Online Optimal State Feedback Control of Linear Systems Over Wireless MIMO Fading Channels

Songfu Cai, Member, IEEE, and Vincent K. N. Lau, Fellow, IEEE

Abstract—In this article, we consider the optimal control of linear systems over wireless multiple-input multiple-output (MIMO) fading channels, where the wireless MIMO fading and random access of the remote controller may cause intermittent controllability or uncontrollability of the closed-loop control system. We consider the finite-time horizon optimal control first and then check the conditions for the existence and uniqueness of optimal control for the infinite-time horizon case. Specifically, for the finite-time horizon case, we show that the optimal control gain matrix can be computed offline based on the distribution of the wireless MIMO fading and random access of the controller. For the infinite-time horizon case, we show that, when the closed-loop system is almost surely controllable, the optimal control solution always exists and is unique. In the case that MIMO fading channels and the random access of the remote controller destroy the closed-loop controllability, we propose a novel controllable and uncontrollable positive semidefinite cone decomposition induced by the singular value decomposition of the MIMO fading channel contaminated control input matrix. This enables a closed-form characterization of the sufficient condition for both the existence and the uniqueness of the optimal control solution. The closed-form sufficient condition reveals the fact that the optimal control action may still exist even if the closed-loop system suffers from intermittent controllability or almost sure uncontrollability. We further propose a novel stochastic approximation (SA)-based online learning algorithm that can learn the optimal control action on the fly based on the plant state observations. We finally extend the technical results to the case in which the channel state information (CSI) is unavailable at the controller. We show that the sufficient condition for the existence and uniqueness of the optimal control solution is more difficult to satisfy due to the penalty caused by the unknown CSI. The proposed scheme is also compared with various baselines, and we show that significant performance gains can be achieved.

Index Terms—Online learning, optimal control, stochastic approximation (SA), uncontrollable linear systems, wireless multiple-input-multiple-output (MIMO) fading channels.

I. INTRODUCTION

OPTIMAL control has received considerable attention in both academia and industry in recent years. A wide spectrum of applications of optimal control can be found in areas, such as aerospace control, industrial flotation process control, automated vehicle systems, and robotics and manufacturing systems [1], [2], [3]. A typical closed-loop feedback control system consists of a dynamic plant (with potentially unstable dynamics), a remote controller, and an actuator, as illustrated in Fig. 1. The wireless network in-between the remote controller and the actuator will have significant impacts on the closed-loop control performance because it introduces various degradations, such as wireless fading, packet errors, and latency [4], [5], [6]. As a result, it is important to incorporate the impairments in the wireless networks into the optimal control design at the remote controller.

Recently, there have been some works on optimal control over static channels. Specifically, in [7], the authors consider the optimal real-time control of wind turbines. The optimal control action is obtained numerically using dynamic programming by maximizing the wind energy capture. In [8], the authors consider the problem of optimal control with transmission power management. Exploiting the information structure at the controller, the authors show that the linear quadratic regulator (LQR) control law is optimal. In [9], the authors consider optimal control design by iteratively solving the associated Bellman optimality equation. Specifically, under an initial stabilizing control policy, the optimal control action is obtained via policy iteration, whereas the initial stabilizing control policy is obtained by solving a linear matrix inequality (LMI). In [10], the authors propose an adaptive optimal control design using adaptive dynamic programming such that the a priori knowledge of an initial stabilizing control policy is no longer required. In [11], the authors propose a mixed mode value and policy iteration algorithm to obtain the optimal control action, which also avoids the need for an initial stabilizing control. However, in all the aforementioned works [7], [8], [9], [10], [11], the associated optimal control gain is static due to the consideration of the static communication channel between the remote controller and the actuator. Therefore, these existing approaches cannot achieve optimal control over dynamic wireless fading channels and brute-force applications of the existing methods [7], [8], [9], [10], [11] may even cause plant instability.

There are several works that consider estimation [12] and control [13] of continuous dynamic systems over unreliable
Illustration of a closed-loop state feedback control system, where the remote controller transmits its control action to the actuator over an MIMO wireless communication network.

wireless communication channels, where the effect of communication channel is modeled by random sampling of the plant state measurements. Specifically, in [12], the random state observation arrivals at the estimator are modeled by a Poisson stochastic process. The authors proposed a novel filter for estimating the expected value of the state. The authors analyzed the associated state estimation mean square error (MSE) and provided sufficient conditions for estimation stability in terms of the observation arrival rate. In [13], the authors considered the optimal control with Poisson-distributed state observations at the controller. The authors studied both the time-invariant piecewise-constant control and time-variant control that minimize a finite-horizon quadratic cost.

There are also several works considering the optimal control of discrete-time systems over wireless communication channels. The seminal work [14] models the wireless communication channel between the remote controller and the actuator as an independent and identically distributed (i.i.d.) random Bernoulli packet loss channel. By using a separation principle, the authors show that the optimal controller is a standard linear quadratic regulator (LQR) controller. In [15], the optimal control over unreliable communication links is studied. The authors provide a sufficient condition for the existence of the mean-square stabilizing optimal control, which minimizes a quadratic cost subject to disruptions of the information flow between the controller and the plant caused by packet losses. The authors in [16] consider the linear quadratic Gaussian (LQG) optimal control over parallel erasure fading channels. A sufficient requirement on the erasure probability of each individual communication link, which guarantees the existence of optimal control, is provided in terms of an LMI. In [17], a point-to-point packet dropping channel model with a fixed number of packet drops in a certain time interval is considered. The authors provide closed-form characterizations on the optimal control action that minimizes a finite-horizon LQR cost. In [18] and [19], the authors consider the finite burst of consecutive packet dropouts, and the optimal control action is obtained using the potential learning approach. However, such a packet-dropping channel model in [14], [15], [16], [17], [18], [19] is an oversimplification of the impairments introduced in practical wireless multiple-input multiple-output (MIMO) fading channels, where the wireless MIMO fading channel is switching in a continuous state space with uncountably many realizations. The authors in [20] and [21] consider the optimal control subject to multiplicative white Gaussian noises in dynamic systems. The authors propose an uncertainty threshold principle, which states that the optimal control exists if and only if the uncertainty does not exceed a given threshold. In [22], the authors consider the optimal control over continuous fading channels but only diagonal fading channels are considered. The system considered in [20], [21], [22] is required to be almost surely controllable at every time slot, which is a very restrictive requirement. Unfortunately, the MIMO fading channels are far more complicated than the diagonal fading channels in the sense that MIMO fading channels can destroy the system controllability and the resultant system may not be controllable at every time slot. The impacts of intermittent uncontrollability on the optimal control design have not been considered.

There also exist several other works considering the optimal control over wireless communication channels with unknown channel state information (CSI). Specifically, in [23], the authors consider the remote stabilization over packet drop-out channels with unknown instantaneous CSI. The optimal control that guarantees the closed-loop mean square stability is derived, which can be implemented at the controller based on the one-step delayed CSI. In [24], the authors further extend the results in [23] to the general parallel wireless fading channels. The random fading channel coefficients are modeled as model uncertainties. Using the robust control analysis and synthesis techniques, the authors provide a sufficient condition for mean square stability in terms of the feasibility of LMIs. Matveev and Savkin [25] considered the finite-horizon LQG control problem over a digital communication channel with finite channel capacity. The control signal is first encoded by the controller before transmission, and then decoded by the actuator for plant actuation. The authors prove the existence of optimal control action and develop the associated encoding and decoding schemes. Fang et al. [26] also considered the feedback stabilization over noisy digital wireless channels. Using an information-theoretic approach, the authors provide the lower bounds on disturbance power reduction in terms of the channel blurriness and the dynamic plant’s unstable eigenvalues. Ranade and Sahai [27] considered the stabilizability for scalar linear systems with unknown and unpredictable CSI. The authors develop a notion of “control capacity” and show that the closed-loop system is stabilizable if and only if the control capacity is larger than the log of the unstable open-loop eigenvalue. In [28], the authors consider the feedback stabilization over diagonal fading channels, where each fading channel coefficient is modeled as a white random process. The stabilization condition is obtained using Wonham decomposition approach.

We consider the optimal control over wireless MIMO fading channels, where the wireless fading may cause intermittent uncontrollability of the closed-loop control system. The following summarizes the key contributions of this article.

1) Existence and uniqueness of optimal control with intermittent controllability or almost sure uncontrollability: We consider the finite-time horizon optimal control first, and from that analyze the existence and uniqueness of the optimal control solution to the infinite-time horizon problem over wireless MIMO fading channels. For the infinite-time horizon problem, we show that the optimal
control solution always exists and is unique if we have \textit{almost sure controllability}. For the cases of intermittent controllability and almost sure uncontrollability, which are caused by the MIMO wireless fading and random access of the remote controller, we propose a novel control and uncontrollable positive semidefinite (PSD) cone decomposition technique that enables a closed-form characterization of the sufficient condition for the existence and uniqueness of the optimal control action.  

2) Online learning the optimal control action and the convergence analysis: We propose a novel stochastic approximation (SA)-based online learning algorithm that can learn the optimal control action $u_k$ on the fly based on the plant state observations $x_k$. The convergence of the proposed SA-based online learning algorithm is characterized via analyzing the associated limiting ordinary differential equation (ODE). We introduce a virtual fixed-point process for which the state trajectory is arbitrarily close to the trajectory of the limiting ODE. By analyzing the fixed-point operator associated with the virtual fixed-point process, we derive a closed-form sufficient condition for the convergence of the limiting ODE, which in turn renders the almost sure convergence of the online learning algorithm to the optimal control solution.  

3) Optimal control over wireless MIMO fading channels with unknown CSI: We further extend the technical results to the case that the instantaneous CSI is unavailable at the remote controller. We establish a closed-form sufficient condition for the existence and uniqueness of the optimal control solution, which incorporates the penalty caused by the unknown MIMO fading coefficients. We show that the optimal control solution is a static state feedback control, where the constant control gain matrix is a function of the MIMO fading channel statistics. We also propose an online fixed point iteration algorithm to learn the optimal control action.  

\textbf{Notation:} Uppercase and lowercase boldface denote matrices and vectors, respectively. $(\cdot)^T$ is the transpose. $0_{m \times n}$ denotes the $m \times n$ dimensional matrix with all the elements being zero. $||A||$ denotes the spectral norm of matrix $A$. $A_{i,j}$ denotes the element in the $i$th row and $j$th column of matrix $A$. $(A)_i$ denotes the $i$th row and $i$th column of matrix $A$. $(A)_{i,j}$ denotes the $j(i-1) \times (m-l+1)$ dimensional block submatrix of $A$ with the first element being $A_{i,j}$. $S^S$ and $S^P$ denotes the set of $S \times S$ dimensional PSD matrices and positive definite matrices, respectively. $\text{Pr}(A)$ denotes the probability of event $A$. $\otimes$ is the Kronecker product. $\text{vec}(\cdot)$ is the vectorization operator, and $\text{vec}^{-1}(\cdot)$ is the inverse vectorization operator of a vector to a square matrix with appropriate dimensions. $\mathbb{R}^{m \times n}$ represents the set of $m \times n$ dimensional real matrices. $\mathbb{R}(\mathbb{C})$ represents the set of real (complex) numbers.  

II. \textbf{SYSTEM MODEL}  

A. Dynamic Plant Model  

A typical closed-loop feedback control system is a geographically distributed system, wherein a potentially unstable dynamic plant, an actuator, and a remote controller are connected through a wireless communication network, as illustrated in Fig. 1. The dynamic evolution of the plant state $x_k$ is summarized below.  

\textbf{Assumption 1 (Dynamic Plant Model):} The plant state $x_k$ follows the dynamic evolution of $x_{k+1} = Ax_k + Bu_k + w_k$, $k \geq 0$, where $x_k \in \mathbb{R}^{S \times 1}$ is the plant state process, $S$ is the plant state dimension, $x_0$ is the initial state, $\bar{u}_k \in \mathbb{R}^{N_r \times 1}$ is the actuation control input, $A \in \mathbb{R}^{K \times K}$, $B \in \mathbb{R}^{S \times N_r}$, and $w_k \in \mathbb{R}^{S \times 1}$ is the plant noise with zero mean and finite covariance matrix $W$. The plant state transition matrix $A$ contains possibly unstable eigenvalues.  

B. Wireless Communication Model  

We model the wireless communication channel between the multiantenna remote controller and the actuator as a wireless MIMO fading channel. Using multiple-antenna techniques, the $N_r$-antenna controller transmits its control action $u_k$ to the $N_r$-antenna actuator through spatial multiplexing. At the $k$th time slot, the received control signal $\bar{u}_k \in \mathbb{R}^{N_r \times 1}$ at the actuator is given by  

$$\bar{u}_k = \delta_k H_k u_k + v_k$$

(1)  

where $H_k \in \mathbb{R}^{N_r \times N_r}$ is the MIMO channel fading matrix, $\delta_k \in \{0, 1\}$ is the random access variable, and $v_k \sim \mathcal{N}(0, I_{N_r})$ is the additive Gaussian channel noise.  

