# Asymptotic Normality of Cumulative Cost in Linear Quadratic Regulators

Borna Sayedana, Peter E. Caines, Aditya Mahajan

***Abstract*—The central limit theorem is a fundamental result in probability theory that characterizes the distribution of deviation from the mean in the law of large numbers. Similar distributional behavior emerges in other frameworks such as maximum likelihood estimation, least squares estimation, and stochastic approximation. In this paper, we establish a central limit theorem for the cumulative per-step cost incurred by the optimal policy in linear quadratic regulators using first principles. Our proof technique relies on a decomposition of cumulative cost using a completion of square argument, properties of the noise sequence with even density, and a central limit theorem for martingale difference sequences.**

## I. Introduction

### A. Motivation

The Central Limit Theorem (CLT), is one of the most important results in probability theory and mathematical statistics. It establishes that the distribution of deviation from the mean in the law of large numbers asymptotically converges to a normal distribution. Similar asymptotic normality for the deviations emerges in other processes as well. For example, in the parameter estimation framework, the asymptotic normality is established for maximum likelihood estimation (see e.g. [1]–[3]). In regression models, asymptotic normality is established for various estimation and prediction methods (see e.g. [4]–[9], for a list of such results, see [10]). This property is also established in the stochastic approximation framework (see e.g. [11], [12]). The importance of asymptotic normality results become evident when they are used to derive confidence bounds for different frameworks.

In the systems and controls literature, there are various characterization of the law of large numbers (e.g. [13]–[19]) but the distribution of the deviation from the mean is less explored. There are some results on CLT for Markov cost/reward process (e.g. [16]–[19]) which are derived using advanced tools in Markov chain theory including weighted geometric ergodicity and weighted uniform ergodicity. These results imply a CLT for the LQR setting (i.e., systems with linear dynamics and quadratic cost). In this paper, we revisit the distribution of the deviation from the mean for LQR setting and establish asymptotic normality using an elementary proof based on first principles. Our result is different from the existing characterizations in the literature and uses different and much simpler proof techniques.

The sample path behavior of the cumulative cost has recently also been studied in the context of regret analysis for adaptive controllers. These analyses are either in the Bayesian framework (e.g., in [20], [21]) or in terms of high probability guarantees for the frequentist regret (e.g., in [22]–[29]) or almost sure guarantees for the frequentist regret (e.g., in [30]–[32]). However, these bounds are not not sharp enough to characterize the distribution of the cumulative cost.

### B. Contributions

Our main contribution is to establish asymptotic normality of the cumulative cost in the LQR framework using an elementary argument. Under a mild technical assumption on the noise distribution, we show the cumulative cost incurred by the optimal policy converges weakly to a Gaussian distribution. Our analysis uses a completion of square argument to decompose the cumulative cost to bounded terms plus a Martingale Difference Sequence (MDS). The convergence argument follows from this decomposition, properties of the noise sequence with even density, and a version of the CLT for MDS.

### C. Organization

The rest of the paper is organized as follows. In Sec. II, we present the system model, assumptions, and the main results. In Sec. III, we present preliminary results on the cost decomposition, implications of our assumption on the noise process, a preliminary on the CLT for MDS, and the proof of the main result. Our concluding remarks are presented in Sec. IV.

### D. Notation

Given a vector $v$, $v(i)$ denotes its $i$-th component. Given a matrix $A$, $A_{i,j}$ denotes its $(i,j)$-th element and $\lambda_{\max}(A)$ denotes the largest magnitudes of right eigenvalues. For a square matrix $Q$, $\mathrm{Tr}(Q)$ denotes the trace. For a vector $x$, $\|x\|$ denotes the Euclidean norm. $\mathbf{0}$ denotes the zero-vector in the appropriate Euclidean space. For a matrix $A$, $\|A\|$ denotes the spectral norm. If $Q$ is symmetric, $Q \succeq 0$ and $Q \succ 0$ denote that $Q$ is positive semi-definite and positive definite, respectively. Given a sequence of random variables $\{x_t\}_{t\geq 0}$, $x_{0:t}$ is a short hand for $(x_0, \dots, x_t)$ and $\sigma(x_{0:t})$ denotes the sigma field generated by random variables $x_{0:t}$. Convergence in almost sure sense is abbreviated by $a.s$. Convergence in distribution is showed by the notation $\xrightarrow{(d)}$. Notation $\mathcal{N}(0,1)$ denotes a standard Gaussian distribution. $\mathbb{R}$ and $\mathbb{N}$ denote the sets of real and natural numbers and $\mathbb{R}_+$ denotes the set of positive real numbers. Given a sequence of positive numbers $\{a_t\}_{t\geq 0}$, $a_T \asymp T$ means that $\limsup_{T\to\infty} a_T/T < \infty$, and $\liminf_{T\to\infty} a_T/T > 0$.

## II. Problem Formulation and Main Result

### A. System Model

Consider a discrete-time linear time-invariant system with full state observation. Let $x_t \in \mathbb{R}^n$ and $u_t \in \mathbb{R}^d$ denote the state and control input at time $t$. The system starts at a known initial state $x_0$ and it evolves according to the following dynamics:

$$x_{t+1} = Ax_t + Bu_t + Dv_{t+1}, \quad t \geq 0, \tag{1}$$

where $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times d}$, and $D \in \mathbb{R}^{n\times n}$ are the system dynamic matrices and $\{v_t\}_{t\geq 1}$, $v_{t+1} \in \mathbb{R}^n$, is an independent and identically distributed (i.i.d.) zero-mean noise process with unit covariance $I$. At each time $t$, the system incurs a per-step cost of

$$c(x_t, u_t) = x_t^\intercal Q x_t + u_t^\intercal R u_t,$$

where $Q \succeq 0$ and $R \succ 0$.

