Analyzing Linear Mergers

Zeev Dvir* Ran Raz†

Abstract

Mergers are functions that transform k (possibly dependent) random sources into a single random source, in a way that ensures that if one of the input sources has min-entropy rate δ then the output has min-entropy rate close to δ. Mergers have proven to be a very useful tool in explicit constructions of *extractors* and *condensers*, and are also interesting objects in their own right. In this work we give a refined analysis of the merger constructed by [Raz05] (based on [LRVW03]). Our analysis uses min-entropy instead of Shannon's entropy to derive tighter results than the ones obtained in [Raz05].

We show that for every r it is possible to construct a merger that takes as input k strings of length n bits each, and outputs a string of length n/r bits, such that if one of the input sources has min-entropy b, the output will be close to having min-entropy $b/(r+1)$. This merger uses a constant number of additional uniform bits when k and r are constants. One advantage of our analysis is that b (the min-entropy of the 'good' source) can be as small as a constant, while in the analysis given in [Raz05], b is required to be linear in n.

1 Introduction

Consider the following problem: You are given k samples $x_1, \ldots, x_k \in \{0,1\}^n$, taken from k, possibly dependent, n-bit random sources $X_1, \ldots, X_k$. Suppose that one of these k sources (the index of which is not known) has min-entropy[1] $\geq \delta n$. You want to output a string of length n' bits, computed from these k samples and from an additional short string which is uniformly distributed, in a way that will ensure that the min-entropy of your output is at least $\delta' n'$, where δ' is not considerably smaller than δ. Mergers are functions that attempt to solve this problem.

The notion of *merger* was first introduced by Ta-Shma [TS96], in the context of explicit

*Department of Computer Science, Weizmann institute of science, Rehovot, Israel.
Email: zeev.dvir@weizmann.ac.il. Research supported by Israel Science Foundation (ISF) grant.

†Department of Computer Science, Weizmann institute of science, Rehovot, Israel.
Email: ran.raz@weizmann.ac.il. Research supported by Israel Science Foundation (ISF) grant.

[1] A source has min-entropy $\geq b$ if none of its values is obtained with probability larger than 2^{-b}.

constructions of *extractors*[2]. Recently, Lu, Reingold, Vadhan and Wigderson [LRVW03] gave a very simple and beautiful construction of mergers based on Locally-Decodable-Codes. This construction was used in [LRVW03] as a building block in an explicit construction of extractors with nearly optimal parameters. More recently, [Raz05] generalized the construction of [LRVW03], and showed how this construction (when combined with other techniques) can be used to construct *condensers*[3] with constant seed length.

The merger constructed by [LRVW03] takes as input k strings of length n, one of which has min-entropy b, and outputs a string of length n that is close to having min-entropy at least $b/2$. Loosely speaking, the output of the merger is computed as follows: treat each input block as a vector in the vector space F^m, where F is some small finite field, and output a uniformly chosen linear combination of these k vectors. One drawback of this construction is that the min-entropy rate (i.e. the ratio between the min-entropy of a source and its length) of the output is at most $\frac{1}{2}$, even if one of the input blocks is completely uniform. This might be a drawback in cases where the output of the merger is used as an input to some other 'device' that requires the min-entropy rate of its input to be larger than $\frac{1}{2}$.

The construction of [Raz05], which generalizes the construction of [LRVW03], overcomes this problem by slightly modifying the above construction. Instead of treating each input block as one vector in F^m, we treat each input block as r concatenated vectors in the vector space F^l (where $l = m/r$). We then output a randomly chosen linear combination of the $r \cdot k$ vectors obtained from all k sources. An extension of the analysis given in [LRVW03] shows that if one of the k sources has min-entropy b, then the output of the merger is close to having min-entropy at least $b/(r+1)$. Since the length of the output is n/r, the min-entropy rate of the output is $\frac{r}{r+1} \cdot \delta$, where $\delta = \frac{b}{n}$ is the min-entropy rate of the 'good' input block. Thus, the min-entropy rate of the output can approach 1, if one of the inputs is uniform (or close to uniform).

The analysis given by [Raz05] uses Shannon's entropy to derive its results. That is, it shows that the *entropy* of the output is high, and uses this fact to lower bound the min-entropy of the output. In this paper we give an improved analysis of the construction of [Raz05] that directly shows that the min-entropy of the output is high, without using Shannon's entropy. This analysis possesses several advantages over that of [Raz05], the most significant of which is that it shows that the merger works even when the min-entropy of the input is very small (a constant), where the analysis of [Raz05] requires the min-entropy of the input to be linear in the length of the input source. Although this improved analysis doesn't give any qualitative improvements to the end results of [Raz05] (namely, to the construction of condensers), we feel that future applications of these mergers might benefit from our results.

