On iterative system design and separation in control over noisy channels ⋆

L. Bao, M. Skoglund and K. H. Johansson

The ACCESS Linnaeus Center and School of Electrical Engineering
Royal Institute of Technology, Stockholm, Sweden

E-mail: lei.bao@ee.kth.se, skoglund@ee.kth.se, kallej@ee.kth.se

Abstract: We study a closed-loop control system with feedback transmitted over a noisy discrete memoryless channel. We design encoder–controller pairs that jointly optimize the sensor measurement quantization, protection against channel errors, and control. The design goal is to minimize an expected linear quadratic cost over a finite horizon. As a result of deriving optimality criteria for this problem, we present new results on the validity of the separation principle subject to certain assumptions. More precisely, we show that the certainty equivalence controller is optimal when the encoder is optimal and has full side-information about the symbols received at the controller. We then use this result to formulate tractable design criteria in the general case. Finally, numerical experiments are carried out to demonstrate the performance obtained by various design methods.

1. INTRODUCTION

Networked control based on limited sensor and actuator information has attracted increasing attention during the past decade. A significant interest in this research area has been devoted to the analysis and synthesis of quantized feedback control over data-rate limited communication links to stabilize an unstable plant. The work so far has mostly assumed control over error-free communication links, where the only limitation imposed by the channel is the limited data rate, see e.g., Baillieul (2002); Brokett and Liberzon (2000); Elia and Mitter (2001); Fagnani and Zampieri (2003), and the survey in Nair *et al* (2007). More recently, control over noisy channels has motivated a great deal of challenging research topics. Some influential works include Mahajan and Teneketzis (2006); Matveev and Savkin (2007); Sahai and Mitter (2006), and Tatikonda and Mitter (2004). For the purpose of the stability, solutions are often based on a kind of separation principle. The basic concept is to design the stabilizing control (or observation) assuming the channel is noise-free; then, require the control/observation to be reliably transmitted over the noisy channel. The control–channel separation relies essentially on information-theoretic results on coding schemes which ensure reliable communication over noisy channels. But, how to construct those coding schemes is still an open question. Compared to stability properties, optimal designs for general criteria are much less explored in the literature. However, the problem of optimal stochastic control over communication channels is addressed in e.g., Matveev and Savkin (2004); Tatikonda *et al.* (2004). Concerning encoder design, some related results can be found in e.g., Borkar *et al.* (2001); Mahajan and Teneketzis (2006); Tatikonda *et al.* (2004).

The main contribution of the present paper is a practical synthesis technique for joint optimization of the quantization, error protection and control for state observations over a bandlimited noisy channel. We extend our previous work on the subject, cf., Bao *et al.* (2007a), by constructing an iterative optimization algorithm applicable to more general systems, especially in situations where the communication between sensors and the controller is highly constrained such that only a few bits can be transmitted. At low transmission rates, there are many advantages in employing designs that accomplish source coding, protection against channel errors and control simultaneously.

The paper is organized as follows. In Section 2 we define the control system with encoder, controller, and communication channel. Thereafter, a problem statement which concerns a linear quadratic (LQ) objective over a finite horizon is formulated. Important features of the optimal controller and encoder are investigated in Section 4 and Section 5. Based on these results, joint and separate design methods are developed in Section 6. Finally, we present numerical experiments in Section 7 and conclusions in the last section. The following notation will be used in this paper. Bold-faced characters describe a sequence of signals or functions, e.g., $\mathbf{x}_a^b = \{x_a, \ldots, x_b\}$ denotes the evolution of a discrete-time signal x_t from $t = a$ to $t = b$. We use $\mathbf{E}\{\cdot\}$ for the expectation operator, $\mathrm{tr}(\cdot)$ for the trace operator, and $P_r(\cdot)$ the probability. The notations $(\cdot)'$ and $(\cdot)^\dagger$ stand for matrix transpose and matrix pseudoinverse, respectively. To indicate an optimal solution, the notation $(\cdot)^*$ is used.

2. PRELIMINARIES

In the most general case, we consider a control system with a communication channel as depicted in Fig. 1. The multivariable linear plant is governed by the equations

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t + v_t, \\ y_t &= Cx_t + e_t, \end{aligned} \tag{1}$$

where $x_t \in \mathbb{R}^n, u_t \in \mathbb{R}^m, y_t \in \mathbb{R}^p$ are the state, the control, and the measurement, respectively. The matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{p\times n}$, are known; moreover, (A,C) is state observable and (A,B) is state controllable. Process noise $v_t \in \mathbb{R}^n$ and measurement noise $e_t \in \mathbb{R}^p$ are zero-mean, independent and

⋆ This work was partially supported by the Swedish Research Council, the Swedish Strategic Research Foundation, and the Swedish Governmental Agency for Innovation Systems.

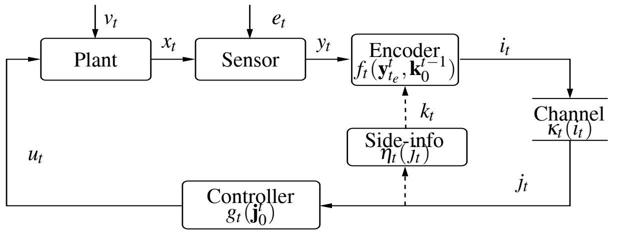

Fig. 1. A general system for feedback control over a discrete memoryless channel. The dashed line indicates potential side-information from the controller to the encoder.

identically distributed (iid) and mutually independent. They are also independent of the system initial state x_0.

