optimal estimator cannot be decomposed into projection followed by nonlinear processing.

This phenomena can be explained by the tendency towards Gaussianity of the sum of independent random variables. Projection, which is a noninvertible linear transformation, makes the residual noise *more Gaussian* and thus less favorable for estimation. A similar phenomena causes *increase of entropy* after noninvertible filtering [9], [16].

In [14], we suggested another way to interpret (29), namely, as an accuracy–quantity tradeoff relation. Notice that $1/\alpha^2$ represents the accuracy (or the resolution) of the measurements. Thus without prefiltering, keeping the quantity/accuracy product n/α^2 fixed keeps the FI constant. The same is true for a Gaussian noise even *after* (appropriate) prefiltering, but not true when the noise is not Gaussian. Thus if prefiltering (projection) is used prior to estimation in the presence of a non-Gaussian noise, *it is better to take few accurate measurements than many noisy ones.*

Acknowledgment

My joint work with Meir Feder on information-theoretic inequalities formed the basis for this correspondence. I also wish to thank Toby Berger, Jean-Francois Cardoso, Hagit Messer, Nadav Shulman, Yossef Steinberg, Tony Weiss, and Arie Yeredor for helpful discussions.

References

[1] A. R. Barron, "Entropy and the central limit theorem," *Ann. Prob.*, vol. 14, no. 1, pp. 336–342, 1986.

[2] A. Bartov and H. Messer, "Lower bound on the achievable DSP performance for localizing step-like continuous signals in noise," *IEEE Trans. Signal Processing*, to be published.

[3] N. M. Blachman, "The convolution inequality for entropy powers," *IEEE Trans. Inform. Theory*, vol. IT-11, pp. 267–271, 1965.

[4] R. E. Blahut, *Principles and Practice of Information Theory.* Reading, MA: Addison Wesley, 1987.

[5] J.-F. Cardoso, "Blind signal separation: A review," *Proc. IEEE*, submitted for publication.

[6] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: Wiley, 1991.

[7] A. Dembo, "Information inequalities and uncertainty principles," Dept. Statistics, Stanford Univ., Tech. Rep. 75, Stanford, CA, July 1990.

[8] A. Dembo, T. M. Cover, and J. A. Thomas, "Information theoretic inequalities," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1501–1518, Nov. 1991.

[9] D. Donoho, "On minimum entropy deconvolution," *Applied Time Series Analysis II.* NY: Academic, 1981, pp. 565–608.

[10] H. L. Van Trees, *Detection Estimation and Modulation theory (Part I).* New York: Wiley, 1968.

[11] A. J. Weiss and B. Friedlander, "Preprocessing for direction finding with minimal variance degradation," *IEEE Trans. Signal Processing*, vol. 42, pp. 1478–1485, June 1994.

[12] R. Zamir, "Universal encoding of signals by entropy coded dithered quantization," Ph.D. dissertation, Tel-Aviv Univ., Tel–Aviv, Israel, 1994.

[13] ——, "A necessary and sufficient condition for equality in the matrix Fisher-information-inequality," Tech. Rep., Tel Aviv Univ., Dept. Elec. Eng.–Syst., 1997.

[14] ——, "It is better to take few accurate measurements than many noisy ones," in *Proc. Information Theory Workshop on Information Theory, Multiple Access and Queueing* (St. Louis, MO, Apr. 1995), p. 35.

[15] R. Zamir and M. Feder, "A generalization of information theoretic inequalities to linear transformations of independent vector," in *Proc. 6th Joint Swedish–Russian Int. Workshop on Information Theory* (Molle, Sweden, Aug. 1993), pp. 254–258.

[16] ——, "A generalization of the Entropy Power Inequality with applications," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1723–1727, Sept. 1993.

