Finite-Length Linear Schemes for Joint Source-Channel Coding over Gaussian Broadcast Channels with Feedback

Yonathan Murin\textsuperscript{1}, Yonatan Kaspi\textsuperscript{2}, Ron Dabora\textsuperscript{1}, and Deniz Gündüz\textsuperscript{3}

\textsuperscript{1}Department of Electrical and Computer Engineering, Ben-Gurion University, Israel
\textsuperscript{2}Department of Electrical and Computer Engineering, UCSD, San Diego, USA
\textsuperscript{3}Department of Electrical and Electronic Engineering, Imperial College London, UK

Abstract

We study the uncoded transmission of a pair of correlated Gaussian sources over a two-user Gaussian broadcast channel with unit-delay noiseless feedback, abbreviated as the GBCF. Each source pair sample is transmitted using a linear transmission scheme in a finite number of channel uses. We investigate three transmission schemes: A scheme based on the Ozarow-Leung (OL) code, a scheme based on the linear quadratic Gaussian (LQG) code of Ardestanizadeh et al., and a novel scheme derived in this work using a dynamic programming (DP) approach. For the OL and LQG schemes we present lower and upper bounds on the minimal number of channel uses needed to achieve a target mean-square error (MSE) pair. For the LQG scheme in the symmetric setting, we identify the optimal scaling of the sources, which results in a significant improvement of its finite horizon performance, and, in addition, characterize the (exact) minimal number of channel uses required to achieve a target MSE. Finally, for the symmetric setting, we show that for any fixed and finite number of channel uses, the DP scheme achieves MSE lower than the MSE achieved by either the LQG or the OL schemes.

1 Introduction

We study the transmission of a pair of correlated Gaussian sources over a two-user memoryless Gaussian broadcast channel (GBC) with correlated noise components at the receivers, when the transmitter has access to noiseless causal feedback (FB) from both receivers. We abbreviate this channel as the GBCF. Motivated by practical broadcast scenarios with strict delay and complexity constraints, e.g., live multimedia broadcast, transmission of critical system parameters in a smart grid, or a body-area sensor network, we focus on linear uncoded transmission schemes, namely, schemes that do not encode over sequences of source symbol pairs.\textsuperscript{1}

Previous studies on GBCFs focused on the channel coding problem which assumes independent and uniformly distributed messages and performance characterization for the infinite horizon regime, i.e., the number of channel uses is unbounded. In the present work we study a lossy joint source-channel coding (JSCC) problem focusing on the finite horizon regime: The sources are assumed to be correlated, and each source is to be reconstructed at its corresponding receiver within a target non-zero mean-square error (MSE) distortion. Our objective is to characterize the minimal number of channel uses required to achieve a target MSE pair.

\textsuperscript{1}This work was supported by the Israel Science Foundation under grant 396/11. Parts of this work were presented at the IEEE Global Conference on Signal and Information Processing (GlobalSIP), December 2014, Atlanta, GA, [1], and at the IEEE Information Theory Workshop (ITW), April 2015, Jerusalem, Israel, [2].

\textsuperscript{2}This is motivated by the work [3] which showed that for a zero-delay source coding problem, with memoryless sources, joint encoding of all the source symbols does not provide any gain.
We focus on \textit{linear and memoryless transmission schemes} \cite[Sec. III]{4}, i.e., the transmitted signal at any time index is restricted to be a linear combination of the channel outputs and the encoder state at the previous time index. In particular, we consider the following three transmission schemes: 1) A scheme based on the Ozarow-Leung (OL) coding scheme developed in \cite{5}, to which we refer as the \textit{OL scheme}; 2) A scheme based on the linear quadratic Gaussian (LQG) code derived in \cite{6}, to which we refer as the \textit{LQG scheme}; and 3) A novel transmission scheme derived in this work whose parameters are obtained using dynamic programming (DP) \cite{7}, to which we refer as the \textit{DP scheme}. While the OL and the DP schemes are time-varying and are designed based on signal processing theoretic arguments, the LQG scheme is time-invariant and is based on a control theoretic approach. As we show in the sequel, these differences lead to different performances in the finite horizon regime.

1.1 Prior Work

It is well known that FB does not increase the capacity of the memoryless point-to-point (PtP) channel \cite{8}. Yet, as shown by Schalkwijk and Kailath in \cite{9}, in a Gaussian PtP channel FB can reduce the complexity and delay required for achieving a target error probability. In fact, the scheme presented in \cite{9} (referred herein as the SK scheme) achieves a \textit{doubly exponential} decay in the probability of error with the number of transmitted symbols, whereas only a single exponential decay can be achieved without feedback. In the SK scheme, the receiver applies minimum MSE (MMSE) estimation to recursively estimate the transmitted source (or message). Using the FB, the transmitter tracks the estimation error at the receiver and sends it at the next channel symbol. Thus, at each channel use, the transmitter sends to the receiver only the “missing information”. The work \cite{10} generalized this idea and presented the posterior matching principle for optimal transmission over memoryless PtP channels with FB: At each channel symbol the transmitter should send only information that is independent of the past transmitted symbols, and is relevant for the reconstruction of the transmitted message.

In contrast to PtP channels, in multiuser channels FB may enlarge the capacity region. For example, the work \cite{11} showed that FB enlarges the capacity region of the multiple-access channel (MAC). Motivated by the optimality of the SK scheme for PtP Gaussian channels, the works \cite{13} and \cite{5} extended it to the two-user Gaussian MAC with FB (GMACF) and to the two-user GBCF, respectively. While for the GMACF this approach achieves the capacity region, for the GBCF this extension is generally suboptimal even though it achieves reliable communication at rates outside the capacity region of the non-degraded GBC.\footnote{We note that feedback does not enlarge the capacity region of degraded GBCFs \cite{12}.} The OL scheme of \cite{5} and the scheme of \cite{13} were later extended to GBCFs and GMACFs with more than two users as well as to Gaussian interference channels with FB (GICFs) in \cite{4}. Recently, in \cite{14}, we extended the OL scheme by using estimators with memory instead of the memoryless estimators used in the original OL scheme of \cite{5}. We note that the extended decoder does not always improve upon the decoder of \cite{5}, in fact, in some situations it may perform worse than the memoryless decoder of \cite{5}. Finally, the work \cite{15} used the scheme of \cite{13} and the OL scheme of \cite{5} to stabilize (in the mean square sense) two linear, discrete-time, scalar and time-invariant systems in closed-loop, via control over GMACFs and GBCFs, respectively. This approach was also used for stabilization over interference channels in \cite{16}.

An alternative for the SK-type schemes is based on control theory. For the Gaussian PtP channel \cite{17} showed that solving an optimal LQG control problem leads to a capacity achieving FB transmission scheme. Moreover, \cite{17} also constituted control-oriented FB transmission schemes for GBCFs, GMACFs and GICFs. In particular, for the two-user GBCF with independent noise components at the receivers, \cite{17} presented a class of coding schemes which were shown to achieve rate pairs outside the achievable rate region of the OL scheme. Later, \cite{6} used the LQG control formulation to develop the FB coding scheme, for the GBCF, referred herein as the \textit{LQG scheme} which does not require the noise components to be independent. The work \cite{6} also showed that when the noise components are independent and
have the same variance, then the LQG scheme achieves rates higher than those achieved by
the OL scheme and the scheme of [4]. In fact, recently, [18] showed that for this scenario
the LQG scheme achieves the maximal sum-rate among all possible linear-feedback schemes.
GBCFs and GICFs were also studied in [19] which presented a (non-linear with memory)
transmission scheme whose sum-rate approaches the corresponding full-cooperation bound,\(^3\)
as the signal-to-noise ratio (SNR) increases to infinity. Lastly, the recent work [20] showed
that the capacity region of the GBCF with independent noises and only a common message
cannot be achieved by linear feedback schemes such as the OL or LQG schemes.

While all the works on GBCFs reviewed above focused on the achievable rates, namely,
bits per channel use that can be transmitted reliably as the number of channel uses goes to
infinity, in the present work we study a JSCC problem. Several works studied JSCC in mul-
tiuser networks with FB. The work [21] presented sufficient conditions for lossy transmission
of DM correlated sources over a DM-MAC with FB which builds upon the hybrid coding
scheme of [22]. Lossy transmission of correlated Gaussian sources over a two-user GMACF
was studied in [23], in which sufficient conditions and necessary conditions for the achievabil-
ity of an MSE pair were derived. In [23] it was also shown that for the symmetric setting, if
the channel SNR is below a certain threshold, then an uncoded transmission scheme is opti-
mal. The works [21]–[23] focused on the scenario in which the source and channel bandwidths
are matched. On the other hand, [24] considered scenarios in which the source and channel
bandwidths are mismatched, and studied the transmission of correlated Gaussian sources
over a two-user GMACF. For the symmetric setting, [24] presented upper and lower bounds
on the energy-MSE tradeoff, i.e., the minimum transmission energy required to communicate
a pair of sources over a noisy channel, such that the sources can be reconstructed within a
specified target MSE. In [25] we study the energy-MSE tradeoff for the symmetric two-user
GBCF.

1.2 Main Contributions

In this work we study the transmission of correlated Gaussian sources over the two-user GBCF
in the finite horizon regime. Specifically, we assume that each source is to be reconstructed at
its corresponding receiver within a target MSE, and aim to characterize the minimal number
of channel uses required to achieve the target MSE pair. Our contributions are listed as
follows:

1. We adapt the OL scheme of [5] to the transmission of Gaussian sources and derive
upper and lower bounds on the minimal number of channel uses needed to achieve a
target pair of MSEs. Then, we consider the symmetric setting, in which the sources
have equal variances, the noise variances at the receivers are equal, and the target
MSEs are equal. For this setting, when the sources are independent, and the noises are
independent, we show that even though OL applies uncoded transmission, in the low
SNR regime, it achieves approximately the same source-channel bandwidth ratio as the
best known separation-based scheme which applies source and channel coding with an
asymptotically large blocklength.\(^4\)

2. We adapt the LQG scheme of [6] to the transmission of Gaussian sources. As the
original LQG may perform poorly in the finite horizon regime, we first derive a new
decoder based on the MMSE criterion, which, in the finite horizon regime, outperforms
the LQG decoder presented in [6]. For the general setting we present lower and upper
bounds on the minimal number of channel uses needed to achieve a target pair of MSEs.
For the symmetric setting, we show that transmitting scaled sources can significantly
improve the finite horizon performance of the LQG scheme. Moreover, we characterize

\(^3\)In the full-cooperation bound each receiver knows the other receiver’s channel output, see [19, Eqn. (1)].

\(^4\)Note that for the considered setting, when the sources are independent, then separate source-channel
coding is optimal.
the optimal scaling (in the MMSE sense),\(^5\) and show that this scaling achieves the same MSE exponents as those of [6]. Thus, we obtain a linear time-invariant transmission scheme with good finite horizon performance and best known infinite horizon performance. Furthermore, for the transmission of scaled sources, we explicitly characterize the minimal number of channel uses required to achieve a target MSE via the roots of a second order polynomial. Finally, we show that the above results can be used to characterize a range of transmit powers, for which the OL scheme outperforms the LQG scheme, i.e., it achieves the target MSE in fewer channel uses. This result is in contrast to the infinite horizon regime in which the LQG scheme strictly outperforms the OL scheme.

3. We present a new linear and memoryless transmission scheme designed based on DP. For a finite number of channel uses, we show that the newly derived DP scheme achieves an MSE lower than those achieved by either the LQG or the OL scheme. Since finding the coefficients of this scheme becomes computationally infeasible as the number of channel uses becomes large, we also propose an approximated low-complexity version of the DP scheme. Simulation results indicate that for moderate to high SNRs this approximate version has negligible or no performance loss compared to the exact DP scheme.

The rest of this paper is organized as follows: The problem formulation is introduced in Section 2. We study JSCC via the OL and LQG schemes in sections 3 and 4, respectively. The DP scheme is introduced in Section 5, and a comparison of the three schemes along with numerical examples are presented in Section 6. Finally, concluding remarks are given in Section 7.

2 Problem Definition

2.1 Notation
We use capital letters to denote random variables (RVs), e.g., \(X\), and boldface letters to denote random column vectors, e.g., \(X\); the \(k\)’th element of a vector \(X\) is denoted by \(X_k\), and we use \(X_k^j\) where \(k < j\), to denote \((X_k, X_{k+1}, \ldots, X_{j-1}, X_j)\). We use sans-serif font to denote deterministic vectors and matrices: boldface letters denote vectors, e.g., \(B\), while regular letters denote matrices, e.g., \(M\). \([M]_{m,n}\) is used to denote the entry at the \(m\)’th row and \(n\)’th column of a matrix \(M\), and \(\det(M)\) denotes the determinant of a square matrix \(M\). We use \(E\{\cdot\}\), \((\cdot)^T\), \(\log(\cdot)\) and \(\mathbb{R}\) to denote expectation, transpose, natural basis logarithm, and the set of real numbers, respectively. Finally, we define \([x]^+ \triangleq \max\{x, 0\}\), \(\operatorname{sgn}(x)\) as the sign of \(x\), where \(\operatorname{sgn}(0) \triangleq 1\), and denote the ceiling function of \(x\) by \(\lceil x \rceil\).

2.2 System Model
The two-user GBCF is depicted in Fig. 1. All the signals are real. The encoder observes a realization of a pair of jointly Gaussian correlated sources, denoted by \(S = [S_1, S_2]^T\), and is required to send the source \(S_i, i = 1, 2\), to the \(i\)'th receiver, denoted by \(R_{x_i}\). Let \(S \sim \mathcal{N}(0, Q_s)\), where the covariance matrix, \(Q_s\), is given by:

\[
Q_s = \begin{bmatrix}
\sigma_1^2 & \rho_{s}\sigma_1\sigma_2 \\
\rho_{s}\sigma_1\sigma_2 & \sigma_2^2
\end{bmatrix}, \quad \sigma_i^2 = E\{S_i^2\}, \quad \rho_s = \frac{E\{S_1S_2\}}{\sigma_1\sigma_2}, \quad |\rho_s| < 1.
\]

Each pair of source symbols is transmitted using \(K\) channel uses, indexed by \(k = 1, 2, \ldots, K\). The channel outputs at the decoders are given by:

\[
Y_{i,k} = X_k + Z_{i,k}, \quad i = 1, 2,
\]

\(^5\)Note that finding this scaling factor is involved as it requires an exact characterization of the average instantaneous transmission power of the LQG scheme. This is explained in detail in Section 4.4.1.
where the noises \([Z_{1,k}, Z_{2,k}]^T \sim \mathcal{N}(0, Q_z)\), are independent and identically distributed (i.i.d.) over \(k = 1, 2, \ldots, K\), with covariance matrix \(Q_z\) given by:

\[
Q_z = \begin{bmatrix}
\sigma_z^2 & \rho_z \sigma_z \sigma_2 \\
\rho_z \sigma_z \sigma_2 & \sigma_2^2
\end{bmatrix}, \quad \rho_z = \frac{\mathbb{E}\{Z_1 Z_2\}}{\sigma_z \sigma_2}, \quad |\rho_z| < 1.
\]

Let \(B = [1, 1]^T, Y_k = [Y_{1,k}, Y_{2,k}]^T\) and \(Z_k = [Z_{1,k}, Z_{2,k}]^T\). The signal model (1) can now be written in the following form:

\[
Y_k = BX_k + Z_k.
\]

At time \(k\), \(\text{Rx}_i, i = 1, 2\), uses its received channel outputs \(Y_{i,1}, Y_{i,2}, \ldots, Y_{i,k}\), to estimate \(S_i\):

\[
\hat{S}_{i,k} = g_{i,k}(Y_{i,1}, Y_{i,2}, \ldots, Y_{i,k}), \quad g_{i,k} : \mathcal{R}^k \rightarrow \mathcal{R}, \quad k = 1, 2, \ldots, K,
\]

and the encoder maps the observed pair of sources and the received FB into a channel input via:

\[
X_k = f_k(S_1, S_2, Y_1, Y_2, \ldots, Y_{k-1}), \quad f_k : \mathcal{R}^{2k} \rightarrow \mathcal{R},
\]

subject to a per-symbol average power constraint defined as:

\[
\mathbb{E}\{X_k^2\} \leq P, \quad \forall k = 1, 2, \ldots, K.
\]

For a specific set of parameters \((\sigma_z^2, \sigma_2^2, \rho_z, \sigma_1^2, \sigma_2^2, \rho_s)\), we define a \((D_1, D_2, K)\) code to be a collection of \(K\) encoding functions each satisfying (5), and two decoding functions such that:

\[
\mathbb{E}\{(S_i - \hat{S}_{i,K})^2\} \leq D_i, \quad 0 < D_i \leq \sigma_i^2, \quad i = 1, 2.
\]

For a given target MSE pair \((D_1, D_2)\), our objective is to characterize the minimal number of channel uses \(K\) such that a \((D_1, D_2, K)\) code exists. In the sequel, we let \(K_{\text{SCHEME}}\) denote the minimal number of channel uses required to achieve an MSE pair \((D_1, D_2)\) by the scheme “SCHEME” \(\in \{\text{OL, LQG, DP}\}\).

Remark 1. Note that in (5) we use a per-symbol average power constraint, see for example [26, Eq. (22)] and [27, Sec. VII]. We emphasize that we focus on the finite horizon regime, hence, there is no point in using an asymptotic power constraint, e.g., [6, Eqn. (2)], as any pair of MSEs can be achieved using a single transmission with arbitrarily (but finite) large power.

Next, we recall some results and definitions from [5], and provide a finite horizon analysis of the OL scheme.
3 The OL Scheme

3.1 The OL Scheme for JSCC

In the OL scheme, prior to the transmission of a channel symbol, the transmitter uses the FB to calculate the estimations at the receivers, from which it obtains the estimation errors at the receivers. The transmitter then sends a linear combination of these estimation errors. Thus, each receiver obtains its estimation error corrupted by a correlated noise term, consisting of the other receiver’s error, and additive noise. Each receiver then updates its estimation accordingly, thereby, decreasing the variance of its estimation error. The scheme is terminated after \( K_{OL} \) channel uses, where \( K_{OL} \) is chosen such that the target MSE for each source is achieved at the corresponding receiver.

**Setup:** Let \( \hat{S}_{i,k} \) be the estimate of \( S_i \) at Rx\(_i\) after the reception of the \( k \)th channel output, \( Y_{i,k} \). Let \( \epsilon_{i,k} \triangleq \hat{S}_{i,k} - S_i \) be the estimation error after \( k \) transmissions, and define \( \hat{\epsilon}_{i,k-1} \triangleq \hat{S}_{i,k-1} - \hat{S}_{i,k} \), which implies that \( \epsilon_{i,k} = \epsilon_{i,k-1} - \hat{\epsilon}_{i,k-1} \). Lastly, define \( \alpha_{i,k} \triangleq \mathbb{E}\{\epsilon_{i,k}^2\} \) to be the MSEs at Rx\(_i\) after \( k \) transmissions, and \( \rho_k \triangleq \frac{\mathbb{E}\{\epsilon_{i,k} \alpha_{i,k}\}}{\sqrt{\alpha_{i,k}\alpha_{i,k}}} \) to be the correlation coefficient between the estimation errors.

**Encoding:** Set \( \hat{S}_{i,0} = 0 \), which implies that \( \epsilon_{i,0} = -S_i \), \( \alpha_{i,0} = \mathbb{E}\{\epsilon_{i,0}^2\} = \sigma^2 \), and \( \rho_0 = \rho_s \). Next, for a given \( P \), let \( g > 0 \) be a constant which controls the tradeoff between the information rate to Rx\(_1\) and to Rx\(_2\), and define \( \Psi_k \triangleq \sqrt{\frac{P}{1 + g^2 + 2g|\rho_k|}} \). At the \( k \)th iteration, \( 1 \leq k \leq K \), the transmitter sends

\[
X_k = \Psi_{k-1} \cdot \left( \frac{\epsilon_{1,k-1}}{\sqrt{\alpha_{1,k-1}}} + \frac{\epsilon_{2,k-1}}{\sqrt{\alpha_{2,k-1}}} \cdot g \cdot \text{sgn}(\rho_{k-1}) \right),
\]

and the corresponding channel outputs are given in (1).

**Remark 2.** Note that (7) implies that in the OL scheme the average per-symbol transmission power is constant. Therefore, the OL scheme inherently satisfies the average per-symbol power constraint (5).

**Decoding:** After the \( k \)th channel use, the estimator which minimizes the instantaneous MSE \( \mathbb{E}\{(S_i - \hat{S}_{i,k})^2\} \) is the conditional expectation, [28, Eqn. (11.10)]:

\[
\hat{S}_{i,k} = \mathbb{E}\{S_i | Y_{i,1}, Y_{i,2}, \ldots, Y_{i,k}^T\}.
\]

However, as successive channel outputs are not independent, the performance analysis of this estimator is highly complicated. For this reason, the work [5] used a simpler and suboptimal approach in which Rx\(_i\) estimates \( \epsilon_{i,k-1} \) based only on \( Y_{i,k} \):\(^6\)

\[
\hat{\epsilon}_{i,k-1} = \mathbb{E}\{\epsilon_{i,k-1} | Y_{i,k}\} Y_{i,k}.
\]

Then, similarly to [29, Eq. (7)], the estimate of the source \( S_i \) is given by:

\[
\hat{S}_{i,k} = \sum_{m=1}^{k} \epsilon_{i,m-1}.
\]

**Remark 3.** In [13] it is shown that for the 2-user GMACF this approach is optimal in the MMSE sense. This follows as in the MAC setup both estimation errors are estimated from the same channel output, thus, they are orthogonal to the previous channel outputs. On the other hand, in the GBCF this approach is sub-optimal since \( [Y_{i,1}, Y_{i,2}, \ldots, Y_{i,k-1}]^T \) is not necessarily orthogonal to \( \epsilon_{2,k-1} \). In [14] we extended the estimator (8) to use \( [Y_{i,k}, Y_{i,k-1}]^T \) instead of using only \( Y_{i,k} \). This resulted in a transmission scheme which is linear but not memoryless.

\(^6\)In [4] this approach is referred to as memoryless linear minimum MSE.
Let $\pi_i \triangleq P + \sigma_{z,i}^2$, $\Sigma \triangleq P + \sigma_{z,1}^2 + \sigma_{z,2}^2 - \rho_z \sigma_{z,1}\sigma_{z,2}$, and $\varsigma_i \triangleq \sigma_{z,i}^2 - \rho_z \sigma_{z,1}\sigma_{z,2}$. In [5] the MSEs of the (memoryless) estimators (8) are shown to be given by the recursive expressions [5, Eqs. (5)-(6)]:

\[
\begin{align*}
\alpha_{1,k} &= \alpha_{1,k-1} \frac{\sigma_{z,1}^2 + \Psi_{k-1}^2 \sigma_{z,2}^2(1 - \rho_{k-1}^2)}{\pi_1}, \\
\alpha_{2,k} &= \alpha_{2,k-1} \frac{\sigma_{z,2}^2 + \Psi_{k-1}^2 (1 - \rho_{k-1}^2)}{\pi_2},
\end{align*}
\]

and $\rho_k$ is given by [5, Eqn. (7)]:

\[
\rho_k = \frac{(\rho_z \sigma_{z,1}\sigma_{z,2}\Sigma + \varsigma_1^2 \varsigma_2^2)\rho_{k-1} - \Psi_{k-1}^2 \Sigma \cdot g(1 - \rho_{k-1}^2) \text{sgn}(\rho_{k-1})}{\sqrt{\pi_1 \pi_2} \sqrt{\sigma_{z,1}^2 + \Psi_{k-1}^2 \sigma_{z,2}^2(1 - \rho_{k-1}^2)}}.
\]

Remark 4. In [5] it was shown that there exists a $\rho \in [0,1]$ such that a steady state is achieved in the sense that if $|\rho_{k-1}| = \rho$, then $\rho_k = -\rho_{k-1}$. This $\rho$ can be obtained by setting $\rho_k = \rho$ and $\rho_{k-1} = -\rho$ in (11), and finding the roots of the resulting sixth-order polynomial. Let $\hat{\rho}$ denote the largest root of this polynomial in $[0,1]$. In [5, pg. 669] it is described how to initialize the transmission to achieve steady-state in (11). This procedure optimizes the asymptotic performance of the OL scheme. In this work we do not apply the initialization procedure of [5, pg. 669], since it may result in higher MSE after a finite number of channel uses. Instead, we set $\epsilon_{i,0} = -S_i$ and $\rho_0 = \rho_s$ in order to take advantage of the correlation among the sources.

