IEEE Control Systems Letters paper presented at
2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Learning to Control under Communication Constraints

Shubham Aggarwal, Raj Kiriti Velicheti, Tamer Başar, *Life Fellow, IEEE*

***Abstract*— How to effectively communicate over wireless networks characterized by link failures is central to understanding the fundamental limits in the performance of a networked control system. In this paper, we study the online remote control of linear-quadratic Gaussian systems over unreliable wireless channels (with random packet drops), where the controller is *a priori* oblivious to the cost parameters. We first reformulate the problem using a semi-definite program and consequently compute a stabilizing policy from its solution. We then derive a $\mathcal{O}(\sqrt{T})$ regret bound (against a best offline policy in hindsight) for a projected online gradient algorithm, where T is the length of the horizon of interest. In the process, we introduce finite-time notions of the classical mean-square stability, which may be of independent interest. Finally, we provide a numerical example to validate the theoretical results, demonstrating the limitations induced by lossy communication on the control performance.**

I. Introduction

With the rapid development of sensing and computing technologies, networked control systems (NCSs) find applications in diverse areas such as internet of things, power grid management, and autonomous vehicles [1]–[3], to name a few. The unifying theme in all these applications is the distributed control of spatially located systems over wireless networks. While networked control allows for information sharing and control via decentralised task execution, it comes with challenges of its own such as transmission delays, bandwidth constraints and link failures in the network which might cause packet drops. NCSs with packet drops and known fixed costs have been well studied in the literature in the context of optimal estimation and control of linear quadratic Gaussian (LQG) systems with random packet drops [2], [4], [5], stabilization of deterministic linear systems with bounded packet losses [6], and $\mathcal{H}_2$ optimal control [7] (see also the references therein), where it is shown that lossy communication restricts the degree of open-loop instability that can be stabilized by a feedback controller, which may not even exist beyond a certain dropout threshold. The complexity of the problem is further elevated when the system under control has varying unknown costs which change, possibly adversarially, due to changes in the environment.

Most of aforementioned references either compute an optimal controller/estimator by using a Riccati-type equation or otherwise solve a linear matrix inequality to provide sufficient conditions for stabilization. Solving the same, however, requires an *a priori* knowledge of the system as well as the involved cost parameters. In this paper, we are concerned with the optimal control of a NCS constituting an unreliable wireless channel with random packet drops, and unknown costs, a prototype for which is shown in Fig. 1. Such scenarios occur in many natural settings such as uplink power control in CDMA networks, control of power grids with costs depending on market auctions (which in turn depend on a number of other uncontrollable factors), manufacturing systems, and automated traffic routing [8], [9], to name a few. The commonality in all these applications is that due to congested and uncertain environments, the cost parameters can only be inferred once a policy has been implemented. Thus, another concern of this work is to circumvent the requirement of these parameters and *learn* a policy which adapts to the (possibly time-varying) parameters, since obliviousness to these precludes the offline computation of an optimal control policy. We utilize techniques from online convex optimisation (OCO) [10], [11] to address this challenge and design an algorithm that suggests a control policy which (possibly) depends on the history of observations. We utilize the standard notion of regret to measure the performance of our algorithm against a best policy in hindsight from a suitable policy class. The objective is to achieve a sublinear (in time horizon) regret bound, which entails that the controller *learns* the aforementioned offline policy in the long run.

Thus, motivated by the challenges of lossy communication and recent developments in online control, we present the main contribution of the work as follows: (a) We pose the problem of optimal control in NCS with packet loss over controller-to-plant wireless link as a semi-definite program (SDP) in steady state distribution, (b) utilizing this convex reformulation, since the costs are unknown, we show that a simple online gradient descent can achieve a regret of the order $\mathcal{O}(\sqrt{T})$, and (c) in the process, we extend the notion of strong stability as in [12] to the mean-square sense, which is of independent interest while dealing with stochasticity in control systems. Finally, we also provide discussions on (i) the case where the plant-to-controller link is also prone to packet drops, and (ii) the inclusion of unknown dynamics alongside unknown cost parameters, both of which augment the presented analysis.

Control of LQG in an online learning framework has been a topic of recent interest. The adaptive control of LQG systems with full and partial state feedback has been studied in [13], [14], and with safety constraints in [15]. The authors in [12] consider the LQG problem with unknown costs and known fixed dynamics, and derive sublinear regret bounds using an SDP formulation [16]. The same has been extended to unknown dynamics but known costs in [17], and distributed control in [18]. Finally, although, slightly orthogonal to, but yet useful in the present context, we mention the setting of OCO-with-memory [19], which has been applied to unknown systems with bounded non-stochastic disturbances

and general convex costs [20]. It is also worth noting that techniques in this work are similar in spirit to those developed in [12], but the additional stochasticity introduced by packet loss requires careful analysis and different tools to achieve a sublinear regret.

This paper is organized as follows. We first formulate the problem (Sec. II) and discuss useful policy characterizations (Sec. III). Then, we present an SDP formulation for the LQ problem (Sec. IV) and perform a regret analysis (Sec. V). Finally, we present some numerical results (Sec VI) and conclude the paper with some useful discussions (Sec. VII), and two appendices.

Notations: The trace of a matrix is denoted by $Tr(\cdot)$. I denotes the identity matrix of appropriate dimensions and $\|\cdot\|$ denotes the Euclidean norm for vectors and induced 2-norm for matrices. For two symmetric matrices A and B of the same dimensions, the notation $A \succeq B$ (resp. $A \succ B$) denotes that the matrix $A-B$ is positive semi-definite (resp. positive definite). $\text{Diag}(A, B)$ denotes a block diagonal matrix with block-diagonal entries A and B.

