Anti-Structure Problems

Ram Zamir
Department of Electrical Engineering-Systems
Tel Aviv University, Israel
zamir@eng.tau.ac.il

***Abstract*—The recent success of structured solutions for a class of information-theoretic network problems, calls for exploring their limits. We show that sum-product channels resist a solution by structured (as well as random) codes. We conclude that the structured approach fails whenever the channel operations do not commute (or for general functional channels, when the channel function is non decomposable).**

Key words: sum-product channels, distributed coding, functional source coding, functional Gelfand-Pinsker problem, non-decomposable functions, commutativity, associativity.

I. Introduction

Structured codes can be effective, and significantly better than random codes, for various multiuser problems. Prominent examples include the well known Korner-Marton (KM) "modulo-two sum" problem [1], as well as more recent setups such as the "dirty" multiple-access channel (MAC) [2], the noisy linear network (along with the compute & forward relaying technique) [3], and more [4]. The effectiveness of structured codes, in particular linear or lattice codes, is due to a good match between their linear structure and the additive nature of the source or channel involved.

In the Korner-Marton problem, for example, the two components (X, Y) of a doubly-symmetric binary source are encoded separately. The joint decoder is not interested in a full reconstruction of X and Y, which are viewed as "helper sources"; rather the decoder is interested in their modulo-two sum $X + Y$ (or Xor). Writing the statistical relation between X and Y as a modulo-additive noise channel

$$Y = X + Z \tag{1}$$

we can recast the problem as that of lossless reconstruction of the noise Z from separate encodings of X and Y.

As shown by Korner and Marton, a linear structured coding scheme, which sends the syndromes of X^n and Y^n with respect to a suitable linear binary code, achieves the minimum possible rate of $H(Z)$ - the entropy of Z - per each encoder. In contrast, a conventional *random* coding scheme cannot do better than to encode at a sum rate equal to the joint entropy of X and Y. This corresponds to the Slepian-Wolf solution [5], i.e., to a lossless reconstruction of both X and Y at the decoder. The resulting sum rate can be therefore significantly higher than $2H(Z)$ for highly correlated sources.

The binary KM problem can be generalized to a q-ary field, in which case a linear q-ary code replaces the linear binary code in the KM solution [6]. And it also has a quadratic-Gaussian version [7].

A dual example with a similar characteristics is that of the "doubly dirty" MAC. This channel extends Costa's "writing on a dirty paper" problem [8], [9] to a MAC; i.e., an additive-noise channel with two inputs X_1 and X_2 and an output Y given by

$$Y = X_1 + X_2 + S_1 + S_2 + \text{noise} \tag{2}$$

where S_1 and S_2 are two interferences, each known as "side information" to one of the encoders. Addition in (2) is over some group in the discrete channel case, or the usual addition in the continuous case. The problem is made interesting by imposing input constraints upon X_1 and X_2, thus the encoders cannot simply subtract the interferences.

Similarly to the Korner-Marton problem, a linear/lattice precoding scheme (which subtracts the interference "modulo the code") achieves the capacity region of this channel [10], [2].[1] And in contrast, the rates achieved by a more conventional random binning scheme *vanish* in the limit of strong interference signals.

This sharp discrepancy is due to the distributive nature of the side-information; if the knowledge of S_1 and S_2 were *centralized* - i.e., they were both known to one encoder or to the joint decoder, then random binning could be effective and (nearly) achieve capacity; see [2].

Sometimes structured codes are *inferior* to random codes. This situation occurs when the linear structure of the code causes ambiguity at the decoder; for example, the symmetric-rates point of the (clean) MAC capacity region, or of the Slepian-Wolf rate region.[2]

In this short note we focus on another, perhaps obvious weakness of structured codes: they are sensitive to the structure of the channel. Specifically, we show that if the additive channel in (1) or in (2) is replaced by one involving both *addition and multiplication*, then structured codes - and in fact, any other coding scheme - are not effective.