We assume that the control inputs are chosen according to a time-homogeneous (and measurable) policy $\pi\colon \mathbb{R}^n \to \mathbb{R}^d$, i.e.,

$$u_t = \pi(x_t).$$

Let $\Pi$ denote the set of all such policies. For a fixed policy $\pi \in \Pi$, let $\{x_t^\pi\}_{t\geq 0}$ and $\{u_t^\pi\}_{t\geq 0}$ denote the sequence of states and control inputs generated over time. Let

$$\mathcal{C}(\pi, T) \coloneqq \sum_{t=0}^{T-1} c(x_t^\pi, u_t^\pi),$$

denote the cumulative cost incurred by policy $\pi$ up to time $T$. Note that our definition of $\mathcal{C}(\pi, T)$ does not include an expectation, so $\mathcal{C}(\pi, T)$ is a random variable.

The long-term average performance of policy $\pi \in \Pi$ is given by

$$J(\pi) \coloneqq \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}[\mathcal{C}(\pi, T)],$$

where the expectation is with respect to the noise process $\{v_t\}_{t\geq 1}$. Let

$$J^* = \inf_{\pi\in\Pi} J(\pi),$$

denote the optimal performance. A policy $\pi^* \in \Pi$ is called optimal if $J(\pi^*) = J^*$.

We impose the following standard assumption on the model.

**Assumption 1.** *The pair of matrices $(A, B)$ is controllable, and the pair of matrices $(A, Q^{1/2})$ is observable.*

It is well known (e.g., see [10]) that under Assumption 1, the optimal policy exists, is unique, and is given by

$$\pi^*(x_t) = -L^* x_t, \tag{2}$$

where the optimal gain $L^*$ is given by

$$L^* = (R + B^\intercal S B)^{-1} B^\intercal S A, \tag{3}$$

where $S$ is the unique fixed point of the Discrete Algebraic Riccati Equation (DARE) given by:

$$P = A^\intercal P A - A^\intercal P B(R + B^\intercal P B)^{-1} B^\intercal P A + Q. \tag{4}$$

Moreover the optimal value $J^*$ is given by:

$$J^* = \mathrm{Tr}(SDD^\intercal). \tag{5}$$

### B. Main Result

The classical result described above characterizes the behavior of the expected value of $\mathcal{C}(\pi^*, T)$; in particular,

$$\lim_{T\to\infty} \frac{1}{T}\mathbb{E}[\mathcal{C}(\pi^*, T)] = \mathrm{Tr}(SDD^\intercal) = J^*. \tag{6}$$

Our main result characterizes a much stronger *sample path* behavior of $\mathcal{C}(\pi^*, T)$. In particular, we will show that under a mild assumption, loosely speaking, the stochastic process $\mathcal{C}(\pi^*, T)$ converges in distribution to a Gaussian random variable. We will present this statement more precisely in this section.

For our analysis, we impose the following additional assumption on the noise process $\{v_t\}_{t\geq 1}$.

**Assumption 2.** *In addition to being i.i.d. across time and having a unit covariance, the noise sequence $\{v_t\}_{t\geq 1}$ satisfies the following conditions for each time $t$:*

- **(A1)** *The components of $v_t$ are independent and admit a density $f_v$ that is even.*
- **(A2)** *$v_t$ is uniformly bounded, that is, there exists a $K_v \in \mathbb{R}_+$ such that $\|v_t\| \leq K_v$ almost surely.*
- **(A3)** *For matrices $D$ and $S$, we have $\mathrm{Var}(v_t^\intercal D^\intercal S D v_t) \neq 0$.*

For the ease of notation, let $\{(x_t^*, u_t^*)\}_{t\geq 0}$ denote the (stochastic) trajectory $\{(x_t^{\pi^*}, u_t^{\pi^*})\}_{t\geq 0}$ of the optimal policy, $w_t = Dv_t$ denote the noise at time $t$, and $A^* = A - BL^*$ denote the closed loop dynamics under the optimal policy. Define:

$$M \coloneqq \mathbb{E}[w_t^\intercal S w_t w_t^\intercal S w_t] - \left(\mathbb{E}[w_t^\intercal S w_t]\right)^2$$

which is a scalar constant. We now define a process $\{N_T\}_{T\geq 1}$ where:

$$N_T \coloneqq \sum_{t=0}^{T-1} \Big[M + 4(A^* x_t^*)^\intercal SDD^\intercal SA^* x_t^*\Big]$$

and let $\{\nu_T\}_{T\geq 1}$ be a stopping time corresponding to $\{N_T\}_{T\geq 1}$ given by

$$\nu_T \coloneqq \min_{\tau\geq 1}\Big\{\tau; \sum_{t=1}^{\tau} N_t \geq T\Big\}. \tag{7}$$

Our main result is the following theorem.

**Theorem 1.** *We have that*

$$\frac{\mathcal{C}(\pi^*, \nu_T) - \nu_T J^*}{\sqrt{T}} \xrightarrow{(d)} \mathcal{N}(0, 1) \text{ as } T \to \infty.$$

The proof is presented in Sec. III.

Above theorem is presented in terms of the stopping time in Eq. (7). In the following lemma, we establish the growth rate of this stopping time in the almost sure sense.

**Lemma 1.** *The stopping time $\{\nu_T\}_{T\geq 1}$ satisfies:*

$$\nu_T \asymp T, \quad a.s.$$

The proof is presented in App. A.

Theorem 1 and Lemma 1 together give a complete picture of distributional behavior of $\mathcal{C}(\pi^*, \nu_T)$, which in the order, matches with the asymptotic normality results in other frameworks.

## III. Proof of Theorem 1

In this section we present the proof of Theorem 1. Our proof relies on three techniques: (i) a completion of square argument to establish a decomposition of the cumulative cost, similar to one used in [33]; (ii) some implications of noise having an even density; and (iii) the CLT for bounded martingale difference sequences [34].

### A. Decomposition of Cumulative Cost

The following lemma provides a decomposition of the cumulative cost of any arbitrary policy $\pi$.

**Lemma 2.** *For any $\pi \in \Pi$, we have*

$$\begin{aligned}\mathcal{C}(\pi, T) = {} & x_0^\intercal S x_0 - (x_T^\pi)^\intercal S x_T^\pi \\ & + \sum_{t=0}^{T-1} \left[(u_t^\pi + L^* x_t^\pi)^\intercal (R + B^\intercal S B)(u_t^\pi + L^* x_t^\pi)\right. \\ & + \sum_{t=0}^{T-1} \left[2(A x_t^\pi + B u_t^\pi)^\intercal S w_{t+1} + w_{t+1}^\intercal S w_{t+1}\right],\end{aligned}$$

*where matrices $L^*$ and $S$ are given by (3) and (4).*

The proof is similar to the decomposition of $\mathbb{E}[\mathcal{C}(\pi, T)]$ presented in [33] and is presented in App. B for completeness.

