

# Volume-Based Lower Bounds to the Capacity of the Gaussian Channel Under Pointwise Additive Input Constraints

Neri Merhav and Shlomo Shamai (Shitz)

The Andrew & Erna Viterbi Faculty of Electrical and Computer Engineering  
Technion - Israel Institute of Technology  
Haifa 3200003, ISRAEL

October 5, 2025

## Abstract

We present a family of relatively simple and unified lower bounds on the capacity of the Gaussian channel under a set of pointwise additive input constraints. Specifically, the admissible channel input vectors  $\mathbf{x} = (x_1, \dots, x_n)$  must satisfy  $k$  additive cost constraints of the form  $\sum_{i=1}^n \phi_j(x_i) \leq n\Gamma_j$ ,  $j = 1, 2, \dots, k$ , which are enforced pointwise for every  $\mathbf{x}$ , rather than merely in expectation. More generally, we also consider cost functions that depend on a sliding window of fixed length  $m$ , namely,  $\sum_{i=m}^n \phi_j(x_i, x_{i-1}, \dots, x_{i-m+1}) \leq n\Gamma_j$ ,  $j = 1, 2, \dots, k$ , a formulation that naturally accommodates correlation constraints as well as a broad range of other constraints of practical relevance.

We propose two classes of lower bounds, derived by two methodologies that both rely on the exact evaluation of the volume exponent associated with the set of input vectors satisfying the given constraints. This evaluation exploits extensions of the method of types to continuous alphabets, the saddle-point method of integration, and basic tools from large deviations theory. The first class of bounds is obtained via the entropy power inequality (EPI), and therefore applies exclusively to continuous-valued inputs. The second class, by contrast, is more general, and it applies to discrete input alphabets as well. It is based on a direct manipulation of mutual information, and it yields stronger and tighter bounds, though at the cost of greater technical complexity. Numerical examples illustrating both types of bounds are provided, and several extensions and refinements are also discussed.

**Index Terms:** Gaussian channel, channel capacity, entropy power inequality, peak-power constraint, volume exponent, saddle-point method, large deviations.

# 1 Introduction

Input-constrained Gaussian channels have been a central topic of study since the earliest days of information theory [24], and they continue to be addressed in numerous papers and textbooks (see, e.g., [3] and references therein, as well as those cited here). The most common constraint is the average power constraint, which reflects the physical power budget of the transmitter and leads to the classical expression for the Gaussian channel capacity [24]. Beyond this standard setting, additional constraints capture further limitations of practical signaling. For instance, the peak-power constraint [26] has yielded fundamental insights, most notably, the result that the capacity-achieving input distribution for the discrete-time memoryless Gaussian channel is discrete with finite support.

In practice, discrete input constellations are employed in virtually all communication systems, regardless of whether they coincide with the true capacity-achieving distribution (which is Gaussian under an average-power-only constraint). Discrete inputs are also used to quantify the performance loss due to practical signaling restrictions, even in cases where Gaussian inputs remain theoretically optimal [19]. From this perspective, the theoretical results of [26] carry clear practical significance. Moreover, constrained Gaussian input models play a central role in the study of optical communication channels, where the key additional constraint is non-negativity of the input signal, imposed by intensity modulation [1, 16, 25]. Similar constrained models have also been considered in other contexts (see, e.g., [11, 21]).

Discrete-time Gaussian channels with filtered inputs, which induce inter-symbol interference (ISI), provide both a natural and practically significant model for communication systems under realistic constraints. These channels have been investigated for many decades, leading to a rich body of results and a variety of bounds, some of which are formulated in terms of equivalent scalar Gaussian channels [15, 23]. Similarly, continuous-time filtered Gaussian channels subject to average- and peak-power constraints have also been extensively studied, motivated by the need to accurately model practical systems (see, e.g., [20] and references therein).

The exact evaluation of capacity under practical constraints, such as peak-power limitations, is notoriously difficult and typically requires intricate numerical procedures. This challenge motivates the study of simple lower and upper bounds on capacity, a line of work

well represented in the literature (see [27] and references therein). Capacity bounds under peak-power and related practical constraints have also been extensively investigated in broader settings, including vector Gaussian channels, multiple-input multiple-output (MIMO) Gaussian channels, and related models [7, 8, 9] and references therein.

As evidenced by the vast literature on discrete-time Gaussian channel models, exact capacity characterizations are rarely attainable. This scarcity strongly motivates the development of relatively simple lower bounds on Gaussian channel capacity under a broad class of constraints, extending beyond the classical average- and peak-power limitations. Examples include, for instance, moment-based constraints studied in [10, 12].

In this work, we present a family of relatively simple and unified lower bounds to the capacity of the Gaussian channel subject to a set of point-wise additive constraints. Specifically, the allowable channel input vectors,  $\mathbf{x} = (x_1, \dots, x_n)$ , must comply with a set of  $k$  additive cost constraints of the form,

$$\sum_{i=1}^n \phi_j(x_i) \leq n\Gamma_j, \quad j = 1, 2, \dots, k, \quad (1)$$

where  $\{\phi_j(\cdot), j = 1, 2, \dots, k\}$  are given cost functions and  $\{\Gamma_j, j = 1, 2, \dots, k\}$  are given numbers. Note that these constraints are imposed point-wise, for every  $\mathbf{x}$ , and not merely in expectation. The most common example is, of course, the average power constraint, corresponding to  $k = 1$  and  $\phi_1(x) = x^2$  and  $\Gamma_1 = P$ . If, in addition, one wishes to add, for example, a peak-power constraint, this can be addressed in this framework, by defining  $k = 2$ ,  $\phi_1(x) = x^2$ , and  $\phi_2(x)$  being defined as  $\phi_2(x) = 0$  for  $|x| \leq A$  and  $\phi_2(x) = \infty$  for  $|x| > A$ . More generally, we can also allow cost functions that depend on a sliding-window of a fixed size,  $m$ , i.e.,

$$\sum_{i=m}^n \phi_j(x_i, x_{i-1}, \dots, x_{i-m+1}) \leq n\Gamma_j, \quad j = 1, 2, \dots, k. \quad (2)$$

These are useful to impose e.g., correlation constraints, where  $\phi_j(x_i, x_{i-\ell}) = x_i x_{i-\ell}$ , as well as a variety of many other practically relevant constraints, as will be discussed in the sequel. We henceforth refer to the constraints of the form (1) as *memoryless constraints*, and to constraints of the form (2) for  $m \geq 2$ , as *constraints with memory*, or as *sliding-window constraints*.

The proposed lower bounds, in their basic forms, depend on  $k$  parameters over which an expression should be minimized. In other words, the number of parameters to be optimized

is equal to the number of constraints. In some of the more sophisticated versions of the proposed lower bounds, there are additional parameters, but these are parameters for maximization, and so, the maximization is not mandatory as an arbitrary choice of the values of those additional parameters are adequate for the purpose of obtaining a valid lower bound.

We propose two classes of lower bounds, which are derived by two methodologies that are based on exact evaluation of the volume exponent associated with the set  $\mathcal{S}_n \subset \mathbb{R}^n$  of input vectors that satisfy the constraints (1), or more generally, (2). This evaluation of the volume exponent is based on extensions of the method of types to continuous alphabets and on saddle-point integration methods [18, Chapters 2 and 3] (see also [4]) as well as elementary large deviations theory [5]. Of course, for a finite channel input alphabet, the ordinary method of types can be used instead. The first class of bounds is based on the entropy-power inequality (EPI) (see, e.g., [3]) and therefore applies to continuous-valued inputs only, and the second class applies also to discrete-alphabet inputs. The second class of bounds, which builds on direct evaluation of the mutual information, is stronger and tighter, but somewhat more complicated. Several extensions and modifications are also discussed.

It is important to emphasize that we do not claim that our bounds are tighter than all bounds reported in the literature for each and every specific model. Our contribution lies in proposing systematic methodologies for deriving good lower bounds in a rather general framework of channel input constraints, including sliding-window constraints (constraints with memory), which are not trivial to handle, in general.

The outline of the remaining part of this article is as follows. In Section 2, we establish our notation conventions (Subsection 2.1), provide a formal description of the setting (Subsection 2.2), and specify our objective (Subsection 2.3). In Section 3, we present the basic EPI lower bound (Subsection 3.1), provide the volume exponent formula, first, for memoryless constraints (Subsection 3.2), demonstrate it in a couple of examples (Subsection 3.3), and finally, extend the scope to constraints with memory (Subsection 3.4) along with some additional examples (Subsection 3.5). In Section 4, we derive alternative lower bounds by direct manipulation of the mutual information where the channel input distribution is set to be uniform across the set of input vectors that comply with all constraints. We do this mostly for memoryless constraints (Subsection 4.1), but we also outline the basics of a possible derivation for constraints with memory (Subsection 4.2). For memoryless constraints,

we also discuss and demonstrate a possible further improvement based on a certain parametric family of non-uniform input distributions. Finally, in Section 5, we summarize and conclude this work, along with an outlook for future work.

## 2 Notation Conventions, Setup, and Objective

### 2.1 Notation Conventions

Throughout this paper, random variables are denoted by capital letters, their realizations – by the corresponding lowercase letters, and their alphabets – by calligraphic letters. Random vectors and their realizations are denoted by boldface capital and lowercase letters, respectively, with their alphabets expressed as Cartesian powers of the underlying single-letter alphabet. For example, the random vector  $\mathbf{X} = (X_1, \dots, X_n)$  may take a realization  $\mathbf{x} = (x_1, \dots, x_n)$  in  $\mathcal{X}^n$ , the  $n$ th power of the single-letter alphabet,  $\mathcal{X}$ . For two positive integers  $i$  and  $j$ , with  $i < j$ , we use the shorthand notation  $x_i^j$  to denote  $(x_i, x_{i+1}, \dots, x_j)$  with the analogous convention for random variables, e.g.,  $X_i^j = (X_i, X_{i+1}, \dots, X_j)$ . When  $i = 1$ , the subscript will be omitted, namely,  $x^j$  and  $X^j$  will stand for  $x_1^j$  and  $X_1^j$ , respectively.

Sources and channels will be denoted by the letters  $p$ ,  $f$ , and  $q$ , subscripted by the names of the relevant random variables or vectors, including conditionings when appropriate, following standard conventions (e.g.,  $q_X$ ,  $p_{Y|X}$ , etc.). When no ambiguity arises, these subscripts will be omitted. The probability of an event  $\mathcal{A}$  will be denoted by  $\Pr\{\mathcal{A}\}$ . The expectation operator with respect to (w.r.t.) a probability distribution  $p$  will be written as  $\mathbf{E}_p\{\cdot\}$ , with the subscript dropped whenever the distribution is clear from context. If the underlying distribution depends on a parameter  $\boldsymbol{\theta}$ , we will denote the expectation by  $\mathbf{E}_{\boldsymbol{\theta}}\{\cdot\}$ . Likewise,  $\Pr_{\boldsymbol{\theta}}\{\mathcal{A}\}$  will denote probability w.r.t. the source parameterized by  $\boldsymbol{\theta}$ . Information measures follow the conventional notation of information theory: for example,  $h(\mathbf{Y})$  is the differential entropy of  $\mathbf{Y}$ ,  $h(\mathbf{Y}|\mathbf{X})$  the conditional differential entropy of  $\mathbf{Y}$  given  $\mathbf{X}$ ,  $I(\mathbf{X};\mathbf{Y})$  is the mutual information between  $\mathbf{X}$  and  $\mathbf{Y}$ , etc. Finally, for a probability function  $q(\mathbf{x})$  (a probability mass function if  $\mathbf{x}$  is discrete, or a probability density function if it is continuous), we denote its support by  $\text{supp}\{q\}$ , that is,  $\text{supp}\{q\} = \{\mathbf{x} : q(\mathbf{x}) > 0\}$ .

For two positive sequences,  $\{a_n\}_{n \geq 1}$  and  $\{b_n\}_{n \geq 1}$ , the notation  $a_n \asymp b_n$  will stand for equality in the exponential scale, that is,  $\lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{a_n}{b_n} = 0$ . Similarly,  $a_n \leq b_n$  means that  $\limsup_{n \rightarrow \infty} \frac{1}{n} \log \frac{a_n}{b_n} \leq 0$ , and so on. The indicator function of an event  $\mathcal{A}$  will be denoted by  $\mathcal{I}\{\mathcal{A}\}$ . The notation  $[x]_+$  will stand for  $\max\{0, x\}$ . Logarithms will be

understood to be taken to the natural base,  $e$ , unless specified otherwise. The Q-function is defined as

$$Q(u) = \frac{1}{\sqrt{2\pi}} \int_u^\infty e^{-x^2/2} dx. \quad (3)$$

## 2.2 Setup

Consider the memoryless Gaussian channel,

$$Y_t = X_t + Z_t, \quad t = 1, 2, \dots \quad (4)$$

where  $\{X_t\}$  is a real-valued channel input signal,  $\{Y_t\}$  is the channel output signal, and where  $\{Z_t\}$  is an i.i.d., zero-mean Gaussian noise process with variance  $\sigma^2$ .

Given a positive integer  $n$ , let us define a set  $\mathcal{S}_n \subset \mathbb{R}^n$  of allowable channel input vectors to be:

$$\mathcal{S}_n = \left\{ \mathbf{x} : \sum_{t=1}^n \phi_j(x_t) \leq n\Gamma_j, \quad j = 1, \dots, k \right\}, \quad (5)$$

where  $k$  is a positive integer,  $\phi_j(\cdot)$  are certain cost constraint functions and  $\Gamma_j$  are given constants,  $j = 1, \dots, k$ . Clearly, equality constraints can be formally incorporated by defining pairs of inequality constraints that differ by their signs, i.e.,

$$\sum_{t=1}^n \phi_j(x_t) \leq n\Gamma_j \quad (6)$$

$$\sum_{t=1}^n [-\phi_j(x_t)] \leq n \cdot [-\Gamma_j]. \quad (7)$$

In some of our derivations and results we will allow more general cost constraint functions, where each  $\phi_j(\cdot)$  (or at least one of them) operates on a sliding-window of  $m$  channel input symbols ( $m$  - positive integer) rather than on a single symbol. In this case, the constraints that define  $\mathcal{S}_n$  would be of form

$$\sum_{t=m}^n \phi_j(x_{t-m+1}^t) \leq n\Gamma_j, \quad j = 1, 2, \dots, k. \quad (8)$$

For the case  $m = 1$ , we refer to the constraints that define  $\mathcal{S}_n$  as *memoryless constraints*, whereas the case  $m \geq 2$ , will be referred to as the case of *constraints with memory* or as *sliding-window constraints*.

