

Remote estimation 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 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.

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 reliability due to packet loss and limited channel capacity. This concern is particularly evident 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 (SNR) ratio [5], [6]. Finally, another well explored approach is the analysis of control systems subject to random packet loss [7], [8], [9], [10], [11] under LQG framework. 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 adding 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 both packet loss and quantization noise. So far mainly packet loss has been considered in the context of remote estimation [12], [13], although there are recent attempts at considering both limitations [14], [15]. 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 make a copy of 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) [16] 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 state is transmitted over the channel, the second named innovation forwarding (IF), in which it is sent the difference between the state and the estimate that a receiver would have if no packet loss has occurred, and the third, named soft innovation forwarding (SIF), which 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.

II. CHANNEL MODELING AND PROBLEM FORMULATION

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

$$x_{t+1} = ax_t + w_t \quad (1)$$

$$y_t = cx_t + v_t \quad (2)$$

This work is supported by the European Community's Seventh Framework Programme [FP7/2007-2013] under grant agreement n. 257462 HYCON2 Network of excellence and by the Australian Research Council grant n. DP120101122.

S. Dey is with the Dept. of Electrical & Electronic Engineering The University of Melbourne, Parkville VIC 3052, Australia, email: sdey@unimelb.edu.au, 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

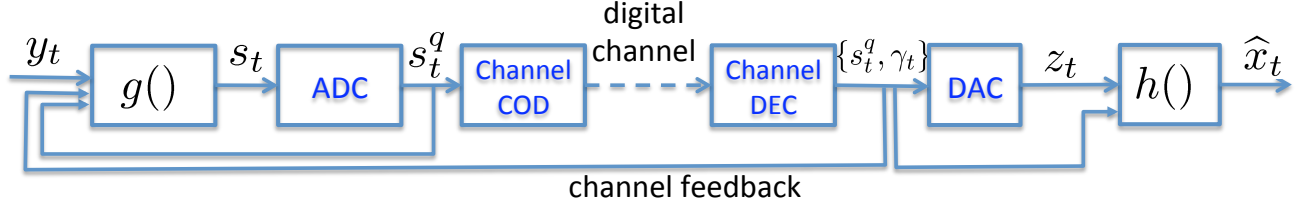


Fig. 1. Equivalent communication model for remote estimation.

where $w_t \sim \mathcal{N}(0, \sigma_w^2)$, $v_t \sim \mathcal{N}(0, \sigma_v^2)$, and $x_0 \sim \mathcal{N}(\bar{x}_0, \sigma_0^2)$ are uncorrelated. 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 binary word s_t^q , 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 $\gamma_t \in \{0, 1\} \equiv \{\text{erased}, \text{decoded}\}$. 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 \hat{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. We now proceed to mathematically model such system.

In the following we will consider the simplified assumption

$$c = 1, \quad \sigma_v^2 = 0, \quad |a| < 1 \quad (3)$$

i.e. $y_t = x_t$ is available at the transmitter. The transmitter can send a signal through a digital noisy erasure channel modelled 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 the fine quantization assumption. The variables satisfy the following assumptions:

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

where Λ is the signal-to-quantization noise ratio (SQNR) of the quantizer. This model for the SQNR noise assumes that the quantizer is matched to the distribution of the incoming signal s_t so as to maintain a constant SQNR value for 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:

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

The first set \mathcal{T}^{CF} corresponds to a scenario with perfect channel feedback, i.e. the transmitter knows 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 send back reliably is just an ACK packet.

The receiver needs to reconstruct the true state signal x_t based on its own information set \mathcal{R}_t given by:

$$\mathcal{R}_t = \{z_t, \dots, z_0, \gamma_t, \dots, \gamma_0\}$$

i.e.

$$\hat{x}_{t|t}^{rx} = h_t(\mathcal{R}_t)$$

where h_t is a measurable function. We are interested in analyzing the performance of the overall system based on the estimation prediction error variance at the receiver, i.e.

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

where the expectation has to be taken also with respect to the packet drop process γ_t besides the noises w_t, n_t .

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

We assume that the decoder has received the last packet at time $t - d$ (i.e. $d = 1$ if no losses have occurred after the last received message). 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 linear encoders

$$s_t := \mathcal{L}_t(y_t, y_{t-1}, \dots, y_{t-d+1}, z_{t-d+1}^-; \gamma_t, \dots, \gamma_0) \quad (4)$$

where $\mathcal{L}_t(y_t, y_{t-1}, \dots, y_{t-d+1}, z_{t-d+1}^-; \gamma_t, \dots, \gamma_0)$ is a linear operator of its arguments $y_t, y_{t-1}, \dots, y_{t-d+1}$ (the samples to be encoded) and z_{t-d+1}^- (the past received signals) which also depends on the packet loss sequence $\gamma_t, \dots, \gamma_0$. The result of this section is summarized in the next proposition. The remaining part of the section proves the result.

Theorem 1: Under the assumption that perfect channel feedback is available (i.e. that $\gamma_t, \dots, \gamma_0$ are known also at the transmitter side), the optimal linear encoder (4) for the linear system (1)-(2) is given by:

$$s_t := \hat{x}_{t|t}^{tx} - \hat{x}_{t|t-d}^{rx} = \mathbb{E}[x_t | y_{t,d}, \mathcal{R}_{t-d}] - \mathbb{E}[x_t | \mathcal{R}_{t-d}] \quad (5)$$

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.

Let us define $y_{t,d} := [y_t, y_{t-1}, \dots, y_{t-d+1}]^\top$ the vector of the last d measurements. We can define

$$e_{t,d} := y_{t,d} - \mathbb{E}[y_{t,d} | \mathcal{R}_{t-d}] \quad (6)$$

as the innovation (i.e. the “new” information) of $y_{t,d}$ w.r.t. \mathcal{R}_{t-d} . Note that, however, only part of this information is necessary to estimate x_t . As a matter of fact $e_{t,d}$ 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) [17]. Since x_t is scalar, the (linear) sufficient statistic in $e_{t,d}$ for x_t has dimension 1 (which is equal to the dimension of the projection of x_t onto the space spanned by the components of $e_{t,d}$).

Hence we seek for a signal $s_t = \alpha^\top e_{t,d}$, $\alpha \in \mathbb{R}^d$, so that the optimal estimation

$$\hat{x}_{t|t}^{rx} := \mathbb{E}[x_t | \mathcal{R}_t]$$

where

$$z_t = s_t^q = s_t + n_t$$

since by assumption $\gamma_t = 1$, and n_t is a zero mean white noise Gaussian process with variance $\frac{\text{Var}(s_t)}{\Lambda}$, has as small variance as possible.

