\text{ctrl}} = \frac{1}{N} \sum_{\mathbf{x} \in \mathcal{X}} \|\boldsymbol{\pi}_{\phi}(\mathbf{x}) - \boldsymbol{\pi}_{\text{BC}}(\mathbf{x})\|^2,$$

is added so that the learned controller is as close to the behavior cloning controller as possible (Dawson et al., 2022). Most of the works use a nominal controller $\boldsymbol{\pi}_{\text{nom}}$ as the behavior cloning controller, which only considers the liveness property without the purpose of collision (Dawson et al., 2022; Qin et al., 2020; Zhang, Garg et al., 2023). However, the safety and the liveness properties encoded here via the means of different loss terms compete with each other in learning and often lead to a sub-optimal solution which either has a high performance with poor safety rate (learned controller close to the nominal controller) or high safety rate with poor performance (Zhang, Garg et al., 2023). Recently, Zhang et al. (2024) proposes to use the controller solved from CBF-QP (8) as the behavior cloning controller, which solves the competition problem between safety and liveness.

Based on these individual loss terms, the parameterized CBF B_{θ} and the control policy $\boldsymbol{\pi}_{\phi}$ are trained using the loss function defined as

$$\mathcal{L}_{\text{CBF}}(\theta, \phi) = \mathcal{L}_{\text{deriv}}(\theta, \phi) + \mathcal{L}_{\text{unsafe}}(\theta) + \mathcal{L}_{\text{safe}}(\theta) + \mathcal{L}_{\text{ctrl}}(\phi). \tag{14}$$

Upon convergence, a CBF and a safe control policy are obtained. The training data for such learning methods are either collected by random sampling in the state space (Jin et al., 2020) or from simulated trajectories (Dai et al., 2022; Yu, Hirayama et al., 2023). These approaches are generalizable to a large class of dynamics and relatively high-dimensional systems in contrast to the limited applicability of non-learning approaches. Also, these methods propose to learn a safe control policy along with the CBF. As a result, there is no need to solve the CBF-QP (8) during the execution, enabling them for real-time implementation. However, since the learned CBF is a neural network, it is difficult, if not impossible, to verify that the learned CBF candidate satisfies the CBF conditions (6) *everywhere* in the state space. In addition, the forward invariant set $\mathcal{X}_C$ used in training is not easy to find. Moreover, as a centralized approach, it still needs global information and hence, it is not scalable for large-scale MAS (Qin et al., 2020; Zhang, Garg et al., 2023).

The data requirements of learned CBFs vary depending on the amount of dynamics information available. There is some work, albeit limited, on learning for robotics systems with safety constraints using hardware data (Baumann, Marco, Turchetta, & Trimpe, 2021; Cosner et al., 2022; Marco et al., 2021). If the system dynamics are known, then the CBF can be trained entirely in simulation with samples from the safe and unsafe regions. If the dynamics are unknown, then the derivative loss $\mathcal{L}_{deriv}$ in Eq. (11) must be computed using state–action pairs collected on hardware, where samples are usually only available from the safe region. In this case, the safety guarantees provided by the CBF are weakened. If the CBF derivative condition is only satisfied in the safe set, the system is still guaranteed to remain safe when starting in the safe region, but we are no longer guaranteed to converge to the safe set if we start in the unsafe set. Note that $\mathcal{L}_{safe}$ and $\mathcal{L}_{unsafe}$ only require evaluating the learned CBF on a state, and hence, they can be evaluated on unsafe states in simulation without requiring unsafe data.

4.2. Distributed CBF

In order to eliminate the need for global information and address the limited scalability of centralized CBF learning methods, researchers are now focusing on distributed CBF approaches (Meng, Qin, & Fan, 2021; Qin et al., 2020; Yu, Yu et al., 2023; Zhang, Garg et al., 2023). These methods assume that agents have access only to local observations and communication within their immediate vicinity. Most of the works suppose that the agents are identical and share the same CBF and control policy. Thus, they use NNs to parameterize *one* CBF B_{θ} and *one* control policy π_{ϕ} that can be used for each agent (Qin et al., 2020; Yu, Yu et al., 2023; Zhang, Garg et al., 2023), or each pair of agents (Meng et al., 2021), and use a similar procedure as discussed in Section 4.1 for learning them. Different from learning centralized CBF works, distributed CBF learning works often adopt the centralized training and distributed execution (CTDE) framework. During centralized training, the agents are trained jointly and the loss of agent i's CBF can be backpropagated to its neighbors $j \in \mathcal{N}_i$ so that the distributed CBF conditions (9) are satisfied for all the agents (Zhang, Garg et al., 2023). During distributed execution, the agents apply the learned controller which only uses the local observations to obtain control inputs.

In contrast to the centralized CBF, where the central computation node has global observation, each agent only has local observation in distributed CBF. Different methods have been used to encode the local observation. For example, Qin et al. (2020) uses a PointNet (Qi, Su, Mo, & Guibas, 2017) so that the observation encoding is permutation-invariant to the observed agents. The recent work (Zhang, Garg et al., 2023) proposes to use graph neural networks (GNNs) with attention mechanism (Li, Gu, Dullien, Vinyals, & Kohli, 2019) to encode the local observation. It utilizes the fact that GNNs with attention can handle a changing number of neighbors. The attention mechanism also addresses the problem of abrupt changes in the CBF value when an agent enters or leaves the sensing region of another agent. For image-based observations, convolutional neural networks (CNNs) and Long Short-Term Memory (LSTM) are used for encoding (Cui et al., 2022).

Since learning a CBF for a large-scale MAS requires sampling from a large state space, it is not computationally tractable to explore the complete state space during learning and hence, the safety rate during execution is generally very low. To this end, an online policy refinement technique is applied during execution to obtain better safety performance (Qin et al., 2020; Zhang, Garg et al., 2023). The online policy refinement step adds a runtime gradient descent process to update the NN controller during execution. At any time step, if the learned control input does not satisfy the CBF descent condition, then the residue $\delta = [-\dot{B} - \alpha(B)]^+$ is computed, and gradient descent is used to update the learned control to minimize this residue. This is a distributed approach since computing $\dot{B}$ needs the knowledge of the control inputs of the neighboring agents also.

Trained with a few tens of agents, the learned distributed CBF has demonstrated impressive generalizability in very large-scale systems constituting thousands of agents (Qin et al., 2020; Zhang, Garg et al., 2023). These approaches are also capable of using realistic and noisy LiDAR observation (Zhang, Garg et al., 2023) or image-based pixels (Cui et al., 2022), instead of assuming that the *actual* relative states are available. Moreover, some approaches also consider contracts among agents in MAS, e.g., agents have different responsibilities for avoiding collisions, and learn this contract (Cosner et al., 2023). While these methods have several advantages as listed above, the learned CBF is hard to verify for correctness, and cannot provide formal guarantees on the safety of the MAS. Another challenge for these myopic distributed methods is deadlocks (Cohen & Belta, 2020; Reis, Aguiar, & Tabuada, 2020), thereby compromising on liveness properties. In particular, as proved in Reis et al. (2020), Lyapunov-CBF-QP methods induce undesirable equilibrium points which lead to deadlocks in MAS, where none of the agents can move with the CBF-based policy without violating the safety of the MAS.

5. Safe multi-agent reinforcement learning

Instead of delegating satisfaction of safety constraints to shielding-based methods or a learned CBF, one can also directly learn a policy to satisfy the safety constraints. One popular method of learning such a control policy is reinforcement learning (RL). In recent years, many of RL's biggest successes have been in its ability to play multi-agent games ranging from two-agent zero-sum games such as Go (Silver et al., 2016, 2017) and Shogi (Schrittwieser et al., 2020), multi-agent zero-sum games such as Poker and Siplomacy, and team-based games such as Starcraft (Vinyals et al., 2019), Dota (Berner et al., 2019), Honor of Kings (Ye et al., 2020) and Football (Liu, Halev et al., 2021).

Despite these successes, there have been relatively few works that explicitly examine safety in Multi-Agent RL (MARL). The numerous approaches to both single-agent and multi-agent reinforcement learning address safety specifications by either terminating the episode when the safety specification is not met (e.g., in Brockman et al. (2016)) or by including reward terms that discourage the *unsafe* behavior (e.g., in Brockman et al., 2016; Hwangbo et al., 2019; Tassa et al., 2018). While these approaches do not provide safety guarantees and can require reward function tuning, they are more popular in comparison to methods that explicitly address safety constraints.

Flavors of safety in RL In the single-agent RL setting, a Constrained Markov Decision Process (CMDP) (Altman, 1999) is the most popular problem formulation that additionally captures safety constraints. A MDP (Markov Decision Process) is defined as the tuple $(\mathcal{X}, \mathcal{U}, \mathbb{P}, r, \rho_0, \gamma)$, where $\mathbb{P}$ denotes the state transition probability, $r : \mathcal{X} \times \mathcal{U} \times \mathcal{X} \to \mathbb{R}$ denotes the reward function, ρ_0 denotes the starting state distribution, and γ denotes the discount factor (Puterman, 2014). In the context of MAS, the reward function may encode liveness properties as in 2.2.2, such as goal reaching (Chen, Liu, Everett and How, 2017; ElSayed-Aly et al., 2021; Zhang et al., 2019) A CMDP $(\mathcal{X}, \mathcal{U}, \mathbb{P}, r, \rho_0, \gamma, C, d)$ extends a MDP by also considering a constraint function $C : \mathcal{X} \to \mathbb{R}$ and constraint bound $d \in \mathbb{R}$. A CMDP seeks a policy π that satisfies the *average cost* constraints

$$\mathbb{E}_\pi \left[\sum_{k=0}^{\infty} \gamma^k C(x_k) \right] \leq d. \tag{15}$$

While average cost constraints (15) are different from the safety constraints in Definition 1, the average cost constraints can be viewed as *chance constraints* under the *discounted* occupation measure q_π of π (Altman, 1999, Chapter 3), where C is taken to be an indicator function $x \mapsto \mathbb{1}\{x \notin \mathcal{S}\}$ and $d \in [0, 1]$ denotes the probability threshold, since

$$\mathbb{E}_\pi \left[\sum_{k=0}^{\infty} \gamma^k \mathbb{1}\{x_k \notin \mathcal{S}\} \right] = \mathbb{E}_{q_\pi} \left[\mathbb{1}\{x \notin \mathcal{S}\} \right] = \Pr(x \notin \mathcal{S}). \tag{16}$$

On the other hand, for general choices of C and d, this becomes a different notion of safety than Definition 1, and the optimal solution of the CMDP could result in exiting the safe set $\mathcal{S}$. Single-agent RL methods that explicitly solve the CMDP label themselves as constrained RL or safe RL (e.g., Altman, 1999; Gu et al., 2022; Yu, Ma, Li, & Chen, 2022) and are commonly tested on benchmarks such as safety gym (Ray, Achiam, & Amodei, 2019).

Another notion of safety is when no constraint violation is allowed at any *any* time-step k, i.e.,

$$\max_{k \geq 0} C(x_k) \leq d \tag{17}$$

This is known as a peak constraint (Geibel, 2006; Geibel & Wysotzki, 2005) or state-wise safety (Zhao, He, Chen, Wei, & Liu, 2023) in the Safe RL literature, and aligns with the notion of safety considered in this paper, where the safe set $\mathcal{S}$ is defined as

$$\mathcal{S} := \{x \in \mathcal{X} \mid C(x) \leq d\}. \tag{18}$$

While peak constraints are not as common for constrained RL methods, some works tackle this problem (Fisac et al., 2019; So & Fan, 2023; Yu et al., 2022) (see Zhao et al. (2023) for a recent survey of some methods). It can be related to the average cost case (15) by taking $\tilde{C}_k := \max(0, C_k - d)$ and $\tilde{d} = 0$ (So & Fan, 2023), yielding the constraints $\tilde{C}_k \leq \tilde{d}$ given as

$$\sum_{k=0}^{\infty} \gamma^k \max(0, C_k - d) \leq 0. \tag{19}$$

For both types of constraints, taking smaller values of d results in a stricter constraint on the total accrued cost and hence a lower total reward, while higher values of d loosen the constraint and allow higher rewards.

