## Pseudo-perturbation-based Broadcast Control of Multi-agent Systems \*

Yuji Ito<sup>a</sup>, Md Abdus Samad Kamal<sup>a</sup>, Takayoshi Yoshimura<sup>a</sup>, Shun-ichi Azuma<sup>b</sup>

<sup>a</sup> TOYOTA CENTRAL R&D LABS., INC., Nagakute-shi, Aichi 480-1192, Japan

<sup>b</sup> Graduate School of Informatics, Kyoto University, Kyoto-shi, Kyoto 606-8501, Japan

#### Abstract

The present paper proposes a novel broadcast control (BC) law for multi-agent coordination. A BC framework has been developed to achieve global coordination tasks with low communication volume. The BC law uses broadcast communication, which transmits an identical signal to all agents indiscriminately without any agent-to-agent communication. Unfortunately, all of the agents are required to take numerous random actions because the BC law is based on stochastic optimization. Such random actions degrade the control performance for coordination tasks and may invoke dangerous situations. In order to overcome these drawbacks, the present paper proposes the pseudo-perturbation-based broadcast control (PBC) law, which introduces multiple virtual random actions instead of the single physical action of the BC law. The following advantages of the PBC law are theoretically proven. The PBC law achieves coordination tasks asymptotically with probability 1. Compared with the BC law, unavailing actions are reduced and agents' states converge at least twice as fast. Increasing the number of multiple actions further improves the control performance because averaging multiple actions reduces unavailing randomness. Numerical simulations demonstrate that the PBC improves the control performance as compared with the BC law.

Key words: Broadcast control; Multi-agent systems; Simultaneous perturbation stochastic approximation; Stochastic control.

## 1 Introduction

Coordinations of multi-agent systems have recently attracted attention in various engineering fields in line with recent technological advancements in computing and communication resources. Swarm robots, groups of vehicles, embedded robotic systems, and electric power networks are examples of multi-agent systems that require coordination in order to achieve desired global behaviors. Some common behaviors desired in multi-agent systems include coverage control by agents deployed over a given space, rendezvous with formation of the agents in a synchronized manner, and assignment at target locations (for examples, see [7, 9, 12, 17]).

Decentralized/distributed control approaches perform multi-agent tasks using agent/supervisor-to-agent communication, e.g., [8,11]. In distributed frameworks, each agent takes action locally and individually, communicating with its neighbor agents. Alternatively, centralized control systems use a supervisor to aggregate information from all agents. Such aggregation is efficient for achieving *global* coordination tasks. The supervisor evaluates a global objective based on information of agents and transmits commands to all of the agents individually. However, such centralized systems require a massive communication volume between a supervisor and agents, i.e., *one-to-all* and *all-to-one* communications.

For achieving global coordination tasks with a low communication volume, the present paper focuses on another structure of multi-agent systems called a broadcast system. The broadcast system can work with a low communication volume because a supervisor broadcasts an identical signal to all of the agents indiscriminately and no agent-to-agent communication is used. Table 1 indicates that the both-way communication vol-

<sup>\*</sup> The material herein was presented in part at the 2018 American Control Conference [6]. Corresponding author is Yuji Ito, Tel. +81-561-71-7512. Md Abdus Samad Kamal is currently with the School of Engineering, Monash University Malaysia, Bandar Sunway, 47500 Selangor Darul Ehsan, Malaysia. Shun-ichi Azuma is currently with the Graduate School of Engineering, Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, Aichi 464-8603, Japan.

Email addresses: ito-yuji@mosk.tytlabs.co.jp (Yuji Ito), maskamal@ieee.org (Md Abdus Samad Kamal), yoshimura@mosk.tytlabs.co.jp (Takayoshi Yoshimura), shunichi.azuma@mae.nagoya-u.ac.jp (Shun-ichi Azuma).

Table 1 Communication volumes of unicast and broadcast protocols. The number of agents is denoted by N.

		one-to-all	all-to-one
	Protocol	communication	communication
	Unicast	O(N)	O(N)
	Broadcast	O(1)	O(N)

ume (O(N + 1)) of the broadcast system is lower than that (O(2N)) of centralized systems employing unicast protocols in one-to-all communication. The broadcast system has the potential to realize global behaviors because the supervisor aggregates the information of all of the agents. Broadcast systems are constructed even if supervisors and all agents use broadcast communication rather than unicast communication, e.g., connected vehicles in traffic coordination [10]. Combining broadcast communication with agent-to-agent communication ensures controllability and/or stability of distributed systems [22, 23].

The concept of broadcast control (BC) [2] has recently been proposed for multi-agent coordination, which is applied to broadcast systems. The concept is related to stochastic/biological systems [5, 21]. The BC law can achieve global objectives, such as coverage, assembly, and assignment tasks [14, 15]. A key scheme of the BC is inspired by simultaneous perturbation stochastic approximation (SPSA) [1, 19, 20]. After each agent randomly moves temporarily, a supervisor broadcasts a signal that corresponds to the achievement degree of an objective, i.e., the value of the corresponding objective function. The agents determine their next actions from the received signal and their random movements. These actions implicitly construct an approximate gradient of the objective function. The BC law employs a stochastic gradient method to minimize the objective function in the manner of SPSA. However, such a process restricts its applicability. The random actions of agents involve significant traveling cost and time in order to achieve an objective. In many systems, e.g., vehicular traffic systems, taking random actions should be avoided in consideration of safety and operational efficiency.

In order to overcome the limitations of the BC law, the present paper proposes a control law called the pseudoperturbation-based broadcast control (PBC) law. In order to avoid unavailing random actions and improve the control performance for coordination tasks, a key concept is employing *multiple virtual* random actions instead of the single physical random action of the BC. The PBC for broadcast systems is operated in a manner similar to the BC and retains the basic features of the BC. The present paper theoretically proves the following advantages and characteristics of the PBC law. Coordination tasks for broadcast systems are asymptotically achieved with probability 1. Unavailing random actions of agents are reduced and coordination tasks are achieved at least twice as fast, as compared to the BC law. The execution of a task is accelerated by increasing the number of multiple actions because the multiple actions improve the approximation accuracy of the gradient of an objective function for the task. The PBC law is demonstrated through numerical simulations of two types of multi-agent coordination tasks, which are a coverage task and a rendezvous task with formation selection.

This paper is an extension of our previous work (under review) [6]. We modified the paper significantly and added new points. Especially, this paper proposes the PBC law based on *multiple* virtual actions of agents, whereas the control law in the conference paper uses a *single* virtual action. To discuss *multiple* actions, the overview of the PBC in Section 3.1 and Theorem 6 were modified. All of Section 3.3 is a new part. Numerical simulations in Sections 4.2 and 4.3 to demonstrate the PBC law are new. All appendices to prove theorems and a proposition are new, where Appendices B and C were a modified version of [6].

The remainder of this paper is organized as follows. Section 2 states the primary problem and associated limitations of the BC law. The PBC law with theoretical analysis is presented in Section 3, which describes the main results of the paper. In order to demonstrate the effectiveness of the PBC law, two types of multi-agent coordination tasks are simulated in Section 4. Conclusions and future research are described in Section 5.

**Notation:** For a vector  $x := [x_1, ..., x_n]^T \in \mathbb{R}^n$  with nonzero elements,  $x^{(-1)}$  represents the element-wise inverse of x, i.e.,  $x^{(-1)} := [1/x_1, ..., 1/x_n]^T \in \mathbb{R}^n$ . The partial derivative of  $f(x) \in \mathbb{R}$  with respect to  $x \in \mathbb{R}^n$ is denoted by  $\partial_x f(x) \in \mathbb{R}^n$ . The operator  $\mathbb{E}[f(\sigma)]$  denotes the expectation of  $f(\sigma) \in \mathbb{R}^n$  with respect to all random variables  $\sigma$ . The operators  $\mathbb{E}[f(\sigma, x(\sigma))|_x]$  and  $\operatorname{Cov}[f(\sigma, x(\sigma))|_x]$  denotes the conditional expectation and the conditional covariance of  $f(\sigma, x(\sigma)) \in \mathbb{R}^n$  with respect to all random variables  $\sigma$  for a fixed x, respectively.

## 2 Problem setting via review of the BC

This section states the main problem and associated drawbacks of the BC law. Section 2.1 introduces the target systems and the main problem. After Section 2.2 reviews the BC law, which is a solution to the main problem, its limitations to be overcome are described in Section 2.3.

#### 2.1 Target systems and the main problem

Let us consider a multi-agent system  $\Sigma$  on an *n*dimensional state space, which is controlled as a broadcast system. The system consists of a global controller (supervisor) G and N agents  $A_i$  equipped with local controllers  $L_i$  (i = 1, ..., N). Note that, although agents are indexed from 1 to N for notational convenience, it is not necessary to discriminate among them.

The dynamics of the agent  $A_i$  at a discrete time step  $t \in \{0, 1, ...\}$  is omni-directional and is given by

$$A_i: x_i(t+1) = x_i(t) + u_i(t), \tag{1}$$

where  $x_i(t) \in \mathbb{R}^n$  and  $u_i(t) \in \mathbb{R}^n$  are the state and the control input of  $A_i$ , respectively. The collective state of all of the agents is denoted by  $x := [x_1^{\mathrm{T}}, ..., x_N^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{nN}$ . Its initial state x(0) is given. The local controller  $L_i$ , with which each agent  $A_i$  is equipped, is described as follows:

$$L_{i}: \begin{cases} \phi_{i}(t+1) &= f_{\phi}(\phi_{i}(t), \nu(t), t), \\ u_{i}(t) &= f_{u}(\phi_{i}(t), \nu(t), t), \end{cases}$$
(2)

where  $\phi_i \in \mathbb{R}^{n_{\phi}}$  is the state of the local controller  $L_i$ . The collective state of all of the local controllers is denoted by  $\phi := [\phi_1^{\mathrm{T}}, ..., \phi_N^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{n_{\phi}N}$ . The initial state is set to  $\phi(0) = 0 \in \mathbb{R}^{n_{\phi}N}$  for simplicity of discussion. The symbol  $\nu(t) \in \mathbb{R}^{n_{\nu}}$  is a broadcast signal, which is sent to all of the agents by the global controller G. The functions  $f_{\phi} : \mathbb{R}^{n_{\phi}} \times \mathbb{R}^{n_{\nu}} \times \mathbb{Z} \to \mathbb{R}^{n_{\phi}}$  and  $f_u : \mathbb{R}^{n_{\phi}} \times \mathbb{R}^{n_{\nu}} \times \mathbb{Z} \to \mathbb{R}^n$  determine the transition of  $\phi_i$  and the input  $u_i(t)$  of the agent  $A_i$ , respectively. The global controller G is expressed as

$$G:\nu(t) = f_{\nu}(x(t),\phi(t),t) \in \mathbb{R}^{n_{\nu}},\tag{3}$$

where  $f_{\nu} : \mathbb{R}^{nN} \times \mathbb{R}^{n_{\phi}N} \times \mathbb{Z} \to \mathbb{R}^{n_{\nu}}$  determines the broadcast signal  $\nu(t)$ .