Note that the “noise” n_t is known at the transmitter side since the transmitter generates z_t starting from s_t . Since $s_t = \alpha^\top e_{t,d}$, and using the fact that both the noise n_t and $e_{t,d}$ are uncorrelated from z_{t-d}^- (see also (6)), also z_t is uncorrelated from z_{t-d}^- . Therefore the estimator $\hat{x}_{t|s}^{rx} := \mathbb{E}[x_t | \mathcal{R}_s]$ satisfies:

$$\begin{aligned} \hat{x}_{t|t}^{rx} &= \mathbb{E}[x_t | \mathcal{R}_{t-d}] + \mathbb{E}[x_t | z_t] \\ &= \hat{x}_{t|t-d}^{rx} + \mathbb{E}[x_t | z_t] \\ &= \hat{x}_{t|t-d}^{rx} + \frac{\mathbb{E}[x_t s_t]}{\mathbb{E}[s_t^2](1 + \frac{1}{\Lambda})} z_t \\ &= \hat{x}_{t|t-d}^{rx} + \frac{1}{1 + \frac{1}{\Lambda}} z_t \end{aligned} \quad (7)$$

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

$$\text{Var}\{\tilde{x}_{t|t}^{rx}\} = \text{Var}\{\tilde{x}_{t|t-d}^{rx}\} - \mathbb{E}\left[\left(\mathbb{E}[x_t | z_t]\right)^2\right]$$

Since the choice of s_t does not affect the first term on the right hand side, minimizing $\text{Var}\{\tilde{x}_{t|t}^{rx}\}$ is equivalent to maximizing

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

where $\bar{s}_t := \frac{s_t}{\sqrt{\mathbb{E}[s_t^2]}}$. Hence we are left with maximizing $\mathbb{E}[x_t \bar{s}_t]$, which is obtained choosing $\alpha \in \mathbb{R}^d$ so that $s_t = \alpha^\top e_{t,d}$ has maximal correlation with x_t . This is achieved when $s_t = \mathbb{E}[x_t | e_{t,d}] = \mathbb{E}[x_t | y_{t,d}, \mathcal{R}_{t-d}] - \mathbb{E}[x_t | \mathcal{R}_{t-d}] = \hat{x}_{t|t}^{tx} - \hat{x}_{t|t-d}^{rx}$.

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

$$s_t := \hat{x}_{t|t}^{tx} - \hat{x}_{t|t-d}^{rx} = \mathbb{E}[x_t | y_{t,d}, \mathcal{R}_{t-d}] - \mathbb{E}[x_t | \mathcal{R}_{t-d}]$$

which concludes the proof. \blacksquare

B. Performance analysis

Based on the analysis in the previous subsection, the optimal linear strategy for remote estimation in the presence of channel feedback graphically represented as in Fig.2. At

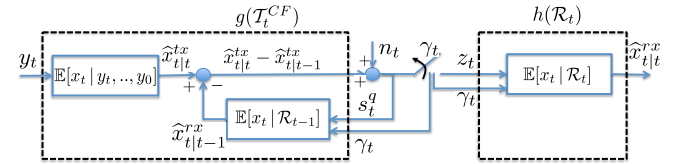


Fig. 2. Remote estimation scheme with perfect channel feedback

the transmitter the measurements are first preprocessed by a standard Kalman filter to obtain the best estimate of the state at the transmitter, that under the assumptions of Eqn. (3), this is simply the state x_t , i.e.

$$\hat{x}_{t|t}^{tx} := \mathbb{E}[x_t | \mathcal{T}_t^{CF}] = \mathbb{E}[x_t | y_t, \dots, y_0] = x_t$$

Since $\mathcal{T}_t^{CF} \supset \mathcal{R}_t$, then the transmitter is able to replicate the estimate at the receiver $\hat{x}_{t|h}^{rx} = h(\mathcal{R}_h)$. Based on the results on the previous subsection, the optimal strategy at the receiver is to construct the minimum mean squared error (MMSE) estimator given its information set, i.e.

$$\hat{x}_{t|h}^{rx} := \mathbb{E}[x_t | \mathcal{R}_h]$$

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

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

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

$$\hat{x}_{t|t-1}^{rx} = a \hat{x}_{t-1|t-1}^{rx} \quad (8)$$

$$\hat{x}_{t|t}^{rx} = \hat{x}_{t|t-1}^{rx} + k_t(z_t - \hat{z}_{t|t-1}) = \hat{x}_{t|t-1}^{rx} + \frac{\Lambda}{1 + \Lambda} z_t \quad (9)$$

where we used the result from Eqn. (7). From the expression it follows that the optimal Kalman filter is given by:

$$k_t = \frac{\Lambda}{\Lambda + 1}$$

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

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

and the corresponding variance:

$$p_{t+1|t}^{rx} = a^2 p_{t|t-1}^{rx} + \sigma_w^2 - (1 - \epsilon) \frac{\Lambda}{1 + \Lambda} a^2 p_{t|t-1}^{rx} \quad (10)$$

where we used the fact that $n_t, w_t, \tilde{x}_{t|t-1}^{rx}, \gamma_t$ are all uncorrelated and $\mathbb{E}[n_t^2] = \frac{1}{\Lambda} \mathbb{E}[\tilde{x}_{t|t-1}^{rx}] = \frac{p_{t|t-1}^{rx}}{\Lambda}$. Since $|a| < 1$ the previous linear equation has a steady state solution given by:

$$p^{CF} = \lim_{t \rightarrow \infty} p_{t+1|t}^{rx} = \frac{\sigma_w^2}{1 - a^2 \frac{1 + \epsilon \Lambda}{1 + \Lambda}} \quad (11)$$

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. $\mathcal{R}_t \not\subseteq \mathcal{T}_t^{NCF}$. As consequence, the transmitter cannot produce a copy of the transmitter estimate $\hat{x}_{t|t-1}^{rx}$. 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 as *state forwarding (SF)* is to simply transmit the current transmitter best estimate of the state x_t , i.e. $s_t = \hat{x}_{t|t}^{tx}$. The second strategy, that we refer as *innovation forwarding (IF)*, is to send the innovation between the best estimate at the transmitter $\hat{x}_{t|t}^{tx}$ and the prediction based on the past quantized transmitted signals $s_t^q = s_t + n_t$, i.e. $s_t = \hat{x}_{t|t}^{tx} - \bar{x}_{t|t-1}^{tx}$ where $\bar{x}_{t|t-1}^{tx} = \mathbb{E}[x_t | s_{t-1}^q, \dots, s_0^q]$. In practice, in this second scenario the transmitter is sending the innovation based on the (incorrect) assumption that all sent packets are received correctly, i.e. assuming $\gamma_t = 1, \forall t$. The rationale behind this strategy is that in a lossless channel, i.e. if $\epsilon = 0$, it provides the optimal strategy. For both transmitter strategies, the receiver will compute the MMSE estimator, i.e. $\hat{x}_{t|t}^{rx} = \mathbb{E}[x_t | \mathcal{R}_t]$. As just mentioned, in general $\hat{x}_{t|t}^{rx} \neq \hat{x}_{t|t}^{tx}$ and $\hat{x}_{t|t}^{rx} \neq \bar{x}_{t|t}^{tx}$. These two strategies can be graphically represented as in Fig 3, where the SF strategy corresponds to $\nu = 1$ and the IF strategy to $\nu = 0$.

