Hypothesis testing via a comparator

Yury Polyanskiy

***Abstract*—This paper investigates the best achievable performance by a hypothesis test satisfying a structural constraint: two functions are computed at two different terminals and the detector consists of a simple comparator verifying whether the functions agree. Such tests arise as part of study of fundamental limits of channel coding, but are also useful in other contexts. A simple expression for the Stein exponent is found and applied to showing a strong converse in the problem of multi-terminal hypothesis testing with rate constraints. Connections to the Gács-Körner common information and to spectral properties of conditional expectation operator are identified. Further tightening of results hinges on finding λ-blocks of minimal weight. Application of Delsarte's linear programming method to this problem is described.**

I. Introduction

A classical problem in statistics and information theory is that of determining which of the two distributions, P or Q, better fit an observed data vector. As shown by Neyman and Pearson, the binary hypothesis testing (in the case of simple hypotheses) admits an optimal solution based on thresholding the relative density of P with respect to Q (a Radon-Nikodym derivative). The asymptotic behavior of the tradeoff between the two types of errors has also been well studied by Stein, Chernoff, Hoeffding and Blahut. Knowledge of this tradeoff is important by itself and is also useful for other parts of information theory, such as channel coding [1, Section III.E] and data compression [2, Section IV.A].

The problem becomes, however, much more complex with the introduction of structural constraints on the allowable tests. For example, it may happen that observations consist of two parts, say $X^n = (X_1, \ldots, X_n)$ and $Y^n = (Y_1, \ldots, Y_n)$, which need to be compressed down to nR bits each before the decision is taken. Even the memoryless case, in which under either hypothesis the pairs (X_i, Y_i) are independent and identically distributed (i.i.d.) according to P_{XY} or Q_{XY}, is a notoriously hard problem with only a handful of special cases solved [3]–[6]. Formally, this problem corresponds to finding the best test of the form

$$T = 1\{(f(X^n), g(Y^n)) \in A\}, \qquad (1)$$

where optimization is over functions f and g with finite co-domains of cardinality 2^{nR} and critical regions A. Here and below $T = 1$ designates the test choosing the distribution P and $T = 0$ the distribution Q.

Another rich source of difficult problems is the distributed case, in which observations are taken by spatially separated sensors (whose measurements are typically assumed to be correlated in space but not in time). The goal is then to optimize the communication cost by designing (single letter) quantizers and a good (single or multi round) protocol for exchanges between the sensors and the fusion center; see [7]–[9] and references therein. These problems can again be restated in the form of constraining the allowable tests similar to (1).

In this paper we consider tests employing a comparator, namely those satisfying the constraint:

$$T = 1\{f(X^n) = g(Y^n)\}, \qquad (2)$$

where the cardinality of the common co-domain of f and g is unrestricted. This constraint is motivated by the meta-converse method [1, Section III.E], which proves a lower bound on probability of error by first using a channel code as a binary hypothesis test and then comparing its performance with that of an optimal (Neyman-Pearson) test. However, so constructed test necessarily satisfies the structural constraint (2) and thus it is natural to investigate whether imposing (2) incurs exponential performance loss.

Another situation in which tests of the form (2) occur naturally is in the analysis of parallel systems, such as in fault-tolerant parallel computers, that under normal circumstances perform a redundant computation of a complicated function with high probability of agreement, while it is required to lower bound the probability of agreement when the fault occurs (modeled as P_{XY} changing to Q_{XY}). Yet another case is in testing hypotheses of biological nature based on the observation of zygosity of cells only (in eukaryotes).

The main result is that in the memoryless setting Stein exponent of tests satisfying (2) can indeed be quite a bit smaller than $D(P_{XY}||Q_{XY})$ and in fact is given by

$$E \triangleq \min_{V_X = P_X, V_Y = P_Y} D(V_{XY}||Q_{XY}), \qquad (3)$$

where $D(\cdot||\cdot)$ is the Kullback-Leibler divergence, and the optimization is over all joint distributions V_{XY} with marginals matching those of P_{XY}. In particular, $E = 0$ if (and only if) the marginals of Q_{XY} coincide with those of P_{XY}.