The random access variable $\delta_k$ is a feature of the wireless networked control system that characterizes the randomness of the remote controller’s access to the wireless MIMO fading channel. Specifically, $\delta_k = 1$ means that the controller is active to access the wireless MIMO fading channel, and the control action $u_k$ is transmitted by the controller to the actuator over the wireless MIMO fading channel, while if $\delta_k = 0$, the controller shuts down and the control action $u_k$ is not transmitted. We have the following assumption on $\delta_k$.  

\textbf{Assumption 2 (Controller Random Access Model):} For any given time slot $k$, $\delta_k$ is Bernoulli distributed with $\text{Pr}(\delta_k = 1) = \delta$ and $\text{Pr}(\delta_k = 0) = 1 - \delta$, where the constant $\delta \in (0, 1)$ denotes the active probability of the remote controller. Furthermore, $\delta_k$ is i.i.d. over the time slots.  

Additionally, we have the following assumption on the wireless MIMO fading channel $H_k$.  

\textbf{Assumption 3 (MIMO Wireless Fading Channel Model):} The random MIMO channel realization $H_k$ remains constant within each time slot and is i.i.d. over the time slots. Each element of $H_k$ is i.i.d. Gaussian distributed with zero mean and unit variance.  

Assumption 3 is quite mild and general in the field of wireless communications. In fact, in the case of flat fading, where the coherence bandwidth of the channel is larger than the bandwidth of the transmitted signal, the transmitted signal will experience the same magnitude of fading [29]. As a result, the realization of $H_k$ will remain constant within each time slot. Since the coherence time of the MIMO fading channel is small relative to the delay requirement of the closed-loop control applications, the fading channel $H_k$ varies fast enough to be represented by a temporally i.i.d. process [30]. Additionally, as pointed out in [30] and [31], in the case of a rich scattering environment with sufficient antenna spacing, the fading channel $H_k$ has no
III. EXISTENCE AND UNIQUENESS OF OPTIMAL CONTROL

A. Finite-Time Horizon Optimal Control

In order to formulate stochastic optimal control for the linear and time varying system, we first extend the system state from \( x_k \in \mathbb{R}^{N_x \times 1} \) to \( S_k = (x_k, H_k, \delta_k) \in \mathbb{R}^{N_x \times 1} \times \mathbb{R}^{N_h \times N_t} \times \{0, 1\} \). In this case, a control policy \( \pi \) is a sequence of mappings \( \pi = \{\Omega_0, \Omega_1, \ldots\} \), where the mapping \( \Omega_k \) at the \( k \)-th time slot is a mapping from the extended state set \( S_k^0 = \{S_0, S_1, \ldots, S_k\} \) to the control action \( u_k \), i.e., \( u_k = \Omega_k(S_k^0) \). We define the per time slot cost as

\[
 r(S_k, u_k) = x_k^T Q x_k + u_k^T R u_k + \mathbb{E} \left[ \mu_k^T M \mu_k | S_k \right] \tag{2}
\]

where \( x_k^T Q x_k \), \( Q \in \mathbb{S}^{N_x}_{+} \), is the state cost, \( u_k^T R u_k \), \( R \in \mathbb{S}^{N_h}_{+} \), is the controller’s transmission power cost, and \( \mathbb{E} \left[ \mu_k^T M \mu_k | S_k \right] \), \( M \in \mathbb{S}^{N_h}_{+} \), is the actuator’s control power cost. Note that in the existing literature of optimal control over wireless channels [15], [16], [21], the per time slot cost includes either only the state cost and the controller’s action generation power cost [13], or only the state cost and the actuator’s control power cost [16], [21]. Compared to the existing literature [15], [16], [21], the formulation of (2) is more comprehensive because it incorporates the power cost of both the controller and the actuator.

The finite-time horizon optimal control over the wireless MIMO fading channels can be formulated as the following Problem 1.

Problem 1 (Finite-Time Horizon Problem): The finite-time horizon optimal control problem is to find a control policy \( \pi^* \) that minimizes the quadratic cost function

\[
 J_k^\pi = \mathbb{E} \left[ x_k^T Q x_k + \sum_{k=0}^{K-1} r(S_k, u_k) \right] \tag{3}
\]

subject to the dynamic plant model in Assumption 1 and the wireless MIMO fading channel model in (1), where \( J_k^\pi \) denotes the finite-time horizon cost over a total number of \( K \) time slots under the control policy \( \pi \).

We have the following Lemma 1 to characterize the structure of the optimal control solution \( u_k^* \) to Problem 1.

Lemma 1 (Optimal Control Solution for Finite-Time Horizon Problem 1): The optimal control action \( u_k^* \) that minimizes the finite-time horizon cost \( J_k^\pi \) in Problem 1 is given by

\[
 u_k^* = -\delta_k \left( H_k^T B^T P_{k+1} B H_k + \delta_k H_k^T M H_k + R \right)^{-1} \cdot H_k^T B^T P_{k+1} A x_k \quad \forall 0 \leq k < K \tag{4}
\]

where \( \{P_k, 0 \leq k \leq K\} \) is given recursively by

\[
 P_k = Q \tag{5}
\]

\[
 P_k = A^T P_{k+1} A - \mathbb{E} \left[ \delta_k A^T P_{k+1} B H_k \left( H_k^T B^T P_{k+1} B H_k + \delta_k H_k^T M H_k + R \right)^{-1} H_k^T B^T \right] P_{k+1} + A + Q \tag{6}
\]

and the expectation \( \mathbb{E} \left[ \cdot \right] \) in (6) is taken with respect to (w.r.t.) the randomness of \( \delta_k \) and \( H_k \).

Proof: Please see Appendix A.

Based on the distribution of \( \delta_k \) and \( H_k \), the matrix sequence \( \{P_k, 0 \leq k < K\} \) can be computed numerically offline according to (5) and (6). The optimal control action \( u_k^* \) in (4) can, thus, be implemented in an online manner based on the precalculated \( P_{k+1} \).

B. Infinite-Time Horizon Optimal Control

We extend the finite-time horizon Problem 1 to the infinite-time horizon optimal control and formulate the following Problem 2.

Problem 2 (Infinite-Time Horizon Problem): The infinite-time horizon optimal control problem is to find a control policy \( \pi^* \) that minimizes the quadratic cost function

\[
 J^\pi = \lim_{K \to \infty} \sup_{\pi} \mathbb{E} \left[ \frac{1}{K} \sum_{k=0}^{K-1} r(S_k, u_k) \right] \tag{7}
\]

subject to the dynamic plant model in Assumption 1 and the wireless MIMO fading channel model in (1), where \( J^\pi \) denotes the infinite-time horizon average cost under the control policy \( \pi \).

Based on the finite-time horizon recursion (6), we reverse the time index similar to that in [15] and [32] and denote the reversed time index by \( i = K - k, \forall 0 \leq k \leq K \). Let \( P_0 = Q \), it follows that the finite-time horizon recursion (6) can be represented as

\[
 P_{i+1} = A^T P_i A - \mathbb{E} \left[ \delta_k A^T P_i B H_k \left( H_k^T B^T P_i B H_k + \delta_k H_k^T M H_k + R \right)^{-1} H_k^T B^T P_i A \right] + Q \quad \forall 0 \leq i \leq K. \tag{8}
\]

Let both the time horizon \( K \) and reversed time index \( i \) in (8) go to infinity, we have the following Theorem 1 on the solution to the infinite-time horizon Problem 2.

Theorem 1 (Optimal Control Solution for Infinite-Time Horizon Problem 2): If the nonlinear matrix equation (NME)

\[
 P = A^T P A - \mathbb{E} \left[ \delta_k A^T P B H_k \left( H_k^T B^T P B H_k + \delta_k H_k^T M H_k + R \right)^{-1} H_k^T B^T P A \right] + Q \tag{9}
\]

has a unique solution \( P^* \in \mathbb{S}^{N_x}_{+} \), where the expectation \( \mathbb{E} \left[ \cdot \right] \) in (9) is taken w.r.t. the randomness of \( \delta_k \) and \( H_k \), the optimal control action \( u_k^* \) that minimizes the infinite-time horizon cost \( J^\pi \) in Problem 2 exists and is unique, and is given by

\[
 u_k^* = -\delta_k \left( H_k^T B^T P^* B H_k + \delta_k H_k^T M H_k + R \right)^{-1} \left( H_k^T B^T P^* A x_k \right) \tag{10}
\]

Proof: Please see Appendix B.

The intuition behind Theorem 1 is illustrated Fig. 2. Specifically, when the time horizon \( K \) is sufficiently large and \( P_i \) attains its limit \( P^* \) at time index \( I \) in the reversed timeline, \( P_i \) will preserve the value of \( P^* \) for all \( i > I \). This means that in the unreversed timeline, \( P_k = P^* \) for all \( k < I \). Taking the limit \( K \to \infty (I \to \infty) \), the optimal control action for the infinite-time horizon case can thus be obtained by replacing...
Theorem 2. The solution to the finite-time horizon Problem 1 and the infinite-time horizon Problem 2.

**C. Sufficient Condition for the Existence and Uniqueness of the Optimal Control Action**

We first have the following Lemma 2 to characterize the impacts of the wireless MIMO fading channel and the controller’s random access on the controllability of the closed-loop system.

**Lemma 2 (Impacts of the MIMO Fading Channel and Random Access on Closed-loop Controllability):** Let the singular value decomposition (SVD) of \( BB^T \) be \( BB^T = U F \Sigma U^T \) with the diagonal elements of \( \Sigma \) in descending order, where \( U \in \mathbb{R}^{S \times S} \) is an unitary matrix. Let \( \text{rank}(B) = \eta_B \). Let the similarity transformation of \( A \) w.r.t. \( U \) be \( \tilde{A} = U^T A U \). Denote the block-wise representation of \( \tilde{A} \) by \( \tilde{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \), where \( A_{11} \in \mathbb{R}^{\eta_B \times \eta_B} \) is the \( \eta_B \)th order leading principal submatrix of matrix \( A \), \( A_{12} \in \mathbb{R}^{\eta_B \times (S - \eta_B)} \), \( A_{21} \in \mathbb{R}^{(S - \eta_B) \times \eta_B} \), and \( A_{22} \in \mathbb{R}^{(S - \eta_B) \times (S - \eta_B)} \) are constant matrices. Assume the pair \((A, B)\) is controllable, the impacts of \( H_k \) and \( \delta_k \) on the controllability of \((A, \delta_k B H_k)\) are given by the following:

1. **Almost Sure Controllability:** If one of the following three conditions (a.1), (a.2), and (a.3) is satisfied, then the closed-loop system is almost surely controllable, i.e., \((A, \delta_k B H_k)\) is controllable with probability 1 (w.p. 1) for any time slot \( k \).
   a.1) \( N_t \geq S \) and \( \Pr(\delta_k = 1) = 1, \forall k \geq 0 \).
   a.2) \( \eta_B \leq N_t < S, \Pr(\delta_k = 1) = 1, \forall k \geq 0 \), and the pair \((A_{22}, \tilde{A}_{12})\) is controllable.
   a.3) \( N_t < \eta_B, \Pr(\delta_k = 1) = 1, \forall k \geq 0 \), and \( \text{Rank}(A - \lambda I) > (S - \eta_B + N_t), \forall \lambda \in \mathbb{C} \).
2. **Intermittent Controllability:** If one of the three conditions (b.1), (b.2), and (b.3) is satisfied, then the closed-loop system is intermittent controllable, i.e., \((A, \delta_k B H_k)\) is almost surely controllable at time slot \( k \) when \( \delta_k = 1 \), and \((A, \delta_k B H_k)\) is uncontrollable at time slot \( k \) when \( \delta_k = 0 \).
   b.1) \( N_t \geq S \) and \( 0 < \Pr(\delta_k = 1) < 1, \forall k \geq 0 \).
   b.2) \( \eta_B \leq N_t < S, 0 < \Pr(\delta_k = 1) < 1, \forall k \geq 0 \), and the pair \((A_{22}, \tilde{A}_{12})\) is controllable.
   b.3) \( N_t < \eta_B, 0 < \Pr(\delta_k = 1) < 1, \forall k \geq 0 \), and \( \text{Rank}(A - \lambda I) > (S - \eta_B + N_t), \forall \lambda \in \mathbb{C} \).
   c.1) \( \eta_B \leq N_t < S \) and the pair \((A_{22}, \tilde{A}_{12})\) is controllable.
   c.2) \( N_t < \eta_B \) and \( \exists \lambda \in \mathbb{C} \) such that \( \text{Rank}(A - \lambda I) \leq (S - \eta_B + N_t) \).