The most common example of a memoryless constraint is the average power constraint, where  $k = 1$  and  $\phi_1(x) = x^2$  and  $\Gamma_1 = P$ , the allowed power. Another important example is the case where, in addition to the average power constraint, a peak-power constraint is

imposed, i.e.,  $|x_t| \leq A$  (for a given positive real  $A$ ) for all  $t = 1, 2, \dots, n$ . In this case,  $k = 2$ ,  $\phi_1$  is as above, and the peak-power constraint can be accommodated within our framework by using the infinite square well (ISW) function,

$$\phi_2(x) = w(x) \triangleq \begin{cases} 0 & |x| \leq A \\ \infty & |x| > A \end{cases} \quad (9)$$

and  $\Gamma_2 = 0$ . Useful examples of cost constraint functions with memory are those associated with correlation constraints, e.g.,

$$\sum_{t=\ell+1}^n x_t x_{t-\ell} \leq nR_\ell, \quad (10)$$

where  $\ell$  is a fixed positive integer and  $R_\ell$  is a given constant. Other useful examples could be associated with limitations on the relative frequency of sign changes along the vector  $\mathbf{x}$ , i.e.,

$$\sum_{t=2}^n \mathcal{I}\{\text{sgn}(x_t x_{t-1}) = -1\} \leq n\alpha, \quad (11)$$

or, for example, a peak-power limitation on a filtered version of  $\mathbf{x}$ , i.e.,

$$\sum_{t=m}^n w \left( \sum_{i=0}^{m-1} h_i x_{t-i} \right) \leq 0, \quad (12)$$

where  $\{h_i\}_{i=0}^{m-1}$  is the filter's impulse response. For binary input channels, sliding-window constraints can also be used to limit (from above and/or below) the number of successive repetitions (or run-lengths) of certain channel input symbols (see, e.g., [22]). To this end, one may define the corresponding cost constraint function to be equal to zero for every allowable channel input pattern and to be equal to infinity for every forbidden pattern.

Let us denote  $\mathbf{\Gamma} = (\Gamma_1, \dots, \Gamma_k)$ , and define the channel capacity subject to the given constraints as

$$C(\mathbf{\Gamma}) \triangleq \liminf_{n \rightarrow \infty} \sup \frac{I(\mathbf{X}; \mathbf{Y})}{n}, \quad (13)$$

where the supremum is taken over all input distributions,  $\{q\}$ , with  $\text{supp}\{q\} \subseteq \mathcal{S}_n$ .

### 2.3 Objective

The objective of this work is to propose two general methodologies for obtaining fairly tight lower bounds to  $C(\mathbf{\Gamma})$ .

The first methodology is based on the entropy power inequality (EPI) and is therefore applicable only when the channel input vector,  $\mathbf{X}$ , takes on continuous values within  $\mathcal{S}_n$ .

Since the EPI lower bound to mutual information depends on the input distribution only via its differential entropy,  $h(\mathbf{X})$ , it is obvious that the maximizing distribution for the EPI lower bound is uniform across  $\mathcal{S}_n$ , namely,  $q(\mathbf{x}) = 1/\text{Vol}\{\mathcal{S}_n\}$  for  $\mathbf{x} \in \mathcal{S}_n$  and  $q(\mathbf{x}) = 0$  elsewhere. Consequently, the corresponding lower bound to  $C(\mathbf{T})$  hinges upon our ability to assess the volume of  $\mathcal{S}_n$ , or more precisely, the asymptotic exponential rate of  $\text{Vol}\{\mathcal{S}_n\}$  as a function of  $n$ , namely, the *volume exponent* of  $\{\mathcal{S}_n, n \geq 1\}$ , defined as

$$v(\mathbf{T}) \triangleq \lim_{n \rightarrow \infty} \frac{\log \text{Vol}\{\mathcal{S}_n\}}{n} \quad (14)$$

for the general form of  $\mathcal{S}_n$ , defined by either memoryless constraints or sliding-window constraints. Note that the limit of (14) exists due to super-additivity of the sequence  $\{\log \text{Vol}\{\mathcal{S}_n\}\}_{n \geq 1}$ . To this end, we invoke tools associated with the extended method of types and saddle-point integration [18, Chapters 2 and 3] (see also [4]) as well as elementary results from large deviations theory [5].

The second methodology is based on direct manipulation of the mutual information,  $I(\mathbf{X}; \mathbf{Y})$ , where instead of maximizing over all input distributions supported by  $\mathcal{S}_n$ , we take the input distribution to be uniform across  $\mathcal{S}_n$ , and once again, the resulting lower bounds will depend on the volume exponent of  $\mathcal{S}_n$ . This class of bounds is somewhat more complicated than the EPI bound, but still reasonably simple and easy to calculate at least for memoryless constraints. More importantly, it is tighter and stronger than the EPI bound, as will be demonstrated in numerical examples. It is also applicable for both discrete and continuous channel inputs, unlike the EPI bound which is valid only for continuous inputs. Moreover, it is easy to extend and strengthen this class of bounds by allowing optimization over certain (parametric) classes of non-uniform input distributions across  $\mathcal{S}_n$ . The caveat, however, is that in order to handle constraints with memory, there is a need to further give up on tightness at a certain point, for reasons that will become apparent in the sequel.

### 3 EPI Volume-Based Bounds

#### 3.1 Elementary Background on EPI Lower Bounds

The idea of deriving lower bounds to the channel capacity by invoking the EPI is simple and not quite new, see, e.g., [13], [14], [20], [21], [27]. Nonetheless, for the sake of completeness, we begin this section by presenting it, and then combine it with our method [18] for evaluating the volume exponent,  $v(\mathbf{T})$ , of  $\{\mathcal{S}_n\}$ .

For every channel input distribution,  $q(\cdot)$ , whose support is within  $\mathcal{S}_n$ , consider the following chain of inequalities:

$$\begin{aligned}
\frac{I(\mathbf{X}; \mathbf{Y})}{n} &= \frac{h(\mathbf{Y}) - h(\mathbf{Y}|\mathbf{X})}{n} \\
&= \frac{h(\mathbf{Y})}{n} - \frac{h(\mathbf{Z})}{n} \\
&\geq \frac{1}{2} \log \left[ e^{2h(\mathbf{X})/n} + e^{2h(\mathbf{Z})/n} \right] - \frac{h(\mathbf{Z})}{n} \\
&= \frac{1}{2} \log \left[ e^{2h(\mathbf{X})/n} + 2\pi e \sigma^2 \right] - \frac{1}{2} \log(2\pi e \sigma^2) \\
&= \frac{1}{2} \log \left[ 1 + \frac{e^{2h(\mathbf{X})/n}}{2\pi e \sigma^2} \right],
\end{aligned} \tag{15}$$

and so,

$$\begin{aligned}
C_n(\boldsymbol{\Gamma}) &\stackrel{\triangle}{=} \sup_{\{q: \text{supp}\{q\} \subseteq \mathcal{S}_n\}} \frac{I(\mathbf{X}; \mathbf{Y})}{n} \\
&\geq \sup_{\{q: \text{supp}\{q\} \subseteq \mathcal{S}_n\}} \frac{1}{2} \log \left[ 1 + \frac{e^{2h(\mathbf{X})/n}}{2\pi e \sigma^2} \right] \\
&= \frac{1}{2} \log \left[ 1 + \frac{\exp\{2\log \text{Vol}\{\mathcal{S}_n\}/n\}}{2\pi e \sigma^2} \right].
\end{aligned} \tag{16}$$

Taking the limit inferior of  $n \rightarrow \infty$ , we arrive at

$$C(\boldsymbol{\Gamma}) \geq C_{\text{EPI}}(\boldsymbol{\Gamma}) \stackrel{\triangle}{=} \frac{1}{2} \log \left[ 1 + \frac{\exp\{2v(\boldsymbol{\Gamma})\}}{2\pi e \sigma^2} \right]. \tag{17}$$

### 3.2 Volume Exponents

As mentioned at the very beginning of this section, this generic EPI lower bound is known for some time. For example, if only an average power constraint is imposed (i.e.,  $k = 1$ ,  $\phi_1(x) = x^2$ , and  $\Gamma_1 = P$ ), then  $\mathcal{S}_n$  is an  $n$ -dimensional Euclidean ball of radius  $\sqrt{nP}$ , whose volume is given by  $\text{Vol}\{\mathcal{S}_n\} = \pi^{n/2}(\sqrt{nP})^n/\Gamma(n/2+1) \doteq (2\pi e P)^{n/2}$  (with  $\Gamma(\cdot)$  being the Gamma function, not to be confused with the vector  $\boldsymbol{\Gamma}$  or its components), and so,  $v(P) = \frac{1}{2} \log(2\pi e P)$ , leading to the tight lower bound,

$$\begin{aligned}
C_{\text{EPI}}(P) &= \frac{1}{2} \log \left( 1 + \frac{e^{2v(P)}}{2\pi e \sigma^2} \right) \\
&= \frac{1}{2} \log \left( 1 + \frac{2\pi e P}{2\pi e \sigma^2} \right) \\
&= \frac{1}{2} \log \left( 1 + \frac{P}{\sigma^2} \right) \\
&= C(P).
\end{aligned} \tag{18}$$

However, for the general case, where  $\mathcal{S}_n$  is defined by arbitrary sets of cost constraint functions, the evaluation of  $v(\mathbf{\Gamma})$  seems to be less trivial. This is exactly the point where our contribution in this section takes place, as we invoke the techniques described in Chapters 2 and 3 of [18].

We begin from the case of memoryless constraints, and later discuss the extension to constraints with memory. Let  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ , where  $\theta_i > 0$  for all  $i = 1, 2, \dots, k$ , and define the function

$$Z(\boldsymbol{\theta}) \triangleq \int_{-\infty}^{\infty} \exp \left\{ - \sum_{j=1}^k \theta_j \phi_j(x) \right\} dx, \quad (19)$$

and assume that  $Z(\boldsymbol{\theta}) < \infty$  for some subset  $\Theta$  of  $\mathbb{R}_+^k$ , which is the set of all  $k$ -vectors,  $\boldsymbol{\theta}$ , with strictly positive components. In the discrete case, the integration over  $\mathbb{R}$  is replaced by a summation over  $\mathcal{X}$ , the alphabet of  $x$ . For shorthand notation, in the sequel, we define the vector function  $\boldsymbol{\phi}(x) = (\phi_1(x), \dots, \phi_k(x))$ , and the inner product  $\boldsymbol{\theta} \bullet \boldsymbol{\phi}(x) = \sum_{j=1}^k \theta_j \phi_j(x)$ , so that we can write  $Z(\boldsymbol{\theta}) \triangleq \int_{-\infty}^{\infty} \exp\{-\boldsymbol{\theta} \bullet \boldsymbol{\phi}(x)\} dx$ . We also denote the inner product  $\boldsymbol{\theta} \bullet \mathbf{\Gamma} = \sum_{j=1}^k \theta_j \Gamma_j$ . Let us define

$$\psi(\boldsymbol{\theta}) \triangleq \log Z(\boldsymbol{\theta}), \quad (20)$$

$$\dot{\psi}(\boldsymbol{\theta}) \triangleq \nabla \psi(\boldsymbol{\theta}) = \left( \frac{\partial \psi(\boldsymbol{\theta})}{\partial \theta_1}, \dots, \frac{\partial \psi(\boldsymbol{\theta})}{\partial \theta_k} \right), \quad (21)$$

$$\ddot{\psi}(\boldsymbol{\theta}) \triangleq \nabla^2 \psi(\boldsymbol{\theta}), \quad (22)$$

where  $\nabla^2 \psi(\boldsymbol{\theta})$  is the  $k \times k$  Hessian matrix whose  $(i, j)$ -th element is given by  $\frac{\partial^2 \psi(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}$ ,  $i, j = 1, \dots, k$ . Finally, define

$$\omega(\mathbf{\Gamma}) \triangleq \inf_{\boldsymbol{\theta} \in \mathbb{R}_+^k} \{ \boldsymbol{\theta} \bullet \mathbf{\Gamma} + \psi(\boldsymbol{\theta}) \}. \quad (23)$$

The following lemma, whose proof appears in the Appendix A, establishes the result that under certain regularity conditions, the volume exponent,  $v(\mathbf{\Gamma})$ , is equal to the function  $\omega(\mathbf{\Gamma})$  defined above.

**Lemma 1** *Let  $\boldsymbol{\phi}(\cdot)$  be defined such that  $Z(\boldsymbol{\theta}) < \infty$  for all  $\boldsymbol{\theta} \in \Theta \in \mathbb{R}_+^k$ . Then,*

1.  $v(\mathbf{\Gamma}) \leq \omega(\mathbf{\Gamma})$ .
2. *Assume, in addition, that for the given  $\mathbf{\Gamma}$ , there exists  $\boldsymbol{\theta}$  such that  $\dot{\psi}(\boldsymbol{\theta}) = -\mathbf{\Gamma}$  and that  $\ddot{\psi}(\boldsymbol{\theta})$  is a positive definite matrix with finite diagonal entries. Then,*

$$v(\mathbf{\Gamma}) \geq \omega(\mathbf{\Gamma}), \quad (24)$$

and therefore, following part 1,

$$v(\mathbf{\Gamma}) = \omega(\mathbf{\Gamma}). \quad (25)$$

Since Lemma 1 plays a central role in our derivations, it is useful to pause and discuss both its significance and technical aspects before applying it to obtain capacity bounds.

1. *Maximum-entropy representation.* As discussed in [18, pp. 40-41], it is not difficult to show that an equivalent expression for  $v(\mathbf{\Gamma})$  is given by the maximum-entropy variational representation as the supremum of the differential entropy,  $h(X)$ , of a random variable  $X$  subject to the simultaneous constraints,  $\mathbf{E}\{\phi_j(X)\} \leq \Gamma_j$ ,  $j = 1, 2, \dots, k$ .
2. *Equality constraints.* As mentioned earlier, an equality constraint can be accommodated by a pair of inequality constraints with the same cost function and cost limit, but with opposite signs (see eqs. (6) and (7) above). This amounts to allowing the corresponding parameter  $\theta_j$  to take on any real value, not just positive values, exactly as in constrained optimization using Lagrangians.
3. *Convex optimization.* It is easy to see that  $\psi(\boldsymbol{\theta})$  is a convex function by observing that its Hessian,  $\ddot{\psi}(\boldsymbol{\theta})$ , is non-negative definite due to the fact that it can be viewed as a covariance matrix of the random vector  $\phi(X)$  under  $f_{\boldsymbol{\theta}}$ , the probability density function (pdf) that is proportional to  $e^{-\boldsymbol{\theta} \bullet \phi(x)}$ . Therefore, the minimization associated with the calculation of  $v(\mathbf{\Gamma})$  is a convex program, and hence can be calculated using standard tools of convex programming.
4. *Redundant constraints and their removal.* Part 2 of Lemma 1 assumes that we can find  $\boldsymbol{\theta}$  such that  $\dot{\psi}(\boldsymbol{\theta}) = -\mathbf{\Gamma}$ . It might happen, however, that this condition is violated at certain instances of the problem. This may be the case when either  $\mathcal{S}_n$  is empty (which is the case when the constraints are contradictory), or when there are redundant constraints, namely, inactive constraints, which are superfluous in the presence of other constraints. Such constraints are characterized by holding with strict inequalities, i.e.,  $\sum_{i=1}^n \phi_j(x_i) < \Gamma_i$ . As an example of a redundant constraint, consider the case  $k = 2$ ,  $\phi_1(x) = |x|$ , and  $\phi_2(x) = x^2$ . Since  $(\frac{1}{n} \sum_{i=1}^n |x_i|)^2$  cannot exceed  $\frac{1}{n} \sum_{i=1}^n x_i^2$ , it is obvious that the constraint  $\frac{1}{n} \sum_{i=1}^n |x_i| \leq \Gamma_1$  becomes redundant whenever  $\Gamma_1 > \sqrt{\Gamma_2}$ . In general, a redundant

constraint,  $\sum_{i=1}^n \phi_j(x_i) < \Gamma_j$ , can be formally removed simply by the assigning  $\theta_j = 0$ .