In fact, for the latter case, the hypothesis testing with constraint (2) turns out to be intimately related to a problem of determining the common information $C(X;Y)$ in the sense of Gács and Körner [10]. Using a technique pioneered by Witsenhausen [11] we show that the error probability cannot decay to zero at all (even subexponentially). Unfortunately, this is only shown under the condition that the confidence level is sufficiently high. Extending to the general case appears to be surprisingly hard. For a special case of binary X and Y we describe a bound based on Delsarte's linear programming method [12] and demonstrate promising numerical results. However, we have not yet been able to identify a convenient

The author is with the Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA 02139 USA. e-mail: yp@mit.edu.

The research was supported by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-09-39370.

polynomial, such as found in [13] for the coding in Hamming space, admitting an asymptotic analysis.

The exponent E has appeared before in the context of hypothesis testing with rate constraints (1), see [4, Theorems 5 and 8], and distributed detection [8, Theorem 2]. We identify the reasons for this below and also use this correspondence to prove the strong converse for the results in [4].

II. Background and notation

Consider a distribution P_{XY} on $\mathcal{X}\times\mathcal{Y}$. We denote a product distribution on $\mathcal{X}^n\times\mathcal{Y}^n$ by P^n_{XY} and by $P_{XY}>0$ the fact that P_{XY} is non-zero everywhere on $\mathcal{X}\times\mathcal{Y}$.

Fix some P_{XY} and Q_{XY}. For each integer $n\ge 1$ and $0\le\alpha\le 1$ the performance of the best possible comparator hypothesis test of confidence level α is given by

$$\tilde\beta_\alpha(P^n_{XY},Q^n_{XY})\stackrel{\triangle}{=}\inf\mathbb{Q}[T=1]\,,$$

where infimum is over all (perhaps, randomized) maps $f:\mathcal{X}^n\to\mathbb{R}$ and $g:\mathcal{Y}^n\to\mathbb{R}$ such that

$$\mathbb{P}[T=1]\ge\alpha\,,$$

where T is defined in (2). Here and below we follow the agreement that $\mathbb{P}$ and $\mathbb{Q}$ denote measures on some abstract spaces carrying random variables (X^n,Y^n) distributed as P^n_{XY} and Q^n_{XY}, resp..

For a finite $\mathcal{X}\times\mathcal{Y}$ and a given distribution P_{XY} we define a bipartite graph with an edge joining $x\in\mathcal{X}$ to $y\in\mathcal{Y}$ if $P_{XY}(x,y)>0$. The connected components of this graph are called components of P_{XY} and the entropy of the random variable indexing the components is called the common information of X and Y, cf. [10]. If the graph is connected, then P_{XY} is called indecomposable. In particular indecomposability implies $P_X>0$ and $P_Y>0$.

We also define a maximal correlation coefficient $S(X;Y)$ between two random variables X and Y as

$$S(X;Y)=\sup_{f,g}\mathbb{E}\left[f(X)g(Y)\right]$$

supremum taken over all zero-mean functions of unit variance. For finite $\mathcal{X}\times\mathcal{Y}$ indecomposability of P_{XY} implies $S(X;Y)<1$ and (under assumption $P_X>0$,$P_Y>0$) is equivalent to it.

Finally, we recall [10] that a pair of sets $A\in\mathcal{X}^n$ and $B\in\mathcal{Y}^n$ is called a λ-block for P^n_{XY} if $P^n_X[A]>0$, $P^n_Y[B]>0$ and

$$\mathbb{P}[X^n\in A|Y^n\in B]\ge\lambda\,,\qquad \mathbb{P}[Y^n\in B|X^n\in A]\ge\lambda\,.$$

An elegant theorem of Gács and Körner states

Theorem 1 ([10]): Let P_{XY} be an indecomposable distribution on a finite $\mathcal{X}\times\mathcal{Y}$. Then for every $\lambda_n\ge\exp\{-o(n)\}$ there exists a sequence $\nu_n=o(n)$ such that for all n any λ_n-block (A,B) for P^n_{XY} satisfies

$$P^n_{XY}[A\times B]\ge\exp\{-\nu_n\}\,.$$

III. Main results

A. Stein exponent

Theorem 2: Consider an indecomposable P_{XY} on a finite $\mathcal{X}\times\mathcal{Y}$. Then for an arbitrary Q_{XY} and any $0<\alpha<1$ we have

$$\lim_{n\to\infty}\frac{1}{n}\log\tilde\beta_\alpha(P^n_{XY},Q^n_{XY})=-E\,,$$

where E is defined in (3). Moreover, if $E=\infty$ then there exists $n_0(\alpha)$ such that $\tilde\beta_\alpha(P^n_{XY},Q^n_{XY})=0$ for all $n\ge n_0$.