Zero-Error Capacity for Models with Memory and the Enlightened Dictator Channel

Rudolf Ahlswede, Ning Cai, and Zhen Zhang, *Senior Member, IEEE*

***Abstract*—We present a general class of zero-error capacity problems with memory covering known cases such as coding for error correction and many new cases. This class can be incorporated into a model of channels with memory, which thus are shown to give a unification of a multitude of seemingly very different coding problems. In this correspondence we analyze a seemingly basic channel in this class.**

***Index Terms*—Finite-state channels, independence number, memory, 0-error capacity.**

I. Introduction

A General 0-Error Problem with Memory

For the space $\mathcal{Z}^n$ of words of length n over alphabet $\mathcal{Z}$, there are several interesting graphs $\mathcal{G} = (\mathcal{Z}^n, \mathcal{E}_n)$ with vertex set $\mathcal{Z}^n$ and an edge set $\mathcal{E}_n$ reflecting string properties.

Examples are, the strong graph product (Shannon's product graph) and the case $\mathcal{Z} = \{0, 1\}$ with $(x^n, x'^n) \in \mathcal{E}$ if and only if (iff) for no two components s, t $x_s = 1 \neq x'_s$ and $x_t = 0 \neq x'_t$.

The product space structure makes it particularly interesting to investigate $\alpha(\mathcal{G}_n)$, the maximal size of cocliques, as a function of n. Then the coclique of the graph in the first example is Shannon's well-known zero-error code and the coclique of the graph in the second example is the well-known Sperner system or antichain (c.f., e.g., [6, Ch. 1]). We propose here a quite general class of such problems, which we term "0-error ∞-memory capacity problems," because they generalize Shannon's well-known zero-error capacity problems and concentrate on a new aspect, namely, memory. Those problems arose for instance in [2].

Definition We call any pair of words from $\mathcal{Z}^\ell$ a separator and any set $\mathcal{S} \subset (\mathcal{Z}^\ell)^2$ of pairs of words of length ℓ a set of separators.

For any $n \geq \ell$ we consider the associated graph $\mathcal{G}^n_{\mathcal{S}} = (\mathcal{Z}^n, \mathcal{E}(\mathcal{S})_n)$, where $(x^n, x'^n) \in \mathcal{E}(\mathcal{S})_n$ iff for no $(s^\ell, s'^\ell) \in \mathcal{S}$ there is an index set $I = \{i_1, \cdots, i_\ell\} \subset \{1, \cdots, n\}$ with $x_{i_j} = s_j$, $x'_{i_j} = s'_j$ $(i_1 < i_2 < \cdots < i_\ell)$.

In the examples above $\mathcal{S}$ is symmetric, that is, $(s^\ell, s'^\ell) \in \mathcal{S}$ implies $(s'^\ell, s^\ell) \in \mathcal{S}$. Here the graphs can be viewed as undirected graphs. In the sequel we assume that $\mathcal{S}$ is symmetric. Thus $\mathcal{S}$ can be viewed as a set of unordered pairs of subsequences.

This covers also t-error correcting codes for $\mathcal{S} = \{(0^{2t+1}, 1^{2t+1})\}$.

II. Consecutive Separating Pairs

Another associated graph $\mathcal{G}^*_{\mathcal{S}} = (\mathcal{Z}^n, \mathcal{E}^*(\mathcal{S})_n)$ is obtained by limiting I in the previous definition to intervals in $\{1, 2, \cdots, n\}$. $\mathcal{S}$ plays here the role of a set of *consecutive* separating pairs of words of length ℓ. Here the problem is to find a maximal $\mathcal{C} \subset \mathcal{Z}^n$ such that for all $c^n, c'^n \in \mathcal{C}$ there is an $\{\alpha, \beta\} \in \mathcal{S}$ and an $i \in \{1, 2, \cdots, n - \ell + 1\}$ such that

$$\{(c_i, c_{i+1}, \cdots, c_{i+\ell-1}), (c'_i, c'_{i+1}, \cdots, c'_{i+\ell-1})\} = \{\alpha, \beta\}.$$

Manuscript received December 12, 1996; revised November 10, 1997.

R. Ahlswede and N. Cai are with Fakultät für Mathematik, Universität Bielefeld, 33501 Bielefeld, Germany.

Z. Zhang is with the Department of Electrical Engineering–Systems, University of Southern California, Los Angeles, CA 90089 USA.

