Feedback Capacity of Gaussian Channels Revisited

Ather Gattami
Bittronics Research
Stockholm, Sweden
Email: atherg@gmail.com

Abstract—In this paper, we revisit the problem of finding the average capacity of the Gaussian feedback channel. First, we consider the problem of finding the average capacity of the analog colored Gaussian noise channel where the noise has an arbitrary spectral density. We introduce a new approach to the problem where we solve the problem over a finite number of transmissions and then consider the limit of the average capacity of the case of infinite number of transmissions. We then consider the important special case of stationary Gaussian noise with finite memory. We show that the channel capacity at stationarity can be found by solving a semi-definite program, and hence computationally tractable. We also give new proofs and structural results of the non stationary solution which bridges the gap between results in the literature for the stationary and non stationary feedback channel capacity.

I. INTRODUCTION

A. Background and Previous Work

We revisit the problem of communication over a Gaussian feedback channel with colored noise \(z = (z_1, z_2, \ldots)\) and \(z_k = 0\) for \(k \leq 0\) (see Figure 1). More precisely, let \(W \in \{1, 2, 3, \ldots, 2^nR\}\) be the message to be transmitted over the Gaussian communication channel

\[y_k = x_k + z_k\]

for the time horizon \(k = 1, \ldots, n\), where \(x^n \triangleq (x_1, x_2, x_3, \ldots, x_n)\) are the transmitted codewords and \(y^n \triangleq (y_1, y_2, y_3, \ldots, y_n)\), \(y^0 \triangleq 0\), are the channel outputs. Each transmitted symbol \(x_k\) is a deterministic function of the message \(W\), the past transmitted symbols \(x^{k-1}\), and the past channel outputs \(y^{k-1}\), which accounts for the channel feedback. Thus, \(x_k = f_k(W, x^{k-1}, y^{k-1})\) for some time-varying function \(f_k\) to be optimized. The transmitted symbols are subject to the average power constraint

\[
\frac{1}{n} \sum_{k=1}^{n} x_k^2 \leq P
\]  

(1)

The noise process is a single-input-single-output (SISO) linear dynamical system given by

\[z_k = H(z)u_k\]

where

\[H(z) = \sum_{l=0}^{\infty} h_l z^{-l}\]

\(h_k \in \mathbb{R}\), \(\{u_k\}\) are independent and identically distributed (i. i. d.) Gaussian variables with \(u_k \sim \mathcal{N}(0, 1)\), and \(z^{-1}\) is the backward shift operator, that is \(z^{-1} u_k = u_{k-1}\).

Let \(W\) be uniformly distributed over \(\{1, 2, 3, \ldots, 2^nR\}\). The decoding map \(\hat{W}_n(y^n)\) is chosen to minimize the average error probability

\[
P_e^{(n)} = \frac{1}{2^nR} \sum_{i=1}^{2^nR} \Pr \left\{ \hat{W}_n(y^n) \neq W | W = i \right\}
\]

\[= \Pr \left\{ \hat{W}_n(y^n) \neq W \right\}
\]

The rate \(R\) is achievable if there exists a sequence of \((2^nR, n)\) codes with \(P_e^{(n)} \rightarrow 0\) as \(n \rightarrow \infty\). The feedback capacity \(C\) is defined as the supremum of all achievable rates.

It is well known that for the case where the Gaussian noise is white (uncorrelated over time, that is \(H(z) = h_0 \in \mathbb{R}\)), feedback does not improve on the capacity of the channel. However, feedback could indeed increase the capacity when the noise is colored. In the seminal work by Cover and Pombra [1], the authors introduced the average capacity

\[C_n = \sup_{f_n} \frac{1}{n} \mathbb{I}(W; y^n)
\]

under the expected value of the average power constraint

\[
\frac{1}{n} \sum_{k=1}^{n} \mathbb{E} \left( x_k^2 \right) \leq P
\]  

(2)

where \(\mathbb{I}(x; y)\) denotes the mutual information between \(x\) and \(y\).

It was shown that for an arbitrary Gaussian stochastic process \(z\), there exists a sequence of \((2^{n(C_n-\epsilon)}, n)\) of feedback codes with \(P_e^{(n)} \rightarrow 0\) as \(n \rightarrow \infty\) for \(\epsilon > 0\). The converse holds also.

It was shown that any sequence of \((2^{n(C_n+\epsilon)}), n)\) codes has \(P_e^{(n)}\) bounded away from zero for all \(n\). Furthermore, it was shown that \(C_n\) can be found by taking the supremum over \(x^n = B_n z^n + v^n\), where \(B_n\) is strictly lower triangular and \(v^n\) is Gaussian white noise. That is, for \(z^n \sim \mathcal{N}(0, Z_n)\), \(Z_n > 0\) (where \(Z_n > 0\) means that \(Z_n\) is positive definite).
and \( v^n \sim \mathcal{N}(0, \mathcal{V}_n) \), \( Z_n \succ 0 \) (where \( \mathcal{V}_n \succeq 0 \) means that \( \mathcal{V}_n \) is positive semi-definite), we have

\[
C_n = \sup_{B_n: \mathcal{V}_n > 0} \frac{1}{2n} \log \frac{\det (\mathcal{V}_n + (B_n + I)Z_n (B_n + I)^T)}{\det (Z_n)}
\]

where \( B_n \) is a matrix. Theorem 6.1 shows that the channel capacity at stationarity is given by

\[
R = \max_{X, Y} \frac{1}{2} \log \frac{\det (X + H \Sigma (X + H)^T)}{\det (\Sigma)}
\]

subject to

\[
\begin{align*}
\mathcal{P} &\geq X \mathcal{X}^T \\
\mathcal{Y} &\geq \mathcal{G} G^T + \mathcal{G} G^T - \mathcal{G} \mathcal{Y} \mathcal{G}^T \\
\mathcal{G} &\geq 0
\end{align*}
\]

The solution relies on considering the stationary problem directly instead of solving the problem over a finite horizon \( n \) and then letting \( n \to \infty \). The stationarity property in turn allows for using problem formulations in the frequency domain with some revealing structure that are not obvious to see in the time domain. However, a solution to the above optimization problem is intractable in practice and one needs another approach in order to get a practical solution.

A related problem is communication over a Gaussian channel with inter-symbol interference that was considered in [4]. The inter-symbol interference was modeled as a finite order filter where the concept of directed information was used to obtain a dynamic programming formulation with constraints.

Beyond the first order filter case, there is no known tractable solution to find the channel capacity numerically.

B. Contributions

First, we consider the problem of finding the average capacity of the analog colored Gaussian noise channel where the noise has an arbitrary spectral density. We introduce a new approach where we solve the problem over a finite number of transmissions and then consider the limit of the average capacity as the number of transmissions tend to infinity. We show that the maximum average feedback capacity \( C \) over an infinite time horizon can be obtained by optimizing over linear strategies \( x_k = B(z)z_k + v_k \) with

\[
B(z) = \sum_{l=1}^{\infty} b_l z^{-l}
\]

satisfying the power constraint

\[
\int_{-\pi}^{\pi} \left( S_x(e^{j\theta}) + |B(e^{j\theta})|^2 S_x(e^{j\theta}) \right) d\theta \leq P
\]

Also, [3] considered the important case of a Gaussian process \( z \) of finite order, given by the state space equations

\[
\begin{align*}
s_{k+1} &= F s_k + G u_k \\
z_k &= H s_k + u_k \\
s_0 &= 0 \\
u_k &\sim \mathcal{N}(0,1)
\end{align*}
\]

where \( F \in \mathbb{R}^{m \times m}, G \in \mathbb{R}^{m}, H \in \mathbb{R}^{1 \times m} \), and \( u_k, s_k, z_k \) take values in \( \mathbb{R}, \mathbb{R}^m \), and \( \mathbb{R} \), respectively. It was shown (Theorem 6.1) that the channel capacity at stationarity is given by

\[
C = \frac{1}{2} \log_2 (Y)
\]

and \( Y \in \mathbb{R} \) is the solution to the nonconvex optimization problem

\[
\max_{X, Y, \Sigma \succeq 0} Y
\]

subject to

\[
\begin{align*}
\mathcal{P} &\geq X \mathcal{X}^T \\
\mathcal{Y} &\geq \mathcal{G} G^T + \mathcal{G} G^T - \mathcal{G} \mathcal{Y} \mathcal{G}^T \\
\mathcal{G} &\geq 0
\end{align*}
\]