Proof: Achievability: Consider functions

$$f(x^n) = 1\{x^n\not\in T^n_{[P_X]}\}\,, \tag{4}$$
$$g(y^n) = 2\cdot 1\{y^n\not\in T^n_{[P_Y]}\}\,, \tag{5}$$

where $T^n_{[P]}$ denotes the set of P-typical sequences [14, Chapter 2] over the alphabet of P. Then, on one hand by typicality:

$$\mathbb{P}[f(X^n)=g(Y^n)] = P^n_{XY}[T^n_{[P_X]}\times T^n_{[P_Y]}] \tag{6}$$
$$\ge 1-o(1)\,. \tag{7}$$

On the other hand, using joint-type decomposition it is straightforward to show that the set $T^n_{[P_X]}\times T^n_{[P_Y]}$ under the product measure Q^n_{XY} satisfies

$$Q^n_{XY}[T^n_{[P_X]}\times T^n_{[P_Y]}]=\exp\{-nE+o(n)\}\,. \tag{8}$$

For the case of $E<\infty$, this has been demonstrated in the proof of [4, Theorem 5]. For the case $E=\infty$, we need to show that for all $n\ge n_0$ we have

$$Q^n_{XY}[T^n_{[P_X]}\times T^n_{[P_Y]}]=0\,.$$

Indeed, assuming otherwise we find a sequence of typical pairs (x^n,y^n) with positive Q_{XY}-probability. But then the sequence of the joint types $V^{(n)}_{XY}$ associated to (x^n,y^n) belongs to the closed set of joint distributions $\{V_{XY}:V_{XY}\ll Q_{XY}\}$ and by compactness must have a limit point $\bar V_{XY}$. By the δ-convention [14, Chapter 2], the accumulation point must have marginals $\bar V_X=P_X$ and $\bar V_Y=P_Y$ and thus $E\le D(\bar V_{XY}||Q_{XY})<\infty$ – a contradiction.

Converse: We reduce to the special case of the theorem, stated as Theorem 3 below. If $E=\infty$ then there is nothing to prove, so assume otherwise and take an arbitrary V_{XY} with $V_X=P_X$, $V_Y=P_Y$ and $D(V_{XY}||Q_{XY})<\infty$. Our goal is to show that

$$\tilde\beta_\alpha(P^n_{XY},Q^n_{XY})\ge\exp\{-nD(V_{XY}||Q_{XY})+o(n)\}\,. \tag{9}$$

If $V_{XY}\not>0$ then we can replace V_{XY} with $(1-\epsilon)V_{XY}+\epsilon P_XP_Y$, which is everywhere positive on $\mathcal{X}\times\mathcal{Y}$, and then take a limit as $\epsilon\to 0$ in (9). Thus we assume $V_{XY}>0$.

Denote

$$A_n\stackrel{\triangle}{=}\{f(X^n)=g(Y^n)\}\,.$$

By the special case of the theorem we have

$$V^n_{XY}[A_n]\ge\exp\{-o(n)\}\,. \tag{10}$$

Then, by a standard change of measure argument, we must have

$$Q^n_{XY}[A_n]\ge\exp\{-nD(V_{XY}||Q_{XY})+o(n)\}\,. \tag{11}$$

Optimizing the choice of V_{XY} in (11) proves (9) and the Theorem. ■

It remains to consider the case of matching marginals:

