New Formulas of Feedback Capacity of AGN Channels Driven by Autoregressive Moving Average Noise

Thesis Submitted in a Partial Fulfillment of the Requirements for the Degree of Magister in Computer Engineering

Edited By

Stelios Louca

The University of Cyprus, Nicosia

Department of Electrical and Computer Engineering

2020

PUBLISHER

Νέες Φόρμουλες Χωρητικότητας Ανατροφοδότησης **AGN** Καναλιών που Καθοδηγούνται από τον Αυτοεκθετικό Κινητό Μέσο Θόρυβο

Διατριβή που κατατίθεται ως μέρος των απαιτήσεων για το πτυχίο Μάστερ στην Μηχανική Υπολογιστών

Έχει εκδόσει ο

Στέλιος Λουκά

Πανεπιστήμιο Κύπρου, Λευκωσία

Τμήμα Ηλεκτρολόγων Μηχανικών Υπολογιστών

2020

ΕΚΔΟΤΗΣ

THESIS APPROVAL SHEET

Thesis Title: *New formulas of Feedback Capacity for AGN Channels Driven by Autoregressive Moving Average Noise*

Student's Name: Stelios Louca

We the members of the Advisory Committee for the above named student verify that the thesis satisfies the requirements of the Graduate School as approved by the Graduate Faculty.

Charalambos D. Charalambous
Professor
Department of Electrical and Computer Engineering
Associate Dean, School of Engineering
Director of Thesis

Christoforos N. Hadjicostis
Professor
Department of Electrical and Computer Engineering

Christos Kourtellaris
Post Doctoral Researcher
Department of Electrical and Computer Engineering

Advisory Committee

Date

Acknowledgement

I would like to express my gratitude to my advisor Professor Charalambos D. Charalambous for his patience and his continuous encouragement and academic support. I appreciate his extensive knowledge and expertise in many areas and his nice treatment and assistance throughout my studies at the University of Cyprus.

I would also like to acknowledge Dr. Christos Kourtellaris, as the second reader of this thesis, and I am gratefully indebted to his for his very valuable comments on this thesis.

A big thank to my family for the support they provided me through my entire life. My special thanks goes to my girlfriend Andria; without her love, patience and encouragement I would not have finished this work.

Abstract

Current telecommunication and information systems are designed based on Shannon's operational definitions of coding-capacity for reliable communication, which utilizes encoders and decoders, to combat communication noise and to remove redundancy in data. Current dynamical control systems are designed by utilizing feedback controllers, actuators and sensors, to ensure stability, robustness, and optimal performance. One of the key challenges in the upcoming years is the development and design of intelligent hierarchical communication and control systems which simultaneously control and transmit information. This thesis focuses on developing universal operational definitions, to any dynamical system with inputs and outputs, called Control-Coding Capacity of dynamical systems, i.e., the designer's ability to develop an controller-encoder pair to simultaneously control the system and encode information, transmit it through the dynamical system to any process attached to it, and reconstruct it using decoders, with arbitrary small error probability.

We choose to work with Additive Gaussian Noise (AGN) channels with finite memory on the noise, because these channels are widely employed to model band-limited channels and are highly realistic since the can adequately describe physical channels under practical scenarios. We begin our analysis by calculating the feedback capacity both for stable and unstable noise, where we show that feedback does not always increase the capacity. We have done the analysis for memory 1, which can be easily generalized to arbitrary finite memory. Subsequently, we have provided achievable rates without feedback which are induced by a causal input process. We have considered both independent and identically distributed (iid) processes and Markov processes.

Πρόλογος

Τα τρέχοντα συστήματα τηλεπικοινωνιών και πληροφοριών έχουν σχεδιαστεί με βάση τους επιχειρησιακούς ορισμούς του Shannon σχετικά με την ικανότητα κωδικοποίησης για αξιόπιστη επικοινωνία, που χρησιμοποιεί κωδικοποιητές και αποκωδικοποιητές, για την καταπολέμηση του θορύβου επικοινωνίας και για την εξάλειψη των πλεονασμάτων στα δεδομένα. Τα τρέχοντα δυναμικά συστήματα ελέγχου έχουν σχεδιαστεί με τη χρήση ελεγκτών ανάδρασης, ενεργοποιητών και αισθητήρων, για τη διασφάλιση σταθερότητας, αντοχής και βέλτιστης απόδοσης. Μία από τις βασικές προκλήσεις τα επόμενα χρόνια είναι η ανάπτυξη και ο σχεδιασμός έξυπνων συστημάτων ιεραρχικής επικοινωνίας και ελέγχου, τα οποία ταυτόχρονα ελέγχουν και μεταδίδουν πληροφορίες. Αυτή η διατριβή, επικεντρώνεται στην ανάπτυξη καθολικών επιχειρησιακών ορισμών, σε οποιοδήποτε δυναμικό σύστημα με εισόδους και εξόδους, που ονομάζεται χωρητικότητα ελέγχου-κωδικοποίησης δυναμικών συστημάτων, π.χ. η ικανότητα του σχεδιαστή να αναπτύξει ζεύγος ελεγκτή-κωδικοποιητή για ταυτόχρονο έλεγχο του συστήματος και κωδικοποίησης πληροφοριών, μετάδοσης μέσω του δυναμικού συστήματος σε οποιαδήποτε διεργασία που συνδέεται με αυτό, και ανακατασκευή του χρησιμοποιώντας αποκωδικοποιητές, με αυθαίρετα μικρή πιθανότητα σφάλματος.

Επιλέγουμε να δουλέψουμε με κανάλια Additive Gaussian Noise (AGN) με πεπερασμένη μνήμη στο θόρυβο, επειδή αυτά τα κανάλια χρησιμοποιούνται ευρέως για τη μοντελοποίηση καναλιών περιορισμένου εύρους και είναι εξαιρετικά ρεαλιστικά, καθώς μπορούν να περιγράψουν επαρκώς τα φυσικά κανάλια σε πρακτικά σενάρια. Ξεκινάμε την ανάλυσή μας υπολογίζοντας την χωρητικότητα με ανάδραση τόσο για σταθερό όσο και για ασταθή θόρυβο, όπου δείχνουμε ότι η ανάδραση δεν αυξάνει πάντα την χωρητικότητα. Έχουμε κάνει την ανάλυση για μνήμη 1, η οποία μπορεί εύκολα να γενικευτεί σε αυθαίρετη πεπερασμένη μνήμη. Στη συνέχεια, έχουμε παράξει εφικτές τιμές χωρίς ανατροφοδότηση που προκαλούνται από μια αιτιώδη διαδικασία εισαγωγής. Εξετάσαμε τόσο ανεξάρτητες και ταυτόσημες κατανεμημένες (iid) διαδικασίες όσο και διαδικασίες Markov.

Contents

List of Figures

List of Acronyms

AGN : Additive Gaussian Noise Channel

PO-SS : Partially Observable State Space

PSD : Power Spectral Density

FTFI : $n-$ Finite Transmission Feedback Information Capacity

DRE : Generalized Difference Riccati Equation

ARE : Generalized Algebraic Riccati Equation

ARMA : Autoregressive Moving Average Model

Notation and List of Symbols

$\mathbb{Z}$: Set of integer numbers

$\mathbb{Z}_+^n \triangleq \{1,2,\ldots,n\}$: n is a finite positive integer

$\mathbb{R} \triangleq (-\infty,\infty)$: Set of real numbers

$\mathbb{R}^m$: Vector space of tuples of the real numbers for an integer $n \in \mathbb{Z}_+$

$\mathbb{C} \triangleq \{a+jb : (a,b) \in \mathbb{R}\times\mathbb{R}\}$: Space of complex numbers

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

$\mathbb{D}_o \triangleq \left\{c \in \mathbb{C} : |c| < 1\right\}$: Open unit disc of the space of complex number $\mathbb{C}$

$\mathbb{S}_+^{n\times n}$, : Set of n by n positive semi definite symmetric matrices with real elements

$\mathbb{S}_{++}^{n\times n}$: Set of n by n positive definite symmetric matrices with real elements

$I_{n\times n} \in \mathbb{S}_{++}^{n\times n}, n \in \mathbb{Z}_+$: Identity matrix

$tr\big(A\big)$: Trace of any matrix $A \in \mathbb{R}^{n\times n}, n \in \mathbb{Z}_+$

$spec(A) \subset \mathbb{C}$: Spectrum of a matrix $A \in \mathbb{R}^{q\times q}, q \in \mathbb{Z}_+$ (the set of all its eigenvalues)

Exponentially stable matrix $A \in \mathbb{R}^{q\times q} \quad \xrightarrow{iff} \quad spec(A) \subset \mathbb{D}_o$

$\Big(\Omega, \mathscr{F}, \mathbb{P}\Big)$: Probability space

$\mathbf{P}_X$: Distribution of a random variable $X : \Omega \to \mathbb{R}^{n_x}, n_x \in \mathbb{Z}_+^n$

Chapter 1

Introduction

An important class of practical problems in Shannon's reliable communication over noisy channels is the additive Gaussian noise (AGN) channel. Such problems are often classified into: (i) memoryless AGN channels, with or without feedback; (ii) AGN channels with memory, with or without feedback. The feedback may be noiseless or noisy.

Two fundamental questions for the sub-class of additive Gaussian noise (AGN) channels with memory and noiseless feedback, are:

(Q1): Feedback and non-feedback capacity of the AGN channel;

(Q2): Feedback coding scheme of communicating a Gaussian random process $\Theta_t \in N(0, \sigma^2_{\Theta_t}), t = 1, \ldots, n$, and the coding scheme of communicating digital messages $w \in \mathscr{M}_n \stackrel{\triangle}{=} \{1, 2, \ldots, \lceil 2^{nR} \rceil\}$, which achieve the feedback, or non-feedback capacity of the channel.

This thesis is focused on the question Q1. We derive the feedback capacity of AGN channels, driven by Autoregressive Moving Average noise and we show our answer, which contradicts several results found in the literature.

1.1 Problem, Motivation, and Main Results

We consider the additive Gaussian noise (AGN) channel defined by

$$Y_t = X_t + V_t, \quad t = 1, \ldots, n, \quad \frac{1}{n}\mathbf{E}\Big\{\sum_{t=1}^{n}(X_t)^2\Big\} \leq \kappa, \quad \kappa \in [0, \infty) \tag{1.1.1}$$

where
$X^n \stackrel{\triangle}{=} \{X_1, X_2, \ldots, X_n\}$ is the sequence of channel input random variables (RVs) $X_t : \Omega \to \mathbb{R}$,

$Y^n \triangleq \{Y_1, Y_2, \ldots, Y_n\}$ is the sequence of channel output RVs $Y_t : \Omega \to \mathbb{R}$,
$V^n \triangleq \{V_1, \ldots, V_n\}$ is the sequence jointly Gaussian distributed RVs $V_t : \Omega \to \mathbb{R}$, with distribution $\mathbf{P}_{V^n}(dv^n)$, not necessarily stationary or ergodic.
We wish to introduce the feedback capacity of the AGN channel (1.1.1) under two distinct formulations of code definition and noise model.

Case I) Formulation. The feedback code does not assume knowledge of the initial state of the noise at the encoder and the decoder (see Definition 1.1.1), and the noise sequence V^n is represented by a partially observable[1] state space realization, with state sequence S^n (see Definition 1.1.2).
Case I) formulation is consistent with the Cover and Pombra formulation of code definition and noise model, for which the optimal channel input with feedback and the "$n-$finite transmission" feedback capacity are derived in [3, eqn(11) and eqn(10)], using the information measure[2],

$$C_n^{fb}(\kappa) \triangleq \sup_{\mathbf{P}_{X_t|X^{t-1},Y^{t-1}}, t=1,\ldots,n:\ \frac{1}{n}\mathbf{E}\left\{\sum_{t=1}^n \left(X_t\right)^2\right\}\leq\kappa} \sum_{t=1}^n H(Y_t|Y^{t-1}) - H(V^n) \tag{1.1.2}$$

provided the supremum exists, and where $H(\cdot)$ denotes differential entropy.
For a feedback code that assumes knowledge of the initial state of the noise or the channel, $S_1 = s$, at the encoder and the decoder (see Definition 1.1.3), it follows from [3, eqn(11) and eqn(10)], that the information measure is

$$C_n^{fb}(\kappa, s) \triangleq \sup_{\mathbf{P}_{X_t|X^{t-1},Y^{t-1},s}, t=1,\ldots,n:\ \frac{1}{n}\mathbf{E}\left\{\sum_{t=1}^n \left(X_t\right)^2 \Big| S_1=s\right\}\leq\kappa} \sum_{t=1}^n H(Y_t|Y^{t-1}, s) - H(V^n|s). \tag{1.1.3}$$

Case II) Formulation. The feedback code assumes knowledge of the initial state of the noise or the channel, $S_1 = s$, at the encoder and the decoder (see Definition 1.1.3), and the noise sequence V^n assumes a state space realization with state sequence S^n, that presupposes the noise V^{t-1} (including the initial state) uniquely defines the noise state sequence S^t and vice-versa for $t = 1, \ldots, n$.
Case II) formulation is consistent with the Yang, Kavcic, and Tatikonda [1], code definition and noise model (see [1, Section II, in particular Section II.C, I)-III)]), for which the optimal channel input with feedback and $n-$finite transmission feedback capacity are derived in [1, Theorem 1], using the information measure,

$$C_n^{fb,S}(\kappa, s) \triangleq \sup_{\mathbf{P}_{X_t|S^t,Y^{t-1},s}, t=1,\ldots,n:\ \frac{1}{n}\mathbf{E}\left\{\sum_{t=1}^n \left(X_t\right)^2 \Big| S_1=s\right\}\leq\kappa} \sum_{t=1}^n H(Y_t|Y^{t-1}, s) - H(V^n|s). \tag{1.1.4}$$

[1] Partially observable means that knowledge of V^{t-1} and initial state do not specify the state S^t.
[2] $C_n^{fb}(\kappa)$ is identified using the converse coding theorem [3].

To clarify the reasons which motivated us to analyze Case I) and II) formulations, we wish to mention two technical issues, which are not clarified in [2,4–7] and lead to *fundamental confusions as well as incorrect interpretation of the results*.
First, to make the transition from the channel input distributions $\mathbf{P}_{X_t|X^{t-1},Y^{t-1}}, t=1,\ldots,n$ of Case I) formulation (1.1.2), to $\mathbf{P}_{X_t|S^t,Y^{t-1},S}, t=1,\ldots,n$ of Case II) formulation, (1.1.4), the conditions stated in (1.1.6), (1.1.7) are necessary (as easily verified from the converse coding theorem).

$$\mathbf{P}_{X_t|X^{t-1},Y^{t-1}} = \mathbf{P}_{X_t|V^{t-1},Y^{t-1}} \quad \text{always holds by channel definition } Y_k = X_k + V_k, k=1,\ldots,n \tag{1.1.5}$$

$$= \mathbf{P}_{X_t|X^{t-1},Y^{t-1},S} \quad \text{if the initial state } S=s \text{ is known to the feedback code} \tag{1.1.6}$$

$$= \mathbf{P}_{X_t|S^t,Y^{t-1},S} \quad \text{if } (V^{t-1}, S=s) \text{ uniquely defines } S^t \text{ and vice-versa.} \tag{1.1.7}$$

Second, the analysis of the asymptotic per unit time limits of (1.1.2)-(1.1.4), and their variants (when the supremum over distributions and limit over $n \longrightarrow \infty$ are interchanged), require certain technical necessary and/or sufficient conditions for the limits to be finite, for the joint process $(X_t, Y_t), t=1,\ldots$ to be stationary or asymptotically stationary, and for the rates to be independent of the initial data, $S_1 = s$.

Our choice of Case I) and II) formulations is further motivated by the believe [2, 4–7], that Kim's characterizations of feedback capacity, in the frequency domain [2, Theorem 4.1], and in the time domain [2, Theorem 6.1], correspond to the Cover and Pombra code definition and noise model. We show this believe is false. We also show the characterization of feedback [2, Theorem 6.1, i.e., C_{FB}] does not correspond to the limit of a jointly stationary or asymptotically stationary process, (X_t, Y_t), one-sided, i.e., $t \in \{1,2,\ldots\}$, or double-sided, i.e., $t \in \{\ldots,-1,0,1,\ldots\}$. In particular, it will become apparent in subsequent parts of this paper that [2, Theorem 6.1, i.e., C_{FB}], presupposed Case II) formulation, and corresponds to the per unit time limit of (1.1.4) (with supremum and limit interchanged). Further, that since feedback capacity in [2, Theorem 6.1, i.e., C_{FB}] is characterized with *zero innovations process of the channel input* (see [2, Lemma 6.1]), it then follows (from the convergence properties of Kalman-filters [8] [9]), that the value of feedback capacity is necessarily zero i.e., $C_{FB} = 0$.

Case I) Formulation of Feedback Code and Noise Definitions. For Case I) formulation we consider the code of Definition 1.1.1 (due to [3]).

Definition 1.1.1. *Time-varying feedback code [3]*

(a) A noiseless time-varying feedback code for the AGN Channel (1.1.1), is denoted by $(2^{nR}, n)$*,* $n = 1, 2, \ldots$*, and consists of the following elements and assumptions.*
(i) The uniformly distributed messages $W : \Omega \to \mathscr{M}_n \triangleq \{1, 2, \ldots, 2^{nR}\}$*.*
(ii) The time-varying encoder strategies, often called codewords of block length n, defined by[3]

$$\mathscr{E}_{[0,n]}(\kappa) \triangleq \Big\{X_1 = e_1(W), X_2 = e_2(W, X_1, Y_1) \ldots, X_n = e_n(W, X^{n-1}, Y^{n-1}) : \\ \frac{1}{n}\mathbf{E}^e\Big(\sum_{t=1}^{n}(X_t)^2\Big) \leq \kappa\Big\}.$$

(iii) The average error probability of the decoder functions $y^n \longmapsto d_n(y^n) \in \mathscr{M}_n$*, defined by*

$$\mathbf{P}_{error}^{(n)} = \mathbb{P}\Big\{d_n(Y^n) \neq W\Big\} = \frac{1}{2^{nR}} \sum_{W=1}^{2^{nR}} \mathbb{P}\Big\{d_n(Y^n) \neq W\Big\}. \tag{1.1.8}$$

(iv) The channel input sequence "$X^n \triangleq \{X_1, \ldots, X_n\}$ *is causally related*[4] *to* V^n*", which is equivalent to the following decomposition of the joint probability distribution of* (X^n, V^n)*:*

$$\mathbf{P}_{X^n, V^n} = \mathbf{P}_{V_n|V^{n-1}, X^n}\, \mathbf{P}_{X_n|X^{n-1}, V^{n-1}} \ldots \mathbf{P}_{V_2|V_1, X^2} \mathbf{P}_{X_2|X_1, V_1} \mathbf{P}_{V_1|X_1} \mathbf{P}_{X_1} \tag{1.1.9}$$

$$= \mathbf{P}_{V^n} \prod_{t=1}^{n} \mathbf{P}_{X_t|X^{t-1}, V^{t-1}}, \quad \text{that is,} \quad \mathbf{P}_{V_t|V^{t-1}, X^t} = \mathbf{P}_{V_t|V^{t-1}}. \tag{1.1.10}$$

That is, $X^t \leftrightarrow V^{t-1} \leftrightarrow V_t$ *is a Markov chain, for* $t = 1, \ldots, n$*. As usual, the messages W are independent of the channel noise* V^n*.*
A rate R is called an achievable rate with feedback coding, if there exists a sequence of codes $(2^{nR}, n), n = 1, 2, \ldots$*, such that* $\mathbf{P}_{error}^{(n)} \longrightarrow 0$ *as* $n \longrightarrow \infty$*. The feedback capacity* $C^{fb}(\kappa)$ *is defined as the supremum of all achievable rates R.*

For Case I) formulation we consider a noise model which is consistent with the Cover and Pombra noise model, i.e., V^n is jointly Gaussian distributed, $\mathbf{P}_{V^n} = \times_{t=1}^{n} \mathbf{P}_{V_t|V^{t-1}}$, and induced by the partially observable state space (PO-SS) realization of Definition 1.1.2.

Definition 1.1.2. *A time-varying PO-SS realization of the Gaussian noise*

[3] The superscript $e(\cdot)$ on $\mathbf{E}^e$ indicates that the distribution depends on the strategy $e(\cdot) \in \mathscr{E}_{[0,n]}(\kappa)$.
[4] A notion found in [3], page 39, above Lemma 5.

A time-varying PO-SS realization of the Gaussian noise $V^n \in N(0, K_{V^n})$ is defined by

$$S_{t+1} = A_t S_t + B_t W_t, \quad t = 1, \ldots, n-1 \tag{1.1.11}$$

$$V_t = C_t S_t + N_t W_t, \quad t = 1, \ldots, n, \tag{1.1.12}$$

$$S_1 \in N(\mu_{S_1}, K_{S_1}), \quad K_{S_1} \succeq 0, \tag{1.1.13}$$

$$W_t \in N(0, K_{W_t}), \quad K_{W_t} \succeq 0, \quad t = 1 \ldots, n \ \ \textit{an indep. Gaussian process}, \quad W^t \ \ \textit{indep. of} \ \ S_1, \tag{1.1.14}$$

$$S_t : \Omega \to \mathbb{R}^{n_s}, \quad W_t : \Omega \to \mathbb{R}^{n_w}, \quad V_t : \Omega \to \mathbb{R}, \quad R_t \stackrel{\triangle}{=} N_t K_{W_t} N_t^T \succ 0, \quad t = 1, \ldots, n \tag{1.1.15}$$

where $(A_t, B_t, C_t, N_t, \mu_{S_1}, K_{S_1}, K_{W_t})$ are nonrandom for all t, and n_s, n_w are finite positive integers. A time-invariant PO-SS realization of the Gaussian noise $V^n \in N(0, K_{V^n})$ is defined by (1.1.11)-(5.2.85), with $(A_t, B_t, C_t, N_t, K_{W_t}) = (A, B, C, N, K_W), \forall t$.

For Case I) formulation we use the terminology "partially observable", which is standard in filtering theory [9], because the noise V^n induces a distribution $\mathbf{P}_{V^n} = \times_{t=1}^n \mathbf{P}_{V_t|V^{t-1}}$, and $\mathbf{P}_{V_t|V^{t-1}}$ cannot be expressed as a function of the state of the noise, i.e., V^{t-1} does not uniquely define S^t. The PO-SS realization is often adopted in many practical problems of engineering and science, to realize jointly Gaussian processes V^n.

We should emphasize that for Case I) formulation, the code of Definition 1.1.1 and the PO-SS realization of Definition 1.1.2, the following two conditions must be respected (to be consistent with [3]):

(A1) *The initial state S_1 of the noise is not known at the encoder and the decoder, and*

(A2) *at each t, the representation of the noise V^{t-1} by the PO-SS realization of Definition 1.1.2, does not uniquely determine the state of the noise S^t and vice-versa, i.e., it is a partially observable realization.*

It is easy to verify that conditions (A1) and (A2) are indeed, consistent with the Cover and Pombra [3] formulation, see for example, the code definition in [3, page 37], the characterization of the $n-$finite transmission feeback capacity given in [3, eqn(11)], and the coding theorems given in [3, Theorem 1].

Case II) Formulation of Feedback Code and Noise Definitions. For Case II) formulation we pressupose:

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

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

By Condition 1 the code is that of Definition 1.1.3, below, which is different from the code of Definition 1.1.1.

Definition 1.1.3. *A code with initial state known at the encoder and the decoder*
A variant of the code of Definition 1.1.3, is a feedback code with the initial state of the noise or channel $S_1 = s$, known to the encoder and decoder strategies, denoted by $(s, 2^{nR}, n)$, $n = 1, 2, \ldots$. The code $(s, 2^{nR}, n)$, $n = 1, 2, \ldots$ is defined as in Definition 1.1.1, with (ii), (iii), (iv) replaced by

$$\mathscr{E}^{s}_{[0,n]}(\kappa) \triangleq \Big\{ X_1 = e_1(W, S_1), X_2 = e_2(W, S_1, X_1, Y_1) \ldots, X_n = e_n(W, S_1, X^{n-1}, Y^{n-1}) : \frac{1}{n+1}\mathbf{E}^e\Big\{ \sum_{i=0}^{n} (X_t)^2 \Big| S_1 = s \Big\} \leq \kappa \Big\}, \quad y^n \longmapsto d^s_n(y^n, v^o_{-\infty}) \in \mathscr{M}_n, \tag{1.1.16}$$

$$\mathbf{P}_{X^n, V^n|S_1} = \mathbf{P}_{V^n|S_1} \prod_{t=1}^{n} \mathbf{P}_{X_t|X^{t-1}, V^{t-1}, S_1}, \quad \text{that is,} \quad \mathbf{P}_{V_t|V^{t-1}, X^t, S_1} = \mathbf{P}_{V_t|V^{t-1}, S_1}. \tag{1.1.17}$$

The initial state may include $S_1 \stackrel{\triangle}{=} (V^0_{-\infty}, Y^0_{-\infty})$, etc.

For Case II) formulation it is obvious (from the converse to the coding theorem), that the optimal channel input conditional distribution is expressed as a function of the state of the noise, S^n, due to (1.1.6), (1.1.7).

Our approach is based on the following two step procedure.

Step # 1. We apply a linear transformation to the Cover and Pombra optimal channel input process [3, eqn(11)] (see (1.2.18)-(1.2.24) which are reproduced from [3] for the convenience of the reader), to equivalently represent it by a linear functional of the past channel noise sequence, the past channel output sequence, and an orthogonal Gaussian process, i.e., an innovations process. That is, X^n is uniquely represented, since it is expressed in terms of the orthogonal process.

Step # 2. We express the optimal input process by a functional of a *sufficient statistic*, which satisfies a Markov recursion, and an *orthogonal innovations process*. It then follows that the Cover and Pombra characterization of the "$n-$block" formula [3, eqn(10)] (see (1.2.18) and (1.2.19)) is equivalently represented by a sequential characterization. The problem of feedback capacity is then expressed as the maximization of the per unit time limit of a sum of (differential) entropies of the innovations process of the channel output process, over two sequences of time-varying strategies of the channel input process. The covariance of the innovations process is a functional of the solutions of two generalized matrix DREs.

1.2 The Cover and Pombra Characterizations of Capacity and Related Literature

First, we recall the Cover and Pombra [3] characterization of feedback capacity, since we use it to derive our new sequential characterizations of the $n-$FTFI capacity (to avoid new independent derivations).

Cover and Pombra applied the converse coding theorem and the maximum entropy principle of Gaussian distributions to identify the characterization of the $n-$FTFI capacity [3, eqn(10)] by[5]

$$C_n^{fb}(\kappa) \stackrel{\triangle}{=} \max_{\left(\mathbf{B}^n, K_{\overline{\mathbf{Z}}^n}\right):\ \frac{1}{n} tr\left\{\mathbf{E}\left(\mathbf{X}^n (\mathbf{X}^n)^T\right)\right\} \leq \kappa} H(Y^n) - H(V^n) \tag{1.2.18}$$

$$= \max_{\left(\mathbf{B}^n, K_{\overline{\mathbf{Z}}^n}\right):\ \frac{1}{n} tr\left\{\mathbf{B}^n\, K_{\mathbf{V}^n}\, (\mathbf{B}^n)^T + K_{\overline{\mathbf{Z}}^n}\right\} \leq \kappa} \frac{1}{2} \log \frac{|\left(\mathbf{B}^n + I_{n\times n}\right) K_{\mathbf{V}^n} \left(\mathbf{B}^n + I_{n\times n}\right)^T + K_{\overline{\mathbf{Z}}^n}|}{|K_{\mathbf{V}^n}|} \tag{1.2.19}$$

where the distribution $\mathbf{P}_{Y^n}$ is induced by a jointly Gaussian channel input process X^n [3, eqn(11)]:

$$X_t = \sum_{j=1}^{t-1} B_{t,j} V_j + \overline{Z}_t, \quad t = 1, \ldots, n, \tag{1.2.20}$$

$$\mathbf{X}^n = \mathbf{B}^n \mathbf{V}^n + \overline{\mathbf{Z}}^n, \quad \mathbf{Y}^n = \left(\mathbf{B}^n + I_{n\times n}\right) \mathbf{V}^n + \overline{\mathbf{Z}}^n, \tag{1.2.21}$$

$$\overline{\mathbf{Z}}^n \ \text{ is jointly Gaussian, } N(0, K_{\overline{\mathbf{Z}}^n}), \quad \overline{\mathbf{Z}}^n \ \text{ is independent of } \ \mathbf{V}^n, \tag{1.2.22}$$

$$\mathbf{X}^n \stackrel{\triangle}{=} \left[\begin{array}{cccc} X_1 & X_2 & \ldots & X_n \end{array}\right]^T \ \text{ and similarly for the rest, } \ \mathbf{B}^n \text{ is a lower diagonal matrix,} \tag{1.2.23}$$

$$\frac{1}{n} \mathbf{E}\Big\{ \sum_{t=1}^{n} (X_t)^2 \Big\} = \frac{1}{n} tr \left\{ \mathbf{E}\left(\mathbf{X}^n (\mathbf{X}^n)^T \right) \right\} \leq \kappa. \tag{1.2.24}$$

The notation $N(0, K_{\overline{\mathbf{Z}}^n})$ means the random variable $\overline{\mathbf{Z}}^n$ is jointly Gaussian with mean $\mathbf{E}\{\overline{\mathbf{Z}}^n\} = 0$ and covariance marix $K_{\overline{\mathbf{Z}}^n} = \mathbf{E}\{\overline{\mathbf{Z}}^n (\overline{\mathbf{Z}}^n)^T\}$, and $I_{n\times n}$ denotes an n by n identity matrix.

The feedback capacity, $C^{fb}(\kappa)$, is characterized by the per unit time limit of the $n-$FTFI capacity [3].

$$C^{fb}(\kappa) \stackrel{\triangle}{=} \lim_{n \longrightarrow \infty} \frac{1}{n} C_n^{fb}(\kappa). \tag{1.2.25}$$

[5] We use $H(X)$ to denote differential entropy of a continuous-valued RV X, hence we indirectly assume the probability density functions exist.

The direct and converse coding theorems, are stated in [3, Theorem 1].

Over the years, considerable efforts have been devoted to compute $C_n^{fb}(\kappa)$ and $C^{fb}(\kappa)$, [1, 2, 4–6, 10], often under simplified assumptions on the channel noise. In addition, bounds are described in [11, 12], while numerical methods are developed in [13], mostly for time-invariant AGN channel, driven by stationary noise. We should mention that most papers considered a variant of (1.2.25), by interchanging the per unit time limit and the maximization operations, under the assumption: the joint process $(X^n, Y^n), n = 1, 2, \ldots$ is either *jointly stationary or asymptotically stationary* (see [2, 4, 5, 10]), and *the joint distribution of the joint process* $(X^n, Y^n), n = 1, 2, \ldots$ *is time-invariant*.

Yang, Kavcic and Tatikonda [1] and Kim [2] analyzed the feedback capacity of the AGN channel (1.1.1) driven by a stationary noise, described the power spectral density (PSD) functions $S_V(e^{j\theta}), \theta \in [-\pi, \pi]$:

$$S_V(e^{j\theta}) \triangleq K_W \frac{\left(1 - \sum_{k=1}^{L} a(k) e^{jk\theta}\right)\left(1 - \sum_{k=1}^{L} a(k) e^{-jk\theta}\right)}{\left(1 - \sum_{k=1}^{L} c(k) e^{jk\theta}\right)\left(1 - \sum_{k=1}^{L} c(k) e^{-jk\theta}\right)}, \ |c(k)| < 1, \ |a(k)| < 1, \ c(k) \neq a(k). \tag{1.2.26}$$

More specifically, the analysis by Yang, Kavcic and Tatikonda considered a specific state space realization of the noise PSD (1.2.26), pressuposed a Case II) formulation (see [1, Section II, in particular Section II.C, I)-III), Theorem 1, Section III]):

The initial state of the noise, $S_1 = s$, is known to the encoder and the decoder, and the initial state and noise (s, V^{t-1}) uniquely define the noise state S^t, and vice versa, for all t.

Kim also analyzed the feedback capacity of the AGN channel (1.1.1) driven by a stationary noise described by the PSD (1.2.26), and by a state space realization of the noise V^n (see [2, Section VI]). A major point of confusion, which should be read with caution is that Kim's characterization of feedback capacity in time-domain [2, Theorem 6.1], does not state the conditions based on which this characterization is derived. The reader, however, can verify from [2, Lemma 6.1 and comments above it], that the characterization of feedback capacity [2, Theorem 6.1], presupposed a Case II) formulation, precisely as Yang, Kavcic and Tatikonda [1].

1.3 Thesis Organization

The rest of this thesis is organized as follows.

In Chapter 2, we derive new equivalent sequential characterizations and formulas of the Cover and Pombra "$n-$block or transmission" feedback capacity formula [3, eqn(11)], that is, for Case I) formulation, $C_n^{fb}(\kappa)$, which have not appeared elsewhere in the literature. In particular, we derive equivalent realizations to the Cover and Pombra optimal channel input process X^n [3, eqn(11)], which are linear functionals of a *finite-dimensional sufficient statistic and an orthogonal innovations process*. From these new realizations, follows the sequential characterizations of the "$n-$block or transmission" feedback capacity formula [3, eqn(11)], henceforth called the "$n-$finite transmission feedback information ($n-$FTFI) capacity", which are expressed as functionals of *two generalized matrix difference Riccati equations (DRE) of filtering theory of Gaussian systems*.

In Chapter 3, we derive results analogous to 1), for Case II) formulation e.g., $C_n^{fb}(\kappa,s)$ and for $C_n^{fb,S}(\kappa,s)$, as special cases of Case I).

In Chapter 4, we analyze the asymptotic per unit time limit of the sequential characterizations of the $n-$FTFI capacity, denoted by $C^{fb,o}(\kappa)$, $C^{fb,o}(\kappa,s)$, $C^{fb,S,o}(\kappa,s)$, when the supremum and limit over $n \longrightarrow \infty$ are interchanged. We identify necessary and/or sufficient conditions for the asymptotic limit to exist, and for the optimal joint process $(X_t,Y_t), t=1,\ldots,$ to be asymptotically stationary, in terms of the convergence properties of two generalized matrix difference Riccati equations (DREs) to their corresponding two generalized matrix algebraic Riccati equations (AREs). Use in made of the so-called detectability and stabilizability conditions of generalized Kalman-filters of Gaussian processes [8, 9]. More specifically, $C_n^{fb,S}(\kappa,s)$ is a functional of one generalized DRE, while $C_n^{fb}(\kappa), C_n^{fb}(\kappa,s)$, are functionals of two generalized DREs. Also, we show that for certain noise models, and under certain conditions, it holds that $C^{fb,o}(\kappa,s) = C^{fb}(\kappa)$, i.e., these values do not depend on the initial state or initial distributions.

In Chapter 5, we calculate the maximum feedback capacity $C^{fb,S,o}(\kappa,s)$, and we show that it doesn't always exist. Otherwise, there is always an achievable rate. Also, for our simplicity, we prefer to use an $ARMA(a,c)$ scalar noise representation.

Finally, in Chapter 6, we highlight the main findings and suggest future potential work.

Chapter 2

Feedback Capacity with Unknown Initial State

In this chapter we derive equivalent sequential characterizations, for $C_n^{fb}(\kappa)$ defined by (1.1.2), i.e., the Cover and Pombra $n-$FTFI capacity characterization (1.2.19) of Case I) formulation,

We organize the presentation of the material as follows:

1) Section 2.1. The main result is Theorem 2.1.1, which gives an equivalent sequential characterization of the Cover and Pombra characterization $C_n^{fb}(\kappa)$, i.e., of (1.2.18), (1.2.19). Our derivation is simple; we apply a linear transformation to the Cover and Pombra Gaussian optimal channel input X^n (1.2.20), to represent X_t, by a linear function of (V^{t-1}, Y^{t-1}) or equivalently (X^{t-1}, Y^{t-1}) and an orthogonal Gaussian innovations process Z_t, which is independent of $(Z^{t-1}, X^{t-1}, V^{t-1}, Y^{t-1})$ for $t = 1, \ldots, n$.
We apply Theorem 2.1.1 to the time-varying PO-SS$(a_t, c_t, b_t^1, b_t^2, d_t^1, d_t^2)$ noise (see Example 2.1.1), and to the nonstationary autoregressive moving average ARMA(a, c), $a \in (-\infty, \infty)$ $c \in (-\infty, \infty)$ noise, and to the stationary ARMA$(a, c), a \in (-1, 1), c \in (-1, 1)$ noise (see Example 2.1.2), which is found in many references, such as, [2].

2) Section 2.2. The main result is Theorem 2.2.1, which gives the sequential characterization of $n-$FTFI capacity for time-varying AGN channel (1.1.1) driven by the PO-SS realization of Definition 1.1.2, for the code of Definition 1.1.1. Our derivation is based on identifying a *finite-dimensional sufficient statistic* to express X_t as a functional of the sufficient statistic, instead of (V^{t-1}, Y^{t-1}) or (X^{t-1}, Y^{t-1}), and an orthogonal Gaussian innovations process. This characterization further simplifies the sequential characterization of $C_n^{fb}(\kappa)$ given in Theorem 2.1.1 (i.e., the

equivalent of (1.2.19)).

In Corollary 2.2.2 we present the application of Theorem 2.2.1 to the ARMA$(a,c), a \in (-\infty,\infty), c \in (-\infty,\infty)$ noise of Example 2.1.2, and show that the $n-$FTFI capacity is expressed in terms of solutions to two DREs.

For Gaussian distributed Random Variables, we use the follows. $\mathbf{P}_X \in N(\mu_X, K_X), K_X \succeq 0$ denotes a Gaussian distributed RV X, with mean value μ_X and covariance matrix $K_X = cov(X,X) \succeq 0$, defined by

$$\mu_X \triangleq \mathbf{E}\{X\}, \quad K_X = cov(X,X) \triangleq \mathbf{E}\Big\{\Big(X - \mathbf{E}\Big\{X\Big\}\Big)\Big(X - \mathbf{E}\Big\{X\Big\}\Big)^T\Big\}. \tag{2.0.1}$$

Given another Gaussian random variables $Y : \Omega \rightarrow \mathbb{R}^{n_y}, n_y \in \mathbb{Z}_+^n$, which is jointly Gaussian distributed with X, i.e., the joint distribution is $\mathbf{P}_{X,Y}$, the conditional covariance of X given Y is defined by

$$K_{X|Y} = cov(X,X\Big|Y) \triangleq \mathbf{E}\Big\{\Big(X - \mathbf{E}\Big\{X\Big|Y\Big\}\Big)\Big(X - \mathbf{E}\Big\{X\Big|Y\Big\}\Big)^T\Big|Y\Big\} \tag{2.0.2}$$

$$= \mathbf{E}\Big\{\Big(X - \mathbf{E}\Big\{X\Big|Y\Big\}\Big)\Big(X - \mathbf{E}\Big\{X\Big|Y\Big\}\Big)^T\Big\} \tag{2.0.3}$$

where the last equality is due to a property of jointly Gaussian distributed RVs.

Given three arbitrary RVs (X,Y,Z) with induced distribution $\mathbf{P}_{X,Y,Z}$, the RVs (X,Z) are called conditionally independent given the RV Y if $\mathbf{P}_{Z|X,Y} = \mathbf{P}_{Z|Y}$. This conditional independence is often denoted by, $X \leftrightarrow Y \leftrightarrow Z$ is a Markov chain.

2.1 Preliminary Characterizations of $n-$FTFI Capacity of AGN Channels Driven by Correlated Noise

We start with preliminary calculations, for the feedback code of Definition 1.1.1, which we use to prove Theorem 2.1.1. These calculations are introduced for the sake of clarity and to establish our notation.

For the feedback code of Definition 1.1.1, by the channel definition (1.1.1), i.e., (1.1.10), the conditional distribution of Y_t given $Y^{t-1} = y^{t-1}, X^t = x^t$, is

$$\mathbb{P}\Big\{Y_t \in dy \Big| Y^{t-1} = y^{t-1}, X^t = x^t\Big\} = \mathbb{P}\Big\{Y_t \in dy \Big| Y^{t-1} = y^{t-1}, X^t = x^t, V^{t-1} = v^{t-1}\Big\}, \text{by (1.1.1)} \tag{2.1.4}$$

$$= \mathbf{P}_{V_t|V^{t-1}}\Big(v_t : x_t + v_t \in dy\Big), \ \ t = 2, \ldots, n, \quad \text{by (1.1.10)} \tag{2.1.5}$$

$$= \mathbf{P}_{Y_t|X_t, V^{t-1}} \tag{2.1.6}$$

$$\equiv \mathbf{P}_t(dy|x_t, v^{t-1}), \tag{2.1.7}$$

$$\mathbb{P}\Big\{Y_1 \in dy \Big| Y^0 = y^0, X^1 = x^1\Big\} = \mathbf{P}_{Y_1|X_1} \equiv \mathbf{P}_1(dy|x_1). \tag{2.1.8}$$

We introduce the set of channel input distributions with feedback, which are consistent with the code of Definition 1.1.1, not necessarily generated by the messages W, as follows:

$$\mathscr{P}_{[0,n]}(\kappa) \triangleq \Big\{P_t(dx_t|x^{t-1}, y^{t-1}) \triangleq \mathbf{P}_{X_t|X^{t-1}, Y^{t-1}}, t = 1, \ldots, n : \frac{1}{n}\mathbf{E}^P\Big(\sum_{t=1}^{n}(X_t)^2\Big) \leq \kappa\Big\}. \tag{2.1.9}$$

By Definition 1.1.1, we have $\mathscr{E}_{[0,n]}(\kappa) \subseteq \mathscr{P}_{[0,n]}(\kappa)$. Moreover, by the channel definition, any pair of the sequence triple (V^t, X^t, Y^t) uniquely defines the remaining sequence. Thus, the identity holds:

$$\overline{\mathscr{P}}_{[0,n]}(\kappa) \triangleq \Big\{\overline{P}_t(dx_t|v^{t-1}, y^{t-1}), t = 1, \ldots, n : \frac{1}{n+1}\mathbf{E}^{\overline{P}}\Big(\sum_{t=1}^{n}(X_t)^2\Big) \leq \kappa\Big\} = \mathscr{P}_{[0,n]}(\kappa). \tag{2.1.10}$$

We also emphasize that, by Definition 1.1.1, for a given feedback encoder strategy $e(\cdot) \in \mathscr{E}_{[0,n]}(\kappa)$, i.e., $x_1 = e_1(w), x_2 = e_2(w, x_1, y_1), \ldots, x_n = e_n(w, x^{n-1}, y^{n-1})$ the conditional distributions of Y_t given $(Y^{t-1}, W) = (y^{t-1}, w)$ depend on the strategies, $e(\cdot)$ as follows:

$$\mathbf{P}^e_{Y_t|W, Y^{t-1}}(dy_t|, y^{t-1}, w) \overset{(a)}{=} \mathbf{P}_t(dy_t|\{e_j(w, x^{j-1}, y^{j-1}) : j = 1, \ldots, t\}, y^{t-1}, w) \tag{2.1.11}$$

$$\overset{(b)}{=} \mathbf{P}_t(dy_t|\{e_j(w, x^{j-1}, y^{j-1}) : j = 1, \ldots, t\}, y^{t-1}, v^{t-1}, w) \tag{2.1.12}$$

$$\overset{(c)}{=} \mathbf{P}_t(dy_t|\{e_j(w, x^{j-1}, y^{j-1}) : j = 1, \ldots, t\}, v^{t-1}, w) \tag{2.1.13}$$

$$\overset{(d)}{=} \mathbf{P}_t(dy_t|\{e_j(w, x^{j-1}, y^{j-1}) : j = 1, \ldots, t\}, v^{t-1}) \tag{2.1.14}$$

$$\overset{(e)}{=} \mathbf{P}_t(dy_t|e_t(w, x^{t-1}, y^{t-1}), v^{t-1}) \tag{2.1.15}$$

(a) is due to knowledge of the distribution of the strategies $e_j(\cdot), j = 1, \ldots, t$, the code definition, and the recursive substitution, $x_1 = e_1(w), x_2 = e_2(w, x_1, y_1), \ldots, e_t(w, x^{t-1}, y^{t-1})$, where x^{t-1} is specified by the knowledge of the strategies, $e_j(\cdot), j = 1, \ldots, t-1$ and the knowledge of (y^{t-2}, w), (b) is due to knowing $x_j = e_j(w, x^{j-1}, y^{j-1}), y_j, j = 1, \ldots, t-1$ specifies $v_j = y_j - x_j, j = 1, \ldots, t -$

1,
(c) is due to the fact that, any pair of the triple (x^t, y^t, v^t) specifies the remaining sequence, i.e., knowing (x^{t-1}, v^{t-1}) specifies y^{t-1}, and hence y^{t-1} is redundant,
(d) is due to the conditional independence $\mathbf{P}_{V_t|V^{t-1},X^t,W} = \mathbf{P}_{V_t|V^{t-1},X^t}$,
(e) is due to (1.1.10), i.e., $\mathbf{P}_{V_t|V^{t-1},X^t} = \mathbf{P}_{V_t|V^{t-1}}$, and the channel definition.

By the channel definition $Y_t = X_t + V_t, t = 1,\ldots,n$, then each $e(\cdot) \in \mathscr{E}_{[0,n]}(\kappa)$ is also expressed as

$$\begin{aligned} x_1 =& e_1(w) = \overline{e}_1(w), \quad x_2 = e_2(w,x_1,y_1) = \tilde{e}_2(w,x_1,v_1,y_1) \overset{(a)}{=} \overline{e}_2(w,v_1,y_1), \quad \ldots, \\ & x_n = e_n(w,x^{n-1},y^{n-1}) = \tilde{e}_n(w,x^{n-1},v^{n-1},y^{n-1}) \overset{(a)}{=} \overline{e}_n(w,v^{n-1},y^{n-1}), \quad w \in \mathscr{M}^{(n)}. \end{aligned} \tag{2.1.16}$$

where (a) is due to the channel definition, i.e., the presence of x^{t-1} in $\tilde{e}_t(\cdot, v^{t-1}, \cdot)$ can be removed, since it is redundant, and specified by (v^{t-1}, y^{t-1}). Consequently, we have the identity

$$\begin{aligned} \overline{\mathscr{E}}_{[0,n]}(\kappa) &\triangleq \Big\{x_1 = \overline{e}_1(w), x_2 = \overline{e}_2(w,v_1,y_1) \ldots, x_n = \overline{e}_n(w,v^{n-1},y^{n-1}) : \frac{1}{n}\mathbf{E}^{\overline{e}}\Big(\sum_{i=1}^{n}(X_t)^2\Big) \leq \kappa\Big\} \\ &= \mathscr{E}_{[0,n]}(\kappa). \end{aligned} \tag{2.1.17}$$

In the next theorem we present our preliminary equivalent sequential characterization of the Cover and Pombra characterization $C_n^{fb}(\kappa)$, i.e., of (1.2.18), under encoder strategies $\mathscr{E}_{[0,n]}(\kappa) = \overline{\mathscr{E}}_{[0,n]}(\kappa)$, and channel input distributions $\mathscr{P}_{[0,n]}(\kappa) = \overline{\mathscr{P}}_{[0,n]}(\kappa)$. Unlike the Cover and Pombra [3] realization of X^n, given by (1.2.20)), at each time t, X_t is driven by an orthogonal Gaussian process Z_t.

