Remote estimation with noisy measurements subject to packet loss and quantization noise

S. Dey, A. Chiuso, L. Schenato

***Abstract*—In this paper we consider the problem of designing coding and decoding schemes to estimate the state of a scalar stable stochastic linear system subject to noisy measurements and in the presence of a wireless communication channel between the sensor and the estimator. In particular, we consider a communication channel which is prone to packet loss and includes quantization noise due to its limited capacity. We study two scenarios: the first with channel feedback and the second with no channel feedback. More specifically, in the first scenario the transmitter is aware of the quantization noise and the packet loss history of the channel, while in the second scenario the transmitter is aware of the quantization noise only. We show that in the first scenario, the optimal strategy among all possible linear encoders corresponds to the transmission of the Kalman filter innovation similarly to the differential pulse-code modulation (DPCM). In the second scenario, we show that there is a critical packet loss probability above which it is better to transmit the state rather than the innovation. We also propose a heuristic strategy based on the transmission of a convex combination of the state and the Kalman filter innovation which is shown to provide a performance close to the one obtained with channel feedback.**

***Index Terms*—Kalman filtering, packet loss, quantization noise, channel feedback, differential encoding**

I. Introduction

Wireless communication has become ubiquitous and wired communication systems are increasingly being replaced with wireless systems thanks to their many advantages such as smaller installation costs, easier maintenance and fewer cumbersome cables. However, wireless communication comes at the price of lower channel capacity which results in higher quantization noise, packet losses and delay. This concern is particularly apparent in industrial applications such as remote sensing and real-time automation, since a very high level of reliability is needed in control systems and safety-critical scenarios. As a consequence, it becomes of paramount importance to understand the impact of realistic channel models in the context of estimation and control. So far most of the works available in the literature have concentrated on stability and control subject to only one specific limitation of wireless communication. For example, in [1], [2] the authors addressed the problem of stabilization of an unstable plant through a rate-limited erasure channel where no performance index is considered besides stability. Other researchers have tried to tackle the channel limitations by using analog models in order to avoid the difficulties associated with explicit design of digital channel encoder/decoder and to optimize some performance metrics among all possible stabilizing controllers subject to packet loss [3], [4] or subject to a maximum signal-to-noise ratio (SNR) [5], [6]. Finally, another well explored approach is the analysis, under an LQG framework, of control systems subject to random packet loss, quantization [7], [8], [9], [10] and possibly delays [11]. All these works have been concerned with stability in control systems. However, there are many applications, such as remote sensing and estimation, where the dynamical system to be controlled is already stable, but the existing communication and feedback performance can be substantially improved. In this work we are interested in exploring the problem of remotely estimating the state of a stable stochastic scalar linear system over a wireless channel. In particular, we want to design coding and decoding strategies that allow good estimation performance in the presence of packet loss, quantization noise and measurement noise. So far, mainly packet loss has been considered in the context of remote estimation [12], [13], although there are recent attempts to consider both limitations [14], [15], [16], [17]. Note that the focus in [16], [17] are on deriving minimum data rates for stabilizability over lossy channels, whereas we focus on the actual estimation error performance in the presence of quantization (data rate constraints) and packet loss. In particular we explore two scenarios. In the first scenario the transmitter has perfect channel feedback, i.e. it is aware of possible packet losses and therefore it is able to replicate the receiver filter. As a result, we show that the optimal transmission strategy is to send the innovation between the best estimate of the state at the filter and the predicted estimate of the state at the receiver. This is reminiscent of differential pulse-code modulation (DPCM) [18] in which a differential signal is sent over a channel with no packet loss. Differently, in the second scenario, we consider the case when the transmitter is not aware of the packet loss history. We propose three strategies: the first named state forwarding (SF) in which the estimated state is transmitted over the channel, the second named innovation forwarding (IF), in which the difference between the state and the estimate that a receiver would have if no packet loss had occurred is sent; the third one, named soft innovation forwarding (SIF), transmits a convex combination of the signals mentioned above and thus includes SF and IF as special cases. For these three strategies we compute their performance and observe that in the low packet loss regime it is better to use strategies that are similar to the IF, while for high packet loss regime it better to use strategies that are similar to the SF. Some preliminary results, which considered the simplified scenario with no measurement

This work is supported by by the European Community's Seventh Framework Programme [FP7/2007-2013] under grant agreement n. 257462 HYCON2 Network of excellence, by the FIRB project Learning meets time (RBFR12M3AC) and by the Australian Research Council grant n. DP120101122.

S. Dey is with the Division of Signals & Systems, Uppsala University, P. O. Box 534, Uppsala SE-75121, Sweden, email: Subhra.Dey@signal.uu.se, while A. Chiuso and L. Schenato are in the Department of Information Engineering, University of Padova, Via Gradenigo 6/b, 35131 Padova, Italy, email: {chiuso|schenato}@dei.unipd.it

noise, can be found in [19].

II. Channel modeling and Problem formulation

We consider the problem of remotely estimating the state of a scalar linear stochastic dynamical system:

$$x_{t+1} = ax_t + w_t \quad (1)$$
$$y_t = cx_t + v_t \quad (2)$$

where $w_t \sim \mathcal{N}(0, \sigma_w^2)$, $v_t \sim \mathcal{N}(0, \sigma_v^2)$ are white, uncorrelated and uncorrelated with the initial condition $x_0 \sim \mathcal{N}(\bar{x}_0, \sigma_0^2)$. More specifically, as graphically depicted in Figure 1, the analogue measurement y_t at the sensor can be pre-processed by the filter $g(\cdot)$ into the analog signal s_t before transmission. The signal is then quantized into a word s_t^q from a finite alphabet, which is then coded and transmitted over a digital noisy channel. At the receiver, the channel decoder either perfectly decodes the word s_t^q or detect an erasure which is modeled by the binary variable $\gamma_t \in \{0,1\} \equiv \{erased, decoded\}$. If correctly decoded, the word s_t^q is converted into the analog signal z_t, which is then processed by the receiver via the filter $h(\cdot)$ to provide the state estimate $\widehat{x}_t$. The transmission protocol might be provided with an ACK-based system that notifies the transmitter whether the packet has been successfully decoded at the receiver. We refer to this scenario as *perfect channel feedback*; if the ACK signal is not available we shall sat that there is *no channel feedback*. We now proceed to mathematically model such system.

In the following we will consider the simplified assumption

$$c = 1, \quad |a| < 1 \quad (3)$$

where the first assumption can be used w.l.o.g. since the case $c \neq 1$ can be easily obtained via a rescaling of the process noise variance σ_w^2, while the second assumption is necessary to guarantee that the stochastic signal y_t is asymptotically stationary with bounded variance. The transmitter can send a signal through a digital noisy erasure channel modeled as follows

$$z_t = \gamma_t s_t^q = \gamma_t(s_t + n_t)$$

where $\gamma_t \in \{0,1\}$ represents the erasure event, $s_t^q \in \mathbb{R}$ is the quantized transmitted signal, $s_t \in \mathbb{R}$ is the signal before quantization, and n_t is the uncorrelated additive noise which models the quantization error under a fine quantization assumption.

Remark 1: The validity of the additive quantization noise model for high rate uniform scalar quantization has been rigorously shown in [20] for continuous input densities, and see also [21] for similar studies. It has been however shown in these papers as well as many other recent literature such as in [22] that although in principle only high rate quantization theory justifies such an additive white quantization noise model, in practice this model holds as a very good approximation for moderate rate quantization. If fact, as shown later in via numerical simulations Section VI, a uniform scalar quantizer with only 3-4 bits of quantization per sample used to quantize the signal s_t provides results that are sufficiently close to the theoretical values based on additive noise model proposed in this work. Note that in a wireless local area network (WLAN) with orders of megabits per second data rates (even when shared amongst various links), it is not unreasonable to expect 3-4 bits per sample with a sampling rate of say 0.1 MHz which is likely to be sufficient for most physical dynamical systems. Thus, this additive white quantization noise model is also suitable for use in practical implementation of estimation over lossy wireless links.

The variables satisfy the following assumptions:

$$\mathbb{P}[\gamma_t = 0] = \epsilon, \qquad n_t \sim \mathcal{N}\left(0, \frac{1}{\Lambda}\mathbb{E}[s_t^2]\right)$$

where Λ is the signal-to-quantization noise ratio (SQNR) of the quantizer; $\{\gamma_t\}$ and $\{n_t\}$ are assumed to be independent. This model for the SQNR noise assumes that the quantizer is matched to the stationary distribution of the incoming signal s_t so as to maintain a constant SQNR value Λ. The transmitter sends a signal according to its available information set, i.e. $s_t = g_t(\mathcal{T}_t)$ where g_t is a measurable function of the information set $\mathcal{T}_t$ which can take the following two forms:

$$\begin{aligned}
\mathcal{T}_t^{CF} &= \{y_t, .., y_0, s_{t-1}, .., s_0, n_{t-1}, .., n_0, \gamma_{t-1}, .., \gamma_0\} \\
&= \{y_t, .., y_0, s_{t-1}, .., s_0, z_{t-1}, .., z_0, \gamma_{t-1}, .., \gamma_0\} \\
\mathcal{T}_t^{NCF} &= \{y_t, .., y_0, s_{t-1}, .., s_0, n_{t-1}, .., n_0\}
\end{aligned}$$

The first set $\mathcal{T}^{CF}$ corresponds to a scenario with perfect channel feedback where the transmitter knows the sequence $\{\gamma_{t-1}, .., \gamma_0\}$, i.e. whether a packet has been received successfully or not, while the second set $\mathcal{T}^{NCF}$ has no such information. The first scenario is realistic in wireless communication where the receiver can transmit back a signal with higher power and therefore very small packet loss probability. Moreover, the information to be sent back reliably is just an ACK packet. For convenience of notation and future use we define the symbol $\mathbb{E}_\gamma$ which denotes expectation taken conditionally on the entire loss sequence γ. Moreover we define

$$\mathcal{Z}_t := \{z_t, .., z_0\} \quad \mathcal{R}_t := \{z_t, \dots, z_0, \gamma_t, \dots \gamma_0\}$$

which correspond to the past history of the received signals. Then the state estimator at the receiver side based on the information $\mathcal{R}_s$ is given by

$$\widehat{x}_{t|s}^{rx} := \mathbb{E}[x_t | \mathcal{R}_s] = \mathbb{E}_\gamma[x_t | \mathcal{Z}_s] \quad (4)$$

Under our Gaussian assumption on the initial condition and noises, $h_\gamma(\mathcal{Z}_t) := \widehat{x}_{t|t}^{rx}$ is a linear function of $\mathcal{Z}_t$ which depends on the loss sequence $\gamma_t, .., \gamma_0$. We are interested in analyzing the performance of the overall system based on the estimation prediction error variance at the receiver, i.e.

$$p_{t+1|t}^{rx} = \mathbb{E}[(x_{t+1} - \widehat{x}_{t+1|t}^{rx})^2]$$

where the expectation has to be taken also with respect to the packet drop process γ_t besides the noises w_t, n_t. As a result, we will assume that the delay necessary to deliver a message from the transmitter to the receiver is smaller or equal that the sampling period, i.e. one time-step.

For future use let us also define the measurement history $\mathcal{Y}_t := \{y_t, .., y_0\}$, the state estimator at the transmitter side

$$\widehat{x}_{t|t}^{tx} := \mathbb{E}[x_t | \mathcal{Y}_t] = a\widehat{x}_{t-1|t-1}^{tx} + \hat{k}_t(y_t - a\widehat{x}_{t-1|t-1}^{tx}) \quad (5)$$

where $\widehat{k}_t$ is the optimal filter gain, and the estimator error

$$\widetilde{x}_{t|t}^{tx} := x_t - \widehat{x}_{t|t}^{tx} \quad (6)$$

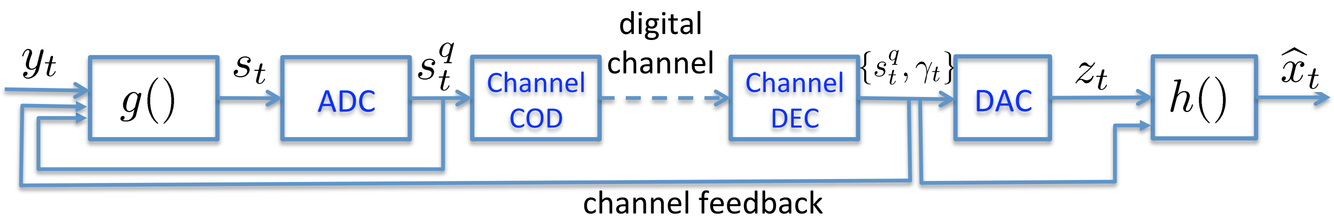

Fig. 1. Equivalent communication model for remote estimation.

