Search with noisy and delayed responses

R. Ahlswede and N. Cai

Abstract

It is well–known that search problems with a stochastic response matrix acting independently for the questions can be equivalently formlated as transmission problems for a discrete memoryless channel (DMC) with feedback.

This is explained in Chapter 3 of the book Search Problems by R. Ahlswede and I. Wegener (Wiley 1987, translation of Suchprobleme, Teubner 1979).

There also Ahlswede's coding scheme for the DMC and also for the arbitrarily varying channel (AVC) achieving the capacities are described. The latter can be viewed as a robust model for search.

In this paper we analyse this robust model with a **time delay** for the noiseless feedback. In the terminology of search this means that the answers are given with delay.

We determine the (asymptotically) optimal performances, that is, **find the capacities**, for the cases where the **delay is constant** and **linear in the blocklength**. Finally we also give the corresponding results for the DMC **with zero–error probability**.

Keywords: Search, noisy responses, liers, delay, feedback, list codes, 0–error capacity

1 Introduction

Delay is an essential property in human interactions and especially also in engineering systems for instance those with control or communication aspects. There have been already in the 70ties studies on delay and overflow in data compression schemes (e.g. Wyner [21], Jelinek [19]). Recently searching with delayed answers was considered by Cicalese, Gargano, and Vaccaro ([11], [12]) in the combinatorial model dealing with lies as considered by Renyi, Berlekamp, and Ulam (see Deppe [16]).

It is well–known that feedback (even without delay) does not increase the capacity of a DMC. This was first proved by Shannon [20], who also found a formlator the sero error capacity of the DMC. This is a special case of the result by Ahlswede [2] on the AVC for maximal probability of error, which was completed by Ahlswede and Cai [7] omitting a convexity assumption and providing a condition for positivity of the capacity (a "trichotomy"). Here feedback (without delay) increases the capacities. This is also true of the easier case of average probability of error, which can be found in [4], [8].

Our first result (in Section 4) concerns the situation where the delayed feedback has a delay time upperbounded by a constant d. With a simple coding scheme for delayed feedback we show that in this case the capacities of all memoryless channels with non–delayed feedback can be achieved. Actually it is relevant here that we can do time sharing.

Next when the delay time increases linearly with the length of codes we obtain characterizations of the zero–error capacity of a DMC (Section 5) and the average–error capacity of an AVC (Section 6).

Finally we draw attention to future study of identification codes for the DMC, for which Ahlswede and Dueck [9] found the capacity in case of delayed feedback. It exceeds the capacity in the absence of feedback.

Furthermore we shall investigate the AVC for maximal probability of error, where feedback increases capacity.

In Section 2 we introduce the necessary notation and definitions. Results, on which this work is based are stated in Section 3.

2 Notation and Definitions

Let $\mathcal{K}$ be an abstract channel. Then we denote by $C_0(\mathcal{K})$, $C_{0,f}(\mathcal{K})$, and $C_{0,\ell}(\mathcal{K}, L)$ its zero–error capacity, zero–error capacity with feedback, and zero–error capacity for list codes with list size L. For a given arbitrarily varying channel $\mathcal{W}$, we denote by $C_{\mathcal{R}}(\mathcal{W})$, $C_a(\mathcal{W})$, $C_{a,f}(\mathcal{W})$, and $C_{a,\ell}(\mathcal{W}, L)$ its average–error capacities for random correlated codes (c.f. [4], [6], or [10], [14] for its definition), ordinary deterministic codes, codes with feedback and list codes with list size L.

Let $\mathcal{X}$ be our input alphabet and let $\mathcal{Y}$ be our output alphabet. We define a code with d time delayed noiseless feedback of length n, or shortly a d–feedback code of length n as a set of functions $\{f_m^{(n,d)} : m \in \mathcal{M}\}$ from $\mathcal{Y}^{n-d}$ to $\mathcal{X}^n$ such that for all $y^{n-d} = (y_1, y_2, \ldots, y_{n-d}) \in$