We then consider the important special case of stationary Gaussian noise with finite memory. We show that the channel capacity at stationarity can be found by solving a semi-definite program, and hence computationally tractable. In particular, we show that the channel capacity is given by

\[
C = \frac{1}{2} \log_2 (Y)
\]
where $Y \in \mathbb{R}$ is the optimal solution to the optimization problem

$$
\begin{align*}
\sup_{K} & Y \\
\text{s.t.} & 0 < \begin{pmatrix} P & K \\ K^T & \Sigma \end{pmatrix} \\
& 0 \preceq \begin{pmatrix} F \Sigma F^T - \Sigma + GG^T & FK^T + F \Sigma H^T + G \\ (FK^T + F \Sigma H^T + G)^T & Y \end{pmatrix} \\
& Y = KH^T + HK^T + H \Sigma H^T + P + 1
\end{align*}
$$

C. Paper Outline

In section II, we introduce the notation used in this paper and give some known results from system theory and information theory. In section III, we formulate the general problem of finding the average capacity of the Gaussian channel with feedback where we derive results similar to those obtained [1] and [3] for the non stationary and stationary Gaussian noise processes, respectively. We then consider the special case of the stationary Gaussian noise process of finite order in sections IV and V. There, we give new proofs that also show that we can find the channel feedback capacity by solving a semi-definite program, and hence computationally tractable. Finally, we provide an example in section VI with Matlab code in the appendix that can be used to compare the results of this paper with existing solutions for the first order Gaussian process case. Most of the proofs are relegated to the appendix.

II. PRELIMINARIES

A. Notation

| Set | Description |
|-----|-------------|
| $\mathbb{N}$ | The set of positive integers. |
| $\mathbb{R}$ | The set of real numbers. |
| $\mathbb{C}$ | The set of complex numbers. |
| $\mathbb{S}^n$ | The set of $n \times n$ symmetric matrices. |
| $\mathbb{S}_+^n$ | The set of $n \times n$ symmetric positive semidefinite matrices. |
| $\mathbb{S}_{++}^n$ | The set of $n \times n$ symmetric positive definite matrices. |

- $\succeq A \succeq B \iff A - B \in \mathbb{S}_+^n$. 
- $\succ A \succ B \iff A - B \in \mathbb{S}_{++}^n$.
- $A^\dagger$ | The Moore-Penrose pseudo-inverse of the square matrix $A$. |
- $s^k = (s_1, s_2, \ldots, s_k)$ and $s^0 \triangleq 0$. 
- $|s| = (s_1, s_2, \ldots, s_k)$, $|s|^2 = \sum_{i=1}^k |s_i|^2$. 

B. System Theory

The material here can be found in [5].

**Definition 1 (Detectability).** Let $F \in \mathbb{R}^{m \times m}$ and $H \in \mathbb{R}^{p \times m}$. The pair of matrices $(H,F)$ is detectable if

$$
\begin{pmatrix} F - \lambda I \\ H \end{pmatrix}
$$

has full column rank for all $\lambda \in \mathbb{C}$ such that $|\lambda| \geq 1$.

**Definition 2 (Controllability).** Let $F \in \mathbb{R}^{m \times m}$ and $G \in \mathbb{R}^{m \times r}$. The pair of matrices $(F,G)$ is controllable if

$$
(F - \lambda I)^2 G
$$

has full row rank for all $\lambda \in \mathbb{C}$.

**Definition 3 (Stability).** Let $F \in \mathbb{R}^{m \times m}$. The matrix $F$ is stable if and only if its eigenvalues have modulus strictly less than 1.

**Proposition 1.** Let $Q \succeq 0$ and $F$ stable. Then, the unique positive semi-definite solution $\Sigma$ to the Lyapunov equation

$$
\Sigma = F \Sigma F^T + Q
$$

is invertible if $(F, Q)$ is controllable.

**Proof:** Consult [5].

**Proposition 2.** Let $F \in \mathbb{R}^{m \times m}$ and $G \in \mathbb{R}^{m \times r}$. Then, the pair $(F, G)$ is controllable if and only if there does not exist a vector $x \neq 0$ and a scalar $\lambda \in \mathbb{C}$ such that $x^T F = \lambda x^T$ and $x^T G = 0$.

**Proof:** Consult [5].

C. Optimal Estimation of Gaussian Processes

Consider a Gaussian process $z$ given by the state space equations

$$
\begin{align*}
s_{k+1} &= F s_k + G u_k \\
z_k &= H s_k + u_k \\
s_1 &= 0 \\
u_k &\sim \mathcal{N}(0,1)
\end{align*}
$$

where $u_k$, $s_k$, and $z_k$ take values in $\mathbb{R}$, $\mathbb{R}^m$, and $\mathbb{R}$, respectively. Let $\hat{s}_k = \mu_k(y^{k-1})$ be an estimate of $s_k$ based on the measurements $y^{k-1}$ and $\hat{s}_k = s_k - \hat{s}_k$. Suppose that we want to minimize the average estimation error

$$
\frac{1}{n} \sum_{k=1}^n \mathbb{E}(\hat{s}_k^2).
$$

It is well known that the optimal estimator is given by $\hat{s}_k = \mathbb{E}(\hat{s}_k|y^{k-1})$ which obeys the optimal Kalman filter recursions [6]

$$
\begin{align*}
S_k &= \mathbb{E}(\hat{s}_k \hat{s}_k^T) \\
S_1 &= 0 \\
S_{k+1} &= F S_k F^T + GG^T \\
&\quad - (FS_k H^T + G)(HS_k H^T + 1)^{-1} \\
&\quad \times (HS_k F^T + G^T) \\
K_k &= (FS_k H^T + G)(HS_k H^T + 1)^{-1} \\
\hat{s}_{k+1} &= F \hat{s}_k + K_k (z_k - H \hat{s}_k) \\
\hat{s}_{k+1} &= (F - K_k H) \hat{s}_k + G u_k - K_k u_k \\
\hat{z}_k &= H \hat{s}_k + u_k
\end{align*}
$$

A property of the Kalman filter is that the innovations $\hat{z}_k = z_k - H \hat{s}_k = H \hat{s}_k + E u_k$ are independent for all $k \in \mathbb{N}$. 
D. Entropy Properties of Gaussian Variables and Processes

The entropy of the Gaussian process given by (3) over a time horizon $k = 1, \ldots, n$ is $h(z^n)$ which may be rewritten as a sum of conditional entropies using the entropy chain rule

$$h(z^n) = \sum_{k=1}^{n} h(z_k | z^{k-1})$$

with $z^k = 0$ if $k = 0$. The entropy rate is given by

$$h(z) = \lim_{n \to \infty} \frac{1}{n} h(z^n)$$

**Proposition 3.** Consider a Gaussian process $z$ with spectral density function $S_z(z)$. Then, the entropy rate of $z$ is given by

$$h(z) = \int_{-\pi}^{\pi} \frac{1}{2} \log (S_z(e^{i\theta})) \, d\theta$$

**Proof:** Consult [7].

**Proposition 4.** Consider a Gaussian process given by (3) over a finite time horizon $k = 1, \ldots, n$. The entropy of $z^n$ is given by

$$h(z^n) = \frac{1}{2} \sum_{k=1}^{n} \log_2 (2\pi e (HS_kH^T + 1))$$

where $S_k$ is given by the recursion

$$S_1 = \mathbf{E}(s_1s_1^T)$$
$$S_{k+1} = FS_kF^T + GG^T - (FS_kH^T + G)(HS_kH^T + 1)^{-1} \times (HS_kF^T + G^T)$$

Furthermore, if $(H, F)$ is detectable, then the stationary entropy rate is given by

$$h(z) = \lim_{n \to \infty} \frac{1}{n} h(z^n) = \frac{1}{2} \log_2 (2\pi e (HSHT + 1))$$

where $S$ is the unique solution to the Riccati equation

$$S = FSF^T + GG^T - (FS^T + G)(HSHT + 1)^{-1} \times (HS^T + G^T)$$

**Proof:** See the appendix.

III. FEEDBACK CAPACITY OF GAUSSIAN CHANNELS WITH COLORED NOISE

In this section, we will study the general problem of finding the feedback capacity with respect to Gaussian noise with an arbitrary spectral density function $H(z)$. More precisely, we consider the following problem.

