LQG control subject to intermittent observations and SNR limitations

F. Parise, L. Dal Col, A. Chiuso, N. Laurenti, L. Schenato, A. Zanella

Abstract— **In this paper we consider the problem controlling unstable stochastic linear systems in the presence of a communication channel between the sensors and the actuators. We propose an LQG architecture that separates the problem of designing suitable regulators for controlling the plant, referred as Plant encoder/decoders, from the problem of designing encoder/decoder for channel transmissions. We provide a mathematical model that takes into account the most important features of today's current communication protocols such as quantization errors, limited channel capacity, decoding delay and packet loss, while still being amenable for analytic treatment. We then restrict to a special class of linear plant encoder/decoders and to a channel with signal-to-noise (SNR) limitations and packet loss only and we derive stability conditions and optimal design parameters for the controller. Through this analysis we are able to recover several results available in the literature that treated packet loss and quantization error separately.**

I. INTRODUCTION

Traditionally control theory and communication theory have been developed independently and have reached considerable success in developing fundamental tools for designing information technology systems. On one side, the major objective of control theory was to develop tools to stabilize unstable plants and to optimize some performance metrics in closed loop under the assumption that the communication channel between sensors and controller and between the controller and the plant were ideal, i.e. without distortion, packet loss and delay. On the other side, the major objective of communication theory was to develop tools to transmit information from a stable source to a receiver though a possibly noisy communication channel where the communication protocols adopted had no feedback on the source. One of the reasons for the success of these theories was that in many control applications the effects of the communication channel was negligible as compared to the effects of noise and uncertainty in the plants, while in many communication applications the rate of changing of the source was negligible as compared to the communication speed of the protocols so that it could be safely assumed to be stationary. With the advent of wireless communication, the Internet and the need for high performance control systems, this sharp separation between control and communication has become to be questioned and a growing body of literature has appeared from both the communication and the control community trying to analyze the interaction between control and communication. One line of research addressed the problem of stabilization of an unstable plant through a rate-limited erasure channel where no performance index is considered besides stability [9], [19], [7]. Another line of research applied information theoretic tools showing connections between structural limitations between feedback performance and channel capacity [8], and showing that Shannon's capacity of a channel is not sufficient to characterize a communication channel from a control perspective [12]. Differently, other researchers have tried to tackle the channel limitations by using analog models in order to avoid the difficulties associated with explicit design of channel encoder/decoder and to optimize some performance metrics among all possible stabilizing controllers [3], [11], [17]. Along these lines, other groups have modeled the channel limited capacity though a constraint on the maximum signal-to-noise (SNR) ratio and found fundamental limits which depend on the unstable eigenvalues of the plant [1], [16], [2]. Finally, another well explored approach is the analysis of control systems subject to random packet loss [18], [4], [5], [14] under LQG framework.

These works are just a partial overview of the literature on control systems subject to communication channel limits which is by no mean complete. Indeed, the current trend is to include multiple channel limitations into the model such as packet loss and quantization [20], [6] which however results is complex optimization problems.

The objective of this work is twofold. The first objective is to provide a more realistic model of a communication channel while still being mathematically amenable to analysis which include packet loss, delay, SNR-limitations and quantization distortion. The second objective is to propose an LQG approach for the design of the control blocks in order to include performance metrics besides stability of the closed loop system. In fact, from a practical standpoint, stability is not sufficient and additional performance criteria need to be satisfied, such as in the LQG framework. Although, we propose a very general architecture for networked control systems, in this work we limit our analysis and design to a simplified channel model which include only packet loss and SNR limitations and to a special class of linear controllers. Nonetheless, we recover several results available in the literature and we find a stability condition that depends on both the packet loss probability and the SNR of the channel.

This work is supported by the PRIN grant n. 20085FFJ2Z "New Algorithms and Applications of System Identification and Adaptive Control" by the Progetto di Ateneo CPDA090135/09 funded by the University of Padova, by the European Community's Seventh Framework Programme [FP7/2007-2013] under agreement n. FP7-ICT-223866-FeedNetBack and under grant agreement n257462 HYCON2 Network of excellence

All authors are in the Department of Information Engineering, University of Padova, Via Gradenigo 6/b, 35131 Padova, Italy XXX@dei.unipd.it

II. PROBLEM FORMULATION

We consider the problem of stabilizing a possibly unstable system across a communication channel. The plant is modeled as a discrete time linear time invariant dynamical systems subject to additive measurement and process noise. More specifically:

$$x_{t+1} = Ax_t + Bu_t + w_t \quad (1)$$
$$y_t = Cx_t + v_t \quad (2)$$

where $x \in \mathbb{R}^n, u \in \mathbb{R}^p, y \in \mathbb{R}^m$, $v_t \sim \mathcal{N}(0, R)$, $w_t \sim \mathcal{N}(0, Q)$, $x_0 \sim \mathcal{N}(0, P_0)$, and $w_t \perp v_t \perp w_t$. We also assume that the pairs (A, B) and (A, Q) are controllable, the pair (A, C) is observable, and $R > 0$.

Stabilization is a necessary requirement in any control systems, but in general additionally performance indices need to be optimized to achieve an acceptable behavior of the whole system. A typical choice is the steady state performance in terms of a quadratic cost index as in the linear-quadratic-gaussian (LQG) framework. More formally, in the contest of finite horizon LQG control the cost function is defined as:

$$J = \frac{1}{T}\sum_{t=0}^{T}\mathbb{E}[x_t^\top W x_t + u_t^\top U u_t]$$

while in the infinite horizon LQG control it is given by:

$$\begin{aligned} J &= \lim_{T\to+\infty}\frac{1}{N}\mathbb{E}[\sum_{t=0}^{T} x_t^\top W x_t + u_t^\top U u_t] \\ &= \lim_{t\to+\infty}\mathbb{E}[x_t^\top W x_t + u_t^\top U u_t] \end{aligned} \quad (3)$$

where the two limits coincide under some ergodicity assumptions.We also assume that the pair (A, W) is observable. Typical choices for the matrices W, U are $W = C^T C, U = \rho I$ so that, for $\rho = 0$, $J = \lim_{t\to+\infty}\mathbb{E}[||y_t||^2] - \text{trace}(R)$ corresponds to steady state output minus the measurement noise power. Hence, with a minor abuse of terminology, when $W = C^\top C$ and $U = \rho I = 0$ we define J as $J := \mathbb{E}\left[\|y_t\|^2\right]$.

