Hypothesis testing via a comparator and hypercontractivity

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 checking if the functions agree. Such tests arise as part of study of fundamental limits of channel coding, and hypothesis testing with communication (rate) constraints. A simple expression for the Stein exponent is found. Connections to the Gács-Körner common information and to hypercontractivity properties of the conditional expectation operator are identified. In the case of zero Stein exponent, a non-vanishing lower bound on probability of error is established by pairing estimates of Ahslwede-Gács and Mossel et al.

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 [1], Hoeffding [2] and Blahut [3]. Knowledge of this tradeoff is important by itself and is also useful for other parts of information theory, such as channel coding [4, Section III.E] and data compression [5, Section IV.A].

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 NSF grant CCF-12-53205 and NSF Center for Science of Information (CSoI) under grant agreement CCF-09-39370. This work (Sections I,II and III) has been presented at the 2012 IEEE International Symposium on Information Theory (ISIT), Cambridge, MA, USA.

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 $X^n = (X_1, \dots, X_n)$ and $Y^n = (Y_1, \dots, Y_n)$, 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 [6]–[9]. Formally, this problem corresponds to finding the best test of the form

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

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 [10]–[12] 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:

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

where the cardinality of the common co-domain of f and g is unrestricted. This constraint is motivated by the meta-converse method [4, 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}).

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

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

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 [13]. Using a technique pioneered by Witsenhausen [14] our conference submission [15] showed that the error probability cannot decay to zero at all (even subexponentially). This was demonstrated in the special case of high confidence level and $Q_{XY} = P_X P_Y$. In this paper we complete the lower bound by invoking hypercontractivity estimates from [16] and [17].

Interestingly, the exponent E has appeared before in the context of hypothesis testing with rate constraints (1), see [7, Theorems 5 and 8], and distributed detection [11, Theorem 2]. We identify the reasons for this below and also use this correspondence to prove the strong converse for the results in [7].

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_{XY}^n 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 $n \geq 1$ and $0 \leq \alpha \leq 1$ the performance of the best possible comparator hypothesis test of confidence level α is given by

$$\tilde{\beta}_\alpha(P_{XY}^n, Q_{XY}^n) \triangleq \inf \mathbb{Q}[T = 1]\,,$$

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

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

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_{XY}^n and Q_{XY}^n, 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. [13]. 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 [13] that a pair of sets $A \in \mathcal{X}^n$ and $B \in \mathcal{Y}^n$ is called a λ-block for P_{XY}^n if $P_X^n[A] > 0$, $P_Y^n[B] > 0$ and

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

An elegant theorem of Gács and Körner states

Theorem 1 ([13]): Let P_{XY} be an indecomposable distribution on a finite $\mathcal{X} \times \mathcal{Y}$. Then for every $\lambda_n \geq \exp\{-o(n)\}$ there exists a sequence $\nu_n = o(n)$ such that for all n any λ_n-block (A, B) for P_{XY}^n satisfies

$$P_{XY}^n[A \times B] \geq \exp\{-\nu_n\} \,. \tag{4}$$

Note that the proof in [13] can be extended to some P_{XY} on infinite spaces $\mathcal{X}$ and $\mathcal{Y}$. In short, the technique of considering a Markov kernel $P_{X|Y} \circ P_{Y|X}$ as in [13] applies to any geometrically ergodic Markov process, of which there are plenty [18]. We omit the details here.

III. Error-exponent analysis

A. Stein exponent

Our main exponential result is

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

$$\lim_{n\to\infty} \frac{1}{n} \log \tilde{\beta}_\alpha(P_{XY}^n, Q_{XY}^n) = -E \,,$$

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

$$f(x^n) = 1\{x^n \notin T^n_{[P_X]}\}\,, \tag{5}$$

$$g(y^n) = 2 \cdot 1\{y^n \notin T^n_{[P_Y]}\}\,, \tag{6}$$

where $T^n_{[P]}$ denotes the set of P-typical sequences [19, Chapter 2] over the alphabet of P. Then, on one hand by typicality:

$$\mathbb{P}[f(X^n) = g(Y^n)] = P^n_{XY}[T^n_{[P_X]} \times T^n_{[P_Y]}] \tag{7}$$

$$\ge 1 - o(1)\,. \tag{8}$$

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

$$Q^n_{XY}[T^n_{[P_X]} \times T^n_{[P_Y]}] = \exp\{-nE + o(n)\}\,. \tag{9}$$

For the case of $E < \infty$, this has been demonstrated in the proof of [7, 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 [19, 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

$$\tilde{\beta}_\alpha(P^n_{XY}, Q^n_{XY}) \ge \exp\{-nD(V_{XY}||Q_{XY}) + o(n)\}\,. \tag{10}$$

If $V_{XY} \not> 0$ then we can replace V_{XY} with $(1-\epsilon)V_{XY} + \epsilon P_X P_Y$, which is everywhere positive on $\mathcal{X} \times \mathcal{Y}$, and then take a limit as $\epsilon \to 0$ in (10). Thus we assume $V_{XY} > 0$.