Safe MARL Given that solving the CMDP problem and obtaining policies that successfully satisfy the safety constraints is already challenging in the single-agent case, it is even more challenging in the MAS setting. Consequently, there have been relatively fewer works that tackle the problem of safe MARL. This is noted in many surveys on MARL, which mention safe MARL to be a direction that is relatively unexplored (Gu et al., 2022; Oroojlooy & Hajinezhad, 2023; Yang & Wang, 2020; Zhang et al., 2021c). Nevertheless, recent years have seen an increasing number of works that tackle this challenging problem. We provide an overview of existing methods in Fig. 5, which we describe in more detail in the coming subsections.

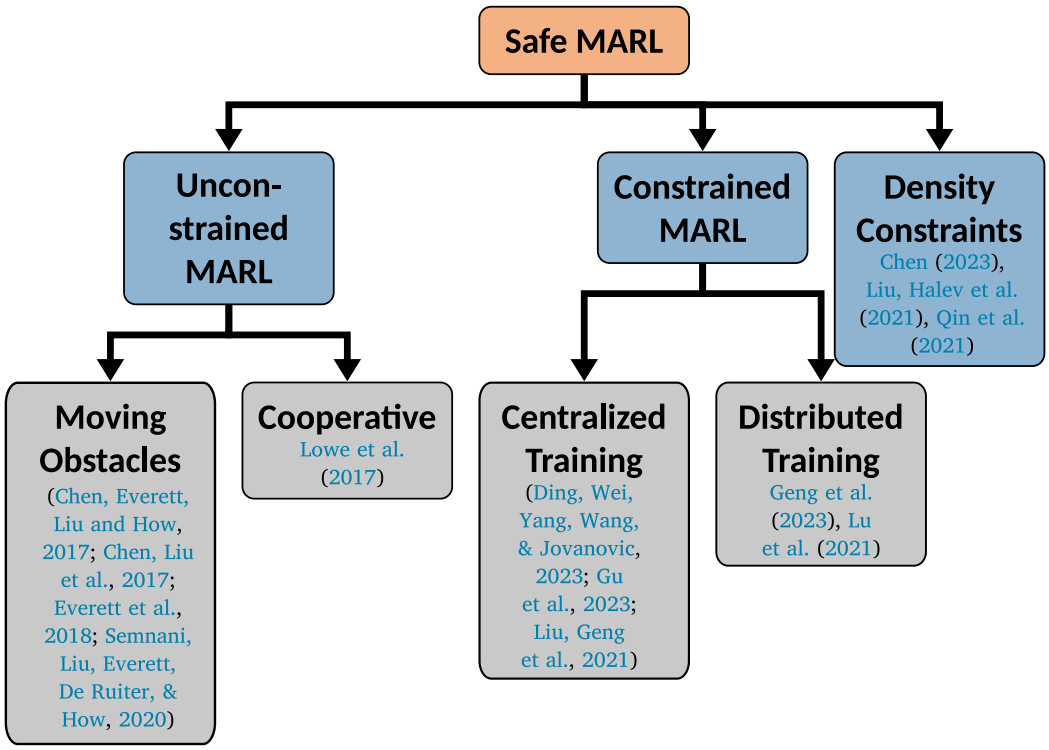

Fig. 5. Overview of Safe MARL-based approaches (Chen, 2023; Geng et al., 2023; Lowe et al., 2017; Qin, Chen, & Fan, 2021).

5.1. Unconstrained MARL

Early works that approached the problem of safety for MARL focused on navigation problems and collision avoidance (Chen, Everett, Liu and How, 2017; Chen, Liu et al., 2017; Everett et al., 2018; Semnani, Liu, Everett, De Ruiter, & How, 2020), where safety is achieved by either a sparse collision penalty (Long et al., 2018), or a shaped reward term that penalizes small distances to obstacles and neighboring agents (Chen, Everett et al., 2017; Chen, Liu et al., 2017; Everett et al., 2018; Semnani et al., 2020). However, the satisfaction of collision avoidance constraints is not necessarily guaranteed by either the final policy or even the optimal policy (Massiani, Heim, Solowjow, & Trimpe, 2023). Consequently, while these methods report 100% safety rates empirically when there are fewer agents, safety violations occur when tested with a larger number of agents (e.g., >3% for 8-agents in Everett et al. (2018), >3% for 20-agents in Long et al. (2018)). These trends are also similar for liveness properties. In Everett et al. (2018), although the use of reward to encourage goal-reaching is sufficient for all agents to eventually reach their goal when there are less than 3 agents, the percentage of deadlocks increases to 1.6% for 10 agents. Improving satisfaction of the liveness properties may require alternate approaches such as imitation learning of Multi-Agent Path Finding (MAPF) algorithms (Damani, Luo, Wenzel, & Sartoretti, 2021; Sartoretti et al., 2019), where success rates are close to 100% even for up to 128 agents.

While there exists some finite scale for the penalty such that the optimal policy is guaranteed to satisfy the constraints (Massiani et al., 2023), too large of a penalty term often results in poorer performance empirically (Shalev-Shwartz, Shammah, & Shashua, 2016). As noted in Shalev-Shwartz et al. (2016), this can be explained by larger penalties resulting in larger variances in the reward which ultimately results in poorer optimization performance.

5.2. Constrained MARL

In contrast to unconstrained MARL methods that penalize safety violations in the reward term and then solve the resulting unconstrained problem, constrained MARL methods explicitly solve the CMDP problem. For the single-agent case, prominent methods for solving CMDPs include primal–dual methods using Lagrange multipliers (Borkar, 2005; Tessler, Mankowitz, & Mannor, 2019) and via trust-region-based approaches (Achiam, Held, Tamar, & Abbeel, 2017). These methods provide guarantees either in the form of asymptotic convergence guarantees to the optimal (safe) solution (Borkar, 2005; Tessler et al., 2019) using stochastic approximation theory (Borkar, 2009; Robbins & Monro, 1951), or recursive feasibility of intermediate policies (Achiam et al., 2017; Satija, Amortila, & Pineau, 2020) using ideas from trust region optimization (Schulman, Levine, Abbeel, Jordan, & Moritz, 2015). The survey (Gu et al., 2022) provides an in-depth overview of the different methods of solving safety-constrained single-agent RL.

MARL algorithms can be broadly divided into two paradigms: centralized and distributed training algorithms (Gronauer & Diepold, 2022). The same holds for their extensions to consider safety constraints.

5.2.1. Centralized training

During centralized training, agent policies are updated at a central node, where information in addition to each agent's local observations is used to update each agent's policies. This is the dominant paradigm for unconstrained MARL (Gronauer & Diepold, 2022) and is referred to as Centralized Training Decentralized Execution (CTDE).

One of the first safe CTDE methods to be proposed is in Gu et al. (2023), where the authors combine Constrained Policy Optimization (CPO) (Achiam et al., 2017), a method for solving CMDPs, with Heterogeneous-Agent Trust Region Policy Optimisation (HATRPO), a MARL method that enjoys a theoretically-justified monotonic improvement guarantee (Kuba et al., 2022). Theoretical analysis guarantees monotonic improvement in reward while theoretically satisfying safety constraints during each iteration, assuming that the initial policy is feasible, the value functions are known and a trust-region optimization problem can be solved exactly. However, neither of these assumptions is guaranteed in the method implemented in practice due to approximation errors in the value function and a quadratic approximation of the trust-region problem (Gu et al., 2023; Kuba et al., 2022).

Another early safe CTDE method is CMIX (Liu, Geng et al., 2021), which extends the value function factorization method QMIX (Rashid et al., 2020) to additionally consider both average constraints and peak constraints. However, no theoretical analysis of the convergence of the proposed algorithm is given (Ding, Wei, Yang, Wang, & Jovanovic, 2023; Gu et al., 2023; Liu, Geng et al., 2021).

5.2.2. Distributed training

In distributed training methods, each agent has a private reward function, policy updates occur locally for each agent, and communication between agents is used to arrive at a policy that minimizes the total reward subject to safety constraints (Lu et al., 2021). In this distributed setting where the reward and constraints are private and each agent has a different policy when the algorithm has not converged, it is not possible to evaluate the performance of each agent's policy. Consequently, these approaches are based on primal–dual optimization and provide asymptotic convergence guarantees to local optima (Lu et al., 2021), but are unable to provide any safety guarantees before convergence.

5.3. Density-based approaches

Finally, a vastly different approach to safe MARL looks at the case when the agents are indistinguishable, the number of agents is very large and applies the mean-field approximation, where they look at the limit as the number of agents goes to infinity, and the quantity of interest is instead the density of the overall swarm of agents (Bensoussan, Frehse, Yam, et al., 2013; Lasry & Lions, 2007). As the concept of individual agents is gone, both inter-agent and agent-obstacle constraints are replaced by density constraints (Lin, Fung, Li, Nurbekyan, & Osher, 2021; Liu, Chen, So, & Theodorou, 2022). Navigation problems when applying the mean-field approximation can be viewed as an optimal transport problem (Liu et al., 2022). When also considering safety constraints via state-dependent costs, the problem becomes a mean field game with distributional boundary constraints that can be solved using RL techniques (Liu et al., 2022).

For example, Lin et al. (2021) and Liu et al. (2022) consider a navigation problem with obstacles, where the density of the multi-agent swarm at the obstacles is constrained to be zero for a given initial and final distribution of agents. In Liu et al. (2022), a penalty on the density of agents is also included to incentivize agents to spread out more, and consequently reduce the risk of inter-agent collisions.

6. Safety verification for learning-enabled MAS

Once a control policy has been learned, it must be checked for correctness before it can be deployed, particularly in safety-critical control contexts. This is particularly true for control policies represented using difficult-to-interpret models such as NNs. This section reviews methods for checking the safety properties of a learned controller or shield for MAS. Broadly speaking, these methods can be organized as shown in Fig. 6, where the primary trade-off is between the ability to provide formal guarantees and the ability to scale to practically-sized problems. Fig. 6 also highlights the specific challenges of MAS verification and summarizes existing approaches to resolving these challenges, as well as open problems regarding the limitations of these methods.

We begin by reviewing the methods from Fig. 6 with reference to the extensive literature for single-agent verification, before highlighting the specific challenges that arise in the multi-agent setting. We then review methods for addressing each of these challenges, concluding with a discussion of open problems.

6.1. Review of single-agent verification tools

Several excellent surveys for centralized verification provide a good starting point for readers interested in a broad introduction to the field; we will review these surveys here, and then devote most of our attention to this survey to the challenges that distinguish multi-agent verification from the centralized case.

The survey (Corso, Moss, Koren, Lee, & Kochenderfer, 2021) covers search- and optimization-based methods for black-box autonomous systems. These methods can sometimes provide formal guarantees, depending on the completeness of the underlying optimizer or search algorithm; for example, some stochastic global optimization methods enjoy asymptotic completeness guarantees under certain assumptions, which may be used to provide formal guarantees, but most black-box optimization methods are best suited for empirical testing. In this survey, we discuss the challenges in scaling these methods to large-scale MAS, as well as how search- and optimization-based methods can be extended to consider issues that are specific to MAS, such as communication noise and cybersecurity.

Centralized reachability analysis using Hamilton–Jacobi (HJ) methods are covered in the survey (Bansal et al., 2017), which are specialized for checking set invariance properties but can provide formal guarantees. The survey (Bansal et al., 2017) discusses centralized verification of MAS (i.e. collapsing the MAS into a single dynamical system) and verification of pairwise safety (i.e. checking for inter-agent collision avoidance), but they do not consider MAS safety more generally. Other methods of centralized reachability analysis (e.g., set propagation), have been extensively covered in previous surveys (Althoff, Frehse, & Girard, 2021; Alur, 2011; Chen & Sankaranarayanan, 2022). These methods use sets to over-approximate the reachable set of a dynamic system and can be used to provide safety guarantees. However, these surveys do not focus on the topics of learned controllers or the reachability of MAS. We restrict our discussion to decomposition-based approaches to scaling HJ methods, along with other reachability analysis tools, for MAS.

