WHEN MULTIPLICATIVE NOISE STYMIES CONTROL

BY JIAN DING[*], YUVAL PERES[†], GIREEJA RANADE[†] AND ALEX ZHAI[‡]

University of Pennsylvania[*], Microsoft Research AI[†] and Stanford University[‡]

We consider the stabilization of an unstable discrete-time linear system that is observed over a channel corrupted by continuous multiplicative noise. Our main result shows that if the system growth is large enough, then the system cannot be stabilized. This is done by showing that the probability that the state magnitude remains bounded must go to zero with time. Our proof technique recursively bounds the conditional density of the system state to bound the progress the controller can make. This sidesteps the difficulty encountered in using the standard data-rate theorem style approach; that approach does not work because the mutual information per round between the system state and the observation is potentially unbounded.

It was known that a system with multiplicative observation noise can be stabilized using a simple memoryless linear strategy if the system growth is suitably bounded. The second main result in this paper shows that while memory cannot improve the performance of a linear scheme, a simple nonlinear scheme that uses one-step memory can do better than the best linear scheme.

1. Introduction. We consider the control and stabilization of a system observed over a multiplicative noise channel. Specifically, we analyze the following system, $\mathcal{S}_a$, with initial state $X_0 \sim \mathcal{N}(0, 1)$:

$$X_{n+1} = a \cdot X_n - U_n,$$
$$Y_n = Z_n \cdot X_n.$$

(1.1)

In the preceding formulation, the system state is represented by X_n at time n, and the control U_n can be any function of the current and previous observations Y_0 to Y_n. The Z_n's are i.i.d. random variables with a known continuous distribution. The realization of the noise Z_n is unknown to the controller, much like the fading coefficient (gain) of a channel might be unknown to the transmitter or receiver in noncoherent communication. The constant a captures the growth of the system. The controller's objective is to stabilize the system in the second-moment sense, that is, to ensure that $\sup_n \mathbb{E}[|X_n|^2] < \infty$. We would like to understand the largest growth factor a for which the system can be stabilized for a given distribution on Z_n. Figure 1 represents a block diagram for this system.

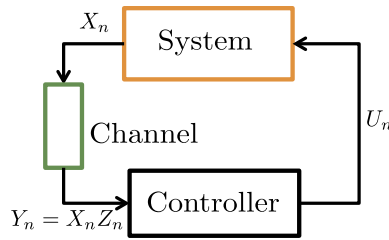Our main theorem provides an impossibility result for stabilizing the system $\mathcal{S}_a$.

FIG. 1.   *The state X_n is observed over a multiplicative noise channel $Y_n = X_n Z_n$.*

THEOREM 1.1.   *Let the Z_n be i.i.d. random variables with finite mean and variance and with bounded density $f_Z(z) = e^{-\phi(z)}$, where $\phi(\cdot)$ is a polynomial of even degree with positive leading coefficient. Then there exists $a \in \mathbb{R}, a < \infty$ such that $|X_n|$ in (1.1) satisfies $\mathbb{P}(|X_n| < M) \to 0$ for all $M < \infty$.*

Theorem 5.1 generalizes this result to a larger class of distributions for Z_n. Note that the conditions on $\phi(\cdot)$ in Theorem 1.1 are satisfied by $Z_n \sim \mathcal{N}(1, \sigma^2)$.

We also discuss a few sufficient conditions for second-moment stability of the system in this paper. When Z_n has mean 1 and variance σ^2, we observe that that a system growth of $a^* = \sqrt{1 + \frac{1}{\sigma^2}}$ can be stabilized in the second-moment sense using a memoryless *linear* strategy (Proposition 3.1). Further, we show that the best *linear* strategy to control the system $\mathcal{S}_a$ in (1.1) is the memoryless linear strategy (Theorem 3.2). Our second main result (Theorem 4.1) shows that a *nonlinear* controller can improve on the performance of the best linear strategy. We state this here for the case where $Z_n \sim \mathcal{N}(1, 1)$.

THEOREM 1.2.   *Let $Z_n \sim \mathcal{N}(1, 1)$. Then the system $\mathcal{S}_a$ in (1.1) with $a \leq \sqrt{2}$ can be stabilized in the second-moment sense by a linear control strategy. Further, there exists $a > \sqrt{2}$ for which a nonlinear controller can stabilize the system in a second-moment sense.*

In particular, there exists a nonlinear strategy with memory that can stabilize $\mathcal{S}_a$ in a second-moment sense with $a = \sqrt{2} + 1.6 \times 10^{-3}$.

We further believe that nonlinear schemes without memory cannot stabilize the system for $1 < a < a^*$, and some evidence in this direction is provided in Theorem 4.5. Finally, in the case where the Z_n have mean zero, a linear strategy cannot stabilize the system in the second-moment sense for any growth factor $a > 1$ (Theorem 3.3), but a nonlinear scheme with memory can stabilize it for some value of the growth factor $a > 1$ (Theorem 4.3).

1.1. *Model motivation.*   Multiplicative noise on the observation channel can model the effects of a fast-fading communication channel (rapidly changing channel gain), as well as the impact of sampling and quantization errors [18, 28].

We illustrate below how synchronization or sampling errors can lead to multiplicative noise, following a discussion from [28]. Consider the nearly trivial continuous-time system,

$$\dot{X}(t) = a \cdot X(t),$$

which is sampled at regular intervals of t_0. The difference equation corresponding to the state at the nth time step is given by $X_{n+1} = e^{at_0} \cdot X_n$. However, in the presence of synchronization error the nth sample, Y_n, might be collected at time $nt_0 + \Delta$ instead of precisely at nt_0. Then

$$Y_n = e^{a(nt_0+\Delta)}X(0) = Z_n X_n,$$

where Z_n is a continuous random variable, since the jitter Δ is a continuous random variable.

1.2. *Proof approach.* We introduce a new converse approach in the proof of Thorem 5.1; instead of focusing on the second moment, our proof bounds the density of the state, and thus shows the instability of any moment of the state. We believe these techniques are a primary contribution of the work.

A key element of the proof is that a "genie" observes the state of the system and provides a quantized version of the logarithm of the state to the controller at each time as extra side-information in addition to the multiplicative noise observation. This side-information bounds the state in intervals of size 2^{-k} (with k increasing as time increases). We know from results on noncoherent communication [14] and carry-free models [24] that only the order of magnitude of the message can be recovered from a transmission with multiplicative noise. As a result, this side-information does not effectively provide much extra information, but it allows us to quantify the rate at which the controller may make progress.

1.3. *Related work.* Our paper builds on many previous ideas in information theory and control that have been discussed in depth in books such as [8, 17, 42]. These works can roughly be divided into two categories, (1) control over data-rate limits or additive noise and SNR limits, and (2) control over multiplicative noise constraints.

1.3.1. *Control with additive noise or data-rate constraints on the observation channel.* Our problem formulation is certainly motivated by the data-rate theorems as studied in [12, 20–22, 36, 38]. These data-rate theorems ask the question: how does limited communication affect our ability to control a system? The results tell us that a noiseless observation data rate $R > \log|a|$ is necessary and sufficient to stabilize a system in the second-moment sense under a few different considerations on the additive noise in the system. Matveev and Savkin [17] have further provided data-rate style theorems for continuous time systems and for different notions of stability such as stability in probability. (Our impossibility result considers

the notion of stability in probability.) Other work has also generalized the data-rate theorems for systems with additive noise to the case when the channel-rate is given by a Markov process, for example, You and Xie [41] and Minero et al. [19].

Building on the data-rate theorems for noiseless channels, there has been a large body of work that considers control over additive noise channels. Sahai and Mitter [31] introduced the notion of anytime capacity as the informational threshold to stabilize a plant over a noisy channel. Nonlinear controllers and encoders/decoders was essential to achieving their results. On the other hand, works such as Braslavsky et al. [3], Silva et al. [33] and Silva and Pulgar [34], also explore control with additive noise and SNR constraints, and focus on understanding what can be achieved using linear controllers, and under what conditions linear controllers might be optimal or close to optimal. Our work focuses on multiplicative noise channels, and shows that nonlinear control strictly outperforms the optimal linear controller for control with multiplicative observation noise.

Other related work includes Freudenberg et al. [9] and Bao et al. [2], which both focus on the finite-horizon setting for control over additive noise channels. These setups are different from our problem in that they allow for explicit encoders and decoders, focus on the impact of additive noise and use channel-state information.

We note that our observation channel model, $Y_n = Z_n X_n$, does not provide an explicit rate limit on the number of bits that can be transmitted across it. Since X_n can be arbitrarily large and Z_n has a fixed mean and variance, an arbitrarily large number of bits could be communicated across the channel by increasing the magnitude of X_n (Lapidoth and Moser [14]). Since a rate-limit cannot be established on our observation channel, the proof approach suggested by the data-rate theorems does not work in our case. Furthermore, our model aims to capture the impact of physical unreliabilities and does not include encoders and decoders, which are also essential to using a data-rate theorem style approach.

1.3.2. *Control with multiplicative noise.* In addition to the data-rate theorems, our problem formulation is inspired by the intermittent Kalman filtering (IKF) problem [25, 35]. The IKF problem considers the estimation of a system over a channel that drops packets, that is, a channel with Bernoulli multiplicative noise. The system cannot use encoding strategies to code over the packet drops. Our paper builds on this setup to consider a general continuous multiplicative noise on the observation.

Chiuso et al. [4] and Dey et al. [5] both extend the work on packet dropping observation channels to consider the further impact of quantization, partial channel feedback and delays. Schenato et al. [32] and Hespanha et al. [11] provide detailed surveys of control over packet dropping networks—for packet drops on both the observation channel as well as the control channel. A key result comes from Elia [7], where the author studies the impact of multiplicative noise on the control action using a robust control perspective. In contrast to Elia [7], our work considers multiplicative noise on the observation instead of the control action and

takes a stochastic control approach. Other related works by Garone et al. [10] and Matveev et al. [16] both consider control actions that are subject to packet drops, but allow the controller and system to use an encoder-decoder pair to treat this limitation as a traditional communication constraint.

Recent work by Xu et al. [40] has considered control with observations transmitted over channels with unknown fading coefficients (i.e., continuous multiplicative observation noise). However, they assume that the unknown multiplicative noise is instantaneously revealed to the decoder. With this assumption the separation of estimation and control holds [42], and they are able to recover matching necessary and sufficient conditions for stability. The lack of channel state information makes the setup in our current paper considerably more challenging, and standard information-theoretic and control-theoretic techniques do not allow for progress. The work of Xu et al. [40] builds on the work of Xiao et al. [39] as well as that of Elia [7], both of which restrict their attention to LTI control strategies. Our formulation has no such restriction and considers general nonlinear strategies.

A related problem is that of estimating a linear system over multiplicative noise. While early work on this problem had been limited to exploring linear estimation strategies [26, 37], some recent work showed a general converse result for the estimation problem over multiplicative noise for both linear and nonlinear strategies [28]. We note that our problem can also be interpreted as an "active" estimation problem for X_0. However, unlike the estimation problem, we cannot describe the distribution of X_n in our problem since the control U_n is arbitrary. Hence, the techniques from [28] do not easily adapt to our current setting, and the current proof uses a different side-information "genie" than used in the converse for [28]. We also cannot bound the range of X_n since U_n is arbitrary, which prevents bounding the rate across the observation channel, and hence the use of a data-rate theorem approach.