Denote

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

By the special case of the theorem we have

$$V_{XY}^n[A_n] \geq \exp\{-o(n)\}\,. \tag{11}$$

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

$$Q_{XY}^n[A_n] \geq \exp\{-nD(V_{XY}||Q_{XY}) + o(n)\}\,. \tag{12}$$

Optimizing the choice of V_{XY} in (12) proves (10) 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

$$\tilde{\beta}_\alpha(P_{XY}^n, Q_{XY}^n) \geq \exp\{-o(n)\}\,. \tag{13}$$

Proof: First we show that any test of level α must contain a λ-block with $\lambda \geq \frac{\alpha}{2}$. Indeed, let

$$\begin{aligned}\lambda_i &= \min\{\mathbb{P}[f(X^n) = i | g(Y^n) = i],\\ &\qquad \mathbb{P}[g(Y^n) = i | f(Y^n) = i]\} &(14)\\ \lambda_{max} &= \max_i \lambda_i\,, &(15)\end{aligned}$$

where i ranges over the (necessarily finite) co-domain of f and g. Clearly,

$$\begin{aligned}\alpha &\leq \mathbb{P}[T = 1] &(16)\\ &= \sum_i \mathbb{P}[f(X^n) = g(Y^n) = i] &(17)\\ &\leq \sum_i \lambda_i \max\{\mathbb{P}[f(X^n) = i], \mathbb{P}[g(Y^n) = i]\} &(18)\\ &\leq \lambda_{max} \sum_i \max\{\mathbb{P}[f(X^n) = i], \mathbb{P}[g(Y^n) = i]\} &(19)\\ &\leq \lambda_{max} \sum_i \mathbb{P}[f(X^n) = i] + \mathbb{P}[g(Y^n) = i] &(20)\\ &= 2\lambda_{max}\,, &(21)\end{aligned}$$

where (18) is by the definition of λ_i and Bayes rule. From (21) we have $\lambda_{max} \geq \frac{\alpha}{2}$ and there must exist an index i such that sets $\{f(X^n) = i\}$ and $\{g(X^n) = i\}$ form a λ-block for the P_{XY}^n with $\lambda \geq \frac{\alpha}{2}$.

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)

$$\mathbb{Q}[f(X^n) = i] \geq \exp\{-o(n)\}\,, \tag{22}$$

$$\mathbb{Q}[g(Y^n) = i] \geq \exp\{-o(n)\}\,. \tag{23}$$

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 (9) it is natural to expect that

$$\mathbb{Q}[f(X^n) = g(Y^n) = i] \geq \exp\{-o(n)\} \tag{24}$$

(note that marginals match and thus $E = 0$ as per (3)). Under the assumption $Q_{XY} > 0$ it is indeed straightforward to show (24) by an application of blowing-up lemma; see [9, Theorem 3].

Finally, (24) 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

$$\tilde{\beta}_1(P^n_{XY}, Q^n_{XY}) \leq 2^{-n}\,, \tag{25}$$

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 [7, Theorem 5] to the entire range $0 < \epsilon < 1$, thereby establishing the full strong converse. This result has been obtained in [9] under different assumptions on P_{XY} and Q_{XY}[1].

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

$$T = 1\{(f(X^n), Y^n) \in A\} \tag{26}$$

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 (26). Thus Theorem 2 establishes the achievability part. Conversely, for any test of the form (26) we may find sets A_0 and A_1 such that

$$\begin{aligned} \mathbb{P}[T = 1] &= \mathbb{P}[\{Y^n \in A_0, f(X^n) = 0\} \\ &\qquad \cup\{Y^n \in A_1, f(X^n) = 1\}] \qquad (27) \\ &\geq 1 - \epsilon\,. \qquad (28) \end{aligned}$$

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

$$g(y^n) = \begin{cases} 0, & y^n \in A_0 \\ 1, & y^n \in A_1 \setminus A_0 \\ 2, & \text{otherwise} \end{cases}$$

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

Then we have

$$\mathbb{P}[f(X^n) = g(Y^n)] \geq \mathbb{P}[y^n \in A_0, f(X^n) = 0] \quad (29)$$
$$\geq \frac{1-\epsilon}{2}, \quad (30)$$

and thus by Theorem 2 we conclude:

$$\mathbb{Q}[T = 1] \geq \mathbb{Q}[f(X^n) = g(Y^n)] \quad (31)$$
$$\geq \exp\{-nE + o(n)\}. \quad (32)$$

■

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 [7, 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 (25) demonstrates the contrary for our setup.