II. Problem Formulation

Consider a networked control system as in Fig. 1 constituting a plant ($\mathcal{P}$) and a remotely located controller ($\mathcal{C}$). Dynamics of the plant $\mathcal{P}$ evolve according to a stochastic linear difference equation:

$$X_{k+1} = AX_k + \alpha_k BU_k + W_k,\ k \geq 0, \tag{1}$$

where $X_k \in \mathbb{R}^n$ and $U_k \in \mathbb{R}^m$ are the state and control input to the plant, respectively, at instant k. The term $W_k \in \mathbb{R}^n$ denotes independent Gaussian noise with zero mean and positive definite covariance Ω. The matrices A and B are time-invariant and have suitable dimensions. The state X_k is

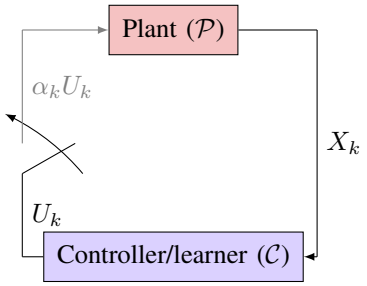

Fig. 1: Closed-loop system with random packet dropouts

sent to the controller over an ideal wireless forward channel while the actuation command from the controller to the plant is sent over an unreliable wireless link, which loses these packets according to a Bernoulli-distributed signal α_k, with probability p, as:

$$\alpha_k = \begin{cases} 1, & \text{w.p. } 1-p \\ 0, & \text{w.p. } p \end{cases}. \tag{2}$$

We assume that W_k are independent of α_k, for all k, and let the initial state X_0 to be 0, without any loss of generality. We also define $\bar{p} := 1-p$ for brevity.

Next, the actions $U_1, U_2, \cdots$, are chosen according to a policy $\mu : \mathbb{R}^n \to \mathbb{P}(\mathbb{R}^m)$, which is a mapping from states to distribution over actions. The cost of following this policy is then given as

$$J_T(\mu) := \mathbb{E}\left[\sum_{k=1}^{T} X_k^\top Q_k X_k + \alpha_k U_k^\top R_k U_k\right], \tag{3}$$

where we assume that $Q_k \geq 0$, $R_k > 0$, and for some $\zeta > 0$, $Tr(Q_k), Tr(R_k) \leq \zeta$, for all k. The multiplicative term α_k denotes that the control cost is incurred only when the control is actually applied to the plant. Finally, the expectation is taken with respect to the random noise, packet drop probability, and the (possibly) randomized policy.

Next, we assume that for any T, the matrices $\{Q_k\}_{k=1}^T$ and $\{R_k\}_{k=1}^T$ are adversarially selected *a priori* and become known to the controller *only after an action has been chosen using some policy* at instant k. In addition to the applications listed in the introduction, another motivating perspective for this setting is explained as follows. Consider cost (3) with Q_k, R_k unknown. Then, we can re-express $\text{Diag}(Q_k, R_k) = (C_k^\top\ D_k^\top)^\top (C_k\ D_k)$, with $Q_k = C_k^\top C_k$, $R_k = D_k^\top D_k$ and $C_k^\top D_k = 0$, without loss of generality. Consequently, we can augment the state dynamics in (1) with an auxiliary controlled output $Z_k = C_k X_k + D_k U_k$. The adversarial control of the costs thus can be interpreted as directly influencing the output coefficients C_k and D_k. The more general case with A, B also unknown, is discussed in the conclusion section.

Note that the obliviousness of the controller to the cost matrices prohibits offline computation using a Riccati-type equation as in [2], [4]. Hence, the objective here is to design a control algorithm $\mathcal{C}$ (for online policy computation) which maps state x_t and previous cost matrices $\{Q_k\}_{k=1}^{t-1}, \{R_k\}_{k=1}^{t-1}$ to a control u_t at time t such that it minimizes its regret ($\mathcal{R}_T(\mathcal{C})$), against a set of benchmark policies in hindsight. More formally, the problem is introduced as follows.

Problem 1.

$$\min \mathcal{R}_T(\mathcal{C})$$
$$\textit{s.t. } (1) \textit{ holds},$$

where $\mathcal{R}_T(\mathcal{C}) := \max_{\{Q_k\}_{k=1}^T \{R_k\}_{k=1}^T} [J_T(\mathcal{C}) - \min_{\mu \in \mathcal{M}} J_T(\mu)]$ *and* $\mathcal{M}$ *is the set of benchmark policies, defined in the next section.*

III. The Benchmark Policy Set $\mathcal{M}$

In this section, we let $\mathcal{M}$ be the set of linear stationary mean-square strongly stable (MSSS) policies, and present the following two-part definition of mean-square strong stability, which is the quantitative analogue of the definition of mean-square stability (MSS) [21].

Definition 1 (Mean-square strong stability)**.** *Let* $\kappa > 0, 0 < \gamma_1, \gamma_2 < 1$. *Then,*

1) *a stationary policy* K *is* $(\kappa, \gamma_1, \gamma_2)$*–MSSS, if there exist matrices* Z, L_1, L_2 *such that*
   (a) $\|Z\|\|Z^{-1}\| \leq \kappa$, $\|K\| \leq \kappa$,
   (b) $A + BK = ZL_1Z^{-1}$, $\|L_1\| \leq 1 - \gamma_1$,

(c) $A = ZL_2Z^{-1}$, $\|L_2\| \leq \frac{1-\gamma_2}{\sqrt{p}}$,

(d) $\upsilon(p,\gamma_1,\gamma_2) := \bar{p}(1-\gamma_1)^2 + (1-\gamma_2)^2 < 1$.

