Deterministic Randomness Extraction from Generalized and Distributed Santha-Vazirani Sources

Salman Beigi[1], Omid Etesami[1(✉)], and Amin Gohari[1,2]

[1] School of Mathematics, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran
salman.beigi@gmail.com, etesami@ipm.ir
[2] Department of Electrical Engineering, Sharif University of Technology, Tehran, Iran
aminzadeh@sharif.edu

Abstract. A Santha-Vazirani (SV) source is a sequence of random bits where the conditional distribution of each bit, given the previous bits, can be partially controlled by an adversary. Santha and Vazirani show that deterministic randomness extraction from these sources is impossible. In this paper, we study the generalization of SV sources for non-binary sequences. We show that unlike the binary case, deterministic randomness extraction in the generalized case is sometimes possible. We present a necessary condition and a sufficient condition for the possibility of deterministic randomness extraction. These two conditions coincide in "non-degenerate" cases.

Next, we turn to a distributed setting. In this setting the SV source consists of a random sequence of pairs $(a_1, b_1), (a_2, b_2), \ldots$ distributed between two parties, where the first party receives a_i's and the second one receives b_i's. The goal of the two parties is to extract common randomness without communication. Using the notion of *maximal correlation*, we prove a necessary condition and a sufficient condition for the possibility of common randomness extraction from these sources. Based on these two conditions, the problem of common randomness extraction essentially reduces to the problem of randomness extraction from (non-distributed) SV sources. This result generalizes results of Gács and Körner, and Witsenhausen about common randomness extraction from i.i.d. sources to adversarial sources.

1 Introduction

Randomized algorithms are simpler and more efficient than their deterministic counterparts in many applications. In some settings such as communication complexity and distributed computing, it is even possible to prove unconditionally that allowing randomness improves the efficiency of algorithms (see e.g., [14,19,30]). However, access to sources of randomness (especially common randomness) may be limited, or the quality of randomness in the source may be far from perfect. Having such an imperfect source of randomness, one may be able to extract

M.M. Halldórsson et al. (Eds.): ICALP 2015, Part I, LNCS 9134, pp. 143–154, 2015.
DOI: 10.1007/978-3-662-47672-7_12

(almost) unbiased and independent random bits using *randomness extractors*. A randomness extractor is a function applied to an imperfect source of randomness whose outcome is an almost perfect source of randomness.

The problem of randomness extraction from imperfect sources of randomness was perhaps first considered by Von Neumann [28]. A later important work in this area is [23] where Santha and Vazirani introduced the imperfect sources of randomness now often called Santha-Vazirani (SV) sources. These sources can easily be defined in terms of an adversary with two coins. Consider an adversary who has two different coins, one of which is biased towards heads (e.g., $\Pr(\text{heads}) = 2/3$) and the other one is biased towards tails (e.g., $\Pr(\text{heads}) = 1/3$). The adversary, in each time step, chooses one of the two coins and tosses it. Adversary's choice of coin may depend (probabilistically) on the previous outcomes of the tosses. The sequence of random outcomes of these coin tosses is called a SV source.

Santha and Vazirani [23] show that randomness extraction from the above sources through a deterministic method is impossible. More precisely, they show that for every deterministic way of extracting one random bit, there is a strategy for the adversary such that the extracted bit is biased, or more specifically, the extracted bit is 0 with probability either $\geq 2/3$ or $\leq 1/3$. Subsequently, other proofs for this result have been found (see e.g., [1,21]). Fig 1 shows a more refined version of this result, which provides a more detailed picture of the limits of what the adversary can achieve.

Despite this negative result, such imperfect sources of randomness are enough for many applications. For example, as shown by Vazirani and Vazirani [25,26], randomized polynomial-time algorithms that use perfect random bits can be simulated using SV sources. This fact can also be verified using the fact that the min-entropy of SV sources is linear in the size of the source (where min-entropy, in the context of extractors, was first introduced by [9]). Indeed, by the later theory of randomness extraction (e.g., see [31]), it is possible to efficiently extract polynomially many almost random bits from such sources with high min-entropy if we are, in addition to the imperfect source, endowed with a perfectly random seed of logarithmic length. (In fact, for the special case of SV sources, a seed of constant length is enough [27, Problem6.6]). For the application of randomized polynomial-time algorithms, we can enumerate in polynomial time over all possible seeds.

