Electronic Colloquium on Computational Complexity, Revision 3 of Report No. 153 (2014)

Communication with Imperfectly Shared Randomness*

Clément L. Canonne†, Venkatesan Guruswami‡, Raghu Meka§, Madhu Sudan¶

January 23, 2024

Abstract

Communication complexity investigates the amount of communication needed for two or more players to determine some joint function of their private inputs. For many interesting functions, the communication complexity can be much smaller than basic information theoretic measures associated with the players' inputs such as the input length, the entropy, or even the conditional entropy. Communication complexity of many functions reduces further when the players share randomness. Classical works studied the communication complexity of functions when the interacting players share randomness perfectly, i.e., they get identical copies of randomness from a common source. This work considers the variant of this question when the players share randomness imperfectly, i.e., when they get noisy copies of the randomness produced by some common source. Our main result shows that any function that can be computed by a k-bit protocol in the perfect sharing model has a 2^k-bit protocol in the setting of imperfectly shared randomness and such an exponential growth is necessary. Our upper bound relies on ideas from locality sensitive hashing while lower bounds rely on hypercontractivity and a new invariance principle tailored for communication protocols.

*An extended abstract of this work appeared in the *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science (ITCS)* [CGMS15].

†Columbia University. Email: ccanonne@cs.columbia.edu. Research supported in part by NSF CCF-1115703 and NSF CCF-1319788. Some of this work was done when the author was an intern at Microsoft Research New England.

‡Carnegie Mellon University. Email: venkatg@cs.cmu.edu. Some of this work was done when the author was visiting Microsoft Research New England. Research supported in part by NSF CCF-0963975.

§University of California, Los Angeles. Email: raghum@cs.ucla.edu .

¶Harvard University. Email: madhu@cs.harvard.edu.

ISSN 1433-8092

Contents

1 Introduction

The availability of shared randomness can lead to enormous savings in communication complexity when computing some basic functions whose inputs are spread out over different communicating players. A basic example of this is Equality Testing, where two players Alice and Bob have inputs $x \in \{0,1\}^n$ and $y \in \{0,1\}^n$ and need to determine if $x = y$. Deterministically this takes n bits of communication. This reduces to $\Theta(\log n)$ bits if Alice and Bob can toss coins and they are allowed some error. But if they share some randomness $r \in \{0,1\}^*$ independent of x and y then the communication cost drops to $O(1)$. (See, for instance, [KN06].)

A more prevalent example of a communication problem is compression with uncertain priors. Here Alice has a distribution P on a universe $[N] = \{1, \ldots, N\}$, and a message $m \in [N]$ chosen according to the distribution P. Alice is allowed to send some bits to Bob and Bob should output m and the goal is to minimize the expected number of bits that Alice sends Bob (over the random choice of m). If Bob knows the distribution P exactly then this is the classical compression problem, solved for example by Huffman coding. In most forms of natural communication (e.g., think about the next email you are about to send), Alice and Bob are not perfectly aware of the underlying context to their exchange, but have reasonably good ideas about each other. One way to model this is to say that Bob has a distribution Q that is *close* to the distribution P that Alice is working with, but is not identical to P. Compressing information down to its entropy in the presence of such uncertainty (i.e., $P \neq Q$) turns out to be possible if Alice and Bob share randomness that is independent of (P, Q, m) as shown by Juba et al. [JKKS11]. However it remains open as to whether such compression can be effected deterministically, without the shared randomness — the best known schemes can only achieve a compression length of roughly $O(H(P) + \log\log N)$, where $H(P) = \sum_{i \in [N]} P(i) \log 1/P(i)$ denotes the entropy of P.[1]

In both examples above it is natural to ask the question: can the (presumed) savings in communication be achieved in the absence of perfect sharing of randomness? The question especially makes sense in the latter context where the essential motivation is that Alice and Bob are not in perfect synchrony with each other: If Alice and Bob are not perfectly aware of the distributions P and Q, why should their randomness be identical?

The question of communication with imperfectly shared randomness was considered recently in the work of Bavarian et al. [BGI14]. They consider the setting where Alice and Bob have randomness r and s respectively, with some known correlation between r and s, and study the implications of imperfectly shared randomness in the simultaneous message communication model (where a referee gets messages from Alice and Bob and computes some joint function of their inputs). Their technical focus is on the different kinds of correlations possible between r and s, but among basic results they show that equality testing has a $O(1)$ communication complexity protocol with imperfectly shared randomness. Similar questions have been considered in other contexts and communities, such as information theory [GK73, Wit75, AC98, KA12, FRKT15, BG15], cryptography [BS93, Mau93, AC93, CN00, RW05], probability theory [MO05, MOR+06, BM11, CMN14], and quantum computing [BBP+96].

[1] We stress that the setting of uncertain compression is completely different from that of compression with the "wrong distribution", a well-studied question in information theory. In the "wrong distribution problem" (see, for instance, [CT91, Theorem 5.4.3]) the sender and receiver agree on the distribution, say P, but both have it wrong and the distribution the message comes from is R. This leads to a compression length of $\mathbb{E}_{m \sim R}[\log(1/P(m))] \approx H(R) + D(R\|P)$. The important aspect here is that while the compression is not as good, there is no confusion between sender and receiver; and the latter is the focus of our problem.

In this work we are concerned with the setting of general communication protocols, where Alice and Bob interact to determine the value of some function. From some perspectives, this setting does not seem to offer a major difference between "private randomness" and "perfectly shared randomness" — Newman [New91] shows that the communication complexity in the former setting can be larger by at most an additive $\log n$ term, where n is the input size (indeed, Newman proves that any protocol with perfectly shared randomness can be converted into one that uses only $O(\log n)$ bits of shared randomness). "Imperfectly shared randomness" being in between the two models cannot therefore be too far from them either. However, problems like compression above highlight a different perspective. There N is the size of the universe of all possible messages, and compression to $\log N$ bits of communication is trivial and uninteresting. Even a solution with $\log \log N$ bits of communication is not completely satisfactory. The real target is $O(H(P))$ bits of communication, which may be a constant independent of the universe size N (and for natural communication, the set of possible messages could be thought of as an infinitely large set). Thus the gap between the communication complexity with perfectly shared randomness and imperfectly shared randomness remains a very interesting question, which we explore in this paper.

We provide a formal description of our models and results in the following section, and here give an informal preview. We consider communication complexity in a simplified setting of imperfectly shared randomness: Alice has a uniform binary string $r \in \{0,1\}^m$ and Bob has a string s obtained by flipping each bit of r independently with some tiny probability. (While this setting is not the most general possible, it seems to capture the most interesting aspects of the "lack of prior agreement" between Alice and Bob.) Our main contributions in this work are the introduction of some new problems of interest in the context of communication complexity, and a comparison of their communication complexity with/without perfect sharing of randomness.

The first problem we study is the complexity of *compression with uncertain priors*. We show that any distribution P can be compressed to $O(H(P))$ bits even when the randomness is not perfectly shared. As in the analogous result of Juba et al. [JKKS11] this protocol sheds some light on natural communication processes, and introduces an error-correcting element that was not previously explained.

The next problem we mention is that of *agreement distillation*. Here Alice and Bob try to agree on a small random string using little communication. This is a natural problem to study in the context of communication complexity with imperfect randomness, since an efficient solution for this problem would allow Alice and Bob to convert any protocol using perfectly shared randomness into one that relies only on imperfectly shared randomness. It turns out that the zero-communication version of this question, where Alice and Bob are not allowed to communicate at all with each other, ant the one-way communication version were studied in the past. Witsenhausen [Wit75] shows for instance that no perfect agreement is possible even for a single bit, i.e. Alice and Bob must fail with positive probability. Ahlswede and Csiszar [AC98] studies the one-way communication version of this question and gives tight bounds on the number of bits that need to be communicated to get k bits of entropy with probability tending to one. Later, Bogdanov and Mossel [BM11] extend this negative result, showing that the probability that Alice and Bob can agree on a k-bit string is exponentially small in k. By a reduction we show that this implies that $o(k)$ bits of communication are insufficient to get agreement on k bits. Conversely, we also show that Alice and Bob can get a constant factor advantage — so they can communicate αk bits for some $\alpha < 1$ to obtain k bits of perfectly shared randomness with high probability. Such a result seems implicit in [BM11].

A sequence of earlier works [GK73, Wit75, AC98] also studied the agreement distillation problem

focusing on the maximum achievable ratio a/c, such that for sufficiently large number r of used correlated samples, Alice and Bob can agree on $a \cdot r$ random bits using $c \cdot r$ bits of communication. Moreover, these works focus on the case where the agreement probability tends to 1 (as $r \to \infty$). It is surprising that despite requiring the number of agreed bits to grow linearly in the number of used samples, Ahlswede and Csiszar [AC98] lose nothing in terms of the best achievable trade-off.

Following our work, Guruswami and Radhakrishnan [GR16] pinpoint the exact trade-off between communication and success probability required in order for Alice and Bob to agree on k bits of common randomness, when an unlimited number of correlated samples are available.

Returning to our work, we next attempt to get a general conversion of communication protocols from the perfectly-shared setting to the imperfectly-shared setting. We introduce a complete promise problem GapInnerProduct which captures two-way communication, and use it to show that any problem with a protocol using k bits of communication with perfectly shared randomness also has a $\min\{\exp(k), k + \log n\}$ bit (one-way) protocol with imperfectly shared randomness. While the protocol is simple, we feel its existence is somewhat surprising; and indeed it yields a very different protocol for equality testing when compared with Bavarian et al. [BGI14].

Lastly, our *main technical result* is a matching lower bound giving a parameterized family of promise problems, SparseGapInnerProduct, where the k'th problem can be solved with k bits of communication with perfect randomness, but requires $\exp(\Omega(k))$ bits with imperfect sharing. This result builds a new connection between influence of variables and communication complexity, which may be of independent interest. Finally we conclude with a variety of open questions.

2 Model, Formal Description of Results and Main Ideas

Throughout the paper, we denote by $\mathbb{Z}^+$ the set of positive integers, and by $[n]$ the set $\{1, \ldots, n\}$. Unless specified otherwise, all logarithms are in base 2. We also recall, for $x \in [0, 1]$, the definition of the binary entropy function $h(x) = -x \log x - (1 - x) \log(1 - x)$; furthermore, for any $p \in [0, 1]$, we will write $\mathrm{Bern}(p)$ for the Bernoulli distribution on $\{0, 1\}$ with parameter p, and $\mathrm{Bern}^n(p)$ for the product distribution on $\{0, 1\}^n$ of n independent Bernoulli random variables. For a distribution P over a domain Ω, we write $H(P) = \sum_{x \in \Omega} P(x) \log(1/P(x))$ for its entropy, and $x \sim P$ to indicate that x is drawn from P. $\mathcal{U}_\Omega$ denotes the uniform distribution over Ω.

Finally, for two elements $x, y \in \{+1, -1\}^n$, their *Hamming distance* $\mathrm{dist}(x, y)$ is defined as the number of coordinates in which they differ (and similarly for $x, y \in \{0, 1\}^n$).

2.1 Model

We use the familiar model of communication complexity, augmented by the notion of imperfectly shared randomness. Recall that in the standard model, two players, Alice and Bob, have access to inputs x and y respectively. A protocol Π specifies the interaction between Alice and Bob (who speaks when and what), and concludes with Alice and Bob producing outputs w_A and w_B respectively. A communication problem P is (informally) specified by conditions on the inputs and outputs (x, y, w_A, w_B). In usual (promise) problems this is simply a relationship on the 4-tuple. In sampling problems, this may be given by requirements on the distribution of this output given x and y. For functional problems, $P = (f_A, f_B)$ and the conditions require that $w_A = f_A(x, y)$ and $w_B = f_B(x, y)$. A randomized protocol is said to solve a functional problem P if the outputs are correct with probability at least 2/3. The (worst-case) complexity of a protocol Π, denoted $\mathrm{cc}(\Pi)$

is the maximum over all x, y of the expected number of bits communicated by Π. This is the main complexity measure of interest to us, although distributional complexity will also be considered, as also any mix. (For instance, the most natural measure in compression is a max-average measure.)

We will be considering the setting where Alice and Bob have access to an arbitrarily long sequence of correlated random bits. For this definition it will be convenient to let a random bit be an element of $\{+1, -1\}$. For $\rho \in [0, 1]$,[2] we say a pair of bits (a, b) are *ρ-correlated (uniform) bits* if $\mathbb{E}[a] = \mathbb{E}[b] = 0$ and $\mathbb{E}[ab] = \rho$. We will consider the performance of protocols when given access to sequences (r, r') where each coordinate pair (r_i, r'_i) are ρ-correlated uniform bits chosen independently for each i. We shall write $r \sim_\rho r'$ for such ρ-correlated pairs.

The *communication complexity of a problem P* with access to ρ-correlated bits, denoted[3] isr-cc$_\rho(P)$ is the minimum over all protocols Π that solve P with access to ρ-correlated bits of cc(Π). For integer k, we let $\mathsf{ISR\text{-}CC}_\rho(k)$ denote the collections of problems P with isr-cc$_\rho(P) \le k$. The one-way communication complexity and simultaneous message complexities[4] are defined similarly (by restricting to appropriate protocols) and denoted isr-cc$^{\mathrm{ow}}_\rho(P)$ and isr-cc$^{\mathrm{sm}}_\rho(P)$ respectively. The corresponding complexity classes are denoted similarly by $\mathsf{ISR\text{-}CC}^{\mathrm{ow}}_\rho(k)$ and $\mathsf{ISR\text{-}CC}^{\mathrm{sm}}_\rho(k)$.

Note that when $\rho = 1$ we get the standard model of communication with shared randomness. We denote this measure by psr-cc(P) = isr-cc$_1(P)$, and write $\mathsf{PSR\text{-}CC}(k)$ for the corresponding complexity class. Similarly, when $\rho = 0$ we get communication complexity with private randomness private-cc(P) = isr-cc$_0(P)$. We note that isr-cc$_\rho(P)$ is non-increasing in ρ. Combined with Newman's Theorem [New91], we obtain:

Proposition 2.1. *For every problem P with inputs $x, y \in \{0,1\}^n$ and $0 \le \rho \le \rho' \le 1$ we have*

$$\text{psr-cc}(P) \le \text{isr-cc}_{\rho'}(P) \le \text{isr-cc}_\rho(P) \le \text{private-cc}(P) \le \text{psr-cc}(P) + O(\log n).$$

The proposition also holds for one-way communication, and (except for the last inequality) simultaneous messages.

2.2 Problems, Results and Techniques

We now define some of the new problems we consider in this work and describe our main results.

2.2.1 Compression

Definition 2.2 (Uncertain Compression)**.** For $\delta > 0$, $\Delta \ge 0$ and integers ℓ, n, the *uncertain compression problem* $\textsc{Compress}^{\ell,n}_{\Delta,\delta}$ is a promise problem with Alice getting as input the pair (P, m), where $P = (P_1, \ldots, P_n)$ is a probability distribution on $[n]$ and $m \in [n]$. Bob gets a probability distribution Q on $[n]$. The promises are that $H(P) \le \ell$ and for every $i \in [n]$, $|\log(P_i/Q_i)| \le \Delta$. The goal is for Bob to output m, i.e., $w_B = m$ with probability at least $1 - \delta$. The measure of interest here is the maximum, over (P, Q) satisfying the promise, of the expected one-way communication complexity when m is sampled according to P.

[2] The definition extends to $\rho \in [-1, +1]$, but in this work we shall without loss of generality only be concerned with non-negative correlations.

[3] All throughout "isr" stands for *imperfectly shared randomness, while* psr *refers to* perfectly *shared randomness.*

[4] Recall that the *simultaneous message passing model* (SMP) [BK97] is defined as a communication game between 3 players: Alice, Bob, and a Referee. Given a function f known to all players, Alice and Bob both receive inputs respectively x and y, and send messages to the Referee who must compute the value $f(x, y)$.

When $\Delta = 0$, this is the classical compression problem and Huffman coding achieves a compression length of at most $\ell + 1$; and this is optimal for "prefix-free" compressions. For larger values of Δ, the work of [JKKS11] gives an upper bound of $\ell + 2\Delta + O(1)$ in the setting of perfectly shared randomness (to get constant error probability). In the setting of deterministic communication or private randomness, it is open if this communication complexity can be bounded by a function of ℓ and Δ alone (without dependence on n). (The work of [HS14] studies the deterministic setting.) Our first result shows that the bound of [JKKS11] can be extended naturally to the setting of imperfectly shared randomness.

Theorem 2.3. *For every $\epsilon, \delta > 0$ and $0 < \rho \leq 1$ there exists $c = c_{\epsilon,\delta,\rho}$ such that for every ℓ, n, we have* $\text{isr-cc}^{\text{ow}}_{\rho}\left(\textsc{Compress}^{\ell,n}_{\Delta,\delta}\right) \leq \frac{1+\epsilon}{1-h((1-\rho)/2)}(H(P) + 2\Delta + c)$.

We stress that the notation $\text{isr-cc}^{\text{ow}}_{\rho}\left(\textsc{Compress}^{\ell,n}_{\Delta,\delta}\right)$ describes the *worst-case* complexity over P with entropy $H(P) \leq \ell$ of the *expected* compression length when $m \sim P$. We first note that one approach would be to initially "distill" perfectly shared randomness from the imperfectly shared one available (by communicating a few bits), before using this perfectly shared randomness to run the protocol of [JKKS11]. Unfortunately, this results in $\Theta(\log n)$ bites of communication, which would be excessively large. Indeed, a naive protocol that ignores P would only require to communicate $\log n$ bits. Instead, to achieve our bound we develop a new protocol based on a simple modification of the protocol of [JKKS11]. Roughly, Alice and Bob use their imperfectly shared randomness to define a "redundant and ambiguous dictionary" with words of every length for every message. Alice communicates using a word of appropriate length given the distribution P, and Bob decodes using maximum likelihood decoding given Q. The main difference in our case is that Alice and Bob work knowing their dictionaries do not match exactly (as if they spelled the same words differently) and so use even longer words during encoding and decoding with some error-correction to allow for spelling errors. Details can be found in Section 3.

2.2.2 Agreement distillation

Next we turn to a very natural problem in the context of imperfect sharing of randomness. Can Alice and Bob communicate to distill a few random bits from their large collection r and r' (of correlated random bits), bits on which they can agree perfectly?

Definition 2.4 (Agreement distillation)**.** In the $\textsc{Agreement-Distillation}^{k}_{\gamma}$ problem, Alice and Bob have no inputs. Their goal is to output w_A and w_B satisfying the following properties:

(i) $\Pr[\, w_A = w_B \,] \geq \gamma$;

(ii) $H_\infty(w_A) \geq k$; and

(iii) $H_\infty(w_B) \geq k$

where $H_\infty(X) = \min_x \log \frac{1}{\Pr[\, X = x \,]}$ denotes the *min-entropy* of X.

The version of this problem where Alice and Bob are not allowed to communicate at all was considered by Bogdanov and Mossel [BM11]. Ahlswede and Csiszar [AC98] consider the setting where Alice and Bob must agree on the same string with high probability as the number of correlated

samples grows, and pinpoint the ratio between communication required and number of random bits agreed upon.

A trivial way to distill randomness would be for Alice to toss random coins and send their outcome to Bob. This would achieve $\gamma = 1$ and communication complexity of k for k bits of entropy. Our first proposition (of which a tighter version was obtained in [AC98]) says that with non-trivial correlation, some savings can always be achieved over this naive protocol.

Proposition 2.5. *For every $\rho > 0$, we have* $\text{isr-cc}^{\text{ow}}_{\rho}(\text{AGREEMENT-DISTILLATION}^{k}_{\gamma}) \leq (h(\frac{1-\rho}{2}) + o_k(1)) \cdot k$ *with* $\gamma = 1 - o_k(1)$. *In particular for every $\rho > 0$ there exists $\alpha < 1$ such that for every sufficiently large k* $\text{isr-cc}^{\text{ow}}_{\rho}(\text{AGREEMENT-DISTILLATION}^{k}_{1/2}) \leq \alpha k$.

For completeness, we prove this proposition in Section 4. Our next theorem says that these linear savings are the best possible: one cannot get away with $o(k)$ communication unless $\rho = 1$. For constant γ and the restriction of one-way communication, this theorem follows from [AC98]; since we are here concerned with the case where $\gamma = o(1)$ even with two-way communication, we give a proof based on Theorem 1 of [BM11] (restated as Lemma 4.1 here) and a reduction that converts protocols with communication to zero-communication protocols with a loss in γ.