2) *a non-stationary policy* $\{K_k\}_{k=1}^T$ *is* $(\kappa,\gamma_1,\gamma_2)$*–MSSS, if there exist sequences of matrices* $\{Z_k\}_{k=1}^T$,$\{L_k^{(1)}\}_{k=1}^T$ *and* $\{L_k^{(2)}\}_{k=1}^T$ *with* $A+BK_k = Z_kL_k^{(1)}Z_k^{-1}$ *and* $A = Z_kL_k^{(2)}Z_k^{-1}$ *satisfying the following:*

(a') $\|Z_k\| \leq \beta_1, \|Z_k^{-1}\| \leq \frac{1}{\beta_2}$, *with* $\kappa = \frac{\beta_1}{\beta_2}$, & $\|K_k\| \leq \kappa$,

(b') $\|L_k^{(1)}\| \leq 1-\gamma_1$, $\|L_k^{(2)}\| \leq \frac{1-\gamma_2}{\sqrt{p}}$

(c') $\|Z_{k+1}^{-1}Z_k\|^2 \leq 1 + \frac{1-\upsilon(p,\gamma_1,\gamma_2)}{2}$

(d') $\upsilon(p,\gamma_1,\gamma_2) < 1$.

With the above definition, the following two lemmas address convergence of the state covariance under a MSSS policy.

Lemma 1. *Let* Σ_k *and* Σ *be the state covariances at instant* k *and at steady state, respectively, and* $\bar{X}_k := \mathbb{E}[X_kX_k^\top]$ *with* $\mathbb{E}[X_0] = 0$*, and* $\Sigma_0 \succeq 0$*, under a given a stationary randomized feedback policy* $\mu(X)$ *with* $\mathbb{E}[\mu(X)] := KX$ *and* $\Sigma^c := Cov(\mu(X)|X) < \infty$*, which is* $(\kappa,\gamma_1,\gamma_2)$*-MSSS. Then, we have that*

$$\|\Sigma_k - \Sigma\| \leq \kappa^2\left[\upsilon(p,\gamma_1,\gamma_2)\right]^k\|\Sigma_0 - \Sigma\|.$$

Proof. Consider the following:

$$\bar{X}_k = pA\bar{X}_{k-1}A^\top + \bar{p}(A+BK)\bar{X}_{k-1}(A+BK)^\top + \bar{p}B\Sigma^cB^\top + \Omega \tag{4a}$$

$$\Sigma = pA\Sigma A^\top + \bar{p}(A+BK)\Sigma(A+BK)^\top + \bar{p}B\Sigma^cB^\top + \Omega. \tag{4b}$$

Then, subtracting (4b) from (4a), taking norms, and noting that $\|\Sigma_k - \Sigma\| \leq \|\bar{X}_k - \Sigma\|$, we get

$$\begin{aligned}\|\Sigma_k - \Sigma\| &\leq \|pA(\bar{X}_{k-1}-\Sigma)A^\top + \\ &\qquad \bar{p}(A+BK)(\bar{X}_{k-1}-\Sigma)(A+BK)^\top\|\|\Sigma_0-\Sigma\| \\ &\leq \kappa^2\sum_{r=0}^{k}\binom{k}{r}\left\{\sqrt{\bar{p}}(1-\gamma_1)\right\}^{2r}(1-\gamma_2)^{2(k-r)}\|\Sigma_0-\Sigma\| \\ &= \kappa^2[\upsilon(p,\gamma_1,\gamma_2)]^k\|\Sigma_0-\Sigma\|,\end{aligned}$$

where the last inequality follows from Definition 1 and the last equality follows using the identity $(a+b)^n = \sum_{r=0}^n\binom{n}{r}a^rb^{n-r}$. Finally, using Definition 1 (d), we get $\|\Sigma_k - \Sigma\| \to 0$ geometrically fast as $k \to \infty$. □

Next, we state a similar result for geometric convergence using non-stationary policies, which can be proved along similar lines as in the proof of Lemma 1 above.

Lemma 2. *Let* $\tilde{\Sigma}_k$ *be the state covariance,* $\bar{X}_k := \mathbb{E}[X_kX_k^\top]$ *at instant* k *with* $\mathbb{E}[X_0] = 0$*, and* $\tilde{\Sigma}_1 \succeq 0$*, under a non-stationary randomized policy* $\mu_k(X)$ *with* $\mathbb{E}[\mu_k(X)] := K_kX$ *and* $\Sigma_k^c := Cov(\mu_k(X)|X) < \infty$*, which is* $(\kappa,\gamma_1,\gamma_2)$*-MSSS. Then, for the non-stationary steady-state covariance* Σ_k'*, if it holds that* $\|\Sigma_k' - \Sigma_{k-1}'\| \leq \eta$, $\forall k$ *and some* $\eta > 0$*, we have that*

$$\|\tilde{\Sigma}_k - \Sigma_k'\| \leq \kappa^2\left[\upsilon(1+\frac{1-\upsilon}{2})\right]^k\|\tilde{\Sigma}_1 - \Sigma_1'\| + \frac{2\kappa^2\eta}{\upsilon^2-3\upsilon+2}.$$

We note that since the cost matrices Q_k, R_k are unknown to the controller in advance, offline minimization of (3) is not possible. This leads to an online policy computation problem, for which we first consider an SDP relaxation [12] of (3), and consequently employ techniques from the OCO literature to derive regret bounds when compared to the benchmark set of MSSS policies for Problem 1.