### 3.2 Finite Horizon Analysis of OL

From (10) it follows that the MSEs at time instance $k$ depend on $\rho_{k-1}$. However, the sequence $\rho_k$ is defined via the non-linear recursion characterized in (11), which implies that an explicit characterization of $K_{OL}$ is rather complicated. Thus, in the following theorem we present upper and lower bounds on $K_{OL}$.

**Theorem 1.** The OL scheme with the decoder in (8) terminates within $K_{OL}^\text{ub} \leq K_{OL} \leq K_{OL}^\text{lb}$ channel uses, where:

\[
\begin{align*}
K_{OL}^\text{ub} &= \left[ \frac{(1 + g^2)}{P} \max \left\{ \frac{\pi_1 \log \left( \frac{\sigma_1^2}{D_1} \right)}{g^2} \log \left( \frac{\sigma_1^2}{D_2} \right) \right\} \right], \\
K_{OL}^\text{lb} &= \left[ \max \left\{ \frac{\pi_1 \log \left( \frac{\sigma_1^2}{D_1} \right)}{P} \frac{\sigma_{z,1}^2}{P} \log \left( \frac{\sigma_{z,1}^2}{D_2} \right) \right\} \right].
\end{align*}
\]

**Proof.** The upper and lower bounds in (12a) and (12b), respectively, are obtained via lower and upper bounding $\rho_k$ in (10). A detailed proof is provided in Appendix A.1. \qed

### 3.3 OL vs. a Separation-Based Scheme

Next, we focus on the symmetric setting in which $\sigma_1^2 = \sigma_z^2 \triangleq \sigma_{z,i}^2$, $\sigma_{z,1}^2 = \sigma_{z,2}^2 \triangleq \sigma_z^2$ and $D_1 = D_2 \triangleq D$; thus, we also set $g = 1$. We further assume that both the sources and the noises are independent, i.e., $\rho_z = 0$ and $\rho_s = 0$. We compare the OL scheme with a transmission scheme which is designed based on source-channel separation, applies coding over multiple samples of source pairs, and uses infinite number of channel uses. Clearly, by coding over multiple samples of source pairs one can obtain MSEs which are at least as low as the MSEs achieved by uncoded schemes.

Consider a coding scheme which requires (on average) $K$ channel uses to send $m$ samples of source pairs in order to achieve a target MSE $D$. We define the source-channel bandwidth ratio of this scheme as $\kappa \triangleq K/m$. 

We first note that as the OL scheme applies uncoded transmission, then \( m = 1 \), and its source-channel bandwidth ratio is given by \( \kappa_{OL} = K_{OL} \). For the symmetric setting with \( \rho_s = 0 \), source-channel bandwidth separation is optimal, see [31, Thm. 2]. Let \( \kappa_{sep} \) denote the source-channel bandwidth ratio of the optimal separation-based scheme which applies optimal source compression followed by a capacity achieving channel code. Since \( \rho_s = 0 \), the optimal source code compresses each of the Gaussian sources separately via an optimal rate-distortion code [32, Thm. 13.3.2], resulting in two independent messages. As the optimal channel code for the GBCF is not known, in the following we consider upper and lower bounds on \( \kappa_{sep} \). A lower bound on \( \kappa_{sep} \) is obtained by using the upper bound on the symmetric achievable rate for the GBCF, stated in [5, pg. 671], i.e., by letting one of the receivers have access to both channel outputs. In [5, pg. 671] it is shown that if \( R \) is a symmetric achievable rate for the GBCF, and \( \rho_s = 0 \), then:

\[
R \leq \frac{1}{2} \log \left( \sqrt{\frac{9}{4} + \frac{2P}{\sigma_2^2}} - 1 \right) \leq \frac{1}{2} \log \left( \sqrt{\frac{9}{4} + 2\text{SNR} - \frac{1}{2}} \right), \tag{13}
\]

where in (a) we set \( \text{SNR} \triangleq \frac{P}{\sigma_2^2} \). Equation (13) is obtained by manipulating the results of [5]. For completeness, this analysis is provided in Appendix A.2. The lower bound on \( \kappa_{sep} \) is thus given by:

\[
\kappa_{sep} \geq \frac{\log \left( \frac{\sigma_2^2}{D} \right)}{\log \left( \sqrt{\frac{9}{4} + 2\text{SNR} - \frac{1}{2}} \right)} \triangleq \kappa_{lb}^{sep},
\]

An upper bound on \( \kappa_{sep} \) is obtained by using the LQG code of [6], which is the best known code for the GBCF:

\[
\kappa_{sep} \leq \frac{\log \left( \frac{\sigma_2^2}{D} \right)}{2 \log |a_1|} \triangleq \kappa_{ub}^{sep},
\]

where \( a_1 \) is defined in [6, Eq. (14)]. A detailed description of the LQG scheme is provided in the following Section 4.1. Recall that when restricting attention only to linear schemes then, when \( \rho_s = 0 \), the LQG code of [6] is the optimal channel code in the sense of maximal sum-rate [18], which motivates focusing on \( \rho_s = 0 \). In the following proposition we upper bound the terms \( K_{OL} - \kappa_{sep}^{ub} \) and \( K_{OL} - \kappa_{sep}^{lb} \).

**Proposition 1.** In the symmetric setting with \( \rho_s = \rho_z = 0 \), \( K_{OL} - \kappa_{sep}^{ub} \) and \( K_{OL} - \kappa_{sep}^{lb} \) are upper bounded by:

\[
K_{OL} - \kappa_{sep}^{ub} \leq \left( 2 \log \left( \frac{\sigma_2^2}{D} \right) \right) \tag{14a}
\]

\[
K_{OL} - \kappa_{sep}^{lb} \leq \left( 2 + \frac{2}{\text{SNR}} - \frac{1}{\sqrt{2\text{SNR}}} \right) \log \left( \frac{\sigma_2^2}{D} \right). \tag{14b}
\]

**Proof.** The proof of (14a) is provided in Appendix A.3. The proof of (14b) is provided in Appendix A.4.

It can be observed that the right-hand side (RHS) of (14a) is independent of the SNR. We further note that both \( K_{OL} \) and \( \kappa_{sep}^{ub} \) increase when SNR decreases. Therefore, (14a) implies that for low enough SNR, \( K_{OL} - \kappa_{sep}^{ub} \ll K_{OL} \), which implies that the gap is negligible compared to \( K_{OL} \) and \( \kappa_{sep}^{ub} \). For instance, let \( \sigma_2^2 = 1 \) and \( D = 10^{-2} \), which implies that \( \left\lfloor 2 \log \left( \frac{\sigma_2^2}{D} \right) \right\rfloor = 10 \). For \( P = 0.001 \) and \( \sigma_2^2 = 1 \), we have \( \kappa_{sep}^{ub} = 9213 \) and therefore \( K_{OL} \leq 9223 \); thus, \( (K_{OL} - \kappa_{sep}^{ub})/K_{OL} \approx 10^{-3} \). An explicit calculation results in \( K_{OL} = 9213 \), thus, here \( K_{OL} = \kappa_{sep}^{ub} \). This supports the observation that the gap in (14a) is negligible compared to \( K_{OL} \) and \( \kappa_{sep}^{ub} \). It should be noted that \( \kappa_{sep}^{ub} \) is achieved by applying source and channel coding with an asymptotically large blocklength. In particular, in contrast to the OL scheme, coding takes place over multiple samples of source pairs, and for \( \kappa_{sep}^{ub} \) to be approached infinitely many channel uses are required, thus, introducing a large delay and a high complexity. On
the other hand, the OL scheme uses finite number of channel uses for the transmission of a single source pair. In spite of this fundamental difference, Prop. 1 shows that in the low SNR regime the performance loss of the OL scheme compared to the separation-based scheme which uses the infinite blocklength LQG channel code is negligible.

On the other hand, it can be observed that in the low SNR regime the RHS of (14b) is approximately given by $\left\lfloor \log \left( \frac{\sigma^2}{\beta^2} \right) \right\rfloor \frac{2}{\text{SNR}}$. Although the RHS of (14b) only constitutes an upper bound, simulation results indicate that indeed, in the low SNR regime, $K_{OL} - \kappa_{lb}^{sep}$ increases proportionally to $\frac{1}{\text{SNR}}$. We conjecture that this negative result is due to the fact that the upper bound (13) based on [5, pg. 671] is not tight.

Next, we discuss a control theoretic approach for the problem of transmitting correlated Gaussian sources over the GBCF.

4 The LQG Scheme

It was observed in [17] and [30] that there is a natural duality between the problem of FB stabilization and communications over the PtP Gaussian channel with FB. Using this duality, results and tools from control theory were used to design FB communications schemes providing desired communication rates. This duality was also exploited to construct communications schemes over multiuser Gaussian channels with FB: In [17] a duality between communications over the GBCF with unit-variance independent noise components, and a FB stabilization problem was established; yet, [17] did not present an explicit FB communications scheme. A scheme which belongs to the class of schemes analyzed in [17], is the LQG scheme of [6], which also supports the case of correlated noise components.

Note that the LQG scheme achieves the best known communication rates over the GBCF. Furthermore, [6, Lemma 1] characterizes a linear relationship between the achievable rates and the achievable MSE exponent, which corresponds to the slope of decay of the logarithm of the MSE for large number of channel uses. Therefore, for large number of channel uses, higher achievable rates correspond to higher rate of decrease of the MSE. This implies that for large enough number of channel uses (or low enough MSEs), the LQG scheme is preferable compared to schemes which achieve lower rates. This, together with the time-invariant property of the LQG scheme, motivates studying the LQG scheme even in the finite horizon regime.

We emphasize here that the LQG scheme is designed to optimize the infinite horizon performance, while in this work we focus on the finite horizon regime. The main challenge in designing a finite horizon LQG scheme, for the studied setting, is the fact that the controllers are constrained, see (5), which leads to minimization problems for which no explicit solutions exist. Furthermore, typically, an LQG scheme is designed for a specific LTI system [7, Ch. 4.1], and it is not clear if and how this system is related to the finite horizon performance. Therefore, in this section we adapt the LQG scheme of [6] to the transmission of Gaussian sources, while applying a finite horizon analysis. Furthermore, for the symmetric setting, we characterize the scaling of the sources which optimizes the finite horizon performance, without changing the infinite horizon performance. In order to obtain this characterization we first derive an exact expression for the per-symbol average transmission power of the LQG scheme, and then deduce the optimal scaling factor such that the per-symbol average power constraint is satisfied.

4.1 The LQG scheme for JSCC

We adapt the LQG scheme of [6], to the transmission of Gaussian sources. The LQG scheme is obtained by mapping the FB control problem into a linear code for the GBCF. The asymptotic performance of this scheme is determined by the eigenvalues of the open-loop matrix of a linear system. These eigenvalues are determined by the minimal power required to stabilize
Figure 2: Control system modeling of transmission over GBCF. The states of the system are denoted by $U_k = [U_1,k, U_2,k]^T$. The controller generates a scalar signal $X_k$, and the noisy channel outputs $Y_1,k$ and $Y_2,k$ represented as the vector $Y_k = [Y_1,k, Y_2,k]^T$ are fed back to the system with a unit delay. $(\hat{S}_{1,k}, \hat{S}_{2,k})$ are the reconstructions of $(S_1, S_2)$ after the $k$th channel use.

The states of the system using the FB. In the finite horizon regime the LQG scheme is terminated after $K$ channel uses when the target MSEs pair is met.

Consider a two-dimensional unstable dynamical system, depicted in Fig. 2, which is stabilized by a controller observing the entire system state vector, $U_k = [U_1,k, U_2,k]^T$. The controller outputs a scalar signal $X_k$, which is corrupted by additive Gaussian noises. Recall that $Y_k$ is the noisy observed control signal at the output of the channel at time $k$, given in (2), and let $A = \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix}$ with $a_i \in \mathbb{R}, a_1 \neq a_2, |a_i| > 1$. The system state vector at time $k$, $U_k$, is recursively given by:

$$U_1 = S, \quad U_k = AU_{k-1} + Y_{k-1}, \quad k = 2, 3, \ldots, K. \quad (15)$$

**Encoding:** In the corresponding communication problem the encoder consists of the system given by (15) and of the controller, see Fig. 2. At each time index the encoder recursively computes $U_k$ and transmits $X_k$ obtained from $U_k$ using the linear controller presented in [6, Lemma 4]:

$$X_k = -C^T U_k,$$

where $C = [c_1, c_2]^T$. The vector $C$ is given by $C = (B^T GB + 1)^{-1}AG^TB$, where $B$ is defined above (2), and $G$ is the unique positive-definite solution of the discrete algebraic Riccati equation (DARE) [6, Eq. (22)]:

$$G = A^T GA - A^T GB (B^T GB + 1)^{-1}B^T GA, \quad (16)$$

such that the magnitudes of both eigenvalues of the matrix $A - BC^T$ are smaller than 1. It follows from [6, Lemma 4] that, as $k \to \infty$, the covariance matrix of $U_k$, $Q_{u,k}$, converges to the solution of the discrete algebraic Lyapunov equation [6, Eq. (23)]:

$$Q_u = (A - BC^T)Q_u(A - BC^T)^T + Q_z, \quad (17)$$

where the solution of (17) is restricted to be a positive definite matrix. Finally, it follows from [6, Lemma 4] that the matrix $A$ can be obtained from the minimum asymptotic average power: $P(A, Q_z) = C^T Q_u C = \text{trace}(GQ_z)$, see [6, Eq. (24)].

**Decoding:** In the work [6] the so-called “zero trajectory” (ZT) detector is used. This detector recursively estimates $\hat{U}_{i,k}$ via [6, Eq. (18)]:

$$\hat{U}_{i,1} = 0, \quad \hat{U}_{i,k} = a_i \hat{U}_{i,k-1} + Y_{i,k-1}, \quad k = 2, 3, \ldots, K + 1. \quad (18)$$

Then, it estimates $\hat{S}_i$ from $\hat{U}_{i,k+1}$ via [6, Subsec. IV.A]:

$$\hat{S}_{i,k} = -a_i^{-k} \hat{U}_{i,k+1}, \quad (19)$$

Note that if $a_1 = a_2$ then the pair $(A, B)$ is not controllable, see [7, Def. 4.1.1 and Prop. 4.4.1].
which results in the MSE \[6, \text{proof of Lemma 3}]:

\[
E\{ (S_i - \hat{S}_{i,k})^2 \} = a_i^{-2k} E\{ U_{i,k+1}^2 \}.
\] (20)

**Remark 5.** Note that in contrast to the OL scheme, in the LQG scheme the encoder and the decoders are decoupled. More precisely, in the OL scheme the transmitted signal at time \(k\) is a linear combination of the estimation errors at time \(k - 1\). Thus, the encoder and the decoders are coupled. On the other hand, in the LQG scheme the transmitted signal at time \(k\), \(X_k\), depends only on \(U_k\) and is not a function of \(\hat{U}_{i,k}\) or \(\hat{S}_{i,k}\).

### 4.2 An Improved LQG Decoder

Note that the decoding rule (19) is not necessarily optimal in the MMSE sense. Let \(M \triangleq A - BC^T\) denote the closed-loop matrix, and let \(Q_{u,k} \triangleq E\{ U_k U_k^T \}\) denote the state covariance matrix at time \(k\), with \(Q_{u,1} = Q_s\). The MMSE estimator of \(S_i\), based on the observation \(\hat{U}_{i,k+1}\) computed via (18), is stated in the following theorem:

**Theorem 2.** The MMSE estimator of \(S_i, i = 1, 2\), at time \(k\), based on the observation \(\hat{U}_{i,k+1}\) is:

\[
\hat{S}_{i,k} = \frac{[M^k Q_s]_{i,i} - \sigma_i^2 a_i^k}{[Q_{u,k+1}]_{i,i} - 2a_i^k [M^k Q_s]_{i,i} + \sigma_i^2 a_i^{2k}} \hat{U}_{i,k+1}.
\] (21)

Furthermore, the MSE of \(\hat{S}_{i,k}\) is given by:

\[
E\{ (S_i - \hat{S}_{i,k})^2 \} = \sigma_i^2 [Q_{u,k+1}]_{i,i} - 2a_i^k [M^k Q_s]_{i,i} + \sigma_i^2 a_i^{2k},
\] (22)

and as \(k \to \infty\) the MSE expression in (22) coincides with the MSE of the decoder in (19).

**Proof.** The proof is provided in Appendix B.1.

**Remark 6.** Since the estimator in (21) is the optimal estimator of \(S_i\), based on the observation \(\hat{U}_{i,k+1}\), it clearly achieves an MSE smaller than or equal to that achieved by the estimator in (19). In particular, (21) outperforms (19) in the finite horizon regime, i.e., for large MSEs, see Fig. 5.

### 4.3 Finite Horizon Analysis of LQG

Next, we discuss the LQG scheme in the finite horizon regime. We begin with the average instantaneous transmission power which we denote by \(P_k\). In contrast to the OL scheme in which \(P_k = P, \forall k\), in the LQG scheme \(P_k\) is time-varying. While the LQG theory implies that \(P_k \to P\) as \(k \to \infty\), it does not restrict \(P_k\) for any finite \(k\) values. It follows that \(P_k\) may be larger than \(P\), thus, violating the per-symbol average power constraint in (5). This implies that for specific \(P, \sigma_1^2\) and \(\sigma_2^2\), there are pairs of sources which cannot be transmitted using the LQG scheme with \(U_1 = S\) as defined in (15). In the following subsection, we present a sufficient condition under which the LQG scheme, with the initialization in (15), satisfies (5). For the symmetric setting we use the same approach to find a condition which is both sufficient and necessary, see Subsection 4.4.

#### 4.3.1 Satisfying the Average Per-Symbol Power Constraint

Let \([\lambda_1, \lambda_2]^T\) denote the eigenvalues of the closed-loop matrix \(M\), and let \(V = \begin{bmatrix} v_1 & v_2 \\ v_3 & v_4 \end{bmatrix}\) be a matrix whose columns are the corresponding eigenvectors of \(M\). Recall that \(C = [c_1, c_2]^T\)
Further define:

\[
\omega_1(s_1, s_2, \rho) \triangleq \frac{c_1(s_1 v_1 v_4 - \rho s_2 v_1 v_2) + c_2(s_1 v_3 v_4 - \rho s_2 v_1 v_3)}{\text{det}(V)}
\]  

(23a)

\[
\omega_2(s_1, s_2, \rho) \triangleq \frac{c_1(\rho s_2 v_1 v_2 - s_1 v_1 v_3) + c_2(\rho s_2 v_1 v_4 - s_1 v_3 v_4)}{\text{det}(V)}
\]  

(23b)

\[
\omega_3(s_1, s_2, \rho) \triangleq \frac{-\omega_2 \sqrt{1 - \rho^2}}{\text{det}(V)}
\]  

(23c)

\[
\omega_4(s_1, s_2, \rho) \triangleq \frac{\omega_2 \sqrt{1 - \rho^2}}{\text{det}(V)}
\]  

(23d)

Further define:

\[
\alpha_1(s_1, s_2, \rho) \triangleq \omega_1^2(s_1, s_2, \rho) + \omega_2^2(s_1, s_2, \rho)
\]  

(24a)

\[
\alpha_2(s_1, s_2, \rho) \triangleq \omega_2^2(s_1, s_2, \rho) + \omega_3^2(s_1, s_2, \rho)
\]  

(24b)

\[
\alpha_3(s_1, s_2, \rho) \triangleq 2\omega_1(s_1, s_2, \rho)\omega_2(s_1, s_2, \rho) + 2\omega_3(s_1, s_2, \rho)\omega_4(s_1, s_2, \rho),
\]  

(24c)

and finally define:

\[
\eta_1(s_1, s_2, \rho) \triangleq \frac{\alpha_1(s_1, s_2, \rho)}{1 - \lambda_1^2}, \quad \eta_2(s_1, s_2, \rho) \triangleq \frac{\alpha_2(s_1, s_2, \rho)}{1 - \lambda_2^2}, \quad \eta_3(s_1, s_2, \rho) \triangleq \frac{\alpha_3(s_1, s_2, \rho)}{1 - \lambda_1 \lambda_2}.
\]  

(25)

The following proposition characterizes source pairs for which the per-symbol average power constraint in (5) is satisfied when the LQG scheme is used:

**Proposition 2.** If the following condition holds for every \( k = 1, 2, 3, \ldots \):

\[
\lambda_1^{2(k-1)}(\alpha_1(s_1, s_2, \rho_s) - \eta_1(s_1, s_2, \rho_s))
\]

\[
+ \lambda_2^{2(k-1)}(\alpha_2(s_1, s_2, \rho_s) - \eta_2(s_1, s_2, \rho_s))
\]

\[
+ (\lambda_1 \lambda_2)^{k-1}(\alpha_3(s_1, s_2, \rho_s) - \eta_3(s_1, s_2, \rho_s)) \leq 0,
\]

(26)

then the LQG scheme satisfies the per-symbol average power constraint in (5).

**Proof.** The proof is provided in Appendix B.2. \(\square\)

**Remark 7.** Note that the sufficient condition in Prop. 2 is implicit. Yet, Prop. 2 can be used to formulate explicit sufficient conditions (on the sources) for the LQG scheme to satisfy the per-symbol average power constraint in (5). For example, if \( \alpha_i(s_1, s_2, \rho_s) < \eta_j(s_1, s_2, \rho_s), j = 1, 2, 3, \) and \( \text{sgn}(\lambda_1 \lambda_2) = 1 \), then \( P_k \leq P, \forall k.\)

### 4.3.2 Analysis of the Termination Time

Let \( K_{\text{LQG}} \) denote the minimal number of channel uses required to achieve an average MSE pair \((D_1, D_2)\) with the LQG scheme using the decoder (21). In this subsection we present upper and lower bounds on \( K_{\text{LQG}} \). An explicit characterization of \( K_{\text{LQG}} \) for the symmetric setting is provided in Thm. 5, see Subsection 4.4.\(^8\)

\(^8\)Note that using the approach of Thm. 5 for the general setting results only in an implicit characterization of \( K_{\text{LQG}} \).
We begin with the following definitions:

\[
\begin{align*}
\tau_1 & \triangleq \frac{\sigma_1(|v_1v_4\lambda_1| + |v_2v_3\lambda_2|) + |\rho_s\sigma_2v_1v_2(|\lambda_2| + |\lambda_1|)}{|\det(V)|} \\
\tau_2 & \triangleq \frac{\sigma_2\sqrt{1 - \rho_s^2} (|v_1v_2|(|\lambda_2| + |\lambda_1|))}{|\det(V)|} \\
\tau_3 & \triangleq \frac{|\sigma_1v_3v_4|(|\lambda_1| + |\lambda_2|) + |\rho_s\sigma_2| (|v_1v_4\lambda_2| + |v_2v_3\lambda_1|)}{|\det(V)|} \\
\tau_4 & \triangleq \frac{\sigma_2\sqrt{1 - \rho_s^2} (|v_1v_4\lambda_2| + |v_2v_3\lambda_1|)}{|\det(V)|} \\
\theta_1 & \triangleq \tau_1^2 + \tau_3^2 + [Q_u]_{1,1} \\
\theta_2 & \triangleq \tau_2^2 + \tau_4^2 + [Q_u]_{2,2} \\
\beta_1 & \triangleq \frac{\sigma_1^2 (|v_1v_4\lambda_1| + |v_2v_3\lambda_2|) + |\rho_s\sigma_1\sigma_2v_1v_2(|\lambda_2| + |\lambda_1|)}{|\det(V)|} \\
\beta_2 & \triangleq \frac{\sigma_2^2 (|v_1v_2\lambda_2| + |v_2v_3\lambda_1|) + |\rho_s\sigma_1\sigma_2v_3v_4(|\lambda_2| + |\lambda_1|)}{|\det(V)|}.
\end{align*}
\]

where \(Q_u\) is the unique solution of (17). Upper and lower bounds on \(K_{LQG}\) are stated in the following theorem:

**Theorem 3.** The LQG scheme with the MMSE decoder in (21) and target MSE values \(D_1\) and \(D_2\) terminates within time \(K_{LQG}^{ub} \leq K_{LQG} \leq K_{LQG}^{lb}\), where:

\[
\begin{align*}
K_{LQG}^{lb} &= \max \left\{ \frac{\log \left( \frac{\rho_s^2}{\tau_1^2} \right)^+}{2 \log |a_1|}, \frac{\log \left( \frac{\rho_s^2}{\tau_2^2} \right)^+}{2 \log |a_2|} \right\} \quad (27a) \\
K_{LQG}^{ub} &= \max \left\{ \frac{\log \left[ \frac{\sigma_1^2(\tau_1^2 - \rho_s^2) - D_1\sigma_2^2}{(2\beta_1 + \sigma_1^2)D_1} \right]^+}{2 \log |a_1|}, \frac{\log \left[ \frac{\sigma_2^2(\tau_2^2 - \rho_s^2) - D_2\sigma_1^2}{(2\beta_2 + \sigma_2^2)D_2} \right]^+}{2 \log |a_2|} \right\}. \quad (27b)
\end{align*}
\]

**Proof.** The proof is provided in Appendix B.3. \(\square\)

### 4.4 Finite Horizon Analysis of LQG for the Symmetric GBCF

In this subsection we study the LQG scheme in the symmetric setting. First, we provide necessary and sufficient conditions on the scenario parameters for the LQG scheme to satisfy the power constraint in (5). Then, based on these conditions, we show how the state initialization in (15) can be scaled such that (5) is satisfied for all time indices \(k\). This also leads to a characterization of the scaling factor (of the state initialization) which minimizes the MSE. Finally, we present an explicit characterization of \(K_{LQG}\) in terms of the roots of a quadratic polynomial. Note that since \(a_1 \neq a_2\), see Subsection 4.1, in the symmetric setting we have \(a_1 = -a_2\), and the elements of the eigenvectors of the matrix \(M\) satisfy \(v_1 = v_4\), and \(v_2 = v_3\).
4.4.1 Satisfying the Average Per-Symbol Power Constraint

First, we define the following terms:

\[
\begin{align*}
\mu_0 &= 2c^2\sigma^2(1 - \rho_s) \\
\mu_1 &= 2c^2\sigma^2\frac{(1 - \rho_z + (1 + \rho_z)a^2_1)}{1 - \lambda^2_1} \\
\mu_2 &= 2c^2\sigma^2(1 + \rho_s)a^2_1 \\
\mu_3 &= \frac{2c^2\sigma^2((1 - \rho_z)\lambda^2_1 + (1 + \rho_z)a^2_1)}{1 - \lambda^2_1}.
\end{align*}
\]  

(28a) (28b) (28c) (28d)

Necessary and sufficient conditions for the power constraint to be satisfied for the LQG scheme, with the initialization in (15), are stated in the following theorem.

