SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

November 2025, Vol. 68, Iss. 11, 210205:1–210205:19
https://doi.org/10.1007/s11432-025-4657-5

Special Topic: Mean-Field Game and Control of Large Population Systems: From Theory to Practice

Stackelberg mean field game-based decentralized collective control for large-scale population of HVACs under grid balancing

Yuexi ZHANG[1], Shaohua YANG[2*] & Tielong SHEN[1]

[1]*School of Control Science and Engineering, Dalian University of Technology, Dalian 116024, China*
[2]*State Key Laboratory of Internet of Things for Smart City, University of Macau, Macao 999078, China*

Received 31 March 2025/Revised 10 July 2025/Accepted 28 October 2025/Published online 5 November 2025

Abstract This article presents a Stackelberg mean-field game (MFG)-based approach for the decentralized collective control of a large-scale population of heating, ventilation, and air conditioning (HVAC) systems, aiming to achieve grid balancing. We first give the thermodynamic model of HVAC systems in detail and build a Stackelberg MFG framework. On this basis, we derive a decentralized control strategy at the number of participants $N \to \infty$ and achieve the $(\varepsilon_1, \varepsilon_2)$-Stackelberg equilibrium. Finally, this paper validates the effectiveness and feasibility of the proposed control strategies through simulations under various weight coefficients and initial conditions.

Keywords Stackelberg mean field game, Stackelberg equilibrium, power grid balance, HVAC loads, collective control

Citation Zhang Y X, Yang S H, Shen T L. Stackelberg mean field game-based decentralized collective control for large-scale population of HVACs under grid balancing. Sci China Inf Sci, 2025, 68(11): 210205, https://doi.org/10.1007/s11432-025-4657-5

1 Introduction

The rapid integration of renewable energy sources (RES), such as photovoltaics and wind power, has significantly transformed global power systems [1]. However, the inherent intermittency of RES introduces substantial power fluctuations, thereby imposing heightened requirements on operating reserves to maintain power balance [2]. In this context, demand response (DR) mechanisms leveraging flexible load resources have emerged as crucial solutions for providing operating reserves through demand-side regulation [3]. Among various flexible resources, the heating, ventilation, and air conditioning (HVAC) systems, due to their widespread deployment and inherent thermal inertia characteristics, are usually used as one of the flexible loads that can be regulated according to the need for grid stable operation. Particularly, in the case where a large-scale population of HVAC units is collectively connected to a grid, the collective behavior of the HVACs will influence grid stability and load balancing [4]. Hence, in order to integrate a large number of HVACs into the power grid for demand response, aggregating the HVAC units to achieve grid scale operating reserves by handling the collective behavior of the HVACs is a challenging issue.

However, current approaches predominantly adopt centralized control architectures, where an aggregator coordinates large HVAC clusters through real-time communication infrastructure [5]. In this way, broadcasting command signals to a large number of HVAC units and getting feedback from them in real-time operation are necessary [6]. When the population of HVAC units becomes extremely large, this traditional broadcasting approach might be resource-intensive and impractical. Instead of broadcasting control signals to every individual HVAC unit by the grid side, in a decentralized control fashion, decision-making is distributed to regional or local controllers [7]; i.e., each HVAC unit makes decisions based on its own local or an embedded local model.

Meanwhile, as a distributed strategy decision theory for managing a large scale population of dynamical agents, the mean-field game (MFG) or mean-field type control theory has been proposed during the last two decades [8, 9]. The foundational work on MFGs, initially proposed by Lasry and Lions [10] and

* Corresponding author (email: shaohuayang2022@gmail.com)

independently by Huang et al. [11], has proven effective and viable for analyzing weakly coupled stochastic controlled systems with mean-field interactions, leading to the establishment of an approximate Nash equilibrium. Recently, several studies have tackled application issues of the MFG theory in decentralized control. In [12], a distributed strategy is developed using MFGs for large-scale stochastic multi-agent systems characterized by coupled cost functions. Moreover, Ref. [13] presented a distributed solution based on the MFG framework to address the collaborative control problem of large-scale temperature-controlled loads participating in grid frequency support. In addition, Ref. [14] proposed a distributed power control algorithm grounded in MFG theory for non-orthogonal multiple access systems. The above research studies have made positive progress.

In general, under an MFG setting, the collective behavior of a large number of agents is obtained by adding a penalty term to the consensus of the state response. In other words, the cost function should involve a term on the difference between the individual agent's state and the expectation of the distribution of the agents. If it is aimed to lead the collective behavior closer to a desired state, then it is necessary to add a penalty related to the difference between the expectation and the reference state. It means that the reference state has to be broadcast to all agents. As a hierarchical MFG, the Stackelberg MFG framework enables only the leader, which is a pre-specified agent with the reference state, and a large number of followers to decide the strategy with the pre-designed leader's strategy or propagated neighboring information. The Stackelberg differential game was originally proposed by [15], and then was extended to (ε_1, ε_2)-Stackelberg MFG or social optimization problem by [16, 17]. Several examples of application in practice have also been reported in [18–20].

In this paper, we address the decentralized collective control problem for a large-scale population of HVACs under grid balancing with the Stackelberg mean-field game setting. A large number of HVACs, here called followers, with a pre-specified HVAC leader, respectively, will be targeted. The desired reference state is the command from the grid for the supply-demand power balance, which is only used by the leader. According to the hierarchical game architecture of Stackelberg MFG, the strategy of the followers is designed under the assumption that the leader prioritizes strategy decisions. Finally, it will be shown that a $(\varepsilon_1, \varepsilon_2)$-Stackelberg equilibrium, between the followers and the leader, and among the followers, can be achieved. As a result, the collective behavior of the overall HVACs is closer to the desired grid response.

The remainder of this paper is organized as follows: Section 2 delineates the system modeling framework and fundamental assumptions. Section 3 derives a suite of centralized Stackelberg equilibrium strategies through rigorous game-theoretic analysis. Subsequently, Section 4 presents a distributed strategy design methodology, accompanied by formal proofs establishing the existence of a Stackelberg equilibrium. The verification of the proposed distributed strategy is given in Section 5, followed by concluding remarks in Section 6.

The notation used in this paper is shown as follows. $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P}, T > 0)$ denotes a complete filtered probability space satisfying the usual conditions. $\mathbb{E}$ denotes the expectation with respect to $\mathbb{P}$. $|\cdot|$ denotes an absolute value or the modulus of a vector. Let $L^2(0,T;\mathbb{R}) := \{h : [0,T] \to \mathbb{R} | \int_0^T |h(t)|^2 \mathrm{d}t < \infty\}$. $L^2_{\mathcal{F}}(0,T;\cdot)$ is the space of all $\mathcal{F}$-adapted processes $f(\cdot)$ satisfying the square-integrability condition: $\mathbb{E}[\int_0^T |f(t)|^2 \mathrm{d}t] \leqslant \infty$. $L^2_{\mathcal{F}_t}(\Omega;\cdot)$ is the space of all $\mathcal{F}_t$-measurable random variables, for $t \in [0,T]$. $B_i,\ i = 1, 2, \cdots, N$ are a sequence of one-dimensional independent Wiener processes defined on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P})$. Let σ-algebra $\mathcal{F}_t := \sigma\left(\{\beta_0^i, B_i(s), s \leqslant t,\ i = 1, 2, \cdots, N\}\right)$ and $\mathcal{F}_t^i := \sigma\left(\{\beta_0^i, B_i(s), s \leqslant t\}\right),\ i = 1, 2, \cdots, N$.

2 Problem formulation

2.1 Motivation

The system considered in this paper can be illustrated as in Figure 1, where a large number of HVAC units are connected to a node of the power grid, and the electric power for operating individual HVAC units is supplied by the grid. As is well-known, the influence of the collective behavior of the HVACs will be non-negligible when the number of HVACs becomes sufficiently large. The collective behavior of a large-scale population of HVACs poses challenges for the power supplier not only in power balance but also in grid stability. However, handling the collective behavior of a large number of HVACs is not an easy task due to the uncertainties in individual HVAC operations that are usually forced by the customer's

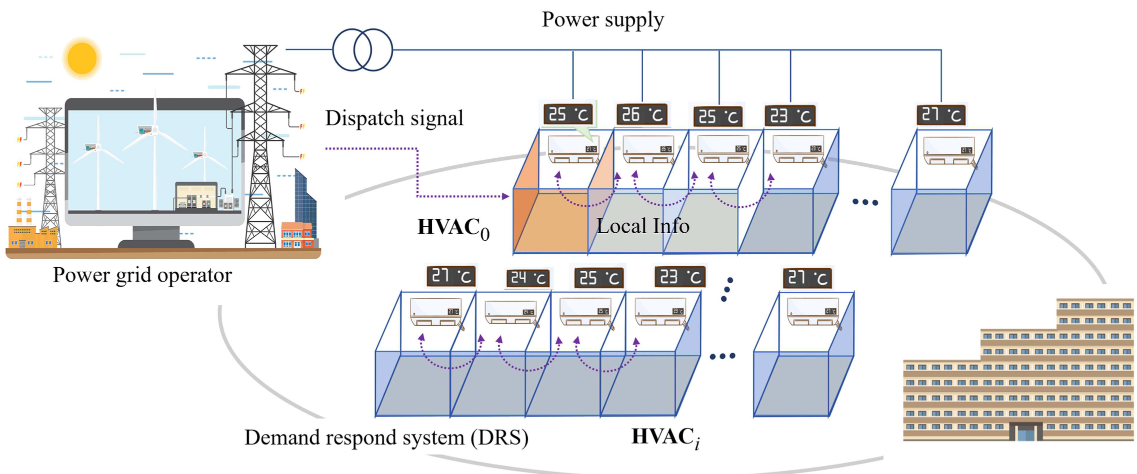

Figure 1 The overall dispatch framework of the HVAC-based DRS.

action under the me-first principle with variety and uncertainties.

As a traditional power grid operation technique, power dispatch primarily relies on centralized regulation by power plants mainly, whose regulation capacity is no longer sufficient to meet the dispatch requirement in the scenario considered in this paper with a large-scale population of HVACs. The decentralized control of each HVAC with consideration of collective behavior is a feasible approach to deal with the dispatch problem for the system involving a large number of HVACs. Of course, under the constraint of power supplier satisfaction, each HVAC's power consumption will conflict with that of others. In other words, for decision-making of dispatch, the power supplier must handle the power demand caused by the collective behavior of the larger-scale population of HVACs. Conversely, the total power consumption of the large-scale population of HVACs should match the constraint of the power supply once the power supplier decides the dispatch plan. This paper will address the latter issue. Namely, the following issues will be challenged. For a given power dispatch schedule, find a decentralized control strategy for individual HVACs under the decentralized operation of HVACs, and the following two goals are achieved.

- Total power consumption of the collective demand of the large-scale population of HVACs is close to the dispatched power in the sense of mean-field approximation.
- The large number of HVACs achieves a Nash equilibrium under the specified reward function, more precisely, the Stackelberg Nash equilibrium.

2.2 Dynamic model of HVACs

Consider a large-scale multi-room system with $N+1$ homogeneous HVAC units where N is a sufficiently large number and goes to infinity in the ideal case. Each room is assigned one HVAC unit for regulating the indoor temperature, and all HVACs are physically connected to the power grid as shown in Figure 1.

The thermodynamic model of HVAC has been constructed in detail in [21]. On this basis, considering the stochastic noise, for the i-th HVAC$_i$ $(i=0,1,\ldots,N)$, the dynamics can be represented by the following stochastic differential equation:

$$\begin{cases} \mathrm{d}T_0(t) = \dfrac{1}{c_A\rho_A V}[H_{\mathrm{g},0}(t) - H_{1,0}(t)]\mathrm{d}t, \\ \mathrm{d}T_i(t) = \dfrac{1}{c_A\rho_A V}[H_{\mathrm{g},i}(t) - H_{1,i}(t)]\mathrm{d}t + \sigma \mathrm{d}B_i(t),\ i=1,2,\cdots,N \end{cases} \tag{1}$$

by choosing the temperature T_i to which the HVAC$_i$ is connected, where c_A is the air heat capacity; ρ_A is the air density; V is the room's volume; B_i denotes the independent Wiener processes which represent the environment noise, and the constant σ diffusion coefficient represents the magnitude of randomness. It should be noted that the unit HVAC$_0$ is detached from others, and it is assumed that there is no stochastic noise in the dynamics. Without loss of generality, the controller is directly connected to the grid, and the assumption of no stochastic noise is for the simplicity of subsequent derivation. We call this HVAC$_0$ the leader and the remaining whole units as followers. The result is easy to extend to the case where the leader's dynamics involve stochastic noise.

This dynamical model of the HVAC units is from the energy conservation law with the heat gain from outside $H_{\mathrm{g},i}$ and the heat loss of the room $H_{\mathrm{l},i}$ $(i = 0, 1, 2, \cdots, N)$. The heat gain can be calculated by heat transfer from air leakages and the building envelope, which can be expressed as follows:

$$H_{\mathrm{g},i}(t) = U_h A_s(T_{\mathrm{o}} - T_i(t)) + c_A \rho_A V n(T_{\mathrm{o}} - T_i(t)), \tag{2}$$

where U_h is the heat transfer coefficient; A_s is the envelope's surface area; T_{o} is the ambient temperature; n denotes the air exchange times; and Vn is the product of V and n. The heat loss can be expressed as

$$H_{\mathrm{l},i}(t) = \eta[\alpha_i(t)R_{\mathrm{P}} + P_0], \tag{3}$$

where η is the coefficient of performance of HVAC, which implies the relationship between the input power and heat supply (cooling or heating); P_0 is the initial power of HVAC; α_i is the variable for online regulation of power used to inject the heat into the room, which is defined as

$$\alpha_i(t) = \frac{\Delta P_i(t)}{R_{\mathrm{P}}}, \tag{4}$$

where $\Delta P_i(t)$ are the actual regulation power of HVAC$_i$; R_{P} is the admissible regulation capacity. In practical terms, the range of $\alpha_i \in [0, 1]$.