IV. Non-vanishing lower bounds via hypercontractivity

By Theorem 3 in the case when marginals of P_{XY} and Q_{XY} coincide the error cannot decay to zero exponentially. In [15] we conjectured that in the cases of matching marginals $\tilde{\beta}_\alpha(P^n_{XY}, Q^n_{XY})$ does not vanish as $n \to \infty$. In this section we prove this conjecture:

Theorem 5: Consider an indecomposable P_{XY} on a finite $\mathcal{X} \times \mathcal{Y}$ and $Q_{XY} > 0$ such that $Q_X = P_X$ and $Q_Y = P_Y$. Then there exists $r \geq 1$ such that for all $n \geq 1$ and all $0 \leq \alpha \leq 1$ we have

$$\tilde{\beta}_\alpha(P^n_{XY}, Q^n_{XY}) \geq \left(\frac{\alpha}{2}\right)^r .$$

Remark 1: Theorem is clearly not extendable to $Q_{XY} \not> 0$. Indeed, if $Q_{XY}(x_0, y_0) = 0$ then a test of sample size $n = 1$ with critical region $\{X = x_0, Y = y_0\}$, cf. (5)-(6), achieves $\tilde{\beta} = 0$ and $\alpha < 1$ with the exception of a trivial case $P_{XY}(x_0, y_0) = 1$.

Previously in [15] we showed Theorem 5 for $Q_{XY} = P_X P_Y$ and $\alpha \geq \alpha_0(P_{XY})$ by invoking the maximal correlation ideas of Witsenhausen [14]. Here we complete the proof by applying methods of hypercontractivity following the steps of Ahslwede and Gács, who improved Gács-Körner's result (4) to a non-vanishing lower bound [16]. This extension (from maximal correlation to hypercontractivity) is very popular in the modern approach in harmonic analysis and concentration of measure, where spectral gap inequality is amplified to a log-Sobolev inequality [20], which is known to be equivalent to hypercontractivity.

Proof: Let f, g be the comparator-based test and let $W = f(X^n), \hat{W} = g(Y^n)$. Then, as in (21) there exists i such that

$$\mathbb{P}[\hat{W} = i | W = i] \geq \frac{\alpha}{2} \tag{33}$$

$$\mathbb{P}[W = i | \hat{W} = i] \geq \frac{\alpha}{2} \tag{34}$$

On the other hand by [16, Theorem 1] we have for some $p > 0$ and $u < 1$

$$\mathbb{P}[\hat{W} = i] \geq \left(\frac{\alpha}{2}\right)^p \mathbb{P}[W = i]^u \,, \tag{35}$$

$$\mathbb{P}[W = i] \geq \left(\frac{\alpha}{2}\right)^p \mathbb{P}[\hat{W} = i]^u \,. \tag{36}$$

Thus, we have

$$\mathbb{P}[W = i], \mathbb{P}[\hat{W} = i] \geq \left(\frac{\alpha}{2}\right)^{\frac{p}{1-u}} \,. \tag{37}$$

On the other hand,

$$\begin{aligned}
\mathbb{Q}[W = \hat{W}] &\geq \mathbb{Q}[W = i, \hat{W} = i] &(38)\\
&\geq \mathbb{Q}[W = i]^{p_1} \mathbb{Q}[\hat{W} = i]^{q_1} &(39)\\
&= \mathbb{P}[W = i]^{p_1} \mathbb{P}[\hat{W} = i]^{q_1} &(40)\\
&\geq \left(\frac{\alpha}{2}\right)^{\frac{(p_1+q_1)p}{1-u}} &(41)
\end{aligned}$$

where (39) holds for some $p_1, q_1 > 0$ by [17, Corollary 8.2], (40) because marginals of P_{XY} and Q_{XY} coincide and (41) is by (37). ■

Acknowledgement

We are grateful to Himanshu Tyagi and Yihong Wu for useful references.

References

[1] H. Chernoff, “A measure of asymptotic efficiency for test of a hypothesis based on a sum of observations,” *Ann. Math. Stat.*, vol. 23, pp. 493–507, 1952.

[2] W. Hoeffding, “Asymptotically optimal tests for multinominal distributions,” *Ann. Math. Stat.*, vol. 36, pp. 369–400, 1965.

[3] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inf. Theory*, vol. 20, no. 4, pp. 405–417, 1974.

[4] 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.

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

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

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

[8] 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.

[9] 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.

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

[11] 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.

[12] 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.

[13] 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.

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

[15] Y. Polyanskiy, "Hypothesis testing via a comparator," in *Proc. 2012 IEEE Int. Symp. Inf. Theory (ISIT)*, Cambridge, MA, Jul. 2012.

[16] R. Ahlswede and P. Gacs, "Spreading of sets in product spaces and hypercontraction of the Markov operator," *Ann. Probab.*, pp. 925–939, 1976.

[17] E. Mossel, K. Oleszkiewicz, and A. Sen, "On reverse hypercontractivity," *Arxiv preprint arXiv:1108.1210v3*, Aug. 2012.

[18] I. Kontoyiannis and S. Meyn, "Spectral theory and limit theorems for geometrically ergodic Markov processes," *Ann. Appl. Probability*, vol. 13, no. 1, pp. 304–362, 2003.

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

[20] M. Ledoux, "Concentration of measure and logarithmic Sobolev inequalities," *Seminaire de probabilites XXXIII*, pp. 120–216, 1999.