2019 18th European Control Conference (ECC)
Napoli, Italy, June 25-28, 2019

The Shannon Capacity of Linear Dynamical Networks

Giacomo Baggio, Vaibhav Katewa, Fabio Pasqualetti, and Sandro Zampieri

***Abstract*— Understanding the fundamental mechanisms enabling fast and reliable communication in the brain is one of the outstanding key challenges in neuroscience. In this work, we address this problem from a systems and information theoretic perspective. Specifically, we first develop a simple and tractable framework to model information transmission in networks driven by linear dynamics. We then resort to the notion of Shannon capacity to quantify the information transfer performance of these networks. Building on this framework, we show that it is possible to increase Shannon capacity via two fundamentally different mechanisms: either by decreasing the degree of stability of the network adjacency matrix, or by increasing its degree of non-normality. We illustrate and validate our findings by means of simple, insightful examples.**

I. INTRODUCTION

The proper functioning of many biological and technological network systems relies on their ability to efficiently process and propagate information across their units [1], [2]. For instance, cortical circuits in the brain are capable of integrating and broadcasting large volumes of sensory data in a remarkably fast and reliable way [3]. However, the fundamental network principles underlying robust and seamless communication in the brain are still poorly understood [4].

In this paper, we seek to understand the impact of network structure on the quality of information transmission through the network. To this end, we focus on networks governed by noisy linear dynamics, which have been employed, e.g., to model the firing-rate evolution of neuronal networks around equilibrium points [5, Chapter 7], [6], [7]. Drawing inspiration from observed patterns of cortical activity in the brain [7], [8], we assume that packets of information are suitably encoded in impulsive inputs. These inputs transiently excite the network dynamics, generating modulated waveforms that propagate the information through the network. The overall transmission process can be modeled as the transmission through a Gaussian channel, whose performance can be measured via the notion of Shannon channel capacity [9, Chapter 7]. Building on this framework, we examine how the network architecture affects the capacity of the network.

Related work Over the past sixty years, many experimental works have addressed the problem of quantifying the information transfer performance of the brain or other biological networks by leveraging information theoretic notions, e.g., see [10] and references therein. However, to the best of our knowledge, only a limited number of works have proposed a systematic and analytical framework to tackle this problem. In particular, in [11], the authors use the notion of Fisher information to measure the memory storage capacity of linear dynamical networks. In [12], [13] the authors investigate flexible mechanisms of information routing in networks governed by oscillatory dynamics. In [14], the authors study the interplay between network dynamics and flow patterns of information. In our earlier work [15], we consider a communication model which includes the contribution of inter-symbol interference. However, due to the intrinsic complexity of the resulting model, only the information transfer performance of a limited class of networks is analytically characterized.

Paper contribution The contribution of this paper is threefold. First, we propose a novel and tractable model of information transmission in networks driven by noisy linear dynamics, which can serve as a simple model of information transfer through neuronal networks. Second, we derive an expression for the Shannon channel capacity of our communication model subject to an input power constraint, and characterize the corresponding optimal power distribution. Third, we link the information transfer performance of the network to its structure, as described by the adjacency matrix of the network, and establish lower and upper bounds on the Shannon capacity for some classes of networks. Our results suggest that both closeness to instability and departure from normality of the network adjacency matrix play a key role in enhancing the Shannon capacity of the network. Further, as minor contributions, we analyze the dependence of Shannon capacity on the transmission time window and noise level, and characterize its behavior for high-dimensional networks.

Mathematical notation Given a matrix $A \in \mathbb{R}^{n\times n}$, we denote with $A^\top$, $\mathrm{tr}(A)$, $\det(A)$, and $\mathrm{diag}(A)$, the transpose, the trace, the determinant, and the diagonal matrix composed by the diagonal entries of A, respectively. Matrix A is said to be Hurwitz stable if all of the eigenvalues of A have strictly negative real part. If $AA^\top = A^\top A$ then A is said to be normal, otherwise A is said to be non-normal. We write $A \geq 0$ ($A > 0$) to mean that the symmetric matrix $A \in \mathbb{R}^{n\times n}$ is positive semidefinite (positive definite, respectively). The space of $n\times n$ positive semidefinite symmetric matrices is denoted by $\mathcal{S}_+^n$. We denote with $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ the (directed) graph with vertex set $\mathcal{V} = \{1, 2, \dots, n\}$ and edge set $\mathcal{E} \subseteq \mathcal{V}\times\mathcal{V}$. The (weighted) adjacency matrix $A \in \mathbb{R}^{n\times n}$ corresponding to the graph $\mathcal{G}$ satisfies $[A]_{ij} \neq 0$ if and only if $(j, i) \in \mathcal{E}$, where $[A]_{ij}$ denotes the (i, j)-th entry of A. We denote with $\mathcal{N}(\mu, \Sigma)$ the n-dimensional Gaussian distribution with mean $\mu \in \mathbb{R}^n$ and covariance $\Sigma \in \mathcal{S}_+^n$, and with $\mathbb{E}[\cdot]$ the expectation of a random variable. $\mathcal{L}_2^p[t_1, t_2]$ stands for the Hilbert space of p-dimensional square integrable functions in $[t_1, t_2]$, $t_2 > t_1$, equipped with the inner product $\langle f, g\rangle_{\mathcal{L}_2} := \int_{t_1}^{t_2} f^\top(t)g(t)\,\mathrm{d}t$ and norm $\|f\|_{\mathcal{L}_2} = \sqrt{\langle f, f\rangle_{\mathcal{L}_2}}$. The i-th canonical vector of $\mathbb{R}^n$ is denoted with e_i, and the Dirac delta function with $\delta(t)$. Finally, we let $x^+ = \max\{0, x\}$, $x \in \mathbb{R}$.

This material is based upon work supported in part by ARO award 71603NSYIP. Giacomo Baggio, Vaibhav Katewa, and Fabio Pasqualetti are with the Department of Mechanical Engineering, University of California at Riverside, USA, {gbaggio,vkatewa,fabiopas}@engr.ucr.edu. Sandro Zampieri is with the Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Padova, Italy, zampi@dei.unipd.it.

