entropy

MDPI

Article

The Capacity Gains of Gaussian Channels with Unstable Versus Stable Autoregressive Noise

Charalambos D. Charalambous [1,*], Christos Kourtellaris [1], Stelios Louka [1] and Sergey Loyka [2]

[1] Department of Electrical and Computer Engineering, University of Cyprus, P.O. Box 20537, CY-1678 Nicosia, Cyprus; christoskourtellaris@yahoo.gr (C.K.); louka.stelios@ucy.ac.cy (S.L.)
[2] School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, ON K1N 6N5, Canada; sergey.loyka@uottawa.ca
* Correspondence: chadcha@ucy.ac.cy

Abstract

In this paper, we consider Cover's and Pombra's formulation of feedback capacity of additive Gaussian noise (AGN) channels, with jointly Gaussian nonstationary and nonergodic noise. We derive closed-form feedback capacity formulas, using Karush–Kuhn–Tucker (KKT) conditions and convergence properties of difference Riccati equations to limiting algebraic Riccati equations of filtering theory, for unstable and stable autoregressive (AR) noise. Surprisingly, the capacity formulas depend on the parameters of the AR noise, its pole $c \in (-\infty, \infty)$ and noise variance $K_W \in (0, \infty)$, and the total transmit power $\kappa \in [0, \infty)$, indicating substantial gains for the unstable noise region $\forall c^2 \in (1, \infty), \forall \kappa > \kappa_{min} \stackrel{\triangle}{=} \frac{K_W\left(1+\sqrt{4c^2-3}\right)}{2\left(c^2-1\right)^2}$ compared to its complement region. In particular, feedback capacity is distinguished by three regimes, as follows. Regime 1, $\forall c^2 \in (1, \infty), \forall \kappa > \kappa_{min}$: the optimal channel input includes an innovations part, the capacity increases as $|c| > 1$ increases, while κ_{min} and the allocated transmit power decrease. Regime 2, $\forall c^2 \in (1, \infty), \forall \kappa \leq \kappa_{min}$, Regime 3, $\forall c \in [-1, 1], \forall \kappa \in [0, \infty)$ (complement of Regime 1): the innovations part of the optimal channel is asymptotically zero and the capacity is fundamentally different compared to Regime 1. The differences of capacity formulas for Regimes 1, 2 and 3 are directly related to their operational meaning: (i) Regime 1 is an ergodic capacity while Regimes 2 and 3 are nonergodic capacities; (ii) Regime 1 is achieved by an asymptotically stationary channel input with a non-zero innovations part, while Regimes 2 and 3 are achieved by an asymptotically zero innovations part. The gains of capacity for Regime 1 are attributed to the high correlation of noise samples compared to stable noise and the use of an informative innovations part by the optimal channel input, which make possible the prediction of future noise samples from past samples, unlike memoryless noise. Our results provide answers to certain open questions regarding the validity of capacity formulas of stable noise that appeared in the literature.

Keywords: capacity gains, Gaussian noise; nonstationary; stationary, closed-form feedback capacity expressions, Riccati equations; detectability; stabilizability; unit circle controllability

check for updates

Academic Editor: Jun Chen

Received: 16 September 2025
Revised: 7 December 2025
Accepted: 8 December 2025
Published: 18 December 2025

Citation: Charalambous, C.D.; Kourtellaris, C.; Louka, S.; Loyka, S. The Capacity Gains of Gaussian Channels with Unstable Versus Stable Autoregressive Noise. *Entropy* **2025**, 27, 1264. https://doi.org/10.3390/e27121264

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The main objective of this paper is to show fundamental gains of feedback and nonfeedback capacity of additive Gaussian noise channels, driven by nonstationary and nonergodic noise compared to stationary noise. A secondary objective is to provide answers

to the open question of the recent comment paper [1] [Abstract, Conclusion] concerning the correctness of the feedback capacity characterization [2] [Theorem 6.1], by identifying conditions of its validity, which are consistent with the technical analysis of our application example, and indicate its implications on the operational meaning of capacity.

1.1. The AGN Channel with Nonstationary and Nonergodic Noise

We consider the additive Gaussian noise (AGN) channel depicted in Figure 1, driven by nonstationary and nonergodic noise as in Cover and Pombra [3],

$$Y_t = X_t + V_t, \quad t = 1, \ldots, n, \quad \frac{1}{n}\mathbf{E}\Big\{\sum_{t=1}^{n}(X_t)^2\Big\} \leq \kappa, \quad \kappa \in [0, \infty) \tag{1}$$

where $X^n = (X_1, X_2, \ldots, X_n)^T$ and $Y^n = (Y_1, Y_2, \ldots, Y_n)^T$ are vectors of channel input and output random variables (RVs), $X_t : \Omega \to \mathbb{R}, t = 1, \ldots, n$ and $Y_t : \Omega \to \mathbb{R}, t = 1, \ldots, n$, respectively, and $V^n = (V_1, \ldots, V_n)^T$ is a vector of nonstationary and nonergodic, jointly Gaussian noise, and $V_t : \Omega \to \mathbb{R}, t = 1, \ldots, n$, with mean μ_{V^n}, and covariance K_{V^n}, denoted by $V^n \in N(\mu_{V^n}, K_{V^n})$. Here, $V_1 \in N(0, K_{V_1})$ is the initial noise RV, with induced distribution $\mathbf{P}_{V_1}$, called the initial state of the channel. Without loss of generality we assume $\mu_{V^n} = 0$.

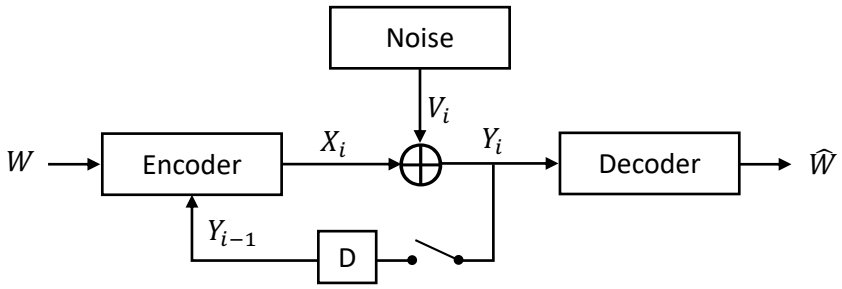


Figure 1. Additive Gaussian noise channel with jointly Gaussian noise $V^n \stackrel{\triangle}{=} \{V_1, V_2, \ldots, V_n\}$. If the switch is open the encoder does not use feedback; if the switch is closed the encoder uses feedback.

Feedback (resp. Nonfeedback) Code. We wish to communicate to anyone the uniformly distributed messages, $W : \Omega \to \mathscr{M}^{(n)} \stackrel{\triangle}{=} \{1, 2, \ldots, M_n\}$, $M_n \stackrel{\triangle}{=} \lceil 2^{nR} \rceil$, by using an encoder with feedback $X_1 = e_1(W), X_2 = e_2(W, X_1, Y_1), \ldots, X_n = e_n(W, X^{n-1}, Y^{n-1})$ (resp., without feedback $X_1 = e_1(W), X_2 = e_2(W, X_1) \equiv \bar{e}_2(W), \ldots, X_n = e_n(W, X^{n-1}) \equiv \bar{e}_n(W)$), and a decoder with arbitrary small average error probability $\mathbf{P}_{error}^{(n)}(\mathbf{P}_{V_1}) = \mathbb{P}\{d_n(Y^n) \neq W\} = \frac{1}{M_n}\sum_{w=1}^{M_n}\mathbf{P}\big(d_n(Y^n) \neq w \big| W = w\big)$, after n sufficiently large transmissions, under the assumption of independence of W and V^n (the dependence of $\mathbf{P}_{error}^{(n)}(\mathbf{P}_{V_1})$ on $\mathbf{P}_{V_1}$ emphasizes the fact that the error probability, in general, depends on the RV $V_1 \in N(0, K_{V_1})$). A feedback rate R is called an *achievable rate*, if there exists a sequence of encoders and decoders satisfying $\lim_{n\longrightarrow\infty}\mathbf{P}_{error}^{(n)}(\mathbf{P}_{V_1}) = 0$ and $\liminf_{n\longrightarrow\infty}\frac{1}{n}\log\lceil M_n \rceil \geq R$. The operational definition of the feedback capacity is $C_{FB}(\kappa, \mathbf{P}_{V_1}) \stackrel{\triangle}{=} \sup\big\{R \big| R \text{ is achievable}\big\}$, i.e., in general, the capacity depends on $\mathbf{P}_{V_1}$. If $C_{FB}(\kappa, \mathbf{P}_{V_1}) = C_{FB}(\kappa), \forall \mathbf{P}_{V_1}$, then it is independent of the choice of the initial distribution $\mathbf{P}_{V_1}$ and $\mathbf{P}_{Y_1}$. Similarly, the nonfeedback capacity is denoted by $C(\kappa, \mathbf{P}_{V_1})$.

In 1989, Cover and Pombra [3], derived a time-domain (TD) information theoretic characterization of feedback capacity $C(\kappa, \mathbf{P}_{V_1})$ (for arbitrary nonstationary and nonergodic noise V^n, not necessarily finite-dimensional), expressed as an asymptotic limit,

$C_{FB}(\kappa, \mathbf{P}_{V_1}) = \lim_{n\longrightarrow\infty} C_{n,FB}(\kappa, \mathbf{P}_{V_1})$, where $C_{n,FB}(\kappa, \mathbf{P}_{V_1})$ is called the $n-$finite transmission or block length feedback information ($n-$FTFI) capacity, given as follows.

$$C_{n,FB}(\kappa, \mathbf{P}_{V_1}) = \sup_{\mathbf{P}_{X_t|X^{t-1},Y^{t-1}}, t=1,\ldots,n, \frac{1}{n}\mathbf{E}\left\{\sum_{t=1}^{n}(X_t)^2\right\}\leq\kappa} \frac{1}{n}\{H(Y^n) - H(V^n)\} \tag{2}$$

$$= \sup_{\left\{(B_n, K_{\overline{Z}^n})\left|\frac{1}{n}Trace\left(B_n\, K_{V^n}\, (B_n)^T + K_{\overline{Z}^n}\right)\leq\kappa\right.\right\}} \frac{1}{2n}\log\frac{\left|\left(B_n + I_n\right)K_{V^n}\left(B_n + I_n\right)^T + K_{\overline{Z}^n}\right|}{|K_{V^n}|} \tag{3}$$

where $B_n \in \mathbb{R}^{n\times n}$ is a lower diagonal matrix, I_n is an $n \times n$ identity matrix, and $K_{\overline{Z}^n}$ is a covariance matrix of a jointly Gaussian process $\overline{Z}^n \in N(0, K_{\overline{Z}^n})$, and $H(X)$ denotes (differential) entropy of RV X.

In [3], it is shown that $C_{n,FB}(\kappa, \mathbf{P}_{V_1})$ is achieved by a channel input process X^n with time-varying coefficients given by

$$X_t = \sum_{j=1}^{t-1} b_{t,j}V_j + \overline{Z}_t, \quad Y_t = \sum_{j=1}^{t-1} b_{t,j}V_j + \overline{Z}_t + V_t, \quad t = 1,\ldots,n, \tag{4}$$

$$X^n = B_nV^n + \overline{Z}^n, \quad Y^n = \left(B_n + I_n\right)V^n + \overline{Z}^n, \tag{5}$$

$$\overline{Z}^n \in N(0, K_{\overline{Z}^n}), \quad K_{\overline{Z}^n} \succeq 0 \tag{6}$$

where $\overline{Z}^n$ is independent of the noise V^n. The reason [3] expressed X_t in terms of V^{t-1} and not (X^{t-1}, Y^{t-1}) is because (X^{t-1}, Y^{t-1}) uniquely defines (V^{t-1}, Y^{t-1}) and vice-versa, $\forall t$, and hence (X^{t-1}, Y^{t-1}) can be expressed in terms of V^{t-1} and the jointly Gaussian process $\overline{Z}_t, t = 1,\ldots,n$ (this should not be confused by an innovations process, which in the standard terminology is an orthogonal process).

Cover and Pombra [3] also derived a TD information theoretic characterization for nonfeedback capacity $C(\kappa, \mathbf{P}_{V_1})$, for the same nonstationary and nonergodic noise V^n, also expressed as an asymptotic limit, $C(\kappa, \mathbf{P}_{V_1}) = \lim_{n\longrightarrow\infty} C_n(\kappa, \mathbf{P}_{V_1})$, where $C_n(\kappa, \mathbf{P}_{V_1})$ is called the $n-$finite transmission or block length without feedback information ($n-$FTwFI) capacity, given as follows.

$$C_n(\kappa, \mathbf{P}_{V_1}) = \sup_{\mathbf{P}_{X_t|X^{t-1}}, t=1,\ldots,n, \frac{1}{n}\mathbf{E}\left\{\sum_{t=1}^{n}(X_t)^2\right\}\leq\kappa} \frac{1}{n}\{H(Y^n) - H(V^n)\} \tag{7}$$

$$= \sup_{\left\{K_{\overline{Z}^n}\left|\frac{1}{n}Trace\left(K_{\overline{Z}^n}\right)\leq\kappa\right.\right\}} \frac{1}{2n}\log\frac{|K_{\overline{Z}^n} + K_{V^n}|}{|K_{V^n}|}. \tag{8}$$

The optimal channel input process X^n that achieves $C_n(\kappa, \mathbf{P}_{V_1})$ is obtained directly from (4)–(6), by letting $B_n = 0$, Ref. [3] [Equations (14)–(16)], i.e., $X^n = \overline{Z}^n, Y^n = \overline{Z}^n + V^n$.

The main objective of this paper is to understand the implications of nonstationary and nonergodic Gaussian noise on feedback and nonfeedback capacity compared to the corresponding capacity of stationary or asymptotically stationary and ergodic noise.

The main result of the paper are the gains on feedback and nonfeedback capacity of nonstationary and nonergodic autoregressive noise compared to the corresponding capacity of stationary or asymptotically stationary and ergodic noise.

Our derivations also provide insight and answers to the recent comment paper by Derpich and Ostergaard [1], which identified gaps in the proofs of the frequency-domain (FD) and TD characterizations of feedback capacity [2] [Theorem 4.1] and [2] [Theorem 6.1] for stationary or asymptotically stationary noise V^n. In fact, from our analysis it follows

that the TD characterization of feedback capacity [2] [Theorem 6.1] is correct, provided certain conditions hold. We identify these conditions.

1.2. Motivation

The current study is motivated by the fundamental limitations of capacity analysis of AGN channels driven by stationary or asymptotically stationary Gaussian noise (stable) $V_t, t = 1, 2, \ldots$. To make our point precise, we recall Shannon's FD information theoretic characterization of nonfeedback capacity of AGN channels driven by stationary or asymptotically stationary (stable) Gaussian noise $V_t, t = 1, 2, \ldots$ with power spectral density (PSD), $S_V(e^{j\theta}), \forall \theta \in [-\pi, \pi]$.

The nonfeedback capacity is described by the well-known *water-filling solution* of the optimal PSD of the channel input $S_X(e^{j\theta}), \forall \theta \in [-\pi, \pi]$, as follows [4–6].

$$\begin{aligned} &C = C^{Shannon} \stackrel{\triangle}{=} \frac{1}{4\pi} \int_{-\pi}^{\pi} \log \Big(\max \{1, \frac{\lambda}{S_V(e^{j\theta})}\} \Big) d\theta \\ &\frac{1}{2\pi} \int_{-\pi}^{\pi} S_X^*(e^{j\theta}) d\theta = \frac{1}{2\pi} \int_{-\pi}^{\pi} \max \{0, \lambda - S_V(e^{j\theta})\} d\theta = \kappa, \qquad \lambda \geq 0, \\ &S_X^*(e^{j\theta}) = \max \{0, \lambda - S_V(e^{j\theta})\} \end{aligned}$$

where λ is the Lagrange multiplier associated with the transmit average power constraint. If the noise PSD is white, i.e., $S_V(e^{j\theta}) = K_V, \forall \theta \in [-\pi, \pi]$, then

$$\begin{aligned} &C = C^{Shannon} \stackrel{\triangle}{=} \frac{1}{2} \log \Big(\max \{1, \frac{\lambda}{K_V}\} \Big), \\ &\max \{0, \lambda - K_V\} = \kappa, \; S_X^*(e^{j\theta}) = K_X^* = \max \{0, \lambda - K_V\}. \end{aligned}$$

The above water-filling formula is restricted to stationary or asymptotically stationary noise, i.e., the pole c of an autoregressive unit memory noise needs to satisfy $|c| < 1$. As $|c|$ approaches 1 the noise power approaches $+\infty$, and this imposes a limitation on the use of formulae $C^{Shannon}$. In this paper, we show that we can use the time-domain formulae for C to compute simple lower bounds on C for nonstationary and nonergodic channels, especially channels with unstable noise (i.e., their poles lie outside the unit disk in the space of complex numbers).

1.3. Literature Review

The literature on feedback and nonfeedback capacity and bounds for AGN channels with stationary or asymptotically stationary noise is very extensive. However, corresponding results for nonstationary and nonergodic noise are not sufficiently developed.

Information Theoretic Bounds on Feedback Capacity. In the early 1970s, Butman, Tienan and Schalkwijk, and Wolfowitz [7–10] derived TD information theoretic bounds on feedback capacity $C_{FB}(\kappa, \mathbf{P}_{V_1})$ of the AGN channel (1), when the channel noise V^n is a stable autoregressive (AR) Gaussian process. These bounds are derived by invoking linear feedback coding schemes of communicating Gaussian random variables (RVs), $\Theta : \Omega \to \mathbb{R}$ and digital messages $W : \Omega \to \{1, 2, \ldots, 2^{nR}\}$, under the assumption $V_1 = v_1$ is known to the encoder and the decoder (see [7–10]).

Information Theoretic Characterization of Feedback and Nonfeedback Capacity for Nonstationary and Nonergodic Noise. In 1989, Cover and Pombra [3] derived a TD information theoretic characterization of feedback and nonfeedback capacity for the AGN channel (1), when the noise V^n is arbitrary nonstationary and nonergodic, as introduced earlier. Although, $C_{n,FB}(\kappa, \mathbf{P}_{Y_1})$ and $C_n(\kappa, \mathbf{P}_{Y_1})$ are compactly represented, to this date their computation for

any finite n, even for the stationary (i.e., stable) AR noise, is very challenging (see [11] for discussion). However, Ref. [3] influenced many subsequent studies [1,2,11–25].

Sequential Information Theoretic Characterization of Feedback for Stationary and Asymptotically Stationary Noise. In 2007, Yang, Kavcic and Tatikonda [20] considered the AGN channel (1), when the noise is stationary or asymptotically stationary described by a power spectral density (PSD) with stable poles and zeros. In [20] [Section Section II.C, Equations (19) and (20)], the authors introduced specific stable state space realizations of the noise, with state $S_t : \Omega \to \mathbb{R}^{n_s}, t = 1, \ldots, n$, which satisfies two important conditions [20] [Section II.C, I–III], specifically,

Condition 1. *The initial state of the noise or channel $S_1 = s$ is known to the encoder and decoder.*

Condition 2. *Given a fixed initial state $S_1 = s$, known to the encoder and decoder, at each t, the state space realization of the noise is such that the noise V^{t-1} uniquely defines the state of the noise S^t and vice-versa, $\forall t$.*

Under Conditions 1 and 2, Ref. [20] [Theorem 1] presents an equivalent TD sequential characterization of Cover's and Pombra's $n-$FTFI capacity, $C_{n,FB}(\kappa, \mathbf{P}_{Y_1})$, which is much simpler. The simplification is attributed to Conditions 1 and 2, which imply the optimal channel input conditional distribution that achieves $C_{n,FB}(\kappa, \mathbf{P}_{Y_1})$ can be expressed causally in terms of the state of the noise and the channel output, i.e., (S^n, Y^n) and an innovations process Z^n. This gave rise to [20] [Theorem 5], which is a sequential characterization of the $n-$FTFI capacity, $C_{n,FB}(\kappa, \mathbf{P}_{Y_1})$, expressed with respect to (w.r.t.) *a single matrix difference Riccati equation (DRE) of Kalman-filtering, parametrized by two variables which need to be optimized*. In [20] [Section IV], the asymptotic limit, $C_{FB}(\kappa, \mathbf{P}_{Y_1}) = \lim_{n \longrightarrow} C_{n,FB}(\kappa, \mathbf{P}_{Y_1})$, is also investigated, called "maximal information rate", Ref. [20] [Theorem 6, I_{max}]; this is expressed w.r.t. *a single matrix algebraic Riccati equation (ARE) of Kalman-filtering, parametrized by the two variables that need to be optimized*.

In 2010, Kim [2] re-visited the AGN channel (1), when the noise is stationary or asymptotically stationary noise described by a power spectral density (PSD) with stable poles and zeros (as in [20]), and by a state space realization [2] [Equation (58)] (the realization is equivalent to the one considered in [20], i.e., both correspond to the same PSD of the noise). The main results of [2] are two intriguing characterizations of feedback capacity, $C_{FB}(\kappa, \mathbf{P}_{Y_1})$, the FD characterization [2] [Theorem 4.1] and the TD characterization [2] [Theorem 6.1] for state space [2] [Equation (58)]. Similarly to [20] [Theorem 6], the TD characterization of feedback capacity $C_{FB}(\kappa, \mathbf{P}_{Y_1})$ in [2] [Theorem 6.1] is also expressed w.r.t. *a single matrix (ARE of Kalman-filtering, but unlike [20] [Theorem 6], it is parametrized by only a single variable that needs to be optimized*. The simplification of $C_{FB}(\kappa, \mathbf{P}_{Y_1})$ in [2] [Theorem 6.1] by a one-parameter optimization problem is based on [2] [Lemma 6.1], which states *the optimal covariance of the innovations part of the channel input is asymptotically zero.* The simplification of the FD characterization, Ref. [2] [Theorem 4.1], is based on the statement that the Gaussian process $\overline{Z}^n \in N(0, K_{\overline{Z}^n})$ in (4) has a zero optimal PSD.

The simplified FD characterization of feedback capacity [2] [Theorem 4.1] is further analyzed in subsequent papers, [21–23].

In 2019, Gattami [24] re-visited Kim's [2] [Theorem 4.1 and Theorem 6.1], i.e., the simplified FD and TD characterizations of feedback capacity, and presented alternative characterizations in slightly different form [24] [Theorems 1–4]. In particular, Ref. [24] [Theorem 1] presented an FD characterization of feedback capacity, in which the optimal PSD of the Gaussian process $\overline{Z}^n \in N(0, K_{\overline{Z}^n})$ in (4) corresponds to white noise, while [24] [Theorem 2] presented a TD characterization of $C_{n,FB}(\kappa, \mathbf{P}_{Y_1})$, expressed w.r.t. *a single matrix DRE of Kalman-filtering*. Moreover, Ref. [24] [Theorem 3] presented a TD characterization of feedback capacity $C_{FB}(\kappa, \mathbf{P}_{Y_1})$, expressed w.r.t. *a single matrix ARE of Kalman-filtering*,

but unlike [2] [Theorem 6.1], the covariance of the innovations part of the optimal input is strictly positive. Moreover, Ref. [24] [Theorem 4] presented a semi-definite programming (SDP) formulation of the optimization problem of $C_{FB}(\kappa, \mathbf{P}_{V_1})$ of [24] [Theorem 3], using linear matrix inequalities (LMI).

The reader may easily verify that the TD and FD characterizations of $C_{FB}(\kappa, \mathbf{P}_{V_1})$ in Kim [2] and Gattami [24], are fundamentally different, although both considered the same AGN channel and models of noise. Moreover, the TD characterizations of [2,24] do not impose Conditions 1 and 2 (at least explicitly), contrary to Yang, Kavcic and Tatikonda [20], who stated explicitly that their TD characterization of feedback capacity is based on Conditions 1 and 2. The confusion and misinterpretation of the TD characterizations of feedback capacity of [2,24] compared to [20] are first clarified in the arXiv paper [26,27] (we will discuss this shortly).

Recently, the comment paper by Derpich and Ostergaard [1] identified gaps in proofs of the TD and FD characterizations [2] [Theorem 4.1 and Theorem 6.1], and many other results in [2], which include Corollary 4.4, Theorems 4.6 and 5.3, Propositions 4.7 and 5.1, as well as Remarks 4.5 and 5.2, and Lemma 6.1. However, as stated in the Abstract and Conclusion in [1], the above results of [2] *may be valid, but a proof is currently missing.*

Sequential Information Theoretic Characterization of Feedback and Nonfeedback Capacity for Nonstationary and Nonergodic Noise. Equivalent TD sequential characterizations of the Cover and Pombra characterization $C_{FB,n}(\kappa, \mathbf{P}_{V_1})$ are presented in [27] [Theorem 1] (and earlier in [26]). Moreover, Ref. [27] [Theorem 3] presented a TD sequential characterization of $C_{FB,n}(\kappa, \mathbf{P}_{V_1})$, when the noise V^n is described by a nonstationary and nonergodic state space realization. A fundamental observation is that $C_{FB,n}(\kappa, \mathbf{P}_{V_1})$ in [27] [Theorem 3], is expressed in terms of *two matrix DREs of Kalman-filtering theory instead of one*, as previously believed in [2,20,24], irrespective of whether the noise is restricted to stationary or asymptotically stationary. This generalization is attributed to the fact that Conditions 1 and 2 are not imposed; hence, the encoder cannot determine the state of the noise from past channel inputs and output sequences or past noise and output sequences. However, when Conditions 1 and 2 are introduced, it is shown in [27] [Corollary 7, Section 2.5], that the characterizations of $C_{FB,n}$ simplify to the TD characterizations previously derived in [20] [Theorem 1]. Similarly, Ref. [27] [Theorem 7], showed that the asymptotic limit $C_{FB}(\kappa, \mathbf{P}_{Y_1}) = \lim_{n\longrightarrow} C_{n,FB}(\kappa, \mathbf{P}_{Y_1})$ is expressed in terms of *two matrix AREs of Kalman-filtering theory instead of one*, as previously obtained in [2,20,24]. However, in [27] [Corollary 12] it is shown that, if Conditions 1 and 2 are imposed, then under the conditions known as [28,29], (i) detectability and (ii) stabilizability (resp., (iii) unit circle controllability), $C_{FB}(\kappa, \mathbf{P}_{Y_1})$ reduces to the characterizations previously obtained in [2,20,24] (which did not impose (i) and (ii) (resp., (iii)). Another major observation is that (i) detectability and (ii) stabilizability (resp., (iii) unit circle controllability) are necessary and sufficient conditions for the convergence of matrix DREs to corresponding matrix AREs [28] and the existence of the asymptotic limit $C_{FB}(\kappa, \mathbf{P}_{Y_1}) = \lim_{n\longrightarrow} C_{n,FB}(\kappa, \mathbf{P}_{Y_1})$. Characterizations of $C_{FB,n}$ and C_{FB} for multiple-input multiple-output (MIMO) Gaussian channels are presented in [30], and follow directly from [26]. Preliminary results on an initial study of unstable autoregressive noise are announced in [31], when the initial state is known to the encoder and the decoder and [32]. For nonfeedback capacity we suggest [33].