IV. Policy Computation using SDP

In this section, we express the LQ problem (1)-(3) via a relaxed SDP and consequently generate a linear MSSS policy from its solution. To this end, let us rewrite the cost (3) as:

$$\begin{aligned}J_T(\Sigma^\mu) &= \mathbb{E}\left[\sum_{k=1}^T X_k^\top Q_kX_k + \alpha_kU_k^\top R_kU_k\right] \\ &= \sum_{k=1}^T \bar{p}Tr\left(\text{Diag}(Q_k,R_k)\Sigma^\mu\right) + \sum_{k=1}^T p\ Tr\left(\text{Diag}(Q_k,0)\Sigma^\mu\right),\end{aligned}$$

where $\Sigma^\mu := \mathbb{E}\begin{bmatrix}X_kX_k^\top & X_kU_k^\top \\ U_kX_k^\top & U_kU_k^\top\end{bmatrix}$ and $U_k = \mu(X_k)$. Further, let Σ_{XX}^K be the steady-state covariance of the state X_k, under a linear feedback policy $\mu(X) = KX$, such that K is a MSS policy. Then, analogous to Σ^μ, we define

$$\Sigma^K := \begin{bmatrix}\Sigma_{XX}^K & \Sigma_{XX}^KK^\top \\ K\Sigma_{XX}^K & K\Sigma_{XX}^KK^\top\end{bmatrix}. \tag{5}$$

where the superscript K denotes the effect of policy K. With the above definitions, we define the SDP problem as follows.

Problem 2 (Relaxed SDP Problem)**.** *Given a tuple* (A,B,Q,R,Ω,p)*, the relaxed SDP corresponding to* (1)-(3) *is given as*

$$\min\ J(\Sigma^{sdp}) := \bar{p}Tr\left(\textit{Diag}(Q,R)\Sigma^{sdp}\right) + pTr\left(Q\Sigma_{XX}^{sdp}\right) \tag{6a}$$

$$\textit{s.t. } \Sigma_{XX}^{sdp} = pA\Sigma_{XX}^{sdp}A^\top + \bar{p}(A\ B)\Sigma^{sdp}(A\ B)^\top + \Omega, \tag{6b}$$

$$Tr(\Sigma^{sdp}) \leq \sigma,\ \Sigma^{sdp} \succeq 0, \tag{6c}$$

where $\Sigma^{sdp} := \begin{bmatrix}\Sigma_{XX}^{sdp} & \Sigma_{XU}^{sdp} \\ (\Sigma_{XU}^{sdp})^\top & \Sigma_{UU}^{sdp}\end{bmatrix} \in \mathbb{R}^{\ell\times\ell},\ \ell = n+m.$

To avoid inconsistent solutions, i.e., the case where the feasible set is empty, we invoke the following condition on system controllability [12] and cost detectability.

Assumption 1. *The pair (A,B) is* (k,κ)*–strongly controllable and the pair* $(A, Q_k^{1/2})$ *is detectable for all* k*.*

We note that while the strong controllability assumption entails a finite cost with explicit cost bounds, the detectability condition is necessary for the existence of an optimal stabilizing policy. Further, given an MSS policy μ satisfying $\mathbb{E}[\|X\|^2 + \|U\|^2] \leq \sigma$, the solution $\Sigma^{sdp} = \Sigma^\mu$ is feasible for Problem 2. Consequently, Problem 2 is indeed a relaxation of the original LQ problem (1)-(3).

Now, we generate a linear policy from the solution to Problem 2 (Σ^{sdp}), which will be shown in the sequel to be (a) MSS with bounded cost, and (b) MSSS. The existence

of Σ^{sdp} is guaranteed by the strong controllability of (A, B) in Assumption 1. Then, we introduce a linear policy as

$$K^{sdp} := K^{sdp}(\Sigma^{sdp}) = (\Sigma^{sdp}_{XU})^\top (\Sigma^{sdp}_{XX})^{-1}, \qquad (7)$$

which is well defined since $\Omega \succ 0$ and (6b). Next, we show that K^{sdp} is MSS with bounded cost.

Proposition 1 (Feasibility of Σ^K)**.** *Suppose that Assumption 1 holds and let $K := K^{sdp}$ be generated using a solution Σ^{sdp} to Problem 2. Then, the policy $\mu(X) = KX$ is MSS with $J(K) = J(\Sigma^K) \leq J(\Sigma^{sdp})$.*

Proof. Define $\Sigma' := \begin{bmatrix} \Sigma^{sdp}_{XX} & \Sigma^{sdp}_{XU} \\ (\Sigma^{sdp}_{XU})^\top & \Sigma^{sdp}_{UU} \end{bmatrix}$, and $\Sigma'' :=$ Diag$(0, \Sigma^{sdp}_{UU} - (\Sigma^{sdp}_{XU})^\top (\Sigma^{sdp}_{XX})^{-1} \Sigma^{sdp}_{XU})$. Then, $\Sigma^{sdp} = \Sigma' + \Sigma''$ follows by substituting (7) in the definition of Σ^{sdp} in Problem 2. Then, we observe that the $(2,2)^{th}$ entry in Σ'' is positive-definite since it is the Schur complement of Σ^{sdp}, which is positive semi-definite by using (6c). Hence, $\Sigma^{sdp} \succeq \Sigma'$.

Next, from (6b), we have that

$$\begin{aligned} \Sigma^{sdp}_{XX} &\succ pA\Sigma^{sdp}_{XX}A^\top + \bar{p}\begin{pmatrix} A & B \end{pmatrix} \Sigma^{sdp} \begin{pmatrix} A & B \end{pmatrix}^\top \\ &\succeq pA\Sigma^{sdp}_{XX}A^\top + \bar{p}\begin{pmatrix} A & B \end{pmatrix} \Sigma' \begin{pmatrix} A & B \end{pmatrix}^\top \\ &= pA\Sigma^{sdp}_{XX}A^\top + \bar{p}(A + BK)\Sigma^{sdp}_{XX}(A + BK)^\top, \end{aligned}$$

where the first inequality follows since $\Omega \succ 0$, and the second one follows since $\Sigma^{sdp} \succeq \Sigma'$, as proved above. Using [21, Theorem 3.9] with $P = \Sigma^{sdp}_{XX}$, we infer that the policy K is MSS.