Note the following three requirements with respect to the communication and controllers in terms of cost and equipment for the multi-agent system  $\Sigma$ . First, there is no agent-to-agent communication. Second, the local controllers  $f_{\phi}$  and  $f_u$  must be identical for all of the agents. Third, the broadcast signal  $\nu(t)$  must be identical for all of the agents (each optimal signal cannot be sent to each agent). Standard centralized, decentralized, and distributed control approaches can no longer be applied due to these requirements, whereas the requirements are crucial for designing controllers for large-scale multi-agent systems. The main problem of realizing coordination tasks under these requirements is stated as follows.

**Main problem:** For the multi-agent system  $\Sigma$ , find global and local controllers  $(G, L_1, L_2, \ldots, L_N)$  (i.e.,  $f_{\nu}$ ,  $f_{\phi}$ , and  $f_u$ ) that satisfy the three above-mentioned requirements and

$$\lim_{t \to \infty} J(x(t)) = \min_{x \in \mathbb{R}^{nN}} J(x), \tag{4}$$

where  $J : \mathbb{R}^{nN} \to \mathbb{R}$  is an objective function that is regarded as the performance index for a coordination task of the system  $\Sigma$ . Convergence to a local minimum of J(x) in (4) is allowed in this problem.

#### 2.2 Review of the BC law

This subsection briefly introduces the BC law [2]. The global controller G of the BC law is defined as

$$\nu(t) = f_{\nu}(x(t), \phi(t), t) := J(x(t)) \in \mathbb{R}.$$
 (5)

Each local controller  $L_i$  is defined by

$$\phi_i(t) := \begin{bmatrix} \phi_{i,1}(t) \\ \phi_{i,2}(t) \end{bmatrix} \in \{-1,1\}^n \times \mathbb{R},\tag{6}$$

$$\phi_i(t+1) = f_\phi(\phi_i(t), \nu(t), t) := \begin{bmatrix} \sigma_i(t) \\ \nu(t) \end{bmatrix}, \tag{7}$$

$$u_{i}(t) = f_{u}(\phi_{i}(t), \nu(t), t)$$
  
$$:= \begin{cases} c(t)\sigma_{i}(t) & (t \in \{0, 2, 4, ...\}) \\ -c(t)\phi_{i,1}(t) - a(t)\frac{\nu(t) - \phi_{i,2}(t)}{c(t)}\phi_{i,1}^{(-1)}(t) \\ (t \in \{1, 3, 5, ...\}), \end{cases}$$
(8)

where  $\phi_{i,1}(t) \in \{-1,1\}^n$  and  $\phi_{i,2}(t) \in \mathbb{R}$ . The symbols  $a(t) \in \mathbb{R}$  and  $c(t) \in \mathbb{R}$  are controller gains. The symbol  $\sigma_i(t) := [\sigma_{i,1}(t), ..., \sigma_{i,n}(t)]^{\mathrm{T}} \in \{-1,1\}^n$  is a random variable, where  $\sigma_{i,j}(t)$   $(i \in \{1, ..., N\}, j \in \{1, ..., n\}, t \in \{0, 1, ...\})$  independently obeys the Bernoulli distribution with outcome  $\pm 1$  equal probabilities. The collection of the random variables is denoted by  $\sigma(t) := [\sigma_1(t)^{\mathrm{T}}, ..., \sigma_N(t)^{\mathrm{T}}]^{\mathrm{T}} \in \{-1, 1\}^{nN}$ . The BC law with these definitions is a solution to the main problem.

**Theorem 1 (BC law [2])** For the multi-agent system  $\Sigma$ , an objective function J(x), and the BC law in (5), (6), (7), and (8), x(t) converges to a (possibly samplepath-dependent) solution to  $\partial_x J(x) = 0$  with probability 1 if the following conditions hold.

(c1) An objective function  $J : \mathbb{R}^{nN} \to \mathbb{R}$  is defined as

$$J(x(t)) := \rho(||x||) J_{obj}(x) + (1 - \rho(||x||)) x^{T} x, \quad (9)$$
  

$$\rho(||x||) := \begin{cases} 1 & (||x|| \le l_{1}), \\ 0 & (||x|| \ge l_{2}), \end{cases} \quad (10)$$

where J(x) and  $J_{obj}(x)$  are nonnegative  $C^2$  continuous on  $\mathbb{R}^{nN}$ , and there exists a solution to  $\partial_x J(x) = 0$  for some  $l_1$  and  $l_2$ .

(c2) The compact connected internally chain transitive invariant sets<sup>1</sup> of a gradient system  $\dot{z}(\tau) = -\partial_z J(z(\tau))$  are included in the solution set to the

<sup>&</sup>lt;sup>1</sup> Consider the system  $\dot{z}(\tau) = f(z(\tau))$  and one of its invariant sets,  $\mathbb{S}_z$ . The invariant set  $\mathbb{S}_z$  is said to be internally

equation  $\partial_z J(z) = 0$  (i.e., to  $\partial_x J(x) = 0$ ), and there exists an asymptotically stable equilibrium for the gradient system, where  $z(\tau) \in \mathbb{R}^{nN}$  and the stability is in the Lyapunov sense.

(c3) The controller gains satisfy a(t) = a(t+1) > 0and c(t) = c(t+1) > 0 for every  $t \in \{0, 2, 4, ...\}$ (i.e., a(0) = a(1) > 0, a(2) = a(3) > 0,... and the same is true of c(t)),  $\lim_{t\to\infty} a(t) = 0$ ,  $\sum_{t=0}^{\infty} a(t) = \infty$ ,  $\lim_{t\to\infty} c(t) = 0$ ,  $\sum_{t=0}^{\infty} (a(t)/c(t))^2 < \infty$ , and  $\sum_{t=0}^{\infty} a(t)c(t)^2 < \infty$ .

**Remark 2** Convergence to a solution to  $\partial_x J(x) = 0$ indicates that (4) (approximately) holds. Thus, the BC law is a solution to the main problem.

**Remark 3** The BC law is based on a two-stage transition of the state x(t). At the time  $t \in \{0, 2, 4, ...\}$ , the agent  $A_i$  receives the broadcast signal  $\nu(t)$  and takes the random action  $u_i(t) = c(t)\sigma_i(t)$ . Such a random action is essential in order to approximate the gradient of J(x)based on SPSA [18]. The second term in (8) corresponds to the approximate gradient (in an average sense) by applying Taylor's theorem [2]

$$\begin{aligned} \forall t \in \{0, 2, 4, \ldots\}, \\ & \mathbf{E}\Big[\frac{\nu(t+1) - \phi_{i,2}(t+1)}{c(t+1)} \phi_{i,1}^{(-1)}(t+1)\Big|_{x(t)}\Big] \\ & = \mathbf{E}\Big[\frac{J(x(t) + c(t)\sigma(t)) - J(x(t))}{c(t)} \sigma_i^{(-1)}(t)\Big|_{x(t)}\Big] \\ & = \partial_{x_i}J(x(t)) + O(c(t)), \text{ as } c(t) \to 0, \end{aligned}$$
(11)

where  $\nu(t+1) = J(x(t+1))$  from (5),  $x_i(t+1) = x_i(t) + c(t)\sigma_i(t)$  from (1) and (8),  $[\phi_{i,1}(t+1)^T, \phi_{i,2}(t+1)] = [\sigma_i(t)^T, J(x(t))]$  from (6) and (7), and c(t+1) = c(t) from the condition (c3) hold. The gradient becomes the moving direction of  $A_i$  for minimizing J(x). At the time  $t+1 \in \{1,3,5,\ldots\}$ ,  $A_i$  moves in such a direction while canceling the random action  $c(t+1)\phi_{i,1}(t+1) = c(t)\sigma_i(t)$ .

**Remark 4** In the condition (c1),  $J_{obj} : \mathbb{R}^{nN} \to \mathbb{R}$  describes the achievement degree of a coordination task to be realized, such as coverage, rendezvous, and assignment tasks. The function  $\rho(||x||)$  switches J(x) such that  $J(x) = J_{obj}(x)$  for  $||x|| \leq l_1$  and  $J(x) = x^T x$  for  $||x|| \geq l_2$ . The term  $x^T x$  is introduced for bounding the size of x. The constants  $l_1$  and  $l_2$  determine the workspace of the coordination. There exists  $\rho(||x||)$  in (10) such that J(x) is  $C^2$  continuous, e.g., a spline type function [2].

**Remark 5** To evaluate J(x(t)) (if it is not directly observed), the global controller G requires the information

of all agents' states  $x(t) \in \mathbb{R}^{nN}$ , via some means, such as cameras, sensors, and direct all-to-one communications.

#### 2.3 Drawbacks of the BC law

For large-scale multi-agent systems, the BC law in Theorem 1 is advantageous because it satisfies the three requirements described in Section 2.2: communications between agents are not used, local controllers are identical for all agents, and the broadcast signal is identical. Such characteristics are efficient for constructing controllers that are not adversely affected by the system size. However, the BC law contains the following limitations.

- The control law using the random action  $c(t)\sigma_i(i)$  in (8) causes an agent to move an unavailing extra distance to approximate the gradient of the objective function J(x). If a broadcast system involves a disturbance (e.g., noise in observing the states of the agents), a large random action may be needed for the system to be robust against the disturbance. Furthermore, it is impractical to enforce random actions of agents in many applications because such a randomness invokes dangerous situations.
- The minimization process in (4) using the BC law can be slow because the state transition is a two-stage transition with randomness, as shown in (8).

In order to overcome the above limitations, the next section presents a novel control law, which is another solution to the main problem, such that the moving distances of agents are reduced and the minimization of J(x) is accelerated as compared with the BC law.

### 3 Proposed method: pseudo-perturbationbased broadcast control law

In this section, we propose the PBC law in order to overcome the drawbacks of the BC law. Section 3.1 presents an overview of the PBC law based on a simple but novel concept. Sections 3.2 and 3.3 derive the theoretical aspects that show the effectiveness of the PBC law as compared with the BC law.

