The Entropy Gain of Linear Systems and Some of Its Implications

Milan S. Derpich [1,*] **, Matias Müller** [1,*] **and Jan Østergaard** [2,*]

1   Department of Electronic Engineering, Universidad Técnica Federico Santa María, Av. España 1680, Valparaíso 2390123, Chile
2   Department of Electronic Systems, Aalborg University, 9220 Aalborg, Denmark
*   Correspondence: milan.derpich@usm.cl (M.S.D.); matias.muller@alumnos.usm.cl (M.M.); jo@es.aau.dk (J.Ø.)

Abstract: We study the increase in per-sample differential entropy rate of random sequences and processes after being passed through a non minimum-phase (NMP) discrete-time, linear time-invariant (LTI) filter G. For LTI discrete-time filters and random processes, it has long been established by Theorem 14 in Shannon's seminal paper that this entropy gain, $\mathcal{G}(G)$, equals the integral of $\log |G(e^{j\omega})|$. In this note, we first show that Shannon's Theorem 14 does not hold in general. Then, we prove that, when comparing the input differential entropy to that of the entire (longer) output of G, the entropy gain equals $\mathcal{G}(G)$. We show that the entropy gain between equal-length input and output sequences is upper bounded by $\mathcal{G}(G)$ and arises if and only if there exists an output additive disturbance with finite differential entropy (no matter how small) or a random initial state. Unlike what happens with linear maps, the entropy gain in this case depends on the distribution of all the signals involved. We illustrate some of the consequences of these results by presenting their implications in three different problems. Specifically: conditions for equality in an information inequality of importance in networked control problems; extending to a much broader class of sources the existing results on the rate-distortion function for non-stationary Gaussian sources, and an observation on the capacity of auto-regressive Gaussian channels with feedback.

Keywords: differential entropy rate; non-minimum phase linear time-invariant systems; entropy loss in linear filters; networked control; rate-distortion for non-stationary sources feedback capacity

1. Introduction

We study the difference between the differential entropy rate of a random process $u_1^\infty = \{u_1, u_2, \ldots\}$ entering a discrete-time *linear time-invariant* (LTI) system G and the differential entropy rate of its (possibly noisy) output y_1^∞, as depicted in Figure 1.
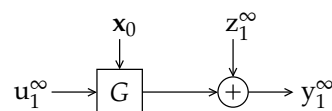
Figure 1. A causal, stable, linear and time-invariant system G with input and output processes, initial state, and output disturbance.

Recall that the differential entropy rate of a random process x_1^∞ is given by $\bar{h}(x_1^\infty) \triangleq \lim_{n\to\infty} n^{-1} h(x_1, x_2, \ldots, x_n)$, provided the limit exists, where $h(x_1, \ldots, x_n) = E[-\log f(x_1, \ldots, x_n)]$ is the differential entropy of the ensemble $x_1, \ldots, x_n$ with *probability density function* (PDF) f [1]. The system G is supposed to satisfy the following:

Assumption 1. *The LTI system G in Figure 1 is causal and stable and such that*

1.   *G has a rational p-th order transfer function $G(z)$ with m zeros $\{\rho_i\}_{i=1}^m$ outside the unit circle, i.e.,* non-minimum-phase *(NMP) zeros, where $m \in \{0, 1, \ldots, p\}$, indexed in non-increasing magnitude order, i.e., $|\rho_1| \geq |\rho_2| \geq \cdots \geq |\rho_m| > 1$.*

2. *The unit-impulse response of G, say, $g_0, g_1, \dots$ satisfies $|g_0| = 1$.*

In this general setup, G may have a random initial state vector $\mathbf{x}_0 \in \mathbb{R}^p$, $p \in \mathbb{N}$, and a real-valued random output disturbance z_1^∞. Our main purpose is to characterize the limit

$$\mathcal{G}(G, \mathbf{x}_0, \mathbf{u}_1^\infty, z_1^\infty) \triangleq \lim_{n \to \infty} \frac{1}{n}(h(\mathbf{y}_1^n) - h(\mathbf{u}_1^n)), \tag{1}$$

evaluating the possible effect produced by $\mathbf{x}_0$ and z_1^∞. This difference can be interpreted as the *entropy gain* (entropy amplification or entropy boost) introduced by the filter G and (as apparent from the other variables in the argument of $\mathcal{G}$) the statistics of $\mathbf{x}_0, \mathbf{u}_1^\infty, z_1^\infty$. We shall refer to the special case in which $\mathbf{x}_0$ and z_1^∞ are both zero (or deterministic) as the noise-less case, and write $\mathcal{G}(G, 0, \mathbf{u}_1^\infty, 0)$ accordingly.

The earliest reference related to this problem corresponds to a noise-less continuous-time counterpart considered by Shannon. In his seminal 1948 paper [2], Shannon gave a formula for the change in differential entropy per degree of freedom that a continuous-time random process $\mathbf{u}_c$, band-limited to a frequency range $[0, B)$ (in Hz), experiences after passing through an LTI continuous-time filter G_c (without considering a random initial state or an output disturbance). Such entropy per degree of freedom is defined in terms of uniformly taken samples as

$$\bar{h}(\mathbf{u}_c) \triangleq \lim_{n \to \infty} \frac{1}{n} h(\mathbf{u}_c(T), \mathbf{u}_c(2T), \dots, \mathbf{u}_c(nT)), \tag{2}$$

with $T \triangleq 1/(2B)$. In this formula, if the LTI filter has frequency response $G_c(\xi)$ (with ξ in Hz), then the resulting differential entropy rate of the output process $\mathbf{y}_c$ is given by the following theorem:

Theorem 1 (Reference [2], Theorem 14). *If an ensemble having an entropy $\bar{h}(\mathbf{u}_c)$ per degree of freedom in band B is passed through a filter with characteristic $G_c(\xi)$ the output ensemble has an entropy*

$$\bar{h}(\mathbf{y}_c) = \bar{h}(\mathbf{u}_c) + \frac{2}{B} \int_0^B \log|G_c(\xi)| d\xi. \tag{3}$$

Shannon arrived at (3) by arguing that an LTI filter can be seen as a linear operator that selectively scales its input signal along infinitely many frequencies, each of them representing an orthogonal component of the source. He then obtained the result by writing down the determinant of the Jacobian of this operator as the product of the squared frequency response magnitude of the filter over n frequency bands, applying logarithm, dividing by n, and then taking the limit as n tends to infinity.

Remark 1. *There is a factor of two in excess in the integral on the right-hand side (RHS) of (3). To see this, consider a filter with a constant gain a over $[0, B)$ (i.e., a simple multiplicative factor). In such case, the entropy rate of $\mathbf{y}_c$ should exceed that of $\mathbf{u}_c$ by $\log|a|$ [1]. However, (3) yields an entropy gain equal to $2\log|a|$. This error arises because the determinant of the Jacobian of the transformation is actually the product of $|G_c|$ over the n frequency bands considered in Shannon's argument. Such excess factor of two is also present in the entropy losses appearing in Reference [2], Table 1.*

Theorem 14 in Reference [2] has found application in works ranging from traditional themes, such as linear prediction [3] and source coding [4], to molecular communication systems [5,6].

The available literature treating the phenomenon itself of the entropy gain (loss, boost, or amplification) induced by LTI systems seems to be rather scarce. This is not surprising

given that (3) was published in Reference [2], Theorem 14, the work which gave birth to Information Theory.

The following publication concerned with this problem is Reference [7], following a time-domain analysis for the corresponding discrete-time problem. In this approach, one can obtain $y_1^n \triangleq \{y(1), y(2), \ldots, y(n)\}$ as a function of u_1^n, for every $n \in \mathbb{N}$, and evaluate the difference between the limits $\bar{h}(y_1^\infty)$ and $\bar{h}(u_1^\infty)$, obtained by letting $n \to \infty$. More precisely, for an LTI discrete-time filter G with impulse response $g_0^\infty = \{g_0, g_1, \ldots\}$, we can write

$$
\mathbf{y}_n^1 = \underbrace{\begin{pmatrix} g_0 & 0 & \cdots & 0 \\ g_1 & g_0 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ g_{n-1} & g_{n-2} & \cdots & g_0 \end{pmatrix}}_{G_n} \mathbf{u}_n^1,
\tag{4}
$$

where we adopt the notation $\mathbf{y}_n^1$ for column vectors to avoid the abuse of notation incurred by treating the sequence y_1^n as a vector, and because, by writing $\mathbf{y}_n^1$, it is easier to remember that its samples are ordered from top to bottom. $\mathbf{y}_n^1 \triangleq [y(1) \; y(2) \; \cdots \; y(n)]^T$ and the random vector $\mathbf{u}_n^1$ is defined likewise. From this, it is clear (see, e.g., the corollary after Theorem 8.6.4 in Reference [1]) that

$$
h(\mathbf{y}_n^1) = h(\mathbf{u}_n^1) + \log|\det(G_n)|,
\tag{5}
$$

where $\det(G_n)$ (or simply $\det G_n$) stands for the determinant of G_n. This result is utilized in Reference [7] to show that no entropy gain is produced by a stable minimum phase LTI system G if and only if the first sample in its impulse response has unit magnitude.

In Reference [8], p. 568, the entropy gain of a discrete-time LTI system G (the noise-less version of the setup depicted in Figure 1) is found to be

$$
\bar{h}(y_1^\infty) = \bar{h}(u_1^\infty) + \frac{1}{2\pi} \int_{-\pi}^{\pi} \log\left|G(e^{j\omega})\right| d\omega,
\tag{6}
$$

where y_1^∞ is the filter's discrete-time output process (without the effect of random initial state or an output disturbance) and

$$
\bar{h}(y_1^\infty) \triangleq \lim_{n \to \infty} \frac{1}{n} h(y_1^n).
\tag{7}
$$

This result was obtained starting from the fact that, for a Gaussian stationary process u_1^∞ with *power spectral density* (PSD) $S_u(e^{j\omega})$, $\bar{h}(u_1^\infty) = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_u(e^{j\omega})d\omega$. If u_1^∞ enters a discrete-time LTI system with frequency response $G(e^{j\omega})$, then the PSD of its output y_1^∞ is $S_y(e^{j\omega}) = S_u(e^{j\omega})|G(e^{j\omega})|^2$; thus, it is argued that (6) follows for Gaussian stationary inputs. Then, Reference [8] extends the result for non-Gaussian inputs with a proof sketch which uses a time-domain relation, like (4), to point out that the filter is a linear operator and, as such, the differential entropy of its output exceeds that of its input by a quantity that is independent of the input distribution. (It is worth noting that (6) is the discrete-time equivalent of (3) (without its wrong factor of 2), which follows directly from the correspondence between sampled band-limited continuous-time systems and discrete-time systems.)

It is in Reference [9], Section II-C, where, for the first time, it is shown that, for a stationary Gaussian input u_1^∞, the full entropy gain predicted by (6) takes place if the system output y_1^∞ is contaminated by an additive output disturbance of length p and positive definite covariance matrix, where p is the order of $G(z)$.

The integral $\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left|G(e^{j\omega})\right|d\omega$ can be related to the structure of the filter G. It is well known (from Jensen's formula) that if G has a causal and stable rational transfer function $G(z)$ and an impulse response with its first sample $g_0 \triangleq \lim_{z\to\infty} G(z)$, then

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left|G(e^{j\omega})\right|d\omega = \log|g_0| + \sum_{i:|\rho_i|>1}\log|\rho_i|, \tag{8}$$

where $\{\rho_i\}$ are the zeros of $G(z)$ (see, e.g., Reference [10,11]). This provides a straightforward formula to evaluate $\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left|G(e^{j\omega})\right|d\omega$ of a given LTI filter with rational transfer function $G(z)$. When combined with (6), this equation also reveals that if the entropy gain $\mathcal{G}(u_1^{\infty}, y_1^{\infty})$ is negative (i.e., if it corresponds to an entropy loss), then $|g_0| < 1$ (with the corresponding change of variables, this is the case in all the examples given by Shannon in Reference [2], Table 1). More importantly, (8) allows us to concentrate, without loss of generality, on LTI systems $G(z)$, whose first impulse-response sample has unit magnitude, as required by Assumption 1. Under the latter condition, (8) shows that the entropy gain is greater than zero if and only if $G(z)$ has zeros outside the unit disk $\mathbb{D} \triangleq \{\rho \in \mathbb{C} : |\rho| \le 1\}$. A system with the latter property is said to be *non-minimum phase* (NMP); conversely, a system with all its zeros inside $\mathbb{D}$ is said to be *minimum phase* (MP) [11].

1.1. Main Contributions of this Paper

The main contributions of this paper can be summarized as follows:

1.　Our first main result is showing that (6) and (3) do not hold for a large class of continuous-time filters and inputs. To see this, notice that

$$|g_0| = 1 \implies |\det(G_n)| = 1, \ \forall n \in \mathbb{N}, \tag{9}$$

which, in view of (5), is equivalent to $h(y_n^1) = h(u_n^1)$, $\forall n \in \mathbb{N}$. In turn, this implies that $\bar{h}(y_1^{\infty}) - \bar{h}(u_1^{\infty}) = 0$, regardless of whether $G(z)$ (i.e., the polynomial $g_0 + g_1 z^{-1} + \cdots$) has zeros with magnitude greater than one (choose, for example, $g_0 = 1$, $g_1 = 2$, and $g_k = 0$ for $k \ge 2$). This reveals that (4) holds if and only $G(z)$ is MP. But (6) and (3) are equivalent (correcting for the in excess factor of 2 discussed in Remark 1); thus, Theorem 14 in Reference [2] also does not hold for a class of continuous-time filters. However, the transfer function $G_c(s)$ of a band-limited continuous-time filter G_c is defined only for imaginary values of s (because the bilateral Laplace transform of $\sin(t)/t$ converges only on the imaginary axis), so one cannot classify such filters as MP or NMP. Instead, we consider a class of continuous-time filters limited to the frequencies in the band $[0, B)$, where $B > 0$ is in [Hz], defined by having a unit-impulse response of the form

$$g(t) \triangleq \sum_{k=0}^{\eta} g_k \phi_k(t), \tag{10}$$

for some absolutely summable sequence of real-valued coefficients $\{g_i\}_{i=0}^{\eta}$, $\eta = 1, 2, \ldots$, where the sinc functions

$$\phi_k(t) \triangleq \frac{\sin(2\pi B[t - k/(2B)])}{\pi[t - k/(2B)]}. \tag{11}$$

Since every such g satisfies $g(k/(2B)) = 0$ for $k < 0$, it makes sense to refer to such filters as "sample-wise causal". For this class of band-limited filters, we show that Theorem 14 holds if and only if the z-transform of $\{g_i\}_{i=0}^{\eta}$ is MP:

Theorem 2. *Suppose G_c is a low-pass continuous-time filter with unit-impulse response as in* (10). *Let the continuous-time random input of G_c be*

$$u_c(t) = \frac{1}{2B} \sum_{k=1}^{\infty} u(k)\phi_k(t), \tag{12}$$

for some random sequence $\{u(k)\}_{k=1}^{\infty}$, with ϕ_k as in (11), *and denote its output as y_c. Then,*

$$\bar{h}(y_c) - \bar{h}(u_c) = \log|g_0| = \log\left|\frac{1}{B}\int_0^B \Re\{G_c(\xi)\}d\xi\right| \overset{(a)}{\leq} \frac{1}{B}\int_0^B \log|G_c(\xi)|d\xi, \tag{13}$$

with equality in (a) *if and only if the polynomial $g_0 + g_1 z^{-1} + g_2 z^{-2} \cdots$ has no roots outside the unit circle.*

2. We show that $\frac{1}{2\pi}\int_{-\pi}^{\pi} \log|G(e^{j\omega})|d\omega$ actually corresponds to the entropy gain introduced by G but considering the new notion of effective differential entropy rate of y_1^{∞} proposed in this paper, defined next.

   Definition 1 (The Effective Differential Entropy). *Let $y \in \mathbb{R}^{\ell}$ be a random vector. If y can be written as a linear transformation $y = Su$, for some $u \in \mathbb{R}^n$ ($n \leq \ell$) with bounded differential entropy, $S \in \mathbb{R}^{\ell \times n}$, then the effective differential entropy of y is defined as*

   $$\check{h}(y) \triangleq h(Ay), \tag{14}$$

   where $S = A^T TC$ is an SVD for S, with $T \in \mathbb{R}^{n \times n}$.

   We can now state our second main result, the proof of which is in Appendix A:

   Theorem 3. *Let u_1^{∞} be the input of an LTI system G with transfer function $G(z)$ without zeros on the unit circle and with an absolutely summable unit impulse response $\{g_i\}_{i=0}^{\eta-1}$, with $\eta = \infty$ if G has an infinite impulse response. Denote the output of G as y_1^{∞}. Suppose $|h(u_1^n)| < \infty$ for every finite n. Then,*

   $$\lim_{n \to \infty} \frac{1}{n}\left(\check{h}(y_1^{n+\eta}(u_1^n)) - h(u_1^n)\right) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log|G(e^{j\omega})|d\omega, \tag{15}$$

   where $y_1^{n+\eta}(u_1^n)$ denotes the entire response of G to the input u_1^n.

   Theorem 3 states that, when considering the full-length output of a system, the effective entropy gain is introduced by the system itself.
   Section 4 provides a geometrical description of the phenomenon behind Definition 1 and Theorem 3.

3. We show that $\frac{1}{2\pi}\int_{-\pi}^{\pi} \log|G(e^{j\omega})|d\omega$ is a tight upper bound to the entropy gain of G (as defined in (1)), when the output is contaminated by some additional additive signal, such as a random initial state (represented by x_0 in Figure 1) or an output disturbance (such as z_1^{∞} in Figure 1), with sufficiently many degrees of freedom (a condition formally stated in Assumption 2 below). Moreover, we show that an entropy gain equal to the latter upper bound can appear even when these disturbances or random initial state have infinitesimally small variances. To the best of our knowledge, the latter phenomenon has been discussed in the literature first (and only) in Reference [9], Section II-C, for Gaussian stationary inputs and an LTI filter. We go beyond the latter result by explicitly and fully characterizing the entropy gain of LTI systems for a large class of not necessarily Gaussian nor stationary random input. We refer to this class as *entropy-balanced* processes, formally specified in the following definition:

Definition 2. *A random process* $\{v(k)\}_{k=1}^{\infty}$ *is said to be entropy balanced if the following two conditions are satisfied:*

(i)   Its sample variances $\sigma^2_{v(n)}$ *are finite for finite n and*

$$\lim_{n\to\infty} \frac{1}{n} \log(\sigma^2_{v(n)}) = 0.$$ (16)

(ii)   For every $\nu \in \mathbb{N}$ *and for every sequence of matrices* $\{\mathbf{\Phi}_n\}_{n=\nu+1}^{\infty}$, $\mathbf{\Phi}_n \in \mathbb{R}^{(n-\nu)\times n}$ *with orthonormal rows,*

$$\lim_{n\to\infty} \frac{1}{n} \left(h(\mathbf{\Phi}_n \mathbf{v}_n^1) - h(\mathbf{v}_n^1) \right) = 0.$$ (17)

The second condition guarantees that projecting an entropy-balanced process onto any subspace having finitely fewer dimensions yields a process with the same differential entropy rate.

The entropy gain induced by finite-length output disturbances is characterized by our next theorem.

Theorem 4. *In the system of Figure 1, let G satisfy Assumption 1 and suppose that* u_1^{∞} *is entropy balanced. Suppose the random output disturbance* z_1^{∞} *is such that* $z(i) = 0$, $\forall i > \kappa$, *and that* $|h(z_1^{\kappa})| < \infty$. *Let* $\bar{\kappa} \triangleq \min\{\kappa, m\}$, *where m is the number of NMP zeros of* $G(z)$. *Then,*

$$\sum_{i=m-\bar{\kappa}+1}^{m} \log|\rho_i| \leq \limsup_{n\to\infty} \frac{1}{n}(h(\mathbf{y}_n^1) - h(\mathbf{u}_n^1)) \leq \sum_{i=1}^{\bar{\kappa}} \log|\rho_i| \overset{(a)}{\leq} \frac{1}{2\pi} \int_{-\pi}^{\pi} \log|G(e^{j\omega})|d\omega$$ (18)

with equality in (a) *if and only if* $\kappa \geq m$.

The proof is presented in Section 6.4, and we provide geometrical insight explaining the phenomenon underlying Definition 2 and Theorem 4 in Section 5.1.

4.   We illustrate the relevance of the results summarized above by applying them to three problems in three areas, namely:

(a)   *Networked Control:* We show that equality holds in the inequality stated in Reference [12], Lemma 3.2 (a fundamental piece for the performance limitation results further developed in Reference [13]), under very general conditions. In addition, we extend the validity of a related equality for the perfect-feedback case, given by Reference [14], Theorem 14, for Gaussian signals, to the much larger class of entropy-balanced processes.

(b)   *The rate-distortion function for non-stationary Gaussian sources:* This problem has been previously solved in References [15–17]. We provide a simpler proof based upon the results described above. This proof extends the result stated in References [16,17] to a broader class of non-stationary sources.

(c)   *Gaussian channel capacity with feedback:* We show that capacity results based on using a short random sequence as channel input and relying on a feedback filter which boosts the entropy rate of the end-to-end channel noise (such as the one proposed in Reference [9]), crucially depend upon the complete absence of any additional disturbance anywhere in the system. Specifically, we show that the information rate of such capacity-achieving schemes drops to zero in the presence of any such additional disturbance. As a consequence, the relevance of characterizing the robust (i.e., in the presence of disturbances) feedback capacity of Gaussian channels, which appears to be a fairly unexplored problem, becomes evident.

1.2. Paper Outline

The remainder of this paper begins with some necessary definitions and preliminary results in Section 2. It continues with our detailed exposition in Section 3 of why Shannon's reasoning fails to yield the right expression for the entropy gain. We present an intuitive discussion leading to the definition of effective differential entropy in Section 4, which is ended by the proof of Theorem 3. Section 5 gives a geometric interpretation of how an arbitrarily small additive perturbation is able to boost the differential entropy rate of the process coming out of an NMP LTI filter. This exposition helps understanding and justifies the introduction of *entropy-balanced* random processes, which are also characterized there. Sections 6 and 7 contain our results for the entropy gain produced by an output disturbance and a random initial state, respectively. Our illustrative application results are presented in Section 8, followed by our conclusions in Section 9. Except when presented right after a statement or in its own section, all proofs are given in Appendix B.

2. Preliminaries

2.1. Notation

The sets of natural, real and complex numbers are denoted $\mathbb{N}$, $\mathbb{R}$, and $\mathbb{C}$, respectively. For a complex x, $\Re\{x\}$ is the real part of x. For a set $\mathcal{S}$, the indicator function $1_{\mathcal{S}}(x)$ equals 1 if $x \in \mathcal{S}$ and 0 otherwise. For any LTI system G, the transfer function $G(z)$ corresponds to the z-transform of the impulse response $g_0, g_1, \ldots$, i.e., $G(z) = \sum_{i=0}^{\infty} g_i z^{-i}$. For a transfer function $G(z)$, we denote by $G_n \in \mathbb{R}^{n \times n}$ the lower triangular Toeplitz matrix having $[g_0 \cdots g_{n-1}]^T$ as its first column. We write x_1^n as a shorthand for the sequence $\{x_1, \ldots, x_n\}$, and, when convenient, we write x_n^1 in vector form as $x_n^1 \triangleq [x_1 \, x_2 \, \cdots \, x_n]^T$, where $()^T$ denotes transposition. Random scalars (vectors) are denoted using non-italic characters, such as x (non-italic and boldface characters, such as **x**). The notation $x \perp\!\!\!\perp y$ means x and y are independent. If x and z are conditionally independent given y, we write $x \longleftrightarrow y \longleftrightarrow z$. For matrices, we use upper-case boldface symbols, such as A. We write $\lambda_i(A)$ to denote the i-th eigenvalue of A sorted in increasing magnitude. If $A \in \mathbb{C}^{m \times n}$, A^H is its conjugate transpose, and $\sigma_i(A) \triangleq \sqrt{\lambda_i(A^H A)}$, if $m \geq n$, and $\sigma_i(A) \triangleq \sqrt{\lambda_i(AA^H)}$, if $m < n$. We define $\sigma_{\min}(A) \triangleq \sigma_1(A)$ and $\sigma_{\max}(A) \triangleq \sigma_{\min\{m,n\}}(A)$. The term $A_{i,j}$ denotes the entry in the intersection between the i-th row and the j-th column. If $A \in \mathbb{C}^{m \times n}$, then A^T and A^* denote the transpose and conjugate transpose of A, respectively. We write $[A]_{i_2}^{i_1}$, with $1 \leq i_1 \leq i_2 \leq m$, to refer to the matrix formed by selecting the rows i_1 to i_2 of A. Likewise, for $1 \leq j_1 \leq j_2 \leq n$, $\overset{j_1 \, j_2}{\sqcap} A$ is the matrix built with columns j_1 to j_2 of A. The expression $^{m_1}[A]_{m_2}$ corresponds to the square sub-matrix along the main diagonal of A, with its top-left and bottom-right corners on A_{m_1,m_1} and A_{m_2,m_2}, respectively. A diagonal matrix whose entries are the elements in a set $\mathcal{D}$ (wherein elements may be repeated) is denoted as $\operatorname{diag} \mathcal{D}$. If $A \in \mathbb{R}^{n \times m_1}$ and $B \in \mathbb{R}^{n \times m_2}$, we write $[A|B] \in \mathbb{R}^{n \times (m_1+m_2)}$ to denote the augmented matrix built by placing the columns of A followed by those of B.

2.2. Mutual Information and Differential Entropy

Let **x**, y, and z be random variables with joint PDF $f_{\mathbf{x},y,z}$, and marginal PDFs $f_{\mathbf{x}}$, f_y, and f_z, respectively. The mutual information between x and y is defined as $I(x; y) \triangleq \int f_{x,y}(x, y) \log\left(\frac{f_{x,y}(x,y)}{f_x(x)f_y(y)}\right) dx dy$. The conditional mutual information between x and y given z is defined as $I(x; y \,|\, z) \triangleq \int f_{x,y,z}(x, y, z) \log\left(\frac{f_{x,y|z}(x,y|z)}{f_{x|z}(x|z)f_{y|z}(y|z)}\right) dx dy dz$, where $f_{x,y \,|\, z}$ is the joint PDF of x and y given z, and $f_{x \,|\, z}$, $f_{y \,|\, z}$ are defined likewise. The conditional differential entropy of x given y is defined as $h(x \,|\, y) \triangleq -\int f_{x,y}(x, y) \log(f_{x \,|\, y}(x|y)) dx dy$.

From these definitions, it is easy to verify the following properties Reference [1], Sections 2.4–2.6 and 8.4–8.6:

- Shift invariance: for every deterministic function f,

$$h(\mathrm{x} + f(\mathrm{y}) \mid \mathrm{y}) = h(\mathrm{x} \mid \mathrm{y}). \tag{19}$$

- Non-negativity:

$$I(\mathrm{x}; \mathrm{y}) \geq 0, \tag{20}$$

with equality if and only if x and y are independent.
- Chain Rule:

$$I(\mathrm{x}; \mathrm{y}, \mathrm{z}) = I(\mathrm{x}; \mathrm{y}) + I(\mathrm{x}; \mathrm{z} \mid \mathrm{y}). \tag{21}$$

- Relationship with entropy:

$$I(\mathrm{x}; \mathrm{y}) = h(\mathrm{x}) - h(\mathrm{x} \mid \mathrm{y}) = h(\mathrm{y}) - h(\mathrm{y} \mid \mathrm{x}). \tag{22}$$

2.3. System Model and Assumptions

Consider the discrete-time system depicted in Figure 1. In this setup, the block G satisfies Assumption 1.

It is worth noting that there is no loss of generality in considering $g_0 = 1$, since one can otherwise write $G(z)$ as $G'(z) = g_0 \cdot (G(z)/g_0)$; thus, the entropy gain introduced by $G'(z)$ would be $\log |g_0|$ plus the entropy gain due to $G(z)/g_0$ (in agreement with (6)), which has an impulse response with its first sample equal to 1.