Enumerating over all seeds may be inefficient for some applications, or does not work at all, e.g., in interactive proofs and one-shot scenarios such as cryptography. Therefore, it is natural to ask whether deterministic randomness extraction from imperfect sources of randomness is possible. For most applications, it is also necessary to require that the extractor be explicit, i.e., extraction can be done efficiently (in polynomial time). Previous to this work, explicit deterministic extractors had been constructed for many different classes of sources, including i.i.d. bits with unknown bias [28], Markov chains [5], affine sources [7,16], polynomial sources [11,12], and sources consisting of independent blocks [6].

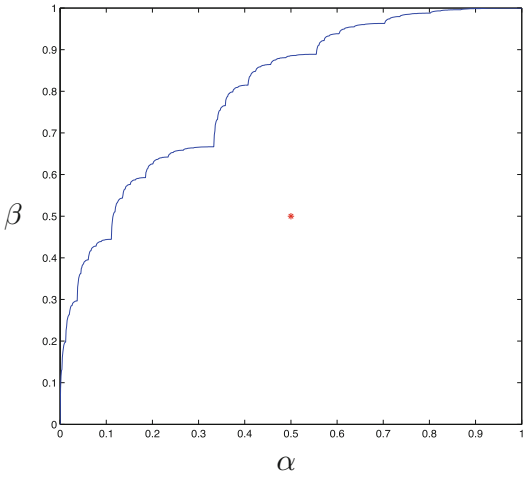

Fig. 1. Given any deterministic extractor, the pair (α, β) is above the curve specified in this figure, where α and β are the minimum and maximum value of probability of the output being zero that the adversary can achieve by choosing its strategy. The plot is for the binary SV source with two coins with probability of heads respectively equal to $1/3$ and $2/3$. The point $(1/2, 1/2)$ is specified by a red star in the figure. The curve has fractal-like self-similiarity: The curve can be split at point $(1/3, 2/3)$ into two curves each of which is a normalized version of the whole curve. To see how the curve is obtained, see Appendix A of the full version [2].

Deterministic Extractors for Generalized SV Sources. Although [23] proves the impossibility of deterministic randomness extraction from SV sources, this impossibility is shown only for binary sources. In this paper we show that if we consider a generalization of SV sources over *non-binary* alphabets, deterministic randomness extraction is indeed possible under certain conditions.

To generalize SV sources over non-binary alphabets, we assume that the adversary, instead of coins, has some multi-faceted (say 6-sided) dice. The numbers written on the faces of different dice are the same, but each die may have a different probability for a given face value. The adversary throws these dice n times, each time choosing a die to throw depending on the results of the previous throws. Again, the outcome is an imperfect source of randomness, for which we may ask whether deterministic randomness extraction is possible or not.

When the dice are non-degenerate, i.e., all faces of all dice have non-zero probability, we give a necessary and sufficient condition for the existence of a deterministic strategy for extracting one bit with arbitrarily small bias. For example, when the dice are 6-sided, the necessary and sufficient condition implies that we can deterministically extract an almost unbiased bit when the adversary has access to any arbitrary set of five non-degenerate dice, but randomness extraction is not possible in general when the adversary has access to six non-degenerate 6-sided dice. More precisely, a set of non-degenerate dice leads

to extractable generalized SV sources if and only if the convex hull of the set of probability distributions associated with the set of dice does not have full dimension in the "probability simplex". We emphasize that when we prove the possibility of deterministic extraction, we also provide an explicit extractor.

Relation to Block-Sources. The generalized SV sources considered in this paper are also a generalization of "block-sources" defined by Chor and Goldreich [9], where the source is divided into several blocks such that each block has min-entropy at least k conditioned on the value of the previous blocks. Such a block-source can be thought as a generalized SV source where the adversary can generate each block (given previous blocks) using any "flat" distribution with support 2^k. Being a special case of generalized SV sources (defined here), block-sources have another difference as well: Since it is impossible to extract from a single block-source deterministically, the common results regarding extraction from block-sources are about either seeded extractors (e.g. [18]) or extraction from at least two independent block-sources (e.g. [20]).