# 3.1 Overview of the PBC law based on a simple but novel concept

The PBC law is derived by extending the BC law as follows. Recall that the agents under the BC law take the random actions  $u_i(t) = c(t)\sigma_i(t)$  in (8) in order to approximate the gradient of an objective function J(x), as explained in Remark 3. In order to avoid the unavailing random actions and improve the approximation accuracy of the gradient, a key concept is the introduction of the following *multiple virtual* random actions rather

chain transitive if, for each  $(z_0, z_f) \in \mathbb{S}_z \times \mathbb{S}_z$ ,  $\epsilon > 0 \in \mathbb{R}$ , and  $T > 0 \in \mathbb{R}$ , there exist  $m \in \mathbb{N}$  and  $(z_1, z_2, ..., z_m) \in \mathbb{S}_z^m$ such that  $||z(\tau, z_i) - z_{i+1}|| < \epsilon$  (i = 0, 1, ..., m) for some  $\tau \in [T, \infty)$ , where  $z_{m+1} := z_f$  and  $z(\tau, z_i)$  represents the state  $z(\tau)$  of the system for the initial state  $z(0) := z_i$  [3].

than the single physical random actions:

$$\hat{x}_{i}^{(k)}(t+1) := x_{i}(t) + \hat{u}_{i}^{(k)}(t), \quad (k = 1, 2, ..., K), \quad (12) 
\hat{u}_{i}^{(k)}(t) := c(t)\sigma_{i}^{(k)}(t), \quad (k = 1, 2, ..., K), \quad (13)$$

where  $\hat{u}_i^{(k)}(t)$  and  $\hat{x}_i^{(k)}(t+1)$  are a virtual input and a virtual predictive state of each agent  $A_i$ , respectively. The global controller G calculates  $\hat{u}_i^{(k)}(t)$  and  $\hat{x}_i^{(k)}(t+1)$  virtually using the information  $\{x_i(t), \sigma_i^{(k)}(t)\}$  sent from the agent  $A_i$ . The symbol K denotes the number of multiple actions. The K virtual inputs  $\hat{u}_i^{(k)}(t) \in \{-1, 1\}^n$ . Each component of the random variables  $\sigma_i^{(k)}(t) \in \{-1, 1\}^n$ . Each component of the random variable  $\sigma_i^{(k)}(t)$  independently obeys the Bernoulli distribution with outcome  $\pm 1$  equal probabilities. Let us define  $\hat{x}^{(k)} := [\hat{x}_1^{(k)T}, ..., \hat{x}_N^{(k)T}]^T \in \mathbb{R}^{nN}$  and  $\sigma^{(k)}(t) := [\sigma_1^{(k)}(t)^T, ..., \sigma_N^{(k)}(t)^T]^T$ .

Based on the multiple virtual random actions, the global controller G of the PBC law is proposed as follows:

$$\nu(t) = f_{\nu}(x(t), \phi(t), t)$$
  
$$:= \begin{bmatrix} J(\hat{x}^{(1)}(t+1)) - J(x(t)) \\ \vdots \\ J(\hat{x}^{(K)}(t+1)) - J(x(t)) \end{bmatrix} \in \mathbb{R}^{K}, \qquad (14)$$

where  $\nu^{(k)}(t) := J(\hat{x}^{(k)}(t+1)) - J(x(t))$  and  $\nu(t) := [\nu^{(1)}(t), ..., \nu^{(K)}(t)]^{\mathrm{T}}$ . Each local controller  $L_i$  with its state  $\phi_i(t)$  of the PBC law is proposed as

$$\phi_i(t) := \begin{bmatrix} \phi_i^{(1)}(t) \\ \vdots \\ \phi_i^{(K)}(t) \end{bmatrix} := \begin{bmatrix} \sigma_i^{(1)}(t) \\ \vdots \\ \sigma_i^{(K)}(t) \end{bmatrix} \in \{-1, 1\}^{nK} \quad (15)$$

$$u_{i}(t) = f_{u}(\phi_{i}(t), \nu(t), t)$$
  
:=  $-a(t) \frac{1}{K} \sum_{k=1}^{K} \frac{\nu^{(k)}(t)}{c(t)} \phi_{i}^{(k)(-1)}(t).$  (16)

Unlike the BC law, a function  $f_{\phi}(\phi_i(t), \nu(t), t)$  as in (7) is not needed because the state  $\phi_i(t)$  of the local controller is nothing but the random variables  $\sigma_i^{(k)}(t)$ (k = 1, ..., K). Note that  $f_{\nu}$  in (14) is indeed a function of x(t) and  $\phi(t)$  because  $\hat{x}^{(k)}(t+1)$  is given from  $x_i(t)$ and  $\phi_i^{(k)}(t) = \sigma_i^{(k)}(t)$ , as shown in (12) and (13). Three important properties of the PBC law are as follows.

First, the PBC is a solution to the main problem and overcomes the limitations of the BC. The objective function J(x) is minimized in a manner similar to the BC law. For the following reasons, the unavailing actions are reduced, and the agents' states are converged at least twice as fast as the BC law under some conditions. The physical random actions  $u_i(t) = c(t)\sigma_i(t)$  in (8) are no longer used for approximating the gradient of J(x). The approximation accuracy of the gradient of J(x) is enhanced by taking multiple actions with large K > 1. In a manner similar to (11), the right of (16) approximates the gradient as follows:

$$\mathbf{E}\left[\frac{1}{K}\sum_{k=1}^{K}\frac{\nu^{(k)}(t)}{c(t)}\phi_{i}^{(k)(-1)}(t)\Big|_{x(t)}\right] = \partial_{x}J(x(t)) + O(c(t)).$$
(17)

as  $c(t) \to 0$ . The operator  $(1/K) \sum_{k=1}^{K} \{\cdot\}$  in (17) reduces the randomness. These properties are shown theoretically in Sections 3.2 and 3.3.

Second, the communication volume of the PBC law is slightly greater than that of the BC law. The global controller G of the PBC law broadcasts K-dimensional signal  $\nu(t)$ , whereas the broadcast signal of the BC law is one-dimensional. In the PBC law, each agent  $A_i$  transmits the signal  $\{x_i(t), \phi_i(t)\} = \{x_i(t), \sigma_i^{(1)}(t), ..., \sigma_i^{(K)}(t)\}$ . The BC law requires that  $A_i$  transmits only  $x_i(t)$  or that G observes  $x_i(t)$  for all  $i \in \{1, ..., N\}$ , in general (see Remark 5). However, these differences between the two laws are not critical. The number K of multiple actions is (significantly) smaller than the number N of agents. The information volume is not so different between  $\{x_i(t), \phi_i(t)\}$  and  $x_i(t)$  because  $\sigma_{i,1}^{(k)}(t) = \pm 1$  is binary (2-bit) and is significantly smaller than a quantized real vector  $x_i(t) \in \mathbb{R}^n$ .

Third, the advantages of the BC law for large-scale multi-agent systems are preserved in the PBC law. The PBC law satisfies the three requirements described in Section 2.1, which are that communications between agents are not used, that local controllers are identical for all of the agents, and that the broadcast signal is identical.

The following subsections derive the theoretical aspects to show that the PBC law improves the control performance, such as the moving distances of agents and the convergence speed for minimizing an objective function J(x), as compared with the BC law. Section 3.2 analyzes the convergence and performance improvement of the PBC law. The effect of taking multiple actions with K > 1 is analyzed in Section 3.3.

## 3.2 Theoretical analysis of the PBC law: convergence and performance improvement

This subsection presents a theoretical analysis of the PBC law proposed in Section 3.1. In the following, Theorem 6 shows that the PBC law is a solution to the main problem. For the case in which the number K of multiple random actions is 1, Theorems 8 and 10 indicate

that the control performance of the PBC law is enhanced compared with the BC law.

The condition (c3) is modified for the PBC law:

(c3') The controller gains satisfy a(t) > 0 and c(t) > 0 for every  $t \ge 0$ ,  $\lim_{t\to\infty} a(t) = 0$ ,  $\sum_{t=0}^{\infty} a(t) = \infty$ ,  $\lim_{t\to\infty} c(t) = 0$ ,  $\sum_{t=0}^{\infty} (a(t)/c(t))^2 < \infty$ , and  $\sum_{t=0}^{\infty} a(t)c(t)^2 < \infty$ .

We first show that the PBC law is a solution to the main problem.

**Theorem 6 (Convergence of the PBC law)** For the multi-agent system  $\Sigma$ , an objective function J(x), and the PBC law in (14), (15), and (16) with any K > 0, if the conditions (c1), (c2), and (c3') hold, then x(t)converges to a (possibly sample-path-dependent) solution to  $\partial_x J(x) = 0$  with probability 1.

**PROOF.** The proof is given in Appendix A.

**Remark 7** As explained in Remark 2, convergence to a solution to  $\partial_x J(x) = 0$  indicates that (4) (approximately) holds, and thus the PBC is a solution to the main problem. The PBC retains the essential feature of the BC law.

In the following, we show that the PBC law is superior to the BC law in terms of the convergence speed and the moving distances of agents in the case of K = 1. The symbols  $\{\cdot\}|_{\text{PBC}}$  and  $\{\cdot\}|_{\text{BC}}$  denote the results obtained by applying the PBC and BC laws, respectively, where the symbols can be omitted if we discuss either the PBC or the BC. The following theorem shows that coordination tasks can be performed twice as fast.

**Theorem 8 (Achieving tasks at twice the speed)** Suppose that K = 1,  $a(t)|_{PBC} = a(2t)|_{BC}$ ,  $c(t)|_{PBC} = c(2t)|_{BC}$ , and  $\sigma^{(1)}(t)|_{PBC} = \sigma(2t)|_{BC}$  hold for all  $t \ge 0$ . Then, the following relations hold

$$x(t)|_{\text{PBC}} = x(2t)|_{\text{BC}}, \quad \forall t \ge 0, \tag{18}$$

$$J(x(t))|_{\text{PBC}} = J(x(2t))|_{\text{BC}}, \quad \forall t \ge 0.$$
 (19)

**PROOF.** The proof is given in Appendix B.

**Remark 9** The speed for achieving the task of the PBC law with K = 1 is twice that of the BC law. An intuitive reason for this result is that the two-stage transition of the BC law explained in Remark 3 is simply combined into a one-stage transition. We next analyze the moving distances of agents. Let us define the total moving distance:

$$D(t) := \sum_{s=0}^{t-1} \sum_{i=1}^{N} \|x_i(s+1) - x_i(s)\| = \sum_{s=0}^{t-1} \sum_{i=1}^{N} \|u_i(s)\|.$$
(20)

The following theorem shows that the moving distance  $D(t)|_{\rm PBC}$  of the PBC law is reduced compared with the moving distance  $D(2t)|_{\rm BC}$  of the BC law in statistical, deterministic, and probabilistic senses.

