Rate of Prefix-free Codes in LQG Control Systems with Side Information

Travis C. Cuvelier
Department of Electrical and Computer Engineering
The University of Texas at Austin, Austin, TX USA
Email: tcuvelier@utexas.edu

Takashi Tanaka
Department of Aerospace Engineering and Engineering Mechanics
The University of Texas at Austin, Austin, TX USA
Email: ttanaka@utexas.edu

Abstract—In this work, we study an LQG control system where one of two feedback channels is discrete and incurs a communication cost. We assume that a decoder (co-located with the controller) can make noiseless measurements of a subset of the state vector (referred to as side information) meanwhile a remote encoder (co-located with a sensor) can make arbitrary measurements of the entire state vector, but must convey its measurements to the decoder over a noiseless binary channel. Use of the channel incurs a communication cost, quantified as the time-averaged expected length of prefix-free binary codeword. We study the tradeoff between the communication cost and control performance. The formulation motivates a constrained directed information minimization problem, which can be solved via convex optimization. Using the optimization, we propose a quantizer design and a subsequent achievability result.

I. INTRODUCTION

In this work we consider discrete-time MIMO LQG control in a system where some measurements incur a communication cost, but others do not. As in [1] and [2], we study the tradeoff between control performance and communication cost, where the latter is measured in terms of the average length of prefix-free codewords. Our principal motivation is a sensing scenario where an energy constrained remote platform (the encoder) must encode, and then wirelessly transmit, its measurements to a joint fusion center/controller (decoder) which contains some sensors of its own. We model the decoder measurements as noiseless observations of a subset of the state vector indices, which we refer to as side information (SI). We consider a setup where both the encoder and decoder have access to the decoder’s measurements. In the remote sensing scenario, it may be reasonable to assume that the decoder has sufficient energy to feed its measurements back to the encoder while the sensor platform could be constrained—under some additional assumptions, minimizing the time-averaged bitrate from the encoder to decoder is a surrogate for minimizing the energy the sensor platform “spends” on communication. We establish a converse bound on the minimum prefix-free codeword length in terms of Massey’s directed information (DI) [3]. The bound applies to the case when the SI is known at both the encoder and decoder, and thus applies when the SI is known at the decoder only. The converse motivates a rate distortion problem where a DI term is minimized subject to a constraint on control performance. The problem is solved optimally via a tractable mathematical program (namely a log-determinant optimization) [4]. We use the optimization to derive an achievability result based on the construction in [2].

Massey’s DI quantifies the flow of information from one stochastic process to another [3]. In [5], the time-averaged bitrate of a prefix-free codec inserted into the feedback loop of a SISO control system was shown to be lower bounded by the DI from the plant output to the control input. Also, [5] motivated the use of entropy dithered quantization (EDQ) in control systems subject to data rate constraints. Extending [5] to the MIMO setting, [1] motivated a rate distortion problem that minimized DI in an LQG control system subject to a constraint on performance. Under standard linear/Gaussian plant dynamics, [1] showed that any optimal measurement and control policy could be implemented via a three-stage separation architecture; namely a linear/Gaussian sensor, a Kalman filter, a certainty equivalence linear feedback controller. The optimization to find the minimum DI (and the minimizing policy) was formulated as a semidefinite (log-determinant) program [1]. [2] gave operational significance to the minimal DI (and minimizing policy) in [1]. In [2], it was shown that a zero-delay source coding scheme, based on quantizing Kalman filter innovations via EDQ, followed by prefix-free coding achieves a DI cost within \(\frac{2}{3} \log \left(\frac{2}{\pi e}\right) + 1\) bits of the minimal DI cost in [1]. Likewise, [6] studied the tradeoff between DI and LQG performance, proved converse bounds applying to plants with non-Gaussian disturbances, and demonstrated achievability without dithering.

The impact of SI (modeled as a decoder-side linear observation of the state vector in additive Gaussian noise) on the tradeoff between DI and LQG performance in LTI SISO systems was investigated in [7]. It was argued that it suffices to consider a rate distortion problem in a related tracking problem and that linear/Gaussian policies were optimal. In [8], an optimization problem was formulated to analyze the minimum attainable DI in a MIMO LQG control system with SI assuming linear feedback policies. Very recently, [4] proved that linear/Gaussian policies conforming to the “three-stage separation” architecture of [1] achieve optimal performance in a MIMO time-varying generalization of the original control problem posed in [7]. It is also argued that it suffices to consider time-invariant policies in the time-invariant infinite horizon setting [4]. In [7], [8], and [4], SI at the encoder does not impact the rate-distortion tradeoff.
In this work, our system model differs slightly from that in \[8\] and is slightly less general than that in \[4\]. Our perspective is quite different. We motivate our rate distortion problem, and demonstrate an achievability result, in terms of digital communications. The achievability approaches in \[8\] and \[4\] are analog in the sense that the feedback channel is continuous. Our contributions are summarized as follows:

1) Assuming that the feedback channel from the encoder to decoder is binary and noiseless, we derive a lower bound on the minimum expected prefix-free codeword length under a constraint on control performance. The converse result motivates a rate distortion formulation.

2) Via the three-stage separation principle (cf. \[4\] (19)), we derive a semidefinite program equivalent to the rate distortion problem.\[^{1}\]

3) We provide a recipe to design both a sensor and a quantizer that nearly achieves the performance of the rate distortion formulation. Namely, we specify both a zero-delay quantizer design and source coding protocol.

A version of this paper with appendices is provided in \[9\].

**Notation:** We denote scalars by lower case letters, vectors by boldface lower-case letters, and matrices by boldface capitols. \(M, M^T\) denotes transpose. We use \(x_{1:t}\) to denote the sequence \((x_1, x_2, \ldots, x_t)\), and \(\{x_t\}\) for \(x_{1:10}\). We define the “time shifted” sequence \(x_{1:t-1}^1\) = \((0, x_1, \ldots, x_{t-1})\). If \(t < 1\), \(x_{1:t} = \emptyset\). Denote the set of finite length binary strings \(\{0, 1\}^*\).

Denote the entropy of a discrete random variable \(R\) \(\mathcal{H}\), differential entropy by \(h\), and mutual information (MI) by \(I\).