5. *Alternative methods.* In the appendix, we prove Lemma 1 by using standard probabilistic arguments. One alternative technique involves the saddle-point integration method (see [18, Chapter 3] and references therein). In this approach, one first expresses the volume of  $\mathcal{S}_n$  as

$$\text{Vol}\{\mathcal{S}_n\} = \int_{\mathbb{R}^n} d\mathbf{x} \prod_{\ell=1}^k U\left(n\Gamma_\ell - \sum_{i=1}^n \phi_\ell(x_i)\right), \quad (26)$$

where  $U(t)$  is the unit step function. Then, one represents each factor,  $U(n\Gamma_\ell - \sum_{i=1}^n \phi_\ell(x_i))$ , of the integrand as the inverse Laplace transform of  $1/s$ , computed at the point  $n\Gamma_j - \sum_{i=1}^n \phi_j(x_i)$ , i.e.,

$$U\left(n\Gamma_\ell - \sum_{i=1}^n \phi_\ell(x_i)\right) = \lim_{T \rightarrow \infty} \frac{1}{2\pi j} \int_{c-jT}^{c+jT} \frac{ds}{s} \exp\left\{s\left(n\Gamma_\ell - \sum_{i=1}^n \phi_\ell(x_i)\right)\right\}, \quad (27)$$

where  $j \triangleq \sqrt{-1}$  and  $c$  is any positive real. Finally, after substituting (27) into (26) and interchanging the order of integrations, one applies the saddle-point approximation to the resulting integral in the (multivariate) complex plane (see Section 3.4 of [18]).

Another technique, that is sometimes applicable, is inspired by large deviations theory (see, e.g., [5]). Suppose, for example, that one of the constraints that define  $\mathcal{S}_n$  involves the ISW function  $w(\cdot)$  defined in (9), which means that  $\mathcal{S}_n \subseteq [-A, A]^n$ , as will be the case in many of our examples in the sequel. Then, one may imagine an auxiliary random vector  $\mathbf{X} = (X_1, \dots, X_n)$ , uniformly distributed within  $[-A, A]^n$ , and then assess the probability,  $\Pr\{\mathbf{X} \in \mathcal{S}_n\} \equiv \text{Vol}\{\mathcal{S}_n\}/(2A)^n$ , using the Chernoff bound, which is exponentially tight under rather general conditions. Then, the estimated volume of  $\mathcal{S}_n$  would be  $(2A)^n$  times the estimated probability of  $\mathcal{S}_n$ , and so,  $v(\boldsymbol{\Gamma}) = \log(2A) - \mathbf{I}(\boldsymbol{\Gamma})$ ,  $\mathbf{I}(\boldsymbol{\Gamma})$  being the large-deviations rate function of the event  $\mathcal{S}_n$ .

6. *Analogy with statistical physics.* Readers familiar with elements of statistical physics (others may skip this comment without loss of continuity), may recognize the resemblance between the formula,

$$v(\boldsymbol{\Gamma}) = \inf_{\boldsymbol{\theta} \in \mathbb{R}_+^k} \{\boldsymbol{\theta} \bullet \boldsymbol{\Gamma} + \psi(\boldsymbol{\theta})\} \quad (28)$$

and the Legendre-Fenchel relationship between the so called specific entropy of the micro-canonical ensemble, associated with a system of  $n$  particles subjected to the constraints

that define  $\mathcal{S}_n$ , and the corresponding specific free energy,  $-\psi(\boldsymbol{\theta})$ , of the equivalent canonical (or Gibbs) ensemble, whose partition function is  $Z(\boldsymbol{\theta})$ . Each component,  $\theta_j$ , of the parameter vector  $\boldsymbol{\theta}$  has the physical significance of a certain external force (one of them being the inverse temperature) that is conjugate to the corresponding macroscopic quantity,  $\sum_{i=1}^n \phi_j(x_i)$ . This external force is applied to the canonical physical system in order to control the expectation of  $\sum_{i=1}^n \phi_j(x_i)$ , so as to keep it in compliance with the constraints of  $\mathcal{S}_n$  (with high probability for large  $n$ ). For more details regarding these relations, see the discussion in the last paragraph of Section 2.2 (page 44) in [17].

### 3.3 Some Examples

In this subsection, we consider a few simple examples.

**Example 1 - simultaneous average power and peak power constraints.** Let  $k = 2$ ,  $\phi_1(x) = x^2$ ,  $\Gamma_1 = P$ ,  $\phi_2(x) = w(x)$ , and  $\Gamma_2 = 0$ . Then,

$$Z(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} \exp\{-\theta_1 x^2 - \theta_2 w(x)\} dx = \int_{-A}^A e^{-\theta_1 x^2} dx = \sqrt{\frac{\pi}{\theta_1}} \cdot [1 - 2Q(A\sqrt{2\theta_1})]. \quad (29)$$

and so, the volume exponent is

$$\begin{aligned} v(P) &= \inf_{\theta_1 > 0} \left\{ \theta_1 P + \frac{1}{2} \log \frac{\pi}{\theta_1} + \log[1 - 2Q(A\sqrt{2\theta_1})] \right\} \\ &= \inf_{s > 0} \left\{ \frac{s}{2} + \frac{1}{2} \log \left( \frac{2\pi P}{s} \right) + \log \left[ 1 - 2Q \left( \sqrt{\frac{A^2 s}{P}} \right) \right] \right\}, \end{aligned} \quad (30)$$

where in the second line we have changed the optimization variable according to  $s = 2\theta_1 P$ , with the benefit that in the resulting expression of  $v(P)$  the dependence upon  $A^2/P$  appears more explicitly. It follows that the corresponding EPI lower bound is given by

$$C_{\text{EPI}}(P, A) = \frac{1}{2} \log \left[ 1 + \frac{e^{2v(P)}}{2\pi e \sigma^2} \right] \quad (31)$$

$$\begin{aligned} &= \frac{1}{2} \log \left[ 1 + \frac{P}{\sigma^2} \cdot \inf_{s > 0} \left\{ \frac{e^{s-1}}{s} \cdot \left[ 1 - 2Q \left( \sqrt{\frac{A^2 s}{P}} \right) \right]^2 \right\} \right] \\ &\triangleq \frac{1}{2} \log \left[ 1 + \frac{P}{\sigma^2} \cdot \lambda \left( \frac{A^2}{P} \right) \right]. \end{aligned} \quad (32)$$

The factor

$$\lambda \left( \frac{A^2}{P} \right) = \inf_{s > 0} \left\{ \frac{e^{s-1}}{s} \cdot \left[ 1 - 2Q \left( \sqrt{\frac{A^2 s}{P}} \right) \right]^2 \right\}, \quad (33)$$

which clearly depends only on the ratio  $A^2/P$ , can be viewed as the effective factor of loss in signal-to-noise ratio (SNR) due to the peak-power constraint, relative to the ordinary

Gaussian channel with average power constraint only. Indeed, it is easy to see that  $\lambda(u)$  never exceeds unity (by setting  $s = 1$  instead of minimizing over  $s$ ) and that  $\lim_{u \rightarrow \infty} \lambda(u) = 1$ .

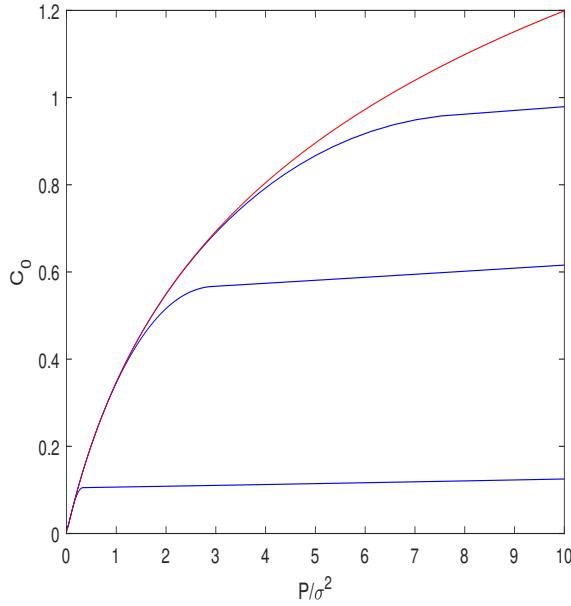


Figure 1:  $C_{\text{EPI}}(P, A)$  as a function of  $P/\sigma^2$  (with  $\sigma^2 = 1$ ) for  $A = 1$ ,  $A = 3$ ,  $A = 5$  (blue curves) and  $A = \infty$  (red curve).

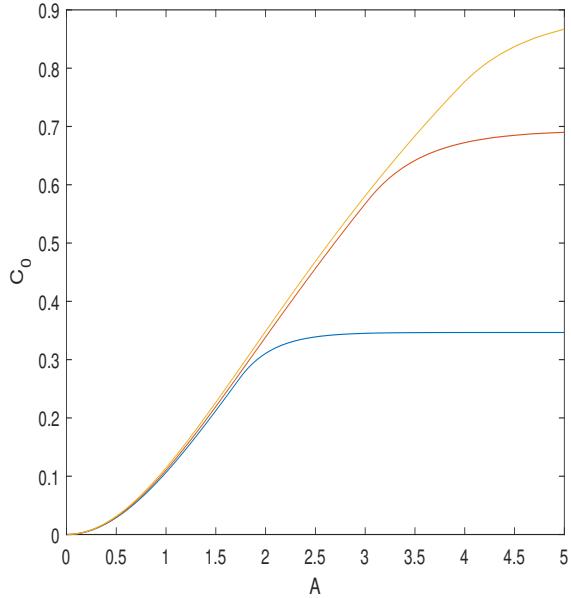


Figure 2:  $C_{\text{EPI}}(P, A)$  as a function of  $A$  for  $\sigma^2 = 1$ ,  $P/\sigma^2 = 1$ ,  $P/\sigma^2 = 3$ , and  $P/\sigma^2 = 5$ .

In Figures 1 and 2, we display plots of  $C_{\text{EPI}}(P, A)$  as functions of  $P$  (in units of  $\sigma^2$ ) for various values of  $A$  and of  $C_{\text{EPI}}(P, A)$  as functions of  $A$  for various values of  $P/\sigma^2$ , respectively. It is interesting to compare the EPI bound with the exact capacity results due to Smith [26] for this model. In Fig. 4 of [26], Smith plots a curve of the exact capacity for  $A = \sqrt{2P}$  as a function of SNR =  $P/\sigma^2$  in dB. Upon reading this curve, one finds that for  $P/\sigma^2 = 10$  dB, the real capacity is in the vicinity of 1.1 nats/channel-use, whereas our EPI bound is 0.8688. Likewise, for  $P/\sigma^2 = 6$  dB, the true capacity is approximately 0.802 nats/channel-use, while the EPI bound is 0.5262. Finally, for  $P/\sigma^2 = 12$  dB, Smith's capacity is nearly 1.412 nats/channel-use, and the EPI bound is 1.0655. It should be stressed that any loss of tightness in the EPI bound is solely due to the EPI, as the volume exponent is exact. Later on we shall suggest several improved lower bounds, which yield results much closer to the exact capacity of Smith's model.

An interesting two-dimensional version of this example concerns the quadrature channel. In this case, suppose that  $n$  is even, divide the components of  $\mathbf{x}$  into  $n/2$  pairs  $(x_{2i-1}, x_{2i})$ ,  $i = 1, 2, \dots, n/2$ , and instead of the peak-power constraint on each  $x_i$ , consider the constraints  $x_{2i-1}^2 + x_{2i}^2 \leq A^2$  for all  $i = 1, 2, \dots, n/2$ . The global average power constraint  $\sum_{i=1}^n x_i^2 \leq nP$  remains intact. In this case, the partition function becomes

$$\begin{aligned}
Z(\theta) &= \int_{\{(x_1, x_2): x_1^2 + x_2^2 \leq A^2\}} e^{-\theta(x_1^2 + x_2^2)} dx_1 dx_2 \\
&= \int_0^{2\pi} \int_0^A e^{-\theta r^2} r dr d\vartheta \\
&= 2\pi \cdot \int_0^A e^{-\theta r^2} r dr \\
&= 2\pi \cdot \int_0^{A^2/2} e^{-2\theta r^2/2} d\left(\frac{r^2}{2}\right) \\
&= 2\pi \cdot \int_0^{A^2/2} e^{-2\theta u} du \\
&= \frac{\pi}{\theta} \cdot (1 - e^{-\theta A^2}). \tag{34}
\end{aligned}$$

The volume exponent is then

$$v(P) = \inf_{\theta \geq 0} \left\{ \theta P + \log\left(\frac{\pi}{\theta}\right) + \log(1 - e^{-\theta A^2}) \right\}, \tag{35}$$

and similarly as above,

$$C_{\text{EPI}}(P, A) = \frac{1}{2} \log \left[ 1 + \frac{P}{\sigma^2} \cdot \inf_{s>0} \frac{e^{s-1}}{s} \cdot \left( 1 - \exp\left\{-\frac{sA^2}{2P}\right\} \right) \right]. \tag{36}$$

Note that whenever  $P \geq A^2/2$ , the infimum is approached for  $s \rightarrow 0$ , which yields  $C_{\text{EPI}}(P, A) = \frac{1}{2} \log(1 + \frac{A^2}{2e\sigma^2}) = \frac{1}{2} \log(1 + \frac{\pi A^2}{2\pi e\sigma^2})$ , independent of  $P$ , as expected, since the average power constraint becomes slack. The “volume” at the numerator,  $\pi A^2$ , is nothing but the area of a circle of radius  $A$ . This concludes Example 1.  $\square$

