This work will be presented at the International Symposium on Information Theory and Its Applications
Honolulu, Hawaii, U.S.A., November 5–8, 2000

Capacity and Lattice-Strategies for Cancelling Known Interference

Uri Erez[†] , Shlomo Shamai[‡] and Ram Zamir[†]

[†]Department of Electrical Engineering-Systems
Tel Aviv University
Tel Aviv 69978 ISRAEL
Email: uri,zamir@eng.tau.ac.il

[‡]Department of Electrical Engineering
Technion
Haifa 3200 ISRAEL
Email: sshlomo@ee.technion.ac.il

Abstract

We derive capacity formulas and strategies for transmission over the the channel $Y = X + S + N$, $\frac{1}{n}E\|\mathbf{X}\|^2 \leq P_X$, where the interference S is a (strong) stochastic process or an arbitrarily varying sequence, known causally or with finite anticipation at the transmitter but not at the receiver. In the causal side information case, we show that strategies associated with entropy constrained quantizers provide lower and upper bounds on the capacity. At high SNR conditions, i.e., if N is weak relative to the power constraint P_X, these bounds coincide, the optimum strategies take the form of scalar lattice translations, and the capacity loss due to not having S at the receiver is shown to be exactly the "shaping gain" 0.254 bit. We also extend these ideas to any SNR and to non-causal side information, by incorporating "MMSE weighting", and by using k-dimensional lattices. For Gaussian N, the capacity loss of this scheme is upper bounded by $0.5\log(2\pi e G_k)$, where G_k is the normalized second moment of the lattice. These results provide an information theoretic framework for the study of common communication problems such as precoding for intersymbol interference channels.

1. Introduction

Consider a channel with side information (SI) at the transmitter as introduced by Shannon [1]. Let $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{S}$ denote the input, output and state alphabets of the channel, respectively, with transition probability $p(y|x,s)$ and with state probabilities given by $p(s)$. The transmitter has access to the side information. This problem classically divides into two categories, according to whether the encoder observes the state process *causally*, or *anticipates future states*. In the causal case, considered by Shannon [1], the encoder maps the message $w \in \{1, 2, ...2^{nR}\}$ into $\mathcal{X}^n$ using functions

$$x_i = f_i(w, s_1^i) \quad 1 \leq i \leq n \tag{1}$$

where $s_1^i = s_1, \ldots, s_i$ are the states up to time i. In the non-causal case, considered by Gelfand and Pinsker [2], the encoder observes the entire state sequence before generating the code sequence, thus

$$x_i = f_i(w, s_1^n) \quad 1 \leq i \leq n. \tag{2}$$

In both cases, the receiver decodes the message w from the whole received vector as $\hat{w} = g(y_1^n)$.

1.1. Shannon's capacity formula

For a general memoryless channel $p(y|x,s)$, with memoryless states, Shannon [1] showed that the capacity is equal to the regular capacity of an *associated* DMC. The input alphabet of the associated channel, denoted $\mathcal{T}$, is the set of all possible mappings

$$t : \mathcal{S} \longrightarrow \mathcal{X}$$

which we refer to as *strategies* or *strategy functions*. Therefore $|\mathcal{T}| = |\mathcal{X}|^{|\mathcal{S}|}$. The output y of the derived channel is related to the input t according to the transition probability

$$p(y|t) = \sum_s p(s) p\Big(y \Big| x = t(s), s\Big) \tag{3}$$

and also

$$p(y_1^n|t_1^n) = \prod_{i=1}^{n} p(y_i|t_i). \tag{4}$$

Thus, the capacity with side information at the transmitter is given by, [1], $C = \max_{p(t)} I(T;Y)$, where the maximization is taken over the distribution $p(t)$ of the random variable $T \in \mathcal{T}$. This result can readily be extended to the case where the alphabets $\mathcal{X}$,$\mathcal{Y}$,$\mathcal{S}$ are the real line and where the transmitter is subject to an average power constraint P_X to yield