The following assumption is made about the output disturbance z_1^∞:

Assumption 2. *The disturbance z_1^∞ is independent of u_1^∞ and belongs to a κ-dimensional linear subspace, for some finite $\kappa \in \mathbb{N}$. This subspace is spanned by the κ orthonormal columns of a matrix $\mathbf{\Phi} \in \mathbb{R}^{|\mathbb{N}| \times \kappa}$ (where $|\mathbb{N}|$ stands for the countably infinite size of $\mathbb{N}$), such that $|h(\mathbf{\Phi}^T \mathbf{z}_\infty^1)| < \infty$. Moreover, $\mathbf{z}_\infty^1 = \mathbf{\Phi} \mathbf{s}_\kappa^1$, where the random vector $\mathbf{s}_\kappa^1 \triangleq \mathbf{\Phi}^T \mathbf{z}_\infty^1$ has finite differential entropy, its covariance matrix $\mathbf{K}_{\mathbf{s}_\kappa^1}$ satisfies $\lambda_{max}(\mathbf{K}_{\mathbf{s}_\kappa^1}) < \infty$, and it is independent of $\mathbf{u}_\infty^1$.*

3. Revisiting Theorem 14 in Reference Shannon et al.

In this section, after presenting the proof of Theorem 2, we develop Shannon's approach into a more detailed and formal exposition. This allows us to explain why, for part of the continuous-time filters considered in Theorem 2, the approach chosen by Shannon to prove Theorem 14 in Reference [2] is unable to predict the correct value for the entropy gain.

3.1. Proof of Theorem 2

To begin with, the Fourier transform of ϕ_k is

$$\Phi_k(\xi) \triangleq \int_{-\infty}^{\infty} \phi_k(t) \, e^{-j2\pi\xi t} \, dt = 1_{[-B,B]}(\xi) \, e^{-j2\pi\xi k/(2B)}. \tag{23}$$

It is easy to verify that the functions ϕ_k satisfy the following orthogonality property:

$$\int_{-\infty}^{\infty} \phi_k(t)\phi_i(t)dt = \begin{cases} 2B & , \ k = i \\ 0 & , \ k \neq i \end{cases} \tag{24}$$

and

$$\phi_k(-t) = \phi_{-k}(t). \tag{25}$$

Notice that $\mathrm{u}(k) = \mathrm{u}_c\left(\frac{k}{2B}\right)$, $k \in \mathbb{N}$.

The output of G_c sampled at time $t = \ell/(2B)$, $\ell \in \mathbb{N}$, is

$$y(\ell) \triangleq y_c(\tfrac{\ell}{2B}) = \int_{-\infty}^{\infty} g(\tau)\, u_c(\tfrac{\ell}{2B} - \tau)d\tau \tag{26}$$

$$= \frac{1}{2B} \sum_{k=1}^{\infty} \sum_{i=0}^{\eta} g_i\, u(k) \int_{-\infty}^{\infty} \phi_i(\tau)\phi_k(\tfrac{\ell}{2B} - \tau)d\tau \tag{27}$$

$$= \frac{1}{2B} \sum_{k=1}^{\infty} \sum_{i=0}^{\eta} g_i\, u(k) \int_{-\infty}^{\infty} \phi_i(\tau)\phi_{\ell-k}(\tau)d\tau \tag{28}$$

$$= \sum_{i=0}^{\eta} g_i\, u(\ell - i), \tag{29}$$

with $u(k) = 0$ for $k \leq 0$. This means that the output samples y_1^{∞} are the discrete-time convolution between u_1^{∞} and the filter coefficients $\{g_i\}_{i=0}^{\eta}$. Therefore, the matrix relation (4) holds. We then obtain that $\bar{h}(y_c) = \bar{h}(u_c) + \log|g_0|$.

The frequency response of G_c is given by

$$G_c(\xi) = \int_{-\infty}^{\infty} g(t)\, e^{-j2\pi\xi t}\, dt = \sum_{k=0}^{\eta} g_k \Phi_k(\xi) = \sum_{k=0}^{\eta} g_k\, e^{-j\pi\xi k/B}, \tag{30}$$

where ξ is in [Hz]. This means that

$$g_0 = \frac{1}{2B} \int_{-B}^{B} G_c(\xi)d\xi = \frac{1}{B} \int_{0}^{B} \Re\{G_c(\xi)\}d\xi, \tag{31}$$

where the last equality holds because $G_c(\xi)$ is conjugate symmetric. Thus, the entropy gain introduced by G_c is the right-hand side of (13), concluding the proof. □

3.2. Formalizing Shannon's Argument

In the approach followed by Shannon, it is argued that the entropy gain is the limit as $n \to \infty$ of $n^{-1}\sum_{r=0}^{n-1} \log|G_c(\xi_r)|$ over uniformly spaced frequencies $\xi_0, \ldots, \xi_{n-1}$. Here, we show that this summation corresponds to $\log|\det(\widetilde{G}_n)|$, where $\widetilde{G}_n$ is an n-by-n Toeplitz circulant matrix. Moreover, the sequences of Hermitian matrices $\{G_n G_n^*\}_{n=1}^{\infty}$ and $\{\widetilde{G}_n \widetilde{G}_n^*\}_{n=1}^{\infty}$ are asymptotically equivalent (as defined in Reference [18], Section 2.3), which would yield $\lim_{n \to \infty} n^{-1}\log|\det(G_n)| = \lim_{n \to \infty} n^{-1}\log|\det(\widetilde{G}_n)|$ if the eigenvalues of $G_n G_n^*$ were bounded between constants $0 < \zeta_m < \zeta_M < \infty$ for all $n \in \mathbb{N}$. However, if $G(z)$ (the z-transform of $\{g_k\}_{k=0}^{\infty}$) has NMP zeros, then $G_n G_n^*$ has eigenvalues tending to zero exponentially as $n \to \infty$, which precludes these two limits to coincide.

To prove the above claims, we first apply the change of variable $\omega \triangleq \pi\xi/B$, with which (30) becomes

$$G_c(B\omega/\pi) = G(e^{j\omega}) \triangleq \sum_{k=0}^{\eta} g_k\, e^{-j\omega k}, \tag{32}$$

where $G(e^{j\omega})$ is the frequency response of the discrete-time filter G with unit-impulse response $\{g_i\}_{i=1}^{\eta}$ and ω is in radians per second. Now, following Shannon's approach, we uniformly sample $G(e^{j\omega})$ at n frequencies

$$\omega_r \triangleq \begin{cases} r\frac{2\pi}{n} & ,\ r/n \leq 0.5 \\ r\frac{2\pi}{n} - 2\pi & ,\ r/n > 0.5 \end{cases} ,\ r = 0, 1, \ldots, n-1, \tag{33}$$

which, from (32), yields the spectral samples

$$G(e^{-j\omega_r}) = \sum_{k=0}^{\eta} g_k\, e^{-j\frac{2\pi}{n}rk}. \tag{34}$$

We will cast the reason why (3) fails to coincide with the correct expression for the entropy gain provided by (5) as a disagreement between the asymptotic behavior of the logarithm of the determinant of two sequences of asymptotically equivalent matrices. For that purpose, since (34) coincides with Reference [18], Equation 4.34, we have that the spectral samples $\{G(e^{-j\omega_r})\}_{r=0}^{n-1}$ are the eigenvalues of the Toeplitz circulant matrix (Reference [18], Chapter 3)

$$\widetilde{G}_n \triangleq \begin{pmatrix} \tilde{g}_{n,0} & \tilde{g}_{n,n-1} & \cdots & \tilde{g}_{n,1} \\ \tilde{g}_{n,1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \tilde{g}_{n,n-1} \\ \tilde{g}_{n,n-1} & \ddots & \ddots & \tilde{g}_{n,0} \end{pmatrix} = U_n^* \operatorname{diag}\left\{ G(e^{-j\omega_0}), \ldots, G(e^{-j\omega_{n-1}}) \right\} U_n, \tag{35}$$

where $U_n \in \mathbb{C}^{n \times n}$ is the n-point *discrete Fourier transform* (DFT) matrix, defined as

$$[U_n]_{k,r} \triangleq \frac{1}{\sqrt{n}} e^{-j\frac{2\pi}{n}kr}, \quad k, r = 0, 1, \ldots, n-1. \tag{36}$$

From Reference [18], Lemma 4.5, $\tilde{g}_{n,k} \triangleq \sum_{i \in \mathbb{N}_0 : k+ni \leq \eta} g_{k+in}$, corresponding to the (possibly) aliased impulse response $g_0, g_1, \ldots, g_\eta$ as a result of sampling in frequency.

We can now see that the discrepancy between the entropy gain predicted by (3) and (5) is the disagreement between the following limits:

$$\lim_{n \to \infty} \frac{1}{2n} \log(\det(G_n G_n^*)) \overset{(31)}{=} \log\left| \frac{1}{B} \int_0^B \Re\{G_c(\xi)\} d\xi \right|, \tag{37a}$$

$$\lim_{n \to \infty} \frac{1}{2n} \log(\det(\widetilde{G}_n \widetilde{G}_n^*)) = \frac{1}{B} \int_0^B \log|G_c(\xi)| d\xi, \tag{37b}$$

where, due to (8), the expressions on both right-hand sides differ if and only if $G(z)$ has NMP zeros. According to Reference [18], Lemma 4.6, the sequences $\{G_n\}_{n=1}^\infty$ and $\{\widetilde{G}_n\}_{n=1}^\infty$ are asymptotically equivalent, which is written as $G_n \sim \widetilde{G}_n$. Then, from Reference [18], Theorem 2.1, the Hermitian matrices $G_n G_n^* \sim \widetilde{G}_n \widetilde{G}_n^*$, which, from Reference [18], Theorem 2.4, implies that

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n f(\lambda_i(G_n G_n^*)) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n f(\lambda_i(\widetilde{G}_n \widetilde{G}_n^*)), \tag{38}$$

for any function f continuous over a finite interval $[\zeta_m, \zeta_M]$ such that

$$\zeta_m \leq \lambda_i(G_n G_n^*), \lambda_i(\widetilde{G}_n \widetilde{G}_n^*) \leq \zeta_M, \quad i = 1, 2, \ldots, n, \ n = 1, 2, \ldots. \tag{39}$$

However, when $G(z)$ has m NMP zeros, Lemma 7 (in Section 6.3) establishes that there are exactly m eigenvalues of G_n that tend to zero exponentially as $n \to \infty$. Crucially, $\log(\cdot)$ is discontinuous at 0, which precludes the limits in (37) from coinciding.

4. The Effective Differential Entropy

Theorem 3 establishes that the effective differential entropy rate of the entire or complete output of an LTI system exceeds that of the (shorter) input sequence by the RHS of (15). This section provides a geometrical interpretation of this problem and intuition about the *effective differential entropy* already introduced in Definition 1.

Consider the random vectors $\mathbf{u} \triangleq [u_1 \ u_2]^T$ and $\mathbf{y} \triangleq [y_1 \ y_2 \ y_3]^T$ related via

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \underbrace{\begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 0 & 2 \end{pmatrix}}_{\triangleq \breve{G}_2} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}. \tag{40}$$

Suppose $\mathbf{u}$ is uniformly distributed over $[0,1] \times [0,1]$. Applying the conventional definition of differential entropy of a random sequence, we would have that

$$h(y_1, y_2, y_3) = h(y_1, y_2) + h(y_3 \mid y_1, y_2) = -\infty \tag{41}$$

because y_3 is a deterministic function of y_1 and y_2:

$$y_3 = [0 \ 2][u_1 \ u_2]^T = [0 \ 2]\begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}^{-1}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

In other words, the problem lies in that, although the output is a three-dimensional vector, it only has two degrees of freedom, i.e., it is restricted to a 2-dimensional subspace of $\mathbb{R}^3$. This is illustrated in Figure 2, where the set $[0,1] \times [0,1]$ is shown (coinciding with the u-v plane), together with its image through $\breve{G}_2$ (as defined in (40)).
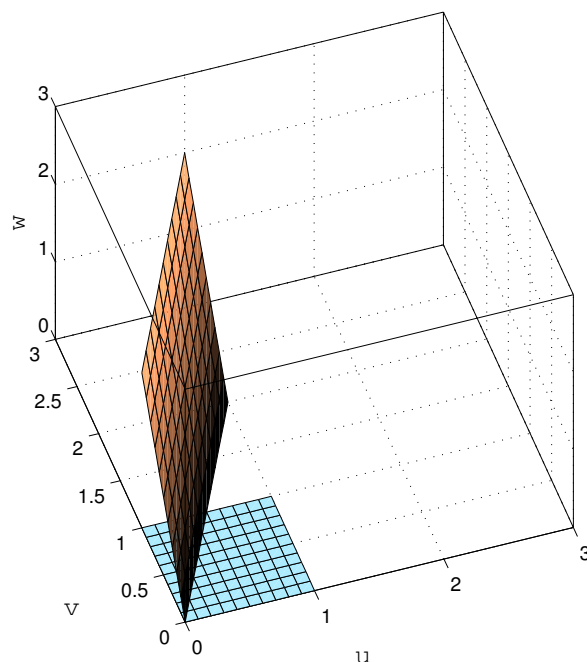
Figure 2. Support of $\mathbf{u}$ (laying in the u-v plane) compared to that of $\mathbf{y} = \breve{G}\mathbf{u}$ (the rhombus in $\mathbb{R}^3$).

As can be seen in this figure, the image of the square $[0,1]^2$ through $\breve{G}_2$ is a 2-dimensional rhombus over which $\{y_1, y_2, y_3\}$ distributes uniformly. Since the intuitive notion of differential entropy of an ensemble of random variables relates to the size of the region spanned by the associated random vector (and determines how difficult it is to compress it in a lossy fashion with a given precision), one could argue that the differential entropy of $\{y_1, y_2, y_3\}$, far from being $-\infty$, should be somewhat larger than that of $\{u_1, u_2\}$ (since the rhombus $\breve{G}_2[0,1]^2$ has a larger area than $[0,1]^2$). So, what does it mean that (and why should) $h(y_1, y_2, y_3) = -\infty$? Simply put, the differential entropy relates to the volume spanned by the support of the probability density function. For $\mathbf{y}$ in our example, the latter (three-dimensional) volume is clearly zero.

From the above discussion, the comparison between the differential entries of $\mathbf{y} \in \mathbb{R}^3$ and $\mathbf{u} \in \mathbb{R}^2$ of our previous example should take into account that $\mathbf{y}$ actually lives in a two-dimensional subspace of $\mathbb{R}^3$. Indeed, since the multiplication by a unitary matrix does not alter differential entries, we could consider the differential entropy of

$$\begin{bmatrix} \tilde{\mathbf{y}} \\ 0 \end{bmatrix} \triangleq \begin{pmatrix} \check{Q} \\ \bar{q}^T \end{pmatrix} \mathbf{y}, \tag{42}$$

where $\check{Q}^T$ is the 3×2 matrix with orthonormal rows in the SVD of $\check{G}_2$

$$\check{G}_2 = \check{Q}^T \check{D} \check{R}, \tag{43}$$

and $\bar{q}$ is a unit-norm vector orthogonal to the rows of $\check{Q}$ (and thus orthogonal to $\mathbf{y}$, as well). We are now able to compute the differential entropy in $\mathbb{R}^2$ for $\tilde{\mathbf{y}}$, corresponding to the rotated version of $\mathbf{y}$ such that its support is now aligned with $\mathbb{R}^2$.

The preceding discussion motivates the use of a modified version of the notion of differential entropy for a random vector $\mathbf{y} \in \mathbb{R}^n$ which considers the number of dimensions actually spanned by $\mathbf{y}$ instead of its length.

It is worth mentioning that Shannon's differential entropy of a vector $\mathbf{y} \in \mathbb{R}^\ell$, whose support's ℓ-volume is greater than zero, arises from considering it as the difference between its (absolute) entropy and that of a random variable uniformly distributed over an ℓ-dimensional, unit-volume region of $\mathbb{R}^\ell$. More precisely, if in this case the *probability density function* (PDF) of $\mathbf{y} = [\mathrm{y}_1 \ \mathrm{y}_2 \ \cdots \ \mathrm{y}_\ell]^T$ is Riemann integrable, then [1], Thm. 9.3.1,

$$h(\mathbf{y}) = \lim_{\Delta \to 0} \left[H(\mathbf{y}^\Delta) + \ell \log \Delta \right], \tag{44}$$

where $\mathbf{y}^\Delta$ is the discrete-valued random vector resulting when $\mathbf{y}$ is quantized using an ℓ-dimensional uniform quantizer with ℓ-cubic quantization cells with volume Δ^ℓ. However, if we consider a variable $\mathbf{y}$ whose support belongs to an n-dimensional subspace of $\mathbb{R}^\ell$, $n < \ell$ (i.e., $\mathbf{y} = S\mathbf{u} = A^T TC\mathbf{u}$, as in Definition 1), then the entropy of its quantized version in $\mathbb{R}^\ell$, say $H_\ell(\mathbf{y}^\Delta)$, is distinct from $H_n((A\mathbf{y})^\Delta)$, the entropy of $A\mathbf{y}$ in $\mathbb{R}^n$. Moreover, it turns out that, in general,

$$\lim_{\Delta \to 0} \left(H_\ell(\mathbf{y}^\Delta) - H_n((A\mathbf{y})^\Delta) \right) \neq 0, \tag{45}$$

despite the fact that A has orthonormal rows. Thus, the definition given by (44) does not yield consistent results for the case wherein a random vector has a support's dimension (i.e., its number of degrees of freedom) smaller that its length (The mentioned inconsistency refers to (45).), which reveals that the asymptotic behavior $H_\ell(\mathbf{y}^\Delta)$ changes if $\mathbf{y}$ is rotated. (If this were not the case, then we could redefine (44) replacing ℓ by n, in a spirit similar to the one behind Renyi's d-dimensional entropy [19].) To see this, consider the case in which $\mathbf{u} \in \mathbb{R}$ distributes uniformly over $[0, 1]$ and $\mathbf{y} = [1 \ \ 1]^T \mathbf{u}/\sqrt{2}$. Clearly, $\mathbf{y}$ distributes uniformly over the unit-length segment connecting the origin with the point $(1, 1)/\sqrt{2}$. Then,

$$H_2(\mathbf{y}^\Delta) = -\left\lfloor \tfrac{1}{\Delta\sqrt{2}} \right\rfloor \Delta\sqrt{2} \log\left(\Delta\sqrt{2}\right) - \left(1 - \left\lfloor \tfrac{1}{\Delta\sqrt{2}} \right\rfloor \sqrt{2}\Delta\right) \log\left(1 - \left\lfloor \tfrac{1}{\Delta\sqrt{2}} \right\rfloor \sqrt{2}\Delta\right). \tag{46}$$

On the other hand, since, in this case, $A\mathbf{y} = \mathbf{u}$, we have that

$$H_1((A\mathbf{y})^\Delta) = H_1(\mathbf{u}^\Delta) = -\left\lfloor \tfrac{1}{\Delta} \right\rfloor \Delta \log \Delta - (1 - \left\lfloor \tfrac{1}{\Delta} \right\rfloor \Delta) \log(1 - \left\lfloor \tfrac{1}{\Delta} \right\rfloor \Delta). \tag{47}$$

Thus, the d-dimensional entropy would not generally be equal to the effective differential entropy, that is:

$$\lim_{\Delta \to 0} \left(H_1((A\mathbf{y})^\Delta) - H_2(\mathbf{y}^\Delta) \right) = \lim_{\Delta \to 0} \left(\left\lfloor \tfrac{1}{\Delta\sqrt{2}} \right\rfloor \Delta\sqrt{2} \log\left(\Delta\sqrt{2}\right) - \left\lfloor \tfrac{1}{\Delta} \right\rfloor \Delta \log \Delta \right) = \log \sqrt{2}. \tag{48}$$

The latter example further illustrates why the notion of effective entropy is appropriate in the setup considered in this section, where the effective dimension of the random sequences does not coincide with their length (it is easy to verify that the effective entropy of $\mathbf{y}$ does not change if one rotates $\mathbf{y}$ in $\mathbb{R}^\ell$).

We finish this section with an example to illustrate the usefulness of the notion of effective differential entropy beyond the context of entropy gain.

Application Example: Shannon Lower Bound

The rate-distortion function (RDF) $R(D)$ is the infimum, among all codes, of the expected number of bits per sample necessary to reconstruct a given random source with distortion not greater than D [1]. Let the source and reconstruction be the vectors $\mathbf{x}_\ell^1$ and $\mathbf{x}_\ell^1 + \mathbf{v}_\ell^1$, respectively, and suppose the distortion is assessed using the *mean-squared error* (MSE) $d(\mathbf{v}_\ell^1) \triangleq \mathrm{E}[\|\mathbf{v}_\ell^1\|^2]$. Then, restricting our attention to uniquely-decodable codes Reference [1], p. 105), the *Shannon Lower Bound* (SLB) [20] establishes that

$$\ell R(D) \geq h(\mathbf{x}_\ell^1) - \max_{d(\mathbf{v}_\ell^1) \leq D} h(\mathbf{v}_\ell^1), \tag{49}$$

provided $h(\mathbf{x}_\ell^1)$ is bounded. Therefore, if $\mathbf{x}_\ell^1$ is the entire forced response of an FIR filter G of order p to an input $\mathbf{u}_1^n$, then $\ell = n + p$ and $h(\mathbf{x}_\ell^1)$ is minus infinity, which precludes one from using (49). We will show next that, in this case, the SLB can still be stated by using the effective differential entropy $\breve{h}(\mathbf{x}_\ell^1)$ instead of $h(\mathbf{x}_\ell^1)$. Following Definition 1, we can write the source vector as $\mathbf{x}_\ell^1 = \boldsymbol{A}^T \boldsymbol{T} \boldsymbol{C} \mathbf{u}_n^1$, where $\boldsymbol{A} \in \mathbb{R}^{n \times \ell}$ has orthonormal rows, $\boldsymbol{T} \in \mathbb{R}^{n \times n}$ is diagonal with non-negative entries, and $\boldsymbol{C} \in \mathbb{R}^{n \times n}$ is unitary. Let $\boldsymbol{H} \triangleq [\boldsymbol{A}^T \mid \overline{\boldsymbol{A}}^T] \in \mathbb{R}^{\ell \times \ell}$ be a unitary matrix, which means that $\overline{\boldsymbol{A}} \boldsymbol{A}^T = \mathbf{0}_{p \times \ell}$. Then,

$$\ell R(D) \overset{(a)}{\geq} I(\mathbf{x}_\ell^1; \mathbf{x}_\ell^1 + \mathbf{v}_\ell^1), \tag{50}$$

$$= I(\boldsymbol{H}\mathbf{x}_\ell^1; \boldsymbol{H}\mathbf{x}_\ell^1 + \boldsymbol{H}\mathbf{v}_\ell^1), \tag{51}$$

$$= I(\boldsymbol{A}\mathbf{x}_\ell^1; \boldsymbol{A}\mathbf{x}_\ell^1 + \boldsymbol{H}\mathbf{v}_\ell^1), \tag{52}$$

$$= I(\boldsymbol{A}\mathbf{x}_\ell^1; \boldsymbol{A}\mathbf{x}_\ell^1 + \boldsymbol{A}\mathbf{v}_\ell^1, \overline{\boldsymbol{A}}\mathbf{v}_\ell^1), \tag{53}$$

$$\overset{(21)}{=} I(\boldsymbol{A}\mathbf{x}_\ell^1; \boldsymbol{A}\mathbf{x}_\ell^1 + \boldsymbol{A}\mathbf{v}_\ell^1), + I(\boldsymbol{A}\mathbf{x}_\ell^1; \overline{\boldsymbol{A}}\mathbf{v}_\ell^1 \mid \boldsymbol{A}\mathbf{x}_\ell^1 + \boldsymbol{A}\mathbf{v}_\ell^1), \tag{54}$$

$$\overset{(20)}{\geq} I(\boldsymbol{A}\mathbf{x}_\ell^1; \boldsymbol{A}\mathbf{x}_\ell^1 + \boldsymbol{A}\mathbf{v}_\ell^1), \tag{55}$$

$$\overset{(22)}{=} h(\boldsymbol{A}\mathbf{x}_\ell^1) - h(\boldsymbol{A}\mathbf{x}_\ell^1 \mid \boldsymbol{A}\mathbf{x}_\ell^1 + \boldsymbol{A}\mathbf{v}_\ell^1), \tag{56}$$

$$\overset{(19)}{=} h(\boldsymbol{A}\mathbf{x}_\ell^1) - h(\boldsymbol{A}\mathbf{v}_\ell^1 \mid \boldsymbol{A}\mathbf{x}_\ell^1 + \boldsymbol{A}\mathbf{v}_\ell^1), \tag{57}$$

$$\overset{(b)}{\geq} h(\boldsymbol{A}\mathbf{x}_\ell^1) - h(\boldsymbol{A}\mathbf{v}_\ell^1), \tag{58}$$

$$\geq h(\boldsymbol{A}\mathbf{x}_\ell^1) - \max_{\mathbf{w}_n^1 : d(\mathbf{w}_n^1) \leq D} h(\mathbf{w}_n^1), \tag{59}$$

$$\overset{(c)}{=} \breve{h}(\mathbf{x}_\ell^1) - \max_{\mathbf{w}_n^1 : d(\mathbf{w}_n^1) \leq D} h(\mathbf{w}_n^1), \tag{60}$$

where (a) stems from Reference [1], Theorems 5.4.1 and 5.5.1 and Equations (10).58-10.61, (b) holds because conditioning does not increase entropy and (c) is from the definition of effective differential entropy.

5. Entropy-Balanced Processes: Geometric Interpretation and Properties

In the first part of this section, we provide a geometric interpretation of the effect that a non-minimum phase LTI system has on its input random process. This will give an intuitive meaning to the notion of an entropy-balanced random process (introduced in Definition 2 above) and provide insights into why and how the entropy gain defined

in (1) arises as a consequence of an output random disturbance or a random initial state (the themes of Sections 6 and 7, respectively).

The second part of this section identifies several entropy-balanced processes and establishes two properties satisfied by this class of processes.

5.1. Geometric Interpretation

We begin our discussion with a simple example.

Example 1. *Suppose that G in Figure 1 is a finite impulse response (FIR) filter with impulse response $g_0 = 1$, $g_1 = 2$, $g_i = 0$, $\forall i \geq 2$. Notice that this choice yields $G(z) = (z - 2)/z$; thus, $G(z)$ has one non-minimum phase zero, at $z = 2$. The associated matrix $\boldsymbol{G}_n$ for $n = 3$ is*

$$\boldsymbol{G}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 2 & 1 \end{pmatrix},$$

whose determinant is clearly one (indeed, all its eigenvalues are 1). Hence, as discussed in the introduction, $h(\boldsymbol{G}_3 \mathbf{u}_3^1) = h(\mathbf{u}_3^1)$; thus, $\boldsymbol{G}_3$ (and $\boldsymbol{G}_n$, in general) does not introduce an entropy gain by itself. However, an interesting phenomenon becomes evident by looking at the SVD of $\boldsymbol{G}_3$, given by $\boldsymbol{G}_3 = \boldsymbol{Q}_3^T \boldsymbol{D}_3 \boldsymbol{R}_3$, where $\boldsymbol{Q}_3$ and $\boldsymbol{R}_3$ are unitary matrices, and $\boldsymbol{D}_3 \triangleq \mathrm{diag}\{d_1, d_2, d_3\}$. In this case, $\boldsymbol{D}_3 = \mathrm{diag}\{0.19394, 1.90321, 2.70928\}$; thus, one of the singular values of $\boldsymbol{G}_3$ is much smaller than the others (although the product of all singular values yields 1, as expected). As will be shown in Section 6, for a stable $G(z)$, such uneven distribution of singular values arises only when $G(z)$ has non-minimum phase zeros. The effect of this can be visualized by looking at the image of the cube $[0, 1]^3$ through $\boldsymbol{G}_3$, shown in Figure 3.
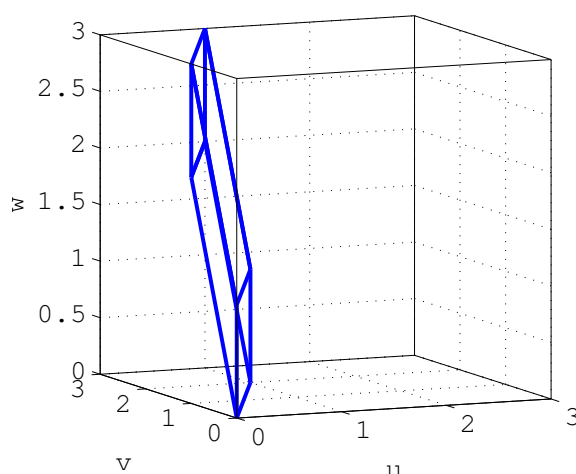

Figure 3. Image of the cube $[0, 1]^3$ through the square matrix with columns $[1\ 2\ 0]^T$, $[0\ 1\ 2]^T$, and $[0\ 0\ 1]^T$.

If the input $\mathbf{u}_3^1$ were uniformly distributed over this cube (of unit volume), then $\boldsymbol{G}_3 \mathbf{u}_3^1$ would distribute uniformly over the unit-volume parallelepiped depicted in Figure 3; hence, $h(\boldsymbol{G}_3 \mathbf{u}_3^1) = h(\mathbf{u}_3^1)$.