1.4. Main Contributions

The main contributions of this paper are the following.

(1) The feedback capacity $C_{FB}(\kappa, \mathbf{P}_{V_1})$ for unstable noise, i.e., nonstationary and nonergodic, is higher compared to the feedback capacity of stable, i.e., stationary or asymptotically stationary noise.

(2) The feedback capacity $C_{FB}(\kappa, \mathbf{P}_{V_1})$ monotonically increases with respect to (w.r.t.) the unstable poles of the noise, while the total transmit power κ monotonically decreases, contrary to the water-filling nonfeedback capacity of stationary or asymptotically stationary noise.

(3) There are achievable lower bounds on nonfeedback capacity of unstable noise, which exhibit the properties of feedback capacity, i.e., the lower bounds monotonically increase w.r.t. the unstable pole of the noise, without increasing the total transmit power.

(4) We provide a proof that the feedback capacity results for stable noise previously presented in [2,20,24] do not hold for asymptotically stationary noise. On the contrary, the capacity formulas in [2,20,24] correspond to stationary noise, and it is not an ergodic capacity (i.e., it is not independent of the channel initial RVs).

This is due to subtle technical issues related to the asymptotic convergence of information theoretic characterizations of capacity, which are not sufficiently explained in [2,20,24], and often lead to misinterpretation of the results therein. We explain these issues throughout the paper, by clarifying the technical conditions and their various implications (i.e., Remark 2, Remark 9 and Remark 10).

(5) With respect to the comment paper [1], we show that the TD feedback capacity [2] [Theorem 6.1] is correct, *provided Conditions 1 and 2 hold, and the optimal innovations part of the channel input is asymptotically zero*. This is attributed to the technical issue of existence of feasible solution to the optimization problem that corresponds to the maximal and stabilizing solution of the algebraic Riccati equation of mean-square estimation. However, Ref. [2] [Theorem 6.1] is not equivalent to Cover and Pombra feedback capacity [3] for general stationary noise. Moreover, the claims in [2,24] that their asymptotic formulas hold for asymptotically stationary noise is not accurate; for their capacity formulas to be correct, the noise should be stationary (due to a technical condition related to asymptotic limits, via the maximal and stabilizing solution of the algebraic Riccati equation). Although in the paper we show using Karush–Kuhn–Tucker (KKT) conditions that for stable AR noise the optimal innovations part of the channel input is asymptotically zero, this does not mean the same property holds for [2] [Theorem 6.1] that corresponds to the general stable state space realization of the noise [2] [Equation (58)]. According to our analysis, Ref. [2] [Theorem 6.1] holds provided the optimal covariance of the innovations part of the channel input is asymptotically zero. This can be performed by addressing the optimization problem, i.e., whether a feasible solution exists and under what conditions, for example, by applying KKT conditions.

The rest of the paper is organized as follows.

In Section 2, we present the closed-form expressions of $C_{FB}(\kappa, \mathbf{P}_{V_1})$ for stable and unstable AR noise, numerical simulations and comparisons to expressions of feedback and nonfeedback capacity that appeared in the literature. We provide an extensive intuitive discussion on many of the technical aspects and their implications on the operational meaning of the capacity.

In Section 3, we present the derivations of asymptotic capacity $C_{FB}(\kappa, \mathbf{P}_{V_1})$ for stable and unstable AR noise, and lower bounds on asymptotic characterizations of nonfeedback capacity. In this section, we also review properties of convergence of difference Riccati equations to corresponding algebraic equations, which are essential in the analysis of asymptotic capacity $C_{FB}(\kappa, \mathbf{P}_{V_1})$. These properties verify our discussion and simulations of Section 2. In this section, we also provide corrections to the TD feedback capacity characterization of [2] [Theorem 6.1].

2. Main Results, Discussion, Simulations and Relations to the Literature

In this section we present the contributions listed in Section 1.4. The derivations are presented in Section 3.

2.1. Notation and Definitions

Throughout the paper, we use the following notation.

$\mathbb{Z} \stackrel{\triangle}{=} \{\ldots,-1,0,1,\ldots\}$, $\mathbb{Z}_+ \stackrel{\triangle}{=} \{1,\ldots\}, \mathbb{Z}_+^n \stackrel{\triangle}{=} \{1,2,\ldots,n\}$ where n is a finite positive integer.

$\mathbb{R} \stackrel{\triangle}{=} (-\infty,\infty)$, and $\mathbb{R}^m, m \in \mathbb{Z}_+$ is the vector space of tuples of the real numbers.

$\mathbb{R}^{n\times m}, (n,m) \in \mathbb{Z}_+ \times \mathbb{Z}_+$ is the set of n by m matrices with entries from the set of real numbers.

For a matrix $A \in \mathbb{R}^{n\times m}$, $(n,m) \in \mathbb{Z}_+ \times \mathbb{Z}_+$, its transpose is denoted by A^T; for $n = m$ the trace of A is denoted by $\mathrm{Trace}(A)$, and $A \succeq 0$ (resp. $A \succ 0$) denotes a symmetric positive semi-definite (resp., positive definite) matrix.

$\Big(\Omega, \mathscr{F}, \mathbb{P}\Big)$ denotes a probability space. Given a random variable (RV) $X : \Omega \to \mathbb{R}^{n_x}, n_x \in \mathbb{Z}_+^n$, its induced distribution on $\mathbb{R}^{n_x}$ is denoted by $\mathbf{P}_X$. $H(X)$ denotes the differential entropy of the RV X, and $I(X;Y)$ denotes the mutual information between RVs X and Y. $X \in N(\mu_X, K_X), K_X \succeq 0$ denotes a Gaussian RV X, with, mean value $\mu_X = \mathbf{E}\{X\}$ and covariance matrix $K_X = cov(X,X) \succeq 0$, defined by $K_X = cov(X,X) \stackrel{\triangle}{=} \mathbf{E}\Big\{\Big(X - \mathbf{E}\Big\{X\Big\}\Big)\Big(X - \mathbf{E}\Big\{X\Big\}\Big)^T\Big\}$. Given another Gaussian random variable $Y : \Omega \to \mathbb{R}^{n_y}, n_y \in \mathbb{Z}_+^n$, which is jointly Gaussian distributed with X, i.e., the joint distribution $\mathbf{P}_{X,Y}$ is Gaussian, the conditional covariance of X given Y, $K_{X|Y} = cov(X, X\Big|Y)$, is defined by

$$\begin{aligned} K_{X|Y} =& cov(X, X\Big|Y) \stackrel{\triangle}{=} \mathbf{E}\Big\{\Big(X - \mathbf{E}\Big\{X\Big|Y\Big\}\Big)\Big(X - \mathbf{E}\Big\{X\Big|Y\Big\}\Big)^T\Big|Y\Big\} \\ \stackrel{(a)}{=}& \mathbf{E}\Big\{\Big(X - \mathbf{E}\Big\{X\Big|Y\Big\}\Big)\Big(X - \mathbf{E}\Big\{X\Big|Y\Big\}\Big)^T\Big\} \end{aligned}$$

where (a) is due to a property of jointly Gaussian RVs.

Next, we introduce standard definitions of nonstationary, asymptotically stationary and stationary autoregressive Gaussian noise [28].

Definition 1. *(Time-varying, time-invariant stable/unstable AR unit memory Gaussian noise models)*

(a) A time-varying (nonstationary and nonergodic) Gaussian unit memory stable/unstable AR noise model, denoted by $AR(c_t), c_t \in (-\infty,\infty), \forall t$, is defined by

$$AR(c_t): \quad \begin{cases} V_t = c_t V_{t-1} + W_t, \quad t = 2,\ldots,n, \\ W_t \in N(0, K_{W_t}), \quad K_{W_t} \geq 0, \quad t = 2,\ldots,n, \quad \textit{indep. Gaussian}, \\ W_t,\ t = 2,\ldots,n \quad \textit{indep. of } V_1 \in N(0, K_{V_1}), K_{V_1} \geq 0, \\ c_t \in (-\infty,\infty), \quad t = 2,\ldots,n. \end{cases} \tag{9}$$

The time-varying $AR(c_t), c_t \in (-\infty,\infty), \forall t$ noise model is called asymptotically time-invariant if $\lim_{n\longrightarrow\infty} c_n = c \in (-\infty,\infty)$ and $\lim_{n\longrightarrow\infty} K_{W_n} = K_W \in (0,\infty)$.

(b) A time-invariant stable/unstable AR noise model denoted by $AR(c), c \in (-\infty,\infty)$ is the restriction of $AR(c_t)$ to $K_{W_t} = K_W > 0, t = 2,\ldots, c_t = c, t = 2,\ldots$.

(c) An unstable (resp., stable) time-invariant AR noise model is the restriction of $AR(c)$ to $c \notin (-1,1)$ (resp., $c \in (-1,1)$).

(d) The stable $AR(c), c \in (-1,1)$ noise model is called asymptotically stationary if the sequence $K_{V_n} \stackrel{\triangle}{=} \mathbf{E}(V_n)^2, n = 1,2,\ldots$ is such that $\lim_{n\longrightarrow\infty} K_{V_n} = K_V \in (0,\infty), \forall K_W > 0, \forall K_{V_1} \geq 0$, and it is called stationary if $K_{V_n} = K_{V_1}$ for all $n = 1,2,\ldots$.

Remark 1. *We recall (see [28]) some known facts.*

Given an $AR(c_t), c_t \in (-\infty,\infty), \forall t$ *noise the value* c_n *is a measure of correlation between the samples* V_n *and* V_{n-1}*, i.e.,* $\mathbb{E}\{V_nV_{n-1}\} = c_nK_{V_{n-1}}, K_{V_{n-1}} \stackrel{\triangle}{=} \mathbb{E}\{(V_{n-1})^2\}$*, and the variance satisfies the recursion,* $K_{V_n} = c_n^2K_{V_{n-1}} + K_{W_n}, K_{V_1} \geq 0, n = 2,3\ldots$.

For an $AR(c)$ *noise,* $K_{V_n} = c^2K_{V_{n-1}} + K_W, K_W > 0, K_{V_1} \geq 0, n = 2,\ldots,n$*. The* $AR(c)$ *is asymptotically stationary, if and only if* $c \in (-1,1)$ *and* $\lim_{n\longrightarrow\infty} K_{V_n} = \frac{K_W}{1-c^2}, \forall K_W > 0, \forall K_{V_1} \geq 0$*. The* $AR(c)$ *is stationary, if and only if* $c \in (-1,1)$ *and* $K_{V_t} = K_{V_1} = \frac{K_W}{1-c^2}$ *for all* $t = 1,2,\ldots$, $c \in (-1,1), K_W > 0$*. Hence, the standard definition [28] of an asymptotically stationary noise does not require* $K_{V_1} > 0$*;* $K_{V_1} = 0$ *means the RV has zero variance, and hence* $V_1 = v_1$ *takes a specific nonrandom value.*

2.2. Closed-Form Feedback Capacity and Nonfeedback Capacity Lower Bounds for Stable/Unstable AR Noise

In Section 3, we show that the characterization of feedback capacity, $C_{FB}(\kappa, \mathbf{P}_{V_1})$ for asymptotically time-invariant $\mathrm{AR}(c_t), \forall c_t \in (-\infty,\infty)$, $\lim_{n\longrightarrow\infty} c_n = c \in (-\infty,\infty), \lim_{n\longrightarrow\infty} K_{W_n} = K_W > 0$, is given by the optimization problem,

$$C_{FB}(\kappa, \mathbf{P}_{V_1}) \stackrel{\triangle}{=} \sup_{\Lambda\in(-\infty,\infty), K_Z\geq 0} \frac{1}{2}\log\Big(\frac{(\Lambda+c)^2K + K_Z + K_W}{K_W}\Big) \tag{10}$$

$$\text{such that (s.t.)} \quad \Lambda^2K + K_Z \leq \kappa, \quad \Lambda \neq -c, \quad \forall c \in (-\infty,\infty), \quad \forall K_W > 0, \tag{11}$$

$$K = c^2K + K_W - \frac{\Big(K_W + cK(\Lambda+c)\Big)^2}{\Big(K_Z + K_W + (\Lambda+c)^2K\Big)}, \quad \text{(Algebraic Riccati eqn)}, \tag{12}$$

$$K \geq 0 \text{ is a stabilizing solution equivalent to } |F^{CL}(K,\Lambda,K_Z)| < 1, \tag{13}$$

$$F^{CL}(K,\Lambda,K_Z) \stackrel{\triangle}{=} c - M(K,\Lambda,K_Z)\Big(\Lambda+c\Big), \quad M(K,\Lambda,K_Z) \stackrel{\triangle}{=} \frac{K_W + cK\Big(\Lambda+c\Big)}{K_Z + K_W + \Big(\Lambda+c\Big)^2K}, \tag{14}$$

$$K = \lim_{n\longrightarrow\infty} K_n, \quad K_n \stackrel{\triangle}{=} \mathbf{E}\Big\{\Big(V_n - \mathbf{E}\Big\{V_n\Big|Y^n\Big\}\Big)^2\Big\}, \; K_n, n = 1,2,\ldots \text{ satisfies recursion (39)}. \tag{15}$$

According to the theory of Riccati equations (presented in Section 3) , the constraint $|F^{CL}(K,\Lambda,K_Z)| < 1$ is necessary and sufficient for existence of capacity $C_{FB}(\kappa, \mathbf{P}_{V_1})$ as an asymptotic limit; this is equivalent to the conditions of detectability and unit circle controllability [28,29]. However, these conditions do not imply uniqueness of non-negative solutions of Riccati equations $K \geq 0$; uniqueness requires the stronger conditions of detectability and stabilizability (instead of unit circle controllability).

We show that $C_{FB}(\kappa, \mathbf{P}_{V_1}) \in (0,\infty)$ is achieved by a channel input parametrized by the optimal values $(\Lambda, K_Z, K) = (\Lambda^*, K_Z^*, K^*)$ of the above optimization problem, as follows.

$$X_n = \Lambda_n^*\Big(V_{n-1} - \mathbf{E}\Big\{V_{n-1}\Big|Y^{n-1}\Big\}\Big) + Z_n, \quad X_1 = Z_1, \quad Z_n \in N(0, K_{Z_n}^*), \quad n = 1,2,\ldots, \tag{16}$$

$$\text{s.t.} \quad \lim_{n\longrightarrow\infty}(\Lambda_n^*, K_{Z_n}^*) = (\Lambda^*, K_Z^*), \quad K_Z^* \geq 0,$$

$$(\Lambda^*, K_Z^*, K^*) \text{ the optimal values of } C_{FB}(\mathbf{P}_{V_1}, \kappa) = C_{FB}(\kappa), \forall \mathbf{P}_{V_1}, \forall K_{V_1} \geq 0.$$

Thus, condition $|F^{CL}(K,\Lambda,K_Z)| < 1$ is necessary and sufficient for the asymptotic convergence of the estimation error $E_n \stackrel{\triangle}{=} V_n - \mathbf{E}\Big\{V_n\Big|Y^n\Big\}, n = 1,2,\ldots$ in a mean-square sense.

Remark 2. *As pointed out in comment paper [1], the feedback capacity* $C_{FB}(\kappa, \mathbf{P}_{V_1})$ *in [2] [Theorem 6.1] for stable noise realization, claimed without a valid proof that* $K_Z = K_Z^* = 0$ *is*

optimal. Thus, for the stable AR noise, i.e., $c \in (-1,1)$, Ref. [2] [Theorem 6.1] corresponds to $C_{FB}(\kappa, \mathbf{P}_{V_1})$ with $K_Z = 0$ (but a proof is missing [1]). If $K_Z = 0$ is assumed then $C_{FB}(\kappa, \mathbf{P}_{V_1})$ is very easy to compute. Our analysis using KKT conditions and convergence properties of Riccati equations illustrates that for $c \in (-1,1)$, the optimal is $K_Z = K_Z^* = 0$ because there is no feasible solution with $K_Z > 0$ of the optimization problem $C_{FB}(\kappa, \mathbf{P}_{V_1})$. Moreover, there are different Regimes of capacity. Further, we also identify conditions for the validity of [2] [Theorem 6.1] for general noise in state space form.

Upon applying the KKT conditions, we show the solution of the optimization $C_{FB}(\kappa, \mathbf{P}_{V_1})$ consists of multiple Regimes, which depend on the values (c, K_W, κ), as follows.

2.2.1. Regime 1: Ergodic Feedback Capacity for Unstable Noise (Theorem 4)

Regime 1. $\forall c^2 \in (1,\infty), \forall \kappa > \kappa_{min} \stackrel{\triangle}{=} \frac{K_W\left(1+\sqrt{4c^2-3}\right)}{2\left(c^2-1\right)^2}$.

The optimal values $(\Lambda, K_Z, K) = (\Lambda^*, K_Z^*, K^*)$ of optimization problem $C_{FB}(\kappa, \mathbf{P}_{V_1})$ in (10) are

$$\Lambda^* = \frac{cK_W}{\kappa\left(c^2-1\right)^2 - K_W} \in (-\infty,\infty), \quad K_Z^* = \frac{\kappa\left(c^2-1\right)\left(\kappa\left(c^2-1\right)^2 - K_W\right) - K_W^2}{\left(c^2-1\right)\left(\kappa\left(c^2-1\right)^2 - K_W\right)} \in (0,\infty), \tag{17}$$

$$K^* = \frac{\kappa\left(c^2-1\right)^2 - K_W}{c^2\left(c^2-1\right)} \in (0,\infty), \quad \left(\Lambda^{\infty,*}\right)^2 + K_Z^* = \kappa. \tag{18}$$

The feedback capacity $C_{FB}(\kappa, \mathbf{P}_{V_1}) \in (0,\infty)$ is an ergodic capacity given by

$$C_{FB}(\kappa, \mathbf{P}_{V_1}) = C_{FB}(\kappa) \stackrel{\triangle}{=} \frac{1}{2}\log\left(\frac{\left(\Lambda^*+c\right)^2 K^* + K_Z^* + K_W}{K_W}\right), \ \forall \mathbf{P}_{V_1}, \forall K_{V_1} \geq 0 \tag{19}$$

$$= \frac{1}{2}\log\left(\frac{c^2\left(\left(c^2-1\right)\kappa + K_W\right)}{\left(c^2-1\right)K_W}\right) \tag{20}$$

$$= \log|c| + \frac{1}{2}\log\left(\frac{1}{c^2-1} + \frac{\kappa}{K_W}\right) \geq \log|c|, \tag{21}$$

$$C_{FB}(\kappa) \simeq \log|c| + \frac{1}{2}\log\gamma, \quad \text{for large } \gamma \stackrel{\triangle}{=} \frac{\kappa}{K_W}, \quad \forall 1 < c^2 < \infty. \tag{22}$$

We also emphasize that for ergodic capacity $C_{FB}(\kappa, \mathbf{P}_{V_1}) = C_{FB}(\kappa), \forall \mathbf{P}_{V_1}$, i.e., it does not depend on the specific distribution of the RV V_1 (see Section 3.3).

Moreover, $C_{FB}(\kappa) \in (0,\infty)$ is achieved by the asymptotically stationary and ergodic optimal input,

$$X_n = \Lambda_n^*\left(V_{n-1} - \mathbf{E}\left\{V_{n-1}\middle|Y^{n-1}\right\}\right) + Z_n, \quad X_1 = Z_1, \quad Z_n \in N(0, K_{Z_n}^*), \quad n = 1,2,\ldots, \tag{23}$$

$$\text{s.t.} \quad \lim_{n\longrightarrow\infty}(\Lambda_n^*, K_{Z_n}^*, K_n^*) = (\Lambda^*, K_Z^*, K^*), \quad K_Z^* > 0, \quad K^* > 0. \tag{24}$$

The optimal parameters of the input (Λ^*, K_Z^*) are such that $K_Z^* > 0$, which imply there is one and only one limiting solution $K = K^* = \lim_{n\longrightarrow\infty} K_n^* \geq 0$ (where K_n^* satisfies recursion (39)) of the algebraic Riccati equation (ARE) (12) which is positive and stabilizing, i.e., $K^* > 0$ and $|F^{CL}(K^*, \Lambda^*, K_Z^*)| < 1$. Moreover, using the theory of Riccati equations (Theorem 2), we deduce that $K_Z^* > 0$ is a necessary and sufficient condition for the feedback capacity to correspond to an ergodic capacity, i.e., $\lim_{n\longrightarrow\infty} C_{n,FB}(\kappa, \mathbf{P}_{V_1}) = C_{FB}(\kappa, \mathbf{P}_{V_1}) = C_{FB}(\kappa), \forall \mathbf{P}_{V_1}, \forall K_{V_1} \geq 0$, that is, the limit does not depend on $\mathbf{P}_{V_1}$, i.e., the convergence is uniform $\forall \mathbf{P}_{V_1}$.

It is well-known (see, Gallager [4]) that for channels with memory there are fundamental differences between ergodic capacity in which the decoding error probability is independent of initial channel states (i.e., V_1 in our case and the value of K_{V_1}) and nonergodic capacity in which the decoding error probability depends on initial channel states (i.e., the value of K_{V_1} in our case).

Comparison to Memoryless Noise. The condition for Regime 1 is alternatively expressed as $|c| > 1$, $\gamma > \gamma_{min} \triangleq \frac{1+\sqrt{4c^2-3}}{2(c^2-1)^2}$ where $\gamma \triangleq \kappa / K_W$ is the signal-to-noise ration (SNR) of the memoryless noise (i.e., corresponding to $c = 0$), and the approximation holds $\gamma_{min} \approx \frac{1}{|c|^3}$ for $|c| \gg 1$. Note that γ_{min} approaches 0 as $|c|$ increases. The feedback capacity in this regime is also expressed as

$$C_{FB}(\kappa) = \log|c| + \frac{1}{2}\log\left(\frac{1}{c^2-1} + \gamma\right) > \frac{1}{2}\log(1+\gamma) \equiv C_{memoryless}$$

where $C_{memoryless}$ is the capacity of the memoryless channel, i.e., corresponding to $c = 0$. Note that for strong memory $|c| \gg 1$; by (20) the capacity can be approximated as

$$C_{FB}(\kappa) \approx \frac{1}{2}\log\left(1 + c^2\gamma\right) \quad \text{for} \quad |c| \gg 1$$

i.e., the combined effect of memory and feedback is to introduce gain of c^2 in the memoryless SNR, $\gamma \triangleq \kappa / K_W$, and this gain can be substantial for large c^2. This shows significant benefits for strong memory noise combined with feedback communication.

By the above expressions the capacity $C_{FB}(\kappa)$ is increased by strong memory combined with feedback and this increase can also be characterized as follows:

$$\Delta C(\kappa) \triangleq C_{FB}(\kappa) - C_{memoryless}(\kappa) = \log|c| + \frac{1}{2}\log\frac{\frac{1}{c^2-1} + \gamma}{1+\gamma} > 0$$

Further note that for sufficiently large $\gamma \gg \max\{1, (c^2-1)^{-1}\}$, this becomes

$$\Delta C(\kappa) \approx \log|c|$$

i.e., $\Delta C(\kappa)$ increases with memory as measured by $|c| > 1$, and $\Delta C(\kappa)$ can be large for large $|c|$. That is, we observe again that strong memory combined with feedback brings significant benefits.

Finally, we note that for unstable noise ($|c| > 1$) with $K_W > 0$ or/and $K_{V_1} > 0$, the average noise power $\overline{K}_V$ is (asymptotically) infinite, $\overline{K}_V \triangleq \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} K_{V_n} = \infty$. Yet, $C_{FB}(\kappa) > 0$, i.e., reliable communication with non-zero rate is possible. This should be contrasted with the memoryless channel, where infinite average noise power implies zero capacity, i.e. reliable communication with non-zero rate is impossible.

2.2.2. Regimes 2 and 3: Nonergodic Feedback Capacity for Stable/Unstable Noise-Complement of Regime 1 (Theorem 5)

For the complement of Regime 1, we deduce from the KKT conditions that $C_{FB}(\kappa, \mathbf{P}_{V_1}) \in (0, \infty)$ if and only if the variance $K^*_{Z_n}, n = 1, 2, \ldots$ is asymptotically zero, i.e., $\lim_{n\longrightarrow\infty} K^*_{Z_n} = K^*_Z = 0$, giving rise to the feedback capacity of Regimes 2, 3 below.

Regime 2. $\forall c^2 \in (1, \infty), \forall \kappa \leq \kappa_{min} \triangleq \frac{K_W\left(1+\sqrt{4c^2-3}\right)}{2\left(c^2-1\right)^2}$, and

Regime 3. $\forall c \in [-1, 1], \forall \kappa \in [0, \infty)$.