We consider an encoder as a mapping from the set of encoder information to a discrete set of symbols. We take each symbol to be represented by an integer index. At time t, the index is $i_t \in \mathscr{L}_I = \{1, \dots, L_I\}$, and L_I is an integer. In particular, we are interested in the class of encoder mappings described by the function

$$i_t = f_t(\mathbf{y}_{t_e}^t, \mathbf{k}_0^{t-1}),\ t_e = t - M_t, \tag{2}$$

where, M_t specifies the memory of past measurements at the encoder. For example, $M_t = 0$ is the case when only the latest measurement is available; while, when $M_t = t$, the encoder knows all past measurements $\mathbf{y}_0^t$. Given the history of side-information $\mathbf{k}_0^{t-1}$, the encoder maps past measurements $\mathbf{y}_{t_e}^t$ to an index $i_t \in \mathscr{L}_I$. The side-information k_t represents the feedback to the encoder about the symbol $j_t \in \mathscr{L}_J = \{1, \dots, L_J\}$ received at the controller. In this paper, we define the *side-information* at the encoder to be

$$k_t = \eta_t(j_t) \in \mathscr{L}_K = \{1, \dots, L_K\}, \quad 1 \le L_K \le L_J, \tag{3}$$

where $\eta_t : \mathscr{L}_J \to \mathscr{L}_K$ is deterministic and memoryless. Accordingly, $k_t = j_t$ and $L_K = L_J$ when full side-information is available, while, $k_t = 0$ and $L_K = 1$ when there is no side-information at the encoder (cf., the Internet User Datagram Protocol). Between the extremes, there are a variety of cases with incomplete side-information, for which $1 < L_K < L_J$. By means of side-information the encoder will be informed about the potential transmission errors. In contrast to the conventional Automatic Repeat re-Quest (ARQ) protocol, no re-transmission will take place, instead, the control is designed to maintain robustness against transmission errors.

Let the discrete memoryless channel have input variable i_t and output j_t, with one channel use defined by

$$j_t = \kappa_t(i_t), \tag{4}$$

where $\kappa_t : \mathscr{L}_I \to \mathscr{L}_J$ is a random memoryless mapping. Conditioned on the transmitted symbol i_t, the mapping to j_t is independent of other parameters in the system, e.g., the process and measurement noise. The fact that the channel is bandlimited is captured by the finite size of the input-alphabet $\mathscr{L}_I$. The expression (4) encompasses many well studied channel models, for example, the Binary Symmetric Channel (BSC), and the Binary Erasure Channel (BEC). The latter is commonly used to model Internet-like channels which suffer packet dropouts. At the receiver side, we consider a controller that causally utilizes all past channel outputs $\mathbf{j}_0^t$, to produce the control command

$$u_t = g_t(\mathbf{j}_0^t) \in \mathbb{R}^m. \tag{5}$$

Note that past controls are completely specified by past received symbols. We denote the conditional mean estimator of the state x_s, based on the history of the received indexes $\mathbf{j}_0^t$, as

$$\hat{x}_{s|t} = \mathbf{E}\left\{x_s|\mathbf{j}_0^t\right\},\ s \le t. \tag{6}$$

We will use $\hat{x}_t$ as a short notation for $\hat{x}_{t|t} = \mathbf{E}\left\{x_t|\mathbf{j}_0^t\right\}$. Let $\tilde{x}_t$ be the error in estimating the state x_t based on $\mathbf{j}_0^t$, that is,

$$\tilde{x}_t = x_t - \hat{x}_t = x_t - \mathbf{E}\left\{x_t|\mathbf{j}_0^t\right\}. \tag{7}$$

Given the channel outputs, we also define the measurement prediction and its corresponding prediction error as $\hat{y}_{t^-} = \hat{y}_{t|t-1} = \mathbf{E}\left\{y_t|\mathbf{j}_0^{t-1}\right\}$ and $\tilde{y}_{t^-} = \tilde{y}_{t|t-1} = y_t - \hat{y}_{t|t-1}$, respectively. In the paper we use "encoder" and "controller" in quite general terms ("the corresponding boxes in Fig. 1"), as well as specific terms ("the components/mappings f_t and g_t at time t").

3. PROBLEM STATEMENT

Our goal is to solve an optimal encoder–controller problem and thereby finding the suitable encoder and controller mappings. The adopted performance measure is the following LQ cost with finite horizon $T > 0$

$$J_T = \sum_{t=1}^{T} \left(x_t' V_t x_t + u_{t-1}' P_{t-1} u_{t-1}\right), \tag{8}$$

where the matrices V_t and P_t are symmetric and positive definite. The objective is to find the encoder–controller mappings which minimize the expected value $\mathbf{E}\{J_T\}$. For ease of reference, we refer to the main design problem as Problem 1, and summarize it below.

Problem 1. Consider the system in Fig. 1. Given the plant (1) *and the memoryless channel* (4)*, find the encoder f_t and controller g_t that minimize the expected value of the cost* (8).