For the temperature T_i, we consider a targeted horizon $[T_{\min}, T_{\max}]$. We normalize T_i by defining

$$\beta_i(t) = \frac{2}{T_{\max} - T_{\min}} T_i(t) - \frac{T_{\max} + T_{\min}}{T_{\max} - T_{\min}}, \tag{5}$$

which maps the horizon $[T_{\min}, T_{\max}]$ map to $[-1, 1]$, and usually β_i is called the comfort state. The lower bound -1 and upper bound 1 represent cold and heat tolerance limits, respectively. Substituting (5) for (1)–(3), the dynamics of the comfort state can be formulated as

$$\begin{cases} \mathrm{d}\beta_0(t) = \left[-\dfrac{(U_h A_s + c_A \rho_A V n)}{c_A \rho_A V} \beta_0(t) - \dfrac{\eta R_{\mathrm{P}}}{\frac{T_{\max} - T_{\min}}{2} c_A \rho_A V} \alpha_0(t) + \dfrac{(U_h A_s + c_A \rho_A V n)(T_{\mathrm{o}} - \frac{T_{\max} + T_{\min}}{2}) + \eta P_0}{\frac{T_{\max} - T_{\min}}{2} c_A \rho_A V} \right] \mathrm{d}t, \\ \mathrm{d}\beta_i(t) = \left[-\dfrac{(U_h A_s + c_A \rho_A V n)}{c_A \rho_A V} \beta_i(t) - \dfrac{\eta R_{\mathrm{P}}}{\frac{T_{\max} - T_{\min}}{2} c_A \rho_A V} \alpha_i(t) + \dfrac{(U_h A_s + c_A \rho_A V n)(T_{\mathrm{o}} - \frac{T_{\max} + T_{\min}}{2}) + \eta P_0}{\frac{T_{\max} - T_{\min}}{2} c_A \rho_A V} \right] \mathrm{d}t \\ \qquad + \sigma \mathrm{d}B_i(t), \ i = 1, 2, \cdots, N. \end{cases} \tag{6}$$

For brevity, in the following, $G_{th} = U_h A_s + c_A \rho_A V n$ and $C_{th} = c_A \rho_A V$ are denoted as the thermal conductance and thermal capacitance coefficient, respectively [21]. On this basis, the state-space equation of the HVAC i can be given as

$$\begin{cases} \mathrm{d}\beta_0(t) = [a\beta_0(t) + b\alpha_0(t) + c]\mathrm{d}t, \\ \mathrm{d}\beta_i(t) = [a\beta_i(t) + +b\alpha_i(t) + c]\mathrm{d}t + \sigma \mathrm{d}B_i(t), \end{cases} \tag{7}$$

where

$$a = -\frac{G_{th}}{C_{th}}, \ b = -\frac{\eta R_{\mathrm{P}}}{\frac{T_{\max} - T_{\min}}{2} C_{th}}, \ c = \frac{G_{th}(T_o - \frac{T_{\max} + T_{\min}}{2}) + \eta P_0}{\frac{T_{\max} - T_{\min}}{2} C_{th}}, \ i = 0, 1, \cdots, N.$$

The initial value of the leader and the ith follower are given as $\beta_0(0) = \beta_0^0$, $\beta_i(0) = \beta_0^i$, $i = 1, 2, \cdots, N$, and $\{\beta_0^i\}$, $i = 1, 2, \cdots, N$ are a sequence of independent and identically distributed (i.i.d., for short) random variables.

Based on this model of system dynamics, it is obvious that if we want to maintain β_i at a desired value β_i^e, the system must be stabilized at the equilibrium forced by the regulation power with a corresponding state value $\alpha_i^{\mathrm{ref}} = \frac{-c - a\beta_i^e}{b}$, $i = 0, 1, \cdots, N$. To design a decentralized regulation strategy, consider the corresponding error dynamics by defining

$$u_i = \alpha_i - \alpha_i^{\mathrm{ref}}, \ x_i = \beta_i - \beta_i^e, \ i = 0, 1, \cdots, N. \tag{8}$$

Then, the dynamical equation (7) becomes the following form:

$$\begin{cases} \mathrm{d}x_0(t) = [ax_0(t) + bu_0(t)]\mathrm{d}t, \ x_0(0) = \zeta_0, \\ \mathrm{d}x_i(t) = [ax_i(t) + bu_i(t)]\mathrm{d}t + \sigma \mathrm{d}B_i(t), \ x_i(0) = \zeta_i, \ i = 1, 2, \cdots, N. \end{cases} \tag{9}$$

2.3 Stackelberg MFG formulation

To achieve the goal of decentralized collective control, we formulate the strategy design problem as the Stackelberg game within the leader-follower framework. In detail, the leader unit HVAC_0 makes policy decisions first, and then the followers make decisions for the best response according to the leader's policy decision. Under this leader-follower structure, the leader decides the leader's strategy by considering grid-friendly power consumption so that the whole system of the large-scale population of HVACs achieves the so-called $(\varepsilon_1, \varepsilon_2)$-Stackelberg equilibrium as an ideal limitation. For ease of reference, we begin with a brief review of the Stackelberg game setting and concepts.

Let

$$b_0 := b_0\left(t, x_0(t), \mu(x_i(t)), u_0(t)\right) : [0,T] \times \mathbb{R} \times \mathcal{P}_2(\mathbb{R}) \times \mathcal{U}_0 \to \mathbb{R},$$

$$\sigma_0 := \sigma_0\left(t, x_0(t), \mu(x_i(t)), u_0(t)\right) : [0,T] \times \mathbb{R} \times \mathcal{P}_2(\mathbb{R}) \times \mathcal{U}_0 \to \mathbb{R},$$

$$b_i := b_i\left(t, x_i(t), \mu(x_i(t)), u_i(t), u_0(t)\right) : [0,T] \times \mathbb{R} \times \mathcal{P}_2(\mathbb{R}) \times \mathcal{U}_i \times \mathcal{U}_0 \to \mathbb{R},$$

$$\sigma_i := \sigma_i\left(t, x_i(t), \mu(x_i(t)), u_i(t), u_0(t)\right) : [0,T] \times \mathbb{R} \times \mathcal{P}_2(\mathbb{R}) \times \mathcal{U}_i \times \mathcal{U}_0 \to \mathbb{R},$$

where $u_i(\cdot) \in \mathcal{U}_i,\ i = 0, 1, \cdots, N$ are permissible strategies; $B_i(t),\ i = 0, 1, \cdots, N$ are independent Wiener processes; μ represent the empirical distribution; $b_i,\ i = 0, 1, \cdots, N$ are drift function; and $\sigma_i,\ i = 0, 1, \cdots, N$ are diffusion function. Consider a leader and the followers with dynamics described by

$$\begin{cases} \mathrm{d}x_0(t) = b_0 \mathrm{d}t + \sigma_0 \mathrm{d}B_0(t),\ x_0(0) = \zeta_0, \\ \mathrm{d}x_i(t) = b_1 \mathrm{d}t + \sigma_1 \mathrm{d}B_i(t),\ x_i(0) = \zeta_i. \end{cases} \tag{10}$$

The cost functions of the leader and followers are defined as follows, respectively:

$$J_0\left(u_0(\cdot), u^N(\cdot)\right) := \mathbb{E}\left\{\int_0^T f_0\left(t, x_0(t), \mu(x_i(t)), u_0(t)\right) \mathrm{d}t + g_0\left(x_0(T), \mu(x_i(T))\right)\right\}, \tag{11}$$

$$J_i\left(u_i(\cdot), u_{-i}(\cdot), u_0(\cdot)\right) := \mathbb{E}\left\{\int_0^T f_1\left(t, x_i(t), \mu(x_i(t)), u_i(t), u_0(t)\right) \mathrm{d}t + g_1\left(x_i(T), \mu(x_i(T))\right)\right\}, \tag{12}$$

where $u^N(\cdot) := (u_1(\cdot), \cdots, u_N(\cdot))$, $u_{-i}(\cdot) := (u_1(\cdot), \cdots, u_{i-1}(\cdot), u_{i+1}, \cdots, u_N(\cdot))$. The optimization problem is as follows:

$$\min_{u_0 \in \mathcal{U}_0} J_0\left(u_0(\cdot), u^N(\cdot)\right), \quad \min_{u_i \in \mathcal{U}_i} J_i\left(u_i(\cdot), u_{-i}(\cdot), u_0(\cdot)\right), \quad i = 1, \ldots, N, \tag{13}$$

which is subject to dynamical constraints (10).

Definition 1 ($(\varepsilon_1, \varepsilon_2)$-Stackelberg equilibrium, see [22])**.** A set of strategies $(u_0^*(\cdot), u_1^*(\cdot), \cdots, u_N^*(\cdot))$ is an $(\varepsilon_1, \varepsilon_2)$-Stackelberg equilibrium with respect to $\{J_i,\ i = 0, 1, \cdots, N\}$ if the following holds.

(1) For a given strategy of the leader $u_0(\cdot) \in \mathcal{U}_0$, $u^{N*}(\cdot) = (u_1^*(\cdot), \cdots, u_N^*(\cdot))$, $u_i^*(\cdot) \in \mathcal{U}_i$ constitutes an ε_1-Nash equilibrium; i.e., there exists a constant $\varepsilon_1 \geqslant 0$ such that for all $i = 0, 1, \cdots, N$,

$$J_i\left(u_i^*(\cdot), u_{-i}^*(\cdot), u_0(\cdot)\right) \leqslant \inf_{u_i(\cdot) \in \mathcal{U}_i} J_i\left(u_i(\cdot), u_{-i}^*(\cdot), u_0(\cdot)\right) + \varepsilon_1. \tag{14}$$

(2) There exists a constant $\varepsilon_2 \geqslant 0$ such that

$$J_0\left(u_0^*(\cdot), u^{N*}\left[\cdot; u_0^*(\cdot)\right]\right) \leqslant \inf_{u_0(\cdot) \in \mathcal{U}_0} J_0\left(u_0(\cdot), u^{N*}\left[\cdot; u_0(\cdot)\right]\right) + \varepsilon_2. \tag{15}$$

Now, our problem can be formulated as the Stackelberg game by defining the cost functionals for the leader and the followers (10) as follows, respectively:

$$J_0(u_0(\cdot), u^N(\cdot)) := \frac{1}{2}\mathbb{E}\left\{\int_0^T \left[r_1\left(x_0(t) - x^{(N)}(t)\right)^2 + r_2\left(x_0(t) - x_0^{\mathrm{ref}}(t)\right)^2 + r_3 u_0^2(t)\right] \mathrm{d}t\right\}, \tag{16}$$

where $x_0^{\mathrm{ref}} = 0$ is determined by α_0^{ref} given by the power grid and for the ith follower, the cost functionals (12) are replaced by

$$J_i(u_i(\cdot), u_{-i}(\cdot), u_0(\cdot)) := \frac{1}{2}\mathbb{E}\left\{\int_0^T \left[q_1\left(x_i(t) - x^{(N)}(t)\right)^2 + q_2\left(x_i(t) - x_0(t)\right)^2 + q_3 u_i^2(t)\right]\mathrm{d}t\right\}, \quad (17)$$

where $x^{(N)}(t) := \frac{1}{N}\sum_{i=1}^N x_i(t)$; $r_i, i = 1,2,3$ and $q_i, i = 1,2,3$ are the weighting coefficients for different terms and we have $r_1, r_2 \geqslant 0$, $r_3 > 0$ and $q_1, q_2 \geqslant 0$, $q_3 > 0$.

It is worth noting that the cost functional J_0 for the leader aims to strike a trade-off between the collective consensus of the whole HVACs and the tracking of the given reference trajectory x_0^{ref} which is pre-specified by power supplier as dispatch plan α^{ref}, according to the HVAC's inverse-dynamics and the scale of the population. On the other hand, the followers aim to reach a consensus on the overall collective behavior, with consideration for consistency with the leader. In the following section, a solution for this Stackelberg game will be derived first and then decentralized into the strategies.

3 Centralized strategies

3.1 The strategies for the followers

In this subsection, we solve the mean-field Nash game for the N followers under an arbitrary given strategy of the leader $u_0(\cdot) \in L^2(0,T;\mathbb{R})$. Once $u_0(\cdot)$ is given, the state response $x_0(\cdot)$ of the leader is determined by the dynamical model (9) according to the initial state ζ_0. The mean-field game strategy for the followers can be found by solving the following optimization problem. For the sake of simplicity, the time variable t will be omitted without ambiguity.

(**P1**): Minimize $J_i, i = 1,\cdots,N$ of (17) over $u_i(\cdot) \in L^2_{\mathcal{F}}(0,T;\mathbb{R})$.