The plant output y_t is measured and possibly preprocessed by a causal Coder/Estimator (COD) which sends data a_t across a channel. At the other hand, a causal Decoder/Controller (DEC) process the data received b_t and compute the control input u_t necessary to stabilize the plant and optimize the performance index J. A pictorial representation is given in top panel of Figure 1.

It is a standard practice to decouple the Coder/Decoder design into two tiers: one associated to the plant(source) and the other associated to the channel, as shown in the bottom panel of Figure 1. The goal of the Plant Coder/Decoder design is stabilize the closed loop system and possibly to optimize some additional performance index. These blocks in control theory framework correspond to filters, estimators, and controllers. Differently, the goal of the Channel Coder/Decoder design is to translate the signal s_t into a signal a_t that is suitable for transmission over the communication channel, in such a way that the signal b_t received from the channel

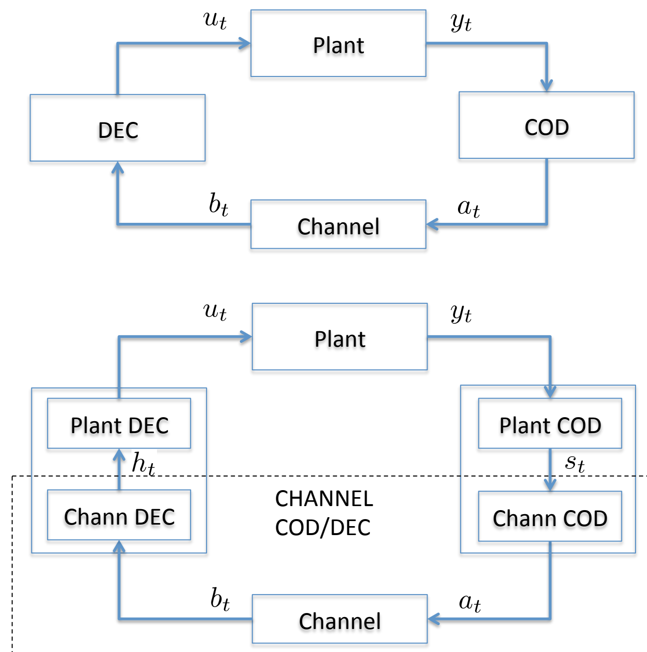

Fig. 1. Scheme of control system across a communication channel: general scheme (top), decoupled scheme (bottom).

can be decoded into a signal h_t that is as close as possible to the original signal s_t, i.e. $h_t \approx s_t$.[1] The decoupling of coding/decoding into Plant Coding/Decoding and Channel Coding/Decoding is not necessarily the optimal strategy in the context of feedback systems with unstable plants. Indeed, some recent work on anytime capacity and coding/decoding for unstable plants suggest that is not optimal. Nonetheless, we will stick to this approach since greatly simplifies the overall design and it is applicable to current communication protocols.

III. HIGH LEVEL CHANNEL MODELING

A fundamental limit to the transmission capability of a physical channel is given by the Shannon theorem on channel capacity, which dictates that in a channel with capacity C, it is possible to transmit at any rate $R < C$ with as small an error probability as desired. For a simple AWGN analog channel, capacity is given by

$$C = W\log_2(1+\Gamma), \quad (4)$$

where Γ is the channel signal-to-noise ratio at the receiver and W the bandwidth of the transmitted signal. Transmissions at rates greater than C, conversely, will inexorably experience decoding errors, irrespective of the coding and modulation technique used to represent the information signal. However, even in the $R < C$ case, when R is close

[1] Note that, the meaning of Plant and Channel Coding/Decoding considered in this work is different from Source and Channel Coding/Decoding considered in classical Information Theory, where the role of Source Coding is to remove the correlation of the signal y_t to reduce its bit rate, whereas Channel coding adds controlled redundancy to the signal before transmission over the channel to increase its robustness to transmit errors.

to C, arbitrarily low error probability is achieved in the limit of infinitely long codewords, which in turn, yield long decoding delays. If shorter codewords are needed, so that the transmission delay τ is limited, then one must allow for a certain *erasure probability* $\epsilon > 0$, i.e., the chance that some codewords (data packets) are not correctly decoded, and hence discarded by an error detection mechanism. The erasure probability, however, can be made very small when R is significantly lower than C.

Unfortunately, a real-valued signal s_t may have an infinite information rate, so that it cannot be transmitted over a finite capacity channel without introducing distortion and/or errors. Prior to transmission, therefore, the signal s_t is quantized to a signal $q_t = \mathcal{Q}(s_t)$ that takes values in a finite alphabet of size L. This operation makes it possible to limit the information rate R_q of the signal to be transmitted, at the cost of the introduction of a *quantization error* $n_t = q_t - s_t$, which is seen as an additive error signal.

Under the assumption of *fine quantization* (i.e., of a large enough L), the quantization error can be effectively modeled as a random process with identically distributed uncorrelated samples, with zero mean and power $\sigma^2 = \mathbb{E}\left[n_t^2\right]$ that depends on the statistics of s_t and the quantizer characteristic $\mathcal{Q}(\cdot)$. The ratio between the signal power $\mathbb{E}\left[s_t^2\right]$ and the quantization noise power σ^2, which will be referred to as Signal-to-Noise Ratio (SNR) in the following,[2] is bound by the rate R_q of the quantized signal. In fact, with properly designed quantizer and source coder, it is possible to prove that [15]

$$R_q \leq \frac{1}{2T_s} \log_2(\mathbb{E}\left[s_t^2\right]/\sigma^2)\,, \tag{5}$$

where T_s is the sample interval of s_t.