Since the system is asymptotically stable[1] the variance $p^{tx}_{t|t} = \mathbb{E}[(\widetilde{x}^{tx}_{t|t})^2]$ has the property that $\lim_{t\to\infty} p^{tx}_{t|t} = p^{tx}_{\infty}$, $\lim_{t\to\infty} \widehat{k}_t = \widehat{k}$ where p^{tx}_{∞} is the unique non-negative solution of the following filter Riccati equation and $\widehat{k}$ its corresponding steady state gain:

$$p^{tx}_{\infty} = a^2 p^{tx}_{\infty} + \sigma_w^2 - \frac{(a^2 p^{tx}_{\infty} + \sigma_w^2)^2}{a^2 p^{tx}_{\infty} + \sigma_w^2 + \sigma_v^2} = \sigma_v^2 \frac{a^2 p^{tx}_{\infty} + \sigma_w^2}{a^2 p^{tx}_{\infty} + \sigma_w^2 + \sigma_v^2} \quad (7)$$

$$\hat{k} = \frac{a^2 p^{tx}_{\infty} + \sigma_w^2}{a^2 p^{tx}_{\infty} + \sigma_w^2 + \sigma_v^2} \quad (8)$$

which shows that for $\sigma_v^2 = 0$, then also $p^{tx}_{\infty} = 0$. In fact, in this scenario $\widehat{x}_{t|t} = x_t$.

III. Optimal Estimation with Perfect Channel Feedback

We now consider the state estimation problem with perfect channel feedback, i.e. also the transmitter is aware of the packet loss sequence incurred across the digital channel. We show that if we restrict our attention to functions $g(\mathcal{T}_t^{CF})$ and $h(\mathcal{R}_t)$ which are linear in the information sets $\mathcal{T}_t^{CF}$ and $\mathcal{R}_t$, then the optimal strategy is to send the state estimate innovation, i.e. the difference between the current best state estimate at the transmitter and the current best prediction of the state at the receiver.

A. Optimal strategy derivation

Our purpose is to find the "optimal" message s_t to be sent through a lossy and SQNR limited channel in order to minimize the state estimation error variance at the receiver, under the assumption that perfect channel feedback is available. We shall look for conditionally linear encoders[2]

$$s_t := \mathcal{L}_\gamma\left(\mathcal{Y}_t, \mathcal{Z}_{t-1}\right) \quad (9)$$

where $\mathcal{L}_\gamma\left(\mathcal{Y}_t, \mathcal{Z}_{t-1}\right)$ is, conditionally on the packet loss sequence $\gamma_{t-1}, ..., \gamma_0$, a linear operator of its arguments $y_t, y_{t-1}, .., y_0$ (the samples to be encoded) and $z_{t-1}, .., z_0$ (the past received signals). The result of this section is summarized in the next theorem. The remaining part of the section proves the result.

Theorem 1: Under the assumption that perfect channel feedback is available (i.e. that $\gamma_{t-1}, ..., \gamma_0$ are known also at the transmitter side), the optimal linear encoder (9) for the linear system (1)-(2) is given by:

$$s_t := \widehat{x}^{tx}_{t|t} - \widehat{x}^{rx}_{t|t-1} = \mathbb{E}[x_t|\mathcal{Y}_t] - \mathbb{E}_\gamma[x_t|\mathcal{Z}_{t-1}] \quad (10)$$

Proof: The encoder has to find a linear function of all available measurements which retains as much information as possible regarding the state to be estimated. We can define

$$e_s := y_s - \mathbb{E}_\gamma[y_s|\mathcal{Z}_{t-1}] \quad (11)$$

which represents the innovation (i.e. the "new" information) in y_s which is not already contained in $\mathcal{R}_{t-1} = \{\gamma_{t-1}, .., \gamma_0, \mathcal{Z}_{t-1}\}$. Then we define $\mathcal{E}_t := \{e_t, ..., e_0\}$.

Note that, however, only part of this information is necessary to estimate x_t. As a matter of fact $\mathcal{E}_t$ can be reduced so as to retain all and only the relevant information on x_t; this reduction has sometimes been called *Sufficient Dimensionality Reduction* (SDR) [23]. Since x_t is scalar, the (linear) sufficient statistic in $\mathcal{E}_t$ for x_t has dimension 1 (which is equal to the dimension of the projection of x_t onto the space spanned by the elements of $\mathcal{E}_t$).

Hence we seek for a signal $s_t = \sum_{i=0}^t \alpha_i e_{t-i}$, $\alpha_i \in \mathbb{R}$, so that the optimal estimation

$$\widehat{x}^{rx}_{t|t} := \mathbb{E}_\gamma[x_t|\mathcal{Z}_t]$$

has as small (conditional) variance as possible.

Note that the "noise" n_t is known at the transmitter side since the transmitter generates s_t^q starting from s_t. Since $s_t = \sum_{i=0}^t \alpha_i e_{t-i}$, and using the fact that both the noise n_t and e_s are uncorrelated from z_s, $s < t$ (see also (11)), also z_t is uncorrelated from z_s, $s < t$. Therefore the estimator $\widehat{x}^{rx}_{t|t} := \mathbb{E}_\gamma[x_t|\mathcal{Z}_t]$, satisfies:

$$\begin{aligned} \widehat{x}^{rx}_{t|t} &= \mathbb{E}_\gamma[x_t|\mathcal{Z}_{t-1}] + \mathbb{E}_\gamma[x_t|z_t] = \widehat{x}^{rx}_{t|t-1} + \mathbb{E}_\gamma[x_t|z_t] \\ &= \widehat{x}^{rx}_{t|t-1} + \frac{\mathbb{E}_\gamma[x_t s_t]}{\mathbb{E}_\gamma[s_t^2](1+\frac{1}{\Lambda})} z_t = \widehat{x}^{rx}_{t|t-1} + \frac{1}{1+\frac{1}{\Lambda}} z_t \end{aligned} \quad (12)$$

Note now that, defining $\tilde{x}^{rx}_{t|t} := x_t - \widehat{x}^{rx}_{t|t}$ we have

$$Var_\gamma\{\tilde{x}^{rx}_{t|t}\} = Var_\gamma\{\tilde{x}^{rx}_{t|t-1}\} - Var_\gamma\{\mathbb{E}[x_t|z_t]\}$$

where the symbol Var_γ denotes the variance conditionally on the sequence $\{\gamma_t\}$. Since the choice of s_t does not affect the first term on the right hand side, minimizing $Var\{\tilde{x}^{rx}_{t|t}\}$ is equivalent to maximizing

$$Var_\gamma\{\mathbb{E}_\gamma[x_t|z_t]\} = \gamma_t \frac{(\mathbb{E}_\gamma[x_t s_t])^2}{\mathbb{E}_\gamma[s_t^2](1+\frac{1}{\Lambda})} = \gamma_t \frac{(\mathbb{E}_\gamma[x_t \bar{s}_t])^2}{(1+\frac{1}{\Lambda})}$$

where $\bar{s}_t := \frac{s_t}{\sqrt{\mathbb{E}_\gamma[s_t^2]}}$. Hence we are left with maximizing $\mathbb{E}_\gamma[x_t \bar{s}_t]$, which is obtained choosing α_i, $i = 0, .., t$ so that

[1] Note that this is not necessary and milder stabilizability and detectability conditions are sufficient for the state estimation error variance to be the unique positive semidefinite and bounded solution of the algebraic Riccati equation (7).

[2] We restrict to linear functionals because the stochastic system is conditionally Gaussian given the loss sequence $\{\gamma_t\}$ and, therefore, the optimal estimator conditionally on $\{\gamma_t\}$ is a linear functional of the observed data.

$s_t = \sum_{i=0}^t \alpha_i e_{t-i}$ has maximal correlation with x_t. This is achieved when[3] $s_t := \mathbb{E}_\gamma[x_t|\mathcal{E}_t] = \mathbb{E}_\gamma[x_t|\mathcal{E}_t, \mathcal{Z}_{t-1}] - \mathbb{E}_\gamma[x_t|\mathcal{Z}_{t-1}] = \mathbb{E}_\gamma[x_t|\mathcal{Y}_t, \mathcal{Z}_{t-1}] - \mathbb{E}_\gamma[x_t|\mathcal{Z}_{t-1}] = \widehat{x}^{tx}_{t|t} - \widehat{x}^{rx}_{t|t-1}$.

Hence, the optimal signal to be sent through the SQNR-limited channel is

$$s_t := \widehat{x}^{tx}_{t|t} - \widehat{x}^{rx}_{t|t-1} = \mathbb{E}[x_t|\mathcal{Y}_t] - \mathbb{E}_\gamma[x_t|\mathcal{Z}_{t-1}]$$

which concludes the proof. ■

The Kalman filter estimation error $\widetilde{x}^{tx}_{t|t}$ in (6) has some interesting uncorrelation properties, which will be useful in the forthcoming analysis, that are summarized in the following Lemma:

Lemma 1: In the perfect channel feedback scenario, the Kalman filter estimation error $\widetilde{x}^{tx}_{t|t}$ is conditionally uncorrelated with respect to both the transmitter and the receiver estimation error, i.e.

$$\mathbb{E}_\gamma[\widetilde{x}^{tx}_{t|t}\widehat{x}^{tx}_{t|t}] = 0, \quad \mathbb{E}_\gamma[\widetilde{x}^{tx}_{t|t}\widehat{x}^{rx}_{t|t}] = 0$$

Proof: A well known property of the optimal estimation error $\widetilde{x}^{tx}_{t|t}$ is that it is uncorrelated to any linear function of the same data based on which it is constructed, i.e. $\mathbb{E}[\widetilde{x}^{tx}_{t|t}\mathcal{L}(\{y_h\}_{h=0}^t)] = 0$, from which it directly follows that $\mathbb{E}[\widetilde{x}^{tx}_{t|t}\widehat{x}^{tx}_{t|t}] = 0$. Conditioned an a specific realization of the packet loss sequence $\{\gamma_t\}$, the estimator at the receiver is a linear function of the received data, i.e. $\widehat{x}^{rx}_{t|t} = \mathcal{L}_\gamma(\{z_h\}_{h=0}^t)$. Since $z_t = \gamma_t(\widehat{x}^{tx}_{t|t} - \widehat{x}^{rx}_{t|t-1} + n_t)$, by linearity we can certainly write the estimator at the receiver as

$$\widehat{x}^{rx}_{t|t} = \mathcal{L}'_\gamma(\{y_h\}_{h=0}^t) + \mathcal{L}''_\gamma(\{n_h\}_{h=0}^t)$$

where $\mathcal{L}'_\gamma$ and $\mathcal{L}''_\gamma$ are linear functions conditionally on the loss sequence. Since $\widetilde{x}^{tx}_{t|t}$ is uncorrelated with the noise sequence $\{n_t\}$, the statement of the first part of the lemma easily follows. ■

B. Performance analysis

Based on the analysis in the previous subsection, the optimal linear strategy for remote estimation in the presence of channel feedback, which is graphically represented as in Fig.2, is the following: at the transmitter the measurements are first

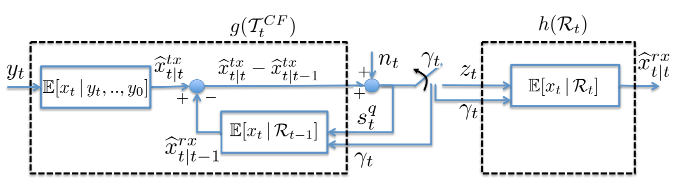

Fig. 2. Remote estimation scheme with perfect channel feedback

preprocessed by a standard Kalman filter to obtain the best estimate of the state at the transmitter $\widehat{x}^{tx}_{t|t}$ (see (5)) as well as to reconstruct the best prediction at the receiver side $\widehat{x}^{rx}_{t|t-1}$, see (4).