Publisher Item Identifier S 0018-9448(98)02370-0.

However, this seemingly different case is a special case of the former: replace the pairs of $\mathcal{S}$ by all pairs of length n by adding a total of $n-\ell$ letters at the beginning and/or the end.

Already by no means easy problems arise in the cases $\mathcal{Z}=\{0,1\}$ and $|\mathcal{S}|=1$. The case $\ell=1$ being trivial let us *start with* $\ell=2$. Here $\mathcal{S}$ contains one pair of 2-length binary sequences. There are totally $\binom{4}{2}=6$ such $\mathcal{S}$'s. However, by interchanging "0" and "1" and "reading" codewords backwards, there are only three nonequivalent cases: $\mathcal{S}_1=\{(00),(11)\}$, $\mathcal{S}_2=\{(01),(10)\}$, and $\mathcal{S}_3=\{(00),(01)\}$.

Denote the maximal code sizes in $\mathcal{Z}^n$ by $\alpha_n(\mathcal{S}_i)$, $1\le i\le 3$.

The codes

$$\mathcal{C}_1=\begin{cases}\{00,11\}^{\frac{n}{2}}, & \text{for } n \text{ even}\\ \{00,11\}^{\frac{n-1}{2}}\times\{0\}, & \text{for } n \text{ odd}\end{cases}$$

$$\mathcal{C}_2=\begin{cases}\{01,10\}^{\frac{n}{2}}, & \text{for } n \text{ even}\\ \{01,10\}^{\frac{n-1}{2}}\times\{0\}, & \text{for } n \text{ odd}\end{cases}$$

show that

$$\alpha_n(\mathcal{S}_i)\ge 2^{\lfloor\frac{n}{2}\rfloor},\qquad \text{for } i=1,2. \tag{2.1}$$

On the other hand, we readily verify that this bound is tight. Indeed, for a code $\mathcal{C}$ use the partition

$$\mathcal{C}=\mathcal{C}'\,\dot{\cup}\,\mathcal{C}'' \tag{2.2}$$

where $\mathcal{C}'$ contains exactly those words of $\mathcal{C}$, which start with 00 or 10.

Now, if $\mathcal{C}$ is an n-length code for $\mathcal{S}_1$ (respectively, $\mathcal{S}_2$), then the codes obtained by deleting the first two bits of $\mathcal{C}'$ and $\mathcal{C}''$ are $(n-2)$-length codes for $\mathcal{S}_1$ (respectively, $\mathcal{S}_2$). So $|\mathcal{C}|\le 2\alpha_{n-2}(\mathcal{S}_1)$ (respectively, $2\alpha_{n-2}(\mathcal{S}_2)$) holds and thus

$$\alpha_n(\mathcal{S}_i)=2^{\lfloor\frac{n}{2}\rfloor},\qquad \text{for } i=1,2. \tag{2.3}$$

The case $\mathcal{S}_3$ is already a little bit more complicated. Here we observe that for any code $\mathcal{C}$ those codewords, which start with 1, can be modified by exchanging this 1 by a 0, because this 1 is useless for the separation.

Therefore, we can assume that all words in $\mathcal{C}$ start with a 0. We call such a code canonical. Let now $\mathcal{C}$ have length n and let us partition it into the subcodes $\mathcal{C}'$, with words starting with 00, and $\mathcal{C}''$, with words starting with 01. Now $\mathcal{C}'_{n-1}$, obtained from $\mathcal{C}'$ by omitting the first 0 in all words, is an $(n-1)$-length code for $\mathcal{S}_3$ and $\mathcal{C}''_{n-2}$, obtained from $\mathcal{C}''$ by omitting the first two bits 01 in all words, is an $(n-2)$-length code for $\mathcal{S}_3$.