**II. System Model and Problem Formulation**

Fig. 1 illustrates our assumed system model. We assume a MIMO plant, a generally randomized sensor/encoder, and two feedback channels (one for SI and one for prefix-free codewords) from the encoder to a possibly randomized decoder/controller. Let \(x_{1:t}^1 \in \mathbb{R}^n\) and \(x_{2:t}^2 \in \mathbb{R}^m\). The state vector is defined as \(x_t = [(x_t^1)^T, (x_t^2)^T]^T\). Let \(A_{11} \in \mathbb{R}^{n \times n}\), \(A_{12} \in \mathbb{R}^{n \times m}\), \(A_{21} \in \mathbb{R}^{m \times n}\), and \(A_{22} \in \mathbb{R}^{m \times m}\) be block partitions of the system matrix \(A\), and define \(W_{11} \in \mathbb{R}^{n \times n}\), and \(W_{22} \in \mathbb{R}^{m \times m}\). The plant dynamics are given by

\[
\begin{bmatrix}
x_{1,t+1}^1 \\
x_{2,t+1}^2
\end{bmatrix} =
\begin{bmatrix}
A_{11} & A_{12} \\
A_{21} & A_{22}
\end{bmatrix}
\begin{bmatrix}
x_{1,t}^1 \\
x_{2,t}^2
\end{bmatrix} + Bu_t + w_t,
\]

(2a)

where \(w_t \sim \mathcal{N}(0, W)\) and \(W = \begin{bmatrix} W_{11} & 0 \\ 0 & W_{22} \end{bmatrix}\).\[^{2}\]

We assume \(W_{11}, W_{22} > 0\) and the \(w_t\) are IID. We assume \(B \in \mathbb{R}^{n+m \times u}\) and that \((A, B)\) is stabilizable. The sensor/encoder policy is a sequence of causally conditioned stochastic kernels denoted \(P(a_t | x_{1:t}^1, x_{1:t}^2)\) = \(\{P(a_t | x_{1:t}^1, a_{t-1})\}\) for any (possibly randomized) control and encoding/decoding policies, we have

\[
\inf \limsup_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}[\ell(a_t)]
\]

s.t. \(\limsup_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}[\|x_{t+1}^1\|_2 + \|u_t\|_2^2] \leq \gamma\) where \(Q > 0\), \(R > 0\). The expectations are taken with respect to the joint measure induced by the policies and plant dynamics.

**III. Converse**

The converse follows from \[2\] and \[1\].

**Theorem III.1** (A converse proof). Consider the model in Fig. 2 Let \(\ell(a_t)\) be the length of the codeword \(a_t\) in bits. For any (possibly randomized) control and encoding/decoding policies, we have

\[
\sum_{t=1}^{T} \mathbb{E}[\ell(a_t)] \geq I(x_{1:T} \rightarrow a_{1:T} | x_{1:T}^1, x_{1:T}^2).
\]

(4)

**Proof.** The model assumes that \(a_t\) is a codeword from a prefix-free code. Let \(A_t = \{a \in \{0, 1\}^* : P(a_t = a) > 0\}\). At every time \(t\), if \(a_1, a_2 \in A_t\), the prefix-free assumption guarantees that \(a_3\) is not a prefix of \(a_2\) and vice-versa. We claim that

\[
\mathbb{E}[\ell(a_t)] \geq H(a_t),
\]

(5)

this follows from a claim that for every \(t\), any function \(C_t : \{0, 1\}^* \rightarrow \{0, 1\}^*\) satisfying \(a = C_t(a)\) for all \(a \in A_t\) is a prefix-free code (in the terminology of \[10\] Ch. 5 ) from \(A_t\) to \(\{0, 1\}^*\). For any prefix-free code \(C_t^*\) (cf. \[10\] Theorem 5.3.1)

\[
\mathbb{E}[\ell(C_t^*(a_t))] \geq H(a_t).
\]

(6)

Since \(C_t\) is identity on \(A_t\), we have \(E[\ell(C_t(a_t))] = E[\ell(a_t)]\), and (5) follows. We discuss (5) in \[9\] Appendix A.
At every time $t$ we have the following chain of inequalities
\[ \mathbb{E}[f(a_t)] \geq H(a_t) \geq H(a_t|a_{t-1}, x_{1:t}) \geq H(a_t|a_{t-1}, x_{1:t}^2) - H(a_t|a_{t-1}, x_{1:t}), \]
(7)
(8)
(9)
Note that (7) is precisely (3) and (8) follows since conditioning reduces entropy, and (9) follows since discrete entropy is positive. The right hand side of (9) is $I(a_t; x_{1:t}|a_{t-1}, x_{1:t}^2)$. Summing over $t$, and applying (11) gives (4).

### IV. Rate Distortion Formulation
Given the converse in Sec. III, the arguments in (1) and (2) suggest attempting the following optimization
\[ \inf_{P(a_{1:n}|x_{1:n})} \limsup_{T \to \infty} I(x_{1:T} \to a_{1:T}|x_{1:T}^2) \]
(10)
\[ \text{s.t. } \limsup_{T \to \infty} \sum_{t=1}^{T} \mathbb{E}[\| x_{t+1} \|^2 | \mathcal{Q} + \| u_t \|^2 ] \leq \gamma, \]
where the infimum is over all possible encoder and decoder policies and all expectations are computed under the measure induced by the policies and the plant dynamics. Let $\{y_t\}$ denote a sequence of (not necessarily discrete) random variables. Define all sets of causally conditioned kernels $P(y_{1:n}|x_{1:n}) = \{P(y_t|x_{t}, y_{1:t-1})\}_{t=1}^{n}$,... and $P(u_{1:n}|y_{1:n}, x_{1:n}) = \{P(u_t|y_{t}, x_{t}, u_{t-1})\}_{t=1}^{n}$. The infimum in (10) is lower bounded by
\[ \inf_{P(y_{1:n}|x_{1:n})} \limsup_{T \to \infty} I(x_{1:T} \to y_{1:T}|x_{1:T}^2) \]
(11)
\[ \text{s.t. } \limsup_{T \to \infty} \sum_{t=1}^{T} \mathbb{E}[\| x_{t+1} \|^2 | \mathcal{Q} + \| u_t \|^2 ] \leq \gamma. \]
(10)
That (11) lower bounds (10) follows from expanding the domain of minimization. In (10) the optimization is restricted to kernels where $a_t$ is a discrete codeword, whereas in (11) we make no such assumption.