**Theorem 10 (Reduction of the moving distance)** Suppose that, K = 1,  $a(t)|_{PBC} = a(2t)|_{BC}$ ,  $c(t)|_{PBC} = c(2t)|_{BC}$ , and  $\sigma^{(1)}(t)|_{PBC} = \sigma(2t)|_{BC}$  hold for all  $t \ge 0$ . Let a binary function  $\mathcal{P}(t)$  be 1 if J(x) is (locally) convex or quasi-convex on the minimum convex set including all possible  $x(t+1)|_{BC}$ , and otherwise let be 0. Then, for all  $t \ge 1$ , the following relations hold:

$$E[D(2t)|_{BC} - D(t)|_{PBC}] \ge \sqrt{n}N \sum_{\tau=0}^{t-1} \mathcal{P}(2\tau)c(2\tau)|_{BC}$$
$$\ge 0, \tag{21}$$

$$D(2t)|_{\rm BC} - D(t)|_{\rm PBC} \ge 0, \ \forall \sigma^{(1)}(0)|_{\rm PBC}, ..., \sigma^{(1)}(t)|_{\rm PBC}, (22)$$

$$\Pr\{D(2t)|_{BC} > D(t)|_{PBC}\} \ge 1 - \prod_{\tau=0}^{t-1} \left(1 - \frac{\mathcal{P}(2\tau)}{2}\right).$$
(23)

**PROOF.** The proof is given in Appendix C.

**Remark 11** The local convexity of J(x) increases the effectiveness of the PBC law because the binary function  $\mathcal{P}(2t)$  is activated from 0 to 1 by the local convexity. If J(x) is convex or quasi-convex for all x, (21) and (23) reduce to  $\mathrm{E}[D(2t)|_{\mathrm{BC}} - D(t)|_{\mathrm{PBC}}] \geq \sqrt{n}N\sum_{\tau=0}^{t-1}c(2\tau)|_{\mathrm{BC}}$  and  $\mathrm{Pr}\{D(2t)|_{\mathrm{BC}} > D(t)|_{\mathrm{PBC}}\} \geq 1 - (1/2^t)$ , respectively.

**Remark 12** In Theorem 10, it is reasonable that the time scales of the PBC and BC laws are normalized as t and 2t, respectively, because  $x(t)|_{PBC} = x(2t)|_{BC}$  and  $J(x(t))|_{PBC} = J(x(2t))|_{BC}$  hold in Theorem 8.

3.3 Theoretical analysis of the PBC law: performance improvement by multiple actions

This subsection presents a theoretical analysis of the PBC law for the case of taking multiple actions, i.e., K > 1. Theorems 13 and 15 and Proposition 17 give results for control performance improvement with K > 1.

To the best of our knowledge, few studies with respect to SPSA using multiple random numbers have been analyzed theoretically. The asymptotic properties of SPSA using multiple random numbers are discussed in [18]. It is also well known that the variance of the average of Ksamples is decreased 1/K times compared to the variance of one sample. This property is applied to the approximate gradient of J(x), i.e., the input in (16):

$$\operatorname{Cov}[u_i(t)|_{x(t)}] = \frac{1}{K} \big( \operatorname{Cov}[u_i(t)|_{x(t)}] \big|_{K=1} \big).$$
(24)

Using a large K enhances the approximation accuracy of the gradient in terms of the covariance.

In the following, some theorems are derived to discuss the influence of K > 1. We derive a theorem to show that increasing the value of K is efficient for enhancing the control performance per step of the PBC law.

## **Theorem 13 (Performance improvement per step)** For a given x(t), if the PBC law is applied, the following relations hold

(i) If J(x) is (locally) convex on the minimum convex set including all possible x(t+1) for the given x(t), the reduction of J(x) per step is enhanced by increasing K

$$\forall 1 \le K_b < K_a < \infty, \\ \mathbf{E}[J(x(t+1))|_{x(t)}]|_{K=K_a} \le \mathbf{E}[J(x(t+1))|_{x(t)}]|_{K=K_b},$$
(25)

where the equality in (25) is possible only if J(x) is not strictly convex but convex. The inequality sign in (25) is reversed if J(x) is concave.

(ii) For any  $\kappa \ge 1 \in \mathbb{R}$ , the power of the moving distance per step is reduced by increasing K

$$\begin{aligned} \forall 1 &\leq K_b < K_a < \infty, \\ & \mathbf{E} \Big[ \sum_{i=1}^{N} \| x_i(t+1) - x_i(t) \|^{\kappa} \Big|_{x(t)} \Big] \Big|_{K=K_a} \\ & \leq \mathbf{E} \Big[ \sum_{i=1}^{N} \| x_i(t+1) - x_i(t) \|^{\kappa} \Big|_{x(t)} \Big] \Big|_{K=K_b}, \end{aligned}$$
(26)

where the equality in (26) is possible only if  $\kappa = 1$ .

#### **PROOF.** The proof is given in Appendix D.

**Remark 14** This theorem means that the per-step control performance in terms of the speed for minimizing J(x) and the moving distance is statistically improved by using a large K. If J(x) is locally concave around x(t), setting K = 1 is suitable for minimizing J(x) quickly. Theorem 13 focuses on the control performance at each step, which is independent of other steps. Based on Theorem 13, we derive a result for a suitable objective function J(x) such that the control performance over multiple steps is improved.

## Theorem 15 (Performance over multiple steps)

For any  $t \geq 1$  and a given x(0), let  $\mathbb{S}_x$  be the minimum convex set including all possible x(0), x(1),..., and x(t). Suppose that, for all  $\sigma^{(k)}(m)$  $(m \in \{0,...,t-1\}, k \in \{1,...,K\})$  and for all  $s \in \{0,...,t\}, J_{s,t}(\tilde{x}) := J(x(t))|_{x(s)=\tilde{x}}$  and  $D_{s,t}(\tilde{x}) :=$  $(D(t) - D(s))|_{x(s)=\tilde{x}}$  are (locally) convex in  $\tilde{x} \in \mathbb{S}_x$ . The PBC law enhances the reductions of J(x(t)) and D(t) by increasing K

$$\begin{aligned} \forall 1 \le K_b < K_a < \infty, \\ \mathbf{E}[J(x(t))]|_{K=K_a} \le \mathbf{E}[J(x(t))]|_{K=K_b}, \quad (27) \\ \forall 1 < K_b < K_a < \infty, \end{aligned}$$

$$E[D(t)]|_{K=K_a} \le E[D(t)]|_{K=K_b},$$
(28)

where the equalities in (27) and (28) are possible only if  $J_{s,t}(\tilde{x})$  and  $D_{s,t}(\tilde{x})$  are not strictly convex but convex in  $\tilde{x}$ , respectively.

**PROOF.** The proof is given in Appendix E.

**Remark 16** Although Theorem 15 considers the time steps from 0 to t, it can be applied to the time steps from  $\tau$  to  $\tau + t$  for a given  $x(\tau)$  and  $\tau \ge 0$  in a similar manner.

Theorem 15 shows that the performance of J(x(t)) and D(t) are enhanced over multiple steps for suitable objective functions J(x). There are some choices of J(x) satisfying the assumptions in Theorem 15. One such class of J(x) is the convex quadratic function class, as described below.

**Proposition 17 (Convex quadratic functions)** If J(x) is convex and quadratic in x, the assumptions in Theorem 15 hold for any x(0) (it is possible that  $J_{s,t}(\tilde{x})$  and  $D_{s,t}(\tilde{x})$  is not strictly convex but convex).

**PROOF.** The proof is given in Appendix F.

**Remark 18** Although Proposition 17 focuses on convex quadratic functions, the PBC law can be successfully applied to various non-convex objective functions, which are locally convex and/or quadratic. Section 4 demonstrate the PBC law for non-convex objective functions.

Theorem 15 and Proposition 17 indicate that convex quadratic objective functions are more suitable than other objective functions. Some locally/globally convex quadratic objective functions for coordination tasks are introduced in the following. (i) Coverage control. The following objective function for coverage control is locally convex and quadratic in x

$$J_{\rm obj}(x) = \frac{\mathbb{V}_q}{N_q} \sum_{j=1}^{N_q} \min_{i \in \{1, \dots, N\}} \|q_j - x_i\|^2, \qquad (29)$$

where  $q_j$  are uniformly distributed on the state space, and  $\mathbb{V}_q := \int_{\mathbb{S}_q} 1 dq$  is the *n*-dimensional volume of a workspace  $\mathbb{S}_q \subset \mathbb{R}^n$   $(q \in \mathbb{S}_q)$ . The operator  $(\mathbb{V}_x/N_q) \sum_{j=1}^{N_q} \{\cdot\}$  approximates  $\int_{\mathbb{S}_q} \{\cdot\} dq$ .

(ii) Rendezvous control with formation selection. The following objective function for a rendezvous control is locally convex and quadratic in x

$$J_{\rm obj}(x) = \min_{\theta \in \mathbb{S}_{\theta}} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \|x_i - x_j - r_{i,j}(\theta)\|^2,$$
(30)

because, for all i and j, the L2-norms are quadratic convex and their sum is quadratic convex. The symbol  $r_{i,j}(\theta) \in \mathbb{R}^n$  is a relative target position between the agents i and j. Various target formations are constructed from  $r_{i,j}(\theta)$  parameterized by a global formation parameter  $\theta \in \mathbb{S}_{\theta}$ . Assuming that  $\mathbb{S}_{\theta}$  is finite, an optimal  $\theta$  minimizing  $J_{obj}(x)$ is selected from among  $\mathbb{S}_{\theta}$ . The PBC law can select such an (local) optimal formation with  $\theta$  by minimizing  $J_{obj}(x)$ , whereas standard distributed control cannot optimize  $\theta$ . If  $\mathbb{S}_{\theta}$  includes only one element,  $J_{obj}(x)$  in (30) reduces to a global convex objective function for well-known rendezvous control.

(iii) Assignment control. For given target locations  $y_i \in \mathbb{R}^n \ (i = 1, ..., N)$ , the following objective function for an assignment task is globally convex and quadratic in x

$$J_{\rm obj}(x,S) = \sum_{i=1}^{N} \|x_i - y_{S_i}\|^2, \qquad (31)$$

where  $S_i \in \{1, ..., N\}$  and  $S_i \neq S_j$  for any  $i \neq j$ . The indices  $S := [S_1, ..., S_N]$  indicate the pairs of target locations and agents to be assigned. Some assignment algorithms determine suitable indices. For example, the Hungarian algorithm gives optimal indices satisfying arg min<sub>S</sub>  $J_{\text{obj}}(x, S)$  with the complexity  $O(N^3)$  [13].