Theorem 2.6. $\forall \rho < 1, \exists \epsilon > 0$ *such that* $\text{isr-cc}_{\rho}(\text{AGREEMENT-DISTILLATION}^{k}_{\gamma}) \geq \epsilon k - \frac{3}{2}\log\frac{1}{\gamma} - O(1)$.

Section 4 contains details of this proof.

2.2.3 General relationships between perfect and imperfect sharing

Our final target in this work is to get some general relationships for communication complexity in the settings of perfect and imperfectly shared randomness. Our upper bounds for communication complexity are obtained by considering a natural promise problem, that we call GAPINNERPRODUCT, which is a "hard problem" for communication complexity. We use a variant, SPARSEGAPINNERPRODUCT, for our lower bounds. We define both problems below.

Definition 2.7 ($\text{GAPINNERPRODUCT}^{n}_{c,s}$, $\text{SPARSEGAPINNERPRODUCT}^{n}_{q,c,s}$)**.** The $\text{GAPINNERPRODUCT}^{n}_{c,s}$ problem has parameters $n \in \mathbb{Z}^{+}$ (dimension), and $c > s \in [0,1]$ (completeness and soundness). Both yes- and no-instances of this problem have inputs $x, y \in \{0,1\}^n$. An instance (x, y) is a yes-instance if $\langle x, y\rangle \geq cn$, and a no-instance if $\langle x, y\rangle < sn$. The $\text{SPARSEGAPINNERPRODUCT}^{n}_{q,c,s}$ is a restriction of $\text{GAPINNERPRODUCT}^{n}_{c,s}$ where both the yes- and the no-instances are sparse, i.e., $\|x\|_2^2 \leq n/q$.

In Proposition 5.5 we show that $\text{GAPINNERPRODUCT}^{n}_{c,s}$ is "hard" for $\mathsf{PSR\text{-}CC}(k)$ with $c = (2/3)2^{-k}$ and $s = (1/3)2^{-k}$. Then in Lemma 5.6 we show that this problem is in $\mathsf{ISR\text{-}CC}^{\text{ow}}_{\rho}(\text{poly}(1/(c-s)))$. Putting the two results together we get the following theorem giving a general upper bound on $\text{isr-cc}^{\text{ow}}_{\rho}(P)$ in terms of $\text{psr-cc}(P)$ for any promise problem P.

Theorem 2.8. $\forall \rho > 0$, $\exists c < \infty$ *such that* $\forall k$, *we have* $\mathsf{PSR\text{-}CC}(k) \subseteq \mathsf{ISR\text{-}CC}^{\text{ow}}_{\rho}(c^k)$.

We prove this theorem in Section 5.2.

Theorem 2.8 is obviously tight already because of known gaps between one-way and two-way communication complexity. For instance, it is well known that the "indexing" problem (where Alice gets a vector $x \in \{0,1\}^n$ and Bob an index $i \in [n]$ and they wish to compute x_i) has one-way communication complexity of $\Omega(n)$ with perfectly shared randomness, while its deterministic two-way communication complexity is at most $\log n + 2$. However one could hope for tighter results capturing

promise problems P with low psr-cc$^{\text{ow}}(P)$, or to give better upper bounds on isr-cc(P) for P with low psr-cc(P). Our next theorem rules out any further improvements to Theorem 2.8 when n is sufficiently large (compared to k). We do so by focusing on the problem SparseGapInnerProduct. In Proposition 5.7 we show that psr-cc$^{\text{ow}}(\text{SparseGapInnerProduct}^n_{q,c,s}) = O(\text{poly}(\frac{1}{q(c-s)})\log q)$ for every q, n and $c > s$. In particular if say $c = 1/(2q)$ and $s = 1/(4q)$ the one-way communication complexity with perfectly shared randomness reduces to $O(\log q)$, in contrast to the poly(q) upper bound on the one-way communication complexity with imperfectly shared randomness from Lemma 5.6.

Our main technical theorem shows that this gap is necessary for every $\rho < 1$. Specifically in Theorem 5.8 we show that isr-cc$_\rho(\text{SparseGapInnerProduct}^n_{q,c=.9/q,s=.6/q}) = \Omega(\sqrt{q})$. Putting the two together we get a strong converse to Theorem 2.8, stated below.

Theorem 2.9. *For every k, there exists a promise problem $P = (P_n)_{n\in\mathbb{Z}^+}$ such that* psr-cc$^{\text{ow}}(P) \leq k$*, but for every $\rho < 1$ it is the case that* isr-cc$_\rho(P) = 2^{\Omega_\rho(k)}$.

Remarks on the proofs. Theorem 2.8 and Theorem 2.9 are the technical highlights of this paper and we describe some of the ideas behind them here.

Theorem 2.8 gives an upper bound for isr-cc$^{\text{ow}}_\rho$ for problems with low psr-cc. As such this ought to be somewhat surprising in that for known problems with low probabilistic communication complexity (notably, equality testing), the known solutions are very sensitive to perturbations of the randomness. But the formulation in terms of GapInnerProduct suggests that any such problem reduces to an approximate inner product calculation; and the theory of metric embeddings, and examples such as locality sensitive hashing, suggest that one can reduce the dimensionality of the problems here significantly and this may lead to some reduced complexity protocols that are also robust to the noise of the ρ-correlated vectors. This leads us to the following idea: To estimate $\langle x, y\rangle$, where $x, y \in \{0,1\}^n$, Alice can compute $a = \langle g_1, x\rangle$ where g_1 is a random n-dimensional spherical Gaussian and send a (or the most significant bits of a) to Bob. Bob can compute $b = \langle g_2, y\rangle$ and $a \cdot b$ is an unbiased estimator (up to normalization) of $\langle x, y\rangle$ if $g_1 = g_2$. This protocol can be easily shown to be robust in that if g_2 is only ρ-correlated with g_1, $a \cdot b$ is still a good estimator, with higher variance. And it is easy to convert a collection of ρ-correlated bits to ρ-correlated Gaussians, so it is possible for Alice and Bob to generate the g_1 and g_2 as desired from their imperfectly shared randomness. A careful analysis (of a variant of this protocol) shows that to estimate $\langle x, y\rangle$ to within an additive error $\epsilon\|x\|_2\|y\|_2$, it suffices for Alice to send about $1/\epsilon^2$ bits to Bob, and this leads to a proof of Theorem 2.8. Next we turn to the proof of Theorem 2.9, which shows a roughly matching lower bound to Theorem 2.8 above. The insight to this proof comes from examining the "Gaussian protocol" above carefully and contrasting it with the protocol used in the perfect randomness setting. In the latter case Alice uses the randomness to pick one (or few) coordinates of x and sends some function of these bits to Bob achieving a communication complexity of roughly $\log(1/\epsilon)$, using the fact that only $O(\epsilon n)$ bits of x are non-zero. In the Gaussian protocol Alice sends a very "non-junta"-like function of x to Bob; this seems robust to the perturbations of the randomness, but leads to $1/\epsilon^2$ bits of communication. This difference in behavior suggests that perhaps functions where variables have low "influence" cannot be good strategies in the setting of perfect randomness, and indeed we manage to prove such a statement in Theorem 6.8. The proof of this theorem uses a variant of the invariance principle that we prove (see Theorem 7.1), which shows that if a communication protocol with low-influences works in a "product-distributional" setting, it will also work with inputs being Gaussian

and with the same moments. This turns out to be a very useful reduction. The reason that SparseGapInnerProduct has nice psr-cc$^{\text{ow}}$ protocols is the asymmetry between the inputs of Alice and the inputs of Bob — inputs of Alice are sparse! But with the Gaussian variables there is no notion of sparsity and indeed Alice and Bob have symmetric inputs and so one can now reduce the "disjointness" problem from communication complexity (where now Alice and Bob hold sets $A, B \subseteq [1/\epsilon]$, and would like to distinguish $|A \cap B| = 0$ from $|A \cap B| = 1$) to the Gaussian inner product problem. Using the well-known lower bound on disjointness, we conclude that $\Omega(1/\epsilon)$ bits of communication are necessary and this proves Theorem 6.8.

Of course, all this rules out only one part of the solution space for the communication complexity problem, one where Alice and Bob use functions of low-influence. To turn this into a general lower bound we note that if Alice and Bob use functions with some very influential variables, then they should agree on which variable to use (given their randomness r and r'). Such agreement on the other hand cannot happen with too high a probability by our lower bound on Agreement-Distillation (from Theorem 2.6). Putting all these ingredients together gives us a proof of Theorem 2.9 (see Section 5.3) for more details).

Organization of the rest of the paper The rest of the paper contains details and proofs of the theorems mentioned in this section. In the next section (Section 3), we prove our isr upper bound for the "Uncertain Compression" problem, namely Theorem 2.3. We then turn, in Section 4, to the matching upper and lower bounds for "Agreement Distillation" as described in Proposition 2.5 and Theorem 2.6. Section 5 contains the details of our main results relating communication with perfectly and imperfectly shared randomness, Theorem 2.8 and Theorem 2.9: we first describe an alternate characterization of communication strategies in Section 5.1, which allows us to treat them as vectors in (carefully defined) convex sets. This enables us to use ideas and machinery from Gaussian analysis: in particular, our lower bound on isr presented in Section 6 relies on a new invariance theorem, Theorem 7.1, that we prove in Section 7.

3 Compression

In this section, we prove Theorem 2.3, restated below:

Theorem 2.3. *For every $\epsilon, \delta > 0$ and $0 < \rho \leq 1$ there exists $c = c_{\epsilon,\delta,\rho}$ such that for every ℓ, n, we have* $\text{isr-cc}^{\text{ow}}_{\rho}\left(\text{Compress}^{\ell,n}_{\Delta,\delta}\right) \leq \frac{1+\epsilon}{1-h((1-\rho)/2)}(H(P) + 2\Delta + c)$.

Proof of Theorem 2.3. Let $\mu = (1-\rho)/2$ and $\epsilon' > 0$ be such that $1/(1-h(\mu+\epsilon')) = (1+\epsilon)/(1-h(\mu))$. Let $c = O(\frac{1}{\epsilon'^2}\ln(1/\delta))$.

We interpret the random strings r and r' as two "dictionaries", i.e., as describing words $\{w_{i,j} \in \{-1,+1\}^j\}_{i\in[n],j\in\mathbb{Z}^+}$ and $\{w'_{i,j} \in \{-1,+1\}^j\}_{i\in[n],j\in\mathbb{Z}^+}$, with the property that for every i, j and coordinate $k \in [j]$, the kth coordinates of $w_{i,j}$ and $w'_{i,j}$ are ρ-correlated.
On input P, m Alice sends $X = w_{m,j}$ to Bob where $j = \max\{c, \frac{1+\epsilon}{1-h(\mu)}(\log(1/P(m)) + 2\Delta + \log(1/\delta))\}$.
On input Q and on receiving X from Alice, Bob computes $j = |X|$ and the set

$$S_X = \left\{ \tilde{m} : \text{dist}(w'_{\tilde{m},j}, X) \leq (\mu + \epsilon')j \right\} ,$$

where dist denotes the Hamming distance between strings. Bob then outputs $\text{argmax}_{\tilde{m}\in S_X}\{Q(\tilde{m})\}$ (so it outputs the most likely message after some error-correction).

It is clear from construction that the expected length of the communication when $m \sim P$ is at most

$$\mathbb{E}_{m\sim P}\left[\frac{1+\epsilon}{1-h(\mu)}\left(\log(1/P(m))+2\Delta+c\right)\right] = \frac{1+\epsilon}{1-h(\mu)}\left(\mathbb{E}_{m\sim P}[\log(1/P(m))]+2\Delta+c\right) = \frac{1+\epsilon}{1-h(\mu)}\left(H(P)+2\Delta+c\right).$$

We finally turn to correctness, i.e., to show that Bob's output $\tilde{m} = m$ with probability at least $1-\delta$. First note that the probability that $m \in S_X$ is at least $(1-\delta/2)$ (by a simple application of Chernoff bounds and the fact that j is sufficiently large compared to ϵ' and δ). Now let $T_m = \left\{\, m' \neq m :\ P(m') \geq P(m)/4^{\Delta} \,\right\}$. Note that $|T_m| \leq 4^{\Delta}/P(m)$. For any fixed $m' \in T_m$, we have that the probability (over the choice of $w'_{m',j}$) that $m' \in S_X$ is at most $2^{-(1-h(\mu+\epsilon'))j}$. Indeed, since $m' \neq m$, the two sets of indices corresponding to $w'_{m',j}$ and $w_{m,j}$ (in the "random bit dictionaries" r, r') are disjoint, and so $w'_{m',j}$, $w_{m,j}$ are independent. Now, for any two independent u.a.r. u, v, the probability that $\mathrm{dist}(u,v) \leq \alpha j$ is given by

$$\sum_{i\leq \alpha j}\binom{j}{i}\frac{1}{2^i}\frac{1}{2^{j-i}} = \frac{1}{2^j}\sum_{i\leq\alpha j}\binom{j}{i} \leq \frac{1}{2^j}\sum_{i\leq\alpha j}\binom{j}{i} \leq 2^{-j}2^{h(\alpha)j}.$$

Taking the union bound over $m' \in T_m$ and plugging in our choice of j, we have that with probability at least $1-\delta/2$, $T_m \cap S_X = \emptyset$. With probability at least $1-\delta$ both events above happen and when they do, as $m \in S_X$ satisfies $Q(m) \geq \frac{P(m)}{2^{\Delta}}$ and any other element $m' \in S_X$ is such that $Q(m') \leq 2^{\Delta}P(m') < 2^{\Delta}P(m)/4^{\Delta}$, we have $\tilde{m} = m$. □

4 Agreement Distillation

In this section we give proofs of Proposition 2.5 and Theorem 2.6 which respectively give upper and lower bounds on the one-way communication complexity of randomness distillation.

We start with the upper bound, which relies on the existence of linear error-correcting codes capable of correcting $\mu \triangleq \frac{1-\rho}{2}$ fraction errors. The fact that such codes have rate approaching $1-h(\mu)$ yields the result that agreement distillation requires $(1+o_k(1))\cdot h(\mu)\cdot k$ communication for $\gamma \to 1$. Details below.

Proof of Proposition 2.5. Let $\epsilon > 0$ be any positive constant and let $\mathrm{Bern}^k(\mu)$ be the distribution on $\{0,1\}^k$ where each bit is independent and is 1 with probability μ. Our proof protocol will rely on the existence of a certain matrix $H \in \{0,1\}^{\ell\times k}$ over $\mathbb{F}_2$ satisfying the following property:

$$\Pr_{e\sim\mathrm{Bern}^k(\mu)}\left[\exists e' \neq e \text{ s.t. } \mathrm{wt}(e') \leq (\mu+\epsilon)k \text{ and } H\cdot e' = H\cdot e\right] \leq \delta/2 \tag{1}$$

(we will establish the existence of such a matrix later, for $\ell = h(\mu+\epsilon)k$).

With this matrix H in hand, Alice and Bob act as follows. Given ρ correlated strings $r, r' \in \{0,1\}^k$, Alice's output is $w_A = r$. She communicates $y = H\cdot r$ to Bob. Bob's output is $w_B = \tilde{r}$ such that (i) $H\cdot\tilde{r} = y$ and (ii) $\mathrm{dist}(\tilde{r}, r') \leq (\mu+\epsilon)k$, provided $\tilde{r}$ with these properties exists and is unique. Else he outputs r'.

It follows that unless $\mathrm{dist}(r, r') > (\mu+\epsilon)k$ or if $\exists e' \neq e \triangleq r - r'$ such that $\mathrm{wt}(e') \leq (\mu+\epsilon)k$ and $H \cdot e' = H \cdot e$, we have $\tilde{r} = r$. (Indeed, this is a consequence of the above property of H, writing $e' \triangleq \tilde{r} - r'$ and observing that $\mathrm{wt}(e') = \mathrm{dist}(\tilde{r}, r')$, and that $He' = He$ if and only if $H \cdot \tilde{r} = H \cdot r$.) The probability of either event above is small (by Chernoff bound for the first, and by the condition on H for the second).

It remains to prove the existence of a matrix H satisfying (1), for small $\ell \in \mathbb{Z}^+$. We will actually show that a random matrix satisfies this condition for $\ell = h(\mu+\epsilon)k$ with probability tending to 1 as k goes to ∞: indeed, fixing any non-zero vector $x \in \{0,1\}^k$, a random matrix H obtained by setting independently each coefficient to be 1 with probability 1/2 satisfies $\Pr\left[Hx = 0^\ell\right] = 2^{-\ell}$. The claim follows from a union bound over the (at most) $\sum_{i \leq (\mu+\epsilon)k} \binom{k}{i} \leq 2^{h(\mu+\epsilon)k - \omega_k(1)}$ vectors e'. □

We now turn towards a proof of Theorem 2.6. We first consider the setting of *zero* communication, i.e., when Alice and Bob are not allowed to communicate at all. Here we use the following lemma due to [BM11] which shows that the agreement probability γ is exponentially small in k.

Lemma 4.1 ([BM11, Theorem 1]). *$\forall \rho < 1, \exists \epsilon > 0$ such that for every zero-communication protocol for* $\textsc{Agreement-Distillation}^k_\gamma$*, we have $\gamma \leq 2^{-\epsilon k}$. (Furthermore, one can take $\epsilon = 1 - O(\rho)$).*

We now derive Theorem 2.6 as a corollary of Lemma 4.1.

Proof of Theorem 2.6. Suppose Π is a c-bit communication protocol for $\textsc{Agreement-Distillation}^k_\gamma$. We claim we can convert Π to a zero-bit communication protocol Π' for $\textsc{Agreement-Distillation}^k_{\gamma'}$, where the agreement γ' is $\mathrm{poly}(\gamma, 2^{-c})$. To do so, let (r, s) denote the randomness inputs to Alice and Bob. Let $\Pi_A(r, s)$ and $\Pi_B(r, s)$ denote Alice's and Bob's outputs, respectively. The zero-communication protocol Π' is obtained as follows.

- $\Pi'_A(r)$: Alice samples s' conditioned on r and outputs $\Pi_A(r, s')$;
- $\Pi'_B(s)$: Bob similarly samples r' conditioned on s and outputs $\Pi_B(r', s)$.

Since the input distributions in the invocation of Π_A and Π_B (by Π'_A and Π'_B, respectively) are exactly the same as in the c-communication protocol, the entropy of the output is unchanged. So it suffices to argue that agreement happens with probability $\mathrm{poly}(\gamma, 2^{-c})$.

Let $P(r, s)$ denote the probability of the input being (r, s). Fix the private randomness $\theta \stackrel{\text{def}}{=} (\theta_A, \theta_B)$ of Alice and Bob, and let γ_θ be the agreement probability for this fixed choice of the randomness of Alice and Bob. From now on we can consider Π to be a deterministic function of r and s. Let $t(r, s)$ denote the transcript under Π on input (r, s), and $Q_r(t)$ (resp. $Q_s(t)$) denote the probability that the transcript of Π equals $t \in \{0,1\}^c$ conditioned on Alice's input being r (resp. Bob's input being s). Let G be the subset of values of (r, s) on which Π has agreement, so that $\sum_{(r,s) \in G} P(r, s) = \gamma_\theta$.

Note that the agreement γ'_θ under Π' (on randomness θ) can be lower bounded by $\gamma'_\theta \geq \sum_{(r,s) \in G} P(r, s) \cdot Q_r(t(r, s)) \cdot Q_s(t(r, s))$, which is the probability with which Alice and Bob generate the same transcript in Π' as they would under Π. Now, had all the $Q_r(t)$'s been equal to 2^{-c}, we would be immediately done with a lower bound of $\gamma_\theta / 4^{-c}$. We shall obtain a slightly worse bound in order to handle non-uniform distributions.

