

Feedback Capacity of MIMO Gaussian Channels

Oron Sabag, Victoria Kostina, Babak Hassibi

Abstract—Finding a computable expression for the feedback capacity of channels with non-white Gaussian, additive noise is a long standing open problem. In this paper, we solve this problem in the scenario where the channel has multiple inputs and multiple outputs (MIMO) and the noise process is generated as the output of a state-space model (a hidden Markov model). The main result is a computable characterization of the feedback capacity as a finite-dimensional convex optimization problem. Our solution subsumes all previous solutions to the feedback capacity including the auto-regressive moving-average (ARMA) noise process of first order, even if it is a non-stationary process. The capacity problem can be viewed as the problem of maximizing the measurements' entropy rate of a controlled (policy-dependent) state-space subject to a power constraint. We formulate the finite-block version of this problem as a *sequential convex optimization problem*, which in turn leads to a single-letter and computable upper bound. By optimizing over a family of time-invariant policies that correspond to the channel inputs distribution, a tight lower bound is realized. We show that one of the optimization constraints in the capacity characterization boils down to a Riccati equation, revealing an interesting relation between explicit capacity formulae and Riccati equations.

I. INTRODUCTION

We consider the feedback capacity of a multiple-input multiple-output (MIMO) Gaussian channel

$$\mathbf{y}_i = \Lambda \mathbf{x}_i + \mathbf{z}_i, \quad (1)$$

where Λ is a known matrix, $\mathbf{y}_i \in \mathbb{R}^m$ is the channel output and $\mathbf{x}_i \in \mathbb{R}^p$ is the channel input. The noise is a non-white Gaussian random process generated by a MIMO state-space model (a hidden Markov model)

$$\begin{aligned} \mathbf{s}_{i+1} &= F\mathbf{s}_i + G\mathbf{w}_i \\ \mathbf{z}_i &= H\mathbf{s}_i + \mathbf{v}_i, \end{aligned} \quad (2)$$

where the sequence $(\mathbf{w}_i, \mathbf{v}_i)$ is i.i.d. Gaussian. For a particular realization of the state-space, well-known random processes can be revealed, e.g., the auto-regressive moving-average (ARMA) random processes that were studied in [1]–[3]. State-space structures are utilized in the fields of control and estimation to obtain explicit policies. In estimation theory, for instance, the celebrated Kalman filter is a closed-form policy obtained using the underlying state-space structure of the signal and its measurements. In this work, we show that imposing a state-space structure on the noise process leads to a computable solution for the feedback capacity problem too.

Our main result is a computable expression for the feedback capacity. Our assumptions on the state-space are mild and

include noise processes that are not necessarily stationary, i.e., when the spectral radius of F is greater than 1. We show that the feedback capacity can be formulated as a finite-dimensional convex optimization problem. The optimization is a standard maximal determinant (max-det) optimization problem subject to linear matrix inequalities (LMIs) constraints [4]–[7]. Maximal determinant formulations previously appeared in the literature, e.g. [8], but it is a recent development that the fundamental limits of *problems with memory* are formulated using this important class of convex optimization problems with sequential methods [9], [10].

Previous literature on the feedback capacity of the (scalar) Gaussian channel is vast, e.g. [11]–[15], and a detailed survey can be found in [16]. The feedback capacity of arbitrary Gaussian processes was characterized by a multi-letter expression in [17]. In [18], a Markov decision processes formulation provided a computational tool, and an explicit lower bound for ARMA noise process of first order was derived. The tightness of the lower bound for the moving average (MA) process was established in [16], concluding the first explicit capacity formula. Later, the same author considered general stationary noise processes and showed that the capacity can be expressed with a variational formula in the frequency domain [19]. This approach does not provide a consistent methodology to compute the feedback capacity, but for the ARMA noise process (of first order), they could obtain a closed-form formula and conclude the tightness of the conjectured lower bound in [18]. Another contribution of [19] was the formulation of the capacity as a *single-letter* (more precisely, a finite-dimensional) but non-convex optimization problem. In [20], a change of variable for this optimization problem showed that it can be reformulated as a convex optimization problem. However, the change of variable relies on the fact that a certain covariance matrix is invertible, an imprecise claim (see Remark 1). We also remark that these results are limited to the case of stable F and $\mathbf{w}_i = \mathbf{v}_i$ which means, effectively, that the encoder has access to the hidden state of the state-space. Our solution generalizes these results to an arbitrary state-space model with possibly non-stationary noise process.