Theorem 2.1.1. *Information structures of maximizing distributions for AGN Channels*
Consider the AGN channel (1.1.1), i.e., with noise distribution $\mathbf{P}_{V^n}$, and the code of Definition 1.1.1. Then the following hold.
(a) The inequality holds,

$$\sup_{\overline{\mathscr{E}}_{[0,n]}(\kappa)} \sum_{t=1}^{n} H^{\overline{e}}(Y_t|Y^{t-1}) \leq \sup_{\overline{\mathscr{P}}_{[0,n]}(\kappa)} \sum_{t=1}^{n} H^{\overline{P}}(Y_t|Y^{t-1}) \tag{2.1.18}$$

where the conditional (differential) entropy $H^{\overline{e}}(Y_t|Y^{t-1})$ is evaluated with respect to the probability distribution $\mathbf{P}_t^{\overline{e}}(dy_t|y^{t-1})$, defined by

$$\mathbf{P}_t^{\overline{e}}(dy_t|y^{t-1}) = \int \mathbf{P}_t(dy_t|\overline{e}_t(w,v^{t-1},y^{t-1}),v^{t-1})\, \mathbf{P}_t^{\overline{e}}(dw,dv^{t-1}|y^{t-1}), \quad t = 0,\ldots,n. \tag{2.1.19}$$

and $H^{\overline{P}}(Y_t|Y^{t-1})$ *is evaluated with respect to the probability distribution* $\mathbf{P}_t^{\overline{P}}(dy_t|y^{t-1})$*, defined by*

$$\mathbf{P}_t^{\overline{P}}(dy_t|y^{t-1}) = \int \mathbf{P}_t(dy_t|x_t, v^{t-1})\, \mathbf{P}_t^{\overline{P}}(dx_t|v^{t-1}, y^{t-1})\, \mathbf{P}_t^{\overline{P}}(dv^{t-1}|y^{t-1}), \quad t = 0, \ldots, n. \tag{2.1.20}$$

(b) The optimal channel input distribution $\{\overline{P}(dx_t|v^{t-1}, y^{t-1}), t = 1, \ldots, n\} \in \overline{\mathscr{P}}_{[0,n]}(\kappa)$*, which maximizes* $\sum_{t=1}^{n} H^{\overline{P}}(Y_t|Y^{t-1})$ *of part (a), i.e., the right hand side of (2.1.18), is induced by a channel input process* X^n*, which is conditionally Gaussian, with linear conditional mean and nonrandom conditional covariance, given by*

$$\mathbf{E}^{\overline{P}}\Big\{X_t\Big|V^{t-1}, Y^{t-1}\Big\} = \begin{cases} \Gamma_t^1 \mathbf{V}^{t-1} + \Gamma_t^2 \mathbf{Y}^{t-1}, & \text{for} \quad t = 2, \ldots, n \\ 0, & \text{for} \quad t = 1, \end{cases} \tag{2.1.21}$$

$$K_{X_t|V^{t-1},Y^{t-1}} \stackrel{\triangle}{=} cov\Big(X_t, X_t\Big|V^{t-1}, Y^{t-1}\Big) = K_{Z_t} \succeq 0, \quad t = 1, \ldots, n \tag{2.1.22}$$

and such that the average constraint holds and (1.1.10) is respected.
(c) The optimal channel input distribution $\{\overline{P}(dx_t|v^{t-1}, y^{t-1}), t = 1, \ldots, n\} \in \overline{\mathscr{P}}_{[0,n]}(\kappa)$ *of part (b), is induced by a jointly Gaussian process* X^n*, with a realization given by*

$$X_t = \sum_{j=1}^{t-1} \Gamma_{t,j}^1 V_j + \sum_{j=1}^{t-1} \Gamma_{t,j}^2 Y_j + Z_t, \quad X_1 = Z_1, \quad t = 2, \ldots, n, \tag{2.1.23}$$

$$= \Gamma_t^1 \mathbf{V}^{t-1} + \Gamma_t^2 \mathbf{Y}^{t-1} + Z_t, \tag{2.1.24}$$

$$Z_t \in N(0, K_{Z_t}), \quad t = 1, \ldots, n \quad \text{a Gaussian sequence}, \tag{2.1.25}$$

$$Z_t \quad \text{independent of} \quad (V^{t-1}, X^{t-1}, Y^{t-1}), \quad t = 1, \ldots, n, \tag{2.1.26}$$

$$Z^n \quad \text{independent of} \quad V^n, \tag{2.1.27}$$

$$\frac{1}{n}\mathbf{E}\Big\{\sum_{t=1}^{n} \big(X_t\big)^2\Big\} \leq \kappa, \tag{2.1.28}$$

$$(\Gamma_t^1, \Gamma_t^2, K_{Z_t}) \in (-\infty, \infty) \times (-\infty, \infty) \times [0, \infty) \quad \text{nonrandom}. \tag{2.1.29}$$

(d) An equivalent characterization of the $n-$*FTFI capacity* $C_n^{fb}(\kappa)$*, defined by (1.2.18), (1.2.19), is given by*

$$C_n^{fb}(\kappa) = \sup_{\frac{1}{n}\mathbf{E}\big\{\sum_{t=1}^{n}\big(X_t\big)^2\big\} \leq \kappa} \sum_{t=1}^{n} H^{\overline{P}}(Y_t|Y^{t-1}) - H(V^n) \tag{2.1.30}$$

where the supremum is over all $(\Gamma_t^1, \Gamma_t^2, K_{Z_t}), t = 1, \ldots, n$ *of the realization of part (c), that induces the distribution* $\overline{P}_t(dx_t|v^{t-1}, y^{t-1}), t = 1, \ldots, n.$

Proof. See Appendix 7.1. □

Theorem 2.1.2. *Converse coding theorem for code of Definition 1.1.1*
Consider the AGN channel (1.1.1).
(a) Any achievable rate R for the code of Definition 1.1.1 satisfies

$$R \leq C^{fb}(\kappa) \stackrel{\triangle}{=} \lim_{n \longrightarrow \infty} \frac{1}{n} C_n^{fb}(\kappa), \tag{2.1.31}$$

$$C_n^{fb}(\kappa) = \sup_{\overline{P}_t(dx_t|v^{t-1},y^{t-1}),t=1,\dots,n:\ \frac{1}{n}\mathbf{E}\left\{\sum_{t=1}^n \left(X_t\right)^2\right\} \leq \kappa} \sum_{t=1}^n H^{\overline{P}}(Y_t|Y^{t-1}) - H(V^n) \tag{2.1.32}$$

provided the supremum exists and the limit exists, where the right hand side of (2.1.32) is given in Theorem 2.1.1.

Proof. Follows from standard arguments, using Fano's inequality (see also [3]) and Theorem 2.1.1. □

Remark 2.1.1.
(a) From the realization of X^n given by (2.1.23), we can recover the Cover and Pombra [3] realization (1.2.20), by recursive substitution of Y^{t-1} into the right hand side of (2.1.23), as follows.

$$X_t = \sum_{j=1}^{t-1} \Gamma_{t,j}^1 V_j + \sum_{j=1}^{t-1} \Gamma_{t,j}^2 Y_j + Z_t \tag{2.1.33}$$

$$= \sum_{j=1}^{t-1} \Gamma_{t,j}^1 V_j + \sum_{j=1}^{t-2} \Gamma_{t,j}^2 Y_j + \Gamma_{t,t-1}^2 \Big(X_{t-1} + Z_{t-1}\Big) + Z_t \tag{2.1.34}$$

$$= \sum_{j=1}^{t-1} B_{t,j} V_j + \overline{Z}_t, \quad \text{by recursive substitution of } X_1, \dots, X_{t-1}, Y_1, \dots, Y_{t-2} \tag{2.1.35}$$

for some $\overline{Z}_t \in (0, K_{\overline{Z}_t})$ which is jointly correlated, and some nonrandom $B_{t,j}$, as given by (1.2.20) and (1.2.21). Clearly, (2.1.35) is precisely (1.2.20).
(b) Unlike the Cover and Pombra [3] realization of X^n, i.e., (1.2.20), the realization of X^n given by (2.1.23), is such that, at each time t, X_t depends on $(\mathbf{V}^{t-1}, \mathbf{Y}^{t-1}, Z_t)$, where Z^t is an innovations or orthogonal process, i.e., (2.1.26) holds.
(c) In subsequent parts of the paper we show that the minimizing sequence X^n given by (2.1.23) is such that $\Gamma_t^2 = -\Gamma_t^1, t = 2, \dots, n$. Then we derive an equivalent sequential characterization of the Cover and Pombra $n-$FTFI capacity (1.2.19), which is simplified further, by the use of a sufficient statistic (that satisfies a Markov recursion).

To characterize $C_n^{fb}(\kappa)$ using Theorem 2.1.1.(d) we need to compute the (differential) entropy $H(V^n)$ of V^n. The following lemma is useful in this respect.

Lemma 2.1.1. *Entropy $H(V^n)$ calculation from generalized Kalman-filter of the PO-SS noise realization.*

Consider the PO-SS realization of V^n of Definition 1.1.2. Define the conditional covariance and conditional mean of S_t given V^{t-1} by

$$\Sigma_t \stackrel{\triangle}{=} cov\big(S_t, S_t \Big| V^{t-1}\big) = \mathbf{E}\Big\{\Big(S_t - \hat{S}_t\Big)\Big(S_t - \hat{S}_t\Big)^T \Big| V^{t-1}\Big\}, \;\; \hat{S}_t \stackrel{\triangle}{=} \mathbf{E}\Big\{S_t \Big| V^{t-1}\Big\}, \;\; t = 2, \ldots, n, \tag{2.1.36}$$

$$\Sigma_1 \stackrel{\triangle}{=} cov\big(S_1, S_1\big) = K_{S_1}, \;\; \hat{S}_1 \stackrel{\triangle}{=} \mu_{S_1}. \tag{2.1.37}$$

Then the following hold.

(a) The conditional distribution of V_t conditioned on V^{t-1} is Gaussian, i.e.,

$$\mathbf{P}_{V_t|V^{t-1}} \in N(\mu_{V_t|V^{t-1}}, K_{V_t|V^{t-1}}), \quad t = 1, \ldots, n \tag{2.1.38}$$

where $\mu_{V_t|V^{t-1}} \stackrel{\triangle}{=} \mathbf{E}\big\{V_t \Big| V^{t-1}\big\}, K_{V_t|V^{t-1}} \stackrel{\triangle}{=} cov\big(V_t, V_t \Big| V^{t-1}\big)$.

(b) The conditional mean and covariance $\mu_{V_t|V^{t-1}}, K_{V_t|V^{t-1}}$ are given by the Generalized Kalman-filter recursions, as follows.

(i) The optimal mean-square error estimate $\hat{S}_t$ satisfies the generalized Kalman-filter recursion

$$\hat{S}_{t+1} = A_t\hat{S}_t + M_t(\Sigma_t)\hat{I}_t, \;\; \hat{S}_1 = \mu_{S_1}, \tag{2.1.39}$$

$$M_t(\Sigma_t) \stackrel{\triangle}{=} \Big(A_t\Sigma_tC_t^T + B_tK_{W_t}N_t^T\Big)\Big(N_tK_{W_t}N_t^T + C_t\Sigma_tC_t^T\Big)^{-1}, \tag{2.1.40}$$

$$\hat{I}_t \stackrel{\triangle}{=} V_t - \mathbf{E}\Big\{V_t \Big| V^{t-1}\Big\} = V_t - C_t\hat{S}_t = C_t\big(S_t - \hat{S}_t\big) + N_tW_t, \;\; t = 1, \ldots, n, \tag{2.1.41}$$

$\hat{I}_t \in N(0, K_{\hat{I}_t}), \;\; t = 1, \ldots, n$ *is an orthogonal innovations process, i.e., $\hat{I}_t$ is independent of $\hat{I}_s$, for all $t \neq s$, and $\hat{I}_t$ is independent of V^{t-1},* (2.1.42)

$$K_{\hat{I}_t} \stackrel{\triangle}{=} cov(\hat{I}_t, \hat{I}_t) = C_t\Sigma_tC_t^T + N_tK_{W_t}N_t^T. \tag{2.1.43}$$

(ii) The error $E_t \stackrel{\triangle}{=} S_t - \hat{S}_t$ satisfies the recursion

$$E_{t+1} = M_t^{CL}(\Sigma_t)E_t + M_t(\Sigma_t)N_tW_t, \;\; E_1 = S_1 - \hat{S}_1, \;\; t = 1, \ldots, n, \tag{2.1.44}$$

$$M_t^{CL}(\Sigma_t) \stackrel{\triangle}{=} A_t - M_t(\Sigma_t)C_t. \tag{2.1.45}$$

(iii) The covariance of the error is such that $\mathbf{E}\{E_t E_t^T\} = \Sigma_t$ *and satisfies the generalized matrix DRE*

$$\Sigma_{t+1} = A_t \Sigma_t A_t^T + B_t K_{W_t} B_t^T - \Big(A_t \Sigma_t C_t^T + B_t K_{W_t} N_t^T\Big)\Big(N_t K_{W_t} N_t^T + C_t \Sigma_t C_t^T\Big)^{-1} \cdot \Big(A_t \Sigma_t C_t^T + B_t K_{W_t} N_t^T\Big)^T, \quad t = 1, \ldots, n, \quad \Sigma_1 = K_{S_1} \succeq 0, \quad \Sigma_t \succeq 0. \tag{2.1.46}$$

(iv) The conditional mean and covariance $\mu_{V_t|V^{t-1}}, K_{V_t|V^{t-1}}$ *are given by*

$$\mu_{V_t|V^{t-1}} = C_t \hat{S}_t, \quad t = 1, \ldots, n, \tag{2.1.47}$$

$$K_{V_t|V^{t-1}} = K_{\hat{I}_t} = C_t \Sigma_t C_t^T + N_t K_{W_t} N_t^T, \quad t = 1, \ldots, n. \tag{2.1.48}$$

(v) The entropy of V^n*, is given by*

$$H(V^n) = \frac{1}{2}\sum_{t=1}^{n} \log\Big(2\pi e\Big[C_t \Sigma_t C_t^T + N_t K_{W_t} N_t^T\Big]\Big) \tag{2.1.49}$$

Proof. (a), (b).(i)-(iv). The generalized Kalman filter of the PO-SS realization of V^n and accompanied statements can be found in many textbooks [9]. However, it is noted that $\hat{I}_t, t = 2, \ldots, n$, $\hat{I}_1 = V_1$ are all independent Gaussian. For example, to show show (2.1.44) we write the recursion for $E_t = S_t - \hat{S}_t$ using part (i) and the realization of S_t. (b).(v) By the chain rule of joint entropy then

$$H(V^n) = H(V_1) + \sum_{t=2}^{n} H(V_t | V^{t-1}) \tag{2.1.50}$$

$$= H(V_1) + \sum_{t=2}^{n} H\Big(V_t - \mathbf{E}\Big\{V_t \Big| V^{t-1}\Big\} | V^{t-1}\Big) \tag{2.1.51}$$

$$= H(V_1) + \sum_{t=2}^{n} H(\hat{I}_t), \quad \text{by orthogonality of } \hat{I}_t \triangleq V_t - \mathbf{E}\Big\{V_t \Big| V^{t-1}\Big\} \text{ and } V^{t-1} \tag{2.1.52}$$

From (2.1.52) and (2.1.48), then follows (2.1.49), from the entropy formula of Gaussian RVs. □

Next we introduce an example of a PO-SS realization of the noise that we often use.

Example 2.1.1. *A time-varying PO-SS$(a_t, c_t, b_t^1, b_t^2, d_t^1, d_t^2)$ noise realization is defined by*

$$S_{t+1} = a_t S_t + b_t^1 W_t^1 + b_t^2 W_t^2, \quad t = 1, 2, \ldots, n-1 \tag{2.1.53}$$

$$V_t = c_t S_t + d_t^1 W_t^1 + d_t^2 W_t^2, \quad t = 1, \ldots, n, \tag{2.1.54}$$

$$S_1 \in N(\mu_{S_1}, K_{S_1}), \quad K_{S_1} \geq 0, \quad W_t^i \in N(0, K_{W_t^i}), \quad K_{W_t^i} \geq 0, \quad i = 1, 2, \quad t = 1, \ldots, n, \tag{2.1.55}$$

$$W^{1,n} \text{ and } W^{2,n} \text{ indep. seq. and indep. of } S_1, \tag{2.1.56}$$

$$a_t \in \mathbb{R}, \quad c_t \in \mathbb{R}, \quad b_t^i \in \mathbb{R}, \quad d_t^i \in \mathbb{R}, \quad i = 1, 2, \forall t \text{ are nonrandom}, \tag{2.1.57}$$

$$b_t \circ b_t \stackrel{\triangle}{=} (b_t^1)^2 K_{W_t^1} + (b_t^2)^2 K_{W_t^2}, \quad b_t \circ d_t \stackrel{\triangle}{=} b_t^1 K_{W_t^1} d_t^1 + b_t^2 K_{W_t^2} d_t^2,$$

$$d_t \circ d_t \stackrel{\triangle}{=} (d_t^1)^2 K_{W_t^1} + (d_t^2)^2 K_{W_t^2} > 0, \ \forall t. \tag{2.1.58}$$

The next corollary is an application of Lemma 2.1.1 to the time-varying PO-SS noise of Example 2.1.1.

Corollary 2.1.1. *The entropy $H(V^n)$ of the PO-SS$(a_t, c_t, b_t^1, b_t^2, d_t^1, d_t^2)$ noise of Example 2.1.1 is computed from Lemma 2.1.1 with the following changes:*

$$C_t \longmapsto c_t, \quad A_t \longmapsto a_t, \quad B_t K_{W_t} N_t^T \longmapsto b_t \circ d_t, \quad B_t K_{W_t} B_t^T \longmapsto b_t \circ b_t, \quad N_t K_{W_t} N_t^T \longmapsto d_t \circ d_t. \tag{2.1.59}$$

Proof. This is easily verified. □

From Corollary 2.1.1 we have the following observations.

Remark 2.1.2. *Consider the PO-SS$(a_t, c_t, b_t^1, b_t^2, d_t^1, d_t^2)$ noise of Example 2.1.1. Then the following hold.*
(a) Consider the code of Definition 1.1.2. At each time t, the optimal channel input process is either realized by the Cover and Pombra process X_t given by (1.2.20), or equivalently by (2.1.23), i.e., $X_t = \sum_{j=1}^{t-1} B_{t,j} V_{t,j} + \overline{Z}_t = \sum_{j=1}^{t-1} \Gamma_{t,j}^1 V_j + \sum_{j=1}^{t-1} \Gamma_{t,j}^2 Y_j + Z_t$. Moreover, X_t cannot be expressed in terms of the state S^t, because by (2.1.53) and (2.1.54) the noise sequence V^{t-1} does not specify S^t, for $t = 1, \ldots, n$.

We also apply our results to various versions the autoregressive moving average (ARMA) noise model, such as, the double-side and single-sided, stationary version of the ARMA noise, previously analyzed in [2] and in many other papers.

Example 2.1.2. *The time-invariant ARMA(a,c) noise*
(a) The time-invariant one-sided, stable or unstable, autoregressive moving average (ARMA$(a,c), a \in (-\infty,\infty), c \in (-\infty,\infty)$) noise, is defined by

$$V_t = cV_{t-1} + W_t - aW_{t-1}, \quad \forall t \in \mathbb{Z}_+ \stackrel{\triangle}{=} \{1,2,\ldots\}, \tag{2.1.60}$$

$$V_0 \in N(0,K_{V_0}), \quad K_{V_0} \geq 0, \quad W_0 \in N(0,K_{W_0}), \quad K_{W_0} \geq 0, \quad W_t \in N(0,K_W), \quad K_W > 0, \tag{2.1.61}$$

$$\{W_0, W_1, \ldots, W_n\} \textit{ indep. seq. and indep. of } V_0, \tag{2.1.62}$$

$$c \in (-\infty,\infty), \quad a \in (-\infty,\infty), \quad c \neq a. \tag{2.1.63}$$

To express the AR(a,c) in state space form we define the state variable of the noise by

$$S_t \stackrel{\triangle}{=} \frac{cV_{t-1} - aW_{t-1}}{c-a}, \quad \forall t \in \mathbb{Z}_+ \tag{2.1.64}$$

Then, the state space realization of V^n is

$$S_{t+1} = cS_t + W_t, \quad \forall t \in \mathbb{Z}_+, \tag{2.1.65}$$

$$V_t = \Big(c-a\Big)S_t + W_t, \quad \forall t \in \mathbb{Z}_+, \tag{2.1.66}$$

$$K_{S_1} = \frac{\big(c\big)^2 K_{V_0} + \big(a\big)^2 K_{W_0}}{\Big(c-a\Big)^2}, \quad K_{V_0} \geq 0, \quad K_{W_0} \geq 0 \quad \textit{both given}. \tag{2.1.67}$$

We note that the AR(a,c) is not necessarily stationary or asymptotically stationary.
A special case of the AR(a,c) is the AR(c) noise (i.e., with $a=0$) defined by

$$V_t = cV_{t-1} + W_t, \quad t = 1,2,\ldots, \quad K_{V_0} \geq 0, \quad K_W > 0. \tag{2.1.68}$$

(b) Double-Sided Wide-Sense Stationary ARMA$(a,c), a \in [-1,1], c \in (-1,1)$ Noise.
A double-sided wide-sense stationary ARMA(a,c) noise is defined by

$$V_t = cV_{t-1} + W_t - aW_{t-1}, \quad \forall t \in \mathbb{Z} \stackrel{\triangle}{=} \{\ldots,-1,0,1,\ldots\}, \quad |a| \leq 1, \, |c| < 1. \tag{2.1.69}$$

where $W_t, \forall t \in \mathbb{Z}$ is an independent and identically distributed Gaussian sequence, i.e., $W_t \in N(0,K_W)$, $\forall t$. The power spectral density (PSD) of the wide-sense stationary noise is (this corresponds to [2, eqn(43) with $L=1$]) is given by

$$S_V(e^{j\theta}) \stackrel{\triangle}{=} K_W \frac{\Big(1-ae^{i\theta}\Big)\Big(1-ae^{-i\theta}\Big)}{\Big(1-ce^{i\theta}\Big)\Big(1-ce^{-i\theta}\Big)}, \quad |c| < 1, \quad |a| \leq 1, \quad c \neq a, \quad K_W > 0. \tag{2.1.70}$$

We define the state process by

$$S_t \triangleq \frac{cV_{t-1} - aW_{t-1}}{c-a}, \quad \forall t \in \mathbb{Z}. \tag{2.1.71}$$

Then the stationary state space realization of $V_t, \forall t \in \mathbb{Z}$ *is*

$$S_{t+1} = cS_t + W_t, \quad \forall t \in \mathbb{Z}, \tag{2.1.72}$$

$$V_t = \Big(c-a\Big)S_t + W_t, \quad \forall t \in \mathbb{Z} \tag{2.1.73}$$

provided the initial covariances, $cov(S_t,S_t), cov(S_t,V_t), cov(V_t,V_t)$ *are chosen appropriately (see Proposition 2.1.1).*

(c) One-sided Wide-Sense Stationary $ARMA(a,c), a \in [-1,1], c \in (-1,1)$.

The one-sided wide-sense stationary $ARMA(a,c)$ *noise is defined as in part (a) with* $\forall t \in \mathbb{Z} \triangleq \{\ldots,-1,0,1,\ldots,\}$ *replaced by* $\forall t \in \mathbb{Z}_+ \triangleq \{1,2,\ldots,\}$ *and (2.1.71)-(2.1.73) hold,* $\forall t \in \mathbb{Z}_+$*, provide the initial covariances are chosen appropriately (see Proposition 2.1.1).*

Remark 2.1.3. *$ARMA(a,c)$ noise of Example 2.1.2*

(a) Consider any of the $AR(a,c)$ of Example 2.1.2. For the code of Definition 1.1.2, as stated in Remark 2.1.2.(a), the channel input process X^n cannot be expressed in terms of the state S^n.

(b) The statement of part (a), also holds for the double-sided and the one-sided wide-sense stationary $AR(a,c), a \in [-1,1], c \in (-1,1)$ of Example 2.1.2.(b), (c).

In the next proposition we state conditions for the stable realizations of Example 2.1.2.(a), i.e., $\text{AR}(a,c), a \in [-1,1], c \in (-1,1)$ to be asymptotically stationary, and for the realizations of Example 2.1.2.(b), (c) to be stationary. For stationary noise we also determine the initial conditions of the generalized Kalman-filter of Lemma 2.1.1.

Proposition 2.1.1. *Asymptotically stationary and stationary $ARMA(a,c)$ noises of Example 2.1.2*

(a) The realization of the double-sided $ARMA(a,c), a \in [-1,1], c \in (-1,1)$ noise of Example 2.1.2.(b) is stationary if the following conditions hold.

$$d_{11} \triangleq cov\big(S_t,S_t\big) = K_{S_t}, \quad d_{12} \triangleq cov\big(S_t,V_t\big) = K_{S_t,V_t}, \quad d_{22} \triangleq cov\big(V_t,V_t\big) = K_{V_t}, \quad \textit{are constants} \tag{2.1.74}$$

where the constants (d_{11}, d_{12}, d_{22}) *are given by*

$$d_{11} = \frac{K_W}{1-c^2}, \quad d_{12} = \frac{\big(c-a\big)K_W}{1-c^2}, \quad d_{22} = \frac{\big(c-a\big)^2 K_W}{1-c^2} + K_W. \tag{2.1.75}$$

Similarly, the one-sided ARMA$(a,c), a \in [-1,1], c \in (-1,1)$ *noise of Example 2.1.2.(c) is stationary if the above equations hold* $\forall t \in \mathbb{Z}_+ \stackrel{\triangle}{=} \{1,2,\ldots\}$.
(b) The realization of the ARMA(a,c) *noise of Example 2.1.2.(a) is asymptotically stationary if* $a \in [-1,1], c \in (-1,1)$.

Proof. See Appendix 7.2. □

2.2 A Sufficient Statistic Approach to the Characterization of $n-$FTFI Capacity of AGN Channels Driven by PO-SS Noise Realizations

The characterization of the $n-$FTFI capacity via (1.2.19), equivalently given in Theorem 2.1.1.(d), although compactly represented, is not very practical, because the input process X^n is not expressed in terms of a *sufficient statistic* that summarizes the information of the channel input strategy. Over the years, such stochastic optimization problems enjoyed much progress via the use of a sufficient statistic [14].

In this section, we wish to identify a *sufficient statistic* for the input process X_t, given by (2.1.23), called the *state of the input*, which summarizes the information contained in (V^{t-1}, Y^{t-1}). It will then become apparent that the characterization of the $n-$FTFI capacity for the Cover and Pombra formulation and code of Definition 1.1.1, can be expressed as a functional of *two generalized matrix DREs*.

First, we invoke Theorem 2.1.1 and Lemma 2.1.1 to show that for each time t, X_t is expressed as

$$X_t = \Lambda_t\Big(\hat{S}_t - \mathbf{E}\Big\{\hat{S}_t\Big|Y^{t-1}\Big\}\Big) + Z_t, \quad t = 1,\ldots,n, \tag{2.2.76}$$

$$\hat{S}_t \stackrel{\triangle}{=} \mathbf{E}\Big\{S_t\Big|V^{t-1}\Big\}, \quad \widehat{\hat{S}}_t \stackrel{\triangle}{=} \mathbf{E}\Big\{\hat{S}_t\Big|Y^{t-1}\Big\}. \tag{2.2.77}$$

For non-feedback, X_t is expressed as

$$X_t = \Lambda_t\overline{S}_t + DZ_t \tag{2.2.78}$$

$$\overline{S}_{t+1} = A\overline{S}_t + BZ_t \tag{2.2.79}$$

The above representation means, at each time t, the state of the channel input process X_t is $\Big(\hat{S}_t, \widehat{\hat{S}}_t\Big)$.
We show that $\widehat{\hat{S}}_t$ satisfies another generalized Kalman-filter recursion.

Now, we prepare to prove (2.2.76) and the main theorem. We start with preliminary calculations.

$$\mathbb{P}\Big\{Y_t\in dy\Big|Y^{t-1},X^t\Big\}=\mathbf{P}_t(dy|X_t,V^{t-1}),\ \ t=2,\dots,n,\ \ \text{by channel definition} \tag{2.2.80}$$
$$=\mathbf{P}_t(dy|X_t,V^{t-1},\hat{S}^t),\ \ \text{by } \hat{S}_t=\mathbf{E}\Big\{S_t\Big|V^{t-1}\Big\} \tag{2.2.81}$$
$$=\mathbf{P}_t(dy|X_t,V^{t-1},\hat{S}_t,\hat{I}^{t-1}),\ \ \text{by (2.1.41), i.e., } V_t=C_t\hat{S}_t+\hat{I}_t \tag{2.2.82}$$
$$=\mathbf{P}_t(dy|X_t,\hat{S}_t),\ \ \text{by } Y_t=X_t+V_t=X_t+C_t\hat{S}_t+\hat{I}_t \text{ and (2.1.42).} \tag{2.2.83}$$

At $t=1$ we also have $\mathbb{P}\Big\{Y_1\in dy\Big|X_1\Big\}=\mathbf{P}_1(dy|X_1)$. By (2.2.83), it follows that the conditional distribution of Y_t given $Y^{t-1}=y^{t-1}$ is

$$\mathbf{P}_t(dy_t|y^{t-1})=\int\mathbf{P}_t(dy|x_t,\hat{s}_t)\mathbf{P}_t(dx_t|\hat{s}_t,y^{t-1})\mathbf{P}_t(d\hat{s}_t|y^{t-1}),\ \ t=2,\dots,n, \tag{2.2.84}$$
$$\mathbf{P}_1(dy_1)=\int\mathbf{P}_1(dy|x_t,\hat{s}_1)\mathbf{P}_1(dx_1|\hat{s}_1)\mathbf{P}_1(d\hat{s}_1). \tag{2.2.85}$$

From the above distributions, at each time t, the distribution of X_t conditioned on (V^{t-1},Y^{t-1}), given in Theorem 2.1.1, is also expressed as a linear functional of $(\hat{S}_t,Y^{t-1})$, for $t=1,\dots,n$.
The next theorem further shows that for each t, the dependence of X_t on Y^{t-1} is expressed in terms of $\mathbf{E}\Big\{\hat{S}_t\Big|Y^{t-1}\Big\}$ for $t=1,\dots,n$, and this dependence gives rise to an equivalent sequential characterization of the Cover and Pombra $n-$FTFI capacity, $C_n^{fb}(\kappa)$.

Theorem 2.2.1. *Equivalent characterization of $n-$FTFI Capacity $C_n^{fb}(\kappa)$ for PO-SS Noise realizations*

Consider the time-varying AGN channel defined by (1.1.1), driven by a noise with the PO-SS realization of Definition 1.1.2, and the code of Definition 1.1.1. Consider also the generalized Kalman-filter of Lemma 2.1.1.
Define the conditional covariance and conditional mean of $\hat{S}_t$ given Y^{t-1}, by

$$K_t\stackrel{\triangle}{=}cov\Big(\hat{S}_t,\hat{S}_t\Big|Y^{t-1}\Big)=\mathbf{E}\Big\{\Big(\hat{S}_t-\widehat{\hat{S}}_t\Big)\Big(\hat{S}_t-\widehat{\hat{S}}_t\Big)^T\Big\},\ \ \widehat{\hat{S}}_t\stackrel{\triangle}{=}\mathbf{E}\Big\{\hat{S}_t\Big|Y^{t-1}\Big\},\ \ t=2,\dots,n, \tag{2.2.86}$$
$$\widehat{\hat{S}}_1\stackrel{\triangle}{=}\mu_{S_1},\ \ K_1\stackrel{\triangle}{=}0. \tag{2.2.87}$$

Then the following hold.
(a) An equivalent characterization of the $n-$FTFI capacity $C_n^{fb}(\kappa)$, defined by (1.2.20)-(1.2.19), is

$$C_n^{fb}(\kappa)=\sup_{\mathscr{P}^{\hat{S}}_{[0,n]}(\kappa)}\sum_{t=1}^n H(Y_t|Y^{t-1})-H(V^n) \tag{2.2.88}$$

where (X^n, Y^n) *is jointly Gaussian, and*

$$H(V^n) \text{ is the entropy of } V^n \text{ given in Lemma 2.1.1, i.e., (2.1.49)}, \tag{2.2.89}$$

$$\hat{I}^n \text{ is the innovations process of } V^n \text{ given in Lemma 2.1.1}, \tag{2.2.90}$$

$$Y_t = X_t + V_t, \quad t = 1, \ldots, n, \tag{2.2.91}$$

$$V_t = C_t \hat{S}_t + \hat{I}_t, \tag{2.2.92}$$

$$\mathbf{P}_t(dy_t|y^{t-1}) = \int \mathbf{P}_t(dy|x_t, \hat{s}_t)\mathbf{P}_t(dx_t|\hat{s}_t, y^{t-1})\mathbf{P}_t(d\hat{s}_t|y^{t-1}), \quad t = 2, \ldots, n, \tag{2.2.93}$$

$$\mathbf{P}_1(dy_1) = \int \mathbf{P}_1(dy|x_t, \hat{s}_1)\mathbf{P}_1(dx_1|\hat{s}_1)\mathbf{P}_1(d\hat{s}_1), \tag{2.2.94}$$

$$\mathbf{P}_t(dy_t|y^{t-1}) \in N(\mu_{Y_t|Y^{t-1}}, K_{Y_t|Y^{t-1}}), \tag{2.2.95}$$

$$\mu_{Y_t|Y_{t-1}} \text{ is linear in } Y^{t-1} \quad \text{and} \quad K_{Y_t|Y^{t-1}} \text{ is nonrandom}, \tag{2.2.96}$$

$$\mathbf{P}_t(dx_t|\hat{s}_t, y^{t-1}) \in N(\mu_{X_t|\hat{S}_t, Y^{t-1}}, K_{X_t|\hat{S}_t, Y^{t-1}}), \tag{2.2.97}$$

$$\mu_{X_t|\hat{S}_t, Y^{t-1}} \text{ is linear in } (\hat{S}_t, Y^{t-1}) \quad \text{and} \quad K_{X_t|\hat{S}_t, Y^{t-1}} \text{ is nonrandom}, \tag{2.2.98}$$

$$\mathscr{P}^{\hat{S}}_{[0,n]}(\kappa) \triangleq \Big\{\mathbf{P}_t(dx_t|\hat{s}_t, y^{t-1}), t = 1, \ldots, n : \frac{1}{n}\mathbf{E}\Big(\sum_{t=1}^{n}(X_t)^2\Big) \leq \kappa\Big\}. \tag{2.2.99}$$

(b) The optimal jointly Gaussian process (X^n, Y^n) *of part (a) is represented by*

$$X_t = \Lambda_t\Big(\hat{S}_t - \widehat{\hat{S}}_t\Big) + Z_t, \quad t = 1, \ldots, n, \tag{2.2.100}$$

$$Z_t \in N(0, K_{Z_t}) \text{ independent of } (X^{t-1}, V^{t-1}, \hat{S}^t, \widehat{\hat{S}}^t, \hat{I}^t, Y^{t-1}), \quad t = 1, \ldots, n, \tag{2.2.101}$$

$$\hat{I}_t \in N(0, K_{\hat{I}_t}) \text{ independent of } (X^{t-1}, V^{t-1}, \hat{S}^t, Y^{t-1}, \widehat{\hat{S}}^t), \quad t = 1, \ldots, n, \tag{2.2.102}$$

$$Y_t = \Lambda_t\Big(\hat{S}_t - \widehat{\hat{S}}_t\Big) + Z_t + V_t, \quad t = 1, \ldots, n, \tag{2.2.103}$$

$$= \Lambda_t\Big(\hat{S}_t - \widehat{\hat{S}}_t\Big) + C_t\hat{S}_t + \hat{I}_t + Z_t, \tag{2.2.104}$$

$$\frac{1}{n}\mathbf{E}\Big\{\sum_{t=1}^{n}(X_t)^2\Big\} = \frac{1}{n}\sum_{t=1}^{n}\Big(\Lambda_t K_t \Lambda_t^T + K_{Z_t}\Big). \tag{2.2.105}$$

where Λ_t *is nonrandom.*

The conditional mean and covariance, $\widehat{\hat{S}}_t$ *and* K_t*, are given by generalized Kalman-filter equations, as follows.*

(i) $\widehat{\hat{S}}_t$ *satisfies the Kalman-filter recursion*

$$\widehat{\hat{S}}_{t+1} = A_t \widehat{\hat{S}}_t + F_t(\Sigma_t, K_t) I_t, \quad \widehat{\hat{S}}_1 = \mu_{S_1}, \tag{2.2.106}$$

$$F_t(\Sigma_t, K_t) \stackrel{\triangle}{=} \Big(A_t K_t \big(\Lambda_t + C_t\big)^T + M_t(\Sigma_t) K_{\hat{I}_t}\Big)\Big\{K_{\hat{I}_t} + K_{Z_t} + \big(\Lambda_t + C_t\big) K_t \big(\Lambda_t + C_t\big)^T\Big\}^{-1} \tag{2.2.107}$$

$$I_t \stackrel{\triangle}{=} Y_t - \mathbf{E}\Big\{Y_t \Big| Y^{t-1}\Big\} = Y_t - C_t \widehat{\hat{S}}_t = \Big(\Lambda_t + C_t\Big)\Big(\hat{S}_t - \widehat{\hat{S}}_t\Big) + \hat{I}_t + Z_t, \quad t = 1, \ldots, n, \tag{2.2.108}$$

$I_t \in N(0, K_{I_t}), \quad t = 1, \ldots, n$ *is an orthogonal innovations process, i.e.,* I_t *is independent of* I_s*, for all* $t \neq s$*, and* I_t *is independent of* V^{t-1}*,* (2.2.109)

$$K_{Y_t|Y^{t-1}} = K_{I_t} \stackrel{\triangle}{=} cov(I_t, I_t) = \Big(\Lambda_t + C_t\Big) K_t \Big(\Lambda_t + C_t\Big)^T + K_{\hat{I}_t} + K_{Z_t}, \tag{2.2.110}$$

$K_{\hat{I}_t}$ *given by (2.1.43).* (2.2.111)

(ii) The error $\widehat{E}_t \stackrel{\triangle}{=} \hat{S}_t - \widehat{\hat{S}}_t$ *satisfies the recursion*

$$\widehat{E}_{t+1} = F_t^{CL}(\Sigma_t, K_t)\widehat{E}_t + F_t(\Sigma_t, K_t)\Big(\hat{I}_t + Z_t\Big), \quad \widehat{E}_1 = \hat{S}_1 - \widehat{\hat{S}}_1 = 0, \quad t = 1, \ldots, n, \tag{2.2.112}$$

$$F_t^{CL}(\Sigma_t, K_t) \stackrel{\triangle}{=} A_t - F_t(\Sigma_t, K_t)\Big(\Lambda_t + C_t\Big). \tag{2.2.113}$$

(iii) $K_t = \mathbf{E}\big\{\widehat{E}_t \widehat{E}_t^T\big\}$ *satisfies the generalized DRE*

$$\begin{aligned} K_{t+1} = {} & A_t K_t A_t^T + M_t(\Sigma_t) K_{\hat{I}_t} \big(M_t(\Sigma_t)\big)^T - \Big(A_t K_t \big(\Lambda_t + C_t\big)^T + M_t(\Sigma_t) K_{\hat{I}_t}\Big)\Big(K_{\hat{I}_t} + K_{Z_t} \\ & + \big(\Lambda_t + C_t\big) K_t \big(\Lambda_t + C_t\big)^T\Big)^{-1} \Big(A_t K_t \big(\Lambda_t + C_t\big)^T + M_t(\Sigma_t) K_{\hat{I}_t}\Big)^T, \\ & K_t \succeq 0, \quad t = 1, \ldots, n, \quad K_1 = 0. \end{aligned} \tag{2.2.114}$$

(c) The characterization of the $n-$*FTFI capacity,* $C_n^{fb}(\kappa)$ *of part (a) is*

$$C_n^{fb}(\kappa) = \sup_{\big(\Lambda_t, K_{Z_t}\big), t=1,\ldots,n: \; \frac{1}{n}\mathbf{E}\big\{\sum_{t=1}^n \big(X_t\big)^2\big\} \leq \kappa} \frac{1}{2}\sum_{t=1}^n \log \frac{K_{Y_t|Y^{t-1}}}{K_{V_t|V^{t-1}}} \tag{2.2.115}$$

$$= \sup_{\big(\Lambda_t, K_{Z_t}\big), t=1,\ldots,n: \; \frac{1}{n}\sum_{t=1}^n \big(\Lambda_t K_t \Lambda_t^T + K_{Z_t}\big) \leq \kappa} \frac{1}{2}\sum_{t=1}^n \log\Bigg(\frac{\Big(\Lambda_t + C_t\Big) K_t \Big(\Lambda_t + C_t\Big)^T + K_{\hat{I}_t} + K_{Z_t}}{K_{\hat{I}_t}}\Bigg). \tag{2.2.116}$$

Proof. See Appendix 7.3. □

Remark 2.2.1. *On the characterization of* $n-$*FTFI capacity of Theorem 2.2.1*
The characterization of $n-$*FTFI capacity* $C_n^{fb}(\kappa)$ *given by (2.2.116), involves the generalized matrix DRE* K_t *which is also a functional of the generalized matrix DRE* Σ_t *of the error covariance*

of the state S^n from the noise output V^n. This feature is not part of the analysis in [2], because the problems treated by the author are fundamentally different from the Cover and Pombra formulation.

In the next corollary we apply Theorem 2.2.1 to obtain the characterization of $n-$FTFI capacity $C_n^{fb}(\kappa)$ of the AGN channel driven by the PO-SS$(a_t, c_t, b_t^1, b_t^2, d_t^1, d_t^2)$ noise.

Corollary 2.2.1. *The $n-$FTFI capacity $C_n^{fb}(\kappa)$ of the AGN channel driven by the PO-SS$(a_t, c_t, b_t^1, b_t^2, d_t^1, d_t^2)$ noise is obtained from Lemma 2.1.1 and Theorem 2.2.1, by using (2.1.59).*

Proof. This is easily verified, as in Corollary 2.1.1. □

In the next corollary we apply Theorem 2.2.1 to the stable and unstable ARMA(a,c) noise, to obtain the characterization of $n-$FTFI capacity $C_n^{fb}(\kappa)$. It is then obvious that for the stable ARMA$(a,c), a \in [-1,1], c \in (-1,1)$ noise, the characterization of $C_n^{fb}(\kappa)$ involves two generalized DREs, contrary to the analysis in [2, 4–7], for the same noise model.

Corollary 2.2.2. *Characterization of $n-$FTFI Capacity $C_n^{fb}(\kappa)$ for the ARMA$(a,c), a \in (-\infty,\infty), c \in (-\infty,\infty)$*

Consider the time-varying AGN channel defined by (1.1.1) and the code of Definition 1.1.1.

(a) For the nonstationary ARMA$(a,c), a \in (-\infty,\infty), c \in (-\infty,\infty)$ noise of Example 2.1.2.(a), the characterization of the $n-$FTFI capacity, $C_n^{fb}(\kappa)$ is

$$C_n^{fb}(\kappa) = \sup_{\left(\Lambda_t, K_{Z_t}\right), t=1,\dots,n: \ \frac{1}{n}\mathbf{E}\left\{\sum_{t=1}^n \left(X_t\right)^2\right\} \le \kappa} \frac{1}{2}\sum_{t=1}^n \log\Big(\frac{\left(\Lambda_t + c - a\right)^2 K_t + K_{\hat{I}_t} + K_{Z_t}}{K_{\hat{I}_t}}\Big) \quad (2.2.117)$$

subject to the constraints

$$\begin{aligned} K_{t+1} =& \left(c\right)^2 K_t + \left(M_t(\Sigma_t)\right)^2 K_{\hat{I}_t} - \Big(cK_t\left(\Lambda_t + c - a\right) + M_t(\Sigma_t) K_{\hat{I}_t}\Big)^2 \\ & \cdot \Big(K_{\hat{I}_t} + K_{Z_t} + \left(\Lambda_t + c - a\right)^2 K_t\Big)^{-1}, \quad K_1 = 0, \quad t = 1, \dots, n, \end{aligned} \quad (2.2.118)$$

$$K_{Z_t} \ge 0, \quad K_t \ge 0, \quad c \ne a, \quad K_W > 0, \quad t = 1, \dots, n \quad (2.2.119)$$

and where

$$M_t(\Sigma_t) \stackrel{\triangle}{=} \Big(c\Sigma_t\big(c-a\big)+K_W\Big)\Big(K_W+\big(c-a\big)^2\Sigma_t\Big)^{-1}, \tag{2.2.120}$$

$$K_{\hat{I}_t} = \big(c-a\big)^2\Sigma_t + K_W, \quad t=1,\ldots,n, \tag{2.2.121}$$

$$\Sigma_{t+1} = \big(c\big)^2\Sigma_t + K_W - \Big(c\Sigma_t\big(c-a\big)+K_W\Big)^2\Big(K_W+\big(c-a\big)^2\Sigma_t\Big)^{-1}, \quad t=1,\ldots,n, \tag{2.2.122}$$

$$\Sigma_1 = K_{S_1} = \frac{\big(c_0\big)^2K_{S_0}+\big(a_0\big)^2K_{W_0}}{\Big(c_0-a_0\Big)^2}. \tag{2.2.123}$$

The optimal jointly Gaussian process (X^n, Y^n) is obtained from Theorem 2.2.1.(b), by invoking,

$$A_t \longmapsto c, \quad C_t \longmapsto c-a, \quad B_t \longmapsto 1, \quad N_t \longmapsto 1, \quad t=1,2,\ldots,n. \tag{2.2.124}$$

(b) For the nonstationary $AR(c), c \in (-\infty,\infty)$ noise of Example 2.1.2.(c), the characterization of the $n-$FTFI capacity $C_n^{fb}(\kappa)$ is obtained from part (a) by setting $a=0$, i.e.,

$$C_n^{fb}(\kappa) = \sup_{\big(\Lambda_t, K_{Z_t}\big), t=1,\ldots,n:\ \frac{1}{n}\mathbf{E}\big\{\sum_{t=1}^n \big(X_t\big)^2\big\}\le\kappa} \frac{1}{2}\sum_{t=1}^{n}\log\Big(\frac{\big(\Lambda_t+c\big)^2K_t+\big(c\big)^2\Sigma_t+K_W+K_{Z_t}}{\big(c\big)^2\Sigma_t+K_W}\Big) \tag{2.2.125}$$

subject to the constraints K_t, Σ_t are the nonnegative solutions of the generalized RDEs:

$$\begin{aligned} K_{t+1} =& \big(c\big)^2K_t + \big(c\big)^2\Sigma_t + K_W - \Big(cK_t\big(\Lambda_t+c\big)+\big(c\big)^2\Sigma_t+K_W\Big)^2 \\ &.\Big(\big(c\big)^2\Sigma_t+K_W+K_{Z_t}+\big(\Lambda_t+c\big)^2K_t\Big)^{-1}, \quad K_1=0, \quad t=1,\ldots,n, \end{aligned} \tag{2.2.126}$$

$$\begin{aligned} \Sigma_{t+1} =& \big(c\big)^2\Sigma_t + K_W - \Big(\big(c\big)^2\Sigma_t+K_W\Big)^2\Big(K_W+\big(c\big)^2\Sigma_t\Big)^{-1}, \\ &\Sigma_1 = K_{S_1} = K_{S_0} \ge 0, \quad t=1,\ldots,n. \end{aligned} \tag{2.2.127}$$

Proof. (a) The first part follows directly from Theorem 2.2.1, by using (2.2.124). □

Chapter 3

Feedback Capacity with Known Initial State

In this chapter we derive equivalent sequential characterizations, for

i) $C_n^{fb}(\kappa,s)$ defined by (1.1.3), as a degenerated case of $C_n^{fb}(\kappa)$, and

ii) $C_n^{fb,S}(\kappa,s)$ defined by (1.1.4) of Case II) formulation, as a degenerated case of $C_n^{fb}(\kappa)$.
We organize the presentation of the material as follows:

1) In Section 3.1, we follow a similar process of Section 2.1 as in Chapter 2, which gives an equivalent sequential characterization of the Cover and Pombra characterization $C_n^{fb}(\kappa,s)$, i.e., of (1.2.18), (1.2.19). We utilize the Definition code 1.1.3 and our derivation is similar; we apply a linear transformation to the Cover and Pombra Gaussian optimal channel input X^n (1.2.20), to represent X_t, by a linear function of (V^{t-1},Y^{t-1},s) or equivalently (X^{t-1},Y^{t-1},s) and an orthogonal Gaussian innovations process Z_t, which is independent of $(Z^{t-1},X^{t-1},V^{t-1},Y^{t-1},s)$ for $t=1,\ldots,n$.
We apply Theorem 3.1.1 to the time-varying PO-SS$(a_t,c_t,b_t^1,b_t^2,d_t^1,d_t^2)$ noise (see Example 2.1.1), and to the nonstationary autoregressive moving average ARMA$(a,c), a\in(-\infty,\infty)$
$c\in(-\infty,\infty)$ noise, and to the stationary ARMA$(a,c), a\in(-1,1), c\in(-1,1)$ noise (see Example 2.1.2), which is found in many references, such as, [2]. However, our characterizations of $n-$FTFI capacity are fundamentally different from past literature, because these depend on whether we consider $C_n^{fb}(\kappa), C_n^{fb}(\kappa,s)$ or $C_n^{fb,S}(\kappa,s)$.

2) Section 3.2. The main result is Corollary 3.2.1, which gives the sequential characterization of $n-$FTFI capacity for time-varying AGN channel (1.1.1) driven by the PO-SS realization of Definition 1.1.2, for the code of Definition 1.1.3. Our derivation is based on identifying a *finite-*

dimensional sufficient statistic to express X_t as a functional of the sufficient statistic, instead of (V^{t-1}, Y^{t-1}) or (X^{t-1}, Y^{t-1}), and an orthogonal Gaussian innovations process. This characterization further simplifies the sequential characterization of $C_n^{fb}(\kappa)$ given in Theorem 3.1.1 (i.e., the equivalent of (1.2.19)).

In Corollary 3.2.3 we present the application of Corollary 3.2.1 to the ARMA$(a,c), a \in (-\infty,\infty), c \in (-\infty,\infty)$ noise of Example 2.1.2, and show that the $n-$FTFI capacity is expressed in terms of solutions to two DREs.

From Corollary 3.2.3, we conclude following:

(i) Neither the time-domain characterization [2, Theorem 6.1] (see [2, Theorem 5.3]) nor the frequency domain characterization [2, Theorem 4.1], correspond to the Cover and Pombra characterization of feedback capacity (when the limit and maximization operations are interchanged) of the nonstationary and stationary ARMA(a,c) noise of Example 2.1.2.