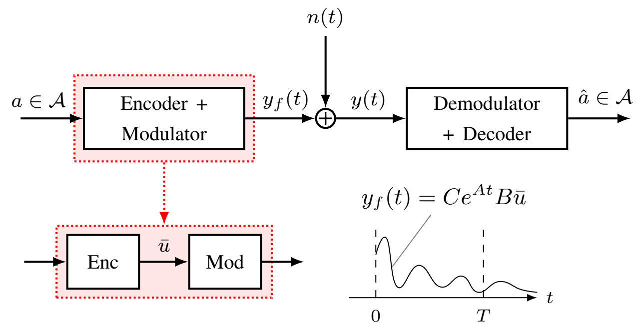

Fig. 1. Schematic of the communication model considered in this paper.

II. Modeling Framework

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a directed graph with weighted adjacency matrix $A \in \mathbb{R}^{n \times n}$. We consider continuous-time, linear, time invariant network dynamics

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t), \\ y(t) = Cx(t) + n(t), \end{cases} \tag{1}$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$, and $y(t) \in \mathbb{R}^p$ denote the vectors of nodes' states, inputs, and outputs at time $t \geq 0$, respectively. The term $n(t) \in \mathbb{R}^p$ represents a white Gaussian noise process, $n(t) \sim \mathcal{N}(0, \sigma^2 I_p)$, $n(t) \perp n(s)$ for $t \neq s$, $t, s \geq 0$. We let $x(0) = 0$, and form $B \in \mathbb{R}^{n \times m}$ and $C \in \mathbb{R}^{p \times n}$ so as to select two subsets of nodes in $\mathcal{V}$, namely,

$$B = [e_{k_1}, \ldots, e_{k_m}], \quad C = [e_{t_1}, \ldots, e_{t_p}]^\top,$$

where $\mathcal{K} := \{k_i\}_{i=1}^m \subseteq \mathcal{V}$ and $\mathcal{T} := \{t_i\}_{i=1}^p \subseteq \mathcal{V}$.

We model information transmission through the network $\mathcal{G}$ as in Fig. 1. In our framework, the packet of information to be transmitted at time $t = 0$ is represented by the symbol a, which belongs to an alphabet $\mathcal{A}$ of finite cardinality. The duration of a transmission is denoted with $T > 0$. The proposed communication protocol consists of two steps:

1) *Encoding and modulation.* The symbol a is first mapped to a vector $\bar{u} \in \mathbb{R}^m$ (*codeword*) which acts as an impulsive input $u(t) = \bar{u}\,\delta(t)$ that transiently excites the linear system in (1). The corresponding forced response of the system (*modulated waveform*)

$$y_f(t) = Ce^{At}B\bar{u}, \ t \in [0, T],$$

propagates the packet of information encoded in $\bar{u}$ across the network.

2) *Demodulation and decoding.* The (noisy) output trajectory of the system in (1), i.e.,

$$y(t) = y_f(t) + n(t), \ t \in [0, T],$$

is demodulated and decoded to recover an estimate $\hat{a} \in \mathcal{A}$ of the transmitted symbol.

In what follows, we will focus on the information transfer performance of a single transmission in the interval $[0, T]$ and will make use of the following two assumptions:

A1) The dynamics in (1) is stable, i.e., A is Hurwitz stable.

A2) The power available at the sender is limited. This implies that all of the input codewords $\bar{u}$'s have bounded norm $\|\bar{u}\| \leq P$, for $P > 0$.

Remark 1: ***(Multiple transmissions and interference)*** When transmissions are performed consecutively, modulated waveforms corresponding to previous symbols may interfere with the current one. A framework taking into account this inter-symbol interference phenomenon has been investigated in [15]. In the present work, the noise term coming from inter-symbol interference is assumed to be negligible. □

III. Shannon Channel Capacity

We use Shannon channel capacity as a measure of information transmission efficiency. In fact, Shannon capacity provides a tight upper bound on the amount of information (measured in bits per channel use) that can be sent reliably, that is, with arbitrarily small decoding error probability, over a communication channel (e.g., see [9] for further details).

Theorem 1: ***(Shannon channel capacity)*** Let $T > 0$. The Shannon capacity of the communication channel in Sec. II is

$$\mathcal{C}_T = \frac{1}{2} \max_{\substack{\Sigma \in \mathcal{S}_+^m \\ \operatorname{tr}\Sigma \leq P}} \log_2 \det \left(I_m + \frac{1}{\sigma^2} \Sigma\, B^\top \mathcal{O}_T B \right), \tag{2}$$

where $\mathcal{O}_T := \int_0^T e^{A^\top t} C^\top C e^{At} \mathrm{d}t$ denotes the observability Gramian of the system (1) over the interval $[0, T]$.

It is worth noting that $B^\top \mathcal{O}_T B$ quantifies the energy of the impulse response of the linear system in (1). Specifically, for a given codeword vector $\bar{u} \in \mathbb{R}^m$, it holds

$$\|y_f(t)\|_{\mathcal{L}_2}^2 = \bar{u}^\top B^\top \mathcal{O}_T B\, \bar{u}.$$

This observation suggests that linear networks featuring a highly energetic impulse response in the interval $[0, T]$ are likely to transmit packets of information more efficiently.

Remark 2: ***(Shannon capacity and transmission window)*** Shannon capacity (2) is a monotonically increasing function of the transmission window T. This follows from the fact that, for any $T_1, T_2 > 0$ such that $T_2 \geq T_1$, it holds $\mathcal{O}_{T_2} \geq \mathcal{O}_{T_1}$. Intuitively, as T increases, a longer portion of the modulated signal $y_f(t)$ reaches the receiver, allowing for a more accurate decoding of the transmitted symbol. □

The following result characterizes the optimal Σ which yields the solution of the maximization problem in (2).