Note that (11) is an optimization over an infinite dimensional policy space and is not computationally amenable. Recently, [3] demonstrated that the minimum in (11) is achievable by a time invariant linear/Gaussian policy conforming to the three-stage separation architecture depicted in Fig. 2, namely, the optimal policy consists of a time-invariant linear/Gaussian sensor, a Kalman filter, and a certainty equivalence controller. Such a structural result allows us to convert (11) into an equivalent finite dimensional optimization. We discuss the optimal architecture in the following subsection.

### A. Three stage test channel (cf. [7], [4])

The feedback loop contains three components:

I. **Time-invariant linear/Gaussian sensor**: Let $C_1 \in \mathbb{R}^{n \times n}$ and $C_2 \in \mathbb{R}^{n \times m}$. The equation governing the sensor output, $y_t$, is assumed to be
\[ y_t = [C_1 \quad C_2] x_t + v_t, \quad \text{where } v_t \sim \mathcal{N}(0, V). \]
(12)

II. **Kalman filter**: The standard Kalman filter (KF) computes the linear minimum mean squared error (LMMSE) estimator, which in the joint Gaussian case is also the MMSE estimator. The estimator is computed by the standard recursion (cf. [1]). The KF computes the estimate $\hat{x}_t$ via a linear (in all arguments), time varying, $C$ and $V$ dependent recursion denoted $\hat{x}_t = \Psi_t(\hat{x}_{t-1}, y_{t}, x_{t}^2, u_t)$.

III. **Certainty equivalence control**: We assume certainty equivalence linear feedback control. Let $S$ be a stabilizing solution to the algebraic Riccati equation [2]
\[ S = A^T SA - A^T SB (B^T SB + R)^{-1} B^T SA + Q. \]
(13)
The feedback control gain $K$ is then given by
\[ K = -(B^T SB + R)^{-1} B^T SA. \]
(14)
Under the three-stage test channel assumption, the design variables are limited to $C$ and $V \geq 0$, converting (11) into a finite-dimensional optimization.

### V. A Convex Programming Approach to the Rate/Control Performance Tradeoff

Via three-stage separation, the minimum in (11) is given by
\[ \inf_{C, V} \limsup_{T \to \infty} \frac{1}{T} I(x_{1:T} \to y_{1:T}|x_{1:T}^2) \]
(15a)
\[ \text{s.t. } \forall t \limsup_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}[\| x_{t+1} \|^2 | \mathcal{Q} + \| u_t \|^2 ] \leq \gamma, \]
(15b)
\[ x_{t+1} = A x_t + B K u_t + w_t, \]
\[ y_t = C x_t + v_t, \quad v_t \sim \mathcal{N}(0, V), \quad V \geq 0, \]
\[ \hat{x}_t = \Psi_t(\hat{x}_{t-1}, y_{t}, x_{t}^2, u_{t-1}), \quad u_t = K \hat{x}_t, \]
where we identify the DI (15a) as the communication cost and the quadratic (15b) as the control cost. All expectations are under the measure induced by $C, V$, and Fig. 2. In this section we derive a convex program from (15). We first simplify the cost (15a) under the assumed architecture, deriving an expression in terms of Kalman filter error covariance matrices.

### A. The rate and control costs in terms of KF variables

Under the architecture in Fig. 2 it can be verified that $x_{1:t-1} \leftrightarrow x_{1:t}^2, x_1^2, y_{1:t-1} \leftrightarrow y_t$. Thus, by the chain rule
\[ I(x_{1:t}; y_t|x_{1:t}^2, y_{1:t-1}) = I(x_t; y_t|x_{t}^2, y_{1:t-1}), \]
and the communication cost (15a) is given by
\[ I(x_{1:T} \to y_{1:T}|x_{1:T}^2) = \sum_{t=1}^{T} I(x_t; y_t|x_{t}^2, y_{1:t-1}). \]
Let \( \tilde{x}_t = \mathbb{E}[x_t | x_{t-1}^2, y_{1:t-1}] \) and \( \tilde{x}_t = \mathbb{E}[x_t | x_{t-1}, y_{1:t-1}] \). Denote the residuals \( \tilde{r}_t = x_t - \tilde{x}_t \) and \( \hat{r}_t = x_t - \tilde{x}_t \). Since \( \tilde{x}_t \) and \( \hat{x}_t \) are measurable functions of \( x_{t-1}^2, y_{1:t-1} \) and \( x_{t-1}, y_{1:t-1} \) respectively, by the definition of MI

\[
I(x_t^2; y_t | x_{t-1}^2, y_{1:t-1}) = h(\tilde{r}_t | x_{t-1}^2, y_{1:t-1}) - h(\hat{r}_t | x_{t-1}^2, y_{1:t-1}).
\]

By the joint Gaussianity of \( x_{t-1} \) and \( y_{1:t} \) and the orthogonality principle, \( \tilde{r}_t \) is Gaussian, has \( \mathbb{E}[\tilde{r}_t] = 0 \), and is independent of \( x_{t-1}^2, y_{1:t-1} \). Likewise \( \hat{r}_t \) is Gaussian, has \( \mathbb{E}[\hat{r}_t] = 0 \), and is independent of \( x_{t-1}^2, y_{1:t} \). Define \( \hat{P}_t = \mathbb{E}[\hat{r}_t \hat{r}_t^T] \) and \( \tilde{P}_t = \mathbb{E}[\tilde{r}_t \tilde{r}_t^T] \). The differential entropy of \( z \sim N(0, \Sigma) \) is \( h(z) = \frac{1}{2} \log \det(\Sigma) + d \log(2\pi e) \). Thus (17) is

\[
I(x_t^2; y_t | x_{t-1}^2, y_{1:t-1}) = \frac{1}{2} \log \det(\tilde{P}_t) - \log \det(\hat{P}_t).
\]

Thus, the rate cost function in (15) may be written

\[
\limsup_{T \to \infty} \frac{1}{T} I(x_{1:T}^1; y_{1:T} | x_{1:T}^{12}) = \limsup_{T \to \infty} \frac{1}{2T} \sum_{t=1}^{T} \log \det(\tilde{P}_t) - \log \det(\hat{P}_t).
\]

Under the present assumptions (cf. [2]), the control cost may also be written in terms of \( \tilde{P}_t \).