The main idea behind our derivation is a novel formulation of the n -letter capacity as a *sequential convex optimization problem* (SCOP). We formulate an optimization problem whose decision variable is a sequence of length n , where at each time three fixed-dimensional matrices are optimized. In the SCOP formulation, each of the LMI constraints may depend on two consecutive times only. While it is well known that the n -letter capacity is a convex optimization problem, it is the sequential property that allows us to obtain a single-letter upper bound for the limiting n -letter capacity. For the lower

This work was supported in part by the National Science Foundation (NSF) under grants CCF-1751356 and CCF-1956386. O. Sabag is partially supported by the ISEF international postdoctoral fellowship. The authors are with California Institute of Technology (e-mails: {oron,vkostina,hassibi}@caltech.edu).

bound, an optimization over a family of time-invariant channel input distributions leads to a non-convex optimization problem that we show to be equivalent to the single-letter upper bound. Our derivation shows the optimality of time-invariant input distributions and therefore extending this important conclusion from [19] to possibly non-stationary noise processes.

The rest of the paper is organized as follows. In Section II, we present the setting and problem formulation. Section III includes our main result on the feedback capacity of the MIMO Gaussian channel. In Section IV, we present the main ideas and the technical lemmas that prove our main result. Due to space limitations, detailed proofs are available at [21].

II. THE SETTING AND PRELIMINARIES

In this section, we define the communication setting and the noise characteristics. We also present some preliminaries on Kalman filtering and Riccati equations.

A. The setting

We consider a MIMO additive Gaussian channel

$$\mathbf{y}_i = \Lambda \mathbf{x}_i + \mathbf{z}_i, \quad (3)$$

where the channel input is $\mathbf{x}_i \in \mathbb{R}^m$, the additive noise is $\mathbf{z}_i \in \mathbb{R}^p$, and $\Lambda \in \mathbb{R}^{p \times m}$ is a fixed known matrix. We assume that the encoder has access to noiseless, instantaneous feedback so that the input $\mathbf{x}_i$ is a function of the message and all previous channel outputs $\mathbf{y}^{t-1} := \mathbf{y}_1, \dots, \mathbf{y}_{t-1}$. For a fixed blocklength n , the channel input has an average power constraint $\frac{1}{n} \sum_{i=1}^n \mathbb{E}[\mathbf{x}_i^T \mathbf{x}_i] \leq P$. The definitions of the average probability of error, an achievable rate and the feedback capacity are standard and can be found in [19], for instance.

We consider a colored Gaussian noise that is generated as the output of a state-space model:

$$\begin{aligned} \mathbf{s}_{i+1} &= F\mathbf{s}_i + G\mathbf{w}_i \\ \mathbf{z}_i &= H\mathbf{s}_i + \mathbf{v}_i, \end{aligned} \quad (4)$$

where $\mathbf{w}_i \sim N(0, W)$ and $\mathbf{v}_i \sim N(0, V)$ are i.i.d. sequences with $\mathbb{E}[\mathbf{w}_i \mathbf{v}_i^T] = L$, and are independent of the initial state $\mathbf{s}_1 \sim N(0, \Sigma_1)$. The class of linear dynamical systems described by the state-space in (4) is rich and captures many known instances like the MIMO ARMA random processes of k th order given by $\mathbf{z}_i = \sum_{j=1}^k A_j \mathbf{z}_{i-j} + B_j \mathbf{u}_{i-j}$, where $\mathbf{u}_i \sim N(0, \Sigma_U)$ is an i.i.d. sequence.

B. The Kalman filter and the Riccati equation

The Kalman filter is a simple, recursive method to compute the maximum likelihood estimation of the hidden state based on the measurements. The predicted-state estimation and the prediction estimation error covariance are defined as

$$\begin{aligned} \hat{\mathbf{s}}_i &= \mathbb{E}[\mathbf{s}_i | \mathbf{z}^{i-1}] \\ \Sigma_i &= \text{cov}(\mathbf{s}_i - \hat{\mathbf{s}}_i). \end{aligned} \quad (5)$$

Then, the standard Kalman filter is given by the recursion

$$\hat{\mathbf{s}}_{i+1} = F\hat{\mathbf{s}}_i + K_{p,i}(\mathbf{z}_i - H\hat{\mathbf{s}}_i), \quad (6)$$

with

$$\begin{aligned} K_{p,i} &= (F\Sigma_i H^T + GL)\Psi_i^{-1} \\ \Psi_i &= H\Sigma_i H^T + V, \end{aligned} \quad (7)$$

and the error covariance is described by the Riccati recursion

$$\Sigma_{i+1} = F\Sigma_i F^T + GWG^T - K_{p,i}(H\Sigma_i H^T + V)K_{p,i}^T. \quad (8)$$

The *innovation process* defined by $\mathbf{e}_i = \mathbf{z}_i - H\hat{\mathbf{s}}_i$ holds the property that it is orthogonal (statistically independent) from previous instances of the measurements $\mathbf{z}^{i-1}$ [22].

Note that in (7), it is assumed that $\Psi_i \succ 0$ for all i . This is a natural assumption in our communication setting since otherwise the capacity is infinite. Namely, if Ψ_i is only positive semidefinite, a coordinate in the noise vector $\mathbf{z}_i$ is a deterministic function of the past noise instances $\mathbf{z}^{i-1}$. Building an infinite-rate scheme is straightforward: the encoder transmits $\mathbf{x}_j = 0$ for $j \leq i-1$ so that $\mathbf{y}_j = \mathbf{z}_j$. Then, based on $\mathbf{z}^{i-1}$, the encoder and the decoder can communicate an infinite number of bits on this vector coordinate (assuming the image of Λ is not degenerated at this particular direction).