Theorem 3 (Special case $E=0$): Let P_{XY} be indecomposable, $Q_{XY}>0$ and $Q_X=P_X$, $Q_Y=P_Y$. Then for any $0<\alpha<1$ we have

$$\tilde\beta_\alpha(P^n_{XY},Q^n_{XY})\ge\exp\{-o(n)\}\,. \tag{12}$$

Proof: First we show that any test of level α must contain a λ-block with $\lambda \geq \frac{\alpha}{2}$. Indeed, each pair $(\{f(X^n) = i\}, \{g(Y^n) = i\})$ is a λ_i-block for some λ_i (chosen to be maximum possible). Then, by the Bayes rule and $\max\{x, y\} \leq x + y$ we get

$$\mathbb{P}[f(X^n) = g(Y^n) = i] \leq \lambda_i(\mathbb{P}[f(X^n) = i] + \mathbb{P}[g(Y^n) = i]) .$$

Summing this over i shows that at least one $\lambda_i \geq \frac{\alpha}{2}$.

By the Gács-Körner effect (Theorem 1) the probability of this λ-block is subexponentially large:

$$\mathbb{P}[f(X^n) = g(Y^n) = i] \geq \exp\{-o(n)\} .$$

Therefore, in particular we have (since the marginals of X^n and Y^n under $\mathbb{P}$ and $\mathbb{Q}$ coincide)

$$\mathbb{Q}[f(X^n) = i] \geq \exp\{-o(n)\} , \tag{13}$$

$$\mathbb{Q}[g(Y^n) = i] \geq \exp\{-o(n)\} . \tag{14}$$

Thus, the sets $\{f(X^n) = i\}$ and $\{g(Y^n) = i\}$ must occupy a subexponential fraction of typical sets $T^n_{[P_X]}$ and $T^n_{[P_Y]}$. In view of (8) it is natural to expect that

$$\mathbb{Q}[f(X^n) = g(Y^n) = i] \geq \exp\{-o(n)\} \tag{15}$$

(note that marginals match and thus $E = 0$ as per (3)). Under the assumption $Q_{XY} > 0$ it is indeed straightforward to show (15) by an application of blowing-up lemma; see [6, Theorem 3].

Finally, (15) completes the proof because

$$\mathbb{Q}[T = 1] \geq \mathbb{Q}[f(X^n) = g(Y^n) = i] .$$

∎

B. Discussion

It should be emphasized that although intuitively one imagines that the behavior of $\tilde{\beta}_\alpha$ should markedly depend on how the connected components of P_{XY} and Q_{XY} relate to each other, Theorem 2 demonstrates that the Stein exponent is not sensitive to the decomposition of Q_{XY}.

The assumption of indecomposability of P_{XY} in Theorem 2, however, is essential. Indeed, consider the case of $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ and $X = Y$ uniform (under P_{XY}) vs X, Y independent uniform (under Q_{XY}). Clearly a test $\{X^n = Y^n\}$ demonstrates

$$\tilde{\beta}_1(P^n_{XY}, Q^n_{XY}) \leq 2^{-n} , \tag{16}$$

while according to the definition (3) we have $E = 0$.

We also remark that the case of $E = \infty$ is possible. For example, let X, Y be binary with $P_{XY}(0, y) = \frac{1}{2} - P_{XY}(1, y) = \frac{p}{2}$ for $p > \frac{1}{2}$, and $Q_{XY}(x, y) = \frac{1}{2}1\{x = y\}$.

C. Hypothesis testing with a 1-bit communication constraint

The exponent E in (3) is related to hypothesis testing under the communication constraint (1). In fact, Theorem 2 extends [4, Theorem 5] to the entire range $0 < \epsilon < 1$, thereby establishing the full strong converse. This result has been obtained in [6] under different assumptions on P_{XY} and Q_{XY}[1].

[1] Namely, we do not require $D(P_{XY}||Q_{XY}) < \infty$ or positivity of Q_{XY}, but require indecomposability of P_{XY}.

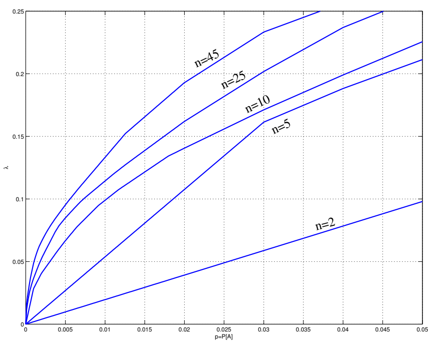

Fig. 1. Linear programming upper bound on λ as a function of p. Uniform X and Y connected by the $BSC(\delta)$, $\delta = 0.3$.