Say the transcript t is *unlikely for* r if $Q_r(t) < \frac{\gamma_\theta}{4} 2^{-c}$. We consider the set of "unlikely randomness" $B \stackrel{\text{def}}{=} \{ (r,s) : t(r,s) \text{ is unlikely for } r \text{ or is unlikely for } s \}$. We first note that $\sum_{(r,s)\in B} P(r,s) < \frac{\gamma_\theta}{2}$. To see this, observe that

$$\begin{aligned}
\sum_{(r,s):t(r,s) \text{ unlikely for } r} P(r,s) &= \sum_r \sum_{t:t \text{ unlikely for } r} \sum_{s:t(r,s)=t} P(r,s) \\
&= \sum_r P(r) \sum_{t:t \text{ unlikely for } r} Q_r(t) \\
&< \sum_r P(r) \sum_{t:t \text{ unlikely for } r} \frac{\gamma_\theta}{4} 2^{-c} \\
&< \frac{\gamma_\theta}{4}
\end{aligned}$$

Similarly, we have $\sum_{(r,s):t(r,s) \text{ unlikely for } s} P(r,s) < \frac{\gamma_\theta}{4}$, yielding the claimed bound. From there, we can write

$$\begin{aligned}
\gamma'_\theta &\geq \sum_{(r,s)\in G} P(r,s) \cdot Q_r(t(r,s)) \cdot Q_s(t(r,s)) \\
&\geq \sum_{(r,s)\in G\setminus B} P(r,s) \cdot Q_r(t(r,s)) \cdot Q_s(t(r,s)) \\
&\geq \sum_{(r,s)\in G\setminus B} P(r,s) \cdot \left(\frac{\gamma_\theta}{4} 2^{-c}\right)^2 \\
&= \frac{\gamma_\theta^2}{16} 4^{-c} \left(\sum_{(r,s)\in G} P(r,s) - \sum_{(r,s)\in B} P(r,s) \right) \\
&\geq \frac{\gamma_\theta^3}{32} 4^{-c}
\end{aligned}$$

Finally, we take expectations over the private randomness θ: this leads to

$$\mathbb{E}_\theta[\gamma'_\theta] \geq \mathbb{E}_\theta \left[\frac{\gamma_\theta^3}{32} \cdot 4^{-c} \right] \geq \frac{\mathbb{E}_\theta[\gamma_\theta]^3}{32} \cdot 4^{-c} = \frac{\gamma^3}{32} \cdot 4^{-c},$$

as claimed. Applying Lemma 4.1, we get that for some constant $\epsilon' = 1 - O(\rho) \in (0,1)$, $2^{-2c} \frac{\gamma^3}{32} \leq 2^{-\epsilon' k}$ and thus $c \geq \frac{\epsilon'}{2} k - \frac{3}{2} \log \frac{1}{\gamma} - O(1)$ as desired. Taking $\epsilon \stackrel{\text{def}}{=} \frac{\epsilon'}{2}$ concludes the proof. □

5 General connection between perfectly and imperfectly shared randomness

In this section we present proofs of Theorem 2.8 and Theorem 2.9. Key to both our upper bound on isr-cc$^{\text{ow}}(P)$ in terms of psr-cc(P), and our lower bound on isr-cc(SparseGapInnerProduct), is a representation of communication strategies as vectors, where the success probability of an interaction is proportional to the inner product of these vectors. We describe this representation in Section 5.1 below. We then use this representation to show that GapInnerProduct is hard for

PSR-CC(k) in Section 5.2. We also give a one-way isr protocol for GapInnerProduct in the same section thus giving a proof of Theorem 2.8. Finally in Section 5.3 we give a one-way psr protocol for SparseGapInnerProduct, and then state our main technical result — an exponentially higher lower bound for it in the two-way isr setting (with the proof deferred to Section 6 modulo an invariance principle which is established in Section 7). The lower bound uses the fact that the space of strategies in the vector representation forms a bounded convex set.

5.1 Communication Strategies: Inner Products and Convexity

We start by formalizing deterministic and probabilistic (private-coin) two-way communication strategies for Alice and Bob. By "strategy" we mean what Alice would do given her input and randomness, as a function of different messages that Bob may send her, and vice versa. We restrict our attention to canonical protocols in which Alice and Bob strictly alternate and communicate one bit per round; and the eventual outcome is a Boolean one, determined after k rounds of communication. (So the only problems that can be solved this way are "promise problems".) Without loss of generality we also assume that the last bit communicated is the output of the communication protocol.

The natural way to define deterministic strategies would be in terms of a triple (f_A, f_B, v) where $f_A = (f_A^{2i}\colon \{0,1\}^{2i} \to \{0,1\})_{0\le i<k/2}$ is a sequence of functions and so is $f_B = (f_B^{2i+1}\colon \{0,1\}^{2i+1} \to \{0,1\})_{0\le i<k/2}$ and $v\colon \{0,1\}^k \to \{0,1\}$. The function $f_A^{2i}(h)$ determines Alice's message bit after $2i$ rounds of communication, with $h \in \{0,1\}^{2i}$ being the transcript of the interaction thus far. Similarly the functions $f_B^{2i+1}(h)$ determine Bob's message bit after $2i+1$ rounds of communication. Finally, v denotes the verdict function. Since we assumed that the last bit transmitted is the output, we have $v(\ell_1,\ldots,\ell_k) = \ell_k$. Thus the output of an interaction is given by $v(\ell)$ where $\ell = (\ell_1,\ldots,\ell_k)$ is given by $\ell_{2i+1} = f_A^{2i}(\ell_1,\ldots,\ell_{2i})$ and $\ell_{2i+2} = f_B^{2i+1}(\ell_1,\ldots,\ell_{2i+1})$ for $0 \le i \le k/2$. The interpretation is that Alice can determine the function f_A from her input and Bob can determine f_B from his input, and this allows both to determine the output after k rounds of interaction.

We will be moving on to the vector representation of strategies shortly, but first we describe probabilistic interactions, where Alice and Bob have private randomness.[5] Such an interaction is also described by a triple (f_A, f_B, v) except that now $f_A = (f_A^{2i}\colon \{0,1\}^{2i} \to [0,1])_{0\le i<k/2}$ and $f_B = (f_B^{2i+1}\colon \{0,1\}^{2i+1} \to [0,1])_{0\le i<k/2}$. The outcome is now the random variable $v(\ell)$ where $\ell = (\ell_1,\ldots,\ell_k)$ is the random variable determined inductively by letting $\ell_{2i+1} = 1$ with probability $f_A^{2i}(\ell_1,\ldots,\ell_{2i})$ and $\ell_{2i+2} = 1$ with probability $f_B^{2i+1}(\ell_1,\ldots,\ell_{2i+1})$ for $0 \le i \le k/2$.

Our vector representation of deterministic interactions is obtained by considering the set of "plausible final transcripts" that a player might see given their own strategy. Recall that the transcript of an interaction is a k-bit string and there are 2^k possible transcripts. In the new representation, we represent Alice's strategy (i.e., the functions f_A) by a vector $\xi_A \in \{0,1\}^{2^k}$ where $\xi_A(\ell) = 1$ if and only if $\ell \in \{0,1\}^k$ is a transcript *consistent* with Alice's strategy. (We give a more formal description shortly.) For probabilistic communication strategies (corresponding to Alice and Bob working with private randomness), we represent them by vectors ξ_A and ξ_B in $[0,1]^{2^k}$. We formalize the set of such strategies, and verdicts, below.

In what follows we describe subsets of $[0,1]^{2^k}$ that are supposed to describe the strategy space for

[5] Since we do not concern ourselves with computational complexity of the protocol, we can assume without loss of generality that the players use fresh randomness at every stage. Indeed, both parties, if willing to rely on some of their respective "previous randomness," can instead sample new random bits conditioned on the past transcript.

Alice and Bob. Roughly, we wish to allow $\xi_A = (\xi_A(i_1,\ldots,i_k))_{i_1,\ldots,i_k\in\{0,1\}}$ to be an "Alice strategy" (i.e., a member of K_A) if for every $i_1,\ldots,i_k$ there exists a Bob strategy such that Alice reaches the transcript $i_1,\ldots,i_k$ with probability $\xi_A(i_1,\ldots,i_k)$. To describe this set explicitly we introduce auxiliary variables $\chi_A(i_1,\ldots,i_j)$ for every $0 \le j \le k$ and $i_1,\ldots,i_j \in \{0,1\}$ where $\chi_A(i_j,\ldots,i_j)$ denotes the probability (again maximized over Bob strategies) of reaching the partial transcript $i_1,\ldots,i_j$. In what follows we first show that the auxiliary variables are linear forms in ξ_A and then show the conditions that the auxiliary variables satisfy. (We warn the reader that the first step — showing that the $\chi_A(\cdots)$'s are linear forms in ξ_A – relies on the constraints imposed later and so some of the definition may be slightly non-intuitive.) Together the two steps allows us to show that the space of strategies is a (closed) convex set.

Definition 5.1. We define the *partial transcript operators* $\mathrm{PT}_A, \mathrm{PT}_B \colon [0,1]^{2^k} \to ([0,1]^{\{0,1\}^j})_{0\le j\le k}$, which map a vector $\xi \in [0,1]^{2^k}$ to respectively Alice and Bob strategies $\mathrm{PT}_A(\xi), \mathrm{PT}_B(\xi)$. For vector $\xi \in [0,1]^{2^k}$, $0 \le j \le k$, and $i_1,\ldots,i_j \in \{0,1\}$ let $(\mathrm{PT}_A(\xi))_j(i_1,\ldots,i_j) = \chi_A(i_1,\ldots,i_j) \in [0,1]$ and $(\mathrm{PT}_B(\xi))_j(i_1,\ldots,i_j) = \chi_B(i_1,\ldots,i_j) \in [0,1]$ be defined as follows:

$$\chi_A(i_1,\ldots,i_j) = \begin{cases} \xi(i_1,\ldots,i_k) & \text{if } j = k \\ \chi_A(i_1,\ldots,i_j,0) + \chi_A(i_1,\ldots,i_j,1) & \text{if } j \text{ is even.} \\ \frac{1}{2}\left(\chi_A(i_1,\ldots,i_j,0) + \chi_A(i_1,\ldots,i_j,1)\right) & \text{if } j \text{ is odd.} \end{cases}$$

$$\chi_B(i_1,\ldots,i_j) = \begin{cases} \xi(i_1,\ldots,i_k) & \text{if } j = k \\ \frac{1}{2}\left(\chi_B(i_1,\ldots,i_j,0) + \chi_B(i_1,\ldots,i_j,1)\right) & \text{if } j \text{ is even.} \\ \chi_B(i_1,\ldots,i_j,0) + \chi_B(i_1,\ldots,i_j,1) & \text{if } j \text{ is odd.} \end{cases}$$

Define

$$\tilde{K}_A = \tilde{K}_A^{(k)} = \left\{ \xi \in [0,1]^{2^k} \,:\, \chi_A() = 1 \text{ and } \forall \text{ odd } j, \forall i_1,\ldots,i_j \in \{0,1\},\ \ \chi_A(i_1,\ldots,i_j,0) = \chi_A(i_1,\ldots,i_j,1) \right\},$$

and

$$\tilde{K}_B = \tilde{K}_B^{(k)} = \left\{ \xi \in [0,1]^{2^k} \,:\, \chi_B() = 1 \text{ and } \forall \text{ even } j, \forall i_1,\ldots,i_j \in \{0,1\},\ \ \chi_B(i_1,\ldots,i_j,0) = \chi_B(i_1,\ldots,i_j,1) \right\}.$$

Let $K_A = \left\{ \xi * v \,:\, \xi \in \tilde{K}_A \right\}$, where $v \in \{0,1\}^{2^k}$ is given by $v(i_1,\ldots,i_k) = i_k$ (and $a * b$ denotes coordinate-wise multiplication of vectors a and b). Let $\tilde{S}_A = \tilde{K}_A \cap \{0,1\}^{2^k}$, $\tilde{S}_B = \tilde{K}_B \cap \{0,1\}^{2^k}$, $S_A = K_A \cap \{0,1\}^{2^k}$, and $S_B = K_B \cap \{0,1\}^{2^k}$.

In what follows we first focus on deterministic communication strategies and show that $\tilde{S}_A$, $\tilde{S}_B$ correspond to the space of deterministic communication strategies for Alice and Bob, while S_A and S_B correspond to outputs computed by such strategies. This step is not strictly needed for this paper since our main focus is on probabilistic strategies and the convex sets K_A and K_B, but we include it for completeness.

Proposition 5.2. . *$\tilde{S}_A$ and $\tilde{S}_B$ correspond to the set of deterministic communication strategies with k bits. For every strategy f_A of Alice there exist vectors $\tilde{\xi}_A \in \tilde{S}_A$ and $\xi_A \in S_A$ and for every strategy f_B of Bob there exist vectors $\tilde{\xi}_B \in \tilde{S}_B$ and $\xi_B \in S_B$ such that if $\ell \in \{0,1\}^k$ is the transcript*

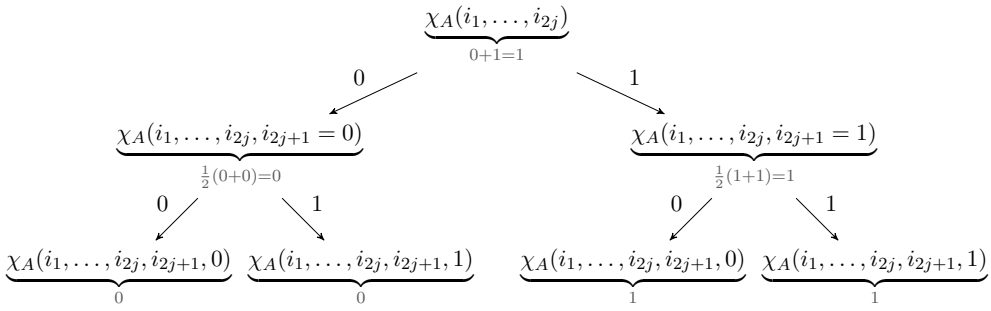

Figure 1: Illustration of the constraints on χ_A (Definition 5.1).

of the interaction between Alice and Bob under strategies f_A and f_B, then ℓ is the unique sequence satisfying $\tilde{\xi}_A(\ell) = \tilde{\xi}_B(\ell) = 1$ and $\langle \xi_A, \xi_B \rangle = 1$ if the interaction accepts and 0 otherwise.

Conversely every vector $\xi_A \in S_A$ corresponds to a strategy f_A for Alice (and similarly for Bob) such that Alice and Bob accept the interaction iff $\langle \xi_A, \xi_B \rangle = 1$.

Proof. Given f_A to construct $\tilde{\xi}_A$, we let $\tilde{\xi}_A(\ell) = 1$ if there exists $f_{B,\ell}$ such that the final transcript of the interaction given by f_A and $f_{B,\ell}$ is ℓ. Furthermore let $\tilde{\chi}_A(i_1,\ldots,i_j) = 1$ if there exists a Bob strategy $f_{B,i_1,\ldots,i_j}$ such that $i_1,\ldots,i_j$ is the partial transcript of the interaction between Alice and Bob; otherwise let $\tilde{\chi}_A(i_1,\ldots,i_j) = 0$. It is now straightforward to verify that the $\tilde{\chi}_A(i_1,\ldots,i_j)$ satisfy the conditions of the definition of $\tilde{\xi}_A$ and the conditions required for membership in $\tilde{K}_A$. In particular we have the following three conditions: (1) $\tilde{\chi}_A() = 1$ since the empty transcript is a legal partial transcript. (2) If j is an even index (and so Alice speaks in round $j+1$) and $\chi_A(i_1,\ldots,i_j) = 0$ (so the partial transcript $i_1,\ldots,i_j$ is not reachable given Alice's strategy), then we must have $\chi_A(i_1,\ldots,i_j,0) = \chi_A(i_1,\ldots,i_j,1) = 0$ (no extension is reachable either). If $\chi_A(i_1,\ldots,i_j) = 1$ then exactly one of the extensions must be reachable (based on Alice's message at this stage) and so again we have $\chi_A(i_1,\ldots,i_j) = \chi_A(i_1,\ldots,i_j,0) + \chi_A(i_1,\ldots,i_j,1)$. (3) If j is odd and it is Bob's turn to speak, then again if $\chi_A(i_1\ldots,i_j) = 0$ we have $\chi_A(i_1,\ldots,i_j,0) = \chi_A(i_1,\ldots,i_j,1) = 0$. On the other hand if $\chi_A(i_1,\ldots,i_j) = 1$ then for each extension there exists a strategy of Bob that permits this extension and so we have $\chi_A(i_1,\ldots,i_j,0) = \chi_A(i_1,\ldots,i_j,1) = 1$ satisfying the condition for odd j. The above three conditions verify membership in $\tilde{K}_A$ and since $\tilde{\xi}_A$ is a 0/1 vector, we also have $\tilde{\xi}_A \in \tilde{S}_A$. The vector $\xi_A = \tilde{\xi}_A * v$ gives the corresponding vector in S_A.

Given a pair of strategies f_A and f_B, let $\tilde{\xi}_A$ and $\tilde{\xi}_B$ be the corresponding vectors representing the strategies, and let $\tilde{\chi}_A(i_1,\ldots,i_j) = (\mathrm{PT}_A(\tilde{\xi}_A)_j)(i_1,\ldots,i_j)$ and $\tilde{\chi}_B(i_1,\ldots,i_j) = (\mathrm{PT}_A(\tilde{\xi}_B)_j)(i_1,\ldots,i_j)$ denote the partial transcripts. We now prove the existence and uniqueness of a leaf ℓ such that $\tilde{\xi}_A(\ell) = \tilde{\xi}_B(\ell) = 1$. We do so by showing, by induction on j, that for every $j \in \{0,\ldots,k\}$ there exists a unique sequence $(i_1,\ldots,i_j)$ such that $\tilde{\chi}_A(i_1,\ldots,i_j) = \tilde{\chi}_B(i_1,\ldots,i_j) = 1$. Using the fact that $\tilde{\chi}_A(i_1,\ldots,i_k) = \tilde{\xi}_A(i_1,\ldots,i_k)$, we get the desired existence and uniqueness (for $\ell = (i_1,\ldots,i_k)$). We refer to a sequence $(i_1,\ldots,i_j)$ as a valid (partial) transcript if $\tilde{\chi}_A(i_1,\ldots,i_j) = \tilde{\chi}_B(i_1,\ldots,i_j) = 1$ and as an invalid transcript otherwise. The base case of the induction is true since there is only one transcript of length 0 and we have $\tilde{\chi}_A() = \tilde{\chi}_B() = 1$ so the empty transcript is the unique valid transcript. Now assume the statement is true for transcripts of length $j-1$ and let $(i_1,\ldots,i_{j-1})$ be the unique valid transcript of length $j-1$. Now for every other sequence $(i'_1,\ldots,i'_{j-1}) \neq (i_1,\ldots,i_{j-1})$ at least one of $\tilde{\chi}_A(i_1,\ldots,i_{j-1})$ or $\tilde{\chi}_B(i_1,\ldots,i_{j-1})$ is zero. Suppose $\tilde{\chi}_A(i_1,\ldots,i_{j-1}) = 0$. Then by the previous paragraph we have both $\tilde{\chi}_A(i_1,\ldots,i_{j-1},0) = 0$ and $\tilde{\chi}_A(i_1,\ldots,i_{j-1},1) = 0$. So none

of the "children" of invalid transcripts are valid. We turn to the unique valid sequence of length $j-1$. Suppose j is an even index (and so Bob speaks in round j). Since $\tilde{\chi}_A(i_1,\ldots,i_{j-1}) = 1$ we must have $\tilde{\chi}_A(i_1,\ldots,i_{j-1},0) = \tilde{\chi}_A(i_1,\ldots,i_{j-1},1) = 1$ (since there exists a Bob strategy making each possible transcript valid). On the other hand there must exist a unique Bob message $b = f_B^{j-1}(i_1, ldots, i_{j-1}) \in \{0,1\}$ in round j given the transcript $(i_1,\ldots,i_{j-1})$ thus far. For this b we have $\tilde{\chi}_A(i_1,\ldots,i_{j-1},b) = 1$ and $\tilde{\chi}_A(i_1,\ldots,i_{j-1},1-b) = 0$, making $(i_1,\ldots,i_{j-1},b)$ the unique valid transcript at level j.