We use the notation

$$\{f_t^*(\mathbf{y}_{t_e}^t, \mathbf{k}_0^{t-1})\}_{t=0}^{T-1} \quad \text{and} \quad \{g_t^*(\mathbf{j}_0^t)\}_{t=0}^{T-1}, \tag{9}$$

for the optimal (not necessarily unique) mappings that solve Problem 1. Problem 1 can be viewed as an extension of the traditional LQ problem that the optimal encoder–controller minimizes the cost function with respect to initial state, process noise, measurement noise, and a bandlimited noisy channel. In general, finding an exact solution to Problem 1 is not feasible, because the optimization problem is highly non-linear and non-convex. In the next two sections, we simplify the overall problem by studying the controller and encoder separately. First, the problem of finding the optimal control strategy for a fixed encoder is addressed in Section 4. Thereafter, in Section 5, we consider the problem of optimizing one single encoder component, assuming the controller and other encoder components are fixed.

4. OPTIMAL CONTROLLER

In this section we investigate the optimal controller mapping g_t, assuming the encoder $\mathbf{f}_0^{T-1}$ is fixed. In the general case, we obtain a recursive equation which is typically difficult to solve, therefore, the solution of the special case of full side-information is utilized to approximate the optimal solution.

4.1 GENERAL CASE

The problem of finding the optimal control assuming the encoder is fixed fits well into the setting of stochastic optimal control, e.g., Aoki (1967), where in our problem the observations at the controller are the integer-valued indexes $\mathbf{j}_0^t$. Applying dynamic programming, the result is summarized in Proposition 2, see also Bao *et al.* (2007a).

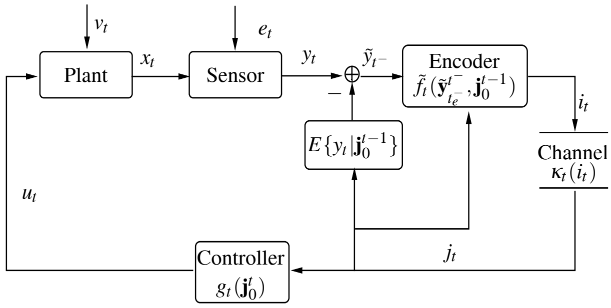

Fig. 2. The measurement prediction error encoder system where a prediction error encoder $\tilde{f}_t(\tilde{\mathbf{y}}_{t_e^-}^{t^-},\mathbf{j}_0^{t-1})$ is employed.

Proposition 2. Consider a fixed encoder $\mathbf{f}_0^{T-1}$. Given the plant (1) and the memoryless channel (4), a controller mapping $u_t = g_t(\mathbf{j}_0^t)$ that minimizes the expected value of (8) fulfills the following recursive relation

$$\begin{aligned} u_{t-1}^* &= \arg\min_{u_{t-1}}\{\gamma_t\}, \\ \gamma_t &= \lambda_t + \mathbf{E}\left\{\gamma_{t+1}^*|\mathbf{j}_0^{t-1}\right\}, \\ \lambda_t &= \mathbf{E}\left\{(Ax_{t-1}+Bu_{t-1}+v_{t-1})'V_t(Ax_{t-1}+Bu_{t-1}+ \right. \\ &\left. v_{t-1})u_{t-1}'P_{t-1}u_{t-1}|\mathbf{j}_0^{t-1}\right\},\ t=1,\dots,T, \end{aligned} \tag{10}$$

initialized with $\gamma_{T+1}^* = 0$.

Unfortunately, it is in general not possible to efficiently solve (10). One main obstruction lies in how $\mathbf{E}\left\{\gamma_{t+1}^*|\mathbf{j}_0^{t-1}\right\}$ is affected by past controls. In the next sub-sections we will first investigate the case of full side-information and demonstrate that this assumption significantly simplifies the problem. Then we will discuss how to apply the conclusions derived assuming full side-information in the general case.

4.2 FULL SIDE-INFORMATION

In this sub-section we look at the special case that the encoder has full side-information. In this case we are able to provide a characterization of the optimal system. Assume full side-information is available, $k_t = j_t$, and the encoder mapping is $f_t(\mathbf{y}_{t_e}^t,\mathbf{j}_0^{t-1})$. Then, consider the new system in Fig. 2 that encodes the measurement prediction errors. In particular, we consider the mapping from prediction errors $\tilde{\mathbf{y}}_{t_e^-}^{t^-}$ (cf., Section 2) and $\mathbf{j}_0^{t-1}$ to index i_t. Define the *measurement prediction error encoder* $\tilde{f}_t : \mathbb{R}^{(M_t+1)\times p}\times\mathscr{L}_J^t \to \mathscr{L}_I$,

$$i_t = \tilde{f}_t(\tilde{\mathbf{y}}_{t_e^-}^{t^-},\mathbf{j}_0^{t-1}). \tag{11}$$

Consequently, we call the system in Fig. 2 the (measurement) prediction error encoder system. For a given prediction error encoder system as in Fig. 2, with the prediction error encoder $\tilde{\mathbf{f}}_0^{T-1}$ and controller $\mathbf{g}_0^{T-1}$ specified, one can construct a system with the original structure of Fig. 1 which has exactly the same performance. The corresponding system in Fig. 1 utilizes the same controller $\mathbf{g}_0^{T-1}$, together with an encoder whose components are determined by $\tilde{\mathbf{f}}_0^{T-1}$ and $\mathbf{g}_0^{T-1}$ as

$$f_t(\mathbf{y}_{t_e}^t,\mathbf{j}_0^{t-1}) = \tilde{f}_t\left(y_{t_e}-\hat{y}_{t_e|t_e-1},\dots,y_t-\hat{y}_{t|t-1},\mathbf{j}_0^{t-1}\right). \tag{12}$$

For the prediction-error-encoder system, we formulate Problem 3.