$$C(P_X) = \max_{p(t): E\{t(s)^2\} \leq P_X} I(T;Y) \tag{5}$$

1.2. Additive noise with a power constraint

We consider the specific channel model

$$Y = X + S + N, \tag{6}$$

where S is known at the encoder, and N is a statistically independent r.v. In [3] Costa studied the case where S and N are Gaussian random variables, with variances P_S and P_N respectively, and where the encoder is *non causal* with unlimited anticipation, i.e., has knowledge of the entire state sequence $S_1, S_2 \ldots, S_n$ at the beginning of transmission. He showed that in this case the capacity is equal to $\frac{1}{2}\log(1 + P_X/P_N)$ where P_X is the power constraint of the transmitter. Therefore the noise S does not incur any loss in capacity in this case.

Finding explicit solutions for the causal side-information case seems a harder problem. Also, it is interesting to know how much is lost in capacity with respect to the non causal case, and with respect to the case where the side information is available also at the receiver side. Willems described in [4] a causal coding scheme in which the encoder uses some of its power to convert the noise S into a discrete r.v. whose support is an equally spaced lattice $(\ldots, -3A, -2A, -A, 0, A, 2A, 3A, \ldots)$ which effectively leaves us with a Gaussian noise channel. However, this scheme entails a power loss due to this "noise concentration" process, equal to $A^2/12$ when $A^2 \ll P_S$. In [5] Willems refers to schemes which circumvent the power loss of "noise concentration".

The paper presents new results regarding the channel (6), where N may or may not be Gaussian, and the side information is either causal or available with anticipation. Section 2 provides explicit capacity formulas and optimum strategies for causal side information. Section 3 proposes efficient schemes for side information known with finite anticipation. Section 4 discusses cancellation of arbitrary interference while Section 5 links the results to the intersymbol interference (ISI) precoding problem.

2. Results for Causal Encoding

We first treat the asymptotic case of strong i.i.d. interference, i.e., $P_S \gg P_X$. This greatly simplifies the treatment, while still incurs only a *finite penalty* relative to the case of $S \equiv 0$ which we shall quantify. To that end, assume S is of the form $S_\beta = \beta S_0$ where S_0 has a uniformly continuous p.d.f. Note that the limit $\beta \to \infty$ corresponds to ("infinitely") strong interference. Define $C(P_X) = \lim_{\beta\to\infty} C_\beta(P_X)$, where $C_\beta(P_X)$ is the capacity with causal side information (5) with $S = \beta S_0$. We begin with a causal encoder and first give a general formula for the capacity. Define

$$\tilde{C}_\beta(P_X) = h(S) - \inf_{f\in\mathcal{F}(P_X)} h(f(S) + N) \tag{7}$$

where $h(\cdot)$ denotes differential entropy, $\mathcal{F}(P_X) = \{f : E[f(S) - S]^2\} \le P_X$ and $S = \beta S_0$.

Theorem 1

$$C(P_X) = \textit{upper convex envelope of } \tilde{C}(P_X). \tag{8}$$

where $\tilde{C}(P_X) = \lim_{\beta\to\infty} \tilde{C}_\beta(P_X)$

The direct part of the theorem follows from the observation that $\tilde{C}(P_X)$ can be achieved by the following scheme. The transmitter uses an input alphabet that consists of strategies belonging to

$$\mathcal{A}_L = \{t : t(s) = t_0(s - c), |c| < L\}. \tag{9}$$

where $t_0(s) + s \in \mathcal{F}(P_X)$. That is, all strategies are a shift of a single strategy of power P_X, with the shift being restricted to some L. For sufficiently large L, a uniform distribution on $\mathcal{A}_L$ induces an output entropy $h(Y) = h(S) + o(1)$ where $o(1) \to 0$ as $\beta \to \infty$. Therefore, to achieve capacity, t_0 should be chosen so that it minimizes $h(Y|T) = h(t_0(S) + S + N)$. Note that $f(s)$ of (7) corresponds to $t_0(s) + s$.