**Theorem 4.** In the symmetric GBCF the LQG scheme satisfies the per-symbol average power constraint (5) if and only if \(\mu_0 \leq \mu_1\) and \(\mu_2 \leq \mu_3\).

**Proof outline.** In Appendix C.1 we show that:

\[
P_k = \begin{cases} 
  P + (\mu_0 - \mu_1)\lambda^2_1(1 - k^{-1}), & \text{if } k \text{ is odd,} \\
  P + (\mu_2 - \mu_3)\lambda^2_1(1 - k^{-1}), & \text{if } k \text{ is even.}
\end{cases}
\]

(29)

Since \(|\lambda_1| < 1\), it follows that (5) is satisfied if and only if \(\mu_0 \leq \mu_1\) and \(\mu_2 \leq \mu_3\).

From Eqn. (29) and from the fact that \(|\lambda_1| < 1\), it follows that if (5) is satisfied for some odd \(k\), then it is satisfied for every odd \(k\). The same observation holds for even values of \(k\). Thus, using (29) we can characterize a scaling factor, for the transmitted source pair, which guarantees that (5) is satisfied. We further note that scaling the sources at the transmitter can be beneficial even if (5) is satisfied for the initialization in (15). As we show next, by scaling the sources we obtain that \(P_k\) is equal to \(P\) in at least (approximately) half of the time indices. Consequently, the available transmission power is used more efficiently. In the following proposition we characterize the scaling factor which minimizes the MSE, for the decoder (21), while satisfying the constraint (5). Before stating the proposition we define \(\nu\) to be:

\[
\nu = \min \left\{ \frac{\sigma^2(1 - \rho_z)}{(1 - \lambda^2_1)(1 - \rho_s)} \sigma^2((1 - \rho_z)\lambda^2_1 + (1 + \rho_z)a^2_1) \frac{(1 - \lambda^2_1)}{(1 + \rho_s)a^2_1} \right\},
\]

(30)

and let \(U_k(\gamma)\) denote the system state vector at time index \(k\), when \(U_1 = \sqrt{\gamma} \cdot S\), for some \(\gamma > 0\). In a similar manner we also define \(\hat{U}_{i,k}(\gamma)\) and \(Q_{u,k}(\gamma)\).

**Proposition 3.** The optimal scaling factor, in the MMSE sense, is \(\sqrt{\gamma} = \sqrt{\frac{\nu}{\pi}}\). Furthermore, when \(U_1 = \sqrt{\gamma} \cdot S\), the MMSE estimator of \(S_i, i = 1, 2\), at time \(k\), based on the observation \(\hat{U}_{i,k+1}(\gamma)\) is:

\[
\hat{S}_{i,k} = \frac{\sqrt{\gamma} (|M^kQ|_{i,i} - \sigma^2 a^2_i)}{|Q_{u,k+1}(\gamma)|_{i,i} - 2\gamma a^2_i |M^kQ|_{i,i} + \gamma \sigma^2 a^2_i} \hat{U}_{i,k+1}(\gamma),
\]

(31)

and the MSE of \(\hat{S}_{i,k}\) is given by:

\[
E\left\{ (S_i - \hat{S}_{i,k})^2 \right\} = \frac{\sigma^2 Q_{u,k+1}(\gamma)|_{i,i} - \gamma (|M^kQ|_{i,i})^2}{|Q_{u,k+1}(\gamma)|_{i,i} - 2\gamma a^2_i |M^kQ|_{i,i} + \gamma \sigma^2 a^2_i}.
\]

(32)

**Proof outline.** First, we show that (30) constitutes an upper bound on the variance of the sources transmitted via an LQG scheme based on (15), which satisfy (5). Explicitly writing the conditions of Thm. 4, i.e., \(\mu_0 \leq \mu_1\) and \(\mu_2 \leq \mu_3\), we obtain:

\[
\sigma^2(1 - \rho_s) \leq \frac{\sigma^2(1 - \rho_z + (1 + \rho_z)a^2_1)}{1 - \lambda^2_1}, \quad \sigma^2(1 + \rho_s)a^2_1 \leq \frac{\sigma^2((1 - \rho_z)\lambda^2_1 + (1 + \rho_z)a^2_1)}{1 - \lambda^2_1}.
\]
This implies that:

\[
\sigma^2_s \leq \min\left\{ \frac{\sigma^2_s(1 - \rho_s + (1 + \rho_z)a_1^2)}{(1 - \lambda_1^2)(1 - \rho_s)}, \frac{\sigma^2_s((1 - \rho_z)\lambda_1^2 + (1 + \rho_z)a_1^2)}{(1 - \lambda_1^2)(1 + \rho_s)a_1^2} \right\}.
\]

Therefore, the maximal possible scaling factor which satisfies (5) is \( \sqrt{\sigma^2_s} \). In Appendix C.2 we also derive the MMSE estimator for scaled transmission, see (31), and obtain its MSE, see (32). Furthermore, we show that scaling by \( \sqrt{\sigma^2_s} \) minimizes the MSE. The detailed proof is provided in Appendix C.2.

Remark 8. As shown in the proof of Prop. 3, the MSE decreases when the scaling factor increases. Therefore, the optimal scaling factor is determined by the per-symbol average power constraint. This implies that when the optimal scaling factor is used, at least one of the following statements hold: 1) \( P_k = P \) for every odd \( k \) and \( P_k \leq P \) for every even \( k \); 2) \( P_k = P \) for every even \( k \) and \( P_k \leq P \) for every odd \( k \).

4.4.2 A Numerical Example

Next, we demonstrate the results of Thm. 4 and Prop. 3. Consider the transmission of a pair of Gaussian sources with variance \( \sigma^2_s \) and correlation coefficient \( \rho_s = 0.4 \), over a GBCF with \( \sigma^2_z = 1.5 \) and \( \rho_z = 0.3 \). We further set \( P = 1 \). Fig. (3a) depicts \( P_k \) vs. \( k \) for the LQG scheme without scaling for \( \sigma^2_s = 1 \) and \( \sigma^2_s = 5 \), and for the LQG scheme with optimal scaling factor specified in Prop. 3. It can be observed that both the non-scaled LQG scheme with \( \sigma^2_s = 1 \) and the scaled LQG scheme satisfy (5); yet, the scaled scheme uses the available power more efficiently. On the other hand, when \( \sigma^2_s = 5 \), then the non-scaled LQG scheme violates the per-symbol average power constraint (5). It can further be observed that in the scaled scheme, \( P_k = P \) for all even values of \( k \), as stated in Remark 8. Fig. 3b illustrates \( \nu \), see (30), as a function of \( \rho_s \). As \( \rho_s = 0.4 \), it follows that the optimal scaling factor for \( \sigma^2_s = 1 \) is \( \sqrt{\gamma} = 2.0227 \), while the optimal scaling factor for \( \sigma^2_s = 5 \) is \( \sqrt{\gamma} = 0.9046 \). Note that in order to maximize the MSE, one should use \( \nu \) values that lie on the boundary of the shaded area in Fig. 3b.

Figure 3: Satisfying the per-symbol average power constraint for \( \sigma^2_s = 1, \rho_s = 0.4, \rho_z = 0.3, \sigma^2_z = 1.5 \) and \( P = 1 \). (a) \( P_k \) vs. \( k \) for scaled and non-scaled LQG schemes. (b) \( \nu \) as a function of \( \rho_s \), see (30).
4.4.3 Analysis of the Termination Time

Next, we explicitly characterize $K_{\text{LQG}}$ for the scaled LQG, i.e., for $U_1 = \sqrt{T} \cdot S$. We first define the following quantities:

$$
\Phi(\varsigma, \rho) \triangleq \frac{\varsigma^2 \left( (v_1^2 + v_2^2 - 2\rho v_1 v_2)^2 + 4(1 - \rho^2)v_1^2v_2^2 \right)}{\det^2(V)}
$$  \hspace{1cm} (33a)

$$
\Psi_0 \triangleq \frac{\sigma_2^2 + \lambda_1^2 \Phi(\sigma_2, \rho_2)}{1 - \lambda_1^4},
$$  \hspace{1cm} (33b)

$$
\Psi_1 \triangleq \frac{\Phi(\sigma_2, \rho_2) + \sigma_2^2 \lambda_1^2}{1 - \lambda_1^4},
$$  \hspace{1cm} (33c)

$$
\Gamma_s \triangleq \frac{\sigma_s^2(v_1^2 + v_2^2 - 2\rho v_1 v_2)}{v_1^2 - v_2^2}
$$  \hspace{1cm} (33d)

$$
\Upsilon_0 \triangleq \Psi_0(D - \sigma_2^2) - D\gamma\sigma_2^4
$$  \hspace{1cm} (33e)

$$
\Upsilon_1 \triangleq \Psi_0(\sigma_2^2 - D) + 2D\gamma\sigma_2^4
$$  \hspace{1cm} (33f)

$$
\Upsilon_2 \triangleq (\gamma \Phi(\sigma_2, \rho_2) - \Psi_1)(\sigma_2^2 - D) - \gamma\Gamma_s^2
$$  \hspace{1cm} (33g)

$$
\Upsilon_3 \triangleq \Psi_0(\sigma_2^2 - D) + 2D\Gamma_s^2
$$  \hspace{1cm} (33h)

Furthermore, let $y$ be a positive integer, and define the functions $f^{(e)}(y) \triangleq 2 \cdot \left\lceil \frac{y}{2} \right\rceil$, and $f^{(o)}(y) \triangleq 2 \cdot \left\lceil \frac{y-1}{2} \right\rceil + 1$. The following theorem explicitly characterizes $K_{\text{LQG}}$:

**Theorem 5.** Let $(x^{(e)}_1, x^{(e)}_2)$ and $(x^{(o)}_1, x^{(o)}_2)$ denote the roots of the polynomials $P^{(e)}(x) \triangleq \Upsilon_0 x^2 + \Upsilon_1 x - D\gamma\sigma_2^4$, and $P^{(o)}(x) \triangleq \Upsilon_2 x^2 + \Upsilon_3 x - D^2\gamma^2\sigma_2^4$, respectively. Furthermore, define:

$$
x^{(e)}_0 \triangleq \begin{cases} 
\min\{x^{(e)}_1, x^{(e)}_2\}, & \Upsilon_0 < 0 \\
\frac{a_1^{-1}}{D\gamma\sigma_2^4}, & \Upsilon_0 \geq 0 
\end{cases}
$$

$$
x^{(o)}_0 \triangleq \begin{cases} 
\min\{x^{(o)}_1, x^{(o)}_2\}, & \Upsilon_2 < 0 \\
\frac{D\gamma\sigma_2^4}{\Upsilon_2}, & \Upsilon_2 \geq 0 
\end{cases}
$$

Then, $K_{\text{LQG}}$ is given by:

$$
K_{\text{LQG}} = \min \left\{ f^{(e)}\left( -\frac{\log x^{(e)}_0}{2\log|a_1|} \right), f^{(o)}\left( -\frac{\log x^{(o)}_0}{2\log|a_1|} \right) \right\}.
$$  \hspace{1cm} (34)

**Proof outline.** The detailed proof is provided in Appendix C.3. The proof consists of the following steps:

1. Recall that the decoder terminates when (see (32)):

$$
\frac{\sigma_s^2[Q_{u,k+1}(\gamma)]_{i,i} - \gamma ([M^k Q_s]_{i,j})^2}{[Q_{u,k+1}(\gamma)]_{i,i} - 2\gamma a_k^2[M^k Q_s]_{i,j} + \gamma\sigma_s^2 a_k^2} \leq D.
$$  \hspace{1cm} (35)

We write $[Q_{u,k+1}(\gamma)]_{i,i}$ and $[M^k Q_s]_{i,j}$ in terms of $v_1, v_2, \lambda_1, \gamma$, and $k$, and note that since $\lambda_1 = -\lambda_2$, then a different analysis should be applied for even and for odd values of $k$.

2. We let $x = a_1^{-2k}$, and recall that $\lambda_1 = \frac{1}{a_1}$ (see [30, Lemma 2.4]).

\hspace{1cm} \text{Footnote:} f^{(e)}(y) \text{ is "round up to the nearest even integer", while } f^{(o)}(y) \text{ is "round up to the nearest odd integer".}
3. We write (35) as a quadratic polynomial in $x$. As we apply separate analysis for even and odd values of $k$, we use $P^{(e)}(x)$ to denote the quadratic polynomial for even values of $k$, and $P^{(o)}(x)$ to denote the quadratic polynomial for odd values of $k$.

4. For even values of $k$, we are interested in the minimal $k$ for which $P^{(e)}(x) \leq 0$ (for odd values of $k$ we look for the minimal $k$ for which $P^{(o)}(x) \leq 0$). Thus, the minimal $k$ is a function of the roots of the polynomials $P^{(e)}(x)$ or $P^{(o)}(x)$.

5. We compute $k$ from the roots of the polynomials via (34).

\[ \square \]

Remark 9. The result of Thm. 5 also holds if $\gamma$ is replaced by any other constant, regardless to whether (5) is satisfied or not.

Remark 10. Consider the expression for $x_0^{(o)}$. It can be observed that if $\Upsilon_2 \geq \frac{-Y_1^2}{4D_2\sigma^2}$, then we choose $x_0^{(o)}$ among the two real roots of $P^{(o)}(x)$, based on the concavity of $P^{(o)}(x)$.

More precisely, we show that if $-\frac{Y_1^2}{4D_2\sigma^2} \leq \Upsilon_2 < 0$ then $P^{(o)}(x)$ is concave with two positive roots: one smaller then 1 and one larger than 1. Hence, we choose the minimal root. On the other hand, if $0 < \Upsilon_2$ then $P^{(o)}(x)$ is convex with one negative root and one positive root smaller than 1. Hence, we choose the maximal root. Finally, when $\Upsilon_2 < -\frac{Y_1^2}{4D_2\sigma^2}$, then $P^{(o)}(x)$ does not have any real roots. Since $\Upsilon_2 < 0$ then $P^{(o)}(x)$ is concave which implies that the condition $P^{(o)}(x) \leq 0$ always hold. Therefore, the required MSE is obtained for every $k$. Hence, we set $x_0^{(e)} = a^{-2}$, which results in $K_{\text{LQG}} = 1$. For $x_0^{(e)}$ we follow similar steps while noting that $\Upsilon_0 < 0$. Hence, for even values of $k$ we only need to analyze the case of a concave polynomial.

Note that both the OL scheme of Section 3 and the LQG-oriented scheme of this section are linear and memoryless. Next, we use the DP approach to formulate the optimal linear and memoryless transmission scheme for the symmetric setting.

### 5 Linear and Memoryless Transmission Scheme Via Dynamic Programming

#### 5.1 Problem Formulation - Revisited

In this section we examine a problem complimentary to the one formulated in Section 2: The number of channel uses is fixed to $K$, and we denote the MSE after $K$ channel uses by $D_K$. Our objective is to find a linear and memoryless transmission scheme which achieves the minimal MSE $D_{K,\min}$ at each receiver.

In the following we adopt (most of) the notations used in Sections 2 and 3, and denote the information sent to the Rx$_i$ at time $k$ by $\epsilon_{i,k-1} = S_{i,k-1} - S_i$, $i = 1, 2$. As we focus on linear and memoryless schemes, $\epsilon_{i,k}$ is given by a linear combination of $\epsilon_{i,k-1}$ and $Y_{i,k}$:

\begin{equation}
\epsilon_{i,k} = \beta_{i,k} (\epsilon_{i,k-1} - b_{i,k}Y_{i,k}), \quad b_{i,k}, \beta_{i,k} \in \mathbb{R},
\end{equation}

where $\epsilon_{i,k-1} = b_{i,k}Y_{i,k}$ is the estimator of $\epsilon_{i,k-1}$, and similarly to (9): $\bar{S}_i = \sum_{m=1}^{K} \epsilon_{i,m-1}$.

Following [4] we let $m_k \in \{1, -1\}$ be a modulation coefficient. To simplify the analysis, we limit our focus to the symmetric setting, and set $|\beta_{1,k}| = |\beta_{2,k}|$, $|b_{1,k}| = |b_{2,k}|$ and let $b_k \triangleq b_{1,k}$. This leads to the following structure of $X_k$, the transmitted signal:

\begin{equation}
X_k \overset{(a)}{=} d_{k-1} (\epsilon_{1,k-1} + m_{k-1} \epsilon_{2,k-1}) \\
\overset{(b)}{=} d_{k-1} ((\epsilon_{1,k-2} - b_{k-1}Y_{1,k-1}) + m_{k-1} (\epsilon_{2,k-2} - m_{k-2}b_{k-1}Y_{2,k-1})),
\end{equation}

\[ \text{With the exception of the special case of } \Upsilon_2 = 0. \]
where in (a) $d_k > 0$ is a gain factor chosen to minimize $D_K$ under the constraint $P_k \leq P$; in (b) setting $b_{2,k} = m_{k-1}b_k$ leads to $E \{ \epsilon_{1,k} \} = E \{ \epsilon_{2,k} \}$, $\forall k$.

Next, similarly to Section 3, we let $\alpha_{1,k} \triangleq E \{ \epsilon_{1,k}^2 \}$, and note that $\alpha_{1,k} = \alpha_{2,k} \triangleq \alpha_k$. With this formulation, similarly to Section 3, $\alpha_k$ is the MSE after $k$ channel uses. Furthermore, we let $r_k \triangleq E \{ \epsilon_{1,k} \epsilon_{2,k} \}$. In Appendix D.2 we show that the optimal choice of $d_k$ for the proposed DP scheme is such that the instantaneous average transmission power obeys $P_k = P$. This results in the following expression for $d_k$:

\[
d_k = \sqrt{\frac{P}{2(\alpha_k + m_k r_k)}}.
\]

(38)

Finally, similarly to Section 3, we initialize the scheme by setting $\epsilon_{i,0} = -S_i, \alpha_0 = \sigma_s^2$ and $r_0 = \rho_s \sigma_s^2$.

Our objective is to minimize the MSE after $K$ channel uses, over all possible vectors of estimation coefficients $\mathbf{b} = [b_1, b_2, \ldots, b_K] \in \mathcal{R}^K$, and over all possible vectors of modulation coefficients $\mathbf{m} = [m_0, m_1, \ldots, m_{K-1}] \in \{1, -1\}^K$. We denote this minimal MSE by $D_{K,\text{min}}$:

\[
D_{K,\text{min}} = \min_{\mathbf{b} \in \mathcal{R}^K, \mathbf{m} \in \{1, -1\}^K} \alpha_K.
\]

(39)

As the joint minimization in (39) is computationally very intensive, we define $\alpha_{K,\text{min}}(\mathbf{m})$ to be the minimal achievable MSE after $K$ channel uses, given a specific modulation vector $\mathbf{m}$:

\[
\alpha_{K,\text{min}}(\mathbf{m}) = \min_{\mathbf{b} \in \mathcal{R}^K} \alpha_K.
\]

(40)

We use DP to calculate $\alpha_{K,\text{min}}(\mathbf{m})$, thereby arriving at the optimization problem:

\[
D_{K,\text{min}} = \min_{\mathbf{m} \in \{1, -1\}^K} \alpha_{K,\text{min}}(\mathbf{m}),
\]

(41)

which can be solved by searching over the possible $2^K$ modulation vectors. In Remark 12 we comment on the practical implementation of this search. In the sequel we denote by the DP scheme the transmission scheme (36)–(37) which uses the optimal $\mathbf{b}$ and $\mathbf{m}$. Next, we present the algorithm for finding the minimizing $\mathbf{b}$ and the minimal $\alpha_{K,\text{min}}(\mathbf{m})$ for a given $\mathbf{m}$.

### 5.2 The Minimizing $\mathbf{b}$ and the Minimal $\alpha_{K,\text{min}}(\mathbf{m})$

Let $\mathbf{m}$ be a given modulation vector. Then, (1), (37) and (38) imply that $\alpha_k$ and $r_k$ are given by (see Appendix D.1 for the details):

\[
\begin{align*}
\alpha_k &= \alpha_{k-1} + b_k^2 \cdot (P + \sigma_s^2) - b_k \sqrt{2P(\alpha_{k-1} + m_{k-1} r_{k-1})} \quad (42a) \\
r_k &= r_{k-1} + b_k^2 m_{k-1} \cdot (P + \rho_s \sigma_s^2) - b_k m_{k-1} \sqrt{2P(\alpha_{k-1} + m_{k-1} r_{k-1})} \quad (42b)
\end{align*}
\]

Therefore, $(\alpha_{k-1}, r_{k-1})$ can be treated as a state variable, which, given $b_k$ and $\mathbf{m}$, evolves deterministically at time $k$. Thus, finding $\alpha_{K,\text{min}}(\mathbf{m})$ can be cast as a DP with state $(\alpha_{k-1}, r_{k-1})$, actions $b_k$, and cost function $\alpha_k$. Note that with this formulation, given $\mathbf{m}$, $b_k$ is a function of the constants $P, \sigma_s^2, \rho_s$, and of $(\alpha_{k-1}, r_{k-1})$. Hence, the last action $b_K$ is the linear MMSE estimation coefficient for estimating $\epsilon_{1,K-1}$ from $Y_{1,K}$.