With a general quantizer, the actual source rate may exceed the bound (5). Nonetheless, by restricting to a given class of quantizers, the source rate R_q can be made a function of the SNR $\mathbb{E}\left[s_t^2\right]/\sigma^2$. For instance, let s_t be a gaussian distributed signal, with zero mean and variance σ_s^2, and consider a simple uniform quantizer $\mathcal{Q}(\cdot)$ with $L = 2^b$ levels that span the interval $[-v_{sat}, +v_{sat}]$, with $v_{sat} = 3\sigma_s$, so that the size of the quantization interval is $\Delta = \frac{2v_{sat}}{L} = \frac{6\sigma_s}{L}$. Under the assumption of fine quantization, the quantization error can be assumed to be uniformly distributed in the interval $[-\Delta/2, \Delta/2]$, with zero mean and power $\sigma^2 = \Delta^2/12$. Therefore, it turns out that $\text{SNR} = \frac{\mathbb{E}\left[s_t^2\right]}{\sigma^2} = \frac{L^2}{3}$, from which we get $L^2 = 3\frac{\mathbb{E}\left[s_t^2\right]}{\sigma^2}$. The resulting transmit rate is finally given by

$$R_q = \frac{b}{T_s} = \frac{\log_2(L^2)}{2T_s} = \frac{\log_2(3) + \log_2(\mathbb{E}\left[s_t^2\right]/\sigma^2)}{2T_s}\,.$$

We notice that with this quantizer the source rate is actually larger than the bound given in (5). Nonetheless, the relation between source rate R_q and SNR $\mathbb{E}\left[s_t^2\right]/\sigma^2$ is preserved.

In summary, the channel capacity C sets an upper bound on the source rate R_q, which shall be chosen in order to cut the more suitable tradeoff between transmission delay τ and erasure probability ϵ. The source rate R_q, in turn, is associated to the maximum achievable signal to noise ratio $\text{SNR}^* = \mathbb{E}\left[s_t^2\right]/\sigma^2$. In particular we define the parameter α as follows:

$$\alpha = \frac{1}{SNR^*} = \frac{\sigma^2}{\mathbb{E}[s_t^2]} \tag{6}$$

Any SNR lower than SNR^* can be sustained by the system without increase of the transmission delay or the erasure probability.

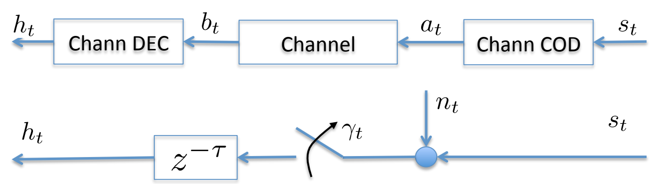

Fig. 2. Equivalent model of Channel COD/DEC using traditional codes

In this paper we adopt a channel model that accounts for all these effects, yet remains amenable to mathematical analysis. More specifically, we consider the model represented in Fig. 2, where n_t represents the quantization noise, whereas $\gamma_t \in \{0, 1\}$ is a Bernoulli process that models the erasure event ($\gamma_t = 0$), occurring with probability ϵ at each packet transmission, independently of previous events. Finally, the delay block $z^{-\tau}$ accounts for the encoding/decoding delay. Therefore, the parameters that characterize this model are $\alpha = 1/SNR^*$, ϵ, and τ. These parameters are clearly related, since reducing the erasure probability ϵ might require increasing the delay τ or reducing the transmission rate R_q, that is decreasing the signal-to-(quantization)-noise SNR^*. Therefore some trade-offs are expected in the context of feedback control systems, since all three terms negatively impact the performance of the closed loop system. Unfortunately, the exact form of the relation among these parameters is not available, though some tight bounds have been derived in [10]. For the ease of mathematical treatment, in our analysis we will assume that these parameters can be independently set, within reasonable constraints such as $R_q < C$, and $\epsilon < 1$. We can thus sort out the impact of each single parameter on the system performance.

IV. LQG ARCHITECTURE

Based on the channel modelling described above, which is independent of the control application, our goal is therefore to optimally design the Plant Coding block $\mathcal{F}_t$ and the Plant Decoding block $\mathcal{G}_t$ as well the optimal channel parameters $(\sigma^2, \epsilon, \tau)$ as depicted in Figure 3 to minimize the performance cost J.

The coder and the decoder can be time-varying but must be causal, i.e. must depend only on the past information set. We define with the upper-script the history of a signal, i.e. $y^t = (y_t, y_{t-1}, \dots, y_0)$. The information set available to the coder always include the plant outputs y^t, however

[2] This SNR is not to be confused with the channel signal-to-noise ratio Γ.

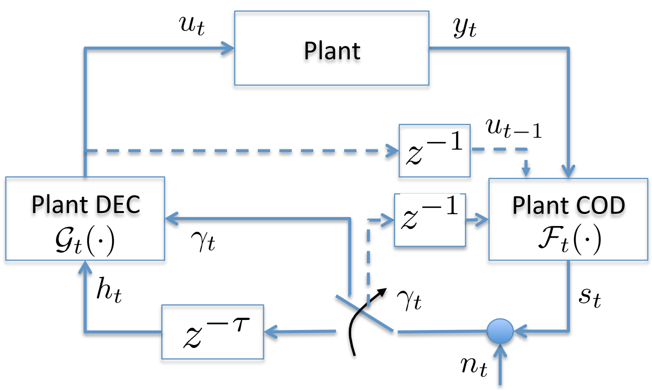

Fig. 3. General scheme of Networked Control System model with implicit channel COD/DEC

sometimes in the information set are also included the one-step delayed plant inputs u^{t-1}'s and/or channel packet loss γ^{t-1}, where $\gamma_t \in \{0, 1\}$ indicates whether a packet has been received correctly ($\gamma_t = 1$) or not ($\gamma_t = 0$). In the literature the following coder information sets are considered

$$s_t = \mathcal{F}_t(y^t, s^{t-1}) \quad (7)$$
$$s_t = \mathcal{F}_t(y^t, \gamma^{t-1}, s^{t-1}) \quad (8)$$
$$s_t = \mathcal{F}_t(y^t, u^{t-1}, s^{t-1}) \quad (9)$$

In particular, the second architecture of Eqn.(8) corresponds to a system with a reliable reception ACK mechanism to the transmitter, while the last architecture of Eqn.(9) corresponds to a system with a perfect communication feedback channel. The information set of the decoder includes, besides its past outputs u^{t-1}, also the output from the channel decoder r^t and packet loss sequence γ^t, i.e.

$$u_t = \mathcal{G}_t(r^t, \gamma^t, u^{t-1}) \quad (10)$$

This type of decoder information set is also referred as information set available in *real erasure channels*. In the framework developed above, the objective is to solve the following optimization problem:

$$\min_{(\sigma,\epsilon,\tau)\in\mathcal{C}} \quad \min_{\mathcal{F}_t, \mathcal{G}_t} J \quad (11)$$
$$\text{s.t.} \quad \mathbb{E}[||x_t||^2] \leq M, \ \forall t \quad (12)$$
$$\lim_{T\to+\infty} \frac{1}{T} \sum_{t=0}^{T} \mathbb{E}[||s_t||^2] \leq P \quad (13)$$

for some $M > 0$. This a formidable optimization problem since it involves many optimization parameters and poses only mild conditions on the possible classes of control functions $\mathcal{F}$ and $\mathcal{G}$ which lead to a large design parameter space. Most of the channel models and control architectures studied in the context of NCS can be cast as a special case of the optimization problem (11)-(13).