The feedback capacity $C_{FB}(\kappa, \mathbf{P}_{V_1}) \in (0, \infty)$ and the corresponding optimal values $(\Lambda, K_Z, K) = (\Lambda^*, K^*_Z, K^*)$ of the optimization problem in (10) are

$$C_{FB}(\kappa, \mathbf{P}_{V_1}) = C_{FB}^{max}(\kappa, \mathbf{P}_{V_1}) \stackrel{\triangle}{=} \log \max\{1, |\Lambda^*|\}, \quad \forall K_{V_1} > 0, \tag{25}$$

$$K_Z^* = 0, \quad |\Lambda^*| > 1, \qquad K^* = K_+^*(\Lambda^*) = \frac{K_W\Big((\Lambda^*)^2 - 1\Big)}{(\Lambda^* + c)^2} \in (0, \infty), \tag{26}$$

$$\text{where } K_+^*(\Lambda^*) \text{ is the maximal stab. sol., } |F^{CL}(K^*, \Lambda^*, K_Z^* = 0) = -\tfrac{1}{\Lambda^*}| < 1 \text{ for } |\Lambda^*| > 1, \tag{27}$$

$$\Lambda^* \text{ the maximal root of quartic eqn, } \; K_W(\Lambda^*)^4 - (K_W + \kappa)(\Lambda^*)^2 - 2c\kappa\Lambda^* - c^2\kappa = 0. \tag{28}$$

We note that from the ARE (12), when $K_Z^* = 0$ there is also a zero solution $K^*(\Lambda^{\infty,*}) = 0$ such that $|\Lambda^*| < 1$, but this is discarded because it implies $C_{FB}(\kappa, \mathbf{P}_{V_1}) = 0$.

Moreover, (using properties of Riccati equations of Theorem 2), $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1}) \in (0, \infty)$ is achieved by the optimal input,

$$X_n = \Lambda_n^*\Big(V_{n-1} - \mathbf{E}\Big\{V_{n-1}\Big|Y^{n-1}\Big\}\Big), \quad n = 2, 3, \ldots, X_1 = Z_1, \quad Z_1 \in N(0, K_{Z_1}^*), \; K_{Z_1}^* > 0, \tag{29}$$

$$\text{s.t. } \lim_{n\longrightarrow\infty}(\Lambda_n^*, K_{Z_n}^*, K_n^*) = (\Lambda^*, K_Z^*, K^*), \; K_Z^* = 0, \; K^* = K_+^* > 0 \text{ the maximal sol. of (12)}. \tag{30}$$

2.2.3. Discussion: Comparison of Regimes 1, 2, 3

Now, we make some observations. We clarify the gaps in the proof of the TD characterizations of feedback capacity in [2] [Theorem 6.1], which are identified in the comment paper [1] but not resolved. We also explain the capacity gains of Regime 1 compared to Regimes 2, 3 (Figure 2).

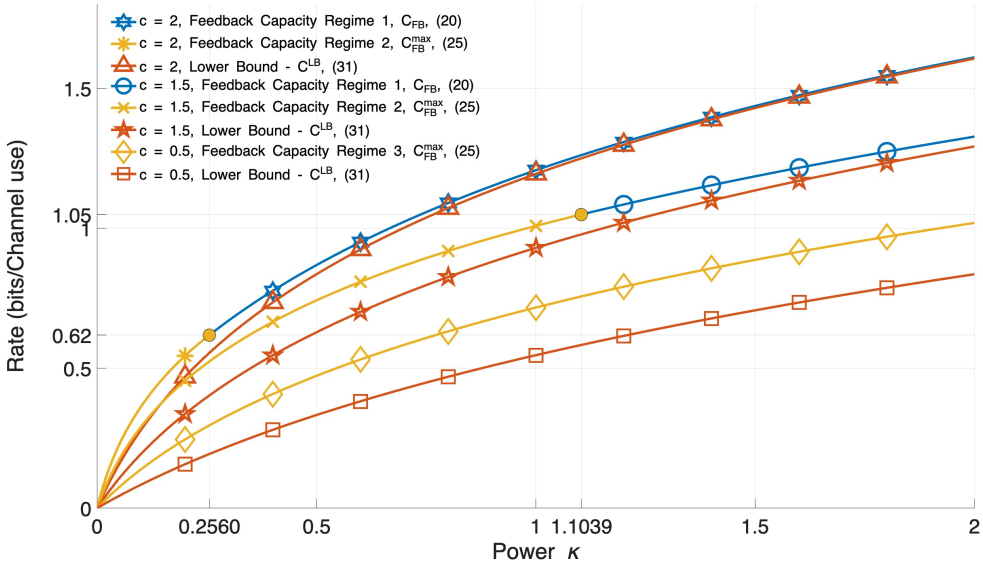


Figure 2. Comparisons of Regime 1, $C_{FB}(\kappa)$, Regimes 2, 3, $C_{FB}^{max}(\mathbf{P}_{V_1}, \kappa) = \log|\Lambda^*|$, and lower bound on nonfeedback capacity $C^{LB}(\kappa)$, for $c = 0.5$, $c = 1.5$, $c = 2$ and $K_W = 1$, as a function of transmit power κ.

(1) For stable noise, i.e., $c \in (-1, 1)$, Ref. [2] [Theorem 6.1] *does not give the capacity for asymptotically stationary noise and is not the ergodic capacity, because* $K_{V_1} = 0$ *should be excluded, unlike Regime 1*. This is contrary to previous beliefs in [2,24], who interpreted [2] [Theorem 6.1] as the feedback capacity of asymptotically stationary noise.

This follows directly from expression $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1}) =$ (25) of Regime 3 with $c \in (-1, 1)$, because this formula is correct provided the condition $K_{V_1} > 0$ holds. However, the condition $K_{V_1} > 0$ contradicts the standard definition of asymptotically stationary noise,

i.e., Definition 1.(d). Moreover, since $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1}) =$ (25) requires $K_{V_1} > 0$, and it follows from our derivations that $K_{V_1} = 0$ implies the asymptotic limit is zero, i.e., $\lim_{n\longrightarrow\infty} C_{n,FB}(\kappa, \mathbf{P}_{V_1}) = 0$, then it depends on the value of K_{V_1} (unlike Regime 1); hence, it is not an ergodic capacity. In fact, the proof of Theorem 5 shows that, when channel input (29) is used and $K_{V_1} = 0$, then the limiting mean-square error converges to zero, i.e., $K = K_n \stackrel{\triangle}{=} \lim_{n\longrightarrow\infty} \mathbf{E}\Big\{\big(V_n - \mathbf{E}\big\{V_n\big|Y^n\big\}\big)^2\Big\} = 0$, ($K_n$ satisfies recursion (39)) and necessarily implies $|\Lambda^*| < 1$ and $C_{FB}(\kappa, \mathbf{P}_{V_1}) = 0$; hence, we require $K_{V_1} > 0$ as stated in (25).

(2) For the stable state space realization of Gaussian noise V^n, which is considered in [2,24], i.e., $S_{t+1} = AS_t + GW_t, V_t = CS_t + W_t, t = 1, \ldots, n$, with stable A, the TD characterizations of feedback capacity in [2] [Theorem 6.1] should have explicitly stated that Conditions 1 and 2 of Yang, Kavcic and Tatikoda [20] [Section II.C, I–III] hold. Moreover, Ref. [2] [Theorem 6.1] did not provide a valid proof (see also comment paper [1]) to support the claim that asymptotically the optimal value is $K_Z = K_Z^* = 0$. A necessary and sufficient condition for such $K_Z = K_Z^* = 0$ is that the matrix DRE converges to the maximal and stabilizing solution $K = K_+$, corresponding to an asymptotic optimal value of $K_Z^* = 0$. This is precisely the reason that for Regimes 2 and 3, the optimal value is $K_Z^* = 0$. Moreover, $K_Z^* = 0$ turned out to be optimal, because for Regimes 2 and 3, which include stable $c \in (-1, 1)$, there is no feasible unique and stabilizing solution of the corresponding matrix ARE, as in Regime 1 with $K_Z^* > 0$, that will ensure positive asymptotic value of capacity. Therefore, Ref. [2] [Theorem 6.1] should be corrected by including $K_Z \geq 0$ in all equations. Then we apply KKT conditions to determine whether the optimal value is $K_Z = K_Z^* = 0$ ([2] [Theorem 6.1] is restricted to A with stable eigenvalues).

(3) The fundamental differences between the feedback capacities of Regime 1 and Regimes 2 and 3, are the following.

For Regime 1, the limit $C_{FB}(\kappa, \mathbf{P}_{V_1}) = \lim_{n\longrightarrow\infty} C_{n,FB}(\kappa, \mathbf{P}_{V_1}) = C_{FB}(\kappa) \in (0, \infty)$ is independent of $\mathbf{P}_{V_1}, K_{V_1} \geq 0$, i.e., the convergence is uniform $\forall \mathbf{P}_{V_1}, \forall K_{V_1} \geq 0$ and the asymptotic variance of the innovations part of the channel input is strictly positive, i.e., $K_Z^* > 0$. The optimal value $K_Z^* > 0$ implies there is a unique optimal value of the asymptotic mean-square error $K^* > 0$. This is attributed to the optimal solution (Λ^*, K_Z^*), which implies that the conditions of detectability and stabilizability hold. Stabilizability implies $K_Z^* > 0$.

For Regimes 2 and 3, the limit $C_{FB}(\kappa, \mathbf{P}_{V_1}) = \lim_{n\longrightarrow\infty} C_{n,FB}(\kappa, \mathbf{P}_{V_1}) = C_{FB}^{max}(\kappa, \mathbf{P}_{V_1}) \in (0, \infty)$ presupposes $K_{V_1} > 0$ and the asymptotic variance of the innovations part of the channel input is zero, i.e., $K_Z^* = 0$. This is attributed to the fact there is no feasible solution (Λ^*, K_Z^*) to the optimization problem with positive capacity when $K_Z^* > 0$. To ensure positive capacity it is necessary that $K_Z^* = 0$; this implies the condition of unit circle controllability condition holds but the condition of Stabilizability does not hold (this requires $K_Z^* > 0$).

The optimal value $K_Z^* = 0$ implies (as easily verified by the quadratic Equation (12)) there are two limiting optimal values of the asymptotic mean-square error $K^* \geq 0$, where $K^* = K_+^* > 0$ is the maximal limiting value corresponding to $C_{FB}(\kappa, \mathbf{P}_{V_1}) = C_{FB}^{max}(\kappa, \mathbf{P}_{V_1}) \in (0, \infty), \forall K_{V_1} > 0$, and $K^* = 0$ is the minimum non-negative limiting value corresponding to $C_{FB}(\kappa, \mathbf{P}_{V_1}) = 0, \forall K_{V_1} \geq 0$.

The main point of the proof of Theorem 5 is that, if $K_{V_1} = 0$, then channel input (29) and (30) implies $C_{FB}(\kappa, \mathbf{P}_{V_1}) = \lim_{n\longrightarrow\infty} C_{n,FB}(\kappa, \mathbf{P}_{V_1}) = 0$.

(4) The capacity gains of Regime 1 compared to Regimes 2 and 3 are attributed to the following two factors.

(a) Unstable noise, i.e., $|c| > 1$ exhibits strong memory dependence, compared to stable noise, i.e., $|c| < 1$. This follows from Remark 1; the value c is a measure of correlation between the samples V_n and V_{n-1}, i.e., $\mathbb{E}\big\{V_n V_{n-1}\big\} = cK_{V_{n-1}}, K_V \stackrel{\triangle}{=} \mathbb{E}\big\{(V_{n-1})^2\big\}$.

As $|c|$ becomes large, $\mathbb{E}\{V_nV_{n-1}\} = cK_{V_{n-1}}$ becomes large in absolute value. On the other hand, as $|c|$ goes to zero, $\mathbb{E}\{V_nV_{n-1}\} = cK_{V_{n-1}}$ goes to zero, and the channel behaves as a memoryless channel $Y_n \sim X_n + W_n$. It is known channels with memory have higher capacity compared to memoryless channels.

(b) Optimal channel inputs should generate new information at each time instant; inputs that make use of an innovations process generate new information.

For Regime 1, the optimal input includes an innovations part, i.e., Z^n, and it is not asymptotically zero, i.e., $K_Z^* > 0$, while for Regimes 2 and 3, the innovations part of the optimal channel input is asymptotically zero, i.e., $K_Z^* = 0$.

For Regime 1, we have $|c| > 1$ and $K_Z^* > 0$, and this explains the reason that the capacity of Regime 1 is much higher than the capacity of Regimes 2, 3, as $|c|$ becomes large.

In fact, our simulations (see Figure 2) show that as $|c|$ increases in the respective regions of Regimes 1, 2, 3, "capacity of Regime 1 $\geq$ capacity of Regime 2 $\geq$ capacity of Regime 3".

In Section 2.3, we present numerical comparisons of $C_{FB}(\kappa)$, $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1})$.

In Section 2.5, we confirm the values of the closed-form expressions of $C_{FB}(\kappa, \mathbf{P}_{V_1})$ for Regimes 1–3, by solving the optimization problem (10) numerically, using a semi-definite programming (SDP) formulation as in [24].

In Section 3.2, we confirm the behavior of the asymptotic limits from the well-known properties of difference Riccati equations of filtering theory, for the convergence of the sequence $K_n \geq 0, n = 1, 2, \ldots$ to $K \geq 0$. It follows that the convergence of $K_{Z_n}, n = 1, 2, \ldots$ to $K_Z > 0$ is a necessary and sufficient condition for convergence of sequence $K_n \geq 0, n = 1, 2, \ldots$ for all initial conditions $K_1 \geq 0, K_{V_1} \geq 0$ to a unique limit $K \geq 0$, and for $\lim_{n\longrightarrow\infty} C_{n,FB}(\kappa, \mathbf{P}_{V_1}) = C_{FB}(\kappa, \mathbf{P}_{V_1}) = C_{FB}(\kappa), \forall \mathbf{P}_{V_1}, \forall K_{V_1} \geq 0$, i.e., the limit is independent of $\mathbf{P}_{V_1}$. This is valid only for feasible Region 1.

2.2.4. Lower Bounds: Lower Bounds on Ergodic Nonfeedback Capacity for Stable/Unstable Noise

In Theorem 7.(2), for Regimes 1–3, i.e., $\forall c \in (-\infty, \infty), \forall \kappa \in [0, \infty)$, we derive the non-trivial lower bound on nonfeedback capacity corresponding to an independent input, $X_1, X_2, \ldots$ given by

$$C_{FB}(\kappa, \mathbf{P}_{V_1}) \geq C(\kappa, \mathbf{P}_{V_1}) \geq C^{LB}(\kappa),$$

$$C^{LB}(\kappa) \stackrel{\triangle}{=} \frac{1}{2} \log \Big(\frac{\kappa\Big(1 + c^2\Big) + K_W + \sqrt{\Big(\kappa\big(1 - c^2\big) + K_W\Big)^2 + 4c^2 K_W \kappa}}{2K_W} \Big), \ \forall \mathbf{P}_{V_1}, K_{V_1} \geq 0, \tag{31}$$

and achieved by (23) with

$$\lim_{n\longrightarrow\infty} \Lambda_n^* = 0, \text{ i.e., } X_n = Z_n \in N(0, K_{Z_n}), \quad \lim_{n\longrightarrow\infty} K_{Z_n}^* = \kappa \tag{32}$$

where $C^{LB}(\kappa)$ is independent of the distribution of V_1, i.e., it holds $\forall K_{V_1} \geq 0$, and hence it is an *ergodic achievable rate*.

Numerical simulations show that the lower bound on nonfeedback capacity $C^{LB}(\kappa)$ for $c \notin (-1, 1)$ is higher than the feedback capacity $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1})$ for $c \in (-1, 1)$. This is consistent with our discussion in Section 2.2.3 under (4). For unstable $c \notin (-1, 1)$, we are not aware of any literature that computed nonfeedback capacity or bounds.

2.3. Numerical Comparisons of $C_{FB}(\kappa)$, $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1})$ and Lower Bound $C^{LB}(\kappa)$ for Stable/Unstable Noise

Figures 2 and 3 illustrate the performance of Regime 1, $C_{FB}(\kappa, \mathbf{P}_{V_1}) = C_{FB}(\kappa), \forall K_{V_1} \geq 0$, Regimes 2, 3, $C_{FB}(\kappa, \mathbf{P}_{V_1}) = C_{FB}^{max}(\mathbf{P}_{V_1}, \kappa) = \log \max\{1, |\Lambda^*|\}, \forall K_{V_1} > 0$, and lower bound

on nonfeedback capacity for Regimes 1-3, $C(\kappa, \mathbf{P}_{V_1}) \geq C^{LB}(\kappa), \forall K_{V_1} \geq 0$, for values $c = 0.5$, $c = 1.5$, $c = 2$ and $K_W = 1$, as a function of transmit power κ.

Some of the conclusions are the following.

(1) Regime 1, $C_{FB}(\kappa), \forall K_{V_1} \geq 0$ is always higher than Regimes 2, 3,

$C_{FB}^{max}(\mathbf{P}_{V_1}, \kappa) = \log \max\{1, |\Lambda^*|\}, \forall K_{V_1} > 0$, and the lower bound $C^{LB}(\kappa)$. This surprising feature is fundamentally different than the behavior of the Shannon capacity $C^{Shannon}$ that is limited by the signal-to-noise ratio.

(2) Regime 1, $C_{FB}(\kappa)$ is an increasing function of the parameter $|c| \in (1, \infty)$, that is, the more unstable the AR(c) noise, the higher the value of $C_{FB}(\kappa)$, while κ_{max} is reduced, i.e., as $|c| \in (1, \infty)$ increases, higher rates are achieved at reduced transmit power κ.

(3) The lower bound $C^{LB}(\kappa)$ approaches Regime 1, $C_{FB}(\kappa)$, for larger values of $|c| \in (1, \infty)$. This is also a surprising result of unstable, not encountered for stable noise.

(4) Regimes 2 and 3, $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1})$, cannot be optimal for Regime 1 because the broken lines of Figure 3 are below Regime 1, $C_{FB}(\kappa)$.

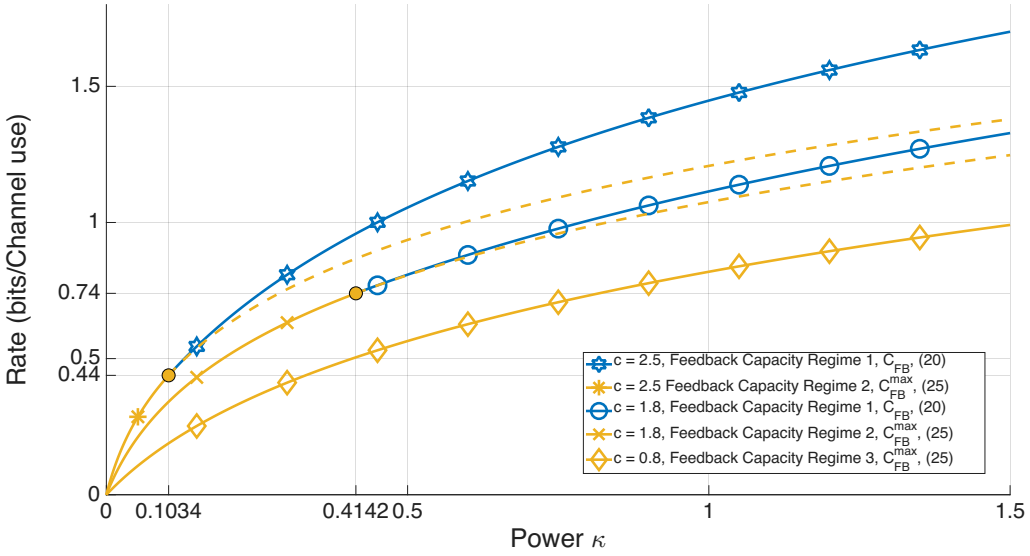


Figure 3. Illustration that Regimes 2 and 3 $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1}) = \log|\Lambda^*|, |\Lambda^*| > 1$ are not optimal compared to Regime 1.

Figure 4 illustrates the tightness of the lower bound on nonfeedback capacity, $C_{LB}^{LB}(\kappa)$ given by (31), and incurred by an IID channel input $Z_t \in N(0, \kappa)$, $K_Z^* = \kappa$, to the optimal water-filling solution of nonfeedback capacity (recall that the water-filling characterization is based on the use of time-invariant or asymptotically time-invariant channel input strategies); Equation (5.514) in Ihara [5] (see also Equation (6) in Butman [10]), given by

$$C(\kappa) = \frac{1}{2}\log\left(1 + \kappa + \frac{c^2}{1-c^2}\right), \quad \kappa > \left(\frac{1}{(1-|c|)^2} - \frac{1}{(1-c^2)}\right), \quad c \in (-1,1), \quad K_W = 1. \tag{33}$$

Contrary to the nonfeedback lower bound Formula (31), which holds for all stable and unstable AR$(c), \forall c \in (-\infty, \infty)$ noise and $\forall \kappa \in [0, \infty)$, the closed form nonfeedback Formula (33) based on water-filling, is restricted to $\kappa > \left(\frac{1}{(1-|c|)^2} - \frac{1}{(1-c^2)}\right)$, and to the stable AR$(c), \forall c \in (-1, 1)$. The maximum difference $C(\kappa) - C^{LB}(\kappa)$, when $K_W = 1$ occurs at $c = 0.75$, and is less than 1.5×10^{-2} bits per channel use.

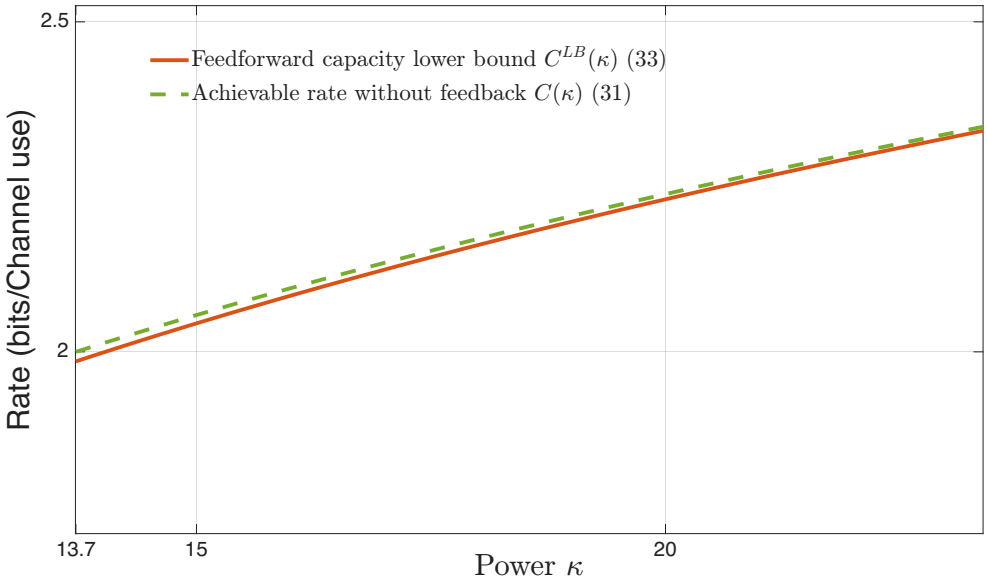


Figure 4. AR(c), noise, $c = 0.75$ and $K_W = 1$. Comparison of nonfeedback capacity $C(\kappa)$ of water-filling Formula (33) (see Equation (5.5.14) in [5] and Equation (6) in [10]), and lower bound $C^{LB}(\kappa) =$ (31) of transmitting an IID channel input $Z_t \in N(0, \kappa)$ for $\kappa \in [\kappa_{min}, \infty)$, where $\kappa_{min} = 13.7$ is the minimum threshold (the values correspond to the maximum difference).

2.4. Time-Sharing Increases Achievable Rates

Figure 5 considers the values $c = 1.5, K_W = 1$ to illustrate the rate of a time-sharing scheme between Regime 1, $C_{FB}(\kappa), \forall \kappa > \kappa_{min} \stackrel{\triangle}{=} \frac{K_W\left(1+\sqrt{4c^2-3}\right)}{2\left(c^2-1\right)^2}$, and the lower bound $C^{LB}(\kappa), \forall \kappa \in [0, \infty)$ on the asymptotic characterization of nonfeedback capacity. This scheme results in higher rates, because it employs a time-varying channel input strategy, that is, two different strategies—one without feedback and one with feedback are applied. The fundamental conclusion is that, for unstable noise, time-varying channel input strategies incur higher rates compared to asymptotically time-invariant channel input strategies.

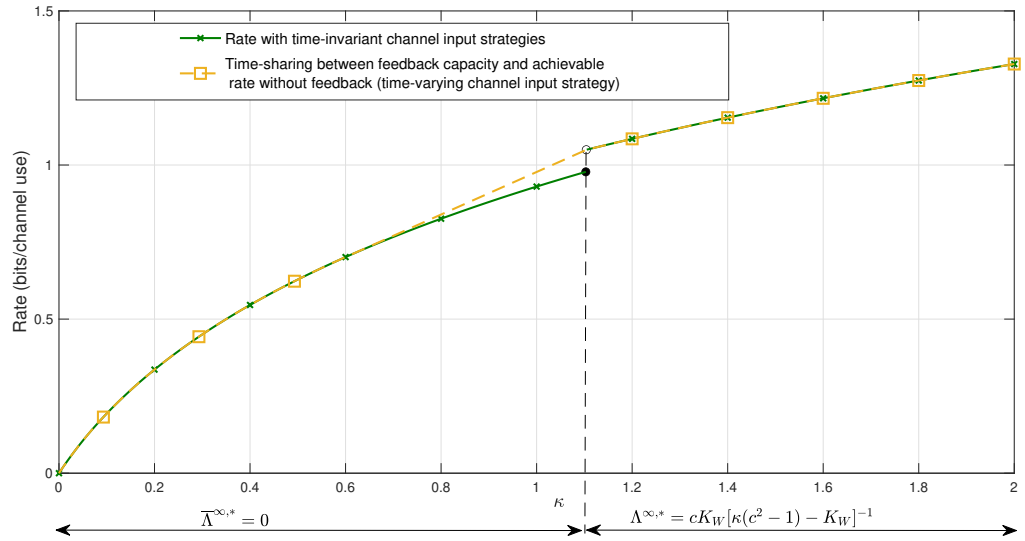


Figure 5. AR(c) noise, with $c = 1.5$ and $K_W = 1$. Time-sharing between the feedback strategy of Regime 1, $X_t^o = \Lambda^{\infty,*}\left(V_{t-1} - \widehat{V}_{t-1}^o\right) + Z_t^o$, with values $(\Lambda^{\infty,*}, K_Z^{\infty,*})$ given by (122) and (123) and the nonfeedback strategy $X_t^o = Z_t^o$, with value $K_Z^{\infty,*} = \kappa$ of the achievable lower bound.