**Example 2 - absolute value constraint.** In this example, we demonstrate that, in contrast to the true channel capacity function,  $C(\Gamma)$ , the lower bound,  $C_{\text{EPI}}(\Gamma)$ , may not necessarily be a concave function. Consider the case  $k = 1$  with  $\phi_1(x) = |x|$ . In this case,  $Z(\theta) = 2/\theta$ , and so,

$$v(\Gamma) = \inf_{\theta > 0} \left\{ \theta\Gamma + \log\left(\frac{2}{\theta}\right) \right\} = \log(2e\Gamma), \quad (37)$$

and so,

$$C_{\text{EPI}}(\Gamma) = \frac{1}{2} \log\left(1 + \frac{4e^2\Gamma^2}{2\pi e\sigma^2}\right) = \frac{1}{2} \log\left(1 + \frac{2e\Gamma^2}{\pi\sigma^2}\right) = \frac{1}{2} \log\left(1 + 1.7305\frac{\Gamma^2}{\sigma^2}\right), \quad (38)$$

which is obviously not concave, as for small  $\Gamma$  it is nearly quadratic (to the first order approximation):

$$C_{\text{EPI}}(\Gamma) \approx \frac{e}{\pi\sigma^2} \cdot \Gamma^2, \quad \Gamma \ll \sigma. \quad (39)$$

More precisely, while the function  $\log(1 + \alpha x^2)$  ( $\alpha > 0$  being a parameter) is concave in  $x > 0$  across the range  $x \geq 1/\sqrt{\alpha}$ , it is actually convex elsewhere. Therefore the lower bound can be tightened by applying the upper concave envelope (UCE) operator, namely,

$$\bar{C}_{\text{EPI}}(\Gamma) = \text{UCE}\{C_{\text{EPI}}(\Gamma)\} \stackrel{\Delta}{=} \sup \left\{ \sum_{i=1}^{k+1} \alpha_i C_{\text{EPI}}(\Gamma_i) \right\}, \quad (40)$$

where the supremum is over all assignments of  $(\Gamma_1, \dots, \Gamma_{k+1})$  ( $\Gamma_i$  being a  $k$ -vector for all  $i = 1, 2, \dots, k+1$ ) and vectors  $\alpha = (\alpha_1, \dots, \alpha_{k+1})$  with non-negative components such that  $\sum_{i=1}^{k+1} \alpha_i = 1$  and  $\sum_{i=1}^{k+1} \alpha_i \Gamma_i = \Gamma$ . The UCE,  $\bar{C}_{\text{EPI}}(\Gamma)$ , can be achieved by time-sharing. In this example, the UCE is obtained by replacing  $C_{\text{EPI}}(\Gamma)$  for small  $\Gamma$  by a linear function, starting at the origin and ending at the point  $(\Gamma_*, C_{\text{EPI}}(\Gamma_*))$ , where its corresponding straight line is tangential to the curve of  $C_{\text{EPI}}(\Gamma)$ , namely, the point pertaining to the non-zero solution to the equation  $\Gamma C'_{\text{EPI}}(\Gamma) = C_{\text{EPI}}(\Gamma)$ ,  $C'_{\text{EPI}}(\Gamma)$  being the derivative of  $C_{\text{EPI}}(\Gamma)$ . The result is

$$\bar{C}_{\text{EPI}}(\Gamma) = \begin{cases} 0.5293 \cdot \frac{\Gamma}{\sigma} & \Gamma \leq 1.5054\sigma \\ \frac{1}{2} \log\left(1 + 1.7305 \cdot \frac{\Gamma^2}{\sigma^2}\right) & \Gamma \geq 1.5054\sigma \end{cases} \quad (41)$$

### 3.4 Constraints with Memory

The EPI lower bound,  $C_{\text{EPI}}$  (as well as its UCE, if applicable), can also be extended to handle constraints with memory, or, sliding-window constraints of the form (2), such as (10), (11), (12) as well as many others, where each constraint function,  $\phi_j$ , manifests a certain limitation on the local behavior of the channel input signal, for example, no more than  $r$  sign changes within each sliding window of length  $m$  ( $r < m$ ), or any other reasonable criterion concerning the signal variability in the time domain.

First and foremost, we need an extended version of Lemma 1 to the case where at least one of the constraint functions has memory  $m \geq 2$ . Similarly as in Subsection 3.2, it is useful to define a parametric exponential family of densities,  $f_{\boldsymbol{\theta}}(\cdot)$ , except that here these densities would no longer be i.i.d., but densities derived from a Markov process of order  $m - 1$ , defined by

$$f_{\boldsymbol{\theta}}(\mathbf{x}) = \frac{\exp\left\{-\boldsymbol{\theta} \bullet \sum_{i=m}^n \phi(x_{i-m+1}^i)\right\}}{Z_n(\boldsymbol{\theta})} = \frac{\prod_{i=m}^n \exp\left\{-\boldsymbol{\theta} \bullet \phi(x_{i-m+1}^i)\right\}}{Z_n(\boldsymbol{\theta})}, \quad (42)$$

where

$$Z_n(\boldsymbol{\theta}) \stackrel{\triangle}{=} \int_{\mathbb{R}^n} \exp\left\{-\boldsymbol{\theta} \bullet \sum_{i=m}^n \phi(x_{i-m+1}^i)\right\} d\mathbf{x}. \quad (43)$$

As before, assume that  $Z_n(\boldsymbol{\theta}) < \infty$  for every  $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}_+^k$  and every positive integer  $n$ . Assume further that the limit,  $\lim_{n \rightarrow \infty} \frac{\log Z_n(\boldsymbol{\theta})}{n}$ , exists and extend the definition of the function  $\psi(\boldsymbol{\theta})$  to be

$$\psi(\boldsymbol{\theta}) \stackrel{\triangle}{=} \lim_{n \rightarrow \infty} \frac{\log Z_n(\boldsymbol{\theta})}{n}. \quad (44)$$

Now, part 1 of Lemma 1 extends straightforwardly under the new definition of  $\psi(\boldsymbol{\theta})$ . For part 2 of Lemma 1 to extend as well, we need to assume that: (i)  $\psi(\boldsymbol{\theta})$  is twice differentiable, (ii) for the given  $\boldsymbol{\Gamma}$ , there exists  $\boldsymbol{\theta}$  such that  $\dot{\psi}(\boldsymbol{\theta}) \stackrel{\triangle}{=} \nabla \psi(\boldsymbol{\theta}) = -(\boldsymbol{\Gamma} - \boldsymbol{\epsilon})$ , for every sufficiently small  $\boldsymbol{\epsilon} > 0$ , and (iii) the underlying Markov process corresponding to  $f_{\boldsymbol{\theta}}$  is (asymptotically) stationary and ergodic, so that by the ergodic theorem, for every sufficiently small  $\boldsymbol{\epsilon} > 0$ ,

$$\lim_{n \rightarrow \infty} \Pr_{\boldsymbol{\theta}} \bigcap_{j=1}^k \left\{ n(\Gamma_j - 2\epsilon) \leq \sum_{i=1}^n \phi_j(X_i) \leq n\Gamma_j \right\} = 1, \quad (45)$$

where  $\Pr_{\boldsymbol{\theta}}\{\cdot\}$  denotes probability under  $f_{\boldsymbol{\theta}}(\cdot)$ ,  $\boldsymbol{\theta}$  being the point where  $\dot{\psi}(\boldsymbol{\theta}) = -(\boldsymbol{\Gamma} - \boldsymbol{\epsilon})$ . In this case, the proof of the second part of Lemma 1 readily generalizes to allow cost functions with memory, except that instead of using the central limit theorem, we use (45).

The main challenge here is to how calculate  $\psi(\boldsymbol{\theta})$ , which is required for the calculation of the volume exponent,  $v(\boldsymbol{\Gamma})$ , according to (28), but under the new definition of  $\psi(\boldsymbol{\theta})$ . Clearly, in the simple special case, where there are only autocorrelation constraints, such as in (10), without an additional peak-amplitude constraint, the calculation of  $Z_n(\boldsymbol{\theta})$  involves a multivariate Gaussian integral (with correlations) and  $C_{\text{EPI}} = \frac{1}{2} \log(1 + \frac{\sigma_u^2}{\sigma^2})$ , where  $\sigma_u^2$  is the innovation variance of an autoregressive process whose of order  $m - 1$  whose first  $m$  autocorrelations are  $R_0, \dots, R_{\ell-1}$  (see [18], Section 2.6). However, the general case of cost functions with memory is more involved.

We now present two general methods for calculating  $\psi(\boldsymbol{\theta})$ .

1. *Integral operators and their spectral radius.* According to this approach, we observe that the calculation of  $Z_n(\boldsymbol{\theta})$  involves an assessment of the exponential order of a multidimensional integral of the form:

$$I_n = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i=m}^n K(x_{i-m+1}^i) d\mathbf{x}, \quad (46)$$

where each factor of the integrand is the same kernel function  $K(x_{i-m+1}^i) = \exp\{-\boldsymbol{\theta} \bullet \phi(x_{i-m+1}^i)\}$  applied to a sliding window of  $m$  variables, and where  $m$  remains fixed as  $n \rightarrow \infty$ . The calculation of  $I_n$  can be viewed as a succession of iterated applications of a sliding-window integral operator of the form

$$(\mathcal{L}g)(x_2^m) = \int_{-\infty}^{\infty} K(x^m) \cdot g(x^{m-1}) dx_1. \quad (47)$$

Accordingly, under mild regularity conditions, one may invoke the Collatz-Wielandt formulas [2], [29] in order to assess the spectral radius of the operator  $\mathcal{L}$ , which coincides with  $e^{\psi(\boldsymbol{\theta})}$ . These formulas are:

$$e^{\psi(\boldsymbol{\theta})} = \inf_g \sup_{x^{m-1}} \frac{(\mathcal{L}g)(x^{m-1})}{g(x^{m-1})} = \sup_g \inf_{x^{m-1}} \frac{(\mathcal{L}g)(x^{m-1})}{g(x^{m-1})}. \quad (48)$$

These expressions can be viewed as continuous-alphabet counterparts of similar formulas for the Perron-Frobenius eigenvalue for finite dimensional positive matrices in the finite-alphabet case. The second formula is somewhat more appealing in the sense that for the purpose of obtaining a lower bound, it is legitimate to select an arbitrary function  $g$  rather than taking the supremum. Moreover, we have the freedom to choose a parametric family of functions, say  $\{g_\alpha\}$ , which are convenient to work with (like multivariate Gaussian functions), and maximize only w.r.t. the parameter  $\alpha$ . A similar comment applies w.r.t.  $x^{m-1}$ .

in the first formula, but optimization over a finite dimensional vector is less problematic than optimization over a function, which involves, in general, calculus of variations.

It should be noted that in the case  $m = 2$ , if  $K(\cdot, \cdot)$  is a symmetric kernel, that is,  $K(x, x') = K(x', x)$  for all  $x$  and  $x'$ , then an alternative formula for the spectral radius is the Rayleigh quotient formula,

$$e^{\psi(\boldsymbol{\theta})} = \sup_{u(\cdot)} \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x)K(x, x')u(x')dx dx'}{\int_{-\infty}^{\infty} u^2(x)dx} \quad (49)$$

$$= \sup_{\{u(\cdot): \|u\|^2=1\}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x)K(x, x')u(x')dx dx'. \quad (50)$$

Once again, for the purpose of obtaining a valid lower bound, it is not mandatory to calculate the supremum, and it is legitimate to pick an arbitrary  $u(\cdot)$  or to maximize within a parametric family,  $\{u_{\alpha}(\cdot)\}$ .

*2. The Donsker-Varadhan variational formula.* Another possible characterization of the spectral radius,  $e^{\psi(\boldsymbol{\theta})}$ , is via the Donsker-Varadhan formula. Consider the Markov process of order  $m - 1$ , whose transition density is given by

$$f_{\boldsymbol{\theta}}(x_m|x^{m-1}) = \frac{K(x^m)}{\int_{-\infty}^{\infty} K(x^m)dx_m} \triangleq \frac{K(x^m)}{\exp\{S(x^{m-1})\}}. \quad (51)$$

Then, neglecting edge effects for  $n \gg m$ ,

$$\begin{aligned} Z_n(\theta) &= \int_{\mathbb{R}^n} \prod_{i=m}^n K(x_{i-m+1}^i) d\mathbf{x} \\ &= \int_{\mathbb{R}^n} \prod_{i=m}^n f_{\boldsymbol{\theta}}(x_i|x_{i-m+1}^{i-1}) \exp\{S(x_{i-m+1}^{i-1})\} d\mathbf{x} \\ &= \mathbf{E}_{\boldsymbol{\theta}} \left\{ \exp \left[ \sum_{i=m}^n S(x_{i-m+1}^{i-1}) \right] \right\} \\ &\stackrel{(a)}{=} \exp \left\{ n \sup_g \mathbf{E}_g [S(X^{m-1}) - D(g(\cdot|X^{m-1}) \| f_{\boldsymbol{\theta}}(\cdot|X^{m-1}))] \right\} \\ &\stackrel{(b)}{=} \exp \left[ n \sup_g (\mathbf{E}_g \{\log K(X^m)\} + h_g(X_m|X^{m-1})) \right] \\ &= \exp \left\{ n \sup_g [-\boldsymbol{\theta} \bullet \mathbf{E}_g \{\phi(X^m)\} + h_g(X_m|X^{m-1})] \right\}, \end{aligned} \quad (52)$$

where (a) is due to the Donsker-Varadhan variational formula [6], (b) is by the definition of  $S(\cdot)$ , and  $h_g(X_m|X^{m-1})$  is the conditional differential entropy of  $X_m$  given  $X^{m-1}$  under  $g$ . Consequently,

$$\psi(\boldsymbol{\theta}) = \sup_g [-\boldsymbol{\theta} \bullet \mathbf{E}_g \{\phi(X^m)\} + h_g(X_m|X^m)], \quad (53)$$

where for a given auxiliary  $(m-1)$ -th order Markov process  $g$ , the expectation,  $\mathbf{E}_g\{\cdot\}$ , is w.r.t. the stationary state distribution associated with  $g$ . Once again, for the purpose of obtaining a lower bound to  $\psi(\boldsymbol{\theta})$ , one may pick a particular  $g$  or maximize within a parametric family,  $\{g_\alpha\}$  to facilitate the calculation at the possible expense of losing tightness. For example, consider the conditional pdf,

$$g^*(x_m|x^{m-1}) = \frac{\exp\{-\boldsymbol{\theta} \bullet \boldsymbol{\phi}(x^m)\}}{Z(\boldsymbol{\theta}, x^{m-1})}, \quad (54)$$

where

$$Z(\boldsymbol{\theta}, x^{m-1}) = \int_{-\infty}^{\infty} \exp\{-\boldsymbol{\theta} \bullet \boldsymbol{\phi}(x^m)\} dx_m. \quad (55)$$

Then,

$$\psi(\boldsymbol{\theta}) \geq \mathbf{E} \log Z(\boldsymbol{\theta}, X^{m-1}), \quad (56)$$

where the expectation is w.r.t. the stationary distribution associated with the Markov process defined by  $g^*$ . The reason that this is just a lower bound is that  $g^*$  may not be the maximizer as its derivation does not take into account the complicated dependence of the stationary distribution upon the conditional densities of the underlying Markov process.