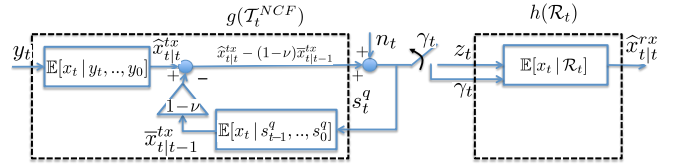


Fig. 3. Remote estimation scheme with no channel feedback

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

In this strategy, and under the simplifying assumption of Eqn. (3), we have that

$$s_t = \hat{x}_{t|t}^{tx} = x_t$$

and $\mathbb{E}[n_t^2] = \frac{1}{\Lambda} \mathbb{E}[x_t^2] = \frac{1}{\Lambda} \frac{\sigma_w^2}{1 - a^2}$, where we assumed that x_t has reached its steady state distribution. In fact, since $\mathbb{E}[x_{t+1}^2] = a^2 \mathbb{E}[x_t^2] + \sigma_w^2$, then $\lim_{t \rightarrow \infty} \mathbb{E}[x_t^2] = \frac{\sigma_w^2}{1 - a^2}$. From this expression it is clear that if $|a| \geq 1$ the state forwarding strategy cannot be used since the quantization noise variance will diverge. The message received at the remote estimator is then

$$z_t = \gamma_t (x_t + n_t)$$

which can be interpreted as a noisy measurement of the 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:

$$\hat{x}_{t|t-1}^{rx} = a\hat{x}_{t-1|t-1}^{rx} \quad (12)$$

$$\hat{x}_{t|t}^{rx} = \hat{x}_{t|t-1}^{rx} + \gamma_t k_t (z_t - \hat{x}_{t|t-1}^{rx}) \quad (13)$$

$$k_t = \frac{\hat{p}_{t|t-1}^{rx}}{\hat{p}_{t|t-1}^{rx} + r} \quad (14)$$

$$\hat{p}_{t+1|t}^{rx} = a^2 \hat{p}_{t|t-1}^{rx} + q - \gamma_t \frac{a^2 (\hat{p}_{t|t-1}^{rx})^2}{\hat{p}_{t|t-1}^{rx} + r} \quad (15)$$

where $q = \sigma_w^2$, $r = \frac{1}{\Lambda} \frac{\sigma_w^2}{1 - a^2}$, and $\hat{p}_{t|t-1}^{rx} = \mathbb{E}[(x_t - \hat{x}_{t|t-1}^{rx})^2 | \mathcal{R}_{t-1}]$. The optimal estimator could be computationally expensive since it needs to keep track of the conditional estimation error covariance $\hat{p}_{t|t-1}^{rx}$ which is a function of the packet loss history $\{\gamma_h\}_{h=0}^{t-1}$. As shown in [18], the previous filter can be replaced with the following constant gain filter:

$$\bar{x}_{t|t-1}^{rx} = a\bar{x}_{t-1|t-1}^{rx} \quad (16)$$

$$\bar{x}_{t|t}^{rx} = \bar{x}_{t|t-1}^{rx} + \gamma_t k (z_t - \bar{x}_{t|t-1}^{rx}) \quad (17)$$

$$k = \frac{p^{SF}}{p^{SF} + r} \quad (18)$$

$$p^{SF} = a^2 p^{SF} + q - (1 - \epsilon) \frac{a^2 (p^{SF})^2}{p^{SF} + r}, p^{SF} > 0 \quad (19)$$

which has the property that its error covariance converge to a steady state which is also an upper bound for the asymptotic error covariance of the optimal estimator $\hat{x}_{t|t-1}^{rx}$, i.e.

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

It has been shown in [18] that the previous inequality is quite tight, i.e. the performance degradation incurred using a constant gain rather than the optimal time-varying gain, is small.

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

In this section we consider the innovation forwarding scheme, that under the simplifying assumption of Eqn. (3) is equivalent of sending

$$s_t = \hat{x}_{t|t-1}^{tx} - \bar{x}_{t|t-1}^{tx} = x_t - \bar{x}_{t|t-1}^{tx}$$

where $\bar{x}_{t|t-1}^{tx} = \mathbb{E}[x_t | s_{t-1}^q, \dots, s_0^q]$ and $s_t^q = s_t + n_t$. The MMSE estimator at the receiver $\hat{x}_{t+1|t}^{rx} = \mathbb{E}[x_{t+1} | \mathcal{R}_t]$ must have the following expression:

$$\begin{aligned} \hat{x}_{t+1|t}^{rx} &= a\hat{x}_{t|t-1}^{rx} + ak_t(z_t - \hat{z}_t) = a\hat{x}_{t|t-1}^{rx} + ak_t z_t \\ z_t &= \gamma_t(s_t + n_t) = \gamma_t(x_t - \bar{x}_{t|t-1}^{tx} + n_t) \end{aligned} \quad (20)$$

where $\hat{z}_t := \mathbb{E}[z_t | \mathcal{R}_{t-1}] = 0$ since $x_t - \bar{x}_{t|t-1}^{tx}$ 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 $\hat{p}_{t+1|t}^{rx} := \mathbb{E}[(x_{t+1} - \hat{x}_{t+1|t}^{rx})^2 | \mathcal{R}_t]$.