Finally, the DP solution [7, Ch. 1.3] implies that $\alpha_k$ can be written as $\alpha_k = \eta_k \alpha_{k-1} + \theta_k m_{k-1} r_{k-1}$, where the sequences $\eta_k$ and $\theta_k, k = 1, 2, \ldots, K - 1$, are obtained using backwards recursion (in time). The minimizing $\mathbf{b}$ and the sequences $\eta_k$ and $\theta_k$ are given in the following theorem:

---

\[\text{Note that since } \epsilon_{1,k-1} \text{ and } Y_{1,k} \text{ are jointly Gaussian, then in this case the linear MMSE is the full MMSE.}\]
Theorem 6. For a fixed $\mathbf{m}$, the sequences $\eta_k$ and $\theta_k$, $k = 1, 2, \ldots, K - 1$, are defined through the backwards recursion (in time):

\[
\eta_{k-1} = \eta_k - \frac{P (\eta_k + \theta_k m_k m_{k-1})^2}{2 (\eta_k (P + \sigma_z^2) + \theta_k m_k m_{k-1} (P + \rho_z \sigma_z^2))}
\]

\[
\theta_{k-1} = \theta_k m_k m_{k-1} - \frac{P (\eta_k + \theta_k m_k m_{k-1})^2}{2 (\eta_k (P + \sigma_z^2) + \theta_k m_k m_{k-1} (P + \rho_z \sigma_z^2))},
\]

where $\eta_{K-1} = \left(1 - \frac{P}{2 (P + \sigma_z^2)}\right)$ and $\theta_{K-1} = - \frac{P}{2 (P + \sigma_z^2)}$. Furthermore, the coefficients $b_k, k = 1, 2, \ldots, K$, are given by:

\[
b_k = \begin{cases} 
\sqrt{\frac{P (\alpha_{k-1} + m_{k-1} r_{k-1})}{2 (\eta_k (P + \sigma_z^2) + \theta_k m_k m_{k-1} (P + \rho_z \sigma_z^2))}}, & k = 1, 2, \ldots, K - 1 \\
\sqrt{\frac{P (\alpha_{K-1} + m_{K-1} r_{K-1})}{2 (P + \sigma_z^2) \rho_z}}, & k = K.
\end{cases}
\]

The corresponding MSE at time $K$ is the minimal MSE given $\mathbf{m}$.

Proof. The proof is provided in Appendix D.1. \qed

Thm. 6 can be used for calculating the optimal $\mathbf{b}$ for a given $\mathbf{m}$. The procedure is summarized in Alg. 1:

**Algorithm 1** Calculating the Minimizing $\mathbf{b}$ and $\alpha_{K, \text{min}}(\mathbf{m})$

1: Initialization: $\eta_{K-1} \leftarrow \left(1 - \frac{P}{2 (P + \sigma_z^2)}\right), \theta_{K-1} \leftarrow - \frac{P}{2 (P + \sigma_z^2)}$
2: Compute the sequences $\eta_k$ and $\theta_k$ using the backwards recursions (43)
3: $\alpha_0 \leftarrow \sigma_z^2, r_0 \leftarrow \rho_z$
4: for $k = 1, 2, \ldots, K$ do
5: Calculate $b_k$ as in Thm. 6
6: Calculate $\alpha_k, r_k$ via (42)
7: end for
8: Output: $\mathbf{b}, \alpha_{K, \text{min}}(\mathbf{m})$

Remark 11. As we aim at minimizing $\alpha_k$ for a given $\mathbf{m}$, then $b_k$ is the MMSE estimation coefficient for estimating $\epsilon_{1, K-1}$ from $Y_{1, K}$, given $\mathbf{m}$. On the other hand, for $k < K$, using the MMSE estimation coefficient is not necessarily optimal as the $b_k$'s affect the future time indices. With this observation, it is clear why the OL scheme, which applies the MMSE estimator for all $k$'s, is not optimal, even among the memoryless linear transmission schemes.

Remark 12. Note that any choice of $\mathbf{m}$ will result in an upper bound on $D_{K, \text{min}}$. While finding $D_{K, \text{min}}$ requires searching over all $2^K$ possible $\mathbf{m}$ sequences, in practice, the search can be shortened at the expense of possibly achieving a larger MSE. Motivated by the alternating sign of $\rho_k$ in the OL and LQG schemes, for $k \to \infty$, (see [6, Eqs. (23), (36)-(37)]), we can choose $\mathbf{m}$ to be a sequence with alternating signs after some $L \ll K$ channel uses, thereby searching only over the first possible $2^L$ sequences. Numerical simulations show that when the SNR is not too low, then this approach performs well, as shown in the following section.

6 Comparisons and Numerical Examples

In this section we compare the different transmission schemes, i.e., OL, LQG and DP, and demonstrate our results via numerical examples.
6.1 DP Outperforms OL and LQG

The following proposition states that the DP scheme outperforms both the OL and the LQG schemes:

Proposition 4. For any fixed number of channel uses $K$, the DP scheme achieves MSE at least as low as the MSEs achieved by the OL and LQG schemes.

Proof outline. As stated in Remark 11, DP outperforms OL. Now, recall that choosing $P_k = P$ in the DP scheme is optimal. Thus, the DP scheme is the optimal scheme (in the sense of minimizing the MSE after $K$ channel uses) among the class of schemes which can be formulated via (36)–(37), and satisfy the constraint $P_k \le P$. In Appendix E we show that the LQG scheme can be written in the form of (36)–(37). Furthermore, we show that all three LQG decoders, (19), (21), and (32), have the same structure as the decoder applied by the DP scheme. We conclude that any LQG code which satisfies the per-symbol average power constraint (5) is within the search range of the DP scheme, and therefore DP achieves MSE at least as low as LQG. \[ \Box \]

6.2 Numerical Examples

Consider the transmission of a pair of Gaussian sources with $\sigma_s^2 = 1$ and $\rho_s = 0.4$, over a GBCF, and let $\sigma_z^2 = 1.5$, $\rho_z = 0.3$, and $P = 1$. Fig. 4 depicts the MSE values corresponding to (10), (32) and the approximation of (41) described in Remark 12 with $L = 15$, for the scenario considered above. Our simulations indicate that setting $L = 15$ does not result in any difference in the MSE compared to the exact solution of (41).\[ 12 \]

\[ 12 \]This was verified for $15 \le K \le 30$.

\[ 13 \]In [2, Sec. VI] we used the fact that for independent noises the LQG scheme is linear-FB sum-rate optimal to argue that for $\rho_z = 0$, and for $K$ large enough, the slope corresponding to the DP scheme is the same as the slope corresponding to the LQG scheme.\[ 13 \]

\[ 20 \]
As stated above, setting $L = 15$, does not degrade the performance of the DP scheme for the scenario considered in Fig. 4. However, in the low SNR regime the sequence $m$ starts alternating only after a relatively large $L$. In such a case, choosing $L$ too small may result in MSEs higher than the MSEs obtained by the OL scheme. This is demonstrated in Fig. 5 which considers the same scenario as Fig. 4 with $P = 0.03$. Note that when $L = 15$, OL outperforms DP. On the other hand, when $L = 25$, DP outperforms OL. Our simulations indicate that the optimal $m$ sequence start alternating for $L \geq 25$. Finally, note that Fig. 5 also depicts the MSEs for the non-scaled LQG scheme and for the LQG scheme with the original zero trajectory (ZT) decoder, see (19). It can be observed that for low values of $K$ the improved decoder in (21) significantly improves upon the ZT decoder, while for large values of $K$ the two decoders achieve approximately the same MSE. Furthermore, Fig. 5 shows that scaling can significantly improve the performance in the low SNR regime.

6.3 When does OL Outperform LQG?

Recall that in the infinite horizon regime LQG outperforms OL. On the other hand, Figs. 4 and 5 indicate that in the finite horizon regime OL can outperform LQG. This leads to the question: When does OL outperform LQG? To answer this question we focus on the symmetric setting, and note that Figs. 4 and 5 imply that the answer changes for different target MSEs. More precisely, using Thm. 1 and Thm. 5 one can answer the question which scheme (OL or LQG) achieves the target MSE with the least number of channel uses. For instance, consider the scenario illustrated in Fig. 5, i.e., $\sigma_s^2 = 1, \rho_s = 0.4, \sigma_z^2 = 1.5, P = 0.03$ and let $D = 10^{-2}$. Here, $K_{\text{LQG}} = 498$ and $K_{\text{OL}} = 470$. Thus, OL outperforms LQG. In fact, using the upper bound presented in Thm. 1, it can be shown that for $D = 10^{-2}$, OL outperforms LQG for all $P < 0.1978$.

7 Conclusions

In this work we studied the transmission of a pair of correlated Gaussian sources over the two-user GBCF focusing on linear and memoryless transmission schemes in the finite horizon regime. We focused on analyzing the minimal number of channel uses required to achieve a non-zero target pair of MSEs. We considered three finite horizon transmission schemes: An adaptation of the OL scheme of [5] to the transmission of a pair of correlated Gaussian sources, an adaptation of the LQG scheme of [6] to the transmission of a pair of correlated Gaussian sources, and a novel scheme derived in this work designed using the DP approach.

For the OL scheme we presented upper and lower bounds on the number of channel uses required to achieve a target pair of MSEs. Then, for the symmetric setting with indepen-
dent sources and independent noise components, we showed that even though OL applies uncoded transmission, in the low SNR regime, it achieves approximately the same source-channel bandwidth ratio as the best known separation-based scheme which applies source and channel coding with an asymptotically large blocklength. More precisely, the gap between the source-channel bandwidth ratio achieved by the OL scheme and the source-channel bandwidth ratio achieved by the separation-based scheme is bounded, where this bound is independent of the SNR.

We adopted the LQG scheme to the finite horizon regime and introduced a new decoder based on the MMSE criterion, which achieves MSE values smaller than or equal to those achieved by the original decoder proposed in [6]. For the general setting, we presented lower and upper bounds on the number of channel uses required to achieve a target pair of MSEs, while for the symmetric setting we explicitly characterized this number of channel uses. In the LQG scheme the instantaneous average transmission power varies over time. This implies that: 1) Some sources must be scaled before transmission in order to satisfy the average per-symbol power constraint; and 2) By scaling the sources before transmission one can use the available transmission power more efficiently. For the symmetric setting we explicitly characterized the optimal scaling factor and showed that this scaling is beneficial in reducing the MSE even in cases for which the average per-symbol power constraint is satisfied without scaling. Thus, by transmitting the optimally scaled sources, we obtained a linear time-invariant transmission scheme with good finite horizon performance.

Lastly, we used DP to derive the optimal linear and memoryless transmission scheme in the symmetric setting. This scheme requires finding a vector of modulation coefficients and a vector of estimation coefficients which minimize the MSE after $K$ channel uses. We showed that this minimization problem can be simplified into the problem of searching only over the possible modulation vectors, while the optimal vector of estimation coefficients, per modulation vector, was formulated as a DP problem whose solution is obtained using a recursive deterministic relationship. For a finite number of channel uses, the DP scheme achieves an MSE lower than both OL and LQG. As finding the optimal modulation vector becomes computationally infeasible as the number of channel uses becomes large, we also proposed an approximate solution which is computationally feasible, and performs well when the SNR is not very low.

The comparisons of the three schemes indicate that they are fundamentally different: The OL scheme is time-varying, its performance analysis is quite complicated, it reaches steady state relatively quickly, but it is suboptimal in the infinite horizon regime. The LQG scheme is time-invariant, its performance analysis is less complicated, it reaches steady state relatively slowly, and in the infinite horizon regime it is the best known scheme. By applying the proposed scaling and using the improved decoder the finite horizon performance of the LQG scheme can be significantly improved, e.g., Fig. 5. We emphasize that even though the LQG scheme achieves MSE exponents higher than those achieved by the OL scheme, even when scaling is applied along with using the improved decoder, in the finite horizon regime the OL scheme can outperform the LQG scheme. In fact, simulations indicate, e.g., Fig 4b, that the LQG scheme outperforms the OL scheme only for very low MSEs. Finally, the DP scheme is time-varying, its performance is highly complicated to analyze, and it outperforms both OL and LQG in the finite horizon regime. As the DP scheme applies a backwards recursion it can be applied only for the finite horizon regime.

The results presented in this work are important in identifying efficient and simple coding schemes for the transmission of correlated Gaussian sources over multiuser channels with FB when strict delay constraints are imposed.

---

14Recall that for the considered setting source-channel separation is optimal [31, Thm. 2].
A OL in the Finite Horizon Regime - Proofs

A.1 Proof of Theorem 1

Recall that \( \alpha_{i,0} = \sigma_i^2, i = 1, 2 \), and that \( \alpha_{i,k} \) is the MSE at Rx \( i \) after the \( k \)th transmission. From (10) we have:

\[
\log\left( \frac{\alpha_{1,K}}{\sigma_1^2} \right) = \sum_{k=1}^{K} \log\left( \frac{\sigma_{1,1}^2 + \Psi_{1,k-1}^2 g^2 (1 - \rho_{k-1}^2)}{\sigma_1^2} \right).
\]

As \( |\rho_k| \in [0, 1] \), it follows that:

\[
\Psi_{1,k-1}^2 g^2 (1 - \rho_{k-1}^2) = \frac{P g^2 (1 - \rho_{k-1}^2)}{1 + g^2 + 2g|\rho_{k-1}^2 - 1|} \leq \frac{P g^2}{1 + g^2}.
\]

Thus, we obtain the upper bound \( \frac{\sigma_{1,1}^2 + \pi_1 g^2}{\pi_1} \leq \frac{\sigma_{1,1}^2 + \pi_1 g^2}{\pi_1} \). Next, we use the fact that \( \log(x) \leq x - 1 \) and write:

\[
\log\left( \frac{\sigma_{1,1}^2 + \pi_1 g^2}{\pi_1 + \pi_1 g^2} \right) \leq \frac{\sigma_{1,1}^2 + \pi_1 g^2}{\pi_1 + \pi_1 g^2} - 1 = -\frac{P}{\pi_1 + \pi_1 g^2}.
\]

Thus, it follows that \( \log\left( \frac{\alpha_{1,K}}{\sigma_1^2} \right) = \log\left( \frac{D_1^1}{\sigma_1^2} \right) \leq -\frac{K P}{\pi_1 + \pi_1 g^2} \), which implies that:

\[
K_{OL}^{ub} = \left[ \frac{(1 + g^2)}{P} \max \left\{ \pi_1 \log\left( \frac{\sigma_{1,1}^2}{D_1^1} \right), \frac{\pi_2}{g} \log\left( \frac{\sigma_{1,2}^2}{D_2^2} \right) \right\} \right].
\]

To obtain \( K_{OL}^{lb} \), we note that \( 0 \leq \Psi_{1,k-1}^2 g^2 (1 - \rho_{k-1}^2) \) where equality is obtained by setting \( \rho_{k-1} = 1 \). Then, we use the inequality \( 1 - \frac{1}{2} \leq \log x \) to obtain:

\[
\log\left( \frac{\sigma_{1,1}^2}{\sigma_{1,1}^2 + P} \right) \geq \frac{\sigma_{1,1}^2 + P}{\sigma_{1,1}^2} = -\frac{P}{\sigma_{1,1}^2}.
\]

Thus, we have \( \log\left( \frac{D_1^1}{\sigma_1^2} \right) \geq -\frac{K P}{\sigma_{1,1}^2} \), which results in the following lower bound:

\[
K_{OL}^{lb} = \left[ \max \left\{ \frac{\sigma_{1,1}^2}{P} \log\left( \frac{\sigma_{1,1}^2}{D_1^1} \right), \frac{\sigma_{1,2}^2}{P} \log\left( \frac{\sigma_{1,2}^2}{D_2^2} \right) \right\} \right].
\]

A.2 Proof of (13)

From [5, pg. 671] it follows that if \( R \) is an achievable symmetric rate for the GBCF, and \( \rho_2 = 0 \), then \( R < \frac{1}{2} \log \left( 1 + \frac{2\chi_0 P}{\sigma_1^2} \right) \), where \( \chi_0 \) is the unique positive root of the polynomial (in \( \chi \)):

\[
\chi^2 + \frac{3\sigma_1^2}{2SNR} \chi - \frac{\sigma_1^2}{2SNR} = \chi^2 + \frac{3}{2SNR} = \frac{1}{2SNR}.
\]

The roots of this polynomial are given by:

\[
\chi_{1,2} = \frac{1}{2} \left( -\frac{3}{2SNR} \pm \sqrt{\frac{9}{4SNR^2} + \frac{2}{SNR}} \right).
\]

Hence, \( \chi_0 \) is given by \( \chi_0 = \frac{1}{2} \left( -\frac{3}{2SNR} + \sqrt{\frac{9}{4SNR^2} + \frac{2}{SNR}} \right) \). Plugging \( \chi_0 \) into the upper bound on \( R \) we write:

\[
R < \frac{1}{2} \log \left( 1 + 2\chi_0 SNR \right)
\]

\[
= \frac{1}{2} \log \left( 1 + SNR \left( -\frac{3}{2SNR} + \sqrt{\frac{9}{4SNR^2} + \frac{2}{SNR}} \right) \right)
\]

\[
= \frac{1}{2} \log \left( \sqrt{\frac{9}{4} + 2SNR} - \frac{1}{2} \right).
\]
A.3 Proof of (14a)

First, we obtain an upper bound on $2\log |a_1|$. Following steps similar to those described in [6, Section IV.C] for the symmetric GBCF with independent noises, we conclude that $a_1^2 = x_0$, where $x_0$ is the unique real positive root\(^{15}\) of the equation:

$$a_1^2x^3 + a_2^2x^2 - (a_3^2 + 2P)x - a_4^2 = 0.$$  

Rewriting this equation equivalently as:

$$x^3 + x^2 - \left(1 + \frac{2P}{\sigma_z^2}\right)x - 1 = 0,$$  

we upper bound $x_0$ using Budan’s theorem [33]:

**Theorem.** (Budan’s theorem) Let $p(x) = a_0 + a_1x + \cdots + a_nx^n$ be a polynomial of degree $n$, and let $p^{(j)}(x)$ be its $j$’th derivative. Define the function $V(\alpha)$ as the number of sign variations in the sequence $p(\alpha), p^{(1)}(\alpha), \ldots, p^{(n)}(\alpha)$. Then, the number of roots of the polynomial $p(x)$ in the open interval $(a, b)$ is either equal to $V(a) - V(b)$, or less by an even number.

Let $p(x)$ be the polynomial in (A.1). Then we have:

$$p^{(0)}(x) = x^3 + x^2 - \left(1 + \frac{2P}{\sigma_z^2}\right)x - 1,$$  

$$p^{(1)}(x) = 3x^2 + 2x - \left(1 + \frac{2P}{\sigma_z^2}\right),$$  

$$p^{(2)}(x) = 6x + 2,$$  

$$p^{(3)}(x) = 6.$$  

For $x = 1$ we have $V(1) = 1$. Note that $\text{sgn}(p^{(0)}(1))$ depends on the term $\frac{2P}{\sigma_z^2}$, however, since $\text{sgn}(p^{(0)}(1)) = -1$ and $\text{sgn}(p^{(2)}(1)) = 1$, in both cases we have $V(1) = 1$. Next, we let $\chi = \frac{P}{2\sigma_z^2}$, and set $x = 1 + \chi$ to obtain:

$$p^{(0)}(1+\chi) = \chi^3, \quad p^{(1)}(1+\chi) = 3\chi^2 + 4\chi + 4, \quad p^{(2)}(1+\chi) = 6\chi + 8, \quad p^{(3)}(1+\chi) = 6,$$

all larger than zero. Therefore, $V(1 + \chi) = 0$. Thus, Budan’s theorem implies that the number of roots of (A.1) in the interval $(1, 1 + \chi)$ is 1. From Descartes’ rule we know that there is a unique positive root, therefore $1 + \chi$ is an upper bound on $x_0$: $x_0 < 1 + \frac{P}{2\sigma_z^2}$.

Next, recall that $a_1^2 = x_0$, which implies that $2\log(|a_1|) = \log(x_0) \leq \log \left(1 + \frac{P}{2\sigma_z^2}\right)$. Using the fact that $\log(x) \leq x - 1$ we have the following bound on $2\log(|a_1|)$:

$$2\log(|a_1|) \leq \frac{P}{2\sigma_z^2}.$$  

(A.3)

Next, we explicitly upper bound $K_{OL} - \kappa_{sep}$ in the symmetric setting (we set $g = 1$ in (12a)):

$$K_{OL} - \kappa_{sep} \leq K_{OL} - \kappa_{sep} \frac{2(P + \sigma_z^2)}{P}\log \left(\frac{\sigma_z^2}{D}\right) = \frac{1}{2\log |a_1|}\log \left(\frac{\sigma_z^2}{D}\right)$$

$$\leq \log \left(\frac{\sigma_z^2}{D}\right) \left(\frac{2(P + \sigma_z^2)}{P} - \frac{2\sigma_z^2}{P}\right)$$

$$\leq \frac{2\log \left(\frac{\sigma_z^2}{D}\right)}{2\log (|a_1|)},$$  

(A.4)

where (a) follows from specializing Thm. 1 to the symmetric setting, and (b) follows from the bound $2\log(|a_1|) \leq \frac{P}{2\sigma_z^2}$.

---

\(^{15}\)The uniqueness of a real positive root follows from Descartes’ rule [34, Subsection 1.6.3.2].
A.4 Proof of (14b)

Recall that \( \kappa_{\text{sep}}^{ib} \triangleq \frac{\log \left( \frac{\sigma^2_s}{D} \right)}{\log \left( \sqrt{\frac{9}{4} + 2\text{SNR} - \frac{1}{2}} \right)} \). Thus, we write:

\[
K_{\text{OL}} - \kappa_{\text{sep}}^{ib} \leq K_{\text{OL}}^{ib} - \kappa_{\text{sep}}^{ib} = 2 \left( \frac{P + \sigma^2_z}{P} \right) \log \left( \frac{\sigma^2_s}{D} \right) - \frac{1}{\log \left( \sqrt{\frac{9}{4} + 2\text{SNR} - \frac{1}{2}} \right)} \log \left( \frac{\sigma^2_s}{D} \right)
\]

\[
= \log \left( \frac{\sigma^2_s}{D} \right) \left( 2 + \frac{2}{\text{SNR}} - \frac{1}{\sqrt{2\text{SNR}}} \right)
\]

\[
\leq \left[ \left( 2 + \frac{2}{\sqrt{2\text{SNR}}} \right) \log \left( \frac{\sigma^2_s}{D} \right) \right],
\]

where (a) follows from the fact that \( \log \left( \sqrt{\frac{9}{4} + 2\text{SNR} - \frac{1}{2}} \right) \leq \sqrt{\frac{9}{4} + 2\text{SNR} - \frac{1}{2}} \leq \sqrt{2\text{SNR}} \).