$\mathcal{Y}^{n-d}$,

$$f_m^{(n,d)}(y^{n-d}) = \big(f_{m,1}^{(d)}, f_{m,2}^{(d)}, \dots, f_{m,d}^{(d)}, f_{m,d+1}^{(d)}(y_1), f_{m,d+2}^{(d)}(y^2), \dots, f_{m,n}^{(d)}(y^{n-d})\big), \tag{1}$$

where $y^i = (y_1, y_2, \dots, y_i)$ and $\mathcal{M}$ is a finite set correslonding to the set of messages. That is, for all $m \in \mathcal{M}$ $f_m^{(d)}$ is a vector valued function and its first d components are constant in $\mathcal{X}$, independent of y^{n-d} and for $t = d+1, d+2, \dots, n$, its t–th component $f_{m,t}^{(d)}$ is a function mapping y^{t-d} to $\mathcal{X}$. The information theoretical meaning is the following. At time t, the encoder sends a letter from the input alphabet $\mathcal{X}$ according to the value of the t–th component of the function $f_m^{(n,d)}$, if he wants to send the message m to the receiver (decoder), and at the same time the channel outputs a letter $y_t \in \mathcal{Y}$ according to the probabilistic rule giverning the channel $\mathcal{K}$ and the inputs. This output y_t arrives via a noiseless channel at the encoder at time $t+d$. Thus at time $t = 1, 2, \dots, d$, there is no feedback available, and so the encoder only can choose an input letter according to the message, which he wants to send. At time $d+1$, the feedback starts to arrive at the encoder. At time $t = d+1, d+2, \dots, n$, the encoder has received the first $t-d$ outputs $y^{t-d} = (y_1, y_2, \dots, y_{t-d})$ and he may associate them with the message to choose the input letter.

We shall consider two cases. In one case the delay time d is a constant and in the other it increases linearly with the code length. It seems to us that the first is more meaningful from a practical point of view and that the second is more interesting from a mathematical point of view. The capacities are defined in the standard way and we denote them in the two cases by $C_f^{(d)}(\mathcal{K})$ and $\tilde{C}_f^{(\delta)}(\mathcal{K})$, respectively, where $\delta = \frac{d}{n}$ for the second case. Analogously the zero–error capacities in these two cases are denoted by $C_{0,f}^{(d)}$ and $\tilde{C}_{0,f}^{(\delta)}(\mathcal{K})$, respectively. These 4 capacities are non–increasing in d or δ. In particular, when $d = 0$, a d–delay feedback code is an ordinary feedback code. Similarly we define $\tilde{C}_{a,f}^{(\delta)}$ for AVC $\mathcal{W}$.

3 Known Results

In this section we report a few known results, which we will use in our proofs in Sections 5 and 6. The average–error capacity for AVC was determined by R. Ahlswede and the key tool in his proof is the following elimination technique.

Lemma 1. (Ahlswede [4]) *For an AVC $\mathcal{W}$, an integer $\gamma \in [n^2, \infty)$, any $\varepsilon, \bar{\lambda} > 0$ for sufficiently large n, there exists a random correlated code of length n assigned to a set of codes of cardinality γ with average probability of error smaller than λ and rate larger than $C_{\mathcal{R}}(\mathcal{W}) - \varepsilon$.*

Next there are results for list decoding.

Lemma 2. (Elias [17]) *Given a discrete memoryless channel W and denote by $C_{f,0}(\mathcal{W})$ its zero–error capacity with (non–delayed) feedback. Then for all $\varepsilon > 0$ and sufficiently large n, there exists a code with list decoding of list size $L = L(\varepsilon, W)$ (depending on ε, W but independent of n) such that the rate of code is larger than $C_{f,0}(W) - \varepsilon$.*

Lemma 3. (Ahlswede–Cai [6]) *For AVC $\mathcal{W}$, and $\varepsilon, \bar{\lambda} > 0$, there exists an $L = L(\varepsilon, \lambda, \mathcal{W})$ such that for all sufficiently large n (independent of L), there exists a code with average* *probability of error smaller than $\bar{\lambda}$ and rate larger than $C_{\mathcal{R}}(\mathcal{W}) - \varepsilon$.*

The previous lemma was remarkably improved.