We move on to present our assumptions on the state-space model. The stability of F is significant since it determines the stationarity of the noise process.

Definition 1. The matrix F is stable if its spectral radius satisfies $\rho(F) < 1$.

Without further assumptions, our results hold for the stationary case, i.e., when F is stable. If $L \neq 0$, the stability of F should be replaced with the stability of $F_s = F - GLV^{-1}H$.

For the general case where F is not stable, we need three additional assumptions. Since the assumptions are satisfied for the case where F is stable, a reader whose interest is limited to the stationary case may skip these assumptions.

Assumption 1. The pair (F, H) is detectable. That is, there exists a matrix K such that $\rho(F - KH) < 1$.

Assumption 2. The pair $(F_s, GW^{1/2})$ is stabilizable. That is, for any x and λ with $|\lambda| \geq 1$ such that $xF = x\lambda$, $xGW^{1/2} \neq 0$.

Assumption 3. The matrix F does not have eigenvalues on the unit circle.

The first two assumptions are made to guarantee that the Riccati recursion in (8) converges to a matrix that solves a Riccati equation. More specifically, consider the function

$$f(\Sigma) = F\Sigma F^T - \Sigma + W - K_p(\Sigma)\Psi(\Sigma)K_p^T(\Sigma), \quad (9)$$

where $K_p(\Sigma) = (F\Sigma H^T + GL)\Psi(\Sigma)^{-1}$ and $\Psi(\Sigma) = H\Sigma H^T + V$. The Riccati equation is defined as $f(\Sigma) = 0$. Under Assumptions 1–2, the Riccati recursion converges to a unique *stabilizing solution* of the Riccati equation [23, Ch. 14]. That is, there exists a unique $\Sigma_s \succeq 0$ such that $f(\Sigma_s) = 0$ and $F - K_p(\Sigma_s)H$ is stable. Moreover, for any initial condition $\Sigma_1, \Sigma_i \rightarrow \Sigma_s$ at exponential rate. From now on, we refer to the constants

$$K_p = (F\Sigma_s H^T + GL)\Psi_s^{-1}$$

$$\Psi = H\Sigma H^T + V \quad (10)$$

as the ones evaluated at the stabilizing solution to the Riccati equation.

The solution to the Riccati equation also characterizes the entropy rate of the Gaussian noise process:

$$\begin{aligned} \frac{1}{n}h(\mathbf{z}^n) &= \frac{1}{n} \sum_{i=1}^n h(\mathbf{z}_i | \mathbf{z}^{i-1}) \\ &= \frac{1}{n} \sum_{i=1}^n \log \det(H\Sigma_i H^T + V) \\ &\rightarrow \log \det(\Psi), \end{aligned} \quad (11)$$

as $n \rightarrow \infty$

III. MAIN RESULT AND DISCUSSION

In this section we present the feedback capacity of the MIMO channel, its particularization to the scalar case and an explicit computation of the feedback capacity for the MA process. The following is our main result.

Theorem 1. *The feedback capacity of the MIMO Gaussian channel in (3)-(4) is given by the convex optimization problem*

$$\begin{aligned} C_{fb}(P) &= \max_{\Pi, \hat{\Sigma}, \Gamma} \frac{1}{2} \log \det(\Psi_Y) - \frac{1}{2} \log \det(\Psi) \\ \text{s.t.} \quad &\Psi_Y = \Lambda \Pi \Lambda^T + H \hat{\Sigma} H^T + \Lambda \Gamma H^T + H \Gamma^T \Lambda^T + \Psi \\ &\begin{pmatrix} \Pi & \Gamma \\ \Gamma^T & \hat{\Sigma} \end{pmatrix} \succeq 0, \quad \text{Tr}(\Pi) \leq P, \\ &\begin{pmatrix} F \hat{\Sigma} F^T + K_p \Psi K_p^T - \hat{\Sigma} & F(\Gamma^T \Lambda^T + \hat{\Sigma} H^T) + K_p \Psi \\ (\Lambda \Gamma + H \hat{\Sigma}) F^T + \Psi K_p^T & \Psi_Y \end{pmatrix} \succeq 0, \end{aligned} \quad (12)$$

where K_p and Ψ are constants given in (10).

The objective structure is the difference between the entropy rates of the channel outputs and the channel noise random processes. The entropy rate of the noise process $\{\mathbf{z}_i\}_{i \geq 1}$ is the constant given in (11), while the entropy rate of the channel outputs process is $\frac{1}{2} \log \det(\Psi_Y)$ and is part of the optimization. The auxiliary decision variable Π along with its power constraint and the first LMI in (12) ensure that the channel input power does not exceed P . In Lemma 1 below, the decision variables Γ and the error covariance matrix $\hat{\Sigma}$ will be given a straightforward interpretation as part of the policy.