2.1. A bound via entropy constrained quantization

From Theorem 1 we see that the capacity formula involves finding an optimal $f(\cdot)$, minimizing $h(f(S) + N)$ subject to the power constraint, $f \in \mathcal{F}(P_X)$. The following theorem links this problem to that of finding the optimal entropy constrained quantizer of S. Let

$$H_{min}(S, D) = \inf H(Q(S))$$

denote the minimum entropy in quantizing S with mean squared distortion D, where $H(\cdot)$ denotes regular entropy and the infimum is over all quantizers Q satisfying $E[Q(S) - S]^2 \le D$.

Lemma 1 *Suppose that $N = \gamma N_0$ and that N_0 has a density. Then*

$$\tilde{C}_\beta(P_X) \ge h(S) - H_{min}(S, P_X) - h(N) \tag{10}$$

On the other hand, for any $a > 0$

$$\tilde{C}_\beta(P_X) \le h(S) - H_{min}(S, [\sqrt{(P_X)} + a/2]^2) - h(N) + I(N; N + Z_a) \tag{11}$$

where the r.v. Z_a is independent of N and is uniformly distributed over $(-a/2, +a/2)$.

Clearly, as the variance of N goes to zero, $I(N;N+Z_a)\to 0$ for any $a>0$, implying

$$\tilde{C}_\beta(P_X)=h(S)-H_{min}(S,P_X)-h(N)+\epsilon(\gamma) \quad (12)$$

where $\epsilon(\gamma)\to 0$ as $\gamma\to 0$.

2.2. High SNR case

We now restrict attention to the case $P_N \ll P_X \ll P_S$. By the well known result of Gish and Pierce from high resolution quantization theory [6, 7], the optimum entropy constrained quantizer is asymptotically uniform, implying that $H_{min}(S,D)=h(S)-\log\sqrt{12D}$ for $D/P_S\to 0$. Hence, by Lemma 1 the capacity with causal side information is given asymptotically by

$$C(P_X)=\frac{1}{2}\log 12P_X-h(N)+o(1), \quad (13)$$

where $\lim_{\gamma\to 0}\lim_{\beta\to\infty}o(1)=0$. Therefore the asymptotic rate loss with respect to the no-interference case $P_S=0$ (or equivalently, to having S also at the receiver), is equal to the "shaping gain", $\frac{1}{2}\log\frac{2\pi e}{12}\approx 0.254$ bit. Note that this result holds for any noise N.

Using a uniform quantizer as $f(\cdot)$ in (7), the direct part of Theorem 1 indicates an explicit encoding scheme. We choose $t_0(s)=Q_\Delta(s)-s$ where $Q_\Delta(\cdot)$ is a uniform scalar quantizer with step size $\Delta=\sqrt{12P_X}$. We now apply a uniform distribution upon the class of strategies which are shifts of t_0

$$\begin{aligned} t_v &= Q_\Delta(s-v)+v-s \\ &= [v-s] \bmod \Delta \quad (14)\end{aligned}$$

where due to periodic nature of t_0, the shifts may be limited to the interval $v\in(-\frac{\Delta}{2},+\frac{\Delta}{2})$. Therefore, $X=Q_\Delta(S-V)+V-S$. It is known , [8], that $X\sim\mathcal{U}(-\frac{\Delta}{2},+\frac{\Delta}{2})$ as well, so $EX^2=P_X$ as desired. Using this input distribution, the decoder receives

$$Y=X+S+N=Q_\Delta(S-V)+V+N \quad (15)$$

Invoking a modulo operation to the basic interval $(-\frac{\Delta}{2},+\frac{\Delta}{2}]$ (which is an information lossless operation when S is strong), the resulting output is $Y_{mod}\approx V+N$, where the approximation follows since $P_N\ll P_X$. This gives rise to the rate

$$\begin{aligned} I(V;Y_{mod}) &\approx h(V)-h(N) \quad (16)\\ &= \log\Delta-h(N) \quad (17)\\ &= \frac{1}{2}\log 12P_X-h(N). \quad (18)\end{aligned}$$

Hence, in light of (13) this scheme is asymptotically optimal.