Lemma 4. (Blinovsky–Narayan–Pinsker [10]) *For the AVC $\mathcal{W}$, there exists a constant $L = L(\mathcal{W})$ such that the average–error capacity for list codes with size of list L $C_{a,\ell}(\mathcal{W}, L) = C_{\mathcal{R}}(\mathcal{W})$.*

4 Codes with delayed constant time feedback

In this section we present a simple coding scheme to show that the constant time delay for feedback does not effect capacities or rate regions for all memoryless channels for regardless of the error concepts (zero–error, maximal–error, average error). This includes for instance arbitrarily varying channels, two way channels, multiple access channels, broadcast channels, interference channels and all other channels in the books [13] and [14], regardless whether their capacities or rate–regions are known or unknown! The coding scheme is based on the folowing two observations.

1. For memoryless channels the optimal rates of codes converge to capacity. That is, the rates of optimal codes are arbitrarily close to the capacity if the lengths of codes are sufficiently long. (For multi–user channels, all points in the capacity regions can be approached by codes of sufficiently long lengths.)

2. For memoryless channels output statistics at time t only depends on the inputs at time t (and the state at time t for arbitrarily varying channels).

For simplicity of notation we present the coding scheme for two terminal channels and leave its obvious extension to any multi–user channel to the reader.

Let $\{f_m^n : m \in \mathcal{M}\}$ be a code with non–delayed feedback of length n whose rate is close to the capacity. Now we construct a code $\left\{f_{m^d}^{(n^d,d)} : m^d = (m_1, m_2, \dots, m_d) \in \mathcal{M}^d\right\}$ by concatening the code $\{f_m^n : m \in \mathcal{M}\}$ as follows. For $t = \tau n + i, 0 \leq \tau \leq d_1$, $1 \leq i \leq n$ and $f_{m^d,t}^{(n^d,d)}$ in (1) (i.e. the tth component of $f_{m^d}^{(n^d,d)}$) we set $f_{m^d,t}^{(n^d,d)} = f_{m_{\tau+1},i}$ for $f_m^n = (f_{m,1}, f_{m,2}, \dots, f_{m,n})$.

Then the new code has the same rate as the original code with no time delayed feedback and by the observation 2 its probability of error is not larger than d times the probability of error of the original code.

Obviously the memorylessness assumption, which implies observation 2, is essential and in general the coding scheme cannot be applied to a channel with memory.

5 Zero–error capacity for a DMC with linear increasing delay time for feedback

In this section we consider the case were for a $\delta \in [0,1)$ the output of a given DMC at time t ($t \leq n - \lfloor n\delta \rfloor$) arrives at the encoder at time $t + \lfloor n\delta \rfloor$ via noiseless feedback.

Theorem 1. *For all DMC's W, and $\delta \in [0,1)$,*

$$\tilde{C}_{f,0}^{(\delta)}(W) = (1-\delta)C_{f,0}(W) + \delta C_0(W). \tag{2}$$

Proof:

a) **Converse Part:**

Let us consider zero–error codes for the following communication system and denote the capacity by $C_{f,0}^{*(\delta)}(W)$. An encoder sends messages to a decoder with a zero–error code with feedback of length n via W and for the time $t < n - \lfloor n\delta \rfloor$, the output of the channel at time t immediately arrives at the encoder via noiseless feedback and the feedback is shut down at time $n - \lfloor n\delta \rfloor$. Obviously the output of the channel at time $t \geq n - \lfloor n\delta \rfloor$ can never arrive at the encoder if the feedback delays $\lfloor n\delta \rfloor$ units of time. Therefore $\lfloor n\delta \rfloor$–feedback code may always be simulated by a code for the above communication system. Consequently

$$\tilde{C}_{f,0}^{(\delta)}(W) \leq C_{f,0}^{*(\delta)}(W). \tag{3}$$

On the other hand we may apply Shannon's wellknown approach in the converse proof of his zero–error coding theorem with feedback [20] and conclude that in the worst case the encoder may not reduce the messages to a list of size smaller than $M2^{-n(1-\delta)C_{0,f}(W)}$ by sending the first $n - \lfloor n\delta \rfloor$ components of the input, if the initial message set size is M. Since the feedback is shut down at time $n - \lfloor n\delta \rfloor$ and the decoder has to determine the message in the last $\lfloor n\delta \rfloor$ units of time without error, $M2^{-n(1-\delta)C_{f,0}(W)}$ must not be larger than $2^{n\delta C_0(W)}$. That is

$$C_{f,0}^{*(\delta)}(W) \leq (1-\delta)C_{f,0}(W) + \delta C_0(W), \tag{4}$$

which together with (3) yields the converse.

b) **Direct Part:**

We prove the direct part by the following coding scheme which consists of three blocks. Let ε be an arbitrarily small but positive constant.