[2]An extractor is a function that transforms a source with min-entropy b into a source which is close to uniform, with the aid of an additional random seed. For a more detailed description of extractors see [LRVW03].

[3]A condenser is a function that transforms a source with min-entropy rate δ into a source which is close to having min-entropy rate $\delta' > \delta$, with the aid of an additional random seed.

1.1 Somewhere-Random-Sources

An *n-bit random source* is a random variable X that takes values in $\{0,1\}^n$. We denote by $\text{supp}(X) \subset \{0,1\}^n$ the support of X (i.e. the set of values on which X has non-zero probability). For two n-bit sources X and Y, we define the statistical distance (or simply distance) between X and Y to be

$$\Delta(X,Y) \triangleq \frac{1}{2} \sum_{a \in \{0,1\}^n} \left| \mathbf{Pr}[X=a] - \mathbf{Pr}[Y=a] \right|.$$

We say that a random source X (of length n bits) has min-entropy $\geq b$ if for every $x \in \{0,1\}^n$ the probability for $X = x$ is at most 2^{-b}.

Definition 1.1 (Min-entropy). *Let X be a random variable distributed over $\{0,1\}^n$. The min-entropy of X is defined as* [4]

$$H^{\infty}(X) \triangleq \min_{x \in supp(X)} \log\left(\frac{1}{\mathbf{Pr}[X=x]}\right).$$

Definition 1.2 ((n,b)-Source). *We say that X is an (n,b)-source, if X is an n-bit random source, and $H^{\infty}(X) \geq b$.*

A *somewhere-(n,b)-source* is a source comprised of several blocks, such that at least one of the blocks is an (n,b)-source. Note that we allow the other source blocks to depend arbitrarily on the (n,b)-source, and on each other.

Definition 1.3 ($(n,b)^{1:k}$-Source)). *A k-places-somewhere-(n,b)-source, or shortly, an $(n,b)^{1:k}$-source, is a random variable $X = (X_1, \ldots, X_k)$, such that every X_i is of length n bits, and at least one X_i is of min-entropy $\geq b$. Note that $X_1, \ldots, X_k$ are not necessarily independent.*

1.2 Mergers

A merger is a function transforming an $(n,b)^{1:k}$-source into a source which is γ-close (i.e. it has statistical distance $\leq \gamma$) to an (m,b')-source. Naturally, we want b'/m to be as large as possible, and γ to be as small as possible. We allow the merger to use an additional small number of truly random bits, called a *seed*. A Merger is *strong* if for almost all possible assignments to the seed, the output is close to be an (m,b')-source. A merger is *explicit* if it can be computed in polynomial time.

Definition 1.4 (Merger). *A function $M : \{0,1\}^d \times \{0,1\}^{n \cdot k} \to \{0,1\}^m$ is a $[d, (n,b)^{1:k} \mapsto (m,b') \sim \gamma]$-merger if for every $(n,b)^{1:k}$-source X, and for an independent random variable Z uniformly distributed over $\{0,1\}^d$, the distribution $M(Z,X)$ is γ-close to a distribution of an (m,b')-source. We say that M is **strong** if the average over $z \in \{0,1\}^d$ of the minimal distance between the distribution of $M(z,X)$ and a distribution of an (m,b')-source is $\leq \gamma$.*

[4] All logarithms in this paper are taken base 2.

1.3 Our Results

The main result of this paper is the following theorem, which shows the existence of explicit strong mergers. Notice that the size of the seed used by the merger does not depend on the length of each input block.

Theorem 1 (Strong Merger). *For any constants $\alpha, \gamma > 0$, and for every integers $k, r \geq 1$, there exists a constant $b_0 > 0$, such that for every $b \geq b_0$, and for every $n \geq b$, there exists an explicit $[d, (n, b)^{1:k} \mapsto (m, b') \sim \gamma]$-strong merger, such that,*

$$m = \left\lceil \frac{n}{r} \right\rceil,$$

$$b' = \frac{b}{r + 1 + \alpha},$$

$$d = k \cdot r \cdot \left\lceil \log\left(\frac{2r}{\gamma}\right) \right\rceil.$$

1.4 Organization

In Section 2 we describe the construction of mergers of [Raz05] (based on [LRVW03]). We then proceed, in Section 3, to give our improved analysis for these mergers, and to prove Theorem 1. In our analysis we will mostly follow the notations of [Raz05].