In the following Lemma, we use Lemma 2 to characterize the cumulative cost function induced by the optimal policy $\mathcal{C}(\pi^*, T)$.

**Lemma 3.** *For the optimal policy $\pi^*$, we have*

$$\begin{aligned}\mathcal{C}(\pi^*, T) = {} & x_0^\intercal S x_0 - (x_T^*)^\intercal S x_T^* \\ & + \sum_{t=0}^{T-1} \left[2(A^* x_t^*)^\intercal S w_{t+1} + w_{t+1}^\intercal S w_{t+1}\right].\end{aligned}$$

*Proof.* The result follows by substituting $u_t^* = -L^* x_t^*$ in Lemma 2, and substituting $x_t^{\pi^*}$ with $x_t^*$. ∎

### B. Implications of the Assumption on the Noise

The assumed symmetry on the components of $v_t$ (i.e., the components of $v_t$ admitting a density $f_v$ that is even) has important implications in our analysis. We show this structure implies that certain cubic transformation of the noise has zero mean. Following lemma summarizes these structures.

**Lemma 4.** *Under Assumption 2, we have the following for any time $t$:*

1) *For any odd $k \in \mathbb{N}$ and any component $i \in \{1, \dots, n\}$, $\mathbb{E}[v_t(i)^k] = 0$.*
2) *For any $i, j \in \{1, \dots, n\}$, $i \neq j$, $\mathbb{E}[v_t(i) v_t(j)^2] = 0$.*
3) *For any arbitrary matrix $M$, let $y_t = M v_t$, then $\mathbb{E}[y_t y_t^\intercal y_t] = \mathbf{0}$.*

Proof is presented in App. C.

Furthermore, the boundedness assumption on the noise sequence $\{v_t\}_{t\geq 1}$ implies the boundedness of optimal state trajectory $\{x_t^*\}_{t\geq 0}$. This is presented in the following lemma.

**Lemma 5.** *Under Assumption 2, there exists a universal constant $K_x \in \mathbb{R}_+$ (which depends only on $K_v$) such that*

$$\|x_t^*\| \leq K_x, \quad a.s., \quad \forall t \geq 0.$$

This is a classic result and its proof exists in many resources. We included a proof in App. D for completeness.

### C. CLT for Martingale Difference Sequences

The usual CLT for martingale difference sequences is the Lindeberg-Levy CLT for triangular array of martingale difference sequences. In our analysis, we use an implication of Lindeberg-Levy CLT stated in [34]. Since this version of the CLT is not as well known, we restate it below for completeness.

Let $\{\delta_t\}_{t\geq 1}$, $\delta_t \in \mathbb{R}$, be a martingale difference sequence adapted to some filtration sequence $\{\mathcal{G}_t\}_{t\geq 0}$, i.e.:

$$\mathbb{E}[\delta_t | \mathcal{G}_{t-1}] = 0.$$

In addition, for all $t \geq 1$, let $\Delta_t := \sum_{\tau=1}^t \delta_\tau$ denote the martingale process corresponding to $\{\delta_t\}_{t\geq 1}$. Let $\rho_t^2 := \mathbb{E}[\delta_t^2 | \mathcal{G}_{t-1}]$ denote the conditional variance of $\delta_t$. For any $T \geq 0$, define the stopping time $\mu_T$ as:

$$\mu_T = \min_{\tau \geq 1} \left\{\tau; \sum_{t=1}^{\tau} \rho_t^2 \geq T\right\}.$$

The following theorem states a version of central limit theorem for the martingale sequence $\{\Delta_t\}_{t\geq 1}$.

**Theorem 2** (see [34, Theorem 35.11])**.** *Suppose the martingale difference sequence $\{\delta_t\}_{t\geq 1}$ satisfies the following conditions:*

**(C1)** *For all $t \geq 1$, $|\delta_t|$ is uniformly bounded, i.e., there exists a $K_\delta \in \mathbb{R}_+$, such that:*

$$|\delta_t| \leq K_\delta, \quad a.s.$$

**(C2)** *We have:*

$$\sum_{t=1}^{\infty} \mathbb{E}[\delta_t^2 | \mathcal{G}_{t-1}] = \infty.$$

*Then we have:*

$$\frac{\Delta_{\mu_T}}{\sqrt{T}} \xrightarrow{(d)} \mathcal{N}(0, 1) \text{ as } T \to \infty.$$

In the subsequent subsection, we show some of the terms in the cumulative cost $\mathcal{C}(\pi^*, T)$ satisfy martingale difference property, we then use Theorem 2 to derive the distribution of the cumulative cost.

## D. Preliminary Results

Define the filtration to be the sigma field generated by the sequence of states and control actions, i.e., $\mathcal{F}_t := \sigma(x^*_{0:t}, u^*_{0:t})$. Using Lemma 3 and the fact that $J^* = \mathbb{E}[w_{t+1}^\intercal S w_{t+1}]$, we rewrite $\mathcal{C}(\pi^*, T) - TJ^*$ as following:

$$\mathcal{C}(\pi^*, T) - TJ^* = x_0^\intercal S x_0 - (x_T^*)^\intercal S x_T^* + \sum_{t=0}^{T-1} \Big[ 2(A^* x_t^*)^\intercal w_{t+1} + w_{t+1}^\intercal S w_{t+1} - \mathbb{E}[w_{t+1}^\intercal S w_{t+1}] \Big].$$

To simplify the algebra, we define following intermediate variables for $t \geq 0$.

$$a_{t+1} := w_{t+1}^\intercal S w_{t+1}, \tag{8}$$
$$b_{t+1} := 2(A^* x_t^*)^\intercal S w_{t+1}, \tag{9}$$
$$c_{t+1} := \mathbb{E}[w_{t+1}^\intercal S w_{t+1}], \tag{10}$$
$$z_{t+1} := a_{t+1} + b_{t+1} - c_{t+1}. \tag{11}$$

As a result of above reparametrization, we have:

$$\mathcal{C}(\pi^*, T) - TJ^* = \sum_{t=0}^{T-1} z_{t+1} + (x_0)^\intercal S(x_0) - (x_T^*)^\intercal S(x_T^*).$$

We show that the sequence $\{z_t\}_{t\geq 1}$ is a martingale difference sequence satisfying conditions (C1) and (C2) in Theorem 2. We first establish the properties of variables $a_{t+1}, b_{t+1}$, and $c_{t+1}$ in the following proposition.