The uncertainty threshold principle [1] was one of the first works to consider the control of systems with multiplicative noise. This work provided the second-moment stability threshold when both the system growth factor (coefficient of X_n) and the control gain (coefficient of U_n) were drawn i.i.d. from a Gaussian distribution. We believe our work complements this result by considering multiplicative uncertainty on the observation gain.

Following the data-rate theorems and the uncertainty-threshold principle result, Martins et al. [15], Okano et al. [23] and Kostina et al. [13] have provided stability-limits for rate-limited control with multiplicative noise on the system gain.

The formulation in [30] is closely related to the formulation in the current paper, with the multiplicative uncertainty on the control gain instead of the observation gain. However, while linear control strategies are optimal in the setting in [30], our current paper shows they are suboptimal when the multiplicative noise is on the observation gain.

Some of our results and methods are summarized in [6].

2. Problem statement. Consider the system $\mathcal{S}_a$ in (1.1). For simplicity, let the initial state X_0 be distributed as $X_0 \sim \mathcal{N}(0, 1)$. Let Z_n be i.i.d. random variables with finite second moment and bounded density $f_Z(z) = e^{-\phi(z)}$. Without loss of generality, we will use the scaling $\mathbb{E}Z_n = 1$ and $\text{Var}(Z_n) = \sigma^2$. The notation Z_n, f_Z, ϕ, and σ defined here will be used throughout the paper.

We introduce two definitions for stability of the system. The first is the notion of stability that is most commonly studied in control theory, that is, second-moment stability.

DEFINITION 2.1. The system $\mathcal{S}_a$ in (1.1) is said to be *second-moment* stabilizable if there exists an adapted control strategy $U_0, \ldots, U_n$ (a control strategy where U_k is a function of $Y_0, \ldots, Y_k$ for each $0 \le k \le n$) such that

$$\sup_n \mathbb{E}[|X_n|^2] < \infty.$$

DEFINITION 2.2. We say the controller can keep the system $\mathcal{S}_a$ in (1.1) *tight* if for every ε and for every n there exists an adapted control strategy $U_0, \ldots, U_n$, and there exist $M_\varepsilon, N_\varepsilon < \infty$ such that

$$\mathbb{P}(|X_n| < M_\varepsilon) \ge 1 - \varepsilon,$$

for $n > N_\varepsilon$. This notion has also been called stability in probability in the literature [42].

3. Linear schemes. This section first provides a simple memoryless linear strategy that can stabilize the system in a second-moment sense in Proposition 3.1. We show in Theorem 3.2 that this strategy is optimal among linear strategies. In Theorem 3.3, we highlight the limitations of linear strategies by showing that when $\mathbb{E}Z_n = 0$, linear strategies cannot stabilize the system for any growth factor $a > 1$. Finally, we consider stability in the sense of keeping the system tight and provide a scheme that achieves this in Theorem 3.4.

PROPOSITION 3.1 (A linear memoryless strategy). *The controller given by* $U_n = d^* Y_n$ *where* $d^* = \frac{a}{1+\sigma^2}$, *can stabilize the system* $\mathcal{S}_a$ *in* (1.1) *in a second-moment sense (Definition* 2.1) *if* $a \le a^*$, *where* $a^* = \sqrt{1 + \frac{1}{\sigma^2}}$.

PROOF. The above strategy gives us $X_{n+1} = (a - d^* Z_n) X_n$. Since Z_n is independent of X_n, we can write

$$\mathbb{E}[X_{n+1}^2] = \mathbb{E}[(a - d^* Z_n)^2]\mathbb{E}[X_n^2] = (a^2 - 2ad^* + (d^*)^2(1 + \sigma^2))\mathbb{E}[X_n^2]$$

$$= \frac{a^2\sigma^2}{1+\sigma^2} \cdot \mathbb{E}[X_n^2].$$

Under this control strategy $\sup_n \mathbb{E}[|X_n|^2]$ is bounded if and only if $a^2 \leq 1 + \frac{1}{\sigma^2}$. $\square$

Note that the above controller is linear in that U_n is a linear function of the Y_i and memoryless in that U_n depends only on Y_n and not Y_i for $i < n$. We might expect an improvement in the achievable performance of a linear strategy if we also allow it to use memory, that is, the past Y_n's. However, it turns out that the optimal linear strategy is in fact memoryless.

THEOREM 3.2. *The control strategy given by* $U_n = d^* Y_n$ *where* $d^* = \frac{a}{1+\sigma^2}$ *is the optimal linear strategy to stabilize* $\mathcal{S}_a$ *in a second-moment sense, in particular, for all* $a > \sqrt{1 + \frac{1}{\sigma^2}}$ *the system* $\mathcal{S}_a$ *in* (1.1) *cannot be second-moment stabilized* (*Definition* 2.1) *using a linear strategy.*

PROOF. Suppose the system $\mathcal{S}_a$ evolves following some linear strategy of the form $U_n = \sum_{i=1}^{n} \alpha_{n,i} Y_i$.

We define a system $\widetilde{\mathcal{S}}$ such that $\widetilde{X}_n$ that evolves in parallel with X_n and tracks the behavior of the strategy $U_n = d^* Y_n$. Formally, $\widetilde{\mathcal{S}}$ is defined as

$$\widetilde{X}_0 = X_0,$$

$$\widetilde{Y}_n = Z_n \widetilde{X}_n,$$

$$\widetilde{X}_{n+1} = a\widetilde{X}_n - d^* \widetilde{Y}_n,$$

where the Z_n's are the same as those acting on X_n. Then, we can write

$$\widetilde{X}_{n+1} = \widetilde{X}_n (a - d^* Z_n)$$

(3.1)
$$= \widetilde{X}_{n-m} \prod_{i=n-m}^{n} (a - d^* Z_i).$$

We will show that $\mathbb{E}[|\widetilde{X}_n|^2]$ is the minimum achievable second moment at any time n. Since $\mathbb{E}[|\widetilde{X}_n|^2] < \infty$ only when $a \leq \sqrt{1 + \frac{1}{\sigma^2}}$, we are done once we show this.

Our approach is to inductively show that $\mathbb{E}[(X_n - \widetilde{X}_n)\widetilde{X}_n] = 0$ for all n and for any linear control strategy applied to the system $\mathcal{S}$, from which it follows that

$$\mathbb{E}[X_n^2] = \mathbb{E}[\widetilde{X}_n^2] + \mathbb{E}[(X_n - \widetilde{X}_n)^2] \geq \mathbb{E}[\widetilde{X}_n^2].$$

Base case: $n = 0$ is trivially true, since $X_0 = \widetilde{X}_0$. Assume that our hypothesis is true for $n = k$. Now consider $n = k + 1$:

$$\mathbb{E}[(X_{k+1} - \widetilde{X}_{k+1})\widetilde{X}_{k+1}]$$

(3.2)
$$= a\mathbb{E}[(X_k - \widetilde{X}_k)\widetilde{X}_{k+1}] - \mathbb{E}\left[\left(\sum_{i=0}^{k} \alpha_{k,i} Y_i\right)\widetilde{X}_{k+1}\right] + \mathbb{E}[d^* \widetilde{Y}_k \widetilde{X}_{k+1}].$$

We will show that all three expectations in the final expression are zero. The first term in (3.2) is

$$\mathbb{E}[(X_k - \widetilde{X}_k)\widetilde{X}_{k+1}] = \mathbb{E}[(X_k - \widetilde{X}_k)(a\widetilde{X}_k - d^*\widetilde{Y}_k)]$$
$$= -d^*\mathbb{E}[(X_k - \widetilde{X}_k)\widetilde{Y}_k],$$

by the induction hypothesis. Because Z_k is independent of X_k and $\widetilde{X}_k$, we may compute the above expectation as

$$\begin{aligned}
\mathbb{E}[(X_k - \widetilde{X}_k)\widetilde{Y}_k] &= \mathbb{E}[(X_k - \widetilde{X}_k)Z_k\widetilde{X}_k] \\
&= \mathbb{E}[Z_k]\mathbb{E}[(X_k - \widetilde{X}_k)\widetilde{X}_k] \\
&= 0.
\end{aligned}$$
(3.3)

To handle the second term, for each $1 \le i \le k$ we can apply (3.1) to obtain

$$\begin{aligned}
\mathbb{E}[Y_i\widetilde{X}_{k+1}] &= \mathbb{E}\left[Z_i X_i \widetilde{X}_i \prod_{j=i}^{k+1}(a - d^* Z_j) \right] \\
&= \mathbb{E}[Z_i(a - d^* Z_i)]\mathbb{E}\left[X_i \widetilde{X}_i \prod_{j=i+1}^{k+1}(a - d^* Z_j) \right] \\
&= 0,
\end{aligned}$$
(3.4)

where again we have used the independence of Z_i from the other terms in the product, and $\mathbb{E}[Z_i(a - d^* Z_i)] = 0$ from the definition of d^*. Finally, the last term may be computed in a similar manner as

$$\begin{aligned}
\mathbb{E}[\widetilde{Y}_k\widetilde{X}_{k+1}] &= \mathbb{E}[Z_k\widetilde{X}_k(a\widetilde{X}_k - d^* Z_k\widetilde{X}_k)] \\
&= \mathbb{E}[Z_k(a - d^* Z_k)]\mathbb{E}[\widetilde{X}_k^2] \\
&= 0.
\end{aligned}$$
(3.5)

by the definition of d^*.

Equations (3.3), (3.4) and (3.5), establish that all three terms in (3.2) are zero. Hence, $\mathbb{E}[(X_n - \widetilde{X}_n)\widetilde{X}_n] = 0$ for all n, and we are done. $\square$

A similar analysis illustrates the limitations of linear strategies when $\mathbb{E}Z_n = 0$, in contrast with nonlinear strategies to be described in the next section.

THEOREM 3.3. *Suppose that instead of $\mathbb{E}Z_n = 1$, we have $\mathbb{E}Z_n = 0$. Then, for all $a > 1$, the system $\mathcal{S}_a$ in (1.1) cannot be second-moment stabilized using a linear strategy. In other words, linear strategies cannot tolerate any growth in the system.*

PROOF. Suppose the system S_a evolves following some linear strategy of the form $U_n = \sum_{i=1}^n \alpha_{n,i} Y_i$.

We will show by induction that for each n, we may write $X_n = W_n X_0$, where W_n is a function of $Z_0, Z_1, \ldots, Z_{n-1}$, and $\mathbb{E} W_n = a^n$. Clearly, this holds for $n = 0$ with $W_0 = 1$. For the inductive step, note that

$$X_{n+1} = aX_n - U_n = \left(aW_n - \sum_{i=1}^n \alpha_{n,i} Z_i W_i \right) X_0,$$

so we may take $W_{n+1} = aW_n - \sum_{i=1}^n \alpha_{n,i} Z_i W_i$. Since Z_i is independent of W_i for each i, we have

$$\mathbb{E} W_{n+1} = a\mathbb{E} W_n - \sum_{i=1}^n \alpha_{n,i} (\mathbb{E} Z_i)(\mathbb{E} W_i) = a^{n+1},$$

completing the induction. It follows that

$$\mathbb{E}[X_n^2] = \mathbb{E}[W_n^2] \cdot \mathbb{E}[X_0^2] \geq (\mathbb{E} W_n)^2 \cdot \mathbb{E}[X_0^2] = a^{2n} \cdot \mathbb{E}[X_0^2],$$

and so $\mathbb{E}[X_n^2]$ grows without bound when $a > 1$. $\square$

Finally, the next theorem considers the weaker sense of stability of keeping the system tight, which is the sense of stability that the impossibility results in Section 5 use.