Common Randomness Extractors. Common random bits, shared by distinct parties, constitute an important resource for distributed algorithms; common random bits can be used by the parties to synchronize the randomness of their local actions. We may ask the question of randomness extraction in this setting too. Assuming that the parties are provided with an imperfect source of common randomness, the question is whether perfect common randomness can be extracted from this source or not.

Gács and Körner [15] and Witsenhausen [29] have looked at the problem of extraction of common random bits from a very special class of imperfect sources, namely i.i.d. sources. In this case, the *bipartite* source available to the parties is generated as follows: In each time step, a pair (A, B) with some predetermined distribution (known by the two parties) and independent of the past is generated; A is revealed to the first party and B is revealed to the second party. After receiving arbitrarily many repetitions of random variables A and B, the two parties aim to extract a common random bit. It is known that in this case, the two parties (who are not allowed to communicate) can generate a common random bit if and only if A and B have a common data [29]. This means that common randomness generation is possible if A and B can be expressed as $A = (A', C)$ and $B = (B', C)$ for a nonconstant common part C, i.e., there are nonconstant functions f, g such that $C = f(A) = g(B)$. Observe that when a common part exists, common randomness can be extracted by the parties by applying the same extractor on the sequence of C's. That is, the problem of common randomness extraction in the i.i.d. case is reduced to the problem of ordinary randomness extraction. These results are obtained using a measure of correlation called *maximal correlation*. The key feature of this measure of correlation that helps proving the above result is the *tensorization property*, i.e., the maximal correlation between random variables A and B is equal to that of A^n and B^n for any n, where A^n and B^n denote n i.i.d. repetitions of A and B.

In this paper we consider the problem of common randomness extraction from *distributed SV sources* defined as follows. In a distributed SV source, the adversary again has some multi-faceted dice, but here, instead of a single number, a pair of numbers (A, B) is written on each face. As before, the set of values written on the faces of the dice is the same, but the probabilities of face values may differ in different dice. In each time step, the adversary depending on the results of the previous throws, picks a die and throws it. If (A, B) is the result of the throw, A is given to the first party and B to second party. Thus, the two parties will observe random variables A and B whose joint distribution depends on the choice of die by the adversary. An application of this distributed case would be a key-agreement scenario under tampering.

Again consider the non-degenerate case where all faces on all the dice of the adversary have positive probability. We show that in this case, we can extract a common random bit from the distributed SV source if and only if it is possible to extract randomness from the common part of A and B. That is, similar to the i.i.d. case, the problem of common randomness extraction from distributed SV sources is reduced to the problem of randomness extraction from non-binary generalized SV sources. Since by our results, we know when randomness extraction from generalized SV sources is possible, we obtain a complete answer to the problem in the distributed case too.

In cases more general than non-degenerate cases we have the following: If C is the common data of A and B, then if there does not exist a nonzero real function of C which has zero expectation under all the different dice of the adversary, then common randomness extraction is impossible. This shows that the relation between the problem of common randomness extraction and the problem of randomness extraction from the common part holds also in some settings other than non-degenerate cases. For example, it resolves the problem of common randomness extraction from the following interesting distributed SV source.

Example. A concrete example of a distributed SV source is as follows. Let us start with the original source considered by Santha and Vazirani with two coins. Assume that the adversary chooses coin $S \in \{1, 2\}$ (where coin 1 is biased towards heads and coin 2 is biased towards tails) and let the outcome of the throw of the coin be denoted by random variable C. The first party, Alice, is assumed to observe both the identity of the coin chosen by the adversary, i.e., S, and the outcome of the coin, which is C. The second party, Bob, observes the outcome of the coin C, but only gets to see the choice of the adversary with probability 0.99. That is, Bob gets $B = (C, \tilde{S})$ where $\tilde{S}$ is the result of passing S through a binary erasure channel with erasure probability 0.01. Here the common part of $A = (C, S)$ and $B = (C, \tilde{S})$ is just C. Our result (Theorem 3) then implies that Alice and Bob cannot benefit from their knowledge of the actions of adversary, and should only consider the C sequence. But then from the result of [23], we can conclude that common random bit extraction is impossible in this example.

Proof Techniques. We briefly explain the techniques used in the proof of the above results. For the full proofs, we refer the reader to the full version of the paper [2].