The Schur complement of the second LMI constraint in (12) implies the Riccati inequality

$$\hat{\Sigma} \preceq F \hat{\Sigma} F^T + K_p \Psi K_p^T - K_Y \Psi_Y K_Y^T, \quad (13)$$

with $K_Y = (F(\Gamma^T \Lambda^T + \hat{\Sigma} H^T) + K_p \Psi) \Psi_Y^{-1}$. In Lemma 6 in Section IV below, it is shown that optimal decision variables $(\Pi, \hat{\Sigma}, \Gamma)$ satisfy that the Riccati inequality (13) with equality, i.e., it is a Riccati equation. This fact reveals that the origin for explicit capacity formulae expressed as function of roots to some polynomials in the literature, e.g., [16], [18], [19] is the Riccati equation. We demonstrate this interesting fact in Section III-B for the MA noise process.

A. The scalar case:

If the channel outputs, inputs, and the additive noise are scalars, but the hidden state of the noise $\mathbf{s}_i$ is possibly a vector, the capacity in Theorem 1 can be simplified.

Theorem 2. *The feedback capacity of the scalar Gaussian channel (3)-(4) with $\Lambda = 1$ is given by the following convex optimization problem*

$$\begin{aligned} C_{fb}(P) &= \max_{\hat{\Sigma}, \Gamma} \frac{1}{2} \log \left(1 + \frac{P + H \hat{\Sigma} H^T + 2\Gamma H^T}{\Psi} \right) \\ \text{s.t.} \quad &\begin{pmatrix} P & \Gamma \\ \Gamma^T & \hat{\Sigma} \end{pmatrix} \succeq 0, \end{aligned} \quad (14)$$

$$\begin{pmatrix} F \hat{\Sigma} F^T + K_p \Psi K_p^T - \hat{\Sigma} & F \Gamma^T + F \hat{\Sigma} H^T + K_p \Psi \\ \Gamma F^T + H \hat{\Sigma} F^T + \Psi K_p^T & P + H \hat{\Sigma} H^T + 2\Gamma H^T + \Psi \end{pmatrix} \succeq 0,$$

where K_p and Ψ are constants defined in (10).

Choosing $H = 0$ in (14) recovers the capacity formula of an additive white Gaussian noise channel

$$C_{fb}(P) = \frac{1}{2} \log \left(1 + \frac{P}{V} \right).$$

Remark 1. *The state-space that was studied in [20] can be recovered by choosing $W = V = L = 1$. In this case, the constants are $\Sigma = 0, K_p = G, \Psi = 1$ and the capacity in (14) and that in [20, Th. 4] are almost in full agreement. Specifically, there is a difference in the sign of the first LMI in (14) which reads as a strict LMI ($\succ$) in [20]. A strict LMI implies that the Schur complement satisfies $P - \Gamma \hat{\Sigma}^{-1} \Gamma^T > 0$. However, it can be shown that the optimum is achieved with equality in the Schur complement at least for particular instances like the MA noise process in Section III-B. The claim that the LMI constraint is positive definite was also utilized in [20] to show their main argument that $\hat{\Sigma}$ is invertible, and thus should be read with care.*

B. Moving average noise

In [16], the feedback capacity of the MA noise process of first order was shown to be

$$C_{fb}(P) = -\log x_0, \quad (15)$$

where x_0 is the unique positive root of $Px^2 = (1 - |\alpha|x)^2(1 - x^2)$. As this noise realization corresponds to the special case $F = 0, H = \alpha, G = W = V = L = 1$, we illustrate the simplicity of computing such expressions from Theorem 2.

Theorem 3. *The feedback capacity of the scalar Gaussian channel with first-order MA noise process is*

$$C_{fb}(P) = \frac{1}{2} \log(1 + \text{SNR}), \quad (16)$$

where SNR is the positive root of the polynomial $\text{SNR} = \left(\sqrt{P} + |\alpha| \sqrt{\frac{\text{SNR}}{1 + \text{SNR}}} \right)^2$.

The capacity expressions in Theorem 3 and that in [16] are different, but it can be shown that they are equal. Specifically, a change of variable in (16) leads to the following equivalent

expression $C_{fb}(P) = -\log x_0$ where x_0 solves $1 - x^2 = x^2 \left(\sqrt{P} + |\alpha| \sqrt{1 - x^2} \right)^2$. Interestingly, the new polynomial and the polynomial in (15) are fundamentally different but it can be shown that they share a unique positive root meaning that the capacity is the same.

Proof of Theorem 3. In Lemma 6 in Section IV, it is shown that the Schur complement of the Riccati LMI (13) is always achieved with equality. We can show the same property for the first LMI in (14) using contradiction. Assume that $P - \Gamma^2 \hat{\Sigma}^{-1} = X$ for some $X > 0$. Then, one can choose $\Gamma' = \Gamma \sqrt{1 + \Gamma^{-2} X \hat{\Sigma}}$ to show that the objective is increased. The Riccati LMI can be verified to be satisfied with this substitution.