**Proof:** Please see Appendix C.

The boundaries that distinguish almost sure controllability, intermittent controllability, and almost sure uncontrollability are visualized in Fig. 3. Lemma 2 delivers several key insights into the use of multiple antennas being able to enhance the closed-loop controllability. Specifically, a larger number of transmit antennas equipped at the remote controller \( N_t \) is more favorable for the closed-loop controllability. Suppose the controller is always active to transmit, i.e., \( \Pr(\delta_k = 1) = 1 \). If the controller has sufficiently many transmit antennas such that \( N_t > S \), the closed-loop system is automatically almost surely controllable. However, if \( N_t \) becomes smaller, an extra condition (a.2) on the plant state transition matrix \( A \) or (a.3) on the control input matrix \( B \) will be required to ensure the controllability. Moreover, a larger number of transmit antennas \( N_t \) is more favorable for the closed-loop controllability of larger dimensional plant systems.

We introduce a controllable and uncontrollable PSD cone decomposition of matrix \( P \) to analyze the solution of the NME (9) subject to the intermittent controllability or almost sure uncontrollability caused by \( \delta_k \) and \( H_k \). We first have the following definition of the controllable and uncontrollable cones of PSD matrices.
**Definition 1 (Controllable and Uncontrollable PSD Cones):**

Given a certain realization of $(\delta_k, H_k)$, the controllable PSD cone $C^c$ and uncontrollable PSD cone $C^{uc}$ associated with $\delta_kBH_k$ are defined by

$$C^c = \{ T \in S^S | \ker (\delta_kBH_k H_k^T B^T T) = \ker (T) \}$$

$$C^{uc} = \{ T \in S^S | \delta_kBH_k H_k^T B^T T = 0 \}.$$

Let the SVD

$$\delta_kBH_k \left( \delta_kH_k^T MH_k + R \right)^{-1} H_k^T B^T T = V_k^T A_k V_k$$

with the diagonal elements of $A_k$ in descending order. Let $\text{rank}(\delta_kBH_k H_k^T B^T T) = \gamma_k$ and $\Pi_k = [I, 0, 0] |S \times S$. The closed-form controllable and uncontrollable PSD cone decomposition of $P$ is characterized by the following Theorem.

**Theorem 2 (Closed-form PSD Cone Decomposition of $P$):**

Given a certain realization of $(\delta_k, H_k)$, the matrix $P$ defined by the NME (9) can be decomposed into two parts: $P = P^c_k + P^{uc}_k$, where $P^c_k \in C^c$ and $P^{uc}_k \in C^{uc}$. The closed-form expressions of $P^c_k$ and $P^{uc}_k$ are given by

$$P^c_k = V_k^T \left[ \sum_k (V_k P_{kk}^c)_{\gamma_k} \sum_k (V_k P_{kk}^c)_{\gamma_k} \right] V_k$$

$$P^{uc}_k = V_k^T (I - \Pi_k) V_k P_{kk}^c (I - \Pi_k) V_k$$

$$- V_k^T \text{diag} \left( 0_{\gamma_k}, \sum_k (V_k P_{kk}^c)_{\gamma_k} \right) V_k \quad (14)$$

where $\gamma_k = \{\gamma_k : 1 \leq \gamma_k \leq \gamma_k + 1 : S\}$.

**Proof:** Please see Appendix D.

Based on Theorem 2, the NME (9) can be represented in a more fine-grained form as

$$P = Q + A^T P \begin{bmatrix} P_{0k}^c & A + A^T P_{0k}^{uc} - \delta_kA^T P_{0k} BH_k \end{bmatrix}$$

$$\cdot \left( \delta_kH_k^T B^T P_{0k}^c BH_k + \delta_kH_k^T MH_k + R \right)^{-1} H_k^T B^T P_{0k}^c A. \quad (16)$$

The closed-form sufficient condition for the existence and uniqueness of $P$ that satisfies (9) can be obtained via analyzing the decomposed NME (16), which is summarized in the following Theorem 3.

**Theorem 3 (Sufficient Condition for the Existence and the Uniqueness of Optimal Control):** If one of the three conditions (a.1), (a.2), and (a.3) in Lemma 2 is satisfied, or condition (17)

$$\| \mathbb{E} \left[ A^T V_k^T \left( I - \Pi_k \right) V_k A \right] \| < 1$$

(17)

is satisfied, then the solution $P^*$ to the NME (9) exists and is unique. Moreover, the optimal control action $u_k^*$ that solves the infinite-time horizon Problem 2 exists and is unique, and is given by (10).

**Proof:** Please see Appendix E.

In the existing literature of control over wireless channels [15], [16], [20], [21], the almost sure controllability or intermittent controllability is an indispensable prerequisite for the existence of optimal control. However, our results in Theorem 3 reveal a key fact that in order to guarantee the existence of optimal control action, the closed-loop system does not necessarily need to be almost surely controllable or intermittently controllable. The optimal control action could still exist even if the system is almost surely uncontrollable at every time slot. As long as the “average” realization of the pair $(A, \delta_kBH_k)$ has good controllability behavior, i.e., condition (17) in Theorem 3 is satisfied, the optimal control $u_k^*$ still exists and is unique. Additionally, in the case of almost sure uncontrollability, the closed-loop system is also stable under the unique optimal control action $u_k^*$ (10) in the sense that $\lim\sup_{K \to \infty} \frac{1}{K} \sum_{k=0}^{K-1} \mathbb{E}[x_k^T Q x_k] < \operatorname{Tr}(M + P^* W + B^T P^* B)$, where $P^*$ is the unique solution to the NME (9).

**IV. ONLINE LEARNING OF OPTIMAL CONTROL OVER WIRELESS MIMO FADING CHANNELS**

**A. Learning of the Optimal Control Action**

We utilize the SA theory to construct an online learning algorithm to estimate the unknown $P$. Specifically, we first rewrite the NME (9) into the standard form of $f(P) = 0$ and apply the SA technique to estimate the root of the equation, where

$$f(P) = A^T P A - \delta_kA^T P BH_k (\delta_kH_k^T B^T P BH_k + \delta_kH_k^T MH_k + R)^{-1} H_k^T B^T P A - P + Q. \quad (18)$$

1) **Online Learning of $P^*$:** To obtain the root of $f(P) = 0$, we can apply the SA iteration

$$P_{k+1} = P_k + \alpha_k \hat{f}(P_k) \quad \forall k \geq 0 \quad (19)$$

where $P_0 \in \mathbb{S}^n_+$ is a bounded constant positive definite matrix, $(\alpha_k, k \geq 0)$ is the step-size sequence satisfying

$$\sum_{k=1}^{\infty} \alpha_k = \infty, \sum_{k=1}^{\infty} \alpha_k^2 < \infty \quad (20)$$

and $\hat{f}(P_k)$ is an unbiased estimator of $f(P_k)$, i.e., $\hat{f}(P_k) = \mathbb{E}[\hat{f}(P_k)] | P_k$, and is given by

$$\hat{f}(P_k) = A^T P_k A - \delta_kA^T P_k BH_k (\delta_kH_k^T B^T P_k BH_k + \delta_kH_k^T MH_k + R)^{-1} H_k^T B^T P_k A - P_k + Q. \quad (21)$$

2) **Online Learning of Control Action $u_k$:** At the $k$th time slot, the control action can be computed based on $P_k$

$$u_k = - \delta_k(\delta_kH_k^T B^T P_k BH_k + \delta_kH_k^T MH_k + R)^{-1} H_k^T B^T P_k A x_k. \quad (22)$$

In the iteration (19), the realization of $H_k$ will be required at the remote controller, which can be obtained by standard channel estimation at the actuator based on the received pilot symbols from the remote controller and channel feedback to the controller.

The following lemma summarizes several key properties of the proposed SA iteration (19).

**Lemma 3 (Properties of the SA Iteration (19))**:

1) **(Lipschitz Continuity):** The matrix-valued function $f(P)$ is Lipschitz continuous with Lipschitz constant
\[ \|A\|^2, \text{i.e.}, \|f(P^{(1)}) - f(P^{(2)})\| \leq (1 + \|A\|^2)\|P^{(1)} - P^{(2)}\|, \forall P^{(1)}, P^{(2)} \in \mathbb{R}_+^2. \]

2) (Martingale Difference Noise): Denote the estimation noise of \( f(P_k) \) in (19) by \( N_k = (\hat{f}(P_k) - f(P_k)) \). The sequence \( \{N_k, k \geq 0\} \) is a martingale difference sequence w.r.t. the filtration \( \{\mathcal{F}_k \} \). Consequently, the convergence of the state trajectory of the limiting ODE (23), if in (24), if \( \tilde{P}_k \) converges to \( P^* \), then the limiting convergent point \( P^* \) must be the root of \( f(P) = 0 \), i.e., \( f(P^*) = 0 \). As a result, if \( P_k \) in the proposed SA iteration (19) converges, it will also converge to the root of \( f(P) = 0 \). The full convergence results are formally summarized in the following Theorem.

**Theorem 4** (Almost Sure Convergence of the Proposed Online Learning Algorithm): If one of the three conditions (a.1), (a.2), and (a.3) in Lemma 2 is satisfied, or condition (17) in Theorem 3 is satisfied, denote the unique root of \( f(P) = 0 \) by \( P^* \), then

1) **Convergence of the Virtual Fixed-point Process:** \( P_k \) in the proposed fixed-point iteration (24) converges to \( P^* \) almost surely, i.e., \( \text{Pr}(\lim_{k \to \infty} P_k = P^*) = 1 \).

2) **Convergence of the SA Iteration:** \( P_k \) in the proposed SA iteration (19) converges to \( P^* \) almost surely, i.e., \( \text{Pr}(\lim_{k \to \infty} P_k = P^*) = 1 \).

3) **Convergence of the Control Action:** The learned control action \( u_k \) in (22) converges to the optimal control action \( u_k^* \) (10) in Theorem 1 almost surely, i.e.,

\[ \text{Pr}\left(\lim_{k \to \infty} u_k = u_k^*\right) = 1. \] (27)

**Proof:** Please see Appendix H.

V. EXTENSION TO OPTIMAL CONTROL OVER MIMO FADING CHANNELS WITH UNKNOWN CSI

A. Structure of Optimal Control Action With Unknown CSI

In this section, we consider the case where there is no feedback channel between the actuator and the remote controller, and the actuator cannot estimate and feedback the CSI \( H_k \) to the controller. As a result, the instantaneous CSI \( H_k \) is completely unknown at the controller. We assume the controller only knows the channel statistics \( \mathbb{E}[H_k] = H \in \mathbb{R}^{N_t \times N_r} \) and \( \mathbb{E}[H_k \otimes H_k'] = \Phi \in \mathbb{R}^{N_t \times N_t'} \). The extended system state in this case is \( \bar{S}_k = (x_k, \bar{u}_k) \in \mathbb{R}^{N_t+1} \times \{0, 1\} \). Due to the unavailability of \( H_k \), the control policy \( \bar{\pi} \) in this case is a sequence of mappings \( \{\Phi^1, \Phi^2, \ldots\} \), where \( \Phi^k \) at the \( k \)-th timeslot is a mapping from the extended system state \( \bar{S}_k = (x_k, \bar{u}_k) \) to the control action \( u_k \), i.e., \( u_k = \Phi^k(\bar{S}_k) \). Since the CSI is completely unknown, the per time slot cost should not depend on the realization of \( H_k \), and is, thus, given by

\[ \bar{r}(\bar{S}_k, u_k) = x_k^T Q x_k + u_k^T R u_k + \mathbb{E} \left[ u_k^T M u_k \right] \] (28)
where the expectation $\mathbb{E}[\cdot]$ is taken w.r.t. the randomness of the MIMO fading channel $H_k$ and the channel noises.

The infinite-time horizon optimal control can thus be formulated as the following Problem 3.

**Problem 3 (Infinite-Time Horizon Problem with Unknown CSI):** The infinite-time horizon optimal control problem with unknown CSI is to find a control policy $\tilde{\pi}^*$ that minimizes the quadratic cost function

$$J^{\tilde{\pi}} = \limsup_{K \to \infty} \frac{1}{K} \sum_{k=0}^{K-1} \mathbb{E} \left[ \tilde{J}_k (S_k, \gamma_k) \right]$$

subject to the dynamic plant model in Assumption 1 with unknown instantaneous CSI $H_k$ and known channel statistics $H$ and $\Phi$, where $J^{\tilde{\pi}}$ denotes the infinite-time horizon average cost under the control policy $\tilde{\pi}$ with unknown instantaneous CSI.

We have the following Lemma 6 to characterize the structural properties of the optimal control solution $u_k^*$ to Problem 3.