THEOREM 3.4. *Suppose that the density function f_Z of Z_n is bounded, and consider linear memoryless strategies of the form $U_n = ad \cdot Y_n$ for a constant $d > 0$. Let $d^\star = \operatorname{argmin}_d \mathbb{E}[\log|1 - d \cdot Z_n|]$ and $a^\star = e^{-\mathbb{E}[\log|1-d^\star \cdot Z_n|]}$. If $d = d^\star$, then the system S_a in (1.1) can be kept tight (Definition 2.2) provided that $|a| < a^\star$. Further, no such strategy can keep the system tight if $|a| \geq a^\star$.*

PROOF. Applying the control law $U_n = adY_n$, we calculate that

$$X_n = a(1 - dZ_{n-1})X_{n-1}$$

$$= a^n \prod_{i=0}^{n-1} (1 - dZ_i) X_0.$$

Let $W_i = \log|1 - dZ_i|$, and let $S_n = \sum_{i=1}^n (W_i + \log|a|)$. Taking logarithms gives us

$$\log|X_n| = S_n + \log|X_0|.$$

Note that

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}} \log|X_n| = \lim_{n\to\infty} \frac{1}{\sqrt{n}} S_n$$

almost surely, so as will be seen shortly, it suffices to analyze S_n.

Take C to be an upper bound on the density of Z_i. Then we have

$$\mathbb{P}(W_i < -t) = \mathbb{P}(|1 - dZ_i| < e^{-t}) \leq \frac{C \cdot e^{-t}}{d},$$

so that W_i has an exponentially decaying left tail. Similarly,

$$\mathbb{P}(W_i > t) = \mathbb{P}(|1 - dZ_i| > e^t) \leq \frac{\mathbb{E}(1 - dZ_i)^2}{e^{2t}},$$

so W_i also has an exponentially decaying right tail.

Thus, W_i has finite first and second moments. Let μ_d and σ_d denote the mean and variance of W_i, respectively. Defining $\widetilde{S}_n = \sum_{i=1}^{n}(W_i - \mu_d)$, the central limit theorem gives us that

$$(3.6) \qquad\qquad \frac{1}{\sqrt{n}}\widetilde{S}_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_d^2) \qquad \text{as } n \to \infty.$$

If $|a| < a^\star$ and we take $d = d^\star$, then we see that $\log|a| < \log a^\star = -\mu_{d^\star} = -\mu_d$. Thus, there exists $\varepsilon > 0$ such that $\log|a| + \mu_d < -2\varepsilon$. Using the union bound, we then have

$$\mathbb{P}(\log|X_n| \geq -n\varepsilon) \leq \mathbb{P}(\log|X_0| > n\varepsilon) + \mathbb{P}(S_n \geq -2n\varepsilon).$$

We have that $\mathbb{P}(\log|X_0| > n\varepsilon) \to 0$ as $n \to \infty$, and also by the law of large numbers $\mathbb{P}(S_n \geq -2n\varepsilon) \to 0$ almost surely. Hence, $\mathbb{P}(\log|X_n| < -n\varepsilon) \to 1$ and the system is kept tight.

On the other hand, suppose that $|a| \geq a^\star$. Then we have $\log|a| \geq \log a^\star = -\mu_{d^\star} \geq -\mu_d$, so $S_n \geq \widetilde{S}_n$. Consider $\delta > 0$. For n large enough, we have that

$$\begin{aligned}
\mathbb{P}(\log|X_n| \leq n^{\frac{1}{4}}) &\leq \mathbb{P}(\log|X_n| \leq \delta\sqrt{n}) \\
&\leq \mathbb{P}(\log|X_0| \leq -\delta\sqrt{n}) + \mathbb{P}(S_n \leq 2\delta\sqrt{n}) \\
&\leq \mathbb{P}(\log|X_0| \leq -\delta\sqrt{n}) + \mathbb{P}(\widetilde{S}_n \leq 2\delta\sqrt{n}),
\end{aligned}$$

where we used the union bound and the fact that $S_n \geq \widetilde{S}_n$ to get the two inequalities. Now $\mathbb{P}(\log|X_0| \leq -\delta\sqrt{n}) \to 0$ as $n \to \infty$ and $\mathbb{P}(\widetilde{S}_n \leq 2\delta\sqrt{n}) \to \Phi(\frac{2\delta}{\sigma_d})$, by (3.6). Hence,

$$\limsup_{n\to\infty} \mathbb{P}(\log|X_n| \leq n^{\frac{1}{4}}) \leq \Phi\left(\frac{2\delta}{\sigma_d}\right),$$

which gives that

$$\limsup_{n\to\infty} \mathbb{P}(\log|X_n| > n^{\frac{1}{4}}) \geq \frac{1}{2}.$$

Thus, in this case the system is not kept tight. $\quad\square$

4. Nonlinear schemes. In the previous section, we focused on linear strategies, where U_n is taken to be a linear combination of Y_i for $0 \leq i \leq n$. We now consider whether more general strategies can do better. Theorem 4.1 shows that when Z_n is Gaussian, a perturbation of the linear strategy indeed does better in the second-moment sense. (The same result should hold for rather general Z_n; see Remark 4.1.) In the setting where $\mathbb{E}Z_n = 0$, Theorem 4.3 exhibits a nonlinear strategy that achieves a nontrivial growth factor $a > 1$. This contrasts with Theorem 3.3, which showed that linear strategies cannot achieve any gain in this setting. In both Theorem 4.1 and Theorem 4.3, improvement is achieved by taking into account information from the previous round while choosing the control.

On the other hand, Theorem 4.5 shows that when $a > a^* = \sqrt{1 + \frac{1}{\sigma^2}}$, for any memoryless strategy (in the sense that U_n is a function of only Y_n), we cannot guarantee for all distributions of X_n that $\mathbb{E}[X_{n+1}^2] \leq \mathbb{E}[X_n^2]$. This suggests that in the memoryless setting, the linear strategy from the previous section may be optimal. However, it does not rule out the possibility for an increase in second moment after one round to be compensated by a larger decrease later.

THEOREM 4.1. *Let $a^* = \sqrt{1 + \frac{1}{\sigma^2}}$ be as in Proposition 3.1. Suppose that our multiplicative noise Z_n has a Gaussian law $Z_n \sim \mathcal{N}(1, \sigma^2)$. Then there exists $a > a^*$ for which a (nonlinear) controller can stabilize the system in a second-moment sense.*

We first establish an elementary inequality for Gaussian variables. In what follows, we define the signum function $\mathrm{sgn}(x)$ to be 1 if $x \geq 0$ and -1 otherwise.

LEMMA 4.2. *Let $Z \sim \mathcal{N}(1, \sigma^2)$, with $\sigma > 0$. We have*

$$\mathbb{E}\left[\mathrm{sgn}(Z)\left(1 - \frac{Z}{1 + \sigma^2} \right) \right] > 0.$$

PROOF. It is convenient to write $Z = 1 - \sigma \widetilde{Z}$, where $\widetilde{Z} \sim \mathcal{N}(0, 1)$. Let $s = \frac{1}{\sigma}$, and let γ denote the standard Gaussian density. Note that

$$\gamma(x) \geq \frac{1}{\sqrt{2\pi}} \max\left(1 - \frac{x^2}{2}, 0 \right)$$

for all x. Hence,

(4.1)
$$\sqrt{2\pi} \int_0^s \gamma(x)\,dx \geq \int_0^s \max\left(1 - \frac{x^2}{2}, 0 \right) dx$$
$$= \begin{cases} s - \dfrac{s^3}{6} & \text{if } s \leq \sqrt{2}, \\ \dfrac{2\sqrt{2}}{3} & \text{if } s > \sqrt{2}. \end{cases}$$

We also have

$$(4.2) \qquad s\sqrt{2\pi}\int_s^\infty x\gamma(x)\,dx = s\int_s^\infty xe^{-\frac{x^2}{2}}\,dx = se^{-\frac{s^2}{2}}.$$

It can be checked by elementary calculations that for any $s > 0$, (4.1) is always strictly greater than (4.2). Indeed, for $s < \sqrt{2}$ use $e^{-\frac{s^2}{2}} < 1 - \frac{s^2}{2} + \frac{s^4}{8} < 1 - \frac{s^2}{6}$, and for $s > \sqrt{2}$ note that $se^{-\frac{s^2}{2}}$ is decaying. Thus,

$$\frac{1}{2} - \mathbb{E}[\mathbb{1}_{\widetilde{Z}\geq s}] = \int_0^s \gamma(x)\,dx > s\int_s^\infty x\gamma(x)\,dx = s\cdot\mathbb{E}[\mathbb{1}_{\widetilde{Z}\geq s}\cdot\widetilde{Z}].$$

Let us rewrite the above equation in terms of Z and σ, noting that $\mathbb{1}_{\widetilde{Z}\geq s} = \frac{1}{2}(1 - \text{sgn}(Z))$. We obtain

$$\mathbb{E}[\text{sgn}(Z)] > \frac{1}{\sigma}\cdot\mathbb{E}\left[(1 - \text{sgn}(Z))\cdot\frac{1 - Z}{\sigma}\right].$$

Rearranging, we have

$$\mathbb{E}\left[\text{sgn}(Z)\left(1 + \frac{1 - Z}{\sigma^2}\right)\right] > \frac{1}{\sigma^2}\mathbb{E}[1 - Z] = 0.$$

Finally, multiplying both sides by $\frac{1+\sigma^2}{\sigma^2}$ yields

$$\mathbb{E}\left[\text{sgn}(Z)\left(1 - \frac{Z}{1 + \sigma^2}\right)\right] > 0. \qquad \square$$

PROOF OF THEOREM 4.1.   To show second-moment stability, it suffices to exhibit controls U_n and U_{n+1} which ensure that $\mathbb{E}X_{n+2}^2 \leq \mathbb{E}X_n^2$ for all possible distributions of X_n. For a positive ε to be specified later, choose

$$a = (1 + \varepsilon^2)a^* = (1 + \varepsilon^2)\sqrt{1 + \frac{1}{\sigma^2}}.$$

For our controls, we take

$$U_n = \frac{a}{1 + \sigma^2}Y_n \quad\text{and}\quad U_{n+1} = \frac{a}{1 + \sigma^2}Y_{n+1} + \varepsilon Y_{n+1}\cdot\left|\frac{Y_n}{Y_{n+1}}\right|.$$

Note that the expression for U_n and the first term in the expression for U_{n+1} are the same as in the linear strategy from Proposition 3.1. However, here we have added a small perturbation to U_{n+1}. For convenience, define the function $g(x) = 1 - \frac{x}{1+\sigma^2}$. Then

$$X_{n+1} = a\cdot g(Z_n)X_n,$$

$$(4.3)\quad X_{n+2} = a\cdot g(Z_{n+1})X_{n+1} - \varepsilon\cdot aZ_{n+1}g(Z_n)\cdot\left|\frac{Z_n}{aZ_{n+1}g(Z_n)}\right|X_n$$

$$= a^2\cdot g(Z_{n+1})g(Z_n)X_n - \varepsilon\cdot\text{sgn}(Z_{n+1})\cdot g(Z_n)\left|\frac{Z_n}{g(Z_n)}\right|X_n.$$

We will compute the second moment of (4.3). Let

$$A = g(Z_{n+1})g(Z_n), \qquad B = \mathrm{sgn}(Z_{n+1}) \cdot g(Z_n)\left|\frac{Z_n}{g(Z_n)}\right|.$$

Then we have

$$\mathbb{E}[A^2] = \mathbb{E}[g(Z_{n+1})] \cdot \mathbb{E}[g(Z_n)] = \frac{\sigma^4}{(1+\sigma^2)^2},$$

$$\mathbb{E}[B^2] = \mathbb{E}Z_n^2 = 1 + \sigma^2,$$

$$\mathbb{E}[AB] = \mathbb{E}\left[g(Z_n)^2 \cdot \left|\frac{Z_n}{g(Z_n)}\right|\right] \cdot \mathbb{E}[\mathrm{sgn}(Z_{n+1})g(Z_{n+1})] > 0,$$

where the inequality in the last line follows from Lemma 4.2 and the fact that $g(Z_n)^2 \cdot |\frac{Z_n}{g(Z_n)}|$ is almost surely positive.