Next, we show that $\Sigma^K \preceq \Sigma'$, for which it suffices to show that $\Delta = \Sigma^{sdp}_{XX} - \Sigma^K_{XX} \succeq 0$. To this end, consider the following from (6b):

$$\begin{aligned} \Sigma^K_{XX} + \Delta \succeq\, & pA(\Sigma^K_{XX} + \Delta)A^\top \\ & + \bar{p}\begin{pmatrix} A & B \end{pmatrix} (\Sigma^K_{XX} + \Delta) \begin{pmatrix} A & B \end{pmatrix}^\top + \Omega, \end{aligned}$$

which yields $\Delta \succeq pA\Delta A^\top + \bar{p}(A+BK)\Delta(A+BK)^\top$. Since K is MSS, we obtain $\Delta \succeq 0$. Next, since $\Sigma^K \preceq \Sigma' \preceq \Sigma^{sdp}$ and Σ^{sdp} is feasible, then so is Σ^K. Further, we have that

$$\begin{aligned} J(\Sigma^K) &= \bar{p}Tr\left(\text{Diag}(Q, R)\Sigma^K\right) + p\ Tr\left(Q\Sigma^K_{XX}\right) \\ &\leq \bar{p}Tr\left(\text{Diag}(Q, R)\Sigma^{sdp}\right) + pTr\left(Q\Sigma^{sdp}_{XX}\right) = J(\Sigma^{sdp}). \end{aligned}$$

The proof is thus complete. ☐

Next, we state the following lemmas which show that both the stationary and the non-stationary policies generated using (7) are MSSS.

Lemma 3. *Suppose that Assumption 1 holds, $\Omega \succeq \omega^2 I$, and $\upsilon(p, \frac{1}{2\kappa^2}, \frac{1}{2\kappa^2}) < 1$. Define $\hat{\kappa} := \sqrt{\sigma}/\omega$. Then, the policy in (7) is $(\hat{\kappa}, \frac{1}{2\hat{\kappa}^2}, \frac{1}{2\hat{\kappa}^2})$–MSSS for* (1).

Lemma 4. *Suppose that the hypotheses of Lemma 3 hold. Let $\Sigma^{sdp}_1, \Sigma^{sdp}_2, \cdots$, be a feasible sequence for the SDP in Problem 2. Suppose further that $\|\Sigma^{sdp}_{k+1} - \Sigma^{sdp}_k\| \leq \eta, \forall t$ and $\eta \leq \omega^2(1 - \upsilon)/2$. Then, the sequence $\{K_k\}^\infty_{k=1}$ is $(\hat{\kappa}, \frac{1}{2\hat{\kappa}^2}, \frac{1}{2\hat{\kappa}^2})$–MSSS for* (1).

We now have all the pieces in place to derive appropriate regret bounds for Problem 1, as done in the next section.

V. Regret Analysis

In this section, we first present the online projected gradient based algorithm (Algorithm 1) motivated from [12] and consequently present regret guarantees afforded by it. To this end, we start by defining the projection set $\mathcal{S}$ as

$$\left\{ \Sigma^{sdp} \in \mathbb{R}^{\ell \times \ell} \middle| \begin{array}{l} \Sigma^{sdp}_{XX} = pA\Sigma^{sdp}_{XX}A^\top + \bar{p}(A\ B)\Sigma^{sdp}(A\ B)^\top + \Omega \\ \Sigma^{sdp} \succeq 0, \quad Tr(\Sigma^{sdp}) \leq \sigma. \end{array} \right\} \qquad (8)$$

Algorithm 1 Online Erasure-based LQ Controller

1: Parameters: $\eta, \sigma > 0$, $p \geq 0$
2: Initialize $\Sigma^{sdp}_1 = I$
3: **for** k=1 to T **do**
4: Receive state X_k
5: Compute $K_k = (\Sigma^{sdp}_{k,XU})^\top (\Sigma^{sdp}_{k,XX})^{-1}$, $M_k = (\Sigma^{sdp}_{k,UU}) - K_k(\Sigma^{sdp}_{k,XX})K^\top_k$
6: Predict $U_k \sim \mathcal{N}(K_k X_k, M_k)$; receive Q_k, R_k
7: Update: $\Sigma^{sdp}_{k+1} = \text{Proj}_\mathcal{S}\left[\Sigma^{sdp}_k - \eta \text{Diag}(Q_k, \bar{p}R_k)\right]$, where $\text{Proj}_\mathcal{S}$ is projection onto the set $\mathcal{S}$ defined in (8).
8: **end for**

We are now ready to state and prove the main result of the paper in the following Theorem.

Theorem 1. *Suppose that $Tr(W_p) \leq \lambda^2$ and $\Omega \succeq \omega^2 I$. Then, given $\kappa > 0$, $0 < \gamma_1 < 1$, $0 < p \leq 1$ with $\upsilon(p, \frac{1}{2\hat{\kappa}^2}, \frac{1}{2\hat{\kappa}^2}) < 1$, and letting $\sigma = 2\kappa^4 Tr(W_p)\Gamma$ and $\eta = \frac{\omega^3}{4\sqrt{\sigma T}}$ in Algorithm 1, the expected regret of Algorithm 1 against a $(\kappa, \gamma_1, \gamma_2)$–MSSS policy K^* is given as*

$$J_T(\mathcal{C}) - J_T(K^*) = \mathcal{O}\left(poly(\kappa, \gamma_1, \gamma_2, \lambda, \omega, \zeta)\Gamma\sqrt{(1 + \bar{p}^2)T}\right)$$

for $T \geq 8\kappa^4\lambda^2\Gamma/\omega^2$, and Γ is defined in Lemma 5.