In general, the channel parameters $(\sigma^2, \epsilon, \tau)$ are assumed to be given and not partially designable. Moreover, they are studied singularly. For example, great attention has been given to lossy communication where only packet loss parameter ϵ is considered and the SNR constrain given in Eqn. (13) is neglected [18], [13], [4], [5]. Another area of active research is the SNR-constrained control which correspond to the problem where the channel model include only the quantization noise σ^2 and the channel power constrain Eqn. (13) [1], [11], [16], [17], [2]. Only recently, there is an attempt to consider more realistic channel models, for instance by including of both packet loss and quantization distortion [20], [6].

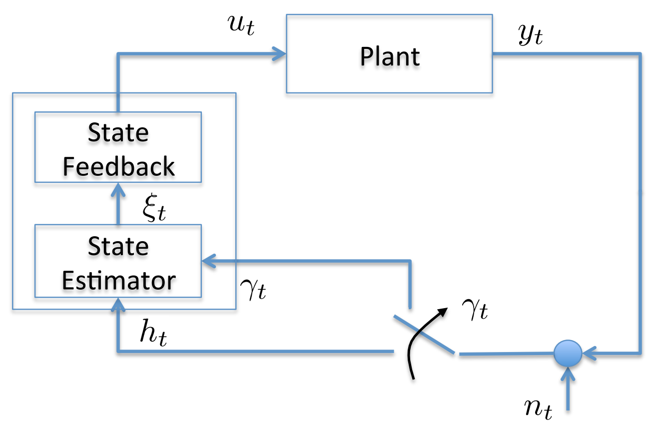

Fig. 4. Special scheme of Networked Control System for scalar output plants

In this work, in fact, we address a special case (shown in Fig. 4) of the general optimization problem (11)-(13) by considering a channel model that includes simultaneously i.i.d packet loss, quantization noise and a limited SNR (hence a finite channel capacity C), but assumes no decoding delay, i.e.

$$\tau = 0, \quad \mathbb{P}[\gamma_t = 0] = \epsilon, \quad \sigma^2 = \alpha\,\mathbb{E}\big[s_t^2\big] \quad (14)$$

Moreover, we restrict our design space to the Plant Decoding block and we consider no Plant Coding, i.e.

$$s_t = \mathcal{F}_t(y^t, s^{t-1}) = y_t \quad (15)$$

Finally, we restrict the Plant Decoder to be a regular obtained with the cascade of a linear state estimator and a state feedback, i.e.

$$\xi_t = A\xi_{t-1} + Bu_{t-1} + \gamma_t K\big(h_t - C(A\xi_{t-1} + Bu_{t-1})\big) \quad (16)$$
$$u_t = L\xi_t \quad (17)$$
$$h_t = \gamma_t(y_t + n_t) \quad (18)$$

Note that the estimator is time-varying since it depends on the sequence γ_t. In fact, if a packet is not received correctly, i.e. $\gamma_t = 0$, then the state estimator updates its state using the model only, while if it is received, i.e. $\gamma_t = 1$, the a correction term based on the output innovation similarly to a Kalman filter. This scheme is the same scheme proposed in [13], that does not coincide with the true optimal Kalman filter as in [18], but has the advantage that it is computationally simpler and allows explicit computation of the performance J, as it will be shown in the next section.

V. Dynamical equations

We now derive the dynamical equation which governs the state as well as the error evolution for the estimator in equations (16). In order to do so it is convenient to consider the "predictor" $\hat{x}_t$ so that

$$\begin{aligned} \hat{x}_{t+1} &= A\hat{x}_t + Bu_t + \gamma_t G\left[h_t - C\hat{x}_t\right] \\ \xi_t &= \hat{x}_t + \gamma_t K\left[h_t - C\hat{x}_t\right] \\ u_t &= L\xi_t = L\left[I - \gamma_t KC\right]\hat{x}_t + \gamma_t LKh_t \end{aligned} \tag{19}$$

where $G := AK$. This system has the form of a "Kalman-like" estimator with constant gain K. If the gain K is chosen to be the "optimal" Kalman gain [3], which we shall denote by K^*, then $\hat{x}_t = \hat{x}_{t|t-1}$ and $\xi_t = \hat{x}(t|t)$, which are respectively the optimal (constant gain) one-step-ahead predictor and estimator of the state x_t.

Consider the error $\tilde{x}_t := x_t - \hat{x}_t$. The "error" equations are therefore:

$$\tilde{x}_{t+1} = x_{t+1} - \hat{x}_{t+1} = Ax_t + Bu_t +$$

$$+v_t - A\hat{x}_t - Bu_t - \gamma_t G\left[z_t - C\hat{x}_t\right] =$$

$$= A\tilde{x}_t + v_t - \gamma_t G\left[C\tilde{x}_t + w_t + n_t\right] =$$

$$= (A - \gamma_t GC)\tilde{x}_t + v_t - \gamma_t G(w_t + n_t)$$

Substituting the input given by the controller in the predictor equations:

$$\hat{x}_{t+1} = A\hat{x}_t + Bu_t + \gamma_t G\left[z_t - C\hat{x}_t\right] =$$

$$= A\hat{x}_t + BL\left[I - \gamma_t KC\right]\hat{x}_t + \gamma_t BLKz_t +$$

$$+\gamma_t G\left[z_t - C\hat{x}_t\right]$$

$$= \left[A + BL\right]\hat{x}_t + \gamma_t\left[BLK + G\right]\left[z_t - C\hat{x}_t\right] =$$

$$= \left[A + BL\right]\hat{x}_t + \gamma_t\left[BL + A\right]K\left[C\tilde{x}_t + w_t + n_t\right]$$