**Proposition 1.** *For all $t \geq 0$, we have:*

**(P1)** $\mathbb{E}[b_{t+1} | \mathcal{F}_t] = 0$.
**(P2)** $\mathbb{E}[a_{t+1} | \mathcal{F}_t] = c_{t+1}$.
**(P3)** $\mathbb{E}[a_{t+1}^2 | \mathcal{F}_t] = \mathbb{E}[a_{t+1}^2]$.
**(P4)** $\mathbb{E}[c_{t+1} a_{t+1} | \mathcal{F}_t] = c_{t+1}^2$.
**(P5)** $\mathbb{E}[c_{t+1} b_{t+1} | \mathcal{F}_t] = 0$.
**(P6)** $\mathbb{E}[a_{t+1} b_{t+1} | \mathcal{F}_t] = 0$.

*Proof.* These properties are the consequences of the assumption on the noise process.

**(P1)** Follows by the fact that $x_t^*$ is $\mathcal{F}_t$-measurable and based on Assumption 2, $w_{t+1} = D v_{t+1}$ is zero mean and independent of $\mathcal{F}_t$.

**(P2)** Follows from independence of $w_{t+1}$ from $\mathcal{F}_t$, and the definition of $c_{t+1}$.

**(P3)** Follows from independence of $w_{t+1}$ from $\mathcal{F}_t$.

**(P4)** Follows from following equations:

$$\mathbb{E}[c_{t+1} a_{t+1} | \mathcal{F}_t] \overset{(a)}{=} c_{t+1} \mathbb{E}[a_{t+1} | \mathcal{F}_t] \overset{(b)}{=} c_{t+1}^2,$$

where $(a)$ follows from the fact that $c_{t+1}$ is not a random variable and $(b)$ follows from Property (P2).

**(P5)** Follows from following equations:

$$\mathbb{E}[c_{t+1} b_{t+1} | \mathcal{F}_t] \overset{(c)}{=} c_{t+1} \mathbb{E}[b_{t+1} | \mathcal{F}_t] \overset{(d)}{=} 0,$$

where $(c)$ follows from the fact that $c_{t+1}$ is not a random variable and $(d)$ follows from Property (P1).

**(P6)** Follows from Lemma 4. To show this, let:

$$y_{t+1} := S^{1/2} D v_{t+1} = S^{1/2} w_{t+1}$$

we have:

$$\begin{aligned}
&\mathbb{E}[a_{t+1} b_{t+1} | \mathcal{F}_t] \\
&\overset{(e)}{=} \mathbb{E}[2(x_t^*)^\intercal (A^*)^\intercal S^{1/2} S^{1/2} w_{t+1} w_{t+1}^\intercal S^{1/2} S^{1/2} w_{t+1} | \mathcal{F}_t] \\
&\overset{(f)}{=} 2(x_t^*)^\intercal (A^*)^\intercal S^{1/2} \mathbb{E}[y_t y_t^\intercal y_t] \overset{(g)}{=} 0,
\end{aligned}$$

where $(e)$ follows from the fact that $S \succ 0$, $(f)$ follows from the fact that $S^{1/2}$ is symmetric, and $(g)$ follows from Lemma 4 part (3).

□

## E. Proof of Theorem 1

To prove the theorem, we first verify the conditions of Theorem 2 for the sequence $\{z_t\}_{t\geq 1}$. First, recall that by definition, $z_{t+1} = a_{t+1} + b_{t+1} - c_{t+1}$. We have:

$$\mathbb{E}[z_{t+1} | \mathcal{F}_t] = \mathbb{E}[a_{t+1} - c_{t+1} | \mathcal{F}_t] + \mathbb{E}[b_{t+1} | \mathcal{F}_t] \overset{(a)}{=} 0,$$

where $(a)$ follows from Properties (P1) and (P2) in Proposition 1. We now verify conditions (C1) and (C2) in Theorem 2.

*1) Verifying (C1):* We know $a_{t+1}$ and $c_{t+1}$ are uniformly bounded by (A2) in Assumption 2. By Lemma 5 and (A2) in Assumption 2, we know $|b_{t+1}|$ is uniformly bounded. As a result, $|z_{t+1}|$ is uniformly bounded almost surely.

*2) Verifying (C2):* We compute the conditional expectation of $z_{t+1}^2$ given the filtration $\mathcal{F}_t$ as following:

$$\begin{aligned}
&\mathbb{E}[z_{t+1}^2 | \mathcal{F}_t] = \mathbb{E}[(a_{t+1} + b_{t+1} - c_{t+1})^2 | \mathcal{F}_t] \\
&= \mathbb{E}[a_{t+1}^2 | \mathcal{F}_t] + \mathbb{E}[b_{t+1}^2 | \mathcal{F}_t] + \mathbb{E}[c_{t+1}^2 | \mathcal{F}_t] \\
&+ 2\mathbb{E}[a_{t+1} b_{t+1} | \mathcal{F}_t] - 2\mathbb{E}[c_{t+1} a_{t+1} | \mathcal{F}_t] - 2\mathbb{E}[c_{t+1} b_{t+1} | \mathcal{F}_t] \\
&\overset{(b)}{=} \mathbb{E}[a_{t+1}^2 | \mathcal{F}_t] + \mathbb{E}[b_{t+1}^2 | \mathcal{F}_t] + \mathbb{E}[c_{t+1}^2 | \mathcal{F}_t] - 2\mathbb{E}[a_{t+1} c_{t+1} | \mathcal{F}_t] \\
&\overset{(c)}{=} \mathbb{E}[a_{t+1}^2] - c_{t+1}^2 + \mathbb{E}[b_{t+1}^2 | \mathcal{F}_t]
\end{aligned} \tag{12}$$

