Capacity Bounds Via Duality: A Phase Noise Example

Amos Lapidoth
Signal and Information Processing Laboratory
Swiss Federal Institute of Technology (ETH) Zurich
CH-8092 Zurich, Switzerland
e-mail: lapidoth@isi.ee.ethz.ch

Abstract — **Lapidoth and Moser have recently proposed a general technique for obtaining upper bounds on channel capacity via a dual expression in which the maximization over probability distributions on the channel input alphabet is replaced with a minimization over probability distributions on the channel output alphabet. They have also introduced the notion of "capacity achieving input distributions that escape to infinity" in order to study channel capacity at high signal-to-noise (SNR) ratios.**

In this partly tutorial paper we shall demonstrate the use of these ideas by applying them to the study of communication over discrete-time channels impaired by additive Gaussian noise and phase noise.

I. Introduction

Lapidoth and Moser [1] have recently proposed a general approach to obtain upper bounds on channel capacity. The approach is based on a dual expression for channel capacity in which maximization (of mutual information) over the space of probability distributions on the input alphabet is replaced with a minimization (of average relative entropy) over the space of probability distributions on the output alphabet. While this dual expression had been known for a while [2, Section 2.3], [3, Exercise 4.17], [4] it had been — to the best of our knowledge — previously mostly used to derive a termination criterion for iterative numerical calculations of the capacity of discrete memoryless channels (DMCs) and to derive connections between the redundancy in universal source coding and channel capacity. In [1] it was proposed to use this expression to derive closed form upper bounds on the capacity of channels with infinite alphabets.

The key to the method is the inequality

$$I(Q;W) \leq \sum_{x\in\mathcal{X}} Q(x) D\big(W(\cdot|x)\|R(\cdot)\big), \quad R \in \mathcal{P}(\mathcal{Y}) \tag{1}$$

which upper bounds the mutual information $I(Q;W)$ between the terminals of a DMC $W(y|x)$ under the input distribution Q in terms of the average (over Q) relative entropy $D\big(W(\cdot|x)\|R(\cdot)\big)$ between the channel output distribution $W(\cdot|x)$ corresponding to the input x and some *arbitrary* distribution $R(\cdot)$ on the channel output alphabet. Here $\mathcal{X}$ and $\mathcal{Y}$ denote the finite channel input and output alphabets, and $\mathcal{P}(\mathcal{Y})$ denotes the set of all probability distributions on the output alphabet $\mathcal{Y}$. While choosing $R(\cdot)$ to be the output distribution $(QW)(y) = \sum_{x'\in\mathcal{X}} Q(x')W(y|x')$ that corresponds to the input distribution Q will yield an equality in (1), other output distributions may lead to more tractable expressions. In [5] we extended this inequality to infinite alphabets and proposed that by a judicious choice of the probability measure $R(\cdot)$ this inequality can lead to useful upper bounds on channel capacity. The proposed approach was used in order to study multi-antenna fading channels [1] [5] and in the study of constrained communication over finite state channels [6].

In [7] it was noticed that for many channels with power constraints, capacity achieving input distributions escape to infinity. Loosely speaking, this means that the asymptotic behavior of channel capacity can be achieved even if the inputs are subjected to an additional constraint that requires them to be bounded away arbitrarily far from zero. This was then used extensively in order to study the fading number of multi-antenna systems operating over flat fading channels [8].

In this paper we shall demonstrate how these two ideas can be used in the study of the asymptotic behavior of channel capacity. We shall illustrate this approach by studying discrete-time channels with additive white Gaussian noise and phase noise impairments.