Now, if we add to $\boldsymbol{G}_3 \mathbf{u}_3^1$ a disturbance $\mathbf{z}_3^1 = \boldsymbol{\Phi} \mathrm{s}$, with scalar s uniformly distributed over $[-0.5, 0.5]$ independent of $\mathbf{u}_3^1$, and with $\boldsymbol{\Phi} \in \mathbb{R}^{3 \times 1}$, the effect would be to "thicken" the support over which the resulting random vector $\mathbf{y}_3^1 = \boldsymbol{G}_3 \mathbf{u}_3^1 + \mathbf{z}_3^1$ is distributed, along the direction pointed by $\boldsymbol{\Phi}$. If $\boldsymbol{\Phi}$ is aligned with the direction along which the support of $\boldsymbol{G}_3 \mathbf{u}_3^1$ is thinnest (given by $\boldsymbol{q}_{3,1}$, the first row of $\boldsymbol{Q}_3$), then the resulting support would have its volume significantly increased, which can be associated with a large increase in the differential entropy of $\mathbf{y}_3^1$ with respect to $\mathbf{u}_3^1$. Indeed, a relatively small variance of s and an approximately aligned $\boldsymbol{\Phi}$ would still produce a significant entropy gain.

The above example suggests that the entropy gain from $\mathbf{u}_n^1$ to $\mathbf{y}_n^1$ appears as a combination of two factors. The first of these is the uneven way in which the random vector $G_n\mathbf{u}_n^1$ is distributed over $\mathbb{R}^n$. The second factor is the alignment of the disturbance vector $\mathbf{z}_n^1$ with respect to the span of the subset $\{q_{n,i}\}_{i\in\Omega_n}$ of columns of Q_n, associated with the smallest singular values of G_n, indexed by the elements in the set Ω_n. As we shall discuss in the next section, if G has m non-minimum phase zeros, then, as n increases, there will be m singular values of G_n going to zero exponentially. Since the product of the singular values of G_n equals 1 for all n, it follows that $\prod_{i\notin\Omega_n} d_{n,i}$ must grow exponentially with n, where $d_{n,i}$ is the i-th diagonal entry of D_n. This implies that $G_n\mathbf{u}_n^1$ expands with n along the span of $\{q_{n,i}\}_{i\notin\Omega_n}$, compensating its shrinkage along the span of $\{q_{n,i}\}_{i\in\Omega_n}$, thus keeping $h(G_n\mathbf{u}_n^1) = h(\mathbf{u}_n^1)$ for all n. Thus, as n grows, any small disturbance distributed over the span of $\{q_{n,i}\}_{i\in\Omega_n}$, added to $G_n\mathbf{u}_n^1$, will keep the support of the resulting distribution from shrinking along this subspace. Consequently, the expansion of $G_n\mathbf{u}_n^1$ with n along the span of $\{q_{n,i}\}_{i\notin\Omega_n}$ is no longer compensated, yielding an entropy increase proportional to $\log(\prod_{i\notin\Omega_n} d_{n,i})$.

The above analysis allows one to anticipate a situation in which no entropy gain would take place even when some singular values of G_n tend to zero as $n \to \infty$. Since the increase in entropy is made possible by the fact that, as n grows, the support of the distribution of $G_n\mathbf{u}_n^1$ shrinks along the span of $\{q_{n,i}\}_{i\in\Omega_n}$, no such entropy gain should arise if the support of the distribution of the input $\mathbf{u}_n^1$ expands accordingly along the directions pointed by the rows $\{r_{n,i}\}_{i\in\Omega_n}$ of R_n.

An example of such situation can be easily constructed as follows: Let $G(z)$ in Figure 1 have non-minimum phase zeros and suppose that u_1^∞ is generated as $G^{-1}\tilde{\mathrm{u}}_1^\infty$, where $\tilde{\mathrm{u}}_1^\infty$ is an i.i.d. random process with bounded entropy rate. Since the determinant of G_n^{-1} equals 1 for all n, we have that $h(\mathbf{u}_n^1) = h(\tilde{\mathbf{u}}_n^1)$, for all n. On the other hand, $\mathbf{y}_n^1 = G_n G_n^{-1}\tilde{\mathbf{u}}_n^1 + \mathbf{z}_n^1 = \tilde{\mathbf{u}}_n^1 + \mathbf{z}_n^1$. Since $\mathbf{z}_n^1 = [\mathbf{\Phi}]_n^1\mathbf{s}_\kappa^1$ for some finite κ (recall Assumption 2), it is easy to show that $\lim_{n\to\infty}\frac{1}{n}h(\mathbf{y}_n^1) = \lim_{n\to\infty}\frac{1}{n}h(\tilde{\mathbf{u}}_n^1) = \lim_{n\to\infty}\frac{1}{n}h(\mathbf{u}_n^1)$; thus, no entropy gain appears.

The preceding discussion reveals that the entropy gain produced by G in the situation shown in Figure 1 **depends on the distribution of the input and on the support and distribution of the disturbance**. This stands in stark contrast with the well known fact that the increase in differential entropy produced by an invertible linear operator depends only on its Jacobian, and not on the statistics of the input [2]. We have also seen that the distribution of a random process along the different directions within the Euclidean space which contains it plays a key role, as well. This motivates the need to specify a class of random processes which distribute more or less evenly over all directions. This is precisely the intuitive meaning of an entropy-balanced process.

The following section identifies a large family of processes belonging to this class, as well as two properties which greatly expands this family.

5.2. Characterization of Entropy-Balanced Processes

We have defined the notion of an "entropy-balanced" process in Section 1.1. In words, the first condition in this definition allows one to guarantee that the orthogonal projection of an entropy-balanced process onto any ν-dimensional linear subspace has a differential entropy whose magnitude remains bounded or grows at most sub-linearly with n. The second condition states that the projection of an entropy-balanced process v_1^∞ onto any linear subspaces having ν fewer dimensions has the same differential entropy rate as the original process. This condition is equivalent to requiring that every unitary transformation on v_1^n yields a random sequence y_1^n such that $\lim_{n\to\infty}\frac{1}{n}h(\mathrm{y}_{n-\nu+1}^n \mid \mathrm{y}_1^{n-\nu}) = 0$. This property of the resulting random sequence y_1^n means that one cannot predict its last ν samples with arbitrary accuracy by using its previous $n - \nu$ samples, even if n goes to infinity.

We now characterize a large family of entropy-balanced random processes and establish some of their properties. Although intuition may suggest that most random processes (such as i.i.d. or stationary processes) should be entropy balanced, that statement seems rather difficult to prove. In the following, we show that the entropy-balanced condi-

tion is met by i.i.d. processes with per-sample *probability density function* (PDF) being uniform, piece-wise constant or Gaussian. It is also shown that adding to an entropy-balanced process an independent random processes independent of the former yields another entropy-balanced process, and that filtering an entropy-balanced process by a stable and minimum phase filter yields an entropy-balanced process, as well. The proofs can be found in Appendix B.

Lemma 1. *Let u_1^∞ be a Gaussian random process with independent elements having positive and bounded variance, i.e., there exist $0 < \check{\sigma}^2 \leq \hat{\sigma}^2 < \infty$ such that $\check{\sigma}^2 \leq \sigma_{u(n)}^2 \leq \hat{\sigma}^2$, $n \in \mathbb{N}$. Then, u_1^∞ is entropy balanced.*

Lemma 2. *Let u_1^∞ be a random process with independent elements satisfying Condition i) in Definition 2, in which each u_i is distributed according to a (possibly different) piece-wise constant PDF such that each interval where this PDF is constant has measure less than θ and greater than ϵ, for some constants $0 < \epsilon < \theta < \infty$. Then, u_1^∞ is entropy balanced.*

Lemma 3. *Let u_1^∞ and v_1^∞ be mutually independent random processes. If u_1^∞ is entropy balanced, and $w_1^\infty \triangleq u_1^\infty + v_1^\infty$ satisfies $\sigma_{w(n)}^2 < \infty$ for finite n and $\lim_{n\to\infty} n^{-1} \log(\sigma_{w(n)}^2) = 0$, then w_1^∞ is also entropy balanced.*

The proof of Lemma 3 is on page 33. The working behind this lemma can be interpreted intuitively by noting that adding to a random process another independent random process can only increase the "spread" of the distribution of the former, which tends to balance the entropy of the resulting process along all dimensions in Euclidean space. In addition, it follows from Lemma 3 that all i.i.d. processes having a per-sample PDF which can be constructed by convolving uniform, piece-wise constant or Gaussian PDFs as many times as required are entropy balanced. It also implies that one can have non-stationary processes which are entropy balanced, since Lemma 3 imposes no requirements for the process v_1^∞.

The next lemma related to the properties of entropy-balanced processes shows that filtering by a stable and minimum phase LTI filter preserves the entropy balanced condition of its input.

Lemma 4. *Let u_1^∞ be an entropy-balanced process and G an LTI stable and minimum-phase filter. Then, the output $w_1^\infty \triangleq G u_1^\infty$ is also an entropy-balanced process.*

This result implies that any stable moving-average auto-regressive process constructed from entropy-balanced innovations is also entropy balanced, provided the coefficients of the averaging and regression correspond to a stable MP filter.

The last lemma of this section states a crucial property of entropy-balanced processes (the proof is in Appendix B, page 34).

Lemma 5. *Let u_1^∞ be an entropy balanced process. Consider a disturbance z_1^∞ satisfying Assumption 2 and define $y_1^\infty \triangleq u_1^\infty + z_1^\infty$. Then, $\lim_{n\to\infty} n^{-1}(h(y_1^n) - h(u_1^n)) = 0$.*

We finish this section by pointing out two examples of processes which are non-entropy-balanced, namely the output of a NMP-filter to an entropy-balanced input and the output of an unstable filter to an entropy-balanced input. The first of these cases plays a central role in the next section.

6. Entropy Gain Due to External Disturbances

In this section, we formalize the ideas which were qualitatively outlined in the previous section. Specifically, for the system shown in Figure 1 we will characterize the entropy gain $\mathcal{G}(G, \mathbf{x}_0, u_1^\infty, z_1^\infty)$ defined in (1) for the case in which the initial state $\mathbf{x}_0$ is zero (or

deterministic) and there exists a random disturbance of (possibly infinite length) z_1^∞ which satisfies Assumption 2.

6.1. Input Disturbances Do Not Produce Entropy Gain

In this section, we show that random disturbances satisfying Assumption 2, when added to the *input* u_1^∞ (i.e., before G), do not introduce entropy gain. This result can be obtained from Lemma 6, as stated in the following theorem:

Theorem 5 (Input Disturbances do not Introduce Entropy Gain). *Let G and z_1^∞ satisfy Assumptions 1 and 2 , respectively. Suppose that u_1^∞ is entropy balanced and consider the output*

$$y_1^\infty = G\,(u_1^\infty + z_1^\infty). \tag{61}$$

Then,

$$\lim_{n\to\infty}\frac{1}{n}(h(y_1^n) - h(u_1^n)) = 0 \tag{62}$$

Proof. From Lemma 5, the differential entropy rate of u_1^∞ equals that of $u_1^\infty + z_1^\infty$. The proof is completed by recalling that G yields no entropy gain for its input $u_1^\infty + z_1^\infty$ because it corresponds to the noise-less scenario.　□

6.2. The Entropy Gain Introduced by Output Disturbances when G is MP is Zero

The results from the previous section yield the following corollary, which states that an LTI system with transfer function $G(z)$ without zeros outside the unit circle (i.e., an MP transfer function) cannot introduce entropy gain.

Corollary 1 (Minimum Phase Filters do not Introduce Entropy Gain). *Consider the system shown in Figure 1 wherein the input u_1^∞ is an entropy-balanced random process and the output disturbance z_1^∞ satisfies Assumption 2. Besides Assumption 1, suppose that $G(z)$ is minimum phase. Then,*

$$\lim_{n\to\infty}\frac{1}{n}(h(y_1^n) - h(u_1^n)) = 0. \tag{63}$$

Proof. Since $G(z)$ is minimum phase and stable, the result follows directly from Lemmas 4 and 5.　□

6.3. The Entropy Gain Introduced by Output Disturbances when G(z) is NMP

We show here that the entropy gain of an LTI system with transfer function $G(z)$ and an output disturbance is at most the sum of the logarithm of the magnitude of the zeros of $G(z)$ outside the unit circle.

The following lemma will be instrumental for that purpose.

Lemma 6. *Consider the system in Figure 1, and suppose z_1^∞ satisfies Assumption 2, and that the input process u_1^∞ is entropy balanced. Let $G_n = Q_n^T D_n R_n$ be the SVD of G_n, where $D_n = \mathrm{diag}\{d_{n,1},\ldots,d_{n,n}\}$ are the singular values of G_n, with $d_{n,1} \le d_{n,2} \le \cdots \le d_{n,n}$, such that $|\det G_n| = 1\ \forall n$. Let m be the number of these singular values which tend to zero exponentially as $n \to \infty$. Then,*

$$\lim_{n\to\infty}\frac{1}{n}(h(y_1^n) - h(u_1^n)) = \lim_{n\to\infty}\frac{1}{n}\left(-\sum_{i=1}^{m}\log d_{n,i} + h\Big([D_n]_m^1 R_n u_n^1 + [Q_n]_m^1 z_n^1\Big)\right). \tag{64}$$

The proof of this lemma can be found on page 34, in Appendix B.

Lemma 6 leaves the need to characterize the asymptotic behavior of the singular values of G_n. This is accomplished in the following lemma, which relates these singular

values to the zeros of $G(z)$. It is a generalization of the unnumbered lemma in the proof of Reference [16], Theorem 1 (restated in Appendix C as Lemma A3), which holds for FIR transfer functions, to the case of *infinite-impulse response* (IIR) transfer functions (i.e., transfer functions having poles).

Lemma 7. *For a transfer function $G(z)$ satisfying Assumption 1, where its zeros $\{\rho_i\}_{i=1}^p$ satisfy $|\rho_1| \geq \cdots \geq |\rho_m| > 1 \geq |\rho_{m+1}| \geq \cdots \geq |\rho_p|$. Then,*

$$\lambda_l(\boldsymbol{G}_n \boldsymbol{G}_n^T) = \begin{cases} \alpha_{n,l}^2 |\rho_l|^{-2n} & , \text{if } l \leq m, \\ \alpha_{n,l}^2 & , \text{otherwise }, \end{cases} \tag{65}$$

where the elements in the sequence $\{\alpha_{n,l}\}$ are positive and increase or decrease at most polynomially with n.

(The proof of this lemma can be found in Appendix B, page 36).

Lemma 6 also precisely formulates the geometric idea outlined in Section 5.1. To see this, notice that no entropy gain is obtained if the output disturbance vector $\mathbf{z}_n^1$ becomes orthogonal (with probability 1) to the space spanned by the first m columns of $\boldsymbol{Q}_n$ sufficiently fast as $n \to \infty$. Recalling from Assumption 2 that

$$\mathbf{z}_n^1 = [\boldsymbol{\Phi}]_n^1 \mathbf{s}_\kappa^1, \tag{66}$$

where the matrix $\boldsymbol{\Phi}$ has κ orthonormal columns of infinite length, such orthogonality condition can be formally stated by defining

$$\kappa_n \triangleq \mathrm{rank}([\boldsymbol{Q}_n]_m^1 [\boldsymbol{\Phi}]_n^1) \tag{67}$$

$$\hat{\kappa}_\infty \triangleq \limsup_{n \to \infty} \kappa_n \tag{68}$$

$$\check{\kappa}_\infty \triangleq \liminf_{n \to \infty} \kappa_n \tag{69}$$

as $\hat{\kappa} = 0$.

If this were the case, then the disturbance would not be able fill the subspace along which $\boldsymbol{G}_n \mathbf{u}_n^1$ is shrinking exponentially. Indeed, if $\kappa_n = 0$ for all n, then $h([\boldsymbol{D}_n]_m^1 \boldsymbol{R}_n \mathbf{u}_n^1 + [\boldsymbol{Q}_n]_m^1 \mathbf{z}_n^1) = h(^1[\boldsymbol{D}_n]_m [\boldsymbol{R}_n]_m^1 \mathbf{u}_n^1) = \sum_{i=1}^m \log d_{n,i} + h([\boldsymbol{R}_n]_m^1 \mathbf{u}_n^1)$, and the latter sum cancels out the one on the RHS of (64), while $\lim_{n\to\infty} \frac{1}{n} h([\boldsymbol{R}_n]_m^1 \mathbf{u}_n^1) = 0$ since u_1^∞ is entropy balanced. On the contrary (and loosely speaking), if the projection of the support of $\mathbf{z}_n^1$ onto the subspace spanned by the first m rows of $\boldsymbol{Q}_n$ is of dimension m (i.e., if $\kappa_n = m$) for all n, then $h([\boldsymbol{D}_n]_m^1 \boldsymbol{R}_n \mathbf{u}_n^1 + [\boldsymbol{Q}_n]_m^1 \mathbf{z}_n^1)$ remains bounded for all n, and the entropy limit of the sum $\lim_{n\to\infty} \frac{1}{n}(-\sum_{i=1}^m \log d_{n,i})$ on the RHS of (64) yields the largest possible entropy gain. Notice that $-\sum_{i=1}^m \log d_{n,i} = \sum_{i=m+1}^n \log d_{n,i}$ (because $\det(\boldsymbol{G}_n) = 1$); thus, this entropy gain stems from the uncompensated expansion of $\boldsymbol{G}_n \mathbf{u}_n^1$ along the space spanned by the rows of $[\boldsymbol{Q}_n]_n^{m+1}$. Beyond these extreme cases (i.e., for general values of $\check{\kappa}$ and $\hat{\kappa}$), the following theorem provides tight bounds on the entropy gain.

Theorem 6. *In the system of Figure 1, suppose that u_1^∞ is entropy balanced, and that $G(z)$ and z_1^∞ satisfy Assumptions 1 and 2, respectively, where the zeros $\{\rho_i\}_{i=1}^p$ of $G(z)$ satisfy $|\rho_1| \geq \cdots \geq |\rho_m| > 1 \geq |\rho_{m+1}| \geq \cdots \geq |\rho_p|$. For each $n \in \mathbb{N}$, let $\boldsymbol{Q}_n^T \in \mathbb{R}^{n \times n}$ be the unitary matrix holding the left singular vectors of $\boldsymbol{G}_n \in \mathbb{R}^{n \times n}$ (as in Lemma 6), where $\boldsymbol{G}_n$ is as defined in (4).*

1.   Then,

$$0 \leq \liminf_{n \to \infty} \frac{1}{n}(h(\mathrm{y}_1^n) - h(\mathrm{u}_1^n)) \leq \limsup_{n \to \infty} \frac{1}{n}(h(\mathrm{y}_1^n) - h(\mathrm{u}_1^n)) \leq \sum_{i=1}^{\hat{\kappa}_\infty} \log |\rho_i| \overset{(8)}{\leq} \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \left| G(e^{j\omega}) \right| d\omega. \tag{70}$$

The bounds on both extremes are tight. Moreover, the lower bound is reached if $\hat{\kappa}_\infty = 0$.

2. *If $\liminf_{n\to\infty} \frac{1}{n}\log(\sigma_{min}([Q_n]_m^1[\Phi]_n^1) = 0$, then*

$$\sum_{i=m-\check{\kappa}_\infty+1}^{m} \log|\rho_i| \leq \liminf_{n\to\infty} \frac{1}{n}(h(\mathbf{y}_1^n) - h(\mathbf{u}_1^n)). \tag{71}$$

Thus, the rightmost upper bound in (70) is achieved if $\check{\kappa}_\infty = m$.

Proof. See Appendix B, page 37. $\square$

The next technical result is very useful for finding conditions under which the requirements of point 2 in Theorem 6 are satisfied (the proof is in Appendix B, page 39).

Lemma 8. *Let F be an FIR LTI causal system of order m such that the m zeros of $F(z)$ are NMP, and $\mathbf{F}_n = Q_n^T D_n R_n$ be an SVD for $\mathbf{F}_n$, for every $n \in \{m, m+1, \ldots\}$. For each $\kappa \in \{1, \ldots, n\}$, define*

$$\kappa_n \triangleq \text{rank}\left(\overset{1\quad\overset{\kappa}{\frown}}{[Q_n]_m^1}\right), \tag{72}$$

and $\bar{\kappa} \triangleq \min\{m, \kappa\}$. Then,

$$\lim_{n\to\infty} \sigma_{min}\left(\overset{1\quad\overset{\kappa}{\frown}}{[Q_n]_m^1}\right) > 0, \tag{73}$$

and $\lim_{n\to\infty} \kappa_n = \bar{\kappa}$.

Now, we can prove Theorem 4.

6.4. Proof of Theorem 4

Factorize $G(z)$ as $G(z) = F(z)\tilde{G}(z)$, where $\tilde{G}(z)$ is stable and minimum phase and $F(z)$ is a stable FIR transfer function with all the m non-minimum-phase zeros of $G(z)$. Letting $\tilde{\mathbf{u}}_n^1 \triangleq \tilde{G}_n \mathbf{u}_n^1$, we have that $h(\mathbf{y}_n^1) = h(\mathbf{F}_n\tilde{\mathbf{u}}_n^1 + \mathbf{z}_n^1)$, $h(\tilde{\mathbf{u}}_n^1) = h(\mathbf{u}_n^1)$, and that $\tilde{\mathbf{u}}_1^\infty$ is entropy balanced (from Lemma 4). Thus,

$$h(\mathbf{y}_n^1) - h(\mathbf{u}_n^1) = h(G_n\mathbf{u}_n^1 + \mathbf{z}_n^1) - h(\mathbf{u}_n^1) = h(\mathbf{F}_n\tilde{\mathbf{u}}_n^1 + \mathbf{z}_n^1) - h(\tilde{\mathbf{u}}_n^1). \tag{74}$$

This means that the entropy gain of G due to the output disturbance $\mathbf{z}_1^\infty$ corresponds to the entropy gain of F due to the same output disturbance.

Clearly, $\mathbf{u}_1^\infty$, $F(z)$, and $\mathbf{z}_1^\infty$ satisfy the assumptions of Theorem 6 with $\Phi = [\,I_\kappa\,|\,\mathbf{0}\,]^T$ (see Assumption 2). Therefore,

$$[Q_n]_m^1[\Phi]_n^1 = \overset{1\quad\overset{\kappa}{\frown}}{[Q_n]_m^1}. \tag{75}$$

Combining this with Lemma 8, it readily follows that, for every $\kappa \geq 1$, the condition in point 2 of Theorem 6 is met, and also $\lim_{n\to\infty} \kappa_n = \bar{\kappa}$. The proof is then completed by substituting $\liminf_{n\to\infty} \kappa_n = \limsup_{n\to\infty} \kappa_n = \bar{\kappa}$ into (70) and (71).

7. Entropy Gain Due to a Random Initial State

Here, we analyze the scenario illustrated by Figure 1 for the case in which there exists a random initial state $\mathbf{x}_0$ independent of the input $\mathbf{u}_1^\infty$, and zero (or deterministic) output disturbance.

The treatment of an initial state of the LTI system G requires one to first define an internal model for it. For this purpose, in this section, we consider the state-space realization of G in the Kalman canonical form, given by

$$\mathbf{x}(k) \triangleq \begin{bmatrix} \mathbf{x}_{co}(k) \\ \mathbf{x}_{\bar{c}o}(k) \\ \mathbf{x}_{c\bar{o}}(k) \\ \mathbf{x}_{\bar{c}\bar{o}}(k) \end{bmatrix} = \begin{pmatrix} A_{co} & A_{12} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & A_{\bar{c}o} & \mathbf{0} & \mathbf{0} \\ A_{31} & A_{32} & A_{c\bar{o}} & A_{34} \\ \mathbf{0} & A_{42} & \mathbf{0} & A_{\bar{c}\bar{o}} \end{pmatrix} \begin{bmatrix} \mathbf{x}_{co}(k-1) \\ \mathbf{x}_{\bar{c}o}(k-1) \\ \mathbf{x}_{c\bar{o}}(k-1) \\ \mathbf{x}_{\bar{c}\bar{o}}(k-1) \end{bmatrix} + \begin{bmatrix} \mathbf{b}_{co} \\ \mathbf{0} \\ \mathbf{b}_{c\bar{o}} \\ \mathbf{0} \end{bmatrix} \mathbf{u}(k) \tag{76a}$$

$$\mathbf{y}(k) = \begin{bmatrix} \mathbf{c}_{co}^T & \mathbf{c}_{\bar{c}o}^T & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{x}(k-1) + \mathbf{u}(k), \tag{76b}$$

(see, e.g., Reference [21] or Reference [22], Chapter 6) where the column state vectors $\mathbf{x}_{co}(k), \mathbf{x}_{\bar{c}o}(k), \mathbf{x}_{c\bar{o}}(k), \mathbf{x}_{\bar{c}\bar{o}}(k)$ are, respectively, *controllable and observable*, *non-controllable and observable*, *controllable and non-observable*, and *non-controllable and non-observable*. There is no loss of generality in choosing this state-space representation, because every state-space representation consistent with a rational transfer function $G(z)$ can be written in this form (Reference [22], Theorem 6.7).

Since our interest is on the effect of the random initial state of G on its output, we only need to consider the observable subsystem within (76) and without its input, given by

$$\mathbf{x}_o(k) \triangleq \begin{bmatrix} \mathbf{x}_{co}(k) \\ \mathbf{x}_{\bar{c}o}(k) \end{bmatrix} = \underbrace{\begin{pmatrix} A_{co} & A_{12} \\ \mathbf{0} & A_{\bar{c}o} \end{pmatrix}}_{A_o} \begin{bmatrix} \mathbf{x}_{co}(k-1) \\ \mathbf{x}_{\bar{c}o}(k-1) \end{bmatrix}, \tag{77a}$$

$$\tilde{\mathbf{y}}(k) = \underbrace{\begin{bmatrix} \mathbf{c}_{co}^T & \mathbf{c}_{\bar{c}o}^T \end{bmatrix}}_{\mathbf{c}_o^T} \mathbf{x}_o(k-1), \tag{77b}$$

where $\tilde{\mathbf{y}}$ is the natural response of G to its initial state $\mathbf{x}_o(0)$ and $\mathbf{x}_{co} \in \mathbb{R}^p$ and $\mathbf{x}_{\bar{c}o} \in \mathbb{R}^q$. We shall decompose $\tilde{\mathbf{y}}$ as

$$\tilde{\mathbf{y}}_1^\infty = [\tilde{\mathbf{y}}_{\bar{c}o}]_1^n + [\tilde{\mathbf{y}}_{co}]_1^\infty, \tag{78}$$

where $\tilde{\mathbf{y}}_{\bar{c}o}$ and $\tilde{\mathbf{y}}_{co}$ are the natural responses of G to initial states $[\mathbf{0}_{1 \times p} \quad \mathbf{x}_{\bar{c}o}(0)^T]^T$ and $[\mathbf{x}_{co}(0)^T \quad \mathbf{0}_{1 \times q}]^T$, respectively. The natural response component $\tilde{\mathbf{y}}_{co}$ can be generated by the following minimal state-space representation of $G(z)$, without the effect of its input u:

$$\underbrace{\begin{bmatrix} \mathbf{x}_{co,1}(k) \\ \mathbf{x}_{co,2}(k) \\ \mathbf{x}_{co,3}(k) \\ \vdots \\ \mathbf{x}_{co,p}(k) \end{bmatrix}}_{\mathbf{x}_{co}(k)} = \underbrace{\begin{pmatrix} b_1 & b_2 & b_3 & \cdots & b_p \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}}_{A_{co}} \underbrace{\begin{bmatrix} \mathbf{x}_{co,1}(k-1) \\ \mathbf{x}_{co,2}(k-1) \\ \mathbf{x}_{co,3}(k-1) \\ \vdots \\ \mathbf{x}_{co,p}(k-1) \end{bmatrix}}_{\mathbf{x}_{co}}, \tag{79}$$

$$\tilde{\mathbf{y}}_{co}(k) = \underbrace{\begin{bmatrix} a_1 & a_2 & \cdots & a_p \end{bmatrix}}_{\mathbf{a}^T} \mathbf{x}_{co}(k-1) + \underbrace{\begin{bmatrix} b_1 & b_2 & \cdots & b_p \end{bmatrix}}_{\mathbf{p}^T} \mathbf{x}_{co}(k-1). \tag{80}$$

Now, we can state and prove the main result of this section:

Theorem 7. *Suppose G satisfies Assumption 1 and $\mathbf{u}_1^\infty$ is entropy balanced. Assume that $\mathbf{x}_o(0)$ (the observable part of the initial state of G) is independent of the input $\mathbf{u}_1^\infty$, $|h(\mathbf{x}_o(0))| < \infty$ and that $\mathrm{tr}\{K_{\mathbf{x}_o(0)}\} < \infty$. Then,*

$$\lim_{n \to \infty} \frac{1}{n}(h(\mathbf{y}_1^n) - h(\mathbf{u}_1^n)) = \sum_{i=1}^m \log|\rho_i|. \tag{81}$$

Proof. Both G and u_1^∞ satisfy the conditions of Theorem 6. Thus, as in its statement, we write $G(z) = F(z)\tilde{G}(z)$, where $\tilde{G}(z)$ is stable and minimum phase and $F(z)$ is a stable FIR transfer function with only the m non-minimum-phase zeros of $G(z)$.