The above-mentioned functions in (29) and (30) include the operator min $\{\cdot\}$ , which invokes indifferentiable points on the state space. If the  $C^2$  continuity in the condition (c1) is strictly concerned, a modified log-sumexp form<sup>2</sup> can approximate the operator min<sub>j</sub> $\{\cdot\}$  by a smooth function  $(1/\varepsilon) \ln \sum_{i} \exp \varepsilon \{\cdot\}$ 

$$\frac{1}{\varepsilon}\ln n_j \le \frac{1}{\varepsilon}\ln\sum_{j=1}^{n_j}\exp\varepsilon f_j - \min_{j\in\{1,\dots,n_j\}}f_j \le 0, \quad (32)$$

for any  $f_j \in \mathbb{R}$   $(j = 1, 2, ..., n_j)$ , where  $\varepsilon < 0 \in \mathbb{R}$ . The term  $(1/\varepsilon) \ln n_j \leq 0$  is the approximation residual. Sufficiently small  $\varepsilon$  realizes a better approximation, i.e.,  $(1/\varepsilon) \ln n_j \to 0$  as  $\varepsilon \to -\infty$ .

#### 4 Numerical example

In this section, the effectiveness of the PBC law is evaluated in comparison with the BC law. The two types of multi-agent coordination tasks are demonstrated.

#### 4.1 Settings

The number of agents is set to N = 15 in the twodimensional state space, i.e., n = 2. The parameters in (10) were respectively set to  $l_1 = 100$  and  $l_2 = 101$ , which determine the size of the workspace. The terminal time of the simulation is set to 300. The controller gains are set such that the conditions (c3) and (c3') are satisfied

$$a(t)|_{\text{PBC}} = \frac{a_0}{(t+t_v)^{a_p}},$$
 (33)

$$c(t)|_{\text{PBC}} = \frac{c_0}{(t+t_v)^{c_p}},$$
 (34)

where  $a(2t)|_{BC} = a(2t + 1)|_{BC} = a(t)|_{PBC}$  and  $c(2t)|_{BC} = c(2t+1)|_{BC} = c(t)|_{PBC}$  in the case of the BC law. Based on conditions for the convergence/divergence of the *p*-series (see Section 8.1.2 [16]), the conditions (c3) and (c3') are satisfied if  $t_v > 0$ ,  $0 < a_p \le 1$ ,  $c_p > 0$ ,  $2a_p - 2c_p > 1$ , and  $a_p + 2c_p > 1$  hold. According to these conditions, the coefficients were set to  $a_0 = 2$ ,  $a_p = 0.7$ ,  $c_0 = 0.003$ ,  $c_p = 0.16$ , and  $t_v = 20$ .

## 4.2 Broadcast rendezvous control with formation selection

This subsection evaluates a rendezvous control task with the objective function  $J_{obj}(x)$  in (30). Recall that minimizing  $J_{obj}(x)$  in (30) automatically selects a best formation parameterized by  $\theta \in \mathbb{S}_{\theta}$ . The state trajectories are shown in Fig. 1. The  $\times$  and  $\circ$  symbols denote the initial states  $x_i(0)$  and the terminal states  $x_i(300)$ , respectively. The colored lines represent the trajectories of

 $<sup>^2\,</sup>$  The log-sum-exp form approximates the max function:

 $<sup>\</sup>max_{j} \varepsilon f_{j} \leq \ln \sum_{j=1}^{n_{j}} \exp \varepsilon f_{j} \leq \max_{j} \varepsilon f_{j} + \ln n_{j}$  (see Section 3.1.5 in [4]). For  $\varepsilon < 0$ , this inequality is transformed as  $\varepsilon \min_{j} f_{j} \leq \ln \sum_{j=1}^{n_{j}} \exp \varepsilon f_{j} \leq \varepsilon \min_{j} f_{j} + \ln n_{j}$ , which corresponds to (32).

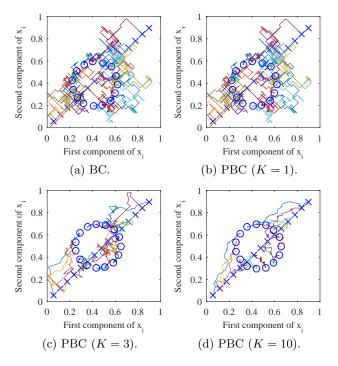


Fig. 1. State transitions at the terminal time t = 300 for the rendezvous control

the agents. As illustrated in Fig. 1, the initial state was set to  $x_i(0) = (0.9i/N)[1,1]^{\mathrm{T}}$ . The target formations were set such that  $r_{i,j}(\theta) = y_i(\theta) - y_j(\theta)$  in (30), where  $y_i(\theta) := 0.2[\cos(2\pi(i+\theta)/N), \sin(2\pi(i+\theta)/N)]^{\mathrm{T}}$  and  $\theta \in \mathbb{S}_{\theta} := \{1, ..., N\}$ . In Fig. 1, the BC law and the PBC law with K = 1 caused the large random actions for the agents. Applying large  $K \in \{3, 10\}$  yields smooth trajectories, reducing the unavailing random actions. Figure 2 shows the transitions of the objective function J(x(t)) and the moving distance D(t) with respect to the mean and the standard deviation (SD) for 500 trials (with different random seeds). The both J(x(t)) and D(t) were successfully reduced by the PBC law. Their means and SDs were decreased more by using large  $K \in \{3, 10\}$  as compared with those of the BC law and the PBC law with K = 1.

#### 4.3 Broadcast coverage control

A coverage control task is evaluated with the objective function  $J_{obj}(x)$  in (29), where  $q_j$  are sampled on the space  $[0, 1] \times [0, 1]$  at 0.01 intervals in each direction. Figure 3 shows the state trajectories of the PBC law with various  $K \in \{1, 3, 10\}$  compared with the BC law. The initial state was set to  $x_i(0) = [0.5+0.2\cos(2\pi i/N), 0.5+$  $0.2\sin(2\pi i/N)]^{T}$ , as illustrated in Fig. 3. The black lines indicate the Voronoi diagram composed by the agents. We see that increasing K reduces the unavailing random actions of the agents. Figure 4 shows J(x(t)) and D(t)in terms of the mean and the SD for 500 trials (with different random seeds). In Fig. 4(a), J(x(t)) was almost the same regardless of the value of K. It appears that

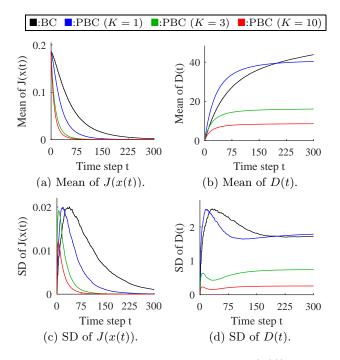


Fig. 2. Transitions of the objective function J(x(t)) and the moving distance D(t) for the rendezvous control

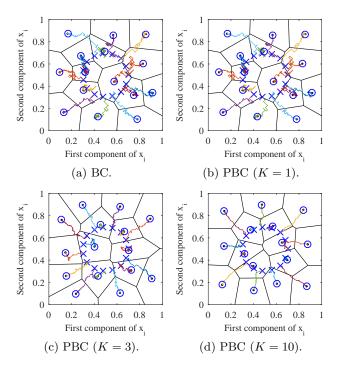


Fig. 3. State transitions at the terminal time t = 300 for the coverage control

J(x(t)) is not convex around the initial state x(0). Recall Theorem 13, which indicates that increasing K does not ensure the improvement of J(x(t)) for non-convex J(x(t)). Nevertheless, the PBC law reduces both J(x(t)) and D(t) compared with the existing BC law.

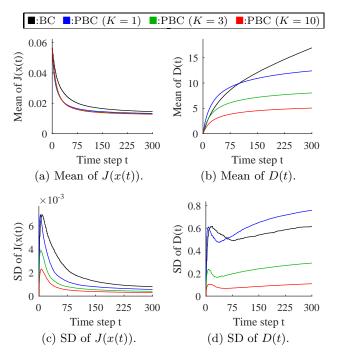


Fig. 4. Transitions of the objective function J(x(t)) and the moving distance D(t) for the coverage control

#### 5 Conclusion

In the present paper, we proposed the PBC law for achieving multi-agent coordination tasks with low communication volume without any agent-to-agent communication. To address drawbacks of the BC law discussed in Section 2.3, a simple but efficient solution was proposed in Section 3.1. The solution is to take multiple virtual actions of agents rather than a single physical random action. Sections 3.2 and 3.3 proved the theoretical aspects of the PBC law. Theorem 6 proved asymptotic convergence of multi-agent coordination with probability 1. Theorems 8 and 10 show that the PBC law is superior to the BC law in terms of the convergence speed and the moving distance in the case of taking a single action (K = 1). In the case of multiple actions (K > 1), Theorem 13 showed that the performance improvements per step become more effective by increasing the number Kof multiple actions. Theorem 15 proved that the performance improvements are retained over multiple steps for suitable objective functions. Such functions are convex quadratic in the state shown in Proposition 17. Section 4 demonstrated that the PBC law is superior to the BC law through two types of coordination tasks.

In the future, we intend to extend the PBC law for nonlinear and/or stochastic multi-agent systems. The PBC law will be applied to merging tasks of multiple vehicles on congested roads.

#### A Proof of Theorem 6

This theorem is proven in a manner similar to the case of the BC law [2]. Unfortunately, the difficulty in the case of the PBC law arises from taking multiple trials (K > 1). The proof of the convergence of the BC law in Appendices A.1 and A.2 of [2] is summarized (with slight modification) as follows.

#### Lemma 19 (Stochastic systems for the BC [2])

Let us define  $t_s := 2s$  for  $s \in \{0, 1, ...\}$ . Supposing that the conditions (c1), (c2), and (c3) hold,  $x(t_s)$  converges to a (possibly sample-path-dependent) solution to  $\partial_x J(x) = 0$  with probability 1 if the following conditions hold:

(c4) For all  $t_s$ ,  $x(t_s)$  obeys

$$x(t_{s+1}) = x(t_s) - a(t_s) \{\partial_x J(x(t_s)) + e_{\mathcal{M}}(t_{s+1}) + e_{\mathcal{B}}(t_{s+1}) \},$$
(A.1)

where  $e_{\mathbf{M}}(t_s) \in \mathbb{R}^{nN}$  is a random vector and  $e_{\mathbf{B}}(t_s) \in \mathbb{R}^{nN}$  is a bounded random vector satisfying  $e_{\mathbf{B}}(t_s) \to 0$  as  $t_s \to \infty$ .