Problem 3. Consider a prediction error encoder system as in Fig. 2. Given the plant (1) *and the memoryless channel* (4)*, find the prediction error encoder and controller mappings*

$$\{\tilde{f}_t^*(\tilde{\mathbf{y}}_{t_e^-}^{t^-},\mathbf{j}_0^{t-1})\}_{t=0}^{T-1} \quad \textit{and} \quad \{g_t^*(\mathbf{j}_0^{t-1})\}_{t=0}^{T-1}, \tag{13}$$

that minimize the expected value of (8).

Given the plant, memoryless channel and design criterion, the solutions to the original Problem 1 and the corresponding Problem 3 are closely related, as revealed by Proposition 4.2.

Proposition 4. I. Consider a solution $\{f_t^*(\mathbf{y}_{t_e}^t,\mathbf{j}_0^{t-1}),g_t^*\}_{t=0}^{T-1}$ *to Problem 1. The same controller* $\mathbf{g}_0^{*T-1}$ *and the prediction error encoder whose components are specified by* $\{\tilde{f}_t(\mathbf{y}_{t_e}^t,\mathbf{j}_0^{t-1}),g_t^*\}_{t=0}^{T-1}$ *according to*

$$\tilde{f}_t(\tilde{\mathbf{y}}_{t_e^-}^{t^-},\mathbf{j}_0^{t-1}) = f_t^*\left(\tilde{y}_{t_e^-}+\hat{y}_{t_e^-},\dots,\tilde{y}_{t^-}+\hat{y}_{t^-},\mathbf{j}_0^{t-1}\right), \tag{14}$$

jointly solve Problem 3.

II. Consider a solution $\{\tilde{f}_t^*(\tilde{\mathbf{y}}_{t_e^-}^{t^-},\mathbf{j}_0^{t-1}),g_t^*\}_{t=0}^{T-1}$ *to Problem 3. The same controller* $\mathbf{g}_0^{*T-1}$ *and the encoder* $\{f_t(\mathbf{y}_{t_e}^t,\mathbf{j}_0^{t-1})\}_{t=0}^{T-1}$ *specified by* $\{\tilde{f}_t^*(\mathbf{y}_{t_e}^t,\mathbf{j}_0^{t-1}),g_t^*\}_{t=0}^{T-1}$ *according to*

$$f_t(\mathbf{y}_{t_e}^t,\mathbf{j}_0^{t-1}) = \tilde{f}_t^*\left(y_{t_e}-\hat{y}_{t_e^-},\dots,y_t-\hat{y}_{t^-},\mathbf{j}_0^{t-1}\right), \tag{15}$$

jointly solve Problem 1.

Proof. I. If the prediction error encoder $\{\tilde{f}_t(\tilde{\mathbf{y}}_{t_e^-}^{t^-},\mathbf{j}_0^{t-1})\}_{t=0}^{T-1}$ derived according to (14), and $\mathbf{g}_0^{*T-1}$ do not jointly solve Problem 3, then another solution to Problem 3 provides a cost lower than the one given by $\{\tilde{f}_t(\tilde{\mathbf{y}}_{t_e^-}^{t^-},\mathbf{j}_0^{t-1}),g_t^*\}_{t=0}^{T-1}$. If this is the case, using the encoder specified by the solution to Problem 3 according to (12), jointly with the controller of the same solution, must lead to a lower cost than the one resulting from $\{f_t^*(\mathbf{y}_{t_e}^t,\mathbf{j}_0^{t-1}),g_t^*\}_{t=0}^{T-1}$. This contradicts the statement that $\{f_t^*(\mathbf{y}_{t_e}^t,\mathbf{j}_0^{t-1}),g_t^*\}_{t=0}^{T-1}$ is a solution to Problem 1. Hence, the statement in Proposition 4 must be true.

II. Follow similar arguments as the proof of part I. □

Proposition 4 indicates that a solution to Problem 3 specifies a solution to Problem 1, and vice versa. As shown later in the paper, it turns out that when using a joint training approach, Problem 3 is easier to solve than Problem 1. Hence, in the special case of full side-information we will focus on finding a solution to Problem 3, and then derive a corresponding solution to Problem 1 according to (15).

Consider now a fixed sequence of prediction error encoder mappings $\tilde{\mathbf{f}}_0^{T-1}$. Notice that for any such fixed mappings, the transmitted indexes $\mathbf{i}_0^{T-1}$ and the received indexes $\mathbf{j}_0^{T-1}$ do not depend on the controls $\mathbf{u}_0^{T-1}$. We demonstrate this fact by first inspecting the prediction error $\tilde{y}_{t|t-1}$ which equals to

$$CA^t x_0 + \sum_{s=0}^{t-1} CA^{t-1-s}v_s + e_t - \mathbf{E}\left\{CA^t x_0 + \sum_{s=0}^{t-1} CA^{t-1-s}v_s|\mathbf{j}_0^{t-1}\right\}, \tag{16}$$

since the controls are completely specified by the received symbols $\mathbf{j}_0^{t-1}$. By (16), $\tilde{y}_{t|t-1}$ does not depend on $\mathbf{u}_0^{t-1}$ if the received symbols $\mathbf{j}_0^{t-1}$ do not depend on $\mathbf{u}_0^{t-1}$. This is the case since $i_0 = \tilde{f}_0(\tilde{y}_{0|-1}) = \tilde{f}_0(x_0+Ce_0)$, $i_1 = \tilde{f}_1(\tilde{\mathbf{y}}_{t_e^-}^{1^-}, j_0)$, $i_2 = \tilde{f}_2(\tilde{\mathbf{y}}_{t_e^-}^{2^-}, j_0, j_1)$, and etc., and j_t depends only on i_t and potential channel errors. Then, consider the state estimation error

$$\tilde{x}_t = A^t x_0 + \sum_{s=0}^{t-1} A^{t-1-s} v_s - \mathbf{E}\left\{ A^t x_0 + \sum_{s=0}^{t-1} A^{t-1-s} v_s | \mathbf{j}_0^{t-1} \right\}. \quad (17)$$