To obtain the capacity expression, we use $\Gamma^2 = P \hat{\Sigma}$ and the Riccati equation which simplifies to $\hat{\Sigma} = 1 - \Psi_Y^{-1}$. Substituting these equations into the objective gives the fixed-point equation $\Psi_Y = 1 + P + \alpha^2(1 - \Psi_Y^{-1}) + 2|\alpha| \sqrt{P(1 - \Psi_Y^{-1})}$, where the sign of Γ is chosen to maximize Ψ_Y . By the variable change $\Psi_Y = 1 + \text{SNR}$, we get the polynomial $\text{SNR} = \left(\sqrt{P} + |\alpha| \sqrt{\frac{\text{SNR}}{1 + \text{SNR}}} \right)^2$. $\square$

Note that the proof is a straightforward computation for all values of α , regardless of whether noise process is stationary.

IV. PROOF SKETCH OF THE MAIN RESULT

In this section we outline the proof of the main result of this paper in Theorem 1. We structure the proof as three parts.

1. **Sequential convex optimization problem (SCOP):** Define the n -letter capacity as

$$C_n(P) = \max_{P(x^n || y^n): \frac{1}{n} \sum_{i=1}^n \mathbb{E}[\mathbf{x}_i^T \mathbf{x}_i] \leq P} h(Y^n) - h(Z^n). \quad (17)$$

The first three lemmas formulate the n -letter capacity as a SCOP. Since the objective of $C_n(P)$ is directed information (e.g. [24], [25]), it is easy to show that it is concave in its decision variable $P(x^n || y^n)$, but the challenge is to formulate it as a convex optimization problem that enables one to explicitly compute the limit of $C_n(P)$ thereafter. To this end, we realize a SCOP whose fundamental LMI constraint has a sequential structure.

2. **Upper bound via convexity:** The second part of the proof utilizes the SCOP structure to show that the capacity expression in Theorem 1 is an upper bound on the capacity. Since the optimization constraints contain decision variables at consecutive times, the standard time-sharing random variable argument does not apply here, and we use a different technique to show that these constraints are *asymptotically satisfied* and not satisfied at all times.

3. **Lower bound using Time-invariant inputs:** The last part constructs a time-invariant policy whose parameters' optimization leads to a lower bound that is expressed as the upper bound optimization problem with additional constraints. We show that the additional constraints are redundant, concluding the proof of the main result.

A. Sequential convex optimization problem

Define the estimators

$$\begin{aligned} \hat{\mathbf{s}}_i &\triangleq \mathbb{E}[\mathbf{s}_i | \mathbf{z}^{i-1}] \\ \hat{\hat{\mathbf{s}}}_i &\triangleq \mathbb{E}[\hat{\mathbf{s}}_i | \mathbf{y}^{i-1}], \end{aligned} \quad (18)$$

The first lemma identifies an optimal structure for the input distribution using these estimators.

Lemma 1 (The optimal policy structure). *For a fixed n , it is sufficient to optimize (17) with inputs of the form*

$$\mathbf{x}_i = \Gamma_i \hat{\Sigma}_i^\dagger (\hat{\mathbf{s}}_i - \hat{\hat{\mathbf{s}}}_i) + \mathbf{m}_i, \quad i = 1, \dots, n \quad (19)$$

where $\mathbf{m}_i \sim N(0, M_i)$ is independent of $(\mathbf{x}^{i-1}, \mathbf{y}^{i-1})$, $\hat{\Sigma}_i^\dagger$ is the Moore-Penrose pseudo-inverse of

$$\hat{\Sigma}_i = \text{cov}(\hat{\mathbf{s}}_i - \hat{\hat{\mathbf{s}}}_i), \quad (20)$$

Γ_i is a matrix that satisfies

$$\Gamma_i (I - \hat{\Sigma}_i^\dagger \hat{\Sigma}_i) = 0, \quad (21)$$

and the power constraint is

$$\frac{1}{n} \sum_{i=1}^n \text{Tr}(\Gamma_i \hat{\Sigma}_i^\dagger \Gamma_i^T + M_i) \leq P. \quad (22)$$

Lemma 1 simplifies the optimization (17) by showing that the optimization domain is over the sequence of matrices $(\Gamma_i, M_i)_{i=1}^n$. Note that $\hat{\Sigma}_i$ is a deterministic function of the policy up to time $i-1$ and thus is not part of the policy. The main insight is that the input has two signaling components. The first component is a scaled version of the estimation error at the decoder $(\hat{\mathbf{s}}_i - \hat{\hat{\mathbf{s}}}_i)$, and its purpose is to refine the decoder's knowledge of the *channel state* $\hat{\mathbf{s}}_i$. The other component is an additive Gaussian corresponding to the new information sent to the decoder. For instance, if the noise is white, the entire power is dedicated to the new information encapsulated in $\mathbf{m}_i$.