Theorem 1. Let $u_0 \in L^2(0,T;\mathbb{R})$ be given. For the initial value $\zeta_i,\ i = 1,\cdots,N$, if (**P1**) admits an optimal control $\hat{u}_i \in L^2_{\mathcal{F}}(0,T;\mathbb{R}), i = 1,\cdots,N$, then the adapted solution $(\hat{x}_i, \hat{p}_i, \hat{q}_i^j, i = 1,\cdots,N, j = 1,\cdot,N)$ to the Hamilton system

$$\begin{cases} \mathrm{d}\hat{x}_i = [a\hat{x}_i - b^2 q_3^{-1}\hat{p}_i]\mathrm{d}t + \sigma \mathrm{d}B_i, \\ \mathrm{d}\hat{p}_i = -\left[a\hat{p}_i + \left[q_1\left(1 - \dfrac{1}{N}\right) + q_2\right]\hat{x}_i - q_1\left(1 - \dfrac{1}{N}\right)\hat{x}^{(N)} - q_2 x_0\right]\mathrm{d}t + \displaystyle\sum_{j=1}^N \hat{q}_i^j \mathrm{d}B_j, \\ \hat{x}_i(0) = \zeta_i, \quad \hat{p}_i(T) = 0, \ i = 1,2,\cdots,N \end{cases} \quad (18)$$

satisfies

$$\hat{u}_i = -q_3^{-1} b\hat{p}_i, \quad \text{a.e., a.s.,} \quad i = 1,\cdots,N. \quad (19)$$

See Appendix A for the proof of Theorem 1.

Due to the difficulty in directly solving the Hamilton system (18), we consider the following parameterization for $\hat{p}_i(\cdot)$:

$$\hat{p}_i(\cdot) = P_N(\cdot)\hat{x}_i(\cdot) + K_N(\cdot)\hat{x}^{(N)}(\cdot) + \hat{\phi}_N(\cdot), \quad (20)$$

where $P_N(\cdot)$, $K_N(\cdot)$ are differential functions with $P_N(T) = 0$, $K_N(T) = 0$. By Itô's formula, we have

$$\begin{aligned} \mathrm{d}\hat{p}_i =& \dot{P}_N\hat{x}_i\mathrm{d}t + \left[aP_N\hat{x}_i - P_N q_3^{-1} b^2\left(P_N\hat{x}_i + K_N\hat{x}^{(N)} + \hat{\phi}_N\right)\right]\mathrm{d}t + P_N\sigma\mathrm{d}B_i + \dot{K}_N\hat{x}^{(N)}\mathrm{d}t \\ &+ \left[aK_N x^{(N)} - K_N q_3^{-1} b^2\left(P_N\hat{x}^{(N)} + K_N\hat{x}^{(N)} + \hat{\phi}_N\right)\right]\mathrm{d}t + K_N\frac{1}{N}\sum_{j=1}^N \sigma\mathrm{d}B_j + \mathrm{d}\hat{\phi}_N. \end{aligned} \quad (21)$$

Instituting (20) into the right side of the second equation in (18) obtains

$$\mathrm{d}\hat{p}_i = -\left[a\left[P_N\hat{x}_i + K_N\hat{x}^{(N)} + \hat{\phi}_N\right] + \left[q_1\left(1 - \frac{1}{N}\right) + q_2\right]\hat{x}_i - q_1\left(1 - \frac{1}{N}\right)\hat{x}^{(N)} - q_2 x_0\right] + \sum_{j=1}^N q_i^j \mathrm{d}B_j. \quad (22)$$

Comparing the coefficients of the right hand side of (21) and (22), yields

$$q_i^i = P_N\sigma + \frac{K_N}{N}\sigma, \quad q_i^j = \frac{K_N}{N}\sigma, \quad i \neq j, \tag{23}$$

$$\dot{P}_N + 2aP_N - q_3^{-1}b^2P_N^2 + \left[q_1\left(1-\frac{1}{N}\right)+q_2\right] = 0, \quad P_N(T) = 0, \tag{24}$$

$$\dot{K}_N + 2aK_N - P_Nq_3^{-1}b^2K_N - K_Nq_3^{-1}b^2\left(P_N+K_N\right) - q_1\left(1-\frac{1}{N}\right) = 0, \quad K_N(T) = 0, \tag{25}$$

$$\dot{\Pi}_N + 2a\Pi_N - q_3^{-1}b^2\Pi_N^2 + q_2 = 0, \quad \Pi_N(T) = 0, \tag{26}$$

and

$$\mathrm{d}\hat{\phi}_N = -\left[\left(a - q_3^{-1}b^2\left(\Pi_N\right)\right)\hat{\phi}_N + q_2x_0\right]\mathrm{d}t, \quad \hat{\phi}_N(T) = 0, \tag{27}$$

where $\Pi_N = P_N + K_N$. Note that Eqs. (24) and (26) are symmetric Riccati differential equations, and $q_i > 0, i = 1,2,3$; then there exists a unique solution for the Riccati differential equations (24) and (26), respectively, which gives the solution of Hamilton equation (18). An alternative derivation of the Riccati equations is provided in Appendix D. Consequently, we immediately have the following conclusion.

Theorem 2. For given $u_0(\cdot) \in L^2(0,T;\mathbb{R})$, then the (**P1**) admits a unique solution

$$\hat{u}_i = -q_3^{-1}b\left(P_N\hat{x}_i + K_N\hat{x}^{(N)} + \phi_N\right), \tag{28}$$

where P_N and K_N are the solution of Riccati differential equations (24) and (26), respectively; ϕ_N is given by (27), and $\hat{x}_i$ and $\hat{x}^{(N)}$ satisfy

$$\begin{cases} \mathrm{d}\hat{x}_i = \left[(a - q_3^{-1}b^2P_N)\hat{x}_i - q_3^{-1}b^2K_N\hat{x}^{(N)} - q_3^{-1}b^2\hat{\phi}_N\right]\mathrm{d}t + \sigma\mathrm{d}B_i, \ x_i(0) = \zeta_i, \ i = 1,2,\cdots,N, \\ \mathrm{d}\hat{x}^{(N)} = \mathrm{d}\dfrac{1}{N}\displaystyle\sum_{i=1}^N \hat{x}_i = \left[(a - q_3^{-1}b^2P_N)\dfrac{1}{N}\displaystyle\sum_{i=1}^N\hat{x}_i - q_3^{-1}b^2K_N\hat{x}^{(N)} - q_3^{-1}b^2\hat{\phi}_N\right]\mathrm{d}t + \sigma\dfrac{1}{N}\displaystyle\sum_{i=1}^N\mathrm{d}B_i \\ \qquad = \left[\left(a - q_3^{-1}b^2\Pi_N\right)\hat{x}^{(N)} - q_3^{-1}b^2\hat{\phi}_N\right]\mathrm{d}t + \sigma\mathrm{d}B^{(N)}, \quad \hat{x}^{(N)}(0) = \zeta^{(N)}, \end{cases} \tag{29}$$

where $\hat{x}^{(N)}(\cdot) = \frac{1}{N}\sum_{i=1}^N \hat{x}_i(\cdot)$, $\mathrm{d}B^{(N)}(\cdot) := \frac{1}{N}\sum_{i=1}^N \mathrm{d}B_i(\cdot)$, and $\zeta^{(N)} := \frac{1}{N}\sum_{i=1}^N \zeta_i$.

3.2 The strategy of the leader

Upon implementing the strategies of followers $\hat{u}_i(\cdot), i = 1,\cdots,N$ according to (28), we derive the best strategy for the leader by solving the following optimization problem:

(**P2**): Minimize $J_0^N\left(u_0(\cdot)\right)$ over $u_0(\cdot) \in \mathcal{U}_0[0,T]$, where

$$J_0(u_0(\cdot), u^N(\cdot)) = \frac{1}{2}\mathbb{E}\left\{\int_0^T\left[r_1\left(x_0 - \hat{x}^{(N)}\right)^2 + r_2\left(x_0 - x_0^{\mathrm{ref}}\right)^2 + r_3u_0^2\right]\mathrm{d}t\right\} \tag{30}$$

subject to the following Forward-backward stochastic differential equations (FBSDEs):

$$\begin{cases} \mathrm{d}x_0 = \left[ax_0 + bu_0\right]\mathrm{d}t, \\ \mathrm{d}\hat{x}_i = \left[(a - q_3^{-1}b^2P_N)\hat{x}_i - q_3^{-1}b^2K_N\hat{x}^{(N)} - q_3^{-1}b^2\hat{\phi}_N\right]\mathrm{d}t + \sigma\mathrm{d}B_i(t), \\ \mathrm{d}\hat{\phi}_N = -\left[\left(a - q_3^{-1}b^2\Pi_N\right)\hat{\phi}_N - q_2x_0\right]\mathrm{d}t, \quad \hat{\phi}_N(T) = 0, \\ x_0(0) = \zeta_0, \quad x_i(0) = \zeta_i, \ i = 1,2,\cdots,N, \end{cases} \tag{31}$$

where $\hat{x}_i$ and $\hat{x}^{(N)}(\cdot)$ are the state response of the followers forced by their strategies given in Theorem 2 and the average value of the states of N followers, respectively. The result of this subsection is summarized as follows.

Theorem 3. Let the followers adopt the optimal strategy (28). If (**P2**) admits an optimal control $\hat{u}_0(\cdot) \in L^2(0,T;\mathbb{R})$, then the adapted solution $\left(\hat{x}_0(\cdot), \hat{x}^{(N)}(\cdot), \hat{\phi}_N(\cdot), \hat{y}_0(\cdot), \hat{y}^{(N)}(\cdot), \hat{\psi}_N(\cdot)\right)$ to the Hamilton system

$$\begin{cases} \mathrm{d}\hat{x}_0 = \left[a\hat{x}_0 - r_3^{-1}b^2\hat{y}_0\right]\mathrm{d}t, \quad \hat{x}_0(0) = \zeta_0, \\ \mathrm{d}\hat{x}^{(N)} = \left[\left(a - q_3^{-1}b^2\Pi_N\right)\hat{x}^{(N)} - q_3^{-1}b^2\hat{\phi}_N\right]\mathrm{d}t + \sigma \mathrm{d}B^{(N)}, \quad \hat{x}^{(N)}(0) = \zeta^{(N)}, \\ \mathrm{d}\hat{\phi}_N = -\left[\left(a - q_3^{-1}b^2\Pi_N\right)\hat{\phi}_N - q_2\hat{x}_0\right]\mathrm{d}t, \quad \hat{\phi}_N(T) = 0, \\ \mathrm{d}\hat{y}_0 = -\left[a\hat{y}_0 + (r_1 + r_2)\hat{x}_0 - r_1\hat{x}^{(N)} - r_2x_0^{\mathrm{ref}} + q_2\hat{\psi}_N\right]\mathrm{d}t, \quad \hat{y}_0(T) = 0, \\ \mathrm{d}\hat{y}^{(N)} = -\left[\left(a - q_3^{-1}b^2\Pi_N\right)\hat{y}^{(N)} - r_1\left(\hat{x}_0 - \hat{x}^{(N)}\right)\right]\mathrm{d}t + g_1\mathrm{d}B^{(N)}, \quad \hat{y}^{(N)}(T) = 0, \\ \mathrm{d}\hat{\psi}_N = \left[\left(a - q_3^{-1}b^2\Pi_N\right)\hat{\psi}_N + q_3^{-1}b^2\hat{y}^{(N)}\right]\mathrm{d}t, \quad \hat{\psi}_N(0) = 0 \end{cases} \tag{32}$$

satisfies

$$\hat{u}_0 = -r_3^{-1}b\hat{y}_0, \qquad \text{a.e., a.s.} \tag{33}$$

The proof of the theorem is in Appendix B.

The results obtained in this section exhibit the essential characteristics of the Stackelberg leader-follower game: the follower's strategy decision is made according to the leader's strategy, and then under the correspondence between the follower's and the leader's strategy, the latter makes the decision of strategy that minimizes the cost functional J_0. In the next section, the strategies will be decentralized and it will be shown that the strategies will achieve the $(\varepsilon_1, \varepsilon_2)$-Stackelberg equilibrium.