Finally we prove the converse showing that every vector $\xi_A \in S_A$ corresponds to a strategy f_A. The key step here is to show that there exists a vector $\tilde{\xi}_A \in \tilde{S}_A$ such that $\xi_A = \tilde{\xi}_A * v$. The strategy f_A can then be read off from $\tilde{\chi}_A = \mathrm{PT}_A(\tilde{\xi}_A)$. By definition of membership in $\xi_A \in S_A \subseteq K_A$ we have that there exists $\tilde{\xi}' \in \tilde{K}_A$ such that $\xi_A = \tilde{\xi}' * v$. Let $\tilde{\chi}_{A,\tilde{\xi}'} = \mathrm{PT}_A(\tilde{\xi}')$. We show how to use this to create a Boolean vector with the same property. We first define a function $\chi_1(i_1,\ldots,i_j) \in \{0,1,?\}$ as follows: We define $\chi_1(i_1,\ldots,i_k) = \xi_A(i_1,\ldots,i_k)$ if $v(i_1,\ldots,i_k) = 1$ and $\chi_1(i_1,\ldots,i_k) =?$ if $v(i_1,\ldots,i_k) = 0$. For j going down from $k-1$ to 0 we proceed as follows: If j is even (and it is Alice's turn to speak), then we set $\chi_1(i_1,\ldots,i_j) = 1$ if $\chi_1(i_1,\ldots,i_j,b) = 1$ for some $b \in \{0,1\}$, we set $\chi_1(i_1,\ldots,i_j) = 0$ if $\chi_1(i_1,\ldots,i_j,b) = 0$ for every $b \in \{0,1\}$, and we set $\chi_1(i_1,\ldots,i_j) =?$ otherwise. If j is odd (and it is Bob's turn to speak), then we set $\chi_1(i_1,\ldots,i_j) = 1$ if $\chi_1(i_1,\ldots,i_j,b) = 1$ for some $b \in \{0,1\}$, else we set $\chi_1(i_1,\ldots,i_j) = 0$ if $\chi_1(i_1,\ldots,i_j,b) = 0$ for some $b \in \{0,1\}$, and we set $\chi_1(i_1,\ldots,i_j) =?$ otherwise. We now assert, by downward induction on j that if $\chi_1(i_1,\ldots,i_j) \in \{0,1\}$ then $\chi_1(i_1,\ldots,i_j) = \tilde{\chi}_{A,\tilde{\xi}'}(i_1,\ldots,i_j)$. This is true trivially for $j = k$. For odd $j < k$, if $\chi_1(i_1,\ldots,i_j,b) \in \{0,1\}$ for some $b \in \{0,1\}$, then $\tilde{\chi}_{A,\tilde{\xi}'}(i_1,\ldots,i_j) = \tilde{\chi}_{A,\tilde{\xi}'}(i_1,\ldots,i_j,1-b) = \tilde{\chi}_{A,\tilde{\xi}'}(i_1,\ldots,i_j,b)\chi_1(i_1,\ldots,i_j,b)$ by induction and so we have $\tilde{\chi}_{A,\tilde{\xi}'}(i_1,\ldots,i_j) = \chi_1(i_1,\ldots,i_j)$. In all other cases $\chi_1(i_1,\ldots,i_j) =?$. Note also that we always have $\chi_1(i_1,\ldots,i_j,b) = \chi_1(i_1,\ldots,i_j,1-b)$ when both are in $\{0,1\}$. For even $j < k$, we reason similarly. We first note that both $\chi_1(i_1,\ldots,i_j,0)$ and $\chi_1(i_1,\ldots,i_j,1)$ can't be 1, since then we would have $\tilde{\chi}_{A,\tilde{\xi}'}(i_1,\ldots,i_j,0) = \tilde{\chi}_{A,\tilde{\xi}'}(i_1,\ldots,i_j,1) = 1$ and this would imply $\tilde{\chi}_{A,\tilde{\xi}'}(i_1,\ldots,i_j) = 2$. We conclude that if $\chi_1(i_1,\ldots,i_j) = 1$ then $\chi_1(i_1,\ldots,i_j,b) = 1$ for exactly one $b \in \{0,1\}$ and so $\tilde{\chi}_{A,\tilde{\xi}'}(i_1,\ldots,i_j,b) = 1$ for at least one b as well in which case we would again have $\tilde{\chi}_{A,\tilde{\xi}'}(i_1,\ldots,i_j) = 1$. The case of $\chi_1(i_1,\ldots,i_j) = 0$ is simpler since this arises when $\chi_1(i_1,\ldots,i_j,0) = \chi_1(i_1,\ldots,i_j,1) = 0$ and in this case we have $\tilde{\chi}_{A,\tilde{\xi}'}(i_1,\ldots,i_j,0) = \tilde{\chi}_{A,\tilde{\xi}'}(i_1,\ldots,i_j,1) = 0$ and so $\tilde{\chi}_{A,\tilde{\xi}'}(i_1,\ldots,i_j) = 0$ as well. Finally we note that $\chi_1(i_1,\ldots,i_j) =?$ only if both $\chi_1(i_1,\ldots,i_j,1) =?$ and $\chi_1(i_1,\ldots,i_j,0) =?$ or j is even and $\chi_1(i_1,\ldots,i_j,b) = 0$ for some $b \in \{0,1\}$ and $\chi_1(i_1,\ldots,i_j,1-b) =?$.

We now use these properties of χ_1 to "complete" to a $\{0,1\}$-valued function χ_2 as follows. Set $\chi_2() = 1$ and now for j increasing from 0 to $k-1$ proceed as follows: If $\chi_1(i_1,\ldots,i_j,b) \in \{0,1\}$ then let $\chi_2(i_1,\ldots,i_j,b) = \chi_2(i_1,\ldots,i_j,b)$. For the remaining choices of $(i_1,\ldots,i_j,b)$ assign $\chi_2(i_1,\ldots,i_j,b)$ as follows: If j is odd then set $\chi_2(i_1,\ldots,i_j,b) = \chi_2(i_1,\ldots,i_j)$. If j is even and $\chi_1(i_1,\ldots,i_j,1-b) \in \{0,1\}$ then set $\chi_2(i_1,\ldots,i_j,b) = \chi_2(i_1,\ldots,i_j) - \chi_2(i_1,\ldots,i_j,1-b)$. Else if we have $\chi_1(i_1,\ldots,i_j,b) =?$ for both $b \in \{0.1\}$ then set $\chi_2(i_1,\ldots,i_j,0) = 0$ (arbitrarily) and $\chi_2(i_1,\ldots,i_j,1) = \chi_2(i_1,\ldots,i_j)$.

Now let $\tilde{\xi}_A(i_1,\ldots,i_k) = \chi_2(i_1,\ldots,i_k)$. It can be verified that $\tilde{\chi}_A = \mathrm{PT}_A(\tilde{\xi}_A)$ satisfies $\tilde{\chi}_A(i_1,\ldots,i_j) = \chi_2(i_1,\ldots,i_j)$ for every $i_1,\ldots,i_j$ and thus satisfies the conditions of membership in $\tilde{S}_A$.

Now to derive the strategy f_A, for even j, if $\tilde{\chi}_A(i_1,\ldots,i_j) = 1$ we set $f_A^j(i_1,\ldots,i_j) = b$ such that $\tilde{\chi}_A(i_1,\ldots,i_j,b) = 1$ (such a b must exist); and set $f_A^j(i_1,\ldots,i_j) = 1$ (arbitrarily) otherwise. It can be verified that for this strategy the vector representation leads exactly to vectors $\tilde{\xi}_A \in \tilde{S}_A$

and $\xi_A = \tilde{\xi}_A * v \in S_A$. □

More significantly for us, the above equivalence also holds for probabilistic communication (i.e., with private randomness), as defined in the third paragraph of Section 5.1. Recall then that a (private-coin) probabilistic strategy is defined by two sequences $(f_A^{2i})_{1\le i<k/2}$, $(f_B^{2i+1})_{1\le i<k/2}$ of functions taking values in $[0,1]$, and a verdict function v. Here the fact that the set of strategies forms a convex space is important to us.

Proposition 5.3. *K_A and K_B are closed convex sets that correspond to the set of probabilistic communication (and decision) strategies with k bits. More precisely, for every probabilistic strategy f_A of Alice there exists a vector $\tilde{\xi}_A \in \tilde{K}_A$ and $\xi_A \in K_A$ and for every strategy f_B of Bob there exists a vector $\tilde{\xi}_B \in \tilde{K}_B$ and $\xi_B \in K_B$ such that $\tilde{\xi}_A(\ell) \cdot \tilde{\xi}_B(\ell)$ is the probability that $\ell \in \{0,1\}^k$ is the transcript of the interaction between Alice and Bob under strategies f_A and f_B and $\langle \xi_A, \xi_B \rangle$ is the acceptance probability of the interaction. Conversely every vector $\xi_A \in K_A$ corresponds to a probabilistic strategy f_A for Alice (and similarly for Bob, with $\langle \xi_A, \xi_B \rangle$ being the acceptance probability of the protocol).*

Proof. The fact that K_A and K_B are closed and convex sets is straightforward from their definition.

We first show that strategies f_A and f_B can be converted into vectors in $\tilde{K}_A$ and $\tilde{K}_B$ respectively. We define $\tilde{\chi}_A$ inductively as follows. Let $\tilde{\chi}_A() = 1$. Further let $\tilde{\chi}_A(i_1,\ldots,i_j,1) = \tilde{\chi}_A(i_1,\ldots,i_j,0) = \tilde{\chi}_A(i_1,\ldots,i_j)$ if j is j is odd. Finally let $\tilde{\chi}_A(i_1,\ldots,i_j,1) = \tilde{\chi}_A(i_1,\ldots,i_j) \cdot f_A^j(i_1,\ldots,i_j)$ and $\tilde{\chi}_A(i_1,\ldots,i_j,1) = \tilde{\chi}_A(i_1,\ldots,i_j) \cdot (1 - f_A^j(i_1,\ldots,i_j))$ if j is even. Now let $\tilde{\xi}_A(i_1,\ldots,i_k) = \tilde{\chi}_A(i_1,\ldots,i_k)$. It can be verified that $\tilde{\xi}_A \in \tilde{K}_A$ and indeed the vector $\chi_A = \mathrm{PT}_A(\tilde{\xi}_A)$ (as given by Definition 5.1) is the vector $\tilde{\chi}_A$. Taking $\xi_A = \tilde{\xi}_A * v$ gives us the vector $\xi_A \in K_A$ as needed. (The definition of $\xi_B \in K_B$ is similar.)

Now for the more important direction, we claim that every vector $\xi_A \in K_A$ corresponds to a strategy f_A. Note by definition that since $\xi_A \in K_A$, there exists $\tilde{\xi}_A \in \tilde{K}_A$ such that $\xi_A = \tilde{\xi}_A * v$. Let $\tilde{\chi}_A = \mathrm{PT}_A(\tilde{\xi}_A)$ be the vector obtained from $\tilde{\xi}_A$. For even j and $i_1,\ldots,i_j \in \{0,1\}$. let $f_A^j(i_1,\ldots,i_j) = \tilde{\chi}_A(i_1,\ldots,i_j,1)/\tilde{\chi}_A(i_1,\ldots,i_j)$ for even j. (If $\tilde{\chi}_A(i_1,\ldots,i_j) = 0$, we define $f_A^j(i_1,\ldots,i_j) = 1$ Note that since $\tilde{\chi}_A(i_1,\ldots,i_t) \in [0,1]$ for every t and $\tilde{\chi}_A(i_1,\ldots,i_j) = \tilde{\chi}_A(i_1,\ldots,i_j,0) + \tilde{\xi}_A(i_1,\ldots,i_j,1)$, we have that $f_A^j(i_1,\ldots,i_j) \in [0,1]$ and so f_A represents a strategy for Alice. It can further be verified that if we apply the transformation of the previous paragraph to this strategy f_A, we recover $\tilde{\xi}_A$ and so this correspondence is indeed bidirectional.

Finally we verify that the acceptance probabilities are given by the inner product function. Consider strategies given by vectors $\xi_A \in K_A$ and $\xi_B \in K_B$ with corresponding vectors $\tilde{\xi}_A \in \tilde{K}_A$ and $\tilde{\xi}_B \in \tilde{K}_B$ such that $\xi_A = \tilde{\xi}_A * v$ and $\xi_B = \tilde{\xi}_B * v$. Let $\tilde{\chi}_A = \mathrm{PT}_A(\tilde{\xi}_A)$ and $\tilde{\chi}_B = \mathrm{PT}_B(\tilde{\xi}_B)$; and let f_A and f_B be the probabilistic strategies corresponding to ξ_A and ξ_B respectively as given by the previous paragraph. We claim, by induction on j, that for every $i_1,\ldots,i_j$ the probability that the interaction reaches the partial transcript $i_1,\ldots,i_j$ under strategies f_A, f_B is $\tilde{\chi}_A(i_1,\ldots,i_j) \cdot \tilde{\chi}_B(i_1,\ldots,i_j)$. This is certainly true for $j = 0$ where $\tilde{\chi}_A() = \tilde{\chi}_B() = 1$. Now consider a partial transcript $i_1,\ldots,i_j$. By induction the probability of reaching this transcript is $p = \tilde{\chi}_A(i_1,\ldots,i_j) \cdot \tilde{\chi}_B(i_1,\ldots,i_j)$. Suppose it is Alice's turn to speak. Then $\tilde{\chi}_B(i_1,\ldots,i_j,0) = \tilde{\chi}_B(i_1,\ldots,i_j,1) = \tilde{\chi}_B(i_1,\ldots,i_j)$. And $\tilde{\chi}_A(i_1,\ldots,i_j,0) + \tilde{\chi}_A(i_1,\ldots,i_j,1) = \tilde{\chi}_A(i_1,\ldots,i_j)$. Thus the probability of reaching the partial transcript $i_1,\ldots,i_j,1$ is $p \cdot f_A^j(i_1,\ldots,i_j) = \tilde{\chi}_A(i_1,\ldots,i_j) \cdot \tilde{\chi}_B(i_1,\ldots,i_j) \cdot \tilde{\chi}_A(i_1,\ldots,i_j,1)/\tilde{\chi}_A(i_1,\ldots,i_j) = \tilde{\chi}_A(i_1,\ldots,i_j,1) \cdot \tilde{\chi}_B(i_1,\ldots,i_j) = \tilde{\chi}_A(i_1,\ldots,i_j,1) \cdot \tilde{\chi}_B(i_1,\ldots,i_j,1)$. The calculation for the extension $i_1,\ldots,i_j,0$ is similar and uses the fact that

$\tilde{\chi}_A(i_1,\ldots,i_j) = \tilde{\chi}_A(i_1,\ldots,i_j,0) + \tilde{\chi}_A(i_1,\ldots,i_j,1)$, and the probability that the transcript extends to $i_1,\ldots,i_j,0$ conditioned on being at $i_1,\ldots,i_j$ is $1 - f_A^j(i_1,\ldots,i_j)$.

From the above, we conclude that the probability of reaching a final transcript ℓ is $\tilde{\chi}_A(\ell)\cdot\tilde{\chi}_B(\ell) = \tilde{\xi}_A(\ell)\cdot\tilde{\xi}_B(\ell)$. Thus the acceptance probability equals $\sum_{\ell:v(\ell)=1}\tilde{\xi}_A(\ell)\cdot\tilde{\xi}_B(\ell) = \sum_{\ell\in\{0,1\}^k}\tilde{\xi}_A(\ell)\cdot\tilde{\xi}_B(\ell)\cdot v(\ell)^2 = \sum_{\ell\in\{0,1\}^k}\xi_A(\ell)\cdot\xi_B(\ell) = \langle\xi_A,\xi_B\rangle$. □

5.2 Upper bound on ISR in terms of PSR

In this section we prove Theorem 2.8. Our first step is to prove that the GapInnerProduct problem (with the right parameters) is hard for all problems with communication complexity k. But first we define what it means for a promise problem to be hard for some class of communication problems.

Recall that a promise problem $P = (P_n)_n$ is given by a collection of yes-instances $P_n^{\mathsf{yes}} \subseteq \{0,1\}^n \times \{0,1\}^n$ and no-instances $P_n^{\mathsf{no}} \subseteq \{0,1\}^n \times \{0,1\}^n$ with $P_n^{\mathsf{yes}} \cap P_n^{\mathsf{no}} = \emptyset$. We define below what it means for a promise problem P to reduce to a promise problem Q.

Definition 5.4. For promise problems $P = (P_n)_n$ and $Q = (Q_n)_n$ we say that P reduces to Q if there exist functions $\ell\colon \mathbb{Z}^+ \to \mathbb{Z}^+$ and $f_n, g_n\colon \{0,1\}^n \to \{0,1\}^{\ell(n)}$ such that if $(x,y) \in P_n^{\mathsf{yes}}$ then $(f_n(x), g_n(y)) \in Q_{\ell(n)}^{\mathsf{yes}}$ and if $(x,y) \in P_n^{\mathsf{no}}$ then $(f_n(x), g_n(y)) \in Q_{\ell(n)}^{\mathsf{no}}$. We say Q is hard for a class $\mathcal{C}$ if for every $P \in \mathcal{C}$ we have that P reduces to Q.

In other words Alice can apply f_n to her input, and Bob can apply g_n to his input and get a new pair that is an instance of the Q-problem. In particular if Q has communication complexity k, then so does P. This can be extended to functions $k(n)$ also: if Q has communication complexity $k(n)$, then P has complexity $k(\ell(n))$.

Since we are mostly interested in k being an absolute constant, we do not strictly care about the length stretching function ℓ. However, we note that in the following proposition we only need a polynomial blowup (so ℓ is a polynomial).

Proposition 5.5. *For every positive integer k,* $\textsc{GapInnerProduct}_{(2/3)2^{-k},(1/3)2^{-k}}$ *is hard for* $\mathsf{PSR\text{-}CC}(k)$.

Proof. Specifically we show that for any problem P with inputs of length n and psr-cc$(P) \le k$, there exist $N = \mathrm{poly}(n)$ and transformations f_n and g_n such that (x,y) is a yes-instance of P if and only if $(f_n(x), g_n(y))$ is a yes-instance of $\textsc{GapInnerProduct}^N_{(2/3)2^{-k},(1/3)2^{-k}}$.

Given $x \in \{0,1\}^n$ and random string R, let $X_R \in S_A^{(k)}$ describe the communication strategy of Alice with input x and randomness R. Similarly let Y_R denote the strategy of Bob. Recall that $\langle X_R, Y_R\rangle = 1$ if the interaction accepts on randomness R and $\langle X_R, Y_R\rangle = 0$ otherwise. Let $f_n(x) = X$ be the concatenation of the strings $\{X_R\}_R$ and let $g_n(y) = Y$ be the concatenation of $\{Y_R\}_R$. By Newman's Theorem we have that the number of random strings R that we need to consider is some polynomial $N' = \mathrm{poly}(n)$. Letting $N = 2^k \cdot N'$, we get that $X, Y \in \{0,1\}^N$ and $\langle X, Y\rangle \ge (2/3)N' = (2/3)\cdot 2^{-k}\cdot N$ if (x,y) is a yes-instance of P and $\langle X, Y\rangle \le (1/3)N' = (1/3)\cdot 2^{-k}\cdot N$ if (x,y) is a no-instance of P. This gives the desired reduction. □

Next we give an upper bound on isr-cc(GapInnerProduct). In fact we give an upper bound on isr-cc$^{\mathrm{ow}}$(GapInnerProduct).

Lemma 5.6. *For all $0 \le s < c \le 1$ and $\rho > 0$,* isr-cc$^{\mathrm{ow}}_\rho(\textsc{GapInnerProduct}^n_{c,s}) = O(1/\rho^2(c-s)^2)$.

Proof. Let $X \in \{0,1\}^n$ and $Y \in \{0,1\}^n$ be the inputs to Alice and Bob. Recall that Alice and Bob want to distinguish the case $\langle X, Y\rangle \geq c \cdot n$ from the case $\langle X, Y\rangle \leq s \cdot n$.

We shall suppose without loss of generality that Alice and Bob have access to a source of ρ-correlated random spherical Gaussian vectors $g, g' \in \mathbb{R}^n$. We can enforce this in the limit by sampling several ρ-correlated random bit vectors $r_i, r'_i \in \{0,1\}^n$ for $i \in [N]$ and setting $g = \sum_{i=1}^N r_i/\sqrt{N}$ and $g' = \sum_{i=1}^N r'_i/\sqrt{N}$. We leave out the details for this technical calculation (involving an appropriate use of the central limit theorem) here.

Let t be a parameter to be chosen later and let $(g_1, g'_1), (g_2, g'_2), \ldots, (g_t, g'_t)$ be t independent ρ-correlated spherical Gaussian vectors chosen from the source as above. By the rotational invariance of the Gaussian distribution, we can assume without loss of generality that $g'_i = \rho g_i + \sqrt{1-\rho^2} g''_i$, where the g''_i's are independent spherical Gaussian vectors.

As $g_1, \ldots, g_t$ are independent spherical Gaussians, by standard tail bounds (e.g., see Ledoux and Talagrand [LT91]), with probability at least $1 - 1/6$,

$$\max_{i \in [t]} \langle X, g_i\rangle = (\alpha\sqrt{\log t} \pm O(1)) \cdot \sqrt{\langle X, X\rangle}$$

for some universal constant α.
The protocol then proceeds as follows:

- Alice computes $\ell = \arg\max_{i\in[t]} \langle X, g_i\rangle$ and m such that $\langle X, X\rangle \in ((m-1)\cdot \frac{(c-s)}{100}n, m \cdot \frac{(c-s)}{100}n]$ and sends (ℓ, m) to Bob (note that this implies $m = O(1/(c-s))$).
- Bob accepts if $m \geq \frac{100c}{c-s}$ and $\langle Y, g'_\ell\rangle \geq \alpha\rho\sqrt{\log t} \cdot \frac{(c+s)n}{2\sqrt{m(c-s)(n/100)}}$ and rejects otherwise.