Note that x_0, v_s, j_s for $s = 0, \ldots, t-1$ do not depend on past controls, the estimation error $\tilde{x}_t$ consequently does not depend on past controls either. Because of this fact, we will be able to solve (10), as given in Proposition 5. The proof follows Proposition 2 in Bao *et al.* (2007a).

Proposition 5. Consider a prediction error encoder system, with a fixed prediction error encoder $\tilde{\mathbf{f}}_0^{T-1}$. Given the plant (1) *and the memoryless channel* (4)*, the controller component $u_t = g_t(\mathbf{j}_0^t)$ that minimizes the expected value of* (8) *is given by*

$$\begin{aligned} u_t &= \ell_t \hat{x}_t, \\ \ell_t &= -(P_t + B'(V_{t+1} + I_{T-t-1})B)^\dagger B'(V_{t+1} + I_{T-t-1})A, \\ I_{T-t-1} &= A'(V_t + I_{T-t-2})A - \pi_{T-t-1}, \\ \pi_{T-t-1} &= A'(V_t + I_{T-t-2})B(P_{t-1} + B'(V_t + I_{T-t-2})B)^\dagger \\ &B'(V_t + I_{T-t-2})A, \end{aligned} \quad (18)$$

initialized at $I_1 = A'V_T A - A'V_T B(P_{T-1} + B'V_T B)^\dagger B'V_T A$.

Since the optimal control (18) can be decomposed into a separate decoder and controller, the *separation property* holds, e.g., Aoki (1967). Moreover, (18) is a so-called *certainty equivalence* (CE) controller. Note that, given any fixed encoder $\mathbf{f}_0^{T-1}$, the CE controller in (18) is not necessarily optimal for the original system in Fig. 1. Still, in the jointly optimal pair $\{\mathbf{f}_0^{*T-1}, \mathbf{g}_0^{*T-1}\}$ that solves Problem 1, the controller $\mathbf{g}_0^{*T-1}$ is a CE controller, as stated in Proposition 6.

Proposition 6. There exits $\{f_t^*(\mathbf{y}_{t_e}^t, \mathbf{j}_0^{t-1}), g_t^*\}_{t=0}^{T-1}$ solving Problem 1 in which the controller $\mathbf{g}_0^{*T-1}$ is the CE controller given by (18) *for $\mathbf{f}_0^{T-1} = \mathbf{f}_0^{*T-1}$.*

Proof. For the given system, one can find an optimal solution $\{\tilde{f}_t^*(\tilde{\mathbf{y}}_{t_e^-}^{t^-}, \mathbf{j}_0^{t-1}), g_t^*\}_{t=0}^{T-1}$ to Problem 3. By Proposition 5, $\mathbf{g}_0^{*T-1}$ is a CE controller. Then, generate the optimal pair $\{f_t^*(\mathbf{y}_{t_e}^t, \mathbf{j}_0^{t-1}), g_t^*\}_{t=0}^{T-1}$ according to Proposition 4. Observe that $\tilde{f}_t^*(\tilde{\mathbf{y}}_{t_e^-}^{t^-}, \mathbf{j}_0^{t-1})$ and $f_t^*(\mathbf{y}_{t_e}^t, \mathbf{j}_0^{t-1})$ produce exactly the same i_t. Therefore, $\hat{x}_t$ and consequently the CE controller are identical for both systems in Problem 1 and Problem 3. □

Proposition 6 shows that there exist solutions such that the optimal controller corresponding to the optimal encoder $\mathbf{f}_0^{*T-1}$ is a CE controller. In the general case, the optimal controller corresponding to any encoder $\mathbf{f}_0^{T-1}$ does not necessarily satisfy the separation principle. For the prediction error encoder, the optimality of the CE controller is attributed to the fact that the dependence of past controls is removed before encoding. There are other classes of encoders for which the separation principle also applies, e.g., the open-loop encoder in Bao *et al.* (2007a). Similar architectures have also been investigated in, e.g., Tatikonda and Mitter (2004); Nair *et al* (2007), for some special cases of the general system in Fig. 1, and assuming noiseless finite-rate transmission. Our result is more general, and is directly applicable to iterative encoder–controller design. The corresponding separation result in Nair *et al* (2007) cannot be used for iterative design, since the controller is explicitly utilized to specify the encoder in that result.