Defining $\mathbf{w}_n^1 \triangleq \tilde{G}_n \mathbf{u}_n^1$, we have

$$\mathbf{y}_n^1 = F_n \tilde{G}_n \mathbf{u}_n^1 + \tilde{\mathbf{y}}_n^1 = F_n \mathbf{w}_n^1 + \tilde{\mathbf{y}}_n, \tag{82}$$

$$h(\mathbf{w}_n^1) = h(\mathbf{u}_n^1), \tag{83}$$

and $\tilde{\mathbf{y}}_n^1 \perp\!\!\!\perp \mathbf{w}_n^1$. In addition, the fact that G is stable guarantees that the sample second moment of $\tilde{\mathbf{y}}_1^\infty$ decays exponentially, which means that $\tilde{\mathbf{y}}_1^\infty$ satisfies Assumption 2. Thus, the conditions of Lemma 6 are met considering $G_n = F_n$, where now $F_n = Q_n^T D_n R_n$ is the SVD for F_n, and $d_{n,1} \leq d_{n,2} \leq \cdots \leq d_{n,n}$. Consequently, the proof would be completed if we can show that $\lim_{n\to\infty} \frac{1}{n} h([D_n]_m^1 R_n \mathbf{w}_n^1 + [Q_n]_m^1 \tilde{\mathbf{y}}_n^1) = 0$. But all the involved variables have bounded variance, while R_n is unitary, $[Q_n]_m^1$ has orthonormal rows and the entries of $[D_n]_m^1$ decay exponentially with n. This implies that $\lim_{n\to\infty} \frac{1}{n} h([D_n]_m^1 R_n \mathbf{w}_n^1 + [Q_n]_m^1 \tilde{\mathbf{y}}_n^1) \leq 0$. Therefore, it is only left to prove that

$$\lim_{n\to\infty} \frac{1}{n} h([D_n]_m^1 R_n \mathbf{w}_n^1 + [Q_n]_m^1 \tilde{\mathbf{y}}_n^1) \geq 0. \tag{84}$$

Recalling (78), let us decompose $[\tilde{\mathbf{y}}_{co}]_n^1$ so that

$$\tilde{\mathbf{y}}_n^1 = F_n \tilde{P}_n \mathbf{x}_{co}(0) + P_n \mathbf{x}_{co}(0) + [\tilde{\mathbf{y}}_{\bar{c}o}]_n^1, \tag{85}$$

where $\tilde{P}_n, P_n \in \mathbb{R}^{n \times (p+q)}$, the sequences $F_n \tilde{P}_n \mathbf{x}_{co}(0)$ and $P_n \mathbf{x}_{co}(0)$, respectively, are the natural responses of $\tilde{G}$ and F to the controllable and observable initial state $\mathbf{x}_{co}$, and $[\tilde{\mathbf{y}}_{\bar{c}o}]_n^1$ is the natural response of G to the non-controllable and observable initial state $\mathbf{x}_{\bar{c}o}(0)$. Then,

$$h([D_n]_m^1 R_n \mathbf{w}_n^1 + [Q_n]_m^1 \tilde{\mathbf{y}}_n^1) \overset{(a)}{\geq} h([Q_n]_m^1 \tilde{\mathbf{y}}_n^1) \overset{(85)}{=} h([Q_n]_m^1 (F_n \tilde{P}_n \mathbf{x}_{co}(0) + P_n \mathbf{x}_{co}(0) + [\tilde{\mathbf{y}}_{\bar{c}o}]_n^1)), \tag{86}$$

$$\overset{(b)}{=} h([Q_n]_m^1 (F_n \tilde{P}_n \mathbf{x}_{co}(0) + P_n \mathbf{x}_{co}(0)) \mid \mathbf{x}_{\bar{c}o}(0)) = h([Q_n]_m^1 (F_n \tilde{P}_n + P_n) \mathbf{x}_{co}(0) \mid \mathbf{x}_{\bar{c}o}(0)), \tag{87}$$

where (a) is from the entropy-power inequality [1] and (b) holds because conditioning does not increase entropy and $[\tilde{\mathbf{y}}_{\bar{c}o}]_n^1$ is a deterministic function of $\mathbf{x}_{\bar{c}o}(0)$. Let the SVD of $[Q_n]_m^1 (F_n \tilde{P}_n + P_n)$ be

$$[Q_n]_m^1 (F_n \tilde{P}_n + P_n) = S_n T_n H_n, \quad n = m, m+1, \ldots, \tag{88}$$

where $S_n \in \mathbb{R}^{m \times m}$ is unitary, $T_n = \text{diag}\{t_1, t_2, \ldots, t_m\}$ holds the singular values of $[Q_n]_m^1 (F_n \tilde{P}_n + P_n)$ and $H_n \in \mathbb{R}^{m \times p}$ has orthonormal rows. Substituting this SVD into (87) we obtain

$$h([D_n]_m^1 R_n \mathbf{w}_n^1 + [Q_n]_m^1 \tilde{\mathbf{y}}_n^1) \geq h(S_n T_n H_n \mathbf{x}_{co}(0) \mid \mathbf{x}_{\bar{c}o}(0)) = \log(\det(T_n)) + h(H_n \mathbf{x}_{co}(0) \mid \mathbf{x}_{\bar{c}o}(0)). \tag{89}$$

This last differential entropy is bounded because $|h(\mathbf{x}_o)| < \infty$ and $\text{tr}\{K_{\mathbf{x}_o}\} < \infty$, which implies (thanks to Proposition A1) that $|h(H_n \mathbf{x}_{co}, \mathbf{x}_{\bar{c}o})| < \infty$, and by the chain rule of entropy,

$$h(H_n \mathbf{x}_{co}, \mathbf{x}_{\bar{c}o}) = h(\mathbf{x}_{\bar{c}o}) + h(H_n \mathbf{x}_{co}(0) \mid \mathbf{x}_{\bar{c}o}(0)), \tag{90}$$

so $|h(H_n \mathbf{x}_{co}(0) \mid \mathbf{x}_{\bar{c}o}(0))| < \infty$ because $|h(\mathbf{x}_{\bar{c}o}(0))| < \infty$ (again from Proposition A1). Thus, in view of (89) and (84), all that remains to prove is that

$$\lim_{n\to\infty} \sigma_{\min}\left([Q_n]_m^1 (F_n \tilde{P}_n + P_n)\right) > 0, \tag{91}$$

For that purpose, notice that $[Q_n]_m^1 (F_n \tilde{P}_n + P_n) = {}^1[D_n]_m [R_n]_m^1 \tilde{P}_n + [Q_n]_m^1 P_n$. Therefore, from Lemma A4 (in Appendix C), it follows that (91) holds if

$$\lim_{n \to \infty} \sigma_{\min} \left([Q_n]_m^1 P_n \right) > 0, \tag{92}$$

and

$$\lim_{n \to \infty} \sigma_{\max} \left({}^1[D_n]_m [R_n]_m^1 \tilde{P}_n \right) = 0. \tag{93}$$

To prove (93), recall that the entries in the diagonal matrix ${}^1[D_n]_m$ decay exponentially with n. On the other hand, the rows of $[R_n]_m^1$ are orthonormal. Finally, the fact that $\tilde{G}$ is stable implies that the $p + q$ columns of $\tilde{P}_n$ have norms which are bounded for all n. These three observations readily yield that (93) holds.

To prove that (92) holds, write the rational transfer function of G (described by (80)) as

$$G(z) = \frac{1 + a_1 z^{-1} + \cdots + a_p z^{-p}}{1 + b_1 z^{-1} + \cdots + b_p z^{-p}} = \underbrace{(1 + f_1 z^{-1} + \cdots + f_m z^{-m})}_{F(z)} \underbrace{\frac{1 + \tilde{a}_1 z^{-1} + \cdots + \tilde{a}_{\tilde{m}} z^{-\tilde{m}}}{1 + b_1 z^{-1} + \cdots + b_p z^{-p}}}_{\tilde{G}(z)}, \tag{94}$$

where $\tilde{m} \triangleq p - m$. The coefficients in the numerator of $G(z)$ are related to those of $F(z)$ and $\tilde{G}(z)$ by the convolution

$$a_i = \sum_{j=0}^m f_j \tilde{a}_{i-j}, \quad i = 1, \ldots, p, \tag{95}$$

where $\tilde{a}_0 = f_0 = 1$.

Denote the natural response of F (up to time n) to its initial state $\mathbf{x}_F(0)$ (which is a linear function of $\mathbf{x}_{co}(0)$) as

$$\ddot{\mathbf{y}}_n^1 \triangleq P_n \mathbf{x}_{co}(0). \tag{96}$$

Let $\tilde{\mathbf{w}}_n^1 \triangleq \tilde{P}_n \mathbf{x}_{co}(0)$ be the natural response of $\tilde{G}$ to its initial state $\mathbf{x}_{co}(0)$. Following the structure of (80), $\tilde{w}(k)$ can be written as

$$\tilde{w}(k) = [\tilde{a}_1 \; \cdots \; \tilde{a}_{\tilde{m}} \; 0 \; \cdots \; 0] \mathbf{x}_{co}(k-1) + \mathbf{p}^T \mathbf{x}_{co}(k-1), \quad k = 1, 2, \ldots, \tag{97}$$

where $\mathbf{x}_{co}$ satisfies (79). Considering the following minimal state-space representation of F

$$\mathbf{x}_F(k) \triangleq \begin{bmatrix} x_{F,1}(k) \\ x_{F,2}(k) \\ x_{F,3}(k) \\ \vdots \\ x_{F,m}(k) \end{bmatrix} = \underbrace{\begin{pmatrix} 0 & 0 & \cdots & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}}_{A_F} \begin{bmatrix} x_{F,1}(k-1) \\ x_{F,2}(k-1) \\ x_{F,3}(k-1) \\ \vdots \\ x_{F,m}(k-1) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} w(k), \tag{98}$$

$$\tilde{y}_{co}(k) = \underbrace{[f_1 \; f_2 \; \cdots \; f_m]}_{c_F^T} \mathbf{x}_F(k-1) + \tilde{w}(k), \tag{99}$$

it can be seen that the natural response of F to its own initial state $\mathbf{x}_F(0)$ can be written as

$$\ddot{y}(k) = \tilde{y}_{co}(k) - \tilde{w}(k) - \text{"the effect of } f_1, \ldots, f_{k-1}\text{"}. \tag{100}$$

But, from (80) and (95),

$$\tilde{y}_{co}(k) = \tilde{\boldsymbol{a}}^T \operatorname{diag}\{[f_1 \ \cdots \ f_m]\}_{\tilde{m}+1}\mathbf{x}_{co}(k-1) + \underbrace{[\tilde{a}_1 \ \cdots \ \tilde{a}_{\tilde{m}} \ 0 \ \cdots \ 0]\mathbf{x}_{co}(k-1) + \boldsymbol{p}^T\mathbf{x}_{co}(k-1)}_{\tilde{w}(k)}, \tag{101}$$

where $\tilde{\boldsymbol{a}} \triangleq [1 \ \tilde{a} \ \cdots \ \tilde{a}_{\tilde{m}}]$, and

$$\operatorname{diag}\{[f_1 \ \cdots \ f_m]\}_{\ell} \triangleq \left. \begin{pmatrix} f_1 & \cdots & f_m & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & f_1 & \cdots & f_m \end{pmatrix} \right\} \ell \text{ rows} \quad \ell = 1, 2, \ldots; \tag{102}$$

therefore,

$$\ddot{y}(1) = \tilde{\boldsymbol{a}}^T \operatorname{diag}\{[f_1 \ f_2 \ \cdots \ f_m]\}_{\tilde{m}+1}\mathbf{x}_{co}(0), \tag{103}$$

$$\ddot{y}(2) = \tilde{\boldsymbol{a}}^T \operatorname{diag}\{[0 \ f_2 \ \cdots \ f_m]\}_{\tilde{m}+1}\boldsymbol{A}_{co}\mathbf{x}_{co}(0), \tag{104}$$

$$= \tilde{\boldsymbol{a}}^T \operatorname{diag}\{[f_2 \ \cdots \ f_m \ 0]\}_{\tilde{m}+1}\mathbf{x}_{co}(0) \tag{105}$$

$$\vdots \tag{106}$$

$$\ddot{y}(m) = \tilde{\boldsymbol{a}}^T \operatorname{diag}\{[0 \ \cdots \ 0 \ f_m]\}_{\tilde{m}+1}\boldsymbol{A}_{co}^{m-1}\mathbf{x}_{co}, \tag{107}$$

$$= \tilde{\boldsymbol{a}}^T \operatorname{diag}\{[f_m \ 0 \ \cdots \ 0]\}_{\tilde{m}+1}\mathbf{x}_{co}(0), \tag{108}$$

with $\ddot{y}(k) = 0$ for $k > m$. Therefore,

$$\ddot{\mathbf{y}}_m^1 = \boldsymbol{E}\mathbf{x}_{co}(0) = \underbrace{[\boldsymbol{M} \mid \boldsymbol{N}]}_{\boldsymbol{E}}\mathbf{x}_{co}(0), \tag{109}$$

where $\boldsymbol{M} \in \mathbb{R}^{m \times (p-m)}$ and $\boldsymbol{N} \in \mathbb{R}^{m \times m}$ is a lower anti-triangular Toeplitz matrix with $\tilde{a}_{\tilde{m}}f_m$ along its main anti diagonal.

This implies that $\boldsymbol{P}_n = [\boldsymbol{E}^T \mid \boldsymbol{0}_{p \times (n-m)}]^T$ and

$$\sigma_{\min}(\boldsymbol{E}) > 0. \tag{110}$$

Thus, resuming the reasoning before (94), we have that

$$[\boldsymbol{Q}_n]_m^1 \boldsymbol{P}_n = {}^1[\boldsymbol{Q}_n^{(p)}]_m \boldsymbol{E}. \tag{111}$$

It then follows from (110) and Lemma 8 that

$$\lim_{n\to\infty} \sigma_{\min}\left([\boldsymbol{Q}_n]_m^1 \boldsymbol{P}_n\right) = \lim_{n\to\infty} \sigma_{\min}\left({}^1[\boldsymbol{Q}_n]_m \boldsymbol{E}\right). > 0. \tag{112}$$

Hence, (91) is satisfied. Substituting (91) into (89) and the latter into (84) yields

$$\lim_{n\to\infty} \frac{1}{n} h\left([\boldsymbol{D}_n]_m^1 \boldsymbol{R}_n \mathbf{w}_n^1 + [\boldsymbol{Q}_n]_m^1 \bar{\mathbf{y}}_n^1\right) = 0. \tag{113}$$

The proof is completed by invoking Lemma (6). □

Theorem 7 allows us to formalize the effect that the presence or absence of a random initial state has on the entropy gain using arguments similar to those utilized in Section 6.

8. Some Implications

The purpose of this section is to illustrate how the results obtained in the previous section can be applied to other problems. To do so, we present next some of the implications

of these results on three different problems previously addressed in the literature, namely finding the rate-distortion function for non-stationary processes, an inequality in networked control theory, and the feedback capacity of Gaussian stationary channels. The common feature in these three problems is that, in all of them, non-minimum phase transfer functions play a role (either explicitly or implicitly).

8.1. Networked Control

The analysis developed in Reference [13] considers an LTI system P within a noisy feedback loop, as the one depicted in Figure 4. In this scheme, C represents a causal feedback channel which combines the output of P with an exogenous (noise) random process c_1^∞ to generate its output. The process c_1^∞ is assumed independent of the initial state of P, represented by the random vector $\mathbf{x}_0$, which has finite differential entropy.
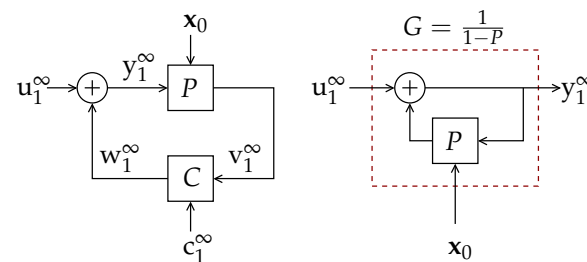

Figure 4. (**Left**): LTI system P within a noisy feedback loop. (**Right**): equivalent system when the feedback channel is noiseless and has unit gain.

For this system, it is shown in Reference [13], Theorem 4.2, that

$$\bar{h}(\mathbf{y}_1^\infty) \geq \bar{h}(\mathbf{u}_1^\infty) + \lim_{n\to\infty} \frac{1}{n} I(\mathbf{x}_0; \mathbf{y}_1^n), \tag{114a}$$

where $I(\mathbf{x}_0; \mathbf{y}_1^n)$ is the mutual information (see Reference [1], Section 8.5) between $\mathbf{x}_0$ and $\mathbf{y}_1^n$, with equality if w is a deterministic function of v. Furthermore, it is shown in Reference [12], Lemma 3.2, that, if $|h(\mathbf{x}_0)| < \infty$ and the steady state variance of system P remains asymptotically bounded as $k \to \infty$, then

$$\lim_{n\to\infty} \frac{1}{n} I(\mathbf{x}_0; \mathbf{y}_1^n) \geq \sum_{p_i:|p_i|>1} \log|p_i|, \tag{114b}$$

where $\{p_i\}$ are the poles of P. Thus, for the (simplest) case in which w $=$ v, the output $\mathbf{y}_1^\infty$ is the result of filtering $\mathbf{u}_1^\infty$ by a filter $G = \frac{1}{1-P}$ (as shown in Figure 4 right), and the resulting entropy rate of $\mathbf{y}_1^\infty$ will exceed that of $\mathbf{u}_1^\infty$ only if there is a random initial state with bounded differential entropy (see (114a)). Moreover, if w $=$ v and $G(z)$ is stable, (114) (as well as Reference [13], Lemma 4.3) implies that this entropy gain is lower bounded by the *right-hand side* (RHS) of (8), which is greater than zero if and only if G is NMP. However, both [12,13] do not provide conditions under which this lower bound is reached.

In Reference [14], Theorem 14, it is shown that, when there is perfect feedback (i.e., when v $=$ w), as in Figure 4 right, with P being the concatenation of a stabilizing LTI controller and an LTI plant, and assuming $\mathbf{u}_1^\infty$ is Gaussian i.i.d. and a Gaussian initial state, then

$$\bar{h}(\mathbf{y}_1^\infty) - \bar{h}(\mathbf{u}_1^\infty) = \sum_{p_i:|p_i|>1} \log|p_i|. \tag{115}$$

Notice that this implies reaching equality in both (114a) and (114b).

By using the results obtained in Section 7 we show next that equality holds in (114b) provided the feedback channel satisfies the following assumption:

Assumption 3. *The feedback channel in Figure 4 can be written as*

$$\mathrm{w} = AB\,\mathrm{v} + BF(\mathrm{c}), \tag{116}$$

where:

1. *A and B are stable rational transfer functions such that AB is biproper, ABP has the same unstable poles as P, and the feedback AB stabilizes the plant P.*
2. *F is any (possibly non-linear) operator such that $\tilde{\mathrm{c}} \triangleq F(\mathrm{c})$ has finite variance $\sigma^2_{\tilde{\mathrm{c}}(n)}$ for finite n, $\lim_{n\to\infty} n^{-1}\log(\sigma^2_{\tilde{\mathrm{c}}(n)}) = 0$, and*
3. *$\mathrm{c}_1^\infty \perp\!\!\!\perp \mathbf{x}_0$.*

We also extend Reference [14], Theorem 14, to situations including a feedback channel satisfying Assumption 3. For the perfect-feedback case, this extends the validity of (115) to a much larger class of distributions for u_1^∞.

An illustration of the class of feedback channels satisfying this assumption is depicted on top of Figure 5. Trivial examples of channels satisfying Assumption 5 are a Gaussian additive channel preceded and followed by linear operators [23]. Indeed, when F is an LTI system with a strictly causal transfer function, the feedback channel that satisfies Assumption 3 is widely known as a *noise shaper with input pre and post filter*, used in, e.g., Reference [24–27].
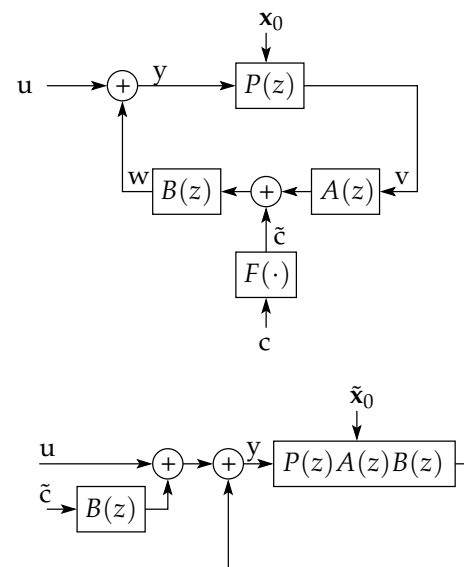

Figure 5. (**Top**): The class of feedback channels described by Assumption 3. (**Bottom**): an equivalent form.

Theorem 8. *In the networked control system of Figure 4, suppose that the feedback channel satisfies Assumption 3, that the plant $P(z)$ has poles $\{p_i\}_i^p$, and that the input u_1^∞ is entropy balanced. If the random initial states of AB and P, namely $\mathbf{s}_0 \in \mathbb{R}^q$ and $\mathbf{x}_0 \in \mathbb{R}^p$, respectively, are independent, have finite variance and $|h(\mathbf{x}_0)| < \infty$, then*

$$\lim_{n\to\infty} \frac{1}{n} I(\mathbf{x}_0; \mathrm{y}_1^n) = \sum_{|p_i|>1} \log|p_i|. \tag{117a}$$

Moreover,

$$\lim_{n\to\infty} \frac{1}{n}(h(\mathrm{y}_1^n) - h(\tilde{\mathrm{u}}_1^n)) = \sum_{|p_i|>1} \log|p_i|, \tag{117b}$$

where $\tilde{\mathrm{u}} \triangleq \mathrm{u} + B\tilde{\mathrm{c}}$ (see Figure 5 bottom).

Proof. Let $P(z) = N(z)/G(z)$ and $T(z) \triangleq A(z)B(z) = \Gamma(z)/\Theta(z)$. We will first show that the output $\mathbf{y}_n^1$ can be written as

$$\mathbf{y}_n^1 = \boldsymbol{G}_n \tilde{\boldsymbol{G}}_n \tilde{\mathbf{u}}_n^1 + \boldsymbol{G}_n \tilde{\boldsymbol{P}}_n [\mathbf{x}_0^T \ \ \mathbf{s}_0^T]^T + \boldsymbol{P}_n \mathbf{x}_0, \tag{118}$$

where $\tilde{G}$ is the stable LTI system with biproper and MP transfer function

$$\tilde{G}(z) \triangleq \frac{\Theta(z)}{\Theta(z)G(z) + N(z)\Gamma(z)}, \tag{119}$$

with $\mathbf{s}_0 \in \mathbb{R}^q$, $\mathbf{x}_0 \in \mathbb{R}^p$ and $[\mathbf{x}_0^T \ \ \mathbf{s}_0^T]^T$ being the random initial states of T, G, and $\tilde{G}$, respectively, and

$$\tilde{\mathbf{u}} \triangleq \mathbf{u} + B\tilde{\mathbf{c}} \tag{120}$$

(see Figure 5 bottom). The matrices $\tilde{\boldsymbol{P}}_n \in \mathbb{R}^{n \times p}$ and $\boldsymbol{P}_n \in \mathbb{R}^{n \times (p+q)}$. From Figure 5, it is clear that the transfer function from $\tilde{\mathbf{u}}$ to $\mathbf{y}$ is $G(z)\frac{\Theta(z)}{\Theta(z)G(z)+N(z)\Gamma(z)}$, validating the first term on the RHS of (118). In addition, it is evident that the initial state of $\tilde{G}$ is a linear combination of $\mathbf{x}_0$ and $\mathbf{s}_0$, justifying the term $\tilde{\boldsymbol{P}}_n [\mathbf{x}_0^T \ \ \mathbf{s}_0^T]^T$ as the natural response of $\tilde{G}$. Thus, it is only left to prove that the initial state of G is $\mathbf{x}_0$. For that purpose, let $G(z) = 1 - \sum_{i=1}^p g_i z^{-i}$ and $N(z) = \sum_{i=1}^p n_i z^{-i}$. Define the following variables:

$$\mathbf{o} \triangleq \frac{1}{G}\mathbf{y}, \quad \mathbf{w} \triangleq N\mathbf{o}. \tag{121}$$

Then, the recursion corresponding to $P(z)$ is

$$\mathbf{o}_k = \sum_{i=1}^p g_i \mathbf{o}_{k-i} + \mathbf{y}_k, \quad k \geq 1, \tag{122}$$

$$\mathbf{w}_k = \sum_{i=1}^p n_i \mathbf{o}_{k-i}, \quad k \geq 1. \tag{123}$$

This reveals that the initial state of $P(z)$ corresponds to

$$\mathbf{x}_0 = [\mathbf{o}_{1-p} \ \ \mathbf{o}_{2-p} \ \ \cdots \ \ \mathbf{o}_0]. \tag{124}$$

But, from (121), $\mathbf{o}$ is also the output of $\tilde{G}$ to the input $\tilde{\mathbf{u}}$, and

$$\mathbf{y}_k = \mathbf{o}_k - \sum_{i=1}^p f_i \mathbf{o}_{k-i}, \quad k \geq 1, \tag{125}$$

which means that the initial state of G is $\mathbf{x}_0$.

Now, using (118), we have that

$$I(\mathbf{x}_0; \mathbf{y}_n^1) = h(\mathbf{y}_n^1) - h(\mathbf{y}_n^1 | \mathbf{x}_0), \tag{126}$$

$$= h(\mathbf{y}_n^1) - h(\boldsymbol{F}_n [\tilde{\boldsymbol{G}}_n \tilde{\mathbf{u}}_n^1 + \tilde{\boldsymbol{P}}_n \mathbf{s}_0]), \tag{127}$$

$$= h(\boldsymbol{F}_n \bar{\mathbf{u}}_n^1 + \boldsymbol{P}_n \mathbf{x}_0) - h(\bar{\mathbf{u}}_n^1), \tag{128}$$

where the first equality is because $\mathbf{s}_0 \perp\!\!\!\perp \mathbf{x}_0$ and $\bar{\mathbf{u}}_n^1 \triangleq \tilde{\boldsymbol{G}}_n \tilde{\mathbf{u}}_n^1 + \tilde{\boldsymbol{P}}_n \mathbf{s}_0$. The last equality holds since the first sample of the unit-impulse response of G is 1. Since $\mathbf{u}_1^\infty$ is entropy balanced, $\tilde{G}(z)$ is biproper, stable, and MP, and both $\tilde{\mathbf{c}}_1^\infty$ and $\tilde{\boldsymbol{P}}_n \mathbf{s}_0$ have finite variance, it follows from Lemmas 3 and 4 that $\bar{\mathbf{u}}_1^\infty$ is entropy balanced, as well. Thus, the proof of the first claim is completed by direct application of Theorem 7.

For the second claim,

$$h(\mathbf{y}_n^1) - h(\tilde{\mathbf{u}}_n^1) \stackrel{(a)}{=} h(\mathbf{y}_n^1) - h(\tilde{\boldsymbol{G}}_n \tilde{\mathbf{u}}_n^1) = h(\mathbf{y}_n^1) - h(\bar{\mathbf{u}}_n^1) + (h(\tilde{\boldsymbol{G}}_n \tilde{\mathbf{u}}_n^1) - h(\tilde{\mathbf{u}}_n^1)), \tag{129}$$

where (a) holds because the first sample of the unit-impulse response of $\tilde{G}$ is $\tilde{g}_0 = \lim_{z \to \infty} \tilde{G}(z) = 1$. Then,

$$\lim_{n \to \infty} \frac{1}{n}(h(\mathbf{y}_n^1) - h(\tilde{\mathbf{u}}_n^1)) = \lim_{n \to \infty} \frac{1}{n}(h(\mathbf{y}_n^1) - h(\bar{\mathbf{u}}_n^1)) + \lim_{n \to \infty} \frac{1}{n}(h(\tilde{G}_n\tilde{\mathbf{u}}_n^1) - h(\tilde{\mathbf{u}}_n^1)), \quad (130)$$

$$\overset{(a)}{=} \lim_{n \to \infty} \frac{1}{n}(h(\mathbf{y}_n^1) - h(\bar{\mathbf{u}}_n^1)), \quad (131)$$

$$\overset{(b)}{=} \sum_{|p_i| > 1} \log|p_i|, \quad (132)$$

where (a) holds because $\tilde{G}\tilde{u}$ is entropy balanced (from Lemma 4), and $\tilde{P}_n\mathbf{s}_0$ has finite variance, allowing us to apply Proposition A3. In turn, (b) follows from (128) and (117a). This completes the proof.