Now, write $Y = aX + bX^\perp$ for some vector $X^\perp$ with $a\langle X, X\rangle = \langle X, Y\rangle$ and $\langle X, X^\perp\rangle = 0$. Then,

$$\langle Y, g'_\ell\rangle = a\rho \langle X, g_\ell\rangle + b\rho \left\langle X^\perp, g_\ell\right\rangle + \sqrt{1-\rho^2}\langle Y, g''_\ell\rangle \ .$$

As $\langle X, g_\ell\rangle$ is independent of $\langle X^\perp, g_\ell\rangle$ and $\langle Y, g''_\ell\rangle$ (the former from the fact that if G is a spherical Gaussian and $u, v \in \mathbb{R}^n$ are orthogonal vectors then $\langle u, G\rangle$ and $\langle v, G\rangle$ are independent one-dimensional Gaussians), it follows from a simple tail bound for univariate Gaussians that with probability at least $1 - 1/6$, $\left|\langle X^\perp, g_\ell\rangle\right|$, $|\langle Y, g''_\ell\rangle| = O(\sqrt{n})$. By combining the above inequalities, we get that with probability at least $2/3$,

$$\langle Y, g'_\ell\rangle = \alpha\rho\sqrt{\log t}\,\langle X, Y\rangle / \sqrt{\langle X, X\rangle} \pm O(\sqrt{n}).$$

To finish the proof observe that for yes-instances, $\langle X, Y\rangle \geq cn$ (so that $m \geq \frac{100c}{c-s}$) and $\langle X, Y\rangle / \sqrt{\langle X, X\rangle} \geq \beta_1 \triangleq c\cdot n/\sqrt{m(c-s)(n/100)}$; while for no-instances, $\langle X, Y\rangle / \sqrt{\langle X, X\rangle} \leq \beta_2 \triangleq s \cdot n/\sqrt{(m-1)(c-s)(n/100)}$. Hence, the protocol works correctly if $\alpha\rho\sqrt{\log t}(\beta_1 - \beta_2) \gg O(\sqrt{n})$.

It follows from the settings of parameters that this indeed happens for some $\log t = \Theta(1/(\rho^2(c-s)^2))$. In particular, we have

$$\beta_1 - \beta_2 = \frac{cn - sn}{\sqrt{m(c-s)(n/100)}} - \frac{sn}{\sqrt{(c-s)(n/100)}}\left(\frac{1}{\sqrt{m-1}} - \frac{1}{\sqrt{m}}\right).$$

By the condition $m \geq \frac{100c}{c-s}$ we have $\frac{1}{\sqrt{m-1}} - \frac{1}{\sqrt{m}} \leq \frac{c-s}{2s}$ and thus

$$\beta_1 - \beta_2 \geq \frac{1}{2}\frac{cn - sn}{\sqrt{m(c-s)(n/100)}} = \frac{\sqrt{25(c-s)n}}{\sqrt{m}}.$$

And so when $\log t \gg \Omega(1/(\alpha^2\rho^2(c-s)^2))$ we find $\alpha\rho\sqrt{\log t}(\beta_1 - \beta_2) \gg O(\sqrt{n})$ as required. □

The above lemma along with the hardness of GapInnerProduct gives us Theorem 2.8.

Proof of Theorem 2.8. By Proposition 5.5, for every promise problem P such that psr-cc$(P) \le k$, P reduces to GapInnerProduct$_{c,s}$ with $c = (2/3)2^{-k}$ and $s = (1/3)2^{-k}$. By Lemma 5.6 we get that the reduced instance of GapInnerProduct$_{c,s}$ has a one-way isr communication protocol of with $O_\rho(1/(c-s)^2) = O_\rho(2^{2k})$ bits of communication. The theorem follows. □

5.3 ISR lower bound for SparseGapInnerProduct

In this section, we consider the promise problem SparseGapInnerProduct$^n_{.99q,.9q^{-1},.6q^{-1}}$ and show that it has a one-way psr protocol with $O(\log q)$ bits of communication, and then give a two-way isr lower bound of $q^{\Omega(1)}$ for this problem. Together this proves Theorem 2.9.

Proposition 5.7. *$\forall c > s$ and $\forall q, n$, we have*

$$\text{psr-cc}^{\text{ow}}(\text{SparseGapInnerProduct}^n_{q,c,s}) \le O\left(\frac{1}{q^2(c-s)^2}\left(\log\frac{1}{c} + \log\frac{1}{q(c-s)} + \log\log\frac{c}{c-s}\right)\right).$$

Proof (Sketch). We first show that there exists an atomic one-way communication protocol for the problem SparseGapInnerProduct$^n_{q,c,s}$ with the following features (where $\gamma = \Theta((c-s)/c)$):

1. the length of communication is $O(\log 1/c + \log 1/(q(c-s)) + \log\log 1/\gamma)$.
2. yes-instances are accepted with probability at least $(1-\gamma)\cdot\frac{c}{c-s}\cdot\frac{100}{m}$ and no-instances with probability at most $\frac{s}{c-s}\cdot\frac{100}{m-1}$ for some $m = \Omega(c/(c-s))$ known by both parties. In particular, the difference between completeness and soundness is $\Omega(1/m)$.

The atomic protocol lets the shared randomness determine a sequence of $t \stackrel{\text{def}}{=} -\log(1/\gamma)/\log(1-c)$ indices $i_1, i_2, \ldots, i_t$ in $[n]$. Alice first computes $m = O(1/(c-s))$ such that $\|x\|_2^2 \in ((m-1)\cdot\frac{(c-s)}{100}n, m\cdot\frac{(c-s)}{100}n]$, and picks the smallest index ℓ such that $x_{i_\ell} \neq 0$. Then she sends (ℓ, m) to Bob, or $(0,0)$ if no such index was found. (Note that by sparsity of x, we have $m = O(1/(q(c-s)))$). Bob outputs 0 if he received $(0,0)$ or if $m < \frac{100c}{c-s}$, and the value of y_{i_ℓ} otherwise.

The completeness follows from the fact that, for yes-instances, $\|x\|_2^2 \ge cn$ (implying $m \ge \frac{100c}{c-s}$) and one expects an index ℓ such that $x_{i_\ell} \neq 0$ among the first roughly $1/c$ choices of ℓ, so that this will be the case with high probability among the first t; conditioned on this (which happens with probability at least $1-\gamma$ by our choice of t, as the probability no ℓ is found is upper bounded by $(1-\|x\|_2/n)^t \le (1-c)^t$), y_{i_ℓ} is 1 with probability at least $\frac{cn}{\|x\|_2^2} \ge \frac{c}{c-s}\frac{100}{m}$. As for the soundness, observe that a no-instance for which Alice does not send 0 to Bob will have $y_{i_\ell} = 1$ with probability at most $\frac{sn}{\|x\|_2^2} < \frac{s}{c-s}\cdot\frac{100}{m-1}$. Now, since $m \ge \frac{100c}{c-s}$, $\frac{100s}{c-s}\left(\frac{1}{m-1} - \frac{1}{m}\right) \le \frac{100}{3m}$; and by the choice of $\gamma \le \frac{c-s}{3c}$ we also have $\gamma\frac{c}{c-s}\frac{100}{m} \le \frac{100}{3m}$. This implies the difference in acceptance probability between completeness and soundness is at least $\frac{100}{3m}$.
Repeating this protocol $O(m^2) = O(1/(q^2(c-s)^2))$ times and thresholding yields the final result. □

We now state our main lower bound theorem.

Theorem 5.8. *There exists $\epsilon > 0$ such that $\forall\rho \in [0,1)$ and $\forall q$, there exists N for which the following holds. For every $n \ge N$, we have*

$$\text{isr-cc}_\rho(\text{SparseGapInnerProduct}^n_{.99q,.9q^{-1},.6q^{-1}}) \ge \epsilon\cdot\sqrt{q}.$$

We prove Theorem 5.8 in Section 6, but we now note that Theorem 2.9 follows immediately from Proposition 5.7 and Theorem 5.8.

Proof of Theorem 2.9. The promise problem is $P = \textsc{SparseGapInnerProduct}_{.99\cdot 2^k,.9\cdot 2^{-k},.6\cdot 2^{-k}}$. By Proposition 5.7 we have $\text{psr-cc}^{\text{ow}}(P) \leq O(k)$ and by Theorem 5.8 we have $\text{isr-cc}(P) \geq 2^{\Omega(k)}$. □

6 Proof of Theorem 5.8

Our goal for this section is to prove, modulo some technical theorems, that SparseGapInnerProduct has high communication complexity in the imperfectly shared randomness setting. Before jumping into the proof we give some overview first.

6.1 Proof setup

To prove Theorem 5.8, we will show that for every "strategy" of Alice and Bob, there is a pair of distributions $\mathcal{Y}$ and $\mathcal{N}$ supported (mostly) on yes and no instances, respectively, such that the strategies do not have much "success" in distinguishing them. We note that in contrast to typical lower bounds for perfectly shared randomness, we cannot hope to fix a distribution that works against every strategy. Indeed for every pair of distributions, by virtue of the protocol given in Proposition 5.7 and the Yao min-max principle we have even a deterministic strategy (let alone randomized strategy with imperfect sharing) that succeeds in distinguishing them with high probability. So instead we have to fix the strategies first and then give a pair of distributions that does not work for that strategy. We define the notion of strategy and success more formally below, and then work towards the proof of Theorem 5.8.

Strategy: We now use Section 5.1 to formalize what it would mean to have a k-bit communication protocol for any communication problem. For aesthetic reasons we view Alice and Bob's strategies as probabilistic ones. Recall, by Proposition 5.3, that k-bit probabilistic communication strategies for Alice can be described by elements of $K_A^{(k)} \subseteq [0,1]^{2^k}$ and similarly by elements of $K_B^{(k)} \subseteq [0,1]^{2^k}$ for Bob. So, on randomness r we have that Alice's communication strategy can be described by a function $f^{(r)} \colon \{0,1\}^n \to K_A^{(k)}$. Similarly for randomness s, Bob's communication strategy can be described by a function $g^{(s)} \colon \{0,1\}^n \to K_B^{(k)}$.

Thus, a *strategy* for a game is a pair of sets of functions $\mathcal{F} = (f^{(r)})_r, \mathcal{G} = (g^{(s)})_s$, where

$$f^{(r)} \colon \{0,1\}^n \to K_A^{(k)}$$
$$\text{and } g^{(s)} \colon \{0,1\}^n \to K_B^{(k)}.$$

We consider a pair of distributions $D = (\mathcal{Y}, \mathcal{N})$ to be *valid* if $\mathcal{Y}$ is mostly (say with probability .9) supported on yes-instances and $\mathcal{N}$ mostly on no-instances. For valid D, we define the *success* as

$$\begin{aligned}
\text{succ}_D(f,g) &\stackrel{\text{def}}{=} \mathbb{E}_{(x,y)\sim\mathcal{Y}}\left[\langle f(x), g(y)\rangle\right] - \mathbb{E}_{(x,y)\sim\mathcal{N}}\left[\langle f(x), g(y)\rangle\right] \\
\text{succ}_{D,\rho}(\mathcal{F},\mathcal{G}) &\stackrel{\text{def}}{=} \left|\mathbb{E}_{r\sim_\rho s}\left[\text{succ}_D(f^{(r)}, g^{(s)})\right]\right| \\
\text{succ}_\rho(\mathcal{F},\mathcal{G}) &\stackrel{\text{def}}{=} \min_{\text{valid } D} \text{succ}_{D,\rho}(\mathcal{F},\mathcal{G})
\end{aligned}$$

We note that any strategy that distinguishes yes-instances of SparseGapInnerProduct from no-instances with probability ϵ must have success $\epsilon - .1$ on every valid distribution as well (with the difference of .1 coming up due to the fact that valid distributions are not entirely supported on the right instances). In what follows we will explain why strategies (with small k) do not have sufficiently positive success.

6.2 Overview of proof of Theorem 5.8.

To prove Theorem 5.8 we need to show that if a pair of strategies $(\mathcal{F}, \mathcal{G})$ achieves $\mathrm{succ}_\rho(\mathcal{F}, \mathcal{G}) > .01$ then k must be large. Roughly our strategy for showing this is as follows: We first define two simple distributions $\mathcal{Y}$ and $\mathcal{N}$ (independent of the strategy $(\mathcal{F}, \mathcal{G})$) and show that any fixed pair of functions (f, g) that are successful in distinguishing $\mathcal{Y}$ from $\mathcal{N}$ must have a few influential variables and furthermore at least one of these variables must be common to both f and g (see Theorem 6.8). Our proof of this theorem, is based on the "invariance principle" [Mos10] and Theorem 6.8 is a variant of it which is particularly suited for use in communication complexity. The proof of this theorem is deferred to Section 7.

We use this theorem to design agreement distillation strategies for two new players Charlie and Dana as follows: Given shared random pair (r, s), Charlie picks a random influential variable x_i of the function $f^{(r)}$ used by Alice on random string r and outputs the index $i \in [n]$. Dana similarly picks a random influential variable y_j of the function $g^{(s)}$ used by Bob and outputs j. Theorem 6.8 assures us that with non-trivial probability $i = j$ and this gives an agreement protocol.

If we could argue that $i = j$ has high min-entropy, then we would be done (using Lemma 4.1 which asserts that it is not possible to distill agreement with high-entropy and high probability). But this step is not immediate (and should not be since we have not crafted a distribution specific to $(\mathcal{F}, \mathcal{G})$). To show that this strategy produces indices of high min-entropy, we consider the distribution of indices that is produced by Charlie as we vary r and let BAD_C denote the indices that are produced with too high a probability. Similarly we let BAD_D denote the indices that are produced with too high a probability by Dana. We now consider a new distribution $\mathcal{Y}'$ supported on yes-instances of the SparseGapInnerProduct problem. In $\mathcal{Y}'$ the (x, y) pairs are chosen so that when restricted to coordinates in $\mathrm{BAD}_C \cup \mathrm{BAD}_D$ they look like they come from $\mathcal{N}$ while when restricted to coordinates outside $\mathrm{BAD}_C \cup \mathrm{BAD}_D$ they look like they come from $\mathcal{Y}$ (see Definition 6.13 below for a precise description). Since $\mathrm{BAD}_C \cup \mathrm{BAD}_D$ is small, the distribution $\mathcal{Y}'$ remains supported mostly on yes-instances, but strategies that depend mainly on coordinates from $\mathrm{BAD}_C \cup \mathrm{BAD}_D$ would not have much success in distinguishing $\mathcal{Y}'$ from $\mathcal{N}'$ (which remains the original $\mathcal{N}$).

We use this intuition to argue formally in Lemma 6.14 that a slightly modified sampling protocol of Charlie and Dana, where they discard i, j from $\mathrm{BAD}_C \cup \mathrm{BAD}_D$, leads to agreement with noticeably high probability on a high-entropy random variable, yielding the desired contradiction.

In the rest of this section we first present the main definitions needed to state Theorem 6.8. We then prove Theorem 5.8 assuming Theorem 6.8. We prove the latter in Section 7, along with the main technical ingredient it relies on, the invariance principle of Theorem 7.1.

6.3 Background on influence of variables

We now turn to defining the notion of influential variables for functions and related background material for functions defined on product probability spaces.

Recall that a finite probability space is given by a pair (Ω, μ) where Ω is a finite set and μ is a probability measure on Ω. We will begin with the natural probabilistic definition of influence of a variable on functions defined on product spaces, and then relate it to a more algebraic definition which is needed for the notion of low-degree influence.

Definition 6.1 (Influence and variance)**.** Let (Ω, μ) be a finite probability space, and let $h\colon \Omega^n \to \mathbb{R}$ be a function on product probability space. The *variance* of h, denoted $\mathsf{Var}(h)$, is defined as the variance of the random variable $h(x)$ for $x \in \Omega^n \sim \mu^{\otimes n}$, i.e., $\mathsf{Var}(h) = \mathbb{E}_x[h(x)^2] - (\mathbb{E}_x[h(x)])^2$. For $i \in [n]$, the *i-th influence of h* is defined as

$$\mathsf{Inf}_i(h) = \mathbb{E}_{x^{(-i)} \sim \mu^{\otimes(n-1)}}\left[\mathsf{Var}_{x_i \sim \mu}[h(x)]\right]$$

where $x^{(-i)}$ denotes all coordinates of x except the i'th coordinate.

To define the notion of *low-degree* influence, we need to work with a multilinear representation of functions $h\colon \Omega^n \to \mathbb{R}$. Let $b = |\Omega|$ and $\mathcal{B} = \{\chi_0, \chi_1, \ldots, \chi_{b-1}\}$ be a basis of real-valued functions over Ω. Then, every function $h\colon \Omega^n \to \mathbb{R}$ has a unique multilinear expansion of the form

$$h(x) = \sum_{\boldsymbol{\sigma}=(\sigma_1,\ldots,\sigma_n)\in\{0,1,\ldots,b-1\}^n} \hat{h}_{\boldsymbol{\sigma}} \chi_{\boldsymbol{\sigma}}(x) \tag{2}$$

for some real coefficients $\hat{h}_{\boldsymbol{\sigma}}$, where $\chi_{\boldsymbol{\sigma}}$ is given by $\chi_{\boldsymbol{\sigma}}(x) \stackrel{\text{def}}{=} \prod_{i\in[n]} \chi_{\sigma_i}(x_i)$.

When the ensemble $\mathcal{B}$ is a collection of *orthonormal* random variables, namely $\chi_0 = 1$ and $\mathbb{E}_{a\sim\mu}[\chi_{j_1}(a)\chi_{j_2}(a)] = \delta_{j_1 j_2}$, it is easy to check that $\mathsf{Var}(h) = \sum_{\boldsymbol{\sigma}\neq\mathbf{0}} \hat{h}^2_{\boldsymbol{\sigma}}$ and also that

$$\mathsf{Inf}_i(h) = \sum_{\boldsymbol{\sigma}\colon \sigma_i \neq 0} \hat{h}^2_{\boldsymbol{\sigma}} \ .$$

One can also take the above as the algebraic definition of influence, noting that it is independent of the choice of the orthonormal basis $\mathcal{B}$ and thus well-defined. The degree of a multi-index $\boldsymbol{\sigma}$ is defined as $|\boldsymbol{\sigma}| = |\{\, i : \sigma_i \neq 0 \,\}|$, and this leads to the definition of low-degree influence.

Definition 6.2 (Low-degree influence)**.** For a function $h\colon \Omega^n \to \mathbb{R}$ with multilinear expansion as in (2) with respect to any orthonormal basis, the *i-th degree d influence of h* is the influence of the truncated multilinear expansion of h at degree d, that is

$$\mathsf{Inf}^d_i(h) \stackrel{\text{def}}{=} \sum_{\substack{\boldsymbol{\sigma}\colon \sigma_i \neq 0 \\ |\boldsymbol{\sigma}|\leq d}} \hat{h}^2_{\boldsymbol{\sigma}} \ .$$

Remark 6.3 (Functions over size 2 domain). When $|\Omega| = 2$, and $\{1, \chi\}$ is an orthonormal basis of real-valued functions over Ω, the expansion (2) becomes the familiar Fourier expansion $h(x) = \sum_{S\subseteq[n]} \widehat{h}(S) \prod_{i\in S} \chi(x_i)$, and we have $\mathsf{Inf}_i(h) \stackrel{\text{def}}{=} \sum_{S\ni i} \widehat{h}(S)^2$ and $\mathsf{Inf}^d_i(h) \stackrel{\text{def}}{=} \sum_{\substack{S\ni i \\ |S|\leq d}} \widehat{h}(S)^2$.

We will make use of the following simple upper bound on the number of low-degree influential coordinates (which follows immediately, for instance, from [Mos10, Proposition 3.8])

Proposition 6.4. *For every $\tau > 0$ and $d \in \mathbb{Z}^+$ there exists $t = t(\tau, d)$ such that for all n and all functions $h\colon \Omega^n \to [-1, 1]$, we have $\left|\left\{\, i \in [n] : \mathsf{Inf}^d_i(h) > \tau \,\right\}\right| \leq t$. (Furthermore, one can take $t = d/\tau$).*

For the invariance principle, we will understand the behavior of a function when its domain is replaced by a different probability space with matching second moments. For this purpose, we will view functions as multilinear polynomials as follows.