Theorem 2: ***(Optimal power allocation and decomposition into independent subchannels)*** Let $\{\mu_i\}_{i=1}^m$ be the eigenvalues of $B^\top \mathcal{O}_T B$ and let $\{u_i\}_{i=1}^m$ be the orthonormal set of corresponding eigenvectors. Let $U := [u_1, \ldots, u_m] \in \mathbb{R}^{m \times m}$. The optimal Σ in (2) has the form

$$\Sigma^\star = U \operatorname{diag}(P_1, \ldots, P_m) U^\top,$$

where

$$P_i = \begin{cases} \left(\nu - \frac{\sigma^2}{\mu_i} \right)^+, & \text{if } \mu_i \neq 0, \\ 0, & \text{if } \mu_i = 0, \end{cases} \tag{3}$$

and $\nu > 0$ is chosen such that $\sum_{i=1}^m P_i = P$. Moreover, the Shannon capacity in (2) can be equivalently rewritten as

$$\mathcal{C}_T = \frac{1}{2} \sum_{i=1}^m \log_2 \left(1 + \frac{P_i \mu_i}{\sigma^2} \right). \tag{4}$$

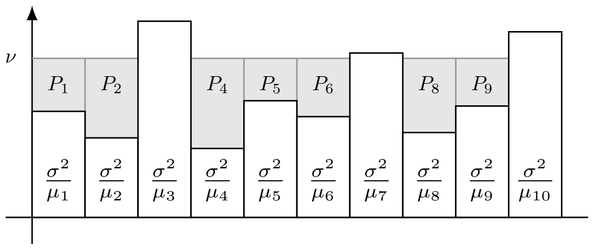

Fig. 2. Example of optimal power allocation for $m = 10$. Note that more power (gray bars) is alloted to the subchannels with higher ratio μ_i/σ^2, whereas no power is assigned to the most noisy subchannels (3, 7, and 10).

Intuitively, Equations (3) and (4) suggest that the communication channel of Section II can be decoupled into m parallel independent subchannels, and the optimal power distribution assigns more power to the subchannels with higher "signal-to-noise" ratio μ_i/σ^2. Following a standard waterfilling argument in information theory [9, Chapter 9], ν can be thought of as the *waterfill level* that marks the height of the power that is poured into the *water vessel* composed by m cylinders with heights equal to σ^2/μ_i (see also Fig. 2).

Remark 3: ***(Shannon capacity in high/low SNR regimes)*** Define the overall channel signal-to-noise as $\text{SNR} = P/\sigma^2$. In the high SNR regime ($\text{SNR} \approx \infty$), the optimal power distribution allocates approximately equal power to the subchannels associated with the non-zero eigenvalues of $\mathcal{O}_T$. In this case, the capacity in (4) can be approximated as

$$\mathcal{C}_T \approx \frac{1}{2}\sum_{i=1}^{r} \log_2\left(1 + \text{SNR}\,\frac{\mu_i}{r}\right) \approx \frac{r}{2}\log_2 \text{SNR} + \frac{1}{2}\sum_{i=1}^{r}\log_2\left(\frac{\mu_i}{r}\right) \approx \frac{r}{2}\log_2 \text{SNR},$$

where r denotes the rank of $B^\top \mathcal{O}_T B$. Therefore, for large values of SNR, the capacity depends only weakly on the network architecture (described by the adjacency matrix A).[1] In the low SNR regime ($\text{SNR} \approx 0$), in view of (3), the optimal power distribution allocates all power to the subchannel with largest ratio μ_i/σ^2. Thus, from (4), we have

$$\mathcal{C}_T \approx \frac{1}{2}\log_2(1 + \text{SNR}\,\mu_{\max}) \approx \frac{\mu_{\max}}{2\ln 2}\text{SNR},$$

where $\mu_{\max} := \max_i \mu_i$ and we used the approximation $\ln(1+x) \approx x$ for $x \approx 0$. From the above expression, for low SNR, it is apparent that the capacity depends on the network structure via the largest eigenvalue of $B^\top \mathcal{O}_T B$. □

Remark 4: ***(Shannon capacity of high-dimensional networks)*** Consider a sequence of directed graphs of increasing dimensions $\{\mathcal{G}_n\}_{n>0}$, $\mathcal{G}_n = (\mathcal{V}_n, \mathcal{E}_n)$, and assume that the input and output nodes subsets of $\mathcal{G}_n$ coincide with $\mathcal{V}_n$ for all n. Let $A_n \in \mathbb{R}^{n\times n}$ be the adjacency matrix of $\mathcal{G}_n$, $\mathcal{O}_{T,n}$ be the $[0, T]$ observability Gramian of the pair (A_n, I_n), and $P_n > 0$ be the total input power associated with $\mathcal{G}_n$. We assume that the input power scales linearly with the network size, i.e., $P_n = nP$, with $P > 0$, and the density of the eigenvalues of $\{\mathcal{O}_{T,n}\}_{n>0}$ converges to a continuous function $f\colon [0,\infty) \to \mathbb{R}_+$. Then, for large n, $\mathcal{C}_T$ grows linearly with n and can be approximated as (see also [17])

$$\mathcal{C}_T \approx \frac{n}{2}\int_0^\infty \log_2\left(1 + \frac{P(\mu)\mu}{\sigma^2}\right) f(\mu)\,\mathrm{d}\mu, \tag{5}$$

where $P\colon [0,\infty) \to \mathbb{R}_+$ is defined as $P(\mu) = \left(\nu - \sigma^2/\mu\right)^+$, and $\nu > 0$ is determined by the integral constraint $\int_0^\infty P(\mu)f(\mu)\,\mathrm{d}\mu = P$. Furthermore, it is interesting to note that, if $A_n \in \mathbb{R}^{n\times n}$ is symmetric, then $\mu_i = (e^{2\lambda_i T} - 1)/(2\lambda_i)$, where $\{\lambda_i\}_{i=1}^n$ and $\{\mu_i\}_{i=1}^n$ denote the eigenvalues of A_n and $\mathcal{O}_{T,n}$, respectively. Using the above relation, it is possible to rewrite the asymptotic capacity in (5) and the corresponding optimal power density $p(\mu) := P(\mu)f(\mu)$ in terms of the eigenvalue density of $\lim_{n\to\infty} A_n$ (assuming it exists). We illustrate this in the next examples.□

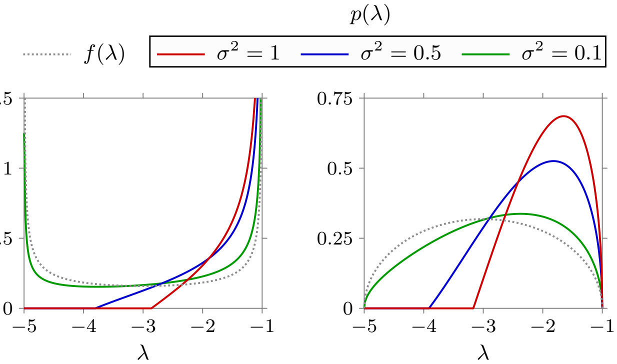

Fig. 3. Eigenvalue density $f(\lambda)$ (dotted lines) and optimal asymptotic power density $p(\lambda)$ (solid lines) for the Toeplitz line network of Example 1 with $\gamma = 3$ and $\beta = 1$ (left panel), and the Erdős–Rényi random network of Example 2 with $\varepsilon = 3$ (right panel). In all plots, we fixed $P = T = 1$.