(ii) For the characterizations given in [2, Theorem 6.1 and Theorem 4.1], to be correct, it is necessary that Conditions 1 and 2 hold.

3) Section 3.3. The main result is Proposition 3.3.1, which further clarifies that the formulation of [1] and the formulation that let to [2, Theorem 6.1], are based on Case II) formulation.

3.1 Sequential Characterization of $n-$FTFI Capacity for Case II) Formulation

In this section we consider Case II) formulation, and we derive the characterization of feedback capacity, $C_n^{fb,S}(\kappa,s)$, of the AGN channel (1.1.1) driven by a noise V^n of Definition 1.1.2, i.e., for the code of Definition 1.1.3, $(s, 2^{nR}, n)$, $n = 1,2,\ldots$, when Conditions 1 and 2 of Section 1.1 hold.

Definition 3.1.1. *AGN channels driven by noise with invertible PO-SS realizations*

The PO-SS realization of the noise of Definition 1.1.2 is called invertible if it satisfies the condition:

(A1) Given the initial state $S_1 = S_1^s = s$, the noise V^{t-1} uniquely specifies the state S^t, for $t = 1,\ldots,n$, and vice versa.

Corollary 3.1.1. *Characterization of $n-$FTFI Capacity for Case II) formulation*

Consider the AGN channel (1.1.1) driven by a noise V^n of Definition 3.1.1, and the code of Definition 1.1.3, $(s, 2^{nR}, n)$, $n = 1,2,\ldots$, that is, Conditions 1 and 2 of Section 1.1 hold.

Define the n−FTFI Capacity for a fixed initial state $S_1 = S_1^s = s$*, by*

$$C_n^{fb,S}(\kappa,s) = \sup_{\mathscr{P}^s_{[0,n]}(\kappa)} H^P(Y^n|s) - H(V^n|s) \tag{3.1.1}$$

where the set $\mathscr{P}^s_{[0,n]}(\kappa)$ *is defined by*

$$\mathscr{P}^s_{[0,n]}(\kappa) \triangleq \Big\{P_t(dx_t|x^{t-1},y^{t-1},s), t=1,\dots,n : \frac{1}{n}\mathbf{E}_s^P\Big(\sum_{t=1}^n (X_t)^2\Big) \leq \kappa\Big\}. \tag{3.1.2}$$

and where $\mathbf{E}_s^P$ *means* $S_1 = S_1^s = s$ *is fixed, and the joint distribution depends on the elements of* $\mathscr{P}^s_{[0,n]}(\kappa)$.

Then the following hold.

(a) The n−FTFI capacity, for a fixed $S_1 = s$ *is characterized by*

$$C_n^{fb,S}(\kappa,s) = \sup_{\overline{\mathscr{P}}^{s,M}_{[0,n]}(\kappa)} \sum_{t=1}^n H^{\overline{P}^M}(Y_t|Y^{t-1},s) - \sum_{t=1}^n H(V_t|V^{t-1},s) \tag{3.1.3}$$

where the $\overline{\mathscr{P}}^M_{[0,n]}(\kappa)$ *is defined by*

$$\overline{\mathscr{P}}^{s,M}_{[0,n]}(\kappa) \triangleq \Big\{\overline{P}_t^M(dx_t|s_t,y^{t-1},s), t=1,\dots,n : \frac{1}{n+1}\mathbf{E}_s^{\overline{P}^M}\Big(\sum_{t=1}^n (X_t)^2\Big) \leq \kappa\Big\} \tag{3.1.4}$$

and where (1.1.10) is respected, $\overline{P}_t^M(dx_t|s_t,y^{t-1},s)$*, is conditionally Gaussian, with linear conditional mean and nonrandom conditional covariance, given by*[1]

$$\mathbf{E}\Big\{X_t\Big|S_t^s,Y^{t-1},S_1^s=s\Big\} = \begin{cases} \Lambda_t\Big(S_t^s - \mathbf{E}\Big\{S_t^s\Big|Y^{t-1},S_1^s=s\Big\}\Big) & \text{for} \quad t=2,\dots,n \\ 0, & \text{for} \quad t=1, \end{cases} \tag{3.1.5}$$

$$K_{X_t|S_t^s,Y^{t-1},S_1^s=s} \triangleq cov\big(X_t,X_t\Big|S_t^s,Y^{t-1},S_1^s=s\big) = K_{Z_t} \succeq 0, \quad t=1,\dots,n. \tag{3.1.6}$$

and $H^{\overline{P}}(Y_t|Y^{t-1},s)$ *is evaluated with respect to the probability distribution* $\mathbf{P}_t^{\overline{P}^M}(dy_t|y^{t-1},s)$*, defined by*

$$\mathbf{P}_t^{\overline{P}^M}(dy_t|y^{t-1},s) = \int \mathbf{P}_t(dy_t|x_t,s_t)\,\mathbf{P}_t^{\overline{P}^M}(dx_t|s_t,y^{t-1},s)\,\mathbf{P}_t^{\overline{P}^M}(ds_t|y^{t-1},s), \quad t=1,\dots,n. \tag{3.1.7}$$

(b) Define the conditional means and conditional covariance for a fixed $S_1 = s$*, by*

$$K_t^s \triangleq cov\big(S_t^s,S_t^s\Big|Y^{t-1},S_1^s=s\big) = \mathbf{E}^{\overline{P}^M}\Big\{\Big(S_t^s-\hat{S}_t^s\Big)\Big(S_t^s-\hat{S}_t^s\Big)^T\Big\}, \tag{3.1.8}$$

$$\hat{S}_t^s \triangleq \mathbf{E}^{\overline{P}^M}\Big\{S_t^s\Big|Y^{t-1},S_1^s=s\Big\}, \quad t=2,\dots,n, \quad K_1^s \triangleq cov(S_1^s,S_1^s|S_1^s=s)=0, \quad \hat{S}_1^s \triangleq s. \tag{3.1.9}$$

[1]The notation $S_t = S_t^s, t=2,\dots,n$ means this sequence is generated from (1.1.11), when the initial state is fixed, $S_1 = S_1^s = s$.

The optimal channel input distribution of part (a) is induced by a jointly Gaussian process process X^n, with a realization given by

$$X_t = \Lambda_t\Big(S_t^s - \widehat{S}_t^s\Big) + Z_t, \quad X_1 = Z_1, \quad t = 2,\ldots,n, \tag{3.1.10}$$

$$Z_t \in N(0, K_{Z_t}) \;\; \textit{independent of} \;\; (S_1, X^{t-1}, V^{t-1}, Y^{t-1}), \quad t = 1,\ldots,n, \tag{3.1.11}$$

$$Y_t = \Lambda_t\Big(S_t^s - \widehat{S}_t^s\Big) + Z_t + V_t, \quad t = 1,\ldots,n, \tag{3.1.12}$$

$$= \Lambda_t\Big(S_t^s - \widehat{S}_t^s\Big) + C_t S_t^s + N_t W_t + Z_t, \tag{3.1.13}$$

$$\frac{1}{n}\mathbf{E}_s^{\overline{P}^M}\Big\{\sum_{t=1}^n (X_t)^2\Big\} = \frac{1}{n}\sum_{t=1}^n \Big(\Lambda_t K_t^s \Lambda_t^T + K_{Z_t}\Big) \le \kappa \tag{3.1.14}$$

where Λ_t is nonrandom.

The conditional means and conditional covariance $\widehat{S}_t^s$ and K_t^s are given by the generalized Kalman-filter, as follows equations.

(i) $\widehat{S}_t^s$ satisfies the Kalman-filter recursion

$$\widehat{S}_{t+1}^s = A_t \widehat{S}_t^s + F_t(K_t^s) I_t^s, \quad \widehat{S}_1 = s, \tag{3.1.15}$$

$$F_t(K_t^s) \stackrel{\triangle}{=} \Big(A_t K_t^s (\Lambda_t + C_t)^T + B_t K_{W_t} N_t^T\Big)\Big\{N_t K_{W_t} N_t^T + K_{Z_t} + (\Lambda_t + C_t) K_t^s (\Lambda_t + C_t)^T\Big\}^{-1}, \tag{3.1.16}$$

$$I_t^s \stackrel{\triangle}{=} Y_t - C_t \widehat{S}_t^s = (\Lambda_t + C_t)(S_t^s - \widehat{S}_t^s) + N_t W_t + Z_t, \quad t = 1,\ldots,n, \tag{3.1.17}$$

$$I_t^s \in N(0, K_{I_t^s}), \quad t = 1,\ldots,n \;\; \textit{an orthogonal innovations process, i.e., } I_t^s \textit{ is independent of}$$
$$I_k^s, \textit{ for all } t \neq k, \textit{ and } I_t^s \textit{ is independent of } Y^{t-1}, \tag{3.1.18}$$

$$K_{Y_t|Y^{t-1},s} = K_{I_t^s} \stackrel{\triangle}{=} cov(I_t, I_t | S_1^s = s) = \Big(\Lambda_t + C_t\Big) K_t^s \Big(\Lambda_t + C_t\Big)^T + N_t K_{W_t} N_t^T + K_{Z_t}. \tag{3.1.19}$$

(ii) The error $E_t^s \stackrel{\triangle}{=} S_t^s - \widehat{S}_t^s$ satisfies the recursion

$$E_{t+1}^s = F_t^{CL}(K_t^s) E_t^s + F_t(K_t^s)\Big(N_t W_t + Z_t\Big), \quad E_1^s = S_1^s - \widehat{S}_1^s = 0, \quad t = 1,\ldots,n, \tag{3.1.20}$$

$$F_t^{CL}(K_t^s) \stackrel{\triangle}{=} A_t - F_t(K_t^s)\Big(\Lambda_t + C_t\Big). \tag{3.1.21}$$

(iii) $K_t^s = \mathbf{E}\big\{E_t^s (E_t^s)^T\big\}$ satisfies the generalized DRE

$$\begin{aligned} K_{t+1}^s = & A_t K_t A_t^T + B_t K_{W_t} B_t^T - \Big(B_t K_{W_t} N_t^T + A_t K_t^s (\Lambda_t + C_t)^T\Big)\Big\{N_t K_{W_t} N_t^T + K_{Z_t} \\ & + (\Lambda_t + C_t) K_t^s (\Lambda_t + C_t)^T\Big\}^{-1}\Big(B_t K_{W_t} N_t^T + A_t K_t^s (\Lambda_t + C_t)\Big)^T, \\ & K_t^s \succeq 0, \quad K_1^s = 0, \quad t = 1,\ldots,n. \end{aligned} \tag{3.1.22}$$

(c) The characterization of the n−FTFI capacity of part (a) is

$$
\begin{aligned}
& C_n^{fb,S}(\kappa,s) \\
&= \sup_{\left(\Lambda_t,K_{Z_t}\right),t=1,\ldots,n:\ \frac{1}{n}\mathbf{E}_s\left\{\sum_{t=1}^n \left(X_t\right)^2\right\}\leq\kappa} \sum_{t=1}^n \log \frac{K_{Y_t|Y^{t-1},s}}{K_{V_t|V^{t-1},s}} \qquad (3.1.23)\\
&= \sup_{\left(\Lambda_t,K_{Z_t}\right),t=1,\ldots,n:\ \frac{1}{n}\mathbf{E}_s\left\{\sum_{t=1}^n \left(X_t\right)^2\right\}\leq\kappa} \frac{1}{2}\sum_{t=1}^n \log\Big(\frac{\left(\Lambda_t+C_t\right)K_t^s\left(\Lambda_t+C_t\right)^T+N_tK_{W_t}N_t^T+K_{Z_t}}{N_tK_{W_t}N_t^T}\Big). \qquad (3.1.24)
\end{aligned}
$$

Proof. See Appendix 7.4. □

Remark 3.1.1. *Comments on the per unit time limit of $C_n^{fb,S}(\kappa,s)$*
(b) The asymptotic analysis of $C^{fb,o}(\kappa)$ and $C^{fb,o}(\kappa,s)$ of Chapter 4, i.e., based on Definition 4.1.1, applies naturally to Corollary 3.1.1, by considering $C^{fb,S,o}(\kappa,s)$.

In the next remark we clarify the relation of Corollary 3.1.1 and the analysis of [1] and [2].

Remark 3.1.2. *Relations of Corollary 3.1.1 and [1, 2]*
(a) The problem analyzed in [1] is precisely $C_n^{fb,S}(\kappa,s)$, when the noise is stationary and Gaussian, i.e., it corresponds to Case II) formulation. Corollary 3.1.1 is derived in [1] for the degenerate case of a time-invariant realization of the noise V^n, i.e., of Definition 3.1.1. However, the asymptotic analysis of [1, Section VI] should be read with caution, because it did not account for the necessary and/or sufficient conditions for convergence of the sequence $\Sigma_t^s, t=1,2,\ldots$ generated by the time-invariant version of the generalized DRE (3.1.22) i.e., $\lim_{n\longrightarrow\infty}\Sigma_n^s=\Sigma^\infty\succeq 0$, where $\Sigma^\infty\succeq 0$ is the unique and stabilizing solution of a corresponding generalized ARE.
(b) The problem analyzed [2] that let to [2, Theorem 6.1, C_{FB}], is the per unit time limit of $C_n^{fb,S}(\kappa,s)$, when the noise is stationary, two-sided or one-sided (asymptotically stationary) and Gaussian, i.e., it corresponds to Case II) formulation. The characterization of feedback capacity presented in [2, Theorem 6.1, C_{FB}] presupposed the following hold ((i)-(iii) are also assumed in [1, Section VI]).
(i) The feedback code is Definition 1.1.3, i.e., $(s,2^{nR},n)$.
(ii) The noise is time-invariant and stable, and the PO-SS realization of the noise is invertible, as

presented in Definition 3.1.1.

(iii) The definition of rate is $C^{fb,S,o}(\kappa,s)$*, with supremum and per unit time limit interchanged, and the supremum taken over using time-invariant channel input distributions.*

(iv) the innovations covariance of the channel input process is zero, i.e., $K_{Z_t} = K_Z = 0, \forall t$*.*

Items (i)-(iv) are confirmed from [2, Lemma 6.1] (and comments above), which is used to derive [2, Theorem 6.1, C_{FB}*].*

However, the characterization of feedback capacity in [2, Theorem 6.1, C_{FB}*] should be read with caution, because the stabilizability condition is violated, because Theorem 5.2.1.(1) is not accounted for. When Theorem 5.2.1.(1) is accounted for, then the only unique and stabilizing solution of the generalized ARE presented in [2, Theorem 6.1,* C_{FB}*], is the zero solution, which then implise* $C_{FB} = 0$*. That is, the rate as defined in [2] exists if and only if* $K_Z > 0$*.*

The above technical matters are discussed extensively in [15], for the case of the $AR(c), c \in (-\infty,\infty)$*, where it is also shown that feedback does not increase capacity for* $c \in (-1,1)$*, i.e., for the stationary* $AR(c)$ *noise.*

Notation 3.1.1. *For the feedback code of Definition 1.1.3, with initial state* $S_1 = s$*, known to the encoder and the decoder, all the sets from Section 2.1* $\mathscr{P}_{[0,n]}(\kappa), \overline{\mathscr{P}}_{[0,n]}(\kappa), \mathscr{E}_{[0,n]}, \overline{\mathscr{E}}_{[0,n]}$ *are replaced by* $\mathscr{P}^s_{[0,n]}(\kappa), \overline{\mathscr{P}}^s_{[0,n]}(\kappa), \mathscr{E}^s_{[0,n]}, \overline{\mathscr{E}}^s_{[0,n]}$*, to indicate the distributions and codes are* $\overline{P}_t(dx_t|v^{t-1},y^{t-1},s)$*,* $t = 1,\ldots, x_1 = \overline{e}_1(w,s), x_2 = \overline{e}_2(w,v_1,y_1,s)\ldots, x_n = \overline{e}_n(w,v^{n-1},y^{n-1},s)$*, etc. i.e., these depend on* s*.*

Theorem 3.1.1. *Information structures of maximizing distributions for AGN Channels*

Consider the time-varying AGN channel defined by (1.1.1), driven by a noise with the PO-SS realization of Definition 1.1.2, and the code of Definition 1.1.3, with initial state $S_1 = S_1^s = s$ *fixed. Then the following hold.*

(a) The $n-$*FTFI capacity* $C_n^{fb}(\kappa,s)$ *is given by*

$$C_n^{fb}(\kappa,s) \triangleq \sup_{\frac{1}{n}\mathbf{E}_s\left\{\sum_{t=1}^n \left(X_t\right)^2 \Big| S_1\right\} \leq \kappa} H^{\overline{P}}(Y^n|s) - H(V^n|s). \tag{3.1.25}$$

$$X_t = \Gamma^0 s + \sum_{j=1}^{t-1} \Gamma^1_{t,j} V_{j-1} + \sum_{j=1}^{t-1} \Gamma^2_{t,j} Y_j + Z_t, \quad t = 1,\ldots,n. \tag{3.1.26}$$

where the supremum is over all $(\Gamma^0, \Gamma^1_{t,j}, \Gamma^2_{t,j}, K_{Z_t}), j = 1,\ldots,t-1, t = 1,\ldots,n$ *of the realization of* X^n*, that induces the distribution* $\overline{P}_t(dx_t|v^{t-1},y^{t-1},s), t = 1,\ldots,n$*, and all statements of Theorem 3.1.1 and Lemma 2.1.1 hold, with the conditional distribitions, expectations, and entropies*

replaced by the corresponding expressions with fixed $S_1 = s$.
(b) A necessary condition for Condition 2 of Section 1.1 to hold is
(i) N_tW_t uniquely defines $C_{t+1}B_tW_t, \forall t$.
Moreover, if (i) holds then the entropy $H(V^n|s)$ of part (a) is given by

$$H(V^n|s) = \frac{1}{2}\sum_{t=1}^{n} \log\left(2\pi e N_t K_{W_t} N_t^T\right). \tag{3.1.27}$$

Proof. See Appendix 7.5. □

Remark 3.1.3. *It is easy to verify that for the code of Definition 1.1.3 that assumes knowledge of the initial state $S_1 = s$, then $C_n^{fb}(\kappa, s)$ is directly obtained from Theorem 3.1.1, as a degenerate case (an independent derivation is easily produced following the derivation of Corollary 3.1.1, with slight variations).*

By utilizing Theorem 3.1.1 we can derive the converse coding theorems stated below for the feedback codes of Definition 1.1.3.

Theorem 3.1.2. *Converse coding theorems for code of Definition 1.1.3*
Consider the AGN channel (1.1.1).
(a) Any achievable rate R for the code of Definition 1.1.3 (with initial state $S_1 = s$) satisfies

$$R \leq C^{fb}(\kappa, s) \stackrel{\triangle}{=} \lim_{n\longrightarrow\infty} \frac{1}{n} C_n^{fb}(\kappa, s), \tag{3.1.28}$$

$$C_n^{fb}(\kappa, s) = \sup_{\overline{P}_t(dx_t|v^{t-1}, y^{t-1}, s), t=1,\ldots,n:\ \frac{1}{n}\mathbf{E}_s\left\{\sum_{t=1}^n \left(X_t\right)^2 \middle| S_1\right\} \leq \kappa} \sum_{t=1}^{n} H^{\overline{P}}(Y_t|Y^{t-1}, s) - H(V^n|s). \tag{3.1.29}$$

where $\mathbf{E}_s\{\cdot\}$ means the expectation is for a fixed $S_1 = s$, provided the supremum exists and the limit exists, and where the right hand side of (3.1.29) is obtained from Theorem 3.1.1, by replacing all conditional distributions, entropies, etc, for fixed initial state $S_1 = s$ (see Notation 3.1.1).

Proof. Follows from standard arguments, using Fano's inequality (see also [3]) and Theorem 3.1.1. □

Proof. From (2.1.52) and (2.1.48), then follows (2.1.49), from the entropy formula of Gaussian RVs. □

From Lemma 2.1.1 in Chapter 2, follows directly the next corollary of the entropy $H(V^n|s)$, when $S_1 = s$ is fixed.

Corollary 3.1.2. *Conditional entropy $H(V^n|s), S_1 = s$ of the PO-SS noise realization.*

Consider the PO-SS realization of V^n of Definition 1.1.2, for fixed $S_1 = s$, and denote the state process generated by recursion (1.1.11), by[2]*, $S_t = S_t^s, t = 2, \ldots, n, S_1 = S_1^s = s$. Replace the conditional covariance and conditional mean (2.1.36) and (2.1.37), by*

$$\Sigma_t^s \triangleq cov\big(S_t^s, S_t^s \big| V^{t-1}, S_1^s\big) = \mathbf{E}\Big\{\Big(S_t^s - \hat{S}_t^s\Big)\Big(S_t^s - \hat{S}_t^s\Big)^T \Big| V^{t-1}, S_1^s\Big\}, \tag{3.1.30}$$

$$\hat{S}_t^s \triangleq \mathbf{E}\Big\{S_t^s \Big| V^{t-1}, S_1^s\Big\}, \quad t = 2, \ldots, n, \qquad S_1^s = s, \quad \hat{S}_1^s \triangleq s, \quad \Sigma_1^s \triangleq cov\big(S_1^s, S_1^s | S_1^s\big) = 0. \tag{3.1.31}$$

Then all statements of Lemma 2.1.1 hold, with the changes,

$$\Sigma_t \longmapsto \Sigma_t^s, \quad \Sigma_1^s = 0, \quad \mathbf{P}_{V_t|V^{t-1}} \longmapsto \mathbf{P}_{V_t|V^{t-1}, S_1^s}, \quad \hat{S}_t \longmapsto \hat{S}_t^s, \quad \hat{S}_1^s = s, \quad etc,\ t = 1, \ldots, n. \tag{3.1.32}$$

In particular, the conditional entropy of V^n conditioned on $S_1 = S_1^s = s$, is given by

$$H(V^n|s) = \frac{1}{2}\sum_{t=1}^{n} \log\Big(2\pi e\Big[C_t \Sigma_t^s C_t^T + N_t K_{W_t} N_t^T\Big]\Big) \tag{3.1.33}$$

where $\Sigma_t^s, t = 2, \ldots, n$ satisfies the generalized DRE (2.1.46) with initial condition $\Sigma_1^s = 0$.

Proof. Follows directly from Lemma 2.1.1 and (3.1.30),(3.1.31) . □

Next we introduce an example of a PO-SS realization of the noise that we often use.

Remark 3.1.4. *Consider the PO-SS$(a_t, c_t, b_t^1, b_t^2, d_t^1, d_t^2)$ noise of Example 2.1.1. Then the following hold.*

(a) Consider the code of Definition 1.1.3, i.e., with a fixed initial state $S_1 = S_1^s = s$. By Corollary 3.1.2 using (2.1.59), then $H(V^n|s)$ is computed from Lemma 2.1.1, with $\Sigma_1 = \Sigma_1^s = 0$, and (3.1.33) reduces to

$$H(V^n|s) = \frac{1}{2}\sum_{t=1}^{n} \log\Big(2\pi e\Big[\big(c_t\big)^2 \Sigma_t^s + n_t \circ n_t\Big]\Big) \tag{3.1.34}$$

where Σ_t^s is the solution of (2.1.46) with $\Sigma_1 = \Sigma_1^s = 0$ (using (2.1.59)).

Remark 3.1.5. *ARMA(a, c) noise of Example 2.1.2*

(a) Consider the nonstationary AR$(a, c), a \in (-\infty, \infty), c \in (-\infty, \infty)$ of Example 2.1.2.(a).

(i) Assume the code of Definition 1.1.3, with initial state $V_0 = v_0$ known to the encoder. By (2.1.64),

$$S_1 = S_1^{v_0} = \frac{cv_0 - aW_0}{c - a}, \quad V_0 = v_0 \tag{3.1.35}$$

[2] We often use the notation $S_t = S_t^s$ to emphasize that the S_t process is generated for $S_1 = S_1^s = s$ fixed.

and hence knowledge of $V_0 = v_0$ at the encoder does not determine $S_1^{v_0}$, because for this to hold the encoder requires knowledge of W_0. It then follows that $H(V^n|v_0)$ is computed Corollary 3.1.2,

$$\Sigma_1 = \Sigma_1^{v_0} = \frac{(a)^2 K_{W_0}}{\left(c-a\right)^2}, \quad \text{and (3.1.33) reduces to} \tag{3.1.36}$$

$$H(V^n|v_0) = \frac{1}{2}\sum_{t=1}^{n} \log\left(2\pi e\left[\left(c\right)^2 \Sigma_t^{v_0} + K_W\right]\right) \tag{3.1.37}$$

where $\Sigma_t^{v_0}$ is the solution of (2.1.46) with initial data $\Sigma_1 = K_{S_1} = \Sigma_1^{v_0} = \frac{\left(a\right)^2 K_{W_0}}{\left(c-a\right)^2}, K_{W_0} \geq 0.$

(ii) Assume the code of Definition 1.1.3, with initial state $S_1 = s$ or $(V_0, W_0) = (v_0, w_0)$ are known to the encoder. Then by Corollary 3.1.2,

$$H(V^n|v_0, w_0) = \frac{1}{2}\sum_{t=1}^{n} \log\left(2\pi e K_W\right). \tag{3.1.38}$$

By (2.1.64),, $S_1 \triangleq \frac{cV_0 - aW_0}{c-a}$, and a necessary condition for Conditions 1 of Section 1.1 to hold, i.e., $S_1 = s$ is known the the encoder and the decoder is: both $(V_0, W_0) = (v_0, w_0)$ are known to the encoder and the decoder.

(b) The statements of parts (a), (b) also hold for the double-sided and the one-sided wide-sense stationary $AR(a,c), a \in [-1,1], c \in (-1,1)$ of Example 2.1.2.(b), (c).

(c) Case II) formulation discussed in Section 1.1, requires Conditions 1 and 2 to hold. For any of the $AR(a,c)$ noise models, then Conditions 1 and 2 hold if and only if $S_1 = s_1$ or $(V_0, W_0) = (v_0, w_0)$ are known to the encoder. Clearly, the values of $H(V^n)$ under Case I) formulation is fundamentally different from the value of $H(V^n|s), S_1 = s$ under Case II) formulation. Consequently, in general, $C_n^{fb}(\kappa)$ given by (2.1.30) is fundamentally different from $C_n^{fb}(\kappa, s)$ i.e., that corresponds to a fixed initial state $S_1 = s$, known to the encoder and the decoder, and to the channel input distribution.

(d) From parts (a)-(c) it is clear that Kim's characterization of feedback capacity for the stationary $ARMA(a,c), a \in [-1,1], c \in (-1,1)$ that uses [2, Theorem 6.1, C_{FB}] (which is derived based on [2, Lemmas 6.1]) presupposed that the encoder and the decoder assumed knowledge of $S_1 = S_1^s = s$.

Remark 3.1.6. *Consider the stationary double-sided or one-sided $ARMA(a,c), a \in [-1,1], c \in (-1,1)$ of Example 2.1.2. From in Proposition 2.1.1, and in particular the initial data $\hat{S}_1, \Sigma_1$, it is clear that even if the encoder and the decoder know the initial state V_0, then $H(V^n|v_0) \neq \frac{1}{2}\sum_{t=1}^{n} \log\left(2\pi e K_W\right)$. In this case, the value of $C_n^{fb}(\kappa, v_0)$ defined by (3.1.29) is fundamentally different from the formulation in [1] and [2] that let to the characterization of feedback capacity [2, Theorem 6.1].*

In the next corollary we further clarify the difference between Case I) formulation and Case II) formulation, by stating the analogue of Theorem 3.1.1 for the code of Definition 1.1.3, i.e., when $S_1 = S_1^s = s$ is fixed.

In the next remark we illustrate that $H(V^n|s)$ given by (3.1.27) follows directly from Lemma 2.1.1, by fixing $S_1 = S_1^s = s$, and assuming $N_t W_t$ uniquely defines $C_{t+1} B_t W_t, \forall t$.

Remark 3.1.7. *The $n-$FTFI capacity for code of Definition 1.1.1 versus code of Definition 1.1.3.*

Consider the generalized Kalman-filter of the PO-SS noise realization, of Lemma 2.1.1, and assume the initial state of the noise S_1 is known, i.e., $S_1 = S_1^s = s$ or $S_1 = S_1^s = s = 0$, and $N_t W_t$ uniquely defined $C_{t+1} B_t W_t, \forall t$. Then all statements of Lemma 2.1.1 hold, by replacing $(\Sigma_t, \hat{S}_t)$ by $(\Sigma_t^s, \hat{S}_t^s)$ for $t = 1, 2, \ldots,$. Since, Σ_t^s satisfies the generalized DRE (2.1.46) with initial condition $\Sigma_1^s = 0$, then it is easy to deduce that $\Sigma_t^s = 0$, for $t = 1, 2, \ldots, n$ is a solution. Substituting $\Sigma_t^s = 0, t = 1, 2, \ldots, n$ in (2.1.49) we obtain (3.1.27), as expected.

3.2 A Sufficient Statistic Approach

In this section, we wish to identify a *sufficient statistic* for the input process X_t, given by (3.1.26), where the initial state is known and fixed to the encoder and the decoder $S_1 = s$, which summarizes the information contained in (V^{t-1}, Y^{t-1}, S_1). On the other hand, for a code that assumes knowledge of the initial state and the state of the noise, and Conditions 1 and 2 hold, the characterization of the $n-$FTFI capacity is expressed as a functional of one generalized DRE (see [1]).

Corollary 3.2.1. *Equivalent characterization of $n-$FTFI Capacity $C_n^{fb}(\kappa, s)$ for PO-SS Noise realizations Consider the time-varying AGN channel defined by (1.1.1), driven by a noise with the PO-SS realization of Definition 1.1.2, and the code of Definition 1.1.3, with initial state $S_1 = S_1^s = s$ fixed, and replace (2.2.86), (2.2.87) by*

$$K_t = K_t^s \stackrel{\triangle}{=} cov\Big(\hat{S}_t^s, \hat{S}_t^s \Big| Y^{t-1}, S_1 = s\Big) = \mathbf{E}\Big\{\Big(\hat{S}_t^s - \widehat{\hat{S}_t^s}\Big)\Big(\hat{S}_t^s - \widehat{\hat{S}_t^s}\Big)^T\Big\}, \tag{3.2.39}$$

$$\widehat{\hat{S}}_t = \widehat{\hat{S}_t^s} \stackrel{\triangle}{=} \mathbf{E}\Big\{\hat{S}_t^s \Big| Y^{t-1}, S_1 = s\Big\}, \quad t = 2, \ldots, n, \quad \widehat{\hat{S}}_1 = \widehat{\hat{S}_1^s} \stackrel{\triangle}{=} s, \quad K_1 = K_1^s = 0. \tag{3.2.40}$$

Then

$$C_n^{fb}(\kappa,s) = \sup_{\mathscr{P}_{[0,n]}^{\hat{S}^s}(\kappa)} \sum_{t=1}^{n} H(Y_t|Y^{t-1},s) - H(V^n|s), \tag{3.2.41}$$

$$\mathscr{P}_{[0,n]}^{\hat{S}^s}(\kappa) \triangleq \Big\{ \mathbf{P}_t(dx_t|\hat{s}_t^s, y^{t-1}, s), t = 1,\ldots,n : \frac{1}{n}\mathbf{E}\Big(\sum_{t=1}^{n} \big(X_t\big)^2 \Big| S_1^s = s \Big) \le \kappa \Big\} \tag{3.2.42}$$

where $H(V^n|s)$ is given by Corollary 3.1.2, and the statements of Theorem 2.2.1 hold with the above changes, and all conditional entropies, distributions, expectations, etc, defined for fixed $S_1 = S_1^s = s$.

Proof. It is easily verified from the derivation of Theorem 2.2.1, by fixing $S_1 = S_1^s = s$. □

In the next corollary we apply Theorem 2.2.1 to obtain the characterization of $n-$FTFI capacity $C_n^{fb}(\kappa,s)$ of the AGN channel driven by the PO-SS$(a_t,c_t,b_t^1,b_t^2,d_t^1,d_t^2)$ noise.

Corollary 3.2.2. *In the next corollary we apply Theorem 2.2.1 to the stable and unstable ARMA(a,c) noise, to obtain the characterization of $n-$FTFI capacity $C_n^{fb}(\kappa,s)$. It is then obvious that for the stable ARMA$(a,c), a\in[-1,1], c\in(-1,1)$ noise, the characterization of $C_n^{fb}(\kappa,s)$ involves one generalized DRE, contrary to the analysis in [2, 4–7], for the same noise model.*

Corollary 3.2.3. *Characterization of $n-$FTFI Capacity $C_n^{fb}(\kappa,s)$ for the ARMA$(a,c), a\in(-\infty,\infty), c\in(-\infty,\infty)$*

Consider the time-varying AGN channel defined by (1.1.1) and the code of Definition 1.1.3.
(a) For the nonstationary ARMA$(a,c), a\in(-\infty,\infty), c\in(-\infty,\infty)$ noise of Example 2.1.2.(a), the characterization of the $n-$FTFI capacity, $C_n^{fb}(\kappa,s)$ is,
The optimal jointly Gaussian process (X^n,Y^n) is obtained from Theorem 2.2.1.(b), by invoking,

$$A_t \longmapsto c, \;\; C_t \longmapsto c-a, \;\; B_t \longmapsto 1, \;\; N_t \longmapsto 1, \;\; t = 1,2,\ldots,n. \tag{3.2.43}$$

If $\Sigma_1 = 0$ or the initial state is fixed, $S_1 = S_1^s = s$, then

$$\Sigma_t = \Sigma_t^s = 0, \;\; K_{\hat{I}_t} = K_W, \;\; M_t(\Sigma_t) = M_t(\Sigma_t^s) = 1, \;\; t = 1,2,\ldots \tag{3.2.44}$$

$$C_n^{fb}(\kappa,s) = \sup_{\big(\Lambda_t, K_{Z_t}\big), t=1,\ldots,n: \;\; \frac{1}{n}\sum_{t=1}^{n}\Big(\big(\Lambda_t\big)^2 K_t^s + K_{Z_t}\Big) \le \kappa} \frac{1}{2}\sum_{t=1}^{n} \log\Big(\frac{\big(\Lambda_t + c - a\big)^2 K_t^s + K_W + K_{Z_t}}{K_W} \Big) \tag{3.2.45}$$

subject to the constraints

$$K^s_{t+1} = (c)^2 K^s_t + K_W - \Big(cK^s_t(\Lambda_t + c - a) + K_W\Big)^2 \\ .\Big(K_{Z_t} + (\Lambda_t + c - a)^2 K^s_t + K_W\Big)^{-1}, \quad K^s_1 = 0, \quad K^s_t \geq 0, \;\; K_{Z_t} \geq 0, \;\; t = 1, \dots, n. \tag{3.2.46}$$

(c) For the nonstationary AR$(c), c \in (-\infty, \infty)$ *noise of Example 2.1.2.(c), with* $\Sigma_1 = 0$ *or a fixed initial state* $S_1 = S^s_1 = s$*, then* $\Sigma_t = \Sigma^s_t = 0, K_{\hat{I}_t} = K_W, M_t(\Sigma_t) = M_t(\Sigma^s_t) = 1, t = 1, 2, \dots$*, and the characterization of the* $n-$*FTFI capacity is given by*

$$C^{fb}_n(\kappa, s) = \sup_{\big(\Lambda_t, K_{Z_t}\big), t=1,\dots,n:\; \frac{1}{n}\sum_{t=1}^n \Big(\big(\Lambda_t\big)^2 K^s_t + K_{Z_t}\Big) \leq \kappa} \frac{1}{2}\sum_{t=1}^n \log\Big(\frac{\Big(\Lambda_t + c\Big)^2 K^s_t + K_W + K_{Z_t}}{K_W}\Big) \tag{3.2.47}$$

subject to the constraint

$$K^s_{t+1} = (c)^2 K^s_t + K_W - \Big(cK^s_t(\Lambda_t + c) + K_W\Big)^2 \Big(K_W + K_{Z_t} + (\Lambda_t + c)^2 K^s_t\Big)^{-1}, \\ K^s_1 = 0, \; t = 1, \dots, n. \tag{3.2.48}$$

Remark 3.2.1. *By Corollary 3.2.3.(a) it is obvious that, if* $\Sigma_1 = 0$*, i.e.,* $K_{S_0} = K_{W_0} = 0$*, which means* $S_1 = S^s_1 = s$ *is fixed, and hence* $(V_0, W_0) = (v_0, w_0)$ *is fixed (and known to the encoder and the decoder), see (2.1.64), then* $\Sigma_1 = \Sigma^s_1 = 0$*, and* $C^{fb}_n(\kappa) = C^{fb}_n(\kappa, s)$*, which depends on the initial state* $S_1 = S^s_1 = s$*. To ensure for large enough n the rate* $\frac{1}{n}C_n(\kappa, s)$ *is independent of s, it is necessary to identify conditions for convergence of solutions* $K^s_t, t = 1, 2, \dots$ *of generalized DRE (3.2.46) to a unique limit,* $\lim_{n\longrightarrow\infty} K^s_n = K^\infty \geq 0$*, that does not depend on the initial data* $K^s_1 = 0$*. We address this problem in Chapter 4.*

3.2.1 Case II) Formulation: A Degenerate of Case I) Formulation

Theorem 2.2.1 gives the $n-$FTFI capacity for Case I) formulation. However, since Case II) formulation is a special case of Case I) formulation, we expect that from Theorem 2.2.1 we can recover the characterization of the $n-$FTFI capacity for Case II) formulation, i.e., when the code is $(s, 2^{nR}, n)$, $n = 1, 2, \dots$, and Conditions 1 and 2 of Section 1.1 hold. We show this in the next corollary.

Corollary 3.2.4. *The degenerate $n-$FTFI Capacity $C_n^{fb}(\kappa)$ of Theorem 2.2.1 for Case II) formulation*

Consider the time-varying AGN channel defined by (1.1.1), driven by a noise with PO-SS realization of Definition 1.1.2, and suppose the following hold.

1) The code is $(s, 2^{nR}, n)$, $n = 1,2,\ldots$, and

2) Conditions 1 and 2 of Section 1.1 hold.

Then the following hold.

(a) Corollary 3.1.2 holds, i.e., all statements of Lemma 2.1.1 hold with $(\Sigma_t, \hat{S}_t)$ replaced by $(\Sigma_t^s, \hat{S}_t^s)$ as defined by (3.1.30), (3.1.31). In particular, $(\Sigma_t^s, \hat{S}_t^s) = (0, S_t^s)$ for $t = 1,2,\ldots$, and $H(V^n) = H(V^n|s)$ is given by (3.1.27).

(b) All statements of Theorem 2.2.1 hold with $(\Sigma_t, \hat{S}_t)$ replaced by $(\Sigma_t^s, \hat{S}_t^s)$, as in part (a), and $(K_t, \widehat{S}_t)$ defined by (2.2.86), (2.2.87) reduce to

$$K_t = K_t^s = cov\Big(S_t^s, S_t^s\Big|Y^{t-1}, S_1^s = s\Big), \tag{3.2.49}$$

$$\widehat{S}_t = \widehat{S}_t^s = \mathbf{E}\Big\{S_t^s\Big|Y^{t-1}, S_1^s = s\Big\},\ K_1^s = 0, \widehat{S}_1^s = s,\ t = 2,\ldots,n$$

In particular, the optimal input process X^n of Theorem 2.2.1.(c) degenerates to

$$X_t = \Lambda_t\Big(S_t^s - \widehat{S}_t^s\Big) + Z_t, \quad X_1 = Z_t, \quad t = 2,\ldots,n. \tag{3.2.50}$$

(c) The characterization of $n-$FTFI capacity, $C_n^{fb}(\kappa)$ of Theorem 2.2.1 degenerates to $C_n^{fb,S}(\kappa, s)$ defined by

$$C_n^{fb}(\kappa) = C_n^{fb,S}(\kappa, s) \stackrel{\triangle}{=} \sup_{\left(\Lambda_t, K_{Z_t}\right), t=1,\ldots,n:\ \frac{1}{n}\mathbf{E}_s\left\{\sum_{t=1}^n \left(X_t\right)^2\right\} \le \kappa} \sum_{t=1}^n \log \frac{K_{Y_t|Y^{t-1},s}}{K_{V_t|V^{t-1},s}} \tag{3.2.51}$$

$$= \sup_{\left(\Lambda_t, K_{Z_t}\right), t=1,\ldots,n:\ \frac{1}{n}\mathbf{E}\left\{\sum_{t=1}^n \left(X_t\right)^2\right\} \le \kappa} \frac{1}{2}\sum_{t=1}^n \log\Big(\frac{\Big(\Lambda_t + C_t\Big)K_t^s\Big(\Lambda_t + C_t\Big)^T + N_t K_{W_t} N_t^T + K_{Z_t}}{N_t K_{W_t} N_t^T}\Big). \tag{3.2.52}$$

$K_t = K_t^s = \mathbf{E}_s\{E_t^s(E_t^s)^T\}$ satisfies the generalized DRE

$$\begin{aligned} K_{t+1}^s =& A_t K_t A_t^T + B_t K_{W_t} B_t^T - \Big(B_t K_{W_t} N_t^T + A_t K_t^s(\Lambda_t + C_t)^T\Big)\Big\{N_t K_{W_t} N_t^T + K_{Z_t} \\ &+ (\Lambda_t + C_t)K_t^s(\Lambda_t + C_t)^T\Big\}^{-1}\Big(B_t K_{W_t} N_t^T + A_t K_t^s(\Lambda_t + C_t)\Big)^T, \\ & K_t^s \succeq 0, \quad K_1^s = 0, \quad t = 1,\ldots,n. \end{aligned} \tag{3.2.53}$$

and the optimal input process

Proof. (a) The statements about Lemma 2.1.1 follow from Remark 3.1.7. (b) The statements about Theorem 2.2.1 are easily verified by replacing all conditional expectations, distributions, etc, for a fixed initial state $S_1 = S_1^s = s$, and using part (a), i.e., $(\Sigma_t^s, \hat{S}_t^s) = (0, S_t^s)$, $t = 1, 2, \ldots$. (c) Follows from parts (a), (b). □

3.3 Comments on the Formulation of [1] and [2]

It is easily verified that Yang, Kavcic and Tatikonda [1] analyzed $C_n^{fb}(\kappa, s)$ defined by (3.1.29), under Case II) formulation. This is further discussed in the following remark.

Remark 3.3.1. *Prior literature on the time-invariant stationary noise of PSD (1.2.26)*
Yang, Kavcic and Tatikonda [1] analyzed the AGN channel driven by a stationary noise with PSD defined by (1.2.26) (see [1, Theorem 1]). The special case of (2.1.70) is found in [1, Section VI.B, Theorem 7].
The analysis in [1] presupposed the following formulation:
(i) the code is $(s, 2^{nR}, n)$, $n = 1, 2, \ldots$, where $S_1 = S_1^s = s$ is the initial state of the noise, known to the encoder and the decoder, as discussed in Definition 1.1.3,
(ii) Conditions 1 and 2 of Section 1.1, hold, and
(iii) the $n-$FTFI capacity formula is $C_n^{fb}(\kappa, s)$ defined by (3.1.29).
It is important to emphasize that in [1, Section II.C] a specific realization of the PSD is considered to ensure Conditions 1 and 2 hold, i.e., the analysis in [1] presupposed a stationary noise and Case II) formulation.

Now, we ask:
Given the PSD of the noise defined by (1.2.26), and the double-sided realization [2, eqn(58)], i.e., the analog of time-invariant version of the PO-SS realization of Definition 1.1.2, or its analogous one-sided realization, what are the necessary conditions for the feedback capacity of [2, Theorem 6.1] to be valid?
The answer to this question is: Conditions 1 and 2 of Section 1.1 are necessary conditions. We show this in the next proposition.

Proposition 3.3.1. *Conditions for validity of the feedback capacity characterization of [2, Theo-*

rem 6.1]

Consider the AGN channel (1.1.1) driven by a stationary noise with PSD defined by (1.2.26) with the double-sided or one-sided realization [2, eqn(58)], (i.e., analog of time invariant of Definition 1.1.2).

Then a necessary condition for [2, Theorem 6.1] to hold is

$$\mathbf{P}_{X_t|X^{t-1},Y_{-\infty}^{t-1}} = \mathbf{P}_{X_t|S^t,Y_{-\infty}^{t-1}}, \quad t = 1,\ldots, \tag{3.3.54}$$

Further, Conditions 1 and 2 of Section 1.1 are necessary and sufficient for equality (3.3.54) to hold.

Proof. See Section 7.6. □

Remark 3.3.2. *Comparison of Cover and Pombra Characterization and [2, Theorem 6.1, C_{FB}] By Proposition 3.3.1, it follows that the characterization [2, Theorem 6.1, C_{FB}] corresponds to Case II) formulation and not to Case I) formulation. It is also noted that the optimization problem of [2, Theorem 6.1, C_{FB}] is precisely the optimization problem investigated by Yang, Kavcic, and Tatikonda [1, Section VI], with the additional restriction that the innovations part of the channel input is taken to be zero in [2, Theorem 6.1, C_{FB}], i.e., see [2, Lemma 6.1 and comments above it].*

Chapter 4

Asymptotic Analysis

In this chapter we address the asymptotic per unit time limit of the $n-$FTFI capacity. Our analysis includes the following.

1) In Section 4.1, we mention the fundamental differences of entropy rates of jointly Gaussian stable versus unstable noise processes.

2) In Section 4.2, we give necessary and/or sufficient conditions expressed in terms of detectability and stabilizability conditions of generalized DREs [8, 9], for existence of entropy rates, and asymptotic stationarity of the joint process $(X^n, Y^n), n = 1, 2, \ldots$.

3) In Section 4.3, we represent additional oversights of the characterizations of feedback capacity or rates, of the formulas presented in [2, Theorem 6.1], which are related to the convergence properties of generalized DREs.

This chapter also reconfirms that, in general, the asymptotic analysis of the $n-$FTFI capacity of a feedback code that depends on the initial state of the channel, i.e., $S_1 = S_1^s = s$, is fundamentally different from a code that does not depend on the initial state. The analysis of the asymptotic per unit time limit of $C_n^{fb}(\kappa, s)$ of AGN channels driven by $\mathrm{AR}(c), c \in (-\infty, \infty)$ noise, i.e., stable and unstable, is found in [15]. We consider the following definition of rate.

4.1 Fundamental Differences of Entropy Rates of Jointly Gaussian Stable Versus Unstable Noise

Definition 4.1.1. *Per unit time limit of $C_n^{fb,o}(\kappa)$ and $C_n^{fb,o}(\kappa,s)$*

Consider the AGN channel defined by (1.1.1), driven by the time-invariant PO-SS realization of Definition 1.1.2, and the code of Definition 1.1.1. Define the per unit time limit of the $n-$FTFI capacity with the limit and supremum operations interchanged, by

$$C^{fb,o}(\kappa) \stackrel{\triangle}{=} \sup_{\lim_{n\longrightarrow\infty}\frac{1}{n}\mathbf{E}\Big\{\sum_{t=1}^{n}\big(X_t\big)^2\Big\}\leq\kappa} \lim_{n\longrightarrow\infty}\frac{1}{n}\Big\{H(Y^n)-H(V^n)\Big\} \leq C^{fb}(\kappa) \stackrel{\triangle}{=} \lim_{n\longrightarrow\infty}\frac{1}{n}C_n^{fb}(\kappa) \tag{4.1.1}$$

where the supremum is taken over all time-invarinat distributions with feedback $\mathbf{P}^o_{X_t|X^{t-1},Y^{t-1}} = \mathbf{P}^o_{X_t|V^{t-1},Y^{t-1}}, t=1,2,\ldots$, such that $(X^n,Y^n), n=1,2,\ldots$, is jointly asymptotically stationary Gaussian.

For code $(s,2^{nR},n)$, $n=1,2,\ldots$, with initial state $S_1=S_1^s=s$ of Definition 1.1.3, $C^{fb,o}(\kappa)$ is replaced by $C^{fb,o}(\kappa,s)$, which is defined by (4.1.1), with differential entropies, conditional expectations, conditional distributions, defined for fixed $S_1^s=s$.

We should exphasize that our definition of rate is consistent with the definition of rates considered in [2,4–7], i.e., the interchange of limit and supremum. However, unlike [2,4–7] we treat the general time-invariant stable and unstable, PO-SS noise realization of Definition 1.1.2, not necessarily stationary or asymptotically stationary.

We should emphasize that, in general, and irrespective of whether the noise is stable or unstable, the entropy rates that appear in (4.1.1) may not exist. To show existence of the limits $C^{fb,o}(\kappa)$ and $C^{fb,o}(\kappa,s)$ we need to identify necessary and/or sufficient conditions, using the characterization of Theorem 2.2.1, when the channel input strategies are restricted to the time-invariant strategies $\Lambda_t=\Lambda^\infty, K_{Z_t}=K_Z^\infty, t=1,2,\ldots$. Clearly, by (4.1.1), whether the limit as $n\longrightarrow\infty$ exists, and supremum over channel input distributions exists, depend on the convergence properties of the coupled generalized matrix DREs, $\Sigma_n, K_n^0 \equiv K_n(\Lambda^\infty,K_Z^\infty,\Sigma)$, as $n\longrightarrow\infty$.

First, we recall the following definition, which is standard and it is found in many textbooks.