2.5. Comparison of Closed-Form Formulas $C_{FB}(\kappa)$ and $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1})$ with Numerical Solutions Produced by the Semi-Definite Program of [24]

Tables 1 and 2 present comparisons of the closed-form expressions of feedback capacity $C_{FB}(\kappa) =$ (20) for Regime 1, $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1}) =$ (25) for Regimes 2 and 3, and lower bound $C^{LB}(\kappa) =$ (31) for Regimes 1–3, to the numerical solutions produced by the semi-definite program (SDP) formulation [24] of the optimization problem $C_{FB}(\kappa, \mathbf{P}_{V_1})$ of (10).

Regime 1: $c^2 \in (1, \infty), \kappa > \kappa_{min}$. The Table 1 second block illustrates that the numerical solution of $C_{FB}(\kappa, \mathbf{P}_{V_1})$ produced by SDP coincides with the closed-form solution of $C_{FB}(\kappa)$ of Regime 1 given by (17)–(20), while $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1})$ of Regimes 2 and 3 is sub-optimal. The Table 2 second block illustrates that the optimal parameters (Λ^*, K_Z^*) produced by SDP of the optimization problem $C_{FB}(\kappa, \mathbf{P}_{V_1})$ coincide with the optimal parameters of the closed-form solution of $C_{FB}(\kappa)$ of Regime 1.

Regime 2, $c^2 \in (1, \infty), \kappa \leq \kappa_{min}$. The Table 1 first block illustrates that the numerical solution of $C_{FB}(\kappa, \mathbf{P}_{V_1})$ produced by SDP coincides with the closed-form solution of $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1})$ of Regime 2 given by (25), which further implies that $C_{FB}(\kappa)$ of Regimes 1 cannot be optimal (actually there is no feasible solution under the constraint $K_Z^* > 0$). The Table 2 first block illustrates that the optimal parameters (Λ^*, K_Z^*) produced by SDP of the optimization problem $C_{FB}(\kappa, \mathbf{P}_{V_1})$ coincide with the optimal parameters of the closed-form solution of $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1})$ of Regime 2.

Regime 3, $c \in [-1, 1], \kappa \in [0, \infty)$. The numerical solution of $C_{FB}(\kappa, \mathbf{P}_{V_1})$ produced by SDP coincides with the closed-form solution of $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1})$ of Regime 3 given by (25), hence we omit the details.

The above numerical solution is consistent with Theorem 4, where we show that the optimization problem (10)–(12) for Regimes 2 and 3 does not have a nontrivial feasible solution that corresponds to an optimal value $K_Z^* > 0$.

Table 1. $c = 2, K_W = 1, \kappa \in [0, \infty), \kappa_{min} = 0.2558639597$. Comparisons of (i) Regime 1, feedback capacity $C_{FB}(\kappa)$ of (20), and Regime 2, lower bound on nonfeedback capacity $C^{LB}(\kappa)$ of (31) incurred by independent channel inputs, with (ii) solution of semi-definite program of [24], and (iii) $C_{FB}^{max} = \log|\Lambda^*|, |\Lambda^*| > 1$ of (25) . Numbers are rounded to the nearest decimal points.

Comp. of C_{FB}, C^{max} and C^{LB} with Numerical Sol. of Semi-Definite Programming (SDP) [24]				
$\kappa \leq \kappa_{min}$	*Reg.* 1, C_{FB}	Reg. 1–3, C^{LB}	SDP [24]	*Reg.* 2, C_{FB}^{max} optimal
0.01	no feasible set	0.0356094122	0.0600925515	0.0600925523
0.03	no feasible set	0.1000526016	0.1539595604	0.1539595679
0.09	no feasible set	0.2589485246	0.3413268000	0.3413268001
0.19	no feasible set	0.4559930447	0.5319371043	0.5319371060
0.25	no feasible set	0.5547462343	0.6183330358	0.6183330413
$\kappa > \kappa_{min}$	Reg. 1, C_{FB} opt.	Reg. 1–3, C^{LB}	SDP [24]	Reg. 2, C_{FB}^{max} not optimal
0.30	0.6740406521	0.6193243103	0.6740406486	0.6726584307
0.33	0.7075908808	0.6575848570	0.7075908775	0.7041839537
1.51	1.4432047194	1.4361084033	1.4432047188	1.2924606576
3.13	1.8963683737	1.8942662100	1.8963683715	1.6247312241
6.76	2.4140154130	2.4135010032	2.4140154125	2.0077561174

Table 2. $c = 2, K_W = 1, \kappa \in [0, \infty), \kappa_{min} = 0.2558639597$. Comparisons of optimal values (Λ^*, K_Z^*) of (i) Regime 1, $C_{FB}(\kappa)$ and Regime 2, $C^{LB}(\kappa)$ incurred by independent channel inputs, i.e., with $\Lambda^* = 0$ with (ii) solution of semi-definite program of [24], and (iii) $C_{FB}^{max} = \log|\Lambda^*|, |\Lambda^*| > 1$ of [2]. Numbers are rounded to the nearest decimal points.

Comparison of (Λ^*, K_Z^*) of C_{FB}, C^{max} and C^{LB} and Numerical Sol. of SDP [24]						
	Reg. 1–3, C^{LB}		SDP [24]		Regimes 2, 3, C_{FB}^{max} opt.	
$\kappa \leq \kappa_{min}$	Λ^*	K_Z^*	Λ^*	K_Z^*	Λ^*	K_Z^*
0.01	0	0.0102	1.0425	3.9172×10^{-10}	1.0425	0
0.03	0	0.0304	1.1126	5.5728×10^{-9}	1.1126	0
0.09	0	0.0910	1.2669	3.2046×10^{-10}	1.2669	0
0.19	0	0.1920	1.4458	1.9058×10^{-8}	1.4459	0
0.25	0	0.2558	1.5347	8.6363×10^{-5}	1.5351	0
	of Reg. 1, C_{FB} opt.		SDP [24]		Reg. 2, 3, C_{FB}^{max} not opt.	
$\kappa > \kappa_{min}$	Λ^*	K_Z^*	Λ^*	K_Z^*	Λ^*	K_Z^*
0.30	1.1575	0.1102	1.1574	0.1102	1.5940	0
0.33	0.9997	0.1668	0.9995	0.1668	1.6292	0
1.51	0.1583	1.4888	0.1582	1.4888	2.4494	0
3.13	0.0736	3.1191	0.0735	3.1191	3.0839	0
6.76	0.0334	6.7621	0.0333	6.7621	4.0216	0

3. Asymptotic Characterizations of Capacity for Stable and Unstable AR Noise

In this Section, we present the mathematical tools and derivations of the expressions presented in Section 2. Our derivations are generic and applicable to other more general noise models. In fact, it will become apparent that the analysis of the AR unit memory noise model is rich enough and brings out all technical issues that one will be encounter, if more general noise models are considered.

3.1. Sequential Characterizations of Feedback Capacity and Lower Bounds on Achievable Nonfeedback Capacity for Gaussian Channels Driven by Time-Varying AR Noise

Theorem 1 is the starting point prior to the analysis of the asymptotic feedback capacity, $\lim_{n \longrightarrow \infty} C_{n,FB}(\kappa, \mathbf{P}_{V_1})$. It was first presented in [27] (and earlier in [26]).

Theorem 1. *[26,32] Sequential characterization of $C_{n,FB}(\kappa, \mathbf{P}_{V_1})$ for Gaussian channels driven by $AR(c_t)$ Noise*

Consider the feedback code of [3] and the $AR(c_t)$ noise of Definition 1. Define the conditional means and covariances,

$$\widehat{X}_t \stackrel{\triangle}{=} \mathbf{E}\Big\{X_t \Big| Y^t\Big\}, \quad \widehat{V}_t \stackrel{\triangle}{=} \mathbf{E}\Big\{V_t \Big| Y^t\Big\}, \qquad t = 1, \ldots, n, \tag{34}$$

$$K_t \stackrel{\triangle}{=} \mathbf{E}\Big\{\Big(X_t - \widehat{X}_t\Big)^2 \Big| Y^t\Big\} = \mathbf{E}\Big\{\Big(V_t - \widehat{V}_t\Big)^2 \Big| Y^t\Big\}, \qquad t = 1, \ldots, n. \tag{35}$$

(1) The characterization of the $n-$FTFI capacity $C_{n,FB}(\kappa, \mathbf{P}_{V_1})$ is given by

$$C_{n,FB}(\kappa, \mathbf{P}_{V_1}) = \sup_{\mathbf{P}_{X_t|X^{t-1},Y^{t-1}}, t=1,\ldots,n, \frac{1}{n}\mathbf{E}\left\{\sum_{t=1}^{n}\left(X_t\right)^2\right\} \le \kappa} \frac{1}{n}\Big\{H(Y^n) - H(V^n)\Big\} \in [0,\infty] \tag{36}$$

$$= \sup_{\mathscr{P}_{n,FB}(\kappa)} \Big\{\frac{1}{2n}\log\Big(\frac{K_{Z_1}+K_{V_1}}{K_{V_1}}\Big) + \frac{1}{2n}\sum_{t=2}^{n}\log\Big(\frac{\left(\Lambda_t + c_t\right)^2 K_{t-1} + K_{Z_t} + K_{W_t}}{K_{W_t}}\Big)\Big\} \tag{37}$$

$$\mathscr{P}_{n,FB}(\kappa) \stackrel{\triangle}{=} \Big\{(\Lambda_t, K_{Z_t}) \in (-\infty,\infty)\times[0,\infty), t=1,\ldots,n \Big| \; K_t \ge 0,\ t=1,\ldots,n \text{ satisfies (39)},$$
$$\frac{1}{n}K_{Z_1} + \frac{1}{n}\sum_{t=2}^{n}\Big\{\left(\Lambda_t\right)^2 K_{t-1} + K_{Z_t}\Big\} \le \kappa\Big\}, \tag{38}$$

$$K_t = c_t^2 K_{t-1} + K_{W_t} - \frac{\Big(K_{W_t} + c_t K_{t-1}\Big(\Lambda_t + c_t\Big)\Big)^2}{\Big(K_{Z_t} + K_{W_t} + \Big(\Lambda_t + c_t\Big)^2 K_{t-1}\Big)}, \quad K_t \ge 0,\ K_1 = \frac{K_{V_1}K_{Z_1}}{K_{V_1}+K_{Z_1}} \ge 0 \tag{39}$$

where $H(X)$ denotes the entropy (differential) of RV X.

(2) $C_n(\kappa, \mathbf{P}_{V_1})$ in part (1) is achieved by the time-varying jointly Gaussian channel input process X^n, with a representation

$$X_t = \overline{\Lambda}_t\Big(X_{t-1} - \widehat{X}_{t-1}\Big) + Z_t, \quad X_1 = Z_1, \quad t=2,\ldots,n, \tag{40}$$
$$= \Lambda_t\Big(V_{t-1} - \widehat{V}_{t-1}\Big) + Z_t, \quad \overline{\Lambda}_t = -\Lambda_t, \tag{41}$$
$$Z_t \in N(0, K_{Z_t}), \quad K_{Z_t} \ge 0, \quad t=1,\ldots,n \quad \textit{indep. Gaussian sequence}, \tag{42}$$
$$Z_t \quad \textit{independent of} \quad (V^{t-1}, X^{t-1}, Y^{t-1}), \quad t=1,\ldots,n, \quad Z^n \quad \textit{independent of} \quad V^n, \tag{43}$$
$$V_t = c_t V_{t-1} + W_t, \quad c_t \in (-\infty,\infty), \quad t=2,\ldots,n, \tag{44}$$
$$Y_t = X_t + V_t = \overline{\Lambda}_t\Big(X_{t-1} - \widehat{X}_{t-1}\Big) + c_t\Big(Y_{t-1} - X_{t-1}\Big) + W_t + Z_t, \quad t=2,\ldots,n \tag{45}$$
$$= \Lambda_t\Big(V_{t-1} - \widehat{V}_{t-1}\Big) + c_t V_{t-1} + W_t + Z_t, \tag{46}$$
$$Y_1 = Z_1 + V_1, \tag{47}$$
$$\frac{1}{n}\mathbf{E}\Big\{\sum_{t=1}^{n}\left(X_t\right)^2\Big\} = \frac{1}{n}K_{Z_1} + \frac{1}{n}\sum_{t=2}^{n}\Big\{\Big(\Lambda_t\Big)^2 K_{t-1} + K_{Z_t}\Big\} \le \kappa, \tag{48}$$
$$(\Lambda_t, K_{Z_t}) \in (-\infty,\infty)\times[0,\infty) \quad \textit{scalar-valued, non-random}, \tag{49}$$
$$\widehat{V}_1 \stackrel{\triangle}{=} \mathbf{E}\Big\{V_1\Big|Y_1\Big\} = \frac{K_{V_1}}{K_{V_1}+K_{Z_1}}Y_1, \quad K_{V_1} \ge 0. \tag{50}$$

and its distribution satisfies, $\mathbf{P}_{X_t|X_{t-1},Y^{t-1}} = \mathbf{P}_{X_t|V_{t-1},Y^{t-1}}, t=1,\ldots,n$.

Further, $(\widehat{V}_t, K_t), t=1,\ldots,n$ are determined by the generalized (unlike [20], we use the term generalized because the conditions for the asymptotic analysis to hold are fundamentally different from those of asymptotic analysis of classical Kalman-filter equations) time-varying Kalman-filter and generalized time-varying difference Riccati equation (DRE) of estimating V^n from Y^n, given below.

Generalized Kalman-filter Recursion for (44)–(47) (see application of Kalman filter in [28,29]):

$$\widehat{V}_t = c_t \widehat{V}_{t-1} + M_t(K_{t-1}, \Lambda_t, K_{Z_t}) I_t, \quad \widehat{V}_1 = (50), \tag{51}$$

$$= F_t^{CL}(K_{t-1}, \Lambda_t, K_{Z_t}) \widehat{V}_{t-1} + M_t(K_{t-1}, \Lambda_t, K_{Z_t}) Y_t, \tag{52}$$

$$I_t \stackrel{\triangle}{=} Y_t - \mathbf{E}\Big\{ Y_t \Big| Y^{t-1} \Big\} = Y_t - c_t \widehat{V}_{t-1}, \quad t = 2, \ldots, n, \quad I_1 = Z_1 + V_1, \tag{53}$$

$$= \Big(\Lambda_t + c_t \Big) \Big(V_{t-1} - \widehat{V}_{t-1} \Big) + Z_t + W_t, \tag{54}$$

$$M_t(K_{t-1}, \Lambda_t, K_{Z_t}) \stackrel{\triangle}{=} \Big(K_{W_t} + c_t K_{t-1} \Big(\Lambda_t + c_t \Big) \Big) \Big(K_{Z_t} + K_{W_t} + \Big(\Lambda_t + c_t \Big)^2 K_{t-1} \Big)^{-1}, \tag{55}$$

$$F_t^{CL}(K_{t-1}, \Lambda_t, K_{Z_t}) \stackrel{\triangle}{=} c_t - M_t(K_{t-1}, \Lambda_t, K_{Z_t}) \Big(\Lambda_t + c_t \Big), \tag{56}$$

$$I_t, \quad t = 1, \ldots, n, \qquad \textit{an orthogonal innovations process.} \tag{57}$$

Error Recursion of the Generalized Kalman filter, $E_t \stackrel{\triangle}{=} V_t - \widehat{V}_t, t = 1, \ldots, n$ *:*

$$E_t = F_t^{CL}(K_{t-1}, \Lambda_t, K_{Z_t}) E_{t-1} - M_t(K_{t-1}, \Lambda_t, K_{Z_t}) \Big(Z_t + W_t \Big) + W_t, \quad E_1 = V_1 - \widehat{V}_1. \tag{58}$$

Generalized Time-Varying Difference Riccati Equation: $K_t = \mathbf{E}\{(E_t)^2\} \geq 0, t = 1, \ldots, n$ *satisfies (39).*

Theorem 1 part (2) is presented because it is this part that gives rise to part (1) and provides fundamental insight.

Remark 3. *The special case of Theorem 1 with* $(c_t, K_{W_t} = (c, K_W), c \in (-1, 1), \forall t$*, and initial state* $V_1 = v_1$ *fixed and known to the encoder and the decoder can be obtained from [2,24], while bounds on* $C_{n,FB}(\kappa, \mathbf{P}_{V_1})$ *are given in [8–10]. However, there is a subtle difference; if* $V_1 = v_1$ *fixed and known to the encoder, then* $C_{n,FB}(\kappa, \mathbf{P}_{V_1}) = C_{n,FB}(\kappa, v_1) =$*, i.e., depends on the initial state* $V_1 = v_1$ *not its distribution* $\mathbf{P}_{V_1}$*, because the following adjustments are needed in Theorem 1,* $K_1 = 0, \widehat{V}_1 = v_1$ *(because* $V_1 = v_1$ *is fixed). This corresponds to Theorem 1, by replacing* $K_{V_1} \geq 0$ *by* $K_{V_1} = 0$*. The subtle issue of dependence of* $C_{n,FB}(\kappa, \mathbf{P}_{V_1})$ *on* $\mathbf{P}_{V_1}$ *will emerge often in the subsequent analysis of asymptotic limit* $C_{FB}(\kappa, \mathbf{P}_{V_1})$*.*

Now, we discuss a subtle issue of $K_{V_1} = 0$. By (37), if $K_{V_1} = 0$ then $\frac{1}{2} \log \Big(\frac{K_{Z_1} + K_{V_1}}{K_{V_1}} \Big) \Big|_{K_{V_1} = 0} = +\infty$. To exclude $+\infty$ values of mutual information, we use Gelfand and Yaglom [34] (see Equation (2.8′) and Chapter II), as stated in Remark 4.

Remark 4. *By Theorem 1 if* $K_{V_1} = 0$ *and* $K_{Z_1} > 0$*, since* $X_1 = Z_1$ *and* $Y_1 = Z_1$*, we have at time* $t = 1$ *a value of mutual information* $I(X_1; Y_1) = I(Z_1; Z_1) = +\infty$*, which is well-known and allowed, for the mutual information between two jointly Gaussian RVs; a detailed discussion on this point is given in Gelfand and Yaglom [34] (see Equation (2.8′) and Chapter II). The value* $I(X_1; Y_1) = +\infty$ *is precisely the value of* $\frac{1}{2} \log \Big(\frac{K_{Z_1} + K_{V_1}}{K_{V_1}} \Big) \Big|_{K_{V_1} = 0} = +\infty$*, i.e., the first right hand side term in (37). As discussed in Gelfand and Yaglom [34] (see Equation (2.8′) and Chapter II), to avoid values of* $+\infty$*, it is a standard practice to remove such RVs from the evaluation of rates. A transmission policy which avoids values of* $I(X_1; Y_1) = +\infty$ *is the following.*

Step # 1. Initialize $K_n, n = 2, 3, \ldots$ *at* $K_1 = 0$*, i.e., corresponding to* $K_{V_1} = 0$*.*

Step # 2. Omit transmission at time $t = 1$*, hence, the term* $\frac{1}{2} \log \Big(\frac{K_{Z_1} + K_{V_1}}{K_{V_1}} \Big) \Big|_{K_{V_1} = 0} = +\infty$ *in (37) is excluded.*

Step # 3. Apply Step #2 and evaluate $\frac{1}{n} C_n(\kappa, \mathbf{P}_{V_1})$ *and its limit, with* $K_{V_1} = 0$*, i.e.,* $K_1 = 0$*.*

By Steps #1–#3, we can treat the case when the noise initial variance is $K_{V_1} = 0$, and hence we can also consider asymptotically stationary noise, since such noise is defied for all $K_{V_1} \geq 0$.

We should emphasize the following subtle issue: Steps #1–#3 do not apply to the channel input of Regimes 2, 3, (29) and (30), because by the proof of Theorem 5, if $K_{V_1} = 0$ then $K_1 = 0$, and $C_{FB}(\kappa, \mathbf{P}_{V_1}) = \lim_{n \longrightarrow \infty} C_{n,FB}(\kappa, \mathbf{P}_{V_1}) = 0$.

Next, we give a lower bound on the characterization of $n-$FTwFI Capacity for non-stationary and nonergodic noise, i.e., for the nonfeedback code [3], which is based on a sequential characterization similar to Theorem 1. In the paper, we evaluate this lower bound against the feedback capacity $C_{FB}(\kappa, \mathbf{P}_{V_1})$.

Corollary 1. *(Lower bound on sequential characterization of $C_n(\kappa, \mathbf{P}_{V_1})$ for Gaussian channels driven by AR Noise)*

Consider the code without feedback in [3], and the AR(c_t) noise of Definition 1. Define

$$K_{X_t} \stackrel{\triangle}{=} \mathbf{E}\Big(X_t\Big)^2, \quad \widehat{X}_t \stackrel{\triangle}{=} \mathbf{E}\Big\{X_t\Big|Y^t\Big\}, \quad \widehat{V}_t \stackrel{\triangle}{=} \mathbf{E}\Big\{V_t\Big|Y^t\Big\}, \quad E_t^{nfb} \stackrel{\triangle}{=} V_t - \widehat{V}_t, \tag{59}$$

$$K_t \stackrel{\triangle}{=} \mathbf{E}\Big\{\Big(X_t - \widehat{X}_t\Big)^2\Big|Y^t\Big\} = \mathbf{E}\Big\{\Big(V_t - \widehat{V}_t\Big)^2\Big|Y^t\Big\}, \quad t = 1, \ldots, n. \tag{60}$$

(1) A lower bound on the characterization of the $n-$FTwFI capacity $C_n(\kappa, \mathbf{P}_{V_1})$ is given by

$$C_n(\kappa, \mathbf{P}_{V_1}) = \sup_{\mathbf{P}_{X_t|X^{t-1}, t=1,\ldots,n}, \frac{1}{n}\mathbf{E}\{\sum_{t=1}^n (X_t)^2\} \leq \kappa} \frac{1}{n}\Big\{H(Y^n) - H(V^n)\Big\} \in [0, \infty] \tag{61}$$

$$\geq C_n^{LB}(\kappa, \mathbf{P}_{V_1}) \stackrel{\triangle}{=} \sup_{\mathscr{P}_n^{LB}(\kappa)} \Big\{\frac{1}{2n}\log\Big(\frac{K_{Z_1} + K_{V_1}}{K_{V_1}}\Big)$$

$$+ \frac{1}{2n}\sum_{t=2}^n \log\Big(\frac{\big(\overline{\Lambda}_t - c_t\big)^2 K_{t-1} + K_{Z_t} + K_{W_t}}{K_{W_t}}\Big)\Big\}, \tag{62}$$

$$\mathscr{P}_n^{LB}(\kappa) \stackrel{\triangle}{=} \Big\{\big(\overline{\Lambda}_t, K_{Z_t}\big) \in (-\infty, \infty) \times [0, \infty), t = 1, \ldots, n\Big| \quad K_t \geq 0,\ K_{X_t} \geq 0,\ t = 1, \ldots, n$$

$$\text{satisfy (64) and (65)}, \quad \frac{1}{n}K_{Z_1} + \frac{1}{n}\sum_{t=2}^n \big\{\big(\overline{\Lambda}_t\big)^2 K_{X_{t-1}} + K_{Z_t}\big\} \leq \kappa\Big\}, \tag{63}$$

$$K_t = \overline{\Lambda}_t^2 K_{t-1} + K_{Z_t} - \frac{\Big(K_{Z_t} + \overline{\Lambda}_t K_{t-1}\Big(\overline{\Lambda}_t - c_t\Big)\Big)^2}{\Big(K_{Z_t} + K_{W_t} + \Big(\overline{\Lambda}_t - c_t\Big)^2 K_{t-1}\Big)},\ K_1 = \frac{K_{V_1} K_{Z_1}}{K_{V_1} + K_{Z_1}} \geq 0,\ K_t \geq 0, \tag{64}$$

$$K_{X_t} = \overline{\Lambda}_t^2 K_{X_{t-1}} + K_{Z_t}, \quad K_{X_t} \geq 0, \quad K_{X_1} = K_{Z_1}, t = 2, \ldots, n. \tag{65}$$

(2) $C_n^{LB}(\kappa, \mathbf{P}_{V_1})$ of part (1) is achieved by the time-varying jointly Gaussian channel input process X^n, with a representation

$$X_t = \overline{\Lambda}_t X_{t-1} + Z_t, \quad X_1 = Z_1, \qquad t = 2, \ldots, n, \tag{66}$$

$$Z_t \in N(0, K_{Z_t}), \quad K_{Z_t} \geq 0, \quad t = 1, \ldots, n \quad \textit{indep. Gaussian sequence,} \tag{67}$$

$$Z_t \quad \textit{independent of} \quad (V^{t-1}, X^{t-1}, Y^{t-1}), \quad t = 1, \ldots, n, \quad Z^n \quad \textit{independent of} \quad V^n, \tag{68}$$

$$V_t = c_t V_{t-1} + W_t, , \qquad c_t \in (-\infty, \infty), \quad t = 2, \ldots, n, \tag{69}$$

$$Y_t = X_t + V_t = \Big(\overline{\Lambda}_t - c_t\Big) X_{t-1} + c_t Y_{t-1} + W_t + Z_t, \quad t = 2, \ldots, n, \tag{70}$$

$$Y_1 = Z_1 + V_1, \tag{71}$$

$$\frac{1}{n}\mathbf{E}\Big\{\sum_{t=1}^{n} (X_t)^2\Big\} = \frac{1}{n} K_{Z_1} + \frac{1}{n}\sum_{t=2}^{n}\Big\{\Big(\overline{\Lambda}_t\Big)^2 K_{X_{t-1}} + K_{Z_t}\Big\} \leq \kappa, \tag{72}$$

$$(\overline{\Lambda}_t, K_{Z_t}) \in (-\infty, \infty) \times [0, \infty) \qquad \textit{scalar-valued, non-random.} \tag{73}$$

and its distribution satisfies $\mathbf{P}_{X_t|X^{t-1}} = \mathbf{P}_{X_t|X_{t-1}}, t = 1, \ldots, n$.