**Problem 1.** Let $W$ be the message to be transmitted and consider the Gaussian communication channel

$$y_k = x_k + z_k$$

over a time horizon $n$, where $x_k = f_k(W, x^{k-1}, y^{k-1})$ is the transmitted signal over the channel with the average power constraint

$$\frac{1}{n} \sum_{k=1}^{n} \mathbf{E}(x_k^2) \leq P$$

$y$ is the measurement signal at the receiver, and $z$ is the Gaussian measurement noise process with spectral density function $H(z)$. Find the average channel capacity

$$C_n = \sup_{f^n} \frac{1}{n} \mathbf{I}(W; y^n)$$

where $f^n$ are deterministic functions. In particular, find $C = \lim_{n \to \infty} C_n$.

It has been shown in [1] that for a noise sequence $z^n$, the mutual information $I(W; y^n)$ is given by

$$I(W; y^n) = h(y^n) - h(z^n)$$

To make this paper self contained, we state this result.

**Proposition 5.** Consider the Gaussian feedback channel as described in Problem 1. Then,

$$I(W; y^n) = h(y^n) - h(z^n)$$

**Proof:** See the appendix.

For a Gaussian noise sequence $z^n$ with zero mean and covariance $Z_n$, it was further shown in [1] that the optimal input sequence $x^n$ has the form $x^n = B_n z^n + v^n$ where $B_n \in \mathbb{R}^{n \times n}$ is strictly lower triangular, $v_n$ is a Gaussian sequence with covariance $V_n \in \mathbb{S}_+$, and the pair $(B_n, V_n)$ satisfies the power constraint

$$\text{Tr}(B_n Z_n B_n^T + V_n) = \mathbf{E}(|x^n|^2) \leq nP$$

Now the mutual information between $v^n$ and $y^n$ is given by

$$I(v^n; y^n) = \frac{1}{2} \log_2 \frac{\det(V_n + (B_n + I)Z_n(B_n + I)^T)}{\det(Z_n)}$$

$$= \frac{1}{2} \log_2 \det(V_n + (B_n + I)Z_n(B_n + I)^T)$$

$$- \frac{1}{2} \log_2 \det(Z_n)$$

$$\leq \mathbf{I}(B_n, V_n, Z_n)$$

The average feedback capacity over the time horizon $n$ is given by

$$C_n = \sup_{B_n, V_n} \frac{1}{n} \mathbf{I}(B_n, V_n, Z_n)$$

where the maximum is taken over $B_n \in \mathbb{R}^{n \times n}$ being strictly lower triangular.

**Theorem 1.** The maximum average feedback capacity $C$ in Problem 1 over an infinite time horizon can be obtained by optimizing over linear strategies $x_k = B(z) z_k + v_k$ with

$$B(z) = \sum_{l=1}^{\infty} b_l z^{-l},$$
subject to 

\[ V = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{\sqrt{2\pi \sigma_x^2}} \exp \left( -\frac{(e^{i\theta} - \mu_x)^2}{2\sigma_x^2} \right) \, d\theta \]

\[ \text{s.t. } V + \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{|B(e^{i\theta})|^2 S_x(e^{i\theta})}{2\pi} \, d\theta \leq P \]

and

\[ \Sigma_{k+1} = (F - \Gamma_k(X_k + H))\Sigma_k(F - \Gamma_k(X_k + H))^\top \\
+ (G - \Gamma_k)(G - \Gamma_k)^\top + \Gamma_k \Lambda_k \Gamma_k^\top \]

\[ = F\Sigma_k F^\top + GG^\top - \Gamma_k \Lambda_k \Gamma_k^\top \]

Proof: See the appendix.

Remark 1. The recursive optimization problem in Theorem 2 given by equations (9)-(12) can be solved by introducing a Lagrange multiplier corresponding to the power constraint and use the bisection method with respect to that Lagrange multiplier and then solve the optimization problem using dynamic programming for each fixed value of the Lagrange multiplier. For further details, consult [8].

V. FEEDBACK CAPACITY AT STATIONARITY

Now consider a stationary finite order Gaussian noise process \( z \) given by

\[ s_{k+1} = Fs_k + Gu_k \]

\[ z_k = Hs_k + u_k \]

\[ u_k \sim \mathcal{N}(0,1) \]

where \( u_k, s_k, \) and \( z_k \) take values in \( \mathbb{R}, \mathbb{R}^m, \) and \( \mathbb{R}, \) respectively.

In [3], by using the results of [1] above, it was shown that the optimal affine strategy of the transmitted symbols \( x_k \) to maximize the capacity at stationarity (that is, as \( n \to \infty \)), is given by

\[ x_k = X(s_k - E(s_k|z^{k-1})) + v_k \]

where \( v \) is a white Gaussian process independent of \( u \) and \( w \) with \( v_k \sim \mathcal{N}(0,V) \), and \( X \) is some vector in \( \mathbb{R}^m \). In fact, it was shown that \( v = 0 \) is optimal, but we will keep it as an optimization parameter as it will simplify the optimization problem considerably.

We give here and in section V new (and shorter) proofs that summarize the results of [1] and [3] on the finite time horizon average capacity \( C_n \) and the (non-stationary) infinite time horizon capacity \( C \).

Theorem 2. The optimal average feedback capacity in Problem 1 over a finite horizon \( k = 1, \ldots, n \) is achieved by \( x_k = Xs_k + v_k \) for some set of vectors \( \{X_k\} \), where \( s_k = E(s_k|z^{k-1}) \), \( s_k = s_k - \hat{s}_k \), and \( \{v_k\} \) is white Gaussian process with \( v_k \sim \mathcal{N}(0,V_k) \), independent of \( u \). The capacity is given by

\[ C_n = \sup_{V_1, \ldots, V_n, \gamma_{k=1}^n} \frac{1}{2n} \sum_{k=1}^n \log_2(Y_k) \] subject to \( \Sigma_1 = 0, \)

\[ P \geq X_k \Sigma_k X_k^\top + V_k, \]

\[ Y_k = (X_k + H)\Sigma_k(X_k + H)^\top + V_k + 1, \]

\[ \Gamma_k = (F\Sigma_k(X_k + H)^\top + G)Y_k^{-1} \]

and

\[ \Sigma_{k+1} = (F - \Gamma_k(X_k + H))\Sigma_k(F - \Gamma_k(X_k + H))^\top \\
+ (G - \Gamma_k)(G - \Gamma_k)^\top + \Gamma_k \Lambda_k \Gamma_k^\top \]

\[ = F\Sigma_k F^\top + GG^\top - \Gamma_k \Lambda_k \Gamma_k^\top \]

Proof: See the appendix.
and we arrive at the following optimization problem for finding the maximum capacity at stationarity:

$$\sup_{V > 0} \sup_{X \Sigma > 0} \frac{1}{2} \log_2(Y)$$

s.t. $P \geq X \Sigma X^T + V$

$$\Sigma = F \Sigma F^T + G G^T - \Gamma Y \Gamma^T$$
$$\Gamma = (F \Sigma (X + H)^T + G)Y^{-1}$$
$$Y = (X + H)\Sigma (X + H)^T + V + 1$$

We have now the following result.