Another recent review article (Dawson et al., 2023) provides a survey of neural certificates, but they focus largely on the centralized case. We extend the discussion of certificates to consider decentralized, distributed, and compositional approaches checking to certificates. Due to the extensive coverage of certificate learning in Section 4, in this section we only discuss the challenges involved in checking the validity of these certificates in the multi-agent case.

In Liu, Arnon et al. (2021), the authors provide a survey of techniques for verifying input–output properties of neural networks, some of which have been applied to verify the soundness of shields like barrier and Lyapunov functions (Abate, Ahmed, Edwards, Giacobbe, & Peruffo, 2021; Dai, Landry, Yang, Pavone, & Tedrake, 2021). Due to the extreme fundamental complexity of neural network verification (an NP-complete problem (Liu, Arnon et al., 2021)) and several open technical challenges that we discuss in the sequel, these methods have yet to be applied to MAS.

Similarly, other surveys cover verification for safe learning and control (Brunke et al., 2022) and autonomous aerial systems (Brat et al., 2023); while these provide a good overview of their respective fields, they do not specifically address MAS.

6.2. Challenges of MAS verification

While many of the concepts from centralized verification extend to the multi-agent case, several challenges are unique to MAS. We discuss these challenges here, and the rest of this section is devoted to reviewing proposed methods for addressing these issues.

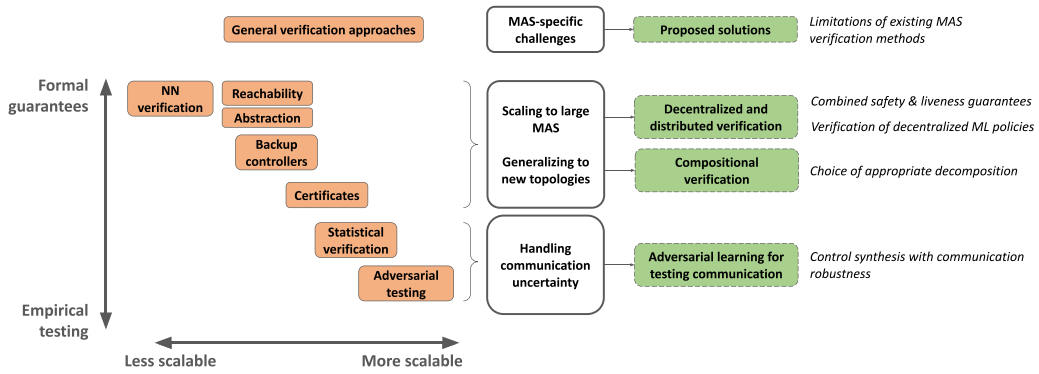

Fig. 6. An overview of verification methods, comparing the ability to provide formal guarantees with the ability to scale in practice.

6.2.1. Scalability to high dimensions

The most immediate challenge in verifying learned control systems for MAS is the scale. Formal verification using methods like HJ (Bansal et al., 2017) and abstraction (Alur, Henzinger, Lafferriere, & Pappas, 2000) are notoriously difficult to scale to systems with large state dimensions, so verifying a MAS by naively concatenating the states of individual agents to form a centralized verification problem is generally not tractable. Similarly, neural network verification is generally intractable for networks with more than a few hundred neurons (Liu, Arnon et al., 2021), so the formal guarantees that might be derived using these methods are hard to apply in practice without additional decomposition of the problem.

Even for empirical testing methods, such as those relying on stochastic optimization, it can be challenging to efficiently explore high-dimensional space. The challenge of scalability has led to a range of specialized techniques for MAS and other large-scale systems, including decentralized certificates, distributed search and optimization, and compositional verification methods, which we discuss in the following section.

6.2.2. Handling changing system topology

The second specific challenge for verifying MAS is that, unlike single-agent systems, MAS can be dynamically reconfigured at runtime. For example, the connectivity of agents might change during execution, or the exact number of agents in the system might be uncertain at test time. As a result, to be most useful for MAS, verification methods must be robust to both the number and topology of agents in the system; in particular, formal guarantees should generalize to different system topology, and empirical testing should achieve sufficient coverage of the range of topology that we expect to see at runtime. In the following, we survey several methods that meet this requirement, such as compositional approaches and adversarial testing, and we identify several important open questions in this area.

6.2.3. Handling communication delay and error

An additional unique feature of MAS verification is that there is a class of disturbances — those affecting inter-agent communication — that are typically not present in single-agent settings, and thus typically not considered by single-agent verification methods (Brunke et al., 2022; Corso et al., 2021). Communication disturbances come in many forms, including both naturally occurring factors like communication delay and noisy or lossy communication and adversarial effects like non-cooperative or malicious agents that can send arbitrary messages to other agents (this latter category is important in considering the cybersecurity of MAS).

6.3. Decentralized and distributed verification

To address the first two challenges (scalability and generalizing to changing system configurations), a natural question is how centralized verification algorithms can be adapted to the distributed or decentralized setting. Decentralized (Qin et al., 2020; Wang et al., 2017) and distributed (Zhang, Garg et al., 2023) barrier certificates are one example of this approach, which we discuss extensively in Sections 3 and 4. Other examples include decentralized reachability analysis, e.g. only considering pairwise interactions (Bansal et al., 2017; Wang, Leung and Pavone, 2020), and distributed optimization to guarantee safety via a multi-agent safe model-predictive control (Muntwiler et al., 2020).

There are several important considerations for decentralized verification. First, it is not possible to verify some MAS safety and liveness properties in a fully decentralized setting. For example, obstacle avoidance (involving only individual agents and the environment) can be verified in a decentralized manner, but inter-agent collision avoidance requires considering pairs or small cliques of agents (Bansal et al., 2017; Qin et al., 2020; Zhang, Garg et al., 2023). Other properties are more nuanced; for example, connectivity maintenance can be verified using only pairwise communication if the communication graph topology is fixed (this results in a pairwise maximum distance constraint), but if the topology is allowed to vary then agents must be able to compute eigenvalues of the communication graph Laplacian in a distributed manner (Cavorsi, Capelli, Sabattini, & Gil, 2022).

A second consideration concerns the common strategy of guaranteeing safety by constructing a *safety filter* using either barrier certificates (Qin et al., 2020; Wang et al., 2017; Zhang, Garg et al., 2023) or reachability analysis (Muntwiler et al., 2020), as discussed in Section 3. This safety filter limits the range of actions that individual agents may take and often provides formal safety guarantees, but this limit often reduces the optimality of the filtered policy, and this reduced optimality can be more severe for decentralized safety filters. For example, Chen, Shih, and Tomlin (2016) shows that a centralized safe controller achieves a higher task completion rate than a controller that only considers pairwise interactions.

Finally, the application of existing neural network verification (NNV) tools to decentralized learned control policies or certificates remains an open problem; even if more performant NNV tools are developed that can scale to large policy networks, existing NNV algorithms cannot handle architectures like graph neural networks (GNNs) commonly used in learning for decentralized and distributed control (Sälzer & Lange, 2023).

6.4. Compositional verification methods

An alternative to the distributed and decentralized methods discussed above is a more scalable form of centralized verification where agent-level guarantees are hierarchically composed to yield guarantees for the entire MAS; this class of approach is known as *compositional*

verification. Compositional certificates have been used to prove MAS stability properties by composing individual-agent stability certificates that are either learned (Zhang, Xiu, Qu and Fan, 2023) or found using sum-of-squares optimization (Shen & Tedrake, 2018). The benefit of compositional approaches to certificate learning is that they require much less data to train; they can be trained at the individual agent level and then generalized to a large number of agents to provide system-level guarantees (Zhang, Xiu et al., 2023).

A similar compositional approach has been applied to reachability-based verification, for example using reachability to certify pairwise collision avoidance (Bansal et al., 2017), which is composed to give system-level inter-agent collision avoidance guarantees under certain assumptions about the sparsity of agents, or composing reachable sets for subsystems to certify properties of networked systems (Althoff, 2014). It is important to note that decomposition of system-level properties to the agent level necessarily constrains the space of safe control policies, potentially introducing conservatism; e.g. Chen et al. (2016) show that composing pairwise collision-free policies leads to more conservative behavior than a more general N-agent decomposition. An important open question in compositional verification is understanding how the choice of decomposition affects the conservatism (or even feasibility) of the resulting control strategy.

6.5. Handling communication uncertainty

A unique feature of MAS (compared to centralized systems) is that they often rely on communication links between agents to function. In a fully decentralized setting, agents might only require local observations of neighbors (e.g. relative position) without explicit communication, but distributed systems can rely on multiple rounds of message passing to achieve consensus or accomplish a distributed optimization task (Espina et al., 2020; Molzahn et al., 2017). If this communication is disrupted, or if adversarial agents inject malicious communication packets, then the safety of the overall MAS can be compromised.

Several surveys consider the problem of communication uncertainty and cybersecurity for MAS; common strategies for ensuring robustness to communication issues include fault detection and isolation (FDI), secure consensus algorithms (e.g. where a non-majority subset of agents can be compromised while still maintaining the integrity of the non-compromised agents (Zhang, Feng, Shi and Srinivasan, 2021)), and safety filters that maintain sufficient connectivity of the communication graph (Cavorsi et al., 2022). Here, we review how these security methods can be extended using learning-based methods. For example, Garg, Dawson, Xu, Ornik, and Fan (2023) learns a model-free FDI policy for single-agent systems with multiple faults; this method can be extended for model-free learning-based FDI in MAS. Other works in this vein include using representation learning to classify the network robustness of MAS (Wang et al., 2018) or using RL to detect intrusions in a networked system (Servin & Kudenko, 2008) or generate adversarial attacks on MAS (Yamagata, Liu, Akazaki, Duan, & Hao, 2021).

A related communication issue arising in MAS is delay, which in the worst case can prevent the system from achieving consensus or even destabilize the system (Papachristodoulou, Jadbabaie, & Münz, 2010). Several methods have been proposed to handle communication effects, including delay and uncertainty, in the reinforcement learning literature (Das et al., 2019; Jiang & Lu, 2018; Kim, Park, & Sung, 2021; Wang, He et al., 2020). However, verifying (either formally or empirically) the robustness of a learned controller to these effects remains a challenging open problem.

7. Open problems

Given the state of the art reviewed in this survey, a few themes stand out as areas for future work. The rest of this section discusses these themes.

7.1. Combined safety and liveness guarantees

While learning-based methods for MAS have seen immense progress in the past decade, much work is still needed when it comes to safety, provable guarantees, and scalability. In the context of learning, there exists a trade-off between liveness properties, such as goal reaching, and safety properties, such as collision avoidance. It is still an open problem as to how to achieve high safety rates or provable safety guarantees along with high performance or guarantees on deadlock resolutions, especially in partially observable systems and while applying distributed algorithms (Qin et al., 2020; Zhang, Garg et al., 2023).

7.2. Decomposition

For any single-agent method, one of the main challenges for their extension to MAS is how to perform a suitable decomposition that balances the performance and scalability of the method.

Appropriate choice of decomposition for shielding. For shielding-based approaches, this decomposition has been done via distributed computation techniques to efficiently compute a centralized shielding function (Pereira et al., 2022), factoring the state space (ElSayed-Aly et al., 2021; Xiao, Lyu et al., 2023), distributed decomposition of the safety constraints (i.e., Zhang et al. (2019)), or performing a completely decentralized factorization of the shielding function such that the communication is not needed (Cai et al., 2021; Melcer et al., 2022). A major open question with these approaches is to explore how the type of decomposition affects the conservatism of the resulting safety filter and its ability to scale and how different choices of shield type (e.g., PSF, CBF,...) can change this tradeoff between ease of construction, scalability, and conservatism.