Recall that the Z_n and Z_{n+1} are both independent of X_n, so taking second moments in (4.3), we have

$$\begin{aligned}
\mathbb{E}X_{n+2}^2 &= (a^4 \cdot \mathbb{E}A^2 - 2\varepsilon a^2 \cdot \mathbb{E}AB + \varepsilon^2 \cdot \mathbb{E}B^2)\mathbb{E}X_n^2 \\
&= \left(\frac{a^4\sigma^4}{(1+\sigma^2)^2} - 2\varepsilon a^2 \cdot \mathbb{E}AB + O(\varepsilon^2)\right)\mathbb{E}X_n^2 \\
&= \left((1+\varepsilon^2)^4 - \frac{2\varepsilon(1+\varepsilon^2)(1+\sigma^2)}{\sigma^2} \cdot \mathbb{E}AB + O(\varepsilon^2)\right)\mathbb{E}X_n^2 \\
&= \left[1 - \varepsilon \cdot \frac{2(1+\sigma^2)}{\sigma^2}\mathbb{E}AB + O(\varepsilon^2)\right]\mathbb{E}X_n^2.
\end{aligned}$$

Since $\mathbb{E}AB > 0$, when ε is a sufficiently small positive number, this gives $\mathbb{E}X_{n+2}^2 \le \mathbb{E}X_n^2$, showing second-moment stability. $\square$

REMARK 4.1. We actually suspect that Theorem 4.1 applies to all continuous distributions of Z_n. Indeed, the above analysis can be carried out for a more general class of control strategies. Consider instead

$$U_n = \frac{a}{1+\sigma^2}Y_n \quad \text{and} \quad U_{n+1} = \frac{a}{1+\sigma^2}Y_{n+1} + \varepsilon Y_{n+1} \cdot h\left(\frac{Y_n}{Y_{n+1}}\right),$$

where h is any function [above, we used $h(x) = |x|$]. Then we would carry out the same analysis except with

$$B = aZ_{n+1}g(Z_n) \cdot h\left(\frac{Z_n}{aZ_{n+1}g(Z_n)}\right).$$

The crucial properties we needed were that $\mathbb{E}B^2 < \infty$ and $\mathbb{E}AB \neq 0$. Thus, for all distributions of Z_n, as long as there exists some function h verifying those two properties, the conclusion of Theorem 4.1 applies.

The next theorem shows that a perturbation can also improve upon linear strategies when $\mathbb{E}Z_n = 0$.

THEOREM 4.3. *Suppose that instead of $\mathbb{E}Z_n = 1$, we have $\mathbb{E}Z_n = 0$. Then, as long as Z_n has finite second moment, there exists $a > 1$ for which a (nonlinear) controller can stabilize the system in a second-moment sense.*

We first prove a technical lemma.

LEMMA 4.4. *Let Z be a random variable with $\mathbb{E}Z = 0$ and finite first moment. Then, for all sufficiently small $\varepsilon > 0$, we have*

$$\mathbb{E}\left[Z\left|\frac{\varepsilon}{Z} - 1\right|\right] < 0.$$

PROOF. For $0 \le t \le \frac{1}{2}$, define the function

$$f(x, t) = x\left|\frac{t}{x} - 1\right|.$$

Note that for each $x \ne 0$ and each t, we have

$$\left|\frac{f(x, t) - f(x, 0)}{t}\right| \le 1 \quad \text{and} \quad \lim_{t \to 0} \frac{f(x, t) - f(x, 0)}{t} = -1.$$

Thus, letting $F(t) = \mathbb{E}f(Z, t)$, the dominated convergence theorem implies

$$\lim_{t \to 0} \frac{F(t) - F(0)}{t} = -1.$$

Consequently, for all sufficiently small t, we have $F(t) < 0$, as desired. □

PROOF OF THEOREM 4.3. We take an approach similar to the proof of Theorem 4.1. Again, it suffices to exhibit controls U_n and U_{n+1} which ensure that $\mathbb{E}[X_{n+2}^2] \le \mathbb{E}[X_n^2]$ for all possible distributions of X_n. By Lemma 4.4, take a small enough $\varepsilon_0 > 0$ so that

$$(4.4) \qquad \mathbb{E}\left[Z_n\left|\frac{\varepsilon_0}{Z_n} - 1\right|\right] < 0.$$

Let $\varepsilon > 0$ be another small number to be specified later, and take $a = 1 + \varepsilon^2$. For our controls, we take

$$U_n = a\varepsilon_0^{-1}Y_n,$$

$$U_{n+1} = -a^2\varepsilon_0^{-1}Y_n - \varepsilon Y_n \cdot \left|\frac{Y_{n+1}}{Y_n}\right|.$$

Then

$$X_{n+1} = a X_n - a \varepsilon_0^{-1} Z_n X_n,$$

$$X_{n+2} = a X_{n+1} + a^2 \varepsilon_0^{-1} Y_n + \varepsilon Y_n \cdot \left| \frac{Y_{n+1}}{Y_n} \right|$$

$$= a^2 X_n + \varepsilon Z_n \cdot \left| \frac{a Z_{n+1}(1 - \varepsilon_0^{-1} Z_n)}{Z_n} \right| X_n$$

$$= a^2 X_n + a \varepsilon_0^{-1} \varepsilon \cdot |Z_{n+1}| \cdot Z_n \left| \frac{\varepsilon_0}{Z_n} - 1 \right| X_n.$$

For convenience, let $A = \varepsilon_0^{-1} \cdot |Z_{n+1}| \cdot Z_n |\frac{\varepsilon_0}{Z_n} - 1|$, and note that $\mathbb{E} A^2 < \infty$ since Z_n and Z_{n+1} have finite second moments. Substituting this definition for A, we calculate

$$\mathbb{E}[X_{n+2}^2] = a^2 \cdot \mathbb{E}(a + \varepsilon A)^2 \cdot \mathbb{E}[X_n^2]$$

$$= (1 + \varepsilon^2)^2 \cdot \mathbb{E}(1 + \varepsilon^2 + \varepsilon A)^2 \cdot \mathbb{E}[X_n^2]$$

$$= (1 + 2\varepsilon \cdot \mathbb{E} A + O(\varepsilon^2)) \mathbb{E}[X_n^2].$$

By (4.4), we have that $\mathbb{E} A$ is strictly negative. Thus, for small enough positive ε, we obtain $\mathbb{E}[X_{n+2}^2] \le \mathbb{E}[X_n^2]$, as desired. $\square$

The next theorem pertains to schemes of the form $U_n = h(Y_n)$, where $h : \mathbb{R} \to \mathbb{R}$ is *any* fixed function.

THEOREM 4.5. *Consider any $a > a^* = \sqrt{1 + \frac{1}{\sigma^2}}$ and any measurable function $h : \mathbb{R} \to \mathbb{R}$. Then there exists a random variable X with finite second moment for which*

$$\mathbb{E}[a^2(X - h(X Z_n))^2] > \mathbb{E} X^2.$$

In particular, we cannot guarantee $\mathbb{E} X_{n+1}^2 \le \mathbb{E} X_n^2$ for the scheme $U_n = h(Y_n)$.

PROOF. Let M be a large parameter to be specified later. Consider the probability density

$$\rho(x) = \begin{cases} \left(1 - \dfrac{1}{M^2}\right)^{-1} |x|^{-3} & \text{if } 1 \le |x| \le M, \\ 0 & \text{otherwise.} \end{cases}$$

We will take X to have density ρ, and for appropriate M, we will find that

$$\mathbb{E}[a^2(X - h(X Z_n))^2] > \mathbb{E} X^2.$$

Recall our notation $f_Z(x) = e^{-\phi(x)}$ for the density of Z_n. To aid in our calculations, for each integer $k \geq 0$ and real number $y \neq 0$, we consider the quantity

$$
\begin{aligned}
\alpha_k(y) &= \int_{-\infty}^{\infty} \frac{x^k \rho(x) f_Z(y/x)}{|x|} \, dx \\
&= \int_1^M \frac{x^k \rho(x) f_Z(y/x) + (-x)^k \rho(-x) f_Z(-y/x)}{|x|} \, dx \\
&= \left(1 - \frac{1}{M^2}\right)^{-1} \int_1^M \frac{x^k f_Z(y/x) + (-x)^k f_Z(-y/x)}{x^4} \, dx \\
&= \left(1 - \frac{1}{M^2}\right)^{-1} \int_y^{y/M} \frac{y^k s^{-k} f_Z(s) + y^k (-s)^{-k} f_Z(-s)}{y^4 s^{-4}} \left(-\frac{y}{s^2}\right) ds \\
&= \left(1 - \frac{1}{M^2}\right)^{-1} \int_{y/M}^y \frac{s^{2-k} f_Z(s) + (-s)^{2-k} f_Z(-s)}{y^{3-k}} \, ds,
\end{aligned}
$$

where we have made the substitution $x = y/s$. Let $\varepsilon > 0$ be a small parameter. Consider a fixed t with $\varepsilon \leq t \leq 1 - \varepsilon$, and set $y = \pm M^t$. We find that

$$(4.5) \quad \lim_{M \to \infty} |y| y^{2-k} \alpha_k(y) = \int_0^{\infty} (s^{2-k} f_Z(s) + (-s)^{2-k} f_Z(-s)) \, ds = \mathbb{E}[Z_n^{2-k}]$$

uniformly over $\varepsilon \leq t \leq 1 - \varepsilon$, where we have taken care to ensure that the above holds for both possible signs of y. Let $\delta > 0$ also be a small parameter. We now choose M to be sufficiently large so that

$$
\left(1 - \frac{1}{M^2}\right)^{-1} \leq 1 + \delta,
$$

and also for all y with $M^\varepsilon \leq |y| \leq M^{1-\varepsilon}$ [in light of (4.5)],

$$
\begin{aligned}
(4.6) \quad \alpha_2(y) - \frac{\alpha_1(y)^2}{\alpha_0(y)} &\geq (1 - \delta)\left(\frac{1}{|y|} - \frac{|y|^{-2} y^{-2} (\mathbb{E}Z_n)^2}{|y|^{-1} y^{-2} \mathbb{E}Z_n^2}\right) \\
&= (1 - \delta)\frac{1}{|y|}\left(1 - \frac{1}{1 + \sigma^2}\right) = (1 - \delta)\frac{\sigma^2}{|y|(1 + \sigma^2)}.
\end{aligned}
$$