II. Sum-Product Korner-Marton

Consider a generalization of the KM problem (1), where the statistical relation between the component sources is given by

[1] In the discrete noiseless case the linear coding scheme is exactly optimal, while in the continuous case it is asymptotically optimal in the continuous high SNR case.

[2] This problem can be resolved by using two *different* (linearly independent) linear codes [11].

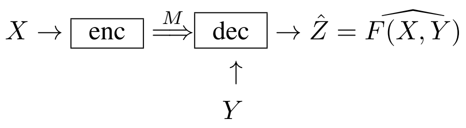

Fig. 1. Functional source coding with side information at the decoder.

the channel

$$Y = A + B' \times B'' \times C \tag{3}$$

where A, B', B'' and C are statistically independent, and $B', B'' \neq 0$ with probability one. All variables in (3) belong to a finite field F_q of size q, and the $+$ and $\times$ are the sum and product operations over F_q. Here A and Y are viewed as the "helper sources", C as the desired source, and B' and B'' as the "channel states". The classical KM problem (1) corresponds to the case where the channel states are deterministic; specifically, $q = 2$, $A = X$ is uniform over $\{0, 1\}$, $B' = B'' = 1$, and $C = Z$.

In the *centralized state* case, the equivalent channel state $B = B' \times B''$ is either known to the joint decoder (i.e., one encoder observes A, the other encoder observes Y, and the decoder has access to B), or to both encoders (i.e., one encoder observes (A, B), and the other encoder observes (Y, B)). It is not hard to show that in this case the compression rate is $H(Z)$ per each encoder, independent of the state distribution, as in the classical KM problem.[3]

Our focus is, however, on the *de-centralized state* case, where each of the channel states B' and B'' is available to a *different* encoder. The performance in this case can be bounded by the simplified setup shown in Fig. 1, where one of the helper sources is available (un-coded) as "side information" at the decoder. That is, there is only one encoder which observes $X = (A, B)$, while

$$Y = A + B \times C$$

is available at the decoder, who wishes to reconstruct $Z = C$.

This latter problem is a special case of *functional source coding* [6], where a function $Z = F(X, Y)$ needs to be reconstructed from separate coded versions of X and Y. The setup of Fig. 1 corresponds to the case where Y is given un-coded as "side information" to the decoder (or it is encoded at a rate greater than or equal to $H(Y)$), and where

$$X = (A, B), \quad Y = A + B \times C, \quad Z = C$$
$$\text{and} \quad F(X, Y) = (Y - A)/B. \tag{4}$$

A precise definition of functional source coding with side information at the decoder is as follows. The encoding function is $f : F_q^n \rightarrow \mathcal{M}$, where the size of the message space $\mathcal{M}$ is 2^{nR}, with n being the code block length and R being the coding rate. The decoding function is $g : \mathcal{M} \times F_q^n \rightarrow F_q^n$. The probability of error P_e is the probability that $g(M, Y^n)$ is not equal to the vector Z^n, where $Z_i = F(X_i, Y_i)$, $i = 1 \ldots n$, and where $M = f(X^n)$ is the encoded message. For a given memoryless double source $(X_1, Y_1), (X_2, Y_2), \ldots$ and a function $F(.,.)$, a rate R is said to be "achievable" if we can make P_e as small as desired for some functions f and g, and large enough n. Finally, R^* denotes the minimum achievable rate.

Clearly, the minimum achievable rate R* satisfies

$$H(X|Y) \geq R^* \geq H(F(X, Y)|Y) \tag{5}$$

where the LHS corresponds to the case where the decoder fully reconstructs X before computing $F(X, Y)$, while the RHS corresponds to the case where the encoder also has access to Y, so it can first compute $F(X, Y)$ and then compress it. In the sum-product case (4), if A is uniform over F_q, then the bounds (5) become

$$H(B) + H(C) \geq R^* \geq H(C). \tag{6}$$

Note that in the classical KM problem $B = 1$, i.e, $H(B) = 0$; thus the bounds coincide, and the coding rate is merely the entropy of the desired variable C.

