Better lossless condensers through derandomized curve samplers

Amnon Ta-Shma*
Computer Science Department
Tel-Aviv University
Tel Aviv, Israel 69978

Christopher Umans†
Computer Science Department
California Institute of Technology
Pasadena, CA 91125

Abstract

Lossless condensers are unbalanced expander graphs, with expansion close to optimal. Equivalently, they may be viewed as functions that use a short random seed to map a source on n bits to a source on many fewer bits while preserving all of the min-entropy. It is known how to build lossless condensers when the graphs are slightly unbalanced [3]. The highly unbalanced case is also important but the only known construction does not condense the source well. We give explicit constructions of lossless condensers with condensing close to optimal, and using near-optimal seed length.

Our main technical contribution is a randomness-efficient method for sampling $\mathbb{F}^D$ (where $\mathbb{F}$ is a field) with low-degree curves. This problem was addressed before [2, 6] but the solutions apply only to degree one curves, i.e., lines. Our technique is new and elegant. We use sub-sampling and obtain our curve samplers by composing a sequence of low-degree manifolds, starting with high-dimension, low-degree manifolds and proceeding through lower and lower dimension manifolds with (moderately) growing degrees, until we finish with dimension-one, low-degree manifolds, i.e., curves. The technique may be of independent interest.

1. Introduction

Expanders are sparse graphs with the property that every "not too large" set of vertices has many neighbors. One can view expanders as *balanced* bipartite graphs, where u on one side is connected to v on the other iff (u, v) is an edge in the original graph. Typically one is interested in constant-degree expanders, which give rise to constant degree balanced bipartite graphs. Explicit constructions of expanders have numerous applications in computer science and combinatorics.

A number of applications demand a different variant, *unbalanced bipartite expanders*, which are sparse bipartite graphs $G = (V, W, E)$ where every "not too large" subset of V has many neighbors in W, and W is much smaller than V. These objects retain the original "expansion" property, while simultaneously mapping elements of V into a much smaller domain. Often this last feature is crucial (e.g., in the error correcting codes of [12]). Unbalanced expanders are often called *condensers*.

More precisely, a condenser is a function[1] $C : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$ with the property that for every distribution X on $\{0,1\}^n$ with min-entropy k_1, the distribution $C(X, U_t)$ is ϵ-close to a distribution with min-entropy k_2. One typically wants to maximize k_2 (and bring it close to $k_1 + t$) while minimizing m (it can be as small as $k_1 + t + O(\log(\frac{1}{\epsilon}))$) and t (it can be as small as $\log((n-k)/(m-k)) + \log(1/\epsilon) + O(1)$). We call a condenser *lossless* if $k_2 = k_1 + t$.

Lossless condensers have some special properties not possessed by lossy condensers. In particular they have the *unique neighbor property*: for every "not too large" subset of V, a constant fraction of the nodes in the subset have unique neighbors in W. Some applications require this property – see the introduction to [3] for a nice outline of many applications in routing, error-correcting codes, fault tolerance, and others.

Capalbo et al. [3] give constructions of lossless condensers with optimal seed length, but the construction time is doubly-exponential in the shrinking factor $n - m$ ([3], Thm 7.2). This gives an explicit construction for the *slightly unbalanced* case where $n - m$ is small, and in particu-

*amnon@post.tau.ac.il. Supported by the Israel Science Foundation, by the Binational Science Foundation, and by the EU Integrated Project QAP.

†umans@cs.caltech.edu. Supported by NSF Grant CCF-0346991, BSF Grant 2004329, and an Alfred P. Sloan Research Fellowship.

[1] As is standard, we use the functional notation, which implicitly describes a bipartite graph. Namely, $V = \{0,1\}^n$, $W = \{0,1\}^m$, and $(v, w) \in E$ iff there exists a y for which $C(v, y) = w$.

Proceedings of the 47th Annual IEEE Symposium
on Foundations of Computer Science (FOCS'06)
0-7695-2720-5/06 $20.00

lar a constant seed length, t, when $n - m$ is a constant. Solving the problem for this restricted regime of parameters supplied the right unbalanced expander for many important applications (e.g., the error correcting codes mentioned above).