To show the possibility of deterministic extraction, we use a nonzero real function of the die face values that has zero expectation under all distributions induced by the different dice of the adversary. Then as we throw the dice several times, we consider the sum of the value of this function applied to the outcome of the dice throws. This sum forms a martingale. We stop the martingale once its absolute value exceeds a particular bound. Since the function used was nonzero, the martingale has large variance after a few throws, and therefore the martingale will be stopped with high probability. Also by the theorem of stopping times, the martingale has zero mean whenever we stop it. Then the extracted bit, determined by whether the stopped martingale is positive or is negative, would be unbiased.

To show the impossibility of deterministic extraction, we view a deterministic extractor that extracts one bit from a generalized SV source as labeling the leaves of a rooted tree with zeros and ones. Each sequence of dice throws corresponds to a path from the root to one of the leaves, and at each node, the adversary has some limited control of which branch to take while moving from the root towards the leaves. We need to show that either the minimum or the maximum of the probability of the output bit being zero, over all adversary's strategies, is far from $1/2$. Our idea is to track these maximum and minimum probabilities in a recursive way, i.e., to find these probabilities for any node of the tree in terms of these values for its children. We then by induction show that for each node of the tree either the minimum probability or the maximum probability is far from $1/2$.

To be more precise, given a deterministic extractor, let α be the minimum probability of output bit being zero (over all strategies of the adversary). Similarly, let β be the maximum probability of output bit being zero (over all strategies of the adversary). Then we show that under certain conditions, there exists a *continuous* function $g(\cdot)$ on the interval $[0, 1]$, such that $\beta \geq g(\alpha)$ and furthermore $g(1/2) > 1/2$. We prove $\beta \geq g(\alpha)$ inductively using the tree structure discussed above. This implies the desired impossibility result, as by the continuity of $g(\cdot)$, both α and β cannot be close to $1/2$. For instance, for the binary SV source with two coins having probability of heads respectively equal to $1/3$ and $2/3$, Figure 1 shows a curve where (α, β) always lies above it. This curve is clearly isolated from $(1/2, 1/2)$.

We follow similar ideas for proving our impossibility result for common randomness extraction from a distributed SV source; again we construct a continuous function, which somehow captures not only the minimum and maximum of the probability of the extracted common bit being zero, but also the probability that the two parties agree on their extracted bits. The construction of this function is more involved in the distributed case; it has two terms one of which is similar to the function in the non-distributed case, and the other is inspired by the definition of maximal correlation mentioned above.

Contributions to Information Theory. As mentioned above, the problem of common randomness extraction from i.i.d. sources has been studied in the information theory community. Then our work provides a generalization and an alternative proof of known results in the i.i.d. case. In particular, we give a new proof of Witsenhausen's result [29] on the impossibility of common randomness extraction from certain i.i.d. sources.

We also would like to point out that a generalized SV source as we define, is indeed an arbitrarily varying source (AVS) [10,13] with a causal adversary. These sources are studied in the information theory literature from the point of view of source coding [4].

Notations. In this paper we consider functions $X : \mathcal{C} \to \mathbb{R}$. Such a function can be thought of as a random variable $X = X(C)$. We sometimes for simplicity use the notation $X(c) = x_c$. The expected value and variance of X are denoted by $\mathbb{E}[X]$ and $\mathrm{Var}[X]$ respectively.

We sometimes have several distributions over the same set $\mathcal{C}$ which are indexed by elements $s \in \mathcal{S}$. In this case to avoid confusions, the expectation value and variance are specified by a subscript s.

For simplicity of notation a sequence $C_1, \ldots, C_n$ of (not necessarily i.i.d.) random variables is denoted by C^n. Similarly for $c_1, \ldots, c_n \in \mathcal{C}$ we use $c^n = (c_1, \ldots, c_n)$. We also use the notation $c_{[k:k+\ell]} = (c_k, c_{k+1}, \ldots, c_{k+\ell})$.