- (c5)  $\sup_{t_s} ||x(t_s)|| < \infty$  holds with probability 1.
- (c6) For any  $m \in \{0, 1, ...\}$ , the stochastic process  $\{\sum_{s=0}^{m} a(t_s)e_{M}(t_{s+1})\}\$  is a square integrable martingale, and  $\sum_{s=0}^{m} a(t_s)^2 \mathbb{E}[\|e_{M}(t_{s+1})\|^2 |\mathcal{F}_s] < \infty$ holds with probability 1 for the filtration  $\mathcal{F}_s$  generated by  $\{x(t_0), x(t_{s'}), e_{M}(t_{s'}), e_{B}(t_{s'})| 1 \le s' \le s\}.$

We apply Lemma 19 to the PBC law. Let us define the following functions for brief notation:

$$g(t,\sigma) := \frac{J(x(t) + c(t)\sigma) - J(x(t))}{c(t)}\sigma^{(-1)},$$
 (A.2)

$$e(t) := \frac{1}{K} \sum_{k=1}^{K} g(t, \sigma^{(k)}(t)) - \mathbf{E} \Big[ \frac{1}{K} \sum_{k=1}^{K} g(t, \sigma^{(k)}(t)) \Big|_{x(t)} \Big].$$
(A.3)

By substituting (14), (15), (16), (17), and (A.3) into (1), the transition of x(t) under the PBC law is given by

$$x(t+1) = x(t) - a(t) \frac{1}{K} \sum_{k=1}^{K} g(t, \sigma^{(k)}(t))$$
  
=  $x(t) - a(t) \{\partial_x J(x(t)) + e(t) + O(c(t))\},$   
(A.4)

as  $c(t) \to 0$ . If the condition (c5) in Lemma 19 holds, then the term O(c(t)) is bounded and  $O(c(t)) \to 0$  as  $t \to \infty$  from the condition (c3'). The pair of the condition (c3) and the definition  $t_s := 2s$  in Lemma 19 can be replaced by the pair of the condition (c3') and  $t_s :=$ s. The condition (c4) then holds under the condition (c5) by regarding t, t + 1, e(t), and O(c(t)) as  $t_s, t_{s+1}, e_M(t_{s+1})$ , and  $e_B(t_{s+1})$ . We next prove the condition (c5). The system dynamics in (A.4) is transformed as follows:

$$\|x(t+1)\|^{2} = \left\|\frac{1}{K}\sum_{k=1}^{K} \{x(t) - a(t)g(t,\sigma^{(k)}(t))\}\right\|^{2}$$
  
$$\leq \|x(t) - a(t)g(t,\sigma^{(k_{*}(t))}(t))\|^{2}, \quad (A.5)$$

where  $k_*(t) := \arg \max_{k \in \{1,...,K\}} ||x(t) - a(t)g(t, \sigma^{(k)}(t))||^2$ By virtue of the above inequality (A.5), the following property in Appendix A.3 of [2] can be applied to (A.5).

**Lemma 20 (Boundedness of the state [2])** Let us define  $t_s := 2s$  for  $s \in \{0, 1, ...\}$ . Suppose that the conditions (c1), (c2), and (c3) hold. If the following relation

$$||x(t_{s+1})||^2 \le ||x(t_s) - a(t_s)g(t_s, \sigma(t_s))||^2, \qquad (A.6)$$

is satisfied, then  $\sup_{t_s} ||x(t_s)|| < \infty$  surely holds, where the pair of the condition (c3) and the definition  $t_s := 2s$ can be replaced by the pair of the condition (c3') and  $t_s := s$ .

Lemma 20 proves the condition (c5) in Lemma 19 by regarding t, t + 1, and  $\sigma^{(k_*(t))}(t)$  in (A.5) as  $t_s, t_{s+1}$ , and  $\sigma(t_s)$  in (A.6). The condition (c6) holds in a manner similar to the case of the BC law (see Appendix A.2 of [2] or Proposition 1 of [18]). Lemma 19 is thus applied to the PBC law. This completes the proof.  $\Box$ 

## B Proof of Theorem 8

This theorem is proven by mathematical induction. In the case of t = 0, (18) and (19) clearly hold. With (A.2), x(2(t+1)) by employing the BC law is given as follows [2]:

$$x(2(t+1))|_{\rm BC} = x(2t)|_{\rm BC} - a(2t)g(2t,\sigma(2t))|_{\rm BC}.$$
(B.1)

Using (A.4), x(t+1) under the PBC law is obtained as

$$x(t+1)|_{PBC} = x(t)|_{PBC} - a(t)g(t,\sigma^{(1)}(t))|_{PBC}.$$
 (B.2)

Since  $a(t)|_{\text{PBC}} = a(2t)|_{\text{BC}}$ ,  $c(t)|_{\text{PBC}} = c(2t)|_{\text{BC}}$ , and  $\sigma^{(1)}(t)|_{\text{PBC}} = \sigma(2t)|_{\text{BC}}$ , if (18) and (19) hold for t, they are also satisfied for t + 1. Therefore, (18) and (19) hold for all  $t \ge 0$ . This completes the proof.  $\Box$ 

#### C Proof of Theorem 10

Let us define the following function in order to simplify the notation:

$$\Delta J(t,\sigma) := J(x(t) + c(t)\sigma) - J(x(t)), \qquad (C.1)$$

Since  $\sigma_i = \sigma_i^{(-1)}$  and  $\|\sigma_i\| = \sqrt{n}$  hold for any  $i, D(2t)|_{BC}$  is given by

$$D(2t)|_{BC} = \sum_{\tau=0}^{t-1} \sum_{i=1}^{N} \|c(2\tau)\sigma_i(2\tau)\|\Big|_{BC} + \left\|c(2\tau)\sigma_i(2\tau) + a(2\tau)\frac{\Delta J(2\tau,\sigma(2\tau))}{c(2\tau)}\sigma_i^{(-1)}(2\tau)\right\|\Big|_{BC} = \sum_{\tau=0}^{t-1} \sqrt{n}N\Big(c(2\tau) + \left|c(2\tau) + a(2\tau)\frac{\Delta J(2\tau,\sigma(2\tau))}{c(2\tau)}\right|\Big)\Big|_{BC}.$$
(C.2)

Here, for any positive scalar coefficients  $\omega_1 > 0$  and  $\omega_2 > 0$ , the following expectation is derived:

$$\begin{split} \mathbf{E}[|\omega_{1}\Delta J(2\tau,\sigma) + \omega_{2}|] \\ &= \Pr\{\Delta J(2\tau,\sigma) \geq 0\}\mathbf{E}[|\omega_{1}\Delta J(2\tau,\sigma) + \omega_{2}||_{\Delta J(2\tau,\sigma) \geq 0}] \\ &+ \Pr\{\Delta J(2\tau,\sigma) < 0\}\mathbf{E}[|\omega_{1}\Delta J(2\tau,\sigma) + \omega_{2}||_{\Delta J(2\tau,\sigma) < 0}] \\ &\geq \Pr\{\Delta J(2\tau,\sigma) \geq 0\}\mathbf{E}[\omega_{1}\Delta J(2\tau,\sigma) + \omega_{2}|_{\Delta J(2\tau,\sigma) \geq 0}] \\ &- \Pr\{\Delta J(2\tau,\sigma) < 0\}\mathbf{E}[\omega_{1}\Delta J(2\tau,\sigma) + \omega_{2}|_{\Delta J(2\tau,\sigma) \geq 0}] \\ &= \mathbf{E}[|\omega_{1}\Delta J(2\tau,\sigma) < 0]\mathbf{E}[\omega_{1}\Delta J(2\tau,\sigma) + \omega_{2}|_{\Delta J(2\tau,\sigma) < 0}] \\ &+ (\Pr\{\Delta J(2\tau,\sigma) \geq 0\} - \Pr\{\Delta J(2\tau,\sigma) < 0\})\omega_{2} \\ &= \mathbf{E}[|\omega_{1}\Delta J(2\tau,\sigma)|] + (2\Pr\{\Delta J(2\tau,\sigma) \geq 0\} - 1)\omega_{2}, \end{split}$$

where note that  $|\omega_1 \Delta J(2\tau, \sigma) + \omega_2| \ge -(\omega_1 \Delta J(2\tau, \sigma) + \omega_2)$  and  $\Pr{\{\Delta J(2\tau, \sigma) < 0\}} = 1 - \Pr{\{\Delta J(2\tau, \sigma) \ge 0\}}$ hold. In order to clarify the probability  $\Pr{\{\Delta J(2\tau, \sigma) \ge 0\}}$  in (C.3), we show the relation

$$\Pr\{\Delta J(2\tau, \sigma(2\tau))|_{BC} \ge 0\} \ge \frac{\mathcal{P}(2\tau)}{2}, \qquad (C.4)$$

where  $\Delta J(2\tau, \sigma(2\tau)) = J(x(2\tau) + c(2\tau)\sigma(2\tau)) - J(x(2\tau))$  in (C.1). If a random vector  $\sigma$  is included in the set  $\{-1, 1\}^{nN}$ ,  $-\sigma$  is also included in  $\{-1, 1\}^{nN}$  because each component of  $\sigma$  is -1 or 1. If we assume that  $\Delta J(2\tau, \sigma) < 0$  and  $\Delta J(2\tau, -\sigma) < 0$  hold, then the relation

$$\max\{J(x(2\tau) + c(2\tau)\sigma), J(x(2\tau) - c(2\tau)\sigma)\} < J(x(2\tau)) = J(\frac{x(2\tau) + c(2\tau)\sigma}{2} + \frac{x(2\tau) - c(2\tau)\sigma}{2})$$
(C.5)

is satisfied. This inequality (C.5) contradicts the necessary condition for the (quasi-)convexity of J(x)on the minimum convex set including all possible  $x(2\tau + 1)|_{BC}$ . Thus, for any  $x(2\tau)|_{BC}$  and any  $\sigma \in \{-1,1\}^{nN}$ , if J(x) is convex or quasi-convex and  $\Delta J(2\tau,\sigma)|_{BC} < 0$  holds, then  $\Delta J(2\tau,-\sigma)|_{BC} \ge 0$  holds. Since the probability of  $\sigma$  is constant on  $\{-1,1\}^{nN}$ ,  $\Pr\{\Delta J(2\tau, \sigma(2\tau))|_{BC} \geq 0\} \geq \Pr\{\Delta J(2\tau, \sigma(2\tau))|_{BC} < 0\}, \text{ i.e., } \Pr\{\Delta J(2\tau, \sigma(2\tau))|_{BC} \geq 0\} \geq 1/2 \text{ holds if } J(x) \text{ is locally convex/quasi-convex, which yields (C.4).}$ 