II. Channel Model

We study a channel whose time-k output $Y_k \in \mathbb{C}$ is a complex random variable given by

$$Y_k = x_k \cdot e^{i\Theta_k} + Z_k \tag{2}$$

where $x_k \in \mathbb{C}$ denotes the time-k power-$|x_k|^2$ channel input, $\{Z_k\}$ is an IID sequence of circularly symmetric zero-mean variance-σ^2 Gaussian random variables, and $\{\Theta_k\}$ is a stationary and ergodic phase noise sequence of finite entropy rate

$$h\big(\{\Theta_k\}\big) > -\infty. \tag{3}$$

We assume throughout that the process $\{Z_k\}$ is independent of the process $\{\Theta_k\}$ and that their joint law does not depend on the input sequence. The channel inputs are assumed to be power limited so that in considering a blocklength-n transmission we require

$$\frac{1}{n}\sum_{k=1}^{n} \mathsf{E}\left[|X_k|^2\right] \leq \mathcal{E}_{\mathrm{s}}. \tag{4}$$

Henceforth we shall assume $\sigma^2 = 2$ so that the signal-to-noise ratio (SNR) is

$$\mathrm{SNR} = \frac{\mathcal{E}_{\mathrm{s}}}{2}. \tag{5}$$

If the sequence $\{\Theta_k\}$ is IID, we shall say that the phase noise is memoryless and we shall drop all time indices. If, in addition, Θ is uniformly distributed over $[-\pi,\pi)$ we shall say that communication is non-coherent. In this case

$$T = |Y|^2 = \left|xe^{i\Theta} + Z\right|^2 \tag{6}$$

is a sufficient statistic. Conditional on $|x|^2$, the distribution of T is a non-central chi-square distribution with non-centrality $|x|^2$ and two degrees of freedom. We therefore write

$$T \Big| \left(|X|^2 = |x|^2\right) \;\sim\; \chi_2'^2(|x|^2) \tag{7}$$

where $\chi'^2_\nu(\lambda)$ denotes the non-central chi-square distribution with $\nu \geq 1$ degrees of freedom and non-centrality λ, i.e., the distribution that results from the addition of the squares of ν independent unit-variance real Gaussian random variables whose squared means sum to λ. For future reference we note the mean, variance, and entropy estimates of such distributions:

- Mean: $\nu + \lambda$
- Variance: $2(\nu + 2\lambda)$
- Differential entropy:

$$h\big(\chi'^2_\nu(\lambda)\big) \leq \frac{1}{2}\log\big(4\pi e(\nu+2\lambda)\big), \quad \nu \in \mathbb{N}, \lambda \in \mathbb{R}. \quad (8)$$

$$\lim_{\lambda\to\infty} h\big(\chi'^2_\nu(\lambda)\big) - \frac{1}{2}\log\big(8\pi e\lambda\big) = 0, \quad \nu \in \mathbb{N}. \quad (9)$$

III. Non-Coherent Case — Upper Bounds

The non-coherent channel can be viewed as a channel whose output T takes value in $\mathbb{R}^+$. For such channels it has been proposed in [5] to employ (1) with the choice of the output distribution $R(\cdot)$ having the Gamma density:

$$\frac{t^{\alpha-1}e^{-t/\beta}}{\beta^\alpha \Gamma(\alpha)}, \quad t \geq 0. \quad (10)$$

Here $\alpha > 0$ and $\beta > 0$ are parameters that will be optimized later. (Somewhat tighter results can be obtained by choosing the output law to be a modified Gamma distribution [5], but for our present purposes the above suffices.) This output distribution and (1) lead to the bound:

$$I(Q;W) \leq \alpha\log\beta + \log\Gamma(\alpha) - h_Q(T|X) + (1-\alpha)\mathsf{E}[\log T] + \frac{1}{\beta}\mathsf{E}[T], \quad \alpha,\beta > 0$$

where $h_Q(T|X)$ is the conditional differential entropy of T given X when X is distributed according to Q and all expectations are with respect to the law on T induced by law Q on X. Choosing $\beta = \mathsf{E}[T]/\alpha$ leads to the bound:

$$I(Q;W) \leq \alpha - \alpha\log\alpha + \log\Gamma(\alpha) + \alpha\log\mathsf{E}[T] + (1-\alpha)\mathsf{E}[\log T] - h_Q(T|X), \quad \alpha > 0. \quad (11)$$

Notice that (11) is not specific to our channel. It holds for any channel $W(\cdot|\cdot)$ taking value in the non-negative reals.