2 Randomness Extraction from Generalized SV Sources

Definition 1 (Generalized SV source). *Let $\mathcal{C}$ be a finite alphabet set. Consider a* finite *set of distributions over $\mathcal{C}$ indexed by a set $\mathcal{S}$. That is, assume that for any $s \in \mathcal{S}$ we have a distribution over $\mathcal{C}$ determined by numbers $p_s(c)$ for all $c \in \mathcal{C}$. A sequence $C_1, C_2, \cdots$ of random variables, each over alphabet set $\mathcal{C}$, is said to be a* generalized SV source *with respect to distributions $p_s(c)$, if the sequence is generated as follows: Assume that $C_1, \ldots, C_{i-1}$ are already generated. In order to determine C_i, an adversary chooses $S_i = s_i \in \mathcal{S}$, depending only on $C_1, \ldots, C_{i-1}$. Then C_i is sampled from the distribution $p_{s_i}(c)$.*

We can think of specifying s as choosing a particular multi-faceted die, and c as the facet that results from throwing the die. The joint probability distribution $p(c_1, c_2, \cdots, c_n, s_1, s_2, \cdots, s_n)$ of random variables $C_1, \ldots, C_n$ and $S_1, \ldots, S_n$ in a generalized SV source factorizes as follows:

$$q(s_1)p_{s_1}(c_1)q(s_2|c_1)p_{s_2}(c_2)\cdots q(s_n|c_1\cdots c_{n-1})p_{s_n}(c_n),$$

where $q(s_i|c_1\cdots c_{i-1})$ describes the action of the adversary at time i. Here, first the adversary chooses $S_1 = s_1$ with probability $q(s_1)$, and then $C_1 = c_1$ is generated with probability $p_{s_1}(c_1)$. Then the adversary chooses $S_2 = s_2$ with probability $q(s_2|c_1)$ and then $C_2 = c_2$ is generated with probability $p_{s_2}(c_2)$, and so on.

Generalized SV sources can be alternatively characterized as follows: Given i and $C_1 = c_1, \ldots, C_{i-1} = c_{i-1}$, the distribution of C_i should be a convex combination of the set of $|\mathcal{S}|$ distributions $\{p_s(\cdot) : s \in \mathcal{S}\}$.

We emphasize that even after fixing distributions $p_s(c)$, the generalized SV source (similar to ordinary SV sources) is not a fixed source, but rather a class of sources. This is because in each step s_i is chosen arbitrarily by the adversary as a (probabilistic) function of $C_1, \ldots, C_{i-1}$. Nevertheless, once we fix adversary's strategy, the generalized SV source is fixed in that class of sources.

Definition 2 (Deterministic extraction). *We say that deterministic randomness extraction from the generalized SV source determined by distributions $p_s(c)$ is possible if for every $\epsilon > 0$ there exist n and $\Gamma_n : \mathcal{C}^n \to \{0,1\}$ such that for every strategy of the adversary, the distribution of $\Gamma_n(C^n)$ is ϵ-close, in total variation distance, to the uniform distribution. That is, independent of adversary's strategy, $\Gamma_n(C^n)$ is an almost uniform bit.*

In the following we present a necessary condition and separately a sufficient condition for the existence of deterministic extractors for generalized SV sources. In the non-degenerate case, i.e., when $p_s(c) > 0$ for all s, c, these two conditions coincide. Thus we fully characterize the possibility of deterministic randomness extraction from generalized SV sources in the non-degenerate case.

2.1 A Sufficient Condition for the Existence of Randomness Extractors

Theorem 1. *Consider a generalized SV source with alphabet $\mathcal{C}$, set of dice $\mathcal{S}$, and probability distributions $p_s(c)$. Suppose that there exists $\psi : \mathcal{C} \to \mathbb{R}$ such that for every $s \in \mathcal{S}$ we have $\mathbb{E}_{(s)}[\psi(C)] = 0$ and $\mathrm{Var}_{(s)}[\psi(C)] > 0$, where $\mathbb{E}_{(s)}$ and $\mathrm{Var}_{(s)}$ are expectation and variance with respect to the distribution $p_s(\cdot)$. Then randomness can be extracted from this SV source.*

Observe that if $p_s(c) > 0$ for all s, c, then this theorem can equivalently be stated as follows: Thinking of each distribution $p_s(\cdot)$ as a point in the probability simplex, if the convex hull of the set of points $\{p_s(\cdot) : s \in \mathcal{S}\}$ in the probability simplex does not have full dimension, then deterministic randomness extraction is possible. For instance if $|\mathcal{S}| < |\mathcal{C}|$ this condition is always satisfied and then we can deterministically extract randomness.

Remark 1. The analysis of the proof of Theorem 1 would show that the bias could be polynomially small, namely a bias of $\Theta(n^{-1/3})$.