**Lemma 6 (Optimal Control Solution for Infinite-Time Horizon Problem 3 with Unknown CSI):** If the deterministic NME

$$P = A^T PA - \delta A^T PBH (\text{vec}^{-1} (\Phi \text{vec} (M + B^T PB))) + R)^{-1} H^T B^T PA + Q$$

(30)

has a unique solution $P^* \in S_+^S$, the optimal control solution $u_k^*$ to Problem 3 exists and is unique, and is given by

$$u_k = - \delta_k A^T PBH (\text{vec}^{-1} (\Phi \text{vec} (M + \delta_k B^T P^* B))) + R)^{-1} H^T B^T P^* A x_k.$$  

(31)

**Proof:** Please see Appendix I.

Let $\text{rank}(\delta_k BH) = \tilde{\gamma}_k$. Let the SVD $\delta_k BH (\text{vec}^{-1} (\Phi \text{vec} (M + \delta_k B^T B))) - HB^T BH + R)^{-1}$.

$$H^T B^T = \tilde{V}_k^T \tilde{A}_k \tilde{V}_k.$$  

(32)

with the diagonal elements of $\tilde{A}_k = \text{diag}(\eta_1, \eta_2, \ldots, \eta_{\tilde{\gamma}_k}, 0, \ldots, 0)$ in descending order, i.e., $\eta_i \geq \eta_j > 0, \forall 1 \leq i < j \leq \tilde{\gamma}_k$ and $\tilde{V}_k = [V_k, 0, 0, \ldots]^{S \times S}$. We have the following Theorem 5 on the existence and uniqueness of $P$ that satisfies the deterministic NME (30).

**Theorem 5 (Sufficient Condition for the Existence and the Uniqueness of Optimal Control with Unknown CSI):** If the condition

$$\left\| \mathbb{E} \left[ A^T \tilde{V}_k^T (I - \tilde{\Pi}_k) \tilde{V}_k A \right] \right\| + \| A \|^2 \mathbb{E} \left[ \sum_{i=1}^{\tilde{\gamma}_k} \eta_i^{-1} \right] < 1$$

(33)

is satisfied, where $\mathbb{E}[\cdot]$ in (33) is taken w.r.t. the randomness $\delta_k$, then the solution $P^*$ to the deterministic NME (30) exists and is unique. Moreover, the optimal control action $u_k^*$ that solves the infinite-time horizon Problem 3 exists and is unique, and is given by (31).

**Proof:** Please see Appendix J.

Compared to the case that the CSI $H_k$ is completely known at the remote controller, the sufficient condition (33) is more difficult to satisfy due to the penalty term $\| A \|^2 \mathbb{E} \left[ \sum_{i=1}^{\tilde{\gamma}_k} \eta_i^{-1} \right]$ caused by the unknown CSI $H_k$. Therefore, we can conclude that the knowledge of the fading coefficients at the controller is beneficial and favorable for the existence and uniqueness of the optimal control over wireless MIMO fading channels.

**B. Online Learning of the Optimal Control Action With Unknown CSI**

Note that the deterministic NME (9) is a standard fixed point equation w.r.t. $P$. As a result, we can use a fixed point iteration method to solve the unknown variable $P$.

1) **Online learning of $P^*$:** Initialize $P_0 = Q, u_0 = 0$ and $k = 0$. At the $k$th time slot, the remote controller updates $P_k$ as

$$P_{k+1} = A^T P_k A - \delta A^T P_k B H (\text{vec}^{-1} (\Phi \text{vec} (M + B^T P_k B))) + R)^{-1} H^T B^T P_k A + Q.$$  

(34)

The convergence of the proposed online learning scheme (34) and (35) is characterized in the following Theorem 6.

**Theorem 6 (Online Learning Convergence with Unknown CSI):** If the sufficient condition (33) in Theorem 5 is satisfied, $u_k$ in (35) converges to the ground-truth optimal control action $u_k^*$ in (31) as the time index $k$ goes to infinity.

**Proof:** The proof of the convergence of $P_k$ in (34) to the unique $P^*$ that satisfies the deterministic NME (30) follows from the monotonicity in the proof of Lemma 6 and the monotone convergence theorem similar to that in the proof of Theorem 1 and, hence, is omitted for brevity.

**VI. NUMERICAL RESULTS**

We compare the performance of the proposed online optimal control scheme with the following baselines via numerical simulations.

1) **Baseline 1 (Existing Q-learning-based LQR for Static Channels) [34], [35], [36], [37]:** The remote controller adopts the existing Q-learning-based LQR solution that is designed for static channels to generate the control actions. Specifically, the Q-function is given by $Q_{BL1}(x_k, u_k) = x_k^T u_k^T \Psi_{BL1} [x_k; u_k]^T$, where $\Psi_{BL1} \in \mathbb{R}^{(S+N+1) \times (S+N+1)}$ is the kernel matrix. The remote controller uses the Q-learning method in [34], [35], [36], [37] to obtain $\Psi_{BL1}$. Based on the learned kernel $\Psi_{BL1}$, the remote controller generates the control action $u_{BL1} = \arg \min_{u_k} Q_{BL1}(x_k, u_k)$.

2) **Baseline 2 (Brute-force Q-learning-based LQR over Wireless Channels without State Reduction):** The remote controller brute-force applies the existing
Comparison of the CPU time versus the plant state dimension is assumed known. For the learned by the proposed scheme converges to the ground truth $u$ dimensions, which is otherwise, with $\because$ in Baseline 2 not to be optimal. For Baseline 3, $-\delta_k = S = 0 \leq S = 0$ using the $1$ function of the $-2 = S = 0 \leq S = 0$ is illustrated, $\kappa = S = 0$ is chosen as $\leq Q = 2 \cdot R$ with $Q = 1) = 0$. Comparison of the accuracy of learned control actions versus (Genie-aided optimal LQR control): $\gamma = 0$, $\deltau = Q$, $\delta = 0$. $\delta_k = 0$, $\delta = 0$ are state variables in the $3$ $Q$ $Q$ and the $t$ state transition matrix $T$. It can be shown that the Q-learning-based LQR approach for closed-loop control over wireless fading channels without state reduction. Both the CSI $H_k$ and the controller random access state $\delta_k$ are state variables in the Q-function. Specifically, the $Q$-function is given by $Q_{BL2}(x_k; \delta_k, H_k, u_k) = [x_k^T, \delta_k, \text{vec}(H_k)^T, u_k]^T$, $u_k \Psi_{BL2}[x_k; \delta_k, \text{vec}(H_k); u_k]^T$, where $\Psi_{BL2} \in \mathbb{R}^{(1+S+N_r+N_l)+(1+S+N_r+N_l)}$. The remote controller uses the Q-learning method in [34], [35], [36], and [37] to obtain $\Psi_{BL2}$. Based on the learned kernel $\Psi_{BL2}$, the remote controller generates the control action $u_{k, BL2} = \arg \min_{u_k} Q_{BL2}(x_k, \delta_k, H_k, u_k)$.

3) Baseline 3 (Genie-aided optimal LQR control): The remote controller adopts the genie-aided optimal control solution. Specifically, the remote controller is assumed to know the $P^*$ that satisfies the NME (9). The remote controller generates the optimal control action $u_k^*$ according to (10).

A. Comparison of the Accuracy of Learned Control Actions

The accuracy of the learned control action $u_k$ versus the time index $k$, i.e., $E[\|u_k - u_k^*\|^2]$, is illustrated Fig. 4. It can be observed that with the increase of time, the gap between the learned control action and the optimal control action becomes prohibitively large for both Baseline 1 and Baseline 2. This is because Baseline 1 is designed for static channels and the impacts of random fading channels and random access of the controller are imprudently ignored. Baseline 2 assumes that the Q-function corresponding to the extended state $(x_k, \delta_k, H_k, u_k)$ is in the form of $Q_{BL2}(x_k; \delta_k, H_k, u_k) = [x_k^T, \delta_k, \text{vec}(H_k)^T, u_k]^T$. However, the ground truth $Q$-function is $Q(x_k; \delta_k, H_k, u_k) = [x_k^T, u_k]^T$.

The computational complexity (the CPU time for $10^4$ simulation runs) versus the plant state dimension $S$. The system parameters are configured as follows: the $(i,j)$th element $A$ is chosen as $(A)_{ij} = -0.1$ when $i = j - 1$, $(A)_{ij} = -0.2$ when $i = j + 1$, $(A)_{ij} = 1.01$ when $i = j$, and $(A)_{ij} = 0$ otherwise, with $1 \leq i, j \leq S$. The $(i,j)$th element $B$ is chosen as $(B)_{ij} = (i + j)^{-1}$ with $1 \leq i \leq S$ and $1 \leq j \leq 2$. $W = 0.003 \delta_k S$, $Q = I_S$, $M = I_2$, $N_i = 3$, $N_r = 2$, and $Pr(\delta_k = 1) = 0.5$.

B. Comparison of the Computational Complexity

The computational complexity versus the plant state dimension $S$. The computational complexity gap between the proposed scheme and Baseline 3 is due to the computation of $P^*$, where the proposed scheme needs to compute $P^*$ using the proposed SA iteration (19).

VII. Conclusion

In this article, we considered online optimal control over wireless MIMO fading channels. We consider the finite-time horizon optimal control first, and then analyze the existence and uniqueness of the control solution to the infinite-time horizon optimal control. We further propose a novel SA-based online algorithm to learn the optimal control action. We derived a closed-form sufficient condition that guarantees the almost sure convergence of the proposed SA-based online learning algorithm to the optimal control solution. We also extended the results to the optimal control with unknown instantaneous CSI.
the remote controller. The proposed scheme was compared with various baselines, and we showed that significant performance gains can be achieved.

APPENDIX

A. Proof of Lemma 1

The optimal solution for the finite-time horizon optimal control can be obtained by backward induction. We define the value functions as $V_K = x_k^T Q x_K$ and

$$V_k = \min_{u_k} \mathbb{E} [r(S_k, u_k) + V_{k+1}], 0 \leq k < K. \quad (36)$$

Let $k = K - 1$. Substitute $V_K$ into (36) and note that $x_K = A x_{K-1} + \delta_{K-1} B^H K_{-1} u_{K-1} + B v_{K-1} + w_{K-1}$, it follows that

$$V_{K-1} = \mathbb{E} [x_{K-1}^T (A^T Q A + Q) x_{K-1}] + \text{Tr} (M + Q W) + \min_{u_{K-1}} \mathbb{E} [u_{K-1}^T (\delta_{K-1} H_{K-1}^T B^T Q B H_{K-1} + \delta_{K-1} H_{K-1}^T R)] u_{K-1} + 2 x_{K-1}^T A^T Q \delta_{K-1} B^H K_{-1} u_{K-1}. \quad (37)$$

Let $P_K = Q$, it follows that

$$u_{K-1} = -\delta_{K-1} (\delta_{K-1} H_{K-1}^T B^T P_K B H_{K-1} + \delta_{K-1} H_{K-1}^T R)^{-1} H_{K-1}^T B^T p_K A x_{K-1}. \quad (38)$$

Substitute (38) back into (37), $V_{K-1}$ can be represented as

$$V_{K-1} = \text{Tr} (M + Q W) + x_{K-1}^T P_K x_{K-1}. \quad (39)$$

with

$$P_{K-1} = A^T P_K A - \mathbb{E} [\delta_{K-1} A^T P_K B H_{K-1} (H_{K-1}^T B^T P_K B H_{K-1} + \delta_{K-1} H_{K-1}^T R)]^{-1} H_{K-1}^T B^T H_{K-1} + M + Q W. \quad (40)$$

Let $k = K - 2$. Substitute $V_{K-1}$ in (39) into (36) and note that $x_{K-1} = A x_{K-2} + \delta_{K-2} B^H K_{-2} u_{K-2} + B v_{K-2} + w_{K-2}$, and continue on. We can show that (38) and (39) hold true for all $0 \leq k < K$. As a result, the optimal control solution $u^*_k$ to the finite-time horizon Problem 1 is characterized by (4), (5) and (6). Therefore, Lemma 1 is proved.

B. Proof of Theorem 1

Based on [15, Th. 3] or [32, Th. 6.9], it suffices to show that $P_i (8)$ in the reversed timeline will converge to $P^*$ as the reversed time index $i$ (and the time horizon $K$) goes to infinity. We prove this using the monotone convergence theorem.