where $(b)$ follows from properties (P5) and (P6) in Proposition 1 and $(c)$ follows from properties (P3) and (P4). Now the term $\mathbb{E}[a_{t+1}^2] - c_{t+1}^2$ is independent of $t$ and depends only on the density $f_v$; therefore, by Jensen's inequality and (A3) in Assumption 2, we know that there exists an $\underline{\epsilon} > 0$, such that:

$$\mathbb{E}[a_{t+1}^2] - c_{t+1}^2 > \underline{\epsilon}, \tag{13}$$

for all $t \geq 0$. By definition we know $\mathbb{E}[b_{t+1}^2 | \mathcal{F}_t] \geq 0$ for all $t \geq 0$. As a result, we have:

$$\sum_{t=0}^{T-1} z_{t+1} \geq T\underline{\epsilon}.$$

Implying that: $\lim_{T\to\infty} \sum_{t=0}^{T-1} \mathbb{E}[z_{t+1}^2 | \mathcal{F}_t] = \infty$, almost surely, verifying the condition (C2).

*3) Concluding the proof:* Since the conditions (C1) and (C2) hold for the sequence $\{z_t\}_{t\geq 1}$, by Theorem 2, we have:

$$\frac{\sum_{t=1}^{\nu_T} z_t}{\sqrt{T}} \xrightarrow{(d)} \mathcal{N}(0, 1).$$

By Lemma 5, we know $(x_T^*)^\intercal S(x_T^*)$ is almost surely bounded for all $T \geq 0$. Moreover $x_0^T S x_0$ is a constant. Therefore, we have:

$$\lim_{T\to\infty} \frac{x_0^\intercal S x_0 - (x_T^*)^\intercal S x_T^*}{\sqrt{T}} \to 0, \quad a.s.$$

As a result, by using Slutsky's Theorem (see [35, Theorem 7.7.3]), we get:

$$\frac{\mathcal{C}(\nu_T, \pi^*) - \nu_T J^*}{\sqrt{T}} \xrightarrow{(d)} \mathcal{N}(0,1).$$

**Remark 1.** In the proof of Theorem 1, each of the two sequences $\{a_{t+1} - c_{t+1}\}_{t\geq 0}$ and $\{b_{t+1}\}_{t\geq 0}$ is a martingale difference sequence. However, these two sequences are dependent, and therefore, the fact that each of them converges in distribution does not trivially imply that their summation also converges in distribution. As a result, applying Theorem 2 on each of these sequences individually would not imply the desired result. Therefore, characterizing the behavior of the sequence $\{a_{t+1} + b_{t+1} - c_{t+1}\}_{t\geq 0}$ similar to the approach in our proof is necessary.

## IV. Conclusion

In this paper we have established the asymptotic normality of the cumulative cost in the LQR framework. We have shown that under mild assumptions on the noise process, asymptotic normality holds for the distribution of the cumulative cost only using first principles. Our result gives a complete description of the cost distribution induced by the optimal policy. We believe this analysis opens new doors to understanding the distributional behavior of the cumulative cost and may pave the way to derive confidence bounds for this framework. These confidence bounds can be used in risk-averse or distributional reinforcement learning within this setup. A natural extension of this work is to derive similar results for larger classes of policies or to weaken the assumption on the noise sequence to be Gaussian or sub-Gaussian.

## References


[1] R. A. Fisher, "On the mathematical foundations of theoretical statistics," *Philosophical transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character*, vol. 222, no. 594-604, pp. 309–368, 1922.

[2] H. Cramér, *Mathematical methods of statistics*. Princeton university press, 1999, vol. 26.

[3] P. J. Huber *et al.*, "The behavior of maximum likelihood estimates under nonstandard conditions," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, no. 1. Berkeley, CA: University of California Press, 1967, pp. 221–233.

[4] F. Eicker, "Asymptotic normality and consistency of the least squares estimators for families of linear regressions," *The annals of mathematical statistics*, vol. 34, no. 2, pp. 447–456, 1963.

[5] T. L. Lai and C. Z. Wei, "Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems," *Ann. Statist.*, vol. 10, no. 1, pp. 154–166, 1982.

[6] L. Lennart and P. E. Caines, "Asymptotic normality of prediction error estimators for approximate system models," *Stochastics*, vol. 3, no. 1-4, pp. 29–46, 1980.

[7] L. Ljung and P. E. Caines, "Asymptotic normality of prediction error estimators for approximate system models," in *1978 IEEE Conference on Decision and Control including the 17th Symposium on Adaptive Processes*, 1978, pp. 927–932.

[8] T. W. Anderson and N. Kunitomo, "Asymptotic distributions of regression and autoregression coefficients with martingale difference disturbances," *Journal of Multivariate Analysis*, vol. 40, no. 2, pp. 221–243, 1992.

[9] S. Borovkova, H. P. Lopuhaä, and B. N. Ruchjana, "Consistency and asymptotic normality of least squares estimators in generalized star models," *Statistica Neerlandica*, vol. 62, no. 4, pp. 482–508, 2008.

[10] P. E. Caines, *Linear stochastic systems*. SIAM, 2018.

[11] V. Fabian, "On asymptotic normality in stochastic approximation," *The Annals of Mathematical Statistics*, pp. 1327–1332, 1968.

[12] J. Sacks, "Asymptotic distribution of stochastic approximation procedures," *The Annals of Mathematical Statistics*, vol. 29, no. 2, pp. 373–405, 1958.

[13] P. H. Algoet, "The strong law of large numbers for sequential decisions under uncertainty," *IEEE Transactions on Information Theory*, vol. 40, no. 3, pp. 609–633, 1994.

[14] R. van Handel, "Ergodicity, decisions, and partial information," *Séminaire de Probabilités XLVI*, pp. 411–459, 2014.

[15] B. Hajek, "Ergodic process selection," in *Open Problems in Communication and Computation*. Springer, 1987, pp. 199–203.