2.2 A Necessary Condition for the Existence of Randomness Extractors

The main result of this subsection is the following theorem.

Theorem 2. *Consider a generalized SV source with alphabet $\mathcal{C}$, set of dice $\mathcal{S}$, and probabilities $p_s(c)$. Suppose that there is no non-zero function $\psi : \mathcal{C} \to \mathbb{R}$*

such that for all $s \in \mathcal{S}$ we have $\mathbb{E}_{(s)}[\psi(C)] = 0$. Then deterministic randomness extraction from this generalized SV source is impossible.

Again, let us consider the case where $p_s(c) > 0$ for all s, c. In this case ψ being non-zero is equivalent to $\mathrm{Var}_{(s)}[\psi] > 0$ for all s. Then comparing to Theorem 1 we find that the necessary and sufficient condition for the possibility of deterministic extraction is the existence of a non-zero ψ with $\mathbb{E}_{(s)}[\psi] = 0$.

Corollary 1. *Consider a generalized SV source with alphabet $\mathcal{C}$, set of dice $\mathcal{S}$, and probabilities $p_s(c)$. Let $\mathcal{S}'$ be a subset of $\mathcal{S}$ and let $\mathcal{C}'$ be the set of all c for which there exists some $s \in \mathcal{S}'$ such that $p_{s'}(c) > 0$. Suppose that there is no non-zero function $\psi : \mathcal{C} \to \mathbb{R}$ such that (i) ψ is zero on $\mathcal{C} - \mathcal{C}'$, and (ii) for all $s \in \mathcal{S}'$ we have $\mathbb{E}_{(s)}[\psi(C)] = 0$. Then deterministic randomness extraction from this generalized SV source is impossible.*

3 Distributed SV Sources

Distributed SV sources can be defined similarly to generalized SV sources except that in this case, the outcome in each time step is a pair that is distributed between two parties.

Definition 3. *Fix finite sets $\mathcal{A}, \mathcal{B}, \mathcal{S}$. Let $p_s(ab)$ define a probability distribution over $\mathcal{A} \times \mathcal{B}$ for any $s \in \mathcal{S}$. The distributed SV source with respect to distributions $p_s(ab)$ is defined as follows. The adversary in each time step i, depending on the previous outcomes $(A_1, B_1) = (a_1, b_1), \ldots, (A_{i-1}, B_{i-1}) = (a_{i-1}, b_{i-1})$ chooses some $S_i = s_i$. Then $(A_i, B_i) = (a_i, b_i)$ is sampled from the distribution $p_{s_i}(a_i b_i)$. The sequence of random variables $(A_1, B_1), (A_2, B_2), \ldots$, is called a distributed SV source.*

Here we assume that the outcomes of this SV source are distributed between two parties, say Alice and Bob. That is, in each time step i, A_i is revealed to Alice and B_i is revealed to Bob. So Alice receives the sequence $A_1, A_2, \ldots$, and Bob receive the sequence $B_1, B_2, \ldots$.

In this section we are interested in whether two parties can generate a common random bit from distributed SV sources. To be more precise, let us first define the problem more formally.

Definition 4. *We say that common randomness can be extracted from the distributed SV source $(A_1, B_1), (A_2, B_2), \ldots$ if for every $\epsilon > 0$ there is n and functions $\Gamma_n : \mathcal{A}^n \to \{0, 1\}$ and $\Lambda_n : \mathcal{B}^n \to \{0, 1\}$ such that for every strategy of adversary, the distributions of $K_1 = \Gamma_n(A^n)$ and $K_2 = \Lambda_n(B^n)$ are ϵ-close (in total variation distance) to uniform distribution, and that $\Pr[K_1 \neq K_2] < \epsilon$.*

In the above definition we considered only deterministic protocols for extracting a common random bit. We could also consider probabilistic protocols where Γ_n and Λ_n are random functions depending on *private* randomnesses of Alice and Bob respectively. More precisely, we could take $K_1 = \Gamma_n(A^n, R_1)$ and

$K_2 = \Lambda_n(B^n, R_2)$ with the above conditions on K_1, K_2, where R_1 and R_2 are private randomnesses of Alice and Bob respectively, which are independent of the SV source and of each other. Nevertheless, if a common random bit can be extracted with probabilistic protocols, then common randomness extraction with deterministic protocols is also possible.