Further, $(\widehat{X}_t, K_t), t = 1, \ldots, n$ *are determined by the generalized time-varying Kalman-filter and generalized time-varying difference Riccati equation (DRE) of estimating* X^n *from* Y^n*, and* $K_{X_t}, t = 1, \ldots, n$ *is determined by the time-varying Lyapunov difference equation, given below. Generalized Kalman-filter Recursion for (66)–(71):*

$$\widehat{X}_t = \overline{\Lambda}_t \widehat{X}_{t-1} + M_t^{nfb}(K_{t-1}, \overline{\Lambda}_t, K_{Z_t}) I_t, \quad \widehat{X}_1 = (50), \quad t = 2, \ldots, n \tag{74}$$

$$= F_t^{nfb}(K_{t-1}, \overline{\Lambda}_t, K_{Z_t})\widehat{X}_{t-1} + M_t^{nfb}(K_{t-1}, \overline{\Lambda}_t, K_{Z_t})\Big(Y_t - c_t Y_{t-1}\Big), \tag{75}$$

$$I_t \stackrel{\triangle}{=} Y_t - \Big(\overline{\Lambda}_t - c_t\Big)\widehat{X}_{t-1} - c_t Y_{t-1}, \quad I_1 = Z_1 + W_1, \quad t = 2, \ldots, n, \tag{76}$$

$$= \Big(\overline{\Lambda}_t - c_t\Big)\Big(X_{t-1} - \widehat{X}_{t-1}\Big) + Z_t + W_t, \tag{77}$$

$$M_t^{nfb}(K_{t-1}, \overline{\Lambda}_t, K_{Z_t}) \stackrel{\triangle}{=} \Big(K_{Z_t} + \overline{\Lambda}_t K_{t-1}\Big(\overline{\Lambda}_t - c_t\Big)\Big)\Big(K_{Z_t} + K_{W_t} + \Big(\overline{\Lambda}_t - c\Big)^2 K_{t-1}\Big)^{-1}, \tag{78}$$

$$F_t^{CL,nfb}(K_{t-1}, \overline{\Lambda}_t, K_{Z_t}) \stackrel{\triangle}{=} \overline{\Lambda}_t - M_t^{nfb}(K_{t-1}, \overline{\Lambda}_t, K_{Z_t})\Big(\overline{\Lambda}_t - c_t\Big) \tag{79}$$

$$I_t, \quad t = 1, \ldots, n, \qquad \textit{an orthogonal innovations process.} \tag{80}$$

Error Recursion of the Generalized Kalman-filter, $E_t^{nfb} \stackrel{\triangle}{=} X_t - \widehat{X}_t, t = 1, \ldots, n$ *:*

$$E_t^{nfb} = F_t^{CL,nfb}(K_{t-1}, \overline{\Lambda}_t, K_{Z_t})E_{t-1} - M_t^{nfb}(K_{t-1}, \overline{\Lambda}_t, K_{Z_t})\Big(Z_t + W_t\Big) + Z_t, \; E_1^{nfb} = V_1 - \widehat{V}_1. \tag{81}$$

Generalized Time-Varying Difference Riccati Equation: $K_t = \mathbf{E}\big\{\big(E_t^{nfb}\big)^2\big\} \geq 0, t = 1, \ldots, n$ *satisfies (64).*

Time-Varying Difference Lyapunov Equation: $K_{X_t} \geq 0, t = 1, \ldots, n$ *satisfies (65) .*

Proof. From Cover and Pombra [3], we know that a jointly Gaussian X^n achieved the supremum in $C_n(\kappa, \mathbf{P}_{V_1})$, i.e., its distribution is $\mathbf{P}_{X_t|X^{t-1}}, t = 1, \ldots, n$. Hence, a Markovian input, i.e., $\mathbf{P}_{X_t|X_{t-1}}, t = 1, \ldots, n$ gives a lower bound. A realization of such Markovian input is $X_t = \overline{\Lambda}_t X_{t-1} + Z_t, X_1 = Z_1, t = 2, \ldots, n$. This implies the recursion (65). The rest of the statements are easily obtained by calculating the $n-$FTwFI capacity of Cover and Pombra [3] for this specific input, using the generalized Kalman-filter recursions following [28,29]. Some of the details are presented in [33], hence we omit them. □

Remark 5. *The solution to the optimization problem* $C_{n,FB}(\kappa, \mathbf{P}_{V_1})$ *of (37) for arbitrary finite n remains to this date a challenging open problem because the optimization needs to be carried out over the time-varying parameters* $(\Lambda_n, K_{Z_n}), K_{Z_n} \geq 0, n = 1, 2 \ldots$. *For the rest of the paper we consider*

the asymptotic feedback capacity problem $\lim_{n\longrightarrow\infty} C_{n,FB}(\kappa, \mathbf{P}_{V_1})$ *for asymptotically time-invariant stable and unstable* $AR(c)$ *noise.*

In Section 3.2 (to be introduced shortly) we present necessary and sufficient conditions to address the limit $C_{FB}(\kappa, \mathbf{P}_{V_1}) = \lim_{n\longrightarrow\infty} C_{n,FB}(\kappa, \mathbf{P}_{V_1})$ and the convergence properties of the error recursion (58).

Although recursion (58) is linear, $\lim_{n\longrightarrow\infty} K_n$ is not expected to exist, for arbitrary $(F_t^{CL}(K_{t-1}, \Lambda_t, K_{Z_t}), M_t(K_{t-1}, \Lambda_t, K_{Z_t})), t = 2, \ldots$. Previous papers [2,20,24] did not address this point. Indeed, by Section 3.2, the convergence properties of the sequence $K_2, \ldots, K_n$ with $K_1 \geq 0$, as $n \longrightarrow \infty$, are characterized by the detectability and stabilizability or unit circle controllability conditions [28,29]. The detectability and stabilizability (resp., unit circle controllability) conditions ensure existence of the limit, $\lim_{n\longrightarrow\infty} K_n = K$, such that $K \geq 0$ is the unique solution (resp., maximal solution) of a generalized ARE and satisfies the stability property, $\lim_{n\longrightarrow\infty} F_n^{CL}(K_{n-1}, \Lambda_n, K_{Z_n}) = F^{CL}(K, \Lambda, K_Z) \in (-1, 1)$, for all $K_{V_1} \geq 0$ (resp., for some $K_{V_1} > 0$). We show that the solution of the asymptotic feedback capacity optimization problem is such that detectability and stabilizability conditions hold, but this is only for Regime 1. For the complement of Regime 1, we show there exists an optimal solution, provided the stabilizability condition is replaced by the more relaxed unit circle controllability condition, and this gives rise to the feedback capacity of Regimes 2, 3. In the sequel, these detectability and stabilizability or unit circle controllability conditions are encountered in the asymptotic feedback capacity optimization problem.

Remark 6. *In Section 3.2 we introduce the properties of generalized DREs and AREs for time-invariant strategies* $(\Lambda_n, K_{Z_n}) = (\Lambda, K_Z)\ \forall n$*; we note that these also hold for asymptotically time-invariant strategies,* $\lim_{n\longrightarrow\infty}(\Lambda_n, K_{Z_n}) = (\Lambda, K_Z)$*, due to the continuity properties of DREs with respect to their coefficients.*

3.2. Convergence Properties of Generalized DREs

First, we emphasize that all materials of this section have natural generalizations to the finite-dimensional state space realizations of the noise V^n, which are considered in [2,20,24] (these are found in [27] and earlier arXiv paper [26]). Unfortunately, [2,20,24] did not relate their asymptotic feedback capacity, $C_{FB}(\kappa, \mathbf{P}_{V_1}) = \lim_{n\longrightarrow\infty} C_{n,FB}(\kappa, \mathbf{P}_{V_1})$, to the material of this section, and this omission leads to various confusions, including the statements found in the comment paper [1] regarding the TD feedback capacity given in [2] [Theorem 6.1]. The complexity will become apparent below.

Consider the generalized DREs and AREs for time-invariant parameters and strategies $(c_n, K_{W_n}) = (c, K_W), (\Lambda_n, K_{Z_n}) = (\Lambda, K_Z)\ \forall n$. Let $\{K_t, t = 1, 2, \ldots, n\}$ denote a sequence that satisfies the time-invariant generalized DRE with arbitrary initial condition

$$K_t = c^2 K_{t-1} + K_W - \frac{\Big(K_W + cK_{t-1}\Big(\Lambda + c\Big)\Big)^2}{\Big(K_Z + K_W + \Big(\Lambda + c\Big)^2 K_{t-1}\Big)}, \qquad K_1 = \text{given}, \qquad t = 2, \ldots, n. \tag{82}$$

The generalized algebraic Riccati equation (ARE) corresponding to (82) is

$$K = c^2 K + K_W - \frac{\Big(K_W + cK\Big(\Lambda + c\Big)\Big)^2}{\Big(K_Z + K_W + \Big(\Lambda + c\Big)^2 K\Big)}, \qquad K \geq 0. \tag{83}$$

We note that a solution of (82) is a functional of the parameters of the right hand side, that is, $K_t \equiv K_t(c, K_W, \Lambda, K_Z, K_1), t = 2, \ldots, n$. To discuss the properties of the generalized

DRE (82), we introduce, as is often performed in the analysis of generalized DREs [28] and [29] [Section 14.7, page 540], the following standard definitions.

$$A \stackrel{\triangle}{=} c, \quad C \stackrel{\triangle}{=} \Lambda + c, \quad A^* \stackrel{\triangle}{=} c - K_W R^{-1} C, \quad B^{*,\frac{1}{2}} \stackrel{\triangle}{=} K_W^{\frac{1}{2}} B^{\frac{1}{2}} \tag{84}$$

$$R \stackrel{\triangle}{=} K_Z + K_W > 0, \quad B \stackrel{\triangle}{=} 1 - K_W \left(K_Z + K_W\right)^{-1}, \tag{85}$$

$$K_W > 0, \quad c \in (-\infty, \infty), \quad \Lambda \in (-\infty, \infty), \quad K_Z \geq 0. \tag{86}$$

By (55) and (57), we also define

$$M(K, \Lambda, K_Z) \stackrel{\triangle}{=} \Big(K_W + AKC\Big)\Big(R + \big(C\big)^2 K\Big)^{-1}, \quad F^{CL}(K, \Lambda, K_Z) = A - M(K, \Lambda, K_Z)C. \tag{87}$$

Next, we introduce the definition of asymptotic stability of the error recursion (58) for $(c_n, K_{W_n}, \Lambda_n, K_{Z_n}) = (c, K_W, \Lambda, K_Z)\ \forall n$.

Definition 2. *Asymptotic stability [28,29]*

A solution $K \geq 0$ to the generalized ARE (83), assuming it exists, is called

(a) stabilizing if $|F^{CL}(K, \Lambda, K_Z)| < 1$,

(b) maximal and stabilizing solution, denoted by K_+, if K_+ is stabilizing, and in addition $K_+ - K \geq 0$, where K is any other solution.

For such solutions, K or $K = K_+$, we say $F^{CL}(K, \Lambda, K_Z)$ is asymptotically stable, i.e., $|F^{CL}(K, \Lambda, K_Z)| < 1$,

Next, we define the important notions of detectability, unit circle controllability, and stabilizability, which are directly related to asymptotic stability of the error recursion (58), in a mean-square sense.

Definition 3. *Detectability, Stabilizability, Unit Circle controllability [28,29]*

(a) The pair $\{A, C\}$ is called detectable if there exists a $G \in \mathbb{R}$ such that $|A - GC| < 1$ (stable).

(b) The pair $\{A^*, B^{*,\frac{1}{2}}\}$ is called unit circle controllable if there exists a $G \in \mathbb{R}$ such that $|A^* - B^{*,\frac{1}{2}} G| \neq 1$.

(c) The pair $\{A^*, B^{*,\frac{1}{2}}\}$ is called stabilizable if there exists a $G \in \mathbb{R}$ such that $|A^* - B^{*,\frac{1}{2}} G| < 1$.

The next lemma characterizes detectability, unit circle controllability, and stabilizability [29,35].

Lemma 1. *[29,35] Necessary and sufficient conditions for detectability, unit circle controllability, and stabilizability*

(a) The pair $\{A, C\}$ is detectable if and only if there exists no eigenvalue, eigenvector $\{\lambda, x\}$, of A, i.e., $Ax = \lambda x$ such that $|\lambda| \geq 1$, and such that $Cx = 0$

(b) The pair $\{A^*, B^{*,\frac{1}{2}}\}$ is unit circle controllable if and only if there exists no eigenvalue, eigenvector $\{\lambda, x\}$, $xA^* = x\lambda$, such that $|\lambda| = 1$, and such that that $xB^{*,\frac{1}{2}} = 0$.

(c) The pair $\{A^*, B^{*,\frac{1}{2}}\}$ is stabilizable if and only if there exists no eigenvalue, eigenvector $\{\lambda, x\}$, $xA^* = x\lambda$ such that $|\lambda| \geq 1$, and such that $xB^{*,\frac{1}{2}} = 0$.

(d) If $C \neq 0$ then the pair $\{A, C\}$ is detectable. If $B^{*,\frac{1}{2}} \neq 0$ then the pair $\{A^*, B^{*,\frac{1}{2}}\}$ is stabilizable.

In the next theorem we summarize known results on sufficient and necessary conditions for convergence of solutions $\{K_t, t = 1, 2, \ldots, n\}$ of the generalized time-invariant DRE, as $n \longrightarrow \infty$, to a non-negative K, which is stabilizing (not necessarily unique, i.e., it may be a maximal solution $K = K_+$), and unique and stabilizing. First, we recall that the

pair $\{A, C\}$ is detectable—*is a necessary condition but not sufficient condition for convergence of the sequence* $\{K_t, t = 1, 2, \ldots, n\}$ generated by the generalized DRE, as $n \longrightarrow \infty$, to a non-negative $K \geq 0$ which is a stabilizing solution of a corresponding generalized ARE. *The sufficient condition is* that the pair $\{A^*, B^{*,\frac{1}{2}}\}$ *is unit circle controllable*. However, under the conditions where the pair $\{A, C\}$ is detectable and the pair $\{A^*, B^{*,\frac{1}{2}}\}$ is unit circle controllable, the limiting K is not necessarily unique, and there may be multiple stabilizing solutions (i.e, such as K_+, where $K_+ - K \geq 0$, and $K \geq 0$ is any other solution) depending on the initial condition $K_1 \geq 0$. Uniqueness of $K \geq 0$ is ensured by replacing *the pair* $\{A^*, B^{*,\frac{1}{2}}\}$ *that is unit circle controllable* by *the pair* $\{A^*, B^{*,\frac{1}{2}}\}$ *is stabilizable*.

Theorem 2. *[28,29] Convergence of time-invariant generalized DRE to stabilizing solutions*

Let $\{K_t, t = 2, \ldots, n\}$ *denote a sequence that satisfies the time-invariant generalized DRE (82) with arbitrary initial condition,* K_1 *and* $(A, C, A^*, B^{*,\frac{1}{2}})$ *defined by (84) and (85), with* $R \stackrel{\triangle}{=} K_Z + K_W > 0$. *Then the following hold.*

(1) Consider the generalized RDE (82) with zero initial condition, i.e., $K_1 = 0$, *and assume the pair* $\{A, C\}$ *is detectable and the pair* $\{A^*, B^{*,\frac{1}{2}}\}$ *is unit circle controllable.*

Then the sequence $\{K_t : t = 2, \ldots, n\}$ *that satisfies the generalized DRE (82), with zero initial condition* $K_1 = 0$, *converges to K, i.e.,* $\lim_{n\longrightarrow\infty} K_n = K$, *where K satisfies the ARE (83) if and only if the pair* $\{A^*, B^{*,\frac{1}{2}}\}$ *is stabilizable.*

(2) Assume the pair $\{A, C\}$ *is detectable and the pair* $\{A^*, B^{*,\frac{1}{2}}\}$ *is unit circle controllable. Then there exists a unique stabilizing solution* $K \geq 0$ *to the generalized ARE (83), i.e., such that* $|F^{CL}(K, \Lambda, K_Z)| < 1$, *if and only if* $\{A^*, B^{*,\frac{1}{2}}\}$ *is stabilizable.*

(3) If $\{A, C\}$ *is detectable and* $\{A^*, B^{*,\frac{1}{2}}\}$ *is stabilizable, then any solution* $K_t, t = 2, \ldots, n$ *to the generalized DRE (82) with arbitrary initial condition,* $K_1 \geq 0$ *is such that* $\lim_{n\longrightarrow\infty} K_n = K$, *where* $K \geq 0$ *is the unique solution of the generalized ARE (83) with* $|F^{CL}(K, \Lambda, K_Z)| < 1$, *i.e., it is stabilizing.*

(4) $\{A, C\}$ *is detectable and* $\{A^*, B^{*,\frac{1}{2}}\}$ *unit circle controllable are necessary and sufficient conditions for any solution* $K_t, t = 2, \ldots, n$ *to the generalized DRE (82) to converge,* $\lim_{n\longrightarrow\infty} K_n = K$, *from some initial condition,* $K_1 \geq 0$ *where* $K \geq 0$ *is a stabilizing solution of the generalized ARE (83), but it may not be unique, i.e., Ref. (83) may have multiple solutions* $K \geq 0$, *where one of the solutions is the maximal and stabilizing solution, denoted by* $K_+ \geq 0$, *such that* $K_+ - K \geq 0$.

Theorem 2.(1) follows by combining [29] [Lemma 14.2.1, page 507] of classical DREs and AREs with [29] [Section 14.7] of generalized DREs and AREs. Theorem 2.(2) is given in [29] [Theorem E.6.1, page 784]. Theorem 2.(3) is obtained from [28] [Theorem 4.2, page 164], and also [29], etc.

From Theorem 2, follows the next lemma.

Lemma 2. *Properties of Solutions of time-invariant DREs and AREs for different cases*

Let $(A, C, A^*, B^{*,\frac{1}{2}})$ *be defined by (84) and (85), with* $R \stackrel{\triangle}{=} K_Z + K_W > 0$, *i.e.,* $K_W > 0$.

(1.a) Suppose $c \in (-1, 1)$. *Then the pair* $\{A, C\}$ *is detectable.*

(1.b) Suppose $c \notin (-1, 1)$. *Then the pair* $\{A, C\}$ *is detectable if* $C \stackrel{\triangle}{=} \Lambda + c \neq 0$.

(1.c) Suppose $c \in (-\infty, \infty)$. *Then the pair* $\{A^*, B^{*,\frac{1}{2}}\}$ *is stabilizable if* $K_Z > 0$.

(2) Suppose $K_Z = 0$. *Then the pair* $\{A^*, B^{*,\frac{1}{2}}\}$ *is unit circle controllable if and only if* $|\Lambda| \neq 1$.

(3) Suppose $K_Z = 0$. *Then the pair* $\{A^*, B^{*,\frac{1}{2}}\}$ *is stabilizable if and only if* $|\Lambda| < 1$.

(4) Suppose $c \in (-1, 1), K_Z = 0$. *The sequence* $\{K_t, t = 1, 2, \ldots, n\}$ *that satisfies the generalized DRE with zero initial condition* $K_1 = 0$ *(reps. non-zero initial conditions* $K_1 > 0$*)*

$$K_t = c^2 K_{t-1} + K_W - \frac{\Big(K_W + cK_{t-1}(\Lambda + c)\Big)^2}{\Big(K_W + (\Lambda + c)^2 K_{t-1}\Big)}, \quad K_1 = 0 \ \ (resp.\ K_1 > 0), \quad t = 2, \ldots, n \tag{88}$$

converges to $K \geq 0$, *i.e.,* $\lim_{n \longrightarrow \infty} K_n = K$, *where* K *satisfies the generalized ARE (83) with* $K_Z = 0$ *if and only if the* $\{A^*, B^{*,\frac{1}{2}}\}$ *is stabilizable, equivalently,* $|\Lambda| < 1$ *(resp., if and only if the* $\{A^*, B^{*,\frac{1}{2}}\}$ *is unit circle controllable, equivalently,* $|\Lambda| \neq 1$*).*

(5) Suppose $K_Z = 0$, *and* $|\Lambda| \neq 1$, *with the corresponding generalized ARE (83),*

$$K = c^2 K + K_W - \frac{\Big(K_W + cK(\Lambda + c)\Big)^2}{\Big(K_W + (\Lambda + c)^2 K\Big)}. \tag{89}$$

Then

$$F^{CL}(K, \Lambda, K_Z = 0) = c - M(K, \Lambda, K_Z = 0)(\Lambda + c), \tag{90}$$

$$M(K, \Lambda, K_Z) = \Big(K_W + cK(\Lambda + c)\Big)\Big(K_W + (\Lambda + c)^2 K\Big)^{-1}. \tag{91}$$

Moreover, the solutions $K \geq 0$ *of the ARE (89) are*

$$\begin{cases} K(\Lambda) = 0, \quad unique, stabilizing, |F^{CL}(K, \Lambda, K_Z = 0) = -\Lambda| < 1, iff\ |\Lambda| < 1, \\ K_+(\Lambda) = \frac{K_W\Big((\Lambda)^2 - 1\Big)}{(\Lambda + c)^2} \in (0, \infty), maximal, stabil., |F^{CL}(K, \Lambda, K_Z = 0) = -\frac{1}{\Lambda}| < 1, \\ iff\ |\Lambda| > 1, c \neq -\Lambda. \end{cases} \tag{92}$$

Proof. See Appendix A.1. □

Proposition 1. *Properties of solutions of asymptotically time-invariant DREs and AREs*

Consider the generalized DRE (82) coefficients of (c, K_W, Λ, K_Z) *replaced by time-varying* $(c_t, K_{W_t}, \Lambda_t, K_{Z_t})$, $t = 2, 3, \ldots, n$,

$$K_t = c_t^2 K_{t-1} + K_{W_t} - \frac{\Big(K_{W_t} + c_t K_{t-1}\Big(\Lambda_t + c_t\Big)\Big)^2}{\Big(K_{Z_t} + K_{W_t} + \Big(\Lambda_t + c_t\Big)^2 K_{t-1}\Big)}, \quad K_1 = given, \quad t = 2, \ldots, n, \tag{93}$$

$$such\ that \quad \lim_{n \longrightarrow \infty} (c_n, K_{W_n}, \Lambda_n, K_{Z_n}) = (c, K_W, \Lambda, K_Z), \quad K_W > 0,\ K_Z \geq 0. \tag{94}$$

Then Theorem 2 and Lemma 2 remain valid.

Proof. This is due to the well-known continuity properties of DREs with respect to their coefficients, i.e., the convergence properties are characterized by the limiting pairs, $\lim_{n \longrightarrow \infty} \{A_n, C_n\} = \{A, C\}$ and $\lim_{n \longrightarrow \infty} \{A_n^*, B_n^{*,\frac{1}{2}}\} = \{A^*, B^{*,\frac{1}{2}}\}$, where $A_n \stackrel{\triangle}{=} c_n, C_n \stackrel{\triangle}{=} \Lambda_n + c_n, A_n^* \stackrel{\triangle}{=} c_n - K_{W_n} R_n^{-1} C_n, B_n^{*,\frac{1}{2}} \stackrel{\triangle}{=} K_{W_n}^{\frac{1}{2}} B^{\frac{1}{2}}, R_n \stackrel{\triangle}{=} K_{Z_n} + K_{W_n} > 0, B_n \stackrel{\triangle}{=} 1 - K_{W_n}\big(K_{Z_n} + K_{W_n}\big)^{-1}, K_{W_n} > 0, c_n \in (-\infty, \infty), \Lambda_n \in (-\infty, \infty), K_{Z_n} \geq 0.$ □

3.3. Asymptotic Characterizations of Feedback Capacity Under Detectability and Stabilizability Versus Unit Circle Controllability

Now, we are ready to address $C_{FB}(\kappa, \mathbf{P}_{V_1}) = \lim_{n\longrightarrow\infty} C_{n,FB}(\kappa, \mathbf{P}_{V_1})$, by making use of the properties of generalized DREs and AREs of Section 3.2, under detectability and stabilizability or unit circle controllability. By invoking Theorem 2, we distinguish the following two cases.

(1) The sufficient and necessary conditions for existence of the limit such that $C_{FB}(\kappa, \mathbf{P}_{V_1}) = \lim_{n\longrightarrow\infty} C_{n,FB}(\kappa, \mathbf{P}_{V_1}) = C_{FB}(\kappa), \forall \mathbf{P}_{V_1}$ are detectability and stabilizability, provided there exists a feasible solution to the optimization problem $C_{FB}(\kappa), \forall \mathbf{P}_{V_1}$. This leads to the solution of Regime 1.

(2) The sufficient and necessary conditions for existence of a limit such that $C_{FB}(\kappa, \mathbf{P}_{V_1}) = \lim_{n\longrightarrow\infty} C_{n,FB}(\kappa, \mathbf{P}_{V_1}) = C_{FB}^{max}(\kappa, \mathbf{P}_{V_1})$ for some $\mathbf{P}_{V_1}$ are detectability and unit circle controllability (relaxation of stabilizability). This leads to the solutions of Regimes 2, 3.

First, we state the main theorem for (1).

Theorem 3. *(Uniform asymptotic limit of feedback capacity for asymptotically time-invariant AR noise-detectability and stabilizability)*

Consider Theorem 1 with asymptotically time-varying parameters of the stable and unstable noise, AR$(c_n), c_n \in (-\infty, \infty), \forall n$ *of Definition 1, i.e.,* $\lim_{n\longrightarrow\infty}(c_n, K_{W_n}) = (c, K_W), K_W > 0$.