**Theorem 3.** The average feedback channel capacity is then given by

$$C = \frac{1}{2} \log_2(Y)$$

where $Y$ is the solution to the optimization problem

$$\sup_{V > 0} \sup_{X \Sigma > 0} Y$$

s.t. $P \geq X \Sigma X^T + V$

$$\Sigma = F \Sigma F^T + G G^T - \Gamma Y \Gamma^T$$
$$\Gamma = (F \Sigma (X + H)^T + G)Y^{-1}$$
$$Y = (X + H)\Sigma (X + H)^T + V + 1$$

**Proof:** Since $Y$ is a scalar, we may maximize $Y$ instead of its logarithm and we get the equivalent optimization problem (17).

Note also that taking $V = 0$ renders the same (nonconvex) optimization problem as that in [3]. However, this case is not achievable since the mutual information vanishes for $V = 0$.

**Lemma 1.** Optimization problem (16) is equivalent to

$$\sup_{V > 0} \sup_{X \Sigma > 0} Y$$

s.t. $P \geq X \Sigma X^T + V$

$$\Sigma \leq F \Sigma F^T + G G^T - \Gamma Y \Gamma^T$$
$$\Gamma = (F \Sigma (X + H)^T + G)Y^{-1}$$
$$Y = (X + H)\Sigma (X + H)^T + V + 1$$

**Proof:** See the appendix.

Now we turn to the Riccati equation (15) and utilize that $V > 0$ to show that the pair

$$(F - \Gamma (X + H), (G - \Gamma)(G - \Gamma)^T + \Gamma V \Gamma^T)$$

is controllable.

**Lemma 2.** Suppose that $(F, G)$ is controllable and $V > 0$. Then, the pair

$$(F - \Gamma (X + H), (G - \Gamma)(G - \Gamma)^T + \Gamma V \Gamma^T)$$

is controllable.

**Proof:** See the appendix.

Now we can use Lemma 2 to prove the following.

**Lemma 3.** Suppose that $(F, G)$ is controllable and $V > 0$ and consider the Riccati equation

$$\Sigma = (F - \Gamma (X + H))\Sigma (F - \Gamma (X + H))^T + (G - \Gamma)(G - \Gamma)^T + \Gamma V \Gamma^T$$

Then, $F - \Gamma (X + H)$ is stable.

**Proof:** See the appendix.

**Lemma 4.** Suppose that $(F, G)$ is controllable and $V > 0$ and consider the Riccati equation

$$\Sigma = (F - \Gamma (X + H))\Sigma (F - \Gamma (X + H))^T + (G - \Gamma)(G - \Gamma)^T + \Gamma V \Gamma^T$$

Then, $\Sigma > 0$.

**Proof:** See the appendix.

Since Lemma 4 established the invertibility of $\Sigma$, we see that optimization problem (17) is equivalent to that of optimizing over strictly positive definite matrices $\Sigma$, that is

$$\sup_{V > 0} \sup_{X \Sigma > 0} Y$$

s.t. $P \geq X \Sigma X^T + V$

$$\Sigma \leq F \Sigma F^T + G G^T - \Gamma Y \Gamma^T$$
$$\Gamma = (F \Sigma (X + H)^T + G)Y^{-1}$$
$$Y = (X + H)\Sigma (X + H)^T + V + 1$$

(19)

Optimization problem (19) can now be transformed to a semi-definite program. The trick is to first eliminate the dependence on $V$ and obtain desired inequalities instead of equalities. Then, by making a variable substitution according to $K = X \Sigma$, which is possible since $\Sigma$ is invertible, we will be able to use a Schur complement argument in order to transform the constraints in (19) into a set of linear matrix inequalities (LMI:s).

We are now ready to state the main result of this paper.

**Theorem 4.** The feedback capacity of the Gaussian channel is given by

$$C = \frac{1}{2} \log_2(Y)$$

where $Y$ is the optimal solution of

$$\sup_{K \Sigma > 0} Y$$

s.t.

$$0 \prec \begin{pmatrix} P & K \\ K^T & \Sigma \end{pmatrix}$$

$$0 \preceq \begin{pmatrix} F \Sigma F^T - \Sigma + G G^T & F K^T + F \Sigma H^T + G \\ F K^T + F \Sigma H^T + G & Y \end{pmatrix}$$

$$Y = KH^T + HK^T + H \Sigma H^T + P + 1$$

(20)

**Proof:** See the appendix.
VI. NUMERICAL EXAMPLE

Consider a communication feedback channel with Gaussian noise given by a first order process according to

\[ z_k + \beta z_{k-1} = u_k + \alpha u_{k-1} \]

with state space representation

\[
\begin{align*}
    s_{k+1} &= -\beta s_k + u_k \\
    z_k &= (\alpha - \beta) s_k + u_k \\
    y_k &= x_k + z_k \\
\end{align*}
\]

\( \alpha \in [-1, 1], \beta \in (-1, 1), u_k \sim N(0, 1), \) and

\[
\sigma = \text{sign}(\beta - \alpha) = \begin{cases} 
1 & \text{if } \beta > \alpha \\
0 & \text{if } \beta = \alpha \\
-1 & \text{if } \beta < \alpha 
\end{cases}
\]

In [3], it was shown that the feedback capacity of the above channel with a power constraint \( E(x_k^2) \leq P \) is given by

\[-\log_2(r) \]

where \( r \) is the unique positive real root of the polynomial equation

\[
(\alpha^2 + \beta^2) r^4 + 2\sigma(\alpha + \beta P)r^3 + (P + 1 - \alpha^2)r^2 - 2\sigma \alpha r - 1 = 0
\]

This can be compared to the solution of the semi-definite optimization problem (20). By noting that \( F = -\beta, G = 1, \) and \( H = \alpha - \beta, \) one can verify numerically that the channel capacity \( \frac{1}{2} \log_2(Y) \) coincides with that of the polynomial solution \(-\log_2(r)\) (see the Matlab code in the appendix, where we used CVX, a package for specifying and solving convex programs [9], [10]).

VII. CONCLUSION

We considered the problem of finding the average capacity of the analog colored Gaussian noise channel where the noise has an arbitrary spectral density. We introduced a new approach to the problem where we solved the problem over a finite number of transmissions and then considered the limit of the average capacity as the number of transmissions went to infinity. We also provided new proofs and structural results of the non stationary solution which bridges the gap between results in the literature [1], [3]. For the stationary Gaussian noise, we showed that the maximum average feedback capacity \( C \) over an infinite time horizon can be obtained by optimizing over linear strategies \( x_k = B(z)z_k + v_k \) with

\[
B(z) = \sum_{l=1}^{\infty} b_l z^{-l}
\]

\( v_k \sim N(0, V), \) and \( V > 0. \) The capacity \( C \) is the optimal value of

\[
\sup_{V > 0, B(z) = \sum_{l=1}^{\infty} b_l z^{-l}} \int_{-\pi}^{\pi} \frac{1}{2} \log \left( V + |B(e^{i\theta})| + 1\right)^2 |S_z(e^{i\theta})| d\theta
\]

s.t. \( V + \int_{-\pi}^{\pi} |B(e^{i\theta})|^2 |S_z(e^{i\theta})| d\theta \leq P \)

We also considered the special case of stationary Gaussian noise with finite memory. We showed that the channel capacity at stationarity can be found by solving a semi-definite program, and hence computationally tractable. In particular, we showed that the channel capacity is given by

\[
C = \frac{1}{2} \log_2(Y)
\]

where \( Y \in \mathbb{R} \) is the optimal solution to the optimization problem

\[
\sup_{Y \geq 0} \left\{ \begin{array}{l}
    Y \\
    s.t. \\
    0 < \left( \begin{array}{c} P \\ K \end{array} \right) \\
    0 \preceq \left( \begin{array}{c} F \Sigma F^T - \Sigma + G G^T \\ F K T + F \Sigma H^T + G^T \end{array} \right) Y \\
    Y = K H^T + H K^T + H \Sigma H^T + P + 1
\end{array} \right. \]