$\square$

Remark 2. *If $A(z)$ had poles outside the unit circle, then Theorem 8 can still be applied by associating those poles to P.*

Remark 3. *Under the conditions of Theorem 8, one has that if either $\bar{h}(\mathbf{u}_1^\infty)$ or $\bar{h}(\tilde{\mathbf{c}}_1^\infty)$ exists, then the other entropy rate exists too. In that case, if $\mathbf{c} \perp\!\!\!\perp \mathbf{u}$ and defining $\bar{\mathbf{c}} \triangleq B\tilde{\mathbf{c}}$, (117) yields*

$$\bar{h}(\mathbf{y}_1^\infty) - \bar{h}(\mathbf{u}_1^\infty) - \bar{h}(\bar{\mathbf{c}}_1^\infty) = \lim_{n \to \infty} \frac{1}{n}I(\mathbf{x}_0; \mathbf{y}_1^n) = \sum_{|p_i| > 1} \log|p_i|, \quad (133)$$

revealing that the gap in (114a) is exactly $\bar{h}(\bar{\mathbf{c}}_1^\infty)$. In addition, in the perfect-feedback scenario, Theorem 8 extends the validity of (115) from the Gaussian i.i.d. $\mathbf{u}$ and Gaussian $\mathbf{x}_0$ considered in Reference [14], Theorem 14, to an entropy-balanced $\mathbf{u}$ and an $\mathbf{x}_0$ with finite variance and finite differential entropy.

8.2. Rate Distortion Function for Non-Stationary Processes

In this section, we obtain a simpler proof of a result by Gray, Hashimoto and Arimoto [15–17], which compares the rate distortion function (RDF) of a non-stationary auto-regressive Gaussian process $\mathbf{x}_1^\infty$ (of a certain class to be defined shortly) to that of a corresponding stationary version, under MSE distortion. Our proof is based upon the ideas developed in the previous sections, and extends the class of non-stationary sources for which the results in Reference [15–17] are valid.

To be more precise, let $\{a_i\}_{i=1}^\infty$ and $\{\tilde{a}_i\}_{i=1}^\infty$ be, respectively, the impulse responses of two linear time-invariant filters A and $\tilde{A}$ with rational transfer functions

$$A(z) = \frac{z^M}{\prod_{i=1}^M (z - p_i)}, \quad (134)$$

$$\tilde{A}(z) = \frac{z^M}{\prod_{i=1}^M |p_i^*|(z - 1/p_i^*)}, \quad (135)$$

where $|p_i| > 1, \forall i = 1, \dots, M$. From these definitions, it is clear that $A(z)$ is unstable, $\tilde{A}(z)$ is stable, and

$$|A(e^{j\omega})| = |\tilde{A}(e^{j\omega})|, \quad \forall \omega \in [-\pi, \pi]. \quad (136)$$

Notice also that $\lim_{|z| \to \infty} A(z) = 1$ and $\lim_{|z| \to \infty} \tilde{A}(z) = 1/\prod_{i=1}^M |p_i|$; thus,

$$a_0 = 1, \qquad\qquad \tilde{a}_0 = \prod_{i=1}^M |p_i|^{-1}. \quad (137)$$

Consider the non-stationary random sequence (source) x_1^∞ and the asymptotically stationary source $\tilde{x}_1^\infty$ generated by passing a stationary Gaussian process w_1^∞ through $A(z)$ and $\tilde{A}(z)$, respectively, which can be written as

$$x_n^1 = A_n w_1^n, \quad n = 1, \dots, \tag{138}$$

$$\tilde{x}_n^1 = \tilde{A}_n w_1^n, \quad n = 1, \dots. \tag{139}$$

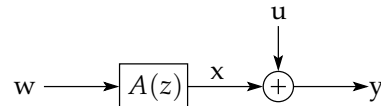(A block-diagram associated with the construction of x is presented in Figure 6.)

Figure 6. Block diagram representation of how the non-stationary source x_1^∞ is built and then reconstructed as $y = x + u$.

Define the rate-distortion functions for these two sources as

$$R_x(D) \triangleq \lim_{n \to \infty} R_{x,n}(D), \qquad R_{x,n}(D) \triangleq \min \frac{1}{n} I(x_1^n; x_1^n + u_1^n), \tag{140}$$

$$R_{\tilde{x}}(D) \triangleq \lim_{n \to \infty} R_{\tilde{x},n}(D), \qquad R_{\tilde{x},n}(D) \triangleq \min \frac{1}{n} I(\tilde{x}_1^n; \tilde{x}_1^n + \tilde{u}_1^n), \tag{141}$$

where, for each n, the minima are taken over all the conditional probability density functions $f_{u_1^n | x_1^n}$ and $f_{\tilde{u}_1^n | \tilde{x}_1^n}$ yielding $E\left[\|u_n^1\|^2\right]/n \leq D$ and $E\left[\|\tilde{u}_n^1\|^2\right]/n \leq D$, respectively.

The above rate-distortion functions have been characterized in Reference [15–17] for the case in which w_1^∞ is an i.i.d. Gaussian process. In particular, it is explicitly stated in Reference [16,17] that, for that case,

$$R_x(D) - R_{\tilde{x}}(D) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log |A^{-1}(e^{j\omega})| d\omega = \sum_{i=1}^{M} \log |p_i|. \tag{142}$$

We will next provide an alternative and simpler proof of this result, and extend its validity for general (not-necessarily stationary) Gaussian w_1^∞, using the entropy gain properties of non-minimum phase filters established in Section 6. Indeed, the approach in References [15–17] is based upon asymptotically-equivalent Toeplitz matrices in terms of the signals' covariance matrices. This restricts w_1^∞ to be Gaussian and i.i.d. and $A(z)$ to be an all-pole unstable transfer function, and then, the only non-stationarity allowed is that arising from unstable poles. For instance, a cyclo-stationary innovation followed by an unstable filter $A(z)$ would yield a source which cannot be treated using Gray and Hashimoto's approach. By contrast, the reasoning behind our proof lets w_1^∞ be any entropy-balanced Gaussian process with bounded differential entropy rate, and then let the source be $A w$, with $A(z)$ having unstable poles (and possibly zeros and stable poles, as well).

The statement is as follows:

Theorem 9. *Let w_1^∞ be any Gaussian entropy-balanced process with bounded differential entropy rate, and let x_1^∞ and $\tilde{x}_1^\infty$ be as defined in (138) and (139), respectively. Then, (142) holds.*

Thanks to the ideas developed in the previous sections, it is possible to give an intuitive outline of the proof of this theorem (given in Appendix B, page 40) by using a sequence of block diagrams. More precisely, consider the diagrams shown in Figure 7.
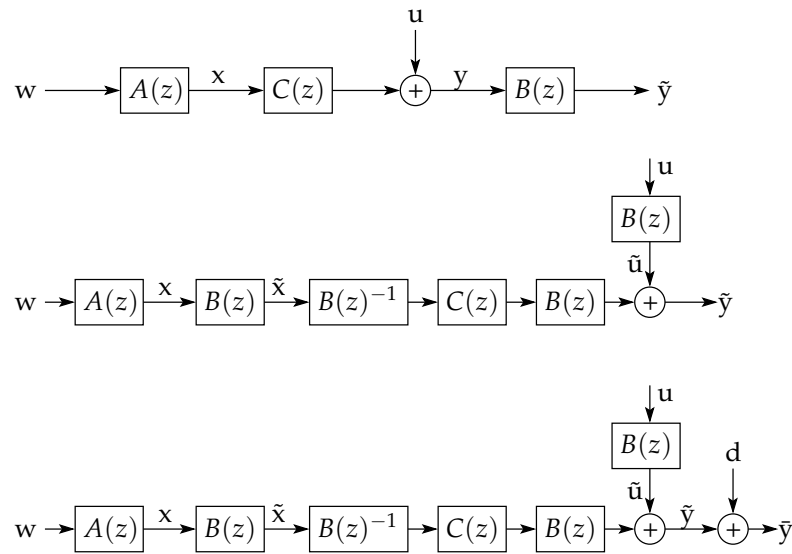
Figure 7. Block-diagram representation of the changes of variables in the proof of Theorem 9.

In the top diagram in this figure, suppose that $y = Cx + u$ realizes the RDF for the non-stationary source x. The sequence u is independent of **x**, and the linear filter $C(z)$ is such that the error $(y - x) \perp\!\!\!\perp y$ (a necessary condition for minimum MSE optimality). The filter $B(z)$ is the Blaschke product of $A(z)$ (see (A83) in Appendix B) (a stable, NMP filter with unit frequency response magnitude such that $\tilde{x} = Bx$).

If one moves the filter $B(z)$ towards the source, then the middle diagram in Figure 7 is obtained. By doing this, the stationary source $\tilde{x}$ appears with an additive error signal $\tilde{u}$ that has the same asymptotic variance as u, reconstructed as $\tilde{y} = C\tilde{x} + \tilde{u}$. From the invertibility of $B(z)$, it also follows that the mutual information rate between $\tilde{x}$ and $\tilde{y}$ equals that between x and y. Thus, the channel $\tilde{y} = C\tilde{x} + \tilde{u}$ has the same rate and distortion as the channel $y = Cx + u$.

However, if one now adds a short disturbance d to the error signal $\tilde{u}$ (as depicted in the bottom diagram of Figure 7), then the resulting additive error term $\bar{u} = \tilde{u} + d$ will be independent of $\tilde{x}$ and will have the same asymptotic variance as $\tilde{u}$. Nonetheless, the differential entropy rate of $\bar{u}$ will exceed that of $\tilde{u}$ by the RHS of (142). This will make the mutual information rate between $\tilde{x}$ and $\bar{y}$ to be less than that between $\tilde{x}$ and $\tilde{y}$ by the same amount. Hence, $R_{\tilde{x}}(D)$ is at most $R_x(D) - \sum_{i=1}^{m} \log|p_i|$. A similar reasoning can be followed to prove that $R_x(D) - R_{\tilde{x}}(D) \leq \sum_{i=1}^{m} \log|p_i|$.

8.3. The Feedback Channel Capacity of (Non-White) Gaussian Channels

Consider a non-white additive Gaussian channel of the form

$$y_k = x_k + z_k, \tag{143}$$

where the input x is subject to the power constraint

$$\lim_{n \to \infty} \frac{1}{n} E[\|x_n^1\|^2] \leq P, \tag{144}$$

and z_1^∞ is a stationary Gaussian process.

The feedback information capacity of this channel is realized by a Gaussian input x, and is given by

$$C_{\text{FB}} = \lim_{n \to \infty} \max_{K_{x_n^1} : \frac{1}{n} \text{tr}\{K_{x_n^1}\} \leq P} I(x_n^1; y_n^1), \tag{145}$$

where $K_{\mathbf{x}_n^1}$ is the covariance matrix of $\mathbf{x}_n^1$, and, for every $k \in \mathbb{N}$, the input $\mathbf{x}_k$ is allowed to depend upon the channel outputs $\mathbf{y}_1^{k-1}$ (since there exists a causal, noise-less feedback channel with one-step delay).

In Reference [9], it was shown that if z is an auto-regressive moving-average process of M-th order, then C_{FB} can be achieved by the scheme shown in Figure 8. In this system, B is a strictly causal and stable finite-order filter and v_1^∞ is Gaussian with $v_k = 0$ for all $k > M$ and such that $\mathbf{v}_n^1$ is Gaussian with a positive-definite covariance matrix $K_{\mathbf{v}_M^1}$.
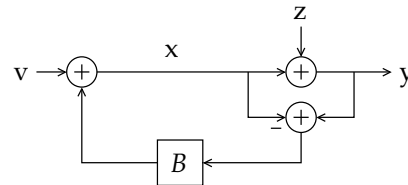

Figure 8. Block diagram representation a non-white Gaussian channel y $=$ x $+$ z and the coding scheme considered in Reference [9].

Here, we use the ideas developed in Section 6 to show that **the information rate achieved by the capacity-achieving scheme proposed in Reference [9] drops to zero if there exists any additive disturbance of length at least M and finite differential entropy affecting the output, no matter how small**.

To see this, notice that, in this case, and for all $n > M$,

$$I(\mathbf{x}_1^n; \mathbf{y}_1^n) = I(\mathbf{v}_1^M; \mathbf{y}_1^n) = h(\mathbf{y}_n^1) - h(\mathbf{y}_n^1 | \mathbf{v}_n^1), \tag{146}$$

$$= h(\mathbf{y}_n^1) - h((\mathbf{I}_n + \mathbf{B}_n)\mathbf{z}_n^1 + \mathbf{v}_n^1 | \mathbf{v}_M^1), \tag{147}$$

$$= h(\mathbf{y}_n^1) - h((\mathbf{I}_n + \mathbf{B}_n)\mathbf{z}_n^1 | \mathbf{v}_M^1), \tag{148}$$

$$= h(\mathbf{y}_n^1) - h((\mathbf{I}_n + \mathbf{B}_n)\mathbf{z}_n^1) = h(\mathbf{y}_n^1) - h(\mathbf{z}_n^1), \tag{149}$$

$$= h((\mathbf{I}_n + \mathbf{B}_n)\mathbf{z}_n^1 + \mathbf{v}_n^1) - h(\mathbf{z}_n^1), \tag{150}$$

since $\det(\mathbf{I}_n + \mathbf{B}_n) = 1$. From Theorem 4, this gap between differential entries is precisely the entropy gain introduced by $\mathbf{I}_n + \mathbf{B}_n$ to an input $\mathbf{z}_n^1$ when the output is affected by the disturbance $\mathbf{v}_M^1$. Thus, from Theorem 4, the capacity of this scheme will correspond to $\frac{1}{2\pi}\int_{-\pi}^{\pi} \log|1 + B(e^{j\omega})|d\omega = \sum_{|\rho_i|>1} \log|\rho_i|$, where $\{\rho_i\}_{i=1}^M$ are the zeros of $1 + B(z)$, which is precisely the result stated in Reference [9], Theorem 4.1.

However, if the output is now affected by an additive disturbance d_1^∞ not passing through $B(z)$ such that $d_k = 0, \forall k > M$ and $|h(\mathbf{d}_M^1)| < \infty$, with $d_1^\infty \perp\!\!\!\perp (v_1^M, z_1^\infty)$, then we will have

$$\mathbf{y}_n^1 = \mathbf{v}_n^1 + (\mathbf{I}_n + \mathbf{B}_n)\mathbf{z}_n^1 + \mathbf{d}_n^1. \tag{151}$$

In this case,

$$I(\mathbf{x}_1^n; \mathbf{y}_1^n) = I(\mathbf{v}_1^M; \mathbf{y}_1^n) = h(\mathbf{y}_n^1) - h(\mathbf{y}_n^1 | \mathbf{v}_n^1), \tag{152}$$

$$= h(\mathbf{y}_n^1) - h((\mathbf{I}_n + \mathbf{B}_n)\mathbf{z}_n^1 + \mathbf{v}_n^1 + \mathbf{d}_n^1 | \mathbf{v}_M^1), \tag{153}$$

$$= h(\mathbf{y}_n^1) - h((\mathbf{I}_n + \mathbf{B}_n)\mathbf{z}_n^1 + \mathbf{d}_n^1 | \mathbf{v}_M^1), \tag{154}$$

$$= h(\mathbf{y}_n^1) - h((\mathbf{I}_n + \mathbf{B}_n)\mathbf{z}_n^1 + \mathbf{d}_n^1). \tag{155}$$

But $\lim_{n\to\infty} \frac{1}{n}(h((\mathbf{I}_n + \mathbf{B}_n)\mathbf{z}_n^1 + \mathbf{v}_n^1 + \mathbf{d}_n^1) - h((\mathbf{I}_n + \mathbf{B}_n)\mathbf{z}_n^1 + \mathbf{d}_n^1)) = 0$, which follows directly from applying Theorem 4 to each of the differential entries. Notice that this result holds irrespective of how small the power of the disturbance may be.

Thus, the capacity-achieving scheme proposed in Reference [9] (and further studied in Reference [28]), although of groundbreaking theoretical importance, would yield zero

rate in any practical situation, since in every physically implemented scheme, signals are unavoidably affected by some amount of noise.

9. Conclusions

We have provided an intuitive explanation and a rigorous characterization of the entropy gain of a *linear time-invariant* (LTI) system, defined as the difference between the differential entropy rates of its output and input random signals. The continuous-time version of this problem, considered by Shannon in Theorem 14 of his 1948 landmark paper, involves an LTI system G_c band limited to B [Hz]. For this scenario, we restricted our attention to systems such that the samples of its unit-impulse response, taken $(2B)^{-1}$ seconds apart, correspond to the unit-impulse response $g_0, g_1, \ldots$ of a causal and stable discrete-time system G. We show that the entropy gain in this case is $\log|g_0|$, which implies that, for this class of systems, Shannon's Theorem 14 holds if and only if G_c has a corresponding discrete-time G that is *minimum phase* (MP).

For the discrete-time case, we introduced a new notion referred to as effective differential entropy, which quantifies the amount of uncertainty in vector signals that are confined to subspaces of lower dimensionality than that of the signals themselves. (Note that this is not possible by the conventional notion of differential entropy, which simply diverges to minus infinity.) It turns out that the difference in effective differential entropy rate between an n-length input to an LTI discrete-time system with frequency response $G(e^{j\omega})$, and its full length output, as n tends to infinity, equals $\frac{1}{2\pi}\int_{-\pi}^{\pi}|G(e^{j\omega})|d\omega$.

When comparing input and output sequences of equal length, our analysis revealed that, in the absence of external random disturbances, the entropy gain of a discrete-time LTI system G with unit-impulse response $g_0, g_1, \ldots$ is simply $\log|g_0|$. An entropy gain greater than $\log|g_0|$ can be obtained only if a random signal is added to the output of G and if such output process has statistical properties that make it susceptible to the added random signal. In order to characterize the role of G, its input has been assumed to be *entropy balanced* (EB), a notion introduced herein. Crucially, the differential entropy rate of an EB process is not susceptible to random signals. EB processes constitute a large family that includes Gaussian processes with bounded, non-vanishing variance. We also show that (i) the sum of an EB process and any bounded variance process is EB, too, and (ii) passing an EB process by a stable MP filter yields an EB process. When the input is EB, we show that if G has NMP zeros $\rho_1, \rho_2, \ldots, \rho_m$, then the largest possible entropy gain is $|g_0| + \sum_{i=1}^{m}\log|\rho_i|$, which equals $\frac{1}{2\pi}\int_{-\pi}^{\pi}|G(e^{j\omega})|d\omega$. This upper bound is achieved by adding a finite-length output disturbance with finite variance and bounded differential entropy if and only if its length is at least m, no matter how tiny its variance may be. The same entropy gain is also obtained if G has a random initial state with bounded differential entropy and finite variance.

We used these fundamental insights about when the entropy gain occurs in order to establish a new and more general proof of the quadratic rate-distortion function for non-stationary Gaussian sources. Moreover, we demonstrated that the information rate of the capacity-achieving scheme proposed in Reference [9] for the auto regressive Gaussian channel with feedback drops to zero in the presence of any additive disturbance in the channel input or output of sufficient (finite) length, no matter how small it may be. This has crucial implications in any physical setup, where noise is unavoidable.

Author Contributions: Conceptualization, M.S.D. and M.M.; Investigation, M.S.D., M.M. and J.Ø.; Writing—original draft, M.S.D.; Writing—review and editing, M.M. and J.Ø. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proof of Theorem 3

The total length of the output ℓ, will grow with the length n of the input, if G is FIR, and will be infinite, if G is IIR. Letting $\eta + 1$ be the length of the impulse response of G in the FIR case, we define the *output-length function*

$$\ell(n) \triangleq \text{length of } \mathbf{y} \text{ when input is } \mathbf{u}_n^1 = \begin{cases} n + \eta & \text{, if } G \text{ is FIR,} \\ \infty & \text{, if } G \text{ is IIR.} \end{cases} \tag{A1}$$

It is also convenient to define the sequence of matrices $\{\breve{G}_n\}_{n=1}^{\infty}$, where $\breve{G}_n \in \mathbb{R}^{\ell(n) \times n}$ is Toeplitz with $[\breve{G}_n]_{i,j} = 0, \forall i < j, [\breve{G}_n]_{i,j} = g_{i-j}, \forall i \geq j$. This allows one to write the *entire* output y_1^{ℓ} of a causal LTI filter G with impulse response $\{g_k\}_{k=0}^{\eta}$ to an input u_1^{∞} as

$$\mathbf{y}_{\ell(n)}^1(\mathbf{u}_n^1) = \breve{G}_n \mathbf{u}_n^1. \tag{A2}$$

Let the SVD of $\breve{G}_n$ be $\breve{G}_n = \breve{Q}_n^T \breve{D}_n \breve{R}_n$, where $\breve{Q}_n \in \mathbb{R}^{n \times \ell(n)}$ has orthonormal rows, $\breve{D}_n \in \mathbb{R}^{n \times n}$ is diagonal with positive elements, and $\breve{R}_n \in \mathbb{R}^{n \times n}$ is unitary.

The effective differential entropy of $y_1^{\ell(n)}(u_1^n)$ exceeds the differential entropy of u_1^n by

$$\breve{h}(\mathbf{y}_{\ell(n)}^1(\mathbf{u}_n^1)) - h(\mathbf{u}_n^1) = h(\breve{Q}_n \breve{G}_n \mathbf{u}_n^1) - h(\mathbf{u}_n^1) = h(\breve{D}_n \breve{R}_n \mathbf{u}_n^1) - h(\mathbf{u}_n^1) = \log \det(\breve{D}_n). \tag{A3}$$

The determinant of D_n can be related to that of $\breve{G}_n^T \breve{G}_n$ by noticing that

$$\breve{G}_n^T \breve{G}_n = (\breve{Q}_n^T \breve{D}_n \breve{R}_n)^T (\breve{Q}_n^T \breve{D}_n \breve{R}_n) = \breve{R}_n^T \breve{D}_n \breve{Q}_n \breve{Q}_n^T \breve{D}_n \breve{R}_n = \breve{R}_n^T \breve{D}_n^2 \breve{R}_n. \tag{A4}$$

Since $\breve{R}_n$ is unitary, it follows that $\det \breve{D}_n^2 = \det \breve{G}_n^T \breve{G}_n$, which from (A3) means that

$$\breve{h}(\mathbf{y}_{\ell(n)}^1(\mathbf{u}_n^1)) - h(\mathbf{u}_n^1) = \frac{1}{2} \log(\det(\breve{G}_n^T \breve{G}_n)). \tag{A5}$$

The product $\boldsymbol{H}_n \triangleq \breve{G}_n^T \breve{G}_n$ is a symmetric Toeplitz matrix, with its first column, $[h_0 \, h_1 \, \cdots \, h_{n-1}]^T$, given by

$h_i = \sum_{k=0}^{n} g_k g_{k-i}$. Thus, the sequence $\{h_i\}_{i=0}^{n-1}$ corresponds to the samples 0 to $n-1$ of those resulting from the complete convolution $g * g^-$, even when the filter G is IIR, where g^- denotes the time-reversed (possibly infinitely long) response g. Consequently, and since $G(z)$ has no zeros on the unit circle, and g is absolutely summable, we can use the Grenander and Szegö's theorem [29], and Reference [18], Theorem 4.2, to obtain that

$$\lim_{n \to \infty} \log\left(\det(\breve{G}_n^T \breve{G}_n)^{1/n}\right) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log\left|G(e^{j\omega})\right|^2 d\omega. \tag{A6}$$

In order to finish the proof, we divide (A5) by n, take the limit as $n \to \infty$, and replace (A6) in the latter.

Appendix B. Proofs of Results Stated in the Previous Sections

Proof of Lemma 1. Let $\sigma_{u(k)}^2$ be the variance of $u(k)$. Thus, $h(\mathbf{u}_n^1) = \frac{1}{2} \log((2\pi e)^n \det(\text{diag} \{\sigma_{u(k)}^2\}_{k=1}^n))$. Let $\mathbf{y}_n^{\nu+1} \triangleq \boldsymbol{\Phi}_n \mathbf{u}_n^1$. Then, $\boldsymbol{K}_{\mathbf{y}_n^{\nu+1}} = \boldsymbol{\Phi}_n \text{diag}\{\sigma_{u(k)}^2\}_{k=1}^n \boldsymbol{\Phi}_n^T$. As a consequence,

$$h(\mathbf{y}_n^{\nu+1}) = \frac{1}{2} \log\left((2\pi e)^{[n-\nu]} \det\left(\boldsymbol{\Phi}_n \text{diag}\{\sigma_{u(k)}^2\}_{k=1}^n \boldsymbol{\Phi}_n^T\right)\right). \tag{A7}$$

But from the Courant-Fischer theorem [30],

$$\log\Big(\det(\mathrm{diag}\{\sigma_{\mathrm{u}(k)}^2\}_{k=1}^n)\Big) - \nu\log(\hat{\sigma}^2) \leq \log\Big(\det\Big(\boldsymbol{\Phi}_n\,\mathrm{diag}\{\sigma_{\mathrm{u}(k)}^2\}_{k=1}^n\boldsymbol{\Phi}_n^T\Big)\Big) \leq \log\Big(\det(\mathrm{diag}\{\sigma_{\mathrm{u}(k)}^2\}_{k=1}^n)\Big) - \nu\log(\check{\sigma}^2); \tag{A8}$$

thus, $\lim_{n\to\infty}\frac{1}{n}(h(\mathbf{y}_n^{\nu+1}) - h(\mathbf{u}_n^1)) = 0$, satisfying Condition ii) in Definition 2. Adding to this the fact that, in this case, $\sigma_{\mathrm{u}(n)}^2 \leq \hat{\sigma}^2 < \infty$ for all n, Condition i) in Definition 2 is satisfied, as well, completing the proof. $\quad\square$

Proof of Lemma 2. Let $\{b_{i,\ell}\}_{\ell=1}^{\infty}$ be the intervals (bins) in $\mathbb{R}$ where the sample $\mathrm{u}(i)$ has constant PDF. Define the discrete random process $\mathbf{c}_1^{\infty}$, where $\mathrm{c}(i) = \ell$ if and only if $\mathrm{u}(i) \in b_{i,\ell}$. Let $\mathbf{y}_n^{\nu+1} \triangleq \boldsymbol{\Phi}_n\mathbf{u}_n^1$ where $\boldsymbol{\Phi}_n \in \mathbb{R}^{(n-\nu)\times n}$ has orthonormal rows. Then,

$$h(\mathbf{y}_n^{\nu+1}) = h(\mathbf{y}_n^{\nu+1}|\mathbf{c}_n^1) + I(\mathbf{c}_n^1; \mathbf{y}_n^{\nu+1}) \tag{A9}$$

$$\leq h(\mathbf{y}_n^{\nu+1}|\mathbf{c}_n^1) + I(\mathbf{c}_n^1; \mathbf{u}_n^1), \tag{A10}$$

where the inequality is due to the fact that $\mathbf{u}_n^1$ and $\mathbf{y}_n^{\nu+1}$ are deterministic functions of $\mathbf{u}_n^1$; hence, $\mathbf{c}_n^1 \longleftrightarrow \mathbf{u}_n^1 \longleftrightarrow \mathbf{y}_n^{\nu+1}$. Subtracting $h(\mathbf{u}_n^1)$ from (A9) we obtain

$$h(\mathbf{y}_n^{\nu+1}) - h(\mathbf{u}_n^1) \leq h(\mathbf{y}_n^{\nu+1}|\mathbf{c}_n^1) + I(\mathbf{c}_n^1; \mathbf{u}_n^1) - h(\mathbf{u}_n^1) \tag{A11}$$

$$= h(\mathbf{y}_n^{\nu+1}|\mathbf{c}_n^1) - h(\mathbf{u}_n^1|\mathbf{c}_n^1). \tag{A12}$$

Hence,

$$\lim_{n\to\infty}\frac{1}{n}\Big(h(\mathbf{y}_n^{\nu+1}) - h(\mathbf{u}_n^1)\Big) \leq \lim_{n\to\infty}\frac{1}{n}\Big(h(\mathbf{y}_n^{\nu+1}|\mathbf{c}_n^1) - h(\mathbf{u}_n^1|\mathbf{c}_n^1)\Big) = 0, \tag{A13}$$

where the last equality follows from Lemma A1 (in Appendix C) whose conditions are met because, given $\mathbf{c}_n^1$, the sequence $\mathbf{u}_n^1$ has independent entries each of them distributed uniformly over a possibly different interval with finite and positive measure. The opposite inequality is obtained by following the same steps as in the proof of Lemma A1, from (A124) onwards, which completes the proof. $\quad\square$