B LQG in the Finite Horizon Regime - Proofs

B.1 Proof of Theorem 2

The MMSE estimator of \( S_i \) based on \( \hat{U}_{i,k} \) is the conditional expectation \( E\{ S_i | \hat{U}_{i,k} \} \), [28, Eqn. (11.10)]. Now, from (15) we can write:

\[
U_k = A U_{k-1} + Y_{k-1}
\]

\[
= A U_{k-1} - B C^T U_{k-1} + Z_{k-1}
\]

\[
= (A - B C^T) U_{k-1} + Z_{k-1}, \quad \text{(B.1)}
\]

and from (18) we have:

\[
\hat{U}_k = A \hat{U}_{k-1} + Y_{k-1}
\]

\[
= A^{k-1} \hat{U}_1 + \sum_{m=1}^{k-1} A^{k-m-1} Y_m
\]

\[
= \sum_{m=1}^{k-1} A^{k-m-1} (-B C^T U_{k-1} + Z_{k-1}). \quad \text{(B.2)}
\]

From the fact that \( Z_k \) is a zero-mean Gaussian vector, from the linear relationship in (B.2), and from the fact that \( U_1 = S \), it follows that for \( i = 1, 2, \hat{U}_{i,k+1} \) and \( S_i \) are jointly Gaussian, both with zero mean. From [28, Eqn. (10.16)] it follows that \( E\{ S_i | \hat{U}_{i,k+1} \} = \frac{E\{ S, \hat{U}_{i,k+1} \}}{E\{ \hat{U}_{i,k+1} \}} \hat{U}_{i,k+1} \). Next, we expand (15) as:

\[
U_k = A U_{k-1} + Y_{k-1}
\]

\[
= A^{k-1} S + \sum_{m=1}^{k-1} A^{k-m-1} Y_m. \quad \text{(B.3)}
\]

Therefore, combining (B.3) and (B.2) we have \( U_{k+1} - \hat{U}_{k+1} = A^k S \Rightarrow U_{k+1} = U_{k+1} - A^k S \), and since \( A \) is a diagonal matrix it follow that \( \hat{U}_{i,k+1} = U_{i,k+1} - a_i^k S_i \). At time \( k + 1 \), the
MMSE estimate of $S_i$ based on $\hat{U}_{i,k+1}$ is given by:

$$
\hat{S}_{i,k} = \frac{E\{S_i(U_{i,k+1} - a_k^i S_i)\}}{E\{(U_{i,k+1} - a_k^i S_i)^2\}} \hat{U}_{i,k+1}
$$

$$
= \frac{E\{S_i U_{i,k+1}\} - a_k^i \sigma_i^2}{E\{U_{i,k+1}^2\} - 2a_k^i E\{S_i U_{i,k+1}\} + a_k^i \sigma_i^2} \hat{U}_{i,k+1}.
$$

(B.4)

From the independence of $S$ and $Z_k$ we have $E\{U_{k+1}S^T\} = (A - BC^T)E\{U_k S^T\}$, and since $U_1 = S$ it follows that $E\{U_{k+1}S^T\} = (A - BC^T)^k S$. Recalling the definition $M \triangleq A - BC^T$ we conclude that:

$$
E\{S_i U_{i,k+1}\} = [M^k Q_s]_{i,i}.
$$

(B.5)

Using the definition of $Q_{u,k}$ in Subsection 4.2 and plugging (B.5) into (B.4) we obtain (21). Next, we use (21) to obtain a recursive expression for the MSE. By plugging the expression for $\hat{S}_{i,k}$ in (B.4) into $E\{(S_i - \hat{S}_{i,k})^2\}$ we obtain that:

$$
E\{(S_i - \hat{S}_{i,k})^2\} = \sigma_i^2 - \frac{(\sigma_i^2 a_k^i)^2}{[Q_{u,k+1}]_{i,i} - 2a_k^i [M^k Q_s]_{i,i} + \sigma_i^2 a_k^{2i}}
$$

$$
= \frac{\sigma_i^2 [Q_{u,k+1}]_{i,i} - (\sigma_i^2 a_k^i)^2}{[Q_{u,k+1}]_{i,i} - 2a_k^i [M^k Q_s]_{i,i} + \sigma_i^2 a_k^{2i}}
$$

(B.6)

which is Eqn. (22). Finally, we consider $\sigma_i^2 a_k^{2i}$ for $k \to \infty$. As the magnitudes of eigenvalues of the matrix $M$ are smaller than unity it follows that $\lim_{k \to \infty} ([M^k Q_s]_{i,i})^2 = 0$ and $\lim_{k \to \infty} [M^k Q_s]_{i,i} = 0$. Furthermore, since $|a_i| > 1$ and since $\lim_{k \to \infty} Q_{u,k} = Q_u$ it follows that:

$$
\lim_{k \to \infty} a_k^{2i} \left( \frac{[Q_{u,k+1}]_{i,i} - 2 [M^k Q_s]_{i,i} + \sigma_i^2}{[Q_{u,k+1}]_{i,i} - 2a_k^i [M^k Q_s]_{i,i} + \sigma_i^2 a_k^{2i}} \right) \approx a_k^{2i} [Q_{u,k+1}]_{i,i}.
$$

Therefore, for $k$ large enough we have:

$$
\frac{\sigma_i^2 [Q_{u,k+1}]_{i,i} - (\sigma_i^2 a_k^i)^2}{[Q_{u,k+1}]_{i,i} - 2a_k^i [M^k Q_s]_{i,i} + \sigma_i^2 a_k^{2i}} \approx a_k^{-2i} [Q_{u,k+1}]_{i,i}
$$

$$
= a_k^{-2i} E\{U_{i,k+1}^2\}.
$$

(B.7)

### B.2 Proof of Proposition 2

We begin with explicitly writing $P_k$ using $U_k$:

$$
P_k = E\{X_k^2\} = E\{C^T U_k U_k^T C\} = C^T E\{U_k U_k^T\} C,
$$

where (a) follows from the structure of the controller. Now, recalling that $M = (A - BC^T)$, we use (B.1) and the fact that $U_k$ and $Z_k$ are independent and write:

$$
E\{U_k U_k^T\} = ME\{U_{k-1} U_{k-1}^T\} M^T + Q_z
$$

$$
= M (ME\{U_{k-2} U_{k-2}^T\} M^T + MQ_z M^T + Q_z
$$

$$
= M^{k-1} Q_s (M^T)^{k-1} + \sum_{l=0}^{k-2} M^l Q_s (M^T)^l.
$$

(B.7)

Therefore, we have:

$$
P_k = C^T E\{U_k U_k^T\} C
$$

$$
= C^T M^{k-1} Q_s (M^T)^{k-1} C + \sum_{l=0}^{k-2} C^T M^l Q_s (M^T)^l C.
$$

(B.8)
Therefore, it follows that:

Next, we define $R \triangleq VD^kV^{-1}$, where $V$ is a function of $k$, yet, to reduce clutter we omit this notation. This implies that:

$$C^T M^k Q_s (M^T)^k C = C^T R R^T C$$

(B.10)

Writing $VD^kV^{-1}$ explicitly we have:

$$VD^kV^{-1} = \left[ \begin{array}{c} v_1 \\ v_2 \\ v_3 \\ v_4 \end{array} \right] \left[ \begin{array}{cc} \lambda^k_1 & 0 \\ 0 & \lambda^k_2 \end{array} \right] \left[ \begin{array}{c} v_1 \\ v_2 \\ v_3 \\ v_4 \end{array} \right]^{-1}$$

$$= \frac{1}{\det(V)} \left[ \begin{array}{cc} v_1 v_4 \omega^k_1 - v_2 v_3 \omega^k_2 & v_1 v_2 (\lambda^k_2 - \lambda^k_1) \\ v_3 v_4 (\lambda^k_1 - \lambda^k_2) & v_1 v_4 \lambda^k_2 - v_2 v_3 \lambda^k_1 \end{array} \right].$$

(B.11)

Therefore, it follows that:

$$R = VD^kV^{-1}L = \frac{1}{\det(V)} \left[ \begin{array}{cc} v_1 v_4 \omega^k_1 - v_2 v_3 \omega^k_2 & v_1 v_2 (\lambda^k_2 - \lambda^k_1) \\ v_3 v_4 (\lambda^k_1 - \lambda^k_2) & v_1 v_4 \lambda^k_2 - v_2 v_3 \lambda^k_1 \end{array} \right] \left[ \begin{array}{c} \sigma_1 \\ \sigma_2 \sqrt{1 - \rho^2_s} \end{array} \right],$$

which implies that:

$$r_1 = \frac{1}{\det(V)} \left( \sigma_1 (v_1 v_4 \omega^k_1 - v_2 v_3 \omega^k_2) + \rho_s \sigma_2 v_1 v_2 (\lambda^k_2 - \lambda^k_1) \right)$$

(B.12a)

$$r_2 = \frac{1}{\det(V)} \sigma_2 \sqrt{1 - \rho^2_s} \cdot v_1 v_2 \cdot (\lambda^k_2 - \lambda^k_1)$$

(B.12b)

$$r_3 = \frac{1}{\det(V)} (\sigma_1 v_3 v_4 (\lambda^k_1 - \lambda^k_2) + \rho_s \sigma_2 (v_1 v_4 \lambda^k_2 - v_2 v_3 \lambda^k_1))$$

(B.12c)

$$r_4 = \frac{1}{\det(V)} \sigma_2 \sqrt{1 - \rho^2_s} (v_1 v_4 \lambda^k_2 - v_2 v_3 \lambda^k_1).$$

(B.12d)

Next, we explicitly write $c_1 r_1 + c_2 r_3$:

$$c_1 r_1 + c_2 r_3 = \frac{c_1}{\det(V)} (\sigma_1 (v_1 v_4 \lambda^k_1 - v_2 v_3 \lambda^k_2) + \rho_s \sigma_2 v_1 v_2 (\lambda^k_2 - \lambda^k_1))$$

$$+ \frac{c_2}{\det(V)} (\sigma_1 v_3 v_4 (\lambda^k_1 - \lambda^k_2) + \rho_s \sigma_2 (v_1 v_4 \lambda^k_2 - v_2 v_3 \lambda^k_1))$$

$$= \lambda^k_1 c_1 (\sigma_1 v_1 v_4 - \rho_s \sigma_2 v_1 v_2) + c_2 (\sigma_1 v_3 v_4 - \rho_s \sigma_2 v_3 v_4)$$

$$+ \lambda^k_2 c_1 (\rho_s \sigma_2 v_1 v_2 - \sigma_1 v_2 v_3) + c_2 (\rho_s \sigma_2 v_1 v_4 - \sigma_1 v_3 v_4)$$

$$= \lambda^k_1 \omega_1 (\sigma_1, \sigma_2, \rho_s) + \lambda^k_2 \omega_2 (\sigma_1, \sigma_2, \rho_s).$$

(B.13a)
Similarly, we explicitly write $c_1 r_2 + c_2 r_4$:

$$c_1 r_2 + c_2 r_4 = \lambda_1^4 - 2\sigma_2 \sqrt{1 - \rho_2^2 (c_1 v_2 + c_2 v_3)} + \lambda_2^4 \sigma_2 \sqrt{1 - \rho_2^2 (c_1 v_2 + c_2 v_4)}$$

$$= \lambda_1^4 \omega_3 (c_1, \sigma_2, \rho_3) + \lambda_2^4 \omega_4 (c_1, \sigma_2, \rho_4),$$  \hspace{1cm} (B.13b)

where $\omega_j(c_1, \sigma_2, \rho_3), j = 1, 2, 3, 4,$ are defined in (23). Hence, squaring (B.13a) and (B.13b), summing and using the expressions $\alpha_j, j = 1, 2, 3,$ defined in (24) we obtain:

$$(c_1 r_2 + c_2 r_3)^2 + (c_1 r_2 + c_2 r_4)^2$$

$$= \lambda_1^4 \omega_3 (c_1, \sigma_2, \rho_3) + \lambda_2^4 \omega_4 (c_1, \sigma_2, \rho_4))^2 + (\lambda_1^4 \omega_3 (c_1, \sigma_2, \rho_3) + \lambda_2^4 \omega_4 (c_1, \sigma_2, \rho_4))^2$$

$$= \lambda_1^4 \omega_3^2 (c_1, \sigma_2, \rho_3) + \lambda_2^4 \omega_4^2 (c_1, \sigma_2, \rho_4) + 2 \lambda_1^4 \lambda_2^4 \omega_3 (c_1, \sigma_2, \rho_3) \omega_4 (c_1, \sigma_2, \rho_4)$$

$$+ \lambda_1^4 \lambda_2^4 \omega_3 (c_1, \sigma_2, \rho_3) + \lambda_2^4 \omega_4 (c_1, \sigma_2, \rho_4) + 2 \lambda_1^4 \lambda_2^4 \omega_3 (c_1, \sigma_2, \rho_3) \omega_4 (c_1, \sigma_2, \rho_4)$$

$$= \lambda_1^4 \omega_3 (c_1, \sigma_2, \rho_3) + \lambda_2^4 \omega_4 (c_1, \sigma_2, \rho_4) + 2 \omega_3 (c_1, \sigma_2, \rho_3) \omega_4 (c_1, \sigma_2, \rho_4)$$

$$= \lambda_1^4 \lambda_2^4 \omega_3 (c_1, \sigma_2, \rho_3) + \lambda_2^4 \omega_4 (c_1, \sigma_2, \rho_4) + \lambda_1^4 \lambda_2^4 \omega_3 (c_1, \sigma_2, \rho_4).$$  \hspace{1cm} (B.14)

We conclude that:

$$C^T M^k Q_z (M^T)^k C = \lambda_1^4 \alpha_1 (c_1, \sigma_2, \rho_3) + \lambda_2^4 \alpha_2 (c_1, \sigma_2, \rho_4) + \lambda_1^4 \lambda_2^4 \alpha_3 (c_1, \sigma_2, \rho_3).$$  \hspace{1cm} (B.15)

Next, we focus on the second term in (B.8): $\sum_{i=0}^{k-2} C^T M^i Q_z (M^T)^i C$. Following identical steps to those leading to (B.15), and recalling that $|\rho_2| < 1$, we write:

$$C^T M^i Q_z (M^T)^i C = \lambda_1^4 \alpha_1 (c_1, \sigma_2, \rho_3) + \lambda_2^4 \alpha_2 (c_1, \sigma_2, \rho_4) + \lambda_1^4 \lambda_2^4 \alpha_3 (c_1, \sigma_2, \rho_3).$$

Therefore, summing over $l$ we obtain:

$$\sum_{i=0}^{k-2} C^T M^i Q_z (M^T)^i C$$

$$= \sum_{i=0}^{k-2} \lambda_1^4 \alpha_1 (c_1, \sigma_2, \rho_3) + \lambda_2^4 \alpha_2 (c_1, \sigma_2, \rho_4) + \lambda_1^4 \lambda_2^4 \alpha_3 (c_1, \sigma_2, \rho_3)$$

$$= 1 - \lambda_1^4 \alpha_1 (c_1, \sigma_2, \rho_3) + 1 - \lambda_2^4 \alpha_2 (c_1, \sigma_2, \rho_4)$$

$$\frac{1 - \lambda_1^4 \lambda_2^4 \alpha_3 (c_1, \sigma_2, \rho_3)}{1 - \lambda_1 \lambda_2}.$$  \hspace{1cm} (B.16)

Combining (B.15) and (B.16) and using the expressions $\eta_j(c_1, \sigma_2, \rho_3), j = 1, 2, 3,$ defined in (25), results in:

$$P_k = C^T M^{k-1} Q_z (M^T)^{k-1} C + \sum_{i=0}^{k-2} C^T M^i Q_z (M^T)^i C$$

$$= \lambda_1^4 \alpha_1 (c_1, \sigma_2, \rho_3) + \lambda_2^4 \alpha_2 (c_1, \sigma_2, \rho_4) + (\lambda_1 \lambda_2)^{k-1} \alpha_3 (c_1, \sigma_2, \rho_3)$$

$$+ \frac{1 - \lambda_1 \lambda_2}{1 - \lambda_1 \lambda_2} \alpha_1 (c_1, \sigma_2, \rho_3) + \frac{1 - \lambda_1 \lambda_2}{1 - \lambda_1 \lambda_2} \alpha_2 (c_1, \sigma_2, \rho_4)$$

$$+ \frac{1 - (\lambda_1 \lambda_2)^{k-1}}{1 - \lambda_1 \lambda_2} \alpha_3 (c_1, \sigma_2, \rho_3)$$

$$= \eta_1 (c_1, \sigma_2, \rho_3) + \eta_2 (c_1, \sigma_2, \rho_4) + \eta_3 (c_1, \sigma_2, \rho_3)$$

$$+ (\lambda_1 \lambda_2)^{k-1} \left(\alpha_1 (c_1, \sigma_2, \rho_3 - \eta_1 (c_1, \sigma_2, \rho_3))\right)$$

$$+ \lambda_1^4 \alpha_1 (c_1, \sigma_2, \rho_3) - \eta_2 (c_1, \sigma_2, \rho_4)$$

$$+ (\lambda_1 \lambda_2)^{k-1} \left(\alpha_3 (c_1, \sigma_2, \rho_3 - \eta_3 (c_1, \sigma_2, \rho_3))\right).$$  \hspace{1cm} (B.17)
Recall that the LQG scheme is designed such that the asymptotic average transmit power is $P$. This implies that:

$$
\eta_1(\sigma_{z,1}, \sigma_{z,2}, \rho_z) + \eta_2(\sigma_{z,1}, \sigma_{z,2}, \rho_z) + \eta_3(\sigma_{z,1}, \sigma_{z,2}, \rho_z) = P.
$$

For the power constraint in (5) to be satisfied for every $k = 1, 2, 3, \ldots$, we should have $P_k \leq P$. From (B.17) we conclude that this condition can be equivalently stated as follows:

$$\lambda_1^{2(k-1)}(\alpha_1(\sigma_{z,1}, \sigma_{z,2}, \rho_z) - \eta_1(\sigma_{z,1}, \sigma_{z,2}, \rho_z))
+ \lambda_2^{2(k-1)}(\alpha_2(\sigma_{z,1}, \sigma_{z,2}, \rho_z) - \eta_2(\sigma_{z,1}, \sigma_{z,2}, \rho_z))
+ (\lambda_1 \lambda_2)^{k-1}(\alpha_3(\sigma_{z,1}, \sigma_{z,2}, \rho_z) - \eta_3(\sigma_{z,1}, \sigma_{z,2}, \rho_z)) \leq 0.$$

### B.3 Proof of Theorem 3

We begin with $K_{LQG}^n$. Since (21) is the optimal estimator based on the observation $\hat{U}_{i,k+1}$, it follows that we can upper bound $K_{LQG}$ by upper bounding the number of channel uses required to achieve a target MSE pair using the decoder in (19). Recall that the MSE of the decoder in (19) is given by (20): $E\{(\hat{S}_i - \hat{S}_k)^2\} = a_i^{-2k}E\{U_{i,k+1}^2\}$. Let $E\{(\hat{S}_i - \hat{S}_k)^2\} \triangleq D_{i,k}$ be the MSE after $k$ channel uses, i.e., at time instance $k+1$. We upper bound $D_{i,k}$ via upper bounding $E\{U_{i,k+1}^2\}$.

Since the eigenvalues of $M$ are inside the unit circle, it follows that $[M^{k}Q_a(M^T)^k]_{i,i} \rightarrow 0$ as $k \rightarrow \infty$, and therefore, $\lim_{k \rightarrow \infty} \sum_{l=0}^{k-1} M^lQ_a(M^T)^l[i,i] = [Q_a]_{i,i}$. Since $Q_a$ is a covariance matrix then the diagonal elements of $M^{k}Q_a(M^T)^k$ are non-negative and we can write:

$$E\{U_{i,k+1}^2\} \leq [M^{k}Q_a(M^T)^k]_{i,i} + [Q_a]_{i,i}.$$

Next, we derive an upper bound on $[M^{k}Q_a(M^T)^k]_{i,i}$. Following the arguments leading to (B.10) we can write $M^{k}Q_a(M^T)^k = RR^T$, again omitting the dependence in $k$ from the matrix $R$, we write:

$$[M^{k}Q_a(M^T)^k]_{1,1} = r_1^2 + r_2^2,$$

where $r_1$ and $r_2$ are given in (B.12). For ease of reference we repeat the expressions for $r_1$ and $r_2$:

$$r_1 = \frac{1}{\det(V)} \left( \sigma_1(v_1v_2\lambda_1^k - v_2v_3\lambda_2^k) + \rho_s \sigma_2v_1v_2(\lambda_2^k - \lambda_1^k) \right),$$

$$r_2 = \frac{1}{\det(V)} \sigma_2 \sqrt{1 - \rho_s^2} \cdot v_1v_2 \cdot (\lambda_2^k - \lambda_1^k).$$

16Recall that $C = (B^TGB + 1)^{-1}AG^TB$ where $G$ is the unique positive-definite solution of the DARE (16). Now, from [30, Lemma 2.4, item (iv)] it follows that the eigenvalues of the closed-loop matrix $M = A - BC^T$ are given by $\lambda_i = \frac{1}{V_i}$. Note that [30, Lemma 2.4] assumes a DARE of the form (16) and studies the properties of the matrix $A - BC^T$, for $C = (B^TGB + 1)^{-1}AG^TB$, see [30, Equation below (11)]. Therefore, it follows that $\lambda_1 = \frac{1}{V_1}$.

17Note that $\sum_{l=0}^{k-1} [M^{l}Q_a(M^T)^l]_{i,i} \geq 0, k = 1, 2, \ldots, i = 1, 2$, since the diagonal elements are sum of the variances of the noise.
Next, we upper bound \( [M^k Q\sigma(M^T)^k]_{1,1} \) via upper bounding \( r_1^2 \) and \( r_2^2 \):

\[
|r_1| \leq \frac{1}{\lvert \text{det}(V) \rvert} \left( \sigma_1 (\lvert v_1 v_4 \rvert \lvert \lambda_1 \rvert^k + \lvert v_2 v_3 \rvert \lvert \lambda_2 \rvert^k) + \rho_1 \cdot \sigma_2 \cdot \lvert v_1 v_2 \rvert (\lvert \lambda_1 \rvert^k + \lvert \lambda_1 \rvert^k) \right)
\]

\[
\leq \sigma_1 (\lvert v_1 v_4 \rvert \lvert \lambda_1 \rvert + \lvert v_2 v_3 \rvert \lvert \lambda_2 \rvert) + \rho_1 \sigma_2 v_1 v_2 (\lvert \lambda_2 \rvert + \lvert \lambda_1 \rvert)
\]

\[
\triangleq \tau_1
\]

where (a) follows from the fact that \( \lvert \lambda_i \rvert < 1, i = 1, 2 \). Using similar arguments we bound \( |r_2| \) as follows:

\[
|r_2| \leq \frac{\sigma_2 \sqrt{1 - \rho_2^2} v_1 v_2 (\lvert \lambda_2 \rvert + \lvert \lambda_1 \rvert)}{\lvert \text{det}(V) \rvert} \triangleq \tau_2.
\]

Hence, we have \( [M^k Q\sigma(M^T)^k]_{1,1} \leq \tau_1^2 + \tau_2^2 \), and this implies that:

\[
\mathbb{E} \{ U_{1,k}^2 \} \leq \tau_1^2 + \tau_2^2 + \lvert Q_u \rvert_{1,1} \triangleq \vartheta_1.
\]

Following similar arguments we have \( [M^k Q\sigma(M^T)^k]_{2,2} \leq \tau_3^2 + \tau_4^2 \), where:

\[
\tau_3 \triangleq \frac{|\sigma_1 v_3 v_4| (|\lambda_1| + |\lambda_2|) + |\rho_1 \sigma_2| (|v_1 v_4 \lambda_1| + |v_2 v_3 \lambda_2|)}{\lvert \text{det}(V) \rvert},
\]

\[
\tau_4 \triangleq \frac{\sigma_2 \sqrt{1 - \rho_2^2} (|v_1 v_4 \lambda_2| + |v_2 v_3 \lambda_1|)}{\lvert \text{det}(V) \rvert},
\]

and therefore:

\[
\mathbb{E} \{ U_{2,k}^2 \} \leq \tau_3^2 + \tau_4^2 + \lvert Q_u \rvert_{2,2} \triangleq \vartheta_2.
\]

To conclude, we have:

\[
K_{LQG} \leq \frac{\log \left( \frac{\vartheta_i}{\vartheta} \right)}{2 \log |a_i|}, \quad i = 1, 2.
\]

To lower bound \( K_{LQG} \) we first lower bound the MSE in (22) as follows:

\[
\mathbb{E} \left\{ (S_i - \tilde{S}_{i,k})^2 \right\} = \sigma_i^2 \frac{|Q_{u,k+1}|_{i,i} - \frac{1}{\sigma_i} \left( [M^k Q\sigma]_{i,i} \right)^2}{|Q_{u,k+1}|_{i,i} - 2 \kappa_i [M^k Q\sigma]_{i,i} + \sigma_i^2 \delta_i^2} \geq \sigma_i^2 \frac{|Q_{u,k+1}|_{i,i} - 2 \kappa_i [M^k Q\sigma]_{i,i} + \sigma_i^2 \delta_i^2}{|Q_{u,k+1}|_{i,i} - 2 \kappa_i [M^k Q\sigma]_{i,i} + \sigma_i^2 \delta_i^2}, \tag{B.19}
\]

To see why step (a) holds we note that:

\[
0 \leq \frac{|Q_{u,k+1}|_{i,i} - \frac{1}{\sigma_i} \left( [M^k Q\sigma]_{i,i} \right)^2}{|Q_{u,k+1}|_{i,i} - 2 \kappa_i [M^k Q\sigma]_{i,i} + \sigma_i^2 \delta_i^2} \leq 1.
\]

This follows as \( \mathbb{E} \left\{ (S_i - \tilde{S}_{i,k})^2 \right\} = \sigma_i^2 - \mathbb{E} \{ \tilde{S}_{i,k}^2 \} \), and therefore we have:

\[
\sigma_i^2 \left( 1 - \frac{|Q_{u,k+1}|_{i,i} - \frac{1}{\sigma_i} \left( [M^k Q\sigma]_{i,i} \right)^2}{|Q_{u,k+1}|_{i,i} - 2 \kappa_i [M^k Q\sigma]_{i,i} + \sigma_i^2 \delta_i^2} \right) = \mathbb{E} \{ \tilde{S}_{i,k}^2 \} \geq 0,
\]

which implies that \( \frac{|Q_{u,k+1}|_{i,i} - \frac{1}{\sigma_i} \left( [M^k Q\sigma]_{i,i} \right)^2}{|Q_{u,k+1}|_{i,i} - 2 \kappa_i [M^k Q\sigma]_{i,i} + \sigma_i^2 \delta_i^2} \leq 1 \). Next, consider the function \( f(x) = \frac{a+x}{b+x}, x > 0, a, b \in \mathbb{R} \). We now show that if \( 0 \leq f(x) \leq 1 \), then \( f(x) \) is an increasing function. The derivative of \( f(x) \) is given by: \( f'(x) = \frac{b-a}{(x+b)^2} \). Consider the following cases:

30
• \( a, b \geq 0 \): From the fact that \( f(x) \leq 1 \), it follows that \( b \geq a \), and therefore \( f'(x) \geq 0 \).
• \( b \geq 0, a \leq 0 \): For this case, it is clear that \( f'(x) \geq 0 \).
• \( a, b \leq 0 \): From the fact that \( f(x) \leq 1 \), it follows that \( a \leq b \), and therefore \( f'(x) \geq 0 \).
• \( b < 0, a > 0 \): This assignment is not valid since \( f(x) \leq 1 \).