Appropriate choice of decomposition for verification. Similar to shielding methods, verification methods also often rely on decomposition. An important open problem discussed in Section 6 is how this choice of decomposition affects the conservatism and completeness of the resulting verification scheme. Furthermore, care must be taken to ensure that the decomposition is valid; for example, if a system is only analyzed pairwise for collision avoidance, how does the verification method account for conflicts between more than two agents?

7.3. Practical issues

Safety under communication uncertainty. Another challenge in designing safe methods for MAS is handling dynamically changing MAS configuration, communication delays, package losses, and adversarial communication. Designing control synthesis and verification tools that are both scalable and robust to time-varying topology as well as communication-related issues remains an open problem, as many existing methods either ignore communication effects or rely on domain-specific information.

Strict requirements of CBF-QP. While using CBF for shielding, a crucial practical issue is the feasibility of the CBF-QP. Real-world systems often have input constraints, e.g., torque limit, acceleration limit, etc. The CBF-QP may become infeasible when input constraints are included. Guaranteeing the feasibility of the CBF-QP is an open problem (Agrawal & Panagou, 2021; Chen et al., 2020; Cortez, Tan, & Dimarogonas, 2021), although it is likely that future work will address these issues.

Safe methods are complex and unpopular in practice. When it comes to RL, one of the main challenges for safety in RL in both the single-agent and multi-agent cases is the tradeoff between the simplicity of the algorithm, practical performance, and safety guarantees. With unconstrained RL, in general, neither the resulting policy after training nor the theoretically optimal policy is guaranteed to satisfy the safety constraints (Massiani et al., 2023; Tasse, Love, Nemecek, James, & Rosman, 2023). On the other hand, while some constrained RL methods

have convergence and safety guarantees (e.g., Gu et al. (2021)), these methods have more components and thus are more difficult to implement and use in practice as compared to unconstrained variants. As a result, these safe methods are relatively unpopular among practitioners compared to unconstrained methods.

Value of methods without practical safety guarantees. Another challenge for safety in MAS is the question of whether learned CBFs or safe MARL methods should be used instead of shielding-based methods when safety guarantees are desired. Although constrained RL can provide per-iteration safety guarantees, this assumes both an initially feasible policy and access to the true value function, neither of which is guaranteed to hold in practice (Gu et al., 2023). Similarly, a learned CBF, if verified to be an actual CBF, inherits the theoretical safety guarantees. However, the training process for learning CBFs does not guarantee that this will always be true. On the other hand, shielding-based methods shift this complexity from the algorithm to the user, who must first construct a valid shielding function to provide provable safety guarantees during both training and deployment. It is not clear whether this tradeoff between scalability and practically relevant safety guarantees will be attractive for safety-critical applications.

8. Conclusions

MAS are ubiquitous in today's world, with potential for applications ranging from robotics to power systems, and there is a large body of literature on control of MAS. However, the safe control design of large-scale robotic MAS is a challenging problem. In this survey, we have reviewed how various learning-based methods have shown promising results in addressing some of the aspects of safe control of MAS, such as the safety guarantees of shielding-based methods, the generalizability of learning CBF methods, and the wide applicability of MARL-based methods. Despite their advantages, no existing method has all the desired properties of being provably safe, scalable, computationally tractable, and implementable to a variety of MAS problems. While certificate learning and safe MARL-based methods can provide theoretical safety guarantees, these guarantees do not hold in practice due to unrealizable assumptions. We have identified a range of open problems covering these concerns, and we hope that our review of the state-of-the-art in this field provides a springboard for further research to address these issues and realize the full potential of safe learning-based control for MAS.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

Funding

This work was supported by the National Science Foundation (NSF), USA CAREER Award #CCF-2238030 and Air Force Office of Scientific Research (AFOSR, USA grant FA9550-23-1-0099.

References

Abate, Alessandro, Ahmed, Daniele, Edwards, Alec, Giacobbe, Mirco, & Peruffo, Andrea (2021). FOSSIL: A software tool for the formal synthesis of Lyapunov functions and barrier certificates using neural networks. In *Proceedings of the 24th international conference on hybrid systems: computation and control*. New York, NY, USA: Association for Computing Machinery.

Achiam, Joshua, Held, David, Tamar, Aviv, & Abbeel, Pieter (2017). Constrained policy optimization. In *International conference on machine learning* (pp. 22–31). PMLR.

Adaldo, Antonio, Liuzza, Davide, Dimarogonas, Dimos V, & Johansson, Karl H (2016). Multi-agent trajectory tracking with self-triggered cloud access. In *2016 IEEE 55th conference on decision and control* (pp. 2207–2214). IEEE.

Agrawal, Devansh R., & Panagou, Dimitra (2021). Safe control synthesis via input constrained control barrier functions. In *2021 60th IEEE conference on decision and control* (pp. 6113–6118). IEEE.

Ahmadi, Amir Ali, & Majumdar, Anirudha (2016). Some applications of polynomial optimization in operations research and real-time decision making. *Optimization Letters*, *10*, 709–729.

Alshiekh, Mohammed, Bloem, Roderick, Ehlers, Rüdiger, Könighofer, Bettina, Niekum, Scott, & Topcu, Ufuk (2018). Safe reinforcement learning via shielding. *Vol. 32*, In *Proceedings of the AAAI conference on artificial intelligence*.

Althoff, Matthias (2014). Formal and compositional analysis of power systems using reachable sets. *IEEE Transactions on Power Systems*, *29*(5), 2270–2280.

Althoff, Matthias, Frehse, Goran, & Girard, Antoine (2021). Set propagation techniques for reachability analysis. *Annual Review of Control, Robotics, and Autonomous Systems*, *4*, 369–395.

Altman, E. (1999). Constrained Markov decision processes. In *Stochastic modeling series*, Taylor & Francis, ISBN: 9780849303821.

Alur, Rajeev (2011). Formal verification of hybrid systems. In *Proceedings of the ninth ACM international conference on embedded software* (pp. 273–278).

Alur, R., Henzinger, T. A., Lafferriere, G., & Pappas, G. J. (2000). Discrete abstractions of hybrid systems. *Proceedings of the IEEE*, *88*(7), 971–984.

Ames, Aaron D, Coogan, Samuel, Egerstedt, Magnus, Notomista, Gennaro, Sreenath, Koushil, & Tabuada, Paulo (2019). Control barrier functions: Theory and applications. In *18th European control conference* (pp. 3420–3431). IEEE.

Ames, Aaron D, Galloway, Kevin, Sreenath, Koushil, & Grizzle, Jessy W (2014). Rapidly exponentially stabilizing control Lyapunov functions and hybrid zero dynamics. *IEEE Transactions on Automatic Control*, *59*(4), 876–891.

Ames, Aaron D., Grizzle, Jessy W., & Tabuada, Paulo (2014). Control barrier function based quadratic programs with application to adaptive cruise control. In *53rd IEEE conference on decision and control* (pp. 6271–6278). IEEE.

Ames, Aaron D, Xu, Xiangru, Grizzle, Jessy W, & Tabuada, Paulo (2016). Control barrier function based quadratic programs for safety critical systems. *IEEE Transactions on Automatic Control*, *62*(8), 3861–3876.

Atınç, Gökhan M, Stipanović, Dušan M, & Voulgaris, Petros G (2020). A swarm-based approach to dynamic coverage control of multi-agent systems. *Automatica*, *112*, Article 108637.

Baier, Christel, & Katoen, Joost-Pieter (2008). *Principles of model checking*. MIT Press.

Bansal, Somil, Chen, Mo, Herbert, Sylvia L., & Tomlin, Claire J. (2017). Hamilton-Jacobi reachability: A brief overview and recent advances. In *2017 IEEE 56th annual conference on decision and control* (pp. 2242–2253).

Bansal, Somil, & Tomlin, Claire J. (2021). DeepReach: A deep learning approach to high-dimensional reachability. In *2021 IEEE international conference on robotics and automation* (pp. 1817–1824). IEEE.

Bastani, Osbert (2021). Safe reinforcement learning with nonlinear dynamics via model predictive shielding. In *2021 American control conference* (pp. 3488–3494). IEEE.

Baumann, Dominik, Marco, Alonso, Turchetta, Matteo, & Trimpe, Sebastian (2021). Gosafe: Globally optimal safe robot learning. In *2021 IEEE international conference on robotics and automation* (pp. 4452–4458). IEEE.

Bensoussan, Alain, Frehse, Jens, Yam, Phillip, et al. (2013). *Mean field games and mean field type control theory*: *vol. 101*, Springer.

Berner, Christopher, Brockman, Greg, Chan, Brooke, Cheung, Vicki, Dębiak, Przemysław, Dennison, Christy, et al. (2019). Dota 2 with large scale deep reinforcement learning. arXiv preprint arXiv:1912.06680.

Blanchini, Franco (1999). Set invariance in control. *Automatica*, *35*(11), 1747–1767.

Bloem, Roderick, Könighofer, Bettina, Könighofer, Robert, & Wang, Chao (2015). Shield synthesis: Runtime enforcement for reactive systems. In *International conference on tools and algorithms for the construction and analysis of systems* (pp. 533–548). Springer.

Borkar, Vivek S. (2005). An actor-critic algorithm for constrained Markov decision processes. *Systems & Control Letters*, *54*(3), 207–213.

Borkar, Vivek S. (2009). *Stochastic approximation: A dynamical systems viewpoint*: *vol. 48*, Springer.

Borrmann, Urs, Wang, Li, Ames, Aaron D., & Egerstedt, Magnus (2015). Control barrier certificates for safe swarm behavior. *IFAC-PapersOnLine*, *48*(27), 68–73.

Brat, Guillaume P, Yu, Huafeng, Atkins, Ella, Sharma, Prashin, Cofer, Darren, Durling, Michael, et al. (2023). *Autonomy verification & validation roadmap and vision 2045*: *Tech. rep. NASA/TM-20230003734*.

Brezis, Haïm (1970). On a characterization of flow-invariant sets. *Communications on Pure and Applied Mathematics*, *23*(2), 261–263.

Brockman, Greg, Cheung, Vicki, Pettersson, Ludwig, Schneider, Jonas, Schulman, John, Tang, Jie, et al. (2016). Openai gym. arXiv preprint arXiv:1606.01540.

Brunke, Lukas, Greeff, Melissa, Hall, Adam W, Yuan, Zhaocong, Zhou, Siqi, Panerati, Jacopo, et al. (2022). Safe learning in robotics: From learning-based control to safe reinforcement learning. *Annual Review of Control, Robotics, and Autonomous Systems*, *5*, 411–444.

Cai, Zhiyuan, Cao, Huanhui, Lu, Wenjie, Zhang, Lin, & Xiong, Hao (2021). Safe multi-agent reinforcement learning through decentralized multiple control barrier functions. arXiv preprint arXiv:2103.12553.

Canese, Lorenzo, Cardarilli, Gian Carlo, Di Nunzio, Luca, Fazzolari, Rocco, Giardino, Daniele, Re, Marco, et al. (2021). Multi-agent reinforcement learning: A review of challenges and applications. *Applied Sciences*, *11*(11), 4948.

Cavorsi, Matthew, Capelli, Beatrice, Sabattini, Lorenzo, & Gil, Stephanie (2022). Multi-robot adversarial resilience using control barrier functions. In *Proceedings of robotics: science and systems*. New York City, NY, USA.

Chen, Yongxin (2023). Density control of interacting agent systems. *IEEE Transactions on Automatic Control*.

Chen, Yu Fan, Everett, Michael, Liu, Miao, & How, Jonathan P. (2017). Socially aware motion planning with deep reinforcement learning. In *2017 IEEE/RSJ international conference on intelligent robots and systems* (pp. 1343–1350). IEEE.