### 3.5 More Examples

We now provide a few examples for the case of cost functions with memory.

#### Example 3 – Peak-power limitation combined with a correlation constraint.

Consider the case of peak-power limitation combined with a specified one-lag empirical autocorrelation  $\sum_{t=2}^n x_t x_{t-1} = nR_1$ . In this case,  $m = 2$  and the partition function pertaining to the volume exponent is given by

$$\begin{aligned} Z_n(\theta) &= \int_{\mathbb{R}^n} \exp[-\phi_1(x_1)] \prod_{t=2}^n \exp\{-w(x_t) - \theta x_t x_{t-1}\} d\mathbf{x} \\ &= \int_{[-A, A]^n} \prod_{t=2}^n e^{-\theta x_t x_{t-1}} d\mathbf{x}, \end{aligned} \quad (57)$$

The exponential growth rate of  $Z_n(\theta)$  is according to the largest eigenvalue of the kernel  $K(x, x') = e^{-\theta x x'}$  defined on the square  $(x, x') \in [-A, A]^2$ . Since the kernel  $K(x, x')$  is symmetric, the largest eigenvalue can be calculated using the Rayleigh quotient formula,

$$e^{\psi(\boldsymbol{\theta})} = \sup_{\{u: \|u\|^2=1\}} \int_{-A}^A \int_{-A}^A u(x) e^{-\theta x x'} u(x') dx dx'. \quad (58)$$

To estimate  $e^{\psi(\theta)}$ , we evaluate the Rayleigh quotient for the constant test function  $u(x) = \frac{1}{\sqrt{2A}}$ ,  $x \in [-A, A]$ , which is normalized in  $L^2([-A, A])$ . Thus,

$$e^{\psi(\theta)} \geq \int_{-A}^A \int_{-A}^A \frac{1}{\sqrt{2A}} e^{-\theta xx'} \frac{1}{\sqrt{2A}} dx dx' = \frac{1}{2A} \int_{-A}^A \int_{-A}^A e^{-\theta xx'} dx dx'. \quad (59)$$

To evaluate the double integral

$$I = \int_{-A}^A \int_{-A}^A e^{-\theta xx'} dx dx', \quad (60)$$

we can simplify it to a single integral as

$$\int_{-A}^A \int_{-A}^A e^{-\theta xx'} dx dx' = \frac{4}{|\theta|} \int_0^{A^2|\theta|} \frac{\sinh u}{u} du, \quad (61)$$

as can be shown by carrying out explicitly one of the two integrations of the exponential function, and then changing the other integration variable. Substituting into our earlier inequality, we find:

$$e^{\psi(\theta)} \geq \frac{1}{2A} \cdot \frac{4}{|\theta|} \int_0^{A^2|\theta|} \frac{\sinh u}{u} du = \frac{2}{A|\theta|} \int_0^{A^2|\theta|} \frac{\sinh u}{u} du. \quad (62)$$

We conclude that

$$e^{\psi(\theta)} \geq \frac{2}{A|\theta|} \cdot \text{Shi}(A^2|\theta|), \quad (63)$$

where  $\text{Shi}(z) = \int_0^z \frac{\sinh u}{u} du$  is the hyperbolic sine integral function. This approximation is expected to be tight due to the positivity and symmetry of the kernel, and the choice of the constant function as a good candidate for the leading eigenfunction. The resulting lower bound to the capacity is then given by

$$C_{\text{EPI}} = \frac{1}{2} \log \left[ 1 + \frac{\inf_{\theta \in \mathbb{R}} e^{2R_1\theta} \frac{2}{A|\theta|} \cdot \text{Shi}(A^2|\theta|)}{2\pi e \sigma^2} \right]. \quad (64)$$

More generally, we may bound the spectral radius of  $K$  from below by considering the parametric family

$$u_\alpha(x) = \sqrt{\frac{\alpha}{\sinh(2\alpha A)}} \cdot e^{-\alpha x}, \quad |x| \leq A, \quad (65)$$

and then

$$e^{\psi(\theta)} \geq \sup_{\alpha \in \mathbb{R}} \frac{2\alpha}{\sinh(2\alpha A)} \int_{-A}^A \frac{\sinh(\theta x + \alpha)}{\theta x + \alpha} dx. \quad (66)$$

The above example can easily be extended to accommodate also an additional average power constraint (thus extending Example 1). In this case, the kernel becomes  $K(x, \hat{x}) =$

$\exp\{-\theta_1(x^2 + \hat{x}^2)/2 - \theta_2 x \hat{x}\}$ , which is still symmetric. This concludes Example 3.  $\square$

**Example 4 - Average power constraint combined with a correlation constraint,**  
 Consider the case  $m = k = 2$  with  $\phi_1(x_1, x_2) = \frac{x_1^2 + x_2^2}{2}$ ,  $\Gamma_1 = P$ ,  $\phi_2(x_1, x_2) = x_1 x_2$ , and  $\Gamma_2 = \rho P$ , where  $|\rho| < 1$ . In this case, denoting  $s = \theta_2/\theta_1$ , we have

$$Z(\boldsymbol{\theta}.x_1) = \int_{-\infty}^{\infty} \exp\{-\theta_1(x_1^2 + x_2^2)/2 - \theta_2 x_1 x_2\} = \sqrt{\frac{2\pi}{\theta_1}} \cdot \exp\{-\theta_1(1 - s^2)x_1^2/2\}. \quad (67)$$

Taking the natural logarithm, then the expectation, and finally exponentiating again, we obtain:

$$\exp[\mathbf{E}\{\log Z(\boldsymbol{\theta}, X_1)\}] \geq \sqrt{\frac{2\pi}{\theta_1}} \cdot \exp\{-\theta_1(1 - s^2)P/2\}, \quad (68)$$

where we have used the fact that  $\mathbf{E}_g\{X_1^2\} \leq P$  (combined with the conditions  $\theta_1 \geq 0$  and  $s^2 \leq 1$ ), since  $g$  must satisfy the moment constraints, as can be deduced from the equivalent maximin problem of  $\sup_g \inf_{\theta}\{\cdot\}$ . Thus, by standard optimization methods,

$$\lim_{n \rightarrow \infty} [\text{Vol}\{\mathcal{S}_n\}]^{2/n} \geq \inf_{\theta_1 > 0, s} \frac{2\pi}{\theta_1} \cdot \exp\{-\theta_1(1 - s^2)P\} \cdot \exp\{2\theta_1 P + 2\theta_1 s \rho P\} = 2\pi e P (1 - \rho^2), \quad (69)$$

and so,  $C_{\text{EPI}} \geq \frac{1}{2} \log \left[ 1 + \frac{P(1 - \rho^2)}{\sigma^2} \right]$ , which is in fact, the exact capacity under these constraints.  $\square$

**Example 5 - Peak amplitude limitation at the output of a linear system.** Consider the case where the transmitter includes a linear filter just before the antenna, and then the peak-amplitude limitation applies to the filter output, namely, one of the constraints is (12). In this case the body defined by the constraints in the channel input space is linearly transformed into an image body that resides in the filter output signal space, and so the uniform distribution within the input body is transformed into a uniform distribution across the image body. The volume of the image body is given by the volume of the input body, multiplied by the Jacobian of the transformation matrix. For a causal filter, this Jacobian is given by  $|h_0|^n$ . Consequently, since the body at the filter output space is the hypercube  $[-A, A]^n$ , whose volume is  $(2A)^n$ , then the volume of its inverse image, at the filter input space is  $(2A)^n / |h_0|^n = (2A / |h_0|)^n$ . Note that, if this causal filter is also minimum phase, then an alternative expression for the Jacobian, in the frequency domain, is given by

$$\exp \left\{ \frac{n}{2\pi} \int_{-\pi}^{\pi} \log |H(e^{j\omega})| d\omega \right\}. \quad (70)$$

## 4 Direct Manipulation of the Mutual Information

The EPI lower bounds are relatively simple and easy to calculate, provided that the calculation of the volume exponent is reasonably easy. Lemma 1 proposes a simple formula for the volume exponent,  $v(\mathbf{T})$ , which once computed, it can be simply substituted into the expression  $\frac{1}{2} \log \left( 1 + \frac{e^{2v(\mathbf{T})}}{2\pi e \sigma^2} \right)$  and the tightness of this bound depends solely on the tightness of the EPI. However, the EPI is not always tight, especially not in the range of low and moderate SNR. Furthermore, one of the severe limitations of the EPI is that it applies merely to continuous-valued input vectors, and not to discrete ones.

In this section, we derive alternative families of lower bounds, which are not based on the EPI, but rather on direct manipulation of the mutual information,  $I(\mathbf{X}; \mathbf{Y})$ . The calculations of these bounds are somewhat more involved than that of the EPI bound, but it has the following advantages compared to the EPI bound: (i) it is typically tighter, (ii) it applies to both continuous and discrete channel inputs (with integrations over the channel input space simply being replaced by summations), and (iii) it is easy to extend to Gaussian channels with memory as well as to memoryless non-Gaussian channels.

### 4.1 Memoryless Constraints

As in Section 3, we start from the case of memoryless cost functions. Consider the following analysis of the mutual information, where instead of maximizing the channel input pdf over the support  $\mathcal{S}_n$ , we set the uniform input pdf across  $\mathcal{S}_n$ .

$$I(\mathbf{X}; \mathbf{Y}) = h(\mathbf{Y}) - \frac{n}{2} \log(2\pi e \sigma^2) \quad (71)$$

$$= -\mathbf{E} \left\{ \log \left[ \frac{1}{\text{Vol}(\mathcal{S}_n)} \int_{\mathcal{S}_n} p(\mathbf{Y}|\mathbf{x}) d\mathbf{x} \right] \right\} - \frac{n}{2} \log(2\pi e \sigma^2) \quad (72)$$

$$= -\mathbf{E} \left\{ \log \left[ \int_{\mathcal{S}_n} p(\mathbf{Y}|\mathbf{x}) d\mathbf{x} \right] \right\} + \log \text{Vol}(\mathcal{S}_n) - \frac{n}{2} \log(2\pi e \sigma^2). \quad (73)$$

Now, applying the Chernoff bounding technique, we have

$$\begin{aligned}
& \mathbf{E} \left\{ \log \left[ \int_{\mathcal{S}_n} p(\mathbf{Y} | \mathbf{x}) d\mathbf{x} \right] \right\} \\
& \leq \inf_{\boldsymbol{\theta} \in \mathbb{R}_+^k} \mathbf{E} \left\{ \log \left[ \int_{\mathbb{R}^n} \exp \left\{ n[\boldsymbol{\theta} \bullet \boldsymbol{\Gamma} - \boldsymbol{\theta} \bullet \sum_{i=1}^n \boldsymbol{\phi}(x_i)] \right\} p(\mathbf{Y} | \mathbf{x}) d\mathbf{x} \right] \right\} \\
& = \inf_{\boldsymbol{\theta} \in \mathbb{R}_+^k} \mathbf{E} \left\{ \log \left[ \int_{\mathbb{R}^n} \exp \left\{ n[\boldsymbol{\theta} \bullet \boldsymbol{\Gamma} - \boldsymbol{\theta} \bullet \sum_{i=1}^n \boldsymbol{\phi}(x_i)] \right\} \frac{\exp \{-\|\mathbf{Y} - \mathbf{x}\|^2/(2\sigma^2)\}}{(2\pi\sigma^2)^{n/2}} d\mathbf{x} \right] \right\} \\
& = \inf_{\boldsymbol{\theta} \in \mathbb{R}_+^k} \left[ n\boldsymbol{\theta} \bullet \boldsymbol{\Gamma} + \mathbf{E} \left\{ \log \left( \prod_{i=1}^n \left[ \int_{\mathbb{R}} \frac{\exp \{-\boldsymbol{\theta} \bullet \boldsymbol{\phi}(x) - (Y_i - x)^2/(2\sigma^2)\}}{\sqrt{2\pi\sigma^2}} dx \right] \right) \right\} \right] \\
& \triangleq \inf_{\boldsymbol{\theta} \in \mathbb{R}^k} \left[ n\boldsymbol{\theta} \bullet \boldsymbol{\Gamma} + \mathbf{E} \left\{ \log \left( \prod_{i=1}^n \zeta_{\boldsymbol{\theta}}(Y_i) \right) \right\} \right] \\
& = n \cdot \inf_{\boldsymbol{\theta} \in \mathbb{R}_+^k} (\boldsymbol{\theta} \bullet \boldsymbol{\Gamma} + \mathbf{E} \{ \log \zeta_{\boldsymbol{\theta}}(Y) \}), \tag{74}
\end{aligned}$$

and so,

$$\begin{aligned}
I(\mathbf{X}; \mathbf{Y}) & \geq \log \text{Vol}(\mathcal{S}_n) - \frac{n}{2} \log(2\pi e \sigma^2) - n \cdot \inf_{\boldsymbol{\theta} \in \mathbb{R}^K} (\boldsymbol{\theta} \bullet \boldsymbol{\Gamma} + \mathbf{E} \{ \log \zeta_{\boldsymbol{\theta}}(Y) \}) \\
& = \frac{n}{2} \log \frac{P}{\sigma^2} - n \cdot \inf_{\boldsymbol{\theta} \in \mathbb{R}^+} (\boldsymbol{\theta} \bullet \boldsymbol{\Gamma} + \mathbf{E} \{ \log \zeta_{\boldsymbol{\theta}}(Y) \}). \tag{75}
\end{aligned}$$

We have therefore arrived at the following lower bound:

$$C(\boldsymbol{\Gamma}) \geq v(\boldsymbol{\Gamma}) - \frac{1}{2} \log(2\pi e \sigma^2) - \inf_{\boldsymbol{\theta} \in \mathbb{R}_+^k} (\boldsymbol{\theta} \bullet \boldsymbol{\Gamma} + \mathbf{E} \{ \log \zeta_{\boldsymbol{\theta}}(Y) \}), \tag{76}$$

where

$$\zeta_{\boldsymbol{\theta}}(y) = \int_{\mathbb{R}} \frac{\exp \{-\boldsymbol{\theta} \bullet \boldsymbol{\phi}(x) - (y - x)^2/(2\sigma^2)\}}{\sqrt{2\pi\sigma^2}} dx. \tag{77}$$

Note that the second term of the lower bound to  $C(\boldsymbol{\Gamma})$  requires knowledge of the asymptotic marginal pdf of a single channel output symbol,  $Y$ , which is given by the convolution between the pdf of a single noise variable, namely,  $\mathcal{N}(0, \sigma^2)$ , and the marginal of a single component of the vector  $\mathbf{X}$ , induced from the uniform pdf of  $\mathbf{X}$  across  $\mathcal{S}_n$  (for  $n \rightarrow \infty$ ). We will address this issue in the sequel through examples.