The idea behind the scheme in (14) resembles that of the technique for information embedding in [9].

2.3. General SNR case

Dropping the high SNR assumption, we propose an efficient strategy based on an "inflated" lattice. This scheme uses a scaling coefficient, $0<\alpha\le 1$, effectively producing at the receiver end a lattice with cells of power P_X/α^2, at the expense of adding an additional noise component with variance $\left(\frac{1-\alpha}{\alpha}\right)^2 P_X$. The scaling factor should be chosen so as to maximize the corresponding mutual information. A detailed exposition of the scheme is presented in the next section, taking $k=1$. A tentative choice of $\alpha=\frac{P_x}{P_x+P_N}$ is close to optimal for Gaussian N, yielding in the limit as $P_X/P_N\to 0$ a rate of $0.2P_X/P_N$ nats, indicating that the rate loss due to causality is bounded by 7dB.

3. Results for Noncausal Side Information

We can link our result to Costa's by allowing the encoder to anticipate k states ahead. We generalize the scheme above by employing an (optimal) lattice vector quantizer $Q_\Lambda(\cdot)$ instead of the scalar one of (15). That is, we now apply a uniform distribution over the family of strategies

$$\begin{aligned} \mathbf{t_v} &= Q_\Lambda(\alpha\mathbf{s}-\mathbf{v})+\mathbf{v}-\alpha\mathbf{s} \\ &= [\mathbf{v}-\alpha\mathbf{s}] \bmod \Lambda \quad (19)\end{aligned}$$

where all vectors are k-dimensional and where $\mathbf{v}$ is a member of a basic Voronoi region of the (optimal) lattice Λ having a second moment P_X. Upon reception, $\mathbf{Y}$ is multiplied by α and then reduced modulo Λ resulting in

$$\mathbf{Y}'=\alpha\mathbf{Y} \bmod \Lambda \quad (20)$$

The resulting channel is a modulo-Λ additive noise channel described by the following lemma:

Lemma 2 (Inflated lattice lemma) *The channel defined by (6),(19) and (20) satisfies*

$$\mathbf{Y}'=\mathbf{V}+\mathbf{N}'. \quad (21)$$

with

$$\mathbf{N}'=\Big[(1-\alpha)\mathbf{U}+\alpha\mathbf{N}\Big]. \quad (22)$$

where $\mathbf{U}$ *is a r.v. distributed uniformly over the Voronoi region of* Λ *and addition is modulo-*Λ*.*

The corresponding rate is

$$\begin{aligned} \mathrm{Rate}(k) &= I(\mathbf{V};\mathbf{Y}) \quad (23)\\ &= h(\mathbf{Y}')-h(\mathbf{N}') \quad (24)\\ &= \frac{1}{2}\log(P_X/G_k)-h(\mathbf{N}'). \quad (25)\end{aligned}$$

For optimum lattices $G_k \to 1/2\pi e$ as $k \to \infty$, and we also have that the noise $\mathbf{U}$ tends to a Gaussian distribution . Choosing $\alpha = \frac{P_X}{P_X+P_N}$ we have

$$\begin{aligned} \mathrm{Var}(N') &= (1-\alpha)^2\mathrm{Var}(\mathbf{U}) + \alpha^2\mathrm{Var}(\mathbf{N}) \quad (26) \\ &= \frac{P_N P_X}{P_N + P_X} \end{aligned}$$

When the noise $\mathbf{N}$ is Gaussian, $\mathbf{N}'$ is also Gaussian (as $k \to \infty$) and we have

$$\begin{aligned} \mathrm{Rate}(k=\infty) &= \frac{1}{2}\log(2\pi e P_X) - \log\left(2\pi e \frac{P_N P_X}{P_N + P_X}\right) \\ &= \frac{1}{2}\log(1 + \frac{P_X}{P_N}). \end{aligned} \quad (27)$$

Therefore for Gaussian noise as $k \to \infty$ there is no loss at all. This coincides with the results of [3]. Note that this result holds at any SNR.