The system output is therefore:

$$y_t = Cx_t + w_t = C\left[\tilde{x}_t + \hat{x}_t\right] + w_t$$

it follows that the equation of the feedback loop system are:

$$\begin{bmatrix} \hat{x}_{t+1} \\ \tilde{x}_{t+1} \end{bmatrix} = \begin{bmatrix} (A+BL) & \gamma_t(A+BL)KC \\ 0 & A(I-\gamma_t KC) \end{bmatrix} \begin{bmatrix} \hat{x}_t \\ \tilde{x}_t \end{bmatrix} +$$

$$+ \begin{bmatrix} 0 \\ I \end{bmatrix} v_t + \begin{bmatrix} \gamma_t(A+BL)K \\ -\gamma_t AK \end{bmatrix} \begin{bmatrix} w_t + n_t \end{bmatrix}$$

$$y_t = \begin{bmatrix} C & C \end{bmatrix} \begin{bmatrix} \hat{x}_t \\ \tilde{x}_t \end{bmatrix} + w_t$$

where we use $G = AK$. Defining

$$P := Var\{[\hat{x}_t^\top, \tilde{x}_t^\top]^\top\} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}$$
$$\bar{A}_\gamma := \begin{bmatrix} (A+BL) & \gamma(A+BL)KC \\ 0 & A(I-\gamma KC) \end{bmatrix}$$

[3]Within the class of constant gain linear estimators.

and using the fact that $\begin{bmatrix} \hat{x}_t \\ \tilde{x}_t \end{bmatrix}$, v_t e $[w_t + n_t]$ are uncorrelated, it follows that:

$$\begin{aligned} P = \; & (1-\epsilon)\bar{A}_1 P \bar{A}_1^\top + \epsilon \bar{A}_0 P \bar{A}_0^\top + \\ & + \begin{bmatrix} 0 \\ I \end{bmatrix} Q \begin{bmatrix} 0 & I \end{bmatrix} + \\ & +(1-\epsilon) \begin{bmatrix} (A+BL)K \\ -AK \end{bmatrix} \begin{bmatrix} R+N \end{bmatrix} \begin{bmatrix} (A+BL)K \\ -AK \end{bmatrix}^\top \end{aligned} \tag{20}$$

$$N = \alpha P_y \qquad P_y = \begin{bmatrix} C & C \end{bmatrix} P \begin{bmatrix} C^\top \\ C^\top \end{bmatrix} + R$$

Substituting the expression for P_y in (20) we obtain:

$$\begin{aligned} P = \; & (1-\epsilon)\bar{A}_1 P \bar{A}_1^\top + \epsilon \bar{A}_0 P \bar{A}_0^\top + \\ & + \begin{bmatrix} 0 \\ I \end{bmatrix} Q \begin{bmatrix} 0 & I \end{bmatrix} + \\ & +(1-\epsilon)(1+\alpha) \begin{bmatrix} (A+BL)K \\ -AK \end{bmatrix} R \begin{bmatrix} (A+BL)K \\ -AK \end{bmatrix}^\top + \\ & +\alpha(1-\epsilon)\bar{\Phi} P \bar{\Phi}^\top \end{aligned} \tag{21}$$

where

$$\bar{\Phi} := \begin{bmatrix} (A+BL)KC & (A+BL)KC \\ -AKC & -AKC \end{bmatrix}$$

For ease of notation we define the operator on the right hand side of (21) as $\mathcal{M}(K, L, P)$, so that (21) can be written in compact form as

$$P = \mathcal{M}(K, L, P)$$

With this architecture the cost function J in (3), with the choices $W := C^\top C$ and $U = \rho I$ takes the form:

$$\begin{aligned} J \; = \; & \mathbb{E}\left[x_t^\top C^\top C x_t\right] + \rho \mathbb{E}\left[u_t^\top u_t\right] \\ = \; & \begin{bmatrix} C & C \end{bmatrix} P \begin{bmatrix} C^\top \\ C^\top \end{bmatrix} + \\ & +\rho(1-\epsilon) \begin{bmatrix} L & LKC \end{bmatrix} P \begin{bmatrix} L^\top \\ C^\top K^\top L^\top \end{bmatrix} + \\ & \rho\epsilon L P_{11} L^\top + \rho(1-\epsilon) LK(R + \alpha P_y) K^\top L^\top \end{aligned} \tag{22}$$

where

$$P_y = \begin{bmatrix} C & C \end{bmatrix} P \begin{bmatrix} C^\top \\ C^\top \end{bmatrix} + R$$

Hence, the LGQ-type optimal control problem can be written as:

$$\begin{aligned} J^* := \min_{K,L} \quad & J \\ s.t. \quad & P = \mathcal{M}(K, L, P) \\ & P \geq 0 \end{aligned} \tag{23}$$

VI. Analysis of the scalar case

In the scalar case with $b = c = 1$ the equations become:

$$\begin{bmatrix} \hat{x}_{t+1} \\ \tilde{x}_{t+1} \end{bmatrix} = \begin{bmatrix} (a+l) & \gamma_t(a+l)k \\ 0 & a(1-\gamma_t k) \end{bmatrix} \begin{bmatrix} \hat{x}_t \\ \tilde{x}_t \end{bmatrix} +$$

$$+ \begin{bmatrix} 0 \\ 1 \end{bmatrix} v_t + \begin{bmatrix} \gamma_t(a+l)k \\ -\gamma_t ak \end{bmatrix} \begin{bmatrix} w_t + n_t \end{bmatrix}$$

$$y_t = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \hat{x}_t \\ \tilde{x}_t \end{bmatrix} + w_t$$

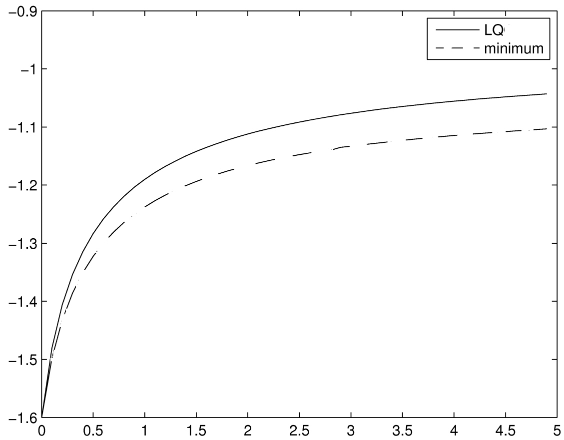

Fig. 5. Comparison between the optimal LQ gain (solid) and the optimal value of l which solves (23) with cost function J in (25), (dashed) as a function of ρ ($\rho = 0$ corresponds to "cheap control", i.e. no penalty on the input).