\[
\limsup_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}[|u_t|^2 | \tilde{x}_t] = \limsup_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \text{Tr}(\Theta \hat{P}_t) + \text{Tr}(SW).
\]

The matrix inversion lemma demonstrates that \( C_2 \) is a Riccati difference equation [12]. Given an initial condition, the recursion (25) converges under a variety of circumstances [12] [11]. If it exists, the steady state equation \( \tilde{P}_\infty \) solves the discrete algebraic Riccati equation

\[
\tilde{P}_\infty = W_{11} + A_{11} (\tilde{P}_\infty + F)^{-1} A_{11}^T.
\]

Denote the \textit{a priori} state estimate for time \( t \) as \( \hat{x}_{t-1} \). Given the linear feedback control, \( \hat{x}_{t-1} \) is a linear function of \( \hat{x}_{t-1} \) and is precisely the MMSE estimator \( \hat{x}_{t-1} = \mathbb{E}[x_{t-1} | x_{t-1}^2, y_{1:t-1}] \). Denote the \textit{a posteriori} residual process \( \tilde{r}_{t-1} = x_{t-1} - \hat{x}_{t-1} \). Note that in contrast to the definition of \( \hat{r}_{t-1} \), \( \tilde{r}_{t-1} \) contains residuals from estimating (predicting) both \( x_1^1 \) and \( x_2^1 \). It can be shown that \( \mathbb{E}[\tilde{r}_{t-1}] = 0 \). Denote the covariance matrix \( \tilde{P}_{t-1} = \mathbb{E}[\tilde{r}_{t-1} \tilde{r}_{t-1}^T] \). Let \( \tilde{A} = [A_{11}^T, A_{21}^T]^T \).

By direct substitution \( \tilde{P}_{t-1} = \tilde{A}P_t \tilde{A}^T + W_t \), where \( \tilde{P}_{t-1} \) is covariance of \( \tilde{r}_{t-1} \) defined in V-A.

The estimator after the SI update (the noiseless observation of \( x_2^1 \)) at time \( t \) is given by \( \tilde{x}_t = [\tilde{x}_t^1, (\tilde{x}_t^2)^T]^T \) (cf. Sec. V).

Again, \( \tilde{x}_t = \mathbb{E}[x_t | x_{t-1}^2, y_{1:t-1}] \) and is a linear function of \( x_{t-1}^2, y_{1:t-1} \). The residual, \( \hat{r}_t \), is again defined with respect to the error estimating \( x_1^1 \) only (as in Sec. V).

Let \( P_{t-1}^{11} \in \mathbb{R}^{n \times m}, P_{t-1}^{12} \in \mathbb{R}^{m \times m} \), \( P_{t-1}^{21} \in \mathbb{R}^{m \times n}, P_{t-1}^{22} \in \mathbb{R}^{m \times m} \) be such that

\[
P_{t-1} = P_{t-1}^{11} + P_{t-1}^{12}(P_{t-1}^{22} + P_{t-1}^{21})^{-1} P_{t-1}^{21}.
\]

Finally, the sensor measurement update computes the posterior state estimate at time \( t \). It can be shown that \( \hat{P}_t \) is given by

\[
\hat{P}_t = \hat{P}_t^{11} + C_1^T V_1^{-1} C_1,
\]

which demonstrates that \( C_2 \) is completely arbitrary.

Let \( F = A_{21}^T W_{22}^{-1} A_{21} \). Using (22), the matrix inversion lemma gives

\[
\hat{P}_{t+1} = \hat{P}_t + C_1^T V_1^{-1} C_1 + F - A_{21}^T W_{22}^{-1} A_{21}.
\]

Substituting (23) into (24) gives a recursion for \( \hat{P}_t \)

\[
\hat{P}_{t+1} = W_{11} + A_{11} (\hat{P}_{t+1} + C_1^T V_1^{-1} C_1 + F)^{-1} A_{11}^T.
\]

The matrix inversion lemma demonstrates that (25) is a Riccati difference equation [12]. Given an initial condition, the recursion (25) converges under a variety of circumstances [12] [11]. If it exists, the steady state equation \( \hat{P}_\infty \) solves the discrete algebraic Riccati equation

\[
\hat{P}_\infty = W_{11} + A_{11} (\hat{P}_\infty + C_1^T V_1^{-1} C_1 + F)^{-1} A_{11}^T.
\]

In particular, [12] Theorem 4.1 establishes convergence to a unique, positive definite solution when \( (A_{11} W_{22}^{-1}) \) is stabilizable and \( ([C_1^T, A_{21}^T] , A_{11}) \) is detectable [12]. The stabilizability is immediate as \( W_{11} > 0 \). Furthermore, in the present setting, the existence of a positive definite solution to (26) can be shown to imply that \( ([C_1^T, A_{21}^T] , A) \) is detectable via a discrete time Liapunov equation. We restrict our attention to the case that \( ([C_1^T, A_{21}^T] , A) \) is detectable.

Convergence of \( \{P_t\} \) implies that \( \{\tilde{P}_t\} \) also converges. The limits of \( \{P_t\} \) and \( \{\tilde{P}_t\} \) must satisfy both

\[
\tilde{P}_\infty = P_\infty + C_1^T V_1^{-1} C_1,
\]

(27a)

\[
\text{and (27b)}
\]
\[
\tilde{P}_\infty = W_{11} + A_{11} (\tilde{P}_\infty + F)^{-1} A_{11}^T.
\] (27b)

Given \( C \) and \( V \), if such a \( \tilde{P}_\infty > 0 \) and \( \tilde{P}_\infty > 0 \) can be found, the resulting \( \tilde{P}_\infty \) will satisfy (26). If for some \( C \) and \( V \) there exists \( \tilde{P}_\infty \) (necessarily positive definite) satisfying (27), it follows that [(\( C_1^T, A_2^T \), \( A \)] is detectable and that both \( P_1 \to \tilde{P}_\infty \) and \( P_\infty \to \tilde{P}_\infty \). Using [19], a standard Cesáro mean (cf. [10]) argument gives that

\[
\limsup_{T \to \infty} \frac{1}{T} I(x_{1:T} : y_{1:T}) = 1 - \frac{1}{2} \log \det \tilde{P}_\infty \quad \text{det} \tilde{P}_\infty. \tag{28}
\]

Similarly, using (29) we have that

\[
\limsup_{T \to \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E} \left[ x_{t+1}^2_q + \left[ u_t^2 \right]_R \right] = \text{Tr}(\Theta \tilde{P}_\infty + SW). \tag{29}
\]

In the following subsection, we use these results to derive a convex program for the rate distortion problem (15).