1) Our first block of the coding scheme has length $n\left(1-\delta-\frac{\varepsilon}{3}\right)$. By Lemma 2, there exists a constant L depending only on the channel and ε such that for sufficiently large n, there exists a zero–error code of length $n\left(1-\delta-\frac{\varepsilon}{3}\right)$ and rate $C_{f,0}(W) - \frac{\varepsilon}{3}$ with list decoding of list size L. The encoder uses such a code in the first block. Then the decoder knows that the message falls in a list of size L after the transformation. But at this moment the encoder does not know the list and to learn the list he has to wait for the feedback.

2) The second block has length $\lfloor n\delta \rfloor$. During the time he is waiting for the feedback, the encoder may use a zero–error code of length $\lfloor n\delta \rfloor$ to sent $n\delta\left(C_0(W) - \frac{\varepsilon}{3}\right)$ bits to the decoder.
3) After $\lfloor n\delta \rfloor$ units of time, the outputs of the whole first block arrive at the encoder and so he learns the list. Now the time for the last block only leaves $n\varepsilon$ units. But it is sufficient, if n is sufficiently large, because the size of the list is a constant L. So the encoder may use a zero–error code of length $\frac{\varepsilon}{3}$ to inform the decoder about the message in the list he sends to the decoder.

In the first two blocks the encoder sends $n\left(1 - \delta - \frac{\varepsilon}{3}\right)$ $\left(C_{f,0}(W) - \frac{\varepsilon}{3}\right)$ bits and $n\delta$ $\left(C_0(W) - \frac{\varepsilon}{3}\right)$ bits, respectively. So totally the rate of the code is

$$\begin{aligned}&\left(1 - \delta - \frac{\varepsilon}{3}\right)\left(C_{f,0}(W) - \frac{\varepsilon}{3}\right) + \delta\left(C_0(W) - \frac{\varepsilon}{3}\right)\\ &= (1-\delta)C_{f,0}(W) + \delta C_0(W) - \frac{\varepsilon}{3}\left[C_{f,0}(W) + 1 - \frac{\varepsilon}{3}\right].\end{aligned}$$

This completes the proof of the direct part.

6 Average–error capacity for an AVC with linear increasing delayed time feedback

We have learnt from [8] that the average–error capacity of codes for an AVC with non–delayed feedback is equal to the capacity of random correlated codes and one cannot expect that a code with delayed feedback is better than a code with non–delayed feedback.

So for all AVC $\mathcal{W}$

$$\tilde{C}_{a,f}^{(\delta)}(\mathcal{W}) \le C_{a,f}(\mathcal{W}) = C_{\mathcal{R}}(\mathcal{W}). \tag{5}$$

We shall show that $\tilde{C}_{a,f}^{(\delta)}$ actually is equal to $C_{\mathcal{R}}(\mathcal{W})$ and consequently the linear time delay makes no difference for feedback. That is

Theorem 2. *For all AVC, $\delta \in [0, 1)$*

$$\tilde{C}_{a,f}^{(\delta)}(\mathcal{W}) = C_{\mathcal{R}}(\mathcal{W}). \tag{6}$$

In order to prove Theorem 2, it is sufficient for us to present a coding scheme asymptotically achieving $C_{\mathcal{R}}(\mathcal{W})$. Before presenting it, let us briefly review the idea in the proof of the direct part of the coding theorem for AVC [4]. R. Ahlswede first reduced the size of a domain of a random correlated code for an AVC to $O(n^2)$ by the elimination technique. Then the encoder may randomly choose a code from this domain and inform the decoder about this choice by a code of size $n^2 \lesssim 2^{no(1)}$ in the case that the channel has positive capacity. This gives a deterministic code, since we may regard the choice as a message sint to the decoder. In the case that the capacity of a channel is zero it still works if the sender and decoder have other resources to obtain common randomness e.g., the (non–delayed) feedback [8]. Along this line in our coding scheme the encoder should use an arbitrarily short block to generate randomness at the output. However the randomness does not arrive at the encoder before $\lfloor n\delta \rfloor$ units of time later. During the waiting time, the encoder may send more messages by a list code. To wait for the list via the feedback, the encoder needs another $\lfloor n\delta \rfloor$ units of time. So totally the waiting time is $2n\delta$. Consequently this naive coding scheme requires the assumption that $\delta < \frac{1}{2}$.