Definition 6.5 (Functions on product spaces as multilinear polynomials)**.** The multilinear polynomial associated with a function $h\colon \Omega^n \to \mathbb{R}$ with respect to a basis $\mathcal{B} = \{\chi_0, \chi_1, \ldots, \chi_{b-1}\}$ of real-valued functions over Ω is a polynomial in indeterminates $\mathbf{z} = \{ z_{i,j} : i \in [n], j \in \{0, 1, \ldots, b-1\} \}$ given by

$$H(\mathbf{z}) = \sum_{\boldsymbol{\sigma} \in \{0,1,\ldots,b-1\}^n} \hat{h}_{\boldsymbol{\sigma}} \mathbf{z}_{\boldsymbol{\sigma}},$$

$\mathbf{z}_{\boldsymbol{\sigma}}$ stands for the monomial $\prod_{i=1}^n z_{i,\sigma_i}$ and the coefficients $\hat{h}_{\boldsymbol{\sigma}}$ are given by the multilinear expansion (2) of f w.r.t. $\mathcal{B}$.

Above, we saw how a function can be viewed as a multilinear polynomial w.r.t. a basis of random variables. Conversely, one can view multilinear polynomials as functions by substituting random variables for its indeterminates.

Definition 6.6 (Multilinear polynomials as random variables on product spaces)**.** Given a collection of random variables $\mathcal{X} = \{\chi_0, \ldots, \chi_{b-1}\}$ over a probability space (Ω, μ), one can view a multilinear polynomial P in indeterminates $\mathbf{z} = \{ z_{i,j} : i \in [n], j \in \{0, 1, \ldots, b-1\} \}$ given by

$$P(\mathbf{z}) = \sum_{\boldsymbol{\sigma} \in \{0,1,\ldots,b-1\}^n} \hat{P}_{\boldsymbol{\sigma}} \mathbf{z}_{\boldsymbol{\sigma}},$$

where $\mathbf{z}_{\boldsymbol{\sigma}}$ stands for the monomial $\prod_{i=1}^n z_{i,\sigma_i}$, as a random variable $P(\mathcal{X}^n)$ over the probability space $(\Omega^n, \mu^{\otimes n})$ mapping $x = (x_1, \ldots, x_n)$ to

$$\sum_{\boldsymbol{\sigma} \in \{0,1,\ldots,b-1\}^n} \hat{P}_{\boldsymbol{\sigma}} \prod_{i=1}^n \chi_{\sigma_i}(x_i) \ . \tag{3}$$

6.4 Proof of Theorem 5.8

We start by introducing a few definitions, in particular of the central distributions and the extraction strategy. We begin with the description of the basic distributions $\mathcal{Y}$ and $\mathcal{N}$.

Definition 6.7 ($\mathcal{Y}$, $\mathcal{N}$)**.** We define two distributions B_N and B_Y on $\{0,1\} \times \{0,1\}$ below. The distributions $\mathcal{Y}$ and $\mathcal{N}$ will be product distributions on $(\{0,1\} \times \{0,1\})^n$, given by $\mathcal{Y} = B_Y^{\otimes n}$ and $\mathcal{N} = B_N^{\otimes n}$.

- A pair (x, y) is drawn from B_N by setting $x \sim \mathrm{Bern}(1/q)$ and $y \in \{0,1\}$ uniformly at random. Note that x, y are independent, and $\mathbb{E}[xy] = \frac{1}{2q}$.
- A pair (x, y) is drawn from B_Y by setting

$$(x,y) = \begin{cases} (0,1) & \text{w.p. } \frac{1}{2}\left(1 - \frac{1.95}{q}\right) \\ (0,0) & \text{w.p. } \frac{1}{2}\left(1 - \frac{0.05}{q}\right) \\ (1,1) & \text{w.p. } \frac{1.95}{2q} \\ (1,0) & \text{w.p. } \frac{.05}{2q} \end{cases}$$

  so that the marginals of x, y in B_Y match those of B_N, and $\mathbb{E}[xy] = \frac{1.95}{2q}$.

A straightforward application of tail inequalities for independent, identically distributed (i.i.d.) random variables tells us that $\mathcal{Y}$ is mostly supported on yes-instances of $\textsc{SparseGapInnerProduct}^n_{.99q,0.9q,0.6q}$ with high probability for sufficiently large n. Similarly $\mathcal{N}$ is mostly supported on no-instances.

Our main technical result is the following theorem showing any fixed pair of vector-valued functions (f, g) (corresponding to strategies for Alice and Bob) that succeed in distinguishing $\mathcal{Y}$ from $\mathcal{N}$ must share an influential variable (with non-trivially high influence of non-trivially low-degree).

Theorem 6.8. *There exist functions $k_0 \geq \Omega_\epsilon(\sqrt{q})$, $d(q,\epsilon) < \infty$, and $\tau(q,\epsilon) > 0$, defined for $q \in \mathbb{Z}^+$,and $\epsilon > 0$, such that the following holds: For every $\epsilon > 0$ and $k, q \in \mathbb{Z}^+$ and every sufficiently large n, if $k < k_0(q,\epsilon)$ and $f\colon \{0,1\}^n \to K_A^{(k)}$ and $g\colon \{0,1\}^n \to K_B^{(k)}$ are functions such that $\mathrm{succ}_{(\mathcal{Y},\mathcal{N})}(f,g) \geq \epsilon$, then there exists $i \in [n]$ such that*

$$\min\left\{\max_{j\in[2^k]} \mathsf{Inf}_i^{d(q,\epsilon)}(f_j), \max_{j\in[2^k]} \mathsf{Inf}_i^{d(q,\epsilon)}(g_j)\right\} \geq \tau(q,\epsilon)$$

where f_j and g_j denote the j'th component function of f and g, respectively. (Here, the influence of f_j is w.r.t. to the $\mathrm{Bern}(1/q)$ distribution on $\{0,1\}$, and that of g_j is w.r.t. the uniform distribution on $\{0,1\}$.)

This theorem is proved in Section 7. Building on this theorem, we can try to build agreement distillation protocols $(\mathsf{Ext}_C, \mathsf{Ext}_D)$ that exploit the success of the strategies $(\mathcal{F}, \mathcal{G})$ to distill common randomness. We start by first identifying coordinates that may be influential for too many pairs (r, s) (and thus may be produced with too high a probability by a naive distillation protocol).

For the rest of the section we fix $q \in \mathbb{Z}^+$ and $\epsilon > 0$ and let $d = d(q,\epsilon)$ and $\tau = \tau(q,\epsilon)$ where $d(\cdot,\cdot)$ and $\tau(\cdot,\cdot)$ are the functions from Theorem 6.8.

Definition 6.9 ($\textsc{Bad}_C$, $\textsc{Bad}_D$)**.** Let $\delta = 1/(100\cdot 2^{k_0}t)$ where $t = t(\tau, d)$ as given by Proposition 6.4, and $k_0 = k_0(q,\epsilon)$ is given by Theorem 6.8. Define

$$\textsc{Bad}_C \stackrel{\text{def}}{=} \left\{ i \in [n] : \Pr_r\left[\max_{j\in[2^k]} \mathsf{Inf}_i^d(f_j^{(r)}) > \tau\right] > \frac{1}{\delta n}\right\} \qquad \text{and}$$

$$\textsc{Bad}_D \stackrel{\text{def}}{=} \left\{ i \in [n] : \Pr_s\left[\max_{j\in[2^k]} \mathsf{Inf}_i^d(g_j^{(s)}) > \tau\right] > \frac{1}{\delta n}\right\},$$

where r, s denote the randomness available to Alice and Bob, $f_j^{(r)}$ denotes the j'th component function for Alice's strategy on randomness r, and similarly for $g_j^{(s)}$.

Directly from this definition and Proposition 6.4, we get

Proposition 6.10. $|\textsc{Bad}_C|, |\textsc{Bad}_D| \leq 2^k \cdot t \cdot \delta \cdot n \leq n/100$.

Next, we define the extraction distillation protocols for Charlie and Dana:

Definition 6.11 ($(\mathsf{Ext}_C, \mathsf{Ext}_D)$)**.** For $r \in \{0,1\}^*$, let

$$S_r \stackrel{\text{def}}{=} \left\{ i \in [n] \setminus \textsc{Bad}_C : \max_{j\in[2^k]} \mathsf{Inf}_i^d(f_j^{(r)}) > \tau \right\} \text{ and } T_s \stackrel{\text{def}}{=} \left\{ i \in [n] \setminus \textsc{Bad}_D : \max_{j\in[2^k]} \mathsf{Inf}_i^d(g_j^{(s)}) > \tau \right\}.$$

Then, $\mathsf{Ext}_C(r)$ is defined as follows:

if $S_r = \emptyset$ output $i \sim \mathcal{U}_{[n]}$; otherwise output $i \sim \mathcal{U}_{S_r}$.

$\mathsf{Ext}_D(s)$ is defined similarly:

if $T_s = \emptyset$ output $j \sim \mathcal{U}_{[n]}$; otherwise output $j \sim \mathcal{U}_{T_s}$.

Proposition 6.12. $H_\infty(\mathsf{Ext}_C(r)) \geq \log n - \log(1 + 1/\delta)$.

Proof. Fix $i \in [n] \setminus (\textsc{Bad}_C \cup \textsc{Bad}_D)$. We have

$$\Pr[\, i \text{ is output} \,] \leq \Pr[\, i \in S_r \text{ and } i \text{ is output} \,] + \Pr[\, i \text{ is output} \mid S_r = \emptyset \,] \leq 1/(\delta n) + 1/n$$

where the upper bound on the first term comes from observing that as $i \notin \textsc{Bad}_C$, $\Pr[\, i \in S_r \,] \leq 1/(\delta n)$. The proposition follows. □

Finally we turn to proving that Ext_C and Ext_D do agree with non-trivial probability. To do so we need to consider a new distribution on yes-instances, defined next:

Definition 6.13 ($\mathcal{Y}'$)**.** The distribution $\mathcal{Y}'$ is a product distribution on $(\{0,1\} \times \{0,1\})^n$, where $(x_i, y_i) \sim B_N$ if $i \in \textsc{Bad}_C \cup \textsc{Bad}_D$ and $(x_i, y_i) \sim B_Y$ otherwise.

Using Proposition 6.10 above we have that $\mathbb{E}_{i,x,y}[x_i y_i] \geq .93/q$ (where the expectation is over $(x, y) \sim \mathcal{Y}'$ and i drawn uniformly at random from $[n]$) and so by standard tail inequalities we still have that $\mathcal{Y}'$ is mostly supported on yes-instances. Our main lemma for this section is that if $(\mathcal{F}, \mathcal{G})$ are successful in distinguishing $\mathcal{Y}'$ and $\mathcal{N}$ and k is small, then Ext_C and Ext_D are likely to agree with noticeable probability (which would contradict Lemma 4.1).

Lemma 6.14. *Let $k_0 = k_0(q, \epsilon)$, $d = d(q, \epsilon)$ and $\tau = \tau(q, \epsilon)$ be as given in Theorem 6.8, and let $t = t(\tau, d)$ as given by Proposition 6.4. If $\mathrm{succ}_{(\mathcal{Y}', \mathcal{N}), \rho}(\mathcal{F}, \mathcal{G}) \geq 2\epsilon$, and $k < k_0$ then*

$$\Pr_{r \sim_\rho s}[\mathsf{Ext}_C(r) = \mathsf{Ext}_D(s)] \geq \epsilon/(2^{2k} t^2) \; .$$

Proof. Expanding the definition of $\mathrm{succ}(\cdot, \cdot)$, we have

$$\left| \mathbb{E}_{(r,s)} \left[\mathbb{E}_{(x,y) \sim \mathcal{Y}'} \left[\left\langle f^{(r)}(x), g^{(s)}(y) \right\rangle \right] - \mathbb{E}_{(x,y) \sim \mathcal{N}} \left[\left\langle f^{(r)}(x), g^{(s)}(y) \right\rangle \right] \right] \right| \geq 2\epsilon.$$

Say that a pair (r, s) is $\textsc{Good}$ if

$$\left| \mathbb{E}_{(x,y) \sim \mathcal{Y}'} \left[\left\langle f^{(r)}(x), g^{(s)}(y) \right\rangle \right] - \mathbb{E}_{(x,y) \sim \mathcal{N}} \left[\left\langle f^{(r)}(x), g^{(s)}(y) \right\rangle \right] \right| \geq \epsilon.$$

By a Markov argument we thus have

$$\Pr_{(r,s)} [\, (r, s) \text{ is } \textsc{Good} \,] \geq \epsilon.$$

For any fixed $\textsc{Good}$ (r, s) we now prove that there exists $i \in (S_r \cap T_s) \setminus (\textsc{Bad}_C \cup \textsc{Bad}_D)$. Note that once we have such an i, we have that $\Pr[\mathsf{Ext}_C(r) = \mathsf{Ext}_D(s) = i]$ with probability at least $1/(2^k t(\tau, d))^2$ by Proposition 6.4. Combining this with the probability that (r, s) is good, we have $\Pr_{(r,s)}[\mathsf{Ext}_C(r) = \mathsf{Ext}_D(s)] \geq \epsilon/(2^{2k} t(\tau, d)^2)$ which yields the lemma. So we turn to this claim.

To simplify notation, assume without loss of generality that $\textsc{Bad}_C \cup \textsc{Bad}_D = \{m+1, \ldots, n\}$. Define functions $f_1 \colon \{0,1\}^m \to K_A^{(k)}$ and $g_1 \colon \{0,1\}^m \to K_B^{(k)}$ by letting

$$f_1(x) = \mathbb{E}_{z \sim \mathrm{Bern}^{n-m}(1/q)}[f^{(r)}(x \cdot z)] \quad \text{and} \quad g_1(y) = \mathbb{E}_{w \sim \mathcal{U}(\{0,1\}^{n-m})}[g^{(s)}(y \cdot w)]$$

where $u \cdot v$ denotes the concatenation of u and v. Note that the success of $(f^{(r)}, g^{(s)})$ in distinguishing $\mathcal{Y}'$ from $\mathcal{N}$ turns into the success of (f_1, g_1) in distinguishing $\mathcal{Y}_m$ from $\mathcal{N}_m$ (where $\mathcal{Y}_m = B_Y^{\otimes m}$ and $\mathcal{N}_m = B_N^{\otimes m}$) — this is immediate since $(x \cdot z, y \cdot w) \sim \mathcal{Y}'$ if $(x, y) \sim \mathcal{Y}_m$ and $(x \cdot z, y \cdot w) \sim \mathcal{N}$ if $(x, y) \sim \mathcal{N}_m$.

So we have $\mathrm{succ}_{(\mathcal{Y}_m, \mathcal{N}_m)}(f_1, g_1) \geq \epsilon$. Since $k < k_0$ we have that there must exist a variable $i \in [m]$ and indices $j, j' \in [2^k]$ with $\mathsf{Inf}_i^d(f_{1,j}) > \tau$ and $\mathsf{Inf}_i^d(g_{1,j'}) > \tau$. (Here $f_{1,j}$ is the j'th component function of f_1, and similarly for $g_{1,j'}$.) Indeed, this follows from Theorem 6.8, and Proposition 6.10 (which ensures that $m \geq \frac{98}{100} n$, and therefore sufficiently large for the conclusion of the theorem to hold). But $\mathsf{Inf}_i^d(f_j^{(r)}) \geq \mathsf{Inf}_i^d(f_{1,j})$ and $\mathsf{Inf}_i^d(g_{j'}^{(r')}) \geq \mathsf{Inf}_i^d(g_{1,j'})$. To see this, note that $\widehat{f_{1,j}}(S) = \widehat{f_j^{(r)}}(S)$ for $S \subseteq [m]$ and so

$$\begin{aligned}
\mathsf{Inf}_i^d(f_j^{(r)}) &= \sum_{i \in S \subseteq [n], |S| \leq d} \widehat{f_j^{(r)}}(S)^2 \\
&\geq \sum_{i \in S \subseteq [m], |S| \leq d} \widehat{f_j^{(r)}}(S)^2 \\
&= \sum_{i \in S \subseteq [m], |S| \leq d} \widehat{f_{1,j}}(S)^2 \\
&= \mathsf{Inf}_i^d(f_{1,j}).
\end{aligned}$$

We thus conclude that $i \in S_r \cap T_s \cap [m]$ and this concludes the claim, and thus the lemma. □

Proof of Theorem 5.8. The proof follows easily from Lemma 4.1 and Lemma 6.14. Assume for contradiction that there is a protocol for $\textsc{SparseGapInnerProduct}^n_{.99q,.9q,.6q}$ with communication complexity less than $k_0(q, .05) = \Omega(\sqrt{q})$ that on access to $r \sim_\rho s$ accepts yes-instances with probability at least $2/3$ and no-instances with probability at most $1/3$. This implies that there exist strategies $(\mathcal{F} = \{f^{(r)}\}_r, \mathcal{G} = \{g^{(s)}\}_s)$ such that for every pair of distributions $(\mathcal{Y}, \mathcal{N})$ supported mostly (i.e., with probability .9) on yes and no instances respectively, we have $\mathrm{succ}_{(\mathcal{Y},\mathcal{N}),\rho}(\mathcal{F}, \mathcal{G}) > .1$. In particular, this holds for the distribution $\mathcal{Y}'$ as defined in Definition 6.13 and $\mathcal{N}$ as defined in Definition 6.7.

Let Ext_C, Ext_D be strategies for $\textsc{Agreement-Distillation}$ as defined in Definition 6.11. By Proposition 6.12 we get that $H_\infty(\mathsf{Ext}_C(r)), H_\infty(\mathsf{Ext}_D(s)) \geq \log n - O(1)$. By Lemma 6.14 we also have $\Pr_{r \sim_\rho s}[\mathsf{Ext}_C(r) = \mathsf{Ext}_D(s)] \geq \Omega_q(1)$. But this contradicts Lemma 4.1 which asserts (in particular) that protocols extracting $\omega_n(1)$ bits can only agree with probability $o_n(1)$. □

7 Low-influence communication strategies

The following theorem states that the expected inner product between two multidimensional Boolean functions without common low-degree influential variables when applied to correlated

random strings, is well approximated by the expected inner product of two related functions, this time applied to similarly correlated Gaussians. As, per Section 5.1, the former quantity captures the behavior of communication protocols, this invariance principle enables one to transfer the study to the (more manageable) Gaussian setting. (For convenience, in this section we switch to the equivalent view of Boolean functions as being defined on $\{+1,-1\}^n$).

We denote by $N_{p_1,p_2,\theta}$ the distribution on $\{+1,-1\} \times \{+1,-1\}$ such that the marginals of $(x,y) \sim N_{p_1,p_2,\theta}$ have expectations respectively p_1 and p_2, and correlation θ (see Definition A.1 for an explicit definition).

Theorem 7.1. *Fix any two parameters $p_1, p_2 \in (-1,1)$. For all $\epsilon \in (0,1]$, $\ell \in \mathbb{Z}^+$, $\theta_0 \in [0,1)$ and closed convex sets $K_1, K_2 \subseteq [0,1]^\ell$, there exist $n_0 \in \mathbb{Z}^+$, $d \in \mathbb{Z}^+$ and $\tau \in (0,1)$ such that the following holds. For all $n \geq n_0$, there exist mappings*

$$T_1 \colon \{f \colon \{+1,-1\}^n \to K_1\} \to \{F \colon \mathbb{R}^n \to K_1\}$$
$$T_2 \colon \{g \colon \{+1,-1\}^n \to K_2\} \to \{G \colon \mathbb{R}^n \to K_2\}$$

such that for all $\theta \in [-\theta_0, \theta_0]$, if f, g satisfy

$$\max_{i \in [n]} \min \left(\max_{j \in [\ell]} \mathsf{Inf}_i^d(f_j), \max_{j \in [\ell]} \mathsf{Inf}_i^d(g_j) \right) \leq \tau \tag{4}$$

then, for $F = T_1(f)$ and $G = T_2(g)$, we have

$$\left| \mathbb{E}_{(x,y) \sim N^{\otimes n}} [\langle f(x), g(y) \rangle] - \mathbb{E}_{(X,Y) \sim \mathcal{G}^{\otimes n}} [\langle F(X), G(Y) \rangle] \right| \leq \epsilon. \tag{5}$$

where $N = N_{p_1,p_2,\theta}$ and $\mathcal{G}$ is the Gaussian distribution which matches the first and second-order moments of N, i.e. $\mathbb{E}[x_i] = \mathbb{E}[X_i]$, $\mathbb{E}[x_i^2] = \mathbb{E}[X_i^2]$ and $\mathbb{E}[x_i y_i] = \mathbb{E}[X_i Y_i]$.