4 Decentralized strategies and $(\varepsilon_1, \varepsilon_2)$-Stackelberg equilibrium

Let $N \to \infty$, then $P_N(\cdot) \to \bar{P}(\cdot), K_N(\cdot) \to \bar{K}(\cdot)$, where $\bar{P}(\cdot), \bar{K}(\cdot)$ satisfy

$$\dot{\bar{P}} + 2a\bar{P} - q_3^{-1}b^2\bar{P}^2 + q_1 + q_2 = 0, \quad \bar{P}(T) = 0, \tag{34}$$

$$\dot{\bar{K}} + 2a\bar{K} - \bar{P}q_3^{-1}b^2\bar{K} - \bar{K}q_3^{-1}b^2(\bar{P} + \bar{K}) - q_1 = 0, \quad \bar{K}(T) = 0. \tag{35}$$

Furthermore, let $\bar{\Pi}(\cdot) := \bar{P}(\cdot) + \bar{K}(\cdot)$; then it is easy to confirm that $\bar{\Pi}(\cdot)$ satisfies

$$\dot{\bar{\Pi}} + 2a\bar{\Pi} - q_3^{-1}b^2\bar{\Pi}^2 + q_2 = 0, \quad \bar{\Pi}(T) = 0. \tag{36}$$

Since $q_i > 0,\ i = 1, 2, 3$, it follows that Eqs. (34) and (35) admit a unique solution, respectively. Inspired by (32), we consider

$$\begin{cases} \mathrm{d}\bar{x}_0 = \left[a\bar{x}_0 - r_3^{-1}b^2\bar{y}_0\right]\mathrm{d}t, \quad \bar{x}_0(0) = \zeta_0, \\ \mathrm{d}\bar{x} = \left[\left(a - q_3^{-1}b^2\bar{\Pi}\right)\bar{x} - q_3^{-1}b^2\bar{\phi}\right]\mathrm{d}t, \quad \bar{x}(0) = \bar{\zeta}, \\ \mathrm{d}\bar{\phi} = -\left[\left(a - q_3^{-1}b^2\bar{\Pi}\right)\bar{\phi} - q_2x_0\right]\mathrm{d}t, \quad \bar{\phi}(T) = 0, \\ \mathrm{d}\bar{y}_0 = -\left[a\bar{y}_0 + (r_1 + r_2)\bar{x}_0 - r_1\bar{x} - r_2x_0^{\mathrm{ref}} + q_2\bar{\psi}\right]\mathrm{d}t, \quad \bar{y}_0(T) = 0, \\ \mathrm{d}\bar{y} = -\left[\left(a - q_3^{-1}b^2\bar{\Pi}\right)\bar{y} - r_1(\bar{x}_0 - \bar{x})\right]\mathrm{d}t, \quad \bar{y}(T) = 0, \\ \mathrm{d}\bar{\psi} = \left[\left(a - q_3^{-1}b^2\bar{\Pi}\right)\bar{\psi} + q_3^{-1}b^2\bar{y}\right]\mathrm{d}t, \quad \bar{\psi}(0) = 0. \end{cases} \tag{37}$$

Let $X = (\bar{x}_0, \bar{x}, \bar{\psi})^\top$, $Y = (\bar{\phi}, \bar{y}_0, \bar{y})^\top$, $C = (0, r_2x_0^{\mathrm{ref}}, 0)^\top$ and

$$A_1 := \begin{bmatrix} a & 0 & 0 \\ 0 & a - q_3^{-1}b^2\bar{\Pi} & 0 \\ 0 & 0 & a - q_3^{-1}b^2\bar{\Pi} \end{bmatrix}, B_1 := \begin{bmatrix} 0 & -r_3^{-1}b^2 & 0 \\ -q_3^{-1}b^2 & 0 & 0 \\ 0 & 0 & q_3^{-1}b^2 \end{bmatrix},$$

$$A_2 := \begin{bmatrix} q_2 & 0 & 0 \\ -(r_1+r_2) & r_1 & -q_2 \\ r_1 & -r_1 & 0 \end{bmatrix}, B_2 := \begin{bmatrix} -(a-q_3^{-1}b^2\bar{\Pi}) & 0 & 0 \\ 0 & -a & 0 \\ 0 & 0 & -(a-q_3^{-1}b^2\bar{\Pi}) \end{bmatrix}.$$

With the above notions, we can rewrite (37) as

$$\begin{cases} \mathrm{d}X = [A_1X + B_1Y]\,\mathrm{d}t, \quad X(0) = \left[\zeta_0, \bar{\zeta}, 0\right]^\top, \\ \mathrm{d}Y = [A_2X + B_2Y + C]\,\mathrm{d}t, \quad Y(T) = [0,0,0]^\top. \end{cases} \tag{38}$$

Suppose $(X(\cdot), Y(\cdot))$ is an adapted solution to (38) and due to the coupling between the two state equations in (38), so assume that $X(\cdot)$ and $Y(\cdot)$ are related by the following affine transformation:

$$X(\cdot) = \Phi(\cdot)Y(\cdot) + \Psi(\cdot),$$

where $\Phi(\cdot)$ and $\Psi(\cdot)$ are both differentiable functions, with $\Phi(0) = \mathbf{0}$ and $\Psi(0) = \left[\zeta_0, \bar{\zeta}, 0\right]^\top$. Next, by Itô's formula, we have

$$\mathrm{d}X = [\dot{\Phi}Y + \dot{\Psi}]\mathrm{d}t + \Phi\left[A_2(\Phi Y + \Psi) + B_2Y + C\right]\mathrm{d}t.$$

This together with the first equation in(38) gives

$$\begin{aligned} &\dot{\Phi} + \Phi B_2 + \Phi A_2 \Phi - A_1\Phi - B_1 = 0, \quad \Phi(0) = \mathbf{0}, \\ &\dot{\Psi} + \Phi A_2 \Psi + \Phi C - A_1 \Psi = 0, \quad \Psi(0) = \left[\zeta_0, \bar{\zeta}, 0\right]^\top. \end{aligned} \tag{39}$$

Note that the Riccati equation (39) is non-symmetric. By Theorem 4.1 on page 47 of [23] again, if Eq. (39) admits a solution $\Phi(\cdot)$, then Eq. (38) admits a unique adapted solution $(X(\cdot), Y(\cdot))$. The existence of a unique solution to the first equation in (39) is discussed in Appendix C.

Theorem 4. Motivated by (28) and (33), we design the decentralized strategies below:

$$\begin{cases} u_0^* = -r_3^{-1}b\bar{y}_0, \\ u_i^* = -q_3^{-1}b\left(\bar{P}x_i^* + \bar{K}\bar{x} + \bar{\phi}\right), \quad i = 1, \cdots, N, \end{cases} \tag{40}$$

where $\bar{y}_0(\cdot), \bar{x}(\cdot), \bar{\phi}(\cdot)$ are given by (37), and $x_i^*(\cdot)$ satisfies

$$\mathrm{d}x_i^* = [(a - b^2q_3^{-1}\bar{P})x_i^* - b^2q_3^{-1}\bar{K}\bar{x} - b^2q_3^{-1}\bar{\phi}]\mathrm{d}t + \sigma\mathrm{d}B_i, \quad x_i^*(0) = \zeta_i, \quad i = 1, \cdots, N. \tag{41}$$

Next, we will demonstrate that the decentralized strategies (40) employed by the leader and followers asymptotically form an $(\varepsilon_1, \varepsilon_2)$-Stackelberg equilibrium as the number of agents N approaches infinity. To obtain the result, the following lemma is required.

Lemma 1. Under the control inputs u_0^* and u_i^* that are given in (40), ones have

$$\mathbb{E}\int_0^T \left(x^{*(N)} - \bar{x}\right)^2 \mathrm{d}t = O\left(\frac{1}{N}\right). \tag{42}$$

Proof. By (41), we can get

$$\mathrm{d}x^{*(N)} = [(a - b^2q_3^{-1}\bar{P})x^{*(N)} - b^2q_3^{-1}\bar{K}\bar{x} - b^2q_3^{-1}\bar{\phi}]\mathrm{d}t + \sigma\mathrm{d}B^{(N)}, \quad x^{*(N)}(0) = \zeta^{(N)}. \tag{43}$$

Combining this with the second equation of (37) yields

$$\mathrm{d}Z(t) = [(a - b^2q_3^{-1}\bar{P})Z(t)]\mathrm{d}t + \sigma\mathrm{d}B^{(N)}(t), \quad Z(0) = \zeta^{(N)} - \bar{\zeta}, \tag{44}$$

where $Z(t) := x^{*(N)}(t) - \bar{x}(t)$. Hence, the solution of equation (44) is

$$Z(t) = Z(0)e^{(a-b^2q_3^{-1}\bar{P})t} + \sigma\int_0^t e^{(a-b^2q_3^{-1}\bar{P})(t-s)}\mathrm{d}B^{(N)}(s). \tag{45}$$

If $a-b^2q_3^{-1}\bar{P}<0$, its expectation is $\mathbb{E}\left[Z(t)\right]=Z(0)e^{(a-b^2q_3^{-1}\bar{P})t}$. According to the Itô isometric theorem, we can obtain

$$\begin{aligned}\mathbb{E}\left[Z^2(t)\right]&=\left[Z(0)e^{(a-b^2q_3^{-1}\bar{P})t}\right]^2+\sigma^2\mathbb{E}\left\{\left[\int_0^t e^{(a-b^2q_3^{-1}\bar{P})(t-s)}\mathrm{d}B^{(N)}(s)\right]^2\right\}\\&=\left[Z(0)e^{(a-b^2q_3^{-1}\bar{P})t}\right]^2+\sigma^2\int_0^t e^{2(a-b^2q_3^{-1}\bar{P})(t-s)}\mathbb{E}\left\{\left[\mathrm{d}B^{(N)}(s)\right]^2\right\}\\&=\left[Z(0)e^{(a-b^2q_3^{-1}\bar{P})t}\right]^2+\frac{\sigma^2}{N}\int_0^t e^{2(a-b^2q_3^{-1}\bar{P})(t-s)}\mathrm{d}s\\&=Z^2(0)e^{2(a-b^2q_3^{-1}\bar{P})t}+\frac{\sigma^2}{N}\cdot\frac{e^{2(a-b^2q_3^{-1}\bar{P})t}-1}{2(a-b^2q_3^{-1}\bar{P})}\\&\leqslant Z^2(0)e^{2(a-b^2q_3^{-1}\bar{P})t}+\frac{\sigma^2}{2N|a-b^2q_3^{-1}\bar{P}|},\end{aligned}$$

where the third last step is according to the $\mathrm{d}B^{(N)}(s)=\frac{1}{N}\sum_{i=1}^N\mathrm{d}B_i(s)$. Integral from 0 to t, one can obtain

$$\begin{aligned}\mathbb{E}\int_0^T\left[Z^2(t)\right]\mathrm{d}t&\leqslant\int_0^T Z^2(0)e^{2(a-b^2q_3^{-1}\bar{P})t}\mathrm{d}t+\int_0^T\frac{\sigma^2}{2N|a-b^2q_3^{-1}\bar{P}|}\mathrm{d}t\\&=\int_0^T Z^2(0)e^{2(a-b^2q_3^{-1}\bar{P})t}\mathrm{d}t+O\left(\frac{1}{N}\right).\end{aligned}$$

Under ζ_i that are i.i.d., it is easy to know that $Z(0)$ is i.i.d. and $\mathbb{E}\left[Z^2(0)\right]=O\left(\frac{1}{N}\right)$. Then the above can verify that

$$\mathbb{E}\int_0^T\left[Z^2(t)\right]\mathrm{d}t=O\left(\frac{1}{N}\right).$$

The proof is complete.

Theorem 5. $(u_0^*(\cdot),u_1^*(\cdot),\cdots,u_N^*(\cdot))$ given in (40) constitutes an $(\varepsilon_1,\varepsilon_2)$-Stackelberg equilibrium, where $\varepsilon_1=\varepsilon_2=O\left(\frac{1}{\sqrt{N}}\right)$.

Proof. Firstly, prove (1) in Definition 1. For $i=1,\cdots,N$, let $\tilde{u}_i(\cdot):=u_i(\cdot)-u_i^*(\cdot)$ and $\tilde{x}_i(\cdot):=x_i(\cdot)-x_i^*(\cdot)$. Then $\tilde{x}_i(\cdot)$ satisfies

$$\mathrm{d}\tilde{x}_i=(a\tilde{x}_i+b\tilde{u}_i)\,\mathrm{d}t,\quad\tilde{x}_i(0)=0,\quad i=1,\cdots,N.\tag{46}$$

From (17), we have

$$\begin{aligned}&J_i\left(u_i(\cdot),u_{-i}^*(\cdot),u_0(\cdot)\right)-J_i\left(u_i^*(\cdot),u_{-i}^*(\cdot),u_0(\cdot)\right)\\=&\frac{1}{2}\mathbb{E}\left\{\int_0^T\left\{q_1\left[(x_i-x^{(N)})^2-(x_i^*-x^{*(N)})^2\right]+q_2\left[(x_i-x_0)^2-(x_i^*-x_0)^2\right]+q_3\left[u_i^2-(u_i^*)^2\right]\right\}\mathrm{d}t\right\}\\=&\frac{1}{2}\mathbb{E}\left\{\int_0^T\left\{q_1\left[\tilde{x}_i-\tilde{x}^{(N)}+2x_i^*-2x^{*(N)}\right]\left(\tilde{x}_i-\tilde{x}^{(N)}\right)+q_2(2x_i^*+\tilde{x}_i-2x_0)\tilde{x}_i+q_3(2u_i^*+\tilde{u}_i)\tilde{u}_i\right\}\mathrm{d}t\right\}\\=&\tilde{J}_i\left(\tilde{u}_i(\cdot),u_{-i}^*(\cdot),u_0(\cdot)\right)+I_i,\end{aligned}$$

where

$$\tilde{J}_i\left(\tilde{u}_i(\cdot),u_{-i}^*(\cdot),u_0(\cdot)\right):=\frac{1}{2}\mathbb{E}\left\{\int_0^T\left[q_1\left(\tilde{x}_i-\tilde{x}^{(N)}\right)^2+q_2(\tilde{x}_i)^2+q_3\tilde{u}_i^2\right]\mathrm{d}t\right\},$$

and

$$I_i:=\mathbb{E}\left\{\int_0^T\left[q_1\left(x_i^*-x^{*(N)}\right)\left(\tilde{x}_i-\tilde{x}^{(N)}\right)+q_2\left(x_i^*-x_0\right)\tilde{x}_i+q_3u_i^*\tilde{u}_i\right]\mathrm{d}t\right\}.\tag{47}$$