We conclude that for all valid cases \( f'(x) \geq 0 \) which implies that \( f(x) \) is an increasing function. With this in mind we note that \([Q_{a,k+1},1,i] \geq [Q_{z},1,i]\),\(^{18}\) which concludes the proof of step (a) in (B.19).

Next, we lower bound the numerator of (B.19) and upper bound the denominator of (B.19). Recall that \( M^k = (A - BC^T)k \) and consider upper bounding \([M^kQ_z]_{1,1}\). Similarly to Section B.2 we write \( M^k \) in terms of the eigenvalues matrix \( D \) and the eigenvectors matrix \( V \) of \( M \) as in (B.9): \( M^k = VD^kV^{-1} \). From (B.11) we have:

\[
VD^kV^{-1}Q_z = \frac{1}{\det(V)} \begin{bmatrix}
v_1v_4\lambda_k^1 - v_2v_3\lambda_k^2 & v_1v_2(\lambda_k^2 - \lambda_k^1) \\
v_3v_4(\lambda_k^1 - \lambda_k^2) & v_1v_4\lambda_k^2 - v_2v_3\lambda_k^1 \\
\end{bmatrix} \begin{bmatrix}
\sigma_i^2 & \rho_i\sigma_1\sigma_2 \\
\rho_i\sigma_1\sigma_2 & \sigma_i^2 \\
\end{bmatrix},
\]

from which we compute:

\[
[M^kQ_z]_{1,1} = \frac{\sigma_i^2(v_1v_4\lambda_k^1 - v_2v_3\lambda_k^2) + \rho_i\sigma_1\sigma_2v_1v_2(\lambda_k^2 - \lambda_k^1)}{\det(V)}.
\]

(B.20)

Using the fact that \( |\lambda_i| < 1, i = 1,2 \), we obtain the following upper bound on \([M^kQ_z]_{1,1}\), \( k \geq 1 \):

\[
|[M^kQ_z]_{1,1}| \leq \frac{\sigma_i^2(|v_1v_4\lambda_1| + |v_2v_3\lambda_2|) + |\rho_i\sigma_1\sigma_2v_1v_2(|\lambda_2| + |\lambda_1|)}{|\det(V)|} \triangleq \beta_1.
\]

Similarly, we also bound:

\[
|[M^kQ_z]_{2,2}| \leq \frac{\sigma_i^2(|v_1v_4\lambda_2| + |v_2v_3\lambda_1|) + |\rho_i\sigma_1\sigma_2v_1v_2(|\lambda_2| + |\lambda_1|)}{|\det(V)|} \triangleq \beta_2.
\]

Now, for \( i = 1,2 \), plugging \( \beta_i \) into (B.19) and setting \( D_{i,k} = D_i \) we write:

\[
D_i \geq \frac{\sigma_i^2[Q_z]_{1,i} - \beta_i^2}{[Q_z]_{1,i} + 2|a_i|^k\beta_i + \sigma_i^2|a_i|^{2k}},
\]

which can also be written as:

\[
\frac{D_i[Q_z]_{1,i} - \sigma_i^2[Q_z]_{1,i} + \beta_i^2}{D_i} \geq -D_i \left( 2|a_i|^k\beta_i + \sigma_i^2|a_i|^{2k} \right),
\]

\[
\Rightarrow \frac{\sigma_i^2[Q_z]_{1,i} - \beta_i^2 - D_i[Q_z]_{1,i}}{D_i} \leq 2|a_i|^k\beta_i + \sigma_i^2|a_i|^{2k}.
\]

Next, we recall that \( |a_i| > 1 \) and write:

\[
\frac{\sigma_i^2[Q_z]_{1,i} - \beta_i^2 - D_i[Q_z]_{1,i}}{D_i} \leq (2\beta_i + \sigma_i^2)|a_i|^{2k}.
\]

Applying the log to both sides we have:

\[
\log \left( \frac{\sigma_i^2[Q_z]_{1,i} - \beta_i^2 - D_i[Q_z]_{1,i}}{D_i} \right) \leq \log \left( (2\beta_i + \sigma_i^2)|a_i|^{2k} \right),
\]

\(^{18}\)From (B.7) it follows that \( E(U_{1,k}^2) \geq [Q_z]_{1,i} \).
which can be written as:
\[
\log \left( \frac{\sigma_i^2[Q_z]_{i,i} - \beta_i^2 - D_i[Q_z]_{i,i}}{(2\beta_i + \sigma_i^2)D_i} \right) \leq 2k \log |a_i|.
\]

Thus, we write:
\[
\log \left( \frac{\sigma_i^2[Q_z]_{i,i} - \beta_i^2 - D_i[Q_z]_{i,i}}{(2\beta_i + \sigma_i^2)D_i} \right) \leq 2 \log |a_i| \leq K_{LQG},
\]
which is stated in (27b).

C Proofs for the LQG Scheme for the Symmetric Setting

C.1 Proof of Theorem 4

We begin with the following lemma:

Lemma 1. For symmetric GBCFs \( c_2 = -c_1 \).

Proof. We explicitly express \( c_1 \) in terms of \( a_1 \). Recall the definition of the vector \( C \) in Section 4.1:
\[
C = (B^TGB + 1)^{-1}AGB \tag{C.1}
\]
where \( G \) is the unique positive-definite solution of the DARE \( G = A^TGA - A^TGB(B^TGB + 1)^{-1}B^TGA \), such that all the eigenvalues of the matrix \( A - BC \) have magnitudes smaller than 1. Let \( G = \begin{bmatrix} g_1 & g_2 \\ g_3 & g_4 \end{bmatrix} \). From [35, Prop. 1] we have that for the symmetric case and for
\[
A = \begin{bmatrix} a_1 & 0 \\ 0 & -a_1 \end{bmatrix}
\]
the elements of \( G \) are given by:
\[
g_1 = g_4 = \frac{(a_1^2 - 1)(1 + a_1^2)^2}{4a_1^4}, \quad g_2 = g_3 = \frac{(1 - a_1^2)^2(1 + a_1^2)}{4a_1^4},
\]
and it follows that \( G = G^T \). Writing \( AGB = AGB \) explicitly:
\[
AGB = \begin{bmatrix} a_1 & 0 \\ 0 & -a_1 \end{bmatrix} \begin{bmatrix} g_1 & g_2 \\ g_3 & g_4 \end{bmatrix} = \begin{bmatrix} a_1(g_1 + g_2) \\ -a_1(g_1 + g_2) \end{bmatrix}.
\]

Using the explicit expressions for \( g_1 \) and \( g_2 \) we can write \( g_1 + g_2 = \frac{(1+a_1^2)(a_1^2-1)}{2} \). Next, writing \( B^TGB + 1 \) explicitly we obtain:
\[
B^TGB + 1 = 2(g_1 + g_2) + 1 = (1 + a_1^2)(a_1^2 - 1) + 1.
\]

We now can explicitly compute \( c_1 \), the first element of \( C \) in (C.1):
\[
c_1 = \frac{a_1(g_1 + g_2)}{2(g_1 + g_2) + 1} = \frac{a_1(1 + a_1^2)(a_1^2 - 1)}{2((1 + a_1^2)(a_1^2 - 1) + 1)} = \frac{a_1^4 - 1}{2a_1^4} \tag{C.2}
\]
Computing \( c_2 \) via similar arguments we find \( c_2 = -c_1 \). \( \square \)

Next, we recall (B.8):
\[
P_k = C^T M_{k} Q_z (M^T)^{k-1} C + \sum_{l=0}^{k-2} C^T M_{l} Q_z (M^T)^{l} C,
\]

32
and note that with $c_2 = -c_1$ (B.10) is specialized to:
\[ C^T M^k Q_s(M^T)^k C = c_1^2 ((r_1 - r_3)^2 + (r_2 - r_4)^2). \] (C.3)

In the symmetric setting we also have $\sigma_1 = \sigma_2 = \sigma_s$ and $\sigma_{z,1} = \sigma_{z,2} = \sigma_z$. From the expression for the matrix $M$ and from the expression for $c_1$ it follows that $v_1 = v_4, v_2 = v_3$, and $-\lambda_2 = \lambda_1$. Therefore, (B.12) is specialized to:

\[
\begin{align*}
r_1 &= \frac{\lambda_s^k \sigma_s}{\det(V)} (v_1^2 - v_2^2(-1)^k + \rho_s v_1 v_2((-1)^k - 1)) \tag{C.4a} \\
r_2 &= \frac{\lambda_s^k \sigma_s \sqrt{1 - \rho_s^2}}{\det(V)} (v_1 v_2((-1)^k - 1)) \tag{C.4b} \\
r_3 &= \frac{\lambda_s^k \sigma_s}{\det(V)} (v_1 v_2(1 - (-1)^k) + \rho_s (v_1^2(-1)^k - v_2^2)) \tag{C.4c} \\
r_4 &= \frac{\lambda_s^k \sigma_s \sqrt{1 - \rho_s^2}}{\det(V)} (v_1^2(-1)^k - v_2^2). \tag{C.4d}
\end{align*}
\]

Next, we explicitly write $r_1 - r_3$:

\[
\begin{align*}
r_1 - r_3 &= \frac{\lambda_s^k \sigma_k}{\det(V)} (v_1^2 - v_2^2(-1)^k + \rho_s v_1 v_2((-1)^k - 1) \\
&\quad - v_1 v_2(1 - (-1)^k) - \rho_s (v_1^2(-1)^k - v_2^2)) \\
&= \frac{\lambda_s^k \sigma_k}{\det(V)} (v_1^2(1 - \rho_s(-1)^k) + v_2^2(\rho_s - (-1)^k) + v_1 v_2(\rho_s + 1)((-1)^k - 1)),
\end{align*}
\]

and by squaring we obtain:

\[
\begin{align*}
(r_1 - r_3)^2 &= \frac{\lambda_s^{2k} \sigma_k^2}{\det^2(V)} (v_1^2(1 - \rho_s(-1)^k) + v_2^2(\rho_s - (-1)^k) + v_1 v_2(\rho_s + 1)((-1)^k - 1))^2 \\
&= \frac{\lambda_s^{2k} \sigma_k^2}{\det^2(V)} \left( v_1^2(1 - \rho_s(-1)^k)^2 + v_2^2(\rho_s - (-1)^k)^2 + v_1^2 v_2^2(1 + \rho_s)^2((-1)^k - 1)^2 \\
&\quad + 2v_1^2 v_2^2(1 - \rho_s(-1)^k)(\rho_s - (-1)^k) \\
&\quad + 2v_1^2 v_2(1 - \rho_s(-1)^k)(\rho_s + 1)((-1)^k - 1) \\
&\quad + 2v_1 v_2^3(\rho_s - (-1)^k)(\rho_s + 1)((-1)^k - 1) \right),
\end{align*}
\]

Now, for \textit{even} values of $k$ we have:

\[
\begin{align*}
(r_1 - r_3)^2 &= \frac{\lambda_s^{2k} \sigma_k^2}{\det^2(V)} \left( v_1^2(1 - \rho_s(-1)^k)^2 + v_2^2(\rho_s - (-1)^k)^2 - 2v_1^2 v_2^2(1 - \rho_s)^2 \right) \\
&= \frac{\lambda_s^{2k} \sigma_k^2}{\det^2(V)} ((1 - \rho_s)^2(v_1^2 - v_2^2)^2) \\
&= \lambda_s^{2k} \sigma_k^2(1 - \rho_s)^2, \tag{C.5}
\end{align*}
\]

while for \textit{odd} values of $k$ we have:

\[
\begin{align*}
(r_1 - r_3)^2 &= \frac{\lambda_s^{2k} \sigma_k^2}{\det^2(V)} \left( v_1^2(1 + \rho_s)^2 + v_2^2(1 + \rho_s)^2 + 4v_1^2 v_2^2(1 + \rho_s)^2 \\
&\quad + 2v_1^2 v_2^2(1 + \rho_s)^2 - 4v_1^2 v_2(1 + \rho_s)^2 - 4v_1 v_2^3(1 + \rho_s)^2 \\
&= \lambda_s^{2k} \sigma_k^2(1 + \rho_s)^2 \frac{v_1^4 + v_2^4 + 6v_1^2 v_2^2 - 4v_1 v_2(2v_1^2 + v_2^2)}{\det^2(V)} \\
&\equiv (a) \lambda_s^{2k} \sigma_k^2(1 + \rho_s)^2 a_1^4, \tag{C.6}
\end{align*}
\]

where (a) follows from the following lemma.
Lemma 2. The following equality holds: $\frac{v_1^4 + v_2^4 + 6\lambda_1^2 v_1^2 - 4v_1 v_2 (v_1^2 + v_2^2)}{\det^2(V)} = a_1^4$.

Proof. We begin with expressing $\lambda_1, v_1,$ and $v_2$. From [30, Lemma 2.4] it follows that $\lambda_1 = \frac{1}{a_1}$, see Footnote 16 for a detailed explanation. Next, we explicitly write $M = \begin{bmatrix} a_1 - c_1 & c_1 \\ -c_1 & -(a_1 - c_1) \end{bmatrix}$, and note that an eigenvector $V_0$ of $M$, corresponding to the eigenvalue $\lambda_1$, obeys $M V_0 = \lambda_1 V_0$. This equation can also be written using a matrix form:

$$(M - \lambda_1 I) V_0 = \begin{bmatrix} a_1 - c_1 - \lambda_1 \\ -(a_1 - c_1) - \lambda_1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0.$$ 

Recalling that eigenvectors have unit norm, we obtain an explicit expression for $V_0$:

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} \frac{c_1}{\sqrt{c_1^2 + (a_1 - c_1 - \lambda_1)^2}} \\ \frac{-(a_1 - c_1 - \lambda_1)}{\sqrt{c_1^2 + (a_1 - c_1 - \lambda_1)^2}} \end{bmatrix}.$$ 

Substituting $\lambda_1 = \frac{1}{a_1}$ we obtain:

$$v_1 = \frac{c_1}{\sqrt{c_1^2 + (a_1 - c_1 - \frac{1}{a_1})^2}} = \frac{a_1 c_1}{\sqrt{a_1^2 c_1^2 + ((a_1 - c_1)a_1 - 1)^2}} \quad \text{(C.7a)}$$

$$v_2 = -\frac{a_1 - c_1 - \frac{1}{a_1}}{\sqrt{c_1^2 + (a_1 - c_1 - \frac{1}{a_1})^2}} = \frac{1 - (a_1 - c_1)a_1}{\sqrt{a_1^2 c_1^2 + ((a_1 - c_1)a_1 - 1)^2}} \quad \text{(C.7b)}$$

Note that (C.2) implies that $0 \leq c_1 \leq a_1$. Using the expression for $c_1$ we now write $a_1(a_1 - c_1) - 1$ in terms of $a_1$:

$$a_1(a_1 - c_1) - 1 = a_1 \left( a_1 - \frac{(a_1^4 - 1)}{2a_1^3} \right) - 1 = \frac{a_1^5 + a_1}{2a_1^3} - 1 = \frac{(a_1^2 - 1)^2}{2a_1^2}. \quad \text{(C.8)}$$

Thus, the numerator of (C.7a) equals $a_1 c_1 = \frac{a_1^2 - 1}{2a_1^2}$, while the numerator of (C.7b) equals $1 - (a_1 - c_1)a_1 = \frac{(a_1^2 - 1)^2}{2a_1^2}$. We further note that the denominators of (C.7a) and (C.7b) are the same. Therefore, we write:

$$\frac{v_1^4 + v_2^4 + 6\lambda_1^2 v_1^2 - 4v_1 v_2 (v_1^2 + v_2^2)}{\det^2(V)} = \frac{(a_1^4 - 1)^4 + (a_1^2 - 1)^8 + 6(a_1^4 - 1)^2(a_1^2 - 1)^4 + 4(a_1^4 - 1)(a_1^2 - 1)^2((a_1^4 - 1)^2 + (a_1^2 - 1)^4)}{((a_1^4 - 1)^2 - (a_1^2 - 1)^4)^2} \quad \text{(C.9)}$$

The denominator of (C.9) can be written as:

$$((a_1^2 - 1)^2 - (a_1^2 - 1)^4)^2 = ((a_1^2 - 1)^2(a_1^2 + 1)^2 - (a_1^2 - 1)^4)^2 = 16a_1^4(a_1^2 - 1)^4. \quad \text{(C.10a)}$$

The numerator of (C.9) can be written as:

$$(a_1^4 - 1)^4 + (a_1^2 - 1)^8 + 6(a_1^4 - 1)^2(a_1^2 - 1)^4 + 4(a_1^4 - 1)(a_1^2 - 1)^2((a_1^4 - 1)^2 + (a_1^2 - 1)^4)$$

$$= (a_1^2 - 1)^4(8a_1^4 + 8 + 8(a_1^2 - 1))$$

$$= 16a_1^8(a_1^2 - 1)^4. \quad \text{(C.10b)}$$

Thus, by combining (C.10a) and (C.10b) we obtain:

$$\frac{v_1^4 + v_2^4 + 6\lambda_1^2 v_1^2 - 4v_1 v_2 (v_1^2 + v_2^2)}{\det^2(V)} = \frac{16a_1^8(a_1^2 - 1)^4}{16a_1^8(a_1^2 - 1)^4} = a_1^4. \quad \text{(C.11)}$$

This concludes the proof of the lemma. □
Similarly to (C.6), we write:
\[
(r_2 - r_4)^2 = \frac{\lambda_1^k \sigma_z^2 (1 - \rho_z^2)}{\det^2(V)} \left( v_1 v_2((-1)^k - 1) - v_1^2(-1)^k + v_2^2 \right)^2
\]
\[
= \frac{\lambda_1^k \sigma_z^2 (1 - \rho_z^2)}{\det^2(V)} \left( v_1^2 v_2^2((-1)^k - 1)^2 + v_1^2 + v_2^2 - 2 v_1 v_2((-1)^k - 1) \right)
\]
\[
+ 2 v_1 v_2^2((-1)^k - 1) - 2 v_1^2 v_2^2((-1)^k)
\]
\[
= \begin{cases} 
\lambda_1^k \sigma_z^2 (1 - \rho_z^2), & k \text{ is even}, \\
\lambda_1^k \sigma_z^2 (1 - \rho_z^2) a_1^4, & k \text{ is odd}.
\end{cases}
\tag{C.12}
\]
Hence, combining (C.3) and (C.5)–(C.12) we obtain:
\[
C^T M^{k-1} Q_z(M^T)^{k-1} C = \begin{cases} 
2 \lambda_1^2 \lambda_1^{2(k-1)} \sigma_z^2 (1 - \rho_z), & k - 1 \text{ is even}, \\
2 \lambda_1^2 \lambda_1^{2(k-1)} \sigma_z^2 (1 + \rho_z) a_1^4, & k - 1 \text{ is odd}.
\end{cases}
\tag{C.13}
\]
Next, we focus on \(\sum_{l=0}^{k-2} C^T M^l Q_z(M^T)^l C\). Following the steps leading to (C.13) we write:
\[
C^T M^l Q_z(M^T)^l C = \begin{cases} 
2 \lambda_1^2 \lambda_1^{2l} \sigma_z^2 (1 - \rho_z), & l \text{ is even}, \\
2 \lambda_1^2 \lambda_1^{2l} \sigma_z^2 (1 + \rho_z) a_1^4, & l \text{ is odd}.
\end{cases}
\]
For even values of \(k - 1\) we have:
\[
\sum_{l=0}^{k-2} C^T M^l Q_z(M^T)^l C = \sum_{m=0}^{k-2-1} 2 \lambda_1^2 \sigma_z^2 (1 - \rho_z) \lambda_1^{4m} + \sum_{m=0}^{k-1-1} 2 \lambda_1^2 \sigma_z^2 (1 + \rho_z) a_1^4 \lambda_1^{4m+2}
\]
\[
= 2 \lambda_1^2 \sigma_z^2 ((1 - \rho_z) + (1 + \rho_z) a_1^4 \lambda_1^2) \sum_{m=0}^{k-1-1} \lambda_1^{4m}
\]
\[
= \frac{2 \lambda_1^2 \sigma_z^2 ((1 - \rho_z) + (1 + \rho_z) a_1^4 \lambda_1^2)}{1 - \lambda_1^4}
\]
\[
= \mu_1 \cdot (1 - \lambda_1^{2(k-1)}),
\tag{C.14}
\]
where (a) follows from the fact that \(a_1 = \frac{1}{\lambda_1^4}\). For odd values of \(k - 1\) we have:
\[
\sum_{l=0}^{k-2} C^T M^l Q_z(M^T)^l C
\]
\[
= \sum_{m=0}^{k-2-1} 2 \lambda_1^2 \sigma_z^2 (1 - \rho_z) \lambda_1^{4m} + \sum_{m=0}^{k-2-1} 2 \lambda_1^2 \sigma_z^2 (1 + \rho_z) a_1^4 \lambda_1^{4m+2}
\]
\[
= 2 \lambda_1^2 \sigma_z^2 ((1 - \rho_z) \lambda_1^{2(k-2)} + 2 \lambda_1^2 \sigma_z^2 ((1 - \rho_z) + (1 + \rho_z) a_1^4 \lambda_1^2) \sum_{m=0}^{k-2-1} \lambda_1^{4m}
\]
\[
= \frac{(1 - \lambda_1^4) 2 \lambda_1^2 \sigma_z^2 ((1 - \rho_z) \lambda_1^{2(k-2)} + 2 \lambda_1^2 \sigma_z^2 ((1 - \rho_z) + (1 + \rho_z) a_1^4 \lambda_1^2) (1 - \lambda_1^{2(k-2)})}{1 - \lambda_1^4}
\]
\[
= \frac{2 \lambda_1^2 \sigma_z^2 ((1 - \rho_z) \lambda_1^{2(k-2)})}{1 - \lambda_1^4} - \frac{2 \lambda_1^2 \sigma_z^2 ((1 - \rho_z) \lambda_1^{2k})}{1 - \lambda_1^4}
\]
\[
+ \frac{2 \lambda_1^2 \sigma_z^2 ((1 - \rho_z) + (1 + \rho_z) a_1^4 \lambda_1^2)}{1 - \lambda_1^4} \frac{2 \lambda_1^2 \sigma_z^2 ((1 - \rho_z) + (1 + \rho_z) a_1^4 \lambda_1^2) \lambda_1^{2(k-2)}}{1 - \lambda_1^4}
\]
\[
= \frac{2 \lambda_1^2 \sigma_z^2 ((1 - \rho_z) + (1 + \rho_z) a_1^4 \lambda_1^2)}{1 - \lambda_1^4}
\]
\[
= \mu_1 - \mu_3 \cdot \lambda_1^{2(k-1)}.
\tag{C.15}
\]
Recalling that \( \mu_0 = 2c^2\sigma_\nu^2(1 - \rho_3) \) and \( \mu_2 = 2c^2\sigma_\nu^2(1 + \rho_3)a_1^2 \), we combine (C.13)–(C.15) to obtain (29). Similarly to (B.18) we have that \( \lim_{k \to \infty} P_k = P \), and since \( |\lambda_1| < 1 \) it follows that \( \mu_1 = P \). Therefore, the power constraint (5) is satisfied if and only if \( \mu_0 \leq \mu_1 \) and \( \mu_2 \leq \mu_3 \).