Chen, Mo, Hu, Qie, Mackin, Casey, Fisac, Jaime F., & Tomlin, Claire J. (2015). Safe platooning of unmanned aerial vehicles via reachability. In *2015 54th IEEE conference on decision and control* (pp. 4695–4701).

Chen, Jingkai, Li, Jiaoyang, Fan, Chuchu, & Williams, Brian C. (2021). Scalable and safe multi-agent motion planning with nonlinear dynamics and bounded disturbances. *Vol. 35*, In *Proceedings of the AAAI conference on artificial intelligence* (pp. 11237–11245).

Chen, Yu Fan, Liu, Miao, Everett, Michael, & How, Jonathan P. (2017). Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning. In *2017 IEEE international conference on robotics and automation* (pp. 285–292). IEEE.

Chen, Fei, Ren, Wei, et al. (2019). On the control of multi-agent systems: A survey. *Foundations and Trends® in Systems and Control*, *6*(4), 339–499.

Chen, Xin, & Sankaranarayanan, Sriram (2022). Reachability analysis for cyber-physical systems: Are we there yet? In *NASA formal methods symposium* (pp. 109–130). Springer.

Chen, Mo, Shih, Jennifer C., & Tomlin, Claire J. (2016). Multi-vehicle collision avoidance via Hamilton-Jacobi reachability and mixed integer programming. In *2016 IEEE 55th conference on decision and control* (pp. 1695–1700).

Chen, Yuxiao, Singletary, Andrew, & Ames, Aaron D. (2020). Guaranteed obstacle avoidance for multi-robot operations with limited actuation: A control barrier function approach. *IEEE Control Systems Letters*, *5*(1), 127–132.

Choi, Jason J., Lee, Donggun, Sreenath, Koushil, Tomlin, Claire J., & Herbert, Sylvia L. (2021). Robust control barrier–Value functions for safety-critical control. In *2021 60th IEEE conference on decision and control* (pp. 6814–6821). IEEE Press.

Clark, Andrew (2021). Verification and synthesis of control barrier functions. In *2021 60th IEEE conference on decision and control* (pp. 6105–6112). IEEE.

Cohen, Max H., & Belta, Calin (2020). Approximate optimal control for safety-critical systems with control barrier functions. In *2020 59th IEEE conference on decision and control* (pp. 2062–2067). IEEE.

Conte, Christian, Zeilinger, Melanie N, Morari, Manfred, & Jones, Colin N (2013). Robust distributed model predictive control of linear systems. In *2013 European control conference* (pp. 2764–2769). IEEE.

Corso, Anthony, Moss, Robert, Koren, Mark, Lee, Ritchie, & Kochenderfer, Mykel (2021). A survey of algorithms for black-box safety validation of cyber-physical systems. *Journal of Artificial Intelligence Research*, *72*, 377–428.

Cortes, Jorge, Martinez, Sonia, Karatas, Timur, & Bullo, Francesco (2004). Coverage control for mobile sensing networks. *IEEE Transactions on Robotics and Automation*, *20*(2), 243–255.

Cortez, Wenceslao Shaw, Tan, Xiao, & Dimarogonas, Dimos V. (2021). A robust, multiple control barrier function framework for input constrained systems. *IEEE Control Systems Letters*, *6*, 1742–1747.

Cosner, Ryan K., Chen, Yuxiao, Leung, Karen, & Pavone, Marco (2023). Learning responsibility allocations for safe human-robot interaction with applications to autonomous driving. In *2023 IEEE international conference on robotics and automation* (pp. 9757–9763).

Cosner, Ryan, Tucker, Maegan, Taylor, Andrew, Li, Kejun, Molnar, Tamas, Ubelacker, Wyatt, et al. (2022). Safety-aware preference-based learning for safety-critical control. In *Learning for dynamics and control conference* (pp. 1020–1033). PMLR.

Cui, Yuxiang, Lin, Longzhong, Huang, Xiaolong, Zhang, Dongkun, Wang, Yunkai, Jing, Wei, et al. (2022). Learning observation-based certifiable safe policy for decentralized multi-robot navigation. In *2022 international conference on robotics and automation* (pp. 5518–5524). IEEE.

Cui, Jingjing, Liu, Yuanwei, & Nallanathan, Arumugam (2019). Multi-agent reinforcement learning-based resource allocation for UAV networks. *IEEE Transactions on Wireless Communication*, *19*(2), 729–743.

Dai, Bolun, Krishnamurthy, Prashanth, & Khorrami, Farshad (2022). Learning a better control barrier function. In *2022 IEEE 61st conference on decision and control* (pp. 945–950). IEEE.

Dai, Hongkai, Landry, Benoit, Yang, Lujie, Pavone, Marco, & Tedrake, Russ (2021). Lyapunov-stable neural network control. In *Proceedings of robotics: science and systems*.

Dalal, Gal, Dvijotham, Krishnamurthy, Vecerik, Matej, Hester, Todd, Paduraru, Cosmin, & Tassa, Yuval (2018). Safe exploration in continuous action spaces. arXiv preprint arXiv:1801.08757.

Damani, Mehul, Luo, Zhiyao, Wenzel, Emerson, & Sartoretti, Guillaume (2021). PRIMAL _2: Pathfinding via reinforcement and imitation multi-agent learning-lifelong. *IEEE Robotics and Automation Letters*, *6*(2), 2666–2673.

Das, Abhishek, Gervet, Théophile, Romoff, Joshua, Batra, Dhruv, Parikh, Devi, Rabbat, Mike, et al. (2019). TarMAC: Targeted multi-agent communication. In *Proceedings of the 36th international conference on machine learning* (pp. 1538–1546). PMLR.

Dawson, Charles, Gao, Sicun, & Fan, Chuchu (2023). Safe control with learned certificates: A survey of neural Lyapunov, barrier, and contraction methods for robotics and control. *IEEE Transactions on Robotics*, *39*(3), 1749–1767.

Dawson, Charles, Qin, Zengyi, Gao, Sicun, & Fan, Chuchu (2022). Safe nonlinear control using robust neural Lyapunov-barrier functions. In *Conference on robot learning* (pp. 1724–1735). PMLR.

Ding, Dongsheng, Wei, Xiaohan, Yang, Zhuoran, Wang, Zhaoran, & Jovanovic, Mihailo (2023). Provably efficient generalized Lagrangian policy optimization for safe multi-agent reinforcement learning. In *Learning for dynamics and control conference* (pp. 315–332). PMLR.

Dorri, Ali, Kanhere, Salil S., & Jurdak, Raja (2018). Multi-agent systems: A survey. *IEEE Access*, *6*, 28573–28593.

Du, Wei, & Ding, Shifei (2021). A survey on multi-agent deep reinforcement learning: From the perspective of challenges and applications. *Artificial Intelligence Review*, *54*, 3215–3238.

ElSayed-Aly, Ingy, Bharadwaj, Suda, Amato, Christopher, Ehlers, Rüdiger, Topcu, Ufuk, & Feng, Lu (2021). Safe multi-agent reinforcement learning via shielding. In *Proceedings of the 20th international conference on autonomous agents and multiagent systems* (pp. 483–491). International Foundation for Autonomous Agents and Multiagent Systems.

Espina, Enrique, Llanos, Jacqueline, Burgos-Mellado, Claudio, Cardenas-Dobson, Roberto, Martinez-Gomez, Manuel, & Saez, Doris (2020). Distributed control strategies for microgrids: An overview. *IEEE Access*, *8*, 193412–193448.

Everett, Michael, Chen, Yu Fan, & How, Jonathan P. (2018). Motion planning among dynamic, decision-making agents with deep reinforcement learning. In *2018 IEEE/RSJ international conference on intelligent robots and systems* (pp. 3052–3059). IEEE.

Fiorini, Paolo, & Shiller, Zvi (1998). Motion planning in dynamic environments using velocity obstacles. *International Journal of Robotics Research*, *17*(7), 760–772.

Fisac, Jaime F, Lugovoy, Neil F, Rubies-Royo, Vicenç, Ghosh, Shromona, & Tomlin, Claire J (2019). Bridging hamilton-Jacobi safety analysis and reinforcement learning. In *2019 international conference on robotics and automation* (pp. 8550–8556). IEEE.

Frampton, Kenneth D., Baumann, Oliver N., & Gardonio, Paolo (2010). A comparison of decentralized, distributed, and centralized vibro-acoustic control. *The Journal of the Acoustical Society of America*, *128*(5), 2798–2806.

Funada, Riku, Santos, María, Yamauchi, Junya, Hatanaka, Takeshi, Fujita, Masayuki, & Egerstedt, Magnus (2019). Visual coverage control for teams of quadcopters via control barrier functions. In *2019 international conference on robotics and automation* (pp. 3010–3016). IEEE.

Gao, Yan, Bai, Chenggang, Fu, Rao, & Quan, Quan (2023). A non-potential orthogonal vector field method for more efficient robot navigation and control. *Robotics and Autonomous Systems*, *159*, Article 104291.

Garg, Kunal, Arabi, Ehsan, & Panagou, Dimitra (2022). Fixed-time control under spatiotemporal and input constraints: A quadratic programming based approach. *Automatica*, *141*, Article 110314.

Garg, Kunal, Cosner, Ryan K, Rosolia, Ugo, Ames, Aaron D, & Panagou, Dimitra (2021). Multi-rate control design under input constraints via fixed-time barrier functions. *IEEE Control Systems Letters*, *6*, 608–613.

Garg, Kunal, Dawson, Charles, Xu, Kathleen, Ornik, Melkior, & Fan, Chuchu (2023). Model-free neural fault detection and isolation for safe control. *IEEE Control Systems Letters*, *7*, 3169–3174.

Garg, Kunal, & Panagou, Dimitra (2019a). Control-Lyapunov and control-barrier functions based quadratic program for spatio-temporal specifications. In *2019 IEEE 58th conference on decision and control* (pp. 1422–1429). IEEE.

Garg, Kunal, & Panagou, Dimitra (2019b). Finite-time estimation and control for multi-aircraft systems under wind and dynamic obstacles. *Journal of Guidance, Control, and Dynamics*, *42*(7), 1489–1505.

Garg, Kunal, & Panagou, Dimitra (2021). Robust control barrier and control Lyapunov functions with fixed-time convergence guarantees. In *2021 American control conference* (pp. 2292–2297). IEEE.

Geibel, Peter (2006). Reinforcement learning for MDPs with constraints. In *17th European conference on machine learning* (pp. 646–653). Springer.

Geibel, Peter, & Wysotzki, Fritz (2005). Risk-sensitive reinforcement learning applied to control under constraints. *Journal of Artificial Intelligence Research*, *24*, 81–108.

Geng, Nan, Bai, Qinbo, Liu, Chenyi, Lan, Tian, Aggarwal, Vaneet, Yang, Yuan, et al. (2023). A reinforcement learning framework for vehicular network routing under peak and average constraints. *IEEE Transactions on Vehicular Technology*.

Glotfelter, Paul, Cortés, Jorge, & Egerstedt, Magnus (2017). Nonsmooth barrier functions with applications to multi-robot systems. *IEEE Control Systems Letters*, *1*(2), 310–315.

Glotfelter, Paul, Cortés, Jorge, & Egerstedt, Magnus (2018). Boolean composability of constraints and control synthesis for multi-robot systems via nonsmooth control barrier functions. In *2018 IEEE conference on control technology and applications* (pp. 897–902). IEEE.

Gronauer, Sven, & Diepold, Klaus (2022). Multi-agent deep reinforcement learning: A survey. *Artificial Intelligence Review*, 1–49.

Gu, Shangding, Kuba, Jakub Grudzien, Chen, Yuanpei, Du, Yali, Yang, Long, Knoll, Alois, et al. (2023). Safe multi-agent reinforcement learning for multi-robot control. *Artificial Intelligence*, *319*, Article 103905.

Gu, Shangding, Kuba, Jakub Grudzien, Wen, Munning, Chen, Ruiqing, Wang, Ziyan, Tian, Zheng, et al. (2021). Multi-agent constrained policy optimisation. arXiv preprint arXiv:2110.02793.