C. Derivation of the convex program

Define \( \hat{P} \triangleq \tilde{P} \) and \( \hat{P} \triangleq \tilde{P} \). Substituting (27), (28), and (29) into (15) yields the finite dimensional optimization

\[
\min_{\hat{P}, V} \left[ \frac{1}{2} \log \det \hat{P} - \log \det \hat{P} \right]
\]

s.t. \( V \geq 0, \hat{P} \geq 0, \text{Tr}(\Theta \hat{P} + SW) \leq \gamma, \quad \hat{P}^{-1} - \hat{P}^{-1} + C_1^T V^{-1} C_1, \quad \hat{P} = W_{11} + A_{11} (\hat{P}^{-1} + F)^{-1} A_{11}^T. \) (30)

Since \( W_{11} > 0, \hat{P} > 0 \). The minimum in (30) can be found by the convex optimization

\[
\min_{\hat{P}, \Pi} \left[ \frac{1}{2} \log \det W - \log \det \Pi - \log \det (W_{22} + A_{21} \hat{P} A_{21}^T) \right]
\]

s.t. \( \hat{P} > 0, \Pi \geq 0, \text{Tr}(\Theta \hat{P} + SW) \leq \gamma, \quad W + \hat{P} A \hat{P}^T - \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \hat{P} - \hat{P} A \hat{P}^T - \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \geq 0. \) (31)

Details are given in [9] Appendix B. Let \( \hat{P}_{\text{min}} \) be the minimizer in (31), and let \( \hat{P}_{\text{min}} \) be given by (27b). The minimizers \( C_1 \) and \( V \) are the set of matrices satisfying \( \hat{P}^{-1} - \hat{P}^{-1} = C_1 V^{-1} C_1 \). Without loss of generality, we choose \( V = I, C_1 \) the corresponding minimizer, and \( C_2 = 0 \). We now show that the minimum is nearly achievable in the architecture of Fig. 4.

VI. QUANTIZATION AND PREFIX FREE CODING

The architecture used to demonstrate the achievability result follows from [2] IV, and is shown in Fig. 4. As in [2], we use a predictive elementwise uniform quantizer with subtractive dither. We define an elementwise uniform quantizer with sensitivity \( \Delta \) as a function \( q_\Delta : \mathbb{R}^n \to \mathbb{R}^n \) such that

\[
[q_\Delta(z)]_i = m \Delta \text{ if } [z]_i \in [m \Delta - \Delta^2, m \Delta + \Delta^2], \quad \text{such that}
\]

where \( n_t \) is a zero mean, uniform random vector with IID elements and \( \mathbb{E}[n_t n_t^T] = I \). The decoder side Kalman filter operates analogously to the two stage filter in Fig. 3. Having received the previous measurements \( y_{t-1}^{NG} \) and the SI \( \hat{x}_{1:t}^{NG} \), the decoder can compute \( \hat{x}_{t}^{NG} \) and form a centered measurement \( \tilde{y}_{t}^{NG} = y_{t}^{NG} + C_1 \hat{x}_{1:t}^{NG} \). It clear that,

\[
\hat{x}_{t}^{NG} = \text{The LMMSE estimate of } x_t \text{ given } y_{1:t}^{NG}, \text{ } x_{1:t}^{NG}, \text{ (34)}
\]

Dithering allows the quantization error to manifest as additive uniform noise; it can be shown that \( n = z - q_\Delta(z) \) is independent of \( z \) and that the elements \( n_t \) are IID with \( [n_t] \sim \text{Uniform}[-\frac{\Delta}{2}, \frac{\Delta}{2}] \) [2]. Lemma 1a) [13]. The caption of Fig. 4 outlines the use of dithering this achieveability result.

We now show that when \( \Delta = 2\sqrt{3} \) the system in Fig. 4 achieves an equivalent control performance as the as the architecture in Fig. 2 for equivalent \( C_1 \) and \( V = I \). In Fig. 4, at time \( t \) the decoder observes a dithered quantized measurement of \( x_t \), denoted \( y_{t}^{NG} \) and to be described presently. The measurement is predictive and defined recursively via an encoder KF process. At time \( t \), a KF at the encoder computes

\[
\tilde{x}_{t}^{1,NG} = \text{The LMMSE estimate of } x_{1} \text{ given } y_{1:t}^{NG}, \text{ } x_{1:t}^{NG}.
\]

The encoder’s quantizer computes the discrete \( n_t = q_\Delta(C_1(x_{t-1} - \tilde{x}_{t-1},NG) + d_t) \), and encodes \( \tilde{z}_t \) with a prefix-free lossless Shannon-Fano-Elias (SFE) code. The codeword is sent to the decoder, which (exactly) reconstructs \( \hat{z}_{t}^{NG} \), the decoder signal, the decoder decoder forms \( y_{t}^{NG} = \hat{z}_t - d_t \), or, equivalently \( y_{t}^{NG} = q_\Delta(C_1(x_{t} - \tilde{x}_{t,NG})) \). This gives

\[
y_{t}^{NG} = C_1 x_{t}^{NG} - C_1 \tilde{x}_{t,NG} + n_t
\]

where \( n_t \) is a zero mean, uniform random vector with IID elements and \( \mathbb{E}[n_t n_t^T] = I \). The decoder side Kalman filter operates analogously to the two stage filter in Fig. 3. Having received the previous measurements \( y_{t-1}^{NG} \) and the SI \( \hat{x}_{1:t}^{NG} \), the decoder can compute \( \hat{x}_{t}^{NG} \) and form a centered measurement \( \tilde{y}_{t}^{NG} = y_{t}^{NG} + C_1 \hat{x}_{1:t}^{NG} \). It clear that,

\[
\hat{x}_{t}^{NG} = \text{The LMMSE estimate of } x_t \text{ given } y_{1:t}^{NG}, \text{ } x_{1:t}^{NG}, \text{ (34)}
\]
is the same as the LMMSE of $x_t^1$ given $\tilde{y}_{t:k}^{NG}$ and $x_{t:t}^2$. The controller forms the control input $u_t = K[ (\hat{x}_t^{1,NG})^T, (\hat{x}_t^2)^T ]^T$ where $K$ is as in [14]. A corollary to the proof of [2] Lemma 1a) demonstrates that under this (any) feedback arrangement, the sequence of quantization noises $\{n_t\}$ is temporally white, e.g. $E[n_t n_{t'}^T] = 0$ if $t \neq t'$.