Denote $g(P) = f(P) + P$, where $f(P)$ is given by (18) in Section IV-A. Note that $g(P)$ is monotonically increasing w.r.t. $P$. Further note that there exist two positive constant $\gamma_1, \gamma_2$, such that $0 < \gamma_1 < 1 < \gamma_2$ and $\gamma_1 P^* < P_i |_{i=0} = Q < \gamma_2 P^*$. We now construct two matrix sequences

$$\left\{ P_i^{(1)} : P_{i+1}^{(1)} = g(P_i^{(1)}), P_0^{(1)} = \gamma_1 P^*, i \geq 0 \right\}. \quad (41)$$

$$\left\{ P_i^{(2)} : P_{i+1}^{(2)} = g(P_i^{(2)}), P_0^{(2)} = \gamma_2 P^*, i \geq 0 \right\}. \quad (42)$$

When the “if condition” on the NME in (9) at the beginning of Theorem 1 is satisfied, we have $g(\gamma_1 P^*) > \gamma_1 g(P^*) = \gamma_1 P^*$ and $g(\gamma_2 P^*) < \gamma_2 g(P^*) = \gamma_2 P^*$. It follows that

$$P_i^{(1)} |_{i=0} = g(P_0^{(1)}) = P_0^{(2)} \quad (43)$$

$$P_i^{(2)} |_{i=0} = g(P_0^{(2)}) = P_0^{(1)} \quad (44)$$

Taking the matrix function $g(\cdot)$ on every component of (43) and (44), and using mathematical induction, we have

$$\gamma_1 P^* < P_i^{(1)} |_{i=0} < P_i^{(2)} |_{i=0} < P_i^{(1)} < P_i^{(2)} < \gamma_2 P^* \quad \forall i \geq 1. \quad (45)$$

Therefore, the monotonically increasing sequence $\{P_i^{(1)}, i \geq 0\}$ is bounded from above and the monotonically decreasing sequence $\{P_i^{(2)}, i \geq 0\}$ is bounded from below. It follows from the monotone convergence theorem that $\{P_i^{(1)}, i \geq 0\}$, $\{P_i^{(2)}, i \geq 0\}$ and $\{P_i, i \geq 0\}$ will converge to the unique solution $P^*$ that satisfies $g(P^*) = P^*$. As a result, we conclude that $\lim_{i \to \infty} P_i = P^*$, and hence, Theorem 1 is proved.

C. Proof of Lemma 2

We prove Lemma 2 based on the Popov–Belevitch–Hautus (PBH) test. Specifically, provided that $(A, B)$ is controllable, if there is a vector-scalar pair $(\lambda, v)$, $\lambda \in \mathbb{R}$, $v \in \mathbb{R}^S$, that $\text{Av} = \lambda v$ and $B^T v = 0$, then $v = 0$.

We first prove the following proposition.

Proposition 1: $\delta_k B H_k$ is statistically identical to $\delta_k (B B^T)^{\frac{1}{2}} H_k$, where each element of $H_k \in \mathbb{R}^{S \times N_t}$ is i.i.d. Gaussian distributed with zero mean and unit variance.

Proof: For any realization of $\delta_k$, if $\text{Rank} (B B^T) = S$, we choose $H_k$ to be $H_k = (B B^T)^{\frac{1}{2}} B^T H_k$. Since each element of $H_k$ is i.i.d. Gaussian distributed with zero mean and unit variance, it follows that each element of $H_k$ is also i.i.d. Gaussian distributed with zero mean and unit variance.

In the case that $(B B^T)^{\frac{1}{2}}$ is rank deficient, let $\text{Rank} (B B^T)^{\frac{1}{2}} = \eta_B < S$. Let the SVD of $(B B^T)^{\frac{1}{2}}$ be $(B B^T)^{\frac{1}{2}} = U T(\Lambda) U^T$, where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_{\eta_B}, 0, \ldots, 0)$, $\lambda_i (1 \leq i \leq \eta_B)$ are the nonzero singular values of $(B B^T)^{\frac{1}{2}}$. Let $A = \text{diag}((\lambda_1)^{-1}, \ldots, (\lambda_{\eta_B})^{-1}, 0, \ldots, 0)$, let $[\tilde{H}(k)]_{l,m}$ be the $S \times S$ dimensional matrix with all the elements being 0 except the $l$th row and the $m$th column element which are $(\tilde{H}(k))_{l,m}$. Let $h_{l,m} = (U^T \tilde{A} U BH_k)_{l,m}$. Let $g_{l,m}$ be an i.i.d. Gaussian distributed random variable with zero mean and unit variance. In this case, $\tilde{H}(k)$ is given by:

$$\tilde{H}(k)_{l,m} = \begin{cases} h_{l,m}, & \text{if } h_{l,m}(k) \neq 0 \\ g_{l,m}, & \text{otherwise}. \end{cases} \quad (46)$$

Therefore, Proposition 1 is proved. 

In the following, we prove Lemma 2 based on the aforementioned PBH test and Proposition 1.
Proof of (a.1): Given \( \delta_k = 1 \), suppose that there is a vector-scalar pair \((\lambda, \tilde{v})\) such that \( \tilde{A}\tilde{v} = \lambda\tilde{v} \) and
\[
\left( (BB^T) \frac{2}{3} \tilde{H}_k \right)^T \tilde{v} = \left( \tilde{H}_k \right)^T \left( (BB^T) \frac{2}{3} \right) \tilde{v} = 0. \tag{47}
\]
Since \( N_t \geq S \), it follows that \( \tilde{H}_k (\tilde{H}_k)^T \) is full rank w.p. 1. Multiplying \( (\tilde{H}_k (\tilde{H}_k)^T)^{-1}\tilde{H}_k \) on both sides of (47), it follows that \( (BB^T)^{\frac{2}{3}} \tilde{v} = 0 \). Since \( (BB^T)^{\frac{2}{3}} \) and \( (BB^T)^{\frac{2}{3}} \) have the same null-space, we conclude that \( BB^T \tilde{v} = 0 \), which leads to \( B^T \tilde{v} = 0 \). Since the pair \((A, B)\) is controllable, we conclude that \( \tilde{v} = 0 \). As a result, \((A, BH_k)\) is almost surely controllable. 

Therefore, (a.1) in Lemma 2 is proved.

Proof of (b.1): Note that when \( \delta_k = 0 \), \((A, 0)\) is uncontrollable. Condition (b.1) follows readily from (a.1).

In the case that \( N_t < S \), \( \tilde{H}_k (\tilde{H}_k)^T \) is rank deficient and \( \text{Rank}(\tilde{H}_k) = N_r \), w.p. 1. We have
\[
\left( \tilde{H}_k \right)^T BB^T \tilde{v} = \left( \tilde{H}_k \right)^T U^T \Xi U \tilde{v} = \tilde{H}_k \tilde{A} \tilde{v}
\]
where \( \tilde{H}_k = (U\tilde{H}_k) \in \mathbb{R}^{N_r \times S} \) is a random matrix with each element being i.i.d. Gaussian distributed with zero mean and unit variance, \( \tilde{v}_1 \in \mathbb{R}^{n_B} \), \( \tilde{v}_2 \in (S-\eta_B) \times 1 \) and \( v = [\tilde{v}_1, \tilde{v}_2] \).

Now suppose \( \left( \tilde{H}_k \right)^T BB^T \tilde{v} = 0 \), it follows that \( \tilde{v}_1 \) must lie in the null space of \( \left( \tilde{H}_k \right)^T,1:n_B \cdot \text{diag}(\lambda_1, \ldots, \lambda_n_B) \), whereas \( \tilde{v}_2 \) can take any value provided that \( \tilde{v}_2 \in (S-\eta_B) \times 1 \). Therefore, \( \tilde{v}_1 \) can be expressed as \( \tilde{v}_1 = \lambda_i \tilde{v} \in \mathbb{R}^{S_n} \times 1 \) and \( \tilde{v}_2 = [v_1, v_2] = \tilde{H}_k \tilde{A} \tilde{v} \).

Moreover, note that \( \tilde{A} \tilde{v} = A U^T \tilde{U} \tilde{v} \), it follows that
\[
UA\tilde{v} = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ \tilde{A}_{21} & \tilde{A}_{22} \end{bmatrix} \begin{bmatrix} \tilde{v}_1 \\ \tilde{v}_2 \end{bmatrix} = \begin{bmatrix} \tilde{A}_{12} \tilde{v}_2 \\ \tilde{A}_{22} \tilde{v}_2 \end{bmatrix}. \tag{50}
\]

Therefore, \( \tilde{A} \tilde{v} = \lambda \tilde{v} \) is equivalent to \( \tilde{A}_{12} \tilde{v}_2 = 0 \) and \( \tilde{A}_{22} \tilde{v}_2 = \lambda \tilde{v}_2 \). It follows that \( \tilde{v}_2 = 0 \) if and only if the pair \((\tilde{A}_{22}, \tilde{A}_{12})\) is controllable. As a result, we conclude that when \( N_t = \eta_B \) and \( (\tilde{A}_{22}, \tilde{A}_{12}) \) is controllable, the \( \tilde{v} \) that simultaneously satisfies \( \left( \tilde{H}_k \right)^T BB^T \tilde{v} = 0 \) and \( \tilde{A} \tilde{v} = \lambda \tilde{v} \) is \( \tilde{v} = 0 \), i.e., \((A, BH_k)\) is controllable w.p. 1. Therefore, (a.2) in Lemma 2 is proved.

Proof of (a.3): In the case that \( N_t < \eta_B \), the left inverse of \( \left( \tilde{H}_k \right)^T,1:n_B \cdot \text{diag}(\lambda_1, \ldots, \lambda_n_B) \) exists, and
\[
\left( \tilde{H}_k \right)^T,1:n_B \cdot \text{diag}(\lambda_1, \ldots, \lambda_n_B) \tilde{v}_1 = 0 \iff \tilde{v}_1 = 0. \tag{49}
\]

D. Proof of Theorem 2

We are trying to decompose \( P \) into two PSD components as \( P = \sum_{k=1}^{K} V_k^P V_k^T \), where \( V_k^P \in \mathbb{C}^{n_B \times n_B} \) belongs to the controllable PSD cone \( C^c \) and \( V_k^P \in \mathbb{C}^{n_B \times n_B} \) belongs to the uncontrollable PSD cone \( C^{uc} \). We then specify the detailed expression of \( \tilde{P}_c \) and \( \tilde{P}_c^{uc} \). We then verify that \( V_k^P \in C^c \) and \( V_k^P \in C^{uc} \) indeed hold true.

Applying the matrix inversion lemma in (1b) in [39], it follows that, for given realizations of \( H_k \) and \( \delta_k \), the NME (9) can be represented as
\[
P = A^T P_A A - A^T P_j \Psi_k (\Psi_k^T P_j \Psi_k + I)^{-1} \Psi_k^T P_A + Q \tag{53}
\]
where \( \Psi_k = \delta_k BH_k (\delta_k H_k^T M H_k + R)^{-\frac{1}{2}} \). Note that \( \Psi_k \in C^c \) and \( V_k^P \in C^{uc} \), which follows readily from (a.2), (a.3), (b.2), and (b.3).

We are trying to decompose \( V_k^P V_k^T \) in (54) into two components of \( \tilde{P}_c \) and \( \tilde{P}_c^{uc} \) according to the following two criteria:
1. **Equivalence:** \( V_k^P \in C^c \) and \( V_k^P \in C^{uc} \) is an element of the controllable PSD cone \( C^c \) and uncontrollable PSD cone \( C^{uc} \).
\[ C^\text{nc}, \text{ respectively, as defined in Definition 1. Namely, } \]
\[ V_k^T P_k^c V_k \in C^c \text{ and } V_k^T P_k^\text{nc} V_k \in C^\text{nc}. \]

We first choose
\[ P_k^c = \begin{bmatrix} (V_k P V_k^T)_{(\gamma_k)} & (V_k P V_k^T)_{(\gamma_k+1:S)} \\ \Sigma_k^{T} (V_k P V_k^T)_{(\gamma_k)} & \Sigma_k^{T} (V_k P V_k^T)_{(\gamma_k+1:S)} \end{bmatrix} \]
(55)

\[ \tilde{P}_k^c = (I - \Pi_k) V_k P V_k^T (I - \Pi_k - \begin{bmatrix} 0 & 0 \\ 0 & \Sigma_k^{T} (V_k P V_k^H)_{(\gamma_k)} \end{bmatrix} \cdot \Sigma_k) \cdot \]
(56)

We now verify that \( \tilde{P}_k^c \) in (55) and \( \tilde{P}_k^\text{nc} \) in (56) satisfy the “equivalence” criterion. Substitute \( \Sigma_k = (V_k P V_k^T)_{(\gamma_k)} \) into (55). It follows that
\[ P_k^c = \begin{bmatrix} (V_k P V_k^T)_{(\gamma_k)} & (V_k P V_k^T)_{(\gamma_k+1:S)} \\ \Sigma_k^{T} (V_k P V_k^T)_{(\gamma_k)} & \Sigma_k^{T} (V_k P V_k^T)_{(\gamma_k+1:S)} \end{bmatrix} \]
(57)

Combining (56) and (57), and canceling the diagonal block matrix \( \Sigma_k^{T} (V_k P V_k^T)_{(\gamma_k)} \cdot \Sigma_k \), we have \( \tilde{P}_k^c + \tilde{P}_k^\text{nc} = V_k P V_k^T \).