2 The Construction

In this section we describe the construction of mergers of [Raz05], which is based on the construction of [LRVW03]. Loosely speaking, the construction can be described as follows: Given k input blocks of length n bits each, we pick integers r, p such that $n = p \cdot r \cdot l$, and treat each input block as r vectors in F^l, where F is a field of size 2^p. The output of the merger is then a uniformly chosen linear combination of these $k \cdot r$ vectors (r vectors in each block) in the vector space F^l. We now describe the construction more formally.

Construction 2.1. *Let n, k, r, p be integers such that $r \cdot p$ divides n, and let $l = \frac{n}{r \cdot p}$. We define a function*

$$M : \{0,1\}^d \times \{0,1\}^{n \cdot k} \to \{0,1\}^{\frac{n}{r}},$$

with

$$d = p \cdot k \cdot r,$$

in the following way: Let F be a finite filed of size 2^p. Given $z \in \{0,1\}^d$, we think of z as a vector $(z_{1,1}, \ldots, z_{k,r}) \in F^{k \cdot r}$. Given $x = (x_1, \ldots, x_k) \in \{0,1\}^{n \cdot k}$, we think of each $x_i \in \{0,1\}^n$ as a vector $(x_{i,1}, \ldots, x_{i,r})$, where each $x_{i,j}$ is in $\{0,1\}^{l \cdot p}$. We think of each $x_{i,j} \in \{0,1\}^{l \cdot p}$ as a

vector in F^l. More generally, we think of $\{0,1\}^{l \cdot p}$ as the vector space F^l. The function M is now defined as

$$M(z,x) = \sum_{i=1}^{k} \sum_{j=1}^{r} z_{i,j} \cdot x_{i,j} \in F^l,$$

where the operations are in the vector space F^l. Intuitively, one can think of M as

$$M : F^{k \cdot r} \times \left(F^l\right)^{k \cdot r} \to F^l.$$

In the next section we will show that this construction gives mergers as in Theorem 1, for an appropriate choice of r and p. One technicality is that we require n to be divisible by $r \cdot p$. This technicality can be addressed by padding each input block with at most $r \cdot p$ zeros to obtain the required relation between n, r and p.

3 The Analysis

In this section we present our analysis of the mergers defined in the last section, and in particular prove Theorem 1. We begin with some notations that will be used throughout this section.

Let $X = (X_1, \ldots, X_k) \in \{0,1\}^{n \cdot k}$ be a somewhere (n,b)-source, and let us assume w.l.o.g. that $\mathrm{H}^{\infty}(X_1) \geq b$. Suppose that $n = p \cdot r \cdot l$, and let $M(z,x) : \{0,1\}^d \times \{0,1\}^{n \cdot k} \to \{0,1\}^{\frac{n}{r}}$ be as in Construction 2.1, where $d = p \cdot k \cdot r$. For every $z \in \{0,1\}^d$ we denote by $Y_z \triangleq M(z,X)$ the random variable given by the output of M on the fixed seed value z (recall that, in Construction 2.1, every seed value corresponds to a specific linear combination of the source blocks). Let $u \triangleq 2^d = 2^{pkr}$ be the number of different seed values, so we can treat the set $\{0,1\}^d$ as the set[5] $[u]$. We can now define $Y \triangleq (Y_1, \ldots, Y_u) \in (\{0,1\}^{p \cdot l})^u$. The random variable Y is a function of X, and is comprised of u blocks, each one of length $p \cdot l$, representing the output of the merger on all possible seed values. We will first analyze the distribution of Y as a whole, and then use this analysis to describe the output of M on a uniformly chosen seed.

Definition 3.1. *Let $D(\Omega)$ denote the set of all probability distributions over a finite set Ω. Let $\mathcal{P} \subset D(\Omega)$ be some property. We say that $\mu \in D(\Omega)$ is γ-close to a convex combination of distributions with property $\mathcal{P}$, if there exists constants $\alpha_1, \ldots, \alpha_t, \gamma > 0$, and distributions $\mu_1, \ldots, \mu_t, \mu' \in D(\Omega)$ such that the following three conditions hold [6]:*

1. $\mu = \sum_{i=1}^{t} \alpha_i \mu_i + \gamma \mu'$.
2. $\sum_{i=1}^{t} \alpha_i + \gamma = 1$.

[5] For an integer n, we write $[n] \triangleq \{1, 2, \ldots, n\}$.

[6] In condition 1, we require that the convex combination of the μ_i's will be strictly smaller than μ. This is not the most general case, but it will be convenient for us to use this definition.

3. $\forall i \in [t] \quad , \quad \mu_i \in \mathcal{P}$.

Let Y be the random variable defined above, and let $\mu : (\{0,1\}^{p \cdot l})^u \to [0,1]$ be the probability distribution of Y (i.e. $\mu(y) = \mathbf{Pr}[Y = y]$). We would like to show that μ is exponentially (in b) close to a convex combination of distributions, each having a certain property which will be defined shortly.