Gu, Shangding, Yang, Long, Du, Yali, Chen, Guang, Walter, Florian, Wang, Jun, et al. (2022). A review of safe reinforcement learning: Methods, theory and applications. arXiv preprint arXiv:2205.10330.

Hsu, Shao-Chen, Xu, Xiangru, & Ames, Aaron D. (2015). Control barrier function based quadratic programs with application to bipedal robotic walking. In *2015 American control conference* (pp. 4542–4548). IEEE.

Hu, Yifan, Fu, Junjie, & Wen, Guanghui (2023). Decentralized robust collision-avoidance for cooperative multirobot systems: A Gaussian process-based control barrier function approach. *IEEE Transactions on Control of Network Systems*, *10*(2), 706–717.

Huang, Jie, Zhang, Jiancheng, Tian, Guoqing, & Chen, Yutao (2023). Integrated planning and control for formation reconfiguration of multiple spacecrafts: A predictive behavior control approach. *Advances in Space Research*, *72*(6), 2007–2019.

Hwangbo, Jemin, Lee, Joonho, Dosovitskiy, Alexey, Bellicoso, Dario, Tsounis, Vassilios, Koltun, Vladlen, et al. (2019). Learning agile and dynamic motor skills for legged robots. *Science Robotics*, *4*(26), Article eaau5872.

Ismail, Zool Hilmi, Sariff, Nohaidda, & Hurtado, E. Gorrostieta (2018). A survey and analysis of cooperative multi-agent robot systems: Challenges and directions. *Applications of Mobile Robots*, 8–14.

Jankovic, Mrdjan, & Santillo, Mario (2021). Collision avoidance and liveness of multi-agent systems with CBF-based controllers. In *2021 60th IEEE conference on decision and control* (pp. 6822–6828). IEEE.

Jiang, Chao, & Guo, Yi (2023). Incorporating control barrier functions in distributed model predictive control for multi-robot coordinated control. *IEEE Transactions on Control of Network Systems*.

Jiang, Jiechuan, & Lu, Zongqing (2018). Learning attentional communication for multi-agent cooperation. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Advances in neural information processing systems*: *vol. 31*, Curran Associates, Inc..

Jin, Wanxin, Wang, Zhaoran, Yang, Zhuoran, & Mou, Shaoshuai (2020). Neural certificates for safe control policies. arXiv preprint arXiv:2006.08465.

Khan, Arbaaz, Zhang, Chi, Li, Shuo, Wu, Jiayue, Schlotfeldt, Brent, Tang, Sarah Y, et al. (2019). Learning safe unlabeled multi-robot planning with motion constraints. In *2019 IEEE/RSJ international conference on intelligent robots and systems* (pp. 7558–7565). IEEE.

Kim, Woojun, Park, Jongeui, & Sung, Youngchul (2021). Communication in multi-agent reinforcement learning: Intention sharing. In *International conference on learning representations*.

Könighofer, Bettina, Alshiekh, Mohammed, Bloem, Roderick, Humphrey, Laura, Könighofer, Robert, Topcu, Ufuk, et al. (2017). Shield synthesis. *Formal Methods in System Design*, *51*, 332–361.

Kuba, Jakub Grudzien, Chen, Ruiqing, Wen, Muning, Wen, Ying, Sun, Fanglei, Wang, Jun, et al. (2022). Trust region policy optimisation in multi-agent reinforcement learning. In *International conference on learning representations*.

Lasry, Jean-Michel, & Lions, Pierre-Louis (2007). Mean field games. *Japanese Journal of Mathematics*, *2*(1), 229–260.

Li, Shuo, & Bastani, Osbert (2020). Robust model predictive shielding for safe reinforcement learning with stochastic dynamics. In *2020 IEEE international conference on robotics and automation* (pp. 7166–7172). IEEE.

Li, Yujia, Gu, Chenjie, Dullien, Thomas, Vinyals, Oriol, & Kohli, Pushmeet (2019). Graph matching networks for learning the similarity of graph structured objects. In *International conference on machine learning* (pp. 3835–3845). PMLR.

Li, Xianwei, Tang, Yang, & Karimi, Hamid Reza (2020). Consensus of multi-agent systems via fully distributed event-triggered control. *Automatica*, *116*, Article 108898.

Lin, Alex Tong, Fung, Samy Wu, Li, Wuchen, Nurbekyan, Levon, & Osher, Stanley J (2021). Alternating the population and control neural networks to solve high-dimensional stochastic mean-field games. *Proceedings of the National Academy of Sciences*, *118*(31).

Lindemann, Lars, & Dimarogonas, Dimos V. (2019). Control barrier functions for multi-agent systems under conflicting local signal temporal logic tasks. *IEEE Control Systems Letters*, *3*(3), 757–762.

Lindemann, Lars, & Dimarogonas, Dimos V. (2020). Barrier function based collaborative control of multiple robots under signal temporal logic tasks. *IEEE Transactions on Control of Network Systems*, *7*(4), 1916–1928.

Liu, Changliu, Arnon, Tomer, Lazarus, Christopher, Strong, Christopher, Barrett, Clark, Kochenderfer, Mykel J, et al. (2021). Algorithms for verifying deep neural networks. *Foundations and Trends in Optimization*, *4*(3–4), 244–404.

Liu, Guan-Horng, Chen, Tianrong, So, Oswin, & Theodorou, Evangelos (2022). Deep generalized Schrödinger bridge. *Advances in Neural Information Processing Systems*, *35*, 9374–9388.

Liu, Chenyi, Geng, Nan, Aggarwal, Vaneet, Lan, Tian, Yang, Yuan, & Xu, Mingwei (2021). CMIX: Deep multi-agent reinforcement learning with peak and average constraints. In *Machine learning and knowledge discovery in databases. Research track: European conference* (pp. 157–173). Springer.

Liu, Yongshuai, Halev, Avishai, & Liu, Xin (2021). Policy learning with constraints in model-free reinforcement learning: A survey. In *The 30th international joint conference on artificial intelligence*.

Long, Pinxin, Fan, Tingxiang, Liao, Xinyi, Liu, Wenxi, Zhang, Hao, & Pan, Jia (2018). Towards optimally decentralized multi-robot collision avoidance via deep reinforcement learning. In *2018 IEEE international conference on robotics and automation* (pp. 6252–6259). IEEE.

Lowe, Ryan, Wu, Yi I, Tamar, Aviv, Harb, Jean, Pieter Abbeel, OpenAI, & Mordatch, Igor (2017). Multi-agent actor-critic for mixed cooperative-competitive environments. *Advances in Neural Information Processing Systems*, *30*.

Lu, Songtao, Zhang, Kaiqing, Chen, Tianyi, Başar, Tamer, & Horesh, Lior (2021). Decentralized policy gradient descent ascent for safe multi-agent reinforcement learning. *Vol. 35*, In *Proceedings of the AAAI conference on artificial intelligence* (pp. 8767–8775).

Luo, Wenhao, Sun, Wen, & Kapoor, Ashish (2020). Multi-robot collision avoidance under uncertainty with probabilistic safety barrier certificates. *Advances in Neural Information Processing Systems*, *33*, 372–383.

Lyu, Xueguang, Xiao, Yuchen, Daley, Brett, & Amato, Christopher (2021). Contrasting centralized and decentralized critics in multi-agent reinforcement learning. In *Proceedings of the 20th international conference on autonomous agents and multiAgent systems* (pp. 844–852). International Foundation for Autonomous Agents and Multiagent Systems.

Machida, Manao, & Ichien, Masumi (2021). Consensus-based control barrier function for swarm. In *2021 IEEE international conference on robotics and automation* (pp. 8623–8628). IEEE.

Majumdar, Rupak, Mallik, Kaushik, Salamati, Mahmoud, Soudjani, Sadegh, & Zareian, Mehrdad (2021). Symbolic reach-avoid control of multi-agent systems. In *Proceedings of the ACM/IEEE 12th international conference on cyber-physical systems* (pp. 209–220).

Mali, Pravin, Harikumar, K, Singh, Arun Kumar, Krishna, K Madhava, & Sujit, PB (2021). Incorporating prediction in control barrier function based distributive multi-robot collision avoidance. In *2021 European control conference* (pp. 2394–2399). IEEE.

Marco, Alonso, Baumann, Dominik, Khadiv, Majid, Hennig, Philipp, Righetti, Ludovic, & Trimpe, Sebastian (2021). Robot learning with crash constraints. *IEEE Robotics and Automation Letters*, *6*(2), 1439–1446.

Massiani, Pierre-François, Heim, Steve, Solowjow, Friedrich, & Trimpe, Sebastian (2023). Safe value functions. *IEEE Transactions on Automatic Control*, *68*(5), 2743–2757.

Mazala, René (2002). Infinite games. In *Automata logics, and infinite games: a guide to current research* (pp. 23–38). Springer.

Mehdifar, Farhad, Bechlioulis, Charalampos P, Hashemzadeh, Farzad, & Baradarannia, Mahdi (2020). Prescribed performance distance-based formation control of multi-agent systems. *Automatica*, *119*, Article 109086.

Melcer, Daniel, Amato, Christopher, & Tripakis, Stavros (2022). Shield decentralization for safe multi-agent reinforcement learning. *Advances in Neural Information Processing Systems*, *35*, 13367–13379.

Meng, Yue, Qin, Zengyi, & Fan, Chuchu (2021). Reactive and safe road user simulations using neural barrier certificates. In *2021 IEEE/RSJ international conference on intelligent robots and systems* (pp. 6299–6306). IEEE.

Mesbahi, Mehran, & Egerstedt, Magnus (2010). Graph theoretic methods in multiagent networks. In *Graph theoretic methods in multiagent networks*. Princeton University Press.

Molzahn, Daniel K, Dörfler, Florian, Sandberg, Henrik, Low, Steven H, Chakrabarti, Sambuddha, Baldick, Ross, et al. (2017). A survey of distributed optimization and control algorithms for electric power systems. *IEEE Transactions on Smart Grid*, *8*(6), 2941–2962.

Muntwiler, Simon, Wabersich, Kim P., Carron, Andrea, & Zeilinger, Melanie N. (2020). Distributed model predictive safety certification for learning-based control. *IFAC-PapersOnLine*, *53*(2), 5258–5265, 21st IFAC World Congress.

Nagumo, Mitio (1942). Über die lage der integralkurven gewöhnlicher differentialgleichungen. *Vol. 24*, In *Proceedings of the physico-mathematical society of Japan. 3rd series* (pp. 551–559). THE PHYSICAL SOCIETY OF JAPAN, The Mathematical Society of Japan.

Nedić, Angelia, & Liu, Ji (2018). Distributed optimization for control. *Annual Review of Control, Robotics, and Autonomous Systems*, *1*, 77–103.

Nguyen, Thanh Thi, Nguyen, Ngoc Duy, & Nahavandi, Saeid (2020). Deep reinforcement learning for multiagent systems: A review of challenges, solutions, and applications. *IEEE Transactions on Cybernetics*, *50*(9), 3826–3839.

Nowzari, Cameron, Garcia, Eloy, & Cortés, Jorge (2019). Event-triggered communication and control of networked systems for multi-agent consensus. *Automatica*, *105*, 1–27.

Nweye, Kingsley, Liu, Bo, Stone, Peter, & Nagy, Zoltan (2022). Real-world challenges for multi-agent reinforcement learning in grid-interactive buildings. *Energy and AI*, *10*, Article 100202.

Oh, Kwang-Kyo, Park, Myoung-Chul, & Ahn, Hyo-Sung (2015). A survey of multi-agent formation control. *Automatica*, *53*, 424–440.

Oroojlooy, Afshin, & Hajinezhad, Davood (2023). A review of cooperative multi-agent deep reinforcement learning. *Applied Intelligence*, *53*(11), 13677–13722.

Panagou, Dimitra (2016). A distributed feedback motion planning protocol for multiple unicycle agents of different classes. *IEEE Transactions on Automatic Control*, *62*(3), 1178–1193.