Applying Itô's formula to $\tilde{x}_i(\cdot)\left(\bar{P}(\cdot)x_i^*(\cdot)+\bar{K}(\cdot)\bar{x}(\cdot)+\bar{\phi}(\cdot)\right)$,

$$\mathrm{d}\left[\tilde{x}_i\left(\bar{P}x_i^*+\bar{K}\bar{x}+\bar{\phi}\right)\right]=\mathrm{d}(\tilde{x}_i)\left(\bar{P}x_i^*+\bar{K}\bar{x}+\bar{\phi}\right)+\tilde{x}_i\left[\dot{\bar{P}}x_i^*+\bar{P}\mathrm{d}(x_i^*)+\dot{\bar{K}}\bar{x}+\bar{K}\mathrm{d}(\bar{x})+\mathrm{d}(\bar{\phi})\right].\tag{48}$$

Substitute (34), (35), (37), (41) and (46) into (48) and integrate from 0 to T, and taking expectation, we have

$$\mathbb{E}\left\{\int_0^T \mathrm{d}\left[\tilde{x}_i\left(\bar{P}x_i^*+\bar{K}\bar{x}+\bar{\phi}\right)\right]\right\}=-\mathbb{E}\left\{\int_0^T\left[q_1\left(x_i^*-\bar{x}\right)\tilde{x}_i+q_2\left(x_i^*-x_0\right)\tilde{x}_i+q_3u_i^*\tilde{u}_i\right]\mathrm{d}t\right\}.$$

Due to $\bar{P}(T)=\bar{K}(T)=\bar{\phi}(T)=0$ and $\tilde{x}_i(0)=0$, we have

$$\begin{aligned}&\mathbb{E}\left\{\int_0^T \mathrm{d}\left[\tilde{x}_i\left(\bar{P}x_i^*+\bar{K}\bar{x}+\bar{\phi}\right)\right]\right\}\\ =&\mathbb{E}\left\{\tilde{x}_i(T)\left[\bar{P}(T)x_i^*(T)+\bar{K}(T)\bar{x}(T)+\bar{\phi}(T)\right]-\tilde{x}_i(0)\left[\bar{P}(0)x_i^*(0)+\bar{K}(0)\bar{x}(0)+\bar{\phi}(0)\right]\right\}=0,\end{aligned}$$

hence,

$$\mathbb{E}\left\{\int_0^T\left[q_2\left(x_i^*-x_0\right)\tilde{x}_i+q_3u_i^*\tilde{u}_i\right]\mathrm{d}t\right\}=-\mathbb{E}\left\{\int_0^T\left[q_1\left(x_i^*-\bar{x}\right)\tilde{x}_i\right]\mathrm{d}t\right\}. \tag{49}$$

Then, Eq. (47) can be rewritten as

$$\begin{aligned}I_i=&\mathbb{E}\left\{\int_0^T\left[q_1\left(x_i^*-x^{*(N)}\right)\left(\tilde{x}_i-\tilde{x}^{(N)}\right)+q_2\left(x_i^*-x_0\right)\tilde{x}_i+q_3u_i^*\tilde{u}_i\right]\mathrm{d}t\right\}\\ =&\mathbb{E}\left\{\int_0^T\left[q_1\left(x_i^*-\bar{x}\right)\left(\tilde{x}_i-\tilde{x}^{(N)}\right)+q_1\left(\bar{x}-x^{*(N)}\right)\left(\tilde{x}_i-\tilde{x}^{(N)}\right)+q_2\left(x_i^*-x_0\right)\tilde{x}_i+q_3u_i^*\tilde{u}_i\right]\mathrm{d}t\right\}.\end{aligned} \tag{50}$$

Substituting (49) into (50) yields

$$I_i=\mathbb{E}\left\{\int_0^T\left[q_1\left(\bar{x}-x^{*(N)}\right)\left(\tilde{x}_i-\tilde{x}^{(N)}\right)\right]\mathrm{d}t\right\}. \tag{51}$$

Given the $u_i,\ u_i^*\in L^2_{\mathcal{F}^i}(0,T;\mathbb{R}),\ i=1,\cdots,N$, we establish the square-integrability condition

$$\mathbb{E}\left[\int_0^T\left|\tilde{u}_i(t)\right|^2\mathrm{d}t\right]<\infty. \tag{52}$$

By (46), we can get $\mathbb{E}\left[\int_0^T\left(\tilde{x}_i(t)\right)^2\mathrm{d}t\right]<\infty$. Combining this result with (42) and (51), we derive

$$I_i=O\left(\frac{1}{\sqrt{N}}\right).$$

Thereby,

$$J_i\left(u_i^*(\cdot),u_{-i}^*(\cdot),u_0(\cdot)\right)\leqslant J_i\left(u_i(\cdot),u_{-i}^*(\cdot),u_0(\cdot)\right)+\varepsilon_1. \tag{53}$$

Thus, $(u_1^*(\cdot),\cdots,u_N^*(\cdot))$ is an ε_1-Nash equilibrium, where $\varepsilon_1=O\left(\frac{1}{\sqrt{N}}\right)$.

Next, prove (2) in Definition 1. Let $\hat{X}=(\hat{x}_0,\hat{x},\hat{\psi})^\top$, $\hat{Y}=(\hat{\phi},\hat{y}_0,\hat{y})^\top$, $\hat{C}=(0,r_2x_0^{ref},0)^\top$, $\hat{D}_1=(0,\sigma,0)^\top$, $\hat{D}_2=(0,0,g_1)^\top$ and

$$A_1:=\begin{bmatrix}a&0&0\\0&a-q_3^{-1}b^2\Pi_N&0\\0&0&a-q_3^{-1}b^2\Pi_N\end{bmatrix},B_1:=\begin{bmatrix}0&-r_3^{-1}b^2&0\\-q_3^{-1}b^2&0&0\\0&0&q_3^{-1}b^2\end{bmatrix},$$

$$A_2:=\begin{bmatrix}q_2&0&0\\-(r_1+r_2)&r_1&-q_2\\r_1&-r_1&0\end{bmatrix},B_2:=\begin{bmatrix}-(a-q_3^{-1}b^2\Pi_N)&0&0\\0&-a&0\\0&0&-(a-q_3^{-1}b^2\Pi_N)\end{bmatrix}.$$

With the above notions, we can rewrite (32) as

$$\begin{cases} \mathrm{d}\hat{X} = \left[\hat{A}_1\hat{X} + \hat{B}_1\hat{Y}\right]\mathrm{d}t + \hat{D}_1\mathrm{d}B^{(N)}, \quad \hat{X}(0) = \left[\zeta_0, \zeta^N, 0\right]^\top, \\ \mathrm{d}\hat{Y} = \left[\hat{A}_2\hat{X} + \hat{B}_2\hat{Y} + \hat{C}\right]\mathrm{d}t + \hat{D}_2\mathrm{d}B^{(N)}, \quad \hat{Y}(T) = [0,0,0]^\top. \end{cases}$$

Let $\tilde{X} := \hat{X} - X, \tilde{Y} := \hat{Y} - Y, \tilde{C} := \hat{C} - C$, and $\tilde{A}_1 := \hat{A}_1 - A_1, \tilde{A}_2 := \hat{A}_2 - A_2, \tilde{B}_1 := \hat{B}_1 - B_1$, $\tilde{B}_2 := \hat{B}_2 - B_2$, where $X,\ Y,\ C,\ A_1,\ A_2,\ B_1, B_2$ are given in (38). We have

$$\begin{cases} \mathrm{d}\tilde{X} = \left[A_1\tilde{X} + \tilde{A}_1\hat{X} + B_1\tilde{Y} + \tilde{B}_1\hat{Y}\right]\mathrm{d}t + \hat{D}_1\mathrm{d}B^{(N)}, \\ \mathrm{d}\tilde{Y} = \left[A_2\tilde{X} + \tilde{A}_2\hat{X} + B_2\tilde{Y} + \tilde{B}_2\hat{Y} + \tilde{C}\right]\mathrm{d}t + \hat{D}_2\mathrm{d}B^{(N)}, \\ \tilde{X}(0) = \left[\zeta_0, \zeta^{(N)} - \bar{\zeta}, 0\right]^\top, \quad \tilde{Y}(T) = [0,0,0]^\top. \end{cases} \tag{54}$$

Let $Z_1 := \tilde{A}_1\hat{X} + \tilde{B}_1\hat{Y}, Z_2 := \tilde{A}_2\hat{X} + \tilde{B}_2\hat{Y} + \tilde{C}$. By the continuous dependence of the solution on the parameter in Theorem 4 of [24], we have $\sup_{0\leqslant t\leqslant T}|Z_1|^2 = O\left(\frac{1}{N}\right), \sup_{0\leqslant t\leqslant T}|Z_2|^2 = O\left(\frac{1}{N}\right)$. Then we can rewrite (54) as

$$\begin{cases} \mathrm{d}\tilde{X} = \left[A_1\tilde{X} + B_1\tilde{Y} + Z_1\right]\mathrm{d}t + \hat{D}_1\mathrm{d}B^{(N)}, \\ \mathrm{d}\tilde{Y} = \left[A_2\tilde{X} + B_2\tilde{Y} + Z_2\right]\mathrm{d}t + \hat{D}_2\mathrm{d}B^{(N)}, \\ \tilde{X}(0) = \left[\zeta_0, \zeta^{(N)} - \bar{\zeta}, 0\right]^\top, \quad \tilde{Y}(T) = [0,0,0]^\top. \end{cases} \tag{55}$$

We assume that $\tilde{Y}(\cdot) = \Theta(\cdot)\tilde{X}(\cdot)$, where $\Theta(\cdot)$ is a differential function with $\Theta(T) = 0$. By Itô's formula, we have

$$\mathrm{d}\tilde{Y} = \dot{\Theta}\tilde{X}\mathrm{d}t + \Theta\left[\left(A_1\tilde{X} + B_1\Theta\tilde{X} + Z_1\right)\mathrm{d}t + \hat{D}_1\mathrm{d}B^{(N)}\right]. \tag{56}$$

According to the second equation in (55),

$$\mathrm{d}\tilde{Y} = \left[A_2\tilde{X} + B_2\Theta\tilde{X} + Z_2\right]\mathrm{d}t + \hat{D}_2\mathrm{d}B^{(N)}. \tag{57}$$

Comparing the corresponding coefficients of (56) and (57), we have

$$\dot{\Theta} + \Theta A_1 + \Theta B_1\Theta - A_2 - B_2\Theta = 0, \quad \Theta(T) = 0.$$

Then we achieve

$$\mathrm{d}\tilde{X} = \left[(A_1 + B_1\Theta)\tilde{X} + Z_1\right]\mathrm{d}t + \hat{D}_1\mathrm{d}B^{(N)}, \quad \tilde{X}(0) = \left[\zeta_0, \zeta^{(N)} - \bar{\zeta}, 0\right]^\top.$$

This implies

$$\tilde{X}(t) = e^{(A_1+B_1\Theta)t}\tilde{X}(0) + \int_0^t e^{(A_1+B_1\Theta)(t-s)}\left[Z_1\mathrm{d}s + \hat{D}_1\mathrm{d}B^{(N)}(s)\right],$$

which is similar to the method of obtaining (42). Then we can give

$$\sup_{0\leqslant t\leqslant T}\mathbb{E}|\tilde{X}(t)|^2 = O\left(\frac{1}{N}\right),$$

and since $\tilde{Y}(\cdot) = \Theta(\cdot)\tilde{X}(\cdot)$, then

$$\sup_{0\leqslant t\leqslant T}\mathbb{E}|\tilde{Y}(t)|^2 = O\left(\frac{1}{N}\right),$$

which means

$$\mathbb{E}\int_0^T(\hat{x}_0 - \bar{x}_0)^2\mathrm{d}t = O\left(\frac{1}{N}\right), \quad \mathbb{E}\int_0^T\left(\hat{x}^{(N)} - \bar{x}\right)^2\mathrm{d}t = O\left(\frac{1}{N}\right), \tag{58}$$

and

$$\mathbb{E}\int_0^T (\hat{y}_0-\bar{y}_0)^2\,\mathrm{d}t = O\left(\frac{1}{N}\right), \quad \mathbb{E}\int_0^T \left(\hat{y}^{(N)}-\bar{y}\right)^2\mathrm{d}t = O\left(\frac{1}{N}\right), \quad \mathbb{E}\int_0^T \left(\hat{\psi}-\bar{\psi}\right)^2\mathrm{d}t = O\left(\frac{1}{N}\right). \tag{59}$$

Combining this with (42), we have

$$\mathbb{E}\int_0^T \left(x^{*(N)}-\hat{x}^{(N)}\right)^2\mathrm{d}t = O\left(\frac{1}{N}\right), \tag{60}$$

and $\bar{x}_0(t)=x_0^*(t)$, for all $t\in[0,T]$. Then

$$\mathbb{E}\int_0^T (x_0^*-\hat{x}_0)^2\,\mathrm{d}t = O\left(\frac{1}{N}\right). \tag{61}$$