Corollary 4: Consider a hypothesis testing between an indecomposable P_{XY} and an arbitrary Q_{XY} with structural restriction on tests of the form

$$T = 1\{(f(X^n), Y^n) \in A\} \tag{17}$$

with binary-valued f. Then for any $0 < \epsilon < 1$ we have

$$\inf \mathbb{Q}[T = 1] = \exp\{-nE + o(n)\} ,$$

where infimum is over all tests satisfying $\mathbb{P}[T = 1] \geq 1 - \epsilon$ and E is given by (3).

Proof: Clearly, any test with binary-valued f and g of the form (2) is also a test of the form (17). Thus Theorem 2 establishes the achievability part. Conversely, for any test of the form (17) we may find sets A_0 and A_1 such that

$$\begin{aligned} \mathbb{P}[T = 1] &= \mathbb{P}[\{Y^n \in A_0, f(X^n) = 0\} \\ &\quad \cup \{Y^n \in A_1, f(X^n) = 1\}] \qquad (18) \\ &\geq 1 - \epsilon . \qquad (19) \end{aligned}$$

Then without loss of generality assume that the first set in the union has $\mathbb{P}$-probability larger than $\frac{1-\epsilon}{2}$. Define the following function

$$g(y^n) = 1\{y^n \in A_1 \setminus A_0\} + 21\{y^n \notin A_0 \cup A_1\}$$

Then since $\{f = g\} \supseteq \{y^n \in A_0, f(X^n) = 0\}$ we have

$$\mathbb{P}[f(X^n) = g(Y^n)] \geq \frac{1-\epsilon}{2} , \tag{20}$$

and thus by Theorem 2 we conclude that $\mathbb{Q}[T = 1]$ is at least $\exp\{-nE + o(n)\}$. ∎

We remark that the correspondence between the hypothesis tests with 1-bit compression and those of interest in this paper (2) does not hold in full generality. In particular, it was shown in [4, Theorem 5] that the exponent E in (3) is still optimal in the 1-bit scenario without the requirement of indecomposability of P_{XY}, while example (16) demonstrates the contrary for our setup.

IV. Non-vanishing lower bounds

By Theorem 3 in the case when marginals of P_{XY} and Q_{XY} coincide the error cannot decay to zero exponentially. In fact, we *conjecture* that in the cases of matching marginals $\tilde{\beta}_\alpha(P^n_{XY}, Q^n_{XY})$ does not vanish at all. In this section we prove the conjecture under additional assumptions and discuss potential methods for extending to the general case.

Theorem 5: Consider a P_{XY} and $Q_{XY} = P_X P_Y$ such that $S \triangleq S(X;Y) < \frac{1}{2}$ (under P_{XY}). Then for any $\alpha \in (2S, 1]$ we have

$$\lim_{n\to\infty} \tilde{\beta}_\alpha(P_{XY}^n, Q_{XY}^n) > 0\,.$$

Proof: As in the proof of Theorem 3 given a test $\{f(X^n) = g(Y^n)\}$ of level α we can extract a λ-block (A, B) with $\lambda > \frac{\alpha}{2}$. We want to show that for some constant $p = p(\alpha) > 0$ and all n at least one of the marginals $P_X^n[A]$ or $P_Y^n[B]$ can be bounded away from zero:

$$\max\{P_X^n[A], P_Y^n[B]\} \ge p\,. \tag{21}$$

Indeed, then we have

$$\mathbb{Q}[T=1] \ge P_X^n[A]P_Y^n[B] \ge \frac{\alpha}{2}p^2 \tag{22}$$

which follows because in a λ-block the smaller of the two probabilities in (21) should still be larger than the joint probability $P_{XY}^n[A\times B]$ which is $\ge \lambda p$. Finally, the estimate (21) follows from the next result. ■

Lemma 6: Consider a λ-block (A, B) for P_{XY}^n. Then,

$$\max\{P_X^n[A], P_Y^n[B]\} \ge \min\left\{\frac{1}{2}, \frac{\lambda - S}{1-S}\right\}, \tag{23}$$

whenever $S = S(X;Y) < 1$.