Proof of Lemma 3. Let $\mathbf{y}_n^1 \triangleq [\boldsymbol{\Psi}_n^T|\boldsymbol{\Phi}_n^T]^T\mathbf{w}_n^1$, where $[\boldsymbol{\Psi}_n^T|\boldsymbol{\Phi}_n^T]^T \in \mathbb{R}^{n\times n}$ is a unitary matrix and where $\boldsymbol{\Psi}_n \in \mathbb{R}^{\nu\times n}$ and $\boldsymbol{\Phi}_n \in \mathbb{R}^{(n-\nu)\times n}$ have orthonormal rows.
　　Then,

$$h(\mathbf{y}_n^{\nu+1}) = h(\mathbf{y}_n^1) - h(\mathbf{y}_\nu^1|\mathbf{y}_n^{\nu+1}) = h(\mathbf{w}_n^1) - h(\mathbf{y}_\nu^1|\mathbf{y}_n^{\nu+1}). \tag{A14}$$

We can lower bound $h(\mathbf{y}_\nu^1|\mathbf{y}_n^{\nu+1})$ as follows:

$$h(\mathbf{y}_\nu^1|\mathbf{y}_n^{\nu+1}) = h(\boldsymbol{\Psi}_n\mathbf{u}_n^1 + \boldsymbol{\Psi}_n\mathbf{v}_n^1 \mid \boldsymbol{\Phi}_n\mathbf{u}_n^1 + \boldsymbol{\Phi}_n\mathbf{v}_n^1) \tag{A15}$$

$$\overset{(a)}{\geq} h(\boldsymbol{\Psi}_n\mathbf{u}_n^1 + \boldsymbol{\Psi}_n\mathbf{v}_n^1 \mid \boldsymbol{\Phi}_n\mathbf{u}_n^1 + \boldsymbol{\Phi}_n\mathbf{v}_n^1,\ \mathbf{v}_n^1) \tag{A16}$$

$$= h(\boldsymbol{\Psi}_n\mathbf{u}_n^1 \mid \boldsymbol{\Phi}_n\mathbf{u}_n^1 + \boldsymbol{\Phi}_n\mathbf{v}_n^1,\ \mathbf{v}_n^1) \tag{A17}$$

$$= h(\boldsymbol{\Psi}_n\mathbf{u}_n^1 \mid \boldsymbol{\Phi}_n\mathbf{u}_n^1,\ \mathbf{v}_n^1) \tag{A18}$$

$$\overset{(b)}{=} h(\boldsymbol{\Psi}_n\mathbf{u}_n^1 \mid \boldsymbol{\Phi}_n\mathbf{u}_n^1) \tag{A19}$$

$$\overset{(c)}{=} h(\mathbf{u}_n^1) - h(\boldsymbol{\Phi}_n\mathbf{u}_n^1), \tag{A20}$$

where (a) holds because conditioning does not increase entropy, (b) is from the fact that $\mathbf{u}_n^1 \perp\!\!\!\perp \mathbf{v}_n^1$, and (c) follows from the chain rule of entropy.
　　Substituting this result into (A14), dividing by n, taking the limit as $n \to \infty$, and recalling that u_1^{∞} is entropy balanced, we conclude that $\lim_{n\to\infty}\frac{1}{n}(h(\boldsymbol{\Phi}_n\mathbf{w}_n^1) - h(\mathbf{w}_n^1)) \leq 0$.
　　The opposite bound over $h(\mathbf{y}_\nu^1|\mathbf{y}_n^{\nu+1})$ can be obtained from

$$h(\mathbf{y}_\nu^1|\mathbf{y}_n^{\nu+1}) = h(\boldsymbol{\Psi}_n\mathbf{u}_n^1 + \boldsymbol{\Psi}_n\mathbf{v}_n^1 \mid \boldsymbol{\Phi}_n\mathbf{u}_n^1 + \boldsymbol{\Phi}_n\mathbf{v}_n^1) \leq h(\boldsymbol{\Psi}_n\mathbf{u}_n^1 + \boldsymbol{\Psi}_n\mathbf{v}_n^1) \leq h(\boldsymbol{\Psi}_n(\mathbf{w}_G)_n^1), \tag{A21}$$

where $(\mathbf{w}_G)_n^1$ is a jointly Gaussian sequence with the same second-order moment as $\mathbf{w}_n^1$. Therefore, $h(\boldsymbol{\Psi}_n(\mathbf{w}_G)_n^1) = \frac{1}{2}\log((2\pi\,\mathrm{e})^\nu \det(\boldsymbol{\Psi}_n K_{\mathbf{w}_n^1}\boldsymbol{\Psi}_n^T)) \leq \frac{\nu}{2}\log(2\pi\,\mathrm{e}\,\lambda_{\max}(K_{\mathbf{w}_n^1}))$. But $\mathbf{w}_n^1$ satisfies the assumptions of Proposition A2; thus, $\lim_{n\to\infty} n^{-1}\log(\lambda_{\max}(K_{\mathbf{w}_n^1})) = 0$. Therefore, $\lim_{n\to\infty} n^{-1}h(\boldsymbol{\Psi}_n(\mathbf{w}_G)_n^1) \leq 0$, which substituted in (A14) yields

$$\lim_{n\to\infty} -\frac{1}{n}h(\mathbf{y}_\nu^1|\mathbf{y}_n^{\nu+1}) = \lim_{n\to\infty} \frac{1}{n}(h(\boldsymbol{\Phi}_n\mathbf{w}_n^1) - h(\mathbf{w}_n^1)) \geq 0.$$

Hence, w_1^∞ satisfies Condition ii) of Definition 2. Since w_1^∞ also satisfies Condition i) of Definition 2, it follows that w_1^∞ is entropy balanced, completing the proof. □

Proof of Lemma 4. Pick any $\nu \in \mathbb{N}$ and let $\mathbf{y}_n^1 \triangleq [\boldsymbol{\Phi}_n^T|\boldsymbol{\Psi}_n^T]^T\mathbf{w}_n^1$ where $[\boldsymbol{\Phi}_n^T|\boldsymbol{\Psi}_n^T]^T \in \mathbb{R}^{n\times n}$ is a unitary matrix and the matrices $\boldsymbol{\Psi}_n \in \mathbb{R}^{\nu\times n}$ and $\boldsymbol{\Phi}_n \in \mathbb{R}^{(n-\nu)\times n}$ have orthonormal rows. Since $\mathbf{w}_n^1 = G_n\mathbf{u}_n^1$, we have that

$$\boldsymbol{\Phi}_n\mathbf{w}_n^1 = \boldsymbol{\Phi}_n G_n\mathbf{u}_n^1. \tag{A22}$$

Let $\boldsymbol{\Phi}_n G_n = A_n\boldsymbol{\Sigma}_n B_n$ be the SVD of $\boldsymbol{\Phi}_n G_n$, where $A_n \in \mathbb{R}^{(n-\nu)\times(n-\nu)}$ is an orthogonal matrix, $B_n \in \mathbb{R}^{(n-\nu)\times n}$ has orthonormal rows and $\boldsymbol{\Sigma}_n \in \mathbb{R}^{(n-\nu)\times(n-\nu)}$ is a diagonal matrix with the singular values of $\boldsymbol{\Phi}_n G_n$.

Hence

$$h(\boldsymbol{\Phi}_n\mathbf{w}_n^1) = h(\boldsymbol{\Phi}_n G_n\mathbf{u}_n^1) = h(A_n\boldsymbol{\Sigma}_n B_n\mathbf{u}_n^1) = \log\det(\boldsymbol{\Sigma}_n) + h(B_n\mathbf{u}_n^1). \tag{A23}$$

The singular values of $\boldsymbol{\Phi}_n G_n$ are $\sigma_i(\boldsymbol{\Phi}_n G_n) = \sqrt{\lambda_i(\boldsymbol{\Phi}_n G_n G_n^T\boldsymbol{\Phi}_n^T)}$, $i = 1, 2, \ldots, n-\nu$. Now, notice that

$$\lambda_i([\boldsymbol{\Phi}_n^T|\boldsymbol{\Psi}_n^T]^T G G^T[\boldsymbol{\Phi}_n^T|\boldsymbol{\Psi}_n^T]) = \lambda_i(G G^T) \tag{A24}$$

and that

$$[\boldsymbol{\Phi}_n^T|\boldsymbol{\Psi}_n^T]^T G G^T[\boldsymbol{\Phi}_n^T|\boldsymbol{\Psi}_n^T] = \begin{pmatrix} \boldsymbol{\Phi}_n G G^T\boldsymbol{\Phi}_n^T & \boldsymbol{\Phi}_n G G^T\boldsymbol{\Psi}_n^T \\ \boldsymbol{\Psi}_n G G^T\boldsymbol{\Phi}_n^T & \boldsymbol{\Psi}_n G G^T\boldsymbol{\Psi}_n^T \end{pmatrix}. \tag{A25}$$

Thus, from (A24) and the Cauchy eigenvalue interlacing theorem [30],

$$\lambda_i(G_n G_n^T) \leq \lambda_i(\boldsymbol{\Phi}_n G G^T\boldsymbol{\Phi}_n^T) \leq \lambda_{i+\nu}(G_n G_n^T), \quad i = 1, \ldots, n-\nu. \tag{A26}$$

Hence,

$$-\frac{\nu}{2n}\log\left(\lambda_{\max}(G_n G_n^T)\right) \leq \frac{1}{n}\log(\det(\boldsymbol{\Sigma}_n)) - \frac{1}{n}\log|\det(G_n)| \leq -\frac{\nu}{2n}\log\left(\lambda_{\min}(G_n G_n^T)\right). \tag{A27}$$

Recalling that G is minimum phase (which guarantees that its singular values change at most polynomially with n, due to Lemma 7), we conclude that

$$\lim_{n\to\infty} \frac{1}{n}\log(\det(\boldsymbol{\Sigma}_n)) = \frac{1}{n}\log|\det(G_n)|. \tag{A28}$$

Substituting back into (A23), we arrive to

$$\lim_{n\to\infty} \frac{1}{n}h(\boldsymbol{\Phi}_n\mathbf{w}_n^1) \overset{(a)}{=} \lim_{n\to\infty} \frac{1}{n}\log|\det(G_n)| + \lim_{n\to\infty} \frac{1}{n}h(\mathbf{u}_1^n) = \lim_{n\to\infty} \frac{1}{n}h(\mathrm{w}_1^n), \tag{A29}$$

where (a) holds because u_1^∞ is entropy balanced. This completes the proof. □

Proof of Lemma 5. Let $\{\boldsymbol{\Psi}_n\}_{n=1}^{\infty}$ be a sequence of matrices, each $\boldsymbol{\Psi}_n \in \mathbb{R}^{\kappa \times n}$ with orthonormal rows spanning a subspace of $\mathbb{R}^n$ that contains the span of the columns of $[\boldsymbol{\Phi}]_n^1$. For each $n \in \mathbb{N}$, let $\overline{\boldsymbol{\Psi}}_n \in \mathbb{R}^{(n-\kappa) \times n}$ be such that $H_n \triangleq [\boldsymbol{\Psi}_n^T \mid \overline{\boldsymbol{\Psi}}_n^T]^T$ is a unitary matrix. Then,

$$h(\overline{\boldsymbol{\Psi}}_n \mathbf{y}_n^1) = h(\overline{\boldsymbol{\Psi}}_n \mathbf{u}_n^1). \tag{A30}$$

Thus,

$$\lim_{n \to \infty} \frac{1}{n}\left(h(\mathbf{y}_n^1) - h(\mathbf{u}_n^1)\right) = \lim_{n \to \infty} \frac{1}{n}\left((h(\mathbf{y}_n^1) - h(\overline{\boldsymbol{\Psi}}_n \mathbf{y}_n^1)) - (h(\mathbf{u}_n^1) - h(\overline{\boldsymbol{\Psi}}_n \mathbf{u}_n^1))\right) = 0, \tag{A31}$$

where the last equality holds because u_1^{∞} is entropy balanced and y_1^{∞} is entropy balanced (from Lemma 3). This completes the proof. $\square$

Proof of Lemma 6. Since Q_n is unitary, we have that

$$h(\mathbf{y}_n^1) = h(Q_n \mathbf{y}_n^1) = h(\overbrace{\underbrace{D_n R_n \mathbf{u}_n^1}_{\mathbf{v}_n^1} + \underbrace{Q_n \mathbf{z}_n^1}_{\bar{\mathbf{z}}_n^1}}^{\mathbf{w}_n^1}) = h(\mathbf{w}_n^1), \tag{A32}$$

where

$$\mathbf{w}_n^1 \triangleq Q_n \mathbf{y}_n^1 = \mathbf{v}_n^1 + \bar{\mathbf{z}}_n^1, \tag{A33}$$

$$\mathbf{v}_n^1 \triangleq D_n R_n \mathbf{u}_n^1, \tag{A34}$$

$$\bar{\mathbf{z}}_n^1 \triangleq Q_n \mathbf{z}_n^1. \tag{A35}$$

Thus,

$$h(\mathbf{y}_n^1) = h(\mathbf{w}_1^n) \overset{(a)}{=} h(\mathbf{w}_1^m) + h(\mathbf{w}_{m+1}^n \mid \mathbf{w}_1^m) = h([D_n]_m^1 R_n \mathbf{u}_n^1 + [Q_n]_m^1 \mathbf{z}_n^1) + h(\mathbf{w}_{m+1}^n \mid \mathbf{w}_1^m), \tag{A36}$$

where (a) follows from the chain rule of differential entropy. It only remains to show that the limit of $(1/n)h(\mathbf{w}_{m+1}^n \mid \mathbf{w}_1^m)$ as $n \to \infty$ equals the entropy rate of u_1^{∞}. We will do this by deriving a lower and an upper bounds which converge to the same expression as $n \to \infty$.

A lower bound for $h(\mathbf{w}_{m+1}^n \mid \mathbf{w}_1^m)$ can be obtained by noticing that

$$h(\mathbf{w}_{m+1}^n \mid \mathbf{w}_1^m) = h(\mathbf{v}_{m+1}^n + \bar{\mathbf{z}}_{m+1}^n \mid \mathbf{v}_1^m + \bar{\mathbf{z}}_1^m) \tag{A37}$$

$$\overset{(a)}{\geq} h(\mathbf{v}_{m+1}^n + \bar{\mathbf{z}}_{m+1}^n \mid \mathbf{v}_1^m, \bar{\mathbf{z}}_1^n) \tag{A38}$$

$$\overset{(b)}{=} h(\mathbf{v}_{m+1}^n \mid \mathbf{v}_1^m, \bar{\mathbf{z}}_1^n) \tag{A39}$$

$$\overset{(c)}{=} h(\mathbf{v}_{m+1}^n \mid \mathbf{v}_1^m) \tag{A40}$$

$$\overset{(d)}{=} h(\mathbf{v}_1^n) - h(\mathbf{v}_1^m) \tag{A41}$$

$$\overset{(e)}{=} h(\mathrm{u}_1^n) - h(\mathbf{v}_1^m), \tag{A42}$$

where (a) follows from the fact that conditioning on more information does not increase differential entropy, (b) is due to the fact that $h(x + a) = h(x)$, for any constant a, (c) holds because $\bar{\mathbf{z}}_1^{\infty} \perp\!\!\!\perp \mathbf{v}_1^{\infty}$, (d) is a direct application of the chain rule of differential entropy, and (e) stems from (A34) and the fact that $\det(D_n R_n) = 1$. On the other hand,

$$h(\mathbf{v}_1^m) = h([D_n]_m^1 R_n \mathbf{u}_n^1) = \sum_{i=1}^{m} \log d_{n,i} + h([R_n]_m^1 \mathbf{u}_n^1). \tag{A43}$$

Then, by inserting (A43) and (A42) in (A37), dividing by n, and taking the limit $n \to \infty$, we obtain

$$\lim_{n\to\infty} \frac{1}{n} h(\mathrm{w}_{m+1}^n \mid \mathrm{w}_1^m) \geq \lim_{n\to\infty} \frac{1}{n}\left(h(\mathrm{u}_1^n) - \sum_{i=1}^m \log d_{n,i} - h([\boldsymbol{R}_n]_m^1 \mathbf{u}_n^1) \right) \tag{A44}$$

$$= \bar{h}(\mathrm{u}_1^\infty) - \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^m \log d_{n,i}, \tag{A45}$$

where the last equality is a consequence of the fact that u_1^∞ is entropy balanced (specifically, from Proposition A3).

We now derive an upper bound for $h(\mathrm{w}_{m+1}^n \mid \mathrm{w}_1^m)$. Defining the random vector

$$\mathbf{x}_n^{m+1} \triangleq [\boldsymbol{R}_n]_n^{m+1} \mathbf{u}_n^1,$$

and since $\boldsymbol{D}_n$ is diagonal, we can write

$$\mathbf{v}_m^1 = [\boldsymbol{D}_n]_n^{m+1} \boldsymbol{R}_n \mathbf{u}_n^1 = {}^{m+1}[\boldsymbol{D}_n]_n \mathbf{x}_n^{m+1}, \tag{A46}$$

where

$${}^{m+1}[\boldsymbol{D}_n]_n \triangleq \mathrm{diag}\{d_{n,m+1}, d_{n,m+2}, \ldots, d_{n,n}\}. \tag{A47}$$

Therefore,

$$h(\mathrm{w}_{m+1}^n \mid \mathrm{w}_1^m) \leq h(\mathbf{w}_n^{m+1}) = h\left({}^{m+1}[\boldsymbol{D}_n]_n \mathbf{x}_n^{m+1} + \bar{\mathbf{z}}_n^{m+1}\right) \tag{A48}$$

$$= \log \det\left({}^{m+1}[\boldsymbol{D}_n]_n\right) + h\left(\mathbf{x}_n^{m+1} + \left({}^{m+1}[\boldsymbol{D}_n]_n\right)^{-1} \bar{\mathbf{z}}_n^{m+1}\right). \tag{A49}$$

Notice that, by Assumption 2, $\bar{\mathbf{z}}_n^{m+1} = [\boldsymbol{Q}_n]_n^{m+1} \mathbf{z}_n^1 = [\boldsymbol{Q}_n]_n^{m+1} [\boldsymbol{\Phi}]_n^1 \mathbf{s}_\kappa^1$ and, thus, is restricted to the span of $[\boldsymbol{Q}_n]_n^{m+1} [\boldsymbol{\Phi}]_n^1$ of dimension $\kappa_n \leq \kappa$, for all $n \geq m + \kappa$. Then, for every $n > m + \kappa_n$, one can construct a unitary matrix $\boldsymbol{H}_n \triangleq (\boldsymbol{A}_n^T \mid \boldsymbol{B}_n^T)^T \in \mathbb{R}^{(n-m)\times(n-m)}$, with $\boldsymbol{A}_n \in \mathbb{R}^{\kappa\times(n-m)}$ and $\boldsymbol{B}_n \in \mathbb{R}^{(n-m-\kappa)\times(n-m)}$, such that the rows of $\boldsymbol{A}_n$ span the space spanned by the columns of $\left({}^{m+1}[\boldsymbol{D}_n]_n\right)^{-1} [\boldsymbol{Q}_n]_n^{m+1} [\boldsymbol{\Phi}]_n^1$ and such that $\boldsymbol{B}_n \left({}^{m+1}[\boldsymbol{D}_n]_n\right)^{-1} [\boldsymbol{Q}_n]_n^{m+1} [\boldsymbol{\Phi}]_n^1 = 0$. Therefore, from (A49),

$$h(\mathrm{w}_{m+1}^n \mid \mathrm{w}_1^m) \leq \log\det\left({}^{m+1}[\boldsymbol{D}_n]_n\right) + h\left(\boldsymbol{H}_n \mathbf{x}_n^{m+1} + \boldsymbol{H}_n \left({}^{m+1}[\boldsymbol{D}_n]_n\right)^{-1} \bar{\mathbf{z}}_n^{m+1}\right)$$

$$= \log\det\left({}^{m+1}[\boldsymbol{D}_n]_n\right) + h\left(\boldsymbol{B}_n \mathbf{x}_n^{m+1}\right) + h\left(\boldsymbol{A}_n \mathbf{x}_n^{m+1} + \boldsymbol{A}_n \left({}^{m+1}[\boldsymbol{D}_n]_n\right)^{-1} \bar{\mathbf{z}}_n^{m+1} \mid \boldsymbol{B}_n \mathbf{x}_n^{m+1}\right)$$

$$\leq \log\det\left({}^{m+1}[\boldsymbol{D}_n]_n\right) + h\left(\boldsymbol{B}_n \mathbf{x}_n^{m+1}\right) + h\left(\boldsymbol{A}_n \mathbf{x}_n^{m+1} + \boldsymbol{A}_n \left({}^{m+1}[\boldsymbol{D}_n]_n\right)^{-1} \bar{\mathbf{z}}_n^{m+1}\right)$$

$$\leq \log\det\left({}^{m+1}[\boldsymbol{D}_n]_n\right) + h\left(\boldsymbol{B}_n \mathbf{x}_n^{m+1}\right) + \frac{1}{2}\log\left((2\pi\mathrm{e})^\kappa \det\left(\boldsymbol{K}_{\boldsymbol{A}_n \mathbf{x}_n^{m+1}} + \boldsymbol{K}_{\boldsymbol{A}_n({}^{m+1}[\boldsymbol{D}_n]_n)^{-1} \bar{\mathbf{z}}_n^{m+1}} \right) \right)$$

$$\leq \log\det\left({}^{m+1}[\boldsymbol{D}_n]_n\right) + h\left(\boldsymbol{B}_n \mathbf{x}_n^{m+1}\right) + \frac{1}{2}\log\left((2\pi\mathrm{e})^\kappa \left[\lambda_{\max}(\boldsymbol{K}_{\mathbf{x}_n^{m+1}}) + \frac{\lambda_{\max}(\boldsymbol{K}_{\bar{\mathbf{z}}_n^{m+1}})}{\lambda_{\min}({}^{m+1}[\boldsymbol{D}_n]_n^2)} \right]^\kappa \right),$$

where $\boldsymbol{K}_{\boldsymbol{A}_n \mathbf{x}_n^{m+1}}$ and $\boldsymbol{K}_{\boldsymbol{A}_n({}^{m+1}[\boldsymbol{D}_n]_n)^{-1} \bar{\mathbf{z}}_n^{m+1}}$ are the covariance matrices of $\boldsymbol{A}_n \mathbf{x}_n^{m+1}$ and $\boldsymbol{A}_n \left({}^{m+1}[\boldsymbol{D}_n]_n\right)^{-1} \bar{\mathbf{z}}_n^{m+1}$, respectively, and where the last inequality follows from [31]. The fact that $\lambda_{\max}(\boldsymbol{K}_{\mathbf{x}_n^{m+1}})$ and $\lambda_{\max}(\boldsymbol{K}_{\bar{\mathbf{z}}_n^{m+1}})$ are upper bounded for all n, and the fact that $\lambda_{\min}({}^{m+1}[\boldsymbol{D}_n]_n)$ either grows with n or decreases sub-exponentially (from Lemma 7), imply that

$$\lim_{n\to\infty} \frac{1}{n} h(\mathrm{w}_{m+1}^n \mid \mathrm{w}_1^m) \leq \lim_{n\to\infty} \frac{1}{n} \log\det\left({}^{m+1}[\boldsymbol{D}_n]_n\right) + \lim_{n\to\infty} \frac{1}{n} h\left(\boldsymbol{B}_n \mathbf{x}_n^{m+1}\right). \tag{A50}$$

But the fact that $\det \boldsymbol{D}_n = 1$ implies that $\log\det\left({}^{m+1}[\boldsymbol{D}_n]_n\right) = -\sum_{i=1}^m \log d_{n,i}$. On the other hand, recalling that $\mathbf{x}_n^{m+1} = [\boldsymbol{R}_n]_n^{m+1} \mathbf{u}_n^1$ and noting that $\boldsymbol{B}_n [\boldsymbol{R}_n]_n^{m+1}$ has orthonormal

rows, reveals that $\lim_{n\to\infty} \frac{1}{n} h(\boldsymbol{B}_n \boldsymbol{x}_n^{m+1}) = \bar{h}(\mathrm{u}_1^\infty)$ (from the assumption that u_1^∞ is entropy balanced). Therefore,

$$\lim_{n\to\infty} \frac{1}{n} h(\mathrm{w}_{m+1}^n \mid \mathrm{w}_1^m) \leq \bar{h}(\mathrm{u}_1^\infty) - \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^m \log d_{n,i}, \tag{A51}$$

which coincides with the lower bound found in (A45), completing the proof. □

Proof of Lemma 7. The transfer function $G(z)$ can be factored as $G(z) = \tilde{G}(z)F(z)$, where $\tilde{G}(z)$ is stable and minimum phase and $F(z)$ is stable with all the non-minimum phase zeros of $G(z)$, both being biproper rational functions. From Lemma A2 (in Appendix C), in the limit as $n \to \infty$, the eigenvalues of $\tilde{\boldsymbol{G}}_n^T \tilde{\boldsymbol{G}}_n$ are lower and upper bounded by $\lambda_{\min}(\tilde{\boldsymbol{G}}^T \tilde{\boldsymbol{G}})$ and $\lambda_{\max}(\tilde{\boldsymbol{G}}^T \tilde{\boldsymbol{G}})$, respectively, where $0 < \lambda_{\min}(\tilde{\boldsymbol{G}}^T \tilde{\boldsymbol{G}}) \leq \lambda_{\max}(\tilde{\boldsymbol{G}}^T \tilde{\boldsymbol{G}}) < \infty$. Let $\tilde{\boldsymbol{G}}_n = \tilde{\boldsymbol{Q}}_n^T \tilde{\boldsymbol{D}}_n \tilde{\boldsymbol{R}}_n$ and $\boldsymbol{F}_n = \boldsymbol{Q}_n^T \boldsymbol{D}_n \boldsymbol{R}_n$ be the SVDs of $\tilde{\boldsymbol{G}}_n$ and $\boldsymbol{F}_n$, respectively, with $\tilde{d}_{n,1} \leq \tilde{d}_{n,2} \leq \cdots \leq \tilde{d}_{n,n}$ and $d_{n,1} \leq d_{n,2} \leq \cdots \leq d_{n,n}$ being the diagonal entries of the diagonal matrices $\tilde{\boldsymbol{D}}_n, \boldsymbol{D}_n$, respectively. Then,

$$\boldsymbol{G}_n^T \boldsymbol{G}_n = \boldsymbol{F}_n^T \tilde{\boldsymbol{G}}_n^T \tilde{\boldsymbol{G}}_n \boldsymbol{F}_n = (\tilde{\boldsymbol{D}}_n \tilde{\boldsymbol{R}}_n \boldsymbol{Q}_n^T \boldsymbol{D}_n \boldsymbol{R}_n)^T \tilde{\boldsymbol{D}}_n \tilde{\boldsymbol{R}}_n \boldsymbol{Q}_n^T \boldsymbol{D}_n \boldsymbol{R}_n. \tag{A52}$$

Denoting the i-th row of $\boldsymbol{R}_n$ by $\boldsymbol{r}_{n,i}^T$ be, we have that, from the Courant-Fischer theorem [30] that

$$\lambda_i(\boldsymbol{G}_n^T \boldsymbol{G}_n) \leq \max_{v \in \mathrm{span}\{\boldsymbol{r}_{n,k}\}_{k=1}^i : \|v\|=1} \|\boldsymbol{G}v\|^2 \tag{A53}$$

$$= \max_{v \in \mathrm{span}\{\boldsymbol{r}_{n,k}\}_{k=1}^i : \|v\|=1} \|\tilde{\boldsymbol{D}}_n \tilde{\boldsymbol{R}}_n^T \boldsymbol{Q}_n^T \boldsymbol{D}_n \boldsymbol{R}_n v\|^2 \tag{A54}$$

$$\leq d_{n,i}^2 \tilde{d}_{n,n}^2. \tag{A55}$$

Likewise,

$$\lambda_i(\boldsymbol{G}_n^T \boldsymbol{G}_n) \geq \min_{v \in \mathrm{span}\{\boldsymbol{r}_{n,k}\}_{k=i}^n : \|v\|=1} \|\boldsymbol{G}v\| \tag{A56}$$

$$= \min_{v \in \mathrm{span}\{\boldsymbol{r}_{n,k}\}_{k=i}^n : \|v\|=1} \|\tilde{\boldsymbol{D}}_n \tilde{\boldsymbol{R}}_n^T \boldsymbol{Q}_n^T \boldsymbol{D}_n \boldsymbol{R}_n v\|^2 \tag{A57}$$

$$\geq d_{n,i}^2 \tilde{d}_{n,1}^2. \tag{A58}$$

Thus,

$$\lim_{n\to\infty} \frac{\lambda_i(\boldsymbol{G}_n^T \boldsymbol{G}_n)}{d_{n,i}^2} \in \left(\lambda_{\min}(\tilde{\boldsymbol{G}}^T \tilde{\boldsymbol{G}}), \lambda_{\max}(\tilde{\boldsymbol{G}}^T \tilde{\boldsymbol{G}}) \right). \tag{A59}$$

The result now follows directly from Lemma A3 (in Appendix C). □

Proof of Theorem 6. To begin with, the entropy power inequality [1] gives $h(\mathbf{y}_n^1) = h(\boldsymbol{G}_n \mathbf{u}_n^1 + \mathbf{z}_n^1) \geq h(\boldsymbol{G}_n \mathbf{u}_n^1) = h(\mathbf{y}_n^1)$, proving the lower bound in (70).