This leads to a result analogous to [2] Lemma 2). Having fixed $C_1$ and $V = I$, denote the jointly Gaussian random variables $(x_t^1, \hat{x}_t^1)$ with the joint distribution induced by the architecture in Fig. 2 by $(x_t^{1,NG}, \hat{x}_t^{1,NG})$. Likewise, denote the (generally non-Gaussian) RVs $(x_t^2, \hat{x}_t^2)$ with the joint distribution induced by the architecture in Fig. 4 by $(x_t^{2,NG}, \hat{x}_t^{2,NG})$. We have the following lemma.

**Lemma VI.1.** If RVs describing the initial conditions $x_0^{NG}$ and $x_t^0$ have identical first and second moments, then the processes $\{(x_t^{1,NG}, \hat{x}_t^{1,NG}; x_t^{2,NG}, \hat{x}_t^{2,NG})\}$ are equivalent up to second moments. Regardless of initial conditions, $E[(x_t^{1,NG} - \hat{x}_t^{1,NG}) (x_t^{2,NG} - \hat{x}_t^{2,NG})^T] \to \hat{P}$.

This result follows from computing the measurement model $\tilde{y}_t^{NG} = C_1 x_t^{1,NG} + n_t$ (35) to the linear/Gaussian model [12] under the assumed choices of $C_1$ and $V = I$. While the additive white noise is uniform, rather than Gaussian, it has $E[n_t] = 0$ and $E[n_t n_{t'}^T] = I$. The first statement follows from an induction on $t$. The latter follows as the Riccati recursion relating the covariance matrices of the error processes $x_t^{1,NG} - \hat{x}_t^{1,NG}$ and $x_t^{2,NG} - \hat{x}_t^{2,NG}$ is identical to that derived in [4, Appendix C]. The same control cost is achieved in both systems (cf. [29]).

It remains to bound the codeword length. Recall the discrete RV $\tilde{z}_t$, and define $z_t = C_1(x_t^{1,NG} - \hat{x}_t^{1,NG})$. At every time $t$, by the SFE construction (cf. [10]) there exists a lossless, prefix-free code that encodes $\tilde{z}_t$ with an expected length $E[\ell(a_t)] \leq H(\tilde{z}_t|d_t) + 1$. Consider the joint Gaussian case and define $\tilde{F}_t^G$ as in Sec. [7]. The next lemma is proved in [9, Appendix C].

**Lemma VI.2.** At every time $t$, we have

$$H(\tilde{z}_t|d_t) \leq \frac{n_t}{2} \log_2 \frac{4\pi e}{12} + I(C_1 x_t^1; C_1 \tilde{F}_t^G + v_t).$$  
(36)

Let $k = 1 + \frac{n_t}{2} \log_2 \frac{4\pi e}{12}$. Our main result is the following.

**Theorem VI.3.** When the entropy encoder and decoder in Fig. 2 use SFE coding adapted to the PMF of $\tilde{z}_t$ for all $t$, the architecture achieves

$$\limsup_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} E[\ell(a_t)] \leq k + \frac{\log \det \hat{P}_{\min} - \log \det \hat{P}_{\min}^2}{2},$$  
(37)

which is convenient for computing the bound via (31).

**Proof.** At every time $t$, SFE codeword has a length $E[\ell(a_t)] \leq 1 + H(\tilde{z}_t|d_t)$. Since $\tilde{x}_t^G$ is a measurable function of $x_{1:t}^G, y_{1:t-1}^G$ we have that

$$I(x_t^1; y_t^G | x_{1:t}^G, y_{1:t-1}^G) = I(\tilde{F}_t^G; C_1 \tilde{F}_t^G + v_t | x_{1:t}^G, y_{1:t-1}^G) + I(x_t^1; y_t^G | \tilde{F}_t^G, C_1 \tilde{F}_t^G + v_t).$$  
(38)

Since $x_t^G$ and $v_t$ are independent of $(x_{1:t}^G, y_{1:t-1}^G)$, this implies

$$I(x_t^1; y_t^G | x_{1:t}^G, y_{1:t-1}^G) = I(\tilde{F}_t^G; C_1 \tilde{F}_t^G + v_t).$$  
(39)

Note that both $\tilde{F}_t^G \leftrightarrow C_1 \tilde{F}_t^G \leftrightarrow C_1 \tilde{F}_t^G + v_t$ and $C_1 \tilde{F}_t^G \leftrightarrow \tilde{F}_t^G \leftrightarrow C_1 \tilde{F}_t^G + v_t$ are Markov chains. Applying the standard data processing inequality (twice) to $I(\tilde{F}_t^G; C_1 \tilde{F}_t^G + v_t)$ using both of these chains allows us to conclude that

$$I(x_t^1; y_t^G | x_{1:t}^G, y_{1:t-1}^G) = I(C_1 \tilde{F}_t^G; C_1 \tilde{F}_t^G + v_t).$$  
(40)

Thus, by Lemma VI.2

$$H(\tilde{z}_t|d_t) \leq \frac{n_t}{2} \log_2 \frac{4\pi e}{12} + I(x_t^1; y_t^G | x_{1:t}^G, y_{1:t-1}^G).$$  
(41)

Summing (36) over $t$, and applying (1) gives (37).
Fig. 5. A generalized version of the path from encoder to decoder in the model of Fig. 1. The minimum achievable rate for a system with the additional “virtual” encoder/decoder pair lower bounds the minimum achievable rate for a system without the additional virtual pair.

**APPENDIX A**

**A FORMAL PROOF OF THE CONVERSE BOUND**

Consider the system model in Fig. 1 at time $t$. By assumption (cf. Sec. II), $a_t$ is a prefix-free codeword, so if $a_1 \neq a_2$, $P(a_t = a_1) > 0$ and $P(a_t = a_2) > 0$ then $a_1$ is not a prefix of $a_2$ and vice-versa. The codeword $a_t$ is a discrete variable with countable support chosen by the policy defined by kernel (a conditional PMF)

$$P_S(a_t|x_{1:t}, x_{1:t}^2, u_{1:t-1}),$$

(42)

where we added the subscript $S$ to denote the “encoder/sensor” policy as in Fig 1. The control action is chosen by the policy defined via the probability measure

$$P_C(u_t|a_{1:t}, x_{1:t}^2, u_{1:t-1}),$$

(43)

where we added the subscript $C$ to denote the “decoder/controller” policy as in Fig 1.