Furthermore, $\{0\}\times\mathcal{C}^*\cup\{01\}\times\mathcal{C}^{**}$ is a code for $\mathcal{S}_3$, if $\mathcal{C}^*$ is an $(n-1)$-length canonical code for $\mathcal{S}_3$ and $\mathcal{C}^{**}$ is $(n-2)$-length code for $\mathcal{S}_3$. Thus

$$\alpha_n(\mathcal{S}_3)=\alpha_{n-1}(\mathcal{S}_3)+\alpha_{n-2}(\mathcal{S}_3) \tag{2.4}$$

$(\alpha_2(\mathcal{S}_3)=1, \alpha_3(\mathcal{S}_3)=2)$.

This equation describes the well-known Fibonacci numbers.

Finally, for $\ell\ge 3$ we have no conclusive results. In the next section we settle a seemingly interesting case with $|\mathcal{S}|\ge 1$.

III. The Enlightened Dictator Channel

The problem discussed in Section II can also be viewed as a kind of zero-error capacity problem of finite memory channels.

Indeed, with a set $\mathcal{S}$ of unordered pairs of words of $\mathcal{X}^\ell$ we can associate a stationary ℓ-memory channel W, which satisfies

$$\sum_{y\in\mathcal{Y}}W(y\mid\alpha)W(y\mid\beta)=0 \text{ exactly if } \{\alpha,\beta\}\in\mathcal{S}$$

and is a finite-state channel in the sense of [3], which is discussed also in [4, Secs. 3.5 and 3.6] and in [5, Sec. 4.6].

Just let $\Gamma=\mathcal{X}^{\ell-1}$ be the state space with transition function $f\colon \Gamma\times\mathcal{X}\to\Gamma$ defined by

$$f(x_1,\cdots,x_{\ell-1},x)=(x_2,x_3,\cdots,x_{\ell-1},x).$$

This is the most straightforward generalization of Shannon's zero-error capacity problem for a memoryless channel. Here already for binary alphabets interesting problems arise.

Our investigation has led to a very interesting type of finite memory channel, which we call "enlightened dictator channel." Clearly, a dictator always follows his own opinion. However, we speak of an enlightened dictator, if he responds to the unanimous vote of the people against his opinion to the degree, that in this case he reaches a decision by coin tossing.

M_k $(k\ge 3)$ is a binary input and binary output channel with $(k-1)$-memory with transmission probabilities

$$\begin{aligned}\Pr(y_t\mid x_t,x_{t-1},\cdots,x_1)&\\ =\Pr(y_t\mid x_t,\cdots,x_{t-k+1})&\\ =\begin{cases}\frac{1}{2}, & \text{if } (\alpha)\ x_t\ne x_{t-1}=\cdots=x_{t-k+1} \text{ for } t\ge k\\ 1, & \text{if } y_t=x_t \text{ and not } (\alpha).\end{cases}&\end{aligned} \tag{3.1}$$

For $t\le k-1$ (α) cannot occur, so we are in the second case.

Denote the zero-error capacity of the channel by $C_0(M_k)$.

Theorem

$$C_0(M_k)=\log\lambda^*(k) \tag{3.2}$$

where $\lambda^*(k)$ is the largest real root of the equation

$$\lambda^{k-1}-\sum_{j=0}^{k-2}\lambda^j=0. \tag{3.3}$$

Proof:

Direct Part: By definition M_k can transmit error-free at the first $k-1$ bits. Define code $\mathcal{C}_n$ recursively and start with $\mathcal{C}_{k-1}=\{0,1\}^{k-1}$.

Suppose we have an $(n-1)$-length code $\mathcal{C}_{n-1}$. For the definition of the n-length code $\mathcal{C}_n$ we make use of the function

$$t\colon \mathcal{X}^m\to\mathbb{N}$$

where $t(x^m)$ equals the maximal number ℓ with

$$x_{m-\ell+1}=x_{m-\ell+2}=\cdots=x_m.$$

We call $t(x^m)$ the length of the tail of x^m.