To obtain the other bounds on the entropy gain of $\boldsymbol{G}_n$, we will use Lemma 6. Recalling the structure of z_1^∞ specified in Assumption 2, the random vector whose differential entropy appears on the RHS of (64) takes the form

$$[\boldsymbol{D}_n]_m^1 \boldsymbol{R}_n \mathbf{u}_n^1 + [\boldsymbol{Q}_n]_m^1 \mathbf{z}_n^1 = [\boldsymbol{D}_n]_m^1 \boldsymbol{R}_n \mathbf{u}_n^1 + [\boldsymbol{Q}_n]_m^1 [\boldsymbol{\Phi}]_n^1 \mathbf{s}_\kappa^1. \tag{A60}$$

Notice that, for every $n \geq \kappa$, the columns of the matrix $[\boldsymbol{Q}_n]_m^1 [\boldsymbol{\Phi}]_n^1 \in \mathbb{R}^{m \times \kappa}$ span a space of dimension $\kappa_n \in \{0, 1, \ldots, \bar{\kappa}\}$, with $\bar{\kappa} \triangleq \min\{m, \kappa\}$. If $\kappa_n = 0$ (i.e., $[\boldsymbol{Q}_n]_m^1 [\boldsymbol{\Phi}]_n^1 = \boldsymbol{0}$), then

$$h([\boldsymbol{D}_n]_m^1 \boldsymbol{R}_n \mathbf{u}_n^1 + [\boldsymbol{Q}_n]_m^1 \mathbf{z}_n^1) = h([\boldsymbol{D}_n]_m^1 \boldsymbol{R}_n \mathbf{u}_n^1). \tag{A61}$$

If that is that case for every $n \geq \kappa$, the lower bound in (70) is reached by inserting the latter expression into (64) and invoking Lemma 7.

Let $[\boldsymbol{Q}_n]^1_m[\boldsymbol{\Phi}]^1_n = \boldsymbol{A}_n^T \boldsymbol{T}_n \boldsymbol{B}_n$ be an SVD for $[\boldsymbol{Q}_n]^1_m[\boldsymbol{\Phi}]^1_n$, where $\boldsymbol{A}_n \in \mathbb{R}^{\kappa_n \times m}$ has orthonormal rows,

$$\boldsymbol{T}_n = \mathrm{diag}\{t_1(n), t_2(n), \dots, t_{\kappa_n}(n)\},$$

where $0 < t_1(n) \leq t_2(n) \leq \cdots \leq t_{\kappa_n}(n) \leq 1$ are the singular values of $[\boldsymbol{Q}_n]^1_m[\boldsymbol{\Phi}]^1_n$, and $\boldsymbol{B}_n \in \mathbb{R}^{\kappa_n \times \kappa}$ has orthonormal rows. Construct a unitary matrix $\boldsymbol{H}_n \in \mathbb{R}^{m \times m}$ such that

$$\boldsymbol{H}_n \triangleq \begin{pmatrix} \boldsymbol{A}_n \\ \overline{\boldsymbol{A}}_n \end{pmatrix}, \tag{A62}$$

where $\boldsymbol{A}_n \in \mathbb{R}^{\kappa_n \times m}$ is as before, and $\overline{\boldsymbol{A}}_n \in \mathbb{R}^{(m-\kappa_n) \times m}$ has orthonormal rows, and its row span is the orthogonal complement of that of $\boldsymbol{A}_n$. Thus,

$$\boldsymbol{H}_n[\boldsymbol{Q}_n]^1_m[\boldsymbol{\Phi}]^1_n = \begin{pmatrix} \boldsymbol{A}_n[\boldsymbol{Q}_n]^1_m[\boldsymbol{\Phi}]^1_n \\ \boldsymbol{0}_{(m-\kappa_n) \times \kappa} \end{pmatrix}, \quad n \geq N. \tag{A63}$$

From (A63) and (A60), we obtain

$$h\Big([\boldsymbol{D}_n]^1_m \boldsymbol{R}_n \mathbf{u}^1_n + [\boldsymbol{Q}_n]^1_m \mathbf{z}^1_n\Big) = h\Big([\boldsymbol{D}_n]^1_m \boldsymbol{R}_n \mathbf{u}^1_n + [\boldsymbol{Q}_n]^1_m[\boldsymbol{\Phi}]^1_n \mathbf{s}^1_\kappa\Big) \tag{A64}$$

$$= h\Big(\boldsymbol{H}_n([\boldsymbol{D}_n]^1_m \boldsymbol{R}_n \mathbf{u}^1_n + [\boldsymbol{Q}_n]^1_m[\boldsymbol{\Phi}]^1_n \mathbf{s}^1_\kappa)\Big) \tag{A65}$$

$$= h\Big(\boldsymbol{A}_n[\boldsymbol{D}_n]^1_m \boldsymbol{R}_n \mathbf{u}^1_n + \boldsymbol{A}_n[\boldsymbol{Q}_n]^1_m[\boldsymbol{\Phi}]^1_n \mathbf{s}^1_\kappa \,\Big|\, \overline{\boldsymbol{A}}_n[\boldsymbol{D}_n]^1_m \boldsymbol{R}_n \mathbf{u}^1_n\Big) + (1 - 1_{\{m\}}(\kappa_n))h(\overline{\boldsymbol{A}}_n[\boldsymbol{D}_n]^1_m \boldsymbol{R}_n \mathbf{u}^1_n), \tag{A66}$$

where the indicator function $1_{\{m\}}(\kappa_n) = 1$ if $\kappa_n = m$ and 0 otherwise. The first differential entropy on the RHS of (A66) can be lower bounded as

$$h\Big(\boldsymbol{A}_n[\boldsymbol{D}_n]^1_m \boldsymbol{R}_n \mathbf{u}^1_n + \boldsymbol{A}_n[\boldsymbol{Q}_n]^1_m[\boldsymbol{\Phi}]^1_n \mathbf{s}^1_\kappa \,\Big|\, \overline{\boldsymbol{A}}_n[\boldsymbol{D}_n]^1_m \boldsymbol{R}_n \mathbf{u}^1_n\Big) \overset{(a)}{\geq} h\Big(\boldsymbol{A}_n[\boldsymbol{Q}_n]^1_m[\boldsymbol{\Phi}]^1_n \mathbf{s}^1_\kappa \,\Big|\, \overline{\boldsymbol{A}}_n[\boldsymbol{D}_n]^1_m \boldsymbol{R}_n \mathbf{u}^1_n\Big)$$

$$\overset{(b)}{=} h\Big(\boldsymbol{A}_n[\boldsymbol{Q}_n]^1_m[\boldsymbol{\Phi}]^1_n \mathbf{s}^1_\kappa\Big) = h\Big(\boldsymbol{T}_n \boldsymbol{B}_n \mathbf{s}^1_\kappa\Big) \overset{(c)}{\geq} \kappa_n \log(t_1(n)) + h(\mathbf{s}^1_\kappa) - \frac{\kappa - \kappa_n}{2}\log(\lambda_{\max}(\boldsymbol{K}_{\mathbf{s}^1_\kappa})), \tag{A67}$$

where (a) is from the entropy power inequality [1], (b) holds because $\mathbf{s}^1_\kappa \perp\!\!\!\perp \mathbf{u}^1_n$ and (c) is from Proposition A1. An upper bound can be obtained as

$$h\Big(\boldsymbol{A}_n[\boldsymbol{D}_n]^1_m \boldsymbol{R}_n \mathbf{u}^1_n + \boldsymbol{A}_n[\boldsymbol{Q}_n]^1_m[\boldsymbol{\Phi}]^1_n \mathbf{s}^1_\kappa \,\Big|\, \overline{\boldsymbol{A}}_n[\boldsymbol{D}_n]^1_m \boldsymbol{R}_n \mathbf{u}^1_n\Big) \overset{(a)}{\leq} h\Big(\boldsymbol{A}_n[\boldsymbol{D}_n]^1_m \boldsymbol{R}_n \mathbf{u}^1_n + \boldsymbol{T}_n \boldsymbol{B}_n \mathbf{s}^1_\kappa\Big)$$

$$\overset{(b)}{\leq} \frac{1}{2}\log\Big((2\pi e)^m \det(\boldsymbol{K}_{\boldsymbol{A}_n[\boldsymbol{D}_n]^1_m \boldsymbol{R}_n \mathbf{u}^1_n} + \boldsymbol{K}_{\boldsymbol{T}_n \boldsymbol{B}_n \mathbf{s}^1_\kappa})\Big)$$

$$\overset{(c)}{\leq} \frac{\kappa_n}{2}\log\Big((2\pi e)\big((d_m)^2 \lambda_{\max}(\boldsymbol{K}_{\mathbf{u}^1_n}) + (t_{\kappa_n}(n))^2 \lambda_{\max}(\boldsymbol{K}_{\mathbf{s}^1_\kappa})\big)\Big), \tag{A68}$$

where (a) holds because conditioning does not increase entropy, (b) is because a Gaussian distribution maximizes the differential entropy for a given covariance matrix, and (c) is due to Reference [31]. Notice that $\mathbf{u}^\infty_1$ satisfies the requirements of Proposition A2, implying that $\lim_{n \to \infty} n^{-1} \lambda_{\max}(\boldsymbol{K}_{\mathbf{u}^1_n}) = 0$. Thus, since $t_{\kappa_n}(n) \leq 1$, it follows from (A67), (A68), and (A66) that

$$\lim_{n \to \infty} \frac{1 - 1_{\{m\}}(\kappa_n)}{n} h(\overline{\boldsymbol{A}}_n[\boldsymbol{D}_n]^1_m \boldsymbol{R}_n \mathbf{u}^1_n) + \lim_{n \to \infty} \frac{\kappa_n}{n}\log(t_1(n))$$

$$\leq \lim_{n \to \infty} \frac{1}{n} h\Big([\boldsymbol{D}_n]^1_m \boldsymbol{R}_n \mathbf{u}^1_n + [\boldsymbol{Q}_n]^1_m \mathbf{z}^1_n\Big) \leq \lim_{n \to \infty} (1 - 1_{\{m\}}(\kappa_n))h(\overline{\boldsymbol{A}}_n[\boldsymbol{D}_n]^1_m \boldsymbol{R}_n \mathbf{u}^1_n). \tag{A69}$$

For the last differential entropy on the RHS of (A66), notice that $[D_n]^1_m R_n = {}^1[D_n]_m [R_n]^1_m$. Consider the SVD $\overline{A}_n {}^1[D_n]_m [R_n]^1_m = V_n^T \Sigma_n W_n$, with $V_n \in \mathbb{R}^{(m-\kappa_n)\times(m-\kappa_n)}$ being unitary, $\Sigma_n \in \mathbb{R}^{(m-\kappa_n)\times(m-\kappa_n)}$ being diagonal, and $W_n \in \mathbb{R}^{(m-\kappa_n)\times n}$ having orthonormal rows. We can then conclude that

$$h(\overline{A}_n [D_n]^1_m R_n \mathbf{u}^1_n) = h(\Sigma_n W_n \mathbf{u}^1_n) = \log|\det(\Sigma_n)| + h(W_n \mathbf{u}^1_n). \tag{A70}$$

Now, the fact that

$$\overline{A}_n {}^1[D_n]_m {}^1[D_n]_m \overline{A}_n^T = \overline{A}_n {}^1[D_n]_m [R_n]^1_m (\overline{A}_n {}^1[D_n]_m [R_n]^1_m)^T = V^T \Sigma W W^T \Sigma^T V = V^T \Sigma \Sigma^T V$$

reveals that

$$\log|\det \Sigma| = \frac{1}{2}\log|\det(\overline{A}_n ({}^1[D_n]_m)^2 \overline{A}_n^T)|. \tag{A71}$$

Recalling that $\overline{A}_n = [H_n]^{\kappa_n+1}_m$ and that $H_n \in \mathbb{R}^{m \times m}$ is unitary, it is easy to show (by using the Courant-Fischer theorem [30]) that

$$\sum_{i=1}^{m-\kappa_n} \log d_{n,i} \overset{(a)}{\leq} \frac{1}{2}\log\left|\det(\overline{A}_n ({}^1[D_n]_m)^2 \overline{A}_n^T)\right| \overset{(b)}{\leq} \sum_{i=\kappa_n+1}^{m} \log d_{n,i}, \tag{A72}$$

with equality in (a) and (b) if and only if $\overline{A}_n = [I_{m-\kappa_n} \mid 0]$ and $\overline{A}_n = [0 \mid I_{m-\kappa_n}]$, respectively. Substituting this into (A71) and then the latter into (A70), we arrive to

$$h([W_n]^1_m \mathbf{u}^1_n) + \sum_{i=1}^{m-\kappa_n} \log d_{n,i} \leq h(\overline{A}_n [D_n]^1_m R_n \mathbf{u}^1_n) \leq h([W_n]^1_m \mathbf{u}^1_n) + \sum_{i=\kappa_n+1}^{m} \log d_{n,i}. \tag{A73}$$

Substituting the upper bound from this equation and from (A68) into (A66) and the latter in (64), exploiting the fact that $\mathbf{u}^\infty_1$ is entropy balanced (which ensures that u^∞_1 satisfies Condition i) in Definition 2) and invoking Lemma 7 yields the upper bound in (70).

Doing the same substitutions but with the lower bounds in (A73) and (A67), and using the assumption that $\lim_{n\to\infty} \frac{1}{n}\log(t_1(n)) = 0$, gives the lower bound of (71). This completes the proof. □

Proof of Lemma 8. We will consider first the case $\kappa = m$ and show that $\lim_{n\to\infty} \sigma_{\min}({}^1[Q_n]_m) > 0$, where now Q_n^T is the left unitary matrix in the SVD $F_n = Q_n^T D_n R_n$. We will prove that this is the case by using a contradiction argument. Thus, suppose the contrary, i.e., that

$$\lim_{n\to\infty} \sigma_{\min}({}^1[Q_n]_m) = 0. \tag{A74}$$

Then, there exists a sequence of unit-norm vectors $\{v_n\}^\infty_{n=1}$, with $v_n \in \mathbb{R}^m$ for all n, such that

$$\lim_{n\to\infty} \|v_n^T {}^1[Q_n]_m\| = 0. \tag{A75}$$

For each $n \in \mathbb{N}$, define the n-length unit-norm image vectors $t_n^T \triangleq v_n^T [Q_n]^1_m$. Then,

$$\|F_n^T t_n\| = \|R_n^T D_n Q_n t_n\| = \|D_n Q_n t_n\| = \|{}^1[D_n]_m v_n\|, \tag{A76}$$

where the last equality follows from the fact that, by construction, t_n^T is in the span of the first m rows of Q_n, together with the fact that Q_n is unitary (which implies that $[Q_n]^{m+1}_n t_n = 0$). Since the top m entries in D_n decay exponentially as n increases, we have that

$$\|F_n^T t_n\| \leq \mathcal{O}(\zeta_n |\rho_M|^{-n}), \tag{A77}$$

where ζ_n is a finite-order polynomial of n (from Lemma A3, in Appendix C).

Now, notice that $[\boldsymbol{F}_n]_n^{m+1}([\boldsymbol{F}_n]_n^{m+1})^T$ is a Toeplitz matrix with the convolution of f and f^- (the impulse response of F and its time-reversed version, respectively) on its first row and column. It then follows from Reference [18], Lemma 4.1, that

$$\lim_{n\to\infty} \lambda_{\min}([\boldsymbol{F}_n]_n^{m+1}([\boldsymbol{F}_n]_n^{m+1})^T) = \min_{\omega:\omega\in[-\pi,\pi]} |F(e^{j\omega})|^2 > 0 \tag{A78}$$

(the inequality is strict because all the zeros of $F(z)$ are strictly outside the unit disk). Then, we conclude that

$$\lim_{n\to\infty} \sigma_{\min}([\boldsymbol{F}_n]_n^{m+1}) > 0. \tag{A79}$$

Recall that $\|\boldsymbol{t}_n\| = 1$; thus, from (A75), $\lim_{n\to\infty} \|[\boldsymbol{t}_n]_m^1\| = 0$ and $\lim_{n\to\infty} \|[\boldsymbol{t}_n]_n^{m+1}\| = 1$, which means that $\lim_{n\to\infty} \|\boldsymbol{F}_n^T \boldsymbol{t}_n\| = \lim_{n\to\infty} \|([\boldsymbol{F}_n]_n^{m+1})^T [\boldsymbol{t}_n]_n^{m+1}\| \geq \lim_{n\to\infty} \sigma_{\min}([\boldsymbol{F}_n]_n^{m+1}) \overset{(A79)}{>} 0$, which contradicts (A77). Therefore,

$$\lim_{n\to\infty} \sigma_{\min}(^1[\boldsymbol{Q}_n]_m) > 0. \tag{A80}$$

Now, consider an arbitrary $\kappa \geq 1$. Since

$$\sigma_{\min}\left(\overbrace{[\boldsymbol{Q}_n]_m^1}^{\overset{1\quad\kappa}{\overbrace{}}}\right) \geq \sigma_{\min}\left(^1[\boldsymbol{Q}_n]_m\right), \tag{A81}$$

it follows from (A80) that

$$\lim_{n\to\infty} \sigma_{\min}\left([\boldsymbol{Q}_n]_m^1[\boldsymbol{\Phi}]_n^1\right) = \lim_{n\to\infty} \sigma_{\min}\left(\overbrace{[\boldsymbol{Q}_n]_m^1}^{\overset{1\quad\kappa}{\overbrace{}}}\right) > 0; \tag{A82}$$

thus, $\lim_{n\to\infty} \kappa_n = \bar{\kappa}$. This completes the proof. $\square$

Proof of Theorem 9. Denote the Blaschke product [11] of $A(z)$ as

$$B(z) \triangleq \frac{\prod_{i=1}^m (z - p_i)}{\prod_{i=1}^m p_i^*(z - 1/p_i^*)}, \tag{A83}$$

which clearly satisfies

$$|B(e^{j\omega})| = 1, \quad \forall \omega \in [-\pi, \pi], \tag{A84}$$

$$b_0 \triangleq \lim_{|z|\to\infty} B(z) = \frac{1}{\prod_{i=1}^m p_i^*}, \tag{A85}$$

where b_0 is the first sample in the impulse response of $B(z)$. Notice that (A84) implies that $\lim_{n\to\infty} \frac{1}{n} E[\|\boldsymbol{B}_n \mathbf{u}_n^1\|^2] = \lim_{n\to\infty} \frac{1}{n} E[\|\mathbf{u}_n^1\|^2]$ for every sequence of random variables $\mathbf{u}_1^\infty$ with uniformly bounded variance. Since $B(z)$ has only stable poles and its zeros coincide exactly with the poles of $A(z)$, it follows that $B(z)A(z)$ is an MP stable transfer function. Thus, the asymptotically stationary process $\tilde{\mathbf{x}}_1^\infty$ defined in (139) can be constructed as

$$\tilde{\mathbf{x}}_n^1 \triangleq \boldsymbol{B}_n \mathbf{x}_n^1, \tag{A86}$$

where $\boldsymbol{B}_n$ is a Toeplitz lower triangular matrix with its main diagonal entries equal to b_0. Since $\mathbf{w}_1^\infty$ is entropy balanced, so is $\tilde{\mathbf{x}}_1^\infty$, thanks to Lemma 4.

The fact that $B(z)$ is biproper with b_0 as in (A85) implies that, for any $\mathbf{u}_n^1$ with finite differential entropy,

$$h(B_n\mathbf{u}_n^1) = h(\mathbf{u}_n^1) - n\underbrace{\sum_{i=1}^m \log|p_i|}_{\triangleq \mathcal{G}}, \tag{A87}$$

which will be utilized next.

For any given $n \geq m$, suppose that $C(z)$ is chosen and $\mathbf{x}_n^1$ and $\mathbf{u}_n^1$ are distributed so as to minimize $I(\mathbf{x}_n^1; C_n\mathbf{x}_n^1 + \mathbf{u}_n^1)$ subject to the constraint $\mathrm{E}[\|\mathbf{y}_n^1 - \mathbf{x}_n^1\|^2] = \mathrm{E}[\|(C_n - I)\mathbf{x}_n^1\|^2] + \mathrm{E}[\|\mathbf{u}_n^1\|]^2 \leq D$ (i.e., $\mathbf{x}_n^1, \mathbf{u}_n^1$ is a realization of $R_{\mathrm{x},n}(D)$), yielding the reconstruction

$$\mathbf{y}_n^1 = C_n\mathbf{x}_n^1 + \mathbf{u}_n^1. \tag{A88}$$

Since we are considering mean-squared error distortion, it follows that, for rate-distortion optimality, $\mathbf{u}_n^1$ must be jointly Gaussian with $\mathbf{x}_n^1$. In addition, there is no loss of rate-distortion optimality if u_1^∞ is entropy balanced (otherwise, it would have a lower entropy rate than its entropy-balanced counterpart, which differs from the former only on a finite number of samples and has the same asymptotic MSE). From these vectors, define

$$\tilde{\mathbf{u}}_n^1 \triangleq B_n\mathbf{u}_n^1, \tag{A89}$$

$$\tilde{\mathbf{y}}_n^1 \triangleq B_n\mathbf{y}_n^1 = B_nC_n(B_n)^{-1}\tilde{\mathbf{x}}_n^1 + \tilde{\mathbf{u}}_n^1, \tag{A90}$$

$$\bar{\mathbf{y}}_n^1 \triangleq \tilde{\mathbf{y}}_n^1 + \mathbf{d}_n^1 = B_nC_n(B_n)^{-1}\tilde{\mathbf{x}}_n^1 + \tilde{\mathbf{u}}_n^1 + \mathbf{d}_n^1, \tag{A91}$$

where $\mathbf{d}_n^1$ is a zero-mean Gaussian vector independent of $(\tilde{\mathbf{u}}_n^1, \tilde{\mathbf{x}}_n^1)$ with finite differential entropy and finite variance such that $d_k = 0, \forall k > m$. Then, we have that (the change of variables and the steps in this chain of equations is represented by the block diagrams shown in Figure 7)

$$nR_{\mathrm{x},n}(D) = I(\mathbf{x}_n^1; \mathbf{y}_n^1) \overset{(a)}{=} I(B_n\mathbf{x}_n^1; B_n\mathbf{y}_n^1) = I(\tilde{\mathbf{x}}_n^1; \tilde{\mathbf{y}}_n^1) \tag{A92}$$

$$= h(\tilde{\mathbf{y}}_n^1) - h(\tilde{\mathbf{y}}_n^1|\tilde{\mathbf{x}}_n^1) \tag{A93}$$

$$\overset{(b)}{=} h(\tilde{\mathbf{y}}_n^1) - h(\tilde{\mathbf{u}}_n^1|\tilde{\mathbf{x}}_n^1) \tag{A94}$$

$$\overset{(c)}{=} h(\tilde{\mathbf{y}}_n^1) - h(\tilde{\mathbf{u}}_n^1) \tag{A95}$$

$$\overset{(d)}{=} h(\tilde{\mathbf{y}}_n^1) - \left(h(\tilde{\mathbf{u}}_n^1 + \mathbf{d}_n^1) + [h(\mathbf{u}_n^1) - h(\tilde{\mathbf{u}}_n^1 + \mathbf{d}_n^1)] - n\mathcal{G}\right) \tag{A96}$$

$$\overset{(e)}{=} h(\tilde{\mathbf{y}}_n^1) - h(\tilde{\mathbf{u}}_n^1 + \mathbf{d}_n^1|\tilde{\mathbf{x}}_n^1) + n\mathcal{G} - [h(\mathbf{u}_n^1) - h(\tilde{\mathbf{u}}_n^1 + \mathbf{d}_n^1)] \tag{A97}$$

$$\overset{(f)}{=} h(\tilde{\mathbf{y}}_n^1) - h(\bar{\mathbf{y}}_n^1|\tilde{\mathbf{x}}_n^1) + n\mathcal{G} - [h(\mathbf{u}_n^1) - h(\tilde{\mathbf{u}}_n^1 + \mathbf{d}_n^1)] \tag{A98}$$

$$= h(\tilde{\mathbf{y}}_n^1) - h(\bar{\mathbf{y}}_n^1) + I(\tilde{\mathbf{x}}_n^1; \bar{\mathbf{y}}_n^1) + n\mathcal{G} - [h(\mathbf{u}_n^1) - h(\tilde{\mathbf{u}}_n^1 + \mathbf{d}_n^1)] \tag{A99}$$

$$= I(\tilde{\mathbf{x}}_n^1; \bar{\mathbf{y}}_n^1) + n\mathcal{G} - [h(\mathbf{u}_n^1) - h(\tilde{\mathbf{u}}_n^1 + \mathbf{d}_n^1)] + [h(\tilde{\mathbf{y}}_n^1) - h(\tilde{\mathbf{y}}_n^1 + \mathbf{d}_n^1)], \tag{A100}$$

where (a) follows from B_n being invertible, (b) is due to the fact that $\tilde{\mathbf{y}}_n^1 = P_n\tilde{\mathbf{x}}_n^1 + \tilde{\mathbf{u}}_n^1$, (c) holds because $\mathbf{u}_n^1 \perp\!\!\!\perp \mathbf{x}_n^1$. The equality (d) stems from $h(\tilde{\mathbf{u}}_n^1) = h(\mathbf{u}_n^1) - n\mathcal{G}$ (see (A87)). Equality holds in (e) because $\tilde{\mathbf{x}}_n^1 \perp\!\!\!\perp (\tilde{\mathbf{u}}_n^1, \mathbf{d}_n^1)$ and in (f) because of (A91). But from Theorem 4 and since u_1^∞ is entropy balanced, $\lim_{n\to\infty} \frac{1}{n}(h(\tilde{\mathbf{u}}_n^1 + \mathbf{d}_n^1) - h(\mathbf{u}_n^1)) = 0$. From Lemma 3 and because u_1^∞ is entropy balanced, so is $\tilde{y}_1^\infty$. This guarantees, from Lemma 5, that $\lim_{n\to\infty} n^{-1}[h(\tilde{\mathbf{y}}_n^1) - h(\tilde{\mathbf{y}}_n^1 + \mathbf{d}_n^1)] = 0$. Thus, $R_{\mathrm{x},n}(D) = \lim_{n\to\infty} \frac{1}{n}(\tilde{\mathbf{x}}_n^1; \bar{\mathbf{y}}_n^1) + \mathcal{G} \geq R_{\tilde{\mathbf{x}},n}(D) + \mathcal{G}$.