We then have

$$
\mathbb{E}X^2 = \int_{-\infty}^{\infty} x^2 \rho(x) \, dx \leq 2(1 + \delta) \int_1^M \frac{1}{x} \, dx = 2(1 + \delta) \log M
$$

and

$$
\begin{aligned}
\mathbb{E}(X - h(XZ_n))^2 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \rho(x)(x - h(xz))^2 f_Z(z) \, dz \, dx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\rho(x)(x - h(y))^2 f_Z(y/x)}{|x|} \, dy \, dx \\
&= \int_{-\infty}^{\infty} (\alpha_2(y) - 2h(y)\alpha_1(y) + h(y)^2 \alpha_0(y)) \, dy.
\end{aligned}
$$

Note that the integrand in the last expression is a quadratic function in $h(y)$ whose minimum possible value is $\alpha_2(y) - \frac{\alpha_1(y)^2}{\alpha_0(y)}$, and note also that this quantity is non-negative since $\alpha_2(y)\alpha_0(y) \geq \alpha_1(y)^2$ by the Cauchy–Schwarz inequality. Thus,

$$
\begin{aligned}
\mathbb{E}(X - h(XZ_n))^2 &\geq \int_{-\infty}^{\infty} \left(\alpha_2(y) - \frac{\alpha_1(y)^2}{\alpha_0(y)} \right) dy \\
&\geq \int_{M^\varepsilon}^{M^{1-\varepsilon}} \left(\alpha_2(y) - \frac{\alpha_1(y)^2}{\alpha_0(y)} + \alpha_2(-y) - \frac{\alpha_1(-y)^2}{\alpha_0(-y)} \right) dy \\
&\geq \frac{2(1-\delta)\sigma^2}{1+\sigma^2} \int_{M^\varepsilon}^{M^{1-\varepsilon}} \frac{1}{y} dy = \frac{2(1-\delta)(1-2\varepsilon)\sigma^2}{1+\sigma^2} \log M,
\end{aligned}
$$

where we have plugged in the bound from (4.6). Consequently,

$$
\frac{\mathbb{E}a^2(X - h(XZ_n))^2}{\mathbb{E}X^2} \geq \frac{(1-\delta)(1-2\varepsilon)}{1+\delta} \cdot \frac{a^2\sigma^2}{1+\sigma^2}.
$$

Since $a > \sqrt{1 + \frac{1}{\sigma^2}}$, the right-hand side is strictly greater than 1 when ε and δ are sufficiently small. This completes the proof. $\square$

5. An impossibility result.

THEOREM 5.1. *For the system $\mathcal{S}_a$, suppose that ϕ is differentiable and satisfies $|z \cdot \phi'(z)| \leq C_1 + C_2 \cdot \phi(z)$ for all z, and also $e^{-\phi(z)} \leq |z|^{-1-\delta}$ for some $\delta > 0$. We additionally assume $\phi(\cdot)$ satisfies a doubling condition on $\phi'(\cdot)$ such that if $\frac{z_1}{2} \leq z_2 \leq 2z_1$, then $\phi'(z_2) \leq C_3 \cdot \phi'(z_1)$. Then there exists $a \in \mathbb{R}$, $a < \infty$ such that $\mathbb{P}(|X_n| < M) \to 0$ for all $M < \infty$.*

Note that the conditions on $\phi(\cdot)$ above imply the conditions in Theorem 1.1.

We rewrite the system $\mathcal{S}_a$ from (1.1) here, with state denoted as $X_{a,n}$, to emphasize the dependence on a:

(5.1)
$$
\begin{aligned}
X_{a,n+1} &= a \cdot X_{a,n} - U_{a,n}, \\
Y_{a,n} &= Z_n \cdot X_{a,n}.
\end{aligned}
$$

Now define $U_n := a^{-n} U_{a,n}$, and consider the system $\mathcal{S}$, which is the system $\mathcal{S}_a$ scaled by a:

(5.2)
$$
\begin{aligned}
X_{n+1} &= X_n - U_n, \\
Y_n &= Z_n \cdot X_n.
\end{aligned}
$$

The Z_n's and the initial state $X_0 = X_{a,0}$ are identical in both systems. Then the scaled system satisfies $X_n = a^{-n} X_{a,n}$. Thus we have that

$$
\mathbb{P}(|X_{a,n}| < M) = \mathbb{P}(|X_n| < a^{-n}M).
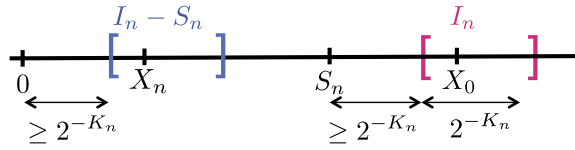$$

FIG. 2.   *A caricature illustrating the intervals I_n and $I_n - S_n$.*

As a result, it suffices to bound the probability that the state of the system $\mathcal{S}$, that is, $|X_n|$, is contained in intervals that are shrinking by a factor of a at each time step. The rest of this section uses the notation X_n to refer to the state of the system $\mathcal{S}$ and $X_{a,n}$ to refer the state of the system $\mathcal{S}_a$.

5.1. *Definitions.*   Let $S_n := \sum_{i=0}^{n-1} U_i$. Hence, $X_n = X_0 - S_n$.

The goal of the controller is to have S_n be as close to X_0 as possible. We will track the progress of the controller through intervals I_n that contain X_0 and are decreasing in length. Figure 2 illustrates the intervals I_n and $I_n - S_n$.

Let $d(I_n, S) := \inf_{x \in I_n} |S - x|$ denote the distance of a point S from the interval I_n.

DEFINITION 5.1.   For all $n \geq 0$ and for $k \in \mathbb{Z}$, there exists a unique integer $h(k)$ such that $X_0 \in [\frac{h(k)}{2^k}, \frac{h(k)+1}{2^k})$. Let $J(k) := [\frac{h(k)}{2^k}, \frac{h(k)+1}{2^k})$. We now inductively define

$$K_0 := \min\{k \geq 0 \mid d(J(k), 0) \geq 2^{-k}\} \quad \text{and}$$

$$K_n := \min\{k \mid k > K_{n-1}, d(J(k), S_n) \geq 2^{-k}\}.$$

Write $H_n := h(K_n)$ and $I_n := J(K_n) = [\frac{H_n}{2^{K_n}}, \frac{H_n+1}{2^{K_n}})$.

Let Y_0^n indicate the observations Y_0 to Y_n, and let $\mathcal{F}_n := \{Y_0^n, K_0^n, H_0^n\}$, which is the total information available to the controller at time n. Let $f_{X_n}(x|\mathcal{F}_n)$ be the conditional density of X_n given $\mathcal{F}_n$.

5.2. *Relationships between I_n, K_n, S_n and X_n.*   We state and prove two lemmas that will be used in the main proof. The first lemma uses K_n to bound how fast S_n approaches X_0.

LEMMA 5.2.

$$2^{-K_n} \leq |X_0 - S_n|,$$

and if $K_n > K_{n-1} + 1$, then

$$|X_0 - S_n| \leq 2^{2-K_n}.$$

PROOF. From the definition of I_n, we know that $d(I_n, S_n) \geq 2^{-K_n}$. This gives $2^{-K_n} \leq |X_0 - S_n|$, since $X_0 \in I_n$.

To show the second half of the inequality, suppose that $|X_0 - S_n| > 2^{2-K_n}$. Then

$$2^{1-K_n} < |X_0 - S_n| - 2^{1-K_n}.$$

Hence, there exists a larger interval $J(K_n - 1)$ that contains X_0 such that

$$2^{1-K_n} < d(J(K_n - 1), S_n),$$

where $J(K_n - 1)$ is an interval of length $2^{1-K_n} > 2^{-K_n}$. Since we also assumed that $K_n > K_{n-1} + 1$, this contradicts the assumption that K_n was the minimal $k > K_{n-1}$ such that $d(J(k), S_n) \geq 2^{-k}$. □

The second lemma bounds the ratio between two points in the interval of interest.

LEMMA 5.3. *For $t \in I_n - S_n$, we have that $\frac{1}{2} \leq \frac{X_n}{t} \leq 2$.*

PROOF. We have from Lemma 5.2 that $2^{-K_n} \leq |X_n|$. The lemma follows since the length of the interval $I_n - S_n$ is 2^{-K_n}. □

5.3. *Preliminary estimates of the Z_i.* We also require some basic estimates for the Z_i, which we record here. Recall that we assumed the existence of a number $\delta > 0$ such that $e^{-\phi(z)} \leq |z|^{-1-\delta}$.

LEMMA 5.4. *Let $\delta' = \delta/(1 + \delta)$. For each i and any $t \geq 0$, we have*

$$\mathbb{P}(\phi(Z_i) \geq t) \leq \frac{2}{\delta'} e^{-\delta' t}.$$

PROOF. Let $s = e^{t/(1+\delta)}$, so that $s^{-1-\delta} = e^{-t}$. We have

$$\mathbb{P}(\phi(Z_i) \geq t) = \mathbb{P}(e^{-\phi(Z_i)} \leq e^{-t}) \leq \int_{-\infty}^{\infty} \min(e^{-t}, |z|^{-1-\delta}) \, dz$$

$$= 2\left(\int_0^s e^{-t} \, dz + \int_s^{\infty} z^{-1-\delta} \, dz\right)$$

$$= 2\left(s \cdot e^{-t} + \frac{s^{-\delta}}{\delta}\right) = 2\left(1 + \frac{1}{\delta}\right)s^{-\delta} = \frac{2}{\delta'} e^{-\delta' t}. \qquad □$$

LEMMA 5.5. *For each i, the random variable $\phi(Z_i)$ has finite moments of all orders.*

PROOF. The condition $|Z_i \phi'(Z_i)| \leq C_1 + C_2 \phi(Z_i)$ implies $\phi(Z_i) \geq -\frac{C_1}{C_2}$. According to Lemma 5.4, we also know that $\phi(Z_i)$ has exponentially decaying upper tails. Thus, $\phi(Z_i)$ has finite moments of all orders. □

5.4. *Proof of the main result.*   The key element of the proof is to provide the interval I_n to the controller at time n as side-information in addition to Y_n. Our strategy is to first bound the density $f_{X_n}(x|\mathcal{F}_n)$ by comparing the change in density from time n to $n+1$. This bound helps us generate bounds for the probabilities of three events that cover the event of interest $\{|X_n| < a^{-n}M\}$. We will show that for large enough a the probabilities of all three of these events go to 0 as $n \to \infty$.