The theorem follows in a straightforward manner from Lemma 7.2 and Theorem 7.3:

Proof of Theorem 7.1. For $\epsilon \in (0,1]$, $\ell \in \mathbb{Z}^+$ and $\theta_0 \in (0,1)$ as above, let $\tau_1 \stackrel{\text{def}}{=} \tau(\epsilon/2, \ell, \theta_0)$ as in Theorem 7.3. Define the operators T_1, T_2 as

$$T_1 = T_1^{(2)} \circ T_1^{(1)}, \quad T_2 = T_2^{(2)} \circ T_2^{(1)}$$

where $T_1^{(1)}$, $T_2^{(1)}$ are the operators from Lemma 7.2 (for $\epsilon/2$, ℓ, θ_0 and τ_1 as above, which yield the values of τ, d and n_0) and $T_1^{(2)}$, $T_2^{(2)}$ are the (non-linear) ones from Theorem 7.3 (with parameters ℓ, θ_0 and $\epsilon/2$). The result follows. □

The first step towards proving the theorem is to convert the expected inner product of Boolean functions with no shared low-degree influential variables into expected inner product of Boolean functions with no influential variables at all.

Lemma 7.2. *Fix any two parameters $p_1, p_2 \in (-1,1)$. For all $\epsilon \in (0,1]$, $\ell \in \mathbb{Z}^+$, $\tau \in (0,1)$, $\theta_0 \in [0,1)$ and convex sets $K_1, K_2 \subseteq [0,1]^\ell$, there exist $n_0 \in \mathbb{Z}^+$, $d \in \mathbb{Z}^+$ and $\tau' \in (0,1)$ such that the following holds. For all $n \geq n_0$ there exist operators*

$$T_1^{(1)} \colon \{f \colon \{+1,-1\}^n \to K_1\} \to \{\tilde{f} \colon \{+1,-1\}^n \to K_1\}$$
$$T_2^{(1)} \colon \{g \colon \{+1,-1\}^n \to K_2\} \to \{\tilde{g} \colon \{+1,-1\}^n \to K_2\}$$

such that for all $\theta \in [-\theta_0, \theta_0]$, *if* f, g *satisfy*

$$\max_{i\in[n]} \min\left(\max_{j\in[\ell]} \mathsf{Inf}_i^d(f_j), \max_{j\in[\ell]} \mathsf{Inf}_i^d(g_j)\right) \le \tau' \tag{6}$$

then, for $\tilde{f} = T_1^{(1)}(f)$ *and* $\tilde{g} = T_2^{(1)}(g)$,

$$\max_{i\in[n]} \max\left(\max_{j\in[\ell]} \mathsf{Inf}_i(\tilde{f}_j), \max_{j\in[\ell]} \mathsf{Inf}_i(\tilde{g}_j)\right) \le \tau \tag{7}$$

and

$$\left|\mathbb{E}_{(x,y)\sim N^{\otimes n}} \langle f(x), g(y)\rangle - \mathbb{E}_{(x,y)\sim N^{\otimes n}} \left\langle \tilde{f}(x), \tilde{g}(y)\right\rangle\right| \le \epsilon. \tag{8}$$

where $N = N_{p_1,p_2,\theta}$.

Proof. The proof uses Lemmas 6.1 and 6.7 in [Mos10] applied to each pair of functions (f_i, g_i), for $i \in [\ell]$ applied with parameter θ_0 and ϵ/ℓ; using when applying the first lemma the fact that the correlation of these $N_{p_1,p_2,\theta}$ is bounded away from 1. The operators given in Lemmas 6.1 and 6.7 in [Mos10] are simple averaging operators (averaging the value of f over some neighborhood of x to get its new value at x) and by the convexity of K_1 we have that the averaged value remains in K_1. Similarly for g and K_2. We omit the details. □

The last ingredient needed is the actual invariance principle, which will take us from the Boolean, low-influence setting to the Gaussian one.

Theorem 7.3. *Fix any two parameters* $p_1, p_2 \in (-1, 1)$. *For all* $\epsilon \in (0, 1]$, $\ell \in \mathbb{Z}^+$, $\theta_0 \in [0, 1)$, *and closed convex sets* $K_1, K_2 \subseteq [0, 1]^\ell$ *there exist* $\tau > 0$ *and mappings*

$$T_1^{(2)} \colon \{f \colon \{+1,-1\}^n \to K_1\} \to \{F \colon \mathbb{R}^n \to K_1\}$$
$$T_2^{(2)} \colon \{g \colon \{+1,-1\}^n \to K_2\} \to \{G \colon \mathbb{R}^n \to K_2\}$$

such that for all $\theta \in [-\theta_0, \theta_0]$, *if* $f \colon \{+1,-1\}^n \to K_1$ *and* $g \colon \{+1,-1\}^n \to K_2$ *satisfy*

$$\max_{i\in[n]} \max\left(\max_{j\in[\ell]} \mathsf{Inf}_i(f_j), \max_{j\in[\ell]} \mathsf{Inf}_i(g_j)\right) \le \tau$$

then for $F = T_1^{(2)}(f) \colon \mathbb{R}^n \to K_1$ *and* $G = T_2^{(2)}(g) \colon \mathbb{R}^n \to K_2$

$$\left|\mathbb{E}_{(x,y)\sim N^{\otimes n}} [\langle f(x), g(y)\rangle] - \mathbb{E}_{(X,Y)\sim \mathcal{G}^{\otimes n}} [\langle F(X), G(Y)\rangle]\right| \le \epsilon\ ,$$

where $N = N_{p_1,p_2,\theta}$ *and* $\mathcal{G}$ *is the Gaussian distribution which matches the first and second-order moments of* N.

Proof. Deferred to Appendix A. □

7.1 Lower bound for Gaussian Inner Product

We now deduce a lower bound on k, the communication complexity of the strategies captured by the range of f and g, needed to achieve sizeable advantage in distinguishing between ξ-correlated and uncorrelated Gaussian inputs. Hereafter, $\mathcal{G}_\rho$ denotes the bivariate normal Gaussian distribution with correlation ρ.

Lemma 7.4. *Let $\xi \in (0, 1/2), \gamma > 0$. There exists a function $k_1(\xi, \gamma) \geq \Omega_\gamma(1/\xi)$ such that for every n the following holds: if there are functions $F\colon \mathbb{R}^n \to K_A^{(k)}$ and $G\colon \mathbb{R}^n \to K_B^{(k)}$ such that*

$$|\mathbb{E}_{(x,y)\sim\mathcal{G}_\xi^{\otimes n}}[\langle F(x), G(y)\rangle] - \mathbb{E}_{(x,y)\sim\mathcal{G}_0^{\otimes n}}[\langle F(x), G(y)\rangle]| \geq \gamma \ ,$$

then $k \geq k_1(\xi, \gamma)$.

We will prove the above theorem by translating the above question to a communication lower bound question.
GaussianCorrelation$_{\xi,n}$: In this (promise) communication game, Alice holds $x \in \mathbb{R}^n$ and Bob holds $y \in \mathbb{R}^n$ from one of two distributions:

- μ_{yes}: each (x_i, y_i) is an independent pair of ξ-correlated standard normal variables.
- μ_{no}: each (x_i, y_i) is an independent copy of uncorrelated standard normal variables.

The goal is for Alice and Bob to communicate with each other, with shared randomness, and distinguish between the two cases with good advantage.

Note that if (X, Y) denotes the random variable each pair (x_i, y_i) is a realization of, estimating $\mathbb{E}[XY]$ within accuracy $< \xi/2$ suffices to solve the above problem. If Alice sends the values of x_i (suitably discretized) for the first $O(1/\xi^2)$ choices of i, then by standard Chebyshev tail bounds Bob can estimate $\mathbb{E}[XY]$ to the desired accuracy, and so this problem can be solved with $O(1/\xi^2)$ bits of (one-way) communication. We now show that $\Omega(1/\xi)$ is a lower bound.

Lemma 7.5. *Let $\xi \in (0, 1/2)$ and n be sufficiently large. Suppose there is a k-bit communication protocol for* GaussianCorrelation *(ξ, n) that distinguishes between μ_{yes} and μ_{no} with advantage $\gamma > 0$. Then $k \geq \Omega_\gamma(1/\xi)$.*

Before we prove the result, note that Lemma 7.4 follows immediately with $k_1(\xi, \gamma) = \Omega_\gamma(1/\xi)$, since by Proposition 5.3 the functions $F\colon \mathbb{R}^n \to K_A^{(k)}$ and $G\colon \mathbb{R}^n \to K_B^{(k)}$ simply correspond to strategies for a k-bit two-way communication protocol with acceptance probability given by $\mathbb{E}_{X,Y}[\langle F(X), G(Y)\rangle]$.

Proof of Lemma 7.5. The lower bound is proved by reducing the Disjointness problem (in particular a promise version of it) to the GaussianCorrelation problem.

Specifically we consider the promise Disjointness problem with parameter m, where Alice gets a vector $u \in \{0,1\}^m$ and Bob gets $v \in \{0,1\}^m$, such that $\|u\|_2^2 = \|v\|_2^2 = \frac{m}{3}$. The yes-instances satisfy $\langle u, v\rangle = 1$ while the no-instances satisfy $\langle u, v\rangle = 0$, where the inner product is over the reals. Håstad and Wigderson [HW07] show that distinguishing yes-instances from no-instances requires $\Omega(m)$ bits of communication, even with shared randomness.

We reduce Disjointness$_m$ to GaussianCorrelation with $\xi = 1/m$ as follows: Alice and Bob share mn independent standard Gaussians $\{ G_{ij} : i \in [n], j \in [m] \}$. Alice generates $x =$

$(x_1, \ldots, x_n)$ by letting $x_i = \sqrt{\frac{3}{m}} \sum_{j=1}^{m} u_j \cdot G_{ij}$ and Bob generates $y = (y_1, \ldots, y_n)$ by letting $y_i = \sqrt{\frac{3}{m}} \sum_{j=1}^{m} v_j \cdot G_{ij}$. It can be verified that x_i and y_i are standard Gaussians[6] with $\mathbb{E}[x_i y_i] = \frac{3}{m} \langle u, v \rangle$. Thus yes-instances of DISJOINTNESS map to yes-instances of GAUSSIANCORRELATION drawn according to μ_{yes} with $\xi = 3/m$, and no-instances map to no-instances drawn according to μ_{no}. The communication lower bound of $\Omega(m)$ for DISJOINTNESS thus translates to a lower bound of $\Omega(1/\xi)$ for GAUSSIANCORRELATION. □

7.2 Putting things together and proof of Theorem 6.8

We now combine the results from the previous two sections to prove Theorem 6.8.

Proof of Theorem 6.8. Postponing the precise setting of parameters for now, the main idea behind the proof is the following. Suppose the conclusion of the theorem does not hold and f, g do not have a common influential variable so that

$$\max_{i \in [n]} \min \left\{ \max_{j \in [2^k]} \mathsf{Inf}_i^d(f_j), \max_{j \in [2^k]} \mathsf{Inf}_i^d(g_j) \right\} \leq \tau \tag{9}$$

for parameters d, τ that can be picked with an arbitrary dependence on q, ϵ.

We now associate the domains of f and g with $\{+1, -1\}^n$ in the natural way by mapping $x \in \{0, 1\} \to 2x - 1 \in \{+1, -1\}$. This defines us functions $f' \colon \{+1, -1\}^n \to K_A^{(k)}$ and $g' \colon \{+1, -1\}^n \to K_B^{(k)}$ which satisfy the same conditions on influence as f. Further, under this mapping, the distribution B_Y is mapped to $N_{\mathcal{Y}} \equiv N_{2/q-1,0,1.9/q}$ and B_N is mapped to $N_{\mathcal{N}} \equiv N_{2/q-1,0,0}$ (for $N_{p_1,p_2,\theta}$ as defined in Theorem 7.1). Let $\mathcal{G}_{\mathcal{Y}}$ and $\mathcal{G}_{\mathcal{N}}$ denote bivariate Gaussian distributions whose first two moments match those of $N_{\mathcal{Y}}$ and $N_{\mathcal{N}}$ respectively.

Since the ranges of f', g' are closed and convex (from Proposition 5.3) we get, by applying Theorem 7.1 to functions f', g' and distributions $N_{\mathcal{Y}}, \mathcal{G}_{\mathcal{Y}}$ and $N_{\mathcal{N}}, \mathcal{G}_{\mathcal{N}}$ respectively, that there exist functions $F \colon \mathbb{R}^n \to K_A^{(k)}$ and $G \colon \mathbb{R}^n \to K_B^{(k)}$ such that

$$\begin{aligned} \left| \mathbb{E}_{(x,y) \sim N_{\mathcal{Y}}^{\otimes n}} [\langle f'(x), g'(y) \rangle] - \mathbb{E}_{(X,Y) \sim \mathcal{G}_{\mathcal{Y}}^{\otimes n}} [\langle F(X), G(Y) \rangle] \right| &\leq \frac{\epsilon}{3} \\ \left| \mathbb{E}_{(x,y) \sim N_{\mathcal{N}}^{\otimes n}} [\langle f'(x), g'(y) \rangle] - \mathbb{E}_{(X,Y) \sim \mathcal{G}_{\mathcal{N}}^{\otimes n}} [\langle F(X), G(Y) \rangle] \right| &\leq \frac{\epsilon}{3} . \end{aligned} \tag{10}$$

Combining the above equations with the hypothesis that $\mathrm{succ}_{(\mathcal{Y},\mathcal{N})}(f, g) \geq \epsilon$, we get

$$\left| \mathbb{E}_{(X,Y) \sim \mathcal{G}_{\mathcal{Y}}^{\otimes n}} [\langle F(X), G(Y) \rangle] - \mathbb{E}_{(X,Y) \sim \mathcal{G}_{\mathcal{N}}^{\otimes n}} [\langle F(X), G(Y) \rangle] \right| \geq \frac{\epsilon}{3} .$$

To finish the argument, we shall appeal to Lemma 7.4. Let $p = 1/q$ and $\theta = .95p/\sqrt{p - p^2} = \Theta(1/\sqrt{q})$. Let $\phi \colon \mathbb{R} \to \mathbb{R}$ be defined by $\phi(z) = 2\sqrt{p - p^2} \cdot z + (2p - 1)$. It is easy to check that

[6] Namely, for any i we have $\mathbb{E}[x_i] = \frac{\sqrt{3}}{\sqrt{m}} \sum_{j=1}^{m} u_j \underbrace{\mathbb{E}[G_{ij}]}_{=0} = 0$, and

$$\mathbb{E}[x_i^2] = \frac{3}{m} \sum_{j=1}^{m} \sum_{\ell=1}^{m} u_j u_\ell \underbrace{\mathbb{E}[G_{ij} G_{i\ell}]}_{=0 \text{ if } j \neq \ell} = \frac{3}{m} \sum_{j=1}^{m} u_j^2 \mathbb{E}[G_{ij}^2] = \frac{3}{m} \|u\|_2^2 = 1.$$

for $(z,w) \sim \mathcal{G}_\theta$, $(\phi(z),w) \sim \mathcal{G}_\mathcal{Y}$ and for $(z,w) \sim \mathcal{G}_0$, $(\phi(z),w) \sim \mathcal{G}_\mathcal{N}$. Therefore, if we define $F' \colon \mathbb{R}^n \to K_A^{(k)}$ by $F'(X) = F(\phi(X_1), \ldots, \phi(X_n))$, then the above equation is equivalent to

$$\left| \mathbb{E}_{(X,Y)\sim\mathcal{G}_\theta^{\otimes n}}[\langle F'(X), G(Y)\rangle] - \mathbb{E}_{(X,Y)\sim\mathcal{G}_0^{\otimes n}}[\langle F'(X), G(Y)\rangle] \right| \geq \frac{\epsilon}{3} \,.$$

We can now conclude from Lemma 7.4 that $k \geq \Omega_\epsilon(1/\theta) = \Omega_\epsilon(\sqrt{q})$. To complete the proof of theorem by a contradiction we set the parameters as follows: choose d, τ in Equation 9 so as to deduce Equation 10 from Theorem 7.1 (with $\epsilon/3$ playing role of ϵ) and set $k_0 = k_1(\theta, \epsilon/3)$ for k_1 as given by Lemma 7.4.

□

8 Conclusions

In this paper we carried out an investigation of the power of imperfectly shared randomness in the context of communication complexity. There are two important aspects to the perspective that motivated our work: First, the notion that in many forms of natural communication, the communicating parties understand each other (or "know" things about each other) fairly well, but never perfectly. This imperfection in knowledge/understanding creates an obstacle to many of the known solutions and new solutions have to be devised, or new techniques need to be developed to understand whether the obstacles are barriers. Indeed for the positive results described in this paper, classical solutions do not work and the solutions that ended up working are even "provably" different from classical solutions. (In particular they work hard to preserve "low influence").

However, we also wish to stress a second aspect that makes the problems here interesting in our view, which is an aspect of scale. Often in communication complexity our main motivation is to compute functions with sublinear communication, or prove linear lower bounds. Our work, and natural communication in general, stresses the setting where inputs are enormous, and the communication complexity one is considering is tiny. This models many aspects of natural communication where there is a huge context to any conversation which is implicit. If this context were known exactly to sender and receiver, then it would play no significant mathematical role. However in natural communication this context is not exactly known, and resolving this imperfection of knowledge before communicating the relevant message would be impossibly hard. Such a setting naturally motivates the need to study problems of input length n, but where any dependence on n in the communication complexity would be impractical.

We note that we are not at the end of the road regarding questions of this form: Indeed a natural extension to communication complexity might be where Alice wishes to compute $f_A(x,y)$ and Bob wishes to compute $f_B(x,y)$ but Alice does not know f_B and Bob does not know f_A (or have only approximate knowledge of these functions). If x and y are n-bits strings, f_A and f_B might require 2^n bits to describe and this might be the real input size. There is still a trivial upper bound of $2n$ bits for solving any such communication problem, but it would be interesting to study when, and what form of, approximate knowledge of f_A and f_B helps improve over this trivial bound.

Turning to the specific questions studied in this paper a fair number of natural questions arise that we have not been able to address in this work. For instance, we stuck to a specific and simple form of correlation in the randomness shared by Alice and Bob. One could ask what general forms of randomness (r, r') are equally powerful. In particular if the distribution of (r, r') is known to both Alice and Bob, can they convert their randomness to some form of correlation in the sense used in this paper (in product form with marginals being uniform)?

In Section 4 we considered the Agreement-Distillation problem where the goal was for Alice and Bob to agree perfectly on some random string. What if their goal is only to generate more correlated bits than they start with? What is possible here and what are the limits?

In the study of perfectly shared randomness, Newman's Theorem [New91] is a simple but powerful tool, showing that $O(\log n)$ bits of randomness suffice to deal with problems on n bit inputs. When randomness is shared imperfectly, such a randomness reduction is not obvious. Indeed for the problem of equality testing, the protocol of [BGI14] uses 2^n bits of randomness, and our Gaussian protocol (which can solve this with one-way communication) uses $\text{poly}(n)$ bits. Do $O(\log n)$ bits of imperfectly shared randomness suffice for this problem? How about for general problems?

Finally almost all protocols we give for imperfectly shared randomness lead to two-sided error. This appears to be an inherent limitation (with some philosophical implications) but we do not have a proof. It would be nice to show that one-sided error with imperfectly shared randomness cannot lead to any benefits beyond that offered by private randomness.

Acknowledgments

We thank Brendan Juba for his helpful notes [Jub09] on the invariance principle. We thank Badih Ghazi for pointing an error in a previous version of this paper and other discussions, and Jayanth Naranambadikalpathy Sadhasivan for pointing out an issue in the proof of Theorem 2.3. We thank the anonymous referee for extensive comments and corrections.