Panagou, Dimitra, Stipanovič, Dušan M, & Voulgaris, Petros G (2013). Multi-objective control for multi-agent systems using lyapunov-like barrier functions. In *52nd IEEE conference on decision and control* (pp. 1478–1483). IEEE.

Papachristodoulou, Antonis, Jadbabaie, Ali, & Münz, Ulrich (2010). Effects of delay in multi-agent consensus and oscillator synchronization. *IEEE Transactions on Automatic Control*, *55*(6), 1471–1477.

Pereira, Marcus A, Saravanos, Augustinos D, So, Oswin, & Theodorou, Evangelos A. (2022). Decentralized safe multi-agent stochastic optimal control using deep FBSDEs and ADMM. In *Proceedings of robotics: science and systems*. New York City, NY, USA.

Peruffo, Andrea, Ahmed, Daniele, & Abate, Alessandro (2021). Automated and formal synthesis of neural barrier certificates for dynamical models. In *International conference on tools and algorithms for the construction and analysis of systems* (pp. 370–388). Springer.

Pnueli, Amir (1977). The temporal logic of programs. In *18th annual symposium on foundations of computer science* (pp. 46–57). IEEE.

Prajapat, Manish, Turchetta, Matteo, Zeilinger, Melanie, & Krause, Andreas (2022). Near-optimal multi-agent learning for safe coverage control. *Advances in Neural Information Processing Systems*, *35*, 14998–15012.

Prajna, Stephen (2006). Barrier certificates for nonlinear model validation. *Automatica*, *42*(1), 117–126.

Prajna, Stephen, & Jadbabaie, Ali (2004). Safety verification of hybrid systems using barrier certificates. In *International workshop on hybrid systems: computation and control* (pp. 477–492). Springer.

Puterman, Martin L. (2014). *Markov decision processes: Discrete stochastic dynamic programming*. John Wiley & Sons.

Qi, Charles R., Su, Hao, Mo, Kaichun, & Guibas, Leonidas J. (2017). Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 652–660).

Qin, Zengyi, Chen, Yuxiao, & Fan, Chuchu (2021). Density constrained reinforcement learning. In *International conference on machine learning* (pp. 8682–8692). PMLR.

Qin, Zengyi, Sun, Dawei, & Fan, Chuchu (2022). Sablas: Learning safe control for black-box dynamical systems. *IEEE Robotics and Automation Letters*, *7*(2), 1928–1935.

Qin, Zengyi, Zhang, Kaiqing, Chen, Yuxiao, Chen, Jingkai, & Fan, Chuchu (2020). Learning safe multi-agent control with decentralized neural barrier certificates. In *International conference on learning representations*.

Queralta, Jorge Pena, Taipalmaa, Jussi, Pullinen, Bilge Can, Sarker, Victor Kathan, Gia, Tuan Nguyen, Tenhunen, Hannu, et al. (2020). Collaborative multi-robot search and rescue: Planning, coordination, perception, and active vision. *IEEE Access*, *8*, 191617–191643.

Rashid, Tabish, Samvelyan, Mikayel, De Witt, Christian Schroeder, Farquhar, Gregory, Foerster, Jakob, & Whiteson, Shimon (2020). Monotonic value function factorisation for deep multi-agent reinforcement learning. *Journal of Machine Learning Research*, *21*(1), 7234–7284.

Ray, Alex, Achiam, Joshua, & Amodei, Dario (2019). Benchmarking safe exploration in deep reinforcement learning. arXiv:1910.01708.

Reis, Matheus F., Aguiar, A. Pedro, & Tabuada, Paulo (2020). Control barrier function-based quadratic programs introduce undesirable asymptotically stable equilibria. *IEEE Control Systems Letters*, *5*(2), 731–736.

Ren, Wei (2007). Distributed attitude alignment in spacecraft formation flying. *International Journal of Adaptive Control and Signal Processing*, *21*(2–3), 95–113.

Ren, Wei, Beard, Randal W., & Atkins, Ella M. (2005). A survey of consensus problems in multi-agent coordination. In *American control conference* (pp. 1859–1864). IEEE.

Ringler, Philipp, Keles, Dogan, & Fichtner, Wolf (2016). Agent-based modelling and simulation of smart electricity grids and markets– A literature review. *Renewable and Sustainable Energy Reviews*, *57*, 205–215.

Rizk, Yara, Awad, Mariette, & Tunstel, Edward W. (2019). Cooperative heterogeneous multi-robot systems: A survey. *ACM Computing Surveys*, *52*(2), 1–31.

Robbins, Herbert, & Monro, Sutton (1951). A stochastic approximation method. *The Annals of Mathematical Statistics*, 400–407.

Robey, Alexander, Hu, Haimin, Lindemann, Lars, Zhang, Hanwen, Dimarogonas, Dimos V, Tu, Stephen, et al. (2020). Learning control barrier functions from expert demonstrations. In *2020 59th IEEE conference on decision and control* (pp. 3717–3724). IEEE.

Roth, Maayan, Simmons, Reid, & Veloso, Manuela (2005). Decentralized communication strategies for coordinated multi-agent policies. In *Multi-robot systems. From swarms to intelligent automata volume III: Proceedings from the 2005 international workshop on multi-robot systems* (pp. 93–105). Springer.

Sabattini, Lorenzo, Secchi, Cristian, Chopra, Nikhil, & Gasparri, Andrea (2013). Distributed control of multirobot systems with global connectivity maintenance. *IEEE Transactions on Robotics*, *29*(5), 1326–1332.

Saim, Muhammad, Ghapani, Sheida, Ren, Wei, Munawar, Khalid, & Al-Saggaf, Ubaid M (2017). Distributed average tracking in multi-agent coordination: Extensions and experiments. *IEEE Systems Journal*, *12*(3), 2428–2436.

Salman, Hadi, Ayvali, Elif, & Choset, Howie (2017). Multi-agent ergodic coverage with obstacle avoidance. *Vol. 27*, In *Proceedings of the international conference on automated planning and scheduling* (pp. 242–249).

Sälzer, Marco, & Lange, Martin (2023). Fundamental limits in formal verification of message-passing neural networks. In *The eleventh international conference on learning representations*.

Salzman, Oren, & Stern, Roni (2020). Research challenges and opportunities in multi-agent path finding and multi-agent pickup and delivery problems. In *Proceedings of the 19th international conference on autonomous agents and multiagent systems* (pp. 1711–1715).

Santos, María, & Egerstedt, Magnus (2018). Coverage control for multi-robot teams with heterogeneous sensing capabilities using limited communications. In *2018 IEEE/RSJ international conference on intelligent robots and systems* (pp. 5313–5319). IEEE.

Sartoretti, Guillaume, Kerr, Justin, Shi, Yunfei, Wagner, Glenn, Kumar, TK Satish, Koenig, Sven, et al. (2019). PRIMAL: Pathfinding via reinforcement and imitation multi-agent learning. *IEEE Robotics and Automation Letters*, *4*(3), 2378–2385.

Satija, Harsh, Amortila, Philip, & Pineau, Joelle (2020). Constrained Markov decision processes via backward value functions. In *International conference on machine learning* (pp. 8502–8511). PMLR.

Saveriano, Matteo, & Lee, Dongheui (2019). Learning barrier functions for constrained motion planning with dynamical systems. In *2019 IEEE/RSJ international conference on intelligent robots and systems* (pp. 112–119). IEEE.

Schrittwieser, Julian, Antonoglou, Ioannis, Hubert, Thomas, Simonyan, Karen, Sifre, Laurent, Schmitt, Simon, et al. (2020). Mastering atari, go, chess and shogi by planning with a learned model. *Nature*, *588*(7839), 604–609.

Schulman, John, Levine, Sergey, Abbeel, Pieter, Jordan, Michael, & Moritz, Philipp (2015). Trust region policy optimization. In *International conference on machine learning* (pp. 1889–1897). PMLR.

Semnani, Samaneh Hosseini, Liu, Hugh, Everett, Michael, De Ruiter, Anton, & How, Jonathan P (2020). Multi-agent motion planning for dense and dynamic environments via deep reinforcement learning. *IEEE Robotics and Automation Letters*, *5*(2), 3221–3226.

Servin, Arturo, & Kudenko, Daniel (2008). Multi-agent reinforcement learning for intrusion detection. In Karl Tuyls, Ann Nowe, Zahia Guessoum, & Daniel Kudenko (Eds.), *Adaptive agents and multi-agent systems iII. Adaptation and multi-agent learning* (pp. 211–223). Berlin, Heidelberg: Springer Berlin Heidelberg.

Shalev-Shwartz, Shai, Shammah, Shaked, & Shashua, Amnon (2016). Safe, multi-agent, reinforcement learning for autonomous driving. arXiv preprint arXiv:1610.03295.

Sheebaelhamd, Ziyad, Zisis, Konstantinos, Nisioti, Athina, Gkouletsos, Dimitris, Pavllo, Dario, & Kohler, Jonas (2021). Safe deep reinforcement learning for multi-agent systems with continuous action spaces. In *ICML workshop on reinforcement learning for real life*.

Shen, Shen, & Tedrake, Russ (2018). Compositional verification of large-scale non-linear systems via sums-of-squares optimization. In *2018 annual American control conference* (pp. 4385–4392).

Silver, David, Huang, Aja, Maddison, Chris J, Guez, Arthur, Sifre, Laurent, Van Den Driessche, George, et al. (2016). Mastering the game of go with deep neural networks and tree search. *Nature*, *529*(7587), 484–489.

Silver, David, Schrittwieser, Julian, Simonyan, Karen, Antonoglou, Ioannis, Huang, Aja, Guez, Arthur, et al. (2017). Mastering the game of go without human knowledge. *Nature*, *550*(7676), 354–359.

Snape, Jamie, Van Den Berg, Jur, Guy, Stephen J, & Manocha, Dinesh (2011). The hybrid reciprocal velocity obstacle. *IEEE Transactions on Robotics*, *27*(4), 696–706.

So, Oswin, & Fan, Chuchu (2023). Solving stabilize-avoid optimal control via epigraph form and deep reinforcement learning. In *Proceedings of robotics: science and systems*. Daegu, Republic of Korea.

So, Oswin, Serlin, Zachary, Mann, Makai, Gonzales, Jake, Rutledge, Kwesi, Roy, Nicholas, et al. (2024). How to train your neural control barrier function: Learning safety filters for complex input-constrained systems. In *2024 international conference on robotics and automation*. IEEE.

Sontag, Eduardo D. (1983). A Lyapunov-like characterization of asymptotic controllability. *SIAM Journal on Control and Optimization*, *21*(3), 462–471.

Srinivasan, Mohit, Abate, Matthew, Nilsson, Gustav, & Coogan, Samuel (2021). Extent-compatible control barrier functions. *Systems & Control Letters*, *150*, Article 104895.

Srinivasan, Mohit, Dabholkar, Amogh, Coogan, Samuel, & Vela, Patricio A (2020). Synthesis of control barrier functions using a supervised machine learning approach. In *IEEE/RSJ international conference on intelligent robots and systems* (pp. 7139–7145). IEEE.

Sun, Dawei, Chen, Jingkai, Mitra, Sayan, & Fan, Chuchu (2022). Multi-agent motion planning from signal temporal logic specifications. *IEEE Robotics and Automation Letters*, *7*(2), 3451–3458.

Tahir, Anam, Böling, Jari, Haghbayan, Mohammad-Hashem, Toivonen, Hannu T, & Plosila, Juha (2019). Swarms of unmanned aerial vehicles—A survey. *Journal of Industrial Information Integration*, *16*, Article 100106.

Tassa, Yuval, Doron, Yotam, Muldal, Alistair, Erez, Tom, Li, Yazhe, Casas, Diego de Las, et al. (2018). Deepmind control suite. arXiv preprint arXiv:1801.00690.