Once again, based on the previous section, the optimal strategy at the transmitter is to send the innovation $s_t = \widehat{x}^{tx}_{t|t} - \widehat{x}^{rx}_{t|t-1}$ from which it follows that the signal received at the remote estimator is

$$z_t = \gamma_t(\widehat{x}^{tx}_{t|t} - \widehat{x}^{rx}_{t|t-1} + n_t) = \gamma_t(x_t - \widetilde{x}^{tx}_{t|t} - \widehat{x}^{rx}_{t|t-1} + n_t).$$

According to the standard MMSE theory for linear systems, the optimal filter equation must be of the form:

$$\widehat{x}^{rx}_{t|t-1} = a\widehat{x}^{rx}_{t-1|t-1} \quad (13)$$
$$\widehat{x}^{rx}_{t|t} = \widehat{x}^{rx}_{t|t-1} + k_t(z_t - \widehat{z}_{t|t-1}) \quad (14)$$

where we used the result from Eqn. (12). The expression of the optimal Kalman gain k_t is given by[4]:

$$k_t = cov_\gamma\{x_t, z_t - \widehat{z}_{t|t-1}\}Var_\gamma^{-1}\{z_t - \widehat{z}_{t|t-1}\} = \frac{\Lambda}{\Lambda+1} \quad (15)$$

which is independent of time and of the packet loss sequence. If we define the estimation error as $\widetilde{x}^{rx}_{t|h} = x_t - \widehat{x}^{rx}_{t|h}$ and its corresponding variance as $p^{rx}_{t|h} = \mathbb{E}[(\widetilde{x}^{rx}_{t|h})^2]$ we get

$$\widetilde{x}^{rx}_{t+1|t} = a(1-\gamma_t k_t)\widetilde{x}^{rx}_{t|t-1} + w_t + \gamma_t a k_t(\widetilde{x}^{tx}_{t|t} - n_t)$$

Note now that, using also Lemma 1, $\mathbb{E}[n_t^2] = \frac{1}{\Lambda}\mathbb{E}[(\widetilde{x}^{rx}_{t|t-1} - \widetilde{x}^{tx}_{t|t})^2] = \frac{1}{\Lambda}\mathbb{E}[(\widetilde{x}^{rx}_{t|t-1})^2 - 2\widetilde{x}^{tx}_{t|t}\widetilde{x}^{rx}_{t|t-1} + (\widetilde{x}^{tx}_{t|t})^2)^2] = \frac{1}{\Lambda}(p^{rx}_{t|t-1} - 2p^{tx}_{t|t} - p^{tx}_{t|t}) = \frac{1}{\Lambda}(p^{rx}_{t|t-1} - p^{tx}_{t|t})$ and $\mathbb{E}[\widetilde{x}^{rx}_{t|t-1}\widetilde{x}^{tx}_{t|t}] = \mathbb{E}[(\widetilde{x}^{tx}_{t|t})^2] = p^{tx}_{t|t}$. Using also that $k_t = \frac{\Lambda}{\Lambda+1}$, then the receiver error (unconditional) variance is given by:

$$p^{rx}_{t+1|t} = a^2 p^{rx}_{t|t-1} + \sigma_w^2 - (1-\epsilon)\frac{a^2\Lambda}{1+\Lambda}(p^{rx}_{t|t-1} - p^{tx}_{t|t})$$

Since $|a| < 1$ the previous linear equation has a steady state solution given by:

$$p^{CF}(\epsilon) = \lim_{t\to\infty} p^{rx}_{t+1|t} = \frac{\sigma_w^2 + (1-\epsilon)\frac{a^2\Lambda}{\Lambda+1}p^{tx}_\infty}{1 - a^2\frac{1+\epsilon\Lambda}{1+\Lambda}} \quad (16)$$

which represents the steady state predictor error variance.

IV. State forwarding vs Innovation forwarding with no channel feedback

In this section we consider the challenging scenario where no channel feedback is present. In this case the information set at the transmitter $\mathcal{T}_t^{NCF}$ does not include the information set at the receiver $\mathcal{R}_t$, i.e. $\mathcal{R}_t \not\subset \mathcal{T}_t^{NCF}$. As consequence, the transmitter cannot produce a copy of the transmitter estimate $\widehat{x}^{rx}_{t|t-1}$. The optimal strategy in this case is not obvious and it is likely to be a non-linear function of the information sets $\mathcal{T}_t^{NCF}, \mathcal{R}_t$. This situation is reminiscent of the loss of separation principle in control systems where the estimator is not aware if the control input has been successfully received by the actuator or not [9].

As a consequence, we explore suboptimal linear strategies for which is it possible to compute the performance. In particular, there are two suboptimal naive strategies that can be proposed. The first strategy, that we refer to as *state forwarding (SF)* is to simply transmit the current transmitter best estimate of the state x_t, i.e. $s_t = \widehat{x}^{tx}_{t|t}$.

[3]The chain of equalities can be obtained recalling that, conditionally on the loss sequence $\{\gamma_t\}$, all random variables are jointly Gaussian and, as such, conditional expectations are linear projections. In addition recall that conditionally on γ, $\mathcal{E}_t$ is uncorrelated with $\mathcal{Z}_t$ given γ and the linear span of $\mathcal{E}_t, \mathcal{Z}_{t-1}$ equals that of $\mathcal{Y}_t, \mathcal{Z}_{t-1}$.

[4]The subscript γ reminds that covariances are taken conditionally on $\{\gamma_t\}$.

To introduce the second strategy, let us now define the state predictor at the transmitter side using the quantized signals as information set, i.e.

$$\overline{x}^{tx}_{t|t-1} := \mathbb{E}[x_t | s^q_{t-1}, s^q_{t-2}, ..., s^q_0] \tag{17}$$

Inspired by the optimal filtering scheme with channel feedback, which requires sending the difference between the state estimator at the transmitter side and the state prediction at the receiver side, we now define the *innovation forwarding (IF)* strategy in which the transmitted signal s_t is given by the difference between the state estimator at the transmitter side and the state predictor computed at the transmitter side assuming (incorrectly) that the all the past quantized transmitted signals $s^q_t = s_t + n_t$ have reached the receiver, i.e. assuming $\gamma_t = 1, \forall t$. More specifically $s_t = \widehat{x}^{tx}_{t|t} - \overline{x}^{tx}_{t|t-1}$. The rationale behind this strategy is that in a lossless channel, i.e. if $\epsilon = 0$, it provides the optimal strategy as discussed in Section III. For both transmitter strategies, the receiver will compute the MMSE estimator, i.e $\widehat{x}^{rx}_{t|t} = \mathbb{E}[x_t \,|\, \mathcal{R}_t]$. As just mentioned, in general $\widehat{x}^{rx}_{t|t} \neq \widehat{x}^{tx}_{t|t}$ and $\widehat{x}^{rx}_{t|t} \neq \overline{x}^{tx}_{t|t}$ These two strategies can be graphically represented as in Fig 3, where the SF strategy corresponds to $\nu = 1$ and the IF strategy to $\nu = 0$.

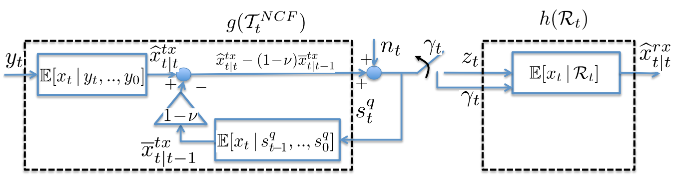

Fig. 3. Remote estimation scheme with no channel feedback

We now state an instrumental lemma which will be useful later on:

Lemma 2: In the scenario with no channel feedback the transmitter state estimation errors $\widetilde{x}^{tx}_{t|t}$ is conditionally uncorrelated with $\widehat{x}^{tx}_{t|t}$, $\overline{x}^{tx}_{t|t-1}$ and $\widehat{x}^{rx}_{t|t-1}$, i.e.

$$\mathbb{E}_\gamma[\widetilde{x}^{tx}_{t|t}\widehat{x}^{tx}_{t|t}] = 0, \;\; \mathbb{E}_\gamma[\widetilde{x}^{tx}_{t|t}\overline{x}^{tx}_{t|t-1}] = 0, \;\; \mathbb{E}_\gamma[\widetilde{x}^{tx}_{t|t}\widehat{x}^{rx}_{t|t}] = 0, \forall \nu$$

and $\widetilde{\overline{x}}^{tx}_{t|t-1} := x_t - \overline{x}^{tx}_{t|t-1}$ is conditionally uncorrelated with, $\overline{x}^{tx}_{t|t-1}$ and $\widehat{x}^{rx}_{t|t-1}$, i.e.

$$\mathbb{E}_\gamma[\widetilde{\overline{x}}^{tx}_{t|t-1}\overline{x}^{tx}_{t|t-1}] = 0, \;\; \mathbb{E}_\gamma[\widetilde{\overline{x}}^{tx}_{t|t-1}\widehat{x}^{rx}_{t|t}] = 0, \forall \nu$$

Proof: Since the estimator $\widetilde{x}^{tx}_{t|t}$ is not influenced by channel feedback, the first statement has been proven in Lemma 1.

The other two statements follow easily using the same arguments as in Lemma 1 since $\overline{x}^{tx}_{t|t-1}$ and $\widehat{x}^{rx}_{t|t-1}$ are linear functions of $\{y_k, k < t\}$ and of $\{n_k, k < t\}$, which are all uncorrelated with $\widetilde{x}^{tx}_{t|t}$ and $\widetilde{\overline{x}}^{tx}_{t|t-1}$. ■

A. State forwarding strategy ($\nu = 1$)

In this section, as seen before, we assume that the transmitted message is a noisy version of the estimated state, i.e. has the form

$$s^q_t = \widehat{x}^{tx}_{t|t} + n_t = x_t - \widetilde{x}^{tx}_{t|t} + n_t$$

where

$$\begin{aligned} \mathbb{E}_\gamma[n_t^2] &= \tfrac{1}{\Lambda}\mathbb{E}_\gamma[(\widehat{x}^{tx}_{t|t})^2] = \tfrac{1}{\Lambda}(\mathbb{E}[x_t^2] - \mathbb{E}_\gamma[(\widetilde{x}^{tx}_{t|t})^2]) \\ &= \tfrac{1}{\Lambda}\left(\tfrac{\sigma_w^2}{1-a^2} - \mathbb{E}_\gamma[(\widetilde{x}^{tx}_{t|t})^2]\right) \end{aligned}$$

and we assume that x_t has reached its steady state distribution. In fact $\lim_{t\to\infty}\mathbb{E}[x_t^2] = \frac{\sigma_w^2}{1-a^2} =: p_{OL}$ holds for $|a| < 1$; this in particular shows that the state forwarding strategy cannot be used for $|a| \geq 1$ since the signal variance and hence the quantization noise variance would diverge. The message received at the remote estimator is then

$$z_t = \gamma_t(\widehat{x}^{tx}_{t|t} + n_t) = \gamma_t(x_t - \widetilde{x}^{tx}_{t|t} + n_t)$$

which can be interpreted as a noisy measurement of the filtered state, where n_t is the measurement noise, subject to intermittent observation. This problem has already been solved in [12] and the solution is given by the following time-varying Kalman filter:

$$\widehat{x}^{rx}_{t|t-1} = a\widehat{x}^{rx}_{t-1|t-1} \tag{18}$$

$$\widehat{x}^{rx}_{t|t} = \widehat{x}^{rx}_{t|t-1} + \gamma_t k_t(z_t - \widehat{x}^{rx}_{t|t-1}) \tag{19}$$

The state estimation error then satisfies the equation

$$\widetilde{x}^{rx}_{t+1|t} = a(1-\gamma_t k_t)\widetilde{x}^{rx}_{t|t-1} + w_t + \gamma_t a k_t(\widetilde{x}^{tx}_{t|t} - n_t)$$

from which the conditional error covariance

$$\begin{aligned} \widehat{p}^{rx}_{t+1|t} = \; & a^2(1-\gamma_t k_t)^2\widehat{p}^{rx}_{t|t-1} + \gamma_t^2 a^2 k_t^2(\mathbb{E}[(\widetilde{x}^{tx}_{t|t})^2 + \mathbb{E}_\gamma[n_t^2]) \\ & + 2a^2\gamma_t k_t(1-\gamma_t k_t)\mathbb{E}_\gamma[\widetilde{x}^{tx}_{t|t}\widetilde{x}^{rx}_{t|t-1}] + \sigma_w^2 \end{aligned}$$

where $\widehat{p}^{rx}_{t+1|t} = \mathbb{E}_\gamma[(\widetilde{x}^{rx}_{t+1|t})^2]$.