Definition 4.1.2. *Entropy rate of continuous-valued random processes*

Let $X_t:\Omega\to\mathbb{R}^{n_z}, n_x\in Z_+$ a random process defined on some probability space $(\Omega,\mathscr{F},\mathbb{P})$. The *entropy rate (differential) is defined by*

$$H_R(X^\infty) \stackrel{\triangle}{=} \lim_{n\longrightarrow\infty} \frac{1}{n} H(X_1, X_2, \ldots, X_n) \tag{4.1.2}$$

when the limit exists.

The next theorem qunatifies the existence of entropy rates of stationary Gaussian processes [9].

Theorem 4.1.1. *The entropy rate of stationary zero mean full rank Gaussian process [9]*
Let $X_t : \Omega \rightarrow \mathbb{R}^{n_x}, n_x \in Z_+, \forall t \in \mathbb{Z}_+$ be an stationary Gaussian process, with zero mean, and full rank covariance of $\mathbf{X}^n$. Let $\mathscr{H}_t^X$ denote the Hilbert space of RVs generated by $\{X_t : s \leq t, s, t \in Z_+\}$, and define the innovations process by

$$\Sigma_t \stackrel{\triangle}{=} \mathbf{E}\Big\{\Big(X_t - \mathbf{E}\Big\{X_t \Big| \mathscr{H}_{t-1}^X\Big\}\Big)\Big(X_t - \mathbf{E}\Big\{X_t \Big| \mathscr{H}_{t-1}^X\Big\}\Big)^T\Big\} \succ 0 \tag{4.1.3}$$

and its limit

$$\Sigma \stackrel{\triangle}{=} \lim_{n\longrightarrow\infty} \Sigma_n \tag{4.1.4}$$

Then the entropy rate is given by

$$H_R(X^\infty) = \frac{n_x}{2} \log\big(2\pi e\big) + \frac{1}{2} \lim_{n\longleftrightarrow\infty} \frac{1}{n} \sum_{t=1}^{n} \log|\Sigma_t| \tag{4.1.5}$$

$$= \frac{n_x}{2\pi} \log\big(2\pi e\big) + \frac{1}{2} \log|\Sigma| \tag{4.1.6}$$

when it exists.

An application of Theorem 4.1.1 is given in the next proposition [16].

Proposition 4.1.1. *Entropy rate of Gaussian process described by PSD (1.2.26)*
Let $V_t, \forall t \in \mathbb{Z}_+$ be a real, scalar-valued, stationary Gaussian noise with PSD (1.2.26), with a corresponding time-invariant stationary realization (similar to Definition 1.1.2). Then the entropy rate is given by

$$H_R(V^\infty) = \frac{1}{2} \log\big(2\pi e K_W\big). \tag{4.1.7}$$

Proof. See Section 7.7. □

The next remark is trivial; it is introduced for subsequent comparison.

Remark 4.1.1. *Let $V_t, \forall t \in \mathbb{Z}_+$ be the nonstationary ARMA$(a,c), a \in (-\infty,\infty), c \in (-\infty,\infty)$ noise of Example 2.1.2. Then the conditional entropy of V^n for fixed initial state $S_1 = S_1^s = s$, is given by*

$$H_R(V^\infty|s) \stackrel{\triangle}{=} \lim_{n\longrightarrow\infty} \frac{1}{n} H(V^n|s) = \lim_{n\longrightarrow\infty} \frac{1}{n} \sum_{t=1}^{n} \frac{1}{2} \log\left(2\pi e K_W\right) = \frac{1}{2} \log\left(2\pi e K_W\right). \tag{4.1.8}$$

The next lemma identifies fundamental conditions for the existence of the entropy rate of the time-varying PO-SS noise realization of Definition 1.1.2 (if $S_1 = S_1^s = s$ is not fixed), and includes the entropy rate $H_R(V^\infty)$ of the nonstationary ARMA$(a,c), a \in (-\infty,\infty), c \in (-\infty,\infty)$ noise of Remark 4.1.1.

Lemma 4.1.1. *Entropy rate of the time-varying PO-SS noise realization of Definition 1.1.2*
Consider the time-varying PO-SS noise realization of Definition 1.1.2. Then the following hold.
(a) The joint entropy of V^n, when it exists, is given by

$$H(V^n) = \sum_{t=1}^{n} H(\hat{I}_t) = \frac{1}{2} \sum_{t=1}^{n} \log\left(2\pi e K_{\hat{I}_t}\right) \tag{4.1.9}$$

where $\hat{I}_t, t = 1,\ldots,n$ is a zero mean covariance $K_{\hat{I}_t} \stackrel{\triangle}{=} cov(\hat{I}_t, \hat{I}_t)$, Gaussian orthogonal innovations process of V^n, defined by

$$\hat{I}_t \stackrel{\triangle}{=} V_t - \mathbf{E}\Big\{V_t \Big| V^{t-1}\Big\}, \qquad t = 1,\ldots,n \tag{4.1.10}$$

that is, $\hat{I}_t$ is independent of $\hat{I}_k, \forall k \neq t$.
(b) Suppose the sequence $K_{\hat{I}_t}, t = 1,2,\ldots,n$, is such that

$$\lim_{n\longrightarrow\infty} K_{\hat{I}_n} = K_{\hat{I}}^\infty > 0. \tag{4.1.11}$$

Then the entropy rate of $V_t, \forall t \in \mathbb{Z}_+$, is given by

$$H_R(V^\infty) = \lim_{n\longrightarrow\infty} \frac{1}{n} \sum_{t=1}^{n} H(\hat{I}_t) = \frac{1}{2} \log\left(2\pi e K_{\hat{I}}^\infty\right). \tag{4.1.12}$$

Proof. See Appendix 7.8. □

Remark 4.1.2. *Entropy rate of nonstationary Gaussian noise*
By Lemma 4.1.1, a necessary condition for existence of the entropy rate of nonstationary Gaussian process V^n is the convergence of the covariance of the Gaussian orthogonal innovations process of V^n, i.e., of $K_{\hat{I}_t} \stackrel{\triangle}{=} cov(\hat{I}_t, \hat{I}_t)$, since $\lim_{n\to\infty} \frac{1}{n} H(V^n) = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} H(\hat{I}_t)$. We can determine such necessary and/or sufficient conditions from the convergence properties of the Generalized Kalman-filter equations [8, 9] of Lemma 2.1.1.

To address the asymptotic properties of estimation errors generated by the recursions of Generalized Kalman-filters, such as, $E_t, t = 1, 2, \ldots$ of Theorem 2.2.1, generated by (2.2.112), we need to introduce the stabilizing solutions of generalized AREs. The next definition is useful in this respect.

4.2 Convergence Properties of Generalized Kalman-Filter Equations

Definition 4.2.1. *Stabilizing solutions of generalized marix AREs*
Let $(A, G, Q, S, R, C) \in \mathbb{R}^{q\times q} \times \mathbb{R}^{q\times k} \times \mathbb{R}^{k\times k} \times \mathbb{R}^{k\times p} \times \mathbb{R}^{p\times p} \times \mathbb{R}^{p\times q}$.
Define the generalized time-invariant matrix DRE

$$P_{t+1} = AP_tA^T + GQG^T - \Big(AP_tC^T + GS\Big)\Big(R + CP_tC^T\Big)^{-1} . \Big(AP_tC^T + GS\Big)^T, \quad P_1 = given, \tag{4.2.13}$$

$$P_t \in \mathbb{S}_+^{q\times q}, t = 1, \ldots, \quad R = R^T \succ 0,$$
$$F^{CL}(P) \stackrel{\triangle}{=} A - \Big(APC^T + GQG^T\Big)\Big(R + CPC^T\Big)^{-1} C.$$

Define also the corresponding generalized matrix ARE

$$P = APA^T + GQG^T - \Big(APC^T + GS\Big)\Big(R + CPC^T\Big)^{-1} . \Big(APC^T + GS\Big)^T, \quad P \in \mathbb{S}_+^{q\times q}. \tag{4.2.14}$$

A solution $P = P^T \succeq 0$ to the generalized matrix ARE (4.2.14), assuming it exists, is called stabilizing if $spec\big(F^{CL}(P)\big) \in \mathbb{D}_o$. In this case, we say $F^{CL}(P)$ is asymptotically stable, that is, the eigevalues of $F^{CL}(P)$ are stable.

With respect to any of the above generalized matrix DRE and ARE, we introduce the important notions of detectability, unit circle controllability, and stabilizability. We use these notions to characterize the convergence properties of solutions of generalized matrix DREs, P_n, as $n \longrightarrow \infty$, to a unique nonnegative stabilizing solution P of the generalized matrix ARE. These notions are used to identify necessary and/or sufficient conditions for the error recursions of generalized Kalman-filters, such as, $E_t, t = 1, 2, \ldots$ of Theorem 2.2.1, generated by (2.2.112), to converge in mean-square sense, to a unique limit.

Definition 4.2.2. *Detectability, Stabilizability, Unit Circle controllability*
Consider the generalized matrix ARE of Definition 4.2.1, and introduce the matrices

$$A^* \stackrel{\triangle}{=} A - GSR^{-1}C, \quad B^* \stackrel{\triangle}{=} Q - SR^{-1}S^T, \quad B^* = B^{*,\frac{1}{2}}\big(B^{*,\frac{1}{2}}\big)^T. \tag{4.2.15}$$

(a) The pair $\{A, C\}$ *is called detectable if there exists a matrix* $K \in \mathbb{R}^{q\times p}$ *such that* $spec\big(A - KC\big) \in \mathbb{D}_o$, *i.e., the eigenvalues* λ *of* $A - KC$ *lie in* $\mathbb{D}_o$ *(stable).*
(b) The pair $\{A^*, GB^{*,\frac{1}{2}}\}$ *is called unit circle controllable if there exists a* $K \in \mathbb{R}^{k\times q}$ *such that* $spec\big(A^* - GB^{*,\frac{1}{2}}K\big) \notin \{c \in \mathbb{C} : |c| = 1\}$, *i.e., all eigenvalues* λ *of* $A^* - GB^{*,\frac{1}{2}}K$ *are such that* $|\lambda| \neq 1$.
(c) The pair $\{A^*, GB^{*,\frac{1}{2}}\}$ *is called stabilizable if there exists a* $K \in \mathbb{R}^{k\times q}$ *such that* $spec\big(A^* - GB^{*,\frac{1}{2}}K\big) \in \mathbb{D}_o$, *i.e., all all eigenvalues* λ *of* $A^* - GB^{*,\frac{1}{2}}K$ *lie in* $\mathbb{D}_o$.
(d) The pair $\{A, C\}$ *is called observable if the rank condition holds,*

$$rank\big(\mathscr{O}\big) = q, \quad \mathscr{O} \stackrel{\triangle}{=} \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{q-1} \end{bmatrix}. \tag{4.2.16}$$

(e) The pair $\{A^*, GB^{*,\frac{1}{2}}\}$ *is called controllable if the rank condition holds,*

$$rank\big(\mathscr{C}\big) = q, \quad \mathscr{O} \stackrel{\triangle}{=} \begin{bmatrix} GB^{*,\frac{1}{2}} & A^*GB^{*,\frac{1}{2}} & \ldots & \big(A^*\big)^{q-1}GB^{*,\frac{1}{2}} \end{bmatrix}. \tag{4.2.17}$$

It is well-known that, if the pair $\{A, C\}$ is observable then it is stabilizable, and if the pair $\{A^*, GB^{*,\frac{1}{2}}\}$ is controllable then it is stabilizable [9].

In the next theorem we summarize known results on sufficient and/or necessary conditions for convergence of solutions $\{P_t, t = 1, 2, \ldots, n\}$ of the generalized time-invariant DRE (4.2.13), as $n \longrightarrow \infty$, to a nonnegative $P \succeq 0$, which is the unique stabilizing solution of a corresponding generalized ARE (4.2.14).

Theorem 4.2.1. *[8, 9] Convergence of time-invariant generalized DRE*
Let $\{P_t, t = 1,2,\ldots,n\}$ denote a sequence that satisfies the time-invariant generalized DRE (4.2.13) with arbitrary initial condition $P_1 \geq 0$.
The following hold.
(1) Consider the generalized DRE (4.2.13) with zero initial condition, i.e., $P_1 = 0$, and assume, the pair $\left\{A, C\right\}$ is detectable, and the pair $\left\{A^*, GB^{*,\frac{1}{2}}\right\}$ is unit circle controllable.
Then the sequence $\{P_t : t = 1,2,\ldots,n\}$ that satisfies the generalized DRE (4.2.13), with zero initial condition $P_1 = 0$, converges to P, i.e., $\lim_{n\longrightarrow\infty} P_n = P$, where P satisfies the generalized marix ARE (4.2.14) if and only if the pair $\left\{A^*, GB^{*,\frac{1}{2}}\right\}$ is stabilizable.
(2) Assume, the pair $\left\{A, C\right\}$ is detectable, and the pair $\left\{A^*, B^{*,\frac{1}{2}}\right\}$ is unit circle controllable. Then there exists a unique stabilizing solution $P \succeq 0$ to the generalized ARE (4.2.14), i.e., such that, $spec\big(F^{CL}(P)\big) \in \mathbb{D}_o$, if and only if $\{A^*, GB^{*,\frac{1}{2}}\}$ is stabilizable.
(3) If $\{A, C\}$ is detectable and $\{A^*, GB^{*,\frac{1}{2}}\}$ is stabilizable, then any solution $P_t, t = 1,2,\ldots,n$ to the generalized matrix DRE (4.2.13) with arbitrary initial condition, $P_1 \succeq 0$ is such that $\lim_{n\longrightarrow\infty} P_n = P$, where $P \succeq 0$ is the unique solution of the generalized matrix ARE (4.2.14) with $spec\big(F^{CL}(P)\big) \in \mathbb{D}_o$ i.e., it is stabilizing.

An application of Theorem 4.2.1 to the generalized Kalman-filter of Lemma 2.1.1 for the time-invariant PO-SS realization), is given in the next corollary; it identifies conditions for existence of the entropy rate $H_R(V^\infty)$.

Corollary 4.2.1. *The entropy rate of PO-SS noise realization based on the generalized Kalman-filter*
Let $\Sigma^o_t = \Sigma_t, t = 1,2,\ldots$ denote the solution of the generalized matrix DRE (2.1.46) of the generalized Kalman-filter of Lemma 2.1.1 of the time-invariant PO-SS realization of V^n of Definition 1.1.2, i.e., $(A_t, B_t, C_t, N_t, K_{W_t}) = (A, B, C, N, K_W), \forall t$, generated by

$$\begin{aligned}\Sigma^o_{t+1} =& A\Sigma^o_t A^T + BK_W B^T - \Big(A\Sigma^o_t C^T + BK_W N^T\Big)\Big(NK_W N^T + C\Sigma^o_t C^T\Big)^{-1} \\ &.\Big(A\Sigma^o_t C^T + BK_W N^T\Big)^T, \quad \Sigma^o_t \succeq 0, \quad t = 1,\ldots,n, \quad \Sigma^o_1 = K_{S_1} \succeq 0.\end{aligned} \tag{4.2.18}$$

$$M^{CL}(\Sigma^o) \stackrel{\triangle}{=} A - M(\Sigma^o)C, \quad M(\Sigma^o) \stackrel{\triangle}{=} \Big(A\Sigma^o C^T + B_t K_W N^T\Big)\Big(NK_W N^T + C\Sigma^o C^T\Big)^{-1}. \tag{4.2.19}$$

Let $\Sigma^\infty = \Sigma^{\infty,T} \succeq 0$ be a solution of the corresponding generalized ARE

$$\Sigma^\infty = A\Sigma^\infty A^T + BK_W B^T - \Big(A\Sigma^\infty C^T + BK_W N^T\Big)\Big(NK_W N^T + C\Sigma^\infty C^T\Big)^{-1} \cdot \Big(A\Sigma^\infty C^T + BK_W N^T\Big)^T. \tag{4.2.20}$$

Define the matrices

$$GQG^T \triangleq BK_W B^T, \quad GS \triangleq BK_W N^T, \quad R \triangleq NK_W N^T \quad \Longrightarrow \quad G \triangleq B, \quad Q \triangleq K_W, \quad S \triangleq K_W N^T \tag{4.2.21}$$

$$A^* \triangleq A - BK_W N^T \big(NK_W N^T\big)^{-1} C, \qquad B^* \triangleq K_W - K_W N^T \Big(NK_W N^T\Big)^{-1} \Big(K_W N^T\Big)^T. \tag{4.2.22}$$

Then all statements of Theorem 4.2.1 hold with (G,Q,S,R) as defined by (4.2.21).
In particular, suppose
(i) $\{A,C\}$ is detectable, and
(ii) $\{A^*, GB^{*,\frac{1}{2}}\}$ is stabilizable.
Then any solution $\Sigma_t^o, t = 1,2,\ldots,n$ to the generalized matrix DRE (4.2.18) with arbitrary initial condition, $\Sigma_1^o \succeq 0$ is such that $\lim_{n\longrightarrow\infty} \Sigma_n^o = \Sigma^\infty$, where $\Sigma^\infty \succeq 0$ is the unique solution of the generalized matrix ARE (4.2.20) with spec$\big(M^{CL}(\Sigma^\infty)\big) \in \mathbb{D}_o$ i.e., it is stabilizing.
Moreover, the entropy rate of V^n is given by

$$H_R(V^\infty) = H(\hat{I}_t^\infty) = \frac{1}{2}\log\Big(2\pi e\Big[C\Sigma^\infty C^T + NK_W N^T\Big]\Big), \quad \forall \Sigma_1^o \succeq 0, \quad \forall t \tag{4.2.23}$$

where

$$\hat{I}_t^\infty \triangleq C\big(S_t - \hat{S}_t^\infty\big) + NW_t \in N(0, C\Sigma^\infty C^T + NK_W N^T), \quad t = 1,2,\ldots, \tag{4.2.24}$$

is the stationary Gaussian innovations process, i.e., with Σ_t^o replaced by Σ^∞, and the entropy rate $H_R(V^\infty)$ is independent of the initial data $\Sigma_1^o \succeq 0$.

Proof. This is a direct application of Theorem 4.2.1. The last part follows from Lemma 4.1.1. □

Next we apply Corollary 4.2.1 to the nonstationary AR$(a,c), a \in (-\infty,\infty), c \in (-\infty,\infty)$ noise.

Lemma 4.2.1. *Properties of solutions of DREs and AREs of AR$(a,c), a \in (-\infty,\infty), c \in (-\infty,\infty)$ noise and entropy rate $H_R(V^\infty)$*

Consider the $AR(a,c), a \in (-\infty,\infty), c \in (-\infty,\infty)$ *noise of Example 2.1.2.(a), and the DRE* $\Sigma_t^o \triangleq \Sigma_t, t = 1,\ldots,n$*, generated by Corollary 3.2.3.(a), i.e.,*

$$\Sigma_{t+1}^o = \big(c\big)^2\Sigma_t^o + K_W - \Big(c\Sigma_t^o\big(c-a\big) + K_W\Big)^2\Big(K_W + \big(c-a\big)^2\Sigma_t^o\Big)^{-1}, \quad t = 1,\ldots,n, \tag{4.2.25}$$

$$\Sigma_1^o = K_{S_1} = \frac{\big(c_0\big)^2 K_{S_0} + \big(a_0\big)^2 K_{W_0}}{\Big(c_0 - a_0\Big)^2} \geq 0. \tag{4.2.26}$$

where $K_W > 0, c \neq a$*,* $K_{S_0} \geq 0, K_{W_0} \geq 0$*. Let* $\Sigma^\infty \geq 0$ *be a solution of the corresponding generalized ARE*

$$\Sigma^\infty = \big(c\big)^2\Sigma^\infty + K_W - \Big(c\Sigma^\infty\big(c-a\big) + K_W\Big)^2\Big(K_W + \big(c-a\big)^2\Sigma^\infty\Big)^{-1}. \tag{4.2.27}$$

Then,

$$\{A, C\} = \{c, c-a\}, \quad \{A^*, GB^{*,\frac{1}{2}}\} = \{a, 0\}. \tag{4.2.28}$$

and the following hold.

(1) The pair $\{A,C\} = \{c,c-a\}$ *is detectable* $\forall c \in (-\infty,\infty), a \in (-\infty,\infty)$ *(the restriction* $c \neq a$ *is always assumed).*

(2) The pair $\{A^*, GB^{*,\frac{1}{2}}\} = \{a, 0\}$ *is unit circle controllable if and only if* $|a| \neq 1$ *(*$\forall c \in (-\infty,\infty)$*).*

(3) The pair $\{A^*, GB^{*,\frac{1}{2}}\} = \{a, 0\}$ *is stabilizable if and only if* $a \in (-1,1)$ *(*$\forall c \in (-\infty,\infty)$*).*

(4) Suppose $c \in (-\infty,\infty)$ *and* $|a| \neq 1$*. The sequence* $\{\Sigma_t^o, t = 1,2,\ldots,n\}$ *that satisfies the generalized DRE with zero initial condition,* $\Sigma_1^o = 0$ *converges to* Σ^∞*, i.e.,* $\lim_{n\longrightarrow\infty}\Sigma_n^o = \Sigma^\infty$*, where* $\Sigma^\infty \geq 0$ *satisfies the ARE (4.2.27) if and only if the* $\{A^*, GB^{*,\frac{1}{2}}\} = \{a, 0\}$ *is stabilizable, equivalently,* $|a| < 1$*. Moreover, the two solutions of the quadratic equation (4.2.27) are, without imposing* $\Sigma^\infty \geq 0$ *are*

$$\Sigma^\infty = \begin{cases} 0 & \textit{the unique, stabilizing, } \Sigma^\infty \geq 0 \textit{ solution of (4.2.27)} \\ \frac{K_W\left(a^2-1\right)}{\left(c-a\right)^2} < 0 & \textit{the non-stabilizing, } \Sigma^\infty < 0 \textit{ solution of (4.2.27)}. \end{cases} \tag{4.2.29}$$

That is, $\lim_{n\longrightarrow\infty}\Sigma_n^0 = \Sigma^\infty = 0$ *is the unique and stabilizing solution* $\Sigma^\infty \geq 0$ *of (4.2.27), i.e., such that* $|M^{CL}(\Sigma^\infty)| < 1$*, if and only if* $|a| < 1$*.*

(5) Suppose $c \in (-\infty,\infty)$ *and* $|a| < 1$*. Then any solution* $\Sigma_t^o, t = 1,2,\ldots,n$ *to the generalized DRE (4.2.25) with arbitrary initial condition,* $\Sigma_1^o \geq 0$ *is such that* $\lim_{n\longrightarrow\infty}\Sigma_n^o = \Sigma^\infty$*, where* $\Sigma^\infty \geq 0$ *is the unique solution of the generalized ARE (4.2.27) with* $M^{CL}(\Sigma^\infty) \in (-1,1)$ *i.e., it is stabilizing, and*

moreover $\Sigma^\infty = 0$.

(6) Suppose $c \in (-\infty, \infty)$ *and* $|a| < 1$. *The entropy rate of* $V_t, \forall t \in \mathbb{Z}_+$, *is given by*

$$H_R(V^\infty) = \lim_{n \longrightarrow \infty} \frac{1}{n} \sum_{t=1}^{n} \frac{1}{2} \log \Big(2\pi e \big[(c-a)^2 \Sigma_t^o + K_W \big] \Big) = \frac{1}{2} \log \big(2\pi e K_W \big), \quad \forall \Sigma_1^o \geq 0. \tag{4.2.30}$$

Proof. See Appendix 7.9. □

To gain additional insight, in the next remark we discuss the application of Lemma 4.2.1 to the $\mathrm{AR}(c), c \in (-\infty, \infty)$ noise.

Remark 4.2.1. *Entropy rate* $H_R(V^\infty)$ *of the* $AR(c), c \in (-\infty, \infty)$ *noise*
From Lemma 4.2.1 we can determine conditions for existence of the entropy rate $H_R(V^\infty)$ *of the nonstationary* $AR(c), c \in (-\infty, \infty)$ *noise defined by (2.1.68), by setting* $a = 0$.
In particular, $\Sigma_t^o, t = 1, \ldots, n$ *is the solution of (4.2.25-4.2.26), with (see Corollary 3.2.3.(b), (2.2.127)), and (4.2.27) degenerates to the ARE,*

$$\Sigma^\infty = \big(c\big)^2 \Sigma^\infty + K_W - \Big(\big(c\big)^2 \Sigma^\infty + K_W \Big)^2 \Big(K_W + \big(c\big)^2 \Sigma^\infty \Big)^{-1} \tag{4.2.31}$$

For $a = 0$, *by (4.2.28) the pair* $\{A, C\} = \{c, c\}$ *is detectable, and the pair* $\{A^*, GB^{*,\frac{1}{2}}\} = \{0, 0\}$ *is stabilizable. The two solutions of the ARE (4.2.31), without imposing* $\Sigma^\infty \geq 0$, *are*

$$\Sigma^\infty = \begin{cases} 0 & \text{the unique, stabilizing, nonnegative solution of the ARE} \\ -\frac{K_W}{c^2} < 0 & \text{the non-stabilizing, negative solution) of the ARE} \end{cases} \tag{4.2.32}$$

That is, $\lim_{n \longrightarrow \infty} \Sigma_n^o = \Sigma^\infty \geq 0$, *where* $\Sigma^\infty = 0$ *is the unique (stabilizing) solution of the ARE, and corresponds to the stable eigenvalue of the error equation (see (2.1.44), i.e.,* $M^{CL}(\Sigma^\infty) = c - \frac{K_W}{K_W} c = 0$.

Next we compute the entropy rate $H_R(V^\infty)$ of the time-invariant nonstationary $\text{PO-SS}(a, c, b^1, b^2, d^1, d^2)$ noise of Corollary 2.1.1 to show fundamental differences from the entropy rate $H_R(V^\infty)$ of the $\mathrm{AR}(a, c)$ noise of Lemma 4.2.1.

Lemma 4.2.2. *Properties of solutions of DREs and AREs of* $PO\text{-}SS(a, c, b^1 = b, b^2 = 0, d^1 = 0, d^2 = d)$ *noise and entropy rate* $H_R(V^\infty)$

Consider the the time-invariant nonstationary PO-SS$(a,c,b^1,b^2=0,d^1=0,d^2=d)$ noise of Example 2.1.1, and the sequence $\Sigma_t^o \stackrel{\triangle}{=} \Sigma_t, t=1,\ldots,n$, generated by the DRE of Lemma 2.1.1, i.e.,

$$\Sigma_{t+1}^o = \big(a\big)^2\Sigma_t^o + \big(b\big)^2 K_{W^1} - \Big(a\Sigma_t^o c\Big)^2\Big(\big(d\big)^2 K_{W^2} + \big(c\big)^2\Sigma_t^o\Big)^{-1}, \quad t=1,\ldots,n,$$
$$\Sigma_1^o = K_{S_1} \geq 0, \quad \Sigma_t^o \geq 0 \tag{4.2.33}$$

where $\big(b\big)^2 K_{W^1} \geq 0, \big(d\big)^2 K_W^2 > 0$. Let $\Sigma^\infty \geq 0$ be the corresponding solution of generalized ARE

$$\Sigma^\infty = \big(a\big)^2\Sigma^\infty + \big(b\big)^2 K_{W^1} - \Big(a\Sigma^\infty c\Big)^2\Big(\big(d\big)^2 K_{W^2} + \big(c\big)^2\Sigma^\infty\Big)^{-1}. \tag{4.2.34}$$

Then

$$\{A,C\} = \{a,c\}, \quad \{A^*, GB^{*,\frac{1}{2}}\} = \{a, b\big(K_{W^1}\big)^{\frac{1}{2}}\}. \tag{4.2.35}$$

and the following hold.

(1) The pair $\{A,C\}=\{a,c\}$ is detectable $\forall c \in (-\infty,\infty), a\in(-\infty,\infty), c\neq 0$. If $c=0$ the pair $\{A,C\}=\{a,0\}$ is detectable if and only if $|a|<1$.

(2) The pair $\{A^*, GB^{*,\frac{1}{2}}\} = \{a, b\big(K_{W^1}\big)^{\frac{1}{2}}\}$ is unit circle controllable if and only if $|b\big(K_{W^1}\big)^{\frac{1}{2}}| \neq 1$, $\forall a \in (-\infty,\infty), c\in(-\infty,\infty)$.

(3) The pair $\{A^*, GB^{*,\frac{1}{2}}\} = \{a, b\big(K_{W^1}\big)^{\frac{1}{2}}\}$ is stabilizable if $b\big(K_{W^1}\big)^{\frac{1}{2}} \neq 0$, $\forall a\in(\infty,\infty), c\in(-\infty,\infty)$. If $b\big(K_{W^1}\big)^{\frac{1}{2}}=0$ the pair $\{A^*, GB^{*,\frac{1}{2}}\}=\{a,0\}$ is stabilizable if and only if $|a|<1$.

(4) Define the set

$$\begin{aligned}\mathscr{L}^\infty \stackrel{\triangle}{=} \Big\{&(a,c,\big(b\big)^2 K_{W^1}) \in (-\infty,\infty)^2 \times [0,\infty): \\ &(i) \text{ the pair } \{A,C\}=\{a,c\} \text{ is detectable, and} \\ &(ii) \text{ the pair} \{A^*, GB^{*,\frac{1}{2}}\} = \{a, b\big(K_{W^1}\big)^{\frac{1}{2}}\} \text{ is stabilizable}\Big\}.\end{aligned} \tag{4.2.36}$$

For any $(a,c,b\big(K_{W^1}\big)^{\frac{1}{2}}) \in \mathscr{L}^\infty$, any solution $\Sigma_t^o, t=1,2,\ldots,n$ to the (classical) DRE (4.2.33) with arbitrary initial condition, $\Sigma_1^o \geq 0$ is such that $\lim_{n\longrightarrow\infty}\Sigma_n^o = \Sigma^\infty$, where $\Sigma^\infty \geq 0$ is the unique solution of the (classical) ARE (4.2.34) with $M^{CL}(\Sigma^\infty) \in (-1,1)$ i.e., it is stabilizing.

(5) For any $(a,c,b^2 K_{W^1}) \in \mathscr{L}^\infty$ of part (4) the entropy rate of $V_t, \forall t \in \mathbb{Z}_+$, is given by

$$\begin{aligned}H_R(V^\infty) &= \lim_{n\longrightarrow\infty}\frac{1}{n}\sum_{t=1}^n \frac{1}{2}\log\Big(2\pi e\big[(c)^2\Sigma_t^o + \big(d\big)^2 K_{W^2}\big]\Big) \\ &= \frac{1}{2}\log\big(2\pi e\big[(c)^2\Sigma^\infty + \big(d\big)^2 K_{W^2}\big]\big), \quad \forall \Sigma_1^o \geq 0.\end{aligned} \tag{4.2.37}$$

Proof. Follow from Theorem 4.2.1. □

Next, we turn our attention to the convergence properties of the entropy rate $H_R(Y^\infty)$, which is needed for the characterization of $C^{fb,o}(\kappa)$ of Definition 4.1.1.

Theorem 4.2.2. *Asymptotic properties of entropy rate $H_R(Y^\infty)$ of Theorem 2.2.1*
Let $K_t^o, t=1,\ldots,$ be the solution of the generalized DRE (3.2.46) of the generalized Kalman-filter of Theorem 2.2.1, corresponding to the time-invariant PO-SS realization of V^n of Definition 1.1.2, $(A_t,B_t,C_t,N_t,K_{W_t})=(A,B,C,N,K_W), \forall t$, with time-invariant strategies $(\Lambda_t,K_{Z_t})=(\Lambda^\infty,K_Z^\infty), \forall t$, generated by

$$\begin{aligned} K_{t+1}^o &= AK_t^oA^T + M(\Sigma_t^o)K_{\hat{I}_t^o}\big(M(\Sigma_t^o)\big)^T - \Big(AK_t^o\big(\Lambda^\infty+C\big)^T + M(\Sigma_t^o)K_{\hat{I}_t^o}\Big)\Big(K_{\hat{I}_t^o}+K_Z^\infty \\ &\quad + \big(\Lambda^\infty+C\big)K_t^o\big(\Lambda^\infty+C\big)^T\Big)^{-1}\Big(AK_t^o\big(\Lambda^\infty+C\big)^T + M(\Sigma_t^o)K_{\hat{I}_t^o}\Big)^T, \\ K_t^o &= K_t^{o,T} \succeq 0, \quad t=1,\ldots,n, \quad K_1^o=0 \end{aligned} \tag{4.2.38}$$

where

$$K_{\hat{I}_t^o} = C\Sigma_t^oC^T + NK_WN^T, \quad \Sigma_t^o \ \text{ is a solution of (4.2.18)}, \quad M(\Sigma^o) \text{ is given by (4.2.19)}, \tag{4.2.39}$$

$$F^{CL}(\Sigma^o,K^o) \stackrel{\triangle}{=} A - F(\Sigma^o,K^o)\Big(\Lambda^\infty + C\Big), \tag{4.2.40}$$

$$F(\Sigma^o,K^o) \stackrel{\triangle}{=} \Big(AK^o\big(\Lambda^\infty+C\big)^T + M(\Sigma^o)K_{\hat{I}^o}\Big)\Big\{K_{\hat{I}^o} + K_{Z^\infty} + \big(\Lambda^\infty+C\big)K^o\big(\Lambda^\infty+C\big)^T\Big\}^{-1}. \tag{4.2.41}$$

Define the corresponding generalized ARE by

$$\begin{aligned} K^\infty &= AK^\infty A^T + M(\Sigma^\infty)K_{\hat{I}^\infty}\big(M(\Sigma^\infty)\big)^T - \Big(AK^\infty\big(\Lambda^\infty+C\big)^T + M(\Sigma^\infty)K_{\hat{I}^\infty}\Big)\Big(K_{\hat{I}^\infty}+K_Z^\infty \\ &\quad + \big(\Lambda^\infty+C\big)K^\infty\big(\Lambda^\infty+C\big)^T\Big)^{-1}\Big(AK^\infty\big(\Lambda^\infty+C\big)^T + M(\Sigma^\infty)K_{\hat{I}^\infty}\Big)^T, \quad K^\infty = K^{\infty,T} \succeq 0. \end{aligned} \tag{4.2.42}$$

where

$$K_{\hat{I}^\infty} = C\Sigma^\infty C^T + NK_WN^T, \quad \Sigma_t^o \ \text{ is a solution of (4.2.20)}, \quad M(\Sigma^\infty) \text{ is given by (4.2.19)}. \tag{4.2.43}$$

Introduce the matrices

$$\begin{aligned} & C(\Lambda^\infty) \stackrel{\triangle}{=} \Lambda^\infty + C, \quad GQG^T \stackrel{\triangle}{=} M(\Sigma^\infty)K_{\hat{I}^\infty}\big(M(\Sigma^\infty)\big)^T, \quad GS \stackrel{\triangle}{=} M(\Sigma^\infty)K_{\hat{I}^\infty}, \\ & R(K_Z^\infty) \stackrel{\triangle}{=} K_{\hat{I}^\infty} + K_Z^\infty. \implies G \stackrel{\triangle}{=} M(\Sigma^\infty), \quad Q \stackrel{\triangle}{=} K_{\hat{I}^\infty}, \quad S \stackrel{\triangle}{=} K_{\hat{I}^\infty}, \\ & A^*(\Lambda^\infty,K_Z^\infty) \stackrel{\triangle}{=} A - M(\Sigma^\infty)K_{\hat{I}^\infty}\Big(K_{\hat{I}^\infty}+K_Z^\infty\Big)^{-1}\Big(\Lambda^\infty+C\Big), \\ & B^*(K_Z^\infty) \stackrel{\triangle}{=} K_{\hat{I}^\infty} - K_{\hat{I}^\infty}\Big(K_{\hat{I}^\infty}+K_Z^\infty\Big)^{-1}K_{\hat{I}^\infty}. \end{aligned} \tag{4.2.44}$$

Suppose the detectability and stabilizability conditions of Lemma 4.2.1.(i) and (ii) hold.
Then, all statements of Theorem 5.2.1 hold with $(C(\Lambda^\infty), G, Q, S, R(K_Z^\infty))$ *as defined by (4.2.44).*
In particular, suppose
(i) $\{A, C(\Lambda^\infty)\} = \{A, \Lambda^\infty + C\}$ *is detectable, and*
(ii) $\{A^*(\Lambda^\infty, K_Z^\infty), GB^{*,\frac{1}{2}}(K_Z^\infty)\}$ *is stabilizable.*
Then any solution $K_t^o, t = 1, 2, \ldots, n$ *to the generalized matrix DRE (4.2.42) with arbitrary initial condition,* $K_1^o \succeq 0$ *is such that* $\lim_{n \longrightarrow \infty} K_n^o = K^\infty$*, where* $K^\infty \succeq 0$ *is the unique solution of the generalized matrix ARE (4.2.42) with* $\text{spec}\big(F^{CL}(K^\infty, \Sigma^\infty)\big) \in \mathbb{D}_o$ *i.e., it is stabilizing.*
Moreover, the entropy rate of Y^n *is given by*

$$H_R(Y^\infty) = H(I_t^\infty) = \frac{1}{2}\log\Big(2\pi e\Big[\Big(\Lambda^\infty + C\Big)K^\infty\Big(\Lambda^\infty + C\Big)^T + K_{\hat{I}^\infty} + K_Z^\infty\Big]\Big), \quad \forall K_1^o \succeq 0, \quad \forall t \tag{4.2.45}$$

where

$$\begin{aligned} &I_t^\infty = \big(\Lambda^\infty + C\big)\big(\hat{S}_t^\infty - \widehat{\hat{S}_t^\infty}\big) + \hat{I}_t^\infty + Z_t \in N(0; \big(\Lambda^\infty + C\big)K^\infty\big(\Lambda^\infty + C\big)^T + K_{\hat{I}^\infty} + K_Z^\infty), \\ &t = 1, 2, \ldots, \end{aligned} \tag{4.2.46}$$

is the stationary Gaussian innovations process, i.e., with (K_t^o, Σ_t^o) *replaced by* $(K^\infty, \Sigma^\infty)$*.*

Proof. Since the detectability and stabilizability conditions of Lemma 4.2.1 hold, then the statements of Corollary 4.2.1 hold. By the continuity property of solutions of generalized difference Riccati equations, with respect to its coefficients (see [9]), and the convergence of the sequence $\lim_{n \longrightarrow \infty} \Sigma_n^\infty = \Sigma^\infty$, where $\Sigma^\infty \succeq 0$ is the unique stabilizing solution of (4.2.20), then the statements of Theorem 4.2.2 hold, as stated. In particular, under the detectability and stabilizability conditions (i) and (ii), then $\lim_{n \longrightarrow \infty} K_n^o = K^\infty$, where $K^\infty \succeq 0$ is the unique and stabilizing solution of (4.2.42). □

In the next lemma we apply Theorem 4.2.2 to the AR$(a, c), a \in (-\infty, \infty), c \in (-\infty, \infty)$ noise of Example 2.1.2.(a), using Lemma 4.2.1.

Lemma 4.2.3. *Consider the* $AR(a, c), a \in (-\infty, \infty), c \in (-\infty, \infty)$ *noise of Example 2.1.2.(a), and the DRE* $\Sigma_t^o \stackrel{\triangle}{=} \Sigma_t, t = 1, \ldots, n$ *and ARE of Lemma 4.2.1, (4.2.25)-(4.2.28).*
Let $K_t^o, t = 1, \ldots, n$ *denote the solution of the DRE of Corollary 3.2.3.(a), when* $\Lambda_t = \Lambda^\infty, K_{Z_t} =$

$K_Z^\infty, K_t = K_t^o, \forall t$, *i.e., given by*

$$K_{t+1}^o = (c)^2 K_t^o + \big(M(\Sigma_t^o)\big)^2 K_{\hat{I}_t^o} - \Big(cK_t^o\big(\Lambda^\infty + c - a\big) + M(\Sigma_t^o)K_{\hat{I}_t^o}\Big)^2 \\ .\Big(K_{\hat{I}_t^o} + K_Z^\infty + \big(\Lambda^\infty + c - a\big)^2 K_t^o\Big)^{-1}, \quad K_1^o = 0, \quad t = 1, \ldots, n, \tag{4.2.47}$$

$$K_Z^\infty \geq 0, \quad K_t^o \geq 0, \quad t = 1, \ldots, n \tag{4.2.48}$$

and where

$$M(\Sigma_t^o) \stackrel{\triangle}{=} \Big(c\Sigma_t^o\big(c-a\big) + K_W\Big)\Big(K_W + \big(c-a\big)^2 \Sigma_t^o\Big)^{-1}, \tag{4.2.49}$$

$$K_{\hat{I}_t^o} = \big(c-a\big)^2 \Sigma_t^o + K_W, \quad t = 1, \ldots, n. \tag{4.2.50}$$

Define the set

$$\mathscr{L}^\infty \stackrel{\triangle}{=} \Big\{(a,c) \in (-\infty,\infty)^2, a \neq c : \ (i) \textit{ the pair } \{A,C\} = \{a, c-a\} \textit{ is detectable, and} \\ (ii) \textit{ the pair} \{A^*, GB^{*,\frac{1}{2}}\} = \{a, 0\} \ \textit{ is stabilizable}\Big\}. \tag{4.2.51}$$

For any $(a,c) \in \mathscr{L}^\infty$, *let* $K^\infty \geq 0$ *be a corresponding solution of the ARE (evaluated at* $\lim_{n \longrightarrow \infty} \Sigma_n^\infty = \Sigma^\infty = 0$*),*

$$K^\infty = (c)^2 K^\infty + K_W - \Big(cK^\infty\big(\Lambda^\infty + c - a\big) + K_W\Big)^2 \Big(K_W + K_Z^\infty + \big(\Lambda^\infty + c - a\big)^2 K^\infty\Big)^{-1}. \tag{4.2.52}$$

$$K_Z^\infty \geq 0, \quad K_W > 0. \tag{4.2.53}$$

and define the pairs

$$\{A, C(\Lambda^\infty\} = \{c, \Lambda^\infty + c - a\}, \tag{4.2.54}$$

$$\{A^*(\Lambda^\infty, K_Z^\infty), GB^{*,\frac{1}{2}}(K_Z^\infty)\} = \\ \{c - K_W\big(K_W + K_Z^\infty\big)^{-1}\big(\Lambda^\infty + c - a\big), \Big(K_W - \big(K_W\big)^2\big(K_W + K_Z^\infty\big)^{-1}\Big)^{\frac{1}{2}}\}. \tag{4.2.55}$$

Then the following hold.

(1) Suppose $\Lambda^\infty + c - a \neq 0$. *Then* $\{A, C(\Lambda^\infty\} = \{c, \Lambda^\infty + c - a\}$ *is detectable* $\forall (a,c) \in (-\infty,\infty)^2$.

(2) Suppose $\Lambda^\infty + c - a = 0$. *Then* $\{A, C(\Lambda^\infty\} = \{c, 0\}$ *is detectable for if and only if* $|c| < 1$ $\forall a \in (-\infty,\infty)$.

(3) Suppose $K_Z^\infty = 0$. *Then the pair* $\{A^*, GB^{*,\frac{1}{2}}\} = \{-\Lambda^\infty + a, 0\}$ *is unit circle controllable if and only if* $|\Lambda - a| \neq 1$ $\forall a \in (-\infty,\infty)$.

(4) Suppose $\Lambda^\infty + c - a \neq 0, |\Lambda - a| \neq 1$ $\forall (a,c) \in (-\infty,\infty)^2$, *and* $K_Z = 0$. *The sequence* $K_t^o, t =$

$1,2,\ldots,n$ that satisfies the generalized DRE (4.2.47) with zero initial condition, $K_1^o = 0$, converges to $K^\infty \geq 0$, i.e., $\lim_{n\longrightarrow\infty} K_n^o = K^\infty$, where K^∞ satisfies the generalized ARE,

$$K^\infty = (c)^2 K^\infty + K_W - \Big(cK^\infty(\Lambda^\infty + c - a) + K_W\Big)^2 \Big(K_W + (\Lambda^\infty + c - a)^2 K^\infty\Big)^{-1}, \quad K^\infty \geq 0 \tag{4.2.56}$$

if and only if the $\{A^*, GB^{*,\frac{1}{2}}\} = \{-\Lambda^\infty + a, 0\}$ is stabilizable, equivalently, $|\Lambda^\infty - a| < 1$. Moreover, the solutions of the ARE (4.2.56), under the stabilizability condition, i.e., $|\Lambda^\infty - a| < 1$, are

$$K^\infty = \begin{cases} 0 & \text{the unique, stabilizing, } K^\infty \geq 0 \text{ solultion of (4.2.56)} \\ \dfrac{K_W\Big(\big(\Lambda^\infty - a\big)^2 - 1\Big)}{\big(\Lambda^\infty + c - a\big)^2} < 0 & \text{the non-stabilizing, } K^\infty < 0 \text{ solution of (4.2.56)}. \end{cases} \tag{4.2.57}$$

That is, $\lim_{n\longrightarrow\infty} \Sigma_n^0 = \Sigma^\infty = 0$ is the unique and stabilizing solution $\Sigma^\infty \geq 0$ of (4.2.56), i.e., such that $|M^{CL}(\Sigma^\infty)| < 1$, if and only if $|\Lambda^\infty - a| < 1$.

Proof. The statements follow from Lemma 4.2.1, Theorem 4.2.2 (and general properties of Theorem 4.2.1). □

Remark 4.2.2. *From Lemma 4.2.3.(4) follows that if $K_Z^\infty = 0$ then the unique and stabilizing solution is $K^\infty = 0$ and corresponds to $|\Lambda^\infty - a| < 1$. This is an application of Theorem 4.2.1.(1).*

In the next theorem we characterize the asymptotic limit of Definition 4.1.1, by invoking Theorem 2.2.1, Corollary 4.2.1, and Theorem 4.2.2.

Theorem 4.2.3. *Feedback capacity $C^{fb,o}(\kappa)$ of Theorem 2.2.1*
Consider $C^{fb,o}(\kappa)$ of Definition 4.1.1 corresponding to Theorem 2.2.1, i.e., the PO-SS realization of V^n of Definition 1.1.2 is time-invariant, $(A_t, B_t, C_t, N_t, K_{W_t}) = (A, B, C, N, K_W), \forall t$, and the strategies are time-invariant, $(\Lambda_t, K_{Z_t}) = (\Lambda^\infty, K_Z^\infty), \forall t$.
Define the set

$$\begin{aligned} \mathscr{P}^\infty \stackrel{\triangle}{=} \Big\{ & (\Lambda^\infty, K_Z^\infty) \in (-\infty, \infty) \times [0, \infty) : \\ & \text{(i) } \{A, C\} \text{ of Corollary 4.2.1 is detectable,} \\ & \text{(ii) } \{A^*, GB^{*,\frac{1}{2}}\} \text{ of Corollary 4.2.1 is stabilizable, } (A^*, B^*) \text{ defined by (4.2.22)} \\ & \text{(iii) } \{A, C(\Lambda^\infty)\} = \{A, \Lambda^\infty + C\} \text{ of Theorem 4.2.2 is detectable,} \\ & \text{(iv) } \{A^*(\Lambda^\infty, K_Z^\infty), GB^{*,\frac{1}{2}}(K_Z^\infty)\} \text{ of Theorem 4.2.2 is stabilizable} \Big\}. \end{aligned} \tag{4.2.58}$$

Then

$$C^{fb,o}(\kappa) = \sup_{\left(\Lambda^\infty, K_Z^\infty\right):\ \lim_{n\longrightarrow\infty}\frac{1}{n}\sum_{t=1}^{n}\left(\Lambda^\infty K_t^o (\Lambda^\infty)^T + K_Z^\infty\right)\leq\kappa} \Bigg\{ \lim_{n\longrightarrow\infty}\frac{1}{2n}\sum_{t=1}^{n}\log\Big(\frac{\Big(\Lambda^\infty + C\Big)K_t^o\Big(\Lambda^\infty + C\Big)^T + C\Sigma_t^o C^T + NK_W N^T + K_Z^\infty}{C\Sigma_t^o C^T + NK_W N^T}\Big)\Bigg\} \tag{4.2.59}$$

$$= \sup_{\left(\Lambda^\infty, K_Z^\infty\right)\in\mathscr{P}^\infty(\kappa)} \frac{1}{2}\log\Big(\frac{\Big(\Lambda^\infty + C\Big)K^\infty\Big(\Lambda^\infty + C\Big)^T + C\Sigma^\infty C^T + NK_W N^T + K_Z^\infty}{C\Sigma^\infty C^T + NK_W N^T}\Big) \tag{4.2.60}$$

where

$$\mathscr{P}^\infty(\kappa) \stackrel{\triangle}{=} \Big\{(\Lambda^\infty, K_Z^\infty) \in \mathscr{P}^\infty : K_Z^\infty \geq 0,\ \ \Lambda^\infty K^\infty (\Lambda^\infty)^T + K_Z^\infty \leq \kappa,$$
$$K^\infty \ \ \text{is the unique and stabiizing solution of (4.2.42), i.e., } |F^{CL}(\Sigma^\infty, K^\infty)| < 1$$
$$\Sigma^\infty \ \ \text{is the unique, stabiizing solution of (4.2.20), i.e., } |M^{CL}(\Sigma^\infty)| < 1\Big\} \tag{4.2.61}$$

provided there exists $\kappa \in [0,\infty)$ such that the set $\mathscr{P}^\infty(\kappa)$ is non-empty.
Moreover, the maximum element $(\Lambda^\infty, K_Z^\infty) \in \mathscr{P}^\infty(\kappa)$, is such that it induces asymptotic stationarity of the joint input and output process, and $C^{fb,o}(\kappa)$ is independent of the initial conditions $K_1^o \succeq 0, \Sigma_1^o \succeq 0$.