Example 1: ***(Optimal power distribution for high-dimensional Toeplitz line networks)*** Consider the line network $\mathcal{G}_n$ described by the symmetric tridiagonal adjacency matrix $A_n \in \mathbb{R}^{n\times n}$ defined as

$$[A_n]_{ij} := \begin{cases} -\gamma, & \text{if } i = j,\\ \beta, & \text{if } i = j \pm 1,\\ 0, & \text{otherwise},\end{cases}$$

for $\gamma > 0$, $\beta > 0$ and $-\gamma + 2\beta < 0$. This ensures that A_n is stable. As $n \to \infty$, it is possible to show that the eigenvalue density of the sequence $\{A_n\}_{n>0}$ converges to (see [18])

$$f(\lambda) = \begin{cases} \frac{1}{\pi\sqrt{4\beta^2 - (\lambda+\gamma)^2}}, & -\gamma - 2\beta \le \lambda \le -\gamma + 2\beta,\\ 0, & \text{otherwise}.\end{cases}$$

Hence, in view of Remark 4, the optimal power density reads as $p(\lambda) = \left(\nu - 2\lambda\sigma^2/(e^{2\lambda T} - 1)\right)^+ f(\lambda)$, where $\nu > 0$ is fixed by $\int_{\gamma-2\beta}^{\gamma+2\beta} p(\lambda)\,\mathrm{d}\lambda = P$. In the left panel of Fig. 3, we plot the profiles of $f(\lambda)$ and $p(\lambda)$ for three different values of the noise variance σ^2. It is interesting to note that, as σ^2 grows, the mean of the optimal power density shifts towards the eigenvalue that is closest to instability. □

Example 2: ***(Optimal power distribution for high-dimensional Erdős–Rényi networks)*** Let $\text{ER}(n, p_n)$ denote the (undirected) Erdős–Rényi graph ensemble with n nodes

[1] More precisely, the dependence is via the rank of $B^\top \mathcal{O}_T B$, which is in turn related to the structural observability properties of the network [16].

and edge probability p_n. Let A_n be the (symmetric) adjacency matrix corresponding to a realization of $\mathrm{ER}(n, p_n)$, and let $\tilde{A}_n := \frac{1}{\sigma} A_n - \mathbb{E}[A_n] - \varepsilon I_n$, with $\sigma = \sqrt{np_n(1-p_n)}$ and $\varepsilon > 2$, denote the centralized and stabilized version of A_n. If $\lim_{n\to\infty} np_n \to \infty$ then the eigenvalue density of the sequence $\{\tilde{A}_n\}_{n>0}$ converges (in distribution) to the Wigner's semicircle distribution (see [19])

$$f(\lambda) = \begin{cases} \frac{1}{2\pi}\sqrt{4-(\lambda+\varepsilon)^2}, & -2-\varepsilon \leq \lambda \leq 2-\varepsilon, \\ 0, & \text{otherwise.} \end{cases}$$

We can apply the formula in Remark 4 to compute the optimal asymptotic power density $p(\lambda)$. In Fig. 3, right panel, we illustrate the profiles of $f(\lambda)$ and $p(\lambda)$ for three different levels of the noise variance σ^2. Similarly to the behavior observed in Example 1, as σ^2 increases, the optimal power density concentrates around the subchannels associated with the eigenvalues that are closer to instability. □

IV. The Role of Network Structure

In this section, we investigate how the network architecture, as described by the adjacency matrix A, affects the performance of information transfer across the network.

Proposition 1: ***(Upper bound on $\mathcal{C}_T$ for normal networks)*** Assume that $A \in \mathbb{R}^{n\times n}$ is normal and Hurwitz stable with eigenvalues $\{\lambda_i\}_{i=1}^n$. For all $T > 0$, it holds

$$\mathcal{C}_T \leq \frac{1}{2}\sum_{i=1}^{n} \log_2\left(1 + \frac{P_i(e^{2\mathrm{Re}\lambda_i T}-1)}{2\sigma^2 \operatorname{Re}\lambda_i}\right), \qquad (6)$$

where $P_i = \left(\nu - (2\sigma^2 \operatorname{Re}\lambda_i)/(e^{2\mathrm{Re}\lambda_i T}-1)\right)^+$, $i = 1,\dots,n$, and $\nu > 0$ is chosen s.t. $\sum_{i=1}^n P_i = P$. Moreover, the inequality in (6) is satisfied with equality if $\mathcal{K} = \mathcal{T} = \mathcal{V}$.

The above proposition sets a fundamental limit on the capacity of normal networks in terms of the spectrum of A. Indeed, from (6), the capacity of a normal network cannot increase unless the real part of some eigenvalues approaches zero. In particular, in the limit $\operatorname{Re}\lambda_i \to 0$ for all $i = 1,\dots,n$, the bound in (6) becomes

$$\mathcal{C}_T \leq \frac{n}{2}\log_2\left(1 + \frac{\mathrm{SNR}\, T}{n}\right),$$

where $\mathrm{SNR} = P/\sigma^2$ and we used the fact that $e^{Tx} - 1 \approx Tx$ for $x \approx 0$. In addition, when $\mathcal{K} = \mathcal{T} = \mathcal{V}$, then the above bound is attained with equality. In this case, for large transmission windows ($T \to \infty$), the capacity grows unbounded as A approaches instability. This is in sharp contrast with the findings of [15], where the contribution of interference generated by previous transmissions is included in the communication model. As a matter of fact, in the latter framework, the information transmission performance of the network always degrades as $\mathrm{tr}(A)$ approaches zero.