Since by Lemma 2 $\mathbb{E}_\gamma[\widetilde{x}^{tx}_{t|t}\widetilde{x}^{rx}_{t|t-1}] = p^{tx}_{t|t}$ and $\mathbb{E}_\gamma[n_t^2] = \frac{1}{\Lambda}\mathbb{E}[(\widehat{x}^{tx}_{t|t})^2] = \frac{1}{\Lambda}(\mathbb{E}[x_t^2] - \mathbb{E}_\gamma[(\widetilde{x}^{tx}_{t|t})^2])$, then the optimal gain obtained by minimizing the right hand side is given by:

$$k_t = \frac{\widehat{p}^{rx}_{t|t-1} - p^{tx}_{t|t}}{\widehat{p}^{rx}_{t|t-1} - \frac{\Lambda+1}{\Lambda}p^{tx}_{t|t} + \frac{\sigma_w^2}{\Lambda(1-a^2)}}.$$

From which it follows:

$$\widehat{p}^{rx}_{t+1|t} = a^2\widehat{p}^{rx}_{t|t-1} + \sigma_w^2 - \gamma_t\frac{(\widehat{p}^{rx}_{t|t-1} - p^{tx}_{t|t})^2}{\widehat{p}^{rx}_{t|t-1} - \frac{\Lambda+1}{\Lambda}p^{tx}_{t|t} + \frac{1}{\Lambda}p^{OL}}$$

The optimal estimator could be computationally expensive since it needs to keep track of the conditional estimation error covariance $\widehat{p}^{rx}_{t|t-1}$ which is a function of the packet loss history $\{\gamma_h\}_{h=0}^{t-1}$. As done in [24], the previous filter can be replaced with the following constant gain filter:

$$\overline{x}^{rx}_{t|t-1} = a\overline{x}^{rx}_{t-1|t-1} \tag{20}$$

$$\overline{x}^{rx}_{t|t} = \overline{x}^{rx}_{t|t-1} + \gamma_t k(z_t - \overline{x}^{rx}_{t|t-1}) \tag{21}$$

$$k = \frac{p^{SF}(\epsilon) - p^{tx}_\infty}{(p^{SF}(\epsilon) - p^{tx}_\infty) + \frac{1}{\Lambda}(p^{OL} - p^{tx}_\infty)}, \;\; p^{SF}(\epsilon) > 0 \tag{22}$$

$$p^{SF}(\epsilon) = a^2 p^{SF}(\epsilon) + \sigma_w^2 - (1-\epsilon)\frac{a^2(p^{SF}(\epsilon) - p^{tx}_\infty)^2}{(p^{SF}(\epsilon) - p^{tx}_\infty) + \frac{1}{\Lambda}(p^{OL} - p^{tx}_\infty)} \tag{23}$$

which has the property that asymptotical its error covariance is also an upper bound for the steady state error covariance $p^{rx}_{t|t-1} := \mathbb{E}[\widehat{p}^{rx}_{t|t-1}]$ of the optimal estimator $\widehat{x}^{rx}_{t|t-1}$, i.e.

$$\limsup_{t\to\infty} p^{rx}_{t|t-1} \leq \lim_{t\to\infty}\mathbb{E}[(x_t - \overline{x}^{rx}_{t|t-1})^2] = p^{SF}(\epsilon)$$

It has been shown in [24] that the previous inequality is quite tight, i.e. the performance degradation incurred using a constant gain rather then the optimal time-varying gain, is small.

B. Innovation forwarding strategy ($\nu = 0$)

In this section we consider the innovation forwarding scheme

$$s_t = \widehat{x}^{tx}_{t|t} - \overline{x}^{tx}_{t|t-1}$$

where $\overline{x}^{tx}_{t|t-1} = \mathbb{E}[x_t \,|\, s^q_{t-1}, \ldots, s^q_0]$ and $s^q_t = s_t + n_t$. The MMSE estimator at the receiver $\widehat{x}^{rx}_{t+1|t} = \mathbb{E}_\gamma[x_{t+1} \,|\, \mathcal{Z}_t]$ is, conditionally on $\{\gamma_t\}$ a linear and finite memory functional of the past received data and must have the following form:

$$\begin{aligned} \widehat{x}^{rx}_{t+1|t} &= a\widehat{x}^{rx}_{t|t-1} + ak_t(z_t - \widehat{z}_t) = a\widehat{x}^{rx}_{t|t-1} + ak_t z_t \\ z_t &= \gamma_t\,(s_t + n_t) = \gamma_t \left(\widehat{x}^{tx}_{t|t} - \overline{x}^{tx}_{t|t-1} + n_t\right) \end{aligned} \tag{24}$$

where $\widehat{z}_t := \mathbb{E}_\gamma[z_t|\mathcal{Z}_{t-1}] = 0$ since $\widehat{x}^{tx}_{t|t} - \overline{x}^{tx}_{t|t-1}$ and n_t are uncorrelated and white. The optimal gain k_t is to be selected, at each step, to minimize the conditional receiver state prediction error covariance $\widehat{p}^{rx}_{t+1|t} := \mathbb{E}_\gamma[(x_{t+1} - \widehat{x}^{rx}_{t+1|t})^2]$.

This is easily achieved writing the equation for the prediction error and differentiating w.r.t k_t. Let us first derive the dynamical equation for $\widetilde{x}^{rx}_{t|t-1} = x_t - \widehat{x}^{rx}_{t|t-1}$, which is obtained by subtracting the state prediction update Eqn. (24) from the state equation Eqn. (1), obtaining

$$\widetilde{x}^{rx}_{t+1|t} \;=\; a(1-\gamma_t k_t)\widetilde{x}^{rx}_{t|t-1} - \gamma_t a k_t(\Delta\widehat{x}_t - \widetilde{x}^{tx}_{t|t} + n_t) + w_t$$

where $\Delta\widehat{x}_t := \widehat{x}^{rx}_{t|t-1} - \overline{x}^{tx}_{t|t-1} = \widetilde{\overline{x}}^{tx}_{t|t-1} - \widetilde{x}^{rx}_{t|t-1}$ and $\widetilde{\overline{x}}^{tx}_{t|t-1} := x_t - \overline{x}^{tx}_{t|t-1}$. This implies that $\widetilde{x}^{rx}_{t|t-1} = \widetilde{\overline{x}}^{tx}_{t|t-1} - \Delta\widehat{x}_t$. Using Lemma 2 we obtain:

$$\begin{aligned} \mathbb{E}_\gamma[\widetilde{x}^{rx}_{t|t-1}\widetilde{\overline{x}}^{tx}_{t|t-1}] &= \mathbb{E}_\gamma(\widetilde{\overline{x}}^{tx}_{t|t-1} - \Delta\widehat{x}_t)\widetilde{\overline{x}}^{tx}_{t|t-1}] \\ &= \mathbb{E}_\gamma[\widetilde{\overline{x}}^{tx}_{t|t-1}\widetilde{\overline{x}}^{tx}_{t|t-1}] =: \widehat{p}^0_t \\ \mathbb{E}_\gamma[\widetilde{x}^{rx}_{t|t-1}\Delta\widehat{x}_t] &= \mathbb{E}_\gamma[\widetilde{x}^{rx}_{t|t-1}(\widetilde{\overline{x}}^{tx}_{t|t-1} - \widetilde{x}^{rx}_{t|t-1})] \\ &= -(\widehat{p}^{rx}_{t|t-1} - \widehat{p}^0_t) \\ \mathbb{E}_\gamma[\Delta\widehat{x}_t\Delta\widehat{x}_t] &= \mathbb{E}_\gamma[(\widetilde{\overline{x}}^{tx}_{t|t-1} - \widetilde{x}^{rx}_{t|t-1})\Delta\widehat{x}_t] \\ &= -\mathbb{E}_\gamma[\widetilde{x}^{rx}_{t|t-1}\Delta\widehat{x}_t] = \widehat{p}^{rx}_{t|t-1} - \widehat{p}^0_t \end{aligned}$$

where $\widehat{p}^0_t = \mathbb{E}_\gamma[(x_t - \overline{x}^{tx}_{t|t-1})^2] = \mathbb{E}[(\widetilde{\overline{x}}^{tx}_{t|t-1})^2]$. Recalling that $\mathbb{E}_\gamma[n_t^2] = \frac{1}{\Lambda}\mathbb{E}_\gamma[(\widehat{x}^{tx}_{t|t} - \overline{x}^{tx}_{t|t-1})^2] = \frac{1}{\Lambda}(p^{tx}_{t|t} - \widehat{p}^0_t)$, and $\mathbb{E}_\gamma[\widetilde{x}^{tx}_{t|t}\widetilde{x}^{rx}_{t|t-1}] = p^{tx}_{t|t}$, then it follows that the receiver conditional variance is given by:

$$\begin{aligned} \widehat{p}^{rx}_{t+1|t} = & (a - \gamma_t a k_t)^2 \widehat{p}^{rx}_{t|t-1} + \sigma_w^2 + \\ & + a^2\gamma_t^2 k_t^2 \left(\widehat{p}^{rx}_{t|t-1} - (\widehat{p}^0_t - p^{tx}_{t|t}) + \frac{\widehat{p}^0_t - p^{tx}_{t|t}}{\Lambda}\right) + \\ & + 2a^2\gamma_t k_t(1-\gamma_t k_t)\left(\widehat{p}^{rx}_{t|t-1} - (\widehat{p}^0_t - p^{tx}_{t|t})\right) \end{aligned} \tag{25}$$

The optimal gain k_t which minimizes the right had side is found by taking the derivative w.r.t. k_t

$$\begin{aligned} \frac{\partial \widehat{p}^{rx}_{t+1|t}}{\partial k_t} = & -2\gamma_t a^2(1-\gamma_t k_t)\widehat{p}^{rx}_{t|t-1} + \\ & + 2a^2\gamma_t^2 k_t \left(\widehat{p}^{rx}_{t|t-1} - (\widehat{p}^0_t - p^{tx}_{t|t}) + \frac{(\widehat{p}^0_t - p^{tx}_{t|t})}{\Lambda}\right) + \\ & + 2a^2\gamma_t(1-2k_t)(\widehat{p}^{rx}_{t|t-1} - (\widehat{p}^0_t - p^{tx}_{t|t})) \end{aligned}$$

which, equated to zero has the unique solution

$$k_t = \frac{\Lambda}{\Lambda + 1}. \tag{26}$$

Inserting k_t back into (25) we obtain:

$$\widehat{p}^{rx}_{t+1|t} = a^2\widehat{p}^{rx}_{t|t-1} + \sigma_w^2 - \gamma_t a^2(\widehat{p}^0_t - p^{tx}_{t|t})\frac{\Lambda}{1+\Lambda}$$

Taking now expectation w.r.t the loss sequence γ_t it follows that the expected error covariance $p^{rx}_{t+1|t} = \mathbb{E}\left[(\widetilde{x}^{rx}_{t+1|t})^2\right]$ is given by

$$p^{rx}_{t+1|t} = a^2 p^{rx}_{t|t-1} + \sigma_w^2 - (1-\epsilon)a^2(p^0_t - p^{tx}_{t|t})\frac{\Lambda}{1+\Lambda} \tag{27}$$

where $p^0_t := \mathbb{E}\left[\widehat{p}^0_t\right]$.

It is interesting to observe that the gain k_t in (26) is time invariant and does not depend on the packet loss probability. In fact k_t is also the Kalman optimal gain for $\epsilon = 0$. Finally, recall that p^0_t is the prediction error covariance with no packet loss, which is given by Eqn. (16) by setting $\epsilon = 0$; then

$$\lim_{t\to\infty} p^0_t =: p^0_\infty = \frac{\sigma_w^2 + a^2\frac{\Lambda}{1+\Lambda}p^{tx}_\infty}{1 - \frac{a^2}{1+\Lambda}} = p^{CF}(0)$$

Note also that the limiting value p^{tx}_∞ of $p^{tx}_{t|t}$ is given in equation (7). Thus it follows that the steady state prediction error covariance is given by:

$$\begin{aligned} p^{IF}(\epsilon) &= \lim_{t\to\infty} p^{rx}_{t+1|t} \\ &= \frac{\sigma_w^2}{1-a^2} - \frac{a^2(1-\epsilon)}{1-a^2+\Lambda}\left(\frac{\sigma_w^2}{1-a^2} - p^{tx}_\infty\right) \\ &= (1-\epsilon)p^{CF}(0) + \epsilon\, p^{OL} \end{aligned} \tag{28}$$

which, remarkably, is a simple linear function of the packet loss probability ϵ.

C. Performance comparison

We now want to compare the performance of the two strategies in terms of the steady state prediction error covariance, which are given by Eqn. (23) for the state forwarding and by Eqn. (28) for the innovation forwarding, as a function of the systems parameters $a, \Lambda, \epsilon, \sigma_w^2, \sigma_v^2$. In particular, we are interested in finding the set $\Phi := \{(a,\Lambda,\epsilon) \,|\, p^{SF}(\epsilon) \le p^{IF}\}$, i.e. the set of parameters where the SF strategy has a better performance than the IF strategy.

Theorem 2: Consider the set $\Phi := \{(a,\Lambda,\epsilon) \,|\, p^{SF}(\epsilon) \le p^{IF}(\epsilon)\}$. Then for $\Lambda > 0$, $0 < |a| < 1$, and $\epsilon < 1$ we have:

$$\Phi := \{(a,\Lambda,\epsilon) \,|\, \epsilon > \epsilon_c(a,\Lambda)\}$$

where $0 \le \epsilon_c < 1$ which is the smallest solution of a quadratic equation of the form

$$\epsilon^2 + \beta_1(\sigma_v,\Lambda,a)\epsilon + \beta_2(\sigma_v,\Lambda,a) = 0$$

and is monotonically decreasing in Λ and $|a|$, and

$$\lim_{\Lambda\to+\infty} \epsilon_c(\Lambda,a) = \lim_{|a|\to 1^-} \epsilon_c(\Lambda,a) = 0$$

The critical probability ϵ_c takes the form

$$\epsilon_c(\Lambda,a) = \frac{(1-a^2)(\Lambda+2)}{2a^2\Lambda}\left(\sqrt{1 + \frac{4a^2\Lambda}{(\Lambda+2)^2(1-a^2)}} - 1\right) \tag{29}$$

Proof: See Appendix A. ■

The previous theorem implies that the IF strategy performs better then the SF strategy only for small packet loss probabilities, and more specifically for $\epsilon < \epsilon_c$. Remarkably, the critical probability is independent of the noise process and measurement variances σ_w^2, σ_v^2. Moreover, the critical probability decreases to zero as the system dynamics becomes less stable, i.e. $|a|$ increases, and as the quantization becomes finer, i.e. Λ increases. In particular, the previous theorem shows that it is always better to use the SF strategy, independently of the systems parameters, if the packet loss probability is greater than one half, i.e. under a high packet loss probability regime.