Proof. By Definition 4.1.1, Theorem 2.2.1, Corollary 4.2.1, and Theorem 4.2.2, then follows (4.2.59). We defined the set $\mathscr{P}^\infty$ using the detectability and stabilizability conditions of Corollary 4.2.1, and Theorem 4.2.2 to ensure convergence of solutions $\{(K_t^o, \Sigma_t^o) : t = 1,2,\ldots,n\}$ of the generalized matrix DREs to unique nonnegative, stabilizing solutions of the corresponding generalized matrix AREs. Then, for any element $(\Lambda^\infty, K_Z^\infty) \in \mathscr{P}^\infty$ both summands in (4.2.59) converge. This establishes the characterization of the right hand side of (4.2.60). □

4.3 Oversights of the characterizations of feedback capacity

Conclusion 4.3.1. *Degenerate version of Theorem 4.2.3 for feedback code of Definition 1.1.3, i.e., $(s, 2^{nR}, n)$, $n = 1,2,\ldots$*

(a) The characterization of feedback capacity $C^{fb,o}(\kappa, s)$ of the AGN channel (1.1.1) driven by a

noise V^n of Definition 1.1.2, for the code of Definition 1.1.3, i.e., $(s, 2^{nR}, n)$, $n = 1, 2, \ldots$, is a degenerate case of Theorem 4.2.3, and corresponds to $\Sigma_t = \Sigma_t^s, t = 1, \ldots, \Sigma_1 = \Sigma_1^s = 0$. In particular, since Theorem 4.2.3 characterizes $C^{fb,o}(\kappa)$ for all initial data $\Sigma_1 \succeq 0$, then includes $\Sigma_1 = \Sigma_1^s = 0$, and follows that $C^{fb,o}(\kappa) = C^{fb,o}(\kappa, s)$, where $C^{fb,o}(\kappa, s)$ independent of the initial state $S_1^s = s$.
(b) The maximal information rate of [1, Theorem 7 and Corollary 7.1], i.e., of Case II) formulation, should be read with caution, because the condition of Theorem 4.2.1.(1) are required for convergence. Similarly, the characterization of feedback capacity of [2, Theorem 6.1] which correspond to Case II) formulation, violates Theorem 4.2.1.(1), because is states that a zero variance of the innovations process is optimal, i.e., $K_Z^\infty = 0$. Consequently, subsequent papers that build on [2] to derive additional results, such as, [4–7], should be read with caution.

We apply Theorem 4.2.3 to obtain $C^{fb,o}(\kappa)$ of AR$(a,c), a \in (-\infty,\infty), c \in (-\infty,\infty)$ noise.

Corollary 4.3.1. *Consider the AR$(a,c), a \in (-\infty,\infty), c \in (-\infty,\infty)$ noise of Example 2.1.2.(a). Define the set*

$$
\begin{aligned}
\mathscr{P}^\infty \triangleq \Big\{ & (\Lambda^\infty, K_Z^\infty) \in (-\infty,\infty) \times [0,\infty) : \\
& \text{(i) } c \in (-\infty,\infty), a \in (-1,1), c \neq a, \\
& \text{(ii) the pair } \{A, C(\Lambda^\infty)\} \triangleq \{c, \Lambda^\infty + c - a\} \text{ is detectable}, \\
& \text{(ii) the pair } \{A^*(\Lambda^\infty, K_Z^\infty), GB^{*,\frac{1}{2}}(K_Z^\infty)\} \text{ is stabilizable, where} \\
& A^*(\Lambda^\infty, K_Z^\infty) \triangleq \; c - K_W \big(K_W + K_Z^\infty\big)^{-1} \big(\Lambda^\infty + c - a\big), \\
& GB^{*,\frac{1}{2}}(K_Z^\infty) \triangleq \Big(K_W - \big(K_W\big)^2 \big(K_W + K_Z^\infty\big)^{-1}\Big)^{\frac{1}{2}} \Big\}.
\end{aligned}
$$

Then,

$$
C^{fb,o}(\kappa) = \sup_{\left(\Lambda^\infty, K_Z^\infty\right) \in \mathscr{P}^\infty(\kappa)} \frac{1}{2} \log\Big(\frac{\big(\Lambda^\infty + c - a\big)^2 K^\infty + K_W + K_Z^\infty}{K_W}\Big) = C^{fb}(\kappa, s), \quad \forall s \tag{4.3.62}
$$

where,

$$
\begin{aligned}
& \mathscr{P}^\infty(\kappa) \triangleq \Big\{ (\Lambda^\infty, K_Z^\infty) \in \mathscr{P}^\infty : \quad \big(\Lambda^\infty\big)^2 K^\infty + K_Z^\infty \leq \kappa, \\
& K^\infty \geq 0 \;\; \text{is the unique and stabiizing solution of} \\
& K^\infty = \big(c\big)^2 K^\infty + K_W - \Big(cK^\infty\big(\Lambda^\infty + c - a\big) + K_W\Big)^2 \Big(K_W + K_Z^\infty + \big(\Lambda^\infty + c - a\big)^2 K^\infty\Big)^{-1} \Big\}
\end{aligned} \tag{4.3.63}
$$

provided there exists $\kappa \in [0, \infty)$ such that the set $\mathscr{P}^{\infty}(\kappa)$ is non-empty.
Moreover, $C^{fb,o}(\kappa)$ and $C^{fb,o}(\kappa, s)$ are independent of $\Sigma_1 \geq 0$ and s, respectively, and the following identities hold.

$$C^{fb,o}(\kappa) = C^{fb,o}(\kappa, s) = C^{fb,S,o}(\kappa, s), \quad \forall s \tag{4.3.64}$$

Proof. The first part is an application of Theorem 4.2.3, Lemma 4.2.1, and Lemma 4.2.3. It remains to show (4.3.64). The equality $C^{fb,o}(\kappa) = C^{fb,o}(\kappa, s), \forall s$ holds by Conclusion 4.3.1.(a). The last equality holds, because for the AR$(a, c), a \in (-\infty, \infty), c \in (-\infty, \infty)$ noise, if the initial state $S_1 = S_1^s = s$ is known to the encoder and the decoder, then Conditions 1 of Section 1.1 holds, and in addition Condition 2 holds, as easily verified from the equations (2.1.65), (2.1.66). □

Remark 4.3.1. *From Corollary 4.3.1 we obtain the degenerate cases, AR$(c), c \in (-\infty, \infty)$, noise i.e., setting $a = 0$. The various implications of the detectability and stabilizability conditions for the AR$(c), c \in (-\infty, \infty)$ noise are found in [15, see Theorem III.1 and Lemma III.2]. The complete analysis of the corresponding $C^{fb,o}(\kappa, s)$ is found in [15], and states that for stable AR(c), and time-invariant strategies, then feedback does not increase capacity.*

Chapter 5

Feedback Capacity of $ARMA(a,c)$noise

5.0.1 Problem Formulation

We introduce the precise mathematical formulation, and the underlying assumptions based on which we derive the results of this chapter. We consider the time-varying AGN channel defined by

$$Y_t = X_t + V_t, \quad t = 1, \ldots, n, \quad \frac{1}{n}\mathbf{E}_s\Big\{\sum_{t=1}^{n}(X_t)^2\Big\} \leq \kappa, \quad \kappa \in [0, \infty) \tag{5.0.1}$$

where

$X^n = \{X_1, X_2, \ldots, X_n\}$ is the sequence of channel input random variables (RVs) $X_t : \Omega \to \mathbb{R}$,

$Y^n = \{Y_1, Y_2, \ldots, Y_n\}$ is the sequence of channel output RVs $Y_t : \Omega \to \mathbb{R}$,

$V^n = \{V_1, \ldots, V_n\}$ conditioned on the initial state $S_1 = s$, is a sequence of jointly Gaussian distributed RVs $V_t : \Omega \to \mathbb{R}$, and $V^n \in N(0, K_{V^n|S_1})$,

$S_1 = s$, is known to the encoder and decoder,

$N(0, K_{V^n|S_1})$ denotes the distribution of the Gaussian RV V^n conditional on S_1, with zero conditional mean, and conditional variance $K_{V^n|S_1}$,

$\mathbf{E}_s\{\cdot\}$ denotes expectation for fixed initial state $S_1 = s$.

Definition 5.0.1. *The time-invariant ARMA(a,c) noise*

A time-invariant autoregressive moving average noise , with initial state $S_1 = s$, is defined by

$$ARMA(a,c): \quad \begin{cases} V_t = cV_{t-1} + W_t - aW_{t-1}, \quad V_0 = v_0, \quad W_0 = w_0, \quad t = 1, \ldots, n, \\ W_t \in N(0, K_{W_t}), \quad t = 1, \ldots, n, \quad \textit{indep. Gaussian, indep. of } V_0 \in N(0, K_{V_0}), \\ K_{V_0} \geq 0, \quad K_{W_t} = K_W > 0, \quad a \in (-\infty, \infty), \quad c \in (-\infty, \infty), \quad c \neq a, \\ t = 1, \ldots, n \quad \textit{non-random.} \end{cases} \tag{5.0.2}$$

Remark 5.0.1. *ARMA(a,c) in state-space representation.*
To express the ARMA(a,c) in state form we define the state variable of the noise by

$$S_t = \frac{cV_{t-1} - aW_{t-1}}{c-a}, \quad t = 1,\dots,n \tag{5.0.3}$$

Then, the state space realization of V_n is

$$S_{t+1} = cS_t + W_t \quad t = 1,\dots,n \tag{5.0.4}$$

$$V_t = (c-a)S_t + W_t, \quad t = 1,\dots,n \tag{5.0.5}$$

$$K_{S_1} = \frac{(c)^2 K_{V_0} + (a)^2 K_{W_0}}{(c-a)^2}, \quad K_{V_0} \geq 0, \quad K_{W_0} \geq 0 \ \textit{both given.} \tag{5.0.6}$$

For the stable noise, ARMA$(a,c), a \in [-1,1], c \in (-1,1)$, the variance defined by $K_{V_t} \stackrel{\triangle}{=} \mathbf{E}\big(V_t\big)^2$, satisfies $K_{V_t} = (c-a)^2 K_{S_t} + K_W, K_{S_1} \geq 0, t = 1,\dots,n$. The stable ARMA$(a,c)$ noise is called asymptotically stationary if $\lim_{n\longrightarrow\infty} K_{V_n} = \frac{(c-a)^2 K_W}{1-c^2}$, for all initial values $K_{V_0} \geq 0$, i.e., $|c| < 1, |a| \leq 1$. ARMA(a,c) without an initial state is defined by (5.0.2), for $t = 1,\dots,n$, with $V_1 \in N(0, K_{V_1}), K_{V_1} \geq 0$, independent of $W_t \in N(0, K_{W_t}), K_{W_t} > 0, t = 1,\dots,n$. Similarly, the stable ARMA(a,c) noise without an initial state is called asymptotically stationary if $K_{V_t} = (c-a)^2 K_{S_t} + K_W, K_{V_1} \geq 0, t = 1,\dots,n$, converges, $\lim_{n\longrightarrow\infty} K_{V_n} = \frac{(c-a)^2 K_W}{1-c^2}$, for all initial values $K_{V_1} \geq 0$, $|c| < 1, |a| \leq 1$. That is, the invariant distribution of the noise is $N(0, \frac{(c-a)^2 K_W}{1-c^2}), c \in (-1,1), a \in [-1,1]$.

At this stage, we introduce the feedback code and non-feedback code of the AGN channel.

Definition 5.0.2. *Feedback and non-feedback codes*
(a) A noiseless time-varying feedback code[1] for the AGN Channel, is denoted by $\mathscr{C}^{fb}_{\mathbb{Z}^+} \stackrel{\triangle}{=} \big\{(n, \lceil M_n \rceil, s, \kappa, \varepsilon_n) : n = 1,2,\dots,\big\}$, and consists of the following elements and assumptions.
(i) The set of uniformly distributed messages $W : \Omega \rightarrow \mathscr{M}^{(n)} \stackrel{\triangle}{=} \{1,2,\dots,\lceil M_n \rceil\}$.
(ii) The set of codewords of block length n, defined by the set[2]

$$\mathscr{E}_{[0,n]}(\kappa) \triangleq \Big\{ X_1 = e_1(W, S_1), X_2 = e_2(W, S_1, X_1, Y_1), \dots, X_n = e_n(W, S_1, X^{n-1}, Y^{n-1}) : \frac{1}{n+1} \mathbf{E}^e_s \Big(\sum_{i=0}^{n} (X_t)^2 \Big) \leq \kappa \Big\}. \tag{5.0.7}$$

(iii) The decoder functions $(s, y^n) \longmapsto d_n(s, y^n) \in \mathscr{M}^{(n)}$, with average error probability

$$\mathbf{P}^{(n)}_{error}(s) = \mathbb{P}\Big\{ d_n(S_1, Y^n) \neq W \Big| S_1 = s \Big\} = \frac{1}{\lceil M_n \rceil} \sum_{w=1}^{\lceil M_n \rceil} \mathbf{P}^e_s \Big(d_n(S_1, Y^n) \neq W \Big) \leq \varepsilon_n. \tag{5.0.8}$$

[1] A time-varying feedback code means the channel input distributions $\mathbf{P}_{X_t|X^{t-1},Y^{t-1},S_1}, t = 1,\dots,n$ are time-varying.
[2] The superscript $e(\cdot)$ on $\mathbf{E}^e_s$ is used to denote that the distribution depends on the strategy $e(\cdot) \in \mathscr{E}_{[0,n]}(\kappa)$.

where $\mathbf{P}_s^e$ means the distribution depends on $e(\cdot) \in \mathscr{E}_{[0,n]}(\kappa)$ and $S_1 = s$ is fixed.

(iv) "X^n is causally related to V^n" [3, page 39, above Lemma 5], which is equivalent to the following decomposition of the joint probability distribution of (X^n, V^n) given S_1.

$$\mathbf{P}_{X^n,V^n|S_1} = \mathbf{P}_{V^n|S_1} \prod_{t=1}^{n} \mathbf{P}_{X_t|X^{t-1},V^{t-1},S_1}. \tag{5.0.9}$$

$$= \mathbf{P}_{V^n|S_1} \prod_{t=1}^{n} \mathbf{P}_{X_t|X^{t-1},Y^{t-1},S_1}, \quad \text{by } Y_t = X_t + V_t. \tag{5.0.10}$$

The coding rate is $r_n \triangleq \frac{1}{n}\log\lceil M_n \rceil$. Given an initial state $S_1 = s$, a rate $R(s)$ is called an achievable rate, if there exists a code sequence $\mathscr{C}_{\mathbb{Z}^+}^{fb}$, satisfying $\lim_{n\longrightarrow\infty} \varepsilon_n = 0$ and $\liminf_{n\longrightarrow\infty} \frac{1}{n}\log\lceil M_n\rceil \geq R(s)$.

The operational definition of the feedback capacity of the AGN channel, for fixed $S_1 = s$, is $C(\kappa, s) \triangleq \sup\{R(s) : R(s) \text{ is achievable}\}$.

(b) A time-varying code without feedback for the AGN Channel, denoted by $\mathscr{C}_{\mathbb{Z}^+}^{nfb}$, is the restriction of the time-varying feedback code $\mathscr{C}_{\mathbb{Z}^+}^{fb}$, to the subset $\mathscr{E}_{[0,n]}^{nfb}(\kappa) \subset \mathscr{E}_{[0,n]}(\kappa)$, defined by

$$\mathscr{E}_{[0,n]}^{nfb}(\kappa) \triangleq \Big\{ X_1 = e_1^{nfb}(W, S_1), X_2 = e_2^{nfb}(W, S_1, X_1), \ldots, X_n = e_n^{nfb}(W, S_1, X^{n-1}) : \frac{1}{n}\mathbf{E}_s^{e^{nfb}}\Big(\sum_{i=1}^{n}(X_t)^2\Big) \leq \kappa \Big\}. \tag{5.0.11}$$

Since the code sequence $\mathscr{C}_{\mathbb{Z}^+}^{fb}$ depends on $S_1 = s$, then in general, the rate $R(s)$, and also $C(\kappa, s)$ depend on s. The Cover and Pombra AGN Channel [3], characterization of feedback capacity, and optimal channel input are recalled in Section 1.2, to emphasize that the assumptions based on which these are derived are fundamentally different from the assumptions based on which [2, Theorem 6.1] (and equivalently [2, Theorem 4.1]) are derived.

Feedback Capacity of Time-Varying Channel Input Strategies. Consider the feedback code of Definition 5.0.2.(a), i.e., $\mathscr{C}_{\mathbb{Z}^+}^{fb}$. Given the elements of the set $\mathscr{E}_{[0,n]}(\kappa)$, by the maximum entropy principle of Gaussian distributions, similar to Cover and Pombra [3], the upper bound holds[3].

$$I^e(W; Y^n|s) \leq H(Y^n|s) - H(V^n|s), \quad \text{if } H(Y^n|s) \text{ is evaluated at a Gaussian } \mathbf{P}_{Y^n|S_1} \tag{5.0.12}$$

where $H(X|s)$ stands for differential entropy of RV X conditioned on the initial state $S_1 = s$. Further, the upper bound in (5.0.12) is achieved [3], if the input X^n is jointly Gaussian for fixed $S_1 = s$,

[3]The superscript e means the underlying distributions are induced by the channel distribution and the elements of the set $e(\cdot) \in \mathscr{E}_{[0,n]}(\kappa)$.

satisfies the average power constraint, and respects (5.0.9). By the chain rule of mutual information, $I^e(W;Y^n|s) = \sum_{t=1}^n I^e(W;Y_t|Y^{t-1},s)$, and the data processing inequality, then the following inequality holds:

$$\sup_{\mathscr{E}_{[0,n]}(\kappa)} I^e(W;Y^n|s) \leq \sup_{\mathbf{P}_{X_t|X^{t-1},Y^{t-1},S_1},t=1,\ldots,n:\ \frac{1}{n}\mathbf{E}_s\left\{\sum_{t=1}^n \left(X_t\right)^2\right\}\leq\kappa} H(Y^n|s) - H(V^n|s), \quad \text{by (5.0.1)} \tag{5.0.13}$$

where the supremum in the right hand side of (5.0.13) is taken over conditionally Gaussian time-varying distributions $\mathbf{P}_{X_t|X^{t-1},Y^{t-1},S_1}, t = 1,\ldots,n$, such that (X^n,Y^n) are jointly Gaussian for fixed $S_1 = s$, and (5.0.9) is respected.

Define, as in [3], the $n-$finite transmission feedback information (FTFI) capacity of code $\mathscr{C}_{\mathbb{Z}^+}^{fb}$, by

$$C_n(\kappa,s) \stackrel{\triangle}{=} \sup_{\mathbf{P}_{X_t|X^{t-1},Y^{t-1},S_1},t=1,\ldots,n:\ \frac{1}{n}\mathbf{E}_s\left\{\sum_{t=1}^n \left(X_t\right)^2\right\}\leq\kappa} H(Y^n|s) - H(V^n|s) \tag{5.0.14}$$

provided the supremum element exists in the set. From the converse and direct coding theorems in [3, Theorem 1], it then follows that the characterization of feedback capacity of code $\mathscr{C}_{\mathbb{Z}^+}^{fb}$, is given by

$$C(\kappa,v_0) = \lim_{n\longrightarrow\infty} \frac{1}{n} C_n(\kappa,s) \tag{5.0.15}$$

provided the limit exists.

Capacity Without Feedback of Time-Varying Channel Input Strategies. Let $C_n^{nfb}(\kappa,s)$ be defined as in (5.0.14), with the time-varying feedback distributions $\mathbf{P}_{X_t|X^{t-1},Y^{t-1},S_1}, t = 1,\ldots,n$, replaced by the time-varying non-feedback distributions $\mathbf{P}_{X_t|X^{t-1},S_1}, t = 1,\ldots,n$, called $n-$finite transmission without feedback information (FTwFI) capacity. The non-feedback capacity of the code $\mathscr{C}_{\mathbb{Z}^+}^{nfb}$ of Definition 5.0.2.(b), is characterized by $C^{nfb}(\kappa,s) = \lim_{n\longrightarrow\infty} \frac{1}{n} C_n^{nfb}(\kappa,s)$, provided the limit is defined.

This brings us to the next definition of capacity, where conditions for existence of the limits of average power and entropy rates are characterized, and they part of our problem formulation.

Feedback Capacity of Time-Invariant Channel Input Strategies. We consider (5.0.14), (5.0.15) with the per unit time limit and supremum operations interchanged, and time-invariant codes and induced distributions, called strategies. To ensure the feedback capacity (to be defined shortly) is well-posed, we introduce the following condition:

(C1) Channel input strategies with feedback are time-invariant, the consistency condition (5.0.9) holds, and the following limits exists and they are finite:
(i) $\lim_{n\longrightarrow\infty}\frac{1}{n}\mathbf{E}_s\Big\{\sum_{t=1}^{n}\big(X_t\big)^2\Big\}\in[0,\infty)$, (ii) $\lim_{n\longrightarrow\infty}\frac{1}{n}\big\{H(Y^n|s)-H(V^n|s)\big\}\in[0,\infty)$.

We define the operational information feedback capacity under condition **(C1)**, as follows.

$$C^{\infty}(\kappa,s)\stackrel{\triangle}{=}\sup_{\lim_{n\longrightarrow\infty}\frac{1}{n}\mathbf{E}\big\{\sum_{t=1}^{n}\big(X_t\big)^2\big\}\leq\kappa,\ \text{subject to }\mathbf{(C1)}}\lim_{n\longrightarrow\infty}\frac{1}{n}\Big\{H(Y^n|s)-H(V^n|s)\Big\} \tag{5.0.16}$$

where the supremum is taken over all jointly Gaussian channel input processes $X^n,n=1,2,\ldots$ with feedback, or distributions with feedback $\mathbf{P}_{X_t|X^{t-1},Y^{t-1},S_1},t=1,2,\ldots$, such that $(X^n,Y^n),n=1,2,\ldots,$ is jointly Gaussian, for $S_1=s$, and **(C1)** holds.

Capacity Without Feedback of Time-Invariant Channel Input Strategies. Similar to (5.0.16), we also analyze the non-feedback capacity analog, under condition (**C1**), which is defined as follows.

$$C^{\infty,nfb}(\kappa,s)\stackrel{\triangle}{=}\sup_{\lim_{n\longrightarrow\infty}\frac{1}{n}\mathbf{E}\big\{\sum_{t=1}^{n}\big(X_t\big)^2\big\}\leq\kappa,\ \text{subject to }\mathbf{(C1)}}\lim_{n\longrightarrow\infty}\frac{1}{n}\Big\{H(Y^n|s)-H(V^n|s)\Big\} \tag{5.0.17}$$

where the supremum is taken over all jointly Gaussian channel input processes $X^n,n=1,2,\ldots,$ without feedback or distributions without feedback, denoted by $\mathbf{P}_{X_t|X^{t-1},S_1},t=1,2,\ldots$, such that $(X^n,Y^n),n=1,2,\ldots$ is jointly Gaussian for $S_1=s$, **(C1)** holds (with $\mathbf{P}_{X_t|X^{t-1},Y^{t-1},S_1}$ replaced by $\mathbf{P}_{X_t|X^{t-1},S_1}$, and (5.0.9) is respected, for $n=1,2,\ldots$. To our knowledge, for AGN channels driven by an unstable noise V^n, no closed form expression of non-feedback capacity is ever reported in the literature.
Given the above formulation, in this paper we obtain answers to the various questions listed under Problem 5.0.1.

Problem 5.0.1. *Main problem*
Given $C^{\infty}(\kappa,s)$ defined by (5.0.16), and $C^{\infty,nfb}(\kappa,s)$ defined by (5.0.17), of the AGN channel driven by a time-invariant stable and unstable, ARMA(a,c) noise, i.e., $c\in(-\infty,\infty)$:
(a) What are necessary and/or sufficient conditions for **(C1)** *to hold?*
(b) What are necessary and/or sufficient conditions for joint asymptotic stationarity of the process $(X^n,Y^n,V^n),n=1,2,\ldots$ or the marginal processes X^n and $Y^n,n=1,2,\ldots$?.
(c) What are the characterizations and closed form formulas of feedback capacity $C^{\infty}(\kappa,s)=C^{\infty}(\kappa),\forall s$?

(d) How do we extract simple lower bounds on non-feedback capacity, $C^{\infty,nfb}(\kappa,s)=C^{\infty,nfb}(\kappa), \forall s$ from the characterizations of feedback capacity?

To address Problem 5.0.1 we make use of the identities

$$\lim_{n\longrightarrow\infty}\frac{1}{n}H(Y^n|s)=\lim_{n\longrightarrow\infty}\frac{1}{n}\sum_{t=1}^{n}H(Y_t|Y^{t-1},s)=\lim_{n\longrightarrow\infty}\frac{1}{n}\sum_{t=1}^{n}H(Y_t-\mathbf{E}\big\{Y_t|Y^{t-1},s\big\}|Y^{t-1},s) \tag{5.0.18}$$

$$=\lim_{n\longrightarrow\infty}\frac{1}{n}\sum_{t=1}^{n}H(I_t),\quad I_t\stackrel{\triangle}{=}Y_t-\mathbf{E}\big\{Y_t|Y^{t-1},s\big\}\ \text{ an indep. innovations process} \tag{5.0.19}$$

Then, we identify necessary and/or sufficient conditions for the limits in (5.0.18) and $\lim_{n\longrightarrow\infty}\frac{1}{n}\mathbf{E}_s\big\{\sum_{t=1}^{n}\big(X_t\big)^2\big\}$ to exist and to be finite, and we provide answers to the questions of Problem 5.0.1, from the properties of the innovations process.

5.0.2 Methodology of the Chapter

Our methodology is based on the following main steps.

Step 1. We characterize $C_n(\kappa,s)$ defined by (5.0.14), i.e., the $n-$FTFI capacity, of the AGN channel driven by a time-varying ARMA(a_t,c_t) noise. We also give a lower bound on the characterization of the $n-$FTwFI capacity $C_n^{nfb}(\kappa,s)$, using a Gaussian channel input process, which is realized by an ARMA(Λ_t) process,

$$X_t=\Lambda_t X_{t-1}+Z_t,\quad X_1=Z_1,\quad \Lambda_t\in(-\infty,\infty),\quad t=2,\ldots,n \tag{5.0.20}$$

where Z^n an independent Gaussian sequence, independent of (V^n,S_1).

Step 2. We characterize the feedback capacity $C^{\infty}(\kappa,s)=C^{\infty}(\kappa),\forall s$ defined by (5.0.16), and we give a lower bound on the characterization of $C^{\infty,nfb}(\kappa,s)$ defined by (5.0.17), of the AGN channel driven by a time-invariant stable or unstable noise, ARMA$(a,c), a\in(-\infty,\infty), c\in(-\infty,\infty)$. Our analysis identifies necessary and/or sufficient conditions for condition **(C1)** to hold, expressed in terms of the convergence properties of *generalized difference Riccati equations (DREs) and algebraic Riccati equations (AREs)*, of estimating the channel state, that is, the noise V^n, from the channel output process Y^n, and the initial state $S_1=s$, for $n=1,2,\ldots$. This step is analogous to [17, Theorem 4.1], although the models considered in [17] involve a classical control DRE and ARE.

Step 3. We derive a closed form formula of feedback capacity $C^{\infty}(\kappa,s)=C^{\infty}(\kappa),\forall s$, that shows there are *multiple regimes* of capacity, and these regimes depend on the parameters (a,c,K_W,κ).

Our feedback capacity formulae $C^{\infty}(\kappa)$ for AGN channels driven by stable noise ARMA$(a,c), a \in [-1,1], c \in (-1,1)$ is fundamentally different from the one obtained using the characterization of feedback capacity in [2, Theorem 6.1]. We show this difference is mainly attributed to the appended detectability and stabilizability conditions on the characterization of our feedback capacity, to ensure the optimal channel input process X^n is such that the limits, $\lim_{n \longrightarrow \infty} \frac{1}{n} \mathbf{E}_s \big\{ \sum_{t=1}^{n} \big(X_t\big)^2 \big\} \in [0,\infty)$, $\lim_{n \longrightarrow \infty} \frac{1}{n} H(Y^n|s) \in [0,\infty)$ exist, and the joint process $(X^n, Y^n), n = 1,2,\ldots$ is asymptotically stationary, which are not accounted for, in [2, Theorem 6.1].
We also give an achievable lower bound on the non-feedback capacity $C^{\infty,nfb}(\kappa,s)$, based on (5.0.20), with $\Lambda_t = 0, \forall t$, i.e., $X_t = Z_t$, $Z^n, n = 1,\ldots$ an independent and identically distributed (IID) sequence, and holds for stable and unstable ARMA$(a,c), a \in (-\infty,\infty), c \in (-\infty,\infty)$ noise.
Step 4. We identify an oversight in the characterization of feedback capacity given in [2, Theorem 6.1, C_{FB}] (i.e., the analog of the limiting expression of $C^{\infty}(\kappa,s)$ without the stabilizability condition), which presupposes a zero variance of the innovations part of the channel input process is optimal. We show that a zero variance of the innovations process, implies the characterization of feedback capacity, based on [2, Theorem 6.1, C_{FB}], is necessarily the zero solution $C_{FB} = 0$, otherwise,C_{FB} does not correspond to the limiting value of (5.0.16).

We structured the chapter as follows.
In Section 5.1, we derive the characterization of the $n-$FTFI capacity, and the lower bound on the characterization of the $n-$FTwFI capacity, for AGN channels driven by the ARMA(a_t, c_t) noise (Section 5.1.1), and present a preliminary elaboration on technical issues that are integral part of capacity definition (5.0.16).
In Section 5.2, we present the derivations of feedback capacity formulas of $C^{\infty}(\kappa,s) = C^{\infty}(\kappa), \forall s$, i.e., (5.0.16), and the achievable lower bounds on the non-feedback capacity $C^{\infty,nfb}(\kappa,s)$, for stable and unstable noise, using the asymptotic analysis of generalized Kalman-filters [8, 9].

5.1 Characterizations of $n-$FTFI and $n-$FTwFI Capacity

In this section we present the following main results.
(1) Theorem 3.1.1 (Section 5.1.1), which gives the characterization of $n-$FTFI capacity for time-varying feedback codes of Definition 5.0.2.(a),
(2) Corollary 5.1.1 (Section 5.1.3), which gives a lower bound on the $n-$FTwFI capacity for time-

varying non-feedback codes of Definition 5.0.2.(b), based on a Markov channel input process without feedback, and follows directly from Theorem 3.1.1.

5.1.1 Characterization of $n-$FTFI Capacity

Below, we introduce the characterization of the $n-$FTFI capacity, for an AGN channel, driven by the time-varying ARMA(a_t,c_t) noise, for the feedback code of Definition 5.0.2.(a). Our presentation, of the next theorem, is based on the degenerate case of the general characterization of the $n-$FTFT capacity of AGN channels, derived in [18]. We should mention that although, [1], treats AGN channels driven by stable noise, some parts of the representation given below can be extracted from the analysis of [1, Section II-V].

Theorem 5.1.1. *Characterization of $n-$FTFI Capacity for AGN Channels Driven by ARMA(a_t,c_t) Noise*

Consider the AGN channel (5.0.1) driven by a time-varying ARMA(a_t,c_t) noise, i.e., (5.0.2), and the code of Definition 5.0.2.(a). Then the following hold.

(a) The optimal time-varying channel input distribution with feedback, for the optimization problem $C_n(\kappa,s)$ defined by (5.0.14), is conditionally Gaussian, of the form

$$\mathbf{P}_{X_t|X_{t-1},Y^{t-1},S_1} = \mathbf{P}_{X_t|S_t,Y^{t-1},S_1}, \qquad t=1,\ldots,n \tag{5.1.21}$$

and it is induced by the time-varying jointly Gaussian channel input process X^n, with a represen-

tation[4]

$$X_t = \Lambda_t\Big(S_t - \widehat{S}_t\Big) + Z_t, \quad t = 1,\ldots,n, \tag{5.1.22}$$

$$X_1 = Z_1, \tag{5.1.23}$$

$$Z_t \in N(0, K_{Z_t}), \quad t = 1,\ldots,n \quad \textit{a Gaussian sequence}, \tag{5.1.24}$$

$$Z_t \quad \textit{independent of} \quad (V^{t-1}, X^{t-1}, Y^{t-1}, S_1), \quad t = 1,\ldots,n, \tag{5.1.25}$$

$$V_t = (c_t - a_t)S_t + W_t, \quad S_1 = s, \quad a_t \in (-\infty,\infty), c_t \in (-\infty,\infty), \quad t = 1,\ldots,n, \tag{5.1.26}$$

$$Y_t = X_t + V_t = \Lambda_t\Big(S_t - \widehat{S}_t\Big) + Z_t + V_t \tag{5.1.27}$$

$$= \Lambda_t\Big(S_t - \widehat{S}_t\Big) + \Big(c_t - a_t\Big)S_t + W_t + Z_t, \tag{5.1.28}$$

$$W_t \in N(0, K_{W_t}) \quad t = 1,\ldots,n \quad \textit{a Gaussian sequence}, \tag{5.1.29}$$

$$Y_1 = Z_1 + \Big(c_1 - a_1\Big)S_1 + W_1, \quad S_1 = s, \tag{5.1.30}$$

$$S_{t+1} = c_t S_t + W_t \quad S_1 = s, \quad t = 2,\ldots,n \tag{5.1.31}$$

$$\frac{1}{n}\mathbf{E}_s\Big\{\sum_{t=1}^{n}(X_t)^2\Big\} = \frac{1}{n}\sum_{t=1}^{n}\Big\{\Big(\Lambda_t\Big)^2 K_t + K_{Z_t}\Big\} \leq \kappa, \tag{5.1.32}$$

$$(\Lambda_t, K_{Z_t}) \in (-\infty,\infty)\times[0,\infty) \quad \textit{scalar-valued, non-random}, \tag{5.1.33}$$

$$\widehat{S}_t \stackrel{\triangle}{=} \mathbf{E}_s\Big\{S_t\Big|Y^{t-1}, S_1 = s\Big\} \tag{5.1.34}$$

$$K_t \stackrel{\triangle}{=} \mathbf{E}_s\Big\{\Big(S_t - \widehat{S}_t\Big)^2\Big\}, \quad K_1 = 0, \quad t = 2,\ldots,n. \tag{5.1.35}$$

Further, $H(Y^n|s) - H(V^n|s)$, $(\widehat{S}_t, K_t), t = 1,\ldots,n$ are determined by the generalized[5] *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 (5.1.26)-(5.1.30) [8, 9]:

$$\widehat{S}_{t+1} = c_t\widehat{S}_t + M_t(K_t, \Lambda_t, K_{Z_t})I_t, \quad \widehat{S}_1 = s, \tag{5.1.36}$$

$$I_t \stackrel{\triangle}{=} Y_t - \mathbf{E}_s\Big\{Y_t\Big|Y^{t-1}\Big\} = Y_t - \Big(c_t - a_t\Big)\widehat{S}_t, \quad I_1 = Z_1 + W_1, \quad t = 1,\ldots,n, \tag{5.1.37}$$

$$= \Big(\Lambda_t + c_t - a_t\Big)\Big(S_t - \widehat{S}_t\Big) + Z_t + W_t, \tag{5.1.38}$$

$$M_t(K_t, \Lambda_t, K_{Z_t}) \stackrel{\triangle}{=} \Big(K_{W_t} + c_t K_t\Big(\Lambda_t + c_t - a_t\Big)\Big)\Big(K_{Z_t} + K_{W_t} + \Big(\Lambda_t + c_t - a_t\Big)^2 K_t\Big)^{-1}, \tag{5.1.39}$$

[4]The fact that $X_1 = Z_1, K_1 = 0, \widehat{S}_1 = s$ is due to the code definition, i.e., $S_1 = s$ is known to the encoder.

[5]Unlike [1], 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.

Generalized Time-Varying Difference Riccati Equation:

$$K_{t+1} = c_t^2 K_t + K_{W_t} - \frac{\Big(K_{W_t} + c_t K_t\Big(\Lambda_t + c_t - a_t\Big)\Big)^2}{\Big(K_{Z_t} + K_{W_t} + \Big(\Lambda_t + c_t - a_t\Big)^2 K_t\Big)}, \quad K_t \geq 0, \ \ K_1 = 0, \quad t = 2, \ldots, n, \tag{5.1.40}$$

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

$$E_{t+1} = F_t(K_t, \Lambda_t, K_{Z_t}) E_t - M_t(K_t, \Lambda_t, K_{Z_t})\Big(Z_t + W_t\Big) + W_t, \ \ E_1 = S_1 - \widehat{S}_1 = 0, \ \ t = 2, \ldots, n. \tag{5.1.41}$$

$$F_t(K_t, \Lambda_t, K_{Z_t}) \stackrel{\triangle}{=} c_t - M_t(K_t, \Lambda_t, K_{Z_t})\Big(\Lambda_t + c_t - a_t\Big) \tag{5.1.42}$$

Entropy of Channel Output Process:

$$H(Y^n|s) = \sum_{t=1}^{n} H(Y_t|Y^{t-1}, s) = \sum_{t=1}^{n} H(Y_t - \mathbf{E}\{Y_t|Y^{t-1}, s\}|Y^{t-1}, s) = \sum_{t=1}^{n} H(I_t|s). \tag{5.1.43}$$

(b) The characterization of the $n-$FTFI capacity $C_n(\kappa, s)$ defined by (5.0.14) is

$$C_n(\kappa, s) \stackrel{\triangle}{=} \sup_{\Big(\Lambda_t, K_{Z_t}\Big), t=1,\ldots,n: \ \frac{1}{n}\sum_{t=1}^{n}\Big\{\Big(\Lambda_t\Big)^2 K_t + K_{Z_t}\Big\} \leq \kappa} \frac{1}{2}\sum_{t=1}^{n} \log\Big(\frac{\big(\Lambda_t + c_t - a_t\big)^2 K_t + K_{Z_t} + K_{W_t}}{K_{W_t}}\Big) \tag{5.1.44}$$

$$\text{subject to: } K_t, t = 1, \ldots, n \text{ satisfies recursion (5.1.40) and } K_{Z_t} \geq 0, t = 1, \ldots, n. \tag{5.1.45}$$

Proof. (a) The information structure (5.1.21) follows, from a degenerate case of [18]. The representation of the jointly Gaussian process X^n, defined by (5.1.22), such that Z^n satisfies (5.1.24) and (2.1.26), is also a degenerate case of [18], where the channel is more general, of the form $Y_t = C_{t-1}Y_{t-1} + D_t X_t + D_{t,t-1}X_{t-1} + V_t$, $V_t = F_t V_{t-1} + W_t$, where $(C_{t,t-1}, D_t, D_{t,t-1}, F_t)$ are non-random, i.e., with past dependence on channel inputs and outputs. Expressions (5.1.26)-(5.1.35) follow directly from (5.1.22), and the channel definition. The generalized Kalman-filter equations follow from standard texbooks, i.e., [9]. (5.1.43) follows from the independent property of the innovations process. (b) Follows from (5.0.14), (5.1.43), $H(V^n|s) = \sum_{t=1} H(W_t)$, and part (a). □

Remark 5.1.1. *By the definition of the innovations process and entropy, (5.1.38) and (5.1.43), it follows that whether the limit exists, $\lim_{n\longrightarrow\infty} \frac{1}{n}\big\{H(Y^n|s) - H(V^n|s)\big\} = \lim_{n\longrightarrow\infty} \frac{1}{n}\sum_{t=1}^{n}\big\{H(I_t) - H(W_t)\big\} \in [0,\infty)$ is determined from the limiting covariance of the innovations process I^n and noise W^n. Similarly, for $\lim_{n\longrightarrow\infty} \frac{1}{n}\mathbf{E}_s\Big\{\sum_{t=1}^{n}\big(X_t\big)^2\Big\} \in [0,\infty)$ by (5.1.32).*

5.1.2 Converse Coding Theorem for AGN Channels

By Theorem 3.1.1, the characterization of $n-$FTFI capacity, $C_n(\kappa,s)$, is expressed in terms of the mean-square error $K_t, t=1,\ldots,n$, that satisfies the time-varying generalized RDE (5.1.40). We recall the error recursion of the generalized Kalman-filter given by (5.1.41). Note that recursion (5.1.41) is linear time-varying, hence its convergence properties, in mean-square sense, i.e., $K_t = \mathbf{E}_s\{(E_t)^2\}$ are determined by the properties of $F_t(K_t,\Lambda_t,K_{Z_t})$ and $M_t(K_t,\Lambda_t,K_{Z_t}),\Lambda_t,K_{Z_t},t=1,2,\ldots$. Hence, in general, $\lim_{n\longrightarrow\infty} K_n = \lim_{n\longrightarrow\infty}\mathbf{E}_s\{(E_n)^2\}$ does not converge, for arbitrary $F_t(K_t,\Lambda_t,K_{Z_t})$ and $M_t(K_t,\Lambda_t,K_{Z_t}),\Lambda_t,K_{Z_t},t=1,2,\ldots$. In view of the error recursion (5.1.41), we have the following theorem.

Theorem 5.1.2. *Converse coding theorem*
Consider the feedback code $\mathscr{C}^{fb}_{\mathbb{Z}^+}$ of Definition 5.0.2.(a).
Converse Coding Theorem. If there exists a feedback code $\mathscr{C}^{fb}_{\mathbb{Z}^+}$, i.e., with $\varepsilon_n \longrightarrow 0$, as $n\longrightarrow\infty$, then the code rate $R(s)$ satisfies:

$$R(s) \leq C(\kappa,s) \stackrel{\triangle}{=} \lim_{n\longrightarrow\infty}\frac{1}{n}C_n(\kappa,s), \quad C_n(\kappa,s) \ \textit{defined in Theorem 3.1.1.(b)} \tag{5.1.46}$$

provided the following conditions hold:
(C1) the maximizing element, denoted by $(\Lambda_t^*,K^*_{Z_t}),t=1,\ldots,n$ which satisfies the average power constraint exists, and
(C2) the limit exists and it is finite.

Proof. Conditions (C1) and (C2) follow from the above discussion, and the converse coding theorem follows from [3]. □

Remark 5.1.2. *By the average power (5.1.32) and optimization problem (5.1.44), it is necessary to identify sufficient and/or necessary conditions such that the maximizing element, $(\Lambda_t^*,K^*_{Z_t}),t=1,\ldots,n$, exists in the set, and to ensure convergence of $K_n=\mathbf{E}_s\{(E_n)^2\}$ (that satisfies the time-varying DRE (5.1.40)), as $n\longrightarrow\infty$, to a finite number, such that the limit in (5.1.46) is finite. However, to ensure $C(\kappa,s)$ is independent of s, it is necessary that the limit is also independent of s. On the other hand, if the limit $C(\kappa,s)$ depends on $S_1=s$, then one needs to consider a formulation based on compound capacity, by taking infimum over all initial states $S_1=s$, as done, for example, in [19], for finite state feedback channels, otherwise different s give rise to different rates.*

5.1.3 Lower Bound on Characterization of $n-$FTwFI Capacity

Next, we give a lower bound on the characterization of $n-$FTwFI Capacity, for the non-feedback code of Definition 5.0.2.(b), which follows directly from Theorem 3.1.1.

Corollary 5.1.1. *Lower bound on characterization of $n-$FTwFI Capacity for AGN Channels Driven by ARMA(a_t,c_t) Noise*

Consider the AGN channel (5.0.1) driven by a time-varying ARMA(a_t,c_t) noise, i.e., (5.0.2), and the code without feedback, of Definition 1.1.1.(b). Define the information theoretic optimization problem of capacity without feedback, i.e., the analog of (5.0.14), by

$$C_n^{nfb}(\kappa,s) \stackrel{\triangle}{=} \sup_{\mathbf{P}_{X_t|X^{t-1},S_1},t=1,\ldots,n:\ \frac{1}{n}\mathbf{E}_s\left\{\sum_{t=1}^n \left(X_t\right)^2\right\}\leq\kappa} H(Y^n|s)-H(V^n|s) \tag{5.1.47}$$

provided the supremum exists. Then the following hold.