Han and Kobayashi [6] give necessary and sufficient conditions for the LHS of (5) to be tight.[4] These conditions are satisfied in the sum-product case.

Lemma 1. *In the sum-product (functional source coding) problem (4), $R^* = H(X|Y)$. Thus, if A is uniform over F_q, then $R^* = H(B) + H(C)$.*

Proof: Follows since two different lines in F_q intersect in at most one point, implying the condition in [6, lem.1]. ∎

This result implies that the minimum coding rate R^* is in general larger than the entropy of the desired variable C, which is the rate in the classical KM setting (1). In fact, the "extra" rate can be as large as $\log(q-1)$, for B which is uniform over $F_q \setminus 0$.

As a corollary from Lemma 1, it follows that for uniform channel states B' and B'' in the distributed coding setup of (3), the rate of each encoder is at least $\log(q-1) + H(C)$. The interpretation is that the introduction of the multiplicative state variables breaks the structure of the classical KM problem; the states B' and B'' must be fully conveyed to the decoder before the linear structure of the channel can be utilized (by means of a linear "syndrome" coding) to encode the desired source C.

III. The Sum-Product Dirty MAC

Consider next a modification of the dirty MAC problem (2), in which the channel output is given by

$$Y = A' + A'' + B \times C \tag{7}$$

[3] Regarding the former case (B available at the decoder), note that a random parity-check matrix H is "good" with high probability for the classical KM problem (1) (i.e., Z^n can be reliably decoded from the syndromes HX^n and HY^n); hence, HB^n is "good" with high probability for the generalized KM problem (3). In the latter case, the encoders simply divide their observations by B, hence get back to the classical KM problem.

[4] They in fact consider a more general case where both X and Y are encoded.

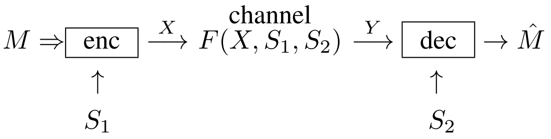

Fig. 2. A deterministic channel with two states, one known to the encoder and another known to the decoder.

where as in (3) all variables belong to a finite field F_q of size q, and the $+$ and $\times$ are the sum and product operations over F_q. The inputs of this MAC are A' and A'' (corresponding to X_1 and X_2 in (2)), while B and C are the channel state variables (corresponding to S_1 and S_2 in (2)). There is no additional noise, nor input constraints.

As in the sum-product KM, the centralized state case is easy: if both state variables B and C are known to one encoder, or are known to the joint decoder, then the product $B \times C$ can be simply subtracted; hence the sum capacity is $\log(q)$, as if the channel was noiseless.

The interesting setup is, again, the de-centralized state case. That is, each encoder has access to only one of the channel states, while the decoder is completely ignorant of the states. The capacity in this case is bounded from above by that of the *single-user* channel shown in Fig. 2, with

$$Y = A + B \times C. \tag{8}$$

Here there is a single encoder that has access to one of the states ($S_1 = B$), while the decoder has access to the second state ($S_2 = C$), where C is independent of both the input $X = A$ and B.[5]

A precise definition of encoding and decoding over such a channel is as follows. The encoding function is $f : M \times F_q^n \to F_q^n$ and the decoding function is $g : F_q^n \times F_q^n \to \mathcal{M}$, where the size of the message space $\mathcal{M}$ is 2^{nR} , R being the coding rate. The error probability P_e is the probability that $g(Y^n, S_2^n) \neq M$, where Y^n depends on X^n, S_1^n and S_2^n, and where $X^n = f(M, S_1^n)$ for $M \in \mathcal{M}$. A rate R is said to be "achievable" if we can make P_e as small as desired for some functions f and g, and large enough n. Finally, the capacity $\mathbf{C}$ is the highest achievable rate.

The sum-product channel (8) is, in fact, a deterministic channel, where the output Y is a function of the input X, and the two states S_1 and S_2:

$$Y = F(X, S_1, S_2). \tag{9}$$

There is no additional noise in the channel, beyond the randomness of the two (known) states S_1 and S_2.