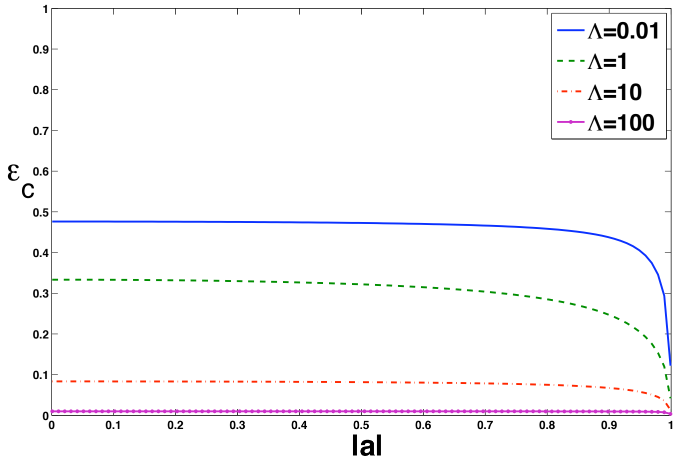

Fig. 4. Critical probability ϵ_c as a function of $|a|$ for different values of the SQNR Λ.

Figure 4 pictures the critical probability ϵ_c as a function of $|a|$ for different values of the SQNR Λ, which shows that such probability is almost equal to $\frac{\Lambda}{\Lambda+1}$ up to $|a| \approx 0.8$ and then rapidly decays to zero.

V. Soft Innovation forwarding with no channel feedback

In this section, we propose an alternative strategy under the no channel feedback scenario, that includes the IF strategy and the SF strategy as special cases. More precisely, we propose a hybrid strategy, where the transmitter sends a convex combination of its best estimate of the state $\widehat{x}^{tx}_{t|t} = \mathbb{E}[x_t \,|\, \mathcal{T}^{NCF}_t]$ and the innovation between its best estimate and the best estimate of the state given the past quantized transmitted signals, i.e. $\Delta\widehat{x}_t = \widehat{x}^{tx}_{t|t} - \overline{x}^{tx}_{t|t-1}$ where $\overline{x}^{tx}_{t|t-1} = \mathbb{E}[x_t \,|\, s^q_{t-1}, \ldots, s^q_0]$. We call this scheme the *soft innovation forwarding (SIF)* scheme. In this case, the transmitted signal is thus given by

$$s_t = \nu\widehat{x}^{tx}_{t|t} + (1-\nu)\Delta\widehat{x}_t = \widehat{x}^{tx}_{t|t} - (1-\nu)\overline{x}^{tx}_{t|t-1} \qquad (30)$$

where $0 \leq \nu \leq 1$ is fixed at the transmitter. This scheme is graphically illustrated in Fig. 3.

A. Transmitter filter design: $g(\mathcal{T}^{NCF}_t)$

In this section, we explicitly compute the transmitter filter function $g(\mathcal{T}^{NCF}_t)$ based on the SIF strategy. Basically, it reduces to the problem of computing the equation for the internal estimator $\overline{x}^{tx}_{t|t-1}$. Since the dynamical systems is linear with additive gaussian noise, then the optimal MMSE estimator is linear in the quantized transmitted signals s^q_t and it is given by the Kalman Filter. However, the equations are somewhat nonstandard since the variance of the quantization noise n_t is not constant but depends on the variance of the transmitted signal. We start by defining the internal estimator error covariance as $\overline{p}_{t|h} = \mathbb{E}[(\widetilde{\overline{x}}^{tx}_{t|h})^2]$, where $\widetilde{\overline{x}}^{tx}_{t|h} = x_t - \overline{x}^{tx}_{t|h}$. Based on this definition, we can compute the power of the transmitted signal s_t as follows:

$$\begin{aligned}
\mathbb{E}[s_t^2] &= \mathbb{E}[(\nu\widehat{x}^{tx}_{t|t} + (1-\nu)\Delta\widehat{x}_t)^2] = \big(\nu^2\mathbb{E}[(\widehat{x}^{tx}_{t|t})^2] + \\
&\quad +(1-\nu)^2\mathbb{E}[\Delta\widehat{x}_t^2] + 2\nu(1-\nu)\mathbb{E}[\widehat{x}^{tx}_{t|t}\Delta\widehat{x}_t]\big) \\
&= \nu^2\left(p^{OL} - p^{tx}_{t|t}\right) + (1-\nu)^2(\overline{p}_{t|t-1} - p^{tx}_{t|t}) + \\
&\quad +2\nu(1-\nu)\mathbb{E}[(\overline{x}^{tx}_{t|t-1} + \Delta\widehat{x}_t)\Delta\widehat{x}_t] \\
&= \nu^2\left(p^{OL} - p^{tx}_{t|t}\right) + (1-\nu)^2(\overline{p}_{t|t-1} - p^{tx}_{t|t}) + \\
&\quad +2\nu(1-\nu)(\overline{p}_{t|t-1} - p^{tx}_{t|t}) \\
&= \nu^2(p^{OL} - \overline{p}_{t|t-1}) + (\overline{p}_{t|t-1} - p^{tx}_{t|t})
\end{aligned}$$

Here we used the fact that $x_t - \overline{x}^{tx}_{t|t-1} = (x_t - \widehat{x}^{tx}_{t|t}) + \Delta\widehat{x}_t$ where $(x_t - \widehat{x}^{tx}_{t|t})$ and $\Delta\widehat{x}_t$ are uncorrelated and that x_t is assumed to be in its steady state distribution. The equations of the filter are given by:

$$\begin{aligned}
\overline{x}^{tx}_{t+1|t} &= a\overline{x}^{tx}_{t|t-1} + k^{tx}_t(s^q_t - \widehat{s}^q_{t|t-1}) \\
\widehat{s}^q_{t|t-1} &= \mathbb{E}[s^q_t \,|\, s^q_{t-1}, \ldots, s^q_0] = \nu\overline{x}^{tx}_{t|t-1} \\
k^{tx}_t &= cov(x_t, s^q_t - \widehat{s}^q_{t|t-1})Var^{-1}\{s^q_t - \widehat{s}^q_{t|t-1}\} \\
&= \frac{a(\overline{p}_{t|t-1} - p^{tx}_{t|t})}{\overline{p}_{t|t-1} - p^{tx}_{t|t} + \mathbb{E}[n_t^2]}
\end{aligned}$$

where

$$\mathbb{E}[n_t^2] = \frac{1}{\Lambda}\mathbb{E}[s_t^2] = \frac{1}{\Lambda}[\nu^2(p^{OL} - \overline{p}_{t|t-1}) + (\overline{p}_{t|t-1} - p^{tx}_{t|t})]$$

For large t such filter will reach a steady state and, therefore, it is possible to consider its steady state implementation which will reach the same steady state performance. The steady state filter is given by:

$$\overline{x}^{tx}_{t+1|t} = (a - \nu\overline{k})\overline{x}^{tx}_{t|t-1} + \overline{k}s^q_t$$

$$\overline{k} = \frac{a(\overline{p} - p^{tx}_{\infty})}{(1+\frac{1}{\Lambda})(\overline{p} - p^{tx}_{\infty}) + \frac{\nu^2}{\Lambda}(p^{OL} - \overline{p})}, \quad \overline{p} > 0 \qquad (31)$$

$$\overline{p} = a^2\overline{p} + \sigma_w^2 - \frac{a^2(\overline{p} - p^{tx}_{\infty})^2}{(1+\frac{1}{\Lambda})(\overline{p} - p^{tx}_{\infty}) + \frac{\nu^2}{\Lambda}(p^{OL} - \overline{p})} \qquad (32)$$

where the last equation is a Riccati-like equation which has a unique stabilizing positive solution $\overline{p}$.

B. Receiver filter design: $h(\mathcal{R}_t)$

In this section we explicitly compute the optimal state estimator at the receiver, i.e. $\widehat{x}^{rx}_{t+1|t} = \mathbb{E}[x_{t+1} \,|\, \mathcal{R}_t]$. We assume that the transmitter filter architecture, and in particular the value of ν, is known at the receiver, therefore it is possible to write the received message $z_t := \gamma_t s^q_t = \gamma_t(\widehat{x}^{tx}_{t|t} - (1-\nu)\overline{x}^{tx}_{t|t-1} + n_t)$ as the output of the following dynamical system:

$$\underbrace{\begin{bmatrix} x_{t+1} \\ \widehat{x}^{tx}_{t+1|t+1} \\ \overline{x}^{tx}_{t+1|t} \end{bmatrix}}_{\xi_{t+1}} = \underbrace{\begin{bmatrix} a & 0 & 0 \\ a\widehat{k} & a(1-\widehat{k}) & 0 \\ 0 & \overline{k} & a-\overline{k} \end{bmatrix}}_{A}\underbrace{\begin{bmatrix} x_t \\ \widehat{x}^{tx}_{t|t} \\ \overline{x}^{tx}_{t|t-1} \end{bmatrix}}_{\xi_t} + \underbrace{\begin{bmatrix} w_t \\ \widehat{k}(w_t + v_{t+1}) \\ \overline{k}n_t \end{bmatrix}}_{\eta_t} \qquad (33)$$

$$z_t = \gamma_t\underbrace{\begin{bmatrix} 0 & 1 & -(1-\nu) \end{bmatrix}}_{C}\begin{bmatrix} x_t \\ \widehat{x}^{tx}_{t|t} \\ \overline{x}^{tx}_{t|t-1} \end{bmatrix} + \gamma_t n_t \qquad (34)$$

where $\widehat{k}$ is the steady state Kalman filtering gain for the transmitter state estimator $\widehat{x}^{tx}_{t|t}$ defined in Eqn. (8).

As a consequence the estimator $\widehat{x}^{rx}_{t+1|t} = \mathbb{E}[x_{t+1} \mid \mathcal{R}_t]$ corresponds of the first component of the optimal estimator $\widehat{\xi}_{t+1|t} = \mathbb{E}[\xi_{t+1} \mid \mathcal{R}_t]$ which turns out to be the optimal Kalman filter with intermittent observations studied in [12]. Such a filter is time-varying since the Kalman gain depends on the packet loss sequence, however, as discussed in Section IV-A, it can be replaced with a constant gain filter with limited performance degradation [24]. The (suboptimal) receiver filter design is then given by:

$$\overline{\xi}_{t+1|t} = (A - \gamma_t KC)\overline{\xi}_{t|t-1} + \gamma_t K z_t \quad (35)$$

$$\overline{x}^{rx}_{t|t-1} = h(\mathcal{R}_{t-1}) = \underbrace{\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}}_{H} \overline{\xi}_{t|t-1} \quad (36)$$

$$K = (APC^T + S)(CPC^T + R)^{-1} \quad (37)$$

$$P = APA^T + Q - (1-\epsilon)K(CPC^T + R)K^T = \Psi(P) \quad (38)$$

$$R = \lim_{t\to\infty} \mathbb{E}[n_t^2] = \frac{1}{\Lambda}[\nu^2(p^{OL} - \overline{p}) + (\overline{p} - p^{tx}_{\infty})]$$

$$Q = \lim_{t\to\infty} \mathbb{E}[\eta_t \eta_t^T] = \begin{bmatrix} \sigma_w^2 & \widehat{k}\sigma_w^2 & 0 \\ \widehat{k}\sigma_w^2 & \widehat{k}^2(\sigma_w^2 + \sigma_v^2) & 0 \\ 0 & 0 & \overline{k}^2 R \end{bmatrix}$$

$$S = \lim_{t\to\infty} \mathbb{E}[\eta_t n_t] = \begin{bmatrix} 0 \\ 0 \\ \overline{k}R \end{bmatrix}$$

The steady state Kalman gain K can therefore be obtained by finding the unique positive definite solution $P > 0$ that solves the modified algebraic Riccati equation (38) and the steady state prediction error has the following upper bound:

$$\limsup_{t\to\infty} \mathbb{E}[(x_t - \widehat{x}^{rx}_{t|t-1})^2] \le p^{SIF} = HPH^T \quad (39)$$

C. Optimal soft innovation forward strategy

The transmitter and receiver filter design proposed in the previous two sections still leave a certain degree of freedom for optimizing the performance $p^{SIF} = p_{11}(\epsilon) = p^{SIF}(\nu, \epsilon)$, where $p_{11}(\epsilon)$ is the $(1,1)$-th element of the receiver estimation error covariance matrix P, and where we explicitly indicate its dependence on the parameters ν, ϵ. If the packet loss probability ϵ is known, then one might optimize for the mixing coefficient ν.