We need two steps. First, by (60), we get

$$\begin{aligned}
&J_0\left(\hat{u}_0(\cdot),\hat{u}^N[\cdot;\hat{u}_0(\cdot)]\right) \leqslant J_0\left(u_0(\cdot),\hat{u}^N[\cdot;u_0(\cdot)]\right)\\
=&\frac{1}{2}\mathbb{E}\left\{\int_0^T\left[r_1\left(x_0-\hat{x}^{(N)}\right)^2+r_2\left(x_0-x_0^{\mathrm{ref}}\right)^2+r_3u_0^2\right]\mathrm{d}t\right\}\\
=&\frac{1}{2}\mathbb{E}\left\{\int_0^T\left[r_1(x_0-x^{*(N)})^2+r_2(x_0-x_0^{\mathrm{ref}})^2+r_3u_0^2\right]\mathrm{d}t\right\}\\
&+\frac{1}{2}\mathbb{E}\left\{\int_0^T r_1\left(x^{*(N)}-\hat{x}^{(N)}\right)^2\mathrm{d}t+2\int_0^T r_1\left(x_0-x^{*(N)}\right)\left(x^{*(N)}-\hat{x}^{(N)}\right)\mathrm{d}t\right\}\\
\leqslant&J_0\left(u_0(\cdot),u^{N*}[\cdot;u_0(\cdot)]\right)+O\left(\frac{1}{N}\right)+O\left(\frac{1}{\sqrt{N}}\right).
\end{aligned}$$

Next, from (58)–(61), we have

$$\begin{aligned}
&J_0\left(u_0^*(\cdot),u^{N*}[\cdot;u_0^*(\cdot)]\right)\\
=&\frac{1}{2}\mathbb{E}\left\{\int_0^T\left[r_1\left(x_0^*-x^{*(N)}\right)^2+r_2\left(x_0^*-x_0^{\mathrm{ref}}\right)^2+r_3(u_0^*)^2\right]\mathrm{d}t\right\}\\
=&\frac{1}{2}\mathbb{E}\left\{\int_0^T\left[r_1\left(x_0^*-x^{*(N)}\right)^2+r_2\left(x_0^*-x_0^{\mathrm{ref}}\right)^2+r_3\left(-r_3^{-1}b\bar{y}_0\right)^2\right]\mathrm{d}t\right\}\\
=&\frac{1}{2}\mathbb{E}\left\{\int_0^T r_1\left(\hat{x}_0-\hat{x}^{(N)}+x_0^*-\hat{x}_0-x^{*(N)}+\hat{x}^{(N)}\right)^2\right\}\\
&+\frac{1}{2}\mathbb{E}\left\{\int_0^T\left[r_2\left(\hat{x}_0-x_0^{\mathrm{ref}}+x_0^*-\hat{x}_0\right)^2-b\left(\hat{y}_0+\bar{y}_0-\hat{y}_0\right)^2\right]\mathrm{d}t\right\}\\
\leqslant&J_0\left(\hat{u}_0(\cdot),\hat{u}^N[\cdot;\hat{u}_0(\cdot)]\right)+O\left(\frac{1}{N}\right)+O\left(\frac{1}{\sqrt{N}}\right).
\end{aligned}$$

Hence,

$$J_0\left(u_0^*(\cdot),u^{*N}[\cdot;u_0^*(\cdot)]\right)\leqslant J_0\left(u_0(\cdot),u^{*N}[\cdot;u_0(\cdot)]\right)+\varepsilon_2.$$

The proof is complete.

5 Numerical simulation

Consider a large scale population of HVACs with the number $N=1000$, and a leader HVAC is additionally appointed. The dynamic equations of HVACs are given in (1), where the physical parameters are detailed in Table 1. Let the ambient temperature $T_{\mathrm{o}}=30$°C, the rated power $R_P=8$ kW, the initial power

Table 1 Typical parameters for HVACs and corresponding rooms.

Symbol	Parameter	Value	Unit
V	Room's volume	300	m^3
A_s	Envelope's surface area	320	m^2
c_A	Heat capacity of air	1.005	$\mathrm{kJ/(kg \cdot ^\circ C)}$
ρ_A	Density of air	1.205	$\mathrm{kg/m^3}$
U_h	Heat transfer coefficient	7.69	$\mathrm{W/(m^2 \cdot ^\circ C)}$
n	Air exchange times	0.5	1/h
η	Coefficient of performance	3	–

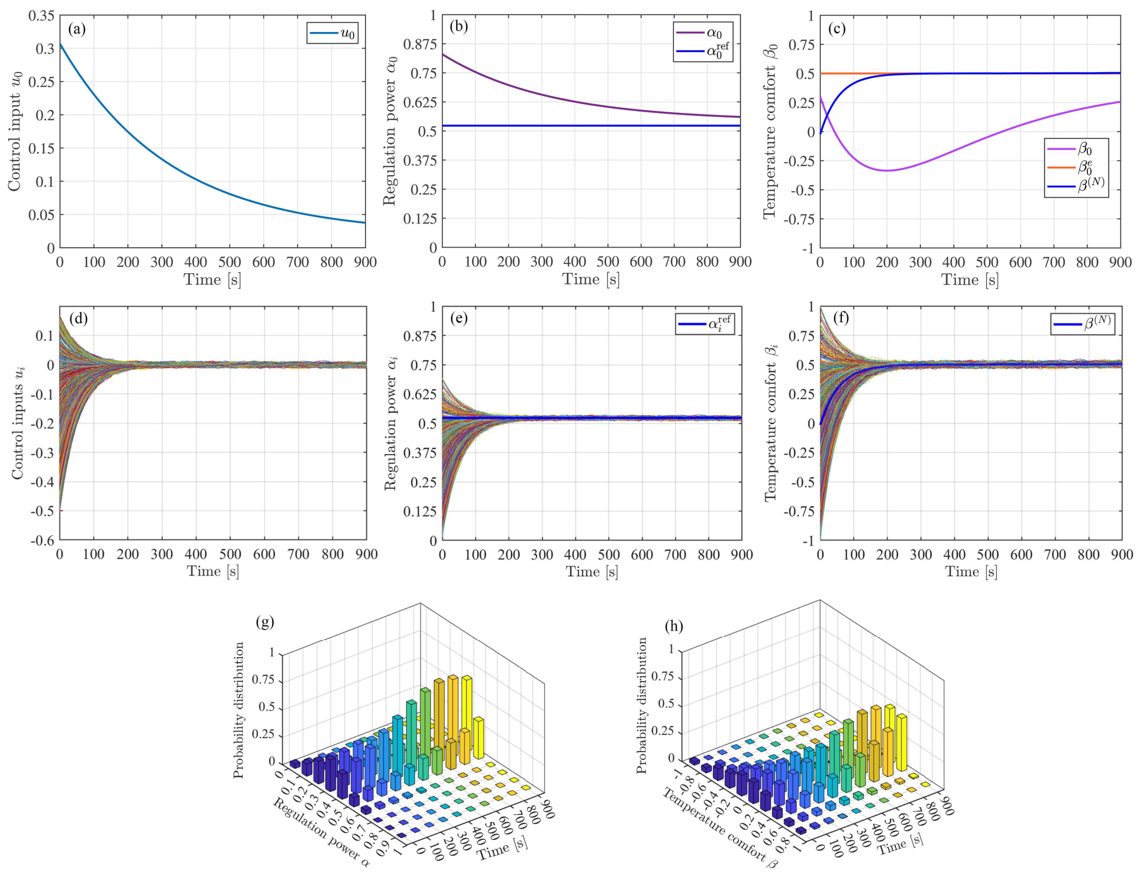

Figure 2 Regulation power and temperature comfort of HVACs and their probability distribution in Case I. (a) The control input u_0; (b) the regulation power α_0; (c) the temperature comfort β_0; (d) the control inputs u_i; (e) the regulation power α_i; (f) the temperature comfort β_i; (g) the probability distribution of α; (h) the probability distribution of β.

$P_0 = 0.837$ kW and the targeted horizon of T_i is $[23°\mathrm{C}, 27°\mathrm{C}]$. Hence, the parameters of the dynamic equations (6) are $a = -0.0069087$, $b = -0.0330296$, and $c = 0.020726$. Suppose $\beta_i^e = 0.5,\ i = 0, 1, \cdots, N$. Then according to (6), we have $\alpha_i^{\mathrm{ref}} = \frac{-c-a\beta_i^e}{b} = 0.52291,\ i = 0, 1, \cdots, N$ that should be the admissible average value of the regulation power expected by the power grid. The simulation time is set to $T = 900$ s.

In the simulation, we examine four different simulation scenarios.

Case I (Benchmark): Let the initial value of the leader's state $\beta_0(0) = 0.3$ ($T_0(0) = 25.6°\mathrm{C}$), and the follower's initial state $\beta_i(0) \sim \mathcal{N}(0, 0.7)$ ($T_i(0) \sim \mathcal{N}(25, 0.7)$), where $\mathcal{N}$ denotes normal distributions. The weighting coefficients in the cost functions (16) and (17) are set as $q_1 = 400$, $q_2 = 0.5$, $q_3 = 50$, $r_1 = 0.01$, $r_2 = 4$, $r_3 = 6$. The results of **Case I** are shown in Figure 2, where Figure 2(a) (or (d)) represents the control input of the leader (or followers), while Figures 2(b) (or (e)) and (c) (or (f)), respectively, represent the state changes of the leader's regulation power α_0 (or α_i) and temperature comfort β_0 (or β_i) under the control input represented in Figure 2(a) (or (d)). The results demonstrate that the collective behaviors α_i and β_i, $i = 0, \cdots, N$ are maintained close to the reference values, confirming the effectiveness of the control algorithm (40). Additionally, Figures 2(g) and (h) display distributions of regulation power

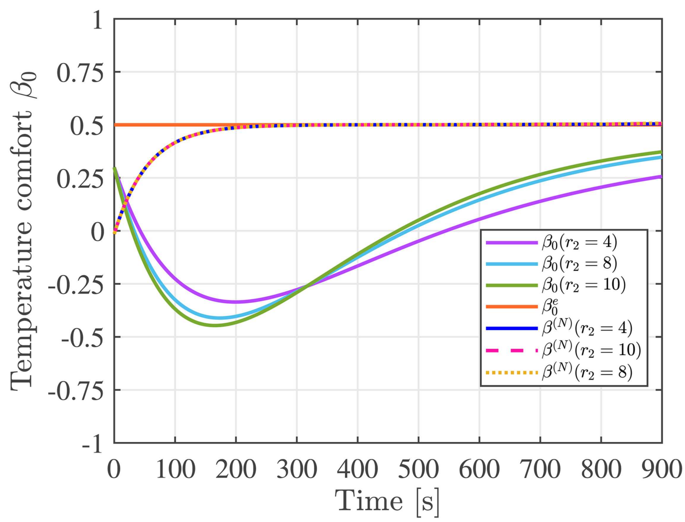

Figure 3 The temperature comfort of the leader in Case II.

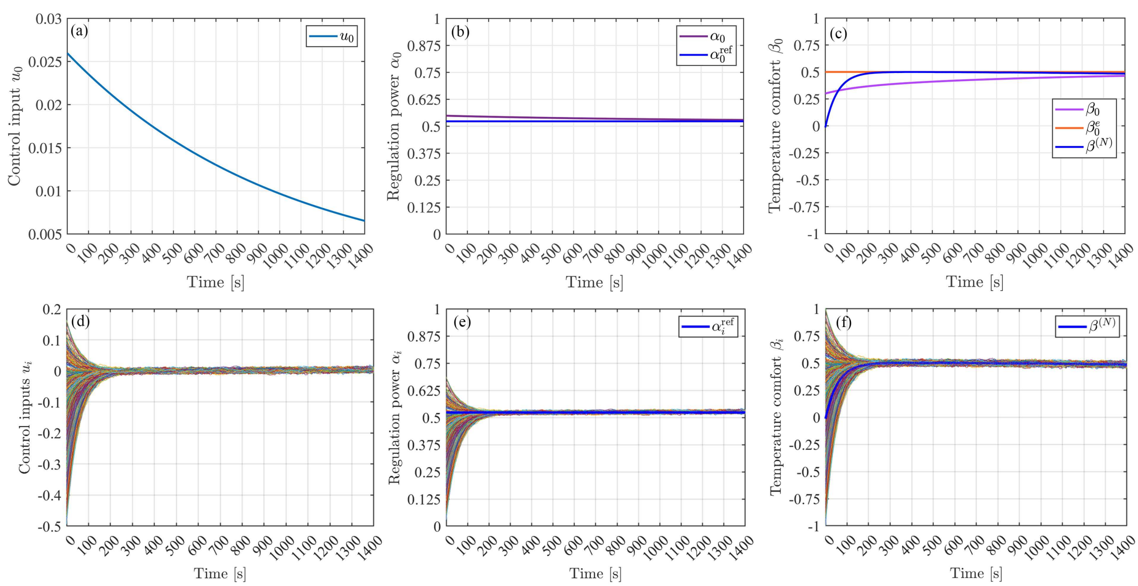

Figure 4 Regulation power and temperature comfort of HVACs in Case III. (a) The control input u_0; (b) the regulation power α_0; (c) the temperature comfort β_0; (d) the control inputs u_i; (e) the regulation power α_i; (f) the temperature comfort β_i.

and temperature comfort. The findings indicate that the probability densities for both the state variable and the control input increase over time, reflecting consistent performance across all agents.

Case II (Parameter changes): The initial value is the same as **Case I**. The parameters are set as $q_1 = 400$, $q_2 = 0.5$, $q_3 = 50$, $r_1 = 0.01$, $r_2 = 8$ (or $r_2 = 10$), $r_3 = 6$, where r_2 is decreased compared with **Case I**, which means prioritizing the leader's own tracking performance over consensus with the collective behavior of the followers. The responses of β_0 and $\beta^{(N)}$ with the difference r_2 are represented in Figure 3. Obviously, the responses show that the intent of the aforementioned is incorporated.