Proposition 2: ***(Lower bound on $\mathcal{C}_T$ for a class of non-normal networks)*** Assume that $A \in \mathbb{R}^{n\times n}$ has the form $A = DSD^{-1}$, where $S \in \mathbb{R}^{n\times n}$ is a Hurwitz stable matrix and $D := \mathrm{diag}(1, \alpha, \dots, \alpha^{n-1})$, $\alpha > 0$. For any $\mathcal{K}$ and $\mathcal{T}$ such that $1 \in \mathcal{K}$ and $n \in \mathcal{T}$, it holds

$$\mathcal{C}_T \geq \frac{1}{2}\log_2\left(1 + \alpha^{2(n-1)}\mathrm{SNR}\,[\tilde{\mathcal{O}}_T]_{11}\right), \qquad (7)$$

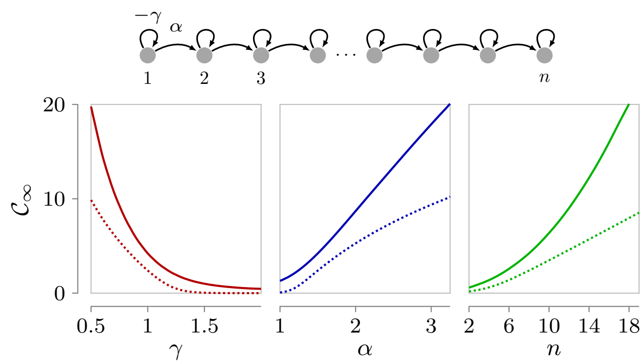

Fig. 4. Shannon capacity $\mathcal{C}_\infty$ for the directed line network topology of Example 3 with $\mathcal{K} = \mathcal{T} = \mathcal{V}$ and $\mathrm{SNR} = P/\sigma^2 = 1$ (solid lines) and the corresponding lower bounds of Proposition 2 (dotted lines) for different values of the parameters γ, $\alpha > 0$ and $n \in \mathbb{N}_{>0}$. In the left plot, we fixed $\alpha = 1.5$, $n = 8$, and let γ vary. In the middle plot, we fixed $\gamma = 1$, $n = 8$, and let α vary. In the right plot, we fixed $\gamma = 1$, $\alpha = 1.5$, and let n vary.

where $\mathrm{SNR} = P/\sigma^2$ and $\tilde{\mathcal{O}}_T$ is the $[0, T]$ observability Gramian of the pair $(S, e_n^\top)$. Moreover, the inequality in (7) is satisfied with equality if $\mathcal{K} = \{1\}$ and $\mathcal{T} = \{n\}$.

The latter result suggests that the "degree of non-normality" of A [20] also enhances the performance of information transmission. In fact, for the class of non-normal networks in Proposition 2, it is possible to increase $\mathcal{C}_T$ by increasing the parameter α and/or the network dimension n (if $[\tilde{\mathcal{O}}_T]_{11}$ does not decrease too quickly with n). These two parameters α, n regulate the degree of non-normality of the network, in the sense that they quantify the spread and intensity of directional paths in the network [21]. In the next example, we illustrate the relation between capacity and degree of non-normality of A for a simple network topology.

Example 3: ***(Shannon capacity of directed line networks)*** Consider a directed line network $\mathcal{G}$ described by the adjacency matrix $A \in \mathbb{R}^{n\times n}$,

$$[A]_{ij} := \begin{cases} -\gamma, & \text{if } i = j, \\ \alpha, & \text{if } i = j+1, \\ 0, & \text{otherwise,} \end{cases}$$

where $\alpha > 0$ and $\gamma > 0$, and assume that $T \to \infty$. Notice that A can be written in the form $A = DSD^{-1}$ where $D = \mathrm{diag}(1, \alpha, \dots, \alpha^{n-1})$ and $S \in \mathbb{R}^{n\times n}$ satisfies $[S]_{ij} = -\gamma$ if $i = j$, $[S]_{ij} = 1$ if $i = j+1$, and $[S]_{ij} = 0$, otherwise. Following [22], it can be shown that the infinite horizon observability Gramian $\tilde{\mathcal{O}}_\infty$ of the pair $(S, e_n^\top)$ admits a closed-form expression in terms of a Pascal matrix, and its first diagonal element can be expressed as $[\tilde{\mathcal{O}}_\infty]_{11} = \frac{1}{(2\gamma)^{2n-1}}\binom{2n-2}{n-1}$. Thus, by Proposition 2, if the input and output nodes $\mathcal{K}$ and $\mathcal{T}$ are such that $1 \in \mathcal{K}$ and $n \in \mathcal{T}$, $n \geq 2$, we have

$$\begin{aligned} \mathcal{C}_\infty &\geq \frac{1}{2}\log_2\left(1 + \mathrm{SNR}\,\frac{2\gamma}{\alpha^2}\left(\frac{\alpha}{2\gamma}\right)^{2n}\binom{2n-2}{n-1}\right) \\ &> \frac{1}{2}\log_2\left(1 + \frac{\mathrm{SNR}\,\gamma}{\alpha^2\sqrt{\pi(n-1)}}\left(\frac{\alpha}{\gamma}\right)^{2n}\right), \end{aligned} \qquad (8)$$

where in the last step we used the bound $\binom{2x}{x} > \frac{1}{2}\frac{4^x}{\sqrt{\pi x}}$, $x \in$

$\mathbb{N}_{>0}$, which follows from Stirling's approximation formula.[2] From (8), it is apparent that the capacity grows unbounded if (i) A approaches instability ($\gamma \to 0$), or (ii) A is strongly non-normal ($\alpha \to \infty$). Interestingly, the bound (8) goes to infinity also when the entries of A are bounded, $\alpha/\gamma > 1$, and the network dimension is very large (in the limit $n \to \infty$). This is actually another way to increase the degree of non-normality of A. In Fig. 4, we illustrate the behavior of the Shannon capacity $\mathcal{C}_\infty$ when $\mathcal{K} = \mathcal{T} = \mathcal{V}$, and the bounds of Proposition 2 for different values of γ, α, and n. □

V. Conclusion

In this paper, we present a theoretical framework for modeling information transfer through networks governed by noisy linear dynamics. We turn to the notion of Shannon channel capacity to measure the performance of information transfer across the network, and derive an expression for the capacity of our communication model. We then investigate how the network structure affects the capacity of the network.