Given a probability distribution μ on $(\{0,1\}^{p \cdot l})^u$ we define for each $z \in [u]$ the distribution $\mu_z : \{0,1\}^{p \cdot l} \to [0,1]$ to be the restriction of μ to the z's block. More formally, we define

$$\mu_z(y) \triangleq \sum_{y_1,\ldots,y_{z-1},y_{z+1},\ldots,y_u \in \{0,1\}^{p \cdot l}} \mu(y_1, \ldots, y_{z-1}, y, y_{z+1}, \ldots, y_u).$$

Let

$$\epsilon \triangleq r \cdot 2^{-p},$$

and let $\alpha > 0$. We say that a distribution $\mu : (\{0,1\}^{p \cdot l})^u \to [0,1]$ is α-good if for at least $(1-\epsilon) \cdot u$ values of $z \in [u]$, μ_z has min-entropy at least $\frac{b}{r+1+\alpha}$. The statement that we would like to prove is that the distribution of Y is close to a convex combination of α-good distributions (see Definition 3.1). As we will see later, this is good enough for us to be able to prove Theorem 1. The following lemma states this claim in a more precise form.

Lemma 3.2 (Main Lemma). *Let $Y = (Y_1, \ldots, Y_u)$ be the random variable defined above, and let μ be its probability distribution. Then, for any constant $\alpha > 0$, μ is $2^{-\Omega(b)}$-close to a convex combination of α-good distributions.*

It is worth noting that Lemma 3.2 gives a stronger result than the one stated in Theorem 1. From Lemma 3.2 we see that the output of the merger contains two kinds of error. One is given by ϵ, and denotes the fraction of 'bad' seeds in every α-good distribution that appears in the convex combination. The second error parameter is exponentially small ($2^{-\Omega(b)}$), and denotes the distance of the output from the convex combination of α-good distributions. This distinction does not appear in Theorem 1, and might be useful in constructions that use this merger as a building block.

We prove Lemma 3.2 in subsection 3.2. The proof of Theorem 1, which follows quite easily from Lemma 3.2, appears in the next subsection.

3.1 Proof of Theorem 1

Fix the constants α, γ, k, r as in the theorem. We choose p to be the smallest integer such that $\epsilon = r \cdot 2^{-p} \leq \frac{1}{2}\gamma$. More precisely, we set $p = \left\lceil \log\left(\frac{2r}{\gamma}\right) \right\rceil$. For every n we let $M : \{0,1\}^d \times \{0,1\}^{n \cdot k} \to \{0,1\}^{\frac{n}{r}}$, be as in Construction 2.1, where $d = p \cdot k \cdot r$, and assume for simplicity that n is divisible by $r \cdot p$, (otherwise we pad each input block with at most $r \cdot p$

zeros, and so the output length will be $\left\lceil \frac{n}{r} \right\rceil$, as required by Theorem 1). Let

$$b' = \frac{b}{r+1+\alpha},$$

$$m = \frac{n}{r},$$

We will show that for values of b larger than some constant b_0, and for values of n larger than b, M is a $[d, (n,b)^{1:k} \mapsto (m, b') \sim \gamma]$-strong merger.

Let $X = (X_1, \ldots, X_k)$ be a somewhere (n,b)-source, and w.l.o.g. assume that X_1 is an (n,b)-source. Let Z be a random variable uniformly distributed over $[u] = \{0,1\}^d$ (as before, we let $u = 2^{pkr} = 2^d$ denote the number of different seed values), and let $Y = (Y_1, \ldots, Y_u)$ and μ be as in Lemma 3.2. Using Lemma 3.2 we can write μ as a convex combination of distributions

$$\mu = \sum_{i=1}^{t} \alpha_i \mu_i + \gamma' \mu', \tag{1}$$

with $\gamma' = 2^{-\Omega(b)}$, and such that for every $i \in [t]$ the distribution μ_i is α-good. That is, for at least $(1-\epsilon) \cdot u$ values of $z \in [u]$, the distribution[7] $(\mu_i)_z$ has min-entropy at least $b' = \frac{b}{r+1+\alpha}$. Next, define for every $z \in [u]$ the set $H_z \subset [t]$ as follows:

$$H_z \triangleq \{i \in [t] \,:\, \mathrm{H}^{\infty}\left((\mu_i)_z \right) < b'\}.$$