## C.2 Proof of Proposition 3

First, we show that the maximal possible scaling which satisfies (5) is: \( \sqrt{\frac{\nu}{\sigma^2}} \). Then, we prove that the optimal estimator and the obtained MSE are given in (31) and (32), respectively. Finally, we show that setting \( \gamma = \frac{\nu}{\sigma^2} \) indeed minimizes the MSE.

### C.2.1 Maximal Scaling

Recall that (30) constitutes an upper bound on the variance of the sources transmitted via an LQG scheme based on (15), which satisfy (5). Explicitly writing the conditions of Thm. 4, i.e., \( \mu_0 \leq \mu_1 \) and \( \mu_2 \leq \mu_3 \), where \( \mu_j, j = 0, \ldots, 3 \) are defined in (28), we obtain:

\[
\sigma_\nu^2(1 - \rho_3) \leq \frac{\sigma_\nu^2(1 - \rho_3 + (1 + \rho_3)a_1^2)}{1 - \lambda_1^2}, \quad \sigma_\nu^2(1 + \rho_3)a_1^4 \leq \frac{\sigma_\nu^2((1 - \rho_3)\lambda_1^2 + (1 + \rho_3)a_1^4)}{1 - \lambda_1^2}.
\]

This implies that:

\[
\sigma_\nu^2 \leq \min \left\{ \frac{\sigma_\nu^2(1 - \rho_3 + (1 + \rho_3)a_1^2)}{1 - \lambda_1^2}(1 - \rho_3), \frac{\sigma_\nu^2((1 - \rho_3)\lambda_1^2 + (1 + \rho_3)a_1^4)}{1 - \lambda_1^2}(1 + \rho_3)a_1^4 \right\},
\]

and therefore, the maximal possible scaling which satisfies (5) is: \( \sqrt{\frac{\nu}{\sigma^2}} \).

### C.2.2 Optimal Estimator and Resulting MSE

Following the same arguments as those applied in Appendix B.1 the optimal estimator of \( S_i \) based on the observation \( \hat{U}_{i,k+1}(\gamma) \) is given by \( \text{E}\{S_i|\hat{U}_{i,k}(\gamma)\} = \text{E}\{\hat{S}_{i,k}(\gamma)|\hat{U}_{i,k}(\gamma)\} \). Letting \( \hat{S} = \sqrt{\gamma} \cdot S \) we can write:

\[
\text{E}\{S_i|\hat{U}_{i,k}(\gamma)\} = \frac{1}{\sqrt{\gamma}} \text{E}\{\hat{S}_{i,k}(\gamma)|\hat{U}_{i,k}(\gamma)\}.
\]

Note that \( \text{E}\{\hat{S}_{i,k}(\gamma)|\hat{U}_{i,k}(\gamma)\} \) can be obtained from (21) by setting \( \sigma_\nu^2 = \gamma \cdot \sigma_\nu^2 \). Let \( Q_\nu \triangleq \text{E}\{\hat{S}\hat{S}^T\} \). Following the arguments leading to (B.4) we write:

\[
\text{E}\{S_i|\hat{U}_{i,k}(\gamma)\} = \frac{1}{\sqrt{\gamma}} \cdot \frac{|M^kQ_{\nu,1}|_{i,i} - \sigma_\nu^2a_1^2}{|Q_{\nu,k+1}(\gamma)|_{i,i} - 2\gamma a_1^2|M^kQ_{\nu,1}|_{i,i} + 2\gamma^2 a_1^4} \hat{U}_{i,k+1}(\gamma)
\]

\[
= \sqrt{\gamma} \left( |M^kQ_{\nu,1}|_{i,i} - \sigma_\nu^2a_1^2 \right) \hat{U}_{i,k+1}(\gamma).
\]

Moreover, by following the arguments leading to (B.6) we obtain (32):

\[
\text{E}\{(S_i - \hat{S}_{i,k})^2\} = \frac{\sigma_\nu^2|Q_{\nu,k+1}(\gamma)|_{i,i} - \gamma |M^kQ_{\nu,1}|_{i,i}^2}{|Q_{\nu,k+1}(\gamma)|_{i,i} - 2\gamma a_1^2|M^kQ_{\nu,1}|_{i,i} + 2\gamma a_1^2}.
\]

### C.2.3 Explicit Expression of the MSE

We now derive an explicit expression for the MSE. From (C.16) it follows that we need to characterize \( |Q_{\nu,k+1}(\gamma)|_{1,1} \) and \( |M^kQ_{\nu,1}|_{1,1} \). Next, we explicitly characterize \( |Q_{\nu,k+1}(\gamma)|_{1,1} \) as a function of \( k \).
C.2.3.1 Analysis of \([Q_{u,k+1}(\gamma)]_{1,1}\): From (B.7) we have:

\[
[Q_{u,k+1}(\gamma)]_{1,1} = \gamma \cdot [M^k Q_z(M^T)^k]_{1,1} + \sum_{l=0}^{k-1} M^l Q_z(M^T)^l]_{1,1}.
\] (C.17)

We now separately analyze the two terms on the RHS of (C.17).

Analysis of \(\gamma \cdot [M^k Q_z(M^T)^k]_{1,1}\): Following arguments similar to those leading to (B.10) we have \([M^k Q_z(M^T)^k]_{1,1} = r_1^2 + r_2^2\), where \(r_1\) and \(r_2\), specialized to the symmetric setting, are given in (C.4). Further simplifying the expressions we obtain:

\[
\begin{align*}
  r_1 &= \frac{\lambda_1^k \sigma_s}{\det(V)} (v_1^2 + v_2^2 - 2 \rho_s v_1 v_2) \\
  r_2 &= \begin{cases} 
  0, & k \text{ is even,} \\
  -2 \lambda_1^k \sqrt{1 - \rho_s^2} v_1 v_2, & k \text{ is odd.}
  \end{cases}
\end{align*}
\]

Thus, recalling that \(\Phi(\sigma, \rho) \triangleq \sqrt{2 (v_1^2 + v_2^2 - 2 \rho_s v_1 v_2)^2 + 4 (1 - \rho_s^2) v_1^2 v_2^2} / \det(V)\), see (33a), we obtain:

\[
\gamma \cdot [M^k Q_z(M^T)^k]_{1,1} = \begin{cases} 
  \lambda_1^{2k} \gamma \sigma_s, & k \text{ is even,} \\
  \lambda_1^{2k} \gamma \Phi(\sigma_s, \rho_s), & k \text{ is odd,}
  \end{cases}
\] (C.18)

where we note that \(\Phi(\sigma_s, \rho_s) \geq 0\). Next, we analyze the second term on the RHS of (C.17).

Analysis of \(\sum_{l=0}^{k-1} M^l Q_z(M^T)^l]_{1,1}\): Following the same arguments used for deriving (C.18), we write:

\[
[M^l Q_z(M^T)^l]_{1,1} = \begin{cases} 
  \lambda_1^{2l} \sigma_s^2, & l \text{ is even,} \\
  \lambda_1^{2l} \Phi(\sigma_s, \rho_s), & l \text{ is odd.}
  \end{cases}
\]

Now, for even \(k\), following similar arguments that led to (C.14), we obtain:

\[
\sum_{l=0}^{k-1} M^l Q_z(M^T)^l]_{1,1} = (\sigma_z^2 + \Phi(\sigma_z, \rho_z) \lambda_1^2) \sum_{l=0}^{k-1} \lambda_1^{4m} = \frac{\sigma_z^2 + \lambda_1^2 \Phi(\sigma_z, \rho_z)}{1 - \lambda_1^4} (1 - \lambda_1^{2k}) = \Psi_0 (1 - \lambda_1^{2k}),
\]

where \(\Psi_0 \triangleq \frac{\sigma_z^2 + \lambda_1^2 \Phi(\sigma_z, \rho_z)}{1 - \lambda_1^4}\) is defined in (33b). Since \(\Phi(\sigma_1, \rho_s) \geq 0\) and \(0 < \lambda_1 < 1\), it follows that \(\Psi_0 > 0\). For odd \(k\), we follow steps similar to those leading to (C.15) to obtain:

\[
\sum_{l=0}^{k-1} M^l Q_z(M^T)^l]_{1,1} = \frac{(\sigma_z^2 + \Phi(\sigma_z, \rho_z) \lambda_1^2)}{1 - \lambda_1^4} \left( \sum_{m=0}^{k-1} \lambda_1^{4m} \right) + \sigma_z^2 \lambda_1^{2(k-1)} - \lambda_1^2 \lambda_1^{2(k-1)} - \lambda_1^2 \lambda_1^{2(k-1)}
\]

\[
= \frac{\sigma_z^2 + \lambda_1^2 \Phi(\sigma_z, \rho_z)}{1 - \lambda_1^4} \Psi_0 - \lambda_1^{2k} - \Psi_1.
\]

Hence, we have:

\[
\sum_{l=0}^{k-1} M^l Q_z(M^T)^l]_{1,1} = \begin{cases} 
  \Psi_0 (1 - \lambda_1^{2k}), & k \text{ is even,} \\
  \Psi_0 - \lambda_1^{2k} - \Psi_1, & k \text{ is odd.}
  \end{cases}
\] (C.19)

Next, we combine (C.18) and (C.19) to obtain:

\[
[Q_{u,k+1}(\gamma)]_{1,1} = \begin{cases} 
  \lambda_1^{2k} (\gamma \sigma_s^2 - \Psi_0) + \Psi_0, & k \text{ is even,} \\
  \lambda_1^{2k} (\gamma \Phi(\sigma_s, \rho_s) - \Psi_1) + \Psi_0, & k \text{ is odd.}
  \end{cases}
\] (C.20)
C.2.3.2 Analysis of $[M^kQ_s]_{1,1}$: Recall the definition of $\Gamma_s \triangleq \frac{\sigma^2_1(v_1^2 + v_2^2 - 2\rho_s v_1 v_2)}{\det(V)}$ in (33d). For the symmetric setting, we rewrite $[M^kQ_s]_{1,1}$, given in (B.20), as follows:

$$[M^kQ_s]_{1,1} = \frac{\sigma^2_1 \lambda^k_s (v_1^2 - v_2^2) ((-1)^k - 1)}{v_1^2 - v_2^2}$$

Thus, for even values of $\lambda$, we have:

$$\gamma \lambda^k_s = \gamma \lambda^2 \lambda^{k-2}$$

Thus, for odd values of $\lambda$, we have:

$$\gamma \lambda^{k-2} \lambda^{k-2}$$

C.2.3.3 An Explicit Expression: By plugging (C.20) and (C.21) into (C.16) we obtain an explicit expression for the MSE:

$$E \left\{ (S_1 - \hat{S}_1, k)^2 \right\} = \begin{cases} 
\frac{\sigma^2_1 (\lambda^2 \gamma^2 \sigma^2_s - \Psi_0) + \Psi_0 \gamma (\lambda^k_s \sigma^2_s)^2}{\lambda^{2k} (\gamma^2 \sigma^2_s - \Psi_0) + \Psi_0 - 2 \gamma^2 \lambda^2 \sigma^2_s + \gamma \lambda^{2k-2} \lambda^k_s} 
& \text{k is even}, \\
\frac{\sigma^2_1 (\lambda^2 \gamma^2 \sigma^2_s - \Psi_0) + \Psi_0 - 2 \gamma (\lambda^2 \sigma^2_s + \gamma \lambda^{2k-2} \lambda^k_s) - \lambda^{2k} (\gamma^2 \sigma^2_s - \Psi_0) + \Psi_0 - \gamma \lambda^{2k} \Gamma^2_s}{\lambda^{2k} (\gamma^2 \sigma^2_s - \Psi_0) + \Psi_0 - 2 \gamma \Gamma^2_s + \gamma \lambda^{2k} \Gamma^2_s} 
& \text{k is odd}.
\end{cases}$$

Next, we show that (C.22) decreases when $\gamma$ increases.

C.2.4 The MSE Decrease with $\gamma$

We begin with the case of even values of $k$:

C.2.4.1 Even values of $k$: Note that $\Psi_0 > 0$. Thus, as $\lambda^2 < 1$, we have that if $\lambda^{2k} - 2 + \lambda^{-2k} > 0$ then the MSE decreases when $\gamma$ increases:

$$\lambda^{2k} - 2 + \lambda^{-2k} = \frac{\lambda^{2k} - 2 \lambda^{2k} + 1}{\lambda^{2k}} = \frac{(\lambda^{2k} - 1)^2}{\lambda^{2k}} > 0, \quad k > 0.$$ 

Thus, for even values of $k$, the MSE decreases with $\gamma$.

C.2.4.2 Odd values of $k$: Recalling the definitions of $\Phi(\sigma_s, \rho_s)$ and $\Gamma_s$ in (33a) and (33d), respectively, we write:

$$\Phi(\sigma_s, \rho_s) = \sigma^2_s \cdot C_0, \quad C_0 \triangleq \frac{(v_1^2 + v_2^2 - 2\rho_s v_1 v_2)^2 + 4(1 - \rho_s^2) v_1^2 v_2^2}{\det(V)},$$

$$\Gamma_s = \sigma^2_s \cdot C_1, \quad C_1 \triangleq \frac{v_1^2 + v_2^2 - 2\rho_s v_1 v_2}{\det(V)}.$$ 

Thus, for odd values of $k$, we write the MSE as follows:

$$E \left\{ (S_1 - \hat{S}_1, k)^2 \right\} = \frac{\gamma \sigma^4_1 \lambda^{2k} (C_0 - C_1^2) + \sigma^2_1 (\Psi_0 - \lambda^{2k} \Psi_1)}{\gamma \sigma^2_1 \lambda^{2k} (C_0 - 2C_1 + \lambda^{-2k}) + \Psi_0 - \lambda^{2k} \Psi_1}$$

Defining $\theta_1 \triangleq \sigma^4_1 \lambda^{2k} (C_0 - C_1^2), \theta_2 \triangleq \sigma^2_1 (\Psi_0 - \lambda^{2k} \Psi_1), \theta_3 \triangleq \sigma^2_1 (\lambda^{2k} C_0 - 2C_1 + \lambda^{-2k})$ and $\theta_4 \triangleq \Psi_0 - \lambda^{2k} \Psi_1$, the MSE is of the form: $\text{MSE}(\gamma) = \frac{\theta_1 + \theta_2}{\theta_3 + \theta_4}$. Clearly, if $\theta_j > 0, j = 1, 2, 3, 4$, and $\theta_3 > \theta_1$, then $\text{MSE}(\gamma)$ decreases with $\gamma$. Thus, we now show that these conditions hold.
**Positivity of \( \theta_1 \):** The positivity of \( \theta_1 \) follows directly from the definitions of \( C_0 \) and \( C_1 \).

**Positivity of \( \theta_2 \) and \( \theta_3 \):** Note that both \( \Psi_0 \) and \( \Psi_1 \) are positive. Furthermore, since \( \lambda_1^4 < 1 \), it is enough to show that \( \Psi_0 - \lambda_1^4 \Psi_1 > 0 \). We have:

\[
\Psi_0 - \lambda_1^4 \Psi_1 = \frac{\sigma_z^2 + \lambda_1^4 \Phi(\sigma_z, \rho_z) - \sigma_z^2 \Phi(\sigma_z, \rho_z) + \sigma_z^2 \lambda_1^4}{1 - \lambda_1^4} = \sigma_z^2 > 0.
\]

**Positivity of \( \theta_3 \):** Let \( \chi = \lambda_1^{2k} \), and write \( \theta_3 = \sigma_z^2 (\chi C_0 - 2C_1 + \chi^{-1}) = \frac{\sigma_z^2}{\chi} (\chi^2 C_0 - 2C_1 \chi + 1) \).

Therefore, as \( \chi > 0 \) and \( C_0 > 0 \), \( \theta_3 > 0 \) if the discriminant of \( \chi^2 C_0 - 2C_1 \chi + 1 \) is negative:

\[
\Delta = 4C_1^2 - 4C_0 = \frac{-16(1 - \rho_z^2)\nu_1^2\nu_2^2}{\det^2(V)} < 0. \tag{23}
\]

Thus, we conclude that \( \theta_3 > 0 \).

**Proving that \( \theta_1 > \theta_3 \):** We have:

\[
\theta_3 - \theta_1 = \sigma_z^2 (\lambda_1^{2k} C_0 - 2C_1 + \lambda_1^{-2k}) - \sigma_z^2 \lambda_1^{2k} (C_0 - C_1^2)
\]

\[
= \sigma_z^2 \lambda_1^{2k} (\lambda_1^{-4k} - 2\lambda_1^{-2k} C_1 + C_1^2)
\]

\[
= \sigma_z^2 \lambda_1^{2k} (\lambda_1^{-2k} - C_1)^2 > 0.
\]

Thus, for odd values of \( k \), the MSE decreases with \( \gamma \). We conclude that the MSE decreases with \( \gamma \) for all values of \( k \). Hence, the optimal \( \gamma \) is determined by the per-symbol average power constraint, and is given by \( \sqrt{\frac{E}{\nu^2}} \), where \( \nu \) is specified in (30).

### C.3 Proof of Theorem 5

To explicitly characterize \( K_{LQG} \) we solve the inequality \( E[(S_1 - \hat{S}_{1,k})^2] \leq D \), where \( E[(S_1 - \hat{S}_{1,k})^2] \) is given in (22). We begin with even values of \( k \).

#### C.3.1 Analysis for even values of \( k \)

For even values of \( k \) we are interested in the minimal even \( k \) such that:

\[
\frac{\sigma_z^2 \Psi_0(1 - \lambda_1^{2k})}{\gamma \sigma_z^2 (\lambda_1^{2k} - 2 + \lambda_1^{-2k})} \leq D, \tag{24}
\]

where \( \sqrt{\gamma} \) is a scaling factor of the transmitted sources, i.e., \( U_1 = \sqrt{\gamma} \cdot S \). This inequality can also be written as:

\[
\sigma_z^2 \Psi_0(1 - \lambda_1^{2k}) \leq D \left( \gamma \sigma_z^2 (\lambda_1^{2k} - 2 + \lambda_1^{-2k}) + \Psi_0(1 - \lambda_1^{2k}) \right).
\]

Next, we multiply both sides of the inequality by \( \lambda_1^{2k} \), and collect common terms to obtain the inequality:

\[
\lambda_1^{4k} (\Psi_0(D - \sigma_z^2) - D\gamma \sigma_z^2) + \lambda_1^{2k} (\Psi_0(\sigma_z^2 - D) + 2D\gamma \sigma_z^2) - D\gamma \sigma_z^2 \leq 0,
\]

which, by using the definitions of \( \Psi_0 \) and \( \Psi_1 \), can also be written as \( \lambda_1^{4k} \Psi_0 + \lambda_1^{2k} \Psi_1 - D\gamma \sigma_z^2 \leq 0 \).

Next, we set \( x = \lambda_1^{2k} \) and note that \( \Psi_0 < 0 \).\(^{20}\) Thus, we obtain the following monic polynomial inequalities:

\[
x^2 + x \frac{\Psi_1}{\Psi_0} - D\gamma \sigma_z^2 \geq 0 \tag{25}
\]

\(^{19}\)Note that \( \gamma \sigma_z^2 (\lambda_1^{2k} - 2 + \lambda_1^{-2k}) + \Psi_0(1 - \lambda_1^{2k}) > 0 \).

\(^{20}\)From the fact that \( \Phi(\sigma_z, \rho_z) \geq 0 \) it follows that \( \Psi_0 \geq 0 \). Furthermore from (6) we have that \( D \leq \sigma_z^2 \).

Therefore, as \( D, \gamma, \sigma_z^2 > 0 \) it follows that \( \Psi_0 = \Psi_0(D_1 - \sigma_z^2) - D_1 \gamma \sigma_z^2 < 0 \).
The discriminant of \( x^2 + x \frac{Y_1}{Y_0} - \frac{D\gamma \sigma^2}{Y_0} \) is equal to \( \frac{1}{Y_0^2} (Y_1^2 + 4D\gamma \sigma^2 Y_0) \). Therefore, if \( Y_0 < \frac{-Y_1^2}{4D\gamma \sigma^2} \), then \( P^{(e)}(x) \) has no real roots. Since \( x^2 + x \frac{Y_1}{Y_0} - \frac{D\gamma \sigma^2}{Y_0} \) is convex, if it has no real roots then it is strictly positive. Hence, in this case the required distortion is achieved for every even \( k \). Therefore, we set \( k = 2 \) and obtain \( x_0^{(e)} = a_1^{-4} \).

On the other hand, if \( Y_0 > \frac{-Y_1^2}{4D\gamma \sigma^2} \) then \( P^{(e)}(x) \) has two real roots. We write \( x^2 + x \frac{Y_1}{Y_0} - \frac{D\gamma \sigma^2}{Y_0} \) as:

\[
x^2 + px + q, \quad p = \frac{Y_1}{Y_0}, \quad q = -\frac{D\gamma \sigma^2}{Y_0}.
\]

The roots of this polynomial are given by \( -\frac{p}{2} \pm \sqrt{\frac{p^2}{4} - q} \). Now, since \( Y_1 > 0 \) and \( Y_0 < 0 \) we have \( -\frac{p}{2} > 0 \). As \( Y_0 < 0 \), then \( q > 0 \). Therefore, \( P^{(e)}(x) \) has two positive roots; one is smaller than \( -\frac{p}{2} \), and the other is larger. In fact, it is easy to see that \( P^{(e)}(1) = 0 \), thus, the larger root equals 1. Let \( \{x_1^{(e)}, x_2^{(e)}\} \) denote the real roots of \( P^{(e)}(x) \). Since \( x = \lambda_1^{(k)} \), and since \( k \geq 1 \), then the root we seek is \( \min\{x_1^{(e)}, x_2^{(e)}\} \). Let \( x_0^{(e)} \) denote the required root. Then we have:

\[
x_0^{(e)} \triangleq \begin{cases} 
\min \{x_1^{(e)}, x_2^{(e)}\}, & \frac{-Y_1^2}{4D\gamma \sigma^2} \leq Y_0 < 0 \\
\min \{x_1^{(e)}, x_2^{(e)}\}, & \text{otherwise}.
\end{cases}
\]

### C.3.2 Analysis of Odd \( k \)'s

For odd \( k \)'s we use (C.22) to write:

\[
\frac{\gamma (\lambda_1^{(k)} \sigma^2 \Phi(s, \rho_s) \sigma - \lambda_1^{(k)} \Gamma_2) + \sigma^2 (\Psi_0 - \lambda_1^{(k)} \Psi_1)}{\gamma (\lambda_1^{(k)} \Phi(s, \rho_s) - 2\Gamma_s + \sigma^2 \lambda_1^{(k)} \Gamma_1) + \Psi_0 - \lambda_1^{(k)} \Psi_1} \leq D. \tag{C.26}
\]

Based on Subsection C.2.3.3, this inequality can also be written as:

\[
\frac{\gamma (\lambda_1^{(k)} \sigma^2 \Phi(s, \rho_s) - \sigma^2 - D) - \gamma \Gamma_2) + \lambda_1^{(k)} (\Psi_0 (\sigma^2 - D) + 2D\gamma \Gamma_s) - D\gamma \sigma^2 \leq 0}.
\]

which, by using the definitions of \( Y_2 \) and \( Y_3 \), can also be written as \( \lambda_1^{(k)} Y_2 + \lambda_1^{(k)} Y_3 - D\gamma \sigma^2 \leq 0 \). Now, similarly to (C.25), we set \( x = \lambda_1^{(k)} \) to obtain the following monic polynomial inequalities:

\[
\begin{aligned}
x^2 + x \frac{Y_1}{Y_2} - \frac{D\gamma \sigma^2}{Y_2} & \leq 0, \quad Y_2 > 0, \\
x^2 + x \frac{Y_1}{Y_2} - \frac{D\gamma \sigma^2}{Y_2} & \geq 0, \quad Y_2 < 0, \\
x Y_3 - D\gamma \sigma^2 & \leq 0, \quad Y_2 = 0.
\end{aligned} \tag{C.27}
\]

The discriminant of \( P^{(o)}(x) = x^2 + x \frac{Y_1}{Y_2} - \frac{D\gamma \sigma^2}{Y_2} \) is given by \( \frac{1}{Y_2^2} (Y_3^2 + 4D\gamma \sigma^2 Y_2) \). Therefore, by applying arguments similar to those applied in Subsection C.3.1 we have:

\[
x_0^{(o)} \triangleq \begin{cases} 
\min \{x_1^{(o)}, x_2^{(o)}\}, & \frac{-Y_1^2}{4D\gamma \sigma^2} \leq Y_2 < 0, \\
\max \{x_1^{(o)}, x_2^{(o)}\}, & \text{otherwise}.
\end{cases}
\]

Note that if \( Y_2 > 0 \), then one of the roots is negative while the other is positive and smaller than 1. In such case we choose the positive root. Furthermore, if \( P^{(o)}(x) \) has no real roots, then we set \( x_0^{(o)} = a_1^{-2} \) which results in the minimal possible odd \( k \); i.e., \( k = 1 \).