However, we observe that the common randomness is not necessarily to be generated by feedback, even if the capacity is zero. In fact the randomness can be sent by a (short) list code in the second block such that the encoder may use a correlated code in the next block until the outputs of the first two blocks come via feedback. Then he may use the common randomness generated by the first block and inform the decoder where the message sent in the second block locates in the list. So, totally the waiting time is around $n\delta$ units.

Proof of Theorem 2: Now we formly prove the theorem by the following coding scheme. For fixed $\varepsilon > 0$, we choose ε', $\eta_i > 0$, $i = 0, 1, 2$ such that

$$1 - \sum_{j=0}^{2} \eta_j > \delta, \tag{7}$$

where ε', η_0, η_2 are chosen arbitrarily small (depending on ε and will be specified later). Then we choose as set of messages a cartesian product

$$\mathcal{M} = \mathcal{M}_r \times \mathcal{M}_1, \tag{8}$$

with

$$|\mathcal{M}_r| = 2^{n\eta_1(C_{\mathcal{R}}(\mathcal{W})-\varepsilon')} \tag{9}$$

and

$$|\mathcal{M}_1| = 2^{n\left(1-\sum\limits_{j=0}^{2}\eta_j\right)(C_{\mathcal{R}}(\mathcal{W})-\varepsilon')}. \tag{10}$$

Then our code consists of four blocks.

1) The first block has length $n\eta_0$. We use it to build common randomness as in [8]. The difference is that in our case the common randomness is built $\lfloor n\delta \rfloor$ units of time later whereas it is built immediately in the model [8].

2) The second block has length $n\eta_1$. In this block the encoder sends message $m_r \in \mathcal{M}_r$ by a list code of constant list size, if the message, which he wants to send is $(m_r, m_1) \in \mathcal{M}_r \times \mathcal{M}_1 = \mathcal{M}$. By (9) and Lemmas 3 or 4, the code with arbitrarily small average probability of error exists.

3) The third block has length $n\left(1 - \sum\limits_{j=0}^{2} \eta_1\right)$. In this block the encoder uses a random correlated code with the domain (of the random code) $\mathcal{M}_r$ and rate $C_{\mathcal{R}}(\mathcal{W}) - \varepsilon'$. The existence of the code follows from Lemma 1. The encoder sends $m_1 \in \mathcal{M}_1$ to the decoder by the m_rth code in the domain $\mathcal{M}_r$, if he wants to send (m_r, m_1) to the decoder. He can do it by (10).

4) There are $n\eta_2$ units of time left for the last block. By (7) the outputs of the first two blocks have arrived at the encoder before the last block is started. So the common randomness generated by the first block has been built and the list of the code in the second block has come to encoding. The encoder can use a code obtained by the elimination technique of Lemma 1 to inform the decoder which message in the list he sends in the second block. Since the size of the list is a constant, η_3 can be chosen arbitrarily small.

5) The encoder knows the message m_r is in a list, but does not know which one it is at the end of the second block. Then he does nothing but waits for the end of the transmission. At end of the last block the decoder learns m_r from the last block. Then he knows which code is used in the third block and so he is able to decode m from the code in the third block. Finally he obtains (m_r, m_1).

   The probability of error, clearly may be arbitrarily small and by (8) – (10) the rate is

$$\eta_1\big(C_{\mathcal{R}}(\mathcal{W})-\varepsilon'\big)+\left(1-\sum_{j=0}^{2}\eta_j\right)(C_{\mathcal{R}}-\varepsilon') = \big[1-(\eta_0+\eta_1)\big]\big(C_{\mathcal{R}}(\mathcal{W})-\varepsilon'\big) < C_{\mathcal{R}}(\mathcal{W})-\varepsilon,$$

   if we choose ε' and η_0, η_2 sufficiently small. So our proof is complete, since the converse trivially follows from (5).