PROOF OF THEOREM 5.1.   Consider

$$
f_{X_n}(x \mid \mathcal{F}_n)
$$
$$
= f_{X_n}(x \mid Y_n, K_n, H_n, \mathcal{F}_{n-1})
$$
$$
= \frac{f_{Y_n, K_n, H_n}(Y_n, K_n, H_n \mid X_n = x, \mathcal{F}_{n-1}) \cdot f_{X_n}(x \mid \mathcal{F}_{n-1})}{f_{Y_n, K_n, H_n}(Y_n, K_n, H_n \mid \mathcal{F}_{n-1})}.
$$

Since $X_0 \in I_n$, the controller knows that $X_n \in I_n - S_n$, where $I_n - S_n$ represents the interval I_n shifted by S_n. We can calculate the ratio of the densities at $x, w \in I_n - S_n$ as

(5.3)
$$
\frac{f_{X_n}(x \mid \mathcal{F}_n)}{f_{X_n}(w \mid \mathcal{F}_n)}
$$
$$
= \frac{f_{X_n}(x \mid \mathcal{F}_{n-1})}{f_{X_n}(w \mid \mathcal{F}_{n-1})} \cdot \frac{f_{Y_n}(Y_n \mid X_n = x, \mathcal{F}_{n-1})}{f_{Y_n}(Y_n \mid X_n = w, \mathcal{F}_{n-1})}.
$$

Since K_n and H_n are defined by I_n, the conditional distributions of K_n and H_n given $X_n = x$ and $X_n = w$ are equal for $x, w \in I_n - S_n$. So these terms cancel when we consider a ratio, giving (5.3).

Taking logarithms and using the triangle inequality gives the following recursive lemma.

LEMMA 5.6.

(5.4)
$$
\left| \log \frac{f_{X_n}(x \mid \mathcal{F}_n)}{f_{X_n}(w \mid \mathcal{F}_n)} \right|
$$
$$
\leq 2^{K_n+1} C_3 |Z_n \cdot \phi'(Z_n)| |x - w| + \left| \log \frac{f_{X_n}(x \mid \mathcal{F}_{n-1})}{f_{X_n}(w \mid \mathcal{F}_{n-1})} \right|.
$$

The proof is deferred to Section 6 to improve readability. This lemma helps us establish the recursive step, since the control law gives us that

$$
f_{X_n}(x \mid \mathcal{F}_{n-1}) = f_{X_{n-1}}(x + U_{n-1} \mid \mathcal{F}_{n-1}),
$$

since U_{n-1} is $\mathcal{F}_{n-1}$ measurable. Substituting this into (5.4) and unfolding recursively gives

(5.5)
$$
\left| \log \frac{f_{X_n}(x \mid \mathcal{F}_n)}{f_{X_n}(w \mid \mathcal{F}_n)} \right|
\leq \sum_{i=1}^{n} 2^{K_i+1} C_3 |Z_i \cdot \phi'(Z_i)| |x - w| + \left| \log \frac{f_{X_0}(x + S_n)}{f_{X_0}(w + S_n)} \right|.
$$

Inequality (5.5) separates the effect of the uncertainty due to X_0 and the subsequent uncertainty due to the observations and control.

Let $\eta_n = \max_{x,w \in I_n - S_n} |\log \frac{f_{X_0}(x+S_n)}{f_{X_0}(w+S_n)}|$. Since I_n is an interval of size at most 2^{-n} which contains X_0, we get that

(5.6)
$$
|\eta_n| \leq \frac{1}{2} |(X_0 + 2^{2-n})^2 - (X_0 - 2^{2-n})^2| \leq 2^{3-n} |X_0|.
$$

Now we define

(5.7)
$$
\Psi_n = \sum_{i=0}^{n} 2^{K_i+1} C_3 |Z_i \cdot \phi'(Z_i)| 2^{-K_n},
$$

and rewrite (5.5) as

$$
\left| \log \frac{f_{X_n}(x \mid \mathcal{F}_n)}{f_{X_n}(w \mid \mathcal{F}_n)} \right| \leq \Psi_n \cdot 2^{K_n} \cdot |x - w| + \eta_n.
$$

We will need the following lemma to bound the crucial quantity Ψ_n.

LEMMA 5.7. *For a sufficiently large constant T, the expectation $\mathbb{E}[e^{\Psi_n 2^{-T}}]$ is uniformly bounded for all n.*

The proof of this lemma is deferred to Section 6. Henceforth, let T denote a constant that is sufficiently large for Lemma 5.7 to apply.

Finally, we are in a position to get a bound on $f_{X_n}(x \mid \mathcal{F}_n)$:

(5.8)
$$
f_{X_n}(x \mid \mathcal{F}_n) \leq \left(e^{\Psi_n 2^{K_n} |x-w| + \eta_n} \right) f_{X_n}(w \mid \mathcal{F}_n).
$$

Now we integrate (5.8) over an interval of length $\gamma = 2^{(-K_n - T)}$ with x at one end point. So $|x - w| \leq 2^{(-K_n - T)}$. Such an interval can be fit into I_n to the left or right of any x depending on where x is in the interval. Assuming without loss of generality that x is the left endpoint of the integration interval we compute that

$$
\int_x^{x+\gamma} f_{X_n}(x|\mathcal{F}_n)\, dw \leq \int_x^{x+\gamma} \left(e^{\Psi_n 2^{K_n} |x-w| + \eta_n} \right) f_{X_n}(w|\mathcal{F}_n)\, dw.
$$

We bound $|x - w|$ on the RHS by $\gamma = 2^{(-K_n - T)}$ to get

$$\gamma \cdot f_{X_n}(x|\mathcal{F}_n) \leq \int_x^{x+\gamma} \left(e^{\Psi_n 2^{K_n} 2^{(-K_n - T)} + \eta_n} \right) f_{X_n}(w|\mathcal{F}_n)\, dw$$

$$\leq \left(e^{\Psi_n 2^{-T} + \eta_n} \right) \cdot 1.$$

The last step follows since the density integrates out to 1. Hence,

$$(5.9) \qquad\qquad f_{X_n}(x|\mathcal{F}_n) \leq e^{\Psi_n 2^{-T} + \eta_n} 2^{K_n + T}.$$

This gives us a bound on the density of X_n in terms of K_n.

It now remains to bound the rate at which the K_n are growing. The following lemma shows that the K_n grow essentially at most linearly.

LEMMA 5.8.   *There exists a constant C such that*

$$\mathbb{P}(K_n - K_0 > C \cdot n) \to 0 \qquad as\ n \to \infty.$$

PROOF.   By construction, $K_{n+1} \geq K_n + 1$. In the case where $K_{n+1} > K_n + 1$, we can apply Lemma 5.2 and get that for $\ell \geq 2$,

$$\mathbb{P}(K_{n+1} - K_n \geq \ell | \mathcal{F}_n) \leq \mathbb{P}\left(|X_{n+1}| \leq 2^{2 - K_n - \ell} | \mathcal{F}_n \right)$$

$$= \mathbb{P}\left(|X_n - U_n| \leq 2^{2 - K_n - \ell} | \mathcal{F}_n \right).$$

This is because the control U_n must have been very close to X_n for K_{n+1} to be much larger than K_n. Then we calculate this probability by integrating out the density as

$$\mathbb{P}\left(|X_n - U_n| \leq 2^{2 - K_n - \ell} | \mathcal{F}_n \right) = \int_{U_n - 2^{2 - K_n - \ell}}^{U_n + 2^{2 - K_n - \ell}} f_{X_n}(t|\mathcal{F}_n)\, dt$$

$$\leq 2 \cdot 2^{2 - K_n - \ell} \left(\max_t f_{X_n}(t|\mathcal{F}_n) \right).$$

Combined with (5.9), this gives us that

$$\mathbb{P}(K_{n+1} - K_n \geq \ell | \mathcal{F}_n) \leq 2 \cdot 2^{2 - K_n - \ell} e^{\eta_n + \Psi_n 2^{-T}} 2^{K_n + T}$$

$$(5.10) \qquad\qquad\qquad = 2^{3 - \ell + T} \cdot e^{\eta_n + \Psi_n 2^{-T}}.$$

Write $D_n = K_{n+1} - K_n$, and let

$$\widetilde{K}_n = \sum_{i=0}^{n-1} (D_i - \mathbb{E}[D_i \mid \mathcal{F}_i]).$$

It is clear that $(\widetilde{K}_n)$ is a martingale with respect to $\mathcal{F}_n$. In addition, (5.10) yields that the conditional distribution of D_n given $\mathcal{F}_n$ is stochastically dominated by the distribution of

$$(5.11) \qquad\qquad G_n + \left(3 + T + \frac{|\eta_n + \Psi_n 2^{-T}|}{\log 2} \right),$$

where G_n is an independent geometric variable with mean 2.

By (5.6) and Lemma 5.7, both η_n and $\Psi_n 2^{-T}$ have bounded second moments, and so for some constant $\widetilde{C}$, we have

$$\mathbb{E}(D_i - \mathbb{E}[D_i \mid \mathcal{F}_i])^2 \leq \mathbb{E}D_i^2 \leq \widetilde{C}.$$

Summing over i, this implies that $\mathbb{E}[\widetilde{K}_n^2] \leq \widetilde{C}n$, and so

(5.12) $$\mathbb{P}(\widetilde{K}_n \geq n) \to 0 \qquad \text{as } n \to \infty.$$

We now turn our attention to terms of the form $\mathbb{E}[D_i \mid \mathcal{F}_i]$. Using (5.11) again, we get that

(5.13) $$\mathbb{E}[D_i \mid \mathcal{F}_i] \leq 5 + T + 2|\eta_i + \Psi_i 2^{-T}|.$$

Observe that from the definition of Ψ_i given in (5.7), we have

$$\sum_{i=0}^{n} \Psi_i \leq C_3 \sum_{i=0}^{n} \sum_{j=0}^{i} 2^{K_j - K_i + 1} |Z_j \cdot \phi'(Z_j)|$$

(5.14) $$= C_3 \sum_{j=0}^{n} \sum_{i=j}^{n} 2^{K_j - K_i + 1} |Z_j \cdot \phi'(Z_j)|$$

$$\leq 4C_3 \sum_{j=0}^{n} |Z_j \cdot \phi'(Z_j)|,$$

where in the last step we have used the fact that the K_i increase by at least 1 in each step, so that $K_i - K_j \geq i - j$. Then, applying the bound $|Z_j \cdot \phi'(Z_j)| \leq C_1 + C_2 \phi(Z_j)$ to (5.14) yields

$$\sum_{i=0}^{n} \Psi_i \leq 4C_3 C_1 n + 4C_3 C_2 \sum_{i=1}^{n} \phi(Z_i).$$

Summing (5.13) over i and applying the above bound gives

$$\sum_{i=1}^{n} \mathbb{E}[D_i \mid \mathcal{F}_i] \leq (5 + T)n + 2\sum_{i=1}^{n} |\eta_i| + 2^{1-T} \sum_{i=1}^{n} \Psi_i$$

$$\leq C_{D,1} \left(n + \sum_{i=1}^{n} |\eta_i| + \sum_{i=1}^{n} \phi(Z_i) \right)$$

for a constant $C_{D,1}$. Now, recalling (5.6), we see that the quantity

$$\sum_{i=1}^{n} |\eta_i| \leq 8|X_0|$$

has mean and variance bounded by a constant, which we call C_η. In addition, by Lemma 5.5, there exists another constant C_ϕ which upper bounds the mean and

variance of $\phi(Z_i)$. We conclude that

$$\mathbb{E}\left(\sum_{i=1}^{n} \mathbb{E}[D_i \mid \mathcal{F}_i]\right) \le C_{D,1}(1 + C_\eta + C_\phi \cdot n),$$

$$\mathrm{Var}\left(\sum_{i=1}^{n} \mathbb{E}[D_i \mid \mathcal{F}_i]\right) \le 2C_{D,1}^2 (C_\eta + C_\phi \cdot n).$$