From our analytical and numerical results, two network properties that can significantly affect the quality of information transmission are the distance to instability and the degree of non-normality of the network adjacency matrix. Specifically, networks that are close to instability or strongly non-normal yield a better information transfer performance. Our findings are also interesting when compared to recent works that examine the interplay between controllability and fragility of networks [24], and the importance of non-normality in neuronal [6], [7] and real-world networks [25].

Appendix

Proof of Theorem 1: We first address the case $\mathcal{K} = \mathcal{V}$, i.e. $B = I_n$, and then extend the argument to the general case $\mathcal{K} \subseteq \mathcal{V}$. With reference to the communication channel of Section II, consider the modulated signal $y_f(t) = Ce^{At}\bar{u}$, $0 \le t \le T$, containing the to-be-transmitted information. Notice that $y_f(t)$ belongs to the finite-dimensional subspace $\mathcal{Q}$ of the space of square integrable function $\mathcal{L}_2^p[0,T]$ generated by the functions $\{Ce^{At}e_i,\ t \in [0,T]\}_{i=1}^n$. Thus, $y_f(t)$ can be written as $y_f(t) = \sum_{i=1}^M y_i f_i(t)$, where $\{f_i(t)\}_{i=1}^M$ is any orthonormal basis in $\mathcal{Q}$ and

$$y_i := \langle f_i(t), y_f(t)\rangle_{\mathcal{L}_2} = \underbrace{\int_0^T f_i^\top(t) C e^{At}\,\mathrm{d}t}_{=:F_i^\top}\,\bar{u}.$$

Let us define $F := [F_1, \dots, F_M]^\top$ and $Y_f := [y_1, \dots, y_M]^\top$. For all $\bar{u} \in \mathbb{R}^n$, it holds

$$\langle y_f(t), y_f(t)\rangle_{\mathcal{L}_2} = Y_f^\top Y_f = \bar{u}^\top F^\top F \bar{u}.$$

This in turn implies that $F^\top F = \mathcal{O}_T$ where $\mathcal{O}_T = \int_0^T e^{A^\top t} C^\top C e^{At}\,\mathrm{d}t$ is the $[0,T]$ observability Gramian of the system (1). The covariance between two components y_h and y_ℓ, $h, \ell = 1, 2, \dots, M$, is given by $\mathbb{E}[y_h y_\ell] = \mathbb{E}\left[F_h^\top \bar{u}\bar{u}^\top F_\ell\right] = F_h^\top \Sigma F_\ell$, where $\Sigma := \mathbb{E}[\bar{u}\bar{u}^\top]$ is the covariance of the codewords distribution. Thus, the covariance of $y_f(t)$ is given by $\Sigma_{y_f} := \mathbb{E}[Y_f Y_f^\top] = F\Sigma F^\top$. The channel noise $n(t)$ can be written as, w.r.t. the previously introduced orthonormal basis $\{f_i(t)\}_{i=1}^M$ of $\mathcal{Q}$, $n(t) = \sum_{i=1}^M n_i f_i(t) + n_\perp(t)$, where $n_\perp(t)$ belongs to the orthogonal complement of $\mathcal{Q}$. The covariance of n_h, n_ℓ, $h, \ell = 1, 2, \dots, M$, is

$$\mathbb{E}[n_h n_\ell] = \sigma^2 \int_0^T \int_0^T f_h^\top(t) f_\ell(\tau)\delta(t-\tau)\,\mathrm{d}t\,\mathrm{d}\tau = \sigma^2 \delta_{h,\ell},$$

where $\delta_{h,\ell}$ denotes the Kronecker delta function. By defining $N := [n_1, \dots, n_M]^\top$, we have that the covariance of the "projected" noise is $\Sigma_n := \mathbb{E}[NN^\top] = \sigma^2 I_n$.

Next, we note that the channel of Section II coincides with an Additive White Gaussian Noise (AWGN) channel with input $y_f(t) \in \mathcal{L}_2^p[0,T]$ and output $y_f(t) + n(t) \in \mathcal{L}_2^p[0,T]$. Thus, by exploiting the closed-form expression of the channel capacity of an AWGN channel [9, Chapter 9] with the power input constraint introduced in Assumption A2), it follows that

$$\begin{aligned}\mathcal{C}_T &= \frac{1}{2} \max_{\substack{\Sigma \in \mathcal{S}_+^n \\ \operatorname{tr}\Sigma \le P}} \log_2 \frac{\det(\Sigma_{y_f} + \Sigma_n)}{\det \Sigma_n}, \\ &= \frac{1}{2} \max_{\substack{\Sigma \in \mathcal{S}_+^n \\ \operatorname{tr}\Sigma \le P}} \log_2 \det\left(I_n + \frac{1}{\sigma^2}\Sigma\,\mathcal{O}_T\right), \end{aligned} \tag{9}$$

where, in the last step, we used the similarity invariance of the determinant, and the fact that $F^\top F = \mathcal{O}_T$. We consider now the general case $\mathcal{K} \subseteq \mathcal{V}$, with $|\mathcal{K}| = m$. In this case, $\bar{u} \in \mathbb{R}^m$ and the search space $\{\Sigma \in \mathcal{S}_+^n,\ \operatorname{tr}\Sigma \le P\}$ in the maximization in (9) must be replaced by $\{B\Sigma B^\top : \Sigma \in \mathcal{S}_+^m,\ \operatorname{tr}\Sigma \le P\}$. By further noting that $\operatorname{tr}(B\Sigma B^\top) = \operatorname{tr}(B^\top B\,\Sigma) = \operatorname{tr}\Sigma$, this yields the equivalent expression

$$\mathcal{C}_T = \frac{1}{2} \max_{\substack{\Sigma \in \mathcal{S}_+^m \\ \operatorname{tr}\Sigma \le P}} \log_2 \det\left(I_m + \frac{1}{\sigma^2} B\Sigma B^\top \mathcal{O}_T\right).$$

To conclude, the channel capacity formula in (2) follows from the determinant identity $\det(I + XY) = \det(I + YX)$, where X and Y are matrices of compatible dimensions. ■