Therefore, the “equivalence” criterion is verified.

We now verify that \( V_k^T \tilde{P}_k^c V_k \in C^c \) and \( V_k^T \tilde{P}_k^\text{nc} V_k \in C^\text{nc} \).

Note that \( \tilde{P}_k^c \in S^S \) because
\[ \tilde{P}_k^c = \begin{bmatrix} (V_k P V_k^T)_{(\gamma_k)}^{T} & (V_k P V_k^T)_{(\gamma_k+1:S)}^{T} \\ \Sigma_k^{T} (V_k P V_k^T)_{(\gamma_k)}^{T} & \Sigma_k^{T} (V_k P V_k^T)_{(\gamma_k+1:S)}^{T} \end{bmatrix} \]
(58)

Further note that
\[ \ker \left( \delta_k B H_k \hat{H}_k^T B \left( V_k^T \tilde{P}_k^c V_k \right) \right) \]
\[ = \ker \left( \begin{bmatrix} (V_k P V_k^T)_{(\gamma_k)}^{T} & (V_k P V_k^T)_{(\gamma_k+1:S)}^{T} \\ \Sigma_k^{T} (V_k P V_k^T)_{(\gamma_k)}^{T} & \Sigma_k^{T} (V_k P V_k^T)_{(\gamma_k+1:S)}^{T} \end{bmatrix} V_k \right) = \ker \left( V_k^T \tilde{P}_k^c V_k \right). \]
(59)

Therefore, \( V_k^T \tilde{P}_k^c V_k \in C^c \) is verified.

Note that \( \tilde{P}_k^\text{nc} \in S^S \) because
\[ \tilde{P}_k^\text{nc} = \begin{bmatrix} -\Sigma_k^{T} & 0 \\ 0 & \Sigma_k \end{bmatrix} \]
(60)

Furthermore
\[ \delta_k B H_k (\delta_k H_k^T M H_k + R)^{-1} H_k^T B \left( V_k^T \tilde{P}_k^\text{nc} V_k \right) \]
\[ = V_k \begin{bmatrix} (A_k)_{(\gamma_k)} & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 \\ 0 & (V_k P V_k^T)_{(\gamma_k+1:S)}^{T} \end{bmatrix} \cdot \Sigma_k \]
\[ \cdot V_k = 0. \]
(61)

Therefore, we have
\[ \delta_k H_k^T B \left( V_k^T \tilde{P}_k^\text{nc} V_k \right) = \delta_k B H_k H_k^T B \left( V_k^T \tilde{P}_k^\text{nc} V_k \right) = 0. \]
(62)

As a result, \( V_k^T \tilde{P}_k^\text{nc} V_k \in C^\text{nc} \) is verified.

We can conclude that \( \tilde{P}_k^c \) can be decomposed into two parts as \( \tilde{P}_k^c = P_k^c + \tilde{P}_k^\text{nc} \) with \( P_k^c = V_k^T \tilde{P}_k^c V_k \) and \( \tilde{P}_k^\text{nc} = V_k^T \tilde{P}_k^\text{nc} V_k \), where \( P_k^c \in C^c \) and \( \tilde{P}_k^\text{nc} \in C^\text{nc} \). Therefore, Theorem 2 is proved.

E. Proof of Theorem 3

Note that \( g(P) = f(P) + P \). It suffices to prove there is a unique \( P^* \) such that \( P^* = g(P^*) \). For proof of the existence, we construct two matrix sequence \( \{P_k^c\}, k \geq 0 \) and \( \{P_k^\text{nc}\}, k \geq 0 \), which is monotonically increasing and bounded from above, and monotonically decreasing and bounded from below, respectively. We prove the existence of \( P^* \) using the monotone convergence theorem. We then prove the uniqueness of \( P^* \) by contradiction.

We first prove the existence of \( P^* \) that satisfies the NME (9) when the closed-loop control system is almost surely controllable, i.e., one of the three conditions (a.1), (a.2), and (a.3) in Lemma 2 is satisfied. Note that there is a \( P_0 = 0 \) such that \( P_1 = 0 < g(P_0) = Q \). Furthermore, for any given realization of \( H_k = H \), \( g(P[H_k = H]) \) can be represented as
\[ g(P[H_k = H]) = A^T (I - K \Psi) P (I - K \Psi)^T A + A^T K K^T A + Q \]
\[ \leq A^T (I - K \Psi) P (I - K \Psi)^T A + A^T K K^T A + Q \]
(63)

where \( \Psi = \Psi_k | H_k = H \), \( K = \Psi P \Psi^T P \Psi + I \) is a constant matrix such that \( A^T (I - K \Psi) \) is Hurwitz. It follows that there exists
\[ P_2 = 0 \]
\[ \geq \sum_{i=0}^{\infty} (A^T (I - K \Psi))^{i} (A^T K K^T A + Q) ((I - K \Psi)^T A) \]
(64)

such that \( P_2 \geq g(P_2 | H_k = H) > 0 \). Further note that
\[ \Psi_k \Psi_k^T = \delta_k B H_k (\delta_k H_k^T M H_k + R)^{-1} H_k^T B \]
\[ = B (M)^{-1} B - B \left( \delta_k M H_k R H_k^T M + M \right)^{-1} B^T. \]
(65)

As a result, if there exist \( H_k^{(1)} \) and \( H_k^{(2)} \) such that \( H_k^{(1)} R (H_k^{(1)})^T > H_k^{(2)} R (H_k^{(2)})^T \), we have \( \Psi_k^{(1)} (\Psi_k^{(1)})^T > \Psi_k^{(2)} (\Psi_k^{(2)})^T \) and \( g(P_2)[H_k = H_k^{(1)}] < g(P_2)[H_k = H_k^{(2)}] \). Therefore
\[ g(P_2) = \mathbb{E} \left[ g(P_2)[H_k R H_k^T \geq H] \right] \]
\[ + \mathbb{E} \left[ g(P_2)[H_k R H_k^T < H] \right] \]
\[ < \mathbb{E} \left[ H R H^T \right] \]
\[ < \mathbb{P}(P_2 | H_k R H_k^T \geq H) \]
\[ + \mathbb{P}(P_2 | H_k R H_k^T < H) \]
\[ \cdot \mathbb{P} \left( H_k R H_k^T < H \right). \]
(66)
By letting $H \to 0$, we have $\Pr(H_k R H_k^T \geq H R H^T) \to 1$. It follows that there exists $P^{(2)} \geq g(P^{(2)})$. We now construct two matrix sequences

$$
\begin{align*}
\{P^{(1)}_k : P^{(1)}_{k+1} = g(P^{(1)}_k), P^{(1)}_0 = P^{(1)}, k \geq 0\} \\
\{P^{(2)}_k : P^{(2)}_{k+1} = g(P^{(2)}_k), P^{(2)}_0 = P^{(2)}, k \geq 0\}.
\end{align*}
$$

(67)

Due to the monotonicity of $g(\cdot)$, it follows that $P^{(1)}_k \leq P^{(1)}_{k+1} \leq P^{(1)}_k$, $\forall k \geq 0$, and $P^{(2)}_k \leq P^{(2)}_{k+1} \leq P^{(2)}_k$, $\forall k \geq 0$. Therefore, we have

$$
P^{(1)}_k \leq P^{(1)}_k \leq P^{(2)}_k \leq P^{(2)}_k.
$$

(68)

The monotonically increasing sequence $\{P^{(1)}_k, k \geq 0\}$ is, therefore, bounded from above, i.e., $P^{(1)}_k \leq P^{(2)}_k$, $\forall k \geq 0$, it follows that the sequence $\{P^{(1)}_k, k \geq 0\}$ is convergent, i.e., there is a $(P^{(1)})^*$ such that

$$
\lim_{k \to \infty} P^{(1)}_k = (P^{(1)})^* = g \left( (P^{(1)})^* \right).
$$

(70)

Therefore, we prove the existence of $P^*$ that satisfies the NME (9) when one of the three conditions (a.1), (a.2), and (a.3) in Lemma 2 is satisfied.

We now prove the existence of $P^*$ such that $P^* = g(P^*)$ under the sufficient condition (17). We substitute $P^*_k$ (14) and $U^{(1)}_k$ (15) into the decomposed NME (16). The $P^*_k$ dependent terms in the decomposed NME (16) can be represented as

$$
A^T E[P^*_k] A = A^T E[V^T_k (I - \Pi_k) V_k P^T (I - \Pi_k) V_k] A
$$

$$
- A^T E[V^T_k \left[ 0_{V} \Sigma_k \sum_{\gamma_k \in \Sigma_k} \left( V_k P^H_k \right)^T \gamma_k \right] V_k] A.
$$

(71)

Moreover, let $A_k = \text{diag}((A_k)_{\gamma_k}, I_{(s_{\gamma_k})})$, it follows that

$$
\delta_k B_k (\delta_k H_k^T M H_k + R)^{-1} H_k^T B^T = V_k^T \tilde{A} \Pi_k \tilde{A} V_k.
$$

(72)

Let $(V_k P^T_k)_{\gamma_k} = \tilde{A}_k$ and $(A_k)_{\gamma_k} (V_k P^T_k)_{\gamma_k} (A_k)_{\gamma_k} = \tilde{A}_k$. The $P^*_k$ dependent terms in the decomposed NME (16), thus, can be represented as

$$
A^T E[P^*_k] A = A^T E[V^T_k \tilde{A}^{-1} \left( \tilde{A}_k V_k P^T_k V_k \tilde{A}_k \Pi_k + I \right)^{-1} \tilde{A}_k V_k P^T_k V_k \tilde{A}_k^{-1} V_k] A
$$

$$
= A^T E[V^T_k \left( I + \tilde{P}^{-1}_{\gamma_k} \right)^{-1} \left( I + \tilde{P}^{-1}_{\gamma_k} \right)^{-1} P_{\gamma_k} \Sigma_k]
$$

$$
\Sigma_k^T P_{\gamma_k} \left( I + \tilde{P}^{-1}_{\gamma_k} \right)^{-1} \Theta_k
$$

where

$$
\tilde{A}_k^{-1} V_k] A.
$$

(73)

Substituting (71) and (73) into (16), it follows that

$$
g(P) = Q
$$

$$
+ A^T E[V^T_k (I - \Pi_k) V_k P^T_k (I - \Pi_k) V_k] A + A^T E[V^T_k \tilde{A}^{-1} \left( \tilde{A}_k V_k P^T_k V_k \tilde{A}_k \Pi_k + I \right)^{-1} \tilde{A}_k V_k P^T_k V_k \tilde{A}_k^{-1} V_k] A
$$

$$
\cdot V_k] A.
$$

(75)

where

$$
\tilde{\Theta}_k = \Sigma_k^T P_{\gamma_k} \left( I + \tilde{P}^{-1}_{\gamma_k} \right)^{-1} \Sigma_k.
$$

(76)

Note that $(I + \tilde{P}^{-1}_{\gamma_k})^{-1} \leq I$, it follows that

$$
g(P) < Q + A^T E[V^T_k (I - \Pi_k) V_k P^T_k (I - \Pi_k) V_k] A
$$

$$
+ ||A||^2 E[Tr((A_k^{-1})) I].
$$

(77)

Therefore, under condition (17), there is $P^{(2)} = 0I$ with

$$
\theta \geq \frac{\|Q\| + ||A||^2 E[Tr((A_k^{-1}))]}{1 - E[A^T V_k (I - \Pi_k) V_k A]}.
$$

(78)

such that $g(P^{(2)}) \leq g(P^{(2)})^*$. Using similar techniques and constructing the two matrix sequences the same way as in (67) and (68), it follows that

$$
\lim_{k \to \infty} P^{(2)}_k = (P^{(2)})^* = g((P^{(2)})^*).
$$

(79)

Therefore, the existence of $P^*$ that satisfies the NME (9) under the sufficient condition (17) in Lemma 2 is proved.