Lemma 1. *In the problem of common random bit extraction, with no loss of generality we may assume that the parties do not have private randomness.*

3.1 Common Data

As discussed in the introduction, the notion of the common data of two random variables A, B first appeared in the problem of common randomness extraction from i.i.d. sources. Briefly speaking, common data of A and B is the finest random variable C that can be computed both as a function $C = C_1(A)$ of A, and as a function $C = C_2(B)$ of B. In the full version of this paper [2], we give a new proof of Witsenhausen's theorem that randomness extraction from i.i.d. repetitions of (A, B) is feasible if and only if common data exists, if and only if maximal correlation is equal to 1.

Here we are interested in common randomness extraction from distributed SV sources. So we need to define common data for such sources. The common data of a distributed SV source (given by distributions $p_s(ab)$ indexed by $s \in \mathcal{S}$) is the finest random variable C that can be computed both as a function $C = C_1(A)$ of A, and as a function $C = C_2(B)$ of B. Here we need $C_1(A) = C_2(B)$ to hold with probability 1 under all distributions $p_s(ab)$.

3.2 Common Random Bit Extraction from Distributed SV Sources

Theorem 3. *Consider a distributed SV source (as in Definition 3) with corresponding sets $\mathcal{S}$, $\mathcal{A}$, and $\mathcal{B}$ and corresponding distributions $p_s(ab)$. Let C be the common data of the distributed SV source. Let $p_s(abc)$ denote the induced joint distribution of A, B, and C. Suppose that there is no non-zero function $\psi : \mathcal{C} \to \mathbb{R}$ such that $\mathbb{E}_{(s)}[\psi(C)] = 0$ for all s. Then common randomness cannot be extracted from this distributed SV source.*

An algorithm to extract common random bits is to focus on the common part C that can be computed by both Alice and Bob. Indeed C itself can be thought of as a generalized SV source. If deterministic randomness extraction from C is possible, then Alice and Bob can obtain a common random bit by individually applying the randomness extraction protocol. Comparing with Theorems 1 and 2, and assuming $p_s(c) > 0$ for all s, c, the above theorem states that a common random bit can be extracted if and only if deterministic randomness extraction from C is possible.

4 Future Work

In this paper we completely characterized the randomness extraction problem for non-degenerate cases. A future work could be to solve this problem for the degenerate cases. In the degenerate cases, for generalized non-distributed sources Corollary 1 gives a mildly stronger necessary condition than Theorem 2, but there is still a gap between this necessary condition and the sufficient condition of Theorem 1.

We note that our randomness extractor in Theorem 1 extracts a bit whose bias is inverse polynomially small in the length of the source sequence. It is interesting to see if this extractor could be improved to yield a bit with an exponentially small bias. Furthermore, if we want to produce more than one bit of randomness, the tradeoff between the number of produced random bits and their quality is open.

Another interesting problem is to look at efficient adversaries, similar to the work of [1]. Our proofs only show existence of inefficient adversaries.

Another way to restrict the adversary is to put limitations on the number of times the adversary can choose a strategy $s \in \mathcal{S}$, i.e. there can be a cost associated to each strategy s.

A different type of limitation can be on the adversary's knowledge about the sequence generated so far. More specifically, the adversary might have *noisy or partial* access to the previous outcomes in the sequence (these sources are called "active sources" [22]). These sources model adversaries with limited memory. Space bounded sources have been studied in [17,24].

Finally, the problem of common randomness extraction can be studied for three or more parties instead of just two parties.

References

1. Austrin, P., Chung, K.-M., Mahmoody, M., Pass, R., Seth, K.: On the impossibility of cryptography with tamperable randomness. In: Garay, J.A., Gennaro, R. (eds.) CRYPTO 2014, Part I. LNCS, vol. 8616, pp. 462–479. Springer, Heidelberg (2014)
2. Beigi, S., Etesami, O., Gohari, A.: Deterministic Randomness Extraction from Generalized and Distributed Santha-Vazirani Sources (2014). arXiv:1412.6641
3. Beigi, S., Tse, D.: under preparation
4. Berger, T.: The source coding game. IEEE Trans. on Information Theory **IT–17**(1), 71–76 (1971)
5. Blum, M.: Independent unbiased coin flips from a correlated biased source - a finite state Markov chain. Combinatorica **6**(2), 97–108 (1986)
6. Bourgain, J.: More on the sum-product phenomenon in prime fields and its applications. International Journal of Number Theory (2005)
7. Bourgain, J.: On the construction of affine extractors. Geometric And Functional Analysis **17**(1), 33–57 (2007)
8. Chor, B., Goldreich, O., Håstad, J., Freidmann, J., Rudich, S., Smolensky, R.: The bit extraction problem of t-resilient functions. In: Proceedings of the 26th Annual Symposium on Foundations of Computer Science, pp. 396–407 (1985)