Proof of Theorem 2: The proof follows a quite standard information-theoretic argument, e.g., see [9, Section 9.4]. The main steps are reported below for the sake of completeness. Consider the function

$$c_T\colon \mathcal{S}_+^m \to \mathbb{R}_+,\ \Sigma \mapsto \log_2 \det\left(I_m + \frac{1}{\sigma^2}\Sigma B^\top \mathcal{O}_T B\right),$$

and let U be the (orthogonal) matrix of eigenvectors of $B^\top \mathcal{O}_T B$. For any $\Sigma \in \mathcal{S}_+^m$, it holds

$$\begin{aligned} c_T(U\Sigma U^\top) &= \log_2 \det\left(I_m + \frac{1}{\sigma^2} U\Sigma U^\top B^\top \mathcal{O}_T B\right) \\ &= \log_2 \det\left(I_m + \frac{1}{\sigma^2}\Sigma U^\top B^\top \mathcal{O}_T B U\right) \\ &= \log_2 \det\left(I_m + \frac{1}{\sigma^2}\Sigma \operatorname{diag}(\mu_1, \dots, \mu_m)\right), \end{aligned}$$

[2]Specifically, using the Stirling's estimate [23] $x! = \sqrt{2\pi x}(x/e)^x e^{\theta(x)}$, with $x \in \mathbb{N}_{>0}$ and $1/(12x+1) \le \theta(x) \le 1/(12x)$, we can write $\binom{2x}{x} = 4^x e^{\theta(2x) - 2\theta(x)}/\sqrt{\pi x}$, where $\theta(x)$ satisfies the previous inequalities. Thus, the desired bound follows from the fact that $\theta(2x) - 2\theta(x) > -\ln 2$.

where $\{\mu_i\}_{i=1}^m$ are the eigenvalues of $B^\top \mathcal{O}_T B$. Further, by Hadamard's inequality [26, Theorem 7.8.1],

$$c_T(U\Sigma U^\top) \leq \sum_{i=1}^m \log_2\left(1 + \frac{\mu_i}{\sigma^2}[\Sigma]_{ii}\right),$$

with equality when Σ is diagonal. Thus, the matrix $\Sigma^\star$ maximizing $c_T(\Sigma)$ must be of the form $\Sigma^\star = U \operatorname{diag}(P_1, \ldots, P_m) U^\top$, with $P_i \geq 0$. Moreover, we have

$$c_T(\Sigma^\star) = \sum_{i=1}^m \log_2\left(1 + \frac{\mu_i P_i}{\sigma^2}\right).$$

Under the constraint $\sum_{i=1}^m P_i \leq P$, the computation of the optimal P_i's boils down to a standard constrained optimization problem that can be solved via Lagrange multipliers [9, Section 9.4]. The solution of the latter problem yields the optimal power distribution of Equation (3). ■

The following lemma is used to prove Proposition 1 and 2.

Lemma 1: ***(Monotonicity of $\mathcal{C}_T$ in $\mathcal{K}$ and $\mathcal{T}$)*** Given any two sets $\mathcal{K}_1, \mathcal{K}_2 \subseteq \mathcal{V}$ such that $\mathcal{K}_1 \subseteq \mathcal{K}_2$ and $\mathcal{T}_1, \mathcal{T}_2 \subseteq \mathcal{V}$ such that $\mathcal{T}_1 \subseteq \mathcal{T}_2$. It holds $\mathcal{C}_T(\mathcal{K}_1, \mathcal{T}_1) \leq \mathcal{C}_T(\mathcal{K}_2, \mathcal{T}_2)$, $\forall T > 0$, where we made explicit the dependence of $\mathcal{C}_T$ on the input and output sets $\mathcal{K}$ and $\mathcal{T}$, respectively.

Proof: We can rewrite the capacity in (2) as

$$\mathcal{C}_T(\mathcal{K}, \mathcal{T}) = \frac{1}{2} \max_{\Sigma \in \mathcal{E}_\mathcal{K}} \log_2 \det\left(I_n + \frac{1}{\sigma^2} \Sigma^{1/2}\, \mathcal{O}_{T,\mathcal{T}}\, \Sigma^{1/2}\right),$$

where $\mathcal{E}_\mathcal{K} := \{B\Sigma B^\top : \Sigma \in \mathcal{S}_+^m,\ \operatorname{tr} \Sigma \leq P\}$ and we used the notation $\mathcal{O}_{T,\mathcal{T}}$ to make explicit the dependence of $\mathcal{O}_T$ on $\mathcal{T}$. Next, we notice that: (i) if $\mathcal{K}_1 \subseteq \mathcal{K}_2$, then $\mathcal{E}_{\mathcal{K}_1} \subseteq \mathcal{E}_{\mathcal{K}_2}$, and (ii) if $\mathcal{T}_1 \subseteq \mathcal{T}_2$, then $\mathcal{O}_{T,\mathcal{T}_1} \leq \mathcal{O}_{T,\mathcal{T}_2}$. Together, the latter two facts and the above expression of the capacity imply that, if $\mathcal{K}_1 \subseteq \mathcal{K}_2$ and $\mathcal{T}_1 \subseteq \mathcal{T}_2$, then $\mathcal{C}_T(\mathcal{K}_1, \mathcal{T}_1) \leq \mathcal{C}_T(\mathcal{K}_2, \mathcal{T}_2)$. ■

Proof of Proposition 1: If $\mathcal{K} = \mathcal{T} = \mathcal{V}$ and A is normal, then $\mathcal{O}_T$ and A are diagonalizable in the same unitary basis, and it holds $\mu_i = (e^{2\mathrm{Re}\lambda_i T} - 1)/(2\mathrm{Re}(\lambda_i))$. Thus, by substituting the latter μ_i's in Equation (4), we obtain

$$\mathcal{C}_T = \frac{1}{2} \sum_{i=1}^n \log_2\left(1 + \frac{P_i(e^{2\mathrm{Re}\lambda_i T} - 1)}{2\sigma^2 \operatorname{Re} \lambda_i}\right),$$

where $P_i = \left(\nu - (2\sigma^2 \operatorname{Re} \lambda_i)/(e^{2\mathrm{Re}\lambda_i T} - 1)\right)^+$, $i = 1, \ldots, n$, and $\nu > 0$ is s.t. $\sum_{i=1}^n P_i = P$. If $\mathcal{K} \subset \mathcal{V}$ or $\mathcal{T} \subset \mathcal{V}$, the bound in (6) follows from Lemma 1. ■