The setup of (9) is an instance of the Gelfand-Pinsker problem [9], i.e., a channel with non-causal side information at the encoder. Hence, it has a single letter solution of the form

$$\mathbf{C} = \max I(U; Y, S_2) - I(U; S_1) \tag{10}$$

where the maximization is over a suitable set of auxiliary random variables U, and functions $X = X(U, S_1)$.[6]

The structure of the function F in (9) plays a key role in determining the capacity $\mathbf{C}$. A favorable case is when F has a *composite form*, where the dependence on the encoder variables (X, S_1) is separate from the decoder state S_2.

Lemma 2. *If the function $F : F_q \times F_q \times F_q \to F_q$ can be decomposed into $F(a, b, c) = \tilde{F}(G(a, b), c)$, where $\tilde{F}$ is invertible with respect to the first argument (i.e., the equation $y = \tilde{F}(t, c)$ has a solution t for every y and c), then (10) is optimized by $U = G(X, S_1)$. If also G is invertible with respect to the first argument, then the capacity is*

$$\mathbf{C} = \log(q)$$

and it is achieved by an input $p(x|s_1)$ that makes $G(X, s_1)$ uniform over F_q for all values of s_1.

Proof: The first part follows from [13, sec. III.F], and the invertibility of $\tilde{F}$. See also [14]. ■

The sum-product channel (8) clearly does not satisfy the first condition of the lemma, as addition and multiplication do *not* commute. In fact, this channel is much worse. To assess its capacity, we shall first establish a relation to a "minimum entropy" problem.

Lemma 3. *The capacity of a deterministic two-state channel of the form*

$$Y = X + F(S_1, S_2) \tag{11}$$

where $+$ denotes addition in F_q, and where S_1 and S_2 are available at the encoder and the decoder, respectively, is given by

$$\mathbf{C} = \log(q) - \inf \frac{1}{n} H\Big(g(S_1^n) + F(S_1^n, S_2^n) | S_2^n\Big) \tag{12}$$

where the second term is the average conditional entropy given S_2^n, and where the infimum is over all code block lengths n, and functions $g : F_q^n \to F_q^n$.

Proof: Easy and will be omitted. ■

Although (12) is not a single-letter expression, it is sometimes easier for analysis than the Gelfand-Pinsker solution (10). Specifically, for the sum-product channel (8) we have:

Theorem 1. (Shany-Zamir [15]) *For the case where $Y = X + S_1 \times S_2$, the minimum average conditional entropy in (12) is bounded by*

$$\log\left(\frac{q}{2}\right) \leq H_{min} \leq \log\left(\frac{q}{2 - 1/q}\right).$$

The upper bound is achieved (for all n) by a quadratic per-letter function $g(s) = s^2$.

As a corollary from this theorem, we conclude that the capacity of the sum-product channel (8) is *at most one bit*,

[5] A continuous version of this setup may be thought of as a channel with a fading interference [12].

[6] The admissible U's are those for which S_2 is independent of (U, X, S_1), and $U \leftrightarrow (X, S_1) \leftrightarrow Y$ form a Markov chain for each value of S_2. Since S_2 is independent of U, the first term in (10) can be written also as $I(U; Y|S_2)$.

for all q. This is quite disappointing when compared to the capacity of $\log(q)$, which is achievable in the centralized-state case. The same statement is true also for the sum-product dirty MAC (7); that is, the rate of each user is at most one bit.[7]

IV. Discussion

The essence of the examples given in this paper is that the order of performing the sum and product operations matters; and in fact, they are very "different". One aspect of this difference - which is related to the sum-product KM problem - is that the expression

$$a + b \times c \tag{13}$$

cannot be decomposed into the form: function(function $(a, b), c)$ (not even approximately). A second aspect - related to the sum-product dirty MAC - is that it is impossible to find a function of $a = a(b)$ such that (13) would be only a function of c (not even approximately). In contrast, these two requirements are easily fulfilled if the expression in (13) is a pure sum $a + b + c$ (by the associativity of summation), or a pure product $a \times b \times c$ (by the associativity of multiplication).[8]

It would be interesting to explore further (and perhaps quantify) the information-theoretic aspects of function decomposition. Note that this question is almost "distribution free" (i.e., nearly independent of the probability distributions of sources and channels). A different aspect of "anti structure", which is due to a "bad" noise distribution, can be found in [16].