With variance:

$$\begin{aligned} P &= (1-\epsilon)\bar{A}_1 P \bar{A}_1^\top + \epsilon \bar{A}_0 P \bar{A}_1^\top + \begin{bmatrix} 0 & 0 \\ 0 & q \end{bmatrix} + \\ &+ (1-\epsilon) \begin{bmatrix} (a+l)k \\ -ak \end{bmatrix} (r + \alpha P_y) \begin{bmatrix} (a+l)k & -ak \end{bmatrix} \end{aligned} \quad (24)$$

$$\begin{aligned} P_y &= \begin{bmatrix} 1 & 1 \end{bmatrix} P \begin{bmatrix} 1 \\ 1 \end{bmatrix} + r \\ &= p_{11} + 2p_{12} + p_{22} + r \end{aligned}$$

The cost J in (22) specializes, in the scalar case, to

$$\begin{aligned} J \;=\; & p_{11} + 2p_{12} + p_{22} + \rho\epsilon l^2 p_{11} \\ & + \rho(1-\epsilon)[l^2 p_{11} + 2l^2 k p_{12} + l^2 k^2 p_{22}] \\ & + \rho(1-\epsilon) l^2 k^2 [r + \alpha(p_{11} + 2p_{12} + p_{22} + r)] \end{aligned} \quad (25)$$

A. *Loss of separation principle*

A natural question to ask is whether the celebrated "separation principle", which guarantees that estimator and control design can be done independently, holds also in this LQG formulation with SNR constraints and packet losses. We now illustrate through a specific case that this is not the case. We consider a scalar system with $a = 1.6$, $q = r = 0.1$, $\alpha = 0.1$ and $\epsilon = 0$. In figure 5 we report, as a function of ρ (see the definition of J in (25)) the optimal LQ gain versus the controller gain l which solves (23) with cost function J in (25), as a function of ρ. It is clear that, while for $\rho = 0$ the solutions coincide (we shall analyze this situation in the next section), for $\rho > 0$ the LQ gain is different from the optimal gain l for J in (25),(25). We can conclude that the separation principle does not hold.

B. *Minimization of the cost $J = E\left[y^2\right]$ ($\rho = 0$)*

Let us now consider the problem of designing k and l so as to minimize the output variance $J = E\left[y^2\right]$. Denoting with

$$P := \begin{bmatrix} p_{11} & p_{12} \\ p_{12} & p_{11} \end{bmatrix}$$

then

$$\begin{aligned} J &= P_y = [1\ 1] P \begin{bmatrix} 1 \\ 1 \end{bmatrix} + r = \\ &= p_{11} + 2p_{12} + p_{22} + r \end{aligned}$$

where p_{11}, p_{12} and p_{22} satisfy (24) which can be written more explicitly as:

$$p_{11} = (a+l)^2 \left[(1-\epsilon)(p_{11} + 2kp_{12} + k^2 p_{22} + k^2(r + \alpha J)) + \epsilon\, p_{11}\right]$$

$$p_{12} = (a+l) \left[(1-\epsilon)(a(1-k)(p_{12} + kp_{22}) - ak^2(\alpha J + r)) + \epsilon a p_{12}\right]$$

$$p_{22} = (1-\epsilon)(a^2(1-k)^2 p_{22} + a^2 k^2 (r + \alpha J)) + \epsilon a^2 p_{22} + q$$

We shall now consider the first order optimality conditions

$$\frac{\partial J}{\partial k} = 0 \quad \frac{\partial J}{\partial l} = 0. \quad (26)$$

and show that $l = -a$ and $k = k^*$ (the optimal Kalman gain which minimizes p_{22} within the class of constant gain estimators) satisfy (26). It is easy to check that, $\forall k$ and $l = -a$,

$$\frac{\partial p_{11}}{\partial k} = 0 \quad \frac{\partial p_{11}}{\partial l} = 0. \quad (27)$$

In addition, choosing $k = k^*$ to be the optimal Kalman gain, so that p_{22} is minimized (within the class of constant gain algorithms), we also have, for $l = -a$ and $k = k^*$,

$$\begin{aligned} \tfrac{\partial p_{12}}{\partial k} &= 0 \\ \tfrac{\partial p_{12}}{\partial l} &= (1-\epsilon)[a(1-k^*)k^* p_{22} - a(k^*)^2(\alpha J + r)]. \end{aligned} \quad (28)$$

In the second equation we have also used the fact that, for $l = -a$, $p_{11} = 0$ and hence also $p_{12} = 0$.

Being $k = k^*$ the optimal Kalman gain, also $\frac{\partial p_{22}}{\partial k} = 0$ must hold for $l = -a$ and $k = k^*$; therefore using (27) and (28), $\frac{\partial J}{\partial k} = 0$ for $k = k^*$, $l = -a$. It follows that, for $k = k^*$, $l = -a$,

$$\frac{\partial p_{22}}{\partial k} = 0 = (1-\epsilon)[2a^2 k^*(\alpha J + r) - 2a^2(1-k^*)p_{22}] \quad (29)$$

Comparing with (28) it follows that also $\frac{\partial p_{12}}{\partial l} = 0$ for $k = k^*$ and $l = -a$). Last,

$$\frac{\partial p_{22}}{\partial l} = (1-\epsilon) \left[a^2(1-k)^2 \frac{\partial p_{22}}{\partial l}\right]$$

which, using (27) and (28) and (29), clearly has $\frac{\partial p_{22}}{\partial l} = 0$ as solution for $k = k^*$ and $l = -a$.