In the following, we shall prove the uniqueness of $P^*$. Suppose there are $(P^{(1)})^*$ and $(P^{(2)})^*$ such that $(P^{(1)})^* = g((P^{(1)})^*)$ and $(P^{(2)})^* = g((P^{(2)})^*)$. Then, there is a positive constant $\phi^* \in (0, 1)$ such that $(P^{(1)})^* \geq \phi^*(P^{(2)})^*$ and $(P^{(1)})^* \not\geq \phi^*(P^{(2)})^*$ for all $\phi > \phi^*$. Note that

$$
g(\phi^*(P^{(2)})) = E[A(\Psi_k^T + (\phi^*(P^{(2)}))^{-1}) A^T] + Q
$$

$$
\geq E[\phi^*(A(\Psi_k^T + (\phi^*(P^{(2)}))^{-1}) A^T] + Q
$$

$$
\geq (1 + \phi^*)^* g(\phi^*(P^{(2)}))
$$

(80)

where $\phi = \frac{\phi^*}{\|g((P^{(2)})^*)\|}$ is a positive constant and $\sigma_Q$ is the minimum singular value of $Q$. Further note that

$$
(P^{(1)})^* \geq g((P^{(1)})^*) \geq g(\phi^*(P^{(2)}))
$$

$$
\geq (1 + \phi^*)^* g(\phi^*(P^{(2)})) = (1 + \phi^*)^* (P^{(2)})^*.
$$

(81)

This means that there is a $\phi = (1 + \phi^*)(\phi^* > \phi^*)$ such that $(P^{(1)})^* \geq \phi(\phi(\phi^*))$, which contradicts the fact that $(P^{(1)})^* \not\geq \phi^*(P^{(2)})^*$. As a result, the uniqueness of $P^*$ that satisfies $(P^*)^* = g((P^*)^*)$ is proved. Therefore, Theorem 3 is proved.
F. Proof of Lemma 3
Define $\mathbf{K}^{(1)}_k$ and $\mathbf{K}^{(2)}_k$ as $\mathbf{K}^{(1)}_k = P^{(1)}_k (\mathbf{Y}_k^T P^{(1)}_k \mathbf{Y}_k + I)^{-1}$ and $\mathbf{K}^{(2)}_k = P^{(2)}_k (\mathbf{Y}_k^T P^{(2)}_k \mathbf{Y}_k + I)^{-1}$, where $\mathbf{Y}_k = \delta_k \mathbf{H}_k (\delta_k \mathbf{H}_k^T \mathbf{M} \mathbf{H}_k + R)^{-\frac{1}{2}}$. Note that
\[
 f(P^{(1)}) \leq P^{(1)} + E[\mathbf{A}_k^T (I - \mathbf{K}_k^{(2)} \mathbf{Y}_k^T) P^{(1)}_k (I - \mathbf{K}_k^{(2)} \mathbf{Y}_k^T)^T \mathbf{A} + \mathbf{A}_k^T \mathbf{K}_k^{(2)} (\mathbf{K}_k^{(2)})^T \mathbf{A} + Q. \tag{82}
\]
Further note that
\[
 f(P^{(2)}) = P^{(2)} + E[\mathbf{A}_k^T (I - \mathbf{K}_k^{(2)} \mathbf{Y}_k^T) P^{(2)}_k (I - \mathbf{K}_k^{(2)} \mathbf{Y}_k^T)^T \mathbf{A} + \mathbf{A}_k^T \mathbf{K}_k^{(2)} (\mathbf{K}_k^{(2)})^T \mathbf{A} + Q. \tag{83}
\]
Subtracting (83) from (82), it follows that
\[
 f(P^{(1)}) - f(P^{(2)}) \leq P^{(1)} - P^{(2)} + E[\mathbf{A}_k^T (I - \mathbf{K}_k^{(2)} \mathbf{Y}_k^T) \mathbf{A}]. \tag{84}
\]
Taking the spectral norm on both sides of (84), and noting that matrix spectral norm is convex and $\|\mathbf{A}_k^T (I - \mathbf{K}_k^{(2)} \mathbf{Y}_k^T)\| \leq \|\mathbf{A}\|$, it follows that
\[
 \left\| f(P^{(1)}) - f(P^{(2)}) \right\| \leq \left\| P^{(1)} - P^{(2)} \right\| + E \left\| \mathbf{A}_k^T (I - \mathbf{K}_k^{(2)} \mathbf{Y}_k^T) \right\| \left\| P^{(1)} - P^{(2)} \right\|. \tag{85}
\]
Note that for any given realization of $\mathbf{P}_k$, $N_{k+1} = (\mathbf{f}(\mathbf{P}_{k+1}) - f(\mathbf{P}_{k+1}))$ is a function of $\{\delta_{k+1}, \mathbf{H}_{k+1}\}$. Moreover, due to the i.i.d. property of $\{\delta_{k+1}, \mathbf{H}_{k+1}\}$, it follows that $E \left[ \mathbf{f}(\mathbf{P}_{k+1}) - f(\mathbf{P}_{k+1}) \right] \mathbf{P}_k = 0$. Therefore,
\[
 E \left[ N_{k+1}^T \mathbf{F}_k \right] = 0, \forall k > 0.
\]
Note that
\[
 \left\| N_{k+1} \right\|^2 \leq E \left[ \left\| \mathbf{f}(\mathbf{P}_{k+1}) - \mathbf{Q} \right\|^2 + \left\| f(\mathbf{P}_{k+1}) - \mathbf{Q} \right\|^2 \right] \leq E \left[ \left\| \mathbf{A} \mathbf{Y}_k \mathbf{Y}_k^T (\mathbf{P}_k + \mathbf{Q})^{-1} \mathbf{A} \right\|^2 \mathbf{P}_k \right] \left\| \mathbf{P}_k \right\| \leq 2 \left\| \mathbf{A} \mathbf{P}_k \mathbf{A}^T \right\|^2 \leq 2 \left\| \mathbf{A} \right\|^2 \left\| \mathbf{P}_k \right\|^2. \tag{86}
\]
Therefore, Lemma 3 is proved.

G. Proof of Lemma 5
We show that the distance between $\mathbf{P}(t)$ and $\mathbf{P}(t)$ on the time interval $t \in [k\xi, (k + N)\xi]$ is $O(\xi)$ under the distance metric of the matrix spectral norm. Then, let $t$ tend to infinity along the trajectory of $\{k\xi\}$. It follows that the gap between the trajectory of $\mathbf{P}(t)$ and that of $\mathbf{P}(t)$ is $O(\xi)$, which can be made arbitrarily small by letting $\xi \to 0$.
Let $L = N\xi$ for some $N > 0$. For $t > 0$, denote $\left| t \right| = \max\{k\xi : n > 0, k\xi < t\}$. For $n \geq 0$ and $1 \leq l \leq L$, we have
\[
 \mathbf{P}(t_{k+l}) = \mathbf{P}(t_k) + \int_{t_k}^{t_{k+l}} f(\mathbf{P}(\left| t \right|)) dt \tag{87}
\]
Subtracting (88) from (87) and noting that
\[
 \left\| \int_{t_k}^{t_{k+l}} (f(\mathbf{P}(t)) - f(\mathbf{P}(\left| t \right|))) dt \right\| \leq \left\| \mathbf{A} \right\|^2 \xi \sum_{m=0}^{l-1} \sup_{j=m}^{m+l} \left| \mathbf{P}(t_{k+j}) - \mathbf{P}(t_{k+j}) \right| \tag{89}
\]
\[
 \left\| \int_{t_{k+l-1}}^{t_{k+l}} (f(\mathbf{P}(t)) - f(\mathbf{P}(\left| t \right|))) dt \right\| \leq c_2 \xi (1 + \mathbf{P}(t_k)) \tag{90}
\]
it follows that
\[
 \sup_{0 \leq j \leq l} \left| \mathbf{P}(t_{k+j}) - \mathbf{P}(t_{k+j}) \right| \leq c_2 \xi (1 + \mathbf{P}(t_k)) \tag{91}
\]
where $c_2$ is a constant. Since both $\sup_{t \leq l \leq t_1} \left| \mathbf{P}(t) - \mathbf{P}(t) \right|^2$ and $\sup_{t \leq l \leq t_1} \left| \mathbf{P}(t) - \mathbf{P}(t) \right|^2$ are $O(\xi)$, it follows that
\[
 \sup_{t \in [0, L]} \left| \mathbf{P}(l + t) - \mathbf{P}(l + t) \right| \leq c_3 \xi \tag{92}
\]
where $c_3$ is a constant. Therefore, Lemma 5 is proved.

H. Proof of Theorem 4
Note that the virtual fixed-point process $\{\mathbf{P}_k, k \geq 0\}$ in (24) corresponds to the fixed-point equation $\mathbf{P} = \tilde{g}(\mathbf{P})$ with $\tilde{g}(\mathbf{P}) = \mathbf{P} + \xi f(\mathbf{P})$. If the fixed-point equation $\mathbf{P} = \tilde{g}(\mathbf{P})$ has a solution $\mathbf{P}^\ast$, $\mathbf{P}^\ast$ must satisfy
\[
 \mathbf{P}^\ast = \tilde{g}(\mathbf{P}^\ast) = \mathbf{P}^\ast + \xi f(\mathbf{P}^\ast). \tag{93}
\]
Since $\xi \in (0, 1)$ is a positive constant, $(94)$ is equivalent to $f(\mathbf{P}) = 0$. Based on the definition of $f(\mathbf{P})$ in (18), it follows that $\mathbf{P}^\ast$ must satisfy the NME (9). Therefore, the existence and uniqueness of the solution to the fixed-point equation $\mathbf{P} = \tilde{g}(\mathbf{P})$ is equivalent to the existence and uniqueness of the solution to the NME (9). As a result, based on Theorem 3, which characterizes the existence of the solution to the NME (9), we can conclude that if one of the three conditions (a.1), (a.2), and (a.3) in Lemma 2 is satisfied, or the condition (17) in Theorem 3 is satisfied, the solution $\mathbf{P}^\ast$ to the fixed-point equation $\mathbf{P}^\ast = \tilde{g}(\mathbf{P}^\ast)$ exists and is unique.
such that $\bar{P}^{(2)} > \tilde{g}(\bar{P}^{(2)})$. We now construct the following two matrix sequences:

$$
\begin{align*}
\{\bar{P}_k^{(1)} : \bar{P}_{k+1}^{(1)} = \tilde{g}(\bar{P}_k^{(1)}), \bar{P}_0^{(1)} = 0, k \geq 0\} \\
\{\bar{P}_k^{(2)} : \bar{P}_{k+1}^{(2)} = \tilde{g}(\bar{P}_k^{(2)}), \bar{P}_0^{(2)} = \bar{P}^{(2)}, k \geq 0\}.
\end{align*}
$$

(95)

(96)

Let the initial condition of the fixed-point process be $0 \leq \bar{P}_0 \leq \bar{P}^{(2)}$, it follows that $\bar{P}_k^{(1)} \leq \bar{P}_k \leq \bar{P}_k^{(2)}$. Let $k \to \infty$ and note that $P^*$ exists and is unique. It follows that

$$
P^* = \lim_{k \to \infty} \bar{P}_k^{(1)} \leq \lim_{k \to \infty} \bar{P}_k \leq \lim_{k \to \infty} \bar{P}_k^{(2)} = P^*.
$$

(97)

Since $\bar{P}^{(2)}$ can be arbitrarily large, it follows that for any bounded initial value $\bar{P}_0$, the virtual fixed-point process $\{\bar{P}_k, k \geq 0\}$ in (24) converges to $P^*$. Based on Lemma 5, the limiting ODE (23), thus, has a unique equilibrium point $P^*$ that is globally asymptotically stable. Moreover, based on Lemma 4, it follows that the $P_k$ obtained by SA iteration (19) converges to $P^*$ almost surely. Based on the structural properties (10) in Theorem 1, it follows that $u_k$ converges to the optimal control action $u^*$ w.p. 1. Therefore, Theorem 4 is proved.

**I. Proof of Lemma 6**

The optimal control solution to infinite-time horizon Problem 3 with unknown CSI can be obtained by considering the finite-time horizon case first and then extending the results to the infinite-horizon case using the technique of reversing time index. Specifically, we can first start from the finite-time horizon problem with unknown CSI to minimize the quadratic cost function $J^* = E[x_k^TQx_k + \sum_{k=0}^{K-1} \tilde{h}(S_k, u_k)]$, and obtain the optimal control action

$$
u_k = -\delta_k \left( E[H_k^T B^T P_{k+1} B H_k + \delta_k H_k^T M H_k + R] \right)^{-1} H_k^T B^T P_{k+1} A x_k \quad \forall 0 \leq k < K
$$

(98)

where the expectation $E[\cdot]$ in (98) is taken w.r.t. the randomness of $H_k$, and $\{P_k, 0 \leq k \leq K\}$ is given recursively by $P_k = Q$ and

$$
P_k = A^T P_{k+1} A - \delta A^T P_{k+1} B H_k E[H_k^T B^T P_{k+1} B H_k] + H_k^T M H_k + R]^{-1} H_k^T B^T P_{k+1} A + Q.
$$

(99)

As a result, it suffices to show that $\bar{P}_i$ in the reversed time index, which corresponds to $P_k$ in (99), converges to the unique solution $P^*$ of the deterministic NME (30). Denote the matrix valued function $h(P)$ by

$$
h(P) = A^T P A - \delta A^T P B H_k E[H_k^T B^T P_{k+1} B H_k] + H_k^T M H_k + R]^{-1} H_k^T B^T P A + Q.
$$

(100)

Since there exist two positive constants $\gamma_1, \gamma_2$, such that $0 < \gamma_1 < 1 < \gamma_2$ and $h(\gamma_1 P^*) > \gamma_1 P^*$ and $h(\gamma_2 P^*) < \gamma_2 P^*$, based on the proof in Appendix B, it suffices to show that $h(P)$ is monotonically increasing w.r.t. $P$. Note that $h(P)$ can be represented as

$$
h(P) = (1-\delta) A^T P A + \delta A^T \left( \Psi(P) \Psi(P) + (P)^{-1} \right) A + Q
$$

(101)

where $\Psi$ is given by

$$
\Psi(P) = BH(\text{vec}^{-1}(\Phi\text{vec}(M)) + E[H_k^T B^T P B H_k] - H^T B^T P B H + R)^{-1/2} = BH(\text{vec}^{-1}(\Phi\text{vec}(M)) + \text{diag}\left(\sum_{i=1}^{N_r} \text{var}\left(\left(H_k^T\right)_i, \left(B^T P B\right)_i\right), \ldots, \sum_{i=1}^{N_r} \text{var}\left(\left(H_k^T\right)_i, \left(B^T P B\right)_i\right) + R\right)^{-1/2}.
$$

(102)

Note that $(B^T P B)_i, i < (B^T P B)_i, i, \forall 1 \leq i \leq N_r$, if $P^{(1)} < P^{(2)}$. Therefore, $\Psi(P^{(1)})^T \Psi(P^{(1)})^T > \Psi(P^{(2)})^T \Psi(P^{(2)})^T$ if $P^{(1)} < P^{(2)}$. As a result, $h(P)$ is monotonically increasing w.r.t. $P$. Therefore, Lemma 6 is proved.