References

[1] R. Ahlswede, A constructive proof of the coding theorem for discrete memoryless channels in case of complete feedback, Sixth Prague Conf. on Inf. Th., Stat. Dec. Fct's and Rand. Proc., 1–22, 1971.

[2] R. Ahlswede, Channels with arbitrarily varying channel probability functions in the presence of noiseless feedback, Z. Wahrscheinlichkeitstheorie und verw. Geb. 25, 239–252, 1973.

[3] R. Ahlswede, Channel capacities for list codes, J. Appl. Probability, lo, 824–836, 1973.

[4] R. Ahlswede, Elimination of correlation in random codes for arbitrarily varying channels, Z. Wahrscheinlichkeitstheorie und verw. Geb. 44, 159–175, 1978.

[5] R. Ahlswede and I. Wegener, Search Problems, Wiley–Interscience Series in Discrete Mathematics and Optimization, R.L. Graham, J.K. Leenstra, R.E. Tarjan, edit., English Edition of the German Edition Suchprobleme, Teubner Verlag, Stuttgart 1979, Russian Edition with Appendix by Maljutov 1981.

[6] R. Ahlswede and N. Cai, Two proofs of Pinsker's conjecture concerning arbitrarily varying channels, IEEE Trans. Inform. Theory, Vol. 37, 1647–1679, 1991.

[7] R. Ahlswede and N. Cai, The AVC with noiseless feedback and maximal error probability: a capacity formula with a trichotomy, in Numbers, Information and Complexity (Festschrift for Rudolf Ahlswede) eds. I. Althöfer etc., Kluwer, 151–176, 2000.

[8] R. Ahlswede and I. Csiszár, Common randomness in information theory and cryptography II, CR. capacity, IEEE Trans. Inform. Theory, Vol. 44, 225–240, 1998.

[9] R. Ahlswede and G. Dueck, Identification in the presence of feedback — a discovery of new capacity formulas, IEEE Trans. on Inf. Theory, Vol. 35, No. 1, 30–39, 1989.

[10] V.M. Blinovsky, P. Narayan, and M.S. Pinsker, The capacity of arbitrarily varying channels for list decoding, Problemy Peredachii Informatsii, Vol. 31, 3–41, 1995.

[11] F. Cicalese and U. Vaccaro, Coping with delay and time–outs in binary search procedures, (ISAAC 2000) D.T. Lee and Shang–Hua Teng eds, Lectures Notes in Comp. Sci., Vol. 1969, 96–107, Springer 2000.

[12] F. Cicalese, L. Gargano and U. Vaccaro, On searching strategies, parallel questions and delayed answers, in. Proc. of Fun. with Algorithms (FUN01), E. Lodi, L. Pagli and N. Santoro eds., 27–42, Carleton Scientific Press, 2001.

[13] T.M. Cover and J.A. Thomas, Elements of Information Theory, John Wiley, 1991.

[14] I. Csiszár and J. Körner, Information Theory: Coding Theorems for discrete memoryless systems, Academic, 1981.

[15] I. Csiszár and P. Narayan, The capacity of the arbitrarily varying channel revised: positivity constraints, IEEE Trans. Inform. Theory, Vol. 34, 81–193, 1988.

[16] C. Deppe, Solution of Ulam's search game with three lies or optimal adaptive strategies for binary three–error–correcting–codes, Dissertation, Universität Bielefeld, 1998.

[17] P. Elias, Zero error capacity under list decoding, IEEE Trans. Inform. Theory, Vol. 34, 1070–1074, 1988.

[18] T.H.E. Ericson, Exponential error bounds for random codes in the arbitrarily varying channel, IEEE Trans. Inform. Theory, Vol. 31, 42–48, 1985.

[19] F. Jelinek, Buffer overflow in variable length coding of fixed rate sources, IEEE Trans. Inform. Theory, Vol. IT–14, No. 3, 490–501, 1968.

[20] C.E. Shannon, The zero error capacity of a noisy channel, IEEE Trans. Inform. Theory, Vol. 2, 8–19, 1956.

[21] A.D. Wyner, On the probability of buffer overflow under an arbitrary bounded input–output distribution, SIAM J. Appl. Math, Vol. 27, No. 4, 544–569, Dec. 1974.