Acknowledgment

I thank Yaron Shany for many interesting discussions about the sum-product channel. Thanks are also due to Aaron Wagner for an interesting discussion about the sum-product KM problems, and to Prakash Iswar and Sandeep Pradhan for pointing out the relation to the work of Han and Kobayashi.

References

[1] J. Korner and K. Marton, "How to encode the modulo-two sum of binary sources," *IEEE Trans. Information Theory*, vol. IT-25, pp. 219–221, 1979.

[2] T. Philosof, R. Zamir, U. Erez, and A. Khisti, "Lattice strategies for the dirty multiple-access channel," *IEEE Trans. Information Theory*, vol. IT-57, pp. 5006–5035, Aug. 2011.

[3] B. Nazer and M. Gastpar, "Reliable physical layer network coding," *Proceedings of the IEEE*, vol. 99(3), pp. 438–460, March 2011.

[4] R. Zamir, "Can structure beat random? the story of lattice codes," *Newsletters of Inf. Th. Society*, pp. 20–29, March 2011.

[5] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Information Theory*, vol. IT-19, pp. 471–480, July 1973.

[6] T. Han and K. Kobayashi, "A dichotomy of functions $F(X, Y)$ of correlated sources (X, Y) from the viewpoint of the achievable rate region," *IEEE Trans. Information Theory*, vol. IT-33, pp. 69–76, Jan. 1987.

[7] D. Krithivasan and S. Pradhan, "Lattices for distributed source coding: Jointly Gaussian sources and reconstruction of a linear function," *IEEE Trans. Information Theory*, vol. 55, pp. 5628–5651, Dec. 2009.

[8] M. Costa, "Writing on dirty paper," *IEEE Trans. Information Theory*, vol. IT-29, pp. 439–441, May 1983.

[9] S. Gelfand and M. S. Pinsker, "Coding for channels with random parameters," *Problemy Pered. Inform. (Problems of Inform. Trans.)*, vol. 9, No. 1, pp. 19–31, 1980.

[10] T. Philosof and R. Zamir, "On the loss of single-letter characterization: The dirty multiple access channel," *IEEE Trans. Information Theory*, vol. IT-55, pp. 2442–2454, June 2009.

[11] E. Haim, Y. Kochman, and U. Erez, "Improving the MAC error exponent using distributed structure," in *Proc. of ISIT*, Aug. 2011, St. Petersburg, pp. 1003–1007.

[12] Y. Avner, B. Zaidel, S. Shamai, and U. Erez, "On the dirty paper channel with fading dirt," in *Proc. of the Electrical and Electronics Engineers in Israel (IEEEI)*, Nov. 2010, Eilat, pp. 000 525–000 529.

[13] R. Barron, B. Chen, and G. Wornell, "The duality between information embedding and source coding with side information, and some applications," *IEEE Trans. Information Theory*, vol. 49, pp. 1159–1180, May 2003.

[14] E. Haim, "Input-cost side-information and broadcast channels," Master's thesis, Tel Aviv University, Nov. 2007.

[15] Y. Shany and R. Zamir, "A lower bound on the average entropy of a function determined up to a diagonal linear map on F_q^n," *Arxiv*, vol. http://arxiv.org/abs/1105.3793, 2011.

[16] A. Cohen and R. Zamir, "Entropy amplification property and the loss for writing on dirty paper," *IEEE Trans. Information Theory*, vol. IT-54, pp. pp. 1477–1487, April 2008.

[7] Perhaps even the stronger statement - that the sum rate is at most one bit - is true.

[8] Note that pure multiplicative versions of the generalized KM and DMAC problems can be solved using linear codes over a "logarithmic" domain.