References

[AC93] Rudolf Ahlswede and Imre Csiszár. Common randomness in information theory and cryptography – Part I: secret sharing. *IEEE Transactions on Information Theory*, 39(4):1121–1132, 1993. 1

[AC98] Rudolf Ahlswede and Imre Csiszár. Common randomness in information theory and cryptography – Part II: CR capacity. *IEEE Transactions on Information Theory*, 44(1):225–240, 1998. 1, 2.2.2, 2.2.2

[BBP+96] Charles H. Bennett, Gilles Brassard, Sandu Popescu, Benjamin Schumacher, John A. Smolin, and William K. Wootters. Purification of noisy entanglement and faithful teleportation via noisy channels. *Phys. Rev. Lett.*, 76:722–725, Jan 1996. 1

[BG15] Salman Beigi and Amin Gohari. On the duality of additivity and tensorization. *arXiv preprint arXiv:1502.00827*, 2015. 1

[BGI14] Mohammad Bavarian, Dmitry Gavinsky, and Tsuyoshi Ito. On the role of shared randomness in simultaneous communication. In Javier Esparza, Pierre Fraigniaud, Thore Husfeldt, and Elias Koutsoupias, editors, *ICALP (1)*, volume 8572 of *Lecture Notes in Computer Science*, pages 150–162. Springer, 2014. 1, 8

[BK97] László Babai and Peter G. Kimmel. Randomized Simultaneous Messages: Solution of a Problem of Yao in Communication Complexity. In *Proceedings of the Twelfth Annual*

IEEE Conference on Computational Complexity, Ulm, Germany, June 24-27, 1997, pages 239–246. IEEE Computer Society, 1997. 4

[BM11] Andrej Bogdanov and Elchanan Mossel. On extracting common random bits from correlated sources. *Information Theory, IEEE Transactions on*, 57(10):6351–6355, Oct 2011. 1, 2.2.2, 2.2.2, 4, 4.1

[BS93] Gilles Brassard and Louis Salvail. Secret-key reconciliation by public discussion. In Tor Helleseth, editor, *Advances in Cryptology - EUROCRYPT '93, Workshop on the Theory and Application of of Cryptographic Techniques, Lofthus, Norway, May 23-27, 1993, Proceedings*, volume 765 of *Lecture Notes in Computer Science*, pages 410–423. Springer, 1993. 1

[CGMS15] Clément Louis Canonne, Venkatesan Guruswami, Raghu Meka, and Madhu Sudan. Communication with imperfectly shared randomness. In Tim Roughgarden, editor, *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science, ITCS 2015, Rehovot, Israel, January 11-13, 2015*, pages 257–262. ACM, 2015. ∗

[CMN14] Siu On Chan, Elchanan Mossel, and Joe Neeman. On extracting common random bits from correlated sources on large alphabets. *IEEE Transactions on Information Theory*, 60(3):1630–1637, 2014. 1

[CN00] Imre Csiszár and Prakash Narayan. Common randomness and secret key generation with a helper. *IEEE Transactions on Information Theory*, 46(2):344–366, 2000. 1

[CT91] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley Publishing, New York, 1991. 1

[FRKT15] Farzan Farnia, Meisam Razaviyayn, Sreeram Kannan, and David Tse. Minimum HGR Correlation Principle: From Marginals to Joint Distribution. *arXiv preprint arXiv:1504.06010*, 2015. 1

[GHM+11] Venkatesan Guruswami, Johan Håstad, Rajsekar Manokaran, Prasad Raghavendra, and Moses Charikar. Beating the random ordering is hard: Every ordering CSP is approximation resistant. *SIAM J. Comput.*, 40(3):878–914, 2011. A

[GK73] Peter Gács and János Körner. Common information is far less than mutual information. *Problems of Control and Information Theory*, 2(2):149–162, 1973. 1

[GR16] Venkatesan Guruswami and Jaikumar Radhakrishnan. Tight bounds for communication-assisted agreement distillation. In *CCC*, volume 50 of *LIPIcs*, pages 6:1–6:17. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2016. 1

[HS14] Elad Haramaty and Madhu Sudan. Deterministic compression with uncertain priors. In *Proceedings of the 5th Conference on Innovations in Theoretical Computer Science*, ITCS '14, pages 377–386, New York, NY, USA, 2014. ACM. 2.2.1

[HW07] Johan Håstad and Avi Wigderson. The Randomized Communication Complexity of Set Disjointness. *Theory of Computing*, 3(11):211–219, 2007. 7.1

[IM12] Marcus Isaksson and Elchanan Mossel. Maximally stable Gaussian partitions with discrete applications. *Israel Journal of Mathematics*, 189:347–396, June 2012. A

[JKKS11] Brendan Juba, Adam Tauman Kalai, Sanjeev Khanna, and Madhu Sudan. Compression without a common prior: an information-theoretic justification for ambiguity in language. In Bernard Chazelle, editor, *ICS*, pages 79–86. Tsinghua University Press, 2011. 1, 2.2.1, 2.2.1

[Jub09] Brendan Juba. 18.177 course project: Invariance principles, 2009. 8

[KA12] Sanmati Kamath and Venkat Anantharam. Non-interactive simulation of joint distributions: The Hirschfeld-Gebelein-Rényi maximal correlation and the hypercontractivity ribbon. In *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*, pages 1057–1064. IEEE, 2012. 1

[KN06] Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, New York, NY, USA, 2006. 1

[LT91] Michel Ledoux and Michel Talagrand. *Probability in Banach spaces: Isoperimetry and processes*. Springer, Berlin, 1991. 5.2

[Mau93] Ueli M. Maurer. Secret key agreement by public discussion from common information. *IEEE Transactions on Information Theory*, 39(3):733–742, 1993. 1

[MO05] Elchanan Mossel and Ryan O'Donnell. Coin flipping from a cosmic source: On error correction of truly random bits. *Random Struct. Algorithms*, 26(4):418–436, 2005. 1

[MOR+06] Elchanan Mossel, Ryan O'Donnell, Oded Regev, JeffreyE. Steif, and Benny Sudakov. Non-interactive correlation distillation, inhomogeneous Markov chains, and the reverse Bonami–Beckner inequality. *Israel Journal of Mathematics*, 154(1):299–336, 2006. 1

[Mos10] Elchanan Mossel. Gaussian bounds for noise correlation of functions. *Geometric and Functional Analysis*, 19(6):1713–1756, 2010. 6.2, 6.3, 7, A

[New91] Ilan Newman. Private vs. common random bits in communication complexity. *Inf. Process. Lett.*, 39(2):67–71, July 1991. 1, 2.1, 8

[RW05] Renato Renner and Stefan Wolf. Simple and tight bounds for information reconciliation and privacy amplification. In Bimal K. Roy, editor, *Advances in Cryptology - ASIACRYPT 2005, 11th International Conference on the Theory and Application of Cryptology and Information Security, Chennai, India, December 4-8, 2005, Proceedings*, volume 3788 of *Lecture Notes in Computer Science*, pages 199–216. Springer, 2005. 1

[Wit75] Hans S. Witsenhausen. On sequences of pairs of dependent random variables. *SIAM Journal on Applied Mathematics*, 28(1):100–113, 1975. 1

A Proofs from Section 7

Our goal in this section is to prove the needed invariance principle, as stated in Theorem 7.3, that allows us to pass from a correlated distribution on $\{+1,-1\}^2$ to a two-dimensional Gaussian distribution with matching moments. We first formally define the discrete distribution of interest to us.

Definition A.1. For parameters $p_1, p_2, \theta \in [-1,1]$, let the distribution $N_{p_1,p_2,\theta}$ on $\{+1,-1\} \times \{+1,-1\}$ be defined as follows:[7]

$$(x,y) = \begin{cases} (+1,+1) & \text{with probability } \frac{1+\theta}{4} + \frac{p_1+p_2}{4} \\ (+1,-1) & \text{with probability } \frac{1-\theta}{4} + \frac{p_1-p_2}{4} \\ (-1,+1) & \text{with probability } \frac{1-\theta}{4} - \frac{p_1-p_2}{4} \\ (-1,-1) & \text{with probability } \frac{1+\theta}{4} - \frac{p_1+p_2}{4} \end{cases}$$

so that $\mathbb{E}[x] = p_1$, $\mathbb{E}[y] = p_2$ and $\mathbb{E}[xy] = \theta$.

The proof of Theorem 7.3 relies on two general ingredients. The first is that replacing f and g by their *smoothened* versions $T_{1-\eta}f$ and $T_{1-\eta}g$ (obtained by applying the Bonami–Beckner noise operator, defined below) does not change the inner product $\langle f(x), g(y)\rangle$ much, due to the fact that the components (x_j, y_j) are sampled independently from a bounded correlation space (namely $N_{p_1,p_2,\theta}$ for $\theta < 1$). The second is a multi-dimensional invariance principle asserting that these smoothened functions behave similarly on Gaussian inputs that have matching moments, with respect to Lipschitz test functions. We then apply this to the Lipschitz function which is the inner product of appropriately rounded versions of inputs, thereby yielding K_1 and K_2 valued functions in the Gaussian domain with inner product close to $\langle f(x), g(y)\rangle$.

Definition A.2 (Bonami–Beckner $T_{1-\eta}$ operator)**.** Let (Ω, μ) be a finite probability space, and $\eta \in (0,1)$. For a function $h\colon \Omega^n \to \mathbb{R}$, the function $T_{1-\eta}h$ is defined as $T_{1-\eta}h(x) = \mathbb{E}_y[h(y)]$, where each coordinate y_i is sampled independently as follows:

- with probability $(1-\eta)$ set $y_i = x_i$; and
- with probability η, pick $y_i \in \Omega$ as a fresh sample according to μ.

For a vector-valued function, $T_{1-\eta}$ acts component-wise, i.e., if $f = (f_1, \ldots, f_\ell)\colon \Omega^n \to \mathbb{R}^\ell$, we define $T_{1-\eta}f = (T_{1-\eta}f_1, \ldots, T_{1-\eta}f_\ell)$.

A useful property of the $T_{1-\eta}$ operator for us is that if h has convex range $K \subseteq [0,1]^\ell$ then so does $T_{1-\eta}h$. As stated below, the action of $T_{1-\eta}$ has a particularly nice form when a function is expanded in an orthonormal basis, but this will not be important for us.

Fact A.3. *If a function $h\colon \Omega^n \to \mathbb{R}$ has multilinear expansion $h(x) = \sum_{\boldsymbol{\sigma}} \hat{h}_{\boldsymbol{\sigma}} \prod_{i=1}^n \chi_{\sigma_i}(x_i)$ w.r.t. an orthonormal ensemble $\mathcal{L} = (\chi_0, \ldots, \chi_{b-1})$ of random variables over Ω, then the multilinear expansion of $T_{1-\eta}h$ is given by $\sum_{\boldsymbol{\sigma}} \hat{h}_{\boldsymbol{\sigma}} (1-\eta)^{|\boldsymbol{\sigma}|} \prod_{i=1}^n \chi_{\sigma_i}(x_i)$.*

[7] We assume that the parameters p_1, p_2, θ are such that each of the probabilities is in $[0,1]$.

We next state the multi-dimensional invariance principle that we rely on. A version similar to the following is stated formally in [GHM+11, Theorem 10.1] (we have renamed some variables to avoid conflict with other uses in this paper) and it follows from Theorem 3.6 in the work of Isaksson and Mossel [IM12].

Theorem A.4. *Let (Ω, μ) be a finite probability space with the least non-zero probability of an atom being at least $\alpha \leq 1/2$. Let $b = |\Omega|$ and let $\mathcal{L} = \{\chi_0 = 1, \chi_1, \chi_2, \ldots, \chi_{b-1}\}$ be a basis for random variables over Ω. Let $\Upsilon = \{\xi_0 = 1, \xi_1, \ldots, \xi_{b-1}\}$ be an ensemble of real-valued Gaussian random variables with first and second moments matching those of the χ_i's; specifically:*

$$\mathbb{E}[\chi_i] = \mathbb{E}[\xi_i] \qquad \mathbb{E}[\chi_i^2] = \mathbb{E}[\xi_i^2] \qquad \mathbb{E}[\chi_i \chi_j] = \mathbb{E}[\xi_i \xi_j] \qquad \forall i, j \in \{1, \ldots, b-1\}$$

Let $h = (h_1, h_2, \ldots, h_t) \colon \Omega^n \to \mathbb{R}^t$ be a vector-valued function such that $\mathsf{Inf}_i(h_\ell) \leq \tau$ and $\mathsf{Var}(h_\ell) \leq 1$ for all $i \in [n]$ and $\ell \in [t]$. For $\eta \in (0,1)$, let H_ℓ, $\ell = 1, 2, \ldots, t$, be the multilinear polynomial associated with $T_{1-\eta} h_\ell$ with respect to the basis $\mathcal{L}$, as per Definition 6.5.

If $\Psi \colon \mathbb{R}^t \to \mathbb{R}$ is a Lipschitz-continuous function with Lipschitz constant Λ (with respect to the L_2-norm), then

$$\left| \mathbb{E}\Big[\Psi\big(H_1(\mathcal{L}^n), \cdots, H_t(\mathcal{L}^n)\big)\Big] - \mathbb{E}\Big[\Psi\big(H_1(\Upsilon^n), \cdots, H_t(\Upsilon^n)\big)\Big] \right| \leq C(t) \cdot \Lambda \cdot \tau^{(\eta/18)\log(1/\alpha)} = o_\tau(1) \quad (11)$$

for some constant $C(t)$ depending on t, where $H_\ell(\mathcal{L}^n)$ and $H_\ell(\Upsilon^n)$, $\ell \in [t]$, denote random variables as in Definition 6.6.

Armed with the above invariance principle, we now turn to the proof of Theorem 7.3, restated below.

Theorem 7.3. *Fix any two parameters $p_1, p_2 \in (-1, 1)$. For all $\epsilon \in (0, 1]$, $\ell \in \mathbb{Z}^+$, $\theta_0 \in [0, 1)$, and closed convex sets $K_1, K_2 \subseteq [0,1]^\ell$ there exist $\tau > 0$ and mappings*

$$T_1^{(2)} \colon \{f \colon \{+1,-1\}^n \to K_1\} \to \{F \colon \mathbb{R}^n \to K_1\}$$
$$T_2^{(2)} \colon \{g \colon \{+1,-1\}^n \to K_2\} \to \{G \colon \mathbb{R}^n \to K_2\}$$

such that for all $\theta \in [-\theta_0, \theta_0]$, if $f \colon \{+1,-1\}^n \to K_1$ and $g \colon \{+1,-1\}^n \to K_2$ satisfy

$$\max_{i \in [n]} \max\left(\max_{j \in [\ell]} \mathsf{Inf}_i(f_j), \max_{j \in [\ell]} \mathsf{Inf}_i(g_j) \right) \leq \tau$$

then for $F = T_1^{(2)}(f) \colon \mathbb{R}^n \to K_1$ and $G = T_2^{(2)}(g) \colon \mathbb{R}^n \to K_2$

$$\left| \mathbb{E}_{(x,y) \sim N^{\otimes n}} \left[\langle f(x), g(y) \rangle \right] - \mathbb{E}_{(X,Y) \sim \mathcal{G}^{\otimes n}} \left[\langle F(X), G(Y) \rangle \right] \right| \leq \epsilon \ ,$$

where $N = N_{p_1, p_2, \theta}$ and $\mathcal{G}$ is the Gaussian distribution which matches the first and second-order moments of N.

Proof of Theorem 7.3. Let $\Omega = \{+1,-1\} \times \{+1,-1\}$ with the measure $N := N_{p_1,p_2,\theta}$. Define the basis $\mathcal{L} = \{\chi_0, \chi_1, \chi_2, \chi_3\}$ of functions on Ω as:

- $\chi_0 = 1$,

- $\chi_1((w_1, w_2)) = w_1$ (where $w_1, w_2 \in \{+1, -1\}$),
- $\chi_2((w_1, w_2)) = w_2$, and
- $\chi_3((w_1, w_2)) = w_1 w_2$.

We will apply the above invariance principle Theorem A.4 with $t = 2\ell$, $h_j = f_j$ and $h_{\ell+j} = g_j$ for $j \in [\ell]$. We note that while f_j, $j \in [\ell]$ are functions on $\{+1, -1\}^n$, we can view them as functions on Ω^n by simply ignoring the second coordinate. (Thus, for $(x, y) \sim \Omega^n$, $f_j(x, y) = f_j(x)$.) The multilinear expansion of f_j w.r.t. $\mathcal{L}$ will only involve χ_0 and χ_1. Similarly, the functions h_j's only depend on the second coordinate of Ω and have a multilinear expansion depending only on χ_0, χ_2. The function $\Psi \colon \mathbb{R}^{2\ell} \to \mathbb{R}$ is defined as

$$\Psi(\mathbf{a}, \mathbf{b}) = \langle \mathsf{Round}_{K_1}(\mathbf{a}), \mathsf{Round}_{K_2}(\mathbf{b}) \rangle$$

for $\mathbf{a}, \mathbf{b} \in \mathbb{R}^\ell$, where for a *closed convex* set $K \subset \mathbb{R}^\ell$, $\mathsf{Round}_K \colon \mathbb{R}^\ell \to \mathbb{R}^\ell$ maps a point to its (unique) closest point (in Euclidean distance) in K – in particular, it is the identity map on K. It is easy to see that by the convexity of K, Round_K is a 1-Lipschitz function,[8] and it follows that the function Ψ is $O(\sqrt{\ell})$-Lipschitz. Also, since $T_{1-\eta} f$ is K_1-valued and $T_{1-\eta} g$ is K_2-valued on $\{+1, -1\}^n$, the Round functions act as the identity on their images, and hence

$$\mathbb{E}\Big[\Psi\big(H_1(\mathcal{L}^n), \cdots, H_t(\mathcal{L}^n)\big)\Big] = \mathbb{E}_{(x,y)}\Big[\langle T_{1-\eta} f(x), T_{1-\eta} g(y) \rangle\Big], \tag{12}$$

where (x, y) is distributed according to $N_{p_1, p_2, \theta}^{\otimes n}$.

For $j \in [\ell]$, define real-valued functions $\tilde{F}_j = H_j(\Upsilon^n)$ and $\tilde{G}_j = H_{\ell+j}(\Upsilon^n)$. Note that as the multilinear expansion of $T_{1-\eta} f_j$ (resp. $T_{1-\eta} h_j$) only involves χ_0, χ_1 (resp. χ_0, χ_2), the multilinear expansion of $\tilde{F}_j$ (resp. $\tilde{G}_j$) only involves ξ_0, ξ_1 (resp. ξ_0, ξ_2). As $\xi_0 = 1$, the functions $\tilde{F}_j$ (resp. $\tilde{G}_j$) are defined on $\mathbb{R}^n$ under a product measure with coordinates distributed as Gaussians with mean p_1 (resp. mean p_2) and second moment 1.

Let $\tilde{F} = (\tilde{F}_1, \ldots, \tilde{F}_\ell)$ and $\tilde{G} = (\tilde{G}_1, \ldots, \tilde{G}_\ell)$, and finally let $F \colon \mathbb{R}^n \to K_1$ be $F(X) = \mathsf{Round}_{K_1}(\tilde{F}(X))$ and $G \colon \mathbb{R}^n \to K_2$ be $G(Y) = \mathsf{Round}_{K_2}(\tilde{G}(Y))$. Note that F (resp. G) depends only on $f = (f_1, \ldots, f_\ell)$ (resp. $g = (g_1, \ldots, g_\ell)$) as required in the statement of Theorem 7.3. By construction, it is clear that

$$\mathbb{E}\Big[\Psi\big(H_1(\Upsilon^n), \cdots, H_t(\Upsilon^n)\big)\Big] = \mathbb{E}_{(X,Y)}\Big[\langle F(X), G(Y) \rangle\Big], \tag{13}$$

for $(X, Y) \sim (\xi_1, \xi_2)^{\otimes n} = \mathcal{G}^{\otimes n}$ where $\mathcal{G}$ is the Gaussian distribution which matches the first and second moments of $N = N_{p_1, p_2, \theta}$.

Combining (12) and (13) with the guarantee (11) of Theorem A.4, we get that

$$\left|\mathbb{E}_{(x,y) \sim N^{\otimes n}}\left[\langle T_{1-\eta} f(x), T_{1-\eta} g(y) \rangle\right] - \mathbb{E}_{(X,Y) \sim \mathcal{G}^{\otimes n}}\left[\langle F(X), G(Y) \rangle\right]\right| \le \epsilon/2 \tag{14}$$

for $\tau > 0$ chosen small enough (as a function of $\epsilon, \ell, p_1, p_2, \theta_0$ and η). We are almost done, except that we would like to be close to the inner product $\langle f(x), g(y) \rangle$ of the original functions, and we

[8] To see why, let a, b be two arbitrary points and $a' = \mathsf{Round}_K(a)$, $b' = \mathsf{Round}_K(b)$. Without loss of generality, we can change the coordinates so that $a' = (0, \ldots, 0)$ and $b' = (c, 0, \ldots, 0)$ for some $c > 0$: by convexity, the segment $[a'b']$ lies within K. Now, by virtue of a' (resp. b') being the closest point to a (resp. b), this implies the first coordinate of a must be non-positive and the first coordinate of b must be at least c; but this in turn means the distance between a and b is at least c.

have the noised versions in (14) above. However, as the correlation of the space $N_{p_1,p_2,\theta}$ is bounded away from 1, applying Lemma 6.1 of [Mos10] implies that for small enough $\eta > 0$ (as a function of $\epsilon, \ell, p_1, p_2, \theta_0$),

$$\left|\mathbb{E}_{(x,y)\sim N^{\otimes n}}\left[\langle T_{1-\eta} f(x), T_{1-\eta} g(y)\rangle\right] - \mathbb{E}_{(x,y)\sim N^{\otimes n}}\left[\langle f(x), g(y)\rangle\right]\right| \leq \epsilon/2 \ .$$

Combining this with (14), the proof of Theorem 7.3 is complete. □

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il