Proof: Consider an operator $T_n : L_2(\mathcal{Y}^n, P_Y^n) \to L_2(\mathcal{X}^n, P_X^n)$ defined as follows:[2]

$$(T_n h)(x^n) \triangleq \mathbb{E}\left[f(Y^n)|X^n = x^n\right], \tag{24}$$

where the expectation is over the distribution P_{XY}^n. Note that the second largest singular value of T_n is precisely the maximal correlation coefficient $S(X;Y)$ (under P_{XY}), see [15]. Thus, for any zero-mean functions $h \in L_2(\mathcal{X}^n)$ and $h' \in L_2(\mathcal{Y}^n)$ we have

$$\mathbb{E}\left[h(X^n)h'(Y^n)\right] = (T_n h', h) \le S(X;Y)||h||_2||h'||_2\,. \tag{25}$$

Denote $p_A = P_X^n[A]$, $p_B = P_Y^n[B]$ and assume $p_B \ge p_A$. If $p_B \ge \frac{1}{2}$ then there is nothing to prove, so assume otherwise. Then, we have

$$\begin{aligned} \lambda p_B &\le P_{XY}^n[A\times B] && (26)\\ &\le p_A p_B + S\sqrt{p_A(1-p_A)p_B(1-p_B)} && (27)\\ &\le p_B^2 + S p_B(1-p_B)\,, && (28) \end{aligned}$$

where (26) is by the definition of a λ-block, (27) is by (25) applied to $h(x^n) = 1\{x^n \in A\} - p_A$ and $h' = 1\{y^n \in B\} - p_B$; and (28) is because $p_A \le p_B \le \frac{1}{2}$. Canceling p_B on both sides in (28) we obtain (23). ■

Next, we discuss what is required to extend Theorem 5 to full generality. To handle a general Q_{XY} one needs a non-vanishing lower bound independent of n on

$$\lambda_{min}(p, Q_{XY}^n) = \min_{A,B} Q_{XY}^n[A\times B]\,,$$

where the minimization is over $Q_X^n[A], Q_Y^n[B] \ge p$. For $Q_{XY} = P_X P_Y$ this problem is void since $\lambda_{min}(p, P_X^n P_Y^n) = p^2$. Nevertheless, even the case of $Q_{XY} = P_X P_Y$ is far from being resolved as we need to extend to the full range $0 < \alpha < 1$. We discuss this second problem further.

A. More on spectral methods

In a nutshell, the proof of Theorem 5 consisted of two steps. First, we identified a Markov chain

$$F \to X^n \to Y^n \to G\,, \tag{29}$$

where we denoted $F \triangleq f(X^n)$, $G \triangleq g(Y^n)$. Note that by the data-processing for maximal correlation we have

$$S(F;G) \le S(X^n;Y^n) = S(X;Y)\,.$$

Second, for large α we showed a lower bound

$$\mathbb{Q}[F=G] = \sum_i \mathbb{P}[F=i]\mathbb{P}[G=i] \ge \text{const} > 0$$

under conditions: a) $\mathbb{P}[F=G] \ge \alpha$ and b) $S(F;G) \le S$. Can a lower bound be tightened so that it does not vanish for all $\alpha > 0$?

The answer is negative. Indeed, consider a distribution P_{FG} on $[M]\times[M]$:

$$P_{FG}(i,j) = \frac{\alpha}{M}1\{i=j\} + \frac{1-\alpha}{M(M-1)}1\{i\ne j\}\,. \tag{30}$$

Then we have $S(F;G) = \alpha - \frac{1-\alpha}{M-1}$. That is, such P_{FG} satisfies the α-constraint and the maximal correlation constraint whenever $\alpha \le S(X;Y)$ and achieves

$$\sum_i \mathbb{P}[F=i]\mathbb{P}[G=i] = \frac{1}{M} \to 0$$

as $M \to \infty$.