We bound the expected length of the prefix-free codeword $a_t$ at every $t$ by bounding the length of lossless prefix-free source code that encodes $a_t$ itself. Consider a modified version of the system model (cf. Fig. 1 shown in Fig. 5). Another, “virtual” encoder/decoder pair has been added between the encoder/sensor and decoder/controller. We assume that at every time $t$, the virtual encoder encodes $a_t$ into a prefix-free codeword $c_t$. We refer to $a_t$ as a “source codeword” and $c_t$ as a “virtual codeword”. At every timestep $t$, the virtual encoder encodes $a_t$ into the virtual codeword $c_t$ via computing

$$c_t = C_t(a_t),$$

(44)

for some deterministic measurable function $C_t$. Likewise, the virtual encoder computes the reconstruction $\hat{a}_t$ by computing

$$\hat{a}_t = D_t(c_t),$$

(45)

where again, for all $t$, $D_t$ is a deterministic measurable function. The virtual encoder and decoder are both memoryless and do not access any SI. We insist that the virtual encoder and decoder are lossless, namely that $a_t = \hat{a}_t$. We think of Fig. 5 as “inserting” the virtual encoder and decoder blocks into Fig. 1.

Note that virtual encoder and decoder policies do not effect the measure induced on the random variables $x_{1:t}, a_{1:t}, x_{1:t}, u_{1:t}$ due to the assumption that $C_t$ and $D_t$ are deterministic and that $\hat{a}_t = a_t$.

The idea is that the insertion of an optimal “virtual” lossless encoder/decoder between the sensor and controller produces a codeword $c_t$ that has a length less than or equal to that of $a_t$. More formally, at every time $t$ we lower bound the length of the codeword $a_t$ by lower bounding the length of the codeword $c_t$ under the optimal zero-error prefix free virtual encoder and virtual decoder policies. If $r$ is a prefix free binary codeword, let $\ell(r)$ denote its length. An “optimal” virtual encoder and decoder policy (there may be more than one), for some fixed sensor and controller policy, is defined as a sequence of deterministic functions $P_t = \{C_t^*, D_t^*\}$ where:

1. $C_t^*$ maps $a_t$ to $c_t$, $D_t^*$ maps $c_t$ to $a'_t$ as in Fig. 6.
2. There is no probability of error, e.g. $D_t^*(C_t^*(a_t)) = a_t$ for all $t$.
3. Let $A_t = \{a \in \{0, 1\}^* : P(a_t = a) > 0\}$. At every time $t$, if $a_1, a_2 \in A_t$ and $a_1 \neq a_2$ then $c_1 = C_t^*(a_1)$ is not a prefix of $c_2 = C_t^*(a_2)$ and vice-versa.
4. At all $t$, $E[\ell(c_t)]$ is minimized among all other policies satisfying (1), (2), and (3) above.

The above expectations and probabilities are taken with respect to the measure induced by the sensor and controller policies $P_S$ and $P_C$ and Fig. 1. In the language of [10, Chapter 5], the constraints (1), (2), and (3) require that at any time $t$, $C_t^*$ is prefix-free code mapping the space $A_t \subset \{0, 1\}^*$ to the space of binary prefix-free codewords in $\{0, 1\}^*$. In the next lemma, we show that under the optimal virtual encoder and decoder policies, we have

$$E[\ell(C_t^*(a_t))] \leq E[\ell(a_t)],$$

(46a)

or, in other words

$$E[\ell(c_t)] \leq E[\ell(a_t)],$$

(46b)

where $c_t = C_t^*(a_t)$.

**Lemma A.1.** For all $t$ there exists a $C_t$ and $D_t$ satisfying (1)-(4) above and $E[\ell(c_t)] \leq E[\ell(a_t)]$.

**Proof.** At every time $t$, the source codewords $a_t$ are codewords of a prefix-free code. Setting $C_t$ and $D_t$ equal to identity, e.g.

$$a_t = C_t(a_t) \quad \text{and} \quad c_t = C_t(c_t),$$

(47)

gives

$$c_t = a_t \quad \text{and} \quad \hat{a}_t = c_t$$

(48)

Under this policy, the virtual encoder sends the input $a_t$ directly and $a_t = c_t = \hat{a}_t$. Thus there exist policies satisfying the constraints that achieve equality in (46) and the result follows. $\square$

Since for every $t$, $C_t^*$ is a prefix-free code from $A_t \to \{0, 1\}^*$, it follows from [10, Theorem 5.3.1] that

$$H(a_t) \leq E[\ell(C_t^*(a_t))]$$

(49)

which gives the result

$$H(a_t) \leq E[\ell(a_t)],$$

(50)

We emphasize that (50) holds for every time $t$.

2This ensures that $C_t^*$, restricted to $A_t$, is injective; a necessary condition for there to exist a deterministic $D_t^*$ such that $D_t^*(C_t^*(a_t)) = a'_t$ (cf. condition (2)). It also ensures that the set of virtual codewords transmitted with nonzero probability at time $t$ are not prefixes of one another; define $C_t^* = \{c \in \{0, 1\}^* : P(C_t(a_t) = c) > 0\}$, and let $c_1, c_2 \in C_t$. If $c_1 \neq c_2$, then $c_1$ is not a prefix of $c_2$ and vice-versa.
APPENDIX B
Proof of Equivalence between (30) and (31)