4.3 PARTIAL SIDE-INFORMATION

In the case of the original system in Fig. 1, where the encoder (2) may not have access to full side-information, it is not possible to convert the original system in Problem 1 to a prediction error encoder system. Therefore, given any fixed encoder, the CE controller in (18) is not necessarily the optimal control strategy. Since we are not able to solve (10) in the general case, we resort to using the CE controller as a suboptimal alternative to the solution to (10). That is, in the case of partial side-information at the encoder, we propose a design based on constraining the controller to be a CE controller. Our numerical experiments in fact demonstrate that subtracting $\hat{y}_{t|t-1}$ does often not influence the result of the encoding (since if the system "works" then $\hat{y}_{t|t-1}$ is "small"). This explains why fairly good results are obtained when using the CE controller also in the case of partial side-information.

5. OPTIMAL ENCODER

In this section, we briefly address the problem of optimizing the encoder component f_t, for a fixed controller $\mathbf{g}_0^{T-1}$ and fixed encoder components $\mathbf{f}_0^{t-1}$, $\mathbf{f}_{t+1}^{T-1}$. The following results are a straightforward consequence of the principle of optimality.

Proposition 7. Consider a fixed controller $\mathbf{g}_0^{T-1}$ and fixed encoder components $\mathbf{f}_0^{t-1}$, $\mathbf{f}_{t+1}^{T-1}$. Given the linear plant (1) *and the memoryless channel* (4)*, the encoder component $f_t(\mathbf{y}_{t_e}^t, \mathbf{k}_0^{t-1})$ that minimizes the expected value of* (8) *is given by*

$$i_t = \arg\min_{i \in \mathscr{L}_I} \mathbf{E}\left\{ \sum_{s=t+1}^{T} (x_s' V_s x_s + u_{s-1}' P_{s-1} u_{s-1}) \middle| \mathbf{y}_{t_e}^t, \mathbf{k}_0^{t-1}, i_t = i \right\}. \quad (19)$$

The analogous rule of the prediction error encoder mapping is similar to (19). However, since the state estimation error $\hat{x}_t$ does not depend on past controls, the following result is useful in practice.

Corollary 8. Consider a prediction error encoder system. There exists a solution to Problem 3 that satisfies the following conditions: The controller mapping is given by $u_t = \ell_t \hat{x}_t$, as in (18)*; and the prediction error encoder mapping $\tilde{f}_t$ is given by*

$$i_t = \arg\min_{i \in \mathscr{L}_I} \mathbf{E}\left\{ \sum_{s=t}^{T-1} \tilde{x}_s' \pi_{T-s} \tilde{x}_s \middle| \tilde{\mathbf{y}}_{t_e^-}^{t^-}, \mathbf{j}_0^{t-1}, i_t = i \right\}, \quad (20)$$

where π_{T-s} can be computed according to (18).

Proof. By Proposition 5 we know that, given a fixed prediction error encoder (in this case the optimal prediction encoder), the CE controller (18) is optimal. Also, by (18), the choice of i_t influences only the term $\mathbf{E}\left\{\sum_{s=t}^{T-1} \tilde{x}_s' \pi_{T-s} \tilde{x}_s | \mathbf{j}_0^{t-1}\right\}$ in the cost-to-go γ_{t+1}, when the CE controller is employed. □

An interpretation of Corollary 8 is that the optimization of the prediction error encoder can be separated from the optimization of the controller. A discussion on the separate design of the encoder–controller will be given later in Section 6.

6. ENCODER–CONTROLLER DESIGN

This section we consider overall design of the encoder–controller. First in Section 6.1, we describe the joint design based on an iterative training method. Thereafter, the complexity reduced separate design is discussed in Section 6.2.

6.1 JOINT DESIGN

As stated previously, the overall encoder–controller optimization problem is typically not tractable, and we therefore optimize the encoder–controller pair iteratively, with the goal of finding locally optimal solutions. Inspired by traditional quantizer design for noisy channels (e.g., Farvardin (1990)), the idea is to fix the encoder and update the controller, then fix the controller and update the encoder etc. Criteria for updating the encoder and controller are developed in Section 4 and Section 5. More details about the encoder–controller operation and practical issues related to the implementation of training algorithm are referred to the full version of the paper Bao *et al.* (2007b). Unfortunately, the iterative optimization algorithm will not guarantee convergence to a global optimum, but by influencing the initial conditions of the design it is possible to search for good locally optimal designs. As discussed in Section 4.3, we constrain the controller to be the CE controller (18) for the case with partial side-information. Note that, for the prediction error encoder system, the joint training converges (to a local optimum), since the CE controller is optimal for any $\tilde{\mathbf{f}}_0^{T-1}$. However, in the general case (with partial encoder side-information) the design does not necessarily converge, since updating the CE controller based on a new $\mathbf{f}_0^{T-1}$ is not guaranteed to result in better performance. Still, in our numerical experiments this has not been a problem, and empirically the design algorithm also appears to converge to a solution in the general case.

6.2 SEPARATE DESIGN

Based on the results in Section 4 and Section 5, we know that a separate design of the prediction error encoder and controller can still achieve optimal performance. In particular, we optimize the prediction error encoder by (20). Thereafter, convert the optimized prediction error encoder and its corresponding CE controller to an encoder–controller pair of the original system, by (15). When the encoder is near-optimal, $\hat{y}_{t|t-1}$ is usually "small". Therefore, in the general case with partial side-information, we can apply the same strategy and expect an insignificant performance degradation.