At the same time, the distortion for the source $\tilde{\mathbf{x}}_n^1$ when reconstructed as $\bar{\mathbf{y}}_n^1$ is

$$\lim_{n\to\infty} \frac{1}{n}\mathrm{E}\left[\|\bar{\mathbf{y}}_n^1 - \tilde{\mathbf{x}}_n^1\|^2\right] = \lim_{n\to\infty} \frac{1}{n}\left(\mathrm{E}\left[\|\tilde{\mathbf{y}} - \tilde{\mathbf{x}}_n^1\|^2\right] + \mathrm{E}\left[\|\mathbf{d}_n^1\|^2\right]\right) \overset{(a)}{=} \lim_{n\to\infty} \frac{1}{n}\mathrm{E}\left[\|\tilde{\mathbf{y}} - \tilde{\mathbf{x}}_n^1\|^2\right] \tag{A101}$$

$$= \lim_{n\to\infty} \frac{1}{n}\mathrm{E}\left[\|B_n(\mathbf{y}_n^1 - \mathbf{x}_n^1)\|^2\right] \overset{(b)}{=} \lim_{n\to\infty} \frac{1}{n}\mathrm{E}\left[\|\mathbf{y}_n^1 - \mathbf{x}_n^1\|^2\right], \tag{A102}$$

where (a) holds because $\|\mathbf{d}_n^1\| = \|\mathbf{d}_m^1\|$ is bounded, and (b) is due to the fact that, in the limit, $B(z)$ is a unitary operator. Recalling the definitions of $R_{\tilde{x}}(D)$ and $R_{\tilde{x}}(D)$, we conclude that $\lim_{n\to\infty} \frac{1}{n}(\tilde{\mathbf{x}}_n^1; \bar{\mathbf{y}}_n^1) \geq R_{\tilde{x},n}(D)$; therefore,

$$R_{\mathrm{x}}(D) - R_{\tilde{x}}(D) \geq \sum_{i=1}^m \log|p_i|. \tag{A103}$$

In order to complete the proof, it suffices to show that $R_{\mathrm{x}}(D) - R_{\tilde{x}}(D) \leq \sum_{i=1}^m \log|p_i|$. For this purpose, consider now the (asymptotically) stationary source $\tilde{\mathbf{x}}_n^1$, and suppose that $\hat{\mathbf{y}}_n^1 = \tilde{\mathbf{x}}_n^1 + \mathbf{u}_n^1$ realizes $R_{\tilde{x},n}(D)$. Again, $\tilde{\mathbf{x}}_n^1$ and $\mathbf{u}_n^1$ will be jointly Gaussian, satisfying $\hat{\mathbf{y}}_n^1 \perp\!\!\!\perp \mathbf{u}_n^1$ (the latter condition is required for minimum MSE optimality). From this, one can propose an alternative realization in which the error sequence is $\tilde{\mathbf{u}} \triangleq B_n \mathbf{u}_n^1$, yielding an output $\tilde{\mathbf{y}}_n^1 = \tilde{\mathbf{x}}_n^1 + \tilde{\mathbf{u}}_n^1$ with $\tilde{\mathbf{y}}_n^1 \perp\!\!\!\perp \tilde{\mathbf{u}}_n^1$. Then,

$$nR_{\tilde{x},n}(D) = I(\tilde{\mathbf{x}}_n^1; \hat{\mathbf{y}}_n^1) = h(\tilde{\mathbf{x}}_n^1) - h(\tilde{\mathbf{x}}_n^1|\hat{\mathbf{y}}_n^1) \tag{A104}$$

$$\overset{(a)}{=} h(\tilde{\mathbf{x}}_n^1) - h(\mathbf{u}_n^1) \tag{A105}$$

$$\overset{(b)}{=} h(\tilde{\mathbf{x}}_n^1) - h(\tilde{\mathbf{u}}_n^1) - n\mathcal{G} \tag{A106}$$

$$\overset{(c)}{=} h(\tilde{\mathbf{x}}_n^1) - h(\tilde{\mathbf{u}}_n^1|\tilde{\mathbf{y}}_n^1) - n\mathcal{G} \tag{A107}$$

$$\overset{(d)}{=} h(\tilde{\mathbf{x}}_n^1) - h(\tilde{\mathbf{x}}_n^1|\tilde{\mathbf{y}}_n^1) - n\mathcal{G} \tag{A108}$$

$$= I(\tilde{\mathbf{x}}_n^1; \tilde{\mathbf{y}}_n^1) - n\mathcal{G} \tag{A109}$$

$$= I(B_n \mathbf{x}_n^1; B_n \mathbf{y}_n^1) - n\mathcal{G} \tag{A110}$$

$$\overset{(e)}{=} I(\mathbf{x}_n^1; \mathbf{y}_n^1) - n\mathcal{G}, \tag{A111}$$

where (a) follows by recalling that $\hat{\mathbf{y}}_n^1 = \tilde{\mathbf{x}}_n^1 + \mathbf{u}_n^1$ and because $\hat{\mathbf{y}}_n^1 \perp\!\!\!\perp \mathbf{u}_n^1$, (b) stems from (A87), (c) is a consequence of $\tilde{\mathbf{y}}_n^1 \perp\!\!\!\perp \tilde{\mathbf{u}}_n^1$, (d) follows from the fact that $\tilde{\mathbf{y}}_n^1 = \tilde{\mathbf{x}}_n^1 + \tilde{\mathbf{u}}_n^1$. Finally, (e) holds because B_n is invertible for all n. Since, asymptotically as $n \to \infty$, the distortion yielded by $\mathbf{y}_n^1$ for the non-stationary source $\mathbf{x}_n^1$ is the same which is obtained when $\tilde{\mathbf{x}}_n^1$ is reconstructed as $\hat{\mathbf{y}}_n^1$ (recall (A84)), we conclude that $R_{\mathrm{x}}(D) - R_{\tilde{x}}(D) \leq \sum_{i=1}^M \log|p_i|$, completing the proof. $\square$

Appendix C. Technical Lemmas and Propositions

Proposition A1. *Let the random vector* $\mathbf{s}_\kappa^1$ *have finite differential entropy, and suppose its covariance matrix* $K_{\mathbf{s}_\kappa^1}$ *satisfies* $\lambda_{max}(K_{\mathbf{s}_\kappa^1}) < \infty$. *Then, for any unitary matrix* $A \in \mathbb{R}^{\kappa\times\kappa}$ *and* $i = 1, 2, \ldots, \kappa$

$$h(\mathbf{s}_\kappa^1) - \frac{\kappa - i}{2}\log(2\pi e\, \lambda_{max}(K_{\mathbf{s}_\kappa^1})) \leq h([A]_i^1 \mathbf{s}_\kappa^1) \leq \frac{i}{2}\log(2\pi e\, \lambda_{max}(K_{\mathbf{s}_\kappa^1})). \tag{A112}$$

Proof. Define $\mathbf{r}_\kappa^1 \triangleq A\mathbf{s}_\kappa^1$. Since A is unitary, it follows that $h(\mathbf{r}_\kappa^1) = h(\mathbf{s}_\kappa^1)$ and that $K_{\mathbf{r}_\kappa^1}$ and $K_{\mathbf{s}_\kappa^1}$ have the same eigenvalues. Therefore,

$$h([A]_i^1 \mathbf{s}_\kappa^1) = h(\mathbf{r}_i^1) \overset{(a)}{\leq} \frac{1}{2}\log((2\pi e)^i \det(K_{\mathbf{r}_i^1})) \overset{(b)}{\leq} \frac{i}{2}\log(2\pi e\, \lambda_{\max}(K_{\mathbf{r}_i^1})) \overset{(c)}{\leq} \frac{i}{2}\log(2\pi e\, \lambda_{\max}(K_{\mathbf{r}_\kappa^1}))$$

$$= \frac{i}{2}\log(2\pi e\, \lambda_{\max}(K_{\mathbf{s}_\kappa^1})), \quad \text{(A113)}$$

where (a) holds because a Gaussian distribution yields the largest differential entropy for a given covariance matrix, (b) is from the fact that $\det(K_{\mathbf{s}_i^1}) = \prod_{k=1}^i \lambda_k(K_{\mathbf{s}_i^1})$ and (c) is due

to the Cauchy interlacing theorem [30]. This proves the upper bound in (A112). For the lower bound, we have

$$h(\mathbf{r}_i^1) \overset{(a)}{\geq} h(\mathbf{r}_\kappa^1) - h(\mathbf{r}_{i\kappa}^{i+1}) \overset{(b)}{\geq} h(\mathbf{r}_\kappa^1) - \frac{\kappa - i}{2}\log(2\pi e\,\lambda_{\max}(\mathbf{K}_{\mathbf{s}_\kappa^1})), \tag{A114}$$

where (a) stems from the fact that $h(\mathbf{a}, \mathbf{b}) \leq h(\mathbf{a}) + h(\mathbf{b})$ and (b) follows from (A113). This completes the proof. □

Proposition A2. *Let* $\mathbf{u}_1^\infty$ *be a random process such that the variance of* $\mathbf{u}(n)$, $\sigma_{\mathbf{u}(n)}^2 < \infty$ *for finite* n*, and*

$$\lim_{n\to\infty} \frac{1}{n}\log(\sigma_{\mathbf{u}(n)}^2) = 0. \tag{A115}$$

Then, $\lim_{n\to\infty} n^{-1}\log(\lambda_{max}(\mathbf{K}_{\mathbf{u}_n^1})) = 0.$ ▲

Proof. The assumptions on $\mathbf{u}_1^\infty$ imply that, for every $\epsilon > 0$, there exists a finite N_ϵ such that, for every $n \geq N_\epsilon$, $\sigma_{\mathbf{u}(n)}^2 < e^{n\epsilon}$ and $S(N_\epsilon) \triangleq \max\{\sigma_{\mathbf{u}(1)}^2, \sigma_{\mathbf{u}(2)}^2, \dots, \sigma_{\mathbf{u}(N_\epsilon)}^2\} < \infty$. Then,

$$\frac{1}{n}\ln(\lambda_{\max}(\mathbf{K}_{\mathbf{u}_n^1})) \overset{(a)}{\leq} \frac{1}{n}\ln\left(\sum_{k=1}^n \sigma_{\mathbf{u}(k)}^2\right) < \frac{1}{n}\ln(N_\epsilon S(N_\epsilon) + (n - N_\epsilon)e^{n\epsilon}) \overset{(b)}{<} \frac{1}{n}\ln((n - N_\epsilon)e^{n\epsilon}) + \frac{N_\epsilon S(N_\epsilon)}{n(n - N_\epsilon)e^{n\epsilon}},$$

where (a) holds because $\sum_{k=1}^n \lambda_k(\mathbf{K}_{\mathbf{u}_n^1}) = \mathrm{tr}\{\mathbf{K}_{\mathbf{u}_n^1}\}$, while (b) stems from the fact that, for every $x, y > 0$, $\ln(x + y) < \ln(x) + y/x$. Thus, for every $\epsilon > 0$, $\lim_{n\to\infty} n^{-1}\log(\lambda_{\max}(\mathbf{K}_{\mathbf{u}_n^1})) < \epsilon$, which means that $\lim_{n\to\infty} n^{-1}\log(\lambda_{\max}(\mathbf{K}_{\mathbf{u}_n^1})) = 0$, completing the proof. □

Proposition A3. *Let* $\mathbf{v}_1^\infty$ *be an entropy-balanced random process. Then, for each* $\nu \in \mathbb{N}$ *and for every sequence of matrices* $\{\mathbf{\Psi}_n\}_{n=\nu}^\infty$, $\mathbf{\Psi}_n \in \mathbb{R}^{\nu \times n}$ *with orthonormal rows,*

$$\lim_{n\to\infty} \frac{1}{n}h(\mathbf{\Psi}_n\mathbf{v}_n^1) = 0. \tag{A116}$$

Proof of Proposition A3. We will first show that

$$\lim_{n\to\infty} \frac{1}{n}h(\mathbf{\Psi}_n\mathbf{v}_n^1) \geq 0. \tag{A117}$$

To see this, notice that, for every $\mathbf{\Psi}_n \in \mathbb{R}^{\nu \times n}$ with orthonormal rows, there exists a matrix $\mathbf{\Phi}_n \in \mathbb{R}^{(n-\nu) \times n}$ with orthonormal rows which are also orthogonal to those of $\mathbf{\Psi}_n$. This means that the matrix $[\mathbf{\Psi}_n^T | \mathbf{\Phi}_n^T]^T \in \mathbb{R}^{n \times n}$ is unitary; thus,

$$h(\mathbf{v}_n^1) = h([\mathbf{\Psi}_n^T | \mathbf{\Phi}_n^T]^T \mathbf{v}_n^1) \overset{(a)}{=} h(\mathbf{\Phi}_n\mathbf{v}_n^1) + h(\mathbf{\Psi}_n\mathbf{v}_n^1 | \mathbf{\Phi}_n\mathbf{v}_n^1) \overset{(b)}{\leq} h(\mathbf{\Phi}_n\mathbf{v}_n^1) + h(\mathbf{\Psi}_n\mathbf{v}_n^1), \quad \text{(A118)}$$

where (a) holds due to the chain-rule of differential entropy and (b) follows because conditioning does not increase differential entropy. Therefore, $h(\mathbf{\Psi}_n\mathbf{v}_n^1) \geq h(\mathbf{v}_n^1) - h(\mathbf{\Phi}_n\mathbf{v}_n^1)$. Dividing this by n, taking the limit when $n \to \infty$ and recalling that $\mathbf{v}_1^\infty$ satisfies (17) yields (A117).

We will now prove that $\lim_{n\to\infty} \frac{1}{n}h(\mathbf{\Psi}_n\mathbf{v}_n^1) \leq 0$. For this purpose, let $\tilde{\mathbf{v}}_1^\infty$ be a jointly Gaussian random process with the same second-order statistics as $\mathbf{v}_1^\infty$. Then,

$$h(\mathbf{\Psi}_n\mathbf{v}_n^1) \leq h(\mathbf{\Psi}_n\tilde{\mathbf{v}}_n^1) = \frac{1}{2}\log\left((2\pi e)^\nu \det(\mathbf{\Psi}_n\mathbf{K}_{\tilde{\mathbf{v}}_n^1}\mathbf{\Psi}_n^T)\right) \leq \frac{1}{2}\log\left((2\pi e)^\nu \lambda_{\max}(\mathbf{K}_{\tilde{\mathbf{v}}_n^1})^\nu\right), \tag{A119}$$

with the inequality due to the fact that $\mathbf{\Psi}_n$ has orthonormal rows. But v_1^∞ meets the requirements of Proposition A2; thus, $\lim_{n\to\infty} \frac{1}{n} h(\mathbf{\Psi}_n \mathbf{v}_n^1) \leq \lim_{n\to\infty} \frac{\nu}{2n}(2\pi\, e\, \lambda_{\max}(K_{\tilde{\mathbf{v}}_n^1})) = 0$. The proof is completed by combining this result with (A117).　□

Lemma A1. *Let* u_1^∞ *be a random process with independent elements, and where each element* u_i *is uniformly distributed over possible different intervals* $[-\frac{a_i}{2}, \frac{a_i}{2}]$, *such that* $a_{max} > |a_i| > a_{min} > 0, \forall i \in \mathbb{N}$, *for some positive and bounded* $a_{min} < a_{max}$. *Then,* u_1^∞ *is entropy balanced.*

Proof. Without loss of generality, we can assume that $a_i \geq 1$, for all i (otherwise, we could scale the input by $1/a_{\min}$, which would scale the output by the same proportion, increasing the input entropy by $n \log(1/a_{\min})$ and the output entropy by $(n - \nu) \log(1/a_{\min})$, without changing the result). The input vector $\mathbf{u}_n^1$ is confined to an n-box $\mathcal{U}_n$ (the support of u_1^n) of volume $\mathcal{V}_n(\mathcal{U}_n) = \prod_{i=1}^n a_i$ and has entropy $\log(\prod_{i=1}^n a_i)$. This support is an n-box which contains $\binom{n}{k} 2^{n-k}$ k-boxes of different k-volume. Each of these k-boxes is determined by fixing $n - k$ entries in $\mathbf{u}_n^1$ to $\pm a_i/2$, and letting the remaining k entries sweep freely over $[-\frac{a_i}{2}, \frac{a_i}{2}]$. Thus, the k-volume of each k-box is the product of the k support sizes a_i of the associated selected free-sweeping entries. But recalling that $a_i > 1$ for all i, the volume of each k-box can be upper bounded by $\prod_{i=1}^n a_i$. With this, the added volume of all the k-boxes contained in the original n-box can be upper bounded as

$$\mathcal{V}_k^\square(\mathcal{U}_n) \leq \binom{n}{k} 2^{n-k} \prod_{i=1}^n a_i. \tag{A120}$$

We now use this result to upper bound the entropy rate of $\mathbf{y}_n^{\nu+1}$.

Let $\mathbf{y}_n^1 \triangleq [\mathbf{\Psi}_n^T | \mathbf{\Phi}_n^T]^T \mathbf{u}_n^1$ where $[\mathbf{\Psi}_n^T | \mathbf{\Phi}_n^T]^T \in \mathbb{R}^{n\times n}$ is a unitary matrix and where $\mathbf{\Psi}_n \in \mathbb{R}^{\nu \times n}$ and $\mathbf{\Phi}_n \in \mathbb{R}^{(n-\nu)\times n}$ have orthonormal rows. From this definition, $\mathbf{y}_n^{\nu+1}$ will distribute over a finite region $\mathcal{Y}_n^{\nu+1} \subseteq \mathbb{R}^{n-\nu}$, corresponding to the projection onto the $(n-\nu)$-dimensional span of the rows of $\mathbf{\Phi}_n$. Hence, $h(\mathbf{y}_n^{\nu+1})$ is upper bounded by the entropy of a uniformly distributed vector over the same support, i.e., by $\log \mathcal{V}_{n-\nu}(\mathcal{Y}_n^{\nu+1})$, where $\mathcal{V}_{n-\nu}(\mathcal{Y}_n^{\nu+1})$ is the $(n-\nu)$-dimensional volume of this support. In turn, $\mathcal{V}_{n-\nu}(\mathcal{Y}_n^{\nu+1})$ is upper bounded by the sum of the volume of all $(\nu - k)$-dimensional boxes contained in the n-box in which $\mathbf{u}_n^1$ is confined, which we already denoted by $\mathcal{V}_{n-\nu}^\square(\mathcal{U}_n)$, and which is upper bounded as in (A120). Therefore,

$$h(\mathbf{y}_n^{1+\nu}) \leq \log \mathcal{V}_{n-\nu}(\mathcal{Y}_n^{\nu+1}) \leq \log \mathcal{V}_{n-\nu}^\square(\mathcal{U}_n) \leq \log\left(\frac{n!}{(n-\nu)!\nu!} 2^\nu \prod_{i=1}^n a_i\right) \tag{A121}$$

$$= \log(n^\nu 2^\nu) + \log\left(\frac{n!}{(n-\nu)! n^\nu \nu!}\right) + \log\left(\prod_{i=1}^n a_i\right). \tag{A122}$$

Recalling that $h(\mathbf{u}_n^1) = \log(\prod_{i=1}^n a_i)$, dividing by n and taking the limit as $n \to \infty$ yields

$$\lim_{n\to\infty} \frac{1}{n}\left(h(\mathbf{y}_n^{\nu+1}) - h(\mathbf{u}_n^1)\right) \leq 0. \tag{A123}$$

On the other hand,

$$h(\mathbf{y}_n^{\nu+1}) = h(\mathbf{y}_n^1) - h(\mathbf{y}_\nu^1|\mathbf{y}_n^{\nu+1}) \overset{(a)}{=} h(\mathbf{u}_n^1) - h(\mathbf{y}_\nu^1|\mathbf{y}_n^{\nu+1}) \geq h(\mathbf{u}_n^1) - h(\mathbf{y}_\nu^1), \tag{A124}$$

where (a) follows because $[\mathbf{\Psi}_n^T | \mathbf{\Phi}_n^T]^T$ is an orthogonal matrix. Letting $(\mathbf{y}_G)_\nu^1$ correspond to the jointly Gaussian sequence with the same second-order moments as $\mathbf{y}_\nu^1$, and recalling that the Gaussian distribution maximizes differential entropy for a given covariance, we obtain the upper bound

$$h(\mathbf{y}_\nu^1) \leq h((\mathbf{y}_G)_\nu^1) \overset{(a)}{=} \frac{1}{2}\log\left((2\pi\, e)^\nu \det(\mathbf{\Psi}_n \operatorname{diag}\{\sigma_{\mathrm{u}_i}^2\}_{i=1}^n \mathbf{\Psi}_n^T)\right) \overset{(b)}{\leq} \frac{\nu}{2}\log\left(2\pi\, e \max\{\sigma_{\mathrm{u}_i}^2\}_{i=1}^n\right), \tag{A125}$$

where (a) follows since the $\{u_i\}_{i=1}^n$ are independent, and (b) stems from the fact that $\mathbf{\Psi}_n \in \mathbb{R}^{\nu \times n}$ has orthonormal rows. Since $\max\{\sigma_{u_i}^2\}_{i=1}^n$ is bounded for all n, we obtain by substituting (A125) into (A124) that $\lim_{n\to\infty} \frac{1}{n}(h(\mathbf{y}_n^{\nu+1}) - h(\mathbf{u}_n^1)) \geq 0$. The combination of this with (A123) yields $\lim_{n\to\infty} \frac{1}{n}(h(\mathbf{y}_n^{\nu+1}) - h(\mathbf{u}_n^1)) = 0$, satisfying Condition ii) in Definition 2. From this, the proof is completed by noting that u_1^∞ satisfies Condition i) in Definition 2. This completes the proof. $\square$

Lemma A2. *Let $A(z)$ be a causal, finite-order, stable and strictly minimum-phase rational transfer function with impulse response $a_0, a_1, \ldots$ such that $a_0 = 1$. Then, $\lim_{n\to\infty} \lambda_1(A_n A_n^T) > 0$ and $\lim_{n\to\infty} \lambda_n(A_n A_n^T) < \infty$.*

Proof of Lemma A2. The fact that $\lim_{n\to\infty} \lambda_n(A_n A_n^T)$ is upper bounded follows directly from the fact that $A(z)$ is a stable transfer function. On the other hand, $A_n A_n^T$ is positive definite, with $\lim_{n\to\infty} \lambda_1(A_n A_n^T) \geq 0$. Suppose that $\lim_{n\to\infty} \lambda_1(A_n A_n^T) = 0$. If this were true, then it would hold that $\lim_{n\to\infty} \lambda_n(A_n^{-1} A_n^{-T}) = \infty$. But A_n^{-1} is the lower triangular Toeplitz matrix associated with $A^{-1}(z)$, which is stable (since $A(z)$ is minimum phase), implying that $\lim_{n\to\infty} \lambda_n(A_n^{-1} A_1^{-T}) < \infty$, thus leading to a contradiction. This completes the proof. $\square$

We re-state here (for completeness and convenience) the unnumbered lemma in the proof of Reference [16], Theorem 1, as follows:

Lemma A3. *Let the transfer function $G(z)$ satisfy Assumption 1 and suppose it has no poles. Then,*

$$\lambda_l(G_n G_n^T) = \begin{cases} \alpha_{n,l}^2 (\rho_l)^{-2n} & ,\text{if } l \leq m, \\ \alpha_{n,l}^2 & ,\text{otherwise}, \end{cases} \quad \text{(A126)}$$

where the elements in the sequence $\{\alpha_{n,l}\}$ are positive and increase or decrease at most polynomially with n.

Lemma A4. *Let A, B be matrices with the same dimensions. Then,*

$$\lambda_{min}\left((A+B)(A+B)^T\right) \geq \lambda_{min}(AA^T) + \lambda_{min}(BB^T) - 2\sigma_{max}(A)\sigma_{max}(B). \quad \text{(A127)}$$

Proof. For every x such that $\|x\| = 1$,

$$x^T(A+B)(A+B)^T x = x^T AA^T x + x^T BB^T x + x^T AB^T x + x^T BA^T x$$
$$\geq \lambda_{\min}(AA^T) + \lambda_{\min}(BB^T) - 2\sigma_{\max}(A)\sigma_{\max}(B), \quad \text{(A128)}$$

where the last inequality holds because AA^T and BB^T are symmetric and because of the Cauchy-Schwartz inequality. The proof is completed by noting that (A128) holds for the x that minimizes $x^T(A+B)(A+B)^T x$, and $\lambda_{\min}((A+B)(A+B)^T) = \min_{x:\|x\|=1} x^T(A+B)(A+B)^T x$. $\square$

References

1. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley-Interscience: Hoboken, NJ, USA, 2006.
2. Shannon, C.E.; Weaver, W. *The Mathematical Theory of Communication*; University of Illinois Press: Urbana, IL, USA, 1949.
3. Itô, H. Principle of the minimum entropy in information theory. *Proc. Jpn. Acad.* **1953**, *29*, 194–197. [CrossRef]
4. O'Neal, J. Bounds on subjective performance measures for source encoding systems. *IEEE Trans. Inf. Theory* **1971**, *17*, 224–231. [CrossRef]
5. Pierobon, M.; Akyildiz, I.F. Capacity of a Diffusion-Based Molecular Communication System With Channel Memory and Molecular Noise. *IEEE Trans. Inf. Theory* **2013**, *59*, 942–954. [CrossRef]
6. Akyildiz, I.F.; Pierobon, M.; Balasubramaniam, S. An Information Theoretic Framework to Analyze Molecular Communication Systems Based on Statistical Mechanics. *Proc. IEEE* **2019**, *107*, 1230–1255. [CrossRef]

7. Aaron, M.R.; McDonald, R.A.; Protonotarios, E. Entropy power loss in linear sampled data filters. *Proc. IEEE* **1967**, *55*, 1093–1094. [CrossRef]

8. Papoulis, A.; Pillai, S.U. *Probability, Random Variables and Stochastic Processes*, 3rd ed.; McGraw-Hill: New York, NY, USA, 1991.

9. Kim, Y.H. Feedback capacity of stationary Gaussian channels. *IEEE Trans. Inf. Theory* **2010**, *56*, 57–85. [CrossRef]

10. Rudin, W. *Real and Complex Analysis*; McGraw-Hill: New York, NY, USA, 1987.

11. Serón, M.M.; Braslavsky, J.H.; Goodwin, G.C. *Fundamental Limitations in Filtering and Control*; Springer: London, UK, 1997.

12. Martins, N.C.; Dahleh, M.M.A. Fundamental limitations of performance in the presence of finite capacity feedback. In Proceedings of the 2005, American Control Conference, Portland, OR, USA, 8–10 June 2005.

13. Martins, N.; Dahleh, M. Feedback control in the presence of noisy Channels: "Bode-like" fundamental limitations of performance. *IEEE Trans. Autom. Control* **2008**, *53*, 1604–1615. [CrossRef]

14. Yu, S.; Mehta, P.G. Bode-Like Fundamental Performance Limitations in Control of Nonlinear Systems. *IEEE Trans. Autom. Control* **2010**, *55*, 1390–1405. [CrossRef]

15. Gray, R.M. Information rates of autorregressive processes. *IEEE Trans. Inf. Theory* **1970**, *IT-16*, 412–421. [CrossRef]

16. Hashimoto, T.; Arimoto, S. On the rate-distortion function for the nonstationary Gaussian autoregressive process. *IEEE Trans. Inf. Theory* **1980**, *IT-26*, 478–480. [CrossRef]

17. Gray, R.M.; Hashimoto, T. A note on rate-distortion functions for nonstationary Gaussian autoregressive processes. *IEEE Trans. Inf. Theory* **2008**, *54*, 1319–1322. [CrossRef]

18. Gray, R.M. Toeplitz and circulant matrices: A review. *Found. Trends Commun. Inf. Theory* **2006**, *2*, 155–239. [CrossRef]

19. Rényi, A. On the dimension and entropy of probability distributions. *Acta Math. Hung.* **1959**, *10*, 193–215. [CrossRef]

20. Shannon, C. Coding theorems for a discrete source with a fidelity criterion. *IRNE Nat. Conv. Rec.* **1959**, *4*, 143–163.

21. Goodwin, G.C.; Graebe, S.F.; Salgado, M.E. *Control System Design*, 1st ed.; Prentice Hall PTR: Upper Saddle River, NJ, USA, 2000.

22. Chen, C.T. *Linear System Theorey and Design*, 3rd ed.; The Oxford Series in Electrical and Computing Engineering; Oxford University Press: Oxford, UK, 1999.

23. Elia, N. When Bode meets Shannon: Control-oriented feedback communication schemes. *IEEE Trans. Autom. Control* **2004**, *49*, 1477–1488. [CrossRef]

24. Silva, E.I.; Derpich, M.S.; Østergaard, J. A framework for control system design subject to average data-rate constraints. *IEEE Trans. Autom. Control* **2011**, *56*, 1886–1899. [CrossRef]

25. Silva, E.I.; Derpich, M.S.; Østergaard, J. An achievable data-rate region subject to a stationary performance constraint for LTI plants. *IEEE Trans. Autom. Control* **2011**, *56*, 1968–1973. [CrossRef]

26. Yüksel, S. Characterization of Information Channels for Asymptotic Mean Stationarity and Stochastic Stability of Nonstationary/Unstable Linear Systems. *IEEE Trans. Inf. Theory* **2012**, *58*, 6332–6354. [CrossRef]

27. Freudenberg, J.S.; Middleton, R.H.; Braslavsky, J.H. Minimum Variance Control Over a Gaussian Communication Channel. *IRE Trans. Comm. Syst.* **2011**, *56*, 1751–1765. [CrossRef]

28. Ardestanizadeh, E.; Franceschetti, M. Control-theoretic approach to communication with feedback. *IEEE Trans. Autom. Control* **2012**, *57*, 2576–2587. [CrossRef]

29. Grenander, U.; Szegö, G. *Toeplitz Forms and Their Applications*; University of California Press: Berkeley, CA, USA, 1958.

30. Horn, R.A.; Johnson, C.R. *Matrix Analysis*; Cambridge University Press: Cambridge, UK, 1985.

31. Fiedler, M. Bounds for the determinant of the sum of Hermitian matrices. *Proc. Am. Math. Soc.* **1971**, *30*, 27–31. [CrossRef]