It follows that

$$(5.15) \qquad \mathbb{P}\left(\sum_{i=1}^{n} \mathbb{E}[D_i \mid \mathcal{F}_i] \ge C_{D,2} \cdot n\right) \to 0 \qquad \text{as } n \to \infty,$$

where $C_{D,2} = C_{D,1}C_\phi + 1$. Finally, setting $C = C_{D,2} + 1$, we have

$$\mathbb{P}(K_n - K_0 > C \cdot n) = \mathbb{P}\left(\sum_{i=1}^{n} D_i > C \cdot n\right)$$

$$= \mathbb{P}\left(\widetilde{K}_n + \sum_{i=1}^{n} \mathbb{E}[D_i \mid \mathcal{F}_i] > (C_{D,2} + 1) \cdot n\right)$$

$$\le \mathbb{P}(\widetilde{K}_n > n) + \mathbb{P}\left(\sum_{i=1}^{n} \mathbb{E}[D_i \mid \mathcal{F}_i] \ge C_{D,2} \cdot n\right),$$

where the last expression goes to 0 as $n \to \infty$ by (5.12) and (5.15). $\qquad \square$

This bound on the growth of the K_n variables allows us to complete the proof of Theorem 5.1.

Let G_n denote the event that $K_n - K_0 > Cn$, and G_n^c its complement. Then we can cover the event of interest by three events, and get that

$$(5.16) \qquad \begin{aligned} &\mathbb{P}(|X_n| < a^{-n}M) \\ &\qquad \le \mathbb{P}(G_n) + \mathbb{P}(K_0 > n) + \mathbb{P}(|X_n| \le a^{-n}M, G_n^c, K_0 \le n). \end{aligned}$$

We evaluate the three terms one by one. For the first term in (5.16), we have $\mathbb{P}(G_n) = \mathbb{P}(K_n - K_0 > Cn) \to 0$ as $n \to \infty$ from Lemma 5.8.

The second term, $\mathbb{P}(K_0 > n)$, captures the case where the initial state X_0 might be very close to zero. However, eventually this advantage dies out for large enough n, since $\mathbb{P}(X_0 < 2^{-n}) \to 0$ as $n \to \infty$.

The last term in (5.16) remains. By the law of iterated expectation,

$$\mathbb{P}(|X_n| < a^{-n}M, G_n^c, K_0 \le n) = \mathbb{E}[\mathbb{P}(|X_n| < a^{-n}M, G_n^c, K_0 \le n \mid \mathcal{F}_n)].$$

We focus on the term conditioned on $\mathcal{F}_n$:

$$(5.17) \qquad \begin{aligned} &\mathbb{P}(|X_n| < a^{-n}M, G_n^c, K_0 \le n \mid \mathcal{F}_n) \\ &\qquad = \mathbb{E}[\mathbb{1}_{\{|X_n|<a^{-n}M\}}\mathbb{1}_{\{G_n^c\}}\mathbb{1}_{\{K_0\le n\}} \mid \mathcal{F}_n] \\ &\qquad = \mathbb{P}(|X_n| < a^{-n}M \mid \mathcal{F}_n) \cdot \mathbb{1}_{\{G_n^c\}}\mathbb{1}_{\{K_0\le n\}}. \end{aligned}$$

Now we can apply (5.9) to get

$$\mathbb{P}(|X_n| < a^{-n}M|\mathcal{F}_n) = \int_{-a^{-n}M}^{a^{-n}M} f_{X_n}(x|\mathcal{F}_n)\,dx$$

$$\leq \int_{-a^{-n}M}^{a^{-n}M} e^{\eta+\Psi_n 2^{-T}} 2^{K_n+T}\,dx$$

$$= 2Ma^{-n} \cdot e^{\eta_n+\Psi_n 2^{-T}} \cdot 2^{K_n+T}.$$

Then we can bound (5.17) as

$$\mathbb{P}(|X_n| < a^{-n}M|\mathcal{F}_n) \cdot \mathbb{1}_{\{G_n^c\}}\mathbb{1}_{\{K_0 \leq n\}} \leq 2Ma^{-n} \cdot e^{\eta_n+\Psi_n 2^{-T}} \cdot 2^{(C+1)n+T},$$

since $K_n \leq Cn + K_0$ and $K_0 \leq n$ implies $K_n \leq (C+1)n$. Taking expectations on both sides, we get

$$(5.18) \quad \mathbb{P}(|X_n| < a^{-n}M, G_n^c, K_0 \leq n) \leq 2Ma^{-n} \cdot 2^{(C+1)n+T} \cdot \mathbb{E}[e^{\eta_n}e^{\Psi_n 2^{-T}}].$$

By Lemma 5.7 and (5.6), the above expression (5.18) tends to 0 for $a > 2^{C+1}$.

Thus, all three probabilities in (5.16) converge to 0 as $n \to \infty$. Hence, if $a > 2^{C+1}$ then $\mathbb{P}(|X_n| < a^{-n}M) \to 0$ for all M. $\square$

6. Bounding the likelihood ratio.
Here we provide the proofs of two lemmas used to bound the term $|\log \frac{f_{X_n}(x|\mathcal{F}_n)}{f_{X_n}(w|\mathcal{F}_n)}|$.

6.1. *Proof of Lemma 5.6.* We take logarithms on both sides of (5.3) and apply the triangle inequality to get

$$(6.1)$$
$$\left|\log \frac{f_{X_n}(x \mid \mathcal{F}_n)}{f_{X_n}(w \mid \mathcal{F}_n)}\right|$$
$$\leq \left|\log \frac{f_{Y_n}(Y_n \mid X_n = x, \mathcal{F}_{n-1})}{f_{Y_n}(Y_n \mid X_n = w, \mathcal{F}_{n-1})}\right| + \left|\log \frac{f_{X_n}(x \mid \mathcal{F}_{n-1})}{f_{X_n}(w \mid \mathcal{F}_{n-1})}\right|.$$

The form of the density of Z gives

$$(6.2) \quad \left|\log \frac{f_{Y_n}(Y_n|X_n = x, \mathcal{F}_{n-1})}{f_{Y_n}(Y_n|X_n = w, \mathcal{F}_{n-1})}\right| = \left|\phi\left(\frac{Y_n}{x}\right) - \phi\left(\frac{Y_n}{w}\right)\right|.$$

We can use the derivatives of the functions to bound the two function differences above. Since $\frac{d}{dx}\phi(\frac{Y_n}{x}) = \frac{Y_n}{x^2}\phi'(\frac{Y_n}{x})$, we bound (6.2) as below. Since $X_n \in I_n - S_n$,

the maximizations are over $t \in I_n - S_n$,

$$(6.3) \qquad \left| \phi\left(\frac{Y_n}{x}\right) - \phi\left(\frac{Y_n}{w}\right) \right| \le \max_{t \in I_n - S_n} \left| \frac{Y_n}{t^2} \phi'\left(\frac{Y_n}{t}\right) \right| |x - w|.$$

For all $t \in I_n - S_n$, by Lemma 5.3, we have $\frac{1}{2} \le \frac{X_n}{t} \le 2$. Using this and $Y_n = Z_n X_n$, we get the following bound on (6.3):

$$(6.4) \qquad \begin{aligned} \left| \phi\left(\frac{Y_n}{x}\right) - \phi\left(\frac{Y_n}{w}\right) \right| &\le \max_{t \in I_n - S_n} \left| \frac{2Z_n}{t} \phi'\left(\frac{Z_n X_n}{t}\right) \right| |x - w| \\ &\le \max_{t \in I_n - S_n} 2C_3 \left| \frac{Z_n}{t} \phi'(Z_n) \right| |x - w|. \end{aligned}$$

(6.4) follows from the doubling property of $\phi'(\cdot)$, since $\frac{Z_n X_n}{t}$ and Z_n are within a factor of two from each other by Lemma 5.3. Now note that

$$\max_{t \in I_n - S_n} \frac{1}{|t|} \le 2^{K_n}.$$

Applying this to the bound from (6.4), we get

$$\left| \phi\left(\frac{Y_n}{x}\right) - \phi\left(\frac{Y_n}{w}\right) \right| \le 2^{K_n+1} C_3 |Z_n \cdot \phi'(Z_n)| |x - w|.$$

This now gives a bound for (6.1) as below:

$$(6.5) \quad \left| \log \frac{f_{X_n}(x \mid \mathcal{F}_n)}{f_{X_n}(w \mid \mathcal{F}_n)} \right| \le 2^{K_n+1} C_3 |Z_n \cdot \phi'(Z_n)| |x - w| + \left| \log \frac{f_{X_n}(x \mid \mathcal{F}_{n-1})}{f_{X_n}(w \mid \mathcal{F}_{n-1})} \right|.$$

6.2. *Proof of Lemma 5.7.* Recall that our goal is to estimate the quantity

$$\Psi_n = \sum_{i=0}^{n} 2^{K_i+1} C_3 |Z_i \cdot \phi'(Z_i)| 2^{-K_n}.$$

Since the K_i's must increase by at least one in each step, we have $K_n - K_i \ge n - i$, and so

$$(6.6) \qquad \begin{aligned} \Psi_n &\le \sum_{i=0}^{n} 2^{1+i-n} C_3 |Z_i \cdot \phi'(Z_i)| \le \sum_{i=0}^{n} 2^{1+i-n} C_3 (C_1 + C_2 \phi(Z_i)) \\ &= 4C_3 C_1 + C_3 C_2 \sum_{i=0}^{n} 2^{1+i-n} \phi(Z_i), \end{aligned}$$

where we have also used the assumption $|z \cdot \phi'(z)| \le C_1 + C_2 \phi(z)$.