9. Chor, B., Goldreich, O.: Unbiased Bits from Sources of Weak Randomness and Probabilistic Communication Complexity. SIAM J. Comput. **17**(2), 230–261 (1988)
10. Dobrusin, R.L.: Individual methods for transmission of information for discrete channels without memory and messages with independent components. Sov. Math. **4**, 253–256 (1963)
11. Dvir, Z.: Extractors for varieties. Computational Complexity **21**(4), 515–572 (2012)
12. Dvir, Z., Gabizon, A., Wigderson, A.: Extractors and rank extractors for polynomial sources. In: FOCS 2007: Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science, pp. 52–62 (2007)
13. Dobrusin, R.L.: Unified methods of optimal quantizing of messages. Sov. Math. **4**, 284–292 (1963)
14. Fischer, M.J., Lynch, N.A.: A lower bound for the time to assure interactive consistency. Information Processing Letters **14**, 183–186 (1982)
15. Gács, P., Körner, J.: Common information is far less than mutual information. Problems of Control and Information Theory **2**(2), 119–162 (1972)
16. Gabizon, A., Raz, R.: Deterministic extractors for affine sources over large fields. In: Proceedings of the 46th FOCS, pp. 407–418 (2005)
17. Kamp, J., Rao, A., Vadhan, S., Zuckerman, D.: Deterministic extractors for small-space sources. In: Proceedings of the thirty-eighth annual ACM symposium on Theory of computing, pp. 691–700 (2006)
18. Nisan, N., Zuckerman, D.: Randomness is Linear in Space. Journal of Computer and System Sciences **52**(1), 43–52 (1996)
19. Rabin, M.O.: Randomized byzantine generals. In: Proceedings of the 24th Annual Symposium on Foundations of Computer Science, pp. 403–409 (1983)
20. Rao, A.: Extractors for a constant number of polynomially small min-entropy independent sources. In: Proceedings of the 38th STOC, pp. 497–506 (2006)
21. Reingold, O., Vadhan, S., Wigderson, A.: A note on extracting randomness from Santha-Vazirani sources. Unpublished manuscript (2004)
22. Palaiyanur, H., Chang, C., Sahai, A.: Lossy compression of active sources. In: IEEE International Symposium on Information Theory, pp. 1977–1981 (2008)
23. Santha, M., Vazirani, U.: Generating quasi-random sequences from slightly-random sources. In: Proceedings of Symposium on the Foundations of Computer Science (1984). Jourdnal of Computer and System Sciences, **33**(1), 75–87 (1986)
24. Vazirani, U.V.: Efficiency considerations in using semi-random sources. In: Proceedings of the Nineteenth STOC, pp. 160–168 (1987)
25. Vazirani, U.V., Vazirani, V.V.: Random polynomial time is equal to slightly-random polynomial time. In: Proc. 26th Annual IEEE Symposium on the Foundations of Computer Science, pp. 417–428 (1985)
26. Vazirani, U.V., Vazirani, V.V.: Sampling a population with a single semi-random source. In: Proc. 6th FST & TCS Conf. (1986)
27. Vadhan, S.: Pseudorandomness. Now Publishers (2012)
28. von Neumann, J.: Various techniques used in connection with random digits. Applied Math Series **12**, 36–38 (1951)
29. Witsenhausen, H.S.: On sequences of pairs of dependent random variables. SIAM Journal on Applied Mathematics **28**(1), 100–113 (1975)
30. Yao, A.C.: Some Complexity Questions Related to Distributed Computing. In: Proc. of 11th STOC, vol. 14, pp. 209–213 (1979)
31. Zuckerman, D.: Randomness-optimal oblivious sampling. Random Structures and Algorithms **11**, 345–367 (1997)