Returning to our channel we note that to use (11) we need an expression for $h_Q(T|X)$, which requires the complicated computation of the differential entropy of a $\chi'^2_2(\lambda)$ random variable. Fortunately, it can be shown that the capacity of our channel can be achieved by input distributions that escape to infinity. Hence, adding the constraint

$$|X| \geq x_{\min} \quad (12)$$

where $x_{\min}$ is any (possibly very large) positive number, does not alter the asymptotic behavior of the channel capacity as $\mathcal{E}_\mathrm{s} \to \infty$. Consequently, we may use (9) to obtain

$$h(T|X=x) = \frac{1}{2}\log|x|^2 + \frac{1}{2}(\log 8\pi e) + o(1) \quad (13)$$

where the correction term $o(1)$ tends to zero as $x_{\min}$ tends to infinity.

As for the term $\mathsf{E}[\log T|X=x]$ we use the exact expression for the expected logarithm of a non-central chi-square random variable [5] to obtain

$$\mathsf{E}[\log T|X=x] = \log|x|^2 + o(1). \quad (14)$$

With $\mathsf{E}[T] = \mathcal{E}_\mathrm{s}+2$ we now have estimates of all the terms in (11) and we thus obtain from (11), (13), and (14):

$$I(Q;W) \leq \alpha\log(\mathcal{E}_\mathrm{s}+2) + \alpha - \alpha\log\alpha - \frac{1}{2}\log(8\pi e) + \log\Gamma(\alpha) + \big(\frac{1}{2}-\alpha\big)\mathsf{E}_Q\big[\log|X|^2\big] + o(1). \quad (15)$$

Finally, upon choosing $\alpha = 1/2$ and using (5), we obtain:

$$\limsup_{\mathrm{SNR}\to\infty}\left\{C(\mathrm{SNR}) - \Big(\frac{1}{2}\log\mathrm{SNR} - \frac{1}{2}\log 2\Big)\right\} \leq 0. \quad (16)$$

(The choice $\alpha = 1/2$ is motivated by considering the max min of the RHS of (15) over $\mathsf{E}_Q\big[\log|X|^2\big] \leq \log\mathcal{E}_\mathrm{s}$ and $\alpha > 0$ respectively.)

IV. Non-Coherent Case — Lower Bounds

The proposed lower bound is again based on the Gamma distribution, but this time applied as an input distribution to the channel. We shall need the fact that if the density of $S \geq 0$ is

$$\frac{s^{\alpha-1}e^{-s}}{\Gamma(\alpha)}, \quad s \geq 0, \quad (17)$$

then

$$\mathsf{E}[S] = \alpha, \quad \mathsf{E}[\log S] = \psi(\alpha) \quad (18)$$

$$h(S) = (1-\alpha)\psi(\alpha) + \alpha + \log\Gamma(\alpha) \quad (19)$$

where $\psi(\alpha) = \Gamma'(\alpha)/\Gamma(\alpha)$ denotes the digamma function.

We now choose

$$|X|^2 = \frac{\mathcal{E}_\mathrm{s}}{\alpha}S. \quad (20)$$

Noting that by circular symmetry

$$\begin{aligned}\log\pi + h(T) &= h\big(Xe^{i\Theta} + Z\big)\\ &\geq h\big(Xe^{i\Theta}\big)\\ &= \log\pi + h\big(|X|^2\big)\end{aligned}$$

we obtain the bound

$$\begin{aligned}h(T) &\geq h\big(|X|^2\big)\\ &= \log\frac{\mathcal{E}_\mathrm{s}}{\alpha} + (1-\alpha)\psi(\alpha) + \alpha + \log\Gamma(\alpha)\end{aligned}$$

which combines with (8) to yield

$$I(X;T) \geq \log\frac{\mathcal{E}_\mathrm{s}}{\alpha} + (1-\alpha)\psi(\alpha) + \alpha + \log\Gamma(\alpha) - \frac{1}{2}\mathsf{E}\big[\log\big(8\pi e(|X|^2+1)\big)\big]. \quad (21)$$