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 $\hat{x}_{t|t-1}^{rx} = x_t - \bar{x}_{t|t-1}^{tx}$, which is obtained by subtracting the state prediction update Eqn. (20) from the state equation Eqn. (1), obtaining

$$\begin{aligned} \hat{x}_{t+1|t}^{rx} &= a\hat{x}_{t|t-1}^{rx} - ak_t z_t + w_t \\ &= a\hat{x}_{t|t-1}^{rx} - \gamma_t ak_t (x_t - \bar{x}_{t|t-1}^{tx} + n_t) + w_t \\ &= a(1 - \gamma_t k_t) \hat{x}_{t|t-1}^{rx} - \gamma_t ak_t (\hat{x}_{t|t-1}^{rx} - \bar{x}_{t|t-1}^{tx} + n_t) + w_t \end{aligned}$$

For future use let us define $\Delta \hat{x}_t := \hat{x}_{t|t-1}^{rx} - \bar{x}_{t|t-1}^{tx} = \hat{x}_{t|t-1}^{rx} - \bar{x}_{t|t-1}^{tx}$ where $\bar{x}_{t|t-1}^{tx} := x_t - \bar{x}_{t|t-1}^{tx}$. This implies that $\hat{x}_{t|t-1}^{rx} = \Delta \hat{x}_t + \bar{x}_{t|t-1}^{tx}$. Since $\Delta \hat{x}_t$ is a function of the past data and $\bar{x}_{t|t-1}^{tx}$ is the error of the optimal estimator using past data, $\mathbb{E}[\Delta \hat{x}_t \bar{x}_{t|t-1}^{tx} | \mathcal{R}_{t-1}] = 0$ so that

$$\begin{aligned} \mathbb{E}[\hat{x}_{t+1|t}^{rx} \bar{x}_{t|t-1}^{tx} | \mathcal{R}_{t-1}] &= \mathbb{E}[(\Delta \hat{x}_t + \bar{x}_{t|t-1}^{tx}) \bar{x}_{t|t-1}^{tx} | \mathcal{R}_{t-1}] \\ &= \mathbb{E}[\bar{x}_{t|t-1}^{tx} \bar{x}_{t|t-1}^{tx} | \mathcal{R}_{t-1}] =: \hat{p}_t^0 \\ \mathbb{E}[\hat{x}_{t+1|t}^{rx} \Delta \hat{x}_t | \mathcal{R}_{t-1}] &= \mathbb{E}[\bar{x}_{t|t-1}^{tx} (\hat{x}_{t|t-1}^{rx} - \bar{x}_{t|t-1}^{tx}) | \mathcal{R}_{t-1}] \\ &= \hat{p}_{t|t-1}^{rx} - \hat{p}_t^0 \\ \mathbb{E}[\Delta \hat{x}_t \Delta \hat{x}_t | \mathcal{R}_{t-1}] &= \mathbb{E}[(\hat{x}_{t|t-1}^{rx} - \bar{x}_{t|t-1}^{tx}) \Delta \hat{x}_t | \mathcal{R}_{t-1}] \\ &= \mathbb{E}[\hat{x}_{t|t-1}^{rx} \Delta \hat{x}_t | \mathcal{R}_{t-1}] = \hat{p}_{t|t-1}^{rx} - \hat{p}_t^0 \end{aligned}$$

where $\hat{p}_t^0 = \mathbb{E}[(x_t - \bar{x}_{t|t-1}^{tx})^2 | \mathcal{R}_{t-1}]$. Using these conditions one obtains:

$$\begin{aligned} \hat{p}_{t+1|t}^{rx} &= (a - \gamma_t ak_t)^2 \hat{p}_{t|t-1}^{rx} + \sigma_w^2 + a^2 \gamma_t^2 k_t^2 (\hat{p}_{t|t-1}^{rx} - \hat{p}_t^0 + \frac{\hat{p}_t^0}{\Lambda}) \\ &\quad + 2a^2 \gamma_t k_t (1 - \gamma_t k_t) (\hat{p}_{t|t-1}^{rx} - \hat{p}_t^0) \end{aligned} \quad (21)$$

Taking the derivative w.r.t. k_t we obtain:

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

which, equated to zero has the unique solution

$$k_t = \gamma_t \frac{\Lambda}{1 + \Lambda} \quad (22)$$

Inserting k_t back into (21) we obtain:

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

from which it follows that the expected error covariance $p_{t+1|t}^{rx} = \mathbb{E}[(\hat{x}_{t+1|t}^{rx})^2]$ is given by

$$p_{t+1|t}^{rx} = a^2 p_{t|t-1}^{rx} + \sigma_w^2 - (1 - \epsilon) a^2 p_t^0 \frac{\Lambda}{1 + \Lambda} \quad (23)$$

where $p_t^0 := \mathbb{E}[\hat{p}_t^0]$. It is interesting to observe that the gain k_t in (22) 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, since p_t^0 is the prediction error covariance with no packet loss, which is given by Eqn. (11) by setting $\epsilon = 0$, then

$$\lim_{t \rightarrow \infty} p_t^0 = \frac{\sigma_w^2}{1 - \frac{a^2}{1 + \Lambda}}$$

from which it follows that the steady state prediction error covariance is given by:

$$p^{IF} = \lim_{t \rightarrow \infty} p_{t+1|t}^{rx} = \frac{(\Lambda + 1)(1 - a^2) + \epsilon a^2 \Lambda}{(1 - a^2)(1 - a^2 + \Lambda)} \quad (24)$$

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. (19) for the state forwarding and by Eqn. (24) for the innovation forwarding, as a function of the systems parameters a, Λ, ϵ . In particular, we are interested in finding the set $\Phi := \{(a, \Lambda, \epsilon) | p^{SF} \leq p^{IF}\}$, i.e. the set of parameters where the SF strategy has a better performance than the IF strategy.

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

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

where

$$\epsilon_c(a, \Lambda) = \frac{(1 - a^2)(\Lambda + 2)}{2a^2 \Lambda} \left(\sqrt{1 + \frac{4a^2 \Lambda}{(\Lambda + 2)^2 (1 - a^2)}} - 1 \right) \quad (25)$$

Moreover, the critical probability $\epsilon_c(a, \Lambda)$ is monotonically decreasing in Λ and $|a|$, and

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

and

$$\epsilon_c < \frac{1}{2}$$

Proof: See Appendix A. ■

The previous theorem implies that the IF strategy performs better than the SF strategy only for small packet loss probabilities, and more specifically for $\epsilon < \epsilon_c$. Such 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 if the packet loss probability is greater than one half.

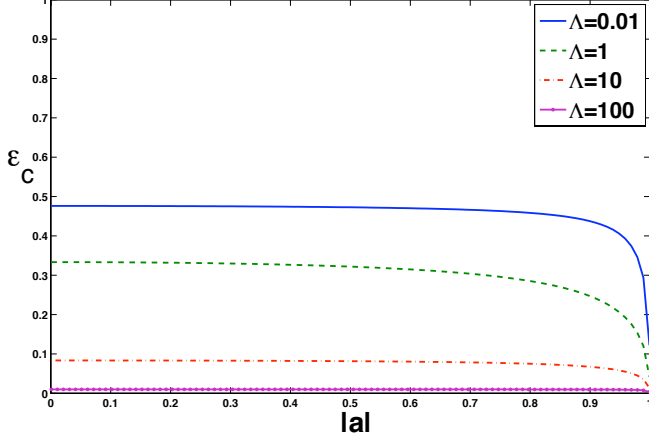


Fig. 4. Critical probability ϵ_c as a function of $|a|$ for different values of the SQNR Λ .