That is, $H_z \subset [t]$ is the set of indices of all distributions among $\{\mu_1, \ldots, \mu_t\}$, for which $(\mu_i)_z$ has min-entropy smaller than b'. Additionally, define for every $z \in [u]$,

$$e_z \triangleq \sum_{i \in H_z} \alpha_i.$$

Claim 3.3. *Let $\Delta(Y_z, (m, b'))$ denote the minimal (statistical) distance between Y_z and an (m, b')-source. Then for every $z \in [u]$*

$$\Delta(Y_z, (m, b')) \leq e_z + \gamma'.$$

Proof. For every $z \in [u]$ let $\mu_z(y) = \mathbf{Pr}[Y_z = y]$ be the probability distribution of Y_z. From Eq.1 we can write μ_z as a convex combination

$$\begin{aligned}
\mu_z &= \sum_{i=1}^{t} \alpha_i \cdot (\mu_i)_z + \gamma' \mu'_z \\
&= \left(\sum_{i \notin H_z} \alpha_i \cdot (\mu_i)_z \right) + \left(\sum_{i \in H_z} \alpha_i \cdot (\mu_i)_z + \gamma' \mu'_z \right) \\
&= (1 - e_z - \gamma') \cdot \mu'' + (e_z + \gamma') \cdot \mu''',
\end{aligned}$$

[7] When writing $(\mu_i)_z$, the first subscript i denotes the index of the distribution, and the second subscript z denotes the restriction of this distribution to the block indexed by z.

where μ'' is the probability distribution of an (m, b') source[8], and μ''' is some other distribution. Clearly, the statistical distance $\Delta(\mu_z, \mu'')$ is at most $e_z + \gamma'$, and since μ'' is an (m, b') source, we have that $\Delta(Y_z, (m, b')) \leq e_z + \gamma'$. □

Claim 3.4. *Let Z be a random variable uniformly distributed over $[u]$. Then, the expectation of e_Z is at most ϵ:*

$$\mathbb{E}[e_Z] \leq \epsilon.$$

Proof. For each $i \in [t]$ define the following indicator random variable

$$\chi_i = \begin{cases} 1, & i \in H_Z; \\ 0, & i \notin H_Z. \end{cases}$$

We can thus write

$$e_Z = \sum_{i=1}^{t} \chi_i \cdot \alpha_i.$$

By linearity of expectation we have

$$\mathbb{E}[e_Z] = \sum_{i=1}^{t} \mathbb{E}[\chi_i] \cdot \alpha_i,$$

and since for each $i \in [t]$ we have that

$$\mathbb{E}[\chi_i] = \mathbf{Pr}_Z[i \in H_Z] < \epsilon$$

(this follows from the fact that each μ_i is α-good), we conclude that

$$\mathbb{E}[e_Z] \leq \epsilon \cdot \sum_{i=1}^{t} \alpha_i \leq \epsilon.$$

□

Combining Claim 3.3 and Claim 3.4, and recalling that $\epsilon \leq \frac{1}{2}\gamma$, and $\gamma' = 2^{-\Omega(b)}$, we see that

$$\mathbb{E}[\Delta(Y_Z, (m, b'))] \leq \mathbb{E}[e_Z] + \gamma' \leq \epsilon + \gamma' \leq \frac{1}{2}\gamma + 2^{-\Omega(b)},$$

where the expectations are taken over Z, which is chosen uniformly in $[u]$. Now, for values of b larger than some constant b_0, this expression is smaller than γ. This completes the proof of Theorem 1. □

[8] A convex combination of (m, b')-sources is an (m, b')-source.

3.2 Proof of Lemma 3.2

In order to prove Lemma 3.2 we prove the following slightly stronger lemma.

Lemma 3.5. *Let $X = (X_1, \ldots, X_k)$ be an $(n, b)^{1:k}$-source, and let $Y = (Y_1, \ldots, Y_u)$ and μ be as in Lemma 3.2. Then for any constant $\alpha > 0$ there exists an integer $t \geq 1$, and a partition of $\{0,1\}^{n \cdot k}$ into $t+1$ sets $W_1, \ldots, W_t, W'$, such that:*

1. $\mathbf{Pr}_X[X \in W'] \leq 2^{-\Omega(b)}$.
2. *For every $i \in [t]$ the probability distribution of $Y \,|\, X \in W_i$ (that is - of Y conditioned on the event $X \in W_i$) is α-good. In other words: for every $i \in [t]$ there exist at least $(1-\epsilon) \cdot u$ values of $z \in [u]$ for which*

$$H^{\infty}(Y_z | X \in W_i) \geq \frac{b}{r+1+\alpha}.$$

Before proving Lemma 3.5 we show how this lemma can be used to prove Lemma 3.2.