However, it turns out that optimizing the encoder, according to (19) or (20), is still computationally intensive and memory demanding, since the update requires both the estimation of the current state and the prediction of the future evolution. The probability distributions of initial state, process noise, measurement noise and transmission errors are all involved in the estimation and the prediction procedure. In practice, the complexity often becomes a limiting factor that requires simple encoder design. In the extreme case, a grid search can be used such that the amount of computations is governed by grid resolution. The search efficiency can be improved by carefully selecting search groups and search criteria. Consider the problem of optimizing the prediction error encoder. A simple but reasonable search group are uniform encoders around the origin, by fact that the $\tilde{\mathbf{y}}_{t_e}^{t^-}$'s are zero-mean random variables. Since in the case of open-loop encoders, Bao *et al.* (2007a), uniform encoders are usually not as good, the prediction-error encoder can be said to be more practical. According to Section 4, we know that an optimal prediction error encoder $\tilde{\mathbf{f}}_0^{T-1}$ minimizes the overall cost $\mathbf{E}\left\{\sum_{s=0}^{T-1}\tilde{x}_s'\pi_{T-s}\tilde{x}_s\right\}$, irrespective of the controller. In the general case with partial side-information, the overall system

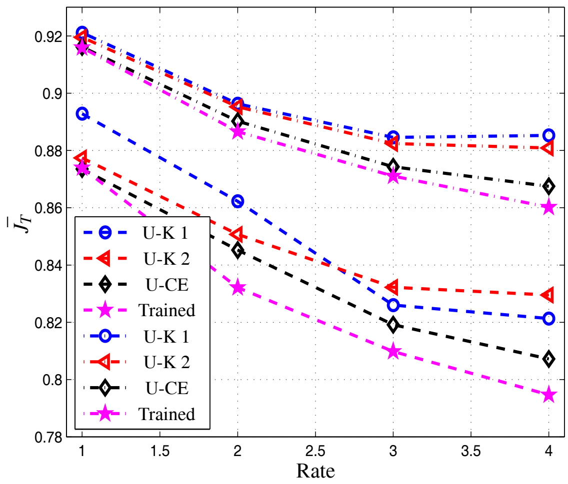

Fig. 3. Performance comparison among various set-ups, while dashed lines for Set-up 1 and dash–dot lines are for Set-up 2. The channel transition probability ε is 0.1.

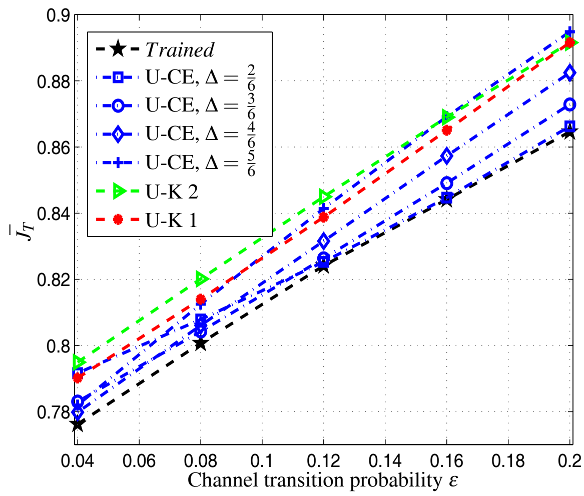

Fig. 4. A performance comparison among jointly and separately designed controller-encoders. The scheme U-CE employs a uniform encoder and a CE controller, where Δ is the step-size of the uniform encoder. The scheme U-K employs a uniform encoder and a Kalman filter.

performance can also be compared according to that cost, for various encoders, but fixed controller. When the encoder is close to the optimal solution, the loss in performance becomes insignificant.

7. NUMERICAL EXAMPLES

Here we present numerical experiments to demonstrate the performance obtained by using various encoder–controller design methods. We study a special case of the general system in Fig. 1, namely, a scalar system in which the current measurement is encoded and transmitted over a Binary Symmetric Channel (BSC). The system equations and the LQ cost are,

$$x_{t+1} = ax_t + u_t + v_t, \quad y_t = x_t + e_t, \quad J_T = x_T^2 + \sum_{t=0}^{T-1} x_t^2 + \rho u_t^2.$$

The initial state x_0, process noise v_t and measurement noise e_t are all zero-mean Gaussian distributed. The encoder only has access to the latest state measurement, i.e., $M_t = 0$. Four types of coding–control schemes are compared. The first type,

Set-up	a	T	ρ	$P_r(x_0)$	$P_r(v_t)$	$P_r(e_t)$	M_t
1	0.7	3	0.5	$\mathscr{N}(0,5)$	$\mathscr{N}(0,1)$	$\mathscr{N}(0,1)$	0
2	0.7	3	0.5	$\mathscr{N}(0,5)$	$\mathscr{N}(0,1)$	$\mathscr{N}(0,4)$	0

Table 1. The set-ups of the experiments.

U-K 1, employs a time-invariant uniform quantizer with a step-length Δ. At the controller, the decoded symbols are fed into a Kalman filter to estimate x_t. The Kalman filter is designed assuming the channel is absent and the system is not exposed to the quantization and transmission errors. Finally, the control is a linear function of the Kalman filter output that the linear coefficient is ℓ_t in (18). The second type, U-K 2 also utilizes a time-invariant uniform quantizer and a Kalman filter. The Kalman filter here is designed assuming the error due to measurement noise, quantization and transmission is white and Gaussian distributed. Still, the control is a linear function of ℓ_t and the Kalman filter output. By this method, the distortion due to quantization and channel error is treated as a part of measurement noise. Note that, the "extended" measurement noise (including quantization error and channel error) is in fact neither Gaussian nor uncorrelated with the state and the process noise, making the Kalman filter a sub-optimal estimator. The third type of encoder–controller is referred to as U-CE, which adopts a time-invariant uniform encoder and a CE controller. The last pair is an encoder–controller trained according to Section 6.1 where the encoder has full side-information.