This implies that the first order optimality conditions (26) are satisfied for $k = k^*$ and $l = -a$.

The value of k^* can be obtained imposing:

$$k^* = arg\min_k p_{22} = \frac{p_{22}}{(1+\alpha)(p_{22} + r)}$$

Substituting this value in p_{22} we get:

$$p_{22} = a^2 p_{22} + q - \frac{1-\epsilon}{1+\alpha} \frac{a^2 p_{22}^2}{p_{22} + r}$$

This is the same modified algebraic Riccati equation (MARE) that appears in [18] with $\bar{\eta} = \frac{1-\epsilon}{1+\alpha}$.

Notice that since the system is scalar the critical probability for the solvability of the MARE is known:

$$\eta_c = 1 - \frac{1}{a^2}$$

Given that the problem has solution, i.e. p_{22} converges, iff:

$$\bar{\eta} = \frac{1-\epsilon}{1+\alpha} > \eta_c = 1 - \frac{1}{a^2}$$

which implies:

$$\alpha < \frac{1-\epsilon a^2}{a^2-1} \quad \Rightarrow \quad \mathrm{SNR}^* > \frac{a^2-1}{1-\epsilon a^2} \tag{30}$$

VII. Analysis for B invertible, scalar output, and $\rho = 0$.

Inspired by the scalar example we are now interested in considering MISO systems where the control B is square and invertible, i.e. we have n independent control inputs, the matrix C is rank-one, i.e. it is a row-vector, and $R = r$ is a positive scalar. We also consider the cheap-control scenario, i.e $\rho = 0$. Similarly as above we choose as control input strategy a control that $A + BL = 0$, i.e.

$$L = -B^{-1}A \tag{31}$$

Under these hypotheses and control gain selection, the steady state variance and cost functions reduces to

$$\begin{aligned}
P_{11} &= 0 && (32)\\
P_{12} &= P_{21}^\top = 0 && (33)\\
P_{22} &= (1-\epsilon)A(I-KC)P_{22}(I-KC)^\top A^\top + \epsilon A P_{22} A^\top + \\
&\quad + Q + (1-\epsilon)AK(\alpha C P_{22} C^\top + (\alpha+1)r)K^\top A^\top && (34)\\
&= \mathcal{L}(K, P_{22})\\
J &= C P_{22} C^\top + r && (35)
\end{aligned}$$

where $\mathcal{L}(K, P_{22})$ is an operator that is affine in P_{22} for any fixed K. Therefore the objective becomes:

$$\begin{aligned}
\overline{J} = \min_{K, P_{22}} \quad & C P_{22} C^\top + r \\
s.t. \quad & P_{22} = \mathcal{L}(K, P_{22}) \\
& P_{22} \geq 0
\end{aligned}$$

After some simple manipulation it is possible to see that Eqn. (34) can be written as:

$$\begin{aligned}
\mathcal{L}(K, P_{22}) \ := \ & \Phi(P_{22}) + (1+\alpha)(1-\epsilon)A(K - K_{P_{22}}) \cdot \\
& \cdot (C P_{22} C^\top + r)(K - K_{P_{22}})^\top A^\top
\end{aligned}$$

where

$$\begin{aligned}
\Phi(P_{22}) &:= A P_{22} A^\top + Q - \eta A P_{22} C^\top (C P_{22} C^\top + R)^{-1} C P_{22} A^\top \\
K_{P_{22}} &:= \frac{1}{1+\alpha} P_{22} C^\top (C P_{22} C^\top + R)^{-1} \\
\eta &= \frac{1-\epsilon}{1+\alpha}
\end{aligned}$$

We are now in the position to prove the following theorem:

Theorem 1: Under the assumption that B is square and invertible, C is rank-one, and $\rho = 0$, then the minimum for the cost function J^* of the optimization problem (23) is upper bounded by:

$$J^* \leq \overline{J} = C P_{22}^* C^\top + r$$

where

$$P_{22}^* = A P_{22}^* A^\top + Q - \eta A P_{22}^* C^\top (C P_{22}^* C^\top + R)^{-1} C P_{22}^* A^\top \tag{36}$$

and is achieved with the following gains:

$$L = -B^{-1}A, \qquad K = \frac{1}{1+\alpha} P_{22}^* C^\top (C P_{22}^* C^\top + R)^{-1}$$

Such matrix P_{22}^* exists and it is unique if and only if

$$\frac{1-\epsilon}{1+\alpha} > 1 - \frac{1}{\prod_i |\lambda_i^u|^2} \tag{37}$$

where λ_i^u represent the unstable eigenvalues of the matrix A.

Proof: The proof follows from the results in [13] and in the interest of space we just provide a brief sketch. From the previous expression it clearly follows that $\Phi(P_{22}) \leq \mathcal{L}(K, P_{22}), \forall K$. Therefore, if $P_{22}^* = \Phi(P_{22}^*), P_{22}^* \geq 0$ and $P_{22} = \mathcal{L}(K, P_{22}), P_{22} \geq 0$, then it must be $P_{22}^* \leq P_{22}$ for any choice of the estimator gain K. Since the existence of a positive definite solution $P_{22}^* = \Phi(P_{22}^*)$ also implies that $P_{22}^* = \mathcal{L}(K_{P_{22}^*}, P_{22}^*)$, than this means P_{22}^* is the smallest among solutions $P_{22} = \mathcal{L}(K, P_{22})$ as K varies, and it is achieved by the gain $K = K_{P_{22}^*}$. This implies that also $C P_{22}^* C^\top + r \leq C P_{22} C^\top + r$, which proves the first part of the theorem. The second part of the theorem follows from the know results the modified algebraic Riccati Equation (MARE) of Eqn.(36) that states that a solution exists if and only if $\eta > 1 - \frac{1}{\prod_i |\lambda_i^u|^2}$ if C is rank-one [18]. ■

A. Discussion and related work

Some considerations is in order based on Theorem 1. The first is that although $\overline{J}$ provides only an upper bound on the achievable optimal cost J^* since we arbitrarily set the control gain to the specific value $L = -B^{-1}A$, we conjecture that the bound is tight non only for the scalar case but also in general, i.e. $J^* = \overline{J}$, and this is part of ongoing research. We also expect that if B is not square and invertible, then the design of the optimal control gain L and the estimator gain K will be coupled and that the optimal cost $\overline{J}$ will provide a lower bound for the true J^*. Moreover also the stability condition of Eqn. (37) is likely to provide only a necessary stability condition but not sufficient.