More specifically we define:

$$\nu^*(\epsilon) := \arg\min_{\nu\in[0,1]} p^{SIF}(\nu, \epsilon) \quad (40)$$

$$p^{OSIF}(\epsilon) := p^{SIF}(\nu^*, \epsilon) \quad (41)$$

where $p^{OSIF}(\epsilon)$ is the optimal soft innovation forward (OSIF) strategy for a given packet loss probability ϵ. It is seen via numerical computations that $p^{SIF}(\nu, \epsilon)$ has a unique minimum in the interval $\nu \in (0,1)$, It is also seen that this optimal value of ν, $\nu^*(\epsilon)$ computed by an exhaustive search, appears to be a monotonically increasing function of ϵ, which implies that as the packet loss probability increases, it is better to place more weight on the state and less on the innovation. Moreover, it is seen that the SF strategy is the optimal strategy when the packet loss probability is very close to 1. Analytically proving these results appears to be difficult in the general noisy measurement case. However, we are able to prove some meaningful results in the noise-free case when $\sigma_v^2 = 0$, that is the sensor has access to full-state observation. In this case, the system description presented in (34) reduces to a 2nd-order system (since in this case $\widehat{x}^{tx}_{t|t} = x_t$). The corresponding descriptions for all the relevant parameters can be found in [19], or also by substituting $\sigma_v^2 = 0, p^{tx}_{\infty} = 0$ in the appropriate equations. With a slight abuse of notation, we use the same notations for this special case to maintain readability. In this special case, we can prove the following two theorems in this noise-free situation at the sensor. The first of these theorems states that for a fixed ϵ there is $\nu \in (0,1)$ that performs better than the SF strategy ($\nu = 1$) and the IF strategy ($\nu = 0$).

Theorem 3: Under the assumption $\sigma_v^2 = 0$, for any arbitrary $\epsilon \in (0,1)$, then $p^{SIF}(\nu, \epsilon)$ is a decreasing function of ν at $\nu = 0$ and an increasing function of ν at $\nu = 1$. This implies that $p^{SIF}(\nu, \epsilon)$ has at least one minimum at some $0 < \nu^* < 1$.

Proof: See Appendix B. ■

Remark 2: It is possible to check numerically via suitable examples that $p^{SIF}(\nu, \epsilon)$ may not be a convex function of ν for a fixed ϵ. Therefore we do not, at this stage, attempt to prove that $p^{SIF}(\nu, \epsilon)$ has a unique minimum with respect to $\nu \in (0,1)$. Instead, the above theorem simply states that there is at least one minimum for $p^{SIF}(\nu, \epsilon)$ at some $0 < \nu^* < 1$. This is not to say that the minimum is not unique (in fact the extensive numerical results indeed suggest uniqueness), but a proof of uniqueness has proved to be elusive so far.
The second theorem states that as the packet loss probability approaches one, then the optimal ν^* approaches one as well, i.e. the SF strategy becomes optimal for large packet loss probabilities, as stated in the following theorem:

Theorem 4: Under the assumption $\sigma_v^2 = 0$, the optimal mixing parameter $\nu^*(\epsilon)$ has the following properties:

$$\nu^*(0) = 0, \quad \lim_{\epsilon\to 1^-} \nu^*(\epsilon) = 1$$

Proof: See Appendix C. ■

VI. Numerical Results

We first illustrate the accuracy of our additive white noise model for the quantization noise. We use a uniform quantizer to quantize s_t given by (30) with a suitable number of quantization levels and saturation thresholds so as to guarantee a SNR equal to Λ. The quantization step Δ_Q is chosen so that the equivalent additive noise variance is $\sigma_n^2 = \frac{\Delta_Q^2}{12}$, where $\sigma_n^2 = \frac{Var\{s_t\}}{\Lambda}$. These latter two expressions combined yield $\Delta_Q = \sqrt{\frac{12Var\{s_t\}}{\Lambda}}$. By setting the saturation thresholds $\pm T_Q$ according to $T_Q = 4\sqrt{Var\{s_t\}}$, the number of quantization levels is given by $N = \left\lceil \frac{2T_Q}{\Delta_Q} \right\rceil = \left\lceil \frac{8\sqrt{Var\{s_t\}}}{\sqrt{\frac{12Var\{s_t\}}{\Lambda}}} \right\rceil = \left\lceil 4\sqrt{\frac{\Lambda}{3}} \right\rceil$, which corresponds to $N_b = \left\lceil log_2\left(\left\lceil 4\sqrt{\frac{\Lambda}{3}} \right\rceil\right) \right\rceil$ bits/sample. We consider now $N_b = 3$ which corresponds to $\Lambda = 12$. We use $a = 0.95$ and the set the packet loss probability equal to $\epsilon = 0.3$. The sample estimation error variance (at the receiver using the soft innovation strategy) and the theoretical variance using the additive white Gaussian noise (AWGN) model are depicted in Figure 5. It can be seen easily that the AWGN

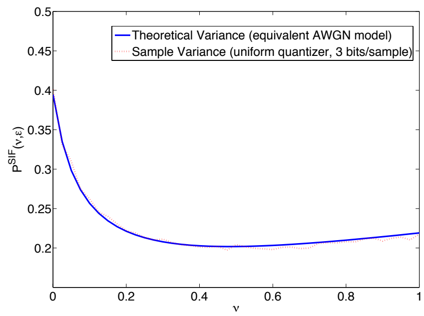

Fig. 5. Sample estimation error variance with uniform quantizer vs. theoretical error variance with AWGN model (3 bits/sample)

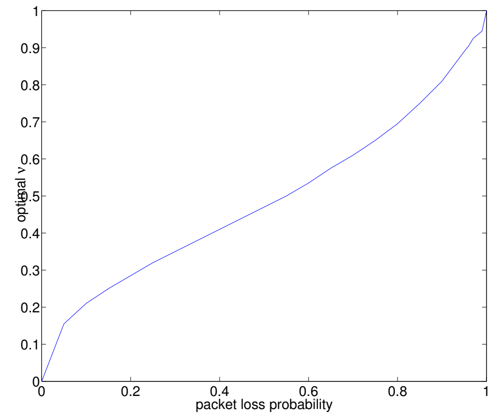

Fig. 7. Optimal mixing coefficient ν^* as a function of the packet loss probability ϵ for the OSIF strategy for $a = 0.95, \Lambda = 3, \sigma_w^2 = 0.1, \sigma_v^2 = 0.05$.

model provides a highly accurate approximation, in fact a very good one for $N_b \geq 2$.

For the rest of the numerical results we use the parameter values $a = 0.95$, $\Lambda = 3$ (2 bits per sample), $\sigma_w^2 = 0.1, \sigma_v^2 = 0.05$. Figure 6 depicts the estimation error performance (normalized by the maximum value p^{OL} at $\epsilon = 1$) of the filters derived so far and the critical probability ϵ_c defined in Eqn. (29). As expected, the performance degrades as the packet loss probability increases for all estimators, but the estimator with channel feedback outperforms all estimators with no channel feedback. The figure also shows that by optimizing ν, the OSIF performs considerably better than the SF and IF strategies, which are just two special cases in the class of the SIF strategies.

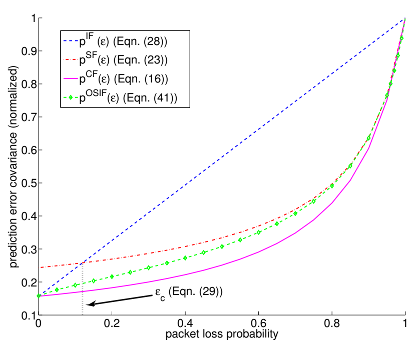

Fig. 6. Prediction error covariance of proposed strategies against packet loss probability ϵ for $a = 0.95, \Lambda = 3, \sigma_w^2 = 0.1$, $\sigma_v^2 = 0.05$.

In Figure 7 below, we plot the optimal mixing coefficient $\nu = \nu^*$ which has been obtained numerically via an exhaustive search. The curve appears to be monotonically increasing from zero to unity, thus confirming that as the packet loss increases, the optimal soft innovation forwarding strategy transits from the IF to the SF strategy.

VII. Discussions and Future Work

In this section we briefly indicate the limitations of the current work and how these results can be generalized in various directions.

Unstable Systems: Suppose one considers an unstable system. Then it is not possible to consider an *uncontrolled* unstable system at the transmitter since regardless of the coder/decoder scheme employed, the power of the signal to be transmitted will grow unbounded. If we were to consider a *controlled* unstable system, where the control action is determined by the receiver, then the very same model of the system will be different from the one used in this manuscript and it is not obvious how the results obtained here can be extended. Recently some of the authors of this paper have looked at the case of controlled unstable systems when no pre-processing is done at the sensor [10], [11]. These two works are complementary and future work will focus on combining these ideas with the current work.

Higher-order and MIMO systems: As for the case of multi-variable systems, the problem is even harder. Suppose in fact that in the scenario with channel feedback we still want to use the same idea of sending the innovation. Even if the dynamical system has a vector state but a scalar output, as soon as two consecutive packets are lost followed by a successful transmission, the innovation that the transmitter has to send is two-dimensional, i.e. two real valued numbers are required to be transmitted across the same scalar channel for the receiver to recover the current estimate when the packet is received. This gives rise to the problem of properly modeling the quantization error when the same number of bits per second are to be sent across the channel, yet two real numbers are to be encoded. Another alternative is to use lattice vector quantization with the same additive white noise quantization model, as used in [22]. This will require the use of a vector channel, and perhaps the use of a vector parameter ν for the soft innovation forwarding with no channel feedback case. Needless to say, the corresponding analysis for the no feedback case will be considerably more difficult if not intractable.

Imperfect feedback channels: In this paper we study the cases of perfect packet acknowledgement feedback or no feedback. A more practical scenario in between these two extreme cases is where the transmitter receives packet acknowledgement but over an imperfect channel, such that the ACK/NACK packets can be also lost with a certain probability. Note that this particular issue has been investigated in a slightly different problem setting in [25]. In this paper, the problem of whether to send a state estimate or the innovation is formulated as a Markov decision problem (MDP) where a long term average estimation error (at the receiver) is minimized. In the case of imperfect ACK/NACK, the problem becomes a partially observed Markov decision problem (POMDP) problem which can be solved using information state techniques that convert the problem to a fully observed MDP problem. This is computationally expensive but suboptimal solutions based on an estimate of the receiver estimation error covariance at the transmitter can be designed in the case of imperfect channel feedback.

Delays: In this work we considered a scenario with a channel delay smaller or equal to the time step. If the delay is larger than unity, the strategies suggested in this work with no channel feedback are still valid since the only difference is that the estimator has to provide the open loop d-time step ahead prediction to reconstruct $\widehat{x}^{rx}_{t|t-d} := \mathbb{E}[x_t|\mathcal{R}_s] = \mathbb{E}_\gamma[x_t|\mathcal{Z}_{t-d}] = a^d \widehat{x}^{rx}_{t-d|t-d}$. However, the results presented in the channel feedback scenario cannot be directly extended since the transmitter requires to know the packet loss sequence with a delay smaller or equal to unity in order to make a perfect copy of the receiver estimator. Therefore alternative strategies are not obvious for $d > 1$.

VIII. Conclusions

In this work we studied the problem of remotely estimating the state of a dynamical stable system based on noisy measurements over a communication channel subject to packet loss and quantization. We showed that with perfect channel feedback it is possible to derive the optimal linear transmitter and receiver filters to minimize the estimation error variance using a strategy that it is reminiscent of DPCM. We also studied the scenario with no channel feedback and we proposed a few heuristic strategies for which we were able to characterize performance and trade-offs.

References

[1] 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.

[2] P. Minero, L. Coviello, and M. Franceschetti, "Stabilization over markov feedback channels: The general case," *IEEE Trans. Autom. Control*, vol. 58, no. 2, pp. 349–362, 2013.

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

[4] E. Silva and S. Pulgar, "Control of LTI plants over erasure channels," *Automatica*, vol. 47, no. 8, pp. 1729–1736, 2011.

[5] 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, pp. 1391–1403, 2007.

[6] E. Silva, G. Goodwin, and D. Quevedo, "Control system design subject to SNR constraints," *Automatica*, vol. 46, no. 2, pp. 428–436, 2010.

[7] 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.

[8] 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.

[9] 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.