Consider the subcodes

$$\mathcal{C}_{n-1,i}=\{x^{n-1}\in\mathcal{C}_{n-1}:t(x^{n-1})=i\} \tag{3.4}$$

and define

$$\begin{aligned}\mathcal{C}_n=&\left\{x^{n-1}x_n:x^{n-1}\in\bigcup_{i=1}^{k-2}\mathcal{C}_{n-1,i};x_n=0,1\right\}\\ &\cup\{x^{n-1}x_n:x^{n-1}\in\mathcal{C}_{n-1,k-1};x_n=1-x_{n-1}\}.\end{aligned} \tag{3.5}$$

Clearly, this is a zero-error code for M_k with length n.

Moreover, by this recursive construction

$$\mathcal{C}_{n,j}=\phi,\qquad \text{for } j\ge k, n\ge k-1$$

and

$$|\mathcal{C}_{n,1}|=\sum_{i=1}^{k-1}|\mathcal{C}_{n-1,i}| \tag{3.6}$$

$$|\mathcal{C}_{n,i}|=|\mathcal{C}_{n-1,i-1}|,\qquad \text{for } i=1,2,\cdots,k-1. \tag{3.7}$$

In terms of the $(k-1)\times(k-1)$-matrix

$$D_k = \begin{pmatrix} 1 & 1 & 0 & \cdot & \cdots & 0 \\ 1 & 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & & & \vdots \\ 1 & 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \tag{3.8}$$

for $\vec{T}_m = (|\mathcal{C}_{m,1}|, \cdots, |\mathcal{C}_{m,k-1}|)$

$$\vec{T}_n = \vec{T}_{n-1} \cdot D_k. \tag{3.9}$$

Therefore,

$$\lim_{n\to\infty} \frac{1}{n} \log |\mathcal{C}_n| = \lambda^*(k) \tag{3.10}$$

the largest real eigenvalue of D_k.

Converse Part: We say a code has prefix x^ℓ, if all its codewords have the common prefix x^ℓ. Denote by $\mathcal{A}_n(x^\ell)$ any optimal (maximal size) n-length zero-error code for M_k with prefix x^ℓ. Obviously, $|\mathcal{A}_n(x^\ell)|$ only depends on $j \triangleq t(x^\ell)$ and $m \triangleq n-\ell$. Therefore, we can write it as $A(j,m)$. Any zero-error code with prefix x^ℓ can be partitioned into two subcodes with prefix $x^\ell 0$ or $x^\ell 1$, respectively. On the other hand, for all x^ℓ with $t(x^\ell) \le k-2$, $\mathcal{A}_n(x^\ell 0) \cup \mathcal{A}_n(x^\ell 1)$ is a zero-error code, because one can seperate any $c^n \in \mathcal{A}_n(x^\ell 0)$ and $c'^n \in \mathcal{A}_n(x^\ell 1)$ at the $\ell + 1$st bit.

Therefore, for $t(x^\ell) \le k-2$, we can write

$$\begin{aligned} &\mathcal{A}_n(x^\ell) = \mathcal{A}_n(x^\ell 0) \cup \mathcal{A}_n(x^\ell 1) \\ &A(j,m) = A(1,m-1) + A(j+1,m-1), \qquad \text{for } j \le k-2. \end{aligned} \tag{3.11}$$

By the definition of M_k in (3.1), when we send a binary sequence $x^n = (x_1, \cdots, x_n)$ over M_k, for $t \ge k$ the tth bit of the output sequence is not uniquely determined. It may be 0 or 1, if x_t agrees with none of its $k-1$ predecessors. One cannot separate x^n from any n-length binary sequence at the tth bit, when this occurs, and we say that the tth bit of x^n is ineffective. It is called effective, if it is not ineffective.

We observe now that for all $x^n \in \mathcal{A}_n(0^{k-1})$ we can change x_t to $1-x_t$ and get a new zero-error code, if

a) x_t is ineffective or, for all $x'^n \in \mathcal{A}_n(0^{k-1})$ with $x'_t = 1-x_t$, x'_t is ineffective

and

b) none of the $k-1$ succesive bits of x_t turns from being effective to being ineffective.