REFERENCES

[1] T. Cover and S. Pombra, “Gaussian feedback capacity,” Information Theory, IEEE Transactions on, vol. 35, no. 1, pp. 37–43, Jan 1989.
[2] L. Vandenberghe, S. Boyd, and S.-P. Wu, “Determinant maximization with linear matrix inequality constraints,” SIAM journal on matrix analysis and applications, vol. 19, no. 2, pp. 499–533, 1998.
[3] Y.-H. Kim, “Feedback capacity of stationary Gaussian channels,” Information Theory, IEEE Transactions on, vol. 56, no. 1, pp. 57–85, Jan 2010.
[4] S. Yang, A. Kavcic, and S. Tatikonda, “Feedback capacity of finite-state machine channels,” Information Theory, IEEE Transactions on, vol. 51, no. 3, pp. 799–810, March 2005.
[5] K. Zhou, J. C. Doyle, and K. Glover, Robust and Optimal Control. Prentice Hall, 1996.
[6] T. Kailath, Linear systems, ser. Prentice Hall information and system sciences series. Englewood Cliffs: Prentice-Hall, 1980. [Online]. Available: http://opac.inria.fr/record=b1077852
[7] T. Cover and J. A. Thomas, Elements of Information Theory. John Wiley & Sons, 2006.
[8] A. Gattami, “Generalized linear quadratic control,” IEEE Trans. Automatic Control, vol. 55, no. 1, pp. 131–136, January 2010.
[9] M. Grant and S. Boyd, “CVX: Matlab software for disciplined convex programming, version 2.1,” http://cvxr.com/cvx, Mar. 2014.
[10] ———, “Graph implementations for nonsmooth convex programs,” in Recent Advances in Learning and Control, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95–110, http://stanford.edu/~boyd/graph_dcp.html.

APPENDIX

Proof of Proposition 4

Note that \( \hat{s}_k = E(z_k|z^{k-1}) \) and \( \delta_k = s_k - \hat{s}_k \) are given by the Kalman filter (4), so \( z_k - E(z_k|z^{k-1}) = \hat{z}_k \) is the Gaussian output estimation error given by the Kalman filter (4). Thus, the entropy chain rule gives the equality

\[
h(z^n) = \sum_{k=1}^{n} h(\hat{z}_k) = \frac{1}{2} \sum_{k=1}^{n} \log_2(2\pi e (H S_k H^T + E E^T))
\]

Since the pair \((H, F)\) is detectable, the recursion (5) has a unique stationary positive semi-definite solution \( S_k = S \). The stationary entropy rate is then given by

\[
h(z) = \lim_{n \to \infty} \frac{1}{n} h(z^n) = \frac{1}{2} \log_2(2\pi e (H S H^T + E E^T))
\]

where \( S \) is the unique positive semi-definite solution to the Riccati equation (4).
Proof of Proposition 5

\[ I(W; y^n) = h(y^n) - h(y^n | W) \]
\[ = h(y^n) - \sum_{k=1}^{n} h(y_k | W, y^{k-1}) \]
\[ = h(y^n) - \sum_{k=1}^{n} h(y_k | W, y^{k-1}, x_k(W, y^{k-1}), x^{k-1}) \]
\[ = h(y^n) - \sum_{k=1}^{n} h(y_k | W, z^{k-1}, x_k, x^{k-1}) \]  \hspace{1cm} (21)
\[ = h(y^n) - \sum_{k=1}^{n} h(z_k | W, z^{k-1}, x_k, x^{k-1}) \]  \hspace{1cm} (22)
\[ = h(y^n) - \sum_{k=1}^{n} h(z_k | W, z^{k-1}) \]  \hspace{1cm} (23)
\[ = h(y^n) - \sum_{k=1}^{n} h(z_k | W, z^{k-1}) \]  \hspace{1cm} (24)
\[ = h(y^n) - h(z^n) \]  \hspace{1cm} (25)

where (21) follows from the fact that \( x_i = f_i(W, x^{i-1}, y^{i-1}) \) and so \( x_i \) is determined by \( (W, y^{i-1}) \), (22) follows from the inequality \( y^{k-1} = x^{k-1} + z^{k-1} \), (23) follows from \( y_k = x_k + z_k \), and (24) follows from the fact that \( x_k \) is determined from \( W \) and \( z^{k-1} \) by the recursion \( x_i = f_i(W, x^{i-1}, x^{i-1} + z^{i-1}) \), and (25) follows from the independence between \( z^n \) and \( W \). \( \square \)

Proof of Theorem 1

According to Proposition 5, we have that

\[ I(W; y^n) = h(y^n) - h(z^n) \]

Let \( \tilde{z}_k = E(\tilde{z}_k | z^{k-1}) \), \( \tilde{x}_k = z_k - \tilde{z}_k \), and \( \tilde{x}_k \) be independent of \( z^{k-1} \) and \( x^{k-1} \). Let \( \tilde{x}_k = E(\tilde{x}_k | y^{k-1}) \), and \( \tilde{y}_k = y_k - \tilde{x}_k - \tilde{z}_k = \tilde{x}_k + \tilde{z}_k + \tilde{z}_k \). Then,

\[ h(y^n) = \sum_{k=1}^{n} h(y_k | y^{k-1}) \]
\[ \leq \sum_{k=1}^{n} h(\tilde{y}_k) \]  \hspace{1cm} (26)
\[ \leq \frac{1}{2} \sum_{k=1}^{n} \log_2(2\pi e E(\tilde{y}_k^2)) \]  \hspace{1cm} (27)

and (29) can be achieved by the transmission strategy \( x_k = L_k \tilde{z}_k + v_k \) for some \( L_k \in \mathbb{R} \) and temporally uncorrelated Gaussian variables \( \{v_k\} \). Let \( E(\tilde{z}_k^2) = \sigma_k \) and

\[ E \left( \begin{pmatrix} \tilde{z}_k \\ \tilde{x}_k \end{pmatrix} \right) = \begin{pmatrix} \zeta_k & \psi_k \\ \psi_k & \xi_k \end{pmatrix} \geq 0 \]  \hspace{1cm} (30)

be an achievable covariance matrix by some strategy \( x_k \). Then,

\[ Y_k = E \left( \tilde{y}_k^2 \right) = E \left( (\tilde{x}_k + \tilde{z}_k + \tilde{z}_k)^2 \right) = E \left( (\tilde{x}_k + \tilde{z}_k)^2 \right) + E(\tilde{z}_k^2) \]
\[ = E \left( \begin{pmatrix} \tilde{z}_k \\ \tilde{x}_k \end{pmatrix} \right) \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \tilde{z}_k \\ \tilde{x}_k \end{pmatrix} + \sigma_k \]
\[ = E \left( \text{Tr} \left( \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \zeta_k & \psi_k \\ \psi_k & \xi_k \end{pmatrix} \right) \right) + \sigma_k \]
\[ = \text{Tr} \left( \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \zeta_k & \psi_k \\ \psi_k & \xi_k \end{pmatrix} \right) + \sigma_k \]

The Schur complement in \( \xi_k \) of

\[ \left( \begin{array}{cc} \zeta_k & \psi_k \\ \psi_k & \xi_k \end{array} \right) \]

is given by

\[ \phi_k = \xi_k - \psi_k^2 \xi_k^\top \geq 0 \]

By taking \( x_k = L_k \tilde{z}_k + v_k \) with \( L_k = \psi_k \zeta_k^\top \) and \( v \) to be a temporally uncorrelated Gaussian process with \( v_k \sim \mathcal{N}(0, \phi_k) \) independent of \( \{\tilde{z}_k\} \), we will have \( \tilde{x}_k = E \left( x_k | y^{k-1} \right) = E \left( L_k \tilde{z}_k + v_k | y^{k-1} \right) = 0 \), \( \tilde{x}_k = x_k \), and we will get a sequence of pairs \( (\tilde{z}_k, x_k) \) that renders the covariance matrix (30). Also, since this strategy is linear and \( v \) is a Gaussian process, \( y \) will be a Gaussian process and hence, \( \tilde{y} \) will be Gaussian such that \( \tilde{y}_k \) is independent of \( y^{k-1} \), and the entropy upper bound (29) is achieved.