We remark that a similar policy has been reported in [16, Section IV] and [20]. Their policy reads $\mathbf{x}_i = \Gamma_i (\hat{\mathbf{s}}_i - \hat{\hat{\mathbf{s}}}_i) + \mathbf{m}_i$, and is missing the scaling $\hat{\Sigma}_i^\dagger$ and the orthogonality constraint in (21). If $\hat{\Sigma}_i$ is invertible, then both policies are equivalent by the variable change $\Gamma'_i = \Gamma_i \hat{\Sigma}_i^{-1}$. However, the invertibility is not always true, and the orthogonality constraint must be introduced prior to the convex optimization formulation (see also Remark 1). In the next lemma, the dynamics of the channel output is formalized as a controlled state space.

Lemma 2 (Channel outputs dynamics). *For a fixed policy $\{(\Gamma_i, M_i)\}_{i=1}^n$, the channel outputs admit the state-space model*

$$\begin{aligned} \hat{\mathbf{s}}_{i+1} &= F \hat{\mathbf{s}}_i + K_{p,i} \mathbf{e}_i, \\ \mathbf{y}_i &= (\Lambda \Gamma_i \hat{\Sigma}_i^\dagger + H) \hat{\mathbf{s}}_i - \Lambda \Gamma_i \hat{\Sigma}_i^\dagger \hat{\hat{\mathbf{s}}}_i + \Lambda \mathbf{m}_i + \mathbf{e}_i, \end{aligned} \quad (23)$$

where $K_{p,i}$ and $\mathbf{e}_i \sim N(0, \Psi_i)$ are defined in (7). The estimator in (18) can be written as

$$\hat{\hat{\mathbf{s}}}_{i+1} = F \hat{\hat{\mathbf{s}}}_i + K_{Y,i} (\mathbf{y}_i - H \hat{\mathbf{s}}_i), \quad (24)$$

and its corresponding error covariance $\hat{\Sigma}_i = \text{cov}(\hat{\mathbf{s}}_i - \hat{\mathbf{s}}_i)$ satisfies the recursion

$$\hat{\Sigma}_{i+1} = F\hat{\Sigma}_iF^T + K_{p,i}\Psi_iK_{p,i}^T - K_{Y,i}\Psi_{Y,i}K_{Y,i}^T \quad (25)$$

with

$$\begin{aligned} \Psi_{Y,i} &= (\Lambda\Gamma_i\hat{\Sigma}_i^\dagger + H)\hat{\Sigma}_i(\Lambda\Gamma_i\hat{\Sigma}_i^\dagger + H)^T + \Lambda M_i\Lambda^T + \Psi_i \\ K_{Y,i} &= (F\hat{\Sigma}_i(\Lambda\Gamma_i\hat{\Sigma}_i^\dagger + H)^T + K_{p,i}\Psi_i)\Psi_{Y,i}^{-1} \end{aligned} \quad (26)$$

and $\hat{\Sigma}_1 = 0$.

Lemma 2 is a consequence of the policy that was derived in Lemma 1. As seen from (23), the encoders' policy translates into an additive measurement noise $\mathbf{m}_i$ and a modification of the observability matrix $\Lambda\Gamma_i\hat{\Sigma}_i^\dagger + H$. Similar state-space structures appeared in [19], [26]. It is interesting to realize that that (23) does not fall into the classical state-space structure since the observability matrix depends on error covariance $\hat{\Sigma}_i$ due to our policy. Lemma 2 already reveals an objective structure that resembles that in Theorem 1. Namely, by (26), we can write the objective at time i as

$$h(\mathbf{y}_i | \mathbf{y}^{i-1}) - h(\mathbf{z}_i | \mathbf{z}^{i-1}) = \frac{1}{2} \log \det(\Psi_{Y,i}) - \frac{1}{2} \log \det(\Psi_i).$$

The next lemma summarizes the SCOP formulation.

Lemma 3 (Sequential convex-optimization formulation). *The n -letter capacity can be bounded by the convex optimization problem*

$$\begin{aligned} C_n(P) &\leq \max_{\{\Gamma_i, \Pi_i, \hat{\Sigma}_{i+1}\}_{i=1}^n} \frac{1}{2n} \sum_{i=1}^n \log \det(\Psi_{Y,i}) - \log \det(\Psi_i) \\ \text{s.t.} \quad &\begin{pmatrix} \Pi_t & \Gamma_t \\ \Gamma_t^T & \hat{\Sigma}_t \end{pmatrix} \succeq 0, \quad \frac{1}{n} \sum_{i=1}^n \text{Tr}(\Pi_i) \leq P, \\ \Psi_{Y,t} &= \Lambda\Pi_t\Lambda^T + H\hat{\Sigma}_tH^T + \Lambda\Gamma_tH^T + H\Gamma_t^T\Lambda^T + \Psi_t \\ K_{Y,t} &= (F\Gamma_t^T\Lambda^T + F\hat{\Sigma}_tH^T + K_{p,t}\Psi_t)\Psi_{Y,t}^{-1} \\ &\begin{pmatrix} F\hat{\Sigma}_tF^T + K_{p,t}\Psi_tK_{p,t}^T - \hat{\Sigma}_{t+1} & K_{Y,t}\Psi_{Y,t} \\ \Psi_{Y,t}K_{Y,t}^T & \Psi_{Y,t} \end{pmatrix} \succeq 0, \end{aligned} \quad (27)$$

where the constraints hold for $t = 1, \dots, n$ and $\hat{\Sigma}_1 = 0$.