Proof. The proof is provided in Appendix B. ☐

Next, we present a numerical example to demonstrate the performance of the algorithm and the effect of packet drops on the system performance.

VI. A Numerical Example

We simulate the system (1)-(3) in Matlab, with the following parameters: $A = [1.1\ 0.3; 0.5\ 1.4]$, $B = [1; 2]$, $p = 0.7$, $\eta = 0.01$, $\sigma = 100$ and $\Omega = \text{Diag}(0.1, 0.1)$. We plot the expected average regret versus $\frac{1}{\sqrt{T}}$ as shown in Fig. 2. In the figure, we observe that Algorithm 1 *learns* the offline policy in the long run, as proved in Theorem 1, since the red curve decays to 0 as T grows large. Next, to demonstrate the effect of packet drops on the regret performance, we provide the following Table I as below. We simulate Algorithm 1

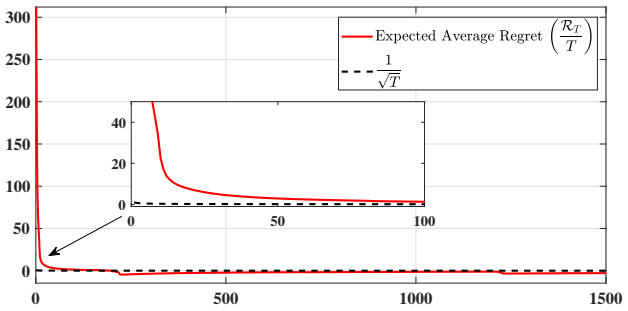

Fig. 2: Comparison of expected average regret $\mathcal{R}_T/T$ (red solid) versus $1/\sqrt{T}$ (black dotted)

for 4 different values of the packet drop probability (p = 0,0.10,0.25,0.60) and compute the expected average regret in each case for two different open-loop unstable systems ($A = \text{Diag}(1.3, 1)$, and $A = \text{Diag}(2, 1)$). The green checks in the table denote that this regret decays down to zero while the red cross marks denote that Algorithm 1 is unable to output a stabilizing policy for the system (1). This is because, in the event that the system state X_k grows unbounded, the SDP formulation breaks down since it is based on the assumption of existence of a steady state covariance, which does not exist in the latter case, as aligned with intuition.

p	$\mathcal{R}_T/T$ ($A = \text{Diag}(1.3, 1)$)	$\mathcal{R}_T/T$ ($A = \text{Diag}(2, 1)$)
0	✔	✔
0.10	✔	✔
0.25	✔	✘
0.60	✘	✘

TABLE I: The table shows the effect of packet drops on the regret performance, where we observe that for higher values of packet drop probabilities and instability in the system matrix, Algorithm 1 is unable to output a stabilizing policy due to non-existence of the steady-state covariance.

VII. Conclusion & Discussions

In this paper, we have studied the effect of packet drops on the performance of an online controller, when the cost parameters are *a priori* unknown to the controller. By reformulating the LQ problem as an SDP, we have used the online gradient descent-based algorithm to derive a sublinear $\mathcal{O}(\sqrt{T})$ regret bound on the online controller cost against a hindsight policy, which is chosen from the set of mean-square stable policies. Finally, we have verified the theoretical results with simulations. Next, we present some immediate extensions that follow from the analysis in the paper.

Forward link failures: One can investigate the case where the forward channel from the plant to the controller is also prone to packet drops, which leads to a partially observed setting, as in Fig. 3. Then, in addition to (1), the observation model is given by $Y_k^o = \beta_k X_k$, where $\beta_k \sim \text{Ber}(q)$, i.e., $\{\beta_k\}$ is a Bernoulli distributed random process and $q \in [0, 1]$ denotes the probability of loss of sensor packets. For the case where the controller has knowledge of the actuation loss instants $\{\alpha_k\}_{k=1}^T$ (which corresponds to the TCP-like protocol [4] with an ACK/NACK signal), the controller can construct a best estimate of the state $Z_k, \forall k$. Consequently, by repeating the presented analysis with an augmented state $[Z_k^\top e_k^\top]^\top$, where $e_k := X_k - Z_k$, we can derive an $\mathcal{O}(\sqrt{T})$ bound on the regret in Problem 1.

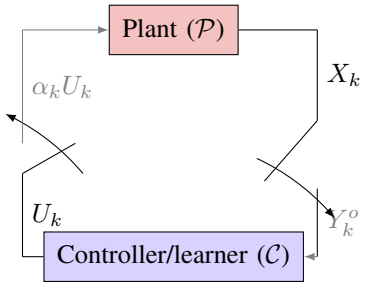

Fig. 3: Closed-loop information flow with random packet drops on forward and backward wireless links, with respective erasure probabilities p and q.

Unknown system dynamics: Another extension of this paper would be to study, in addition to unknown cost parameters Q_k, R_k, the effect of unknown system parameters A, B on the regret performance. In this case, we can use a system identification algorithm similar to that in [17] to obtain an estimate of system parameters $\hat{A}, \hat{B}$, determine the approximation errors between the estimated and actual parameters, and consequently use the presented analysis to derive sublinear regret bounds.

References

[1] X. Fang, S. Misra, G. Xue, and D. Yang, "Smart grid—the new and improved power grid: A survey," *IEEE Communications Surveys & Tutorials*, vol. 14, no. 4, pp. 944–980, 2011.

[2] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. S. Sastry, "Foundations of control and estimation over lossy networks," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 163–187, 2007.