**Example 6 - Example 1 revisited.** Consider again the case where there is both an average power constraint,  $P$ , and a peak-power constraint,  $|x_i| \leq A$ ,  $i = 1, 2, \dots, n$ . As we showed in Example 1,

$$v(P) = \frac{1}{2} \log \left[ 2\pi e P \cdot \inf_{s>0} \left\{ \frac{e^{s-1}}{s} \left[ 1 - 2Q \left( s\sqrt{\frac{A^2}{P}} \right) \right]^2 \right\} \right] = \frac{1}{2} \log \left[ 2\pi e P \cdot \lambda \left( \frac{A^2}{P} \right) \right]. \tag{78}$$

Now, after a straightforward algebraic manipulation, we obtain

$$\begin{aligned}
\log \zeta_\theta(y) &= \log \left( \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-A}^A dx \exp \left\{ -\theta x^2 - \frac{(y-x)^2}{2\sigma^2} \right\} \right) \\
&= -\frac{1}{2} \log(1+2\theta\sigma^2) - \frac{\theta(\sigma_X^2 + \sigma^2)}{1+2\theta\sigma^2} + \\
&\quad \log \left[ 1 - Q \left( \frac{A-m(y)}{s} \right) - Q \left( \frac{A+m(y)}{s} \right) \right], \tag{79}
\end{aligned}$$

where

$$s = \frac{\sigma}{\sqrt{1+2\theta\sigma^2}} \tag{80}$$

$$m(y) = \frac{y}{1+2\theta\sigma^2} \tag{81}$$

$$\sigma_X^2 = \frac{\int_{-A}^A x^2 e^{-x^2/(2P)} dx}{\int_{-A}^A e^{-x^2/(2P)} dx}. \tag{82}$$

Thus,

$$\begin{aligned}
C(\mathbf{T}) &\geq \frac{1}{2} \log \left[ \frac{P}{\sigma^2} \cdot \lambda \left( \frac{A^2}{P} \right) \right] - \\
&\quad \inf_{\theta>0} \left\{ \theta P - \frac{1}{2} \log(1+2\theta\sigma^2) - \frac{\theta(\sigma_X^2 + \sigma^2)}{1+2\theta\sigma^2} + \right. \\
&\quad \left. \mathbf{E} \log \left[ 1 - Q \left( \frac{A-m(Y)}{s} \right) - Q \left( \frac{A+m(Y)}{s} \right) \right] \right\}, \tag{83}
\end{aligned}$$

and the expectation is over the randomness of a single symbol,  $Y$ , given by the convolution between the pdf of a single symbol  $X$  and  $\mathcal{N}(0, \sigma^2)$ . To determine the asymptotic marginal pdf of a single component,  $X$ , consider the following line of thought (which is similar to the derivation of the Boltzmann distribution in statistical mechanics): Given that  $X = x$  ( $|x| \leq A$ ), the marginal pdf is proportional to the volume of the body formed by the intersection between the hypercube  $[-A, A]^{n-1}$  and the hyper-ball  $\mathcal{B}_{n-1}(\sqrt{nP - x^2})$ , which is given exponentially by

$$\exp \left\{ \inf_{\theta \geq 0} \left[ \theta(nP - x^2) + n \log \int_{-A}^A e^{-\theta\xi^2} d\xi \right] \right\},$$

which in turn, is proportional to  $\exp\{-\theta_\star x^2\}$ , where  $\theta_\star$  is the minimizer of  $\theta P + \log \int_{-A}^A e^{-\theta\xi^2} d\xi$ . This minimizer is  $\theta_\star = 0$  if  $P \geq A^2/3$ , and the unique positive solution to the equation

$$\frac{\int_A^A x^2 e^{-\theta x^2} dx}{\int_A^A e^{-\theta x^2} dx} = P$$

if  $P < A^2/3$ . Consequently, the asymptotic marginal of  $X$  is uniform across  $[-A, A]$  whenever  $P \geq A^2/3$  and

$$p_X(x) = \begin{cases} \frac{e^{-\theta_* x^2}}{\sqrt{\pi/\theta_*}[1-2Q(A\sqrt{2\theta_*})]} & |x| \leq A \\ 0 & \text{elsewhere} \end{cases} \quad (84)$$

whenever  $P < A^2/3$ . Thus, defining  $P' = \frac{1}{2\theta_*}$ , the asymptotic marginal of  $Y$  is given by

$$p_Y(y) = \frac{\exp\left\{-\frac{y^2}{2(P'+\sigma^2)}\right\}}{\sqrt{2\pi(P'+\sigma^2)}[1-2Q(A/\sqrt{P'})]} \cdot \left[ Q\left(\frac{\zeta y - A}{\sigma_e}\right) - Q\left(\frac{\eta y + A}{\sigma_e}\right) \right], \quad (85)$$

with  $\eta = \frac{P'}{P'+\sigma^2}$  and  $\sigma_e = \sqrt{\frac{P'\sigma^2}{P'+\sigma^2}}$ . In other words, the last term is given by

$$\begin{aligned} & E \log \left[ 1 - Q\left(\frac{A - m(Y)}{s}\right) - Q\left(\frac{A + m(Y)}{s}\right) \right] \\ &= \int_{-\infty}^{\infty} dy p_Y(y) \cdot \log \left[ 1 - Q\left(\frac{A - m(y)}{s}\right) - Q\left(\frac{A + m(y)}{s}\right) \right]. \end{aligned}$$

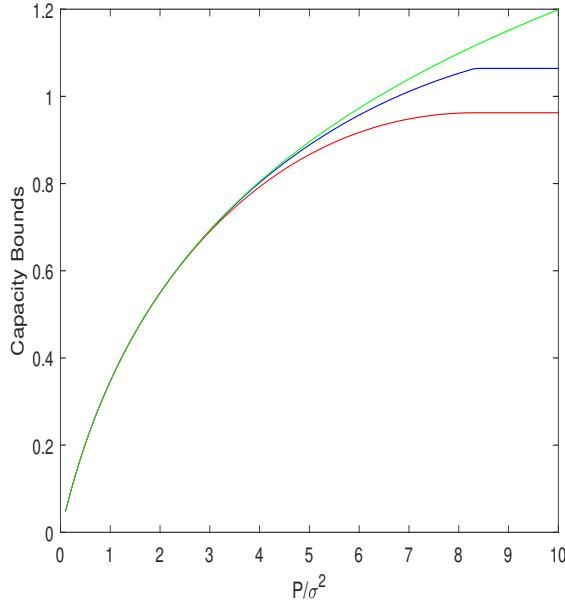


Figure 3: Exact capacity of  $A = \infty$  (green), EPI-based bound (red), and present bound (blue), both for  $A = 5$  and  $\sigma^2 = 1$ .

As can be seen in Fig. 3, the present bound is better than the EPI bound, and the gap becomes visible especially as SNR grows. Note that in the blue graph there is a phase transition at  $P = A^2/3 = 5^2/3 = 8.333\dots$  beyond which the power constraint becomes slack given the amplitude constraint.

Let us also revisit the comparison Smith [26], but this time, also with numerical results on the present bound. As mentioned earlier, Fig. 4 of his paper displays plots of the

capacity as function of the the SNR for  $A = \sqrt{2P}$ . For  $P/\sigma^2 = 10$ dB, the exact capacity is approximately 1.1 nats/channel-use, the EPI bound is 0.8688 and the present bound gives 0.9743. Likewise, for  $P/\sigma^2 = 6$ dB, the true capacity is nearly 0.802 nats/channel-use, the EPI bound is 0.5262 and the new bound gives 0.6316. Finally, for  $P/\sigma^2 = 12$ dB, Smith's capacity is approximately 1.412 nats/channel-use, but the EPI bound is 1.0655 and the current bound gives 1.1626.

Returning to the case  $P/\sigma^2 = 10$ dB, we also extended our present lower bound to correspond to  $q(\mathbf{x}) \propto e^{\alpha \|\mathbf{x}\|^2}$  within  $\mathcal{S}_n$  and zero elsewhere (see derivation in Appendix B), and as expected,  $\alpha > 0$  improves on the uniform pdf within  $\mathcal{S}_n$  ( $\alpha = 0$ ), since higher energy input vectors are preferred. This improved our lower bound from 0.9743 of  $\alpha = 0$  up to 1.0393 for  $\alpha = 0.1$ , as can be seen in Fig. 4. This concludes Example 6.

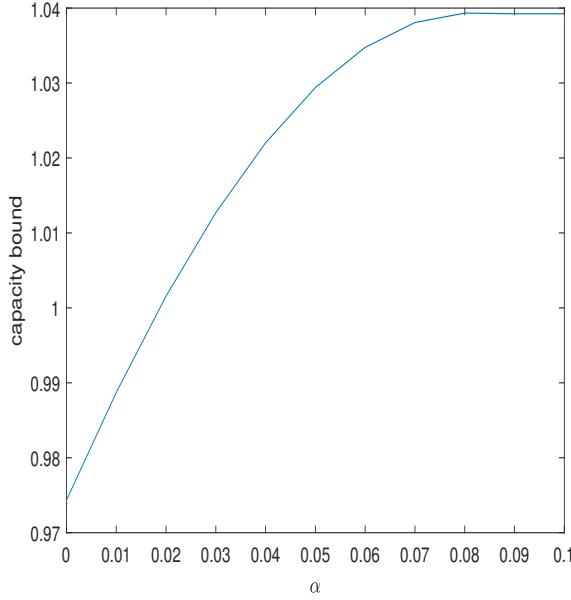


Figure 4: Present bound for  $A = \sqrt{20}$ ,  $P = 10$  and  $\sigma^2 = 1$  as a function of  $\alpha$ .

## 4.2 Constraints with Memory

When dealing with constraint functions with memory, our above derivation is supposed to involve evaluation of an expression of the form

$$\int_{\mathbb{R}^n} \exp \left\{ -\boldsymbol{\theta} \bullet \sum_{i=1}^n \phi(x_{i-m+1}^i) \right\} p(\mathbf{y}|\mathbf{x}) d\mathbf{x} = \int_{\mathbb{R}^n} \prod_{i=m}^n [\exp \{ -\boldsymbol{\theta} \bullet \phi(x_{i-m+1}^i) \} p(y_i|x_i)] d\mathbf{x}.$$

Unfortunately, since the sliding-window kernel depends here on  $y_i$ , and therefore, not fixed as before, there is no apparent single-letter characterization of its exponential growth rate, to the best knowledge of the authors. In order to proceed, it seems necessary to give up on some tightness and to bound the corresponding quantity using manageable expressions. Here we only outline the starting point of the derivation and the continuation will be deferred to future work.

We begin by applying the Jensen inequality:

$$\begin{aligned}
& \mathbf{E} \left\{ \log \left[ \int_{\mathcal{S}_n} p(\mathbf{Y}|\mathbf{x}) d\mathbf{x} \right] \right\} \\
& \leq \log \mathbf{E} \left\{ \int_{\mathcal{S}_n} p(\mathbf{Y}|\mathbf{x}) d\mathbf{x} \right\} \\
& = \log \left[ \int_{\mathcal{S}_n} \frac{d\mathbf{x}'}{\text{Vol}(\mathcal{S}_n)} \int_{\mathbb{R}^n} dy p(\mathbf{y}|\mathbf{x}') \int_{\mathcal{S}_n} p(\mathbf{y}|\mathbf{x}) d\mathbf{x} \right] \\
& = \log \left[ \int_{\mathcal{S}_n^2} d\mathbf{x} d\mathbf{x}' \int_{\mathbb{R}^n} dy p(\mathbf{y}|\mathbf{x}) p(\mathbf{y}|\mathbf{x}') \right] - \log \text{Vol}(\mathcal{S}_n) \\
& = \log \left[ \int_{\mathcal{S}_n^2} d\mathbf{x} d\mathbf{x}' \frac{\exp\{-\|\mathbf{x} - \mathbf{x}'\|^2/(4\sigma^2)\}}{(4\pi\sigma^2)^{n/2}} \right] - \log \text{Vol}(\mathcal{S}_n) \\
& = \log \left[ \int_{\mathcal{S}_n^2} d\mathbf{x} d\mathbf{x}' \exp\left\{-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{4\sigma^2}\right\} \right] - \log \text{Vol}(\mathcal{S}_n) - \frac{n}{2} \log(4\pi\sigma^2) \\
& \leq \log \left[ \inf_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \mathbb{R}_+^k} \int_{\mathbb{R}^{2n}} d\mathbf{x}_1 d\mathbf{x}_2 \exp\{n(\boldsymbol{\theta}_1 + \boldsymbol{\theta}_2) \bullet \boldsymbol{\Gamma} - \boldsymbol{\theta}_1 \bullet \boldsymbol{\phi}(\mathbf{x}_1) - \boldsymbol{\theta}_2 \bullet \boldsymbol{\phi}(\mathbf{x}_2)\} \times \right. \\
& \quad \left. \exp\left\{-\frac{\|\mathbf{x}_1 - \mathbf{x}_2\|^2}{4\sigma^2}\right\} \right] - \log \text{Vol}(\mathcal{S}_n) - \frac{n}{2} \log(4\pi\sigma^2) \\
& = \inf_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \mathbb{R}_+^k} \left( n(\boldsymbol{\theta}_1 + \boldsymbol{\theta}_2) \bullet \boldsymbol{\Gamma} + n \log \left[ \int_{\mathbb{R}^2} dx_1 dx_2 \exp\left\{-\boldsymbol{\theta}_1 \bullet \boldsymbol{\phi}(\mathbf{x}_1) - \boldsymbol{\theta}_2 \bullet \boldsymbol{\phi}(\mathbf{x}_2) - \frac{(x_1 - x_2)^2}{4\sigma^2}\right\} \right] \right) - \log \text{Vol}(\mathcal{S}_n) - \frac{n}{2} \log(4\pi\sigma^2) \\
& \triangleq n \cdot \inf_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \mathbb{R}_+^k} [(\boldsymbol{\theta}_1 + \boldsymbol{\theta}_2) \bullet \boldsymbol{\Gamma} + \log Z(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)] - \log \text{Vol}(\mathcal{S}_n) - \frac{n}{2} \log(4\pi\sigma^2). \tag{86}
\end{aligned}$$