To see that (27) is a standard convex optimization, note that each of the LMI constraints is a linear function of the decision variables. In the next section, we provide the single-letter upper bound on the capacity. The key to the upper bound is the concavity of the objective function along with the linearity of the constraints, along with the crucial property that the Riccati LMI constraint contains decision variables of two consecutive times only.

B. Single-letter upper bound

The next lemma concludes the upper bound in Theorem 1.

Lemma 4 (The upper bound). *The feedback capacity is bounded by the convex optimization problem*

$$C_{fb}(P) \leq \max_{\Pi, \hat{\Sigma}, \Gamma} \frac{1}{2} \log \det(\Psi_Y) - \frac{1}{2} \log \det(\Psi)$$

$$\begin{aligned} \text{s.t.} \quad &\begin{pmatrix} \Pi & \Gamma \\ \Gamma^T & \hat{\Sigma} \end{pmatrix} \succeq 0, \quad \text{Tr}(\Pi) \leq P, \\ \Psi_Y &= \Lambda\Pi\Lambda^T + H\hat{\Sigma}H^T + \Lambda\Gamma H^T + H\Gamma^T\Lambda^T + \Psi \\ K_Y &= (F\Gamma^T\Lambda^T + F\hat{\Sigma}H^T + K_p\Psi)\Psi_Y^{-1} \\ &\begin{pmatrix} F\hat{\Sigma}F^T + K_p\Psi K_p^T - \hat{\Sigma} & K_Y\Psi_Y \\ \Psi_Y K_Y^T & \Psi_Y \end{pmatrix} \succeq 0. \end{aligned} \quad (28)$$

The main idea behind the upper bound is to show that the convex combination of each of the decision variables in Lemma 3 obtains a larger objective. At a high level, this is similar to the time-sharing random variable, but the challenge lies in the constraints. Specifically, one cannot show that the Riccati LMI constraint (28) is satisfied at all times when evaluated at the convex combination of the decision variables. To settle this point, we show that the constraint is satisfied in the asymptotics.

C. Lower bound

In this section, we prove that the upper bound in Lemma 4 is achievable. It will be shown using two lemmas: the first formulates a lower bound as an optimization problem that resembles the upper bound but has two additional constraints. The second lemma shows that in the upper bound optimization problem these two constraints are satisfied.

Lemma 5 (Lower bound). *For time-invariant policies*

$$\mathbf{x}_i = \Gamma(\hat{\mathbf{s}}_i - \hat{\mathbf{s}}_i) + \mathbf{m}_i, \quad i \geq 1 \quad (29)$$

with $\mathbf{m}_i \sim N(0, M)$, the maximization of (17) over (Γ, M) achieves the lower bound

$$\begin{aligned} C_{fb}(P) &\geq \max_{\Gamma, \Pi, \hat{\Sigma}} \log \det(\Psi_Y) - \log \det(\Psi) \\ \text{s.t.} \quad &\begin{pmatrix} \Pi & \Gamma \\ \Gamma^T & \hat{\Sigma} \end{pmatrix} \succeq 0, \quad \text{Tr}(\Pi) \leq P \\ K_Y &= (F\hat{\Sigma}H^T + F\Gamma^T\Lambda^T + K_p\Psi)\Psi_Y^{-1} \\ \Psi_Y &= \Lambda\Pi\Lambda^T + \Lambda\Gamma H^T + H\Gamma^T\Lambda^T + \Psi \\ \hat{\Sigma} &= F\hat{\Sigma}F^T + K_p\Psi K_p^T - K_Y\Psi_Y K_Y^T \\ &\exists K : \rho(F - K(\Lambda\Gamma\hat{\Sigma}^\dagger + H)) < 1. \end{aligned} \quad (30)$$

$$\exists K : \rho(F - K(\Lambda\Gamma\hat{\Sigma}^\dagger + H)) < 1. \quad (31)$$

The optimization problem in (31) is the same as the upper bound in (28) except for the additional constraint (31) and the Riccati equation (30) which appears as an inequality in the upper bound (13). Next, we show that these two conditions can be neglected and conclude the proof of Theorem 1.