It may appear that as a workaround one may consider higher spectral invariants in addition to $S(X;Y)$. Formally, to any joint distribution P_{XY}^n we associate the operator T_n as in (24). Let the singular values of T_n sorted in decreasing order be

$$1 = \sigma_{n,0} \ge \sigma_{n,1} \ge \sigma_{n,2} \ge \cdots \ge 0\,,$$

where $\sigma_{1,1} = S(X;Y)$. Since $T_n = T_1^{\otimes n}$ the singular spectrum of T_n consists of all possible products of the form $\prod_{t=1}^n \sigma_{1,j_t}$ and in particular

$$1 = \sigma_{n,0} \ge \sigma_{n,1} = \cdots = \sigma_{n,n} = S(X;Y)\,.$$

Moreover, it is easy to show that if one has a Markov chain (29) then singular values $\{\mu_j, j = 1,\ldots\}$ associated with P_{FG} are related to those of P_{XY}^n via the following "spectral-processing" inequalities:

$$\prod_{j=1}^k \mu_j \le \prod_{j=1}^k \sigma_{n,j} \qquad k = 1,\ldots\,. \tag{31}$$

Clearly this extends the data-processing for maximal correlation used in the proof of Theorem 5. Does it lead to a lower-bound non-vanishing for all α?

Alas, the answer is negative. Indeed, in the example (30) the singular spectrum associated to P_{FG} consists of 1 and $\frac{1-\alpha}{M-1}$ (of multiplicity $M-1$). This spectrum satisfies (31) as long as $\alpha \le S(X;Y)$ and $M \le n+1$. Thus, for $\alpha \le S(X;Y)$, inequalities (31) can not rule out the possibility that

$$\mathbb{Q}[F=G] \le \frac{1}{n+1}\,.$$

[2] The idea to use the maximal correlation to relate marginals and the joint distribution was first proposed by Witsenhausen [11] in the context of a slightly different problem.

B. λ-blocks of minimal weight

Another method to extend the range of α in Theorem 5 is to find a non-vanishing (as $n \to \infty$) lower bound on the marginal probability $P_{X^n}[A]$ of a λ-block (A, B). In fact, it is enough to consider the case of P_{XY} with $\mathcal{X} = \mathcal{Y}$ and $P_X = P_Y$. Indeed, consider an arbitrary λ-block (A, B) and construct a Markov kernel $W : \mathcal{X} \to \mathcal{X}$ as composition $W = P_{X|Y} \circ P_{Y|X}$, namely

$$W(x_1|x_0) = \sum_{y \in Y} P_{X|Y}(x_1|y) P_{Y|X}(y|x_0) .$$

Then distribution P_X is a stationary distribution of the Markov chain associated with W (and operator of conditional expectation (24) is self-adjoint). Moreover, we clearly have

$$W^n(A|A) \geq P^n_{X|Y} P^n_{Y|X}(B|A) \geq \lambda^2 . \tag{32}$$

And hence, it is enough to lower bound $P^n_X[A]$ among all A with the requirement that (A, A) be a λ^2-block for $W : \mathcal{X} \to \mathcal{X}$. In other words:

> *Problem (λ-blocks of minimal weight):* Given a Markov kernel $W : \mathcal{X} \to \mathcal{X}$ with stationary distribution P_X determine
>
> $$\lambda^*(p) = \lim_{n\to\infty} \max_{A: P^n_X[A] \leq p} W^n(A|A) .$$

In fact, for the purpose of extending Theorem 5 we only need to show $\lambda^*(0+) = 0$.

In the remaining we consider a special case of $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ and $W(0|1) = W(1|0) = \delta$ – a binary symmetric channel, $BSC(\delta)$. First, let us consider sets $A \subset \{0, 1\}^n$ which are linear subspaces, then denoting by Z^n a vector with i.i.d. Bernoulli(δ) components, we can easily argue that

$$W^n(A|A) = P_{Z^n}[A] \leq (1 - \delta)^{n - \dim A} ,$$

whereas on the other hand $P^n_X[A] = 2^{\dim A - n}$. Therefore, the λ-p tradeoff achievable with linear sets satisfies

$$\lambda \leq p^{\log_2 \frac{1}{1-\delta}} .$$

For the general case, consider an arbitrary set $A \subset \{0, 1\}^n$ of cardinality $|A| \leq p2^n$. Define its weight distribution as

$$\alpha_d = \frac{1}{|A|} \cdot |\{(x, y) : x \in A, y \in A, d(x, y) = d\}| ,$$