(a) A lower bound on $C_n^{nfb}(\kappa,s)$ is obtained by the conditionally Gaussian, time-varying channel input distribution without feedback, given by

$$\mathbf{P}_{X_t|X^{t-1},S_1} = \mathbf{P}_{X_t|X_{t-1},S_1}, \quad t=1,\ldots,n \tag{5.1.48}$$

which is induced by the time-varying jointly Gaussian channel input process X^n, with a represen-

tation

$$X_t = \Lambda_t X_{t-1} + Z_t, \quad t = 2,\ldots,n, \tag{5.1.49}$$

$$X_1 = Z_1, \tag{5.1.50}$$

$$Z_t \in N(0, K_{Z_t}), \ \ t = 1,\ldots,n \ \ \textit{a Gaussian sequence}, \tag{5.1.51}$$

$$Z_t \ \ \textit{independent of} \ \ (V^{t-1}, X^{t-1}, Y^{t-1}, S_1), \ \ t = 1,\ldots,n, \tag{5.1.52}$$

$$V_t = \Big(c_t - a_t\Big)S_t + W_t, \quad S_1 = s, \quad a_t \in (-\infty,\infty), c_t \in (-\infty,\infty), \ \ t = 2,\ldots,n, \tag{5.1.53}$$

$$Y_t = X_t + V_t = \Big(\Lambda_t\Big)X_{t-1} + \Big(c_t - a_t\Big)S_t + W_t + Z_t, \ \ t = 1,\ldots,n, \tag{5.1.54}$$

$$Y_1 = Z_1 + \Big(c_1 - a_1\Big)S_1 + W_1, \ \ S_1 = s, \tag{5.1.55}$$

Define a new state, $\overline{S_t} \stackrel{\triangle}{=} \begin{pmatrix} S_t \\ X_{t-1} \end{pmatrix}$*.Then,*

$$\overline{S}_{t+1} = \overline{A}_t \overline{S}_t + \overline{B}_t \overline{W}_t, \ \ \textit{where}, \tag{5.1.56}$$

$$\overline{A}_t \stackrel{\triangle}{=} \begin{pmatrix} c_t & 0 \\ 0 & \Lambda_t \end{pmatrix}, \ \ \overline{S}_t \stackrel{\triangle}{=} \begin{pmatrix} S_t \\ X_{t-1} \end{pmatrix}, \ \ \overline{B}_t \stackrel{\triangle}{=} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \textit{and} \ \ \overline{W}_t \stackrel{\triangle}{=} \begin{pmatrix} W_t \\ Z_t \end{pmatrix}. \tag{5.1.57}$$

$$Y_t = \overline{C}_t \overline{S}_t + \overline{N}_t \overline{W}_t, \ \ \textit{where}, \tag{5.1.58}$$

$$\overline{C}_t \stackrel{\triangle}{=} \begin{pmatrix} c_t - a_t & \Lambda_t \end{pmatrix} \ \ \textit{and}\, \overline{N}_t \stackrel{\triangle}{=} \begin{pmatrix} 1 & 1 \end{pmatrix}. \tag{5.1.59}$$

$$\frac{1}{n}\mathbf{E}_s\Big\{\sum_{t=1}^{n}(X_t)^2\Big\} = \frac{1}{n}\sum_{t=1}^{n}\Big\{\Big(\Lambda_t\Big)^2 K_{X_{t-1}} + K_{Z_t}\Big\} \le \kappa, \tag{5.1.60}$$

$$(\Lambda_t, K_{Z_t}) \in (-\infty,\infty)\times[0,\infty) \quad \textit{scalar-valued, non-random}, \tag{5.1.61}$$

$$K_{X_t} \stackrel{\triangle}{=} \mathbf{E}_s\Big(X_t\Big)^2, \tag{5.1.62}$$

$$\widehat{\overline{S}}_{t+1} \stackrel{\triangle}{=} \mathbf{E}_s\Big\{\overline{S}_{t+1}\Big|Y^t\Big\}, \tag{5.1.63}$$

$$K_t \stackrel{\triangle}{=} \mathbf{E}_s\left\{\Big(\overline{S}_t - \widehat{\overline{S}}_t\Big)\Big(\overline{S}_t - \widehat{\overline{S}}_t\Big)^{\mathrm{T}}\right\}, \quad t = 1,\ldots,n. \tag{5.1.64}$$

Further, $(\overline{\widehat{S}}_t, K_t), t = 1,\ldots,n$ *are determined by the generalized time-varying Kalman-filter and generalized time-varying difference Riccati equation (DRE), of estimating* $\overline{S}^n$ *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 (5.1.49)-(5.1.59) [8, 9]:

$$\widehat{\overline{S}}_{t+1} = \overline{A}_t \widehat{\overline{S}}_t + M_t^{nfb}(K_t) I_t, \quad \widehat{\overline{S}}_1 = \widehat{\overline{s}}_1, \quad t = 2, \ldots, n \tag{5.1.65}$$

$$I_t \stackrel{\triangle}{=} Y_t - \mathbf{E}_s\Big\{Y_t \Big| Y^{t-1}\Big\} \tag{5.1.66}$$

$$= \overline{C}_t \overline{S}_t + \overline{N}_t \overline{W}_t - \overline{C}_t \mathbf{E}_s\Big\{\overline{S}_t \Big| Y^{t-1}\Big\} \tag{5.1.67}$$

$$= \overline{C}_t\Big(\overline{S}_t - \widehat{\overline{S}}_t\Big) + \overline{N}_t \overline{W}_t, \quad I_1 = Z_1 + W_1, \quad t = 1, \ldots, n, \tag{5.1.68}$$

$$M_t^{nfb}(K_t) \stackrel{\triangle}{=} \Big(\overline{A}_t K_t \overline{C}_t^{\mathrm{T}} + \overline{B}_t \overline{K}_{W_t} \overline{N}_t^{\mathrm{T}}\Big)\Big(\overline{N}_t \overline{K}_{W_t} \overline{N}_t^{\mathrm{T}} + \overline{C}_t K_t \overline{C}_t^{\mathrm{T}}\Big)^{-1}, \quad \textit{where}, \tag{5.1.69}$$

$$\overline{K}_{W_t} \stackrel{\triangle}{=} \big(K_{W_t} \quad K_{Z_t}\big). \tag{5.1.70}$$

$$I_t, \quad t = 1, \ldots, n, \qquad \textit{an orthogonal innovations process.} \tag{5.1.71}$$

Generalized Time-Varying Difference Riccati Equation:

$$K_{t+1} = \overline{A}_t K_t \overline{A}_t^{\mathrm{T}} + \overline{B}_t \overline{K}_{W_t} \overline{B}_t^{\mathrm{T}} - \Big(\overline{A}_t K_t \overline{C}_t^{\mathrm{T}} + \overline{B}_t \overline{K}_{W_t} \overline{N}_t^{\mathrm{T}}\Big)\Big(\overline{C}_t K_t \overline{C}_t^{\mathrm{T}} + \overline{N}_t \overline{K}_{W_t} \overline{N}_t^{\mathrm{T}}\Big)^{-1}\Big(\overline{A}_t K_t \overline{C}_t^{\mathrm{T}} + \overline{B}_t \overline{K}_{W_t} \overline{N}_t^{\mathrm{T}}\Big)^{\mathrm{T}}, \tag{5.1.72}$$

$$K_t \geq 0, \quad K_1 = 0, \qquad t = 1, \ldots, n, \tag{5.1.73}$$

Time-Varying Difference Lyapunov Equation:

$$K_{X_t} = \Lambda_t^2 K_{X_{t-1}} + K_{Z_t}, \qquad K_{X_t} \geq 0, \quad K_{X_0} = 0, \qquad t = 1, \ldots, n, \tag{5.1.74}$$

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

$$\begin{aligned} E_{t+1}^{nfb} &= \overline{S}_{t+1} - \widehat{\overline{S}}_{t+1} \\ &= \overline{A}_t \overline{S}_t + \overline{B}_t \overline{W}_t - \overline{A}_t \widehat{\overline{S}}_t - M(K_t) I_t \\ &= \overline{A}_t \overline{S}_t + \overline{B}_t \overline{W}_t - \overline{A}_t \widehat{\overline{S}}_t - M(K_t)\Big[\overline{C}_t\Big(\overline{S}_t - \widehat{\overline{S}}_t\Big) + \overline{N}_t \overline{W}_t\Big] \\ &\Big(\overline{A}_t - M(K_t)\overline{C}_t\Big)\Big(\overline{S}_t - \widehat{\overline{S}}_t\Big) + \Big(\overline{B}_t - M(K_t)\overline{N}_t\Big)\overline{W}_t, \end{aligned} \tag{5.1.75}$$

$$E_1 = \overline{S}_1 - \widehat{\overline{S}}_1 = 0, \qquad t = 1, \ldots, n. \tag{5.1.76}$$

(b) The lower bound characterization of the $n-$*FTwFI capacity* $C_n^{nfb}(\kappa, s)$*, defined by (5.1.47), is*

$$C_n^{nfb}(\kappa, s) \geq C_{n,LB}^{nfb}(\kappa, s) \stackrel{\triangle}{=} \sup_{\left(\Lambda_t, K_{Z_t}\right), t=1,\ldots,n: \ \frac{1}{n}\sum_{t=1}^n \left\{\left(\Lambda_t\right)^2 K_{X_{t-1}} + K_{Z_t}\right\} \leq \kappa} \frac{1}{2}\sum_{t=1}^n \log\Big(\frac{\overline{C}_t K_t \overline{C}_t^{\mathrm{T}} + \overline{N}_t \overline{K}_{W_t} \overline{N}_t^{\mathrm{T}}}{\overline{N}_t K_{W_t} \overline{N}_t^{\mathrm{T}}}\Big) \tag{5.1.77}$$

subject to: $K_t, K_{X_t}, t = 1, \ldots, n$ *satisfy recursions (5.1.73), (5.1.74), and* $K_{Z_t} \geq 0, t = 1, \ldots, n.$ (5.1.78)

Proof. (a) Similar to the feedback capacity of Theorem 5.1.1, by the maximum entropy of Gaussian distributions, the maximizing distributions $\mathbf{P}_{X_t|X^{t-1},S_1}, t=1,\dots,n$ for the optimization problem (5.1.47) are conditionally Gaussian, such that (X^n,Y^n) for $S_1=s$, is jointly Gaussian, the average power constraint is satisfied, and condition (5.0.9) is respected. Clearly, the restriction to distributions that satisfy (5.1.48) result in a lower bound on $C_n^{nfb}(\kappa,s)$ defined by (5.1.47). Note that the restriction to (5.1.48) is precisely the restriction of feedback distributions (5.1.21) to non-feedback distributions. The rest of the equations follow, similarly to Theorem 3.1.1.(a), if the channel is used without feedback, i.e., $X_t=\Lambda_t X_{t-1}+Z_t$. The rest of the expression of part (a) are obtained as in Theorem 3.1.1.(a). (b) The derivation follows from the expressions of part (a). □

Remark 5.1.3. *Corollary 5.1.1 is useful, because the lower bound is much easier to compute, compared to $C_n^{nfb}(\kappa,s)$, defined by (5.1.47), where the supremum is taken over all jointly Gaussian channel input processes $X^n, n=1,2,\dots,$ without feedback or distributions without feedback,* $\mathbf{P}_{X_t|X^{t-1},S_1}, t=1,2,\dots$.

5.2 New Formulas of Feedback Capacity

In this section we derive a closed form formulae for feedback capacity $C^{\infty}(\kappa,s)$, defined by (5.0.16), and lower bounds on capacity without feedback $C^{\infty,nfb}(\kappa,s)$, defined by (5.0.17), of AGN channels driven by ARMA(a,c), stable and unstable noise, when channel input strategies or distributions are time-invariant. This section includes material on basic properties of generalized DREs, AREs, and definitions and implications of the notions of detectability and stabilizability.

5.2.1 Characterization of Feedback Capacity for Time-Invariant Channel Input Distributions

We restrict the class of channel input distributions of Theorem 3.1.1 to the class of time-invariant distributions. We note that our restriction is weaker than the analysis in [2], which presupposes stationarity or asymptotic joint stationarity of the joint Gaussian process $(X^n,Y^n), n=1,2,\dots$ (the author also considers a double sided joint process). However, unlike [1, 2], we do not assume the ARMA(a,c) noise is stable.

By Theorem 3.1.1, and restricting the channel input strategies to the time-invariant channel input strategies, $(\Lambda_t, K_{Z_t}) = (\Lambda^\infty, K_Z^\infty), t=1,\ldots,n$, with corresponding $X_t = X_t^o$, $\widehat{S}_t = \widehat{S}_t^o$, $Y_t = Y_t^o$, $I_t = I_t^o$, $E_t = E_t^o$, $K_t = K_t^o$ (not necessarily stationary), then we have the following representation[6].

$$X_t^o = \Lambda^\infty\Big(S_t - \widehat{S}_t^o\Big) + Z_t^o, \quad X_1^o = Z_1^o, \qquad t=2,\ldots,n, \tag{5.2.79}$$

$$V_t = \Big(c-a\Big)S_t + W_t, \quad S_1 = s, \quad t=2,\ldots,n, \tag{5.2.80}$$

$$Y_t^o = X_t^o + V_t = \Lambda^\infty\Big(S_t - \widehat{S}_t^o\Big) + \Big(c-a\Big)S_t + W_t + Z_t^o, \quad t=1,\ldots,n, \tag{5.2.81}$$

$$Y_1^o = Z_1^o + \Big(c-a\Big)S_1 + W_1, \quad S_1 = s, \tag{5.2.82}$$

$$Z_t^o \sim N(0, K_Z^\infty), \quad t=1,\ldots,n \ \text{ is a Gaussian sequence,} \tag{5.2.83}$$

$$Z_t^o \ \text{ is independent of } \ (V^{t-1}, X^{o,t-1}, Y^{o,t-1}, S_1), \quad t=1,\ldots,n, \tag{5.2.84}$$

$$cov(\begin{bmatrix} W_t \\ W_t + Z_t^o \end{bmatrix}, \begin{bmatrix} W_t \\ W_t + Z_t^o \end{bmatrix}^T) = \begin{bmatrix} K_W & K_W \\ K_W & K_W + K_Z^\infty \end{bmatrix}, \tag{5.2.85}$$

$$\frac{1}{n}\mathbf{E}_s\Big\{\sum_{t=1}^n \big(X_t^o\big)^2\Big\} = \frac{1}{n}\sum_{t=1}^n \big(\Lambda^\infty\big)^2 K_t^o + K_Z^\infty \le \kappa, \tag{5.2.86}$$

$$(\Lambda^\infty, K_Z^\infty) \in (-\infty,\infty)\times[0,\infty) \qquad \text{are non-random,} \tag{5.2.87}$$

$$\mathbf{P}_{X_t^o|S_t, Y^{o,t-1}, S_1} = \mathbf{P}^\infty(dx_t|S_t, y^{t-1}, s), \quad t=1,\ldots,n, \quad \text{that is, the distribution is time-invariant} \tag{5.2.88}$$

where $(\widehat{S}_t^o, K_t^o), t=1,\ldots,n$ satisfy the generalized Kalman-filter and time-invariant DRE, given below.

Generalized Kalman-filter Recursion:

$$\widehat{S}_{t+1}^o = c\widehat{S}_t^o + M(K_t^o, \Lambda^\infty, K_Z^\infty)I_t^o, \quad \widehat{S}_1^o = s, \tag{5.2.89}$$

$$I_t^o \stackrel{\triangle}{=} Y_t^o - \Big(c-a\Big)\widehat{S}_t^o, \quad I_1^o = Z_1^o + W_1, \quad t=1,\ldots,n, \tag{5.2.90}$$

$$= \Big(\Lambda^\infty + c - a\Big)\Big(S_t - \widehat{S}_t^o\Big) + Z_t^o + W_t, \tag{5.2.91}$$

$$M(K_t^o, \Lambda^\infty, K_Z^\infty) \stackrel{\triangle}{=} \Big(K_W + cK_t^o\Big(\Lambda^\infty + c - a\Big)\Big)\Big(K_Z^\infty + K_W + \Big(\Lambda^\infty + c - a\Big)^2 K_t^o\Big)^{-1}, \tag{5.2.92}$$

$$I_t^o, \quad t=1,\ldots,n, \qquad \text{an orthogonal innovations process.} \tag{5.2.93}$$

Generalized Time-Invariant Difference Riccati Equation:

$$K_{t+1}^o = c^2K_t^o + K_W - \frac{\Big(K_W + cK_t^o\Big(\Lambda^\infty + c - a\Big)\Big)^2}{\Big(K_Z^\infty + K_W + \Big(\Lambda^\infty + c - a\Big)^2 K_t^o\Big)}, \qquad K_t^o \ge 0, \quad K_1^o = 0, \qquad t=1,\ldots,n, \tag{5.2.94}$$

[6]The variation of notation is judged necessary to distinguish it from the time-varying channel input strategies (Λ_t, K_{Z_t}) and corresponding distributions $\mathbf{P}_{X_t|X^{t-1}, Y^{t-1}, S_1} = \mathbf{P}_t(dx_t|x^{t-1}, y^{t-1}, s), t=1,\ldots,n$.

We should emphasize that the Kalman-filter recursion (5.2.89) is time-varying, but the DRE (5.2.94) is time-invariant.

Then the analog of the error recursion (5.1.41), for time-invariant strategies, is the following.

Error Recursion of the Generalized Kalman-filter, $E_t^o \stackrel{\triangle}{=} S_t - \widehat{S}_t^o, t = 1,\ldots,n$:

$$E_{t+1}^o = F(K_t^o,\Lambda^\infty,K_Z^\infty)E_t^o - M(K_t^o,\Lambda^\infty,K_Z^\infty)\Big(Z_t^o + W_t\Big) + W_t,\ E_1^o = 0,\ t = 1,\ldots,n, \tag{5.2.95}$$

$$Z_t^o \in N(0,K_Z^\infty),\ t = 1,2,\ldots,n. \tag{5.2.96}$$

$$F(K_t^o,\Lambda^\infty,K_Z^\infty) \stackrel{\triangle}{=} c - M(K_t^o,\Lambda^\infty,K_Z^\infty)\Big(\Lambda^\infty + c - a\Big), \tag{5.2.97}$$

Note that recursion (5.2.95) is linear time-varying. Hence, $\lim_{n\longrightarrow\infty} K_n^o = \lim_{n\longrightarrow\infty} \mathbf{E}_s\big\{\big(E_n^o\big)^2\big\}$ is not expected to exists, and to be bounded, for arbitrary $(F(K_t^o,\Lambda^\infty,K_Z^\infty), M(K_t^o,\Lambda^\infty,K_Z^\infty)), t = 1,2,\ldots$. Indeed, the convergence properties of the sequence $K_1^o, K_2^o, \ldots, K_n^o$ generated by (5.2.94), as $n \longrightarrow \infty$, are characterized by the detectability and stabilizability conditions [8, 9] (which we introduce shortly). These conditions ensure existence of a finite, unique nonnegative limit, $\lim_{n\longrightarrow\infty} K_n^o = K^\infty$, such that the stability property holds: $\lim_{n\longrightarrow\infty} F(K_n^o,\Lambda^\infty,K_Z^\infty) = F(K^\infty,\Lambda^\infty,K_Z^\infty) \in (-1,1)$, and moreover that $K^\infty \geq 0$ is the unique solution of a generalized ARE.

Next, we define the characterization of the $n-$FTFI capacity, its per unit time limit, and the alternative definition, with the per unit time limit and maximization interchanged.

Definition 5.2.1. *Characterizations of asymptotic limits*

Consider the characterization of the $n-$FTFI capacity of Theorem 3.1.1, restricted to the time-invariant strategies $(\Lambda_t = \Lambda^\infty, K_{Z_t} = K^\infty), t = 1,\ldots,n$, as defined by (5.2.79)-(5.2.94).

(a) The characterization of the $n-$FTFI capacity for time-invariant strategies is defined by

$$C_n^o(\kappa,s) \stackrel{\triangle}{=} \sup_{(\Lambda^\infty,K_Z^\infty):\ \frac{1}{n}\sum_{t=1}^n \left(\Lambda^\infty\right)^2 K_t^o + K_Z^\infty \leq \kappa} \sum_{t=1}^n \frac{1}{2}\log\Big(\frac{\left(\Lambda^\infty + c - a\right)^2 K_t^o + K_Z^\infty + K_W}{K_W}\Big) \tag{5.2.98}$$

$$\text{subject to: } K_t^o, t = 1,\ldots,n \text{ satisfies recursion (5.2.94) and } K_Z^\infty \geq 0, t = 1,\ldots,n \tag{5.2.99}$$

provided the supremum exists in the set. The per unit time-limit is then defined by

$$C^o(\kappa,s) \stackrel{\triangle}{=} \lim_{n\longrightarrow\infty} \frac{1}{n} C_n^o(\kappa,s). \tag{5.2.100}$$

provided the supremum exists and the limit exists and it is finite.

(b) The characterization of the $n-$FTFI capacity for time-invariant strategies, with limit and max-

imization interchanged is defined by

$$C^{\infty}(\kappa,s) \triangleq \sup_{\left(\Lambda^{\infty},K_Z^{\infty}\right):\ \lim_{n\longrightarrow\infty}\frac{1}{n}\sum_{t=1}^{n}\left(\Lambda^{\infty}\right)^2 K_t^o+K_Z^{\infty}\leq\kappa} \lim_{n\longrightarrow\infty}\frac{1}{n}\sum_{t=1}^{n}\frac{1}{2}\log\Big(\frac{\left(\Lambda^{\infty}+c-a\right)^2 K_t^o+K_Z^{\infty}+K_W}{K_W}\Big) \tag{5.2.101}$$

provide the limit exists and it is finite and the supremum also exists in the set.

To ensure $C^{\infty}(\kappa,s)$ defined by (5.2.101) is well defined, i.e., that the optimal time-invariant channel input strategy or distribution ensures the limit exits, it is finite, and $C^{\infty}(\kappa,s)$ is independent of s, we shall impose condition **(C1)**. We shall express condition **(C1)** in terms of properties of generalized time-invariant DREs and AREs, introduced in the next section. That is, we shall address Problem 5.0.1.

5.2.2 Convergence Properties of Time-Invariant Generalized DREs

We recall that in the study of mean-square estimation, and in particular, the filtering theory, of time-invariant jointly Gaussian processes described by linear recursions, driven by Gaussian noise processes, and of jointly stationary Gaussian processes, the concepts of detectability and stabilizability, have been very effective [8, 9]. In this section, we summarize these concepts in relation to the properties of generalized DREs and AREs. It is then obvious how these concepts generalize to AGN channels driven by Gaussian noise with limited memory.

Let $\{K_t, t = 1,2,\ldots,n\}$ denote a sequence that satisfies the time-invariant generalized DRE with arbitrary initial condition

$$K_{t+1} = c^2 K_t + K_W - \frac{\Big(K_W + cK_t\Big(\Lambda + c - a\Big)\Big)^2}{\Big(K_Z + K_W + \Big(\Lambda + c - a\Big)^2 K_t\Big)}, \quad K_1 = \text{given}, \quad t = 1,\ldots,n. \tag{5.2.102}$$

We note that a solution of (5.2.102) is a functional of the parameters of the right hand side, that is, $K_t \equiv K_t(a,c,K_W,\Lambda,K_Z,K_1), t = 1,\ldots,n$. To discuss the properties of the generalized DRE (5.2.102), we introduce, as often done in the analysis of generalized DREs [9] and [8, Section 14.7, page 540], the following definitions.

$$A \triangleq c, \quad C \triangleq \Lambda + c - a, \quad A^* \triangleq c - K_W R^{-1} C, \quad B^{*,\frac{1}{2}} \triangleq K_W^{\frac{1}{2}} B^{\frac{1}{2}} \tag{5.2.103}$$

$$R \triangleq K_Z + K_W, \quad B \triangleq 1 - K_W\left(K_Z + K_W\right)^{-1}. \tag{5.2.104}$$

By (5.2.92) and (5.2.97), we also have

$$M(K,\Lambda,K_Z) \triangleq \Big(K_W + AKC\Big)\Big(R + (C)^2 K\Big)^{-1}, \tag{5.2.105}$$

$$F(K,\Lambda,K_Z) = A - M(K,\Lambda,K_Z)C. \tag{5.2.106}$$

The generalized algebraic Riccati equation (ARE) corresponding to (5.2.102) is

$$K = c^2 K + K_W - \frac{\Big(K_W + cK\Big(\Lambda + c - a\Big)\Big)^2}{\Big(K_Z + K_W + \Big(\Lambda + c - a\Big)^2 K\Big)}, \quad K \geq 0. \tag{5.2.107}$$

Next, we introduce the definition of asymptotic stability of the error recursion (5.2.95).

Definition 5.2.2. *Asymptotic stability*
A solution $K \geq 0$ to the generalized ARE (5.2.107), assuming it exists, is called stabilizing if $|F(K,\Lambda,K_Z)| < 1$. In this case, we say $F(K,\Lambda,K_Z)$ is asymptotically stable, that is, $|F(K,\Lambda,K_Z)| < 1$.

With respect to any of the above generalized DRE and ARE, we define the important notions of detectability, unit circle controllability, and stabilizability.

Definition 5.2.3. *Detectability, Stabilizability, Unit Circle controllability*
(a) The pair $\Big\{A, C\Big\}$ is called detectable if there exists a $G \in \mathbb{R}$ such that $|A - GC| < 1$ (stable).
(b) The pair $\Big\{A^*, B^{*,\frac{1}{2}}\Big\}$ 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 $\Big\{A^*, B^{*,\frac{1}{2}}\Big\}$ is called stabilizable if there exists a $G \in \mathbb{R}$ such that $|A^* - B^{*,\frac{1}{2}}G| < 1$.

The next theorem characterizes, detectability, unit circle controllability, and stabilizability [8, 20].

Lemma 5.2.1. *[8, 20] Necessary and sufficient conditions for detectability, unit circle controllability, stabilizability*
(a) The pair $\Big\{A, C\Big\}$ is detectable if and only if there exists no eigenvalue, eigenvector $\{\lambda, x\}$, $Ax = \lambda x$ such that $|\lambda| > 1$, and such that $Cx = 0$
(b) The pair $\Big\{A^*, B^{*,\frac{1}{2}}\Big\}$ is unit circle uncontrollable if and only if there exists no eigenvalue, eigenvector $\{\lambda, x\}$, $xA^* = \lambda x$ such that $|\lambda| = 1$, and such that $B^{*,\frac{1}{2}}x = 0$.

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

In the next theorem we summarize known results on sufficient and/or 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 nonnegative K, which is the unique stabilizing solution of a corresponding generalized ARE.

Theorem 5.2.1. *[8, 9] Convergence of time-invariant generalized DRE*
Let $\{K_t, t = 1, 2, \ldots, n\}$ *denote a sequence that satisfies the time-invariant generalized DRE (5.2.102) with arbitrary initial condition.*
Then the following hold.
(1) Consider the generalized RDE (5.2.102) with zero initial condition, i.e., $K_1 = 0$*, and assume, the pair* $\left\{A, C\right\}$ *is detectable, and the pair* $\left\{A^*, B^{*,\frac{1}{2}}\right\}$ *is unit circle controllable.*
Then the sequence $\{K_t : t = 1, 2, \ldots, n\}$ *that satisfies the generalized DRE (5.2.102), with zero initial condition* $K_1 = 0$*, converges to* K*, i.e.,* $\lim_{n \longrightarrow \infty} K_n = K$*, where* K *satisfies the ARE*

$$K = c^2 K + K_W - \frac{\Big(K_W + cK\big(\Lambda + c - a\big)\Big)^2}{\Big(K_Z + K_W + \big(\Lambda + c - a\big)^2 K\Big)} \tag{5.2.108}$$

if and only if the pair $\left\{A^*, B^{*,\frac{1}{2}}\right\}$ *is stabilizable.*
(2) Assume, the pair $\left\{A, C\right\}$ *is detectable, and the pair* $\left\{A^*, B^{*,\frac{1}{2}}\right\}$ *is unit circle controllable. Then there exists a unique stabilizing solution* $K \geq 0$ *to the generalized ARE (5.2.102), i.e., such that,* $|F(K, \Lambda, K_Z)| < 1$*, 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* $K_t, t = 1, 2, \ldots, n$ *to the generalized RDE (5.2.102) with arbitrary initial condition,* K_1 *is such that* $\lim_{n \longrightarrow \infty} K_n = K$*, where* $K \geq 0$ *is the unique solution of the generalized ARE (5.2.102) with* $|F(K, \Lambda, K_Z)| < 1$*, i.e., it is stabilizing.*

We should mentioned that Theorem 5.2.1.(1) follows by combining [8, Lemma 14.2.1, page 507] of classical DREs and AREs with [8, Section 14.7] of generalized DREs and AREs. Theorem 5.2.1.(2) is given in [8, Theorem E.6.1, page 784]. Theorem 5.2.1.(3) is obtained from [9, Theorem 4.2, page 164], and also [8].

From the properties of generalized DREs and AREs of Theorem 5.2.1, we have the next lemma.

Lemma 5.2.2. *Properties of Solutions of DREs and AREs for different cases*
Consider the definitions (5.2.103) ,(5.2.104).
(1) Suppose $c \in (-1,1)$. Then the pair $\{A, C\}$ is detectable.
(2) Suppose $K_Z = 0$. Then the pair $\{A^*, B^{*,\frac{1}{2}}\}$ is unit circle controllable if and only if $|\Lambda - a| \neq 1$.
(3) Suppose $K_Z = 0$. Then the pair $\{A^*, B^{*,\frac{1}{2}}\}$ is stabilizable if and only if $|\Lambda - a| < 1$.
(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, i.e.,

$$K_{t+1} = c^2 K_t + K_W - \frac{\Big(K_W + cK_t\big(\Lambda + c - a\big)\Big)^2}{\Big(K_W + \big(\Lambda + c - a\big)^2 K_t\Big)}, \quad K_1 = 0, \quad t = 1, \ldots, n \tag{5.2.109}$$

converges to K, i.e., $\lim_{n \longrightarrow \infty} K_n = K$, where K satisfies the generalized ARE (5.2.107) if and only if the $\{A^*, B^{*,\frac{1}{2}}\}$ is stabilizable, equivalently, $|\Lambda - a| < 1$.
(5) Suppose $K_Z = 0$, and $|\Lambda - a| \neq 1$, with the corresponding ARE,

$$K = c^2 K + K_W - \frac{\Big(K_W + cK\big(\Lambda + c - a\big)\Big)^2}{\Big(K_W + \big(\Lambda + c - a\big)^2 K\Big)}. \tag{5.2.110}$$

Then, the two solution, are given by

$$K = 0, \quad K = \frac{K_W\Big(\big(\Lambda - a\big)^2 - 1\Big)}{\Big(\Lambda + c - a\Big)^2}, \;\; \Lambda + c - a \neq 0 \tag{5.2.111}$$

Moreover, $K = 0$ is the unique and stabilizing solution $K \geq 0$ to (5.2.110), i.e., such that $|F(K, \Lambda, K_Z)| < 1$, if and only if $|\Lambda - a| < 1$.

Proof. See Appendix 7.10. □

In the next remark we make some comments on [2, Theorem 6.1], i.e., that a zero variance of the innovations process is not the optimal value.

Remark 5.2.1. *Asymptotic stationarity of optimal process of [2]*
Consider the characterization of feedback capacity given in [2, Theorem 6.1, Σ satisfying eqn(61)], in which the variance of the innovations process is replaced by a zero value (see comment below [2, Theorem 6.1]). Then $\Sigma = 0$ is one solution of [2, Σ satisfying eqn(61)].

We ask: what are necessary and/or sufficient conditions for convergence $\lim_{n\longleftrightarrow}\Sigma_n=\Sigma$, *where* $\Sigma\geq 0$ *is the unique limit that stabilizes the estimation error of the noise?*

By the multidimensional version of Theorem 5.2.1.(1), and Lemma 5.2.2.(3), then the limit $\lim_{n\longrightarrow\infty}\Sigma_n$ *converges if and only if the stabilizability condition holds. For the ARMA(a,c) noise model, since the characterization of feedback capacity given [2, Theorem 6.1], presupposes a zero variance of the innovations process, i.e.,* $K_Z^\infty=0$, *then the value of feedback capacity [2, Theorem 6.1,* $C_{FB}=0,\forall\kappa\in[0,\infty)$*].*

5.2.3 Feedback Capacity of AGN Channels Driven by Time-Invariant Stable/Unstable ARMA(a,c) Noise

In this section we analyze the asymptotic per unit time limit of the $n-$FTFI capacity of Definition 5.2.1, by making use of the properties of generalized DREs and AREs of Section 5.2.2 to identify sufficient and necessary conditions, such that condition **(C1)** holds. Then we derive closed form expressions for $C^\infty(\kappa,s)=C^\infty(\kappa),\forall s$ defined by (5.2.101). We show that, there are multiple regimes of feedback capacity; in some regimes feedback does not increase capacity.

First, we define the main problem of asymptotic analysis.

Problem 5.2.1. *Problem of feedback capacity $C^\infty(\kappa,s)$ for stable/unstable time-invariant ARMA(a,c) noise*

Consider the characterization of the $n-$FTFI capacity of Theorem 3.1.1, and restrict the admissible strategies or distributions to the time-invariant strategies or distributions, defined by (5.2.79)-(5.2.87), which generate $(X^{o,n},Y^{o,n})$.

Define the per unit time limit and maximum by

$$C^\infty(\kappa,s)\stackrel{\triangle}{=}\max_{\mathscr{P}^\infty_{[0,\infty]}(\kappa)}\lim_{n\longrightarrow\infty}\frac{1}{2n}\sum_{t=1}^{n}\log\Big(\frac{\big(\Lambda^\infty+c-a\big)^2K_t^o+K_Z^\infty+K_W}{K_W}\Big)\tag{5.2.112}$$

where the average power constraint is defined by

$$\mathscr{P}^\infty_{[0,\infty]}(\kappa)\stackrel{\triangle}{=}\Big\{(\Lambda^\infty,K_Z^\infty):X_t^o=\Lambda^\infty\big(S_t-\widehat{S}_t^o\big)+Z_t^o,\ X_1^o=Z_1^o,\quad t=2,\ldots,n,$$
$$Z_t^o\in N(0,K_Z^\infty),\quad K_Z^\infty\geq 0,\quad \lim_{n\longrightarrow\infty}\frac{1}{n}\mathbf{E}_s\Big(\sum_{t=1}^{n}\big(X_t^o\big)^2\Big)=\lim_{n\longrightarrow\infty}\frac{1}{n}\sum_{t=1}^{n}(\Lambda^\infty)^2K_t^o+K_Z^\infty\leq\kappa\Big\}.\tag{5.2.113}$$

Determine sufficient and/or necessary conditions such that

(a) the per unit time limit exists, it is finite, i.e., condition **(C1)** *holds, and*

(b) the maximum over $(\Lambda^\infty, K_Z^\infty)$ exists, and the optimal strategy is such that $C^\infty(\kappa,s) = C^\infty(\kappa)$ is independent of the initial state s.

In the next theorem we provide the answer to Problem 5.2.1, by invoking Theorem 5.2.1.

Theorem 5.2.2. *Feedback capacity $C^\infty(\kappa,s)$*

Consider the Problem 5.2.1, defined by (5.2.112), (5.2.113).

Define the set

$$\begin{aligned}\mathscr{P}^\infty \stackrel{\triangle}{=} \Big\{ & (\Lambda^\infty, K_Z^\infty) \in (-\infty,\infty) \times [0,\infty) : \\ & \text{(i) the pair } \{A,C\} \equiv \{A, C(\Lambda^\infty)\} \text{ is detectable,} \\ & \text{(ii) the pair } \{A^*, B^{*,\frac{1}{2}}\} \equiv \{A^*(K_Z^\infty), B^{*,\frac{1}{2}}(K_Z^\infty)\} \text{ is stabilizable} \Big\}. \end{aligned} \tag{5.2.114}$$

Where $\{A,C\}, \{A^*, B^{*,\frac{1}{2}}\}$ are given by (5.2.103), (5.2.104).

Then,

$$C^\infty(\kappa,s) = C^\infty(\kappa) \stackrel{\triangle}{=} \max_{\left(\Lambda^\infty, K_Z^\infty\right) \in \mathscr{P}^\infty(\kappa)} \frac{1}{2} \log\left(\frac{(\Lambda^\infty + c - a)^2 K^\infty + K_Z^\infty + K_W}{K_W} \right) \tag{5.2.115}$$

that is, $C^\infty(\kappa,s)$ is independent of s, where

$$\begin{aligned}\mathscr{P}^\infty(\kappa) \stackrel{\triangle}{=} \Big\{ & (\Lambda^\infty, K_Z^\infty) \in \mathscr{P}^\infty : K_Z^\infty \geq 0, \quad (\Lambda^\infty)^2 K^\infty + K_Z^\infty \leq \kappa, \\ & K^\infty = c^2 K^\infty + K_W - \frac{\Big(K_W + cK^\infty(\Lambda^\infty + c - a)\Big)^2}{\Big(K_Z^\infty + K_W + (\Lambda^\infty + c - a)^2 K^\infty\Big)} \\ & K^\infty \geq 0 \text{ is unique and stabilizable, i.e., } |F(K^\infty, \Lambda^\infty, K_Z^\infty)| < 1 \Big\}, \end{aligned} \tag{5.2.116}$$

$$F(K^\infty, \Lambda^\infty, K_Z^\infty) \stackrel{\triangle}{=} c - M(K^\infty, \Lambda^\infty, K_Z^\infty)\Big(\Lambda^\infty + c - a\Big), \tag{5.2.117}$$

$$M(K^\infty, \Lambda^\infty, K_Z^\infty) \stackrel{\triangle}{=} \Big(K_W + cK^\infty\Big(\Lambda^\infty + c - a\Big)\Big)\Big(K_Z^\infty + K_W + \Big(\Lambda^\infty + c - a\Big)^2 K^\infty\Big)^{-1} \tag{5.2.118}$$

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

Moreover, the maximum element $(\Lambda^\infty, K_Z^\infty) \in \mathscr{P}^\infty(\kappa)$, is such that,

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

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

Proof. The sequence $\{K_t^o : t = 1,2,\ldots,n\}$ satisfies the time-invariant generalized DRE (5.2.94), with zero initial condition, $K_1^o = 0$. Then for elements in the set $\mathscr{P}^\infty$, an application of Theorem 5.2.1.(1), (2), states that the sequence generated by (5.2.94) converges, i.e., $\lim_{n\longrightarrow\infty} K_n^o = K^\infty$, where $K^\infty = K^\infty(\Lambda^\infty, K_Z^\infty) \geq 0$ is the unique stabilizing solution of the generalized ARE given in (5.2.116). Hence, the following summands converge, and so the limits exist and they are finite.

$$\lim_{n\longrightarrow\infty} \frac{1}{n}\sum_{t=1}^{n}\Big(\big(\Lambda^\infty\big)^2 K_t^o + K_Z^\infty\Big) = \big(\Lambda^\infty\big)^2 K^\infty + K_Z^\infty, \tag{5.2.119}$$

$$\lim_{n\longrightarrow\infty} \frac{1}{2n}\sum_{t=1}^{n}\log\Big(\frac{\big(\Lambda^\infty + c - a\big)^2 K_t^o + K_Z^\infty + K_W}{K_W}\Big) = \frac{1}{2}\log\Big(\frac{\big(\Lambda^\infty + c - a\big)^2 K^\infty + K_Z^\infty + K_W}{K_W}\Big). \tag{5.2.120}$$

This establishes the characterization of the right hand side of (5.2.115), and its independence on s. The last part of the theorem follows from the asymptotic properties of the Kalman-filter, i.e., for (i) $E_t^o, t = 1,\ldots$ is asymptotically stationary, which implies $X_t^o = \Lambda^\infty E_t^o + Z_t^o, t = 1,\ldots$, the innovations process $I_t^o, t = 1,\ldots$ and $Y_t^o = X_t^o + V_t, t = 1,\ldots$ are asymptotically stationary, and similarly for (ii) but $Y_t^o = X_t^o + V_t, t = 1,\ldots$ is not asymptotically stationary, because $V_t, t = 1,\ldots$ is unstable. □

Clearly, the set $\mathscr{P}^\infty$, defined in Theorem 5.2.2 characterizes condition **(C1)**, and (5.2.115) characterizes the asymptotic limit of feedback capacity defined by (5.0.16).
In the next remark, we discuss some aspects of Theorem 5.2.2, and we show that $\mathscr{P}^\infty(\kappa) \subseteq \mathscr{P}^\infty$ is non-empty for some values of $\kappa \in [0,\infty)$.

Remark 5.2.2. *Comments on Theorem 5.2.2*
(1) Theorem 5.2.2 characterizes the feedback capacity $C^\infty(\kappa,s) = C^\infty(\kappa)$, independently of s, for AGN channels, driven by stable or unstable $ARMA(a,c)$ noise, i.e., $a \in (-\infty,\infty)$, $c \in (-\infty,\infty)$.
(2) Let $(\Lambda^{\infty,*} K_Z^{\infty,*}) \in \mathscr{P}^\infty(\kappa)$ denote the optimal pair for the optimization problem $C^\infty(\kappa)$. Then we need to characterize the set of all $\kappa \in [0,\infty)$ such that $(\Lambda^{\infty,*} K_Z^{\infty,*}) \in \mathscr{P}^\infty(\kappa)$.
Case 1-Stable. *If $a \in [-1,1]$, $c \in (-1,1)$, then for $K_Z^\infty = 0$, by Lemma 5.2.2, $\{A,C\}$ is detectable and $\{A^*, B^{*\frac{1}{2}}\}$ is stabilizable if and only if $|\Lambda^\infty - a| < 1$. For such a choice of $(\Lambda^\infty, K_Z^\infty) \in \mathscr{P}^\infty$, then $K^\infty = 0$, and $\mathscr{P}^\infty(\kappa)$ in non-empty for all $\kappa \in [0,\infty)$.*
Case 2-Unstable. *If $|a| > 1$, $|c| \geq 1$, then for $K_Z^\infty = 0$, by Lemma 5.2.2, $\{A^*, B^{*\frac{1}{2}}\}$ is stabilizable if and only if $|\Lambda^\infty - a| < 1$, and there exists a $G \in (-\infty,\infty)$ such that $|A - GC| = |c - G(\Lambda^\infty + c - a)| < 1$, i.e., taking $G = 1$, then $\{A,C\}$ detectable if and only if $|\Lambda^\infty - a| < 1$. For such a choice of $(\Lambda^\infty, K_Z^\infty) \in \mathscr{P}^\infty$, then $K^\infty = 0$, and hence $\mathscr{P}^\infty(\kappa)$ in non-empty for all $\kappa \in [0,\infty)$.*

However, from Case 1 and Case 2, if we use $K_Z^{\infty,*}=0$ then $C^{\infty}(\kappa)=0, \forall \kappa \in [0,\infty)$. On the other hand, in Theorem 5.2.3, we show there always exists a non-feedback channel input strategy that induces a strictly positive achievable rate.

In the next lemma we give necessary conditions for the optimization problem $C^{\infty}(\kappa)$ defined by (5.2.115).

Lemma 5.2.3. *Necessary conditions for the optimization problem of Theorem 5.2.2*
Suppose there exists a policy $(\Lambda^{\infty,*},K_Z^{\infty,*}) \in \mathscr{P}^{\infty}(\kappa)$ for the optimization problem $C^{\infty}(\kappa)$ in (5.2.115). Define the Lagrangian by

$$\begin{aligned}
&\mathscr{L}(\Lambda^{\infty},K_Z^{\infty},K^{\infty},\lambda) \triangleq (\Lambda^{\infty}+c-a)^2K^{\infty}+K_Z^{\infty}+K_W \\
&\quad -\lambda_1\Big\{\Big(K^{\infty}-c^2K^{\infty}-K_W\Big)\Big(K_Z^{\infty}+K_W+(\Lambda^{\infty}+c-a)^2K^{\infty}\Big)+\Big(K_W+cK^{\infty}(\Lambda^{\infty}+c-a)\Big)^2\Big\} \\
&\quad -\lambda_2\Big((\Lambda^{\infty})^2K^{\infty}+K_Z^{\infty}-\kappa\Big)-\lambda_3\Big(-K^{\infty}\Big)-\lambda_4\Big(-K_Z^{\infty}\Big), && (5.2.121)\\
&\lambda \triangleq (\lambda_1,\lambda_2,\lambda_3\lambda_4)\in\mathbb{R}^4, \quad c\neq a, \quad K_W>0. && (5.2.122)
\end{aligned}$$

Then the following hold.
(i) Stationarity:

$$\frac{\partial}{\partial K_Z^{\infty}}\mathscr{L}(\Lambda^{\infty},K_Z^{\infty},K^{\infty},\lambda)\Big|_{\Lambda^{\infty}=\Lambda^{\infty,*},K_Z^{\infty}=K_Z^{\infty,*},K^{\infty}=K^{\infty,*},\lambda=\lambda^*}=0, \tag{5.2.123}$$

$$\frac{\partial}{\partial \Lambda^{\infty}}\mathscr{L}(\Lambda^{\infty},K_Z^{\infty},K^{\infty},\lambda)\Big|_{\Lambda^{\infty}=\Lambda^{\infty,*},K_Z^{\infty}=K_Z^{\infty,*},K^{\infty}=K^{\infty,*},\lambda=\lambda^*}=0, \tag{5.2.124}$$

$$\frac{\partial}{\partial K^{\infty}}\mathscr{L}(\Lambda^{\infty},K_Z^{\infty},K^{\infty},\lambda)\Big|_{\Lambda^{\infty}=\Lambda^{\infty,*},K_Z^{\infty}=K_Z^{\infty,*},K^{\infty}=K^{\infty,*},\lambda=\lambda^*}=0. \tag{5.2.125}$$

(ii) Complementary Slackness:

$$\lambda_2^*\Big((\Lambda^{\infty,*})^2K^{\infty,*}+K_Z^{\infty,*}-\kappa\Big)=0, \quad \lambda_3^*K^{\infty,*}=0, \quad \lambda_4^*K_Z^{\infty,*}=0, \tag{5.2.126}$$

$$\lambda_1^*\Big\{\Big(K^{\infty,*}-c^2K^{\infty,*}-K_W\Big)\Big(K_Z^{\infty,*}+K_W+(\Lambda^{\infty,*}+c-a)^2K^{\infty,*}\Big) \tag{5.2.127}$$

$$+\Big(K_W+cK^{\infty,*}(\Lambda^{\infty,*}+c-a)\Big)^2\Big\}=0. \tag{5.2.128}$$

Primal Feasibility:

$$(\Lambda^{\infty,*})^2K^{\infty,*}+K_Z^{\infty,*}\leq\kappa, \quad K_Z^{\infty,*}\geq 0, \quad K^{\infty,*}\geq 0, \tag{5.2.129}$$

$$\Big(K^{\infty,*}-c^2K^{\infty,*}-K_W\Big)\Big(K_Z^{\infty,*}+K_W+(\Lambda^{\infty,*}+c-a)^2K^{\infty,*}\Big)+\Big(K_W+cK^{\infty,*}(\Lambda^{\infty,*}+c-a)\Big)^2=0. \tag{5.2.130}$$

(iii) Dual Feasibility:

$$\lambda_1^* \in (-\infty,\infty), \ \ \lambda_2^* \geq 0, \ \ \lambda_3^* \geq 0, \ \ \lambda_4^* \geq 0. \tag{5.2.131}$$

Further, if either $K_Z^* = 0$ or $K^{\infty,*} = 0$ then $C^\infty(\kappa) = 0, \forall \kappa \in [0,\infty)$.

Proof. The conditions are a consequence of the optimization problem. We show the last part. Suppose $K_Z^{\infty,*} = 0$. Then by the generalized ARE (5.2.130) there are two solutions, $K^{\infty,*} = 0$ and $K^{\infty,*} = \frac{K_W\left(\left(\Lambda^\infty - a\right)^2 - 1\right)}{\left(\Lambda^\infty + c - a\right)^2}$. By Lemma 5.2.2, then a necessary and sufficient condition for stabilizability of the pair $\{A^*, B^{*,\frac{1}{2}}\}$ is $|\Lambda^{\infty,*} - a| < 1$. Hence, the only non-negative stabilizing solution is $K^{\infty,*} = 0$. However, $K^{\infty,*} = 0, K_Z^{\infty,*} = 0$ implies $C^\infty(\kappa) = 0, \forall \kappa \in [0,\infty)$. Now, suppose $K^{\infty,*} = 0$. Then by the generalized ARE (5.2.130), necessarily $K_Z^{\infty,*} = 0$, because $K_W > 0$ and hence $C^\infty(\kappa) = 0, \forall \kappa \in [0,\infty)$. This completes the proof. □

Achievable Rates Without Feedback for Stable and Unstable ARMA(a,c) Noise

If we replace Λ^∞ in the optimization problem of Theorem 5.2.2 by $\Lambda^\infty = 0$, then $C^\infty(\kappa)\Big|_{\Lambda^\infty=0}$, if it exists, it is an achievable rate. For $\Lambda^\infty = 0$, by (5.2.79), the channel input is an independent innovations process $X_t^o = Z_t^o, t = 1,\ldots,n$, and hence the code does not use feedback. The rate $C^\infty(\kappa)\Big|_{\Lambda^\infty=0}$ is indeed achievable, if we ensure the detectability of the pair $\{A,C\}$ and stabilizability of the pair $\{A^*, B^{*,\frac{1}{2}}\}$ are satisfied, when $\Lambda^\infty = 0$. In the next theorem, we show $C^\infty(\kappa)\Big|_{\Lambda^\infty=0}$ is an achievable rate without feedback, and we calculate its value.

Theorem 5.2.3. *Achievable rates without feedback for stable and unstable ARMA(a,c) noise for the case where $c \neq a$*

For $\Lambda^\infty = 0$, define the set

$$\mathscr{P}_0^{\infty,nfb} \triangleq \Big\{ K_Z^\infty \in [0,\infty) :$$

$$\textit{(i) the pair } \{A,C\}\Big|_{\Lambda^\infty=0} \equiv \{A, C(\Lambda^\infty)\}\Big|_{\Lambda^\infty=0} \textit{ is detectable,} \tag{5.2.132}$$

$$\textit{(ii) the pair } \{A^*, B^{*,\frac{1}{2}}\}\Big|_{\Lambda^\infty=0} \equiv \{A^*(K_Z^\infty), B^{*,\frac{1}{2}}(K_Z^\infty)\}\Big|_{\Lambda^\infty=0} \textit{ is stabilizable}\Big\}. \tag{5.2.133}$$

where,

$$A\Big|_{\Lambda^\infty=0} = c, \ \ C\Big|_{\Lambda^\infty=0} = c - a, \ \ A^*\Big|_{\Lambda^\infty=0} = c - \frac{K_W}{K_Z + K_W}(c-a),$$

$$B^{*,\frac{1}{2}}\Big|_{\Lambda^\infty=0} = K_W^{\frac{1}{2}}\left(1 - \frac{K_W}{K_W + K_Z}\right)^{\frac{1}{2}} \tag{5.2.134}$$

For $\Lambda^\infty = 0$, define the channel input and output processes by

$$X_t^o = Z_t^o, \quad t = 1, \ldots, n, \tag{5.2.135}$$

$$V_t = (c-a)S_t + W_t, \quad V_0 = v_0, \quad W_0 = w_0, \quad S_1 = s \tag{5.2.136}$$

$$V_1 = (c-a)S_1 + W_1, \tag{5.2.137}$$

$$Y_t^o = X_t^o + V_t = (c-a)S_t + W_t + Z_t^o, \tag{5.2.138}$$

$$Y_1^o = (c-a)S_1 + W_1 + Z_1^o. \tag{5.2.139}$$

(1) A lower bound on non-feedback capacity $C^{\infty,nfb}(\kappa, s)$ is $C_{LB}^{\infty,nfb}(\kappa)$ given by

$$C^{\infty,nfb}(\kappa,s) \geq C^\infty(\kappa)\Big|_{\Lambda^\infty=0} = C_{LB}^{\infty,nfb}(\kappa) \stackrel{\triangle}{=} \max_{K_Z^\infty \in \mathscr{P}_0^{\infty,nfb}(\kappa)} \frac{1}{2}\log\Big(\frac{(c-a)^2K^\infty + K_Z^\infty + K_W}{K_W}\Big) \tag{5.2.140}$$

where,

$$\mathscr{P}_0^{\infty,nfb}(\kappa) \stackrel{\triangle}{=} \Big\{K_Z^\infty \in \mathscr{P}_0^{\infty,nfb} : K_Z^\infty \geq 0, \quad K_Z^\infty \leq \kappa,$$

$$K^\infty = c^2K^\infty + K_W - \frac{\Big(K_W + cK^\infty(c-a)\Big)^2}{\Big(K_Z^\infty + K_W + (c-a)^2K^\infty\Big)}, \tag{5.2.141}$$

$$K^\infty \geq 0 \text{ is unique and stabilizing, i.e., } |F^{nfb}(K^\infty, K_Z^\infty)| < 1\Big\}, \tag{5.2.142}$$

$$F^{nfb}(K^\infty, K_Z^\infty) \stackrel{\triangle}{=} c - M^{nfb}(K^\infty, K_Z^\infty)c, \tag{5.2.143}$$

$$M^{nfb}(K^\infty, K_Z^\infty) \stackrel{\triangle}{=} \Big(K_W + cK^\infty(c-a)\Big)\Big(K_Z^\infty + K_W + (c-a)^2K^\infty\Big)^{-1} \tag{5.2.144}$$

provided there exists $\kappa \in [0,\infty)$ such that the set $\mathscr{P}_0^{\infty,nfb}(\kappa)$ is non-empty.
Moreover, $C_{LB}^{\infty,nfb}(\kappa)$ is an achievable rate without feedback, i.e., the optimal channel input strategy induces asymptotic stationarity of the joint input and output process, and (5.2.115) with $\Lambda^\infty = 0$ is independent of the initial state $S_1 = s$.
(2) The lower bound on non-feedback capacity of (1) is given by

$$C_{LB}^{\infty,nfb}(\kappa) = \frac{1}{2}\log\Big(\frac{(c-a)^2K^{\infty,*} + \kappa + K_W}{K_W}\Big), \qquad \kappa \in \mathscr{K}^{\infty,nfb}(a,c,K_w) \tag{5.2.145}$$

where the two solutions of Riccati equation (5.2.141) and $K_Z^{\infty,*}$, are given by

$$K^{\infty,*} = \begin{cases} \frac{-\kappa\left(1-c^2\right)-K_W\left(1-a^2\right)+\sqrt{\left(\kappa\left(1-c^2\right)+K_W\left(1-a^2\right)\right)^2+4\left(c-a\right)^2 K_W\kappa}}{2\left(c-a\right)^2} \geq 0, & \kappa \in \mathscr{K}^{\infty,nfb}(c,a,K_W), \\ \frac{-\kappa\left(1-c^2\right)-K_W\left(1-a^2\right)-\sqrt{\left(\kappa\left(1-c^2\right)+K_W\left(1-a^2\right)\right)^2+4\left(c-a\right)^2 K_W\kappa}}{2\left(c-a\right)^2} \leq 0, & \kappa \in \mathscr{K}^{\infty,nfb}(c,a,K_W), \end{cases} \tag{5.2.146}$$

$$K_Z^{\infty,*} = \kappa, \tag{5.2.147}$$

$$\mathscr{K}^{\infty,nfb}(a,c,K_W) \triangleq \left\{\kappa \in [0,\infty): \quad K^{\infty} \geq 0\right\} = [0,\infty) \tag{5.2.148}$$

where, the first solution is the unique and non-negative solution.