Next, we focus on the case of \( Y_2 = 0 \) and note that for \( Y_2 = 0 \) equation (C.27) can be written as: \( x \leq \frac{D\gamma \sigma^2}{Y_3} \). This follows from the fact that \( Y_3 = \Psi_0 (\sigma^2 - D) + 2D\gamma \Gamma_s > 0 \). To see this note that \( \Psi_0 \geq 0, \sigma^2 \geq D \) and \( D, \gamma > 0 \). Furthermore, the numerator of
\( \Gamma_s \) in (33d) is positive, while the positivity of the denominator of \( \Gamma_s \) follows from (C.11) which implies that \( v_1^2 - v_2^2 > 0 \). Next, we show that \( 0 < \frac{D_k \alpha_2^2}{\alpha_1^2} < 1 \). The first inequality follows from the fact that \( \Upsilon_3, D, \gamma, \alpha_2^2 > 0 \). The second inequality follows from the fact that \( \Gamma_s = \frac{\sigma_1^2 v_1^2 + \sigma_2^2 - 2 \rho_s v_1 v_2}{v_2^2 - v_2^2} > \frac{\sigma_2^2}{\alpha_1^2} > 0 \),

where (a) follows from the fact that \( \rho_s < 1 \), and (b) follows from (C.7)–(C.8). We conclude that when \( \Upsilon_2 = 0 \) we can set \( x_0^{(e)} = \frac{D_k \sigma_2^2}{\sigma_1^2} \).

Lastly, as \( x = \lambda t^k = a_1^{-2k} \) then \( k = -\frac{\log(a_1)}{2} \). This implies that we have two candidates for the required \( K: K^{(e)} \) obtained from \( x_0^{(e)} \) and \( K^{(o)} \) obtained from \( x_0^{(o)} \). Since \( K \) is an integer, we use \( \lceil \cdot \rceil \), and the functions \( f^{(e)}(\cdot) \) and \( f^{(o)}(\cdot) \) to round up to the nearest even and odd integers, respectively.

### D Proofs for the Dynamic Programming Scheme

#### D.1 Proof of Theorem 6

Let \( W_k = \{\alpha_k, r_k\} \) be a “state” variable, and let \( m = [m_0, m_1, \ldots, m_{K-1}] \) be a given modulation vector. In the following we show that there exists a deterministic function \( f_{DP} \) such that \( W_k = f_{DP}(W_{k-1}, b_k, m), k = 1, 2, \ldots, K \). Namely, given the action \( b_k \) and the modulation vector \( m \), the state evolves deterministically as in (42). In appendix D.2 we show that \( P_k = P \) is the optimal assignment for the DP scheme which implies that the DP scheme exploits all the available instantaneous average transmission power. For \( \alpha_k \) we write:

\[
\alpha_k = E \left\{ \epsilon_{1,k-1} - b_k Y_{1,k} \right\}^2
\]

\[= \alpha_{k-1} + b_k^2 (P + \sigma_1^2) - 2 b_k E \{ \epsilon_{1,k-1} Y_{1,k} \} \]

\[= \alpha_{k-1} + b_k^2 (P + \sigma_1^2) - 2 b_k d_{k-1} E \{ \epsilon_{1,k-1} (\epsilon_{1,k-1} + m_{k-1} \epsilon_{2,k-1}) \} \]

\[= \alpha_{k-1} + b_k^2 (P + \sigma_1^2) - b_k \sqrt{2P (\alpha_{k-1} + m_{k-1} r_{k-1})}, \]

where (a) follows from the fact that since the transmitted signal and the noises are independent, then when \( P_k = P \) we have \( E \{ Y_{1,K} \} = P + \sigma_2^2 \); (b) follows by noting that \( E \{ \epsilon_{1,k-1} Y_{1,k} \} = E \{ \epsilon_{1,k-1} X_{k} \} = d_k E \{ \epsilon_{1,k-1} (\epsilon_{1,k-1} + m_{k-1} \epsilon_{2,k-1}) \} \). For \( r_k \) we write:

\[r_k = E \{ (\epsilon_{1,k-1} - b_k Y_{1,k}) (\epsilon_{2,k-1} - b_k m_{k-1} Y_{2,k}) \}
\]

\[= r_{k-1} + b_k^2 m_{k-1} (P + \rho_2 \sigma_1^2)
- b_k d_{k-1} m_{k-1} (\alpha_{k-1} + m_{k-1} r_{k-1}) - b_k d_{k-1} (m_{k-1} \alpha_{k-1} + r_{k-1})\]

\[= r_{k-1} + b_k^2 m_{k-1} (P + \rho_2 \sigma_1^2)
- b_k m_{k-1} \sqrt{2P (\alpha_{k-1} + m_{k-1} r_{k-1})}. \]

Therefore, the optimization problem in (41) can be cast as a dynamic program with state \( W_k \), actions \( \{b_k\}_{k=1}^K \) and cost function \( \alpha_{K,\min}(m) \), namely, a cost function that takes into account only the MSE at time \( K \) and ignores all the MSEs at times \( k < K \).

As we aim at minimizing \( \alpha_{K,\min}(m) \), the last action, \( b_K \), should be the MMSE estimator of \( \epsilon_{K-1} \) based on \( Y_{1,K} \), which is given by \( b_K = \frac{E \{ \epsilon_{1,K-1} Y_{1,K} \}}{E \{ Y_{1,K} \}} \).

\[b_k = \begin{cases} \frac{d_{K-1} (\alpha_{K-1} + m_{K-1} r_{K-1})}{P + \sigma_1^2} \quad (D.3a) \\
\sqrt{\frac{P (\alpha_{K-1} + m_{K-1} r_{K-1})}{2 (P + \sigma_1^2)^2}} \quad (D.3b) 
\end{cases} \]
where (a) is obtained by assuming that \( P_k = P \) in evaluating \( E\{Y_{1,k}^2\} \); and (b) is obtained by plugging the expression for \( d_{K-1} \) which is given in (38). In order to find the optimal \( \{b_k\}_{k=1}^K \), we first plug (D.3b) into (42a) and write:

\[
\alpha_K = \frac{P(\alpha_{K-1} + m_K - 1) + r_{K-1}}{2(P + \sigma_k^2)} - \frac{P(\alpha_{K-1} + m_K - 1) + r_{K-1}}{2(P + \sigma_k^2)} \quad (\alpha) \eta_{K-1} - \frac{1}{2(P + \sigma_k^2)} \eta_{K-1} \quad \eta_{K-1} \quad (D.4)
\]

where (a) follows by defining \( \eta_{K-1} \triangleq 1 - \frac{P}{2(P + \sigma_k^2)} \) and \( \theta_{K-1} \triangleq -\frac{P}{2(P + \sigma_k^2)} \). Next, plugging (D.1a) and (D.2a) into (D.4) we write:

\[
\alpha_K = \eta_{K-1} \alpha_{K-1} + \theta_{K-1} m_K - 1 r_{K-1}
\]

Hence, given \( W_{K-2}, \mathbf{m}, \eta_{K-1} \) and \( \theta_{K-1}, \alpha_K \) is a quadratic function of \( b_{K-1} \). This implies that the optimizing \( b_{K-1} \) is given by:

\[
b_{K-1} = \frac{d_{K-2} \left( \eta_{K-1} + \theta_{K-1} m_K - 1 m_{K-2} \right) \left( \alpha_{K-2} + m_K - 2 r_{K-2} \right)}{\eta_{K-1} \left( P + \sigma_k^2 \right) + \theta_{K-1} m_K - 1 m_{K-2} \left( P + \rho_k^2 \right)} \quad (D.6a)
\]

\[
= \sqrt{\frac{P \left( \eta_{K-1} + \theta_{K-1} m_K - 1 m_{K-2} \right) \left( \alpha_{K-2} + m_K - 2 r_{K-2} \right)}{\eta_{K-1} \left( P + \sigma_k^2 \right) + \theta_{K-1} m_K - 1 m_{K-2} \left( P + \rho_k^2 \right)}} \quad (D.6b)
\]

We note that (D.6b) holds with \( K - 1 \) replaced by \( k, k \leq K - 1 \), and \( K - 2 \) replaced by \( k - 1 \). In the following we derive the backwards calculation of \( \eta_k \) and \( \theta_k \). Hence, (D.6b) along with (D.3b) constitute (44). Note that given \( W_{K-2} \) and \( \mathbf{m}, b_{K-1} \) is a function of \( \eta_{K-1} \) and \( \theta_{K-1} \). Next, we plug (D.6b) back into (D.5) to obtain:

\[
\alpha_K = -\frac{d_{K-2} \left( \eta_{K-1} + \theta_{K-1} m_K - 1 m_{K-2} \right)^2 \left( \alpha_{K-2} + m_K - 2 r_{K-2} \right)}{\eta_{K-1} \left( P + \sigma_k^2 \right) + \theta_{K-1} m_K - 1 m_{K-2} \left( P + \rho_k^2 \right)} + \eta_{K-1} \alpha_{K-2} + \theta_{K-1} m_K - 1 r_{K-2}
\]

\[
= -\frac{P \left( \eta_{K-1} + \theta_{K-1} m_K - 1 m_{K-2} \right)^2 \left( \alpha_{K-2} + m_K - 2 r_{K-2} \right)}{2 \left( \eta_{K-1} \left( P + \sigma_k^2 \right) + \theta_{K-1} m_K - 1 m_{K-2} \left( P + \rho_k^2 \right) \right)} + \eta_{K-1} \alpha_{K-2} + \theta_{K-1} m_K - 1 r_{K-2}
\]

\[
= \alpha_{K-2} \left( \eta_{K-1} - \frac{P \left( \eta_{K-1} + \theta_{K-1} m_K - 1 m_{K-2} \right)^2}{2 \left( \eta_{K-1} \left( P + \sigma_k^2 \right) + \theta_{K-1} m_K - 1 m_{K-2} \left( P + \rho_k^2 \right) \right)} \right)
\]

\[
+ m_K - 2 r_{K-2} \left( \theta_{K-1} m_K - 1 m_{K-2} - \frac{P \left( \eta_{K-1} + \theta_{K-1} m_K - 1 m_{K-2} \right)^2}{2 \left( \eta_{K-1} \left( P + \sigma_k^2 \right) + \theta_{K-1} m_K - 1 m_{K-2} \left( P + \rho_k^2 \right) \right)} \right)
\]

\[
\triangleq \eta_{K-2} \alpha_{K-2} + \theta_{K-2} m_K - 2 r_{K-2}. \quad (D.7)
\]
Hence, (D.7) implies that the sequences \( \eta_k \) and \( \theta_k \), for \( k = K-1, K-2, \ldots, 1 \) obey the backwards recursive formulation given in (43), where \( \eta_{K-1} \) and \( \theta_{K-1} \) are provided just below (D.4). We conclude that given the sequences \( \eta_k \) and \( \theta_k \), and given \( \mathbf{m} \) and \( W_{K-1} \), the optimal coefficient \( b_k \) can be calculated via (D.6b), and then, the computed \( b_k \) can be used in the forward calculation (42). The optimal \( \alpha_K \) for the given modulation vector \( \mathbf{m} \) is given by (D.4).

### D.2 Optimality of \( P_k = P \) in the DP Scheme

In this subsection we prove that the optimal scaling \( d_k \), in the MMSE sense, results in \( P_k = P, \forall k \). We begin our analysis with \( d_{K-1} \), and recall that \( \mathbb{E}\{X_k^2\} = 2d_{K-1}^2(\alpha_{K-1} + m_{K-1}r_{K-1} - 1) + \sigma_z^2 \). Thus, rewriting (D.1b) for \( k = K \) we obtain:

\[
\alpha_K = \alpha_{K-1} + b_K^2(2d_{K-1}^2(\alpha_{K-1} + m_{K-1}r_{K-1} - 1) + \sigma_z^2) - 2b_K d_{K-1}(\alpha_{K-1} + m_{K-1}r_{K-1} - 1). \quad (D.8)
\]

Similarly, (D.3a) becomes:

\[
b_K = \frac{d_{K-1}(\alpha_{K-1} + m_{K-1}r_{K-1} - 1)}{2d_{K-1}^2(\alpha_{K-1} + m_{K-1}r_{K-1} - 1) + \sigma_z^2}. \quad (D.9)
\]

Next, we plug (D.9) into (D.8) to obtain:

\[
\alpha_K = \alpha_{K-1} - \frac{d_{K-1}^2(\alpha_{K-1} + m_{K-1}r_{K-1} - 1)^2}{2d_{K-1}^2(\alpha_{K-1} + m_{K-1}r_{K-1} - 1) + \sigma_z^2}. \quad (D.10)
\]

Now, fix \( \alpha_{K-1} \) and let \( r_{K-1} = \rho_{K-1} - k_{K-1} \), \( |rk_{K-1}| \leq 1 \), which implies that \( \alpha_{K-1} + m_{K-1}r_{K-1} - 1 = \alpha_{K-1}(1 + m_{K-1}r_{K-1}) - 1 \geq 0 \). Let \( \zeta_k \triangleq \alpha_k + m_kr_k \) and define \( x \triangleq d_k^2 - 1 \). We write:

\[
\frac{d_{K-1}^2(\alpha_{K-1} + m_{K-1}r_{K-1} - 1)^2}{2d_{K-1}^2(\alpha_{K-1} + m_{K-1}r_{K-1} - 1) + \sigma_z^2} = \frac{\zeta^2_k - 1}{\bar{\zeta}_k + \sigma_z^2} = f_0(x), \quad x, \bar{\zeta}_k \geq 0.
\]

It can be easily shown that \( \frac{d_0(x)}{dx} = \frac{\zeta^2_k - 1}{(\bar{\zeta}_k + \sigma_z^2)^2} \geq 0 \), which implies that \( f_0(x) \) is a monotonic non-decreasing function for \( x \geq 0 \). We conclude that for any given \( \alpha_{K-1} \) (D.10) is minimized when \( d_{K-1}^2 \) is maximized, i.e., \( d_{K-1} = \sqrt{\frac{\rho_{K-1}^2 + m_{K-1}r_{K-1}}{2(\alpha_{K-1} + m_{K-1}r_{K-1})}} \) and therefore \( P_K = P \), thus, satisfying the average per-symbol power constraint with equality.

Next, we consider the case of \( k = K-1 \), and since setting \( P_k = P \) is optimal we can use \( \eta_{K-1} \) and \( \theta_{K-1} \) given below (D.4). Recall that \( \mathbb{E}\{X_{K-1}^2\} = 2d_{K-1}^2(2\alpha_{K-1} + 2m_{K-1}r_{K-1} - 1) = 2d_{K-1}^2\zeta_{K-1} - 1 \). Hence, we rewrite (D.5) with \( P \) replaced by \( 2d_{K-2}^2\zeta_{K-2} \) to obtain:

\[
\alpha_K = b_{K-1}^2(\eta_{K-1}(2d_{K-2}^2\zeta_{K-2} + \sigma_z^2) + \theta_{K-1} - 1 - m_{K-1}r_{K-2} + 2d_{K-2}^2\zeta_{K-2} + \rho_2\sigma_z^2) + b_{K-2}\left(\eta_{K-2}(1 + \theta_{K-2} - 1 - m_{K-2}r_{K-2})\zeta_{K-2} + \eta_{K-1}\alpha_{K-2} + \theta_{K-1} - 1 - m_{K-1}r_{K-2}\right), \quad (D.11)
\]

where the optimal \( b_{K-1} \), in terms of \( d_{K-2}^2 \), is given by, see (D.6a):

\[
b_{K-1} = \frac{d_{K-2}^2(\eta_{K-1} + \theta_{K-1} - 1 - m_{K-1}r_{K-1} - 1)\zeta_{K-2}}{2d_{K-2}^2\zeta_{K-2}(\eta_{K-1} + \theta_{K-1} - 1 - m_{K-1}r_{K-2}) + \sigma_z^2(\eta_{K-1} + \theta_{K-1} - 1 - m_{K-1}r_{K-2})^2}. \quad (D.12)
\]

Plugging (D.12) into (D.11) we write:

\[
\alpha_K = -\frac{d_{K-2}^2(\eta_{K-2} + \theta_{K-2} - 1 - m_{K-2}r_{K-2} - 1)\zeta_{K-2}}{2d_{K-2}^2\zeta_{K-2}(\eta_{K-2} + \theta_{K-2} - 1 - m_{K-2}r_{K-2} - 1) + \sigma_z^2(\eta_{K-2} + \theta_{K-1} - 1 - m_{K-1}r_{K-2} - 1)^2 + \eta_{K-1}\alpha_{K-2} + \theta_{K-1} - 1 - m_{K-1}r_{K-2} - 1}.
\]

Again, \( \eta_{K-1}, \theta_{K-1}, \alpha_{K-1}, m_{K-1}, \) and \( r_{K-2} \) can be viewed as constants. Let \( \gamma_k \triangleq \eta_k + \theta_km_k \), \( \xi_k \triangleq \eta_k + \theta_km_km_k \), and \( \zeta_{K-2} \) can be viewed as constants. Let \( \gamma_k \triangleq \eta_k + \theta_km_k \), \( \xi_k \triangleq \eta_k + \theta_km_km_k \). We now write:

\[
-\frac{d_{K-2}^2(\eta_{K-2} + \theta_{K-2} - 1 - m_{K-2}r_{K-2} - 1)\zeta_{K-2}}{2d_{K-2}^2\zeta_{K-2}(\eta_{K-2} + \theta_{K-2} - 1 - m_{K-2}r_{K-2} - 1) + \sigma_z^2(\eta_{K-2} + \theta_{K-1} - 1 - m_{K-1}r_{K-2} - 1)^2} = \frac{\zeta_{K-2}^2\zeta_{K-1} - 1}{2\zeta_{K-2}\zeta_{K-1} + \xi_{K-1}\sigma_z^2} = f_1(x).
\]
Since \( \frac{dP_2(x)}{dx} = \frac{\xi_{K-1}^2 \sigma^2}{(2\xi_{K-1} - 2\xi_{K-1}^2 + \xi_{K-1}^2 \sigma^2)} \), we conclude that the sign of \( \frac{dP_2(x)}{dx} \) does not depend on \( x \), and therefore \( f_2(x) \) is monotonic. Now, if \( \xi_{K-1} > 0 \) then \( \alpha_K \) is minimized when \( d_{K-2}^* \) is maximized, i.e., \( P_{K-1} = P \). On the other hand, if \( \xi_{K-1} < 0 \) then \( \alpha_K \) is minimized when \( d_{K-2}^* = 0 \). Clearly, the case of \( d_{K-2} = 0 \) implies that \( P_{K-1} = 0 \), which cannot be optimal as it implies that \( \alpha_K = \alpha_{K-1} \). Finally, if \( \xi_{K-1} = 0 \) we have that \( \alpha_K \) is independent of \( d_{K-2} \) which clearly cannot hold. We conclude that the optimal choice of \( d_{K-2} \) is the one which results in \( P_{K-1} = P \). Furthermore, we note that similarly to Subsection D.1, the analysis for \( k = K - 1 \) holds for any \( k < K \), which implies that \( P_k = P \) is optimal for all values of \( K \).

### E Proof of Proposition 4

From Remark 11 it is clear that DP outperforms OL. Next, to compare DP with LQG we show that both schemes have the same structure of state evolution, transmitted signal, and decoders. Therefore, the LQG scheme is in the search range of DP.

**The DP scheme:** In the DP scheme the transmitted signal is given by (37):

\[
X_{k+1} = d_k (\epsilon_{1,k} + m_k \epsilon_{2,k}), \quad (E.1)
\]

where \( \epsilon_{i,k} \) evolve as:

\[
\begin{align*}
\epsilon_{1,k} &= \epsilon_{1,k-1} - b_{1,k} Y_{1,k}, \quad \epsilon_{2,k} = \epsilon_{2,k-1} - b_{2,k} Y_{2,k}. \quad (E.2)
\end{align*}
\]

Here, \( b_{1,k} = b_k \) and \( b_{2,k} = m_{k-1} b_k \). From (9), and similarly to [29, Eq. (7)], it follows that the DP scheme estimates the source \( S_i \) via \( \hat{S}_i = \sum_{m=1}^{K} b_{i,m} Y_{i,m} \). Note that in the DP scheme we optimize over the sequences \( \{d_k\}_{k=1}^K, \{b_k\}_{k=1}^K \) and \( \{m_k\}_{k=0}^{K-1} \).

**The LQG scheme:** We can write the transmitted signal of the LQG scheme in the following form, see Subsection 4.1:

\[
X_k = \tilde{c} (U_{1,k} - U_{2,k}), \quad (E.3)
\]

where \( \tilde{c} = -c_1/a \). Together with (E.3), the states \( U_{i,k} \) evolve as (15):

\[
\begin{align*}
U_{1,k} &= U_{1,k-1} + \frac{1}{a} Y_{1,k}, \quad U_{2,k} = -U_{2,k-1} + \frac{1}{a} Y_{2,k}. \quad (E.4)
\end{align*}
\]

Next, recall that the decoding in the LQG scheme is applied in two stages. First the state \( U_{i,k} \) is estimated as in (18), and then \( S_i \) is estimated from the estimated state. From (B.2) we have that the estimated state \( \hat{U}_{i,k+1} \) obeys:

\[
\hat{U}_{i,k+1} = \sum_{m=1}^{K} a_i^{k-m} Y_{i,m}. \quad (E.5)
\]

Furthermore, we note that for all three decoders (19), (21), and (32), \( S_i \) is estimated from \( \hat{U}_{i,k+1} \) via: \( \tau_{i,k} \hat{U}_{i,k+1} \), where \( \{\tau_{i,k}\}_{k=1}^K \) is a sequence which depended on the decoder in use. We emphasize that the sequences \( \{\tau_{i,k}\}_{k=1}^K \) are given and we do not optimize over their values. This implies that \( \hat{S}_{i,k} \) has the following form:

\[
\hat{S}_{i,k} = \sum_{m=1}^{K} \tau_{i,k} a_i^{k-m} Y_{i,m} = \sum_{m=1}^{K} \hat{\tau}_{i,m} Y_{i,m} \quad (E.6)
\]

for a known sequence \( \{\hat{\tau}_{i,k}\}_{k=1}^K \).

Therefore, as the transmitted signals, the state evolution, and the decoders has the same linear (recursive) structure, and in the DP scheme we optimize over the sequences \( \{d_k\}_{k=1}^K, \{b_k\}_{k=1}^K \) and \( \{m_k\}_{k=0}^{K-1} \), we conclude that DP outperforms LQG with each one of the decoders, as long as (5) is satisfied.
References