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 include 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 $\hat{x}_t^{tx} = \mathbb{E}[x_t | \mathcal{T}_t^{NCF}]$ and the innovation between its best estimate and the best estimate of the state given the past quantised transmitted signals, i.e. $\Delta \hat{x}_t = \hat{x}_t^{tx} - \bar{x}_{t|t-1}^{tx}$ where $\bar{x}_{t|t-1}^{tx} = \mathbb{E}[x_t | s_{t-1}^q, \dots, s_0^q]$. We call this scheme the *soft innovation forwarding (SIF)* scheme. In this case, the transmitted signal is thus given by

$$s_t = \nu \hat{x}_t^{tx} + (1 - \nu) \Delta \hat{x}_t = \hat{x}_t^{tx} - (1 - \nu) \bar{x}_{t|t-1}^{tx} \quad (26)$$

where $0 \leq \nu \leq 1$ is fixed at the transmitter. This scheme is graphically illustrated in Fig. 3. Under the simplifying hypotheses of Eqn. (3), then the previous equation reduces to $s_t = x_t - (1 - \nu) \bar{x}_{t|t-1}^{tx}$.

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

In this section, we explicitly compute the transmitter filter function $g(\mathcal{T}_t^{NCF})$ based on the SIF strategy. Basically, it reduces to the problem of computing the equation for the internal estimator $\bar{x}_{t|t-1}^{tx}$. Since the dynamical systems is

linear with additive gaussian noise, then the optimal MMSE estimator is linear in the quantized transmitted signals s_t^q and it is given by the Kalman Filter. However, the equations are somewhat non-standard 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 $\bar{p}_{t|h} = \mathbb{E}[(\tilde{x}_{t|h}^{tx})^2]$, where $\tilde{x}_{t|h}^{tx} = x_t - \bar{x}_{t|h}^{tx}$. Based on this definition, we can compute the power of the transmitted signal s_t as follows:

$$\begin{aligned} \mathbb{E}[s_t^2] &= \mathbb{E}[(\nu x_t + (1 - \nu) \tilde{x}_{t|t-1}^{tx})^2] \\ &= (\nu^2 \mathbb{E}[x_t^2] + (1 - \nu)^2 \bar{p}_{t|t-1} + \\ &\quad + 2\nu(1 - \nu) \mathbb{E}[(\bar{x}_{t|t-1}^{tx} + \tilde{x}_{t|t-1}^{tx}) \tilde{x}_{t|t-1}^{tx}]) \\ &= \nu^2 \frac{\sigma_w^2}{1 - a^2} + (1 - \nu^2) \bar{p}_{t|t-1} \end{aligned}$$

where we used the fact that $\bar{x}_{t|t-1}^{tx}$ and $\tilde{x}_{t|t-1}^{tx}$ are uncorrelated, and that x_t is assumed to be in its steady state distribution. The equation of the filter are given by:

$$\begin{aligned} \bar{x}_{t+1|t}^{tx} &= a \bar{x}_{t|t-1}^{tx} + k_t^{tx} (z_t - \hat{z}_{t|t-1}) \\ \hat{z}_{t|t-1} &= \mathbb{E}[z_t | s_{t-1}^q, \dots, s_0^q] = \nu \bar{x}_{t|t-1}^{tx} \\ k_t^{tx} &= \frac{a \bar{p}_{t|t-1}}{\bar{p}_{t|t-1} + \mathbb{E}[n_t^2]} = \frac{a \bar{p}_{t|t-1}}{(1 + \frac{1-\nu^2}{\Lambda}) \bar{p}_{t|t-1} + \frac{\nu^2 \sigma_w^2}{\Lambda(1-a^2)}} \end{aligned}$$

Such filter will reach a steady state, therefore, as standards, it is possible to substitute it with its steady state implementation, since it will reach the same steady state performance. The steady state filter is given by:

$$\begin{aligned} \bar{x}_{t+1|t}^{tx} &= (a - \nu \bar{k}) \bar{x}_{t|t-1}^{tx} + \bar{k} z_t \\ \bar{k} &= \frac{a \bar{p}}{(1 + \frac{1-\nu^2}{\Lambda}) \bar{p} + \frac{\nu^2 \sigma_w^2}{\Lambda(1-a^2)}} \quad (27) \\ \bar{p} &= a^2 \bar{p} + \sigma_w^2 - \frac{a^2 \bar{p}^2}{(1 + \frac{1-\nu^2}{\Lambda}) \bar{p} + \frac{\nu^2 \sigma_w^2}{\Lambda(1-a^2)}}, \quad \bar{p} > 0 \quad (28) \end{aligned}$$

where the last equation represents a Riccati-like equation which has a unique stabilizing positive solution \bar{p} .

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

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

$$\begin{aligned} \begin{bmatrix} x_{t+1} \\ \bar{x}_{t+1|t}^{tx} \end{bmatrix} &= \underbrace{\begin{bmatrix} a & 0 \\ \bar{k} & (a - \bar{k}) \end{bmatrix}}_A \underbrace{\begin{bmatrix} x_t \\ \bar{x}_{t|t-1}^{tx} \end{bmatrix}}_{\xi_t} + \underbrace{\begin{bmatrix} w_t \\ \bar{k} n_t \end{bmatrix}}_{\eta_t} \quad (29) \\ z_t &= \underbrace{\gamma_t \begin{bmatrix} 1 & -(1 - \nu) \end{bmatrix}}_C \begin{bmatrix} x_t \\ \bar{x}_{t|t-1}^{tx} \end{bmatrix} + \gamma_t n_t \quad (30) \end{aligned}$$