Proof. (1) Note that by setting $\Lambda^{\infty}=0$, then the representation of channel input $X^{o,n}$ defined by (5.2.79)-(5.2.94) is used without feedback, and this is a lower bound on the non-feedback capacity. Hence, the statements follow from Theorem 5.2.2, as a special case.
Solving (5.2.141) gives (5.2.146). We check the detectability or the pair $\{A,C\}\Big|_{\Lambda^{\infty}=0}$ and the stabilizability or the pair $\{A^*,B^{*,\frac{1}{2}}\}\Big|_{\Lambda^{\infty}=0}$.
(a) The pair $\{A,C\}\Big|_{\Lambda^{\infty}=0} = \{c,c-a\}$ is detectable if there exists a $G\in\mathbb{R}$, such that $|A-GC| = |c-G(c-a)| < 1$.
Since there is G such that, $c-G(c-a)=0$, i.e., $G=\frac{c}{c-a}$, then $\{A,C\}\Big|_{\Lambda^{\infty}=0}$ is detectable.
(b) The pair $\{A^*,B^{*,\frac{1}{2}}\}\Big|_{\Lambda^{\infty}=0}$ given by (5.2.134), is always stabilizable, because $B^{*,\frac{1}{2}} > 0$ and $K_W > 0$ $K_Z^{\infty} > 0$. □

In the next Definition, we discuss a special case of the time-invariant $ARMA(a,c)$ noise, of Definition 5.0.1.

Definition 5.2.4. *Moving Average $MA(a)$ noise*
Suppose $c=0$. Hence, by Definition 5.0.1 we have the time-invariant stable or unstable noise $MA(a), a\in(-\infty,\infty)$, which is defined by,

$$V_t = W_t - aW_{t-1}, \quad \forall t \in \mathbb{Z}_+ \triangleq \{1,2,\ldots\}, \tag{5.2.149}$$

$$W_0 \in N(0,K_{W_0}), \quad K_{W_0} \geq 0, \quad W_t \in N(0,K_W), \quad t=1,2,\ldots, \quad K_W > 0, \tag{5.2.150}$$

$$\{W_0,W_1,\ldots,W_n\} \text{ indep. seq. and indep. of } V_0, \quad a \in (-\infty,\infty) \tag{5.2.151}$$

By Remark 5.0.1, then the state variable is,

$$S_t \stackrel{\triangle}{=} W_{t-1}, \quad \forall t \in \mathbb{Z}_+ \tag{5.2.152}$$

and the state space realization of V^n is

$$S_{t+1} = W_t, \quad S_1 = W_0 = s, \quad \forall t \in \mathbb{Z}_+, \tag{5.2.153}$$

$$V_t = -aS_t + W_t, \quad \forall t \in \mathbb{Z}_+, \tag{5.2.154}$$

$$K_{S_1} = K_{W_0}, \quad K_{W_0} \geq 0. \tag{5.2.155}$$

Corollary 5.2.1. *Achievable rates without feedback for stable and unstable MA(a) noise*
For $\Lambda^\infty = 0$ and $c = 0$, define the set

$$\mathscr{P}_0^{\infty,nfb} \stackrel{\triangle}{=} \Big\{ K_Z^\infty \in [0,\infty) : \\ \text{(i) the pair } \{A,C\}\Big|_{\Lambda^\infty=0,c=0} \equiv \{A, C(\Lambda^\infty)\}\Big|_{\Lambda^\infty=0,c=0} \text{ is detectable,} \tag{5.2.156}$$

$$\text{(ii) the pair } \{A^*, B^{*,\frac{1}{2}}\}\Big|_{\Lambda^\infty=0,c=0} \equiv \{A^*(K_Z^\infty), B^{*,\frac{1}{2}}(K_Z^\infty)\}\Big|_{\Lambda^\infty=0,c=0} \text{ is stabilizable} \Big\}. \tag{5.2.157}$$

where,

$$A\Big|_{\Lambda^\infty=0,c=0} = 0, \ C\Big|_{\Lambda^\infty=0,c=0} = -a, \ A^*\Big|_{\Lambda^\infty=0,c=0} = a\frac{K_W}{K_Z + K_W}, \\ B^{*,\frac{1}{2}}\Big|_{\Lambda^\infty=0,c=0} = K_W^{\frac{1}{2}}\Big(1 - \frac{K_W}{K_W + K_Z}\Big)^{\frac{1}{2}} \tag{5.2.158}$$

For $\Lambda^\infty = 0$, define the channel input and output processes by

$$X_t^o = Z_t^o, \qquad t = 1, \ldots, n, \tag{5.2.159}$$

$$Y_t^o = X_t^o + V_t = -aS_t + W_t + Z_t^o, \tag{5.2.160}$$

$$Y_1^o = -aS_1 + W_1 + Z_1^o. \tag{5.2.161}$$

(1) A lower bound on non-feedback capacity $C^{\infty,nfb}(\kappa,s)$ is $C_{LB}^{\infty,nfb}(\kappa)$ given by

$$C^{\infty,nfb}(\kappa,s) \geq C^\infty(\kappa)\Big|_{\Lambda^\infty=0,c=0} = C_{LB}^{\infty,nfb,a}(\kappa) \stackrel{\triangle}{=} \max_{K_Z^\infty \in \mathscr{P}_0^{\infty,nfb}(\kappa)} \frac{1}{2}\log\Big(\frac{a^2K^\infty + K_Z^\infty + K_W}{K_W}\Big) \tag{5.2.162}$$

where,

$$\mathscr{P}_0^{\infty,nfb}(\kappa) \triangleq \Big\{ K_Z^\infty \in \mathscr{P}_0^{\infty,nfb} : K_Z^\infty \geq 0, \quad K_Z^\infty \leq \kappa,$$

$$K^\infty = K_W - \frac{{K_W}^2}{\Big(K_Z^\infty + K_W + a^2 K^\infty\Big)} \tag{5.2.163}$$

$$K^\infty \geq 0 \text{ is unique and stabilizing, i.e., } |F^{nfb}(K^\infty, K_Z^\infty)| < 1 \Big\} \tag{5.2.164}$$

Further,

$$C_{LB}^{\infty,nfb}(\kappa) = \frac{1}{2}\log\Big(\frac{a^2K^{\infty,*} + \kappa + K_W}{K_W}\Big), \qquad \kappa \in [0,\infty) \tag{5.2.165}$$

When,

$$K^{\infty,*} = \frac{-\kappa - K_W\big(1-a^2\big) + \sqrt{\Big(\kappa + K_W\big(1-a^2\big)\Big)^2 + 4a^2K_W\kappa}}{2a^2}. \tag{5.2.166}$$

Proof. This is a special case of the Theorem 5.2.3, where we replace the variable c with $c = 0$. □

In the next theorem, we derive closed form expressions for the feedback capacity, by solving the optimization problem of Theorem 5.2.2.

First, by the definition of the sets $\mathscr{P}^\infty$ and $\mathscr{P}_0^{\infty,nfb}$ of Theorem 5.2.2 and Theorem 5.2.3, we have

$$\mathscr{P}^\infty = \mathscr{P}_0^{\infty,nfb} \bigcup \mathscr{P}^{\infty,fb}, \qquad \mathscr{P}^{\infty,fb} \triangleq \Big\{(\Lambda^\infty, K_Z^\infty) \in \mathscr{P}^\infty : \Lambda^\infty \neq 0\Big\}. \tag{5.2.167}$$

Thus, if $(\Lambda^\infty, K_Z^\infty) \in \mathscr{P}^{\infty,fb}$ then the channel input process applies feedback, and if $K_Z^\infty \in \mathscr{P}_0^{\infty,nfb}$ then the channel input process does not apply feedback.

Theorem 5.2.4. *Feedback capacity-solution of optimization problem of Theorem 5.2.2*

(1) The non-zero feedback capacity $C^\infty(\kappa)$ defined by (5.2.115), for a stable and unstable ARMA(a,c) noise, i.e, $a \in (-\infty,\infty)$, $c \in (-\infty,\infty)$, with $c \neq 0, c^2 \neq 1, c^2 \neq 2, c \neq a$, occurs in the set $\mathscr{P}^{\infty,fb}$, and

is given, as follows.

$$C^{\infty}(\kappa) = \frac{1}{2}\log\Big(\frac{\big(\Lambda^{\infty,*}+c-a\big)^2 K^{\infty,*}+K_Z^{\infty,*}+K_W}{K_W}\Big), \qquad \kappa \in \mathscr{K}^{\infty}(a,c,K_W) \tag{5.2.168}$$

$$\Lambda^{\infty,*} = \frac{K_W\Big(ac-1\Big)+\kappa\Big(c^2-1\Big)-K^{\infty,*}\Big(a-c\Big)^2}{K^{\infty,*}\Big(c^2-2\Big)\Big(a-c\Big)} \tag{5.2.169}$$

$$K_Z^{\infty,*}+\Big(\Lambda^{\infty,*}\Big)^2 K^{\infty,*} = \kappa, \tag{5.2.170}$$

$$\mathscr{K}^{\infty}(a,c,K_W) \triangleq \Big\{\kappa \in [0,\infty): \quad K^{\infty,*}>0, \quad K_Z^{\infty,*}>0\Big\} \tag{5.2.171}$$

where $K^{\infty,*}$ is the unique positive and stabilizing solution, i.e., $|F(K^{\infty,*},\Lambda^{\infty,*},K_Z^{\infty,*})|<1$. Further, for any $\kappa \in \mathscr{K}^{\infty}(a,c,K_W)$, then

$$K^{\infty,*} = \frac{K_W(2a-c+a^2c^3-2a^2c)+c\kappa(c^2-1)^2}{c(c^2-1)(a-c)^2}, \tag{5.2.172}$$

$$\Lambda^{\infty,*} = \frac{K_W(a-c)^2(1-ac)}{K_W(2a-c+a^2c^3-2a^2c)+c\kappa(c^2-1)^2}, \tag{5.2.173}$$

$$K_Z^{\infty,*} = \frac{\kappa\Big(c(c^2-1)(K_W(2a-c+a^2c^3-2a^2c)+c\kappa(c^2-1)^2\Big)-K_W^2(a-c)^2(1-ac)^2}{c(c^2-1)\Big(K_W(2a-c+a^2c^3-2a^2c)+c\kappa(c^2-1)^2\Big)}. \tag{5.2.174}$$

(2) The non-zero feedback capacity $C^{\infty}(\kappa)$, $\kappa \in \mathscr{K}^{\infty}(a,c,K_W)$ of part (1), is restricted to the regions:

$$\textit{Region A: } c \in \big(1,\sqrt{2}\big)\cup\big(\sqrt{2},\infty\big), a \in \Big[\frac{-c}{c^2-2},\frac{1}{c}\Big], \quad \textit{for } \kappa > \kappa_{min}, \tag{5.2.175}$$

$$\textit{Region B: } c \in \big(-\infty,-\sqrt{2}\big)\cup\big(-\sqrt{2},-1\big), a \in \big(-\infty,\frac{1}{c}\big]\cup\big[\frac{-c}{c^2-2},\infty\big), \quad \textit{for } \kappa > \kappa_{min}. \tag{5.2.176}$$

where,

$$\kappa_{min} = \frac{K_W\big(1-ac\big)\big(2ac-ac^3-c^2+\sqrt{c^3\big(a^2c^3-6ac^2+4a+4c^3-3c\big)}\big)}{2c^2\big(c^2-1\big)^2} \tag{5.2.177}$$

Proof. See Appendix 7.11. □

Theorem 5.2.5. *The non-zero feedback capacity $C^{\infty}(\kappa)$ by Theorem 5.2.2, for a stable and unstable ARMA(a,c) noise, i.e, $a \in (-\infty,\infty)$, $c \in (-\infty,\infty)$, with $c \neq 0, c^2 \neq 1, c^2 \neq 2, c \neq a$, occurs in*

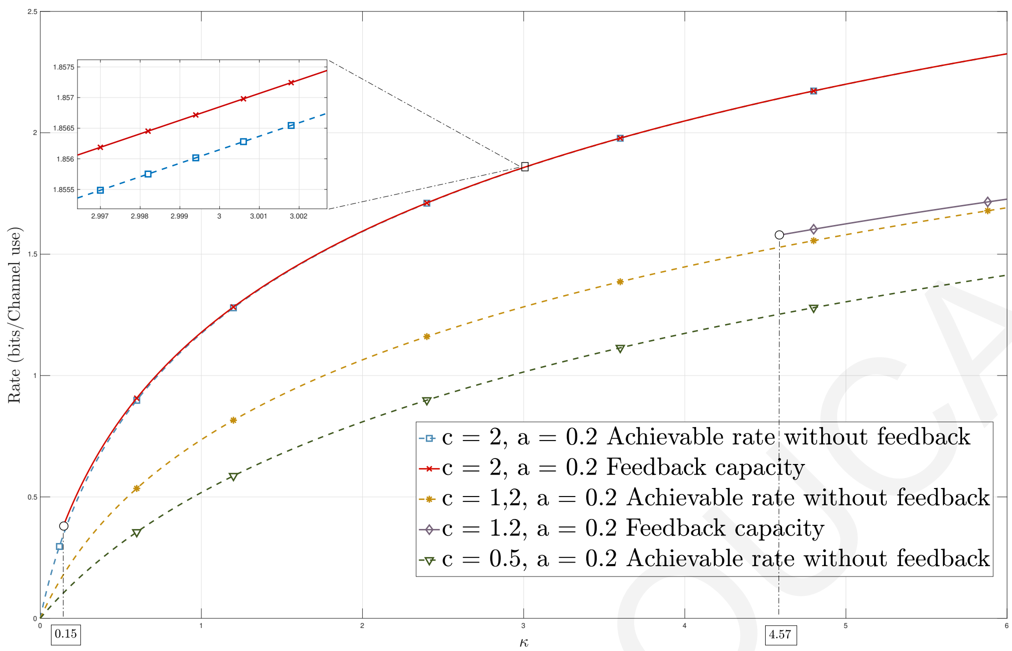

Figure 5.2.1: Feedback capacity $C^{\infty}(\kappa)$ for $\kappa \in \mathscr{K}^{\infty}(a,c,K_W)$ based on (5.2.115) and lower bound on nofeedback capacity $C_{LB}^{\infty,nfb}(\kappa)$ for $\kappa \in [0,\infty)$ based on (5.2.145), of the AGN channel driven by ARMA(a,c) noise, for various values of $a = 0.2$, $c \in (-\infty,\infty)$ and $K_W = 1$.

the set $\mathscr{P}^{\infty,fb}$, and is given, as follows.

$$C^{\infty}(\kappa) = \frac{1}{2}\log\left(\frac{\left(\Lambda^{\infty,*}+c-a\right)^2 K^{\infty,*} + K_Z^{\infty,*} + K_W}{K_W}\right), \qquad \kappa \in \mathscr{K}^{\infty}(a,c,K_W) \tag{5.2.178}$$

$$\Lambda^{\infty,*} = \frac{K_W\left(ac-1\right)+\kappa\left(c^2-1\right)-K^{\infty,*}\left(a-c\right)^2}{K^{\infty,*}\left(c^2-2\right)\left(a-c\right)} \tag{5.2.179}$$

$$K_Z^{\infty,*} + \left(\Lambda^{\infty,*}\right)^2 K^{\infty,*} = \kappa, \tag{5.2.180}$$

$$\mathscr{K}^{\infty}(a,c,K_W) \triangleq \left\{\kappa \in [0,\infty): \quad K^{\infty,*} > 0, \quad K_Z^{\infty,*} > 0\right\} \tag{5.2.181}$$

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

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

$$K^{\infty,*} = \frac{K_W(2a-c+a^2c^3-2a^2c)+c\kappa(c^2-1)^2}{c(c^2-1)(a-c)^2}, \tag{5.2.182}$$

$$\Lambda^{\infty,*} = \frac{K_W(a-c)^2(1-ac)}{K_W(2a-c+a^2c^3-2a^2c)+c\kappa(c^2-1)^2}, \tag{5.2.183}$$

$$K_Z^{\infty,*} = \frac{\kappa\Big(c(c^2-1)(K_W(2a-c+a^2c^3-2a^2c)+c\kappa(c^2-1)^2\Big)-K_W^2(a-c)^2(1-ac)^2}{c(c^2-1)\Big(K_W(2a-c+a^2c^3-2a^2c)+c\kappa(c^2-1)^2\Big)}. \tag{5.2.184}$$

Proof. See Appendix 7.11. □

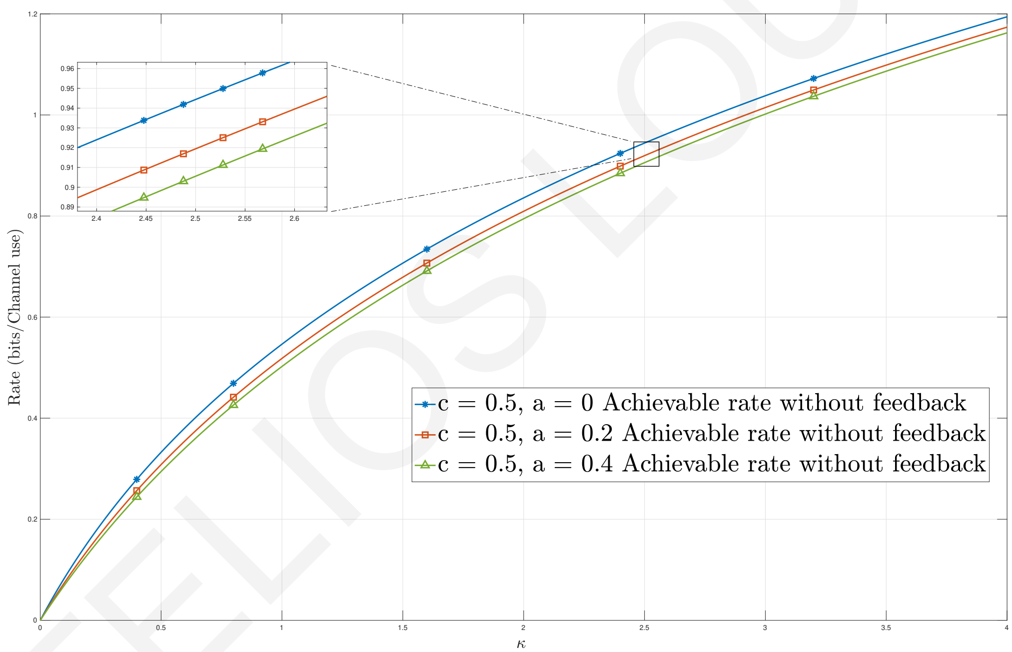

Figure 5.2.2: Lower bound achievable rate for no-feedback $C_{LB}^{\infty,nfb}(\kappa)$ for $\kappa \in [0,\infty)$ based on (5.2.145), of the AGN channel driven by ARMA(a,c) noise, for various values of $a \in (-\infty,\infty)$, $c = 0.5$ and $K_W = 1$.

From the previous theorem it then follows the next theorem, that states feedback does not increase capacity $C^{\infty}(\kappa)$ defined by (5.2.115), for the two regions $\mathscr{K}^{\infty,nfb}(a,c)$ and $\mathscr{K}^{\infty,nfb}(a,c,K_W)$.

Theorem 5.2.6. *Feedback does not increase capacity for certain regions*

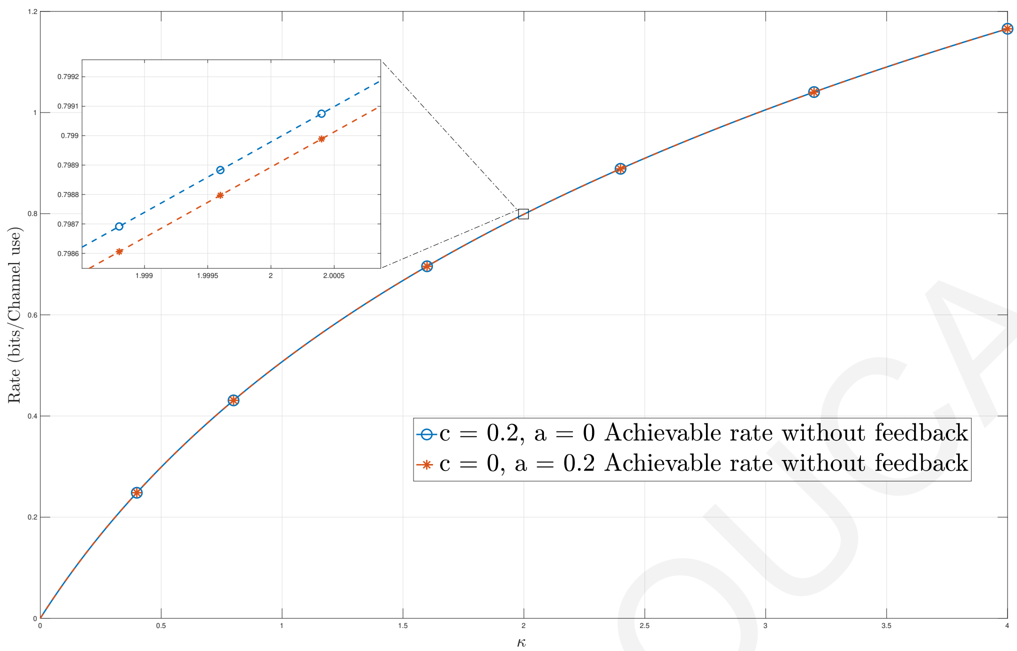

Figure 5.2.3: Comparison between $AR(c)$ noise and $MA(a)$ noise, for various values of $a \in 0, 0.2$, $c \in 0.2, 0$ and $K_W = 1$.

Feedback does not increase capacity $C^{\infty}(\kappa)$ defined by (5.2.115), for the following regions:

$$\text{Region 1: Outside the Region A} \tag{5.2.185}$$

$$e.g.,\ c \in \big(1, \sqrt{2}\big) \cup \big(\sqrt{2}, \infty\big), \quad \alpha \in \Big[\frac{-c}{c^2-2}, \frac{1}{c}\Big], \ \ for\ \kappa \le \kappa_{min}, \tag{5.2.186}$$

$$\text{Region 2: Outside the Region B} \tag{5.2.187}$$

$$e.g.,\ c \in \big(-\infty, -\sqrt{2}\big) \cup \big(-\sqrt{2}, -1\big), \quad \alpha \in \Big(-\infty, \frac{1}{c}\Big] \cup \Big[\frac{-c}{c^2-2}, \infty\Big), \ \ for\ \kappa \le \kappa_{min}, \tag{5.2.188}$$

$$\text{Region 3: } c \in \big[-1, 1\big], \alpha \in \mathbb{R} \ \ for\ \kappa \in \big[0, \infty\big). \tag{5.2.189}$$

Where $|c| \ge 1$ corresponds to the unstable noise, and $|c| < 1$ to the stable.

Proof. By Theorem 5.2.4 we deduce that, if $\Lambda^{\infty} \neq 0$, i.e., if feedback is used, then there does not exists a non-zero value of $C^{\infty}(\kappa)$, for $\kappa \in [0, \infty)$. On the other hand, by Theorem 5.2.3, an achievable rate without feedback exists for all $\kappa \in [0, \infty)$, by letting $\Lambda^{\infty} = 0$, and $C_{LB}^{\infty, nfb}(\kappa)$ is a lower bound on capacity without feedback. □

For the regions of Theorem 5.2.3 it doesn't exist a feedback capacity. However, there is an achievable rate (5.2.165), which is calculated in Theorem 5.2.3. Further, it's important to understand that, for the regions that there exists a feedback capacity, we have an achievable rate too, but it's less than the feedback capacity. In general, we have always an achievable rate, and sometimes a greater feedback capacity. In Figure 5.2.1, we see that, whether $|c| < 1$, then we have only an achievable rate. Otherwise, for $|c| \geq 1$, we have an always an achievable rate, but there exists a greater feedback capacity after a specific power constraint. It's also important to emphasize that we keep the value of a as a constant to show that the feedback capacity and the achievable rate for no-feedback are analogous with the value c. These illustrate that feedback capacity $C^{\infty}(\kappa)$ for $\kappa \in \mathscr{K}^{\infty}(c, K_W)$ is an increasing function of the parameter, $|c| \in [1, \infty)$, that is, the more unstable ARMA(a,c) noise the higher the value of capacity $C^{\infty}(\kappa)$. Further, the lower bound on nofeedback capacity $C_{LB}^{\infty,nfb}(\kappa)$ is achievable for all $\kappa \in [0, \infty)$, for stable and unstable ARMA$(a,c), a \in (-\infty, \infty), c \in (-\infty, \infty)$ noise, because the IID channel input process, induces asymptotic stationarity and ergodicity of the channel output process. As illustrated in Figure 5.2.1, for values of $|c| \geq 1$, feedback capacity occurs at $\kappa > \kappa_{min}$, from (5.2.177).

In Figure 5.2.2, we decided to keep the value of c as a constant, because we need to show that whether the value of a is increasing the achievable rate is decreasing. For $a = 0$, we have the maximum achievable rate that we could have.

We have in mind that, the AGN channel is driven by ARMA(a,c) noise, which consists of $AR(c)$ and $MA(a)$. So, in the Figure 5.2.3 we compare which of the two noise models, gives more Achievable rate and we see that $AR(c)$ noise model is better.

In conclusion, we have always an achievable rate, but feedback capacity only in Regimes (5.2.175), (5.2.176). Further, whether the initial state $S_1 = s$ is known to the encoder and decoder, it's better AGN channel driven by $AR(c)$.

Chapter 6

Conclusion

The n-finite transmission feedback information (FTFI) capacity for additive Gaussian noise (AGN) channels with feedback, is characterized, and lower bounds on the characterization of the n-finite transmission without feedback information (FTwFI) capacity are derived. The channel is driven by unit memory stable and unstable Autoregressive Moving Average Noise, where the initial state is known to the encoder and the decoder. It is shown that closed form feedback capacity formulas are derived, when channel input strategies or distributions are time-invariant, which does not always exist e.g., the noise is stable for certain unstable noise. However, lower bound on the non-feedback capacity is also derived, based on Markov channel input distributions, i.e., induced by a Gaussian Markov channel input process, and also by an independent and identically distributed channel input process. This achievable rate always holds for any autoregressive moving average noise model, whether it is stable or unstable.

6.1 Future Work

In the future, we would be interested to deal with the following topics:
(1) To derive a closed form expression of feedback capacity, when the AGN channel is driven by Autoregressive Moving Average Noise, where the state is not known either to the encoder or the decoder.
(2) To derive a closed form expression of feedback capacity, when the channel is driven by Autoregressive Moving Average Noise with memory.
(3) Work on several problems, with time-varying channel input strategies.

Chapter 7

Appendix

7.1 Proof of Theorem 2.1.1

(a) Consider an element of $\overline{\mathscr{E}}_{[0,n]}(\kappa)$. Then the conditional entropies $H^{\overline{e}}(Y_t|Y^{t-1}), t=1,\ldots,n$ are defined, provided the conditional distributions of Y_t conditioned on Y^{t-1}, i.e., $\mathbf{P}_t^{\overline{e}}(dy_t|y^{t-1})$, for $t=1,\ldots,n$, are determined. By the reconditioning property of conditional distributions, then

$$\mathbf{P}_t^{\overline{e}}(dy_t|y^{t-1}) = \int \mathbf{P}_t^{\overline{e}}(dy_t|y^{t-1},w,v^{t-1})\,\mathbf{P}_t^{\overline{e}}(dw,dv^{t-1}|y^{t-1}), \quad t=0,\ldots,n \tag{7.1.1}$$

$$= \int \mathbf{P}_t(dy_t|\overline{e}_t(w,v^{t-1},y^{t-1}),v^{t-1})\,\mathbf{P}_t^{\overline{e}}(dw,v^{t-1}|y^{t-1}), \qquad \text{by (2.1.15), (2.1.16).} \tag{7.1.2}$$

Hence, (2.1.19) is shown. Similarly, consider an element of $\overline{\mathscr{P}}_{[0,n]}(\kappa)$. Then the conditional entropies $H^{\overline{P}}(Y_t|Y^{t-1}), t=1,\ldots,n$ are defined, provided the conditional distributions of Y_t conditioned on Y^{t-1}, i.e., $\mathbf{P}_t^{\overline{P}}(dy_t|y^{t-1})$ for $t=1,\ldots,n$, are determined. By (2.1.7) and (2.1.8), then (2.1.20) is obtained. Since $\overline{\mathscr{E}}_{[0,n]}(\kappa) \subseteq \overline{\mathscr{P}}_{[0,n]}(\kappa)$ it then follows the inequality (2.1.18).

(b) This part follows by the maximum entropy principle of Gaussian distributions. That is, under the restriction (1.1.10), then a conditional Gaussian element of $\{\overline{P}(dx_t|v^{t-1},y^{t-1}), t=1,\ldots,n\} \in \overline{\mathscr{P}}_{[0,n]}(\kappa)$, with linear conditional mean and nonrandom conditional covariance induces a jointly Gaussian distribution of the process (X^n,Y^n), such that the marginal distribution of Y^n is jointly Gaussian. Below, we provide alternative proof that uses the Cover and Pombra characterization of the $n-$FTFI capacity, given by (1.2.20)-(1.2.19). Consider (1.2.20) and define the process

$$Z_1 \stackrel{\triangle}{=} \overline{Z}_1 - \mathbf{E}\Big\{\overline{Z}_1\Big\}, \tag{7.1.3}$$

$$Z_t \stackrel{\triangle}{=} \overline{Z}_t - \mathbf{E}\Big\{\overline{Z}_t \Big| X^{t-1}, V^{t-1}, Y^{t-1}\Big\}, \quad t = 2,\ldots,n, \tag{7.1.4}$$

$$= \overline{Z}_t - \mathbf{E}\Big\{\overline{Z}_t \Big| V^{t-1}, Y^{t-1}\Big\}, \qquad \text{since } X^{t-1} \text{ is uniquely defined by } (V^{t-1}, Y^{t-1}). \tag{7.1.5}$$

Then, Z_t is a Gaussian orthogonal innovations process, i.e., Z_t is independent of $(X^{t-1}, V^{t-1}, Y^{t-1})$, for $t = 2,\ldots,n$, and $\mathbf{E}\{Z_t\} = 0$, for $t = 1,\ldots,n$. By (1.2.20), we re-write $X_t, t = 1,\ldots,n$ as,

$$X_t = \sum_{j=1}^{t-1} B_{t,j} V_j + \overline{Z}_t, \quad t = 1,\ldots,n, \tag{7.1.6}$$

$$= \sum_{j=1}^{t-1} B_{t,j} V_j + \mathbf{E}\Big\{\overline{Z}_t \Big| V^{t-1}, Y^{t-1}\Big\} + Z_t, \qquad \text{by (7.1.5)} \tag{7.1.7}$$

$$\stackrel{(a)}{=} \sum_{j=1}^{t-1} B_{t,j} V_j + \overline{\Gamma}_t \begin{pmatrix} \mathbf{V}^{t-1} \\ \mathbf{Y}^{t-1} \end{pmatrix} + Z_t, \qquad \text{for some } \overline{\Gamma}_t \text{ nonrandom} \tag{7.1.8}$$

$$= \sum_{j=1}^{t-1} \Gamma^1_{t,j} V_j + \sum_{j=1}^{t-1} \Gamma^2_{t,j} Y_j + Z_t, \qquad \text{for some } (\Gamma^1_{\cdot,\cdot}, \Gamma^2_{\cdot,\cdot}) \tag{7.1.9}$$

$$= \Gamma^1_t \mathbf{V}^{t-1} + \Gamma^2_t \mathbf{Y}^{t-1} + Z_t, \qquad \text{by definition} \tag{7.1.10}$$

where (a) is due to the by joint Gaussianity of (Z^n, X^n, Y^n). From (7.1.10) and the independence of Z_t and $(X^{t-1}, V^{t-1}, Y^{t-1})$, for $t = 2,\ldots,n$, it then follows (2.1.21), and also (2.1.22).

(c) The statements follow directly from the representation of part (b), while the independence of Z^n and V^n is due to the code definition, i.e., Definition 1.1.1.(iv).

(d) The statement follows from (a)-(c).

7.2 Proof of Proposition 2.1.1

(a) The covariances of the realization of the ARMA(a,c) noise of Example 2.1.2.(b) satisfy the recursions

$$K_{S_{t+1}} = c^2 K_{S_t} + K_W, \quad K_{S_t,V_t} = \Big(c-a\Big) K_{S_t}, \quad K_{V_t} = \Big(c-a\Big)^2 K_{S_t} + K_W, \quad \forall t \in \mathbb{Z}. \tag{7.2.11}$$

If the recursion $K_{S_{t+1}} = c^2 K_{S_t} + K_W$ is initiated at the stationary value $K_{S_1} = d_{11} = \frac{K_W}{1-c^2}$, then $K_{S_{t+1}} = d_{11}, \forall t = 2,3,\ldots,$, and hence $S_t, \forall t \in \mathbb{Z}$ is stationary, which then implies stationarity of $V_t, \forall t \in \mathbb{Z}$. Hence, if (2.1.74) holds then $(V_t, S_t), \forall t \in \mathbb{Z}$ is stationary. By simple calculations it then follows (2.1.75). Similarly for the one-sided ARMA(a,c). (b) By the above covariances, for all $K_{S_1} \geq 0$, then $\lim_{n\longrightarrow\infty} K_{S_n} = K_S^\infty$, where $K_S^\infty = c^2 K_S^\infty + K_W$, which then implies $K_S^\infty = d_{11}$. Similarly, $\lim_{n\longrightarrow\infty} K_{S_n,V_n} = K_{S,V}^\infty = d_{12}$, $\lim_{n\longrightarrow\infty} K_{V_n} = K_V^\infty = d_{22}$. The initial data are determined from mean-square estimation of jointly Gaussian RVs, as follows.

$$\begin{aligned}
\hat{S}_1 =& \mathbf{E}\Big\{S_1\Big|V_0\Big\} = \mathbf{E}\Big\{cS_0 + W_0\Big|V_0\Big\} \\
=& \mathbf{E}\Big\{cS_0 + W_0\Big\} + cov(cS_0 + W_0, V_0)\Big\{cov(V_0,V_0)\Big\}^{-1}\Big(V_0 - \mathbf{E}\Big\{V_0\Big\}\Big) && (7.2.12) \\
=& \Big(cd_{12} + K_W\Big)d_{22}^{-1}V_0, && (7.2.13) \\
\Sigma_1 =& cov(S_1,S_1\Big|V_0) = cov(S_1,S_1) - \Big\{cov(S_1,S_1)\Big\}^2\Big\{cov(V_0,V_0)\Big\}^{-1} && (7.2.14) \\
=& d_{11} - d_{11}^2 d_{22}^{-1} && (7.2.15)
\end{aligned}$$

The last part is obvious.

7.3 Proof of Theorem 2.2.1

(a) Clearly, (2.2.104)-(2.2.99), follow directly from Theorem 3.1.1, and the preliminary calculations, prior to the statement of the theorem. However, (2.2.104)-(2.2.99) can also be shown independently of Theorem 3.1.1, by invoking the maximum entropy property of Gaussian distributions, as follows. By Lemma 2.1.1, then $H(V^n) = \sum_{t=1}^n H(\hat{I}_t)$. By the maximum entropy principle, then $H(Y^n)$ is maximized if $\mathbf{P}_{Y^n}$ is jointly Gaussian, the average power constraint holds, and (1.1.10) is respected. By (2.2.83), and (2.2.93), (2.2.94), if (2.2.93)-(2.2.99) hold, then (X^n, Y^n) is jointly Gaussian, and hence $H(Y^n)$ is maximized. This shows (a).

(b) Step 1. By (2.2.97) and (2.2.98), an alternative representation of X^n to the one given in Theorem 3.1.1, and induced by (2.1.23), is

$$\begin{aligned}
& X_t = \Gamma_t^1 \hat{S}_t + \Gamma_t^2 Y^{t-1} + Z_t, \quad t = 1, \ldots, n, && (7.3.16) \\
& Z_t \quad \text{satisfies (2.2.101).} && (7.3.17)
\end{aligned}$$

for some nonrandom $(\Gamma_\cdot^1, \Gamma_\cdot^2)$. Upon substituting (7.3.16) into the channel output Y^n we have

$$\begin{aligned}
Y_t =& \Gamma_t^1 \hat{S}_t + \Gamma_t^2 Y^{t-1} + Z_t + V_t, \quad t = 1, \ldots, n && (7.3.18) \\
=& \Big(\Gamma_t^1 + C_t\Big)\hat{S}_t + \Gamma_t^2 Y^{t-1} + Z_t + \hat{I}_t, \quad \text{by (2.1.41).} && (7.3.19)
\end{aligned}$$

The right hand side of (7.3.19) is driven by two independent processes, $Z_t, t = 1, \ldots, n$ and $\hat{I}_t, t = 1, \ldots, n$, which are also mutually independent. Further, the right hand side of (7.3.19) is a linear function of a state process $\hat{S}_t, t = 1, \ldots, n$, which satisfies the recursion (2.1.39):

$$\hat{S}_{t+1} = A_t \hat{S}_t + M_t(\Sigma_t) \hat{I}_t, \quad \hat{S}_1 = \mu_{S_1}, \tag{7.3.20}$$

Note that the right hand side of (7.3.20) is driven by the orthogonal process $\hat{I}_t$, which is independent of V^{t-1} and hence of $\hat{S}_t$, and also independent of Y^{t-1}. By (2.2.101), Z_t is independent of Y^{t-1} and of $\hat{S}_t$. By (7.3.19) and (7.3.20), it follows that $\widehat{\hat{S}}_t, t = 1, \ldots, n$ satisfies a generalized Kalman-filter recursion, similar to that of Lemma 2.1.1, and hence the entropy $H(Y^n)$ can be computed using the innovations process of Y^n, as in Lemma 2.1.1.

Define the orthogonal Gaussian innovations process I^n of Y^n by

$$\begin{aligned}
I_t &\stackrel{\triangle}{=} Y_t - \mathbf{E}\Big\{Y_t \Big| Y^{t-1}\Big\}, \qquad t = 1, \ldots, n && (7.3.21)\\
&= \Big(\Gamma_t^1 + C_t\Big)\Big(\hat{S}_t - \widehat{\hat{S}}_t\Big) + \hat{I}_t - \mathbf{E}\Big\{\hat{I}_t \Big| Y^{t-1}\Big\} + Z_t, \quad \text{by (7.3.19).} && (7.3.22)\\
&= \Big(\Gamma_t^1 + C_t\Big)\Big(\hat{S}_t - \widehat{\hat{S}}_t\Big) + \hat{I}_t + Z_t, \quad \text{by } \hat{I}_t \text{ indep. of } Y^{t-1}, \mathbf{E}\Big\{\hat{I}_t\Big\} = 0 && (7.3.23)
\end{aligned}$$

The entropy of Y^n is computed as follows.

$$\begin{aligned}
H(Y^n) &= \sum_{t=1}^{n} H(Y_t | Y^{t-1}) && (7.3.24)\\
&= \sum_{t=1}^{n} H(I_t | Y^{t-1}), \quad \text{by (7.3.21) and a property of conditional entropy} && (7.3.25)\\
&= \sum_{t=1}^{n} H(I_t), \quad \text{by orthogonality of } I_t \text{ and } Y^{t-1}. && (7.3.26)
\end{aligned}$$

By (7.3.23) the Gaussian innovations process I^n does not depend on the strategy Γ^2, and consequently by (7.3.26) the entropy $H(Y^n)$ does not depend on the strategy Γ^2.

Step 2. Let $g_t(Y^{t-1}) \stackrel{\triangle}{=} \Gamma_t^2 Y^{t-1}, t = 1, \ldots, n$. By (7.3.16) and (7.3.17), it then follows,

$$\begin{aligned}
\frac{1}{n}\mathbf{E}\Big\{\sum_{t=1}^{n} (X_t)^2\Big\} &= \frac{1}{n}\mathbf{E}\Big\{\sum_{t=1}^{n}\Big(\Gamma_t^1 \hat{S}_t + g_t(Y^{t-1}) + Z_t\Big)^2\Big\} && (7.3.27)\\
&= \frac{1}{n}\mathbf{E}\Big\{\sum_{t=1}^{n}\Big(\Gamma_t^1 \hat{S}_t + g_t(Y^{t-1})\Big)^2\Big\} + \sum_{t=1}^{n} K_{Z_t},\\
&\quad \text{by indep. of } Z_t \text{ and } (V^{t-1}, \hat{S}_t, Y^{t-1}). && (7.3.28)
\end{aligned}$$

By mean-square estimation theory, then the choice of $g(\cdot)$ that minimizes the right hand of (7.3.28) is

$$g_t(Y^{t-1}) = g_t^*(Y^{t-1}) = -\Gamma_t^1 \mathbf{E}\Big\{\hat{S}_t \Big| Y^{t-1}\Big\} = -\Gamma_t^1 \widehat{\hat{S}}_t, \quad t = 1, \ldots, n. \tag{7.3.29}$$

Hence, $\Gamma_t^2 = -\Gamma_t^1, \forall t$. Let $\Lambda_t \triangleq \Gamma_t^1, \forall t$, and substitute into the recursion (7.3.19), to obtain (2.2.104), and into the average power (7.3.28), to obtain (2.2.105). Hence, the derivation of (2.2.100)-(2.2.105) is completed.

It then follows that (2.2.106)-(3.2.46) are the generalized Kalman-filter recursions, of estimating the new state process $\hat{S}_t, t = 1, \ldots, n$ that satisfies recursion (7.3.20), from the channel output process Y_t that satisfies the recursion (2.2.104).

(c) By the entropy of Gaussian RVs, upon substituting (7.3.26), (2.1.49), into (2.2.88), then it follows (3.1.24). By substituting (2.2.110) and (2.2.111) into (3.1.24), then it follows (2.2.116). This completes the proof.

7.4 Proof of Corollary 3.1.1

First, note that the analog of Theorem 3.1.1.(a), for the code $(s, 2^{nR}, n)$, $n = 1, 2, \ldots$ is (3.1.1) and (3.1.2), because $P_t(dx_t|x^{t-1}, y^{t-1}, s) = \overline{P}_t(dx_t|v^{t-1}, y^{t-1}, s), t = 1, \ldots, n$. Define $\overline{\mathscr{P}}^s_{[0,n]}(\kappa)$ as in (3.1.2) with x^{t-1} replaced by $v^{t-1}, t = 1, \ldots, n$.

(a) Then

$$\begin{aligned} P_t(dx_t|x^{t-1}, y^{t-1}, s) =& P_t(dx_t|v^{t-1}, y^{t-1}, s_0), \quad t = 1, \ldots, n, \quad \text{by } Y_t = X_t + V_t && (7.4.30)\\ =& \overline{P}_t(dx_t|s^t, y^{t-1}, s), \qquad \text{Definition 3.1.1.} && (7.4.31) \end{aligned}$$

The PO-SS realization, for fixed $S_1 = s$ is then

$$\begin{aligned} V_t &= C_t S_t + N_t W_t, \quad S_1 = s, \quad t = 1, \ldots, n, && (7.4.32)\\ S_{t+1} &= A_t S_t + B_t W_t, \quad S_1 = s. && (7.4.33) \end{aligned}$$

Then

$$\begin{aligned} \mathbf{P}_t(dy_t\big|x^t, y^{t-1}, s) =& \mathbf{P}_t(dy_t\big|x^t, v^{t-1}, y^{t-1}, s) && (7.4.34)\\ & \mathbf{P}_t(dy_t\big|x^t, s^t, y^{t-1}, s), \qquad \text{by Definition 3.1.1, (A1)} && (7.4.35)\\ =& \mathbf{P}_t(dy_t\big|x_t, s^t), \qquad \text{by } Y_t = X_t + V_t \text{ and (7.4.32)} && (7.4.36)\\ \overset{(a)}{=}& \mathbf{P}_t(dy_t|x_i, s_t, s), \qquad \text{by mutually independence of } (W_1, \ldots, W_n, S_1). && (7.4.37) \end{aligned}$$

The probability distribution $\mathbf{P}_t(dy_t|y^{t-1}, s)$ is then given by

$$\begin{aligned} \mathbf{P}_t^{\overline{P}}(dy_t\big|y^{t-1}, s) = \int & \mathbf{P}_t(dy_t|x_t, s_t)\mathbf{P}_t(dx_t|s_t, y^{t-1}, s)\\ & \otimes \mathbf{P}_t^{\overline{P}}(ds_t\big|y^{t-1}, s), \quad t = 1, \ldots, n, \qquad \text{by reconditioning and (7.4.37).} && (7.4.38) \end{aligned}$$

The pay-off is the sum of conditional entropies $\sum_{t=1}^{n} H(Y_t|Y^{t-1},s)$, and the constraint is (3.1.2). By Definition 1.1.2, the state $S_t, t=2,\ldots,n$ is Markov, that is, $\mathbf{P}_{S_t|S^{t-1}} = \mathbf{P}_{S_t|S_{t-1}}, t=2,\ldots,n$. By (7.4.38) and the Markov property of S^n, then, it can be shown that, at each time t, the input distribution $\mathbf{P}_t^{\overline{P}}(ds_t|y^{t-1},s)$ depends on $\mathbf{P}_j(dx_j|s_j,y^{j-1},s), j=1,\ldots,t-1$ and not on $\overline{P}_j(dx_j|s^j,y^{j-1},s), j=1,\ldots,t-1$. It then follows (3.1.3) and (3.1.4), by letting $\mathbf{P}_t(dx_t|s_t,y^{t-1},s) = \overline{P}_t^M(dx_t|s_t,y^{t-1},s), t=1,\ldots,n$. It should be noted that (3.1.3) and (3.1.4) also follow from a slight variation of the derivation given in [1, Theorem 1]. By the maximum entropy principle of Gaussian distributions it then follows that the distribution $\overline{P}_t^M(dx_t|s_t,y^{t-1},s)$, is conditionally Gaussian, with linear conditional mean and nonrandom conditional covariance, given by

$$\mathbf{E}^{\overline{P}_t^M}\Big\{X_t\Big|S_t,Y^{t-1},S\Big\} = \begin{cases} \Lambda_1 S + \Gamma_t S_t + \Gamma_t^2 Y^{t-1} & \text{for} \quad t=2,\ldots,n \\ 0, & \text{for} \quad t=1, \end{cases} \tag{7.4.39}$$

$$K_{X_t|S_t,Y^{t-1},S} \stackrel{\triangle}{=} cov\Big(X_t,X_t\Big|S_t,Y^{t-1},S\Big) = K_{Z_t} \succeq 0, \quad t=1,\ldots,n. \tag{7.4.40}$$

Then (7.4.39) and (7.4.40) follow by repeating the derivation of the same step in Theorem 2.2.1. This completes the derivation of all statements of part (a).
(b), (c). By part (a) and using the generalized Kalman-filter, as in Theorem 2.2.1, then the statements are shown.