Let $\delta' = \delta/(1 + \delta)$ as in Lemma 5.4. Consider any $\theta < \delta'/2$. Applying Lemma 5.4, we have for each i that

$$
\begin{aligned}
\mathbb{E}\big[e^{\theta\phi(Z_i)}\big] &= \int_{-\infty}^{\infty} \theta e^{\theta t} \cdot \mathbb{P}(\phi(Z_i) \geq t)\, dt \\
&\leq \int_{-\infty}^{0} \theta e^{\theta t}\, dt + \frac{2\theta}{\delta'} \int_{0}^{\infty} e^{(\theta - \delta')t}\, dt \\
&= 1 + \frac{2\theta}{\delta'(\delta' - \theta)} \leq 1 + \frac{4\theta}{\delta'^2}.
\end{aligned}
$$

(6.7)

Now choose T large enough so that $2^{1-T} C_2 C_3 < \delta'/2$. We can then apply (6.7) to each term in (6.6) by taking $\theta = 2^{-T} C_3 C_2 \cdot 2^{1+i-n}$. This yields

$$
\mathbb{E}\big[\exp(2^{-T} C_3 C_2 \cdot 2^{1+i-n}\phi(Z_i))\big] \leq 1 + C \cdot 2^{i-n}
$$

for a constant C not depending on n. We then have

$$
\begin{aligned}
\mathbb{E}\big[e^{2^{-T}\Psi_n}\big] &\leq \mathbb{E}\left[\exp\left(4C_3 C_1 + C_3 C_2 \sum_{i=0}^{n} 2^{1+i-n}\phi(Z_i)\right)\right] \\
&\leq e^{4C_3 C_1} \prod_{i=0}^{n} (1 + C \cdot 2^{i-n}) \\
&\leq e^{4C_3 C_1} \prod_{i=0}^{\infty} (1 + C \cdot 2^{-i}),
\end{aligned}
$$

which is a (finite) constant not depending on n.

7. Conclusion. This paper provides a first proof-of-concept converse for a control system observed over continuous multiplicative noise. However, there is an exponential gap between the scaling behavior of the achievable strategy and the converse.

We note that if the system $\mathcal{S}_a$ in (1.1) is restricted to using linear control strategies, then its performance limit is the same as that of a system with the same multiplicative actuation noise (i.e., the control U_n is multiplied by a random scaling factor) but perfect observations (as in [29]). Previous work has shown how to compute the control capacity for systems with multiplicative noise on the actuation channel [27, 29]. However, computing the control capacity of the system $\mathcal{S}_a$, that is, computing tight upper and lower bounds on the system growth factor a, remains open.

REFERENCES

[1] ATHANS, M., KU, R. and GERSHWIN, S. B. (1977). The uncertainty threshold principle: Some fundamental limitations of optimal decision making under dynamic uncertainty. *IEEE Trans. Automat. Control* **AC-22** 491–495. MR0497220

[2] BAO, L., SKOGLUND, M. and JOHANSSON, K. H. (2011). Iterative encoder-controller design for feedback control over noisy channels. *IEEE Trans. Automat. Control* **56** 265–278. MR2761095

[3] BRASLAVSKY, J. H., MIDDLETON, R. H. and FREUDENBERG, J. S. (2007). Feedback stabilization over signal-to-noise ratio constrained channels. *IEEE Trans. Automat. Control* **52** 1391–1403. MR2341744

[4] CHIUSO, A., LAURENTI, N., SCHENATO, L. and ZANELLA, A. (2014). LQG-like control of scalar systems over communication channels: The role of data losses, delays and SNR limitations. *Automatica J. IFAC* **50** 3155–3163. MR3284150

[5] DEY, S., CHIUSO, A. and SCHENATO, L. (2017). Feedback control over lossy SNR-limited channels: Linear encoder-decoder-controller design. *IEEE Trans. Automat. Control* **62** 3054–3061. MR3660592

[6] DING, J., PERES, Y. and RANADE, G. (2016). A tiger by the tail: When multiplicative noise stymies control. In *IEEE International Symposium on Information Theory*.

[7] ELIA, N. (2005). Remote stabilization over fading channels. *Systems Control Lett.* **54** 237–249. MR2115541

[8] FANG, S., CHEN, J. and ISHII, H. (2016). *Towards Integrating Control and Information Theories. Lecture Notes in Control and Information Sciences* **465**. Springer, Cham. MR3822315

[9] FREUDENBERG, J. S., MIDDLETON, R. H. and BRASLAVSKY, J. H. (2011). Minimum variance control over a Gaussian communication channel. *IEEE Trans. Automat. Control* **56** 1751–1765. MR2856793

[10] GARONE, E., SINOPOLI, B., GOLDSMITH, A. and CASAVOLA, A. (2012). LQG control for MIMO systems over multiple erasure channels with perfect acknowledgment. *IEEE Trans. Automat. Control* **57** 450–456. MR2918748

[11] HESPANHA, J. P., NAGHSHTABRIZI, P. and XU, Y. (2007). A survey of recent results in networked control systems. *Proc. IEEE* **95** 138–162.

[12] KOSTINA, V. and HASSIBI, B. (2016). Rate cost tradeoffs in control. part 1: Lower bounds. arXiv:1612.02126 [cs.IT].

[13] KOSTINA, V., PERES, Y., RÁCZ, M. and RANADE, G. (2016). Rate-limited control of systems with uncertain gain. In *Allerton Conference on Communication, Control, and Computing*.

[14] LAPIDOTH, A. and MOSER, S. M. (2003). Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels. *IEEE Trans. Inform. Theory* **49** 2426–2467. MR2027996

[15] MARTINS, N. C., DAHLEH, M. A. and ELIA, N. (2006). Feedback stabilization of uncertain systems in the presence of a direct link. *IEEE Trans. Automat. Control* **51** 438–447. MR2205681

[16] MATVEEV, A. S. and SAVKIN, A. V. (2004). The problem of LQG optimal control via a limited capacity communication channel. *Systems Control Lett.* **53** 51–64. MR2077188

[17] MATVEEV, A. S. and SAVKIN, A. V. (2009). *Estimation and Control over Communication Networks. Control Engineering.* Birkhäuser, Boston, MA. MR2498757

[18] MEYR, H., MOENECLAEY, M. and FECHTEL, S. A. (1998). *Digital Communication Receivers: Synchronization, Channel Estimation and Signal Processing.* Wiley, New York.

[19] MINERO, P., COVIELLO, L. and FRANCESCHETTI, M. (2013). Stabilization over Markov feedback channels: The general case. *IEEE Trans. Automat. Control* **58** 349–362. MR3023928

[20] MINERO, P., FRANCESCHETTI, M., DEY, S. and NAIR, G. N. (2009). Data rate theorem for stabilization over time-varying feedback channels. *IEEE Trans. Automat. Control* **54** 243–255. MR2490804

[21] NAIR, G. N. and EVANS, R. J. (2004). Stabilizability of stochastic linear systems with finite feedback data rates. *SIAM J. Control Optim.* **43** 413–436. MR2086167

[22] NAIR, G. N., FAGNANI, F., ZAMPIERI, S. and EVANS, R. J. (2007). Feedback control under data rate constraints: An overview. *Proc. IEEE* **95** 108–137.

[23] OKANO, K. and ISHII, H. (2014). Minimum data rate for stabilization of linear systems with parametric uncertainties. arXiv preprint, arXiv:1405.5932.

[24] PARK, S., RANADE, G. and SAHAI, A. (2012). Carry-free models and beyond. In *IEEE International Symposium on Information Theory* 1927–1931. IEEE.

[25] PARK, S. and SAHAI, A. (2011). Intermittent Kalman filtering: Eigenvalue cycles and nonuniform sampling. In *American Control Conference* (*ACC*) 3692–3697.

[26] RAJASEKARAN, P. K., SATYANARAYANA, N. and SRINATH, M. D. (1971). Optimum linear estimation of stochastic signals in the presence of multiplicative noise. *IEEE Trans. Aerosp. Electron. Syst.* **AES-7** 462–468. MR0278815

[27] RANADE, G. (2014). Active systems with uncertain parameters: An information-theoretic perspective. Ph.D. thesis, Univ. California, Berkeley.

[28] RANADE, G. and SAHAI, A. (2013). Non-coherence in estimation and control. In *51st Annual Allerton Conference on Communication, Control, and Computing*.

[29] RANADE, G. and SAHAI, A. (2015). Control capacity. In *IEEE International Symposium on Information Theory*. IEEE, New York.

[30] RANADE, G. and SAHAI, A. (2017). Control capacity. arXiv:1701.04187 [cs.IT].

[31] SAHAI, A. and MITTER, S. (2006). The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link. I. Scalar systems. *IEEE Trans. Inform. Theory* **52** 3369–3395. MR2242355

[32] SCHENATO, L., SINOPOLI, B., FRANCESCHETTI, M., POOLLA, K. and SASTRY, S. S. (2007). Foundations of control and estimation over lossy networks. *Proc. IEEE* **95** 163–187.

[33] SILVA, E. I., GOODWIN, G. C. and QUEVEDO, D. E. (2010). Control system design subject to SNR constraints. *Automatica J. IFAC* **46** 428–436. MR2877090

[34] SILVA, E. I. and PULGAR, S. A. (2013). Performance limitations for single-input LTI plants controlled over SNR constrained channels with feedback. *Automatica J. IFAC* **49** 540–547. MR3004722

[35] SINOPOLI, B., SCHENATO, L., FRANCESCHETTI, M., POOLLA, K., JORDAN, M. I. and SASTRY, S. S. (2004). Kalman filtering with intermittent observations. *IEEE Trans. Automat. Control* **49** 1453–1464. MR2086911

[36] TATIKONDA, S. and MITTER, S. (2004). Control under communication constraints. *IEEE Trans. Automat. Control* **49** 1056–1068. MR2071934

[37] TUGNAIT, J. K. (1981). Stability of optimum linear estimators of stochastic signals in white multiplicative noise. *IEEE Trans. Automat. Control* **26** 757–761. MR0630815

[38] WONG, W. S. and BROCKETT, R. W. (1997). Systems with finite communication bandwidth constraints. I. State estimation problems. *IEEE Trans. Automat. Control* **42** 1294–1299. MR1470572

[39] XIAO, N., XIE, L. and QIU, L. (2012). Feedback stabilization of discrete-time networked systems over fading channels. *IEEE Trans. Automat. Control* **57** 2176–2189. MR2968776

[40] XU, L., MO, Y., XIE, L. and XIAO, N. (2017). Mean square stabilization of linear discrete-time systems over power-constrained fading channels. *IEEE Trans. Automat. Control* **62** 6505–6512. MR3743533

[41] YOU, K. and XIE, L. (2011). Minimum data rate for mean square stabilizability of linear systems with Markovian packet losses. *IEEE Trans. Automat. Control* **56** 772–785. MR2814928

[42] YÜKSEL, S. and BAŞAR, T. (2013). *Stochastic Networked Control Systems: Stabilization and Optimization Under Information Constraints*. Birkhäuser/Springer, New York. MR3088779

J. DING
DEPARTMENT OF STATISTICS
UNIVERSITY OF PENNSYLVANIA
457 JON M. HUNTSMAN HALL
3730 WALNUT STREET
PHILADELPHIA, PENNSYLVANIA 19104
USA
E-MAIL: dingjian@wharton.upenn.edu

Y. PERES
G. RANADE
MICROSOFT RESEARCH AI
BUILDING 99, ONE MICROSOFT WAY
REDMOND, WASHINGTON 98052
USA
E-MAIL: peres@microsoft.com
        gireeja@eecs.berkeley.edu

A. ZHAI
DEPARTMENT OF MATHEMATICS
STANFORD UNIVERSITY
450 SERRA MALL, BUILDING 380
STANFORD, CALIFORNIA 94305
USA
E-MAIL: azhai@stanford.edu