As a consequence the estimator $\hat{x}_{t+1|t}^{rx} = \mathbb{E}[x_{t+1} | \mathcal{R}_t]$ corresponds of the first component of the optimal estimator $\hat{\xi}_{t+1|t} = \mathbb{E}[\xi_{t+1} | \mathcal{R}_t]$ 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 [18]. The (suboptimal) receiver filter design is then given by:

$$\bar{\xi}_{t+1|t} = (A - \gamma_t K C) \bar{\xi}_{t|t-1} + \gamma_t K z_t \quad (31)$$

$$\bar{x}_{t|t-1}^{rx} = h(\mathcal{R}_{t-1}) = \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_H \bar{\xi}_{t|t-1} \quad (32)$$

$$K = (APC^T + S)(CPC^T + R)^{-1} \quad (33)$$

$$P = APA^T + Q - (1 - \epsilon)K(CPC^T + R)K^T = \Psi(P) \quad (34)$$

$$R = \mathbb{E}[n_t^2] = \frac{1}{\Lambda} \left(\nu^2 \frac{\sigma_w^2}{1 - a^2} + (1 - \nu^2) \bar{p} \right)$$

$$Q = \mathbb{E}[\eta_t \eta_t^T] = \begin{bmatrix} \sigma_w^2 & 0 \\ 0 & \bar{k}^2 R \end{bmatrix}$$

$$S = \mathbb{E}[\eta_t n_t] = \begin{bmatrix} 0 \\ \bar{k} R \end{bmatrix}$$

The steady state Kalman gain K can therefore be obtained by finding the unique positive definite solution $P > 0$ that solves the Riccati-like equation (34) and the steady state prediction error has the following upper bound:

$$\limsup_{t \rightarrow \infty} \mathbb{E}[(x_t - \hat{x}_{t|t-1}^{rx})^2] \leq p^{SIF} = HPH^T \quad (35)$$

The expression for the covariance matrix P can be computed from three paired nonlinear equations. 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 \bar{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:

$$\begin{aligned} p_{11}(\epsilon) &= \frac{\sigma_w^2}{1 - a^2} - \frac{a^2}{1 - a^2} \frac{(1 - \epsilon)}{M_\infty(\nu)} (p_{11}(\epsilon) - (1 - \nu)p_{12}(\epsilon))^2 \\ p_{12}(\epsilon) &= \frac{a\bar{k}}{1 - a^2 + a\bar{k}} p_{11}(\epsilon) - \frac{a(1 - \epsilon)}{1 - a^2 + a\bar{k}} (p_{11}(\epsilon) - (1 - \nu)p_{12}(\epsilon)) \frac{L_\infty(\nu)}{M_\infty(\nu)} \\ p_{22}(\epsilon) &= \frac{\bar{k}^2}{1 - (a - \bar{k})^2} p_{11}(\epsilon) + \frac{2\bar{k}(a - \bar{k})}{1 - (a - \bar{k})^2} p_{12}(\epsilon) \\ &\quad + \frac{\bar{k}^2}{1 - (a - \bar{k})^2} R - \frac{(1 - \epsilon)}{1 - (a - \bar{k})^2} \frac{L_\infty^2(\nu)}{M_\infty(\nu)} \end{aligned} \quad (36)$$

where

$$M_\infty(\nu) = p_{11}(\epsilon) - 2p_{12}(\epsilon)(1 - \nu) + p_{22}(\epsilon)(1 - \nu)^2 + R,$$

and

$$\begin{aligned} L_\infty(\nu) &= \bar{k}p_{11}(\epsilon) + (a - \bar{k}(2 - \nu))p_{12}(\epsilon) - \\ &\quad - (a - \bar{k})(1 - \nu)p_{22}(\epsilon) + \bar{k}R \end{aligned}$$

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 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:

$$\nu^*(\epsilon) := \arg \min_{\nu \in [0, 1]} p^{SIF}(\nu, \epsilon) \quad (37)$$

$$p^{OSIF}(\epsilon) := p^{SIF}(\nu^*, \epsilon) \quad (38)$$

where $p^{OSIF}(\epsilon)$ is the optimal soft innovation forward (OSIF) strategy for a given packet loss probability ϵ . As shown in the next section, the numerical computation of $\nu^*(\epsilon)$ via 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 shows that the SF strategy is the optimal strategy for very large packet loss probability. This is consistent with the next theorem 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: 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 1: 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.

It is also possible to formally prove 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: The optimal mixing parameter $\nu^*(\epsilon)$ has the following properties:

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

Proof: See Appendix C. ■

VI. NUMERICAL RESULTS

Figure 5 depicts the performance of the filters derived so far and the critical probability ϵ_c defined in Eqn. (25). 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.

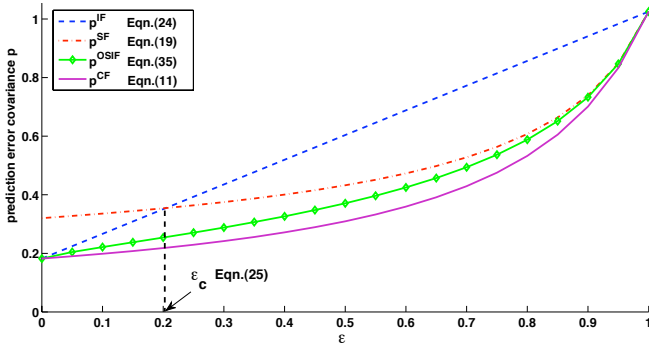


Fig. 5. Prediction error covariance of proposed strategies against packet loss probability ϵ for $a = 0.95$, $\Lambda = 1$, $\sigma_w^2 = 1$.

In Figure 6 below, we plot the optimal mixing coefficient $\nu = \nu^*$ which has been obtained numerically. 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.

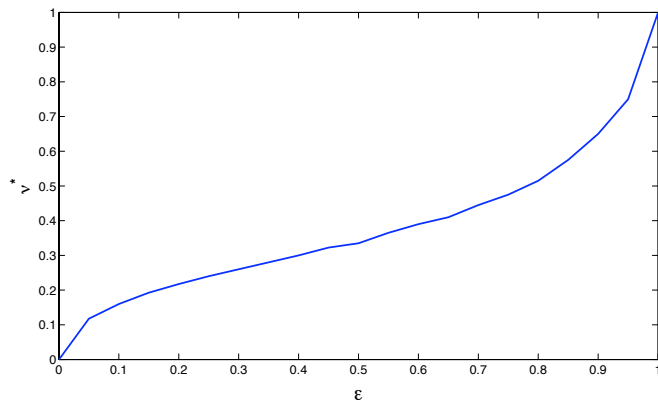


Fig. 6. Optimal mixing coefficient ν^* as a function of the packet loss probability ϵ for the OSIF strategy for $a = 0.95$, $\Lambda = 1$, $\sigma_w^2 = 1$.

VII. CONCLUSIONS AND FUTURE WORK

In this work we studied the problem of remotely estimating the state of a dynamical stable system over a noisy channel subject to packet loss. 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 propose few heuristic strategies for which we were able to characterize performance and trade-offs.