Hence, an optimal strategy has the form

\[ x_k = B_k(z)z_k + v_k \]

with

\[ B_k(z) = \sum_{l=1}^{\infty} b_l(k)z^{-l} \]

and \( v \) is a white Gaussian process.

Now for the stationary case, \( \zeta_k \) and \( \sigma_k \) will be constant.

That is, \( \zeta_k = \zeta \) and \( \sigma_k = \sigma \). This implies that we can take \( B_k \) and \( \phi_k \) to be constant, \( B_k = B \), \( \phi_k = V \), for all \( k \). The spectral density function of the stationary Gaussian process \( y \) is given by

\[ S_y(z) = V + |B(z) + 1|^2S_z(z) \]

where \( V > 0 \). Note that the strict inequality is necessary since \( V = 0 \) implies that \( v \) vanishes (almost everywhere), which implies in turn that the mutual information between \( v \) and \( y \) vanishes and it would render \( C = 0 \). Now the average power of the process \( x \) is given by

\[ V + \int_{-\pi}^{\pi} |B(e^{i\theta})|^2S_z(e^{i\theta})\frac{d\theta}{2\pi} \leq P \]
Furthermore, according to Proposition 3, the entropies of \( y \) and \( z \) are given by
\[
h(y) = \int_{-\pi}^{\pi} \frac{1}{2} \log \left( 2\pi e (V + |B(e^{i\theta})| + 1)^2 S_z(e^{i\theta}) \right) \frac{d\theta}{2\pi}
\]
and
\[
h(z) = \int_{-\pi}^{\pi} \frac{1}{2} \log \left( 2\pi e S_z(e^{i\theta}) \right) \frac{d\theta}{2\pi}
\]
Now the average channel capacity is given by
\[
I(v, y) = \lim_{n \to \infty} \frac{1}{n} I(v^n, y^n)
= \lim_{n \to \infty} \frac{1}{n} h(y^n) - \frac{1}{n} h(z^n)
= h(y) - h(z)
\]
Hence, the average channel capacity \( C \) is the optimal value of the optimization problem
\[
\sup_{V > 0} \int_{-\pi}^{\pi} \frac{1}{2} \log \left( V + |B(e^{i\theta})| + 1 \right)^2 S_z(e^{i\theta}) \frac{d\theta}{2\pi}
\text{s.t. } V + \int_{-\pi}^{\pi} |B(e^{i\theta})|^2 S_z(e^{i\theta}) \frac{d\theta}{2\pi} \leq P
\]