Case III (Parameter and time changes): The initial value is the same as **Case I**. The parameters are set as $q_1 = 400$, $q_2 = 0.5$, $q_3 = 50$, $r_1 = 0.1$, $r_2 = 1$, $r_3 = 25$, and the simulation time is set to $T = 1400$ s. In Figure 4, it can be seen that β_i is close to β_0.

Case IV (Initial value changes): Let the initial value of the leader's state $\beta_0(0) = 0.3$ ($T_0(0) = 25.6$°C), and the follower's initial state $\beta_i(0) \sim \mathcal{U}(-1, 1)$ ($T_i(0) \sim \mathcal{U}(23, 27)$), where $\mathcal{U}$ denotes uniform distributions. The coefficients in the cost functionals are set the same as in **Case I**. In Figure 5, the regulation power and temperature comfort α_i and β_i are close to their desired values. This indicates that the control strategies demonstrate adaptability and robustness in managing various types of initial value distributions.

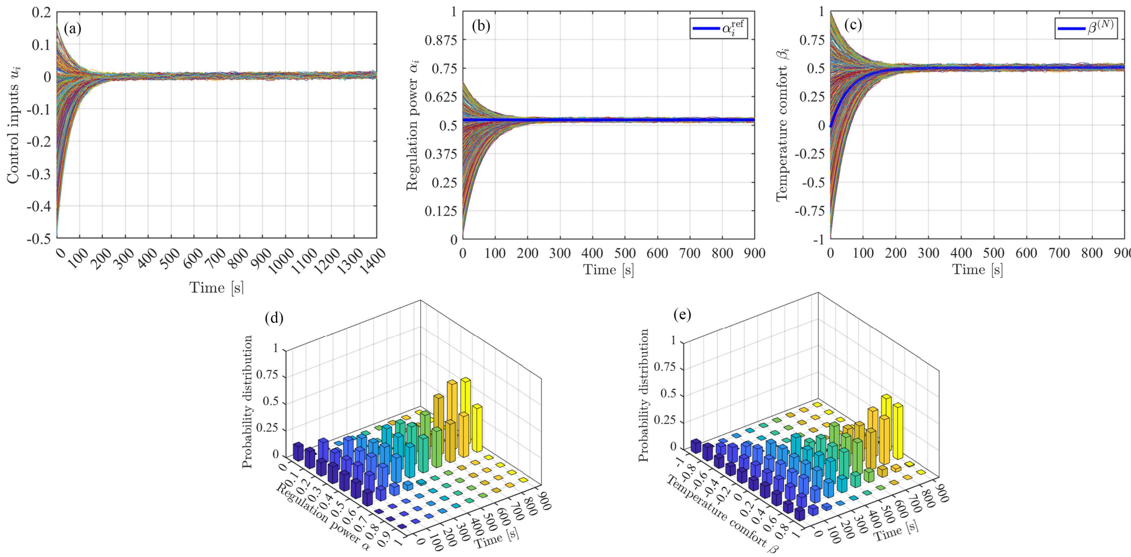

Figure 5 Regulation power and temperature comfort of followers and their probability distribution in Case IV. (a) The control inputs u_i; (b) the regulation power α_i; (c) the temperature comfort β_i; (d) the probability distribution of α; (e) the probability distribution of β.

6 Conclusion

To control a large number of HVACs with consideration of grid balancing, it is generally accepted that a centralized control system should be avoided since the infeasibility of implementation with constraints of limited resources and the uncertainties of the system dynamics, especially when the population becomes sufficiently large. Motivated by the benefit of the mean-field game framework where the collective behavior of a large number of agents is handled, but under decentralized control, to achieve the Nash equilibrium, a Stackelberg MFG-based approach is proposed in this paper for the decentralized control of large-scale HVAC systems, aiming to achieve grid balancing. First, we provide the thermodynamic model of HVAC systems, followed by the formulation of the Stackelberg MFG problem in a general context. We derived centralized strategies from a detailed analysis of agents' dynamics and costs, leading to decentralized strategies that establish a $(\varepsilon_1, \varepsilon_2)$-Stackelberg equilibrium. The derivation process of the proposed strategy is given in detail. Specifically, the method starts by addressing an N-player game problem within a large, finite population framework. It then decouples or reduces high-dimensional systems to derive centralized control based on the state of an individual player and the average state of the population. As the population size N approaches infinity, the development of decentralized strategies becomes feasible. Finally, we present simulation results under various weight coefficients and initial conditions to demonstrate the effectiveness of the proposed control algorithm.

Acknowledgements The authors appreciate Dr. Zhenhui XU, Institute of Science Tokyo, Japan, for the stimulating discussion on the topic addressed in this paper.

References

1 Int Renew Energy Agency. Renewable capacity statistics 2024. 978-92-9260-342-7, 2024. https://www.irena.org/Publications/2024/Mar/Renewable-capacity-statistics-2024
2 Siano P. Demand response and smart grids a survey. Renew Sustain Energy Rev, 2014, 30: 461–478
3 Zhang X, Biagioni D, Cai M, et al. An edge-cloud integrated solution for buildings demand response using reinforcement learning. IEEE Trans Smart Grid, 2021, 12: 420–431
4 Yang S, Lao K W, Hui H, et al. Secure distributed control for demand response in power systems against deception cyber-attacks with arbitrary patterns. IEEE Trans Power Syst, 2024, 39: 7277–7290
5 Hui H, Chen Y, Yang S, et al. Coordination control of distributed generators and load resources for frequency restoration in isolated urban microgrids. Appl Energy, 2022, 327: 120116
6 Huang X, Qi D, Chen Y, et al. Distributed self-triggered privacy-preserving secondary control of VSG-based AC microgrids. IEEE Trans Smart Grid, 2025, 16: 850–862
7 Lian Z, Zhu Y, Guo F, et al. Distributed cyber resilient control strategy for remote DC microgrids under integrated satellite terrestrial networks. IEEE Trans Ind Inf, 2025, 21: 2363–2372
8 Xu Z, Wang B C, Shen T. Mean field LQG social optimization: a reinforcement learning approach. Automatica, 2025, 172: 111924
9 Xu Z, Shen T, Huang M. Model-free policy iteration approach to NCE-based strategy design for linear quadratic Gaussian games. Automatica, 2023, 155: 111162
10 Lasry J M, Lions P L. Mean field games. Jpn J Math, 2007, 2: 229–260

11 Huang M Y, Malhamé R P, Caines P E. Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle. Commun Inf Syst, 2006, 6: 221–251
12 Li T, Zhang J F. Asymptotically optimal decentralized control for large population stochastic multiagent systems. IEEE Trans Automat Contr, 2008, 53: 1643–1660
13 de Paola A, Trovato V, Angeli D, et al. A mean field game approach for distributed control of thermostatic loads acting in simultaneous energy-frequency response markets. IEEE Trans Smart Grid, 2019, 10: 5987–5999
14 Benamor A, Habachi O, Kammoun I, et al. Mean field game-theoretic framework for distributed power control in hybrid NOMA. IEEE Trans Wireless Commun, 2022, 21: 10502–10514
15 von Stackelberg H. The Theory of the Market Economy. London: Oxford University Press, 1952
16 Wang B C, Zhang J F. Hierarchical mean field games for multiagent systems with tracking-type costs: distributed ε-Stackelberg equilibria. IEEE Trans Automat Contr, 2014, 59: 2241–2247
17 Cong W Y, Shi J T. Direct approach of linear-quadratic Stackelberg mean field games of backward-forward stochastic systems. In: Proceedings of 2024 the 43rd Chinese Control Conference (CCC), 2024
18 Moon J, Başar T. Linear-quadratic stochastic differential Stackelberg games with a high population of followers. In: Proceedings of the 54th IEEE Conference on Decision and Control, Osaka, 2015. 2270–2275
19 Lin Y, Jiang X, Zhang W. An open-loop Stackelberg strategy for the linear quadratic mean-field stochastic differential game. IEEE Trans Automat Contr, 2019, 64: 97–110
20 Zhou Z, Xu H. Decentralized adaptive optimal tracking control for massive autonomous vehicle systems with heterogeneous dynamics: a Stackelberg game. IEEE Trans Neural Netw Learn Syst, 2021, 32: 5654–5663
21 Yang S, Lao K W, Chen Y, et al. Resilient distributed control against false data injection attacks for demand response. IEEE Trans Power Syst, 2024, 39: 2837–2853
22 Nourian M, Caines P E, Malhame R P, et al. Mean field LQG control in leader-follower stochastic multi-agent systems: likelihood ratio based adaptation. IEEE Trans Automat Contr, 2012, 57: 2801–2816
23 Ma J, Yong J M. Forward-Backward Stochastic Differential Equations and Their Applications. Beilin: Springer-Verlag, 1999
24 Huang M, Zhou M. Linear quadratic mean field games: asymptotic solvability and relation to the fixed point approach. IEEE Trans Automat Contr, 2020, 65: 1397–1412

Appendix A Proof of Theorem 1

Proof. We consider the ith follower. For given $\zeta_i \in L^2_{\mathcal{F}^i_0}(\Omega; R)$, $u_0(\cdot) \in L^2(0,T;\mathbb{R})$, $\hat{u}_i(\cdot) \in L^2_{\mathcal{F}}(0,T;R)$, suppose that $(\hat{x}_i(\cdot), \hat{p}_i(\cdot), \hat{q}_i^j(\cdot), j=0,1,\cdots,N)$ is an adapted solution to (18). For any $u_i(\cdot) \in L^2_{\mathcal{F}}(0,T;\mathbb{R})$ and $\varepsilon \in R$, let $x_i^\varepsilon(\cdot)$ be the solution to the following perturbed state equation:

$$\begin{cases} \mathrm{d}x_i^\varepsilon = [a x_i^\varepsilon + b u_i^\varepsilon]\mathrm{d}t + \sigma \mathrm{d}B_i, \\ x_i^\varepsilon(0) = \zeta_i. \end{cases} \tag{A1}$$

Then, $\tilde{u}_i = \frac{u_i^\varepsilon(\cdot) - \hat{u}_i}{\varepsilon}$ and $\tilde{x}_i = \frac{x_i^\varepsilon - \hat{x}_i}{\varepsilon}$ are independent of ε and satisfy

$$\begin{cases} \mathrm{d}\tilde{x}_i = [a\tilde{x}_i + b\tilde{u}_i]\mathrm{d}t, \\ \tilde{x}_i(0) = 0, \end{cases} \tag{A2}$$

which depend on Lemma 3.1 in the reference[1)]. Applying Itô's formula to $\hat{p}_i(\cdot)\tilde{x}_i(\cdot)$, integrating from 0 to T, and taking the expectation, we have

$$\begin{aligned} 0 =& \mathbb{E}\left[\hat{p}_i(T)\tilde{x}_i(T) - \hat{p}_i(0)\tilde{x}_i(0)\right] \\ =& -\mathbb{E}\int_0^T \left\{ q_1\left(1 - \frac{1}{N}\right)(\hat{x}_i - \hat{x}^{(N)})\tilde{x}_i + q_2(\hat{x}_i - \hat{x}_0)\tilde{x}_i - b\tilde{u}_i\hat{p}_i \right\} \mathrm{d}t. \end{aligned} \tag{A3}$$

Hence,

$$\begin{aligned} & J_i^N\left(u_i^\varepsilon(\cdot), u_{-i}(\cdot), u_0(\cdot)\right) - J_i^N\left(\hat{u}_i(\cdot), u_{-i}(\cdot), u_0(\cdot)\right) \\ =& \frac{1}{2}\mathbb{E}\left\{\int_0^T \left[q_1\left(x_i^\varepsilon - x^{\varepsilon(N)}\right)^2 + q_2\left(x_i^\varepsilon - x_0\right)^2 + q_3 (u_i^\varepsilon)^2\right]\mathrm{d}t\right\} \\ & - \frac{1}{2}\mathbb{E}\left\{\int_0^T \left[q_1\left(\hat{x}_i - \hat{x}^{(N)}\right)^2 + q_2\left(\hat{x}_i - x_0\right)^2 + q_3 \hat{u}_i^2\right]\mathrm{d}t\right\} \\ =& \frac{1}{2}\varepsilon^2\mathbb{E}\left\{\int_0^T \left[q_1\left(1 - \frac{1}{N}\right)\tilde{x}_i^2 + q_2\hat{x}_i^2 + q_3\tilde{u}_i^2\right]\mathrm{d}t\right\} \\ & + \varepsilon\mathbb{E}\left\{\int_0^T \left[q_1\left(\hat{x}_i - \hat{x}^{(N)}\right)\left(1 - \frac{1}{N}\right)\tilde{x}_i + q_2\left(\hat{x}_i - x_0\right)\tilde{x}_i + q_3\hat{u}_i\tilde{u}_i\right]\mathrm{d}t\right\} \\ =& \underbrace{\frac{1}{2}\varepsilon^2\mathbb{E}\left\{\int_0^T \left[q_1\left(1 - \frac{1}{N}\right)\tilde{x}_i^2 + q_2\tilde{x}_i^2 + q_3\tilde{u}_i^2\right]\mathrm{d}t\right\}}_{\mathcal{N}_1} + \underbrace{\varepsilon\mathbb{E}\left\{\int_0^T \left[q_3\hat{u}_i\tilde{u}_i + b\tilde{u}_i\hat{p}_i\right]\mathrm{d}t\right\}}_{\mathcal{N}_2}, \end{aligned} \tag{A4}$$

where only here $\hat{x}^{(N)}(\cdot) := \frac{\hat{x}_i(\cdot)}{N} + \frac{1}{N}\sum_{j\neq i} x_j(\cdot), \hat{x}^{(N)}(\cdot) := \frac{\hat{x}_i(\cdot)}{N} + \frac{1}{N}\sum_{j\neq i} x_j(\cdot)$. According to $q_1, q_2 \geqslant 0$, $q_3 > 0$ and (19), we get $\mathcal{N}_1 \geqslant 0$ and $\mathcal{N}_2 = 0$. Therefore,

$$J_i^N\left(\hat{u}_i(\cdot), u_{-i}(\cdot), u_0(\cdot)\right) \leqslant J_i^N\left(\hat{u}_i(\cdot) + \varepsilon u_i(\cdot), u_{-i}(\cdot), u_0(\cdot)\right).$$

The proof is complete.