Tasse, Geraud Nangue, Love, Tamlin, Nemecek, Mark, James, Steven, & Rosman, Benjamin (2023). ROSARL: Reward-only safe reinforcement learning. arXiv preprint arXiv:2306.00035.

Tee, Keng Peng, Ge, Shuzhi Sam, & Tay, Eng Hock (2009). Barrier Lyapunov functions for the control of output-constrained nonlinear systems. *Automatica*, *45*(4), 918–927.

Tessler, Chen, Mankowitz, Daniel J., & Mannor, Shie (2019). Reward constrained policy optimization. In *International conference on learning representations*.

Tong, Mukun, Dawson, Charles, & Fan, Chuchu (2023). Enforcing safety for vision-based controllers via control barrier functions and neural radiance fields. In *2023 IEEE international conference on robotics and automation* (pp. 10511–10517).

Tonkens, Sander, & Herbert, Sylvia (2022). Refining control barrier functions through hamilton-Jacobi reachability. In *2022 IEEE/RSJ international conference on intelligent robots and systems* (pp. 13355–13362).

Usevitch, James, Garg, Kunal, & Panagou, Dimitra (2020). Strong invariance using control barrier functions: A clarke tangent cone approach. In *2020 59th IEEE conference on decision and control* (pp. 2044–2049). IEEE.

Vinod, Abraham P, Safaoui, Sleiman, Chakrabarty, Ankush, Quirynen, Rien, Yoshikawa, Nobuyuki, & Di Cairano, Stefano (2022). Safe multi-agent motion planning via filtered reinforcement learning. In *2022 international conference on robotics and automation* (pp. 7270–7276). IEEE.

Vinyals, Oriol, Babuschkin, Igor, Chung, Junyoung, Mathieu, Michael, Jaderberg, Max, Czarnecki, Wojciech M, et al. (2019). Alphastar: Mastering the real-time strategy game starcraft II. *DeepMind Blog*, *2*, 20.

Vorotnikov, Sergey, Ermishin, Konstantin, Nazarova, Anaid, & Yuschenko, Arkady (2018). Multi-agent robotic systems in collaborative robotics. In *Interactive collaborative robotics: third international conference* (pp. 270–279). Springer.

Wabersich, Kim P., & Zeilinger, Melanie N. (2018). Linear model predictive safety certification for learning-based control. In *2018 IEEE conference on decision and control* (pp. 7130–7135). IEEE.

Wang, Li, Ames, Aaron, & Egerstedt, Magnus (2016). Safety barrier certificates for heterogeneous multi-robot systems. In *2016 American control conference* (pp. 5213–5218). IEEE.

Wang, Li, Ames, Aaron D., & Egerstedt, Magnus (2017). Safety barrier certificates for collisions-free multirobot systems. *IEEE Transactions on Robotics*, *33*(3), 661–674.

Wang, Qishao, Duan, Zhisheng, Lv, Yuezu, Wang, Qingyun, & Chen, Guanrong (2020). Distributed model predictive control for linear–quadratic performance and consensus state optimization of multiagent systems. *IEEE Transactions on Cybernetics*, *51*(6), 2905–2915.

Wang, Rundong, He, Xu, Yu, Runsheng, Qiu, Wei, An, Bo, & Rabinovich, Zinovi (2020). Learning efficient multi-agent communication: An information bottleneck approach. *Vol. 119*, In *Proceedings of the 37th international conference on machine learning* (pp. 9908–9918). PMLR.

Wang, Jianrui, Hong, Yitian, Wang, Jiali, Xu, Jiapeng, Tang, Yang, Han, Qing-Long, et al. (2022). Cooperative and competitive multi-agent systems: From optimization to games. *IEEE/CAA Journal of Automatica Sinica*, *9*(5), 763–783.

Wang, Xinrui, Leung, Karen, & Pavone, Marco (2020). Infusing reachability-based safety into planning and control for multi-agent interactions. In *2020 IEEE/RSJ international conference on intelligent robots and systems* (pp. 6252–6259).

Wang, Liang, Wang, Kezhi, Pan, Cunhua, Xu, Wei, Aslam, Nauman, & Hanzo, Lajos (2020). Multi-agent deep reinforcement learning-based trajectory planning for multi-UAV assisted mobile edge computing. *IEEE Transactions on Cognitive Communications and Networking*, *7*(1), 73–84.

Wang, Guang, Xu, Ming, Wu, Yiming, Zheng, Ning, Xu, Jian, & Qiao, Tong (2018). Using machine learning for determining network robustness of multi-agent systems under attacks. In Xin Geng, & Byeong-Ho Kang (Eds.), *PRICAI 2018: trends in artificial intelligence* (pp. 491–498). Cham: Springer International Publishing.

Wang, Lizhi, Zhang, Songyuan, Zhou, Yifan, Fan, Chuchu, Zhang, Peng, & Shamash, Yacov A (2023a). Learning-based, safety and stability-certified microgrid control. In *2023 IEEE power & energy society general meeting* (pp. 1–5). IEEE.

Wang, Lizhi, Zhang, Songyuan, Zhou, Yifan, Fan, Chuchu, Zhang, Peng, & Shamash, Yacov A (2023b). Physics-informed, safety and stability certified neural control for uncertain networked microgrids. *IEEE Transactions on Smart Grid*.

Wei, Caisheng, Luo, Jianjun, Dai, Honghua, & Duan, Guangren (2018). Learning-based adaptive attitude control of spacecraft formation with guaranteed prescribed performance. *IEEE Transactions on Cybernetics*, *49*(11), 4004–4016.

Wieland, Peter, & Allgöwer, Frank (2007). Constructive safety using control barrier functions. *IFAC Proceedings Volumes*, *40*(12), 462–467.

Wu, Guofan, & Sreenath, Koushil (2016). Safety-critical control of a planar quadrotor. In *2016 American control conference* (pp. 2252–2258). IEEE.

Xian, Zhou, Lertkultanon, Puttichai, & Pham, Quang-Cuong (2017). Closed-chain manipulation of large objects by multi-arm robotic systems. *IEEE Robotics and Automation Letters*, *2*(4), 1832–1839.

Xiao, Wenli, Lyu, Yiwei, & Dolan, John (2023). Model-based dynamic shielding for safe and efficient multi-agent reinforcement learning. In *Proceedings of the 2023 international conference on autonomous agents and multiagent systems* (pp. 1587–1596). Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

Xiao, Wei, Wang, Tsun-Hsuan, Hasani, Ramin, Chahine, Makram, Amini, Alexander, Li, Xiao, et al. (2023). Barriernet: Differentiable control barrier functions for learning of safe robot control. *IEEE Transactions on Robotics*.

Xie, Jing, & Liu, Chen-Ching (2017). Multi-agent systems and their applications. *Journal of International Council on Electrical Engineering*, *7*(1), 188–197.

Xu, Xiangru, Waters, Thomas, Pickem, Daniel, Glotfelter, Paul, Egerstedt, Magnus, Tabuada, Paulo, et al. (2017). Realizing simultaneous lane keeping and adaptive speed regulation on accessible mobile robot testbeds. In *2017 IEEE conference on control technology and applications* (pp. 1769–1775). IEEE.

Xuan, Ping, & Lesser, Victor (2002). Multi-agent policies: From centralized ones to decentralized ones. In *Proceedings of the first international joint conference on autonomous agents and multiagent systems: part 3* (pp. 1098–1105).

Xue, Lei, & Cao, Xianghui (2019). Leader selection via supermodular game for formation control in multiagent systems. *IEEE Transactions on Neural Networks and Learning Systems*, *30*(12), 3656–3664.

Yamagata, Yoriyuki, Liu, Shuang, Akazaki, Takumi, Duan, Yihai, & Hao, Jianye (2021). Falsification of cyber-physical systems using deep reinforcement learning. *IEEE Transactions on Software Engineering*, *47*(12), 2823–2840.

Yang, Yaodong, & Wang, Jun (2020). An overview of multi-agent reinforcement learning from game theoretical perspective. arXiv preprint arXiv:2011.00583.

Yang, Tao, Yi, Xinlei, Wu, Junfeng, Yuan, Ye, Wu, Di, Meng, Ziyang, et al. (2019). A survey of distributed optimization. *Annual Reviews in Control*, *47*, 278–305.

Ye, Deheng, Chen, Guibin, Zhang, Wen, Chen, Sheng, Yuan, Bo, Liu, Bo, et al. (2020). Towards playing full moba games with deep reinforcement learning. *Advances in Neural Information Processing Systems*, *33*, 621–632.

Yin, Ji, Dawson, Charles, Fan, Chuchu, & Tsiotras, Panagiotis (2023). Shield model predictive path integral: A computationally efficient robust MPC method using control barrier functions. *IEEE Robotics and Automation Letters*, *8*(11), 7106–7113.

Yu, Hongzhan, Hirayama, Chiaki, Yu, Chenning, Herbert, Sylvia, & Gao, Sicun (2023). Sequential neural barriers for scalable dynamic obstacle avoidance. In *IEEE/RSJ international conference on intelligent robots and systems*.

Yu, Dongjie, Ma, Haitong, Li, Shengbo, & Chen, Jianyu (2022). Reachability constrained reinforcement learning. In *International conference on machine learning* (pp. 25636–25655). PMLR.

Yu, Chenning, Yu, Hongzhan, & Gao, Sicun (2023). Learning control admissibility models with graph neural networks for multi-agent navigation. In *Conference on robot learning* (pp. 934–945). PMLR.

Zavlanos, Michael M., & Pappas, George J. (2008). Distributed connectivity control of mobile networks. *IEEE Transactions on Robotics*, *24*(6), 1416–1428.

Zhang, Wenbo, Bastani, Osbert, & Kumar, Vijay (2019). MAMPS: Safe multi-agent reinforcement learning via model predictive shielding. arXiv preprint arXiv:1910.12639.

Zhang, Dan, Feng, Gang, Shi, Yang, & Srinivasan, Dipti (2021). Physical safety and cyber security analysis of multi-agent systems: A survey of recent advances. *IEEE/CAA Journal of Automatica Sinica*, *8*(2), 319–333.

Zhang, Songyuan, Garg, Kunal, & Fan, Chuchu (2023). Neural graph control barrier functions guided distributed collision-avoidance multi-agent control. In *7th annual conference on robot learning*.

Zhang, Songyuan, So, Oswin, Garg, Kunal, & Fan, Chuchu (2024). Gcbf+: A neural graph control barrier function framework for distributed safe multi-agent control. arXiv preprint arXiv:2401.14554.

Zhang, Songyuan, Xiu, Yumeng, Qu, Guannan, & Fan, Chuchu (2023). Compositional neural certificates for networked dynamical systems. In *Proceedings of the 5th annual learning for dynamics and control conference* (pp. 272–285). PMLR.

Zhang, Kaiqing, Yang, Zhuoran, & Başar, Tamer (2021b). Decentralized multi-agent reinforcement learning with networked agents: Recent advances. *Frontiers of Information Technology & Electronic Engineering*, *22*(6), 802–814.

Zhang, Kaiqing, Yang, Zhuoran, & Başar, Tamer (2021c). Multi-agent reinforcement learning: A selective overview of theories and algorithms. *Handbook of Reinforcement Learning and Control*, 321–384.

Zhang, Kaiqing, Yang, Zhuoran, Liu, Han, Zhang, Tong, & Basar, Tamer (2018). Fully decentralized multi-agent reinforcement learning with networked agents. In *International conference on machine learning* (pp. 5872–5881). PMLR.

Zhao, Weiye, He, Tairan, Chen, Rui, Wei, Tianhao, & Liu, Changliu (2023). State-wise safe reinforcement learning: A survey. In *Proceedings of the thirty-second international joint conference on artificial intelligence* (pp. 6814–6822). Survey Track.

Zhou, Ziyuan, Liu, Guanjun, & Tang, Ying (2023). Multi-agent reinforcement learning: Methods, applications, visionary prospects, and challenges. arXiv preprint arXiv:2305.10091.