**Proof of Theorem 2**

According to Proposition 5, we have that
\[
I(W; y^n) = h(y^n) - h(z^n)
\]
Let \( \hat{z}_k = E (z_k | z^{k-1}) \) and \( \bar{z}_k = z_k - \hat{z}_k \). We have that
\[
h(z^n) = \sum_{k=1}^{n} h(z_k | z^{k-1}) = \sum_{k=1}^{n} h(\bar{z}_k)
\]
Since \( s_1 = 0 \), we get \( \bar{z}_1 = z_1 = u_1 \). We also have that \( E (z_2 | z_1) = Hs_2 \), and so \( \bar{z}_2 = u_2 \). Inductively, we get that \( \bar{z}_k = u_k \) for \( k = 1, \ldots, n \). Thus,
\[
h(z^n) = \sum_{k=1}^{n} h(\bar{z}_k) = \sum_{k=1}^{n} h(u_k) = \frac{1}{2} \sum_{k=1}^{n} \log_2(2\pi e)
\]
Let \( \hat{s}_k = E (s_k | y^{k-1}) \), \( \bar{s}_k = s_k - \hat{s}_k \), and \( \bar{y}_k = y_k - H\hat{s}_k = x_k + H\bar{s}_k + u_k \).
Note that
\[
h(y^n) = \sum_{k=1}^{n} h(y_k | y^{k-1}) \tag{31}
\]
\[
= \sum_{k=1}^{n} h(\bar{y}_k | y^{k-1}) \tag{32}
\]
\[
\leq \sum_{k=1}^{n} h(\bar{y}_k) \tag{33}
\]
\[
\leq \frac{1}{2} \sum_{k=1}^{n} \log_2(2\pi e E (\bar{y}_k^2)) \tag{34}
\]
where (33) follows from the fact that conditioning only reduces entropy (with equality if \( \bar{y}_k \) is independent of \( y^{k-1} \)) and (34) follows from the fact that for a fixed covariance, the Gaussian distribution has the maximum entropy. We will now show that
\[
Y_k \triangleq E (\bar{y}_k^2)
\]
can be achieved by the transmission strategy \( x_k = X_k\hat{s}_k + v_k \). First note that \( E (\bar{y}_k^2) = E \left( (x_k + H\hat{s}_k + u_k)^2 \right) = E \left( (x_k + H\hat{s}_k)^2 \right) + E (u_k^2) \) since \( u_k \) is independent of \( x_k \) and \( \hat{s}_k \). Now let
\[
E \left( \left( \begin{array}{c} \hat{s}_k \\ x_k \end{array} \right) \left( \begin{array}{c} \hat{s}_k \\ x_k \end{array} \right)^\top \right) = \left( \begin{array}{c} \Sigma_k & \Psi_k \\ \Psi_k^\top & \Xi_k \end{array} \right) \tag{35}
\]
be an achievable covariance matrix by some strategy \( x_k \). Then,
\[
Y_k = E (\bar{y}_k^2)
= E \left( (x_k + H\hat{s}_k)^2 \right) + E (u_k^2)
= E \left( \left( \begin{array}{c} \hat{s}_k \\ x_k \end{array} \right) \left( \begin{array}{c} \hat{s}_k \\ x_k \end{array} \right)^\top \right) + 1
= E \left( \text{Tr} \left( \left( \begin{array}{cc} H^\top & H^\top \\ H & 1 \end{array} \right) \left( \begin{array}{c} \hat{s}_k \\ x_k \end{array} \right) \left( \begin{array}{c} \hat{s}_k \\ x_k \end{array} \right)^\top \right) \right) + 1
= \text{Tr} \left( \left( \begin{array}{cc} H^\top & H^\top \\ H & 1 \end{array} \right) \left( \begin{array}{c} \Sigma_k & \Psi_k \\ \Psi_k^\top & \Xi_k \end{array} \right) \right) + 1
\]
The Schur complement in \( \Xi_k \) of
\[
\left( \begin{array}{c} \Sigma_k & \Psi_k \\ \Psi_k^\top & \Xi_k \end{array} \right) \preceq 0
\]
is given by
\[
\Phi_k \triangleq \Xi_k - \Psi_k \Sigma_k^{-1} \Psi_k \succeq 0
\]
By taking \( x_k = X_k\hat{s}_k + v_k \) with \( X_k = \Psi_k \Sigma_k^{-1} \) and \( v_k \sim \mathcal{N}(0, \Phi_k) \) independent of \( \hat{s}_k, u_k \), and \( v_l \) for \( l \neq k \), we will get a sequence of pairs \( (\hat{s}_k, x_k) \) that renders the covariance matrix (35). Also, since this strategy is linear and \( v \) is a Gaussian process, \( \bar{y} \) is Gaussian such that \( \bar{y}_k \) is independent of \( y^{k-1} \), and the entropy upper bound (34) is achieved. The mutual information becomes
\[
I(W; y^n) = h(y^n) - h(z^n)
= \frac{1}{2} \sum_{k=1}^{n} \log_2(2\pi e E (\bar{y}_k^2)) - \frac{1}{2} \sum_{k=1}^{n} \log_2(2\pi e)
= \frac{1}{2} \sum_{k=1}^{n} \log_2(2\pi e Y_k) - \frac{1}{2} \sum_{k=1}^{n} \log_2(2\pi e)
= \frac{1}{2} \sum_{k=1}^{n} \log_2(Y_k)
\]
and the average feedback channel capacity is
\[
C_n = \sup_{\text{f}} \frac{1}{n} I(W; y^n) = \max_{X_1, \ldots, X_n \geq 0} \frac{1}{2n} \sum_{k=1}^{n} \log_2(Y_k)
\]
Now for \( x_k = X_k\hat{s}_k + v_k \), we have that
\[
\bar{y}_k = y_k - H\hat{s}_k = (X_k + H)\hat{s}_k + v_k + u_k
\]
(34) Let
\[
\Gamma_k \triangleq E \left( (F\hat{s}_k + G\hat{u}_k) (\bar{y}_k^2) \right)^{-1}
= (F\Sigma_k (X_k + H)^\top + G)Y_k^{-1}
\]
Then, 
\[ \mathbb{E}(F \bar{\delta}_k + G u_k | \bar{y}_k) = \Gamma_k \bar{y}_k \]
The dynamics of \( \delta_k \) and \( \bar{\delta}_k \) are given by
\[
\begin{align*}
\delta_{k+1} &= \mathbb{E}(F \bar{\delta}_k + G u_k | y^k) \quad (36) \\
\delta_{k+1} &= \mathbb{E}(F \bar{\delta}_k + G u_k | y^{k-1}) \quad (37) \\
\delta_{k+1} &= \mathbb{E}(F \bar{\delta}_k + G u_k | y^{k-1}) \quad (38) \\
\bar{\delta}_{k+1} &= F \bar{\delta}_k + \mathbb{E}(F \bar{\delta}_k + G u_k | \bar{y}_k) \quad (39) \\
\bar{\delta}_{k+1} &= F \bar{\delta}_k + \Gamma_k \bar{y}_k \quad (40)
\end{align*}
\]
which (39) follows from the orthogonality between \((u_k, \bar{\delta}_k, \bar{y}_k)\) and \(y^{k-1}\). Hence, the error dynamics become
\[
\begin{align*}
\delta_{k+1} &= F \bar{\delta}_k - \Gamma_k \bar{y}_k + G u_k \\
\bar{y}_k &= (X_k + H) \hat{s}_k + v_k + u_k
\end{align*}
\]
whence it implies that
\[
Y_k = (X_k + H) \Sigma_k (X_k + H)^	op + V_k + 1
\]
Finally, we have that \( \Sigma_k = \mathbb{E}(\bar{s}_k \bar{s}_k^	op) = \mathbb{E}(s_k s_k^	op) = 0 \) and the recursion of \( \Sigma_k \) is given by
\[
\Sigma_{k+1} = (F - \Gamma_k (X_k + H)) \Sigma_k (F - \Gamma_k (X_k + H))^	op + (G - \Gamma_k) (G - \Gamma_k)^	op + \Gamma_k V_k \Gamma_k^	op
\]
Now let \( S_k = \mathbb{E}(s_k s_k^	op) \) and \( \bar{S}_k = \mathbb{E}(\bar{s}_k \bar{s}_k^	op) \). Then, \( S_k = \bar{S}_k + \Sigma_k \) (since \( \bar{s}_k \) and \( \bar{s}_k \) are independent). Also, since \( S_{k+1} = F \bar{S}_k F^	op + G G^	op \) and \( \bar{S}_{k+1} = F \bar{S}_k F^	op + \Gamma_k Y_k \Gamma_k^	op \), we obtain
\[
\Sigma_{k+1} = S_{k+1} - \bar{S}_{k+1} = (F - \Gamma_k (X_k + H)) \Sigma_k (F - \Gamma_k (X_k + H))^	op + (G - \Gamma_k) (G - \Gamma_k)^	op + \Gamma_k Y_k \Gamma_k^	op
\]
Proof of Lemma 1
Consider a stationary finite order Gaussian noise process \( z \) given by
\[
\begin{align*}
s_{k+1} &= F s_k + G u_k + e_k \\
z_k &= H s_k + u_k \\
u_k &\sim \mathcal{N}(0,1) \\
(42)
\end{align*}
\]
where \( e \) is a white Gaussian process independent of \( u \) with \( e_k \sim \mathcal{N}(0, \epsilon I) \), \( 0 \leq \epsilon \in \mathbb{R} \). Clearly, the channel capacity is maximized when \( \epsilon \to 0 \) since \( \epsilon \) only increases the power of the noise process \( z \). According to Theorem 3, the feedback capacity with the noise process given by (42) is
\[
C = \frac{1}{2} \log_2(Y_c)
\]
where \( Y_c \) is the solution to the optimization problem
\[
\begin{align*}
\sup_{V > 0} \sup_{\Sigma \succeq 0} Y_c \\
\text{s.t.} \quad P \geq X \Sigma X^	op + V \\
\Sigma = F \Sigma F^	op + G G^	op + \epsilon I - \Gamma Y \Gamma^	op \\
\Gamma = (F \Sigma (X + H)^	op + G) Y_c^{-1} \\
Y_c = (X + H) (X + H)^	op + V + 1
\end{align*}
\]
As we noted above, the channel capacity is maximized as \( \epsilon \) approaches 0. Thus, the channel capacity we seek is \( \max_c Y_c = Y_0 \). Hence, (16) is equivalent to
\[
\begin{align*}
\sup_{V > 0} \sup_{\Sigma > 0} Y_c \\
\text{s.t.} \quad P \geq X \Sigma X^	op + V \\
\Sigma = F \Sigma F^	op + G G^	op + \epsilon I - \Gamma Y \Gamma^	op \\
\Gamma = (F \Sigma (X + H)^	op + G) Y_c^{-1} \\
Y_c = (X + H) (X + H)^	op + V + 1
\end{align*}
\]
Now we can regard \( \epsilon \) as a slack variable in the above optimization problem transform the equality
\[
\Sigma = F \Sigma F^	op + G G^	op + \epsilon I - \Gamma Y \Gamma^	op
\]
to the inequality
\[
\Sigma \preceq F \Sigma F^	op + G G^	op - \Gamma Y \Gamma^	op
\]
Hence, we have eliminated \( \epsilon \) from the optimization problem above and transformed an equality into an inequality constraint, and we conclude that (16) is equivalent to (18).

Proof of Lemma 2
Note first that from Proposition 2, the pair
\[
(F - \Gamma (X + H), (G - \Gamma) (G - \Gamma)^	op + \Gamma \Gamma^	op)
\]
is controllable if there does not exist a complex number \( \lambda \in \mathbb{C} \) and a vector \( x \) such that
\[
x \top ((G - \Gamma) (G - \Gamma)^	op + \Gamma \Gamma^	op) = 0
\]
and
\[
x \top (F - \Gamma (X + H)) = \lambda x \top
\]
Now suppose that \( x \) is such that
\[
x \top ((G - \Gamma) (G - \Gamma)^	op + \Gamma \Gamma^	op) = 0
\]
and
\[
x \top (F - \Gamma (X + H)) = \lambda x \top
\]
Then,
\[
x \top ((G - \Gamma) (G - \Gamma)^	op + \Gamma \Gamma^	op) x = 0
\]
Since \( (G - \Gamma) (G - \Gamma)^	op \succeq 0 \) and \( \Gamma \Gamma^	op \succeq 0 \), we must have that \( x \top \Gamma \Gamma^	op x = (\Gamma \Gamma x) \top = 0 \) and \( x \top (G - \Gamma) (G - \Gamma) \top x = 0 \). Since \( V > 0 \), we must have \( \Gamma x = 0 \). Thus, \( \Gamma = (\Gamma x) \top = 0 \). Similarly, \( x \top (G - \Gamma) (G - \Gamma) \top x = 0 \) implies that \( x \top (G - \Gamma) = 0 \) and since \( \Gamma x = 0 \), we get \( x \top G = 0 \).