[3] H. Sandberg, V. Gupta, and K. H. Johansson, "Secure networked control systems," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 5, pp. 445–464, 2022.

[4] O. C. Imer, S. Yüksel, and T. Başar, "Optimal control of LTI systems over unreliable communication links," *Automatica*, vol. 42, no. 9, pp. 1429–1439, 2006.

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

[6] J. Xiong and J. Lam, "Stabilization of linear systems over networks with bounded packet loss," *Automatica*, vol. 43(1), pp. 80–87, 2007.

[7] M. Braksmayer and L. Mirkin, "Discrete-time $\mathcal{H}_2$ optimal control under intermittent and lossy communications," *Automatica*, vol. 103, pp. 180–188, 2019.

[8] T. Alpcan, T. Başar, R. Srikant, and E. Altman, "CDMA uplink power control as a noncooperative game," *Wireless Networks*, vol. 8, pp. 659–670, 2002.

[9] J. R. Marden, H. P. Young, G. Arslan, and J. S. Shamma, "Payoff-based dynamics for multiplayer weakly acyclic games," *SIAM Journal on Control and Optimization*, vol. 48, no. 1, pp. 373–396, 2009.

[10] M. Zinkevich, "Online convex programming and generalized infinitesimal gradient ascent," in *Proceedings of the 20th International Conference on Machine Learning*, 2003, pp. 928–936.

[11] E. Hazan *et al.*, "Introduction to online convex optimization," *Foundations and Trends® in Optim.*, vol. 2, no. 3-4, pp. 157–325, 2016.

[12] A. Cohen, A. Hasidim, T. Koren, N. Lazic, Y. Mansour, and K. Talwar, "Online linear quadratic control," in *International Conference on Machine Learning*. PMLR, 2018, pp. 1029–1038.

[13] 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*. JMLR Workshop and Conference Proceedings, 2011, pp. 1–26.
[14] S. Lale, K. Azizzadenesheli, B. Hassibi, and A. Anandkumar, "Regret bound of adaptive control in linear quadratic Gaussian (LQG) systems," *available on arXiv:2003.05999*, 2020.
[15] Y. Li, S. Das, J. Shamma, and N. Li, "Safe adaptive learning-based control for constrained linear quadratic regulators with regret guarantees," *arXiv preprint arXiv:2111.00411*, 2021.
[16] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
[17] 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.
[18] T.-J. Chang and S. Shahrampour, "Distributed online linear quadratic control for linear time-invariant systems," in *2021 American Control Conference (ACC)*. IEEE, 2021, pp. 923–928.
[19] O. Anava, E. Hazan, and S. Mannor, "Online learning for adversaries with memory: price of past mistakes," *Advances in Neural Information Processing Systems*, vol. 28, 2015.
[20] E. Hazan and K. Singh, "Introduction to online nonstochastic control," *Available on arXiv:2211.09619*, 2022.
[21] O. L. V. Costa, M. D. Fragoso, and R. P. Marques, *Discrete-time Markov Jump Linear Systems*. Springer Sci. & Buss. Media, 2006.

APPENDIX A

Lemma 5. *Suppose that K is a randomized $(\kappa, \gamma_1, \gamma_2)$-MSSS policy and Σ and Σ^u are the steady-state covariances of the state X and control U, respectively, under policy K. Define $\Gamma = \frac{1}{\gamma_2} + \bar{p}\frac{1-\gamma_1}{\gamma_1}$. Then, we have that 1) $Tr(\Sigma) \leq \kappa^2 Tr(W_p)\Gamma$, and 2) $Tr(U) \leq \kappa^4 Tr(W_p)\Gamma$.*

Proof. Define $W_p := \bar{p}B\Sigma^c B^\top + \Omega$. Then, we have that

$$\Sigma = pA\Sigma A^\top + \bar{p}(A+BK)\Sigma(A+BK)^\top + W_p$$

which, using Definition 1, leads to

$$\Sigma = \sum_{k=0}^{\infty} pA^k(W_p){A^\top}^k + \sum_{k=1}^{\infty} \bar{p}(A+BK)^k W_p (A+BK)^{\top k}.$$

Taking trace of both sides of the above equation and using Definition 1, we arrive at 1). Finally, noting that $Tr(\Sigma^u) = Tr(K\Sigma K^\top)$, we arrive at 2) of the statement. □

Lemma 6. *Suppose $\eta \leq \omega^2(1-\upsilon)/4$. Then, the following holds:*

$$\sum_{k=1}^{T} Tr\left(Y_k(\Sigma_k^K - \Sigma_k^{sdp})\right) \leq \frac{2(3\bar{\upsilon} - \bar{\upsilon}^2)\sigma\hat{\kappa}^2}{\bar{\upsilon}^2 - 3\bar{\upsilon} + 2} + \frac{2\hat{\kappa}^2\eta\zeta T\sqrt{1+\bar{p}^2}}{\bar{\upsilon}^2 - 3\bar{\upsilon} + 2}. \tag{9}$$

Proof. Consider the non-stationary randomized policy $\mu_k(X) = K_k X + \nu_k$ with $\nu_k \sim \mathcal{N}(0, M_k)$. Then, we have that $\Sigma_{k,UU}^K = K_k \Sigma_{k,XX}^K K_k^\top + M_k$ and $\Sigma_{k,UU}^{sdp} = K_k \Sigma_{k,XX}^{sdp} K_k^\top + M_k$. Define $\varrho := \eta\zeta\sqrt{1+\bar{p}^2}$. This then yields

$$\begin{aligned}
&\sum_{k=1}^{T} Tr\left(Y_k(\Sigma_k^K - \Sigma_k^{sdp})\right) \\
&\quad\leq \sum_{k=1}^{T} Tr(Q_k + \bar{p}K_k R_k K_k^\top)\|\Sigma_{k,XX}^K - \Sigma_{k,XX}^{sdp}\| \\
&\quad\leq \zeta(1+\bar{p}\hat{\kappa}^2)\sum_{k=1}^{T}\|\Sigma_{k,XX}^K - \Sigma_{k,XX}^{sdp}\|.
\end{aligned} \tag{10}$$