[16] S. P. Meyn and R. L. Tweedie, *Markov chains and stochastic stability*. Springer Science & Business Media, 2012.

[17] O. Hernández-Lerma and J. B. Lasserre, *Further topics on discrete-time Markov control processes*. Springer Science & Business Media, 2012, vol. 42.

[18] M. Duflo, *Random iterative models.* Berlin-Heidelberg: Springer, 1997.

[19] N. Maigret, "Théorème de limite centrale fonctionnel pour une chaîne de markov récurrente au sens de harris et positive," in *Annales de l'institut Henri Poincaré. Section B. Calcul des probabilités et statistiques*, vol. 14, no. 4, 1978, pp. 425–440.

[20] Y. Ouyang, M. Gagrani, and R. Jain, "Learning-based control of unknown linear systems with thompson sampling," *arXiv preprint arXiv:1709.04047*, 2017.

[21] B. Sayedana, M. Afshari, P. E. Caines, and A. Mahajan, "Thompson-sampling based reinforcement learning for networked control of unknown linear systems," in *2022 IEEE 61st Conference on Decision and Control (CDC)*, 2022, pp. 723–730.

[22] Y. Abbasi-Yadkori and C. Szepesvári, "Regret bounds for the adaptive control of linear quadratic systems," in *Proceedings of the 24th Annual Conference on Learning Theory*, 2011, pp. 1–26.

[23] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu, "Regret bounds for robust adaptive control of the linear quadratic regulator," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[24] A. Cohen, T. Koren, and Y. Mansour, "Learning linear-quadratic regulators efficiently with only $\sqrt(T)$ regret," in *International Conference on Machine Learning*. PMLR, 2019, pp. 1300–1309.

[25] H. Mania, S. Tu, and B. Recht, "Certainty equivalence is efficient for linear quadratic control," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[26] M. Simchowitz and D. Foster, "Naive exploration is optimal for online lqr," in *International Conference on Machine Learning*. PMLR, 2020, pp. 8937–8948.

[27] M. K. S. Faradonbeh, A. Tewari, and G. Michailidis, "Optimism-based adaptive regulation of linear-quadratic systems," *IEEE Trans. Autom. Control*, vol. 66, no. 4, pp. 1802–1808, 2020.

[28] A. Cassel, A. Cohen, and T. Koren, "Logarithmic regret for learning linear quadratic regulators efficiently," in *International Conference on Machine Learning*. PMLR, 2020, pp. 1328–1337.

[29] F. Wang and L. Janson, "Exact asymptotics for linear quadratic adaptive control," *The Journal of Machine Learning Research*, vol. 22, no. 1, pp. 12 136–12 247, 2021.

[30] M. K. S. Faradonbeh, A. Tewari, and G. Michailidis, "On adaptive linear–quadratic regulators," *Automatica*, vol. 117, p. 108982, 2020.

[31] Y. Lu and Y. Mo, "Almost surely $\sqrt{T}$ regret bound for adaptive LQR," *arXiv preprint arXiv:2301.05537*, 2023.

[32] B. Sayedana, M. Afshari, P. E. Caines, and A. Mahajan, "Relative almost sure regret bounds for certainty equivalence control of Markov jump systems," in *IEEE Conference on Decision and Control (CDC)*, 2023, pp. 6629–6634.

[33] K. J. Åström, *Introduction to Stochastic Control Theory*. Dover, 1970.

[34] P. Billingsley, *Probability and measure*. John Wiley & Sons, 2017.

[35] R. B. Ash, B. Robert, C. A. Doleans-Dade, and A. Catherine, *Probability and measure theory*. Academic press, 2000.

[36] B. Sayedana, M. Afshari, P. E. Caines, and A. Mahajan, "Strong consistency and rate of convergence of switched least squares system identification for autonomous Markov jump linear systems," *IEEE Transactions on Automatic Control*, pp. 1–8, 2024.

[37] K. J. Åström, *Introduction to stochastic control theory*. Courier Corporation, 2012.

## Appendix A
## Proof of Lemma 1

Using Eq. (12), we have:

$$\mathbb{E}[z_{t+1}^2|\mathcal{F}_t] = \mathbb{E}[a_{t+1}^2] - c_{t+1}^2 + \mathbb{E}[b_{t+1}^2|\mathcal{F}_t].$$

By (A3) in Assumption 2 and Jensen's inequality, we know there exists a $\underline{\epsilon} > 0$ such that $\mathbb{E}[a_{t+1}^2] - c_{t+1}^2 > \underline{\epsilon}$. Since $\mathbb{E}[b_{t+1}^2|\mathcal{F}_t] > 0$, we have:

$$\liminf_{T\to\infty} \frac{N_T}{T} = \liminf_{T\to\infty} \frac{\sum_{t=0}^{T-1} \mathbb{E}[z_{t+1}^2|\mathcal{F}_t]}{T} \geq \underline{\epsilon} > 0, \quad a.s.$$

From the definition of $b_{t+1}$, it is clear that there exists a constant $C \in \mathbb{R}_+$ such that $\mathbb{E}[b_{t+1}^2|\mathcal{F}_t] \leq C\|x_t\|^2$ for all $t \geq 0$. As a result, by following arguments similar to [36, Lemma 5], we have:

$$\limsup_{T\to\infty} \frac{\sum_{t=0}^{T-1} \mathbb{E}[b_{t+1}^2|\mathcal{F}_t]}{T} < \infty, \quad a.s.$$

Since the term $\mathbb{E}[a_{t+1}^2] - c_{t+1}^2$ is independent of $t$ and only depends on the density $f_v$, there exists an $\bar{\epsilon} > 0$, such that:

$$\mathbb{E}[a_{t+1}^2] - c_{t+1}^2 < \bar{\epsilon}.$$

As a result,

$$\limsup_{T\to\infty} \frac{N_T}{T} = \limsup_{T\to\infty} \frac{\sum_{t=0}^{T-1} \mathbb{E}[b_{t+1}^2|\mathcal{F}_t]}{T} + \bar{\epsilon} < \infty,$$

almost surely, implying that $N_T \asymp \mathcal{O}(T)$ and therefore $\nu_T \asymp \mathcal{O}(T)$, almost surely.