The second is that we recover some of the results available in the literature. In fact if we set $\alpha = 0$, then this is equivalent to consider a channel with infinite capacity and we obtain the same stability condition in the lossy network literature [18], [14]. Alternatively, if we assume no packet loss in the channel, i.e. $\epsilon = 0$, and recalling that $\alpha = \frac{1}{\mathrm{SNR}^*}$, then the stability condition can be rewritten as

$$1 - \frac{1}{\prod_i |\lambda_i^u|^2} < \frac{1}{1 + \frac{1}{\mathrm{SNR}^*}} = 1 - \frac{1}{1 + \mathrm{SNR}^*}$$

which lead to

$$\text{SNR}^* > \prod_i |\lambda_i^u|^2 - 1$$

which is the same stability condition presented in the context of SNR-limited control system in [1].

Finally, the bound provided by Eqn. (37) will be useful to compare different communication protocols. In fact, by using a corse quantizer it is possible to reduce the transmission rate R_q, thus allowing more redundant channel coding schemes and consequently a smaller packet loss probability ϵ. On the other hand a corser quantizer gives a smaller SNR^* and consequently a higher α. Therefore, α and ϵ are coupled and cannot be designed separately.

VIII. Conclusions and future work

We have considered an LQG control problem under communication constraints; the model we have proposed accounts for bandwidth limitations, as per the Shannon capacity theorem, as well as for packet drops and delays. We have argued in fact that there is a tight connection between the actual rate at which one can transmit information, the decoding delay (due to long block coding) and the packet-drop probability. While delays do not influence the possibility to stabilize a system, they do play a major role when performance (e.g. measured by the variance of certain error signals) is of interest. However, transmitting close to channel capacity with small delays, will make packet drops non-negligible.

We then restricted our attention to a specific control architecture in which the plant outputs are transmitted via a bandlimited channel and then processed through the cascade of a state estimator followed by a linear (state) feedback controller; for ease of exposition we did not consider delays, while both limited rate and packet drops have been included in our analysis. We have first considered a scalar model and showed that the separation principle does not hold in general. Then we assumed a "cheap control" setup and found that, in the scalar case, the optimal controller has a dead-beat structure and the optimal estimator is a Kalman-like constant gain estimator (which accounts for the packet drop probability). Conditions for stability are derived in terms a modified algebraic Riccati equation and recapture results from the literature as special cases. These results can be extended, to some extent, to the non scalar case provided the system is reachable in one step, i.e. the B matrix is square and invertible.

Future work will include a detailed analysis of the multivariable case as well as the inclusion of delays in our model.

References

[1] J. Braslavsky, R. Middleton, and J. Freudenberg, "Feedback stabilization over signal-to-noise ratio constrained channels," *IEEE Transactions on Automatic Control*, vol. 52, no. 8, 2007.

[2] ——, "Minimum variance control over a gaussian communication channel," *IEEE Transactions on Automatic Control*, vol. 56, no. 8, 2011.

[3] N. Elia, "Remote stabilization over fading channels," *Systems and Control Letters*, vol. 54, pp. 237–249, 2005.

[4] V. Gupta, D. Spanos, B. Hassibi, and R. M. Murray, "Optimal LQG control across a packet-dropping link," *Systems and Control Letters*, vol. 56, no. 6, pp. 439–446, 2007.

[5] O. C. Imer, S. Yuksel, and T. Basar, "Optimal control of dynamical systems over unreliable communication links," *Automatica*, vol. 42, no. 9, pp. 1429–1440, September 2006.

[6] Y. Ishido, K. Takaba, and D. Quevedo, "Stability analysis of networked control systems subject to packet-dropouts and finite-level quantization," *Systems & Control Letters*, vol. 60, pp. 325–332, 2011.

[7] L.Coviello, P. Minero, and M. Franceschetti, "Stabilization over markov feedback channels," *Transactions on Automatic Control (provisionally accepted)*.

[8] N. C. Martins and M. A. Dahleh, "Feedback control in the presence of noisy channels: Bode-like fundamental limitations of performance," *IEEE Transactions on Automatic Control*, vol. 52, p. 16041615, 2008.

[9] G. N. Nair and R. J. Evans, "Exponential stabilisability of finite-dimensional linear systems with limited data rates," *Automatica*, vol. 39, no. 4, pp. 585–593, April 2003.

[10] Y. Polyanskiy, H. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, p. 23072359, 2010.

[11] A. J. Rojas, J. H. Braslavsky, and R. H. Middleton, "Fundamental limitations in control over a communication channel," *Automatica*, vol. 44, no. 12, 2008.

[12] A. Sahai and S. Mitter, "The necessity and sufficiency of anytime capacity for control over a noisy communication link: Part I," *IEEE Transaction on Information Theory*, 2006.

[13] L. Schenato, "Kalman filtering for networked control systems with random delay and packet loss," *IEEE Transactions on Automatic Control*, vol. 53, pp. 1311–1317, 2008.

[14] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. Sastry, "Foundations of control and estimation over lossy networks," *Proceedings of the IEEE*, vol. 95, pp. 163–187, 2007.

[15] C. E. Shannon, "Communication in the presence of noise," *Proceedings of the IRE*, vol. 37, no. 1, p. 10 21, 1949.

[16] E. Silva, G. Goodwin, and D. Quevedo, "Control system design subject to snr constraints," *Automatica*, vol. 46, no. 2, 2010.

[17] E. Silva and S. Pulgar, "Control of lti plants over erasure channels," *Automatica*, vol. 47, no. 8, 2011.

[18] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. Jordan, and S. Sastry, "Kalman filtering with intermittent observations," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1453–1464, September 2004.

[19] S. Tatikonda and S. Mitter, "Control under communication constraints," *IEEE Transaction on Automatic Control*, vol. 49, no. 7, pp. 1056–1068, July 2004.

[20] K. Tsumura, H. Ishii, and H. Hoshina, "Tradeoffs between quantization and packet loss in networked control of linear systems," *Automatica*, vol. 45, no. 12, pp. 2963–2970, 2009.