Future work will include formal proofs about the properties of the OSIF strategy, and the extension to multivariable systems. Also, another important feature is to consider a lossy channel feedback which is more realistic than the two scenarios presented in this work.

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] L. Coviello, P. Minero, and M. Franceschetti, "Stabilization over markov feedback channels," *Transactions on Automatic Control (provisionally accepted)*.
- [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, 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, 2007.
- [6] E. Silva, G. Goodwin, and D. Quevedo, "Control system design subject to snr constraints," *Automatica*, vol. 46, no. 2, 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. Yüksel, 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 - accepted*, July 2013.
- [11] —, "LQG cheap control over SNR-limited lossy channels with delay," University of Padova, technical report, submitted to IEEE CDC, 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] S. Chitode, *Digital communications*. Technical Publications, 2009.
- [17] D. Cook and B. Li, "Dimension reduction for conditional mean in regression," *The Annals of Statistics*, vol. 30, no. 2, pp. 455–474, 2002.
- [18] 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.

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 $\Delta p(a, \Lambda, \epsilon) = p^{SF} - p^{IF}$, where we also make explicit the dependence of the performance in terms of the parameters. Since p^{SF} is a convex function of ϵ for fixed a and Λ , as shown in [12], and since p^{IF} is an affine function of ϵ , then also $\Delta p(a, \Lambda, \epsilon)$ is convex in ϵ . It is straightforward to observe that for $\Delta p(a, \Lambda, 1) = 0$ and

$\Delta p(a, \Lambda, 0) > 0$, therefore if we can show that there exist $\epsilon_c \in (0, 1)$ such that $\Delta p(a, \Lambda, \epsilon_c) = 0$, then this implies that $p^{SF} \leq p^{IF}$ for $\epsilon \geq \epsilon_c$. We now show that this is the case. If $\Delta p(a, \Lambda, \epsilon_c) = 0$, then $p^{SF} = p^{IF} = p$. The points p that satisfy this equality must also satisfy the Eqn. (19) and Eqn. (24), therefore, if we take the difference and recalling that $a \neq 0$ and $\epsilon \neq 1$ we have:

$$\frac{p^2}{p+r} = \frac{p^0}{1 + \frac{1}{\Lambda}}$$

from which it follows:

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

Substituting Eqn. (24) into the previous expression and after some manipulations, which are valid for $\epsilon \neq 0$, $a \neq 0$ 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. (25).

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

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

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:

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

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

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

From this analysis, it follows that

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

which concludes the proof.

B. Proof of Theorem 3

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 $\frac{\partial p_{11}(\epsilon)}{\partial \nu}|_{\nu=1}$. We also use $l_2 = \frac{\sigma_w^2}{1-a^2}$ in order to simplify notations.

Using the equation for p_{11} from (36), and taking partial derivatives, we have (after some algebra):

$$\begin{aligned} \frac{\partial p_{11}(\epsilon)}{\partial \nu}|_{\nu=1} & \left[1 + \frac{a^2(1-\epsilon)p_{11}}{(1-a^2)(p_{11} + \frac{l_2}{\Lambda})} \left(2 - \frac{p_{11}}{(p_{11} + \frac{l_2}{\Lambda})} \right) \right] \\ & = \frac{2a^2}{1-a^2}(1-\epsilon) \frac{p_{11}l_2}{\Lambda(p_{11} + \frac{l_2}{\Lambda})^2} \left[p_{11}(1 - \frac{\bar{p}}{l_2}) - p_{12} \right] \end{aligned}$$

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 $p_{12} < p_{11}(1 - \frac{\bar{p}}{l_2})$. To this end, note that from the equation for p_{12} from (36), it follows that (recall that all values are evaluated at $\nu = 1$), $p_{12} < \frac{a\bar{k}}{1-a^2+a\bar{k}}p_{11}$. It can be shown that for all $0 \leq \nu < 1$ (see Proof of Theorem 4 below)

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

where $\bar{P}_\infty = \frac{\bar{p}}{l_2}$. Hence it follows that $p_{12} < p_{11}(1 - \bar{P}_\infty) = p_{11}(1 - \frac{\bar{p}}{l_2})$. 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 $\begin{bmatrix} x_{t+1}^a \\ x_{t+1}^b \end{bmatrix} = T \begin{bmatrix} x_t \\ \hat{x}_{t+1|t} \end{bmatrix}$, where $T = \begin{bmatrix} 1 & 0 \\ 1 & -(1-\nu) \end{bmatrix}$. It can be easily checked that $x_t^a = x_t$ and $x_t^b = s_t$. Using this transformation, we can write a new state space system as

$$\begin{aligned} \begin{bmatrix} x_{t+1} \\ s_{t+1} \end{bmatrix} & = \bar{A}_1 \begin{bmatrix} x_t \\ s_t \end{bmatrix} + T \begin{bmatrix} w_t \\ \bar{k}n_t \end{bmatrix} \\ z_t & = \gamma_t \bar{C}_1 \begin{bmatrix} x_t \\ s_t \end{bmatrix} + \gamma_t n_t \end{aligned} \quad (39)$$

where $\bar{A}_1 = TAT^{-1} = \begin{bmatrix} a & 0 \\ \nu\bar{k} & (a-\bar{k}) \end{bmatrix}$ and $\bar{C}_1 = \bar{C}T^{-1} = [0 \ 1]$. 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 $\tilde{P} = \begin{bmatrix} \tilde{p}_{11}(\epsilon) & \tilde{p}_{12}(\epsilon) \\ \tilde{p}_{12}(\epsilon) & \tilde{p}_{22}(\epsilon) \end{bmatrix}$ where the individual elements satisfy the following equations:

$$\begin{aligned} \tilde{p}_{11}(\epsilon) & = \frac{\sigma_w^2}{1-a^2} - \frac{a^2(1-\epsilon)}{1-a^2} \frac{\tilde{p}_{12}(\epsilon)}{\tilde{p}_{22}(\epsilon) + R} \\ \tilde{p}_{12}(\epsilon) & = \frac{1}{1-a(a-\bar{k})} \left[a\nu\bar{k}\tilde{p}_{11}(\epsilon) + \sigma_w^2 \right. \\ & \quad \left. - \frac{(1-\epsilon)a\tilde{p}_{12}(\epsilon)}{\tilde{p}_{22}(\epsilon) + R} \bar{V}(\nu) \right] \\ \tilde{p}_{22}(\epsilon) & = \frac{1}{1-(a-\bar{k})^2} \left[\nu^2\bar{k}^2\tilde{p}_{11}(\epsilon) \right. \\ & \quad \left. + 2\nu\bar{k}(a-\bar{k})\tilde{p}_{12}(\epsilon) + \sigma_w^2 \right. \\ & \quad \left. + (1-\nu)^2\bar{k}^2R - (1-\epsilon) \frac{F_\infty^2(\nu)}{\tilde{p}_{22}(\epsilon) + R} \right] \end{aligned} \quad (40)$$