**J. Proof of Theorem 5**

The sufficient condition in Theorem 5 is obtained by applying the proposed controllable and uncontrollable PSD cone decomposition in Theorem 2 to the deterministic NME (30).

Specifically, note that there is a $P^{(1)} = 0$ such that $P^{(1)} = 0 < h(P^{(1)}) = Q$. Furthermore, let $\text{rank}(BH) = \tilde{\gamma}$, let the SVD $\Psi(P) \Psi(P)^T = V(P)^T \Lambda(P) V(P)$ with the diagonal elements of $\Lambda(P) = \text{diag}(\tilde{\gamma}(P), \tilde{\gamma}(P), \ldots, \tilde{\gamma}(P), 0, \ldots, 0)$ in descending order, and let $\tilde{\Pi} = [0 \cdots 0 \cdots 0]_{S \times S}$. Applying the controllable and uncontrollable PSD cone decomposition in Theorem 2 to $h(P)$ in (101), it follows that

$$
h(P) < (1-\delta) A^T P A + \delta A^T \frac{V(P)^T}{\tilde{\Pi}} V(P) A + \|A\|^2 \sum_{i=1}^{\tilde{\gamma}} \tilde{\gamma}_i^{-1} (P) I + Q.
$$

(103)

Therefore, under condition (33), there is a $P^{(2)} = \tilde{\delta} I$ with $\tilde{\delta} \geq \|Q\| \left(1-\left(1-\delta\right)A^T P A + \delta A^T \tilde{\Pi} V(P) A + \|A\|^2 \sum_{i=1}^{\tilde{\gamma}} \tilde{\gamma}_i^{-1} (P) I + Q\right)^{-1}$ such that $h(P^{(2)}) < L(P^{(2)})$. Therefore, the existence of $P^*$ that satisfies the deterministic NME (30) can be proved by constructing two monotonic matrix sequences $\{P_k, k \geq 0\}$ and $\{P_k, k \geq 0\}$ the same way as in (67) and (68).

By noting that $\Psi(\phi^* P) \Psi(\phi^* P)^T < \psi_1^{-1} \Psi(P) \Psi(P)^T$, $\forall \phi^* \in (0, 1)$, the uniqueness of $P^*$ that satisfies the deterministic NME (30) can be obtained by contradiction in the same way as that in (80) and (81). Therefore, Theorem 5 is proved.

**REFERENCES**

[1] J. P. Hespanha, P. Naghshtabrizi, and Y. Xu, "A survey of recent results in networked control systems," Proc. IEEE, vol. 95, no. 1, pp. 138–162, Jan. 2007.

[2] I. S. Han and C. B. Chung, "Dynamic modeling and simulation of a fluidized catalytic cracking process Part I: Process modeling," Chem. Eng. Sci. vol. 56, pp. 1951–1990, Mar. 2001.

[3] Y. Cao, W. Yu, W. Ren, and G. Chen, "An overview of recent progress in the study of distributed multi-agent coordination," IEEE Trans. Ind. Informat., vol. 9, no. 1, pp. 427–438, Feb. 2013.

[4] R. A. Berry and R. G. Gallager, "Communication over fading channels with delay constraints," IEEE Trans. Inf. Theory, vol. 48, no. 5, pp. 1135–1149, May 2002.
[5] D. E. Quevedo, A. Ahlen, and K. H. Johansson, “State estimation over sensor networks with correlated wireless fading channels,” IEEE Trans. Autom. Control, vol. 58, no. 3, pp. 581–593, Mar. 2013.

[6] L. Schenato, “Optimal estimation in networked control systems subject to random delay and packet drop,” IEEE Trans. Autom. Control, vol. 53, no. 6, pp. 1311–1317, Jun. 2008.

[7] Z. Ma, Z. Yan, M. L. Shaltout, and D. Chen, “Optimal real-time control of wind turbine during partial load operation,” IEEE Trans. Control Syst. Technol., vol. 23, no. 6, pp. 2216–2226, Nov. 2015.

[8] K. Gatsis, A. Ribeiro, and G. J. Pappas, “Optimal power management in wireless control systems,” IEEE Trans. Autom. Control, vol. 59, no. 6, pp. 1495–1510, Jun. 2014.

[9] F. L. Lewis and G. Vamvoudakis, “Optimal adaptive control for unknown systems using output feedback by reinforcement learning methods,” in Proc. Conf. Control Autom., 2010, pp. 2138–2145.

[10] T. Bian and Z.-P. Jiang, “Value iteration and adaptive dynamic programming for data-driven adaptive optimal control design,” Automatica, vol. 71, pp. 348–360, Sep. 2016.

[11] B. Luo, Y. Yang, H.-N. Wu, and T. Huang, “Balancing value iteration and policy iteration for discrete-time control,” IEEE Trans. Syst., Man, Cybern., Syst., vol. 50, no. 11, pp. 3948–3958, Nov. 2020.

[12] A. Tanwani and O. Yufereva, “Error covariance bounds for suboptimal filters with Lipschitzian drift and Poisson-sampled measurements,” Automatica, vol. 122, 2020, Art. no. 109280.

[13] M. Ades, P. E. Caines, and R. P. Malhame, “Stochastic optimal control under Poisson-distributed observations,” IEEE Trans. Autom. Control, vol. 45, no. 1, pp. 3–13, Jan. 2000.

[14] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. Sastry, “Foundations of control and estimation over lossy networks,” Proc. IEEE, vol. 95, no. 1, pp. 163–187, Jan. 2007.

[15] O. C. Imer, S. Yuksel, and T. Basar, “Optimal control of dynamical systems over unreliable communication links,” Automatica, vol. 42, no. 9, pp. 1429–1440, Sep. 2006.

[16] E. Garone, B. Sinopoli, A. Goldsmith, and A. Casavola, “LQG control for MIMO systems over multiple erasure channels with perfect acknowledgment,” IEEE Trans. Autom. Control, vol. 58, no. 7, pp. 1835–1841, Jul. 2013.

[17] J. Fan, Q. Wu, Y. Jiang, T. Chai, and F. L. Lewis, “Model-free optimal output regulation for linear discrete-time lossy networked control systems,” IEEE Trans. Syst., Man, Cybern., Syst., vol. 50, no. 11, pp. 4033–4042, Nov. 2020.

[18] Y. Jiang, J. Fan, T. Chai, F. L. Lewis, and J. Li, “Tracking control for linear discrete-time networked control systems with unknown dynamics and dropout,” IEEE Trans. Neural Netw. Learn. Syst., vol. 39, no. 10, pp. 4607–4620, Oct. 2017.

[19] M. Athans, R. Ku, and S. B. Gershwin, “The uncertainty threshold principle, some fundamental limitations of optimal decision making under dynamic uncertainty,” IEEE Trans. Autom. Control, vol. AC-22, no. 3, pp. 491–495, Jun. 1977.

[20] R. Ku and M. Athans, “Further results on the uncertainty threshold principle,” IEEE Trans. Autom. Control, vol. AC-22, no. 5, pp. 491–495, Oct. 1977.

[21] A. S. Leong, S. Dey, and J. Anand, “Optimal LQG control over continuous fading channels,” IFAC Proc. Volumes, vol. 44, no. 1, pp. 6580–6585, Jan. 2011.

[22] N. Elia and J. N. Eisenbeis, “Limitation of linear control over packet drop networks,” in Proc. IEEE Conf. Decis. Control, 2004, pp. 5152–5157.

[23] N. Elia, “Remote stabilization over fading channels,” Syst. Control Lett., vol. 54, no. 3, pp. 237–249, 2005.

[24] A. S. Matveev and A. V. Savkin, “The problem of LQG optimal control via a limited capacity communication channel,” Syst. Control Lett., vol. 53, no. 1, pp. 51–64, 2004.

[25] S. Fang, J. Chen, and H. Ishii, “Intrinsic limits of power reduction in MIMO networked control systems,” in Proc. IEEE Conf. Decis. Control, 2017, pp. 4759–4764.

[26] G. Ranade and A. Sahai, “Control capacity,” IEEE Trans. Inf. Theory, vol. 65, no. 1, pp. 235–254, Jan. 2018.

[27] N. Xiao, L. Xie, and L. Qiu, “Feedback stabilization of discrete-time networked systems over fading channels,” IEEE Trans. Autom. Control, vol. 57, no. 9, pp. 2176–2189, Sep. 2012.

[28] D. Tse and P. Viswanath, Fundamentals of Wireless Communications. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[29] D. Gesbert, H. Bolcskei, D. A. Gore, and A. J. Paulraj, “Outdoor MIMO wireless channels: Models and performance prediction,” IEEE Trans. Commun., vol. 50, no. 12, pp. 1926–1934, Dec. 2002.

[30] A. Hedayat, H. Shah, and A. Nosratinia, “Analysis of space-time coding in correlated fading channels,” IEEE Trans. Wireless Commun., vol. 4, no. 6, pp. 2882–2891, Nov. 2005.

[31] H. Lin, H. Su, P. Shi, Z. Shu, and Z. G. Wu, Estimation and Control for Networked Systems With Packet Losses Without Acknowledgement. Berlin, Germany: Springer, 2017.

[32] Y. S. Borkar, Stochastic Approximation: A Dynamical Systems Viewpoint. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[33] J. Abouabdali, D. Bertsekas, and V. S. Borkar, “Learning algorithms for Markov decision processes with average cost,” SIAM J. Control Optim., vol. 40, no. 3, pp. 681–698, 2001.

[34] S. Wu, X. Ren, Q. Jia, K. H. Johansson, and L. Shi, “Learning optimal scheduling policy for remote state estimation under uncertain channel condition,” IEEE Trans. Control Netw. Syst., vol. 7, no. 2, pp. 579–591, Jun. 2020.

[35] F. Lewis and V. S. Borkar, Optimal Control, 3rd ed. New York, NY, USA: Wiley, 2012.

[36] F. L. Lewis and D. Vrabie, “Reinforcement learning and adaptive dynamic programming for feedback control,” IEEE Circuits Syst. Mag., vol. 9, no. 3, pp. 32–50, Jul./Sep. 2009.

[37] M. Stojnic, W. Xu, and B. Hassibi, “Compressed sensing–probabilistic analysis of a null-space characterization,” in Proc. Int. Conf. Acoust. Speech Signal Process., 2008, pp. 3377–3380.

[38] D. Tylavsky and G. Sohie, “Generalization of the matrix inversion lemma,” Proc. IEEE, vol. AC-74, no. 7, pp. 1050–1052, Jul. 1986.

Dr. Cai was the recipient of the Hong Kong Ph.D. Fellowship (HKPF) in 2013.

Vincent K. N. Lau (Fellow, IEEE) received the B.Eng (Distinction first Hons.) degree in electrical and electronic engineering from the University of Hong Kong, Hong Kong, in 1995. He was with Bell Labs from 1997 to 2004 and the Department of ECE, Hong Kong University of Science and Technology (HKUST), Hong Kong, in 2004. He is currently a Chair Professor and the Founding Director of Huawei-HKUST Joint Innovation Laboratory, HKUST. His current research focus includes wireless communications for 5G systems, content-centric wireless networking, wireless networking for mission-critical control, and cloud-assisted autonomous systems.