1) Andersson D, Djehiche B. A maximum principle for SDEs of mean-field type. Appl Math Optim, 2011, 63: 341–356.

Appendix B Proof of Theorem 3

Proof. Given $\zeta_0 \in L^2_{\mathcal{F}_0^0}(\Omega; R)$ and $\hat{u}_0(\cdot) \in L^2(0,T;\mathbb{R})$, let $\left(\hat{x}_0(\cdot), \hat{x}^{(N)}(\cdot), \hat{\phi}_N(\cdot), \hat{y}_0(\cdot), \hat{y}^{(N)}(\cdot), \hat{\psi}_N(\cdot)\right)$ be an adapted solution to (32). For any $u_0(\cdot) \in L^2(0,T;\mathbb{R})$ and $\varepsilon \in \mathbb{R}$, let $\left(x_0^\varepsilon(\cdot), x^{\varepsilon(N)}(\cdot), \phi_N^\varepsilon(\cdot)\right)$ be the solution to the following perturbed state equation of the leader:

$$\begin{cases} \mathrm{d}x_0^\varepsilon = [a x_0^\varepsilon + b u_0^\varepsilon]\,\mathrm{d}t, \quad x_0^\varepsilon(0) = \zeta_0, \\ \mathrm{d}x^{\varepsilon(N)} = \left[\left(a - q_3^{-1} b^2 \Pi_N\right) x^{\varepsilon(N)} - q_3^{-1} b^2 \phi_N^\varepsilon\right] \mathrm{d}t + \sigma \mathrm{d}B^{(N)}, \quad x^{\varepsilon(N)}(0) = \zeta^{(N)}, \\ \mathrm{d}\phi_N^\varepsilon = -\left[\left(a - q_3^{-1} b^2 \Pi_N\right) \phi_N^\varepsilon - q_2 x_0^\varepsilon\right] \mathrm{d}t, \quad \phi_N^\varepsilon(T) = 0. \end{cases} \tag{B1}$$

Then, let $\left(\tilde{x}_0(\cdot), \tilde{x}^{(N)}(\cdot), \tilde{\phi}_N(\cdot)\right)$ denote the solution to

$$\begin{cases} \mathrm{d}\tilde{x}_0 = [a\tilde{x}_0 + b\tilde{u}_0]\,\mathrm{d}t, \\ \mathrm{d}\tilde{x}^{(N)} = \left[(a - q_3^{-1} b^2 \Pi_N)\tilde{x}^{(N)} - q_3^{-1} b^2 \tilde{\phi}_N\right] \mathrm{d}t, \\ \mathrm{d}\tilde{\phi}_N = -\left[\left(a - q_3^{-1} b^2 \Pi_N\right) \tilde{\phi}_N - q_2 \tilde{x}_0\right] \mathrm{d}t, \quad \tilde{\phi}_N(T) = 0, \\ \tilde{x}_0(0) = 0, \quad \tilde{x}_i(0) = 0, \ i = 1, 2, \cdots, N. \end{cases} \tag{B2}$$

We have $x_0^\varepsilon(\cdot) = \hat{x}_0(\cdot) + \varepsilon \tilde{x}_0(\cdot), x^{\varepsilon(N)}(\cdot) = \hat{x}^{(N)}(\cdot) + \varepsilon \tilde{x}^{(N)}(\cdot), \phi_N^\varepsilon(\cdot) = \hat{\phi}_N(\cdot) + \varepsilon \tilde{\phi}_N(\cdot)$, and applying Itô's formula to $\tilde{x}_0(\cdot)\hat{y}_0(\cdot) + \tilde{x}^{(N)}(\cdot)\hat{y}^{(N)}(\cdot) + \tilde{\phi}_N(\cdot)\hat{\psi}_N(\cdot)$, integrating from 0 to T and taking expectation, we obtain

$$\begin{aligned} 0 &= -\mathbb{E}\left\{\int_0^T \mathrm{d}\left(\tilde{x}_0(\cdot)\hat{y}_0(\cdot) + \tilde{x}^{(N)}(\cdot)\hat{y}^{(N)}(\cdot) + \tilde{\phi}_N(\cdot)\hat{\psi}_N(\cdot)\right)\right\} \\ &= -\mathbb{E}\left\{\int_0^T \left[r_1\left(\hat{x}_0 - \hat{x}^{(N)}\right)\left(\tilde{x}_0 - \tilde{x}^{(N)}\right) + r_2\left(x_0 - x_0^{\mathrm{ref}}\right)\tilde{x}_0 - b\hat{y}_0\tilde{u}_0\right] \mathrm{d}t\right\}. \end{aligned} \tag{B3}$$

Hence,

$$\begin{aligned} & J_0\left(u_0^\varepsilon(\cdot)\right) - J_0\left(\hat{u}_0(\cdot)\right) \\ =& \frac{1}{2}\mathbb{E}\left\{\int_0^T \left[r_1\left(x_0^\varepsilon - x^{\varepsilon(N)}\right)^2 + r_2\left(x_0^\varepsilon - x_0^{\mathrm{ref}}\right)^2 + r_3 (u_0^\varepsilon)^2\right] \mathrm{d}t\right\} \\ & - \frac{1}{2}\mathbb{E}\left\{\int_0^T \left[r_1\left(\tilde{x}_0 - \tilde{x}^{(N)}\right)^2 + r_2\left(\tilde{x}_0 - x_0^{\mathrm{ref}}\right)^2 + r_3 \tilde{u}_0^2\right] \mathrm{d}t\right\} \\ =& \frac{1}{2}\varepsilon^2\mathbb{E}\left\{\int_0^T \left[r_1\left(\tilde{x}_0 - \tilde{x}^{(N)}\right)^2 + r_2 \tilde{x}_0^2 + r_3 \tilde{u}_0^2\right] \mathrm{d}t\right\} \\ & + \varepsilon\mathbb{E}\left\{\int_0^T \left[r_1\left(\hat{x}_0 - \hat{x}^{(N)}\right)\left(\tilde{x}_0 - \tilde{x}^{(N)}\right) + r_2\left(x_0 - x_0^{ref}\right)\tilde{x}_0 + r_3 \tilde{u}_0 \tilde{u}_0\right] \mathrm{d}t\right\} \\ =& \underbrace{\frac{1}{2}\varepsilon^2\mathbb{E}\left\{\int_0^T \left[r_1\left(\tilde{x}_0 - \tilde{x}^{(N)}\right)^2 + r_2 \tilde{x}_0^2 + r_3 \tilde{u}_0^2\right] \mathrm{d}t\right\}}_{Q_1} + \underbrace{\varepsilon\mathbb{E}\left\{\int_0^T \left[b\hat{y}_0\tilde{u}_0 + r_3 \hat{u}_0 \tilde{u}_0\right] \mathrm{d}t\right\}}_{Q_2}. \end{aligned} \tag{B4}$$

Due to $r_1, r_2 \geqslant 0,\ r_3 > 0$ and (33), we have $Q_1 \geqslant 0$ and $Q_2 = 0$. Therefore,

$$J_0^N\left(\hat{u}_0(\cdot)\right) \leqslant J_0^N\left(\hat{u}_0^\varepsilon(\cdot)\right).$$

The proof is complete.

Appendix C

Theorem C6. Let the pair $(\omega_1(\cdot), \omega_2(\cdot))$ be the solution to the following differential equation:

$$\frac{\mathrm{d}}{\mathrm{d}t}\binom{\omega_1(t)}{\omega_2(t)} = \begin{pmatrix} A_1 & B_1 \\ A_2 & B_2 \end{pmatrix}\binom{\omega_1(t)}{\omega_2(t)}, \quad \binom{\omega_1(T)}{\omega_2(T)} = \binom{0_{3\times 3}}{I_3}.$$

If $\omega_2(t)$ is invertible for all $t \in [0,T]$, then the Riccati equation (39) has a unique solution $\Phi(t) = \omega_1(t)\omega_2^{-1}(t)$ for all $t \in [0,T]$.

Proof. Note that

$$\frac{\mathrm{d}\omega_2^{-1}(t)}{\mathrm{d}t} = -\omega_2^{-1}(t)\frac{\mathrm{d}\omega_2(t)}{\mathrm{d}t}\omega_2^{-1}(t) = -\omega_2^{-1}(t)A_2\omega_1(t)\omega_2^{-1}(t) - \omega_2^{-1}(t)B_2,$$

which implies

$$\begin{aligned} \frac{\mathrm{d}\Phi(t)}{\mathrm{d}t} &= \frac{\mathrm{d}\omega_1(t)}{\mathrm{d}t}\omega_2^{-1}(t) + \omega_1(t)\frac{\mathrm{d}\omega_2^{-1}(t)}{\mathrm{d}t} \\ &= A_1\omega_1(t)\omega_2^{-1}(t) + B_1 - \omega_1(t)\omega_2^{-1}(t)A_2\omega_1(t)\omega_2^{-1}(t) - \omega_1(t)\omega_2^{-1}(t)B_2 \\ &= A_1\Phi(t) + B_1 - \Phi(t)A_2\Phi(t) - \Phi(t)B_2, \end{aligned}$$

and the proof is achieved.

Appendix D

If the form of $\hat{p}_i$ set in (20) is different, different Riccati equations can be obtained for solving. When we consider the following parameterization for $\hat{p}_i(\cdot)$,

$$\hat{p}_i(\cdot) = P_N(\cdot)(\hat{x}_i(\cdot) - \hat{x}^{(N)}(\cdot)) + \Pi_N(\cdot)\hat{x}^{(N)}(\cdot) + \hat{\phi}_N(\cdot), \tag{D1}$$

where $P_N(\cdot)$, $\Pi_N(\cdot)$ are differential functions with $P_N(T) = 0$, $\Pi_N(T) = 0$. By Itô's formula, we have

$$\begin{aligned} \mathrm{d}\hat{p}_i =& \dot{P}_N(\hat{x}_i - \hat{x}^{(N)})\mathrm{d}t + \left[aP_N\hat{x}_i - P_N q_3^{-1}b^2\left(P_N(\hat{x}_i - \hat{x}^{(N)}) + \Pi_N\hat{x}^{(N)} + \hat{\phi}_N\right)\right]\mathrm{d}t + P_N\sigma\mathrm{d}B_i \\ & - P_N\frac{1}{N}\sum_{j=1}^{N}\sigma\mathrm{d}B_j + \dot{\Pi}_N\hat{x}^{(N)}\mathrm{d}t + \left[a\Pi_N x^{(N)} - \Pi_N q_3^{-1}b^2\left(P_N(\hat{x}_i - \hat{x}^{(N)}) + \Pi_N\hat{x}^{(N)} + \hat{\phi}_N\right)\right]\mathrm{d}t \\ & + \Pi_N\frac{1}{N}\sum_{j=1}^{N}\sigma\mathrm{d}B_j + \mathrm{d}\hat{\phi}_N. \end{aligned} \tag{D2}$$

Instituting (D1) into the right side of the second equation in (18) obtains

$$\begin{aligned} \mathrm{d}\hat{p}_i =& -\left[a\left[P_N(\hat{x}_i - \hat{x}^{(N)}) + \Pi_N\hat{x}^{(N)} + \hat{\phi}_N\right] + \left[q_1\left(1 - \frac{1}{N}\right) + q_2\right]\hat{x}_i - q_1\left(1 - \frac{1}{N}\right)\hat{x}^{(N)} - q_2x_0\right] \\ & + \sum_{j=1}^{N} q_i^j\mathrm{d}B_j. \end{aligned} \tag{D3}$$

Comparing the coefficients of the right hand side of (D2) and (D3), yields

$$q_i^i = P_N\sigma + \frac{\Pi_N - P_N}{N}\sigma, \quad q_i^j = \frac{\Pi_N - P_N}{N}\sigma, \quad i \neq j, \tag{D4}$$

$$\dot{P}_N + 2aP_N - q_3^{-1}b^2P_N^2 + q_1\left(1 - \frac{1}{N}\right) + q_2 = 0, \quad P_N(T) = 0, \tag{D5}$$

$$\dot{\Pi}_N - \dot{P}_N + 2a\Pi_N - 2aP_N + q_3^{-1}b^2P_N^2 - q_3^{-1}b^2\Pi_N^2 - q_1\left(1 - \frac{1}{N}\right) = 0, \quad K_N(T) = 0, \tag{D6}$$

and

$$\mathrm{d}\hat{\phi}_N = -\left[\left(a - q_3^{-1}b^2\left(\Pi_N\right)\right)\hat{\phi}_N + q_2x_0\right]\mathrm{d}t, \quad \hat{\phi}_N(T) = 0. \tag{D7}$$