We begin by writing (30) in terms of \( \hat{\mathbf{P}} \) only. It can immediately be seen that the design variables \( \mathbf{C} \) and \( \mathbf{V} \) are essentially slack. The constraint \( \hat{\mathbf{P}}^{-1} = \hat{\mathbf{P}}^{-1} + \mathbf{C}_1 \mathbf{V}^{-1} \mathbf{C}_1 \) can be replaced with the constraints \( \hat{\mathbf{P}} \succeq \mathbf{0} \) and \( \hat{\mathbf{P}} > \mathbf{0} \). The new inequality constraint may be readily combined with the equality constraint for \( \hat{\mathbf{P}} \) (cf. (27b)) to derive a linear matrix inequality (LMI) in \( \hat{\mathbf{P}} \). Applying the matrix inversion lemma to (27b) gives
\[
\hat{\mathbf{P}} - \hat{\mathbf{P}} = \mathbf{W}_{11} - \hat{\mathbf{P}} + \mathbf{A}_{11} (\hat{\mathbf{P}} - \hat{\mathbf{P}}A_{21}^T(A_{21}\hat{\mathbf{P}}A_{21}^T + \mathbf{W}_{22})^{-1}A_{21}\hat{\mathbf{P}})A_{11}^T. \tag{51}
\]
The right hand side of (51) is a Shur complement, and the LMI constraint follows directly. Thus \( \hat{\mathbf{P}} - \hat{\mathbf{P}} \succeq \mathbf{0} \) is equivalent to the LMI
\[
\mathbf{W} + \hat{\mathbf{A}}\hat{\mathbf{P}}\hat{\mathbf{A}}^T - \begin{bmatrix} \hat{\mathbf{P}} & 0 \\ 0 & 0 \end{bmatrix} \succeq \mathbf{0}. \tag{52}
\]
The corresponding \( \mathbf{C}_1 \) and \( \mathbf{V} \) are not unique, and can be found by factorizing \( \hat{\mathbf{P}}^{-1} - \hat{\mathbf{P}}^{-1} \).

It remains to simplify the rate cost. Using (27b) and invoking the matrix determinant lemma twice, we
\[
\log \det \hat{\mathbf{P}} - \log \det \hat{\mathbf{P}} = \log \det (\hat{\mathbf{P}}^{-1} + \hat{\mathbf{A}}^T \mathbf{W}^{-1} \hat{\mathbf{A}}) + \log \det \mathbf{W} - \log \det (\mathbf{W}_{22} + \mathbf{A}_{21}\hat{\mathbf{P}}A_{21}^T). \tag{53}
\]
Introduce the slack variable \( \Pi \). We have
\[
\log \det (\hat{\mathbf{P}}^{-1} + \hat{\mathbf{A}}^T \mathbf{W}^{-1} \hat{\mathbf{A}}) = 
\min_{\Pi \succeq (\hat{\mathbf{P}}^{-1} + \hat{\mathbf{A}}^T \mathbf{W}^{-1} \hat{\mathbf{A}})^{-1}} - \log \det \Pi. \tag{54}
\]
Applying the matrix inversion lemma and the Shur complement formula to the constraint \( \Pi \succeq (\hat{\mathbf{P}}^{-1} + \hat{\mathbf{A}}^T \mathbf{W}^{-1} \hat{\mathbf{A}})^{-1} \) gives the equivalent LMI
\[
\begin{bmatrix} \hat{\mathbf{P}} - \Pi & \hat{\mathbf{P}}A^T \\ \hat{\mathbf{A}}\hat{\mathbf{P}} & \mathbf{W} + \hat{\mathbf{A}}\hat{\mathbf{P}}A^T \end{bmatrix} \succeq \mathbf{0}. \tag{55}
\]
The preceding discussion demonstrates that
\[
\min_{\hat{\mathbf{P}}, \Pi} \frac{\log \det \mathbf{W} - \log \det \Pi - \log \det (\mathbf{W}_{22} + \mathbf{A}_{21}\hat{\mathbf{P}}A_{21}^T)}{2} \\
\text{s.t. } \hat{\mathbf{P}} > \mathbf{0}, \Pi \succeq \mathbf{0}, \tr{(\Theta\hat{\mathbf{P}})} + \tr{(\mathbf{S}\mathbf{W})} \leq \gamma, \\
\mathbf{W} + \hat{\mathbf{A}}\hat{\mathbf{P}}A^T - \begin{bmatrix} \hat{\mathbf{P}} & 0 \\ 0 & 0 \end{bmatrix} \succeq \mathbf{0}, \\
\begin{bmatrix} \hat{\mathbf{P}} - \Pi & \hat{\mathbf{P}}A^T \\ \hat{\mathbf{A}}\hat{\mathbf{P}} & \mathbf{W} + \hat{\mathbf{A}}\hat{\mathbf{P}}A^T \end{bmatrix} \succeq \mathbf{0}. \tag{56}
\]
achieves the same minimum as (30). This program is the minimization of a convex objective with convex constraints.

APPENDIX C
Proof of Lemma VI.2

The proof follows closely from [2]. Assume the definitions of Sec. VI. It turns out that \( H(\tilde{z}_i|d_i) \) admits a bound in terms of the squared error rate distortion function of \( z \) [2, Lemma 1]. Let \( D = n\Delta^2/12 = n \). Define the rate distortion function
\[
R_x(D) = \inf_{\mathbf{P}(\mathbf{u}|z):E[|z-u|^2] \leq D} I(\mathbf{x};\mathbf{u}). \tag{57}
\]
We have
\[
H(\tilde{z}_i|d_i) \leq \frac{n}{2} \log_2 \frac{4\pi e}{12} + R_{x_i}(D). \tag{58}
\]
It is known (cf. [10, Problem 10.8]) that if a Gaussian random vector \( x \) has \( \text{cov}(x) = \text{cov}(z) \) then \( R_x(D) \leq R_x(D) \).

We claim that
\[
R_{x_i}(D) \leq I(C_1\tilde{r}_i^G;C_1\tilde{r}_i^G + v_i) \tag{59}
\]
Note that \( \Theta_i = C_1\tilde{r}_i^G \) is Gaussian and that, by Lemma VI.1 we have \( \mathbb{E}[(\Theta_i\Theta_i^T)] = \mathbb{E}[\tilde{z}_i\tilde{z}_i^T] = C_1\tilde{P}_iC_1^T \) and that \( \mathbb{E}[(\Theta_i)] = \mathbb{E}[\tilde{z}_i] = 0 \). Thus \( R_{x_i}(D) \leq R_{\Theta_i}(D) \). By the assumption that \( \mathbf{V} = \mathbf{I} \), we have \( \mathbb{E}[v_i^Tv_i] = n \). Since \( D = n \), \( R_{\Theta_i}(D) \leq I(C_1\tilde{r}_i^G;C_1\tilde{r}_i^G + v_i) \), and (59) follows. Substituting this into (58) establishes the Lemma.