The system performance is determined by a number of parameters. The relations among them are complicated. In Fig. 3 we demonstrate the particular impacts imposed by channel rate and measurement noise. The Performance measure $\bar{J}_T$ is obtained by normalizing $\mathbf{E}\{J_T\}$ with the expected cost obtained when no control action is taken. The system parameters are given in Table 1. For all uniform encoders, we let the boundaries be kept equally spaced between -2 and 2. Accordingly, the maximum quantization error in the saturated region decreases with the increasing transmission rate. It can be seen in the figure that the trained encoder–controller pair outperforms the other three coding–control schemes. Given same encoder, the system employing a CE controller always performs better than the systems employing the Kalman filters. The gain obtained by the trained encoder–controller appears to be mostly attributed to the CE controller. An interesting observation is that U-K 2 is not necessarily always superior to U-K 1. That means, the way U-K 2 handles the quantization distortion and transmission errors may do more harm than good.

In Fig. 4, we show system performance versus the channel transition probability ε. The system parameters are as given in Set-up 1 in Table 1. A comparison of all 4 types of coding–control schemes is depicted. Especially, the U-CE scheme is displayed for several step lengths. From the figure we see that the deterioration in system performance can be insignificant, even when using a simple time-invariant uniform encoder. Note that the time-invariant uniform encoder can provide a performance bound for the encoder with no side-information. However, all uniform encoders in Fig. 4 have near-optimal Δ. When Δ is chosen improperly, can have severe consequences.

8. CONCLUSION

This paper investigates optimization of the encoder and controller in closed-loop control of a linear plant with low-rate feedback over a memoryless noisy channel. We have shown that the CE controller is optimal for the class of encoders referred to as "measurement prediction error encoders." The CE controller can also be optimal for special encoders with full side-information. We used these results to motivate design criteria for the general case of only partial encoder side-information. We performed numerical investigations to compare the joint training design and the separate design of the encoder–controller.

REFERENCES

M. Aoki. Optimization of Stochastic Systems - Topics in Discrete-Time Systems. Academic Press, 1967.

J. Baillieul. Feedback coding for information-based control: operating near the data-rate limit. *Proc. of the 41st IEEE Conference on Decision and Control*, pages 3229 - 3236, 2002.

L. Bao, M. Skoglund and K.H. Johansson. On Optimal System Design for Feedback Control Noisy Channels. *Proc. of 2007 IEEE International Symposium on Information Theory*, pages 2486–2490, 2007.

L. Bao, M. Skoglund and K.H. Johansson. Joint Encoder–Controller Design for Feedback Control over Noisy Channels. *IEEE transactions on Automatic Control*, submitted.

V.S. Borkar, S.K. Mitter and S. Tatikonda. Optimal Sequenctial Vector Quantization of Markov Sources. *SIAM J. Control Optim.*, volume 40, number 1, pages 135–148, 2001.

R.W. Brockett, D. Liberzon. Quantized feedback stabilization of linear systems. *IEEE transactions on Automatic Control*, volume 45, number 7, pages 1279–1289, 2000.

N. Elia, S.K. Mitter. Stabilization of Linear Systems with Limited Information. *IEEE transactions on Automatic Control*, volume 46, number 9, pages 1384–1400, 2001.

F. Fagnani, S. Zampieri. Stability Analysis and Synthesis for Scalar Linear Systems with a Quantized Feedback. *IEEE transactions on Automatic Control*, volume 48, number 9, pages 1569–1584, 2003.

N. Farvardin. A Study of Vector Quantization for Noisy Channels. *IEEE Transactions on Information Theory*, volume 36, number 4, pages 799–809, 1990.

A. Mahajan, D. Teneketzis. Fixed Delay Optimal Joint Source–Channel Coding for Finite–Memory Systems. *Proc. of 2006 IEEE International Symposium on Information Theory*, pages 2319–2323, 2006.

A.S. Matveev, A.V. Savkin. An analogue of Shannon information theory for detection and stabilization via noisy discrete communication channels. *SIAM J. Control Optm.*, volume 46, number 4, pages 1323–1367, 2007.

A.S. Matveev, A.V. Savkin. The problem of LQG optimal control via a limited capacity communication channel. *Systems and Control Letters*, volume 53, pages 51–64, 2004.

G.N. Nair, F. Fagnani, S. Zampieri and R.J. Evans. Feedback Control under Data Rate Constraints: an Overview. *Proc. of the IEEE*, volume 95, number 1, pages 108–137, 2007.

A. Sahai, S. Mitter. The Necessity and Sufficiency of Anytime Capacity for Stabilization of a Linear System Over a Noisy Communication Link Part I: Scalar Systems. *IEEE transactions on Information Theory*, 2006.

S. Tatikonda, S. Mitter. Control over Noisy channels. *IEEE transactions on Automatic Control*, volume 49, number 7, pages 1196-1201, 2004.

S. Tatikonda, A. Sahai, S. Mitter. Stochastic linear control over a communication channel. *IEEE transactions on Automatic Control*, volume 49, number 9, pages 1549-1561, 2004.