Now \( (F, G) \) is controllable, and Proposition 2 implies that there does not exist a number \( \lambda \in \mathbb{C} \) such that \( x \top G = 0 \) and \( \lambda x \top = x \top F = x \top (F - \Gamma (X + H)) \). Thus, we cannot have a vector \( x \) and a number \( \lambda \in \mathbb{C} \) such that
\[
x \top ((G - \Gamma) (G - \Gamma)^	op + \Gamma \Gamma^	op) = 0
\]
and
\[ x^T (F - \Gamma (X + H)) = \lambda x^T \]

Hence, using Proposition 2 again, we conclude that
\((F - \Gamma (X + H), (G - \Gamma)(G - \Gamma)^T + \Gamma \Gamma^T)\)
is controllable.

**Proof of Lemma 3**

Since the pair \((F, G)\) is controllable and \(V > 0\), Lemma 2 implies that
\((F - \Gamma (X + H), (G - \Gamma)(G - \Gamma)^T + \Gamma \Gamma^T)\)
is controllable. Suppose that there exists a vector \(x\) such that
\[ x^T \Sigma x = x^T (F - \Gamma (X + H)) \Sigma (F - \Gamma (X + H))^T x \]
\[ + x^T (G - \Gamma)(G - \Gamma)^T x + x^T \Gamma \Gamma^T x \]
\[ = |\lambda|^2 x^T \Sigma x + x^T (G - \Gamma)(G - \Gamma)^T x + x^T \Gamma \Gamma^T x \]
\[ \geq x^T \Sigma x + x^T (G - \Gamma)(G - \Gamma)^T x + x^T \Gamma \Gamma^T x \]
which implies that \(x^T ((G - \Gamma)(G - \Gamma)^T + \Gamma \Gamma^T) = 0\). But then, we obtain that
\[ x^T (F - \Gamma (X + H) - \lambda I (G - \Gamma)(G - \Gamma)^T + \Gamma \Gamma^T) = 0 \]
which contradicts the fact that the pair
\((F - \Gamma (X + H), (G - \Gamma)(G - \Gamma)^T + \Gamma \Gamma^T)\)
is controllable. Thus, we conclude that \(|\lambda| < 1\) and the closed loop matrix \(F - \Gamma (X + H)\) must be stable.

**Proof of Lemma 4**

Since the pair \((F, G)\) is controllable and \(V > 0\), Lemma 2 implies that
\((F - \Gamma (X + H), (G - \Gamma)(G - \Gamma)^T + \Gamma \Gamma^T)\)
is controllable and Lemma 3 implies that \(F - \Gamma (X + H)\) is stable. Taking \(Q = (G - \Gamma)(G - \Gamma)^T + \Gamma \Gamma^T \geq 0\), we conclude from Proposition 1 that \(\Sigma\) is invertible and thus, \(\Sigma > 0\).

**Proof of Theorem 4**

We can make the variable substitution \(K = X \Sigma\) and maximize with respect to \(V, K\) and \(\Sigma > 0\). This gives the optimization problem
\[
\sup_{V > 0} \sup_{K > 0} Y \\
\text{s.t.} \quad P \geq K \Sigma^{-1} K^T + V \\
\Sigma \preceq F \Sigma F^T + G G^T - \Gamma Y \Gamma^T \\
\Gamma = (F K^T + F \Sigma H^T + G) Y^{-1} \\
Y = K \Sigma^{-1} K^T + H K^T + H \Sigma H^T + \lambda_1 + V + 1
\]
First we show the intuitive result that for any \(K\) and \(\Sigma\), taking \(V\) such that equality is achieved for the power constraint \(P \geq K \Sigma^{-1} K^T + V\) is optimal. Suppose that \(V\) is such that \(P > K \Sigma^{-1} K^T + V\). Introduce \(V' = P - K \Sigma^{-1} K^T > V\). Then, we have that
\[
Y \leq K \Sigma^{-1} K^T + K H^T + H K^T + H \Sigma H^T + V' + 1 = K H^T + H K^T + H \Sigma H^T + P + 1
\]

Furthermore, with \(\Gamma' \triangleq (F K^T + F \Sigma H^T + G)(Y')^{-1}\), the constraint
\[
\Sigma \preceq F \Sigma F^T + G G^T - \Gamma' Y' \Gamma'^T
\]
is satisfied. Thus, increasing \(V\) to \(V'\) increases the value of \(Y\) to \(Y'\). Hence, (43) is equivalent to
\[
\sup_{V > 0} \sup_{K > 0} Y \\
\text{s.t.} \quad P = K \Sigma^{-1} K^T + V \\
\Sigma \preceq F \Sigma F^T + G G^T - \Gamma Y \Gamma^T \\
\Gamma = (F K^T + F \Sigma H^T + G) Y^{-1} \\
Y = K H^T + H K^T + H \Sigma H^T + P + 1
\]

Now we have that
\[
P \geq K \Sigma^{-1} K^T + V > K \Sigma^{-1} K^T
\]

Thus, any parameters satisfying the constraints in (44) also satisfy the constraints in
\[
\sup_{K > 0} Y \\
\text{s.t.} \quad P > K \Sigma^{-1} K^T \\
\Pi = F K^T + F \Sigma H^T + G \\
Y = K H^T + H K^T + H \Sigma H^T + P + 1 \\
\Sigma \preceq F \Sigma F^T + G G^T - \Pi Y^{-1} \Pi^T
\]

Similarly, if \(K, \Sigma\) satisfy the constraints in (45), then they also satisfy the constraints in (44) with \(V = P - K \Sigma^{-1} K^T\). The power inequality
\[
P > K \Sigma^{-1} K^T
\]
is equivalent to the linear matrix inequality (LMI)
\[
\begin{pmatrix} P & K \\ K^T & \Sigma \end{pmatrix} > 0
\]
The error covariance inequality
\[
\Sigma \preceq F \Sigma F^T + G G^T - \Pi Y^{-1} \Pi^T
\]
can be recast as the LMI
\[
0 \preceq \begin{pmatrix} F \Sigma F^T - \Sigma + GG^T & \Pi \\ \Pi^T & Y \end{pmatrix} = \begin{pmatrix} F K^T + F \Sigma H^T + G \\ (FK^T + F \Sigma H^T + G)^T \end{pmatrix} Y
\]
Summing up, the maximum entropy of the channel output is given by the value of the following maxmin problem

\[
\begin{align*}
\sup_{K} & \quad Y \\
\text{s.t.} & \quad 0 \prec \begin{pmatrix} P & K \\ K^T & \Sigma \end{pmatrix} \\
& \quad Y = KH^T + HK^T + H\Sigma H^T + P + 1 \\
& \quad 0 \preceq \begin{pmatrix} F^2 \Sigma F^T - \Sigma + GG^T & FK^T + F\Sigma H^T + G \\ (FK^T + F\Sigma H^T + G)^T & Y \end{pmatrix}
\end{align*}
\]

(Matlab Code)

```matlab
alpha = 0.7;
beta = -0.25;
sigma = sign(beta-alpha);
P = 1;
d = 1;
F = -beta;
G = 1;
H = alpha-beta;
a4 = alpha^2+beta^2*P;
a3 = 2*sigma*(alpha+beta*P);
a2 = P+1-alpha^2;
a1 = -2*sigma*alpha;
a0 = -1;
r = roots([a4 a3 a2 a1 a0]);

cvx_begin sdp
    variable S(d,d) symmetric
    variable K(1,d)
    variable Y
    S > 0
    [P K; K' S] > 0
    [F*S*F' - S + G*G' F*K' + F*S*H' + G; 
     K*F' + H*S*F' + G' Y] > 0
    Y == K*H' + H*K' + H*S*H' + P + 1
    maximize Y;
cvx_end

C_sdp = 0.5*log2(Y)
C_poly = -log2(r(4))
```