Noting now that by [5] the condition $h(S) > -\infty$ implies

$$\lim_{\mathcal{E}_s\to\infty}\Big\{\mathsf{E}\big[\log(|X|^2+1)\big]-\mathsf{E}\big[\log(|X|^2)\big]\Big\}=0 \qquad (22)$$

we obtain

$$I(X;T) \geq \frac{1}{2}\log\mathcal{E}_s - \frac{1}{2}\log\alpha + \big(\frac{1}{2}-\alpha\big)\psi(\alpha)+\alpha + \log\Gamma(\alpha) - \frac{1}{2}\log(8\pi e) + o(1)$$

where we have used (20) and (18) to compute $\mathsf{E}\big[\log|X|^2\big]$ explicitly.

The choice of $\alpha = 1/2$ now demonstrates the achievability of

$$\frac{1}{2}\log\mathrm{SNR} - \frac{1}{2}\log 2 + o(1)$$

which combines with (16) to yield

$$C(\mathrm{SNR}) = \frac{1}{2}\log\left(1+\frac{\mathrm{SNR}}{2}\right)+o(1) \qquad (23)$$

where the $o(1)$ terms tends to zero as the SNR tends to infinity.

It is interesting to note that the choice of $\alpha = 1/2$, which at high SNR asymptotically achieves channel capacity, corresponds to choosing $|X|^2$ to have a central chi-square distribution of *one* degree of freedom. At high SNR the choice of X as a zero-mean Gaussian so that $|X|^2$ has *two* degrees of freedom is thus sub optimal.

V. Memoryless Phase Noise

We now consider the case where Θ is not uniformly distributed over $[-\pi,\pi)$. We assume that

$$h(\Theta) > -\infty \qquad (24)$$

and that the distribution of Θ is fixed and does not vary with the SNR. This latter assumption is reasonable when the source of the phase noise is inaccuracies in the oscillators, and perhaps less so when the source is Gaussian noise in the phase recovery loop.

Since the mutual information across this channel is invariant under a rotation of the input distribution, and since mutual information is concave in the input distribution, it follows that there is no loss in optimality in limiting the input distributions to circularly symmetric input distributions. For such distributions $X = |X|\cdot e^{i\Phi}$ where Φ is independent of $|X|$ and uniformly distributed we have:

$$\begin{aligned}I(X;Y) &= I\big(|X|\,;\,Y\big)+I\big(\Phi\,;\,Y\big|\,|X|\big)\\ &= I\big(|X|\,;\,|Y|\big)+I\big(\Phi\,;\,Y\big|\,|X|\big)\\ &\leq I\big(|X|\,;\,|Y|\big)+I\big(\Phi\,;\,Xe^{i\Theta}\big|\,|X|\big)\\ &= I\big(|X|\,;\,|Y|\big)+I\big(\Phi\,;\,e^{i(\Phi+\Theta)}\big)\\ &= \frac{1}{2}\log\mathrm{SNR}-\frac{1}{2}\log 2+\log(2\pi)-h(\Theta)+o(1).\end{aligned}$$

This bound is actually tight because

$$\lim_{|x|\to\infty} I\big(\Phi\,;\,Y\big|\,|X|=|x|\big) = I\big(\Phi\,;\,e^{i(\Phi+\Theta)}\big) \qquad (25)$$

and because channel capacity is achievable by inputs distributions that escape to infinity so that we may limit ourselves to inputs of such very large magnitudes.