Proof of Lemma 3.2: The lemma follows immediately from Lemma 3.5 and from the following equality, which holds for every partition $W_1, \ldots, W_t, W'$, and for every y.

$$\mathbf{Pr}[Y = y] = \sum_{i=1}^{t} \mathbf{Pr}[X \in W_i] \cdot \mathbf{Pr}[Y = y \,|\, X \in W_i] + \mathbf{Pr}[X \in W'] \cdot \mathbf{Pr}[Y = y \,|\, X \in W'].$$

If the partition $W_1, \ldots, W_t, W'$ satisfies the two conditions of Lemma 3.5 then from Definition 3.1 it is clear that Y is exponentially (in b) close to a convex combination of α-good distributions. □

Proof of Lemma 3.5: Every random variable Y_z is a function of X, and so it partitions $\{0,1\}^{n \cdot k}$ in the following way:

$$\{0,1\}^{n \cdot k} = \bigcup_{y \in \{0,1\}^{p \cdot l}} (Y_z)^{-1}(y),$$

where $(Y_z)^{-1}(y) \triangleq \left\{x \in \{0,1\}^{n \cdot k} \,|\, Y_z(x) = y\right\}$. For each $z \in [u]$ we define the set

$$\begin{aligned} B_z &\triangleq \bigcup_{\left\{y \;\middle|\; \mathbf{Pr}[Y_z = y] > 2^{-\frac{b}{r+1+\alpha/2}}\right\}} (Y_z)^{-1}(y) \\ &= \left\{x' \in \{0,1\}^{n \cdot k} \;\middle|\; \mathbf{Pr}_X[Y_z(X) = Y_z(x')] > 2^{-\frac{b}{r+1+\alpha/2}}\right\}. \end{aligned}$$

Intuitively, B_z contains all values of x that are "bad" for Y_z, where in "bad" we mean that $Y_z(x)$ is obtained with high probability in the distribution $Y_z(X)$.

Next, we define a set $S \subset [u]^{r+1}$ in the following way:

$$S \triangleq \left\{(z_1, \ldots, z_{r+1}) \in [u]^{r+1} \mid Y_{z_1}, \ldots, Y_{z_{r+1}} \ \textit{determine}^9 \ X_1\right\}.$$

The next claim shows that for every $r+1$ seed values $(z_1, \ldots, z_{r+1}) \in S$, the set of x's that are "bad" for all of them is of exponentially small probability. That is, for most values of x at least one of the random variables $Y_{z_1}, \ldots, Y_{z_{r+1}}$ is such that $Y_{z_i}(x)$ is obtained with small enough probability.

Claim 3.6. *For all $(z_1, \ldots, z_{r+1}) \in S$ it holds that*

$$\mathbf{Pr}_X[X \in B_{z_1} \cap \ldots \cap B_{z_{r+1}}] \leq 2^{-\frac{\alpha}{2(r+1)+\alpha} \cdot b}.$$

Proof. For each $i \in [r+1]$ we can partition B_{z_i} (according to the value of Y_{z_i}) into m_i disjoint sets $B_{i,1}, \ldots, B_{i,m_i}$ such that the following three conditions hold:

1. For all $j \in [m_i]$, $\mathbf{Pr}_X[X \in B_{i,j}] > 2^{-\frac{b}{r+1+\alpha/2}}$.
2. For all $j \in [m_i]$, Y_{z_i} is constant on $B_{i,j}$.
3. $m_i \leq 2^{\frac{b}{r+1+\alpha/2}}$.

(1 and 2 follow from the definition of B_z. 3 follows from 1).

For every $(j_1, \ldots, j_{r+1}) \in [m_1] \times \ldots \times [m_{r+1}]$ we know that

$$\mathbf{Pr}_X[X \in B_{1,j_1} \cap \ldots \cap B_{(r+1),j_{r+1}}] \leq 2^{-b},$$

(this is because X_1 is constant on this intersection, and $\mathrm{H}^{\infty}(X_1) \geq b$).

We can now write

$$\begin{aligned}
\mathbf{Pr}_X[X \in B_{z_1} \cap \ldots \cap B_{z_{r+1}}] &= \sum_{(j_1,\ldots,j_{r+1}) \in [m_1] \times \ldots \times [m_{r+1}]} \mathbf{Pr}_X[X \in B_{1,j_1} \cap \ldots \cap B_{(r+1),j_{r+1}}] \\
&\leq m_1 \cdot m_2 \cdot \ldots \cdot m_{r+1} \cdot 2^{-b} \\
&\leq 2^{\frac{r+1}{r+1+\alpha/2} \cdot b} \cdot 2^{-b} = 2^{-\frac{\alpha}{2(r+1)+\alpha} \cdot b}.
\end{aligned}$$

□

The next lemma (rephrased from [Raz05]) and the corollary that follows, show that every set $A \subset [u]$, who's density is larger than ϵ, contains at least one $(r+1)$-tuple from S.