Lemma 6 (Equality between the lower and upper bounds). *An optimal tuple $(\Pi, \hat{\Sigma}, \Gamma)$ for the upper bound optimization problem in (28) satisfies the following:*

- 1) The Schur complement of the Riccati LMI (13) is achieved with equality.
- 2) The pair $(F, \Lambda\Gamma\hat{\Sigma}^\dagger + H)$ is detectable, i.e.,

$$\exists K : \rho(F - K(\Lambda\Gamma\hat{\Sigma}^\dagger + H)) < 1.$$

Consequently, the upper bound in Lemma 4 and the lower bound in Lemma 5 are equal to the feedback capacity.

REFERENCES

- [1] S. Butman, "A general formulation of linear feedback communication systems with solutions," *IEEE Trans. Inf. theory*, vol. 15, pp. 392–400, 1969.
- [2] S. Butman, "Linear feedback rate bounds for regressive channels (corresp.)," *IEEE Trans. Inf. Theory*, vol. 22, no. 3, pp. 363–366, 1976.
- [3] J. Tiernan and J. Schalkwijk, "An upper bound to the capacity of the band-limited gaussian autoregressive channel with noiseless feedback," *IEEE Trans. Inf. Theory*, vol. 20, no. 3, pp. 311–316, 1974.
- [4] L. Vandenberghe, S. Boyd, and S.-P. Wu, "Determinant maximization with linear matrix inequality constraints," *SIAM J. Matrix Anal. Appl.*, vol. 19, no. 2, p. 499–533, Apr. 1998.
- [5] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear matrix inequalities in system and control theory*. SIAM, 1994.
- [6] C. Scherer and S. Weiland, "Linear matrix inequalities in control," *Lecture Notes, Dutch Institute for Systems and Control, Delft, The Netherlands*, vol. 3, no. 2, Jan. 2000.
- [7] R. J. Caverly and J. R. Forbes, "LMI properties and applications in systems, stability, and control theory," 2019, available at arxiv.org/abs/1903.08599.
- [8] 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.
- [9] T. Tanaka, K. K. Kim, P. A. Parrilo, and S. K. Mitter, "Semidefinite programming approach to Gaussian sequential rate-distortion trade-offs," *IEEE Trans. Autom. Control*, vol. 62, no. 4, pp. 1896–1910, Apr. 2017.
- [10] O. Sabag, P. Tian, V. Kostina, and B. Hassibi, "The minimal directed information needed to improve the LQG cost," in *2020 IEEE 59th Conference on Decision and Control (CDC)*, South Korea, 2020.
- [11] P. Ebert, "The capacity of the Gaussian channel with feedback," *Bell Syst. Tech. J.*, pp. 1705–1712, 1970.
- [12] A. Shahar-Doron and M. Feder, "On a capacity achieving scheme for the colored gaussian channel with feedback," in *Int. Symp. Inf. Theory (ISIT)*, 2004, p. 74.
- [13] C. Li and N. Elia, "Youla coding and computation of gaussian feedback capacity," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 3197–3215, 2018.
- [14] T. Liu and G. Han, "Feedback capacity of stationary gaussian channels further examined," *IEEE Trans. Inf. Theory*, vol. 65, no. 4, pp. 2492–2506, 2019.
- [15] S. Fang and Q. Zhu, "A connection between feedback capacity and kalman filter for colored gaussian noises," in *2020 IEEE International Symposium on Information Theory (ISIT)*, 2020, pp. 2055–2060.
- [16] Y.-H. Kim, "Feedback capacity of the first-order moving average Gaussian channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3063–3079, Jul. 2006.
- [17] T. M. Cover and S. Pombra, "Gaussian feedback capacity," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 37–43, 1989.
- [18] S. Yang, A. Kavčić, and S. C. Tatikonda, "On the feedback capacity of power constrained Gaussian channels with memory," *IEEE Trans. Inf. Theory*, vol. 53, no. 3, pp. 929–954, Mar. 2007.
- [19] Y.-H. Kim, "Feedback capacity of stationary Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 57–85, 2010.
- [20] A. Gattami, "Feedback capacity of gaussian channels revisited," *IEEE Trans. Inf. Theory*, vol. 65, no. 3, pp. 1948–1960, 2019.
- [21] O. Sabag, V. Kostina, and B. Hassibi, "Feedback capacity of MIMO Gaussian channels," 2021, available online at <http://www.its.caltech.edu/~oron/Publication.html>.
- [22] T. Kailath, "An innovations approach to least-squares estimation—part i: Linear filtering in additive white noise," *IEEE Trans. Autom. Control*, vol. 13, no. 6, pp. 646–655, 1968.
- [23] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear estimation*. Prentice Hall, 2000.
- [24] H. H. Permuter, T. Weissman, and A. J. Goldsmith, "Finite state channels with time-invariant deterministic feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 644–662, Feb. 2009.
- [25] S. Tatikonda and S. Mitter, "The capacity of channels with feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 323–349, Jan. 2009.
- [26] C. D. Charalambous, C. Kourtellis, and S. Louka, "New formulas of feedback capacity for AGN channels with memory: A time-domain sufficient statistic approach," 2020, available at arxiv.org/abs/2010.06226.