[10] A. Chiuso, N. Laurenti, L. Schenato, and A. Zanella, "LQG cheap control subject to packet loss and SNR limitations," in *Proceedings of European Control Conference*, July 2013.

[11] ——, "LQG cheap control over SNR-limited lossy channels with delay," in *Proceedings of the IEEE Conference on Decision and Control*, December 2013, [Online] Available at http://automatica.dei.unipd.it/tl_files/utenti/lucaschenato/Papers/Conference/ChiusoLSZcdc13TR.pdf.

[12] 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.

[13] L. Shi and H. Zhang, "Scheduling two gauss-markov systems: An optimal solution for remote state estimation under bandwidth constraint," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 2038–2042, 2012.

[14] L. Xie, "State estimation over unreliable network," in *Proceedings of Asian Control Conference*, 2009, pp. 453 – 458.

[15] M. Trivellato and N. Benvenuto, "State control in networked control systems under packet drops and limited transmission bandwidth," *IEEE Transaction on Communications*, vol. 58, no. 2, pp. 611–622, 2010.

[16] K. You and L. Xie, "Minimum data rate for mean square stabilizability of linear systems with markovian packet losses," *IEEE Trans. on Automat. Contr.*, vol. 56, no. 4, pp. 772–785, 2011.

[17] K. Okano and H. Ishii, "Stabilization of uncertain systems with finite data rates and markovian packet losses," in *Proceedings of European Control Conference (ECC)*, 2013, pp. 2368–2373.

[18] S. Chitode, *Digital communications*. Technical Publications, 2009.

[19] S. Dey, A. Chiuso, and L. Schenato, "Remote estimation subject to packet loss and quantization noise," in *Proceedings of the IEEE Conference on Decision and Control*, December 2013.

[20] D. Marco and D. Neuhoff, "The validity of the additive noise model for uniform scalar quantizers," *IEEE Trans. Info. Theory*, vol. 51, no. 5, pp. 1739–1755, 2005.

[21] V. Goyal, "High-rate transform coding: how high is high, and does it matter?" in *Proc. IEEE Intl. Symp. Info. Theory (ISIT)*, 2000.

[22] A. Leong, S. Dey, and G. Nair, "Quantized filtering schemes for multi-sensor linear state estimation: Stability and performance under high rate quantization," *IEEE Trans. Sig. Proc.*, vol. 61, no. 15, pp. 3852–3865, 2013.

[23] D. Cook and B. Li, "Dimension reduction for conditional mean in regression," *The Annals of Statistics*, vol. 30, no. 2, pp. 455–474, 2002.

[24] 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.

[25] M. Nourian, A. Leong, S. Dey, and D. Quevedo, *IEEE Transactions on Control of Network Systems*, 2013, under review.

Appendix

A. Proof of Theorem 2

Without loss of generality we can set $\sigma_w^2 = 1$, since it simply scales the error covariance and therefore it does not affect the set Φ. Let us define $\Delta p(a, \Lambda, \epsilon) = p^{SF}(\epsilon) - p^{IF}(\epsilon)$, where we also make explicit the dependence of the performance in terms of the parameters. It is straightforward to observe that $\Delta p(a, \Lambda, 1) = 0$ and $\Delta p(a, \Lambda, 0) > 0$. Therefore, if we can show that there exists a unique $\epsilon_c \in (0, 1)$ such that $\Delta p(a, \Lambda, \epsilon_c) = 0$, then this implies that $p^{SF}(\epsilon) \leq p^{IF}(\epsilon)$ for $\epsilon \geq \epsilon_c$. We now show that this is the case. If $\Delta p(a, \Lambda, \epsilon_c) = 0$, then $p^{SF}(\epsilon)(\epsilon_c) = p^{IF}(\epsilon_c) = p^*$. The points p that satisfy this equality must also satisfy Eqn. (23) and Eqn. (28), therefore, if we take the difference and recalling that $a \neq 0$ and $\epsilon \neq 1$ we have:

$$\frac{(p^* - p^{tx}_\infty)^2}{(p^* - p^{tx}_\infty) + \frac{1}{\Lambda}(p^{OL} - p^{tx}_\infty)} = \frac{\Lambda(1-a^2)}{1-a^2+\Lambda}(p^{OL} - p^{tx}_\infty)$$

From Eqn. (28) it follows that

$$p^* - p^{tx}_\infty = \left(1 - \frac{a^2\Lambda(1-\epsilon)}{1-a^2+\Lambda}\right)(p^{OL} - p^{tx}_\infty)$$

If we substitute this equation into the previous expression and after some manipulations, which are valid for $a \neq 0$ and $\Lambda \neq 0$, we get:

$$a^2\Lambda\epsilon^2 + (1-a^2)(\Lambda+2)\epsilon - (1-a^2) = 0$$

from which it follows that the only positive feasible solution for ϵ_c is given by Eqn. (29).

We can now study the dependence of ϵ_c in terms of the parameters Λ and a. By rearranging the different terms, we have

$$(a^2\epsilon + 1 - a^2)\epsilon + \frac{1}{\Lambda}(1-a^2)(2\epsilon - 1) = 0$$

from which it follows via root-locus analysis that for fixed a, $\epsilon_c(\Lambda, a)$ is a monotonically decreasing function of Λ where

$$\lim_{\Lambda\to 0^+}\epsilon_c(\Lambda,a)=\frac{1}{2},\quad \lim_{\Lambda\to+\infty}\epsilon_c(\Lambda,a)=0$$

Similarly, by defining $\eta=\frac{a^2}{1-a^2}$ which is a strictly monotonically increasing function of a^2 where $\eta\in(0,+\infty)$, and by rearranging terms we get:

$$\Lambda\epsilon^2+\frac{1}{\eta}\left((\Lambda+2)\epsilon-1\right)=0$$

from which it follows via root-locus analysis that, for fixed Λ, $\epsilon_c(\Lambda, a)$ is a monotonically decreasing function of a^2 where

$$\lim_{|a|\to 0^+}\epsilon_c(\Lambda,a)=\frac{1}{2+\Lambda},\quad \lim_{|a|\to 1^-}\epsilon_c(\Lambda,a)=0$$

From this analysis, it follows that

$$\epsilon_c(\Lambda,a)<\frac{1}{2},\qquad \forall |a|\in(0,1),\Lambda\in(0,+\infty)$$

which concludes the proof.

B. Proof of Theorem 3

In the noise-free case, where the state is fully observed at the sensor, we have a two-dimensional state vector as discussed before. In this case, the expression for the covariance matrix P can be computed from three paired nonlinear equations as shown below (42). Let us denote $P=\begin{bmatrix} p_{11}(\epsilon) & p_{12}(\epsilon)\\ p_{12}(\epsilon) & p_{22}(\epsilon)\end{bmatrix}$, where we have explicitly indicated that P is symmetric and its elements depend on ϵ. Although we will be primarily interested in the behaviour of $p_{11}(\epsilon)=p^{SIF}$ with respect to ν, the properties of $p_{12}(\epsilon), p_{22}(\epsilon)$ will also be useful. In the case when $\epsilon=0$ (i.e, there is no packet loss), it is easy to check that $p_{11}(0)$ satisfies the same equation as the steady-state transmitter Kalman predictor error covariance given by $\overline{p}$, and is clearly minimum when $\nu=0$. Also, $p_{12}(0)=p_{22}(0)=0$.

It can be shown after some algebraic manipulation that the elements of P satisfy the following equations:

$$\begin{aligned}
p_{11}(\epsilon) &= \frac{\sigma_w^2}{1-a^2}-\frac{a^2}{1-a^2}\frac{(1-\epsilon)}{M_\infty(\nu)}\left(p_{11}(\epsilon)-(1-\nu)p_{12}(\epsilon)\right)^2\\
p_{12}(\epsilon) &= \frac{a\overline{k}}{1-a^2+a\overline{k}}p_{11}(\epsilon)-\\
&\quad \frac{a(1-\epsilon)}{1-a^2+a\overline{k}}\left(p_{11}(\epsilon)-(1-\nu)p_{12}(\epsilon)\right)\frac{L_\infty(\nu)}{M_\infty(\nu)}\\
p_{22}(\epsilon) &= \frac{\overline{k}^2}{1-(a-\overline{k})^2}p_{11}(\epsilon)+\frac{2\overline{k}(a-\overline{k})}{1-(a-\overline{k})^2}p_{12}(\epsilon)\\
&\quad +\frac{\overline{k}^2}{1-(a-\overline{k})^2}R-\frac{(1-\epsilon)}{1-(a-\overline{k})^2}\frac{L^2_\infty(\nu)}{M_\infty(\nu)} \qquad (42)
\end{aligned}$$

where

$$\begin{aligned}
M_\infty(\nu) &= p_{11}(\epsilon)-2p_{12}(\epsilon)(1-\nu)+p_{22}(\epsilon)(1-\nu)^2+R,\\
L_\infty(\nu) &= \overline{k}p_{11}(\epsilon)+\left(a-\overline{k}(2-\nu)\right)p_{12}(\epsilon)-\\
&\quad -(a-\overline{k})(1-\nu)p_{22}(\epsilon)+\overline{k}R
\end{aligned}$$

Recall that $p_{11}(\epsilon)=p^{SIF}(\nu,\epsilon)$. Hence we will use $p_{11}(\epsilon)$ to indicate $p^{SIF}(\nu,\epsilon)$ in the following proof. The proof is divided into two parts: (i) showing that $\frac{\partial p_{11}(\epsilon)}{\partial\nu}_{|\nu=1}>0$ and (ii) $\frac{\partial p_{11}(\epsilon)}{\partial\nu}_{|\nu=0}<0$.

(i) For simplicity, we will drop the dependence on the argument ϵ in this part, and make the observation that all values of p_{11}, p_{12} etc. are evaluated at $\nu=1$ in the expression of the partial derivative $\frac{\partial p_{11}(\epsilon)}{\partial\nu}_{|\nu=1}$. Recall also that $p^{OL}=\frac{\sigma_w^2}{1-a^2}$.

Using the equation for p_{11} from (42), and taking partial derivatives, we have (after some algebra):

$$\begin{aligned}
&\frac{\partial p_{11}(\epsilon)}{\partial\nu}_{|\nu=1}\left[1+\frac{a^2(1-\epsilon)p_{11}}{(1-a^2)(p_{11}+\frac{p^{OL}}{\Lambda})}\left(2-\frac{p_{11}}{(p_{11}+\frac{p^{OL}}{\Lambda})}\right)\right]\\
&\quad=\frac{2a^2}{1-a^2}(1-\epsilon)\frac{p_{11}p^{OL}}{\Lambda(p_{11}+\frac{p^{OL}}{\Lambda})^2}\left[p_{11}(1-\frac{\overline{p}}{p^{OL}})-p_{12}\right]
\end{aligned}$$

It is easy to check that the expression in the square brackets on the left had side of the previous equation is positive. All that remains to show therefore is that $p_{12}<p_{11}(1-\frac{\overline{p}}{p^{OL}})$. To this end, note that from the equation for p_{12} from (42), it follows that (recall that all values are evaluated at $\nu=1$), $p_{12}<\frac{a\overline{k}}{1-a^2+a\overline{k}}p_{11}$. It can be shown that for all $0\le\nu<1$ (see Proof of Theorem 4 below)

$$\frac{a\overline{k}}{1-a(a-\overline{k})}=\frac{a^2\bar{P}_\infty}{\bar{P}_\infty+J(\nu)(1-a^2)}=(1-\bar{P}_\infty),$$

where $\bar{P}_\infty=\frac{\overline{p}}{p^{OL}}$. Hence it follows that $p_{12}<p_{11}(1-\bar{P}_\infty)=p_{11}(1-\frac{\overline{p}}{p^{OL}})$. This implies that $\frac{\partial p_{11}(\epsilon)}{\partial\nu}_{|\nu=1}>0$.