4. Arbitrarily Varying Interference

We note that *any* interference sequence known at the transmitter can be "forced" to comply with the strong memoryless interference condition by using dithering techniques. Assume that the transmitter and receiver have access to a common (pseudo) random source. Then the tranmitter may add a very strong i.i.d. dither signal S_i' to S_i, resulting in $S_i'' = S_i + S_i'$. The encoding scheme, i.e. (19), is performed as before with s_i being subsituted with s_i''. The receiver adds this same dither upon recetion, i.e., $Y_i = X_i + S_i + S_i' + N_i$ and proceeds as before. Therefore, $S_i + S_i'$ satisfies the required conditions, so the results of (18) and (27) hold for any interference sequence, even an arbitrarily varying one. In particular, for Gaussian N, the effect of any interference known at the transmitter non causally can be cancelled completely, with *no power loss*.

5. Application to Precoding

We now briefly hint at the connection of the above results to the problem of precoding for ISI channels. Assume a discrete time linear channel model (corresponding to a WMF front end)

$$Y_i = X_i + \sum_{k=1}^{L} h_k X_{i-k} + N_i \quad (28)$$

where all quantities are real valued and N_i is i.i.d. Gaussian. Decision feedback is a well known technique for cancelling the effect of the ISI at the receiver side.

In precoding (see, e.g., [10]), one regards the ISI term, $\sum_{k=1}^{L} h_k X_{i-k}$ as interference that is known to the encoder but ignored by the receiver. This fits our channel model of side information (the ISI term) at the transmitter side, but for the fact that the ISI term is message dependent. However, this flaw can be remedied by adding a strong ("infinite") dither signal as done in Section 4, thereby producing a state sequence $S_i = \mathrm{ISI}_i + \mathrm{dither}_i$, which is message independent. This being done, performance characterization and bounds can be drawn from the previous sections. One observation is that confining the ISI cancellation to be causal, the optimal encoding method at high SNR is through using as "input symbols" (strategies) uniform quantizers of the state sequence. This is exactly equivalent to Tomlinson precoding [11]. This also proves that the loss under these constraints is exactly the shaping gain. The treatment can also be extended to more general ISI channel models than (28) such as the MMSE-DFE front end model and to non causal precoding along the lines of Section 3.

References

[1] C. E. Shannon. Channels with side information at the transmitter. *IBM Journal Research and Development*, 2:289–293, Oct. 1958.

[2] S.I. Gelfand and M. S. Pinsker. Coding for channel with random parameters. *Problemy Pered. Inform. (Problems of Inform. Trans.)*, 9, No. 1:19–31, 1980.

[3] M.H.M. Costa. Writing on dirty paper. *IEEE Trans. Information Theory*, IT-29:439–441, May 1983.

[4] F.M.J.Willems. On gaussian channels with side information at the transmitter. In *Proc. of the Ninth Symposium on Information Theory in the Benelux*, Enschede, The Netherlands, 1988.

[5] F.M.J.Willems. Private communication.

[6] H. Gish and N. J. Pierce. Asymptotically efficient quantization. *IEEE Trans. Information Theory*, IT-14:676–683, Sept. 1968.

[7] T. Linder. Private communication.

[8] R. Zamir and M. Feder. On lattice quantization noise. *IEEE Trans. Information Theory*, pages 1152–1159, July 1996.

[9] B. Chen and G. W. Wornell. Quantization index modulation: A class of provably good methods for digital watermarking and information embedding. In *Proceedings of ISIT 2000, Sorrento, Italy*, page 46, June 2000.

[10] M. V. Eyuboglu and G. D. Forney Jr. Trellis precoding:combined coding, precoding and shaping for intersymbol interference channels. *IEEE Trans. Information Theory*, IT-38:301–314, March,1992.

[11] M. Tomlinson. New automatic equalizer employing modulo arithmetic. *Electronic Lett.*, vol. 7:138–139, Mar. 1971.