Substituting the relation (C.3) with  $\omega_1 = a(2\tau)/c(2\tau)$ ,  $\omega_2 = c(2\tau)$ , and (C.4) into the expectation of  $D(2t)|_{\rm BC}$  in (C.2) yields

$$\begin{split} \mathbf{E}[D(2t)|_{\mathrm{BC}}] \\ \geq \sum_{\tau=0}^{t-1} \sqrt{n} N\Big(c(2\tau) + (\mathcal{P}(2\tau) - 1)c(2\tau) \\ &+ a(2\tau) \frac{\mathbf{E}[|\Delta J(2\tau, \sigma(2\tau))|]}{c(2\tau)}\Big)\Big|_{\mathrm{BC}} \\ = \sqrt{n} N \\ &\times \sum_{\tau=0}^{t-1} \Big(\mathcal{P}(2\tau)c(2\tau) + a(2\tau) \frac{\mathbf{E}[|\Delta J(2\tau, \sigma(2\tau))|]}{c(2\tau)}\Big)\Big|_{\mathrm{BC}}. \end{split}$$

$$(C.6)$$

On the other hand,

$$D(t)|_{\text{PBC}} = \sum_{\tau=0}^{t-1} \sum_{i=1}^{N} ||\sigma_i^{(1)}(\tau)|| - a(\tau) \frac{\Delta J(\tau, \sigma^{(1)}(\tau))}{c(\tau)} ||_{\text{PBC}} = \sqrt{nN} \sum_{\tau=0}^{t-1} |-a(\tau) \frac{\Delta J(\tau, \sigma^{(1)}(\tau))}{c(\tau)} ||_{\text{PBC}}, \quad (C.7)$$
$$E[D(t)|_{\text{PBC}} = \sqrt{nN} \sum_{\tau=0}^{t-1} a(\tau) \frac{E[|\Delta J(\tau, \sigma^{(1)}(\tau))|]}{c(\tau)} |_{\text{PBC}},$$

$$\frac{1}{\tau=0} \qquad c(\tau) \qquad \text{IPBC} \\ (C.8)$$

Recall the relation  $x(t)|_{\text{PBC}} = x(2t)|_{\text{BC}}$  in (18) and the assumptions that  $a(t)|_{\text{PBC}} = a(2t)|_{\text{BC}}$ ,  $c(t)|_{\text{PBC}} = c(2t)|_{\text{BC}}$ , and  $\sigma^{(1)}(t)|_{\text{PBC}} = \sigma(2t)|_{\text{BC}}$  for all  $t \ge 0$ . From (C.6) and (C.8),  $\text{E}[D(2t)|_{\text{BC}} - D(t)|_{\text{PBC}}]$  satisfies (21). The relation (22) is proven because of (C.2) and (C.7). Next, the following relation

$$\begin{aligned} \exists \tau \in \{0, ..., t-1\}, \text{ s.t. } \Delta J(2\tau, \sigma(2\tau))|_{\mathrm{BC}} &\geq 0\\ \Rightarrow \exists \tau \in \{0, ..., t-1\}, \text{ s.t.}\\ c(2\tau)|_{\mathrm{BC}} + \left| c(2\tau) + a(2\tau) \frac{\Delta J(2\tau, \sigma(2\tau))}{c(2\tau)} \right| \Big|_{\mathrm{BC}}\\ &= 2c(2\tau)|_{\mathrm{BC}} + \left| a(2\tau) \frac{\Delta J(2\tau, \sigma(2\tau))}{c(2\tau)} \right| \Big|_{\mathrm{BC}}\\ \Rightarrow D(2t)|_{\mathrm{BC}} > D(t)|_{\mathrm{PBC}} \end{aligned} \tag{C.9}$$

holds for the terms in (C.2). Therefore,  $\Delta J(2\tau, \sigma(2\tau))|_{BC} < 0, \forall \tau \in \{0, ..., t-1\}$  is the necessary condition for  $D(2t)|_{BC} = D(t)|_{PBC}$ . The relation (23) is proven because  $p(\Delta J(2\tau, \sigma(2\tau))|_{BC} < 0) \leq 1 - (\mathcal{P}(2\tau)/2)$  holds by transforming (C.4). This completes the proof.  $\Box$ 

#### D Proof of Theorem 13

We first prove the statement (i) for the case in which  $1 \leq K_b < K_a < \infty$ . For brevity of notation, let us define  $J_g(g) := J(x(t+1))|_{x(t+1)=x(t)-a(t)g}$  for a given  $x(t), g^{(k)} := g(t, \sigma^{(k)}(t))$  in (A.2), and a combination  $\mathcal{C} := {}_{K_a}C_{K_b}$ . For the given set  $\{1, ..., K_a\}$  of the index k in  $g^{(k)}$ , we consider the  $\mathcal{C}$  combinations to select  $K_b$  indexes. Let us define a vector  $\mathcal{I}_i \in \{1, ..., K_a\}^{K_b}$  as the array of the selected indexes in the *i*-th combination

$$\mathcal{I}_{i} := [\mathcal{I}_{i,1}, ..., \mathcal{I}_{i,K_{b}}]^{\mathrm{T}}, \ (i = 1, 2, ..., \mathcal{C}).$$
(D.1)

Each index  $k \in \{1, ..., K_a\}$  is evenly selected  $K_b C/K_a$ times because the total number of selected indexes is  $K_b C$ . Thus, dividing  $g^{(k)}$  by  $K_b C/K_a$  yields the following relation<sup>3</sup>

$$\frac{1}{K_a} \sum_{k=1}^{K_a} g^{(k)} = \frac{1}{K_a} \sum_{i=1}^{\mathcal{C}} \sum_{j=1}^{K_b} \frac{K_a g^{(\mathcal{I}_{i,j})}}{K_b \mathcal{C}} = \sum_{i=1}^{\mathcal{C}} \frac{1}{\mathcal{C}} \sum_{j=1}^{K_b} \frac{g^{(\mathcal{I}_{i,j})}}{K_b}.$$
(D.2)

Note that the function  $J_g(g)$  is (strictly) convex in g because J(x(t+1)) is (strictly) convex in x(t+1) and g = (x(t) - x(t+1))/a(t) is uniquely determined as the linear function of x(t+1). Since  $\sum_{i=1}^{\mathcal{C}} (1/\mathcal{C}) = 1$ , applying Jensen's inequality to convex  $J_g(g)$  yields

$$J_g\left(\sum_{i=1}^{\mathcal{C}} \frac{1}{\mathcal{C}} \sum_{j=1}^{K_b} \frac{g^{(\mathcal{I}_{i,j})}}{K_b}\right) \le \sum_{i=1}^{\mathcal{C}} \frac{1}{\mathcal{C}} J_g\left(\sum_{j=1}^{K_b} \frac{g^{(\mathcal{I}_{i,j})}}{K_b}\right). \quad (D.3)$$

Here, we consider the case in which  $J_g(g)$  is strictly convex. The equality in (D.3) holds only if  $\sum_{j=1}^{K_b} g^{(\mathcal{I}_{i,j})}/K_b$  is constant for all *i*, i.e.,  $g^{(k)}$  is constant for all *k*. Because of the strict convexity,  $g^{(k)} = g(t, \sigma^{(k)}(t)) \neq 0$  in (A.2) and the relation  $g^{(k)} = g^{(k')} \Leftrightarrow \sigma^{(k)} = \sigma^{(k')}$  hold. Thus, if  $J_g(g)$  is strictly convex, taking the expectation in (D.3) with respect to  $\sigma^{(k)}$  (the set of  $\sigma^{(k)}$  for each *k* is finite) excludes the equality from (D.3):

<sup>&</sup>lt;sup>3</sup> For example, if  $K_a = 4$  and  $K_b = 3$  are considered,  $\mathcal{C} = 4$ ,  $\mathcal{I}_i \in \{[1, 2, 3]^{\mathrm{T}}, [1, 2, 4]^{\mathrm{T}}, [1, 3, 4]^{\mathrm{T}}, [2, 3, 4]^{\mathrm{T}}\}, \text{ and } \frac{1}{4}(g^{(1)} + g^{(2)} + g^{(3)} + g^{(4)}) = \frac{1}{4}\{\frac{g^{(1)} + g^{(2)} + g^{(3)}}{3} + \frac{g^{(1)} + g^{(2)} + g^{(4)}}{3} + \frac{g^{(1)} + g^{(2)} + g^{(4)}}{3} + \frac{g^{(2)} + g^{(3)} + g^{(4)}}{3}\}$  hold.

Using (D.2) and (D.4) yields the following inequality:

$$\begin{split} \mathbf{E}[J(x(t+1)|_{x(t)}]|_{K=K_{a}} &= \mathbf{E}\Big[J_{g}\Big(\frac{1}{K_{a}}\sum_{k=1}^{K_{a}}g^{(k)}\Big)\Big|_{x(t)}\Big] \\ &= \mathbf{E}\Big[J_{g}\Big(\sum_{i=1}^{C}\frac{1}{C}\sum_{j=1}^{K_{b}}\frac{g^{(\mathcal{I}_{i,j})}}{K_{b}}\Big)\Big|_{x(t)}\Big] \\ &\leq \mathbf{E}\Big[\sum_{i=1}^{C}\frac{1}{C}J_{g}\Big(\sum_{j=1}^{K_{b}}\frac{g^{(\mathcal{I}_{i,j})}}{K_{b}}\Big)\Big|_{x(t)}\Big] \\ &= \frac{1}{C}\sum_{i=1}^{C}\mathbf{E}\Big[J_{g}\Big(\sum_{k=1}^{K_{b}}\frac{g^{(k)}}{K_{b}}\Big)\Big|_{x(t)}\Big] \\ &= \mathbf{E}\Big[J_{g}\Big(\frac{1}{K_{b}}\sum_{k=1}^{K_{b}}g^{(k)}\Big)\Big|_{x(t)}\Big] = \mathbf{E}[J(x(t+1)|_{x(t)}]|_{K=K_{b}}, \\ (\mathbf{D}.5) \end{split}$$

where the equality of  $\leq$  in (D.5) is possible only if J(x) is not strictly convex but convex. If  $J_g(g)$  is concave in g, the inequality sign in (D.5) is reversed. The statement (i) therefore holds.