For a possible further improvement, one may execute a change of measures, using the following well-known inequality for a positive random variable,  $U$ :

$$\mathbf{E}_p \{ \log U \} \leq \log \mathbf{E}_q \{ U \} + D(p\|q), \tag{87}$$

where  $q$  is an arbitrary distribution and equality is achieved for  $q(u) = \frac{p(u)/u}{\int_0^\infty p(u')/u' du'}$ . In our case,  $p(\mathbf{y}) = \frac{1}{\text{Vol}(\mathcal{S}_n)} \int_{\mathcal{S}_n} p(\mathbf{y}|\mathbf{x}) d\mathbf{x}$ . We will take the auxiliary distribution to be  $q(\mathbf{y}) = \frac{1}{\text{Vol}(\mathcal{S}_n)} \int_{\mathcal{S}_n} q(\mathbf{y}|\mathbf{x}) d\mathbf{x}$ , where  $q(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{x}, s^2 I_n)$  (other possibilities will be considered in

the sequel). Now,

$$\begin{aligned}
& \mathbf{E} \left\{ \log \left[ \int_{\mathcal{S}_n} p(\mathbf{Y}|\mathbf{x}) d\mathbf{x} \right] \right\} \\
\leq & \log \mathbf{E}_q \left\{ \int_{\mathcal{S}_n} p(\mathbf{Y}|\mathbf{x}) d\mathbf{x} \right\} + D(p_{\mathbf{Y}} \| q_{\mathbf{Y}}) \\
\leq & \log \mathbf{E}_q \left\{ \int_{\mathcal{S}_n} p(\mathbf{Y}|\mathbf{x}) d\mathbf{x} \right\} + D(p_{\mathbf{X}\mathbf{Y}} \| q_{\mathbf{X}\mathbf{Y}}) \\
\leq & \log \mathbf{E}_Q \left\{ \int_{\mathcal{S}_n} p(\mathbf{Y}|\mathbf{x}) d\mathbf{x} \right\} + \frac{n}{2} \left( \frac{\sigma^2}{s^2} - \log \frac{\sigma^2}{s^2} - 1 \right) \\
= & \log \left\{ \int_{\mathcal{S}_n^2} d\mathbf{x} d\mathbf{x}' \int_{\mathbb{R}^n} p(\mathbf{y}|\mathbf{x}) q(\mathbf{y}|\mathbf{x}') d\mathbf{y} \right\} - \log \text{Vol}(\mathcal{S}_n) + \frac{n}{2} \left( \frac{\sigma^2}{s^2} - \log \frac{\sigma^2}{s^2} - 1 \right) \\
= & \log \left\{ \int_{\mathcal{S}_n^2} d\mathbf{x} d\mathbf{x}' (2\pi\sigma^2)^{-n/2} (2\pi s^2)^{-n/2} \int_{\mathbb{R}^n} \exp \left[ -\frac{\|\mathbf{y} - \mathbf{x}\|^2}{2\sigma^2} - \frac{\|\mathbf{y} - \mathbf{x}'\|^2}{2s^2} \right] d\mathbf{y} \right\} - \\
& \log \text{Vol}(\mathcal{S}_n) + \frac{n}{2} \left( \frac{\sigma^2}{s^2} - \log \frac{\sigma^2}{s^2} - 1 \right) \\
= & \log \left\{ \int_{\mathcal{S}_n^2} d\mathbf{x} d\mathbf{x}' [2\pi(\sigma^2 + s^2)]^{-n/2} \exp \left[ -\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2(\sigma^2 + s^2)} \right] \right\} - \\
& \log \text{Vol}(\mathcal{S}_n) + \frac{n}{2} \left( \frac{\sigma^2}{s^2} - \log \frac{\sigma^2}{s^2} - 1 \right) \\
= & \log \left\{ \int_{\mathcal{S}_n^2} d\mathbf{x} d\mathbf{x}' \frac{n}{2} \left( \frac{\sigma^2}{s^2} - \log \frac{\sigma^2}{s^2} - 1 \right) \right\} \\
= & \log \left\{ \int_{\mathcal{S}_n^2} d\mathbf{x} d\mathbf{x}' \exp \left[ -\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2(\sigma^2 + s^2)} \right] \right\} - \\
& - \frac{n}{2} \log [2\pi(\sigma^2 + s^2)] - \log \text{Vol}(\mathcal{S}_n) + \frac{n}{2} \left( \frac{\sigma^2}{s^2} - \log \frac{\sigma^2}{s^2} - 1 \right). \tag{88}
\end{aligned}$$

Finally, the first term is handled as before, except that in the definition of  $Z(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ , the denominator of the exponent,  $4\sigma^2$ , is replaced by  $2(\sigma^2 + s^2)$ . Somewhat more generally, if we define  $q(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\alpha\mathbf{x}, s^2 I_n)$ , we end up with

$$\begin{aligned}
& \log \left\{ \int_{\mathcal{S}_n^2} d\mathbf{x} d\mathbf{x}' \exp \left[ -\frac{\|\mathbf{x} - \alpha\mathbf{x}'\|^2}{2(\sigma^2 + s^2)} \right] \right\} - \\
& - \frac{n}{2} \log [2\pi(\sigma^2 + s^2)] - \log \text{Vol}(\mathcal{S}_n) + \frac{n}{2} \left( \frac{\sigma^2}{s^2} - \log \frac{\sigma^2}{s^2} - 1 \right) + \frac{1}{2s^2} \mathbf{E} \|\mathbf{X} - \alpha\mathbf{X}\|^2 \tag{89}
\end{aligned}$$

where the last term is  $\frac{n(1-\alpha)^2 \mathbf{E}\{\mathbf{X}^2\}}{2s^2}$ , which requires the marginal of  $X$  (e.g., the variance  $\sigma_X^2$  of the truncated Gaussian RV in Smith's model).

## 5 Summary and Conclusion

In this work, we have addressed the classical problem of assessing the capacity of the discrete-time Gaussian memoryless channel, focusing on a general framework of pointwise channel input constraints, which are relevant in a wide spectrum of theoretical and practical scenarios, including peak-power constraints, correlation constraints, limitations on the relative frequency of sign changes in the channel input signal, and many others. Our main contribution is in proposing systematic methodologies for deriving good lower bounds to the channel capacity in this general framework.

Two classes of lower bounds are derived based on the precise evaluation of the asymptotic exponential behavior of the volume (that is, the volume exponent) of the set of legitimate input vectors, namely, those that satisfy the aforementioned constraints. As mentioned, the mathematical analysis technique is based on extensions of the method of types to continuous alphabets [18], which can be presented also on the basis of the saddle-point integration method [4] and large deviations theory [5]. The first class of lower bounds applies to continuous-valued channel inputs and relies on the classical entropy-power inequality [3]. The second class provides tighter results at the expense of more complicated expressions to be optimized.

The quality and generality of the bounds is demonstrated in several examples, including the classical peak- and average power constraints, absolute value constraints, correlation constraints, both separate and combined. Also the combination of a linear operation and a peak power constraint is mentioned.

For future work along the same line of thought, several directions seem to be interesting:

1. Further development of the second class of bounds for constraints with memory (see the end of Subsection 4.2).
2. Extending the results for colored Gaussian channels.
3. Extending the results for memoryless channels that are not necessarily Gaussian.
4. Further development of the improved bounds for non-uniform inputs, in continuation to the last part of Subsection 4.1.
5. Dual upper bounds for the rate-distortion function of the Gaussian source subject to multiple constraints on the reproduction vector.

6. Applying similar techniques to the timely problem of Integrated Sensing and Communications (ISAC) [28], [30]. Here the bounds could apply to both sensing and communication: the mutual information is relevant for users that decode the transmitted message, while those who merely know the statistics of the transmitted signal and attempt to estimate it, under say, the minimum mean square error (MMSE) performance measure (considering interesting bounds related to MMSE).

## Appendix A

*Proof of Lemma 1.* Define the exponential family,

$$f_{\boldsymbol{\theta}}(x_i) = \frac{\exp\{-\boldsymbol{\theta} \bullet \phi(x_i)\}}{Z(\boldsymbol{\theta})}, \quad (\text{A.1})$$

and let  $f_{\boldsymbol{\theta}}(\mathbf{x}) = \prod_{i=1}^n f_{\boldsymbol{\theta}}(x_i)$ . Now, for every  $\boldsymbol{\theta} \in \mathbb{R}_+^k$ ,

$$\begin{aligned} 1 &\geq \int_{\mathcal{S}_n} f_{\boldsymbol{\theta}}(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{S}_n} \frac{\exp\{-\boldsymbol{\theta} \bullet \sum_{i=1}^n \phi(x_i)\}}{[Z(\boldsymbol{\theta})]^n} d\mathbf{x} \\ &\geq \int_{\mathcal{S}_n} \frac{\exp\{-n\boldsymbol{\theta} \bullet \boldsymbol{\Gamma}\}}{[Z(\boldsymbol{\theta})]^n} d\mathbf{x} \\ &= \frac{\text{Vol}\{\mathcal{S}_n\} \exp\{-n\boldsymbol{\theta} \bullet \boldsymbol{\Gamma}\}}{[Z(\boldsymbol{\theta})]^n}, \end{aligned} \quad (\text{A.2})$$

and so,

$$\text{Vol}\{\mathcal{S}_n\} \leq \exp\{n\boldsymbol{\theta} \bullet \boldsymbol{\Gamma}\} \cdot [Z(\boldsymbol{\theta})]^n. \quad (\text{A.3})$$

Taking logarithms of both sides, dividing by  $n$ , and finally, passing to the limit of  $n \rightarrow \infty$ , yields

$$v(\boldsymbol{\Gamma}) \leq \boldsymbol{\theta} \bullet \boldsymbol{\Gamma} + \psi(\boldsymbol{\theta}), \quad (\text{A.4})$$

and since the left-hand side does not depend on  $\boldsymbol{\theta}$ , we may minimize the right-hand side over  $\boldsymbol{\theta} \in \mathbb{R}_+^k$ , and obtain

$$v(\boldsymbol{\Gamma}) \leq \omega(\boldsymbol{\Gamma}). \quad (\text{A.5})$$

This completes the proof of part 1.

Moving on to part 2, select  $\boldsymbol{\theta}$  such that  $\mathbf{E}_{\boldsymbol{\theta}}\{\phi(X)\} \equiv -\dot{\psi}(\boldsymbol{\theta}) = \boldsymbol{\Gamma}_i$ , where  $\mathbf{E}_{\boldsymbol{\theta}}\{\cdot\}$  denotes expectation under  $f_{\boldsymbol{\theta}}$ . This is possible by one of the postulates of part 2. Let  $c > 0$  be an

arbitrary constant, and consider the set

$$\begin{aligned}\mathcal{S}'_n &= \left\{ \mathbf{x} : n\Gamma_j - c\sqrt{n} \leq \sum_{i=1}^n \phi_j(x_i) \leq n\Gamma_j \quad j = 1, 2, \dots, k \right\} \\ &= \left\{ \mathbf{x} : 0 \leq \frac{1}{\sqrt{n}} \sum_{i=1}^n [\Gamma_j - \phi_j(x_i)] \leq c \quad j = 1, 2, \dots, k \right\}.\end{aligned}\quad (\text{A.6})$$

Now, by the multivariate version of the central limit theorem, as  $n \rightarrow \infty$ , the probability of  $\mathcal{S}'_n$  tends to the probability of the hypercube  $[0, c]^k$  under the zero-mean multivariate Gaussian distribution with covariance matrix  $\{\text{Cov}_{\boldsymbol{\theta}}\{\phi_i(X), \phi_j(X)\}, i, j = 1, \dots, k\} = \ddot{\psi}(\boldsymbol{\theta})$ , where  $\text{Cov}_{\boldsymbol{\theta}}\{\cdot, \cdot\}$  denotes covariance under  $f_{\boldsymbol{\theta}}$ . Let us denote this probability by  $\Pi_c(\boldsymbol{\theta})$ . Thus, for every  $\epsilon > 0$ , and large enough  $n$ ,

$$\int_{\mathcal{S}'_n} f_{\boldsymbol{\theta}}(\mathbf{x}) d\mathbf{x} \geq \Pi_c(\boldsymbol{\theta}) \cdot (1 - \epsilon). \quad (\text{A.7})$$

On the other hand, consider the following chain of inequalities:

$$\begin{aligned}\int_{\mathcal{S}'_n} f_{\boldsymbol{\theta}}(\mathbf{x}) d\mathbf{x} &= \int_{\mathcal{S}'_n} \frac{\exp\{-\boldsymbol{\theta} \bullet \sum_{i=1}^n \phi(x_i)\}}{[Z(\boldsymbol{\theta})]^n} d\mathbf{x} \\ &\leq \int_{\mathcal{S}'_n} \frac{\exp\{-n\boldsymbol{\theta} \bullet (\boldsymbol{\Gamma} - \mathbf{c}/\sqrt{n})\}}{[Z(\boldsymbol{\theta})]^n} d\mathbf{x} \\ &= \frac{\text{Vol}\{\mathcal{S}'_n\} \cdot \exp\{-n\boldsymbol{\theta} \bullet (\boldsymbol{\Gamma} - \mathbf{c}/\sqrt{n})\}}{[Z(\boldsymbol{\theta})]^n} \\ &\leq \frac{\text{Vol}\{\mathcal{S}_n\} \cdot \exp\{-n\boldsymbol{\theta} \bullet (\boldsymbol{\Gamma} - \mathbf{c}/\sqrt{n})\}}{[Z(\boldsymbol{\theta})]^n},\end{aligned}\quad (\text{A.8})$$

where  $\mathbf{c}$  is the  $k$ -dimensional vector whose components are equal to  $c$ , and the last step is due to the fact that  $\mathcal{S}'_n \subseteq \mathcal{S}_n$ . Combining this with eq. (A.7), we have

$$\text{Vol}\{\mathcal{S}_n\} \geq [Z(\boldsymbol{\theta})]^n \cdot \exp\{n\boldsymbol{\theta} \bullet (\boldsymbol{\Gamma} - \mathbf{c}/\sqrt{n})\} \cdot \Pi_c(\boldsymbol{\theta})(1 - \epsilon). \quad (\text{A.9})$$

Taking logarithms of both sides, dividing by  $n$ , and passing to the limit of  $n \rightarrow \infty$ , we get

$$v(\boldsymbol{\Gamma}) \geq \psi(\boldsymbol{\theta}) + \boldsymbol{\theta} \bullet \boldsymbol{\Gamma} \geq \inf_{\boldsymbol{\theta} \in \mathbb{R}_+^k} \{\psi(\boldsymbol{\theta}) + \boldsymbol{\theta} \bullet \boldsymbol{\Gamma}\} = \omega(\boldsymbol{\Gamma}), \quad (\text{A.10})$$

completing the proof of part 2.