Next, using Lemma 4, we have that

$$\|\Sigma_{k,XX}^K - \Sigma_{k,XX}^{sdp}\| \leq \hat{\kappa}^2\left[\bar{\upsilon}(1 + \frac{1-\bar{\upsilon}}{2})\right]^k \|\Sigma_{1,XX}^K - \Sigma_{1,XX}^{sdp}\| + \frac{2\hat{\kappa}^2\eta\zeta\sqrt{1+\bar{p}^2}}{\bar{\upsilon}^2 - 3\bar{\upsilon} + 2}, \tag{11}$$

where $\bar{\upsilon} = \upsilon(\hat{\kappa}, \frac{1}{2\hat{\kappa}^2}, \frac{1}{2\hat{\kappa}^2})$, and we used the fact that $\|\Sigma_{k+1,XX}^{sdp} - \Sigma_{k,XX}^{sdp}\| \leq \|\Sigma_{k+1}^{sdp} - \Sigma_k^{sdp}\| \leq \varrho$. Substituting (11) in (10), we get

$$\text{LHS} \leq 2\sigma\hat{\kappa}^2 \sum_{k=1}^{T}\left[\bar{\upsilon}(1 + \frac{1-\bar{\upsilon}}{2})\right]^k + \frac{2\hat{\kappa}^2\eta\zeta T\sqrt{1+\bar{p}^2}}{\bar{\upsilon}^2 - 3\bar{\upsilon} + 2},$$

which leads to (9). □

Lemma 7. *The following inequality holds:*

$$\sum_{k=1}^{T} Tr\left(Y_k(\Sigma_k^{sdp} - \Sigma^{K^*})\right) \leq \frac{2\sigma^2}{\eta} + (1-p/2)^2\zeta^2\eta T.$$

Proof. The proof uses Theorem 1 from [10]. We first note that $Tr(\Sigma^{sdp}) \leq \sigma$. Next, we have that $\|Q\| + \bar{p}\|R\| \leq \sqrt{Tr(QQ^\top)} + \bar{p}\sqrt{Tr(RR^\top)} \leq (2-p)\zeta$. Using these bounds in Theorem 1 of [10], we obtain the result. □

Lemma 8. *Given a $(\kappa, \gamma_1, \gamma_2)$–MSSS policy K^*, we have*

$$\sum_{k=1}^{T} Tr\left(Y_k(\Sigma^{K^*} - \Sigma_k^{K^*})\right) \leq 2\sigma\zeta(1+\bar{p}\kappa^2)\kappa^2\frac{\upsilon}{1-\upsilon}. \tag{12}$$

Proof. We first note that $Tr(\Sigma^{K^*}) = \kappa^2 Tr(W_p)\Gamma + \kappa^4 Tr(W_p)\Gamma \leq 2\kappa^4 Tr(W_p)\Gamma$. Thus, for $\sigma = 2\kappa^4 Tr(W_p)\Gamma$ with $Tr(W_p) \leq \lambda^2$, we have that Σ^{K^*} is feasible. Then, consider the following:

$$\begin{aligned}
&\sum_{k=1}^{T} Tr\left(Y_k(\Sigma^{K^*} - \Sigma_k^{K^*})\right) \\
&= \sum_{k=1}^{T} Tr(Q_k + \bar{p}(K^*)^\top R_k K^*(\Sigma_{XX}^{K^*} - \Sigma_{k,XX}^{K^*})) \\
&\leq \zeta(1+\bar{p}\kappa^2)\kappa^2\|\Sigma_{XX}^{K^*} - (\Sigma_{1,XX}^{K^*})\|\sum_{k=1}^{T}\upsilon^k,
\end{aligned}$$

which proves the result. □

APPENDIX B

Proof of Theorem 1. Define $Y_k := \text{Diag}(Q_k, \bar{p}R_k)$. Given K^*, let $\{\Sigma_k^{K^*}\}_{k=1}^T$ be the sequence of covariance matrices as in (5) under K^*. Further, let $\{\Sigma_k^K\}_{k=1}^T$ be the analogous sequence of covariance matrices when the policy of Algorithm 1 is applied, and $\{\Sigma_k^{sdp}\}_{k=1}^T$ be the sequence of covariance matrices generated by the SDP. Then, we have that

$$\begin{aligned}
J_T(\mathcal{C}) - J_T(K^*) &= \sum_{k=1}^{T} Tr\left(Y_k(\Sigma_k^K - \Sigma_k^{K^*})\right) \\
&\leq \sum_{k=1}^{T} Tr\left(Y_k(\Sigma_k^K - \Sigma_k^{sdp})\right) + \sum_{k=1}^{T} Tr\left(Y_k(\Sigma_k^{sdp} - \Sigma^{K^*})\right) \\
&+ \sum_{k=1}^{T} Tr\left(Y_k(\Sigma^{K^*} - \Sigma_k^{K^*})\right).
\end{aligned} \tag{13}$$

Next, using the results of Lemmas 5-8 from Appendix A, $\hat{\kappa} = \sqrt{\sigma}/\omega$ and $\hat{\gamma}_1 = \hat{\gamma}_2 = \omega^2/2\sigma$, and plugging in the values of η and σ, we obtain the desired result. □