We next prove the statement (ii). In the above proof of the statement (i), let us redefine  $J_g(g) := \sum_{i=1}^N ||x_i(t+1) - x_i(t)||^{\kappa}|_{x_i(t+1) = x_i(t) - a(t)g_i} = a(t)^{\kappa} \sum_{i=1}^N ||g_i||^{\kappa}$  for a given x(t), where  $g =: [g_i^T, ..., g_N^T]^T$ . Since the norm  $||g_i||$  is convex in  $g_i$  [4],  $J_g(g)$  is convex in g for  $\kappa = 1$  and is strictly convex in g for  $\kappa > 1$ . The inequality in (D.5) holds in a similar manner, which leads to the statement (ii). This completes the proof.  $\Box$ 

#### E Proof of Theorem 15

We consider the case in which  $1 \leq K_b < K_a < \infty$ . We prove this theorem by focusing on the dynamic programming of stochastic dynamical systems. Let us define

$$V_s(x(s)) := \mathbb{E}[J(x(t))|_{x(s)}],$$
 (E.1)

where  $V_t(x(t)) = J(x(t))$  holds. We obtain

$$V_{t-2}(x(t-2)) = \mathbf{E}[\mathbf{E}[V_t(x(t))|_{x(t-1)}]|_{x(t-2)}]$$
  
=  $\mathbf{E}[V_{t-1}(x(t-1))|_{x(t-2)}],$  (E.2)

Iterating the above transformation yields the recurrence formula of  $V_s(x(s))$ 

$$V_s(x(s)) = \mathbb{E}[V_{s+1}(x(s+1))|_{x(s)}].$$
 (E.3)

Note that for all  $s \in \{0, ..., t-1\}$ ,  $V_{s+1}(x(s+1)) = E[J(x(t))|_{x(s+1)}]$  is convex in x(s+1) because, for all  $\sigma^{(k)}(m)$   $(m \in \{0, ..., t-1\}, k \in \{1, ..., K\}), J_{s+1,t}(\tilde{x}) := J(x(t))|_{x(s+1)=\tilde{x}}$  is convex in  $\tilde{x}$  and the expectation of convex  $J_{s+1,t}(\tilde{x})$  is also convex [4]. Let us define x(s+1) = x(s+1) = x(s+1)

 $1, K_a)|_{x(s)} := x(s+1)|_{x(s),K=K_a}$  and  $V_s(x(s), K_a) := E[J(x(t))|_{x(s)}]|_{K=K_a}$ . Applying Theorem 13 to the expectation with respect to  $\sigma^{(k)}(s)$  yields

for all  $s \in \{0, ..., t-1\}$ . Meanwhile, by applying Theorem 13 to the expectation with respect to  $\sigma^{(k)}(t-1)$ , the following relation holds for all x(t-1, K)

$$V_{t-1}(x(t-1,K),K_a) = \mathbb{E}[J(x(t,K_a))|_{x(t-1,K)}]$$
  

$$\leq \mathbb{E}[J(x(t,K_b))|_{x(t-1,K)}] = V_{t-1}(x(t-1,K),K_b).$$
(E.5)

Thus,

(E.7)

holds. Using (E.4) and (E.6) provides

$$V_{t-2}(x(t-2), K_a) = \mathbb{E}[V_{t-1}(x(t-1, K_a), K_a)|_{x(t-2)}]$$
  

$$\leq \mathbb{E}[V_{t-1}(x(t-1, K_b), K_a)|_{x(t-2)}]$$
  

$$\leq \mathbb{E}[V_{t-1}(x(t-1, K_b), K_b)|_{x(t-2)}]$$
  

$$= V_{t-2}(x(t-2), K_b).$$

Using (E.4) and (E.7) yields

$$V_{t-3}(x(t-3), K_a) = \mathbb{E}[V_{t-2}(x(t-2, K_a), K_a)|_{x(t-3)}]$$

$$\leq \mathbb{E}[V_{t-2}(x(t-2, K_b), K_a)|_{x(t-3)}]$$

$$\leq \mathbb{E}[V_{t-2}(x(t-2, K_b), K_b)|_{x(t-3)}]$$

$$= V_{t-3}(x(t-3), K_b).$$
(E.8)

Iterating the above process from t - 1 to 0 yields

$$V_0(x(0), K_a) = \mathbb{E}[J(x(t))|_{x(0)}]|_{K=K_a}$$
  

$$\leq V_0(x(0), K_b) = \mathbb{E}[J(x(t))|_{x(0)}]|_{K=K_b}.$$
(E.9)

Since Theorem 13 is used, the equalities in (E.4), (E.5), (E.6), (E.7), (E.8), and (E.9) are possible only if  $J_{s,t}(\tilde{x}) := J(x(t))|_{x(s)=\tilde{x}}$  is not strictly convex but convex in  $\tilde{x}$ . Therefore, the relation (E.9) is equivalent to (27). The relation (28) can be proven as described above by applying D(t) and the conditional expectations  $E[\cdot|_{x(0),\ldots,x(s)}]$  rather than J(x(t)) and  $E[\cdot|_{x(s)}]$ , respectively. In this case,  $V_{s+1}(x(s+1)) := E[D(t)|_{x(0),\ldots,x(s+1)}]$  is indeed convex in x(s+1) because  $D_{s+1,t}(\tilde{x}) := (D(t) - D(s+1))|_{x(s+1)=\tilde{x}}$  and  $D(s+1)|_{x(s+1)=\tilde{x}}$  for each fixed  $x(0),\ldots,x(s)$  are convex in  $\tilde{x}$ . This completes the proof.  $\Box$ 

#### F Proof of Proposition 17

The input  $u_i(s)$  including  $\nu^{(k)}(s)$  is linear in x(s) because J(x) is quadratic in x. Therefore, for all  $s \in \{0, ..., t\}$ ,

 $s' \in \{0, ..., t\}$ , and  $\sigma^{(k)}(m)$   $(m \in \{0, ..., t-1\}, k \in \{1, ..., K\}), x(s)$  is linear in x(s'). Since J(x(t)) is convex in  $x(t), J_{s,t}(\tilde{x})$  is convex in  $\tilde{x}$ . Since D(t) is convex in x(s') for all  $s' \in \{0, ..., t\}, D_{s,t}(\tilde{x})$  is convex in  $\tilde{x}$ . This completes the proof.  $\Box$ 

#### References

- [1] Mohd Ashraf Ahmad, M.A. Rohani, Raja Mohd Taufika Raja Ismail, M.F. Mat Jusof, Mohd Helmi Suid, and Ahmad Nor Kasruddin Nasir. A model-free pid tuning to slosh control using simultaneous perturbation stochastic approximation. In Proc. of 2015 IEEE International Conference on Control System, Computing and Engineering, pages 331–335, 2015.
- Shun-ichi Azuma, Ryota Yoshimura, and Toshiharu Sugie. Broadcast control of multi-agent systems. *Automatica*, 49(8):2307–2316, 2013.
- [3] Vivek S Borkar. Stochastic approximation: a dynamical systems viewpoint. Cambridge University Press, New York, NY, USA, 2008.
- [4] Stephen Boyd and Lieven Vandenberghe. Convex Optimization. Cambridge University Press, New York, NY, USA, 2004.
- [5] Ádám Halász, M Selman Sakar, Harvey Rubin, Vijay Kumar, George J Pappas, et al. Stochastic modeling and control of biological systems: the lactose regulation system of escherichia coli. *IEEE Trans. on Automatic Control*, 53(Special Issue):51–65, 2008.
- [6] Yuji Ito, Md Abdus Samad Kamal, Takayoshi Yoshimura, and Shun-ichi Azuma. On pseudo-perturbation-based broadcast control of multi-agent systems. In Proc. of the 2018 American Control Conference, pages 2877–2882, 2018.
- [7] Meng Ji, Shun-ichi Azuma, and Magnus Egerstedt. Role assignment in multi-agent coordination. International Journal of Assistive Robotics and Mechatronics, 7(1):32–40, 2006.
- [8] Meng Ji and Magnus B Egerstedt. Distributed coordination control of multi-agent systems while preserving connectedness. 2007.
- [9] Md Abdus Samad Kamal and Junichi Murata. Coordination in multiagent reinforcement learning systems by virtual reinforcement signals. *International Journal of Knowledge Based Intelligent Engineering Systems*, 11(3):181–191, 2007.
- [10] Kengo Kishimoto, Masaya Yamada, and Masayuki Jinno. Cooperative inter-infrastructure communication system using 700 mhz band. SEI Technical Review, (78):19–23, 2014.
- [11] Magdi S. Mahmoud. Decentralized Systems with Design Constraints. Springer-Verlag London, 2011.
- [12] Sonia Martínez, Jorge Cortés, and Francesco Bullo. Motion coordination with distributed information. *IEEE Control* Systems, 27(4):75–88, 2007.
- [13] James Munkres. Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial* and Applied Mathematics, 5(1):32–38, 1957.
- [14] M. H. Mohamad Nor, Z. H. Ismail, and M. A. Ahmad. Broadcast control of multi-agent systems for assembling at unspecified point with collision avoidance. *Journal of Telecommunication, Electronic and Computer Engineering*, 8(11):75–79, 2016.
- [15] M. H. Mohamad Nor, Z. H. Ismail, and M. A. Ahmad. Broadcast control of multi-agent systems with instability

phenomenon. In Proc. of 2016 IEEE 6th International Conference on Underwater System Technology: Theory & Applications, pages 7–12, 2016.

- [16] Andrei D. Polyanin and Alexander V. Manzhirov. Handbook of Mathematics for Engineers and Scientists. CRC Press, 2006.
- [17] Wei Ren and Yongcan Cao. Distributed Coordination of Multi-agent Networks. Springer-Verlag London, 2011.
- [18] James C Spall. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Trans. on Automatic Control*, 37(3):332–341, 1992.
- [19] James C. Spall. Developments in stochastic optimization algorithms with gradient approximations based on function measurements. In Proc. of the 1994 Winter Simulation Conference, pages 207–214, 1994.
- [20] James C. Spall. An overview of the simultaneous perturbation method for efficient optimization. JOHNS HOPKINS APL TECHNICAL DIGEST, 19(4):482–492, 1998.
- [21] Jun Ueda, Lael Odhner, and H Harry Asada. Broadcast feedback of stochastic cellular actuators inspired by biological muscle control. *The International Journal of Robotics Research*, 26(11-12):1251–1265, 2007.
- [22] Myung-Gon Yoon, Jung-Ho Moon, and Tae Kwon Ha. Broadcasting stabilization for dynamical multi-agent systems. International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering, 8(6):879–882, 2014.
- [23] Myung-Gon Yoon, Peter Rowlinson, Dragoš Cvetković, and Zoran Stanić. Controllability of multi-agent dynamical systems with a broadcasting control signal. Asian Journal of Control, 16(4):1066–1072, 2014.