Define the set

$$\mathscr{P}_{FB}^{\infty} \stackrel{\triangle}{=} \Big\{(\Lambda, K_Z) \in (-\infty, \infty) \times [0, \infty) \Big| \ \textit{(i) the pair } \{A, C\} \equiv \{A, C(\Lambda)\} \textit{ is detectable, } \textit{(ii) the pair } \{A^*, B^{*,\frac{1}{2}}\} \equiv \{A^*(K_Z), B^{*,\frac{1}{2}}(K_Z)\} \textit{ is stabilizable, } A, C, A^*, B^{*,\frac{1}{2}} \textit{ given by (84)} \Big\}. \tag{95}$$

Then

$$C_{FB}(\kappa, \mathbf{P}_{V_1}) = C_{FB}(\kappa) \stackrel{\triangle}{=} \sup_{\mathscr{P}_{FB}^{\infty}(\kappa)} \frac{1}{2} \log\Big(\frac{(\Lambda + c)^2 K + K_Z + K_W}{K_W}\Big), \quad \forall \mathbf{P}_{V_1}, \forall K_{V_1} \geq 0, \tag{96}$$

$$\mathscr{P}_{FB}^{\infty}(\kappa) \stackrel{\triangle}{=} \Big\{(\Lambda, K_Z) \in \mathscr{P}_{FB}^{\infty} \Big| K_Z \geq 0, \quad (\Lambda)^2 K + K_Z \leq \kappa\Big\} \tag{97}$$

s.t. $K \geq 0$ *is the unique, stabilizable sol., i.e.,* $|F^{CL}(K, \Lambda, K_Z)| < 1$ *of the generalized ARE,*

$$K = c^2 K + K_W - \frac{\Big(K_W + cK\big(\Lambda + c\big)\Big)^2}{\Big(K_Z + K_W + \big(\Lambda + c\big)^2 K\Big)}, \tag{98}$$

$$F^{CL}(K, \Lambda, K_Z) \stackrel{\triangle}{=} c - M(K, \Lambda, K_Z)\Big(\Lambda + c\Big), \tag{99}$$

$$M(K, \Lambda, K_Z) \stackrel{\triangle}{=} \Big(K_W + cK\Big(\Lambda + c\Big)\Big)\Big(K_Z + K_W + \Big(\Lambda + c\Big)^2 K\Big)^{-1} \tag{100}$$

provided there exists $\kappa \in [0, \infty)$ *such that the set* $\mathscr{P}_{FB}^{\infty}(\kappa)$ *is non-empty (if the set* $\mathscr{P}_{FB}^{\infty}(\kappa)$ *is empty for some* $\kappa \in [0, \infty)$ *we should replace the pair* $\{A^*, B^{*,\frac{1}{2}}\}$ *that is stabilizable by the pair that is unit circle controllable; this is treated in Corollary 2.*

Moreover, $C_{FB}(\kappa)$ *is achieved by the asymptotically time-invariant channel input,*

$$X_t = \Lambda_t\big(V_{t-1} - \widehat{V}_{t-1}\big) + Z_t, \quad X_1 = Z_1, \quad t = 2, \ldots, \quad Z_t \in N(0, K_{Z_t}), \quad K_{Z_t} \geq 0, \tag{101}$$

$$\lim_{n\longrightarrow\infty}(c_n, K_{W_n}) = (c, K_W), K_W > 0, \quad \lim_{n\longrightarrow\infty}(\Lambda_n, K_{Z_n}) = (\Lambda, K_Z), \quad K_Z \geq 0. \tag{102}$$

and the optimal element $(\Lambda, K_Z) = (\Lambda^*, K_Z^*) \in \mathscr{P}_{FB}^{\infty}(\kappa)$ *of* $C_{FB}(\kappa)$ *is such that*

(i) if the noise is stable, i.e., $c \in (-1,1)$ then the input and the output processes (X_t, Y_t), $t = 1, \ldots$ are asymptotic stationary, and

(ii) if the noise is unstable i.e., $c \notin (-1,1)$ then the input and the innovations processes $(X_t, I_t), t = 1, \ldots$ are asymptotic stationary.

Proof. This follows from Theorem 2, Proposition 1. □

Remark 7. *Clearly, Theorem 3 imposes detectability and stabilizability, and as in all optimization problems with constraints, the set $\mathscr{P}_{FB}^{\infty}(\kappa)$ must be non-empty. However, by Lemma 2, a necessary and sufficient condition for stabilizability is $K_Z > 0$. As we show shortly, for the parameters (c, K_W, κ) of Regime 1 a feasible feedback strategy $(\Lambda, K_Z), \Lambda \neq 0, K_Z > 0$ exists in the set $\mathscr{P}_{FB}^{\infty}(\kappa) \subseteq \mathscr{P}_{FB}^{\infty}$. However, for the parameters (c, K_W, κ) of Regimes 2 and 3, a feasible feedback strategy $(\Lambda, K_Z), \Lambda \neq 0, K_Z > 0$ does not exist in the set $\mathscr{P}_{FB}^{\infty}(\kappa) \subseteq \mathscr{P}_{FB}^{\infty}$. Therefore, for Regimes 2 and 3, we relax the stabilizability to a unit circle controllability, and then show there is a feasible feedback strategy (Λ, K_Z) under this relaxation. However, in this case it is necessary that $K_{V_1} > 0$.*

Corollary 2 is a relaxation of Theorem 3 by replacing stabilizability by unit circle controllability.

Corollary 2. *(Variations of Theorem 3—detectability and unit circle Ccontrollability)*

Consider the statement of Theorem 3, and suppose the set $\mathscr{P}_{FB}^{\infty}(\kappa) \subseteq \mathscr{P}_{FB}^{\infty}$ is empty, i.e., there do not exist parameters (c, K_W, κ) and $(\Lambda, K_Z), \Lambda \neq 0$, such that $K_Z > 0$.

Replace the set $\mathscr{P}_{FB}^{\infty}$ in (95) (i.e., the detectability and the stabilizability conditions) by the larger set $\mathscr{P}_{FB}^{\infty,+} \supseteq \mathscr{P}_{FB}^{\infty}$ of detectability and unit circle controllability conditions, defined by

$$\mathscr{P}_{FB}^{\infty,+} \stackrel{\triangle}{=} \Big\{ (\Lambda, K_Z) \in (-\infty, \infty) \times [0, \infty) \Big| \quad \text{(i) the pair } \{A, C\} \equiv \{A, C(\Lambda^{\infty})\} \text{ is detectable,}$$
$$\text{(ii) the pair } \{A^*, B^{*,\frac{1}{2}}\} \equiv \{A^*(K_Z), B^{*,\frac{1}{2}}(K_Z)\} \text{ is unit circle controllable,}$$
$$A, C, A^*, B^{*,\frac{1}{2}} \text{ given by (84)} \Big\}. \tag{103}$$

Then feedback capacity is given by

$$C_{FB}(\kappa, \mathbf{P}_{V_1}) = C_{FB}^{max}(\kappa, \mathbf{P}_{V_1}) \stackrel{\triangle}{=} \sup_{\mathscr{P}_{FB}^{\infty,+}(\kappa)} \frac{1}{2} \log \Big(\frac{(\Lambda + c)^2 K + K_Z + K_W}{K_W} \Big), \quad \text{for some} \quad K_1 \geq 0 \tag{104}$$

$$\mathscr{P}_{FB}^{\infty,+}(\kappa) \stackrel{\triangle}{=} \Big\{ (\Lambda, K_Z) \in \mathscr{P}_{FB}^{\infty,+} \Big| K_Z \geq 0, \quad (\Lambda)^2 K + K_Z \leq \kappa \Big\}, \tag{105}$$

$$\text{s.t. } K \geq 0 \text{ is a stabilizable solution, i.e., } |F^{CL}(K, \Lambda, K_Z)| < 1 \text{ of ARE (98)–(100)} \tag{106}$$

provided there exists $\kappa \in [0, \infty)$ such that the set $\mathscr{P}_{FB}^{\infty,+}(\kappa)$ is non-empty.

Moreover, $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1})$ is generally neither the capacity of asymptotically stationary noise nor the ergodic capacity.

If the optimal value is $K_Z = K_Z^* = 0$, then $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1}) = C_{FB}(\kappa, \mathbf{P}_{V_1}), \forall K_{V_1} > 0$ and is achieved by the asymptotically time-invariant channel input (101) such that $\lim_{n \longrightarrow \infty} K_{Z_n} = K_Z = 0$.

Proof. Suppose in Theorem 3, the set $\mathscr{P}_{FB}^{\infty}(\kappa) \subseteq \mathscr{P}_{FB}^{\infty}$ is empty for some parameter values. Then by Theorem 2.(4), we can relax the conditions of detectability and stabilizability, and replace the set $\mathscr{P}_{FB}^{\infty}$ by the larger set $\mathscr{P}_{FB}^{\infty,+}$ specified by detectability and unit circle controllability. Under this relaxation, by Theorem 2.(4) the limit exists $\lim_{n \longrightarrow \infty} K_n = K$ for some initial condition, $K_1 \geq 0$ (not for all K_1, i.e., $K_1 = 0$ may not be included), where $K \geq 0$ is a stabilizing solution of the generalized ARE (83), but it may not be unique,

i.e., Ref. (83) may have multiple solutions $K \geq 0$, where one of the solutions is the maximal and stabilizing solution, denoted by $K_+ \geq 0$, such that $K_+ - K \geq 0$ w.r.t. any other solution $K \geq 0$. Since in general, under detectability and unit circle cotrolllability conditions, the limit depends on $K_1 \geq 0$ and by (39), $K_1 = \frac{K_{V_1} K_{Z_1}}{K_{V_1} + K_{Z_1}} \geq 0$ depends on K_{V_1}, then $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1})$ is neither the capacity of asymptotically stationary noise nor the ergodic capacity. The last part is obtained as follows. Suppose $K_Z = K_Z^* = 0$ is optimal for $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1})$. By Lemma 2.(5), when $K_Z = K_Z^* = 0$ there are two solutions of the ARE (98)–(100), where one is the zero solution $K = 0$ and the other the maximal solution $K = K_+ = \frac{K_W\left(\Lambda^2 - 1\right)}{\left(\Lambda + c\right)^2} \in (0, \infty)$, which is stabilizing, $|F^{CL}(K, \Lambda, K_Z = 0) = -\frac{1}{\Lambda}| < 1$, if $|\Lambda| > 1$, $c \neq -\Lambda$. It is also verified in the proof of Theorem 5. Since the zero solution $K = 0, K_Z = 0$ gives zero value of $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1})$, we pick the maximal solution $K = K_+ > 0$ which gives a positive value of $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1}), \forall K_1 > 0$. By Theorem 2.(4) and $K_1 = \frac{K_{V_1} K_{Z_1}}{K_{V_1} + K_{Z_1}} \geq 0$, we require $K_{V_1} > 0, K_{Z_1} > 0$; otherwise, the limit of $K_1, K_2, \ldots, K_n$ as $n \longrightarrow \infty$ would correspond to the zero solution $K = 0$. This implies $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1}) = C_{FB}(\kappa, \mathbf{P}_{V_1}), \forall K_{V_1} > 0$ and is achieved by the asymptotically time-invariant channel input (101) such that $\lim_{n \longrightarrow \infty} K_{Z_n} = K_Z = 0$. □

Remark 8. *The point to be made is that there are fundamental differences between the optimization problems of Theorem 3 under the constraint of detectability and stabilizability, and Corollary 2, under the relaxed constraint of detectability and unit circle controllability. This will become even more clear in the next sections, where we derive the optimal values of capacity for Regimes 1–3.*

3.4. Closed-Form Expressions of Asymptotic Feedback Capacity and Lower Bounds on Nonfeedback Capacity

In this section we apply KKT conditions to derive the closed-form expressions of feedback capacity for Regimes 1–3, i.e., $C_{FB}(\kappa)$ $\forall \mathbf{P}_{V_1}, \forall K_{V_1} \geq 0$ and $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1}), \forall \mathbf{P}_{V_1}$ such that $K_{V_1} > 0$.

3.4.1. Closed-Form Expressions of Ergodic Capacity $C_{FB}(\kappa)$ Under Detectability/Stabilizability

First, we give the KKT conditions for $C_{FB}(\kappa)$ defined by (96).

Lemma 3. *(Optimality conditions for ergodic capacity of optimization problem $C_{FB}(\kappa)$—detectability/stabilizability)*

(a) Consider the optimization problem of feedback capacity $C_{FB}(\kappa)$ of (96), which presupposes that the optimal policy $(\Lambda^*, K_Z^*) \in \mathscr{P}_{FB}^{\infty}(\kappa)$ satisfies the detectability and stabilizability conditions.

Define the Lagrangian by

$$\begin{aligned}\mathscr{L}(\Lambda, K_Z, K, \lambda) &\stackrel{\triangle}{=} (\Lambda + c)^2 K + K_Z + K_W - \lambda_1 \Big\{ \Big(K - c^2 K - K_W\Big) \Big(K_Z + K_W + (\Lambda + c)^2 K\Big) \\ &+ \Big(K_W + cK(\Lambda + c)\Big)^2 \Big\} \\ &- \lambda_2 \Big((\Lambda)^2 K + K_Z - \kappa\Big) - \lambda_3 \Big(-K\Big) - \lambda_4 \Big(-K_Z\Big), \quad \lambda \stackrel{\triangle}{=} (\lambda_1, \lambda_2, \lambda_3, \lambda_4) \in \mathbb{R}^4. \end{aligned} \tag{107}$$

(i) The following hold.

Stationarity:

$$\frac{\partial}{\partial K_Z}\mathscr{L}(\Lambda, K_Z, K, \lambda)\Big|_{\Lambda=\Lambda^*, K_Z=K_Z^*, K=K^*, \lambda=\lambda^*} = 0, \tag{108}$$

$$\frac{\partial}{\partial \Lambda}\mathscr{L}(\Lambda, K_Z, K, \lambda)\Big|_{\Lambda=\Lambda^*, K_Z=K_Z^*, K=K^*, \lambda=\lambda^*} = 0, \tag{109}$$

$$\frac{\partial}{\partial K}\mathscr{L}(\Lambda, K_Z, K, \lambda)\Big|_{\Lambda=\Lambda^*, K_Z=K_Z^*, K=K^*, \lambda=\lambda^*} = 0. \tag{110}$$

Complementary Slackness:

$$\lambda_2^*\Big((\Lambda^*)^2K^* + K_Z^* - \kappa\Big) = 0, \quad \lambda_3^*K^* = 0, \quad \lambda_4^*K_Z^* = 0, \tag{111}$$

$$\lambda_1^*\Big\{\Big(K^* - c^2K^* - K_W\Big)\Big(K_Z^* + K_W + \big(\Lambda^* + c\big)^2K^*\Big) + \Big(K_W + cK^*\big(\Lambda^* + c\big)\Big)^2\Big\} = 0. \tag{112}$$

Primal Feasibility:

$$(\Lambda^*)^2K^* + K_Z^* \le \kappa, \quad K_Z^* \ge 0, \quad K^* \ge 0, \tag{113}$$

$$\Big(K^* - c^2K^* - K_W\Big)\Big(K_Z^* + K_W + \big(\Lambda^* + c\big)^2K^*\Big) + \Big(K_W + cK^*\big(\Lambda^* + c\big)\Big)^2 \le 0. \tag{114}$$

Dual Feasibility:

$$\lambda_1^* \ge 0, \quad \lambda_2^* \ge 0, \quad \lambda_3^* \ge 0, \quad \lambda_4^* \ge 0. \tag{115}$$

(ii) If $K_Z^* = 0$ then $K^* = 0$ and $C_{FB}(\kappa) = 0, \forall \kappa \in [0, \infty)$.
(iii) A necessary condition for existence of $\kappa \in (0, \infty)$ such that $C_{FB}(\kappa) > 0$ is $\lambda_1^* > 0$, $\lambda_2^* > 0, \lambda_3^* = 0, \lambda_4^* = 0$, and (114) holds with equality. (b) Suppose the optimization problem of feedback capacity $C_{FB}(\kappa)$ of (96) of part (a) does not produce a policy $(\Lambda^*, K_Z^*) \in \mathscr{P}_{FB}^{\infty}(\kappa)$ such that $C_{FB}(\kappa) > 0$, i.e., the KKT conditions of part (a) do not produce an optimal $K^* > 0$ and $K_Z^* > 0$.

Then feedback capacity is given by the optimization problem $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1})$ of (104), where the optimization is over the larger set $\mathscr{P}_{FB}^{\infty,+}(\kappa)$.

Proof. See Appendix A.2. □

Theorem 4 gives the complete closed-form expressions of the optimization problem $C_{FB}(\kappa)$.

Theorem 4. *(Closed-form expressions for ergodic capacity $C_{FB}(\kappa)$—detectability/stabilizability)*
(1) The non-zero value of $C_{FB}(\kappa)$ defined by (96), for asymptotically time-invariant parameters of the stable and unstable AR(c) noise, i.e, $c \in (-\infty, \infty)$, with $c \neq 0, c \neq 1$, and strategies of the channel input, is given as follows.

$$C_{FB}(\kappa) = \frac{1}{2}\log\Big(\frac{\big(\Lambda^* + c\big)^2K^* + K_Z^* + K_W}{K_W}\Big), \quad \forall \kappa \in \mathscr{K}_1^{\infty}(c, K_W), \tag{116}$$

$$= \frac{1}{2}\log\Big(\frac{c^2\Big(\big(c^2-1\big)\kappa + K_W\Big)}{\big(c^2-1\big)K_W}\Big) \tag{117}$$

where

$$\Lambda^* = \frac{\kappa\left(1-c^2\right)+K_W+c^2K^*}{c\left(c^2-2\right)K^*} \in (-\infty,\infty), \quad K_Z^* + \left(\Lambda^*\right)^2 K^* = \kappa, \tag{118}$$

$$\mathscr{K}_1^{\infty}(c,K_W) \stackrel{\triangle}{=} \left\{\kappa \in [0,\infty): \quad K^*>0, \quad K_Z^*>0\right\}, \quad c \in (-\infty,\infty), \quad c \neq 0,\ c \neq 1 \tag{119}$$

and where K^* is the unique positive and stabilizing solution, i.e., $|F^{CL}(K^*,\Lambda^*,K_Z^*)|<1$, of the quadratic equation

$$\begin{aligned} &c^4\left(c^2-1\right)\left(K^*\right)^2 + c^4\left(\left(1-c^2\right)\kappa+K_W\right)K^* + \left(\left(1-c^2\right)\kappa+K_W\right)^2 \\ &+4\left(c^2-1\right)K_W\kappa - c^4\kappa K_W = 0. \end{aligned} \tag{120}$$

Further, for any $\kappa \in \mathscr{K}_1^{\infty}(c,K_W)$, then

$$K^* = \frac{\kappa\left(c^2-1\right)^2-K_W}{c^2\left(c^2-1\right)} \in (0,\infty), \tag{121}$$

$$\Lambda^* = \frac{cK_W}{\kappa\left(c^2-1\right)^2-K_W} \in (-\infty,\infty), \tag{122}$$

$$K_Z^* = \frac{\kappa\left(c^2-1\right)\left(\kappa\left(c^2-1\right)^2-K_W\right)-K_W^2}{\left(c^2-1\right)\left(\kappa\left(c^2-1\right)^2-K_W\right)} \in (0,\infty). \tag{123}$$

(2) The non-zero value of $C_{FB}(\kappa)$, $\forall \kappa \in \mathscr{K}_1^{\infty}(c,K_W)$ of part (1), is restricted to the region:

(a) Regime 1. $\mathscr{K}_1^{\infty}(c,K_W) = \left\{\kappa \in [0,\infty)\middle| 1<c^2<\infty, \kappa > \frac{K_W+K_W\sqrt{4c^2-3}}{2\left(c^2-1\right)^2}\right\}$.

Moreover, $C_{FB}(\kappa)$, $\forall \kappa \in \mathscr{K}_1^{\infty}(c,K_W)$ is achieved by the optimal channel input,

$$X_n = \Lambda_n^*\left(V_{n-1} - \mathbf{E}\left\{V_{n-1}\middle|Y^{n-1}\right\}\right) + Z_n, \quad X_1 = Z_1, \quad Z_n \in N(0,K_{Z_n}^*), \quad n=1,2,\ldots, \tag{124}$$

$$\text{such that} \quad \lim_{n\longrightarrow\infty}(\Lambda_n^*,K_{Z_n}^*) = (\Lambda^*,K_Z^*), \quad K_Z^*>0, \ \text{the unique solution of part (1).} \tag{125}$$

(3) A non-zero value of feedback capacity $C_{FB}(\kappa)$, i.e., with $\Lambda \neq 0$, does not exist for the two regions:

(a) Regime 2. $\kappa \in \mathscr{K}_2^{\infty}(c,K_W) \stackrel{\triangle}{=} \left\{\kappa \in [0,\infty)\middle| 1<c^2<\infty, \quad \kappa \leq \frac{K_W+K_W\sqrt{4c^2-3}}{2\left(c^2-1\right)^2}\right\}$.

(b) Regime 3. $\kappa \in \mathscr{K}_3^{\infty}(c) \stackrel{\triangle}{=} \left\{\kappa \in [0,\infty)\middle| 0 \leq c^2 \leq 1\right\}$.

Moreover, for Regions 2 and 3, a non-zero value exists for the feedback capacity optimization problem $C_{FB}^{max}(\kappa,\mathbf{P}_{V_1})$ for some $\mathbf{P}_{V_1}$ such that $K_{V_1}>0$ (with $\Lambda \neq 0$).

Proof. See Appendix A.3. □

3.4.2. Closed-Form Expression of Capacity $C_{FB}(\kappa,\mathbf{P}_{V_1})$ Under Detectability/Unit Circle Controllability

Since for the complement of Regime 1 there does not exist a non-zero value solution $C_{FB}(\kappa,\mathbf{P}_{V_1}) = C_{FB}(\kappa) \in (0,\infty)$ for $K_{V_1}=0$, we have have to consider Corollary 2 to derive the closed-form expression of $C_{FB}(\kappa,\mathbf{P}_{V_1}) = C_{FB}^{max}(\kappa,\mathbf{P}_{V_1}), \forall K_{V_1}>0$, where the optimization is over the set $\mathscr{P}_{FB}^{\infty,+}$ in (95), for Regimes 2 and 3.

Theorem 5. *(Closed-form expression of capacity $C_{FB}^{max}(\kappa,\mathbf{P}_{V_1})$—detectability/unit circle controllability)*

The nonergodic capacity $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1})$ *of Corollary 2, where optimization is over the larger set* $\mathscr{P}_{FB}^{\infty,+}$, *is given by*

$$C_{FB}^{max}(\kappa, \mathbf{P}_{V_1}) = \sup_{\Lambda\in(-\infty,\infty),\ (\Lambda)^2K\le\kappa} \frac{1}{2}\log\Big(\frac{(\Lambda+c)^2K+K_W}{K_W}\Big), \quad \forall K_{V_1}>0, \tag{126}$$

$$s.t.\ |\Lambda|\neq 1,\ K=K_+(\Lambda)\ge 0 \text{ is the maximal and stabilizing sol. of the ARE}, \tag{127}$$

$$K = c^2K^{\infty}+K_W-\frac{\Big(K_W+cK(\Lambda+c)\Big)^2}{K_W+(\Lambda+c)^2K},\ K\ge 0, \tag{128}$$

$$K=\lim_{n\longrightarrow\infty}K_n,\quad K_n\stackrel{\triangle}{=}\mathbf{E}\Big\{\Big(V_n-\mathbf{E}\Big\{V_n\Big|Y^n\Big\}\Big)^2\Big\}. \tag{129}$$

Moreover, a non-zero value of feedback capacity $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1})$ *exists for all* $K_{V_1}>0$, *for Regimes 2 and 3, and it given by*

$$C_{FB}^{max}(\kappa, \mathbf{P}_{V_1}) = \log\max\{1,|\Lambda^*|\},\quad \forall K_{V_1}>0,\quad \forall\kappa\in\mathscr{K}_2^{\infty}(c,K_W)\bigcup\mathscr{K}_3^{\infty}(c), \tag{130}$$

where Λ^* *is the maximal root of the fourth order equation [10],*

$$K_W(\Lambda^*)^4-(K_W+\kappa)(\Lambda^*)^2-2c\kappa\Lambda^*-c^2\kappa=0, \tag{131}$$

$$K^*=K_+^*=\frac{K_W\Big((\Lambda^*)^2-1\Big)}{(\Lambda^*+c)^2},\quad |\Lambda^*|>1 \tag{132}$$

and $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1})$ *is achieved by the optimal channel input,*

$$X_n=\Lambda_n^*\Big(V_{n-1}-\mathbf{E}\Big\{V_{n-1}\Big|Y^{n-1}\Big\}\Big),\quad n=2,3,\ldots,X_1=Z_1,\quad Z_1\in N(0,K_{Z_1}^*),\ K_{Z_1}^*>0, \tag{133}$$

$$\text{such that } \lim_{n\longrightarrow\infty}(\Lambda_n^*,K_{Z_n}^*)=(\Lambda^*,K_Z^*),\quad K_Z^*=0, \tag{134}$$

$$(\Lambda,K_Z,K)=(\Lambda^*,K_Z^*,K_+^*) \text{ is the maximal solution of } C_{FB}(\mathbf{P}_{V_1},\kappa),\forall K_{V_1}>0. \tag{135}$$

Proof. See Appendix A.4. □

In the next remark, we clarify and correct some of the misinterpretations of results in [2,24],

Remark 9. *Previous papers [2,24] presented feedback capacity formula* $C_{FB}=\log\max\{1,|\Lambda^*|\}$ *for stable AR(c) noise, i.e.,* $c\in(-1,1)$, *which is included in Regime 3,* $\kappa\in\mathscr{K}_3^{\infty}(c)\stackrel{\triangle}{=}\Big\{\kappa\in[0,\infty)\Big|0\le c^2\le 1\Big\}$. *However, papers [2,24] claimed that their feedback capacity is* $C_{FB}=\log\max\{1,|\Lambda^*|\}$, *and this is the capacity of asymptotically stationary noise* V^n. *This statement should be read with caution because* $K_{V_1}>0$ *should be imposed, and this excludes asymptotically stationary noise* V^n. *However,* $C_{FB}(\kappa,\mathbf{P}_{V_1})=C_{FB},\forall K_{V_1}>0$ *corresponds to the capacity of stationary noise* V^n. *This means, the probability of error* $\mathbf{P}_{error}^{(n)}(\mathbf{P}_{V_1})=\mathbf{P}_{error}^{(n)}(K_{V_1})$, *i.e., depends on* K_{V_1}, *and if* $K_{V_1}=0$ *then* $\lim_{n\longrightarrow\infty}\mathbf{P}_{error}^{(n)}(\mathbf{P}_{V_1})=1$ *(see Gallager [4] for extensive discussion).*

The clarifications of Remark 9 also apply to the TD feedback capacity formulae given in [2] [Theorem 6.1], for general stable noise described by the state space realization [2] [Equation (58)] and also to [24].

First, we state a theorem that first appeared in [27] (and earlier in an arXiv paper [26]).