## Appendix B
## Proof of Lemma 2

### A. Preliminary Result

The proof of this lemma is similar to the regret decomposition in [32] . Following algebraic lemma is adapted from [37, Lemma 6.1].

**Lemma 6.** *We have following statements:*

1) *(Algebraic completion of square) For $x \in \mathbb{R}^n$ and $u \in \mathbb{R}^d$ and matrices $A, B, S, R$ with appropriate dimensions, we have*

$$\begin{aligned} & u^\intercal Ru + (Ax+Bu)^\intercal P(Ax+Bu) + x^\intercal Qx \\ &= [u+L(P,R,A,B)x]^\intercal [R+B^\intercal PB][u+L(P,R,A,B)x] \\ &+ x^\intercal K(P,A,B,R,Q)x, \end{aligned} \tag{14}$$

*with $L(P,R,A,B) := -[R+B^\intercal PB]^{-1}B^\intercal PA$, and $K(P,A,B,R,Q)$ is defined as:*

$$Q + A^\intercal PA - A^\intercal PB(R+B^\intercal PB)^{-1}B^\intercal PA.$$

2) *The Discrete Algebraic Riccati Equation (DARE) in Eq. (4), i.e. $K(P,A,B,R,Q) = P$ has a unique positive definite fixed point solution $S \succeq 0$. As a result, we have:*

$$\begin{aligned} & u^\intercal Ru + (Ax+Bu)^\intercal S(Ax+Bu) + x^\intercal Qx \\ &= [u+L(S,R,A,B)x]^\intercal [R+B^\intercal SB][u+L(S,R,A,B)x] \\ &+ x^\intercal Sx \end{aligned}$$

### B. Proof of Lemma 2

*Proof.* The proof follows by applying Lemma 6. We start by adding and subtracting the term $(x_T^\pi)^\intercal S(x_T^\pi)$ to the expression. Recall that $\{x_t^\pi\}_{t\geq 0}$ and $\{u_t^\pi\}_{t\geq 0}$ denote the sequence of state and actions induced by the policy $\pi$. We have:

$$\begin{aligned} \mathcal{C}(\pi,T) &= \sum_{t=0}^{T-1} \left[(x_t^\pi)^\intercal Q(x_t^\pi) + (u_t^\pi)^\intercal R(u_t^\pi)\right] \\ &\quad + (x_T^\pi)^\intercal S(x_T^\pi) - (x_T^\pi)^\intercal S(x_T^\pi) \\ &= \sum_{t=0}^{T-2} \left[(x_t^\pi)^\intercal Q(x_t^\pi) + (u_t^\pi)^\intercal R(u_t^\pi)\right] - (x_T^\pi)^\intercal S x_T^\pi \\ &+ \left[(x_{T-1}^\pi)^\intercal Q(x_{T-1}^\pi) + (u_{T-1}^\pi)^\intercal R(u_{T-1}^\pi) + (x_T^\pi)^\intercal S(x_T^\pi)\right] \\ &= \Big[\sum_{t=0}^{T-2} (x_t^\pi)^\intercal Q(x_t^\pi) + (u_t^\pi)^\intercal R(u_t^\pi)\Big] - (x_T^\pi)^\intercal S(x_T^\pi) \\ &+ (x_{T-1}^\pi)^\intercal Q(x_{T-1}^\pi) + (u_{T-1}^\pi)^\intercal R(u_{T-1}^\pi) \\ &+ (Ax_{T-1}^\pi + Bu_{T-1}^\pi + w_T)^\intercal S(Ax_{T-1}^\pi + Bu_{T-1}^\pi + w_T) \\ &\overset{(a)}{=} \Big[\sum_{t=0}^{T-2} (x_t^\pi)^\intercal Q(x_t^\pi) + (u_t^\pi)^\intercal R(u_t^\pi)\Big] \\ &+ (x_{T-1}^\pi)^\intercal S(x_{T-1}^\pi) - (x_T^\pi)^\intercal S(x_T^\pi) \\ &+ \Big[(u_{T-1}^\pi + L^* x_{T-1}^\pi)^\intercal (R + B^\intercal SB)(u_{T-1}^\pi + L^* x_{T-1}^\pi) \\ &\quad + w_T^\intercal S w_T + 2(Ax_{T-1}^\pi + Bu_{T-1}^\pi)^\intercal S w_T\Big], \end{aligned}$$

where $(a)$ follows from Lemma 6, with $L^*$ being the RHS of Eq. (3). By repeating the same argument, we get:

$$\begin{aligned} \mathcal{C}(\pi,T) &= x_0^\intercal S x_0 - x_T^\intercal S x_T \\ &+ \sum_{t=1}^{T-1} \Big[(u_t^\pi + L^* x_t^\pi)^\intercal (R + B^\intercal SB)(u_t^\pi + L^* x_t^\pi) \\ &\quad + 2(Ax_t^\pi + Bu_t^\pi)^\intercal S w_{t+1} + w_{t+1}^\intercal S w_{t+1}\Big]. \end{aligned}$$ □

## Appendix C
## Proof of Lemma 4

For an odd $n$, Assumption 2, implies that for all $1 \leq i \leq n$ and for all $t \geq 0$, we have:

$$\mathbb{E}[v_t(i)^k] = \int_{-K_v}^{K_v} v^k f_v(v) dv.$$

1) Proof of part (1): The PDF $f_v$ is an even function and for odd $k \in \mathbb{N}$, $v^k$ is an odd function . As a result, $v^k f_v$ is an odd function, and integrating an odd function from $-K_v$ to $K_v$ is 0.

2) Proof of part (2): For all $i \neq j$, we have:

$$\mathbb{E}[v_t(i)v_t(j)^2] \overset{(a)}{=} \mathbb{E}[v_t(i)]\mathbb{E}[v_t(j)^2] \overset{(b)}{=} 0,$$

where $(a)$ follows from the independence of the components of $v_t$, and $(b)$ follows from part (1) of this lemma.