7.5 Proof of Theorem 3.1.1

(a) Since we have assumed $S_1 = s$ is fixed, and known to the encoder and the decoder, then Theorem 3.1.1 still holds, by replacing all conditional distributions, expectations and entropies, by corresponding expressions with fixed $S_1 = s$. Hence, (2.1.30) is replaced by (3.1.25), and (2.1.23) is replaced by (3.1.26) (since the code is allowed to depend on $S_1 = s$). (b) From the PO-SS realization of Definition 1.1.2 with $S_1 = s$ fixed, it follows that a necessary condition for Conditions 1 of Section 1.1 to hold is (i). The expression of entropy (3.1.27) is easily obtained by invoking condition (i), and properties of conditional entropy. That is, $H(V_1|s) = H(C_1S_1 + N_1W_1|S_1 = s) = H(N_1W_1|S_1 = s) = H(N_1W_1)$ by independence of W_1 and S_1, and $H(V_2|V_1,s) = H(V_2|V_1,S_1 = s) = H(C_2S_2 + N_2W_2|C_1S_1 + N_1W_1, S_1 = s) = H(C_2S_2 + N_2W_2|N_1W_1, S_1 = s) = H(C_2A_1S_1 + C_2B_1W_1 + N_2W_2|N_1W_1,S_1 = s) = H(N_2W_2|N_1W_1,S_1 = s) = H(N_2W_2)$, etc. This completes the proof.

7.6 Proof of Proposition 3.3.1

Since the proof of [2, Theorem 6.1] is based [2, Lemma 6.1], where the channel input X_t is expressed as $X_t = \Lambda\Big(S_t - \mathbf{E}\Big\{S_t\Big|Y_{-\infty}^{t-1}\Big),t=1,\ldots,$ where Λ is a nonradom vectors, then (3.3.54) is a necessary for [2, Theorem 6.1] to hold. Next, we show Conditions 1 and 2 of Section 1.1 are necessary and sufficient for equality (3.3.54) to hold. To avoid complex notation, we prove the claim for the realization of Example 2.1.2.(a). Suppose the *initial state* S_1 of the noise is $S_1 = s_1$ is known to the encoder and the decoder, and without loss of generality take $s_1 = 0$, which by (2.1.71), implies $V_0 = 0, W_0 = 0$ (as often done in [2]). Then, the following hold.

$$S_1 = 0 \implies V_1 = W_1, \ \ S_2 = W_1 = V_1, \quad \text{by (2.1.72), (2.1.73),} \tag{7.6.41}$$

$$(S_1 = 0, V_1) \text{ uniquely define } S_2 = cS_1 + W_1 = W_1 = V_1, \quad \text{by (2.1.72), (2.1.73),} \tag{7.6.42}$$

$$V_2 = \Big(c-a\Big)S_2 + W_2, \ \ S_3 = cS_2 + W_2 = cV_1 + W_2, \quad \text{by (2.1.72), (2.1.73),} \tag{7.6.43}$$

$$(S_1 = 0, V_1, V_2) \text{ uniquely define } (S_2, S_3), \tag{7.6.44}$$

$$\text{repeating, then } (S_1 = 0, V_1, \ldots, V_{t-1}) \text{ uniquely define } (S_2, S_3, \ldots, S_t), \forall t = 3, 4, \ldots. \tag{7.6.45}$$

From (7.6.41)-(7.6.45) it then follows, that for any $S_1 = s_1$, including, $s_1 = 0$, known the the encoder that the equalities hold:

$$\mathbf{P}_{X_t|X^{t-1},Y_{-\infty}^{t-1},S_1} = \mathbf{P}_{X_t|V^{t-1},Y_{-\infty}^{t-1},S_1}, \quad \text{by } Y_t = X_t + V_t \tag{7.6.46}$$

$$= \mathbf{P}_{X_t|S^t,Y_{-\infty}^{t-1},S_1}, \quad t = 1, \ldots, \tag{7.6.47}$$

We can go one step further to identify the information structure of optimal channel input distributions using (7.6.47), that is, to show $\mathbf{P}_{X_t|S^t,Y_{-\infty}^{t-1},S_1} = \mathbf{P}_{X_t|S_t,Y_{-\infty}^{t-1},S_1}, \quad t = 1, \ldots,$ by repeating to proof of [1, Theorem 1]. However, for the statement of the proposition this is not necessary.
Suppose either $S_1 = s_1$ is not known to the encoder, i.e., $V_0 = v_0, W_0 = w_0$ are not known to the encoder, and $S_1 \neq 0$, while the optimal channel input is expressed as a function of the state of the noise, S^n, that is,

$$\mathbf{P}_{X_t|X^{t-1},Y_{-\infty}^{t-1}} = \mathbf{P}_{X_t|V^{t-1},Y_{-\infty}^{t-1}} = \mathbf{P}_{X_t|S^t,Y_{-\infty}^{t-1}}, \quad t = 1, \ldots, \tag{7.6.48}$$

Then by (2.1.72) and (2.1.73), it follows that $V_1 = \Big(c-a\Big)S_1 + W_1, S_2 = aS_1 + W_1$, hence knowledge of V_1 does not specify S_2, and similarly, V^t does not specify S^t, for $t = 2, 3, \ldots$. Hence, we arrive at the contradiction of equality (7.6.48). This competes the proof.

7.7 Proof of Proposition 4.1.1

(a) This is shown in [16] by using the Szego formula and Poisson's integral formula.

(b) By definition,

$$H(V^n|s_1) = \sum_{t=1}^{n} H(V_t|V^{t-1},s_1) \tag{7.7.49}$$

$$= H((c-a)S_1+W_1|s_1)+H((c-a)S_2+W_2|V_1,s_1)+\ldots+H((c-a)S_n+W_n|V^{n-1},s_1) \tag{7.7.50}$$

$$= H((W_1|s_1)+H((c-a)S_2+W_2|V_1,s_1,W_1)+\ldots+H((c-a)S_n+W_n|V^{n-1},s_1) \tag{7.7.51}$$

$$= H(W_1)+H((W_2)+\ldots+H((c-a)S_n+W_n|V^{n-1},s_1), \quad \text{by } S_2 = aS_1+W_1 \tag{7.7.52}$$

$$= \sum_{t=1}^{n} H(W_t), \quad \text{by repeating the proceedure} \tag{7.7.53}$$

Since $W_t \in N(0,K_W), K_W > 0, t = 1,\ldots,n$, then (4.1.8) is obtained.

7.8 Proof of Lemma 4.1.1

(a) This is due to Lemma 2.1.1.(v).

(b) By taking the per unit time limit (4.1.9), and utilizing the hypothesis (4.1.11), the continuity of the $\log(\cdot)$ and the fact that, for any convergent sequence $a_n, n = 1,2,\ldots$, i.e., $\lim_{n\longrightarrow\infty} a_n = a$, then $\frac{1}{n}\sum_{t=1}^{n} a_n \longrightarrow a$, as $n \longrightarrow \infty$, then it follows (4.1.12).

7.9 Proof of Lemma 4.2.1

From Corollary 3.2.3.(a) we deduce that $\Sigma_t^o \triangleq \Sigma_t, t = 1,\ldots,n$ satisfies (4.2.25) with initial condition (4.2.26). By Definition 5.2.3 the corresponding generalized algebraic Riccati equation is (4.2.27), and pairs $\{A,C\}$ and $\{A^*, GB^{*,\frac{1}{2}}\}$ are given by (4.2.28).

(1) By Definition 5.2.3, for $c \neq a$ the pair $\{A,C\} = \{c,c-a\}$ is observable, and hence detectable.

(2) By Definition 5.2.3, the pair $\{A^*, GB^{*,\frac{1}{2}}\} = \{a,0\}$ is unit circle controllable if and only if $|a| \neq 1$.

(3) By Definition 5.2.3, the pair $\{A^*, GB^{*,\frac{1}{2}}\} = \{a,0\}$ is stabilizable if and only if $a \in (-1,1)$.

(4) This follows from Theorem 5.2.1.(1) and parts (1), (2) and (3). Since (4.2.27) is a quadratic equation we can solved it explicitly to verify the two solutions are $\Sigma^\infty = 0$ and $\Sigma^\infty = \frac{K_W\left(a^2-1\right)}{\left(c-a\right)^2}$, and the statement of (4.2.29).

(5) For values $c \in (-\infty,\infty)$ and $|a| < 1$, the pair $\{A,C\} = \{c,c-a\}$ is detectable and the pair

$\{A^*, GB^{*,\frac{1}{2}}\} = \{a, 0\}$ is stabilizable, and the statement follows from Theorem 5.2.1.(3).
(6) (4.2.30) follows from Lemma 4.1.1.(b), by invoking Corollary 3.2.3.(a), i.e., $K_{\hat{I}_t} = \big(c-a\big)^2 \Sigma_t^o + K_W, t = 1, \ldots, n$, where Σ_t^o is generated by (4.2.25), and part (5).

7.10 proof of Lemma 5.2.2

(1) We utilize the equation 5.2.110 and the scalar representation of the equation 4.2.13, of the Definition 4.2.1,

$$K^\infty = A^2 K^\infty + G^2 Q - \Big(AK^\infty C + GS\Big)^2 \Big(R + C^2 K^\infty\Big)^{-1}, \quad K_1 = \text{given}. \tag{7.10.54}$$

Hence, $A = c, C = c - a + \Lambda$. By $c \in (-1,1)$, then there always exists a $P \in \mathbb{R}$, such that $|A - PC| = |c - P(c - a + \Lambda)| < 1$, i.e., take $P = 1$. This shows (1).
(2) Since $K_Z = 0$, then $B = 1 - K_W\big(K_Z + K_W\big)^{-1} = 0, B^{*,\frac{1}{2}} = K_W^{\frac{1}{2}} B^{\frac{1}{2}} = 0$, $A^* = c - K_W\big(K_W + K_Z\big)^{-1} C = c - \big(c - a + \Lambda\big) = a - \Lambda$, and hence, there exists a $P \in \mathbb{R}$, such that $|A^* - B^{*,\frac{1}{2}} P| = |a - \Lambda| \neq 1$, if and only if $|a - \Lambda| \neq 1$. This shows (2).
(3) Since $K_Z = 0$, similar to the prove in (2), there exists a $P \in \mathbb{R}$, such that $|A^* - B^{*,\frac{1}{2}} P| = |a - \Lambda| < 1$, if and only if $|a - \Lambda| < 1$. 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 $|a - \Lambda| \neq 1$ and by (3) the pair $\{A^*, B^{*,\frac{1}{2}}\}$ is stabilizable, if and only if $|a - \Lambda| < 1$. By Theorem 5.2.1.(1) we deduce the claim. This shows (4).
(5) Clearly, (5.2.110) is equivalent to the quadratic equation

$$K^2 \big(\Lambda + c - a\big)^2 - K\big(\big(\Lambda - a\big)^2 - 1\big) K_W = 0. \tag{7.10.55}$$

Hence, the two solutions are (5.2.111). The last statement is also obtained by applying Theorem 5.2.1.(2), as follows. By (1), $\{A, C\}$ is detectable, by (2), $\{A^*, B^{*,\frac{1}{2}}\}$ is unit circle controllable, if and only if $|a - \Lambda| \neq 1$, and by (3), $\{A^*, B^{*,\frac{1}{2}}\}$ is stabilizable, if and only if $|a - \Lambda| < 1$. By invoking Theorem 5.2.1.(2), the non-negative solution $K = 0$ is unique and stabilizable, if and only if $|a - \Lambda| < 1$.

7.11 proof of Lemma 5.2.3

At this point we calculate the maximum feedback capacity from Theorem 5.2.2, driven by $ARMA(a, c)$ noise of Definition 5.0.1 and we examine for which regions we have a feedback capacity, using the necessary conditions of Lemma 5.2.3.

(i) By the stationarity conditions of Lemma 5.2.3, i.e., (5.2.123)-(5.2.126), with $\lambda=(\lambda_1,\lambda_2,\lambda_3,\lambda_4)$:

$$\frac{\partial}{\partial K_Z^{\infty}}\mathscr{L}(\Lambda^{\infty,*},K_Z^{\infty,*},K^{\infty,*},\lambda^*)=1-\lambda_1^*\Big(K^{\infty,*}-c^2K^{\infty,*}-K_W\Big)-\lambda_2^*+\lambda_4^*=0, \tag{7.11.56}$$

$$\begin{aligned}&\frac{\partial}{\partial \Lambda^{\infty}}\mathscr{L}(\Lambda^{\infty,*},K_Z^{\infty,*},K^{\infty,*},\lambda^*)=\Big(\Lambda^{\infty,*}+c-a\Big)K^{\infty,*}-\lambda_1^*\Big\{\Big(K^{\infty,*}-c^2K^{\infty,*}-K_W\Big)\\&\Big(\Lambda^{\infty,*}+c-a\Big)K^{\infty,*}+\Big(K_W+cK^{\infty,*}(\Lambda^{\infty,*}+c-a)\Big)cK^{\infty,*}\Big\}-\lambda_2^*\Lambda^{\infty,*}K^{\infty,*}=0,\end{aligned} \tag{7.11.57}$$

$$\begin{aligned}&\frac{\partial}{\partial K^{\infty}}\mathscr{L}(\Lambda^{\infty},K_Z^{\infty},K^{\infty},\lambda^*)=\Big(\Lambda^{\infty,*}+c-a\Big)^2-\lambda_1^*\Big\{\Big(1-c^2\Big)\Big(K_Z^*+K_W\\&+(\Lambda^{\infty,*}+c-a)^2K^{\infty,*}\Big)+\Big(K^{\infty,*}-c^2K^{\infty,*}-K_W\Big)\Big(\Lambda^{\infty,*}+c-a\Big)^2+\\&2c\Big(K_W+cK^{\infty,*}(\Lambda^{\infty,*}+c-a)\Big)\Big(\Lambda^{\infty,*}+c-a\Big)\Big\}-\lambda_2^*\Big(\Lambda^{\infty,*}\Big)^2+\lambda_3^*=0,\end{aligned} \tag{7.11.58}$$

First of all, we will find the values of the Lagrangian multipliers $\lambda=(\lambda_1,\lambda_2,\lambda_3,\lambda_4)$, through some steps.
Suppose $\lambda_4^*\neq 0$. Then, by complementary slackness (5.2.126), we have $\lambda_4^*K_Z^{\infty,*}=0$, which implies $K_Z^{\infty,*}=0$, and hence we have $K^{\infty,*}=0$. By complementary slackness (5.2.126), we also have $\lambda_2^*\big((\Lambda^{\infty,*})^2K^{\infty,*}+K_Z^{\infty,*}-\kappa\big)=\lambda_2^*(0-\kappa)=0$, hence for any $\kappa>0$, $\lambda_2^*=0$. This implies $C^{\infty}(\kappa)=0, \kappa\in[0,\infty)$ hence the rate is zero. Similarly, if $\lambda_3^*\neq 0$, then $K_Z^{\infty,*}=0$ and $K^{\infty,*}=0$, which lead to a zero rate. However, by (i), i.e., by Theorem 5.2.3, we know that for $\Lambda^{\infty,*}=0, K_Z^{\infty,*}\neq 0$, we exhibit a non-zero rate, which is a lower bound on the nofeedback rate. Hence, for the rest of the derivation we characterize the set of all values $\kappa\in\mathscr{K}^{\infty}(a,c,K_W)$, that is, we assume $\lambda_3^*=0, \lambda_4^*=0$, and treat the case $\kappa\notin\mathscr{K}^{\infty}(a,c,K_W)$ separately.
(iii) By complementary slackness (5.2.126), then $\lambda_4^*K_Z^{\infty,*}=0$. Suppose $\lambda_4^*=0$. By (7.11.56), then

$$1-\lambda_1^*\Big(K^{\infty,*}-c^2K^{\infty,*}-K_W\Big)-\lambda_2^*=0, \tag{7.11.59}$$

$$K^{\infty,*}\in[0,\infty)\quad \text{iff}\quad K^{\infty,*}=\frac{1-\lambda_2^*+\lambda_1^*K_W}{\lambda_1^*\Big(1-c^2\Big)}\geq 0,\ \ \lambda_1^*\neq 0,\ \ c^2\neq 1. \tag{7.11.60}$$

(iv) By (7.11.57), we have

$$\begin{aligned}&\Big(\Lambda^{\infty,*}+c-a\Big)K^{\infty,*}-\lambda_1^*\Big\{\Big(K^{\infty,*}-c^2K^{\infty,*}-K_W\Big)\Big(\Lambda^{\infty,*}+c-a\Big)K^{\infty,*}\\&+\Big(K_W+cK^{\infty,*}(\Lambda^{\infty,*}+c-a)\Big)cK^{\infty,*}\Big\}-\lambda_2^*\Lambda^{\infty,*}K^{\infty,*}=0.\end{aligned} \tag{7.11.61}$$

By complementary slackness (5.2.126), then $\lambda_3^* K^{\infty,*} = 0$. Suppose $\lambda_3^* = 0$. By (7.11.61), with $K^{\infty,*} > 0$,

$$\Big(\Lambda^{\infty,*} + c - a\Big) - \lambda_1^*\Big\{\Big(K^{\infty,*} - c^2 K^{\infty,*} - K_W\Big)\Big(\Lambda^{\infty,*} + c - a\Big) \\ + \Big(K_W + cK^{\infty,*}(\Lambda^{\infty,*} + c - a)\Big)c\Big\} - \lambda_2^* \Lambda^{\infty,*} = 0, \quad K^{\infty,*} > 0. \tag{7.11.62}$$

Substituting λ_2^* from (7.11.59) into (7.11.62), and solving for λ_1^* we obtain

$$\lambda_1^* = \frac{c-a}{K^{\infty,*}(\Lambda^{\infty,*}c^2 + c - a) + aK_W} \tag{7.11.63}$$

Substituting λ_1^* into (7.11.59) we also obtain

$$\lambda_2^* = 1 - \frac{(a-c)\big(K^{\infty,*}(c^2-1) + K_W\big)}{K^{\infty,*}(\Lambda^{\infty,*}c^2 + c - a) + aK_W} \tag{7.11.64}$$

(v) Now, we show $\lambda_1^* \neq 0$ (although it follows from the above calculations). Suppose $\lambda_3^* = 0, \lambda_4^* = 0$, and $\lambda_1^* = 0$. Then (7.11.57) is given by $2K^{\infty,*}\big(-a + \Lambda^{\infty,*} + c - \Lambda^{\infty,*}\lambda_2^*\big) = 0$, and is satisfied if

$$K^{\infty,*} = 0 \quad \text{or} \quad -a + \Lambda^{\infty,*} + c - \Lambda^{\infty,*}\lambda_2^* = 0 \tag{7.11.65}$$

For $K^{\infty,*} = 0$, then $K_Z^{\infty,*} = 0$, and by (7.11.59) we have $\lambda_2^* = 1$. By the complementary slackness, $\lambda_2^*\big((\Lambda^{\infty,*})^2 K^{\infty,*} + K_Z^{\infty,*} - \kappa\big) = 0$, then $\lambda_2^\infty(-\kappa) = 0$, hence $\lambda_2^* = 0$, unless $\kappa = 0$. This contradicts the value $\lambda_2^* = 1$. Hence, $\lambda_1^* = 0$ and $K^{\infty,*} = 0$ are not possible choices. Suppose $\lambda_1^* = 0$ and $K^{\infty,*} > 0$, hence (7.11.65) holds, and by (7.11.59) we have $\lambda_2^* = 1$. By (7.11.65) then $c = 0$. Since, $c \neq 0$, otherwise the channel is driven by $MA(a)$ noise and in Theorem (na valw to applicaiton) we will show that there isn't the optimum solution. Then, the only choice is $\lambda_1^* \neq 0$.

(vi) Suppose $\lambda_3^* = 0, \lambda_4^* = 0$, and $\lambda_2^* = 0$. Now,we show by contradiction that $\lambda_2^* \neq 0$. By (7.11.59) then

$$1 - \lambda_1^*\Big(K^{\infty,*} - c^2 K^{\infty,*} - K_W\Big) = 0. \tag{7.11.66}$$

Substituting (7.11.66) into (7.11.57) we have

$$\lambda_1^*\Big(K_W + cK^{\infty,*}(\Lambda^{\infty,*} + c - a)\Big)cK^{\infty,*} = 0. \tag{7.11.67}$$

Substituting (7.11.66) into (7.11.58) we have

$$-\lambda_1^*\Big\{\Big(1 - c^2\Big)\Big(K_Z^* + K_W + (\Lambda^{\infty,*} + c - a)^2 K^{\infty,*}\Big) \\ + 2c\Big(K_W + cK^{\infty,*}(\Lambda^{\infty,*} + c - a)\Big)\Big(\Lambda^{\infty,*} + c\Big)\Big\} = 0. \tag{7.11.68}$$

Substituting (7.11.66) into the ARE (5.2.130) we have

$$\frac{1}{\lambda_1^*}\Big(K_Z^{\infty,*}+K_W+(\Lambda^{\infty,*}+c-a)^2K^{\infty,*}\Big)+\Big(K_W+cK^{\infty,*}(\Lambda^{\infty,*}+c-a)\Big)^2=0. \tag{7.11.69}$$

By (7.11.66) then $\lambda_1^*\neq 0$. Since $\lambda_1^*\neq 0$, then by (7.11.67), we have $K^{\infty,*}=0$ or $K_W+cK^{\infty,*}\big(\Lambda^{\infty,*}+c-a\big)=0$. However, either choice, when substituted into (7.11.69) implies $K_W=0$, which contradicts our assumption that $K_W>0$. Hence, we deduce $\lambda_2^*\neq 0$.

(vii) By the above analysis we consider $\lambda_3^*=0,\lambda_4^*=0$, $\lambda_1^*\neq 0,\lambda_2\neq 0$, and (7.11.63), (7.11.64) hold. By (7.11.58), then the following holds:

$$\begin{aligned}&\Big(\Lambda^{\infty,*}+c-a\Big)^2-\lambda_1^*\Big\{\Big(1-c^2\Big)\Big(K_Z^{\infty,*}+K_W+(\Lambda^{\infty,*}+c-a)^2K^{\infty,*}\Big)\\&+\Big(K^{\infty,*}-c^2K^{\infty,*}-K_W\Big)\Big(\Lambda^{\infty,*}+c-a\Big)^2\\&+2c\Big(K_W+cK^{\infty,*}(\Lambda^{\infty,*}+c-a)\Big)\Big(\Lambda^{\infty,*}+c-a\Big)\Big\}-\lambda_2^*\Big(\Lambda^{\infty,*}\Big)^2=0,\quad K^{\infty,*}>0.\end{aligned} \tag{7.11.70}$$

Since we have $\big(\Lambda^{\infty,*}\big)^2K^{\infty,*}+K_Z^{\infty,*}=\kappa$, then $K_Z^{\infty,*}=\kappa-\big(\Lambda^{\infty,*}\big)^2K^{\infty,*}\in(0,\infty)$. Substituting $(\lambda_1^*,\lambda_2^*)$ and $K_Z^{\infty,*}$ into (7.11.70), we obtain

$$\begin{aligned}K^{\infty,*}&=\frac{\kappa\Big(1-c^2\Big)+K_W\Big(1-ac\Big)}{\Big(a-c\Big)\Big(\Lambda^{\infty}(2-c^2)+c-a\Big)}>0\iff\\ \Lambda^{\infty,*}&=\frac{K_W\Big(ac-1\Big)+\kappa\Big(c^2-1\Big)-K^{\infty,*}\Big(a-c\Big)^2}{K^{\infty,*}\Big(c^2-2\Big)\Big(a-c\Big)}\end{aligned} \tag{7.11.71}$$

Substituting $\Lambda^{\infty,*}$ and $K_Z^{\infty,*}$ into the generalized ARE (5.2.130) we obtain:

$$\begin{aligned}&\frac{\Big(K^{\infty,*}c^2-K^{\infty,*}+K_W\Big)\Big(K_W(2a-c)+c\kappa-K^{\infty,*}c(a-c)^2\Big)}{c^2-2}+\\&\frac{\Big(c^2-1\Big)^2\Big(K_W(2a-c)+c\kappa-K^{\infty,*}c(a-c)^2\Big)^2}{\Big(c^2-2\Big)^2\Big(a-c\Big)^2}=0\end{aligned} \tag{7.11.72}$$

We assume that $c\neq a$ and $c^2\neq 2$.

$$\left(K^{\infty,*}\right)^2\left(\frac{c^2(c^2-1)(a-c)^2}{(c^2-2)^2}\right)+$$
$$K^{\infty,*}\left(\frac{c^2\left(K_W(c^2+2a^2-2ac-a^2c^2)+c^2\kappa(1-c^2)\right)}{\left(c^2-2\right)^2}\right)+$$
$$\frac{\left(K_W(2a-c)+c\kappa\right)\left(\left(c^2-1\right)^2\left(K_W(2a-c)+c\kappa\right)+cK_W(c^2-2)^2(a-c)^2\right)}{\left(a-c\right)^2\left(c^2-2\right)^2}=0 \tag{7.11.73}$$

$$K_{1,2}^{\infty,*}=\frac{-c\left(K_W(2a^2+c^2-a^2c^2-2ac)+c^2\kappa(1-c^2)\right)}{2c\left(c^2-1\right)\left(a-c\right)^2}$$
$$\pm\frac{\sqrt{\left(c^2-2\right)^2\left(K_W(c-2a+a^2c)+c\kappa(c^2-1)\right)^2}}{2c\left(c^2-1\right)\left(a-c\right)^2} \tag{7.11.74}$$

Also, we assume that $c \neq 0$ and $c^2 \neq 1$.

Hence, we have two solutions:

$$K_1^{\infty,*}=\frac{-c\left(K_W(2a^2+c^2-a^2c^2-2ac)+c^2\kappa(1-c^2)\right)+\left|c^2-2\right|\left|K_W(c-2a+a^2c)+c\kappa(c^2-1)\right|}{2c\left(c^2-1\right)\left(a-c\right)^2} \tag{7.11.75}$$

$$K_2^{\infty,*}=\frac{-c\left(K_W(2a^2+c^2-a^2c^2-2ac)+c^2\kappa(1-c^2)\right)-\left|c^2-2\right|\left|K_W(c-2a+a^2c)+c\kappa(c^2-1)\right|}{2c\left(c^2-1\right)\left(a-c\right)^2} \tag{7.11.76}$$

The variable c can take any value except $c \neq a$, $c \neq 0$, $c^2 \neq 1$ and $c^2 \neq 2$. Thus, we deduce the following cases of solutions.

Case 1. $c > \sqrt{2}$. In addition, $K_W(c-2a+a^2c)+c\kappa(c^2-1) > 0$, always holds, because $c > 0$, $\kappa \geq 0$, $c^2-1 > 0$ and $K_W > 0$. Also, $c-2a+a^2c$ is positive, because the discriminant of this expression is negative.

Thus, we have $\kappa > \frac{-K_W(c-2a+a^2c)}{c(c^2-1)}$ which always holds.

The first solution is

$$K_1^{\infty,*} = \frac{K_W(2a-c+a^2c^3-2a^2c)+c\kappa(c^2-1)^2}{c(c^2-1)(a-c)^2}, \tag{7.11.77}$$

$$\Lambda_1^{\infty,*} = \frac{K_W(a-c)^2(1-ac)}{K_W(2a-c+a^2c^3-2a^2c)+c\kappa(c^2-1)^2} \tag{7.11.78}$$

$$K_{Z_1}^{\infty,*} = \frac{\kappa\Big(c(c^2-1)(K_W(2a-c+a^2c^3-2a^2c)+c\kappa(c^2-1)^2\Big)-K_W^2(a-c)^2(1-ac)^2}{c(c^2-1)\Big(K_W(2a-c+a^2c^3-2a^2c)+c\kappa(c^2-1)^2\Big)}. \tag{7.11.79}$$

$$\lambda_1^* = \frac{c(a-c)^2}{K_W(c-2a+a^2c)+c\kappa(c^2-1)}, \quad \lambda_2^* = c^2. \tag{7.11.80}$$

The second solution is

$$K_2^{\infty,*} = \frac{K_W(2a-c)+c\kappa}{c(a-c)^2}, \tag{7.11.81}$$

$$\Lambda_2^{\infty,*} = \frac{(a-c)(aK_W+c\kappa)}{K_W(c-2a)-c\kappa}, \tag{7.11.82}$$

$$K_{Z_2}^{\infty,*} = \frac{K_W(K_Wa^2+\kappa c^2)}{c(2aK_W+cK_W-c\kappa)}. \tag{7.11.83}$$

$$\lambda_1^* = -\frac{c(a-c)^2}{K_W(c-2a+a^2c)+c\kappa(c^2-1)}, \quad \lambda_2^* = 0. \tag{7.11.84}$$

The second solution is always rejected, because λ_2 should be $\lambda_2 \neq 0$. Thus, we will continue with the one and only valid solution $K_1^{\infty,*}$. The case 1 and the rest of the cases that we show below, are mentioned to the feedback case. So, we should check for every case whether $K_1^{\infty,*}$ is valid. More specifically, we need to check if $K_1^{\infty,*} > 0$, $K_{Z_1}^{\infty,*} > 0$ and $\lambda_1^* > 0$.
Set,

$$\kappa_1 = \frac{K_W(c-2a+a^2c)-K_Wa^2c(c^2-1)}{c(c^2-1)^2}, \tag{7.11.85}$$

$$\kappa_2 = \frac{K_W(1-ac)(ac^3-2ac+c^2-\sqrt{c^3(a^2c^3-6ac^2+4a+4c^3-3c)})}{2c^2(c^2-1)^2}, \tag{7.11.86}$$

$$\kappa_3 = \frac{K_W(1-ac)(ac^3-2ac+c^2+\sqrt{c^3(a^2c^3-6ac^2+4a+4c^3-3c)})}{2c^2(c^2-1)^2}. \tag{7.11.87}$$

$K_1^{\infty,*} > 0$, for $\kappa > \kappa_1$ and $\kappa \geq 0$, for $a \in \left[\frac{-c}{c^2-2}, \frac{1}{c}\right]$,
$\lambda_1^{\infty,*} > 0$, for $\kappa \in [0,\infty)$ and
$K_{Z_1}^{\infty,*} > 0$, for $\kappa \in [0,\kappa_2) \cup (\kappa_3,\infty)$.

Thus, we consider a set $\mathscr{A}_1(a,c,K_W)$, which is not empty.

$$\mathscr{A}_1(a,c,K_W) = \mathscr{A}_1^1(a,c,K_W) \cap \mathscr{A}_1^2(a,c,K_W) \tag{7.11.88}$$

$$\mathscr{A}_1^1(a,c,K_W) \triangleq \left\{\kappa \in [0,\infty) : K_1^{\infty,*} > 0, \lambda_1^{\infty,*} > 0, c \in (\sqrt{2},\infty), a \in \left[\frac{-c}{c^2-2}, \frac{1}{c}\right], \kappa > \kappa_1\right\} \tag{7.11.89}$$

$$\mathscr{A}_1^2(a,c,K_W) \triangleq \left\{\kappa \in [0,\infty), a \in (-\infty,\infty) : K_{Z_1}^{\infty,*} > 0, c \in (\sqrt{2},\infty), \kappa \in [0,\kappa_2) \cup (\kappa_3,\infty)\right\} \tag{7.11.90}$$

Case 2. $c < -\sqrt{2}$. In addition, $K_W(c-2a+a^2c) + c\kappa(c^2-1) < 0$, always holds, because $c < 0$, $\kappa \geq 0$, $c^2-1 < 0$ and $K_W > 0$. Also, $c-2a+a^2c$ is negative, because the discriminant of this expression is negative.

Thus, we have $\kappa > \frac{-K_W(c-2a+a^2c)}{c(c^2-1)}$ which always holds.

The only valid solution is (7.11.77)-(7.11.80).

$K_1^{\infty,*} > 0$, for $\kappa > \kappa_1$ and $\kappa \geq 0$, for $a \in \left(-\infty, \frac{1}{c}\right] \cup \left[\frac{-c}{c^2-2}, \infty\right)$,

$\lambda_1^{\infty,*} > 0$, for $\kappa \in [0,\infty)$ and

$K_{Z_1}^{\infty,*} > 0$, for $\kappa \in [0,\kappa_2) \cup (\kappa_3,\infty)$.

Thus, we consider a set $\mathscr{A}_2(a,c,K_W)$, which is not empty.

$$\mathscr{A}_2(a,c,K_W) = \mathscr{A}_2^1(a,c,K_W) \cap \mathscr{A}_2^2(a,c,K_W) \tag{7.11.91}$$

$$\mathscr{A}_2^1(a,c,K_W) \triangleq \Big\{\kappa \in [0,\infty) : K_1^{\infty,*} > 0, \lambda_1^{\infty,*} > 0, c \in (-\infty,-\sqrt{2}),$$
$$a \in \left(-\infty, \frac{1}{c}\right] \cup \left[\frac{-c}{c^2-2}, \infty\right), \kappa > \kappa_1\Big\} \tag{7.11.92}$$

$$\mathscr{A}_2^2(a,c,K_W) \triangleq \left\{\kappa \in [0,\infty), a \in (-\infty,\infty) : K_{Z_1}^{\infty,*} > 0, c \in (-\infty,-\sqrt{2}), \kappa \in [0,\kappa_2) \cup (\kappa_3,\infty)\right\} \tag{7.11.93}$$

Case 3. $\sqrt{2} > c > 1$. In addition, $K_W(c-2a+a^2c) + c\kappa(c^2-1) > 0$, always holds, because $c > 0$, $\kappa \geq 0$, $c^2-1 > 0$ and $K_W > 0$. Also, $c-2a+a^2c$ is positive, because the discriminant of this expression is negative.

Thus, we have $\kappa > \frac{-K_W(c-2a+a^2c)}{c(c^2-1)}$ which always holds.

The only valid solution is (7.11.77)-(7.11.80).

$K_1^{\infty,*} > 0$, for $\kappa > \kappa_1$ and $\kappa \geq 0$, for $a \in \left[\frac{-c}{c^2-2}, \frac{1}{c}\right]$,

$\lambda_1^{\infty,*} > 0$, for $\kappa \in [0,\infty)$ and

$K_{Z_1}^{\infty,*} > 0$, for $\kappa \in [0,\kappa_2) \cup (\kappa_3,\infty)$.

Thus, we consider a set $\mathscr{A}_3(a,c,K_W)$, which is not empty.

$$\mathscr{A}_3(a,c,K_W) = \mathscr{A}_3^1(a,c,K_W) \cap \mathscr{A}_3^2(a,c,K_W) \tag{7.11.94}$$

$$\mathscr{A}_3^1(a,c,K_W) \triangleq \Big\{\kappa \in [0,\infty) : K_1^{\infty,*} > 0, \lambda_1^{\infty,*} > 0, c \in (1,\sqrt{2}), a \in \Big[\frac{-c}{c^2-2}, \frac{1}{c}\Big], \kappa > \kappa_1\Big\} \tag{7.11.95}$$

$$\mathscr{A}_3^2(a,c,K_W) \triangleq \Big\{\kappa \in [0,\infty), a \in (-\infty,\infty) : K_{Z_1}^{\infty,*} > 0, c \in (1,\sqrt{2}), \kappa \in [0,\kappa_2) \cup (\kappa_3,\infty)\Big\} \tag{7.11.96}$$

Case 4. $-\sqrt{2} < c < -1$. In addition, $K_W(c-2a+a^2c) + c\kappa(c^2-1) < 0$, always holds, because $c < 0$, $\kappa \geq 0$, $c^2 - 1 < 0$ and $K_W > 0$. Also, $c - 2a + a^2c$ is negative, because the discriminant of this expression is negative.

Thus, we have $\kappa > \frac{-K_W(c-2a+a^2c)}{c(c^2-1)}$ which always holds.

The only valid solution is (7.11.77)-(7.11.80).

$K_1^{\infty,*} > 0$, for $\kappa > \kappa_1$ and $\kappa \geq 0$, for $a \in \Big(-\infty, \frac{1}{c}\Big] \cup \Big[\frac{-c}{c^2-2}, \infty\Big)$,

$\lambda_1^{\infty,*} > 0$, for $\kappa \in [0,\infty)$ and

$K_{Z_1}^{\infty,*} > 0$, for $\kappa \in [0,\kappa_2) \cup (\kappa_3,\infty)$.

Thus, we consider a set $\mathscr{A}_4(a,c,K_W)$, which is not empty.

$$\mathscr{A}_4(a,c,K_W) = \mathscr{A}_4^1(a,c,K_W) \cap \mathscr{A}_4^2(a,c,K_W) \tag{7.11.97}$$

$$\mathscr{A}_4^1(a,c,K_W) \triangleq \Big\{\kappa \in [0,\infty) : K_1^{\infty,*} > 0, \lambda_1^{\infty,*} > 0, c \in (-\infty,\sqrt{2}),$$

$$a \in \Big(-\infty, \frac{1}{c}\Big] \cup \Big[\frac{-c}{c^2-2}, \infty\Big), \kappa > \kappa_1\} \tag{7.11.98}$$

$$\mathscr{A}_4^2(a,c,K_W) \triangleq \Big\{\kappa \in [0,\infty), a \in (-\infty,\infty) : K_{Z_1}^{\infty,*} > 0, c \in (-\infty,\sqrt{2}), \kappa \in [0,\kappa_2) \cup (\kappa_3,\infty)\Big\} \tag{7.11.99}$$

We saw that for any of the sets, $\mathscr{A}_1(a,c,K_W)$, $\mathscr{A}_2(a,c,K_W)$, $\mathscr{A}_3(a,c,K_W)$, $\mathscr{A}_4(a,c,K_W)$, that is, unstable noise, the conditions $K_1^{\infty,*} > 0$, $K_{Z_1}^{\infty,*} > 0$ and $\lambda_1^* > 0$ always hold. In conclusion, for any unstable case (i.e., $|c| > 1$), we have a non-zero feedback capacity. Also, there is a minimum power constraint, which is positive (we take the intersection of all the power constraints for any case). Otherwise, whether the power κ is less than the constraint, it means that we have an achievable rate (see Theorem 5.2.3).

Now, we have to check for feedback capacity, for the stable cases $|c| < 1$, $c \neq 0$.

Earlier we saw that the second absolute value was always positive or negative for each case. In the following cases, this is not obvious, because the discriminant of $c - 2a + a^2c$ is positive. More specifically, for $c \in (-1,0)$, we see that,

Solving the quadratic equation, $c - 2a + a^2c$, we have,

$$\delta = \beta^2 - 4\alpha\gamma = c^2 - 1 > 0$$

$$\alpha_1 = \frac{1-\sqrt{1-c^2}}{c}, \quad \alpha_2 = \frac{1+\sqrt{1-c^2}}{c} \text{ hence,} \tag{7.11.100}$$

$$c - 2a + a^2c > 0, \text{ for } a \in (a_2, a_1) \text{ and } c - 2a + a^2c < 0, \text{ for } a \in (-\infty, a_2) \cup (a_1, \infty) \tag{7.11.101}$$

On the other hand for $c \in (0,1)$, we have,

$$c - 2a + a^2c < 0, \text{ for } a \in (a_1, a_2) \text{ and } c - 2a + a^2c > 0, \text{ for } a \in (-\infty, a_1) \cup (a_2, \infty), \text{(by 7.11.100)} \tag{7.11.102}$$

So, to include all the cases, we will break the cases 5 and 6 in sub-cases, as follow,
Case 5: $c\kappa(c^2-1) + K_W(c - 2a + a^2c) > 0$

1. $c \in (-1,0)$, where, $c\kappa(c^2-1) > 0$

   (a) $a \in (a_2, a_1)$, where, $K_W(c - 2a + a^2c) > 0$, (see 7.11.100)
   Hence, the inequality $c\kappa(c^2-1) + K_W(c - 2a + a^2c) > 0$, implies $\kappa > \frac{-K_W(c-2a+a^2c)}{c(c^2-1)}$, which always holds.
   The only valid solution is (7.11.77)-(7.11.80).
   $K_1^{\infty,*} > 0$, for $\kappa < \frac{K_W(c-2a+a^2c) - K_W a^2 c(c^2-1)}{c(c^2-1)^2}$,
   $\lambda_1^{\infty,*} > 0$, which never holds, because we assumed before that the denominator of (7.11.84) is positive and that implies $c(a-c)^2 > 0$ too. However, it doesn't exist, because c is negative, and $c(a-c)^2 < 0$ gives us a contradiction.

   (b) $a \in (-\infty, a_2) \cup (a_1, \infty)$, where, $K_W(c - 2a + a^2c) < 0$, (see 7.11.100)
   Hence, the inequality $c\kappa(c^2-1) + K_W(c - 2a + a^2c) < 0$, implies $\kappa > \frac{K_W(c-2a+a^2c)}{c(c^2-1)}$.
   The only valid solution is (7.11.77)-(7.11.80).
   $\lambda_1^{\infty,*} > 0$, never holds, like the previous sub-case, because of $c \in (-1,0)$.

2. $c \in (0,1)$, where, $c\kappa(c^2-1) < 0$

   (a) $a \in (a_1, a_2)$, where, $K_W(c - 2a + a^2c) < 0$, (see 7.11.100)
   Hence, the inequality $c\kappa(c^2-1) + K_W(c - 2a + a^2c) > 0$ doesn't exist.

(b) $a \in (-\infty, a_2) \cup (a_1, \infty)$, where, $K_W(c - 2a + a^2c) < 0$, (see 7.11.100)
Hence, the inequality $c\kappa(c^2 - 1) + K_W(c - 2a + a^2c) > 0$, implies $\kappa < \frac{K_W(c-2a+a^2c)}{c(1-c^2)}$.
The only valid solution is (7.11.77)-(7.11.80).
$K_1^{\infty,*} > 0$, for $\kappa < \kappa_1$,
$\lambda_1^{\infty,*} > 0$, for $\kappa \in [0, \infty)$,
$K_{Z_1}^{\infty,*} > 0$, for $\kappa \in [0, \kappa_2) \cup (\kappa_3, \infty)\Big\}$.
This sub-case can't give us feedback capacity, because,

$$\frac{K_W(c - 2a + a^2c)}{c(1 - c^2)} < \frac{K_W(1 - ac)(ac^3 - 2ac + c^2 - \sqrt{c^3(a^2c^3 - 6ac^2 + 4a + 4c^3 - 3c)})}{2c^2(c^2 - 1)^2}.$$

That means there isn't any intersection between the power constraints.

Case 6: $c\kappa(c^2 - 1) + K_W(c - 2a + a^2c) < 0$

1. $c \in (-1, 0)$, where, $c\kappa(c^2 - 1) > 0$

(a) $a \in (a_2, a_1)$, where, $K_W(c - 2a + a^2c) > 0$, (see 7.11.100)
Hence, the inequality $c\kappa(c^2 - 1) + K_W(c - 2a + a^2c) > 0$ doesn't exist.

(b) $a \in (-\infty, a_2) \cup (a_1, \infty)$, where, $K_W(c - 2a + a^2c) < 0$, (see 7.11.100)
Hence, the inequality $c\kappa(c^2 - 1) + K_W(c - 2a + a^2c) < 0$, implies $\kappa < \frac{K_W(c-2a+a^2c)}{c(1-c^2)}$.
The only valid solution is (7.11.77)-(7.11.80).
$K_1^{\infty,*} > 0$, for $\kappa < \kappa_1$,
$\lambda_1^{\infty,*} > 0$, for $\kappa \in [0, \infty)$,
$K_{Z_1}^{\infty,*} > 0$, for $\kappa \in [0, \kappa_2) \cup (\kappa_3, \infty)\Big\}$.
This sub-case can't give us feedback capacity because,

$$\kappa_1 < \kappa_2 \implies \frac{K_W(c - 2a + a^2c) - K_W a^2 c(c^2 - 1)}{c(c^2 - 1)^2} < \frac{K_W(1 - ac)(ac^3 - 2ac + c^2 - \sqrt{c^3(a^2c^3 - 6ac^2 + 4a + 4c^3 - 3c)})}{2c^2(c^2 - 1)^2}.$$

That means there isn't any intersection between the power constraints.

2. $c \in (0, 1)$, where, $c\kappa(c^2 - 1) < 0$

(a) $a \in (a_1, a_2)$ where, $K_W(c-2a+a^2c) < 0$, (see 7.11.100)
Hence, the inequality $c\kappa(c^2-1)+K_W(c-2a+a^2c) < 0$, implies $\kappa > \frac{K_W(c-2a+a^2c)}{c(1-c^2)}$, which always holds.
The only valid solution is (7.11.77)-(7.11.80).
$K_1^{\infty,*} > 0$, for $\kappa < \kappa_1$,
$\lambda_1^{\infty,*} > 0$, which never holds, because we assumed before that the denominator of (7.11.84) is negative and that implies $c(a-c)^2 < 0$ too. However, it doesn't exist, because c is positive, and $c(a-c)^2 > 0$ gives us a contradiction.

(b) $a \in (-\infty, a_2) \cup (a_1, \infty)$, where, $K_W(c-2a+a^2c) > 0$, (see 7.11.100)
Hence, the inequality $c\kappa(c^2-1)+K_W(c-2a+a^2c) < 0$, implies $\kappa > \frac{K_W(c-2a+a^2c)}{c(1-c^2)}$.
The only valid solution is (7.11.77)-(7.11.80).
$\lambda_1^{\infty,*} > 0$, never holds, like the previous sub-case, because of $c \in (0,1)$.

Until here, we see that feedback does not increase capacity for the regime, $c \in \big(-1,1\big)$, $a \in \mathbb{R}$, $\kappa \in \big[0,\infty\big)$. Cases 5 and 6 correspond to the stable noise cases. However, there is an achievable rate, which is calculated in no-feedback capacity Theorem 5.2.3.

Bibliography

[1] S. Yang, A. Kavcic, and S. Tatikonda, "On feedback capacity of power-constrained Gaussian noise channels with memory," *Information Theory, IEEE Transactions on*, vol. 53, no. 3, pp. 929–954, March 2007.

[2] Y.-H. Kim, "Feedback capacity of stationary Gaussian channels," *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 57–85, 2010.

[3] T. Cover and S. Pombra, "Gaussian feedback capacity," *IEEE Transactions on Information Theory*, vol. 35, no. 1, pp. 37–43, Jan. 1989.

[4] T. Liu and G. Han, "Feedback capacity of stationary Gaussian channels further examined," *IEEE Transactions on Information Theory*, vol. 64, no. 4, pp. 2494–2506, April 2019.

[5] A. Gattami, "Feedback capacity of Gaussian channels revisited," *IEEE Transactions on Information Theory*, vol. 65, no. 3, pp. 1948–1960, March 2019.

[6] S. Ihara, "On the feedback capacity of the first-order moving average gaussian channel," *Japanese Journal of Statistics and Data Science*, March 2019. [Online]. Available: https://doi.org/10.1007/s42081-019-00045-z

[7] C. Li and N. Elia, "Youla coding and computation of Gaussian feedback capacity," *IEEE Transactions on Information Theory*, vol. 64, no. 4, pp. 3197–3215, April.

[8] T. Kailauth, A. Sayed, and B. Hassibi, *Linear Estimation*. Prentice Hall, 2000.

[9] P. E. Caines, *Linear Stochastic Systems*, ser. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., New York, 1988.

[10] T. Liu and G. Han, "The ARMAk Gaussian feedback capacity," in *IEEE International Symposium on Information Theory Proceedings (ISIT)*, 25-30, June 2017, pp. 211–215.

[11] H. W. Chen and K. Yanaki, “Refiniements of the half-bit and factor-of-two bounds for capacity in Gaussian channels with feedback,” *IEEE Transactions on Information Theory*, vol. 45, pp. 316–325, January 1999.

[12] ——, “Upper bounds on the capacity of discrete-time blockwise white Gaussian channels with feedback,” *IEEE Transactions on Information Theory*, vol. 43, no. 3, pp. 1125–1131, May 2000.

[13] E. Ordentlich, “A class of optimal coding schmes for moving average additive Gaussian noise channels with feedback,” in *IEEE International Symposium on Information Theory Proceedings (ISIT)*, 1994, p. 467.

[14] P. R. Kumar and P. Varaiya, *Stochastic Systems: Estimation, Identification, and Adaptive Control*. Prentice Hall, 1986.

[15] C. D. Charalambous, C. Kourtellaris, and S. Loyka, “Time-invariant feedback strategies do not increase capacity of AGN channels driven by stable and certain unstable autoregressive noise,” *IEEE Transactions on Information Theory*, submitted for publication, September 2019. [Online]. Available: https://arxiv.org/abs/1907.10991

[16] S. Ihara, *Information theory for Continuous Systems*. World Scientific, 1993.

[17] C. Kourtellaris and C. D. Charalambous, “Information structures of capacity achieving distributions for feedback channels with memory and transmission cost: Stochastic optimal control & variational equalities,” *IEEE Transactions on Information Theory*, vol. 64, no. 7, pp. 4962–4992, July 2018.

[18] C. D. Charalambous, I. Kourtellaris, C., and Loyka, “Control-coding capacity of decision models driven by correlated noise and Gaussian application examples,” in *IEEE Conference on Decision and Control (CDC)*, 17-19, December 2018, pp. 4967–4972.

[19] H. Permuter, T. Weissman, and A. Goldsmith, “Finite state channels with time-invariant deterministic feedback,” *IEEE Transactions on Information Theory*, vol. 55, no. 2, pp. 644–662, Feb. 2009.

[20] J. H. van Schuppen, *Mathematical Control and System Theory of Discrete-Time Stochastic Systems*. Preprint, 2010.