(ii) The proof of this part relies on using a state transformation technique. Denote a new state vector $\begin{bmatrix} x^a_{t+1}\\ x^b_{t+1}\end{bmatrix}=T\begin{bmatrix} x_{t+1}\\ \widehat{x}^{tx}_{t+1|t}\end{bmatrix}$, where $T=\begin{bmatrix} 1 & 0\\ 1 & -(1-\nu)\end{bmatrix}$. It can be easily checked that $x^a_t=x_t$ and $x^b_t=s_t$. Using this transformation, we can write a new state space system as

$$\begin{bmatrix} x_{t+1}\\ s_{t+1}\end{bmatrix}=\bar{A}_1\begin{bmatrix} x_t\\ s_t\end{bmatrix}+T\begin{bmatrix} w_t\\ \overline{k}n_t\end{bmatrix},\ z_t=\gamma_t\bar{C}_1\begin{bmatrix} x_t\\ s_t\end{bmatrix}+\gamma_t n_t \qquad (43)$$

where

$$\bar{A}_1=TAT^{-1}=\begin{bmatrix} a & 0\\ \nu\overline{k} & (a-\overline{k})\end{bmatrix}\qquad \bar{C}_1=\bar{C}T^{-1}=[0\ 1]\,.$$

It is straightforward to show that for this transformed state space system, one can derive a similar suboptimal constant gain Kalman filter which has a steady state stabilizing solution $\tilde{P}(\epsilon)$ whose elements $p_{ij}(\epsilon)$, $i=1,2$, $j=1,2$ satisfy the following equations:

$$\begin{aligned}
\tilde{p}_{11}(\epsilon) &= \frac{\sigma_w^2}{1-a^2}-\frac{a^2(1-\epsilon)}{1-a^2}\frac{\tilde{p}^2_{12}(\epsilon)}{\tilde{p}_{22}(\epsilon)+R}\\
\tilde{p}_{12}(\epsilon) &= \frac{1}{1-a(a-\overline{k})}\left[a\nu\overline{k}\tilde{p}_{11}(\epsilon)+\sigma_w^2-\frac{(1-\epsilon)a\tilde{p}_{12}(\epsilon)}{\tilde{p}_{22}(\epsilon)+R}\bar{V}(\nu)\right]\\
\tilde{p}_{22}(\epsilon) &= \frac{1}{1-(a-\overline{k})^2}\Big[\nu^2\overline{k}^2\tilde{p}_{11}(\epsilon)+2\nu\overline{k}(a-\overline{k})\tilde{p}_{12}(\epsilon)+\sigma_w^2\\
&\quad +(1-\nu)^2\overline{k}^2R-(1-\epsilon)\frac{F^2_\infty(\nu)}{\tilde{p}_{22}(\epsilon)+R}\Big] \qquad (44)
\end{aligned}$$

where $\bar{V}(\nu):=\nu\overline{k}\tilde{p}_{12}(\epsilon)+(a-\overline{k})\tilde{p}_{22}(\epsilon)-(1-\nu)\overline{k}R$ for notational simplicity.

First, it is useful to observe a few facts regarding the steady state stabilizing solution $\tilde{P}$ and its relationship with P. One can easily verify that $\tilde{p}_{11}(\epsilon) = p_{11}(\epsilon)$, $\tilde{p}_{12}(\epsilon) = p_{11}(\epsilon) - (1-\nu)p_{12}(\epsilon)$ and $\tilde{p}_{22}(\epsilon) = p_{11}(\epsilon) - 2(1-\nu)p_{12}(\epsilon) + (1-\nu)^2 p_{22}(\epsilon)$. Also, when $\nu = 0$, the state space description (43) implies that the receiver only receives the transmitter innovation sequence in the presence of white Gaussian noise n_t when a packet is received. However, since the transmitter innovation sequence is also a zero mean *i.i.d.* Gaussian sequence, whether a packet is received or not, the minimum mean square estimate of the state s_t is simply its mean, which is zero. Therefore, the corresponding estimation error $\tilde{p}_{22}(\epsilon)_{|\nu=0} = p_\infty(0)$, where $p_\infty(0)$ is the variance of the transmitter innovation sequence when $\nu = 0$, which can be obtained from (32) by substituting $\nu = 0$, as $\frac{\sigma_w^2}{1-\frac{a^2}{\Lambda+1}}$. Indeed, this can be also verified by solving the corresponding quadratic equation for $\tilde{p}_{22}(\epsilon)$ after substituting $\nu = 0$. Similarly, it can be checked that $\tilde{p}_{12}(\epsilon)_{|\nu=0} = p_\infty(0)$ as well.

In what follows, we will be dropping the dependence on ϵ of the relevant quantities to keep things simple. Also, all values of the relevant quantities are computed at $\nu = 0$ unless otherwise specifically indicated. Using the equations in (44), one can show the following facts:

$$\frac{\partial \tilde{p}_{11}}{\partial \nu}\bigg|_{\nu=0} = -\frac{a^2(1-\epsilon)}{1-a^2}\frac{1}{(1+\frac{1}{\Lambda})^2}\left(2(1+\frac{1}{\Lambda})\frac{\partial \tilde{p}_{12}}{\partial \nu} - \frac{\partial \tilde{p}_{22}}{\partial \nu}\right)\bigg|_{\nu=0}$$

One can also easily show the following rather simple but useful result which states that $\frac{\partial \tilde{p}_{22}}{\partial \nu}_{|\nu=0} = 0$ regardless of the value of ϵ. Therefore, we only need to show that $\frac{\partial \tilde{p}_{12}}{\partial \nu}_{|\nu=0} > 0\ \forall \epsilon > 0$. Note that at $\epsilon = 0$, we have $\tilde{p}_{11}(0) = p_{11}(0) = p_\infty(0)$ and therefore $\frac{\partial p_{11}}{\partial \nu}_{|\nu=0} = 0$ and hence $\frac{\partial \tilde{p}_{12}}{\partial \nu}_{|\nu=0} = 0$ also at $\epsilon = 0$. Using the above facts, from (44), one can evaluate (after some algebra) that for $\epsilon > 0$,

$$\left[1 + \frac{a^2(1-\epsilon)\Lambda}{(1-\frac{a^2}{\Lambda+1})(\Lambda+1)^2}\right]\frac{\partial \tilde{p}_{12}}{\partial \nu}_{|\nu=0} = \frac{a^2 \epsilon p_\infty(0)}{1+\frac{1}{\Lambda}} \qquad (45)$$

which is clearly positive for $\epsilon > 0$. Hence we have $\frac{\partial \tilde{p}_{11}}{\partial \nu}_{|\nu=0} < 0$ for $\epsilon > 0$. Therefore $\tilde{p}_{11}(\epsilon) = p_{11}(\epsilon) = p^{SIF}(\nu, \epsilon)$ is a decreasing function of ν at $\nu = 0$ for $\epsilon > 0$.

C. Proof of Theorem 4

The proof of the first part of the theorem that $\nu^*(0) = 0$ is obvious. In order to prove the second part, we first obtain an $O(\delta)$ approximation of $p_{11}(1-\delta)$, where $\delta = 1-\epsilon \approx 0$ and then show that this approximation is minimized at $\nu^* = 1$. Using the expression for $p_{11}(\epsilon)$ from (42), one can (after some elementary analysis) show that an $O(\delta)$ approximation for $p_{11}(1-\delta)$ can be obtained as

$$\begin{aligned} p_{11}(1-\delta) \;&\approx\; \frac{\sigma_w^2}{1-a^2} - \delta\frac{a^2}{1-a^2}\left(p_{11}(1) - (1-\nu)p_{12}(1)\right)^2 \times \\ &\times \frac{1}{p_{11}(1) - 2p_{12}(1)(1-\nu) + p_{22}(1)(1-\nu)^2 + R} \end{aligned} \qquad (46)$$

One can easily obtain the values of $p_{11}(\epsilon), p_{12}(\epsilon)$ and $p_{22}(\epsilon)$ at $\epsilon = 1$ or $\delta = 0$ as $p_{11}(1) = \frac{\sigma_w^2}{1-a^2} = p^{OL}$, $p_{12}(1) = p^{OL}\frac{a\bar{k}}{1-a^2+a\bar{k}}$ and

$$p_{22}(1) = p^{OL}\left[\frac{\bar{k}^2}{1-(a-\bar{k})^2}(1+J(\nu)) + \frac{2\bar{k}^2 a(a-\bar{k})}{(1-(a-\bar{k})^2)(1-a^2+a\bar{k})}\right]$$

where $J(\nu) = \frac{R}{p^{OL}} = \frac{1}{\Lambda}(\nu^2 + (1-\nu^2)\frac{\overline{p}}{p^{OL}})$.

Substituting these expressions into (46), one can then show that the task of minimizing the $O(\delta)$ approximation of $p_{11}(\epsilon)$ is equivalent to maximizing a function $\bar{U}(\nu)$ of ν given by $\bar{U}(\nu) = \frac{F^2(\nu)}{G(\nu)}$, where $F(\nu) = 1 - \frac{(1-\nu)a\bar{k}}{(1-a^2+a\bar{k})}$ and

$$\begin{aligned} G(\nu) \;&=\; 1 - \frac{2a\bar{k}}{(1-a^2+a\bar{k})}(1-\nu) + \frac{(1-\nu)^2\bar{k}^2}{1-(a-\bar{k})^2}(1+J(\nu)) + \\ &+ \frac{(1-\nu)^2 2\bar{k}^2 a(a-\bar{k})}{\left(1-(a-\bar{k})^2\right)\left(1-a^2+a\bar{k}\right)} + J(\nu) \end{aligned}$$

Numerical examples seem to indicate that $\bar{U}(\nu)$ is an increasing function of ν for $0 \le \nu < 1$. However, it seems to be rather tedious to prove this. We use a different technique by bounding $\bar{U}(\nu)$ from above and showing that this upper bound is an increasing function for $0 \le \nu < 1$, and finally show that the upper bound is tight at $\nu = 1$. Note the expressions for $F(\nu)$ and $G(\nu)$ and that $\bar{U}(\nu) = \frac{F^2(\nu)}{G(\nu)}$. By completing a square in $G(\nu)$, it can be easily shown that $\bar{U}(\nu) \le \frac{1}{1+\frac{\bar{R}(\nu)}{F^2(\nu)}}$, where

$$\bar{R}(\nu) = \frac{(1-a^2)(1-\nu)^2\bar{k}^2}{\left(1-(a-\bar{k})^2\right)\left(1-a(a-\bar{k})\right)} + \left(1 + \frac{(1-\nu)^2\bar{k}^2}{1-(a-\bar{k})^2}\right)J(\nu)$$

Denoting $\bar{P}_\infty = \frac{\overline{p}}{p^{OL}}$ and noting that $\overline{p} < p^{OL} = \frac{\sigma_w^2}{(1-a^2)}$ for $0 \le \nu < 1$ for all $0 \le \epsilon < 1$, we have $\bar{P}_\infty < 1$. After a little algebra, it can be also shown that $\bar{P}_\infty$ satisfies $\bar{P}_\infty + J(\nu)(1-a^2) = \frac{a^2\bar{P}_\infty}{(1-\bar{P}_\infty)}$. Finally, using $\bar{k} = \frac{a\bar{P}_\infty}{\bar{P}_\infty + J(\nu)}$, one can derive that

$$\frac{a\bar{k}}{1-a(a-\bar{k})} = \frac{a^2\bar{P}_\infty}{\bar{P}_\infty + J(\nu)(1-a^2)} = (1-\bar{P}_\infty).$$

Substituting the above equality in the expression for $\bar{R}(\nu)$, one can immediately derive that $\frac{R(\nu)}{F^2(\nu)} \ge \frac{J(\nu)}{F^2(\nu)}$ and $\bar{U}(\nu) \le \frac{1}{1+\frac{J(\nu)}{F^2(\nu)}}$. We will now show that this upper bound is an increasing function of ν by showing that $\frac{J(\nu)}{F^2(\nu)}$ is a decreasing function of ν. Here we will omit the details, but will provide the key ingredients. We will need to use the fact that $\frac{d\bar{P}_\infty}{d\nu}\left[2\bar{P}_\infty(1 + \frac{1-\nu^2}{\Lambda}(1-a^2)) + (1-a^2)\left(\frac{2\nu^2}{\Lambda} - (1+\frac{1}{\Lambda})\right)\right] = \frac{2\nu}{\Lambda}(1-a^2)(1-\bar{P}_\infty)^2$. Using this one can show that $\frac{J(\nu)}{F^2(\nu)}$ is a decreasing function of ν by dividing the range of $\bar{P}_\infty$ into two intervals: $0 < \bar{P}_\infty < \frac{\nu}{1+\nu}$ and $\frac{\nu}{1+\nu} \le \bar{P}_\infty < 1$ and proving the derivative of $\frac{J(\nu)}{F^2(\nu)}$ with respect to ν is negative separately for both intervals.

The next step is to verify that $S(1) = \frac{1}{1+\frac{J(1)}{F^2(1)}} = \frac{1}{1+\frac{1}{\Lambda}}$, that is the bound is tight for $\nu = 1$. Therefore we have $\bar{U}(\nu) \le \frac{1}{1+\frac{J(\nu)}{F^2(\nu)}} \le \frac{1}{1+\frac{1}{\Lambda}} = \bar{U}(1)$ for $0 \le \nu \le 1$. This implies that an $O(1-\epsilon)$ approximation of $p_{11}(\epsilon) = p^{SIF}(\nu, \epsilon)$ is minimized at $\nu^* = 1$ when $\epsilon \to 1$ from below. Since $p^{SIF}(\nu, \epsilon)$ is a continuous function of ϵ, we can say that for ϵ sufficiently close to but less than 1, $p^{SIF}(\nu, \epsilon)$ is minimized at $\nu^* = 1$. Hence the proof of Theorem 4 follows.