VI. Phase Noise with Memory

We next address the case where the phase noise $\{\Theta_k\}$ is not memoryless but rather a stationary and ergodic process of finite entropy rate. Denoting by X^k the random variables $X_1,\ldots,X_k$ and similarly for Y we have

$$\frac{1}{n}I(X^n;Y^n) = \frac{1}{n}\sum_{k=1}^{n} I(X^n;Y_k|Y^{k-1})$$

Inspecting each of the terms on the right hand side of the above separately we have:

$$\begin{aligned}I(X^n;Y_k|Y^{k-1}) &= h(Y_k|Y^{k-1}) - h(Y_k|Y^{k-1},X^n)\\ &\leq h(Y_k) - h(Y_k|Y^{k-1},X^n)\\ &= h(Y_k) - h(Y_k|Y^{k-1},X^{k-1},X_k)\\ &\leq h(Y_k) - h(Y_k|Y^{k-1},X^{k-1},\Theta^{k-1},X_k)\\ &= h(Y_k) - h(Y_k|\Theta^{k-1},X_k)\\ &= I(X_k;Y_k) + I(\Theta^{k-1};Y_k|X_k)\\ &= I(X_k;Y_k) + I(\Theta^{k-1};Y_k,X_k)\\ &\leq I(X_k;Y_k) + I(\Theta^{k-1};\Theta_k).\end{aligned}$$

Here the first term on the right corresponds to the mutual information in the memoryless case, and the second term approaches the difference between $h(\Theta_1)$ and the entropy rate $h(\{\Theta_k\})$.

A lower bound on capacity can be demonstrated by considering IID inputs that achieve the memoryless channel capacity and that have large norms (with probability one), thus guaranteeing that from past inputs and outputs one would be able to reconstruct the past phases with arbitrarily high precision. With this approach we obtain the asymptotic expansion:

$$C = \frac{1}{2}\log\left(1+\frac{\mathrm{SNR}}{2}\right)+\log(2\pi)-h\big(\{\Theta_k\}\big)+o(1) \quad (26)$$

or

$$C = \frac{1}{2}\log\Big(1+2\pi^2 e^{-2h(\{\Theta_k\})}\,\mathrm{SNR}\Big)+o(1) \qquad (27)$$

where the $o(1)$ term tends to zero as the SNR tends to infinity.

Acknowledgments

The assistance of S. M. Moser and the careful reading of the manuscript by İ. E. Telatar are gratefully acknowledged.

References

[1] A. Lapidoth and S. Moser, "Convex-Programming Bounds on the Capacity of Flat-Fading Channels," *Proc. 2001 IEEE Int. Symposium on Inform. Theory*, p. 52, Washington DC, June 2001.

[2] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* Academic Press, 1981.

[3] R. G. Gallager, *Information Theory and Reliable Communication.* John Wiley & Sons, 1968.

[4] F. Topsœ, "An information theoretical identity and a problem involving capacity," *Studia Scientiarum Mathematicarum Hungarica*, vol. 2, pp. 291–292, 1967.

[5] A. Lapidoth and S.M. Moser, "Capacity bounds via duality with applications to multi-antenna systems on flat fading channels," preprint 2002.

[6] P. Vontobel and D. Arnold, "An upper bound on the capacity of channels with memory and constraint input," in *Proceedings of the 2001 IEEE Information Theory Workshop*, (Cairns, Australia), pp. 147–149, September 2 – 7, 2001 2001.

[7] A. Lapidoth, "Capacity-achieving distributions that escape to infinity and the fading number", MSRI Workshop on Information Theory, Berkeley, California February 25 – March 1, 2002.

[8] A. Lapidoth and S.M. Moser, "On the Fading Number of Multi-Antenna Systems," *Proc. 2001 IEEE Inform. Theory Workshop*, Cairns, Australia, Sept. 2001.