## Appendix B

*Derivation of Example 6 for  $q(\mathbf{x}) \propto \exp\{\alpha \|\mathbf{x}\|^2\}$ .*

Consider Smith's model, where

$$q(\mathbf{x}) = \begin{cases} \frac{e^{\alpha\|\mathbf{x}\|^2}}{Z_n(\alpha, A)} & \mathbf{x} \in \mathcal{S}_n \\ 0 & \text{elsewhere} \end{cases} \quad (\text{B.1})$$

where

$$Z_n(\alpha, A) = \int_{\mathcal{S}_n} e^{\alpha\|\mathbf{x}\|^2} d\mathbf{x} \quad (\text{B.2})$$

and  $\mathcal{S}_n = [-A, A]^n \cap \mathcal{B}_n(\sqrt{n}P)$  with  $\mathcal{B}_n(r) = \{\mathbf{x} : \|\mathbf{x}\|^2 \leq r^2\}$ . Let us denote

$$J(a, b) \triangleq \int_{-A}^A e^{ax^2+bx} dx. \quad (\text{B.3})$$

The mutual information is given by

$$I(\mathbf{X}; \mathbf{Y}) = h(\mathbf{Y}) - \frac{n}{2} \log(2\pi e \sigma^2) \quad (\text{B.4})$$

$$= -\mathbf{E} \left\{ \log \left[ \int_{\mathcal{S}_n} q(\mathbf{x}) p(\mathbf{Y}|\mathbf{x}) d\mathbf{x} \right] \right\} - \frac{n}{2} \log(2\pi e \sigma^2) \quad (\text{B.5})$$

$$\begin{aligned} &= -\mathbf{E} \left\{ \log \left[ \int_{\mathcal{S}_n} \exp \left\{ \alpha\|\mathbf{x}\|^2 - \frac{\|\mathbf{Y} - \mathbf{x}\|^2}{2\sigma^2} \right\} d\mathbf{x} \right] \right\} + \log Z_n(\alpha, A) + \\ &\quad \frac{n}{2} \log(2\pi \sigma^2) - \frac{n}{2} \log(2\pi e \sigma^2) \\ &= -\mathbf{E} \left\{ \log \left[ \int_{\mathcal{S}_n} \exp \left\{ \alpha\|\mathbf{x}\|^2 - \frac{\|\mathbf{Y} - \mathbf{x}\|^2}{2\sigma^2} \right\} d\mathbf{x} \right] \right\} + \log Z_n(\alpha, A) - \frac{n}{2} \\ &\triangleq -A_n + B_n - \frac{n}{2}. \end{aligned} \quad (\text{B.6})$$

Now,

$$\begin{aligned} B_n &= \log Z_n(\alpha, A) \\ &= \log \left[ \int_{\mathcal{S}_n} \exp \{ \alpha\|\mathbf{x}\|^2 \} d\mathbf{x} \right] \\ &= \inf_{\theta \geq 0} \log \left[ \int_{[-A, A]^n} \exp \{ n\theta P + (\alpha - \theta)\|\mathbf{x}\|^2 \} d\mathbf{x} \right] \\ &= n \cdot \inf_{\theta \geq 0} \left\{ \theta P + \log \left( \int_{-A}^A e^{(\alpha-\theta)x^2} dx \right) \right\} \\ &= n \cdot \inf_{\theta \geq 0} \{ \theta P + \log J(\alpha - \theta, 0) \}. \end{aligned} \quad (\text{B.7})$$

and

$$\begin{aligned}
A_n &= \mathbf{E} \left\{ \log \left[ \int_{\mathcal{S}_n} \exp \left\{ \alpha \|\mathbf{x}\|^2 - \frac{\|\mathbf{Y} - \mathbf{x}\|^2}{2\sigma^2} \right\} d\mathbf{x} \right] \right\} \\
&\leq \inf_{\theta \geq 0} \left( n\theta P + \mathbf{E} \left\{ \log \left[ \int_{[-A,A]^n} \exp \left\{ (\alpha - \theta) \|\mathbf{x}\|^2 - \frac{\|\mathbf{Y} - \mathbf{x}\|^2}{2\sigma^2} \right\} d\mathbf{x} \right] \right\} \right) \\
&= \inf_{\theta \geq 0} \left( n\theta P + \mathbf{E} \left\{ \log \left[ \prod_{i=1}^n \int_{-A}^A \exp \left\{ (\alpha - \theta) x^2 - \frac{(Y_i - x)^2}{2\sigma^2} \right\} dx \right] \right\} \right) \\
&= \inf_{\theta \geq 0} \left( n\theta P + \sum_{i=1}^n \mathbf{E} \left\{ -\frac{Y_i^2}{2\sigma^2} + \log \left[ \int_{-A}^A \exp \left\{ \left( \alpha - \frac{1}{2\sigma^2} - \theta \right) x^2 + \frac{xY_i}{\sigma^2} \right\} dx \right] \right\} \right) \\
&= n \cdot \inf_{\theta \geq 0} \left[ \theta P - \frac{\mathbf{E}\{Y^2\}}{2\sigma^2} + \mathbf{E} \left\{ \log J \left( \alpha - \frac{1}{2\sigma^2} - \theta, \frac{Y}{\sigma^2} \right) \right\} \right] \\
&= n \cdot \inf_{\theta \geq 0} \left[ \theta P - \frac{\mathbf{E}\{X^2\} + \sigma^2}{2\sigma^2} + \mathbf{E} \left\{ \log J \left( \alpha - \frac{1}{2\sigma^2} - \theta, \frac{Y}{\sigma^2} \right) \right\} \right] \\
&= n \cdot \inf_{\theta \geq 0} \left[ \theta P - \frac{\mathbf{E}\{X^2\}}{2\sigma^2} + \mathbf{E} \left\{ \log J \left( \alpha - \frac{1}{2\sigma^2} - \theta, \frac{Y}{\sigma^2} \right) \right\} \right] - \frac{n}{2}, \tag{B.8}
\end{aligned}$$

where the expectations are w.r.t. the asymptotic marginals of the single symbols  $X$  and  $Y$ , respectively. Let  $\theta_*$  denote the minimizing  $\theta$  in the definition of  $B_n$ . Then,  $p_X(x)$  is given by

$$p_X(x) = \frac{e^{(\alpha-\theta_*)x^2} \mathcal{I}\{|x| \leq A\}}{J(\alpha - \theta_*, 0)} \tag{B.9}$$

where  $\theta_* = 0$  whenever  $P \geq \frac{\int_{-A}^A x^2 e^{\alpha x^2} dx}{\int_{-A}^A e^{\alpha x^2} dx} = \frac{\partial \log J(\alpha, 0)}{\partial \alpha}$ . The marginal of  $Y$  is given by the convolution between  $p_X$  and  $\mathcal{N}(0, \sigma^2)$ , namely,

$$\begin{aligned}
p_Y(y) &= \int_{-A}^A \frac{e^{(\alpha-\theta_*)x^2}}{J(\alpha - \theta_*, 0)} \cdot \frac{e^{-(y-x)^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}} dx \\
&= \frac{\exp\{-y^2/(2\sigma^2)\}}{J(\alpha - \theta_*, 0)\sqrt{2\pi\sigma^2}} \int_{-A}^A \exp \left\{ \left( \alpha - \frac{1}{2\sigma^2} - \theta_* \right) x^2 + \frac{xy}{\sigma^2} \right\} dx \\
&= \frac{\exp\{-y^2/(2\sigma^2)\}}{J(\alpha - \theta_*, 0)\sqrt{2\pi\sigma^2}} \cdot J \left( \alpha - \frac{1}{2\sigma^2} - \theta_*, \frac{y}{\sigma^2} \right). \tag{B.10}
\end{aligned}$$

To summarize,  $C(\mathbf{\Gamma}) \geq C_1$ , where

$$\begin{aligned}
C_1 &= \inf_{\theta \geq 0} \{ \theta P + \log J(\alpha - \theta, 0) \} + \frac{\int_{-A}^A x^2 e^{(\alpha-\theta_*)x^2} dx}{2\sigma^2 J(\alpha - \theta_*, 0)} - \inf_{\vartheta \geq 0} \left\{ \vartheta P + \right. \\
&\quad \left. \int_{-\infty}^{\infty} dy \frac{\exp\{-y^2/(2\sigma^2)\}}{J(\alpha - \theta_*, 0)\sqrt{2\pi\sigma^2}} \cdot J \left( \alpha - \frac{1}{2\sigma^2} - \theta_*, \frac{y}{\sigma^2} \right) \times \right. \\
&\quad \left. \log J \left( \alpha - \frac{1}{2\sigma^2} - \vartheta, \frac{y}{\sigma^2} \right) \right\}. \tag{B.11}
\end{aligned}$$

where  $\theta_*$  is the minimizing  $\theta$  in the first minimization.

## References

- [1] A. Chaaban, Z. Rezki, and M.-S. Alouini, “On the capacity of intensity-modulation direct-detection Gaussian optical wireless communication channels: a tutorial,” *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 455-491, first quarter, 2022.
- [2] L. Collatz, “Einschließungssatz für die charakteristischen Zahlen von Matrizen”, **Mathematische Zeitschrift**, vol. 48, no. 1, pp. 221-226, 1942.
- [3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Second Edition, Wiley, 2006.
- [4] N. G. de Bruijn, *Asymptotic Methods in Analysis*, Dover Publications, Inc., Second Ed., 1981.
- [5] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Springer-Verlag, Second Edition, New York, 1998.
- [6] M. D. Donsker and S. R. S. Varadhan, “Asymptotic evaluation of certain Markov process expectations for large time, IV” *Communications on Pure and Applied Mathematics*, vol. 36, no. 2, pp. 183–212, 1983.
- [7] A. Dytso, M. Al, H. V. Poor, and S. Shamai (Shitz), “On the capacity of the peak power constraint vector Gaussian channel: An estimation theoretic perspective,” *IEEE Trans. Inform. Theory*, vol. 65, no. 6, pp. 3907-3921, June 2019.
- [8] A. Dytso, M. Goldenbaum, H. V. Poor, and S. Shamai (Shitz), “Amplitude constrained MIMO channels: properties of optimal input distributions and bounds on the capacity,” *Entropy*, vol. 21, no. 2, 2019.
- [9] A. Dytso, S. Yagli, H. V. Poor, and S. Shamai (Shitz), “The capacity achieving distribution for the amplitude constrained additive Gaussian channel: An upper bound on the number of mass points,” *IEEE Trans. Inform. Theory*, vol. 66, no. 4, pp. 2006-2022, 2020.
- [10] J. Eisen, R. Mazumdar, and P. Mitran, “Capacity-achieving input distributions of additive vector Gaussian noise channels: even moment constraints and unbounded or compact support,” *Entropy*, 25(8), 1180, 2023.

- [11] J. Fahs and I. Abou-Faycal, “On properties of the support of capacity-achieving distributions for additive noise channel models with input cost constraints,” *IEEE Trans. Inform. Theory*, vol. 64, no. 2, pp. 1178-1198, February 2018.
- [12] A. Girish, E. Telatar and S. Shamai (Shitz), “On entropy-constrained Gaussian channel capacity via the moment problem,” *Proc. 2025 IEEE International Symposium on Information Theory (ISIT 2025)*, June 22-27, 2025, Ann Arbor (Michigan), USA.
- [13] V. Jog and V. Anantharam, “An energy harvesting AWGN channel with a finite battery,” *Proc. 2014 IEEE International Symposium on Information Theory (ISIT 2014)*, pp. 806-810, Honolulu, Hawaii, U.S.A., July 2014.
- [14] V. Jog and V. Anantharam, “A geometric analysis of the AWGN channel with a  $(\sigma, \rho)$ -power constraint,” *IEEE Trans. Inform. Theory*, vol. 62, no. 8, pp. 4413-4438, August 2016.
- [15] W. Hirt and J. L. Massey, “Capacity of the discrete-time Gaussian channel with inter-symbol interference,” *IEEE Trans. Inform. Theory*, vol. 34, pp. 380-388, May 1988.
- [16] A. Lapidoth, S. M. Moser, and M. A. Wigger, “On the capacity of free-space optical intensity channels,” *IEEE Trans. Inform. Theory*, vol. 55, no. 10, pp. 4449-4461, October 2009.
- [17] N. Merhav, *Statistical Physics for Electrical Engineering*, Springer Nature, Cham, Switzerland, 2018.
- [18] N. Merhav and N. Weinberger, “A toolbox for refined information-theoretic analyses”, *Foundations and Trends in Communications and Information Theory*, vol. 22, no. 1, pp. 1-184, January 2025.
- [19] L. H. Ozarow and A. D. Wyner, “On the capacity of the Gaussian channel with finite number of input levels,” *IEEE Trans. Inform. Theory*, vol. 36, no. 6, pp. 1426-1429, November 1990.
- [20] M. Peleg and S. Shamai (Shitz), “On the capacity of the peak-limited and band-limited channel. *Entropy*, December 2024; 26(12):1049.

- [21] S. Shamai (Shitz) and I. Bar-David, “The capacity of average and peak-power-limited quadrature Gaussian channels,” *IEEE Trans. Inform. Theory*, vol. 41, no. 4, pp. 1060–1071, July 1995.
- [22] S. Shamai (Shitz) and Y. Kofman, “On the capacity of binary and Gaussian channels with run-length-limited inputs,” *IEEE Trans. Communications*, vol. 38, no. 5, pp. 584–594, May 1990.
- [23] S. Shamai (Shitz), L. H. Ozarow and A. D. Wyner, “Information rates for a discrete-time Gaussian channel with inter-symbol interference and stationary inputs,” *IEEE Trans. Inform. Theory*, vol. 37, no. 6, pp. 1527–1539, November 1991,
- [24] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, 1949.
- [25] N. Sharma and S. Shamai (Shitz), “Transition points in the capacity-achieving distribution for the peak-power limited AWGN and free-space optical intensity channels,” *Problems of Information Transmission*, vol. 46, no. 4, pp. 283–299, April 2010.
- [26] J. G. Smith, “The information capacity of amplitude- and variance-constrained scalar Gaussian channels,” *Information and Control*, vol. 18, no. 3, 1971.
- [27] A. Thangaraj, G. Kramer, G. Bocherer, “Capacity bounds for discrete-time, amplitude-constrained, additive white Gaussian noise channels,”
- [28] D. Wen, Y. Zhou, X. Li, Y. Shi, K. Huang and K. B. Letaief, “A Survey on Integrated Sensing, Communication, and Computation,” *IEEE Communications Surveys & Tutorials*, to appear, 2025. (arXiv:2408.08074). *IEEE Trans. Inform. Theory*, vol. 63, no. 7, pp. 4172–4182, July 2017.
- [29] H. Wielandt, “Unzerlegbare, nicht negative Matrizen,” *Mathematische Zeitschrift*, vol. 52, no. 1, 642–648, 1950.
- [30] Y. Xiong, F. Liu, Y. Cui, W. Yuan, T. X. Han, and G. Caire, “On the Fundamental Tradeoff of Integrated Sensing and Communications and Gaussian Channels,” *IEEE Trans. Inform. Theory*, vol. 69, no. 9, pp. 5723–5751, September 2023.