3) Proof of part (3): Let $m_{ij}$ denote the $(i,j)$-th component of $M$. Then Recall that we have

$$y_t(i) = [Mv_t](i) = \sum_{j=1}^n m_{ij} v_t(j).$$

It is clear that $\mathbb{E}[y_t(i)] = 0$ for all $t \geq 0$ by the linearity of the expectation operator. We show that for all $i \in \{1, \dots, n\}$ and all $t \geq 0$, we have: $\mathbb{E}[y_t(i)^3] = 0$. By multinomial theorem, we have:

$$\mathbb{E}\Big[y_t(i)^3\Big] = \mathbb{E}\Bigg[\Big(\sum_{j=1}^{n} m_{ij} v_t(j)\Big)^3\Bigg]$$
$$= \mathbb{E}\Bigg[\sum_{k_1+\dots+k_n=3} \binom{3}{k_1, \dots, k_n} (m_{i1} v_t(1))^{k_1} \dots (m_{in} v_t(n))^{k_n}\Bigg].$$

Where the notation $\sum_{k_1+\dots+k_n=3}$ denotes all possible tuples $(k_1, \dots, k_n)$ such that $k_1 + \dots + k_n = 3$. Let the tuple $(k'_1, \dots, k'_n)$ be a decreasing permutation of $(k_1, \dots, k_n)$, i.e.,

$$k'_1 \geq k'_2 \geq \dots \geq k'_n.$$

Since $k_1 + \dots + k_n = 3$, there are only 3 choices for the tuple $(k'_1, \dots, k'_n)$. These choices are $(3, 0, \dots, 0)$ or $(2, 1, \dots, 0)$ or $(1, 1, 1, 0, \dots, 0)$. By Parts (1) and (2), we get:

1) For any $i \in \{1, \dots, n\}$, $\mathbb{E}[v_t(i)^3] = 0$.
2) For any $i, j \in \{1, \dots, n\}$, $i \neq j$, $\mathbb{E}[v_t(i)^2 v_t(j)] = 0$.
3) For any $i, j, k \in \{1, \dots, n\}$, $i \neq j \neq k$, $\mathbb{E}[v_t(i) v_t(j) v_t(k)] = 0$.

This implies that all the permutations which are mapped to the tuples $(3, 0, \cdots, 0)$ or $(2, 1, \cdots, 0)$ or $(1, 1, 1, 0, \cdots, 0)$ have zero expected value; therefore, $\mathbb{E}[y_t(i)^3] = 0$. Next we show for all $i, j \in \{1, \dots, n\}$ such that $i \neq j$, we have: $\mathbb{E}[y_t(i)^2 y_t(j)] = 0$. By using the multinomial theorem, we have:

$$\mathbb{E}\Big[y_t(i)^2\Big] = \mathbb{E}\Bigg[\Big(\sum_{j=1}^{n} m_{ij} v_t(j)\Big)^2\Bigg]$$
$$= \mathbb{E}\Bigg[\sum_{k_1+\dots+k_n=2} \binom{2}{k_1, \dots, k_n} (m_{i1} v_t(1))^{k_1} \dots (m_{in} v_t(n))^{k_n}\Bigg].$$

Again let the tuple $(k'_1, \dots, k'_n)$ be a decreasing permutation of $(k_1, \dots, k_n)$. Since $k_1 + \dots + k_n = 2$, there are only 2 choices for the tuple $(k'_1, \dots, k'_n)$. These choices are $(2, 0, \dots, 0)$ or $(1, 1, 0, \dots, 0)$. Now since $y_t(j) = \sum_{k=1}^{n} m_{jk} v_t(k)$, expanding $y_t(i)^2 y_t(j)$ and ordering the permutations we again end up with 3 choices for $(k'_1, \dots, k'_n)$, i.e., $(3, 0, \dots, 0)$ , $(2, 1, \dots, 0)$ , and $(1, 1, 1, 0, \dots, 0)$. By repeating the arguments similar to the previous part, we have that $\mathbb{E}[y(i)^2 y(j)] = 0$. At last, since

$$\mathbb{E}[y y^\intercal y] = \begin{bmatrix} y(1) \\ \vdots \\ y(n) \end{bmatrix} \Big(y(1)^2 + \cdots y(n)^2\Big). \tag{15}$$

All the terms are either of the form $\mathbb{E}[y(i)^3]$ or $\mathbb{E}[y(i)^2 y(j)]$, $i \neq j$, implying that:

$$\mathbb{E}[y y^\intercal y] = \mathbf{0}.$$

## Appendix D
## Proof of Lemma 5

Given that $\|v_t\| \leq K_v$, we have that $\|w_t\| \leq \|D\| \|v_t\| =: K_w$. Let $\rho_{\max} = \lambda_{\max}(A^*) < 1$ (recall $A^* = A - BL^*$) since $L^*$ is a stabilizing controller gain. Pick an $\varepsilon > 0$ such that $\rho_{\max} + \varepsilon < 1$. Then, by Gelfand's formula, we know that there exists a $T_0$ such that for all $t > T_0$, $\|(A^*)^t\| < \rho_{\max} + \varepsilon$. By the convolutional form of the output, we have that for $T > T_0$,

$$\begin{aligned}
\|x_T\| &= \|(A^*)^T x_0\| + \Big\|\sum_{\tau=1}^{T} (A^*)^\tau w_{T-\tau}\Big\| \\
&\leq \|(A^*)^T\| \|x_0\| + \sum_{\tau=1}^{T} \|(A^*)^\tau\| \|w_{T-\tau}\| \\
&\leq \|(A^*)^T\| \|x_0\| + K_w \sum_{\tau=1}^{T} \|(A^*)^\tau\| \\
&\leq (\rho_{\max} + \varepsilon)^T \|x_0\| + K_w \sum_{\tau=1}^{T} (\rho_{\max} + \varepsilon)^\tau \\
&\overset{(a)}{\leq} (\rho_{\max} + \varepsilon)^{T_0} \|x_0\| + \frac{K_w}{1 - (\rho_{\max} + \varepsilon)} =: K_x
\end{aligned}$$

where $(a)$ uses the fact that $\rho_{\max} + \varepsilon < 1$.