where $d(x, y)$ is the Hamming distance. Then,

$$W^n(A|A) = \sum_{d=0}^{n} \alpha_d (1 - \delta)^{n-d} \delta^d \tag{33}$$

Define $\beta_v(\alpha) = \sum_{x=0}^{n} K_v(x)\alpha_x$, a dual weight distribution of A, with $K_v(x)$ – Krawtchouk polynomials; e.g. [13, Appendix A]. By Delsarte's theorem [12], $\beta_v(\alpha) \geq 0$ and in fact by the cardinality constraint

$$\beta_0(\alpha) \leq p2^n . \tag{34}$$

Thus, we get the following linear-programming bound

$$\lambda^*_n(p) \leq \max \sum_{d=0}^{n} \alpha_d (1 - \delta)^{n-d} \delta^d , \tag{35}$$

where maximum is over all non-negative $\{\alpha_d\}$ such that $\alpha_0 = 1$, $\beta_v(\alpha) \geq 0$ and (34).

To give the dual formulation of (35) say that a polynomial $P(x)$ of degree not larger than n is admissible if

$$P(x) = \sum_{v=0}^{n} p_v K_v(x) ,$$

and $p_v \geq (1 - 2\delta)^v$ for all $v = 0, \ldots, n$. Then, we have

$$\lambda^*_n(p) \leq \min \left(2^{-n} P(0) + (p - 2^{-n}) \max_{x=1,\ldots,n} P(x) \right) ,$$

where minimum is over all admissible polynomials. The bound of Lemma 6 states

$$\lambda^*_n(p) \leq 1 + 2\delta(p - 1) , \tag{36}$$

and corresponds to choosing

$$P(x) = K_0(x) + (1 - 2\delta) \sum_{v=1}^{n} K_v(x) \tag{37}$$

As numerical evaluation of (35) shows, see Fig. 1, the bound (36) can be significantly improved. Finding a suitable admissible polynomial $P(x)$ remains an open problem.

References

[1] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[2] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," *Arxiv preprint arXiv:1102.3944*, 2011.

[3] R. Ahlswede and I. Csiszár, "Hypothesis testing with communication constraints," *IEEE Trans. Inf. Theory*, vol. 32, no. 4, pp. 533–542, Jul. 1986.

[4] T. S. Han, "Hypothesis testing with multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. 33, no. 6, pp. 759–772, Nov. 1987.

[5] T. S. Han and K. Kobayashi, "Exponential-type error probabilities for multiterminal hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 2–14, Jan. 1989.

[6] H. Shalaby and A. Papamarcou, "Multiterminal detection with zero-rate data compression," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 254–267, mar 1992.

[7] J. Tsitsiklis, "Decentralized detection by a large number of sensors," *Math. Contr. Signals, Syst.*, vol. 1, no. 2, pp. 167–182, 1988.

[8] H. Shalaby and A. Papamarcou, "A note on the asymptotics of distributed detection with feedback," *IEEE Trans. Inf. Theory*, vol. 39, no. 2, pp. 633–640, Mar. 1993.

[9] W. Tay and J. Tsitsiklis, "The value of feedback for decentralized detection in large sensor networks," in *Proc. 2011 Int. Symp. Wireless and Pervasive Comp. (ISWPC)*, Hong Kong, China, Feb. 2011, pp. 1–6.

[10] P. Gács and J. Körner, "Common information is far less than mutual information," *Prob. Contr. Inf. Theory*, vol. 2, no. 2, pp. 149–162, 1973.

[11] H. Witsenhausen, "On sequences of pairs of dependent random variables," *SIAM J. Appl. Math.*, vol. 28, pp. 100–113, 1975.

[12] P. Delsarte, "An algebraic approach to the association schemes of coding theory," *Philips Research Rep. Supp.*, no. 10, p. 103, 1973.

[13] R. McEliece, E. Rodemich, H. Rumsey, and L. Welch, "New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities," *IEEE Trans. Inf. Theory*, vol. 23, no. 2, pp. 157–166, 1977.

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

[15] O. V. Sarmanov, "A maximal correlation coefficient," *Dokl. Akad. Nauk SSSR*, vol. 121, no. 1, 1958.