where we have used $\bar{V}(\nu) = \nu\bar{k}\tilde{p}_{12}(\epsilon) + (a-\bar{k})\tilde{p}_{22}(\epsilon) - (1-\nu)\bar{k}R$ 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 (39) 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 (28) 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 (40), one can show the following facts:

$$\frac{\partial \tilde{p}_{11}}{\partial \nu} \Big|_{\nu=0} = -\frac{a^2(1-\epsilon)}{1-a^2} \frac{1}{(1+\frac{1}{\Lambda})^2} \left(2\left(1+\frac{1}{\Lambda}\right) \frac{\partial \tilde{p}_{12}}{\partial \nu} \Big|_{\nu=0} - \frac{\partial \tilde{p}_{22}}{\partial \nu} \Big|_{\nu=0} \right) \quad (41)$$

One can also easily show the following rather simple but useful result which states that $\frac{\partial \tilde{p}_{22}}{\partial \nu} \Big|_{\nu=0} = 0$ regardless of the value of ϵ . Therefore, we only need to show that $\frac{\partial \tilde{p}_{12}}{\partial \nu} \Big|_{\nu=0} > 0$ $\forall \epsilon > 0$. Note that at $\epsilon = 0$, we have $\tilde{p}_{11}(0) = p_{11}(0) = p_{\infty}(0)$ and therefore $\frac{\partial \tilde{p}_{11}}{\partial \nu} \Big|_{\nu=0} = 0$ and hence $\frac{\partial \tilde{p}_{12}}{\partial \nu} \Big|_{\nu=0} = 0$ also at $\epsilon = 0$. Using the above facts, from (40), one can evaluate (after some algebra) that for $\epsilon > 0$,

$$\left[1 + \frac{a^2(1-\epsilon)\Lambda}{(1-\frac{a^2}{\Lambda+1})(\Lambda+1)^2} \right] \frac{\partial \tilde{p}_{12}}{\partial \nu} \Big|_{\nu=0} = \frac{a^2 \epsilon p_{\infty}(0)}{1+\frac{1}{\Lambda}} \quad (42)$$

which is clearly positive for $\epsilon > 0$. Hence we have $\frac{\partial \tilde{p}_{11}}{\partial \nu} \Big|_{\nu=0} < 0$ 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 (36), one can (after some elementary analysis) show that an $O(\delta)$ approximation for $p_{11}(1-\delta)$ can be obtained as

$$p_{11}(1-\delta) \approx \frac{\sigma_w^2}{1-a^2} - \delta \frac{a^2}{1-a^2} (p_{11}(1) - (1-\nu)p_{12}(1))^2 \times \frac{1}{p_{11}(1) - 2p_{12}(1)(1-\nu) + p_{22}(1)(1-\nu)^2 + R} \quad (43)$$

One can easily obtain the values of $p_{11}(\epsilon)$, $p_{12}(\epsilon)$ and $p_{22}(\epsilon)$ at $\epsilon = 1$ or $\delta = 0$ as $p_{11}(1) = \frac{\sigma_w^2}{1-a^2} = l_2$, $p_{12}(1) = l_2 \frac{a\bar{k}}{1-a^2+a\bar{k}}$ and

$$p_{22}(1) = l_2 \left[\frac{\bar{k}^2}{1-(a-\bar{k})^2} (1+J(nu)) + \frac{2\bar{k}^2 a(a-\bar{k})}{(1-(a-\bar{k})^2)(1-a^2+a\bar{k})} \right]$$

where $J(\nu) = \frac{R}{l_2} = \frac{1}{\Lambda}(\nu^2 + (1-\nu^2)\frac{\bar{p}}{l_2})$. Substituting these expressions into (43), 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

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

and Numerical examples seem to indicate that $\bar{U}(\nu)$ is an increasing function of ν for $0 \leq \nu < 1$. 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 $0 \leq \nu < 1$, and finally show that the upper bound is tight at $\nu = 1$. Noting that the expressions for $F(\nu)$ and $G(\nu)$ in $\bar{U}(\nu) = \frac{F^2(\nu)}{G(\nu)}$, and completing a square in $G(\nu)$, it can be easily shown that $\bar{U}(\nu) \leq \frac{1}{1+\frac{R(\nu)}{F^2(\nu)}}$, where

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

Denoting $\bar{P}_{\infty} = \frac{\bar{p}}{l_2}$ and noting that $\bar{p} < l_2 = \frac{\sigma_w^2}{(1-a^2)}$ for $0 \leq \nu < 1$ for all $0 \leq \epsilon < 1$, we have $\bar{P}_{\infty} < 1$. After a little algebra, it can be also shown that \bar{P}_{∞} satisfies the quadratic equation

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

Finally, using the expression for $\bar{k} = \frac{a\bar{P}_{\infty}}{\bar{P}_{\infty}+J(\nu)}$, 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 $\frac{\bar{R}(\nu)}{F^2(\nu)} \geq \frac{J(\nu)}{F^2(\nu)}$ since $(1-\nu)^2 \geq 0$ for all $0 \leq \nu \leq 1$. Recall that this also implies that $\bar{U}(\nu) \leq \frac{1}{1+\frac{J(\nu)}{F^2(\nu)}}$. 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 $\frac{d\bar{P}_\infty}{d\nu} \left[2\bar{P}_\infty \left(1 + \frac{1-\nu^2}{\Lambda} (1-a^2) \right) + (1-a^2) \left(\frac{2\nu^2}{\Lambda} - \left(1 + \frac{1}{\Lambda} \right) \right) \right] = \frac{2\nu}{\Lambda} (1-a^2) (1-\bar{P}_\infty)^2$. 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}_∞ into two intervals: $0 < \bar{P}_\infty < \frac{\nu}{1+\nu}$ and $\frac{\nu}{1+\nu} \leq \bar{P}_\infty < 1$ 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 $\bar{U}(\nu) \leq \frac{1}{1 + \frac{J(\nu)}{F^2(\nu)}} \leq \frac{1}{1 + \frac{1}{\Lambda}} = \bar{U}(1)$ for $0 \leq \nu \leq 1$.

This implies that an $O(1 - \epsilon)$ approximation of $p_{11}(\epsilon) = p^{SIF}(\nu, \epsilon)$ is minimized at $\nu^* = 1$ when $\epsilon \rightarrow 1$ 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.