[9] We say that a random variable Y determines another random variable X if the entropy of X given Y is zero (i.e. X is a function of Y).

Lemma 3.7 (Lemma 4.4 in [Raz05]). *Let F be a finite field, and let $e_1, \ldots, e_r$ denote the first r unit vectors in the standard basis of $F^{k'}$, for some $k' > r$. Let $A \subset F^{k'}$ be a set of size larger than $r \cdot |F|^{k'-1}$. Then, there exists $v \in F^{k'}$, and non-zero $\alpha_1, \ldots, \alpha_r \in F$, such that all $r+1$ vectors $v, v+\alpha_1 e_1, \ldots, v+\alpha_r e_r$ belong to A.*

□

Corollary 3.8. *Let $A \subset [u]$ be a set of density larger than ϵ. Then $A^{r+1} \cap S \neq \emptyset$.*

Proof. We view the set $[u]$ as the vector space $F^{k \cdot r}$ (recall that F is a field of size 2^p). If the density of A is larger than $\epsilon = r \cdot 2^{-p}$ then

$$|A| > \epsilon \cdot u = r \cdot 2^{-p} \cdot u = r \cdot |F|^{k \cdot r - 1}.$$

We can apply Lemma 3.7 to get a vector $v \in F^{k \cdot r}$ such that

$$(v, v+\alpha_1 e_1, \ldots, v+\alpha_r e_r) \in A^{r+1}.$$

Recalling Construction 2.1, we can write $X_1 = (X_{11}, \ldots, X_{1r})$, where each X_{1j} is in F^l. Now, for every $j \in [r]$ the pair of random variables Y_v and $Y_{v+\alpha_j e_j}$ satisfies

$$\frac{1}{\alpha_j}\left(Y_{v+\alpha_j e_j} - Y_v\right) = X_{1j}$$

(where the operations are preformed in the vector space F^l), and so they determine X_{1j}. This means that the set of $r+1$ random variables $Y_v, Y_{v+\alpha_1 e_1}, \ldots, Y_{v+\alpha_r e_r}$ determine X_1. Hence,

$$(v, v+\alpha_1 e_1, \ldots, v+\alpha_r e_r) \in S.$$

□

We now define for each $x \in \{0,1\}^{n \cdot k}$ a vector $\pi(x) \in \{0,1\}^u$ in the following way :

$$\forall z \in [u] \quad , \quad \pi(x)_z = 1 \iff x \in B_z.$$

For a vector $\pi \in \{0,1\}^u$, let $w(\pi)$ denote the weight of π (i.e. the number of 1's in π). Since the weight of $\pi(x)$ denotes the number of seed values for which x is "bad", we would like to somehow show that for most x's $w(\pi(x))$ is small. This can be proven by combining Claim 3.6 with Corollary 3.8, as shown by the following claim.

Claim 3.9.

$$\mathbf{Pr}_X[w(\pi(X)) > \epsilon \cdot u] \leq u^{r+1} \cdot 2^{-\frac{\alpha}{2(r+1)+\alpha} \cdot b}.$$

Proof. If x is such that $w(\pi(x)) > \epsilon \cdot u$ then, by Corollary 3.8, we know that there exists an $(r+1)$-tuple $(z_1, \ldots, z_{r+1}) \in S$ such that $x \in B_{z_1} \cap \ldots \cap B_{z_{r+1}}$. Therefore we have

$$\mathbf{Pr}_X[w(\pi(X)) > \epsilon \cdot u] \leq \mathbf{Pr}_X[\,\exists (z_1, \ldots, z_{r+1}) \in S \ s.t. \ X \in B_{z_1} \cap \ldots \cap B_{z_{r+1}}].$$

Now, using the union bound and Claim 3.6 we can bound this probability by

$$|S| \cdot 2^{-\frac{\alpha}{2(r+1)+\alpha} \cdot b} \leq u^{r+1} \cdot 2^{-\frac{\alpha}{2(r+1)+\alpha} \cdot b}.$$