Notice now that the 1 at the kth bit of any codeword in $\mathcal{A}_n(0^{k-1})$ is ineffective and one can change the kth bit of a codeword x^n from 0 to 1 without violating b), unless $n \ge 2k-1$ and x^n has the $(2k-1)$-length prefix $0^k1^{k-2}0$. However, in the latter case, the 1 at the $k+1$st bit is ineffective and b) is not violated, if we change the prefix $0^k1^{k-2}0$ of x^n to $0^{k-1}101^{k-3}0$ (here, when $k=3$, 1^0 is understood as an empty sequence as usual).

In this way we obtain from $\mathcal{A}_n(0^{k-1})$ a new zero-error code of the same size, but only with a k-prefix $0^{k-1}1$. Therefore,

$$\begin{aligned} A(k-1, n-k+1) &= |\mathcal{A}_n(0^{k-1})| \\ &\le |\mathcal{A}_n(0^{k-1}1)| = A(1, n-k). \end{aligned} \tag{3.12}$$

The opposite inequality is obvious, because $\mathcal{A}_n(0^{k-1}1)$ is a code with prefix 0^{k-1} as well, and thus

$$A(k-1,m) = A(1,m-1), \qquad \text{for } m \ge 1. \tag{3.13}$$

Let now $\vec{A}_m = (A(1,m), \cdots, A(k-1,m))$, then it immediately follows from (3.8), (3.11), and (3.13) that in terms of the transpose D_k^τ of D_k

$$\vec{A}_m = \vec{A}_{m-1} D_k^\tau. \tag{3.14}$$

Substituting the initial condition $\vec{A}_0 = 1^{k-1}$ in (3.14), we get

$$\vec{A}_m = 1^{k-1}(D_k^\tau)^m. \tag{3.15}$$

Finally, since there are exactly $|\mathcal{C}_{k-1,i}|$ (defined in (3.4)) $(k-1)$-length sequences with tail length i, considering the decomposition of an optimal code according to the $(k-1)$-length prefices of its codewords, (3.15) implies that the n-length optimal code has size

$$N(n, M_k) = 1^{k-1}(D_k^\tau)^{n-k+1}\vec{T}_{k-1}^\tau.$$

(Actually we thus not only proved the converse, but also the direct part again.) □

Acknowledgment

The authors wish to thank one of the referees for the comment: "It is interesting to note that the optimal code for the enlightened dictator channel can be viewed as sequences by walks on finite state transition diagram corresponding to a run-length-limited channel [7]. Hence the results on coding for run-length-limited channels suggest coding schemes (message-codeword maps) for the enlightened dictator channel."

Indeed, our optimal code for the enlightened dictator channel is a code with run-length less than k and any code with this constraint for the run-length is a zero-error dictator channel. However, a zero-error code for the enlightened dictator channel does not necessarily meet this constraint. (The simplest example is the code $\{(0,0,\cdots,0),(1,1,\cdots,1)\}$.) Our theorem shows that the optimal rate can be achieved by such a code.

References

[1] C. E. Shannon, "The zero-error capacity of a noisy channel," *IRE Trans. Inform. Theory*, vol. IT-3, pp. 3–15, 1956.

[2] R. Ahlswede and A. Kaspi, "Optimal coding strategies for certain permuting channels," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 310–314, May 1987.

[3] D. L. Blackwell, L. Breiman, and A. J. Thomasian, "Proof of Shannon's transmission theorem for finite-state indecomposable channels," *Ann. Math. Statist.*, vol. 29, pp. 1209–1220, 1958.

[4] J. Wolfowitz, *Coding Theorems of Information Theory*, 3rd ed. Berlin, Germany: Springer-Verlag, 1978.

[5] R. G. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.

[6] I. Anderson, *Combinatorics of Finite Sets.* Oxford, U.K.: Clarendon, 1987.

[7] B. Marcus, P. Siegel, and J. Wolf, "Finite-state modulation codes for data storage," *IEEE J. Select. Areas Commun.*, vol. 10, no. 1, 1992.