Theorem 6. *([26,27] Asymptotic feedback capacity with initial state)*

Consider the Gaussian noise $V_n, n = 1, 2$, described by the stable state space realization (see above and below [2]) [Equation (70), Equation (43)],

$$S_{t+1} = AS_t + BW_t, \quad V_t = CS_t + W_t, \qquad t = 1, 2, \ldots \tag{136}$$

where $S_t : \Omega \to \mathbb{R}^{n_s}$ is the state of the noise (n_s a finite positive integer), A, B, C are specific matrices of dimensions $n_s \times n_s, n_s \times 1, 1 \times n_s$, respectively, $W_t \in N(0, K_W), K_W > 0, t = 1, \ldots, n$ is the independent Gaussian process, independent of $S_1 \in N(0, K_{S_1}), K_{S_1} \succeq 0$, and the eigenvalues of A lie inside the unit disk in the space of complex numbers, i.e., A is a stable matrix, meaning its poles and zeros inside the unit disk ([2], Theorem 6.1 consider the value $K_W = 1$).

Suppose Conditions 1 and 2 hold.

Condition 1. *The initial state of the noise or channel $S_1 = s$ is known to the encoder and decoder.*

Condition 2. *Given a fixed initial state $S_1 = s$, known to the encoder and the decoder, at each t, the state space realization of the noise is such that the noise V^{t-1} uniquely defines the state of the noise S^t and vice-versa, $\forall t$.*

(1) An equivalent sequential characterization of Cover and Pombra [3] feedback capacity $C_{n,FB}(\kappa, \mathbf{P}_{V_1})$ is

$$C_{n,FB}(\kappa, \mathbf{P}_{V_1}) = C_{n,FB}^{AStatNoise}(\kappa, s)$$

$$\stackrel{\triangle}{=} \sup_{\{\Lambda_t, K_{Z_t}\}_{t=1}^n,\ \frac{1}{n}\sum_{t=1}^n \left(\Lambda_t K_t \Lambda_t^T + K_{Z_t}\right) \le \kappa} \frac{1}{2n} \sum_{t=1}^n \log\left(\frac{\Big(\Lambda_t + C\Big) K_t \Big(\Lambda_t + C\Big)^T + K_{Z_t} + K_W}{K_W}\right), \tag{137}$$

$$s.t.\ K_{Z_t} \ge 0, \quad \forall t, \quad K_t \stackrel{\triangle}{=} \mathbf{E}\Big\{\big(S_t - \mathbf{E}\{S_t | Y^{t-1}, S_1 = s\}\big)\big(S_t - \mathbf{E}\{S_t | Y^{t-1}, S_1 = s\}\big)^T\Big\} \tag{138}$$

a sol. of the matrix DRE $K_t \succeq 0$ of (139),

$$K_{t+1} = AK_tA^T + BK_WB^T - \Big(BK_W + AK_t\big(\Lambda_t + C\big)^T\Big)\Big(K_W + \big(\Lambda + C\big)K\big(\Lambda_t + C\big)^T + K_{Z_t}\Big)^{-1}$$
$$\cdot\Big(BK_W + AK\big(\Lambda_t + C\big)^T\Big)^T, \qquad K_1 = 0. \tag{139}$$

where $K_t \in \mathbb{R}^{n_s \times n_s}$ and $K_t \succeq 0, \forall t$ means a symmetric positive semidefinite matrix.

Moreover, the optimal channel input corresponding to $C_{n,FB}^{AStatNoise}(\kappa, \mathbf{P}_{V_1})$ is given by

$$X_t = \Lambda_t\Big(S_t - \mathbf{E}\Big\{S_t\Big|Y^{t-1}, S_1 = s\Big\}\Big) + Z_t, \quad t = 1, 2, \ldots, n. \tag{140}$$

$$Z_t \in N(0, K_{Z_t}), \quad K_{Z_t} \ge 0, \quad t = 1, 2, \ldots, n \quad \textit{orthogoal process and indp. of } (S^t, V^t, Y^{t-1}). \tag{141}$$

(2) Suppose the (i) detectability condition and (ii) unit circle controllability condition of the matrix algebraic Riccati equation (ARE) (144) hold.

The asymptotic feedback capacity is

$$C_{FB}(\kappa, \mathbf{P}_{V_1}) = C_{FB}^{AStaNoise}(\kappa, s) = \lim_{n \longrightarrow \infty} C_{n,FB}^{AStatNoise}(\kappa, s)$$

$$\stackrel{\triangle}{=} \sup_{\Lambda \in \mathbb{R}^{1 \times n_s},\, K_Z,\, \Lambda K \Lambda^T + K_Z \leq \kappa} \frac{1}{2} \log \Big(\frac{\Big(\Lambda + C\Big) K \Big(\Lambda + C\Big)^T + K_Z + K_W}{K_W} \Big), \tag{142}$$

s.t. $K_Z \geq 0$, K *is a positive semi-definite solution* $K \succeq 0$ *of the ARE (144),* (143)

$$K = AKA^T + BK_WB^T - \Big(BK_W + AK\big(\Lambda + C\big)^T\Big)\Big(K_W + \big(\Lambda + C\big)K\big(\Lambda + C\big)^T + K_Z\Big)^{-1}$$

$$.\Big(BK_W + AK\big(\Lambda + C\big)^T\Big)^T \tag{144}$$

where (K_Z, Λ, K) *are the asymptotic limits* $(K_Z, \Lambda, K) \stackrel{\triangle}{=} \lim_{n \longrightarrow \infty}(K_{Z_n}, \Lambda_n, K_n)$. (145)

Moreover, the following hold.

(2.1) If the optimal pair (Λ, K_Z) *is such that* $K_Z > 0$ *then* $C_{FB}^{AStatNoise}(\kappa, s)$ *is independent of the initial state* $S_1 = s$ *and is achieved by the optimal channel input,*

$$X_t = \Lambda\Big(S_t - \mathbf{E}\Big\{S_t \Big| Y^{t-1}, S_1 = s\Big\}\Big) + Z_t, \quad t = 1, 2, \ldots, n. \tag{146}$$

$$Z_t \in N(0, K_Z), \quad K_Z > 0, \quad t = 1, 2, \ldots, n \quad \textit{orthogoal process and indp. of } (S^t, V^t, Y^{t-1}). \tag{147}$$

(2.2) If the optimal pair (Λ, K_Z) *is such that* $K_Z = 0$ *then* $C_{FB}^{AStatNoise}(\kappa, s)$ *depends on the initial state* $S_1 = s$ *and is achieved by*

$$X_1 = Z_1 \in N(0, K_{Z_1}),\ K_{Z_1} > 0, \quad X_t = \Lambda\Big(S_t - \mathbf{E}\Big\{S_t \Big| Y^{t-1}, S_1 = s\Big\}\Big), \quad t = 2, \ldots, n. \tag{148}$$

(3) The asymptotic characterization of feedback capacity [2] [Theorem 6.1] corresponds to $C_{FB}(\kappa, \mathbf{P}_{V_1}) = C_{FB}^{AStaNoise}(\kappa, s)$, *and optimal pair* (Λ, K_Z) *such that* $K_Z = 0$.

Proof. Statement (1) and (2) are shown in [26,27]. Statement (3) is also discussed in [26,27]. It also follows from the simple observation that according to Cover and Pombra [3], the $n-$FTFI capacity $C_{n,FB}(\kappa, \mathbf{P}_{V_1})$ corresponds to an optimal channel input, $X_t = \sum_{j=1}^{t-1} b_{t,j} V_j + \overline{Z}_t$, with corresponding output $Y_t = \sum_{j=1}^{t-1} b_{t,j} V_j + \overline{Z}_t + V_t, t = 1, \ldots, n$, where for some jointly Gaussian process $\overline{Z}^n \in N(0, K_{\overline{Z}^n})$ and nonrandom coefficients $b_{t,j}$. Clearly, Theorem 6 is inconsistent with the Cover and Pombra [3], the $n-$FTFI capacity $C_{n,FB}(\kappa, \mathbf{P}_{V_1})$, unless the optimal input is expressed as a function of the state S^n of the noise instead of the noise S^n. It is now easy to verify that given a state space realization, Conditions 1 and 2 are necessary and sufficient for the validity of Theorem 6. *To verify Conditions 1 and 2, note the following.* If Condition 1 holds, i.e., $S_1 = s$ is known to the encoder, then from the state space realization, (136), $V_1 = Cs_1 + BW_1$, hence $W_1 = V_1 - Cs_1$ is uniquely determined from (V_1, s_1), and this specifies $S_2 = As_1 + BW_1 = As_1 + BV_1 - BCs_1$, etc., therefore Condition 2 holds. From [2] [Theorem 6.1], it is clear that optimal pair (Λ, K_Z) is taken such that $K_Z = 0$ (without a valid proof). □

Remark 10. *We emphasize that* $C_{n,FB}^{AStatNoise}(\kappa, \mathbf{P}_{V_1})$ *is specific to the state space realization, [2]) [Equation (70), Equation (43)], i.e., (136), and does not hold for the alternative realization,* $S_{t+1} = AS_t + BW_t^1, V_t = CS_t + W_t^2, t = 1, \ldots, n$, *where* $W_t^1, t = 1, \ldots, n$ *and* $W_t^2, t = 1, \ldots, n$ *are two independent Gaussian processes.*

3.4.3. Nonfeedback Ergodic Achievable Rates for Stable and Unstable AR Noise

By Theorem 3, for the value $\Lambda = 0$, the pair $\{A, C\}$ is detectable and the pair $\{A^*, B^{*,\frac{1}{2}}\}$ is stabilizable, hence $C_{FB}(\kappa)\Big|_{\Lambda=0}$ is a nonfeedback achievable rate, i.e., by (101), for $\Lambda = 0$, the channel input is an independent innovations process, $X_t = Z_t, t = 1, \ldots, n$, and hence the code does not use feedback. In the next theorem we calculate $C_{FB}(\kappa)\Big|_{\Lambda=0}$.

Theorem 7. *Ergodic achievable rates without feedback for stable and unstable AR(c) noise*

Consider the optimization problem of feedback capacity $C_{FB}(\kappa)$ of Theorem 3, with $\Lambda = 0$, and corresponding channel input and output processes,

$$X_t = Z_t, \quad t = 1, \ldots, n, \tag{149}$$

$$V_t = c_t V_{t-1} + W_t, \quad t = 2, , \ldots, n, \tag{150}$$

$$Y_t = X_t + V_t = c_t V_{t-1} + W_t + Z_t, \quad Y_1 = Z_1 + W_1, \quad t = 2, \ldots, n. \tag{151}$$

(1) A lower bound on the asymptotic characterization of nonfeedback capacity, $C(\kappa, \mathbf{P}_{V_1}) = \lim_{n \longrightarrow \infty} C_n(\kappa, \mathbf{P}_{V_1})$, is

$$C(\kappa, \mathbf{P}_{V_1}) \geq C_{FB}(\kappa)\Big|_{\Lambda=0} = C^{LB}(\kappa) \stackrel{\triangle}{=} \max_{\mathscr{P}_0^{\infty,nfb}(\kappa)} \frac{1}{2} \log\Big(\frac{c^2K + K_Z + K_W}{K_W}\Big), \quad \forall \mathbf{P}_{V_1}, \forall K_{V_1} \geq 0 \tag{152}$$

where

$$\mathscr{P}_0^{\infty,nfb}(\kappa) \stackrel{\triangle}{=} \Big\{K_Z \in \mathscr{P}_0^{\infty,nfb}\Big| \quad K_Z \leq \kappa\Big\}, \tag{153}$$

$$\mathscr{P}_0^{\infty,nfb} \stackrel{\triangle}{=} \Big\{K_Z \in [0,\infty)\Big| \quad \textit{(i) the pair } \{A, C\}\Big|_{\Lambda=0} \equiv \{A, C(\Lambda)\}\Big|_{\Lambda=0} \textit{ is detectable,}$$
$$\textit{(ii) the pair } \{A^*, B^{*,\frac{1}{2}}\}\Big|_{\Lambda=0} \equiv \{A^*(K_Z), B^{*,\frac{1}{2}}(K_Z)\}\Big|_{\Lambda=0} \textit{ is stabilizable}\Big\}, \tag{154}$$

$K \geq 0$ *is the unique and stabilizing of (156), i.e.,* $|F^{nfb}(K, K_Z)| < 1$, (155)

$$K = c^2K + K_W - \frac{\Big(K_W + c^2K\Big)^2}{\Big(K_Z + K_W + c^2K\Big)}, \tag{156}$$

$$F^{CL,nfb}(K, K_Z) \stackrel{\triangle}{=} c - M^{nfb}(K, K_Z)c, \quad M^{nfb}(K, K_Z) \stackrel{\triangle}{=} \Big(K_W + c^2K\Big)\Big(K_Z + K_W + c^2K\Big)^{-1} \tag{157}$$

provided there exists $\kappa \in [0, \infty)$ such that the set $\mathscr{P}_0^{\infty,nfb}(\kappa)$ is non-empty.

Moreover, $C^{LB}(\kappa)$ is an achievable rate without feedback, independent of the initial state $K_1 \geq 0$ and

(i) if the noise is stable, i.e., $c \in (-1, 1)$ then the input and the output processes (X_t, Y_t), $t = 1, \ldots$ are asymptotically stationary;

(ii) if the noise is unstable i.e., $c \notin (-1, 1)$ then the input and the innovations processes $(X_t, I_t \stackrel{\triangle}{=} Y_t - E\{Y_t|Y^{t-1}\}), t = 1, \ldots$ are asymptotic stationary.

(2) The lower bound on $C(\kappa, \mathbf{P}_{Y_1})$ of (1) is given by

$$C^{LB}(\kappa) = \frac{1}{2} \log\Big(\frac{c^2K^* + \kappa + K_W}{K_W}\Big), \quad \forall \kappa \in \mathscr{K}^{\infty,nfb}(c, K_w), \tag{158}$$

$$\mathscr{K}^{\infty,nfb}(c, K_W) \stackrel{\triangle}{=} \Big\{\kappa \in [0,\infty)\Big| \quad K \geq 0\Big\} = [0,\infty) \tag{159}$$

where $K^* \geq 0$ is unique and stabilizing, and K_Z^*, are given by

$$K^* = \begin{cases} \frac{-\kappa\left(1-c^2\right)-K_W+\sqrt{\left(\kappa\left(1-c^2\right)+K_W\right)^2+4c^2K_W\kappa}}{2c^2}, & \forall c \neq 0, \quad \forall \kappa \in \mathscr{K}^{\infty,nfb}(c,K_W), \\ \frac{\kappa K_W}{\kappa+K_W}, & c=0, \quad \forall \kappa \in [0,\infty), \end{cases} \tag{160}$$

$$K_Z^* = \kappa. \tag{161}$$

Proof. (1) The statements follow from Theorem 3 as a special case. (2) This follows from (1), since the optimal K_Z occurs on the boundary, i.e., $K_Z^* = \kappa$. Then by substituting $(\Lambda^*, K_Z^*) = (0, \kappa)$ into the generalized ARE (156), we obtain

$$c^2\left(K^*\right)^2 + K^*\left\{\kappa\left(1-c^2\right)+K_W\right\} - K_W\kappa = 0, \quad c \neq 0. \tag{162}$$

From (162), the unique and non-negative stabilizing solution of the generalized ARE is given by (160), for $c \neq 0$. For $c = 0$, the generalized ARE reduces to $K^*\{\kappa + K_W\} - K_W\kappa = 0$, and hence all equations under (2) are obtained. The validity of the stability condition, i.e., $|F^{CL,nfb}(K, K_Z)| < 1$, although it is ensured by the conditions, can be easily verified. □

Remark 11. *By Theorem 7.(2) the lower bound $C^{LB}(\kappa), \kappa \in [0,\infty)$ holds for asymptotically time-invariant stable and unstable AR(c) noise.*

(1) For $c = 0$, we recover, as expected, the capacity of the AGN channel with memoryless noise.

(2) For $|c| = 1$, we obtain the lower bound $C^{LB}(\kappa)$ on the nonfeedback capacity given by

$$C^{LB}(\kappa) = \frac{1}{2}\log\left(\frac{c^2K^*+\kappa+K_W}{K_W}\right), \quad K^* = \frac{-K_W+\sqrt{K_W^2+4K_W\kappa}}{2}, \quad K_Z^* = \kappa \in [0,\infty). \tag{163}$$

The above choice of a channel input strategy ensures the pair $\{A, C\}\Big|_{\Lambda^*=0} = \{c, c\}$ is detectable, and the pair $\{A^*, B^{*,\frac{1}{2}}\}\Big|_{K_Z^*=\kappa} = \{c - \frac{K_W}{\kappa+K_W}, K_W^{\frac{1}{2}}\sqrt{\left(1-K_W\left(\kappa+K_W\right)^{-1}\right.}\}$ is stabilizable, for any $\kappa \in (0,\infty)$ (since $K_W > 0$), and hence this is an ergodic achievable rate. We note that the nonfeedback capacity, when $|c| \notin (0,1)$, cannot be handled by the frequency-domain water-filling formula [4] because V_n is neither stationary nor asymptotically stationary.

4. Conclusions

This paper considers additive Gaussian noise (AGN) channels when the noise is nonstationary and nonergodic. It presents closed-form feedback capacity expressions when the noise is an autoregressive (AR) unstable process and an AR stable process (i.e., asymptotically stationary or stationary). These are described by three Regimes which depend on the parameters of the AR noise. It is shown that the feedback capacity for unstable AR noise is much higher compared to the feedback capacity for stable AR noise (i.e., asymptotically stationary or stationary). Similarly, it is shown that achievable lower bounds on nonfeedback rates for unstable AR noise are much higher than the nonfeedback capacity for stable AR noise. The derivations make use of the asymptotic properties of difference Riccati equations to their corresponding limits described by unique solutions of algebraic Riccati equations and KKT conditions. From the closed-form expressions, it follows that for unstable noise, the feedback capacity is an ergodic capacity, i.e., it does not depend on the initial distribution of the AR noise, while for stable noise it depends on the initial distribution noise, and hence it does not correspond to the feedback capacity for an asymptotically stationary noise. This observation is also shown to hold for more general

noise described by the state space realization in [2]. This implies that the TD characterization of feedback capacity for stable noise stated in [2] [Theorem 4.1] is correct, under certain assumptions which are identified in this paper (these are precisely the assumptions of the earlier paper [20]), and the additional condition that the optimal covariance of the innovations part of the input is asymptotically zero. Our closed-form feedback capacity expressions provide fundamental insights and answers to the comment paper [1] that identified gaps in the proof of [2] [Theorem 4.1].

Author Contributions: Conceptualization, C.D.C., C.K., S.L. (Stelios Louka) and S.L. (Sergey Loyka); Methodology, C.D.C., C.K., S.L. (Stelios Louka) and S.L. (Sergey Loyka); Writing—review & editing, C.D.C., C.K., S.L. (Stelios Louka) and S.L. (Sergey Loyka). All authors have read and agreed to the published version of the manuscript.

Funding: The work of C.D. Charalambous was co-funded by the European Regional Development Fund and the Republic of Cyprus through the Research and Innovation Foundation (Project: EXCELLENCE/1216/0296).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of this study; in the collection, analyses, or interpretation of data; in the writing of this manuscript; or in the decision to publish the results.

Appendix A

Appendix A.1. Proof of Lemma 2

Recall (84) and (85). (1.a) By definition $A = c, C = \Lambda + c$. By $c \in (-1, 1)$, there always exists a $G \in \mathbb{R}$ such that $|A - GC| = |c - G(\Lambda + c)| < 1$, i.e., take $G = 0$. (1.b) Since $\Lambda \neq -c$, we can easily verify the existence of $G \in \mathbb{R}$ such that $|c - G(\Lambda + c)| < 1$. Of course, it follows from Lemma 1. (1.c) Since $K_Z > 0$ then $B^{*,\frac{1}{2}} \neq 0$ and the statement follows from Lemma 1. This shows (1).

(2) Since $K_Z = 0$, then $B = 1 - K_W\big(K_Z + K_W\big)^{-1} = 0, B^{*,\frac{1}{2}} = K_W^{\frac{1}{2}} B^{\frac{1}{2}} = 0$, $A^* = c - K_W\big(K_W + K_Z\big)^{-1} C = c - \Lambda - c = -\Lambda$, and hence, there exists a $G \in \mathbb{R}$ such that $|A^* - B^{*,\frac{1}{2}} G| = |-\Lambda| \neq 1$ if and only if $|\Lambda| \neq 1$. This shows (2).

(3) Since $K_Z = 0$, similar to the proof in (2), there exists a $G \in \mathbb{R}$ such that $|A^* - B^{*,\frac{1}{2}} G| = |-\Lambda| < 1$ if and only if $|\Lambda| < 1$. This shows (3).

(4) Since $c \in (-1, 1), K_Z = 0$, then, by (1), the pair $\{A, C\}$ is detectable, by (2) the pair $\{A^*, B^{*,\frac{1}{2}}\}$ is unit circle controllable if and only if $|\Lambda| \neq 1$, and by (3) the pair $\{A^*, B^{*,\frac{1}{2}}\}$ is stabilizable if and only if $|\Lambda| < 1$. By Theorem 2.(1), (4), we deduce the claim. This shows (4).

(5) Clearly, (89) is equivalent to the quadratic equation $K^2\big(\Lambda + c\big)^2 + K\big(K_W - K_W\Lambda^2\big) = 0$. Hence, the two solutions are (92). The last statement is also obtained by applying Theorem 2.(2), as follows. By (1), $\{A, C\}$ is detectable, by (2), $\{A^*, B^{*,\frac{1}{2}}\}$ is unit circle controllable if and only if $|\Lambda| \neq 1$, and by (3), $\{A^*, B^{*,\frac{1}{2}}\}$ is stabilizable if and only if $|\Lambda| < 1$. By invoking Theorem 2.(2), the last statement follows. On the other hand, it is easily verified from (92), that the uniqueness of solutions $K \geq 0$ holds if and only if $|\Lambda| < 1$, because for $|\Lambda| > 1$ there are two non-negative solutions. By (87), evaluated at $K_Z = 0, K = 0$, we have $F^{CL}(K, \Lambda, K_Z)\Big|_{K_Z=0, K=0} = -\Lambda$. Hence, the non-negative solution $K = 0$ is unique and stabilizable if and only if $|\Lambda| < 1$. Similarly, it can be shown that $F^{CL}(K, \Lambda, K_Z)\Big|_{K_Z=0, K=K_+}$ is also stabilizable if and only if $|\Lambda| > 1$.

Appendix A.2. Proof of Lemma 3

(a) (i) The conditions are a consequence of the optimization problem; Equation (114) are due to the standard relaxation of the generalized ARE to an inequality $K \leq c^2K + K_W - \frac{\Big(K_W+cK\big(\Lambda+c\big)\Big)^2}{\Big(K_Z+K_W+\big(\Lambda+c\big)^2\Big)}, K \geq 0.$

(ii) Suppose $K_Z^* = 0$. By (114), with $K_Z^* = 0$, then

$$\Big(K^* - c^2K^* - K_W\Big)\Big(K_W + (\Lambda^* + c)^2K^*\Big) + \Big(K_W + cK^*(\Lambda^* + c)\Big)^2 \leq 0, \tag{A1}$$

$$\implies \quad K^*\Big(K^*(\Lambda^* + c)^2 + K_W\Big(1 - (\Lambda^*)^2\Big)\Big) \leq 0. \tag{A2}$$

By Lemma 2, a necessary and sufficient condition for stabilizability of the pair $\{A^*, B^{*,\frac{1}{2}}\}$, when $K_Z^* = 0$, is $|\Lambda^*| < 1$, and therefore (A2) is satisfied if and only if K^* lies in the region

$$\frac{K_W\Big((\Lambda^*)^2 - 1\Big)}{\Big(\Lambda^* + c\Big)^2} \leq K^* \leq 0, \quad \text{for} \quad |\Lambda^*| < 1. \tag{A3}$$

Since it must be that $K^* \geq 0$ then necessarily $K^* = 0$, which implies $C_{FB}^{\infty}(\kappa) = 0, \forall \kappa \in [0, \infty)$.
(iii) By the stationarity conditions (108)–(110), with $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$:

$$\frac{\partial}{\partial K_Z}\mathscr{L}(\Lambda^*, K_Z^*, K^*, \lambda^*) = 1 - \lambda_1^*\Big(K^* - c^2K^* - K_W\Big) - \lambda_2^* + \lambda_4^* = 0, \tag{A4}$$

$$\begin{aligned}\frac{\partial}{\partial \Lambda}\mathscr{L}(\Lambda^*, K_Z^*, K^*, \lambda^*) &= \Big(\Lambda^* + c\Big)K^* - \lambda_1^*\Big\{\Big(K^* - c^2K^* - K_W\Big)\Big(\Lambda^* + c\Big)K^* \\ &+ \Big(K_W + cK^*\big(\Lambda^* + c\big)\Big)cK^*\Big\} - \lambda_2^*\Lambda^*K^* = 0,\end{aligned} \tag{A5}$$

$$\begin{aligned}\frac{\partial}{\partial K}\mathscr{L}(\Lambda^*, K_Z^*, K^*, \lambda^*) &= \Big(\Lambda^* + c\Big)^2 - \lambda_1^*\Big\{\Big(1 - c^2\Big)\Big(K_Z^* + K_W + (\Lambda^* + c)^2K^*\Big) \\ &+ \Big(K^* - c^2K^* - K_W\Big)\Big(\Lambda^* + c\Big)^2 + 2c\Big(K_W + cK^*\big(\Lambda^* + c\big)\Big)\Big(\Lambda^* + c\Big)\Big\} \\ &- \lambda_2^*\Big(\Lambda^*\Big)^2 + \lambda_3^* = 0,\end{aligned} \tag{A6}$$

Suppose $\lambda_4^* \neq 0$. Then, by complementary slackness (111), we have $\lambda_4^*K_Z^* = 0$, which implies $K_Z^* = 0$, and hence by (ii), $K^* = 0$. By complementary slackness (111), we also have $\lambda_2^*\big((\Lambda^*)^2K^* + K_Z^* - \kappa\big) = \lambda_2^*(0 - \kappa) = 0$, hence $\lambda_2^* = 0$. By (ii) it follows that $C(\kappa) = 0, \forall \kappa \in [0, \infty)$, hence the rate is zero. Similarly, if $\lambda_3^* \neq 0$ then $K_Z^* = 0$ and $K^* = 0$, which lead to a zero rate. However, it can be verified from Theorem 5 that in this case the feedback capacity is given by the relaxed optimization problem $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1})$ that imposes the relaxed conditions of detectability and unit circle controllability. Also, by Theorem 7, for $\Lambda^* = 0, K_Z^* \neq 0$, we exhibit a non-zero rate, which is a lower bound on the non-feedback rate. Next, consider the case $\lambda_3^* = 0, \lambda_4^* = 0$ and $\lambda_1^* = 0$. Then, by (A4) $\lambda_2^* = 1$. However, when $\lambda_2^* = 1$ equalities (A5) and (A6) hold only if $c = 0$. Since, $c \neq 0$, otherwise the channel is memoryless, then the only choice is $\lambda_1^* > 0$. Moreover, since $\lambda_1^* > 0$, then in order to satisfy the complementary slackness condition (114), then inequality must hold with equality, i.e.,

$$\Big(K^* - c^2K^* - K_W\Big)\Big(K_Z^* + K_W + (\Lambda^* + c)^2K^*\Big) + \Big(K_W + cK^*(\Lambda^* + c)\Big)^2 = 0. \tag{A7}$$

Finally we consider the case $\lambda_3^* = 0, \lambda_4^* = 0$ and $\lambda_2^* = 0$. Solving the system of Equations (111), (A4)–(A7), the following sets of solutions are obtained.
The first solution is

$$K^* = \frac{K_W + 1}{1 - c^2}, \quad \Lambda^* = -\frac{c^2 + K_W}{c(K_W + 1)}, \quad K_Z^* = -\frac{K_W(c^2 + K_W)}{c^2(K_W + 1)}, \quad \lambda_1^* = 1. \tag{A8}$$

The second solution is

$$K^* = 0, \quad \Lambda^* = -\frac{c^2 + 1}{2c}, \quad K_Z^* = 0, \quad \lambda_1^* = -\frac{1}{K_W}. \tag{A9}$$

The first solution is discarded since for $K_W > 0$, then, $K_Z^* < 0$, while the second solution is discarded due to (ii). Therefore $\lambda_2^* \neq 0$, which by the complementary slackness condition (111) implies that

$$(\Lambda^{\infty,*})^2 K^* + K_Z^* = \kappa \tag{A10}$$

Thus, we have shown that a necessary condition for existence of $\kappa \in (0, \infty)$ such that $C_{FB}^{\infty}(\kappa) > 0$ is $\lambda_1^* > 0$, $\lambda_2^* > 0$, $\lambda_3^* = 0$ and $\lambda_4^* = 0$.