□

From Claim 3.9 we see that every x (except for an exponentially small set) is contained in at most $\epsilon \cdot u$ sets B_z. The idea is now to partition the space $\{0,1\}^{n \cdot k}$ into sets of x's that have the same $\pi(x)$. If we condition the random variable Y on the event $\pi(X) = \pi_0$, where π_0 is of small weight, we will get an α-good distribution. We now explain this idea in more details. Let $\lambda \triangleq \frac{\alpha}{2(r+1+\alpha)(r+1+\alpha/2)}$, and define

$$BAD_1 \triangleq \{\pi' \in \{0,1\}^u \quad | \quad w(\pi') > \epsilon \cdot u\},$$

$$BAD_2 \triangleq \left\{\pi' \in \{0,1\}^u \quad | \quad \mathbf{Pr}_X[\pi(X) = \pi'] < 2^{-\lambda \cdot b}\right\},$$

$$BAD \triangleq BAD_1 \cup BAD_2.$$

The set $BAD \subset \{0,1\}^u$ contains values $\pi' \in \{0,1\}^u$ that cannot be used in the partitioning process described in the last paragraph. There are two reasons why a specific value $\pi' \in \{0,1\}^u$ is included in BAD. The first reason is that the weight of π' is too large (i.e. larger than $\epsilon \cdot u$), these values of π' are included in the set BAD_1. The second less obvious reason for π' to be excluded from the partitioning is that the set of x's for which $\pi(x) = \pi'$ is of extremely small probability. These values of π' are bad because we can say nothing about the min-entropy of Y when conditioned on the event[10] $\pi(X) = \pi'$.

Having defined the set BAD, we are now ready to define the partition required by Lemma 3.5. Let $\{\pi^1, \ldots, \pi^t\} = \{0,1\}^u \backslash BAD$. We define the sets $W_1, \ldots, W_t, W' \subset \{0,1\}^{n \cdot k}$ as follows:

- $W' = \{x \,|\, \pi(x) \in BAD\}$.
- $\forall i \in [t] \quad , \quad W_i = \{x \,|\, \pi(x) = \pi^i\}$.

Clearly, the sets $W_1, \ldots, W_t, W'$ form a partition of $\{0,1\}^{n \cdot k}$. We will now show that this partition satisfies the two conditions required by Lemma 3.5. To prove the first part of the lemma note that the probability of W' can be bounded by (using Claim 3.9 and the union-bound)

$$\begin{aligned} \mathbf{Pr}_X[X \in W'] &\leq \mathbf{Pr}_X[\pi(X) \in BAD_1] + \mathbf{Pr}_X[\pi(X) \in BAD_2] \\ &\leq u^{r+1} \cdot 2^{-\frac{\alpha}{2(r+1)+\alpha} \cdot b} + 2^u \cdot 2^{-\lambda \cdot b} = 2^{-\Omega(b)}. \end{aligned}$$

[10] Consider the extreme case where there is only one $x_0 \in \{0,1\}^{n \cdot k}$ with $\pi(x_0) = \pi'$. In this case the min-entropy of Y, when conditioned on the event $X \in \{x_0\}$, is zero, even if the weight of $\pi(x_0)$ is small.

We now prove that $W_1, \ldots, W_t$ satisfy the second part of the lemma. Let $i \in [t]$, and let $z \in [u]$ be such that $(\pi^i)_z = 0$ (there are at least $(1-\epsilon) \cdot u$ such values of z). Let $y \in \{0,1\}^{p \cdot l}$ be any value. If $\mathbf{Pr}[Y_z = y] > 2^{-\frac{b}{r+1+\alpha/2}}$ then $\mathbf{Pr}[Y_z = y \,|\, X \in W_i] = 0$ (this follows from the way we defined the sets B_z and W_i). If on the other hand $\mathbf{Pr}[Y_z = y] \leq 2^{-\frac{b}{r+1+\alpha/2}}$ then

$$\begin{aligned}
\mathbf{Pr}[Y_z = y \mid X \in W_i] &\leq \frac{\mathbf{Pr}[Y_z = y]}{\mathbf{Pr}[X \in W_i]} \\
&\leq 2^{-\frac{b}{r+1+\alpha/2}} / 2^{-\lambda \cdot b} \\
&= 2^{\left(\lambda - \frac{1}{r+1+\alpha/2}\right) \cdot b} \\
&= 2^{-\frac{b}{r+1+\alpha}}.
\end{aligned}$$

Hence, for all values of y we have $\mathbf{Pr}[Y_z = y \mid X \in W_i] \leq 2^{-\frac{b}{r+1+\alpha}}$. We can therefore conclude that for all $i \in [t]$, $H^\infty(Y_z | X \in W_i) \geq \frac{b}{r+1+\alpha}$. This completes the proof of Lemma 3.5. □

References

[LRVW03] Chi-Jen Lu, Omer Reingold, Salil Vadhan, and Avi Wigderson. Extractors: optimal up to constant factors. In *STOC '03: Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 602–611. ACM Press, 2003.

[Raz05] Ran Raz. Extractors with weak random seeds. *STOC 2005 (to appear)*, 2005.

[TS96] Amnon Ta-Shma. On extracting randomness from weak random sources (extended abstract). In *STOC '96: Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 276–285. ACM Press, 1996.