Proof of Proposition 2: Consider first the case $\mathcal{K} = \{1\}$ and $\mathcal{T} = \{1\}$. Since $B = e_1$, it holds

$$\mathcal{C}_T = \frac{1}{2} \log_2\left(1 - \frac{P}{\sigma^2} e_1^\top \mathcal{O}_T e_1\right).$$

Further, since $C = e_n^\top$ and $A = DSD^{-1}$, with $D = \operatorname{diag}(1, \alpha, \ldots, \alpha^{n-1})$ and $\alpha > 0$, we have

$$\begin{aligned} e_1^\top \mathcal{O}_T e_1 &= e_1^\top \left(\int_0^T e^{A^\top t} e_n e_n^\top e^{At} \mathrm{d}t\right) e_1 \\ &= e_1^\top D^{-1} \left(\int_0^T e^{S^\top t} D e_n e_n^\top D e^{St} \mathrm{d}t\right) D^{-1} e_1 \\ &= \alpha^{2(n-1)} e_1^\top \tilde{\mathcal{O}}_T e_1 = \alpha^{2(n-1)} [\tilde{\mathcal{O}}_T]_{11}, \end{aligned}$$

where $\tilde{\mathcal{O}}_T = \int_0^T e^{S^\top t} e_n e_n^\top e^{St}\, \mathrm{d}t$ is the observability Gramian of the pair $(S, e_n^\top)$. For general $\mathcal{K}$, $\mathcal{T}$ satisfying $1 \in \mathcal{K}$, $n \in \mathcal{T}$, the bound in (7) follows from Lemma 1. ■

REFERENCES

[1] A. Guille, H. Hacid, C. Favre, and D. A. Zighed, "Information diffusion in online social networks: A survey," *ACM Sigmod Record*, vol. 42, no. 2, pp. 17–28, 2013.

[2] G. Tkačik and A. M. Walczak, "Information transmission in genetic regulatory networks: A review," *Journal of Physics: Condensed Matter*, vol. 23, no. 15, p. 153102, 2011.

[3] S. B. Laughlin and T. J. Sejnowski, "Communication in neuronal networks," *Science*, vol. 301, no. 5641, pp. 1870–1874, 2003.

[4] A. Avena-Koenigsberger, B. Misic, and O. Sporns, "Communication dynamics in complex brain networks," *Nature Reviews Neuroscience*, vol. 19, no. 1, pp. 17–33, 2018.

[5] P. Dayan and L. F. Abbott, *Theoretical neuroscience*. Cambridge, MA: MIT Press, 2001.

[6] B. K. Murphy and K. D. Miller, "Balanced amplification: A new mechanism of selective amplification of neural activity patterns," *Neuron*, vol. 61, no. 4, pp. 635–648, 2009.

[7] G. Hennequin, T. P. Vogels, and W. Gerstner, "Optimal control of transient dynamics in balanced networks supports generation of complex movements," *Neuron*, vol. 82, no. 6, pp. 1394–1406, 2014.

[8] M. M. Churchland, J. P. Cunningham, M. T. Kaufman, J. D. Foster, P. Nuyujukian, S. I. Ryu, and K. V. Shenoy, "Neural population dynamics during reaching," *Nature*, vol. 487, no. 7405, p. 51, 2012.

[9] T. M. Cover and J. A. Thomas, *Elements of information theory*, 2nd ed. John Wiley & Sons, 2012.

[10] A. G. Dimitrov, A. A. Lazar, and J. D. Victor, "Information theory in neuroscience," *Journal of Computational Neuroscience*, vol. 30, no. 1, pp. 1–5, 2011.

[11] S. Ganguli, D. Huh, and H. Sompolinsky, "Memory traces in dynamical systems," *Proceedings of the National Academy of Sciences*, vol. 105, no. 48, pp. 18 970–18 975, 2008.

[12] T. Akam and D. M. Kullmann, "Oscillations and filtering networks support flexible routing of information," *Neuron*, vol. 67, no. 2, pp. 308–320, 2010.

[13] C. Kirst, M. Timme, and D. Battaglia, "Dynamic information routing in complex networks," *Nature Communications*, vol. 7, p. 11061, 2016.

[14] U. Harush and B. Barzel, "Dynamic patterns of information flow in complex networks," *Nature Communications*, vol. 8, p. 2181, 2017.

[15] G. Baggio, V. Rutten, G. Hennequin, and S. Zampieri, "Information transmission in dynamical networks: The normal network case," in *IEEE Conference on Decision and Control*, 2018, (to appear).

[16] K. J. Reinschke, *Multivariable control: A graph theoretic approach*. Springer-Verlag, 1988.

[17] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 1, pp. 1–182, 2004.

[18] R. Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices," *IEEE Transactions on Information Theory*, vol. 18, no. 6, pp. 725–730, 1972.

[19] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, "Spectral statistics of Erdős–Rényi graphs I: Local semicircle law," *The Annals of Probability*, vol. 41, no. 3B, pp. 2279–2375, 2013.

[20] L. N. Trefethen and M. Embree, *Spectra and pseudospectra: The behavior of nonnormal matrices and operators*. Princeton University Press, 2005.

[21] G. Baggio and S. Zampieri, "On the relation between non-normality and diameter in linear dynamical networks," in *European Control Conference*, 2018, pp. 1839–1844.

[22] S. Zhao and F. Pasqualetti, "Controllability degree of directed line networks: Nodal energy and asymptotic bounds," in *European Control Conference*, 2018, pp. 1857–1862.

[23] H. Robbins, "A remark on Stirling's formula," *The American Mathematical Monthly*, vol. 62, no. 1, pp. 26–29, 1955.

[24] F. Pasqualetti, C. Favaretto, S. Zhao, and S. Zampieri, "Fragility and controllability tradeoff in complex networks," in *American Control Conference*, 2018, pp. 216–221.

[25] M. Asllani, R. Lambiotte, and T. Carletti, "Structure and dynamical behavior of non-normal networks," *Science Advances*, vol. 4, no. 12, p. eaau9403, 2018.

[26] R. A. Horn and C. R. Johnson, *Matrix analysis*, 2nd ed. Cambridge university press, 2012.