(b) If there does not exist a strategy in the set $\mathscr{P}_{FB}^{\infty}(\kappa)$ such that $C_{FB}(\kappa) > 0$, then we show in Theorem 5 that we can always consider the relaxation $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1})$, where the optimization problem is over the larger set $\mathscr{P}_{FB}^{\infty,+}(\kappa)$ and exhibits a positive rate. This completes the proof.

Appendix A.3. Proof of Theorem 4.

We prove the statements in several steps.

(i) First, we recall Lemma 3.(a).(ii) that states if $K_Z^* = 0$ then $K^* = 0$, and $C_{FB}(\kappa) = 0, \forall \kappa \in [0, \infty)$. In this case we shall consider the relaxed expression of feedback capacity $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1})$ (see Theorem 5), to exhibit a non-zero rate.

(ii) Second, we recall Lemma 3.(a).(iii), which states a necessary condition for existence of a non-zero feedback rate for some $\kappa \in (0, \infty)$ is $\lambda_1^* > 0$, $\lambda_2^* > 0$, $\lambda_3^* = 0$ and $\lambda_4^* = 0$. For the rest of the derivation we characterize the set of all values $\kappa \in \mathscr{K}_1^{\infty}(c, K_W)$ if such exist, and treat the case when $\mathscr{K}_1^{\infty}(c, K_W)$ is empty separately, using the relaxed expression of feedback capacity $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1})$ (see Theorem 5).

(iii) Consider $\lambda_1^* > 0$, $\lambda_2^* > 0$, $\lambda_3^* = 0$ and $\lambda_4^* = 0$. We solve the system of Equations (A4)–(A7) and (A10). First we solve the system of Equations (A4) and (A5) to obtain λ_1^* and λ_2^* as a function of $\{K^*, K_Z^*, \Lambda^*\}$. By substituting λ_1^*, λ_2^* and K_Z^* from (A10) in (A6), we obtain (118). Finally, by substituting K_Z^* and Λ^* in (A7) we deduce the quadratic Equation (120). The two solutions of the the quadratic Equation (120) give rise to the following two solutions. The first solution is

$$K^* = K_1^* = \frac{\kappa\left(c^2-1\right)^2 - K_W}{c^2\left(c^2-1\right)}, \tag{A11}$$

$$\Lambda^* = \Lambda_1^* = \frac{cK_W}{\kappa\left(c^2-1\right)^2 - K_W}, \tag{A12}$$

$$K_Z^* = K_{Z_1}^* = \frac{\kappa\left\{\left(\kappa\left(c^2-1\right)^2 - K_W\right)\left(c^2-1\right)\right\} - K_W^2}{\left(\kappa\left(c^2-1\right)^2 - K_W\right)\left(c^2-1\right)}. \tag{A13}$$

$$\lambda_1^* = \frac{c^2}{K_W - \kappa\left(1-c^2\right)}, \quad \lambda_2^* = c^2. \tag{A14}$$

The second solution is

$$K^* = K_2^* = \frac{\kappa - K_W}{c^2}, \quad \Lambda^* = \Lambda_2^* = \frac{c\kappa}{K_W - \kappa}, \quad K_Z^* = K_{Z_2}^* = \frac{\kappa K_W}{K_W - \kappa}, \tag{A15}$$

$$\lambda_1^* = \frac{c^2}{\kappa - K_W\left(1-c^2\right)}, \quad \lambda_2^* = 0. \tag{A16}$$

Solutions $K^* = K_2^*, \Lambda^* = \Lambda_2^*, K_Z^* = K_{Z_2}^*$, $(\lambda_1^*, \lambda_2^*)$ given by (A16) are not valid solutions, because if $K^* = K_2^* > 0$, then $K_Z^* = K_{Z_2}^* < 0$, and vice-versa. Thus, the only valid solution is Solution 1, from which all statements of (1) are obtained. It remains to show the statements under (2) and (3).

(iv) Consider $c^2 < 1$ and define the set

$$\mathscr{A}_1(c, K_W) \stackrel{\triangle}{=} \left\{\kappa \in [0,\infty) \,\middle|\, K^* = K_1^* > 0,\ K_Z^* = K_{Z_1}^* > 0,\ c^2 < 1, c \neq 0,\ \lambda_1^* > 0, \lambda_2^* > 0\right\}. \tag{A17}$$

Similarly, define the sets for $c^2 > 1$

$$\mathscr{A}_2(c, K_W) \stackrel{\triangle}{=} \left\{\kappa \in [0,\infty) \,\middle|\, K^* = K_1^* > 0,\ K_Z^* = K_{Z_1}^* > 0,\ c^2 > 1, c \neq 0,\ \lambda_1^* > 0, \lambda_2^* > 0\right\} \tag{A18}$$

and the set for $c^2 = 1$

$$\mathscr{A}_3(c, K_W) \stackrel{\triangle}{=} \left\{\kappa \in [0,\infty) \,\middle|\, K^* = K_1^* > 0,\ K_Z^* = K_{Z_1}^* > 0,\ c^2 = 1, c \neq 0,\ \lambda_1^* > 0, \lambda_2^* > 0\right\} \tag{A19}$$

The proof is then completed by determining the values of $\kappa \in \mathscr{A}_1(c, K_W)$ and $\kappa \in \mathscr{A}_2(c, K_W)$, if such exist. If the set is empty then, we need to consider the relaxed expression of feedback capacity $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1})$ (see Theorem 5).

(iv.1) For $c^2 < 1$, we have the following

$$K^* = K_1^* > 0 \implies \kappa < \frac{K_W}{(c^2-1)^2}, \tag{A20}$$

$$\lambda_1^* > 0 \implies \kappa < \frac{K_W}{(1-c^2)}, \tag{A21}$$

$$K_Z^* = K_{Z_1}^* > 0 \implies \kappa^2\left(c^2-1\right)^3 - \kappa K_W\left(c^2-1\right) - K_W^2 > 0, \tag{A22}$$

$$\implies \frac{K_W - K_W\sqrt{4c^2-3}}{2\left(1-c^2\right)^2} < \kappa < \frac{K_W + K_W\sqrt{4c^2-3}}{2\left(1-c^2\right)^2}, \quad 4c^2 \geq 3. \tag{A23}$$

Next, we show the set $\mathscr{A}_1(c, K_W)$ is empty. For $4c^2 < 3$, $\kappa^2\left(c^2-1\right)^3 - \kappa K_W\left(c^2-1\right) - K_W^2 < 0$, thus $K_Z^* < 0$. For $4c^2 \geq 3$, and from (A21) and (A23), it suffices to show that

$$\frac{K_W - K_W\sqrt{4c^2-3}}{2\left(1-c^2\right)^2} > \frac{K_W}{1-c^2}, \quad \text{for} \quad 4c^2 \geq 3. \tag{A24}$$

Clearly, after simple algebra, we can show that (A24) holds if $\left(c^2-1\right)^2 > 0, \quad \text{for} \quad 4c^2 \geq 3$, thus it holds for all $c \neq 1$. Therefore, the set defined by $\mathscr{A}_1(c, K_W)$ is empty. In this case we consider the relaxed expression of feedback capacity $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1})$ in Theorem 5.

(iv.2) For $c^2 > 1$, we have the following

$$K^* = K_1^* > 0 \implies \kappa > \frac{K_W}{(c^2-1)^2}, \tag{A25}$$

$$\lambda_1^* > 0 \implies \kappa > \frac{K_W}{(1-c^2)}, \tag{A26}$$

$$K_Z^* = K_{Z_1}^* > 0 \implies \kappa^2\left(c^2-1\right)^3 - \kappa K_W\left(c^2-1\right) - K_W^2 > 0, \tag{A27}$$

$$\kappa < \frac{K_W - K_W\sqrt{4c^2-3}}{2\left(1-c^2\right)^2} \text{ or } \kappa > \frac{K_W + K_W\sqrt{4c^2-3}}{2\left(1-c^2\right)^2}. \tag{A28}$$

Next, note that for $c^2 > 1$ the following inequalities hold

$$\frac{K_W}{(1-c^2)} < \frac{K_W - K_W\sqrt{4c^2-3}}{2\left(1-c^2\right)^2} < 0 < \frac{K_W}{(c^2-1)^2} < \frac{K_W + K_W\sqrt{4c^2-3}}{2\left(1-c^2\right)^2}. \tag{A29}$$

Then, from (A25)–(A29), we deduce that the set $\mathscr{A}_2(c, K_W)$ is non empty, only if the power κ satisfies $\kappa > \frac{K_W\left(1+\sqrt{4c^2-3}\right)}{2\left(1-c^2\right)^2}$. For values of $\kappa \leq \frac{K_W\left(1+\sqrt{4c^2-3}\right)}{2\left(1-c^2\right)^2}$, since they do not belong in the set $\mathscr{A}_2(c, K_W)$, we will consider the relaxed expression of feedback capacity $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1})$ in Theorem 5.

(iv.3) For $c^2 = 1$ clearly the set $\mathscr{A}_3(c, K_W)$ is empty, thus we need to consider the relaxed expression of feedback capacity $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1})$. Putting all the above together we obtain the statements under (2) and (3), where the last statement of (3) is shown by considering the feedback capacity optimization problem $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1})$ in Theorem 5. The proof is completed.

Appendix A.4. Proof of Theorem 5

By Theorem 4, under the constraint of detectability and stabilizability, there does not exist a non-zero value of the optimization problem $C_{FB}(\kappa)$ for Regimes 2 and 3. Further, stabilizability implies the optimal $K_Z^* > 0$ for Regime 1. However, under the relaxations of detectability and unit circle controllability, the optimization problem $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1})$ is over the set $\mathscr{P}_{FB}^{\infty,+}$, as stated in Corollary 2, and unit circle controllability implies the relaxation $K_Z^* \geq 0$. Since by the proof of Theorem 4 there is no non-zero value of $C_{FB}(\kappa)$ corresponding to Regimes 2 and 3 under the constraint $K_Z^* > 0$, then necessarily $K_Z^* = 0$, which implies the optimization problem is given by (126)–(128), where $|\Lambda| \neq 1$ is due to unit circle controllability. Consequently, the solutions of the ARE (128) are

$$K = 0 \quad \text{unique stab. sol., } |F^{CL}(K,\Lambda,K_Z^*)\Big|_{K_{Z^*}=0} = -\Lambda| < 1, \text{ if } |\Lambda| < 1, \tag{A30}$$

$$K_+(\Lambda) = \frac{K_W\Big((\Lambda)^2-1\Big)}{\Big(\Lambda+c\Big)^2} > 0 \quad \text{maximal stab. sol., } |F^{CL}(K,\Lambda,K_Z^*)\Big|_{K_{Z^*}=0} - \tfrac{1}{\Lambda}| < 1 \text{ if } |\Lambda| > 1. \tag{A31}$$

where $K_+(\Lambda) > 0, |\Lambda^\infty| > 1$ is the maximal and stabilizing solution that gives a positive value of $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1}) > 0$, while $K(\Lambda) = 0, |\Lambda| < 1$ gives a zero value of $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1}) = 0$. Finally, substituting $K_+(\Lambda) > 0, |\Lambda| > 1$ into the average power constraint and $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1})$ we obtain (130)–(132), where $\Lambda = \Lambda^*$ denotes the optimal value. It remains to show (133)–(135) achieves $C_{FB}^{max}(\kappa, \mathbf{P}_{V_1}) \stackrel{\triangle}{=} \log\max\{1, |\Lambda^*|\} > 0, \forall \kappa \in \mathscr{K}_2^\infty(c, K_W) \bigcup \mathscr{K}_3^\infty(c)$, provided $K_{V_1} > 0$, that is, it does not include $K_{V_1} = 0$. To show the claim, we consider (133)–(135), and we determine the limit of $K_n \stackrel{\triangle}{=} \mathbf{E}\Big\{\Big(V_n - \mathbf{E}\Big\{V_n\Big|Y^n\Big\}\Big)^2\Big\}$, as $n \longrightarrow \infty$. By applying the Kalman-filter [28], $K_n, n = 1, 2, \ldots$ satisfy the recursions,

$$K_n = c_n^2 K_{n-1} + K_{W_n} - \frac{\Big(K_{W_n} + c_n K_{n-1}\Big(\Lambda_n + c_n\Big)\Big)^2}{K_{W_n} + \Big(\Lambda_n + c_n\Big)^2 K_{n-1}}, \quad K_n \geq 0, \quad n = 2, 3, \ldots, \quad K_1 = \frac{K_{V_1} K_{Z_1}}{K_{V_1} + K_{Z_1}}, \tag{A32}$$

$$F_n^{CL}(K_{n-1}, \Lambda) \stackrel{\triangle}{=} c_n - M_n(K_{n-1}, \Lambda_n)\Big(\Lambda_n + c_n\Big), \tag{A33}$$

$$M_n(K_{n-1}, \Lambda_n) \stackrel{\triangle}{=} \Big(K_{W_n} + c_n K_{n-1}\Big(\Lambda_n + c_n\Big)\Big)\Big(K_{W_n} + \Big(\Lambda_n + c_n\Big)^2 K_{n-1}\Big)^{-1}. \tag{A34}$$

Fact 1. *$K_{V_1} = 0$ implies $K_n = 0, n = 1, 2, 3, \ldots$, hence $\lim_{n\longrightarrow\infty} K_n = K = 0 \neq K_+(\Lambda)$, and by (126), $C_{FB}^{max} = 0$; consequently, C_{FB}^{max} cannot be the asymptotic feedback capacity.*

Using Fact 1, the convergence $\lim_{n\longrightarrow\infty}(c_n, K_{W_n}, \Lambda_n) = (c, K_W, \Lambda)$, and also when detectability and unit circle controllability are necessary and sufficient conditions for $\lim_{n\longrightarrow\infty} K_n = K$, where K satisfies (128), then we have the following.

$$K = 0, \quad \text{if } |\Lambda| < 1, \ \forall K_{V_1} \geq 0, \text{ the unique, stab. sol., } |F^{CL}(K = 0, \Lambda) = -\Lambda| < 1, \tag{A35}$$

$$K_+ = \frac{K_W\big((\Lambda)^2 - 1\big)}{\Big(\Lambda + c\Big)^2} > 0, \text{ if } |\Lambda| > 1, \ K_{V_1} > 0, \text{ the maximal and stab. sol., } |F^{CL}(K, \Lambda) = -\tfrac{1}{\Lambda}| < 1, \tag{A36}$$

The limit of the average power is

$$\begin{aligned}\lim_{n\longrightarrow\infty}\frac{1}{n}\mathbf{E}\Big\{\sum_{t=1}^n (X_t)^2\Big\} &= \lim_{n\longrightarrow\infty}\Big\{\frac{1}{n}K_{Z_1} + \frac{1}{n}\sum_{t=2}^n \Lambda_n^2 K_{t-1}\Big\} \\ &= \begin{cases} 0 & \text{if } |\Lambda| < 1, \forall K_{V_1} \geq 0, \text{ corresponding to the unique and stab. sol.} \\ (\Lambda)^2\frac{K_W\left((\Lambda)^2-1\right)}{\left(\Lambda+c\right)^2} > 0 & \text{if } |\Lambda| > 1, \ K_{V_1} > 0, \text{ corresponding to the max. and stab. sol.} \end{cases}\end{aligned} \tag{A37}$$

From the above equations we have the following. If $K_{V_1} > 0$ and $|\Lambda| > 1$, the average power constraint is satisfied,

$\lim_{n\longrightarrow\infty}\frac{1}{n}\mathbf{E}\Big\{\sum_{t=1}^n (X_t)^2\Big\} = (\Lambda)^2\frac{K_W\left((\Lambda)^2-1\right)}{\left(\Lambda+c\right)^2} = \kappa$, which gives (131), and follows that $C_{FB}^{max} = \frac{1}{2}\log\Big(\frac{\left(\Lambda+c\right)^2 K_+ + K_W}{K_W}\Big) = \log|\Lambda|, |\Lambda| > 1$. If $K_{V_1} = 0$, then $\lim_{n\longrightarrow\infty} K_n = K(\Lambda) = 0$ and this is the *unique and stabilizing solution*, i.e., $\lim_{n\longrightarrow\infty} F^{CL}(K_{n-1}, \Lambda_n) = F^{CL}(K, \Lambda) = -\Lambda$, and $|F^{CL}(K, \Lambda)| < 1$ if and only if $|\Lambda| < 1$. If $K_{V_1} = 0$ the average power is $\lim_{n\longrightarrow\infty}\frac{1}{n}\mathbf{E}\Big\{\sum_{t=1}^n (X_t)^2\Big\} = 0$. Thus we have verified our statements.

References

1. Derpich, M.S.; Ostergaard, J. Comments on "Feedback Capacity of Stationary Gaussian Channels". *IEEE Trans. Inf. Theory* **2024**, *70*, 1848–1851. [CrossRef]
2. Kim, Y.H. Feedback Capacity of Stationary Gaussian Channels. *IEEE Trans. Inf. Theory* **2010**, *56*, 57–85. [CrossRef]
3. Cover, T.; Pombra, S. Gaussian Feedback Capacity. *IEEE Trans. Inf. Theory* **1989**, *35*, 37–43. [CrossRef]
4. Gallager, R.T. *Information Theory and Reliable Communication*; John Wiley & Sons, Inc.: New York, NY, USA, 1968.
5. Ihara, S. *Information Theory for Continuous Systems*; World Scientific: Singapore, 1993.
6. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2006.
7. Butman, S. A General Formulation of Linear Feedback Communications Systems with Solutions. *IEEE Trans. Inf. Theory* **1969**, *15*, 392–400. [CrossRef]
8. Tienan, J.; Schalkwijk, J.P.M. An Upper Bound to the Capacity of Bandlimited Gaussian Autoregressive Channel with Noiseless Feedback. *IEEE Trans. Inf. Theory* **1974**, *20*, 311–316. [CrossRef]
9. Wolfowitz, J. Signalling over a Gaussian Channel with Feedback and Autoregressive Noise. *J. Appl. Probab.* **1975**, *12*, 713–723. [CrossRef]
10. Butman, S. Linear Feedback Rate Bounds for Regressive Channels. *IEEE Trans. Inf. Theory* **1976**, *22*, 363–366. [CrossRef]
11. Ordentlich, E. A Class of Optimal Coding Schemes for Moving Average Additive Gaussian Noise Channels with Feedback. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Trondheim, Norway, 27 June–1 July 1994; p. 467.
12. Ozarow, L. Upper Bounds on the Capacity of Gaussian Channel with Feedback. *IEEE Trans. Inf. Theory* **1984**, *36*, 156–161. [CrossRef]
13. Ozarow, L.H. Random Coding for Additive Gaussian Channels with Feedback. *IEEE Trans. Inf. Theory* **1984**, *36*, 17–22. [CrossRef]
14. Dembo, A. On Gaussian Feedback Capacity. *IEEE Trans. Inf. Theory* **1989**, *35*, 1072–1076. [CrossRef]
15. Ihara, S.; Yanaki, K. Capacity of Discrete-Time Gaussian Channels with and without Feedback—II. *Jpn. J. Appl. Math.* **1989**, *6*, 245–258. [CrossRef]
16. Yanaki, K. Necessary and Sufficient Conditions for the Capacity of the Discrete-Time Gaussian Channel to Be Increased by Feedback. *IEEE Trans. Inf. Theory* **1992**, *38*, 1788–1791. [CrossRef]
17. Yanaki, K. An Upper Bound on the Discrete-Time Gaussian Channel with Feedback—II. *IEEE Trans. Inf. Theory* **1994**, *40*, 588–593. [CrossRef]
18. Chen, H.W.; Yanaki, K. Refinements of the Half-Bit and Factor-of-Two Bounds for Capacity in Gaussian Channels with Feedback. *IEEE Trans. Inf. Theory* **1999**, *45*, 319–325. [CrossRef]
19. Chen, H.W.; Yanaki, K. Upper Bounds on the Capacity of Discrete-Time Blockwise White Gaussian Channels with Feedback. *IEEE Trans. Inf. Theory* **2000**, *43*, 1125–1131. [CrossRef]
20. Yang, S.; Kavcic, A.; Tatikonda, S. On Feedback Capacity of Power-Constrained Gaussian Noise Channels with Memory. *IEEE Trans. Inf. Theory* **2007**, *53*, 929–954. [CrossRef]
21. Liu, T.; Han, G. The ARMAk Gaussian Feedback Capacity. In Proceedings of the IEEE International Symposium on Information Theory Proceedings (ISIT), Aachen, Germany, 25–30 June 2017; pp. 211–215.
22. Liu, T.; Han, G. Feedback Capacity of Stationary Gaussian Channels Further Examined. *IEEE Trans. Inf. Theory* **2019**, *64*, 2494–2506. [CrossRef]
23. Li, C.; Elia, N. Youla Coding and Computation of Gaussian Feedback Capacity. *IEEE Trans. Inf. Theory* **2019**, *64*, 3197–3215. [CrossRef]
24. Gattami, A. Feedback Capacity of Gaussian Channels Revisited. *IEEE Trans. Inf. Theory* **2019**, *65*, 1948–1960. [CrossRef]
25. Ihara, S. On the Feedback Capacity of the First-Order Moving Average Gaussian Channel. *Jpn. J. Stat. Data Sci.* **2019**, *2*, 491–506. [CrossRef]
26. Charalambous, C.D.; Kourtellaris, C.; Louka, S. New Formulas of Feedback Capacity for AGN Channels with Memory: A Time-Domain Sufficient Statistic Approach. *arXiv* **2020**, arXiv:2010.06226. [CrossRef] [PubMed]
27. Charalambous, C.D.; Kourtellaris, C.; Louka, S. New Formulas of Feedback Capacity for AGN Channels with Memory: A Time-Domain Sufficient Statistic Approach. *Entropy* **2025**, *27*, 207. [CrossRef]
28. Caines, P.E. *Linear Stochastic Systems*; Wiley Series in Probability and Statistics; John Wiley & Sons, Inc.: New York, NY, USA, 1988.
29. Kailath, T.; Sayed, A.; Hassibi, B. *Linear Estimation*; Prentice Hall: Hoboken, NJ, USA, 2000.
30. Charalambous, C.D.; Kourtellaris, C.; Louka, S. Sequential Characterization of Cover and Pombra Gaussian Feedback Capacity: Generalizations to MIMO Channels via a Sufficient Statistic. In Proceedings of the 2021 IEEE Information Theory Workshop (ITW), Kanazawa, Japan, 17–21 October 2021.
31. Kourtellaris, C.; Charalambous, C.D.; Loyka, S. New Formulas for Ergodic Feedback Capacity of AGN Channels Driven by Stable and Unstable Autoregressive Noise. In Proceedings of the International Symposium on Information Theory (ISIT), Los Angeles, CA, USA, 21–26 June 2020; pp. 2073–2078.

32. Charalambous, C.D.; Kourtellaris, C.; Loyka, S. Control-Coding Capacity of Decision Models Driven by Correlated Noise and Gaussian Application Examples. In Proceedings of the IEEE Conference on Decision and Control (CDC), Miami Beach, FL, USA, 17–19 December 2018; pp. 4967–4972.
33. Kourtellaris, C.; Charalambous, C.D.; Loyka, S. From Feedback Capacity to Tight Achievable Rates without Feedback for AGN Channels with Stable and Unstable Autoregressive Noise. In Proceedings of the International Symposium on Information Theory (ISIT), Los Angeles, CA, USA, 21–26 June 2020; pp. 2091–2096.
34. Gelfand, M.I.; Yaglom, M. Calculation of the Amount of Information about a Random Function Contained in Another Such Function. *Am. Math. Soc. Transl.* **1959**, *2*, 199–246.
35. van Schuppen, J.H. *Control and System Theory of Discrete-Time Stochastic Systems*; Springer Nature AG: Cham, Switzerland, 2021.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.