The highly unbalanced case is also of great importance. This is demonstrated by the many extractor and disperser constructions that involve a condenser as a main ingredient (e.g., [7, 13, 15, 4, 9, 16, 5] just to mention a partial list). In fact, in many of these constructions the progress was made by improving the condenser quality and then using the new condenser in a sophisticated way (e.g., the sequence of papers [7, 13, 9, 5]). This is not surprising, as extractors are a special case of condensers (when $m = k_2$). We note, however, that in spite of much effort, most of these condensers are *lossy*, which means that $k_2 < k_1$.

Lossless condensers (as opposed to lossy ones) can be used in a completely modular fashion, and are an important goal because other objects can be easily derived from them. For example, it was pointed out several times (e.g., in [9, 5]) that by applying a lossless condenser to a source on n bits with k min-entropy, one obtains a source on somewhat more than k bits in which the min-entropy has been preserved. An extractor for very high min-entropy can then be applied (and this parameter setting has historically been easier to deal with).

In spite of their usefulness, there are few constructions of lossless condensers. For very high min-entropies, the already mentioned Zig-Zag construction [3] gives nearly-optimal lossless condensers. It is also not too hard to get a lossless condenser for every min-entropy k with seed length $O(\log^3 n)$ (see, e.g., [5], Theorem 7.3). For the very low min-entropy regime ($k = O(\log(n))$) the *lossless extractors* of [13] combined with simple hashing gives a nearly optimal lossless condenser, and this was used several times (e.g., in the extractor-condenser pairs of [8]). Finally, the lossless condensers or [16] work for all min-entropies k, but the output length $m = k^{1+\epsilon}$ is larger than what one might hope for.

We change the picture significantly in this paper. We construct lossless condensers with much smaller output length, for any min-entropy k. We do not get seed length $t = O(\log n)$ but we get close. Specifically, we obtain in Theorems 6 and 5, respectively:

- for any min-entropy k and any constant $\alpha > 0$, a lossless condenser $C : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$ with output length $m = k \cdot (\log n)^{O(1)}$ and seed length $t = O((\log n)^{1+\alpha})$, and,
- for any min-entropy k, a lossless condenser $C : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$ with output length $m = k \cdot O(\log(n))^{2\log\log(n)}$ and seed length $t = O(\log(n)\log\log(n))$.

This should be compared with seed length $t = O(\log^2 n)$ that is required by all previous constructions that have output length smaller than $k^{1+\epsilon}$, and with the lower bound $t = \Omega(\log(n))$.

Our results are obtained using a new and elegant technique that we describe next.

1.1. Derandomized curve samplers

Our main technical contribution is a "derandomized curve sampler," which we believe to be of independent interest. To describe how this enables us to produce lossless condensers, we briefly outline the technique of [16] for obtaining lossless condensers from so-called "reconstructive extractors."

The approach in [16] builds on Trevisan's insight [18] that certain pseudo-random generator (PRG) constructions can be converted to extractor constructions. One way to frame this (as done by [16] and further formalized in [20]) is to observe that certain "reconstructive extractors" come equipped with an additional randomized "advice function" A and a "reconstruction procedure" with the property that for every large enough subset X, with high probability over y, the advice $A(x, y)$ (and knowledge of what is the set X) suffices to reconstruct x using the prescribed reconstruction procedure. In [16] it is proved that in such setups the advice function A is a lossless condenser!

There are really just two basic constructions of "reconstructive extractors." The first, breakthrough, construction is Trevisan's, and it can be described as mainly combinatorial (relying on error-correcting codes, and combinatorial designs). The results in [16] are based on this construction. The second construction is by Shaltiel and Umans [11] and it can be characterized as mainly algebraic (it is based on another reconstructive construction by [17] that has a geometric intuition). Their advice function $A(x, y)$ is the following: first encode $x \in \{0,1\}^n$ as a low-degree polynomial $p : \mathbb{F}^D \to \mathbb{F}$ with $|\mathbb{F}^D| = poly(n)$, then use the randomness y to select a degree $t = O(\log n)$ curve in $\mathbb{F}^D$, and finally output p restricted to m successive "shifts" of that curve. The standard way to select a random degree t curve in $\mathbb{F}^D$ is to select t random points in $\mathbb{F}^D$ and pass a curve through them. This means that the advice function uses $O(t \log n) = O(\log^2 n)$ random bits, which is too many. So the bottleneck preventing us from obtaining a lossless condenser from [11] is that picking a random curve requires too much randomness. If it were not for this bottleneck, the construction in [11] would yield a lossless condenser with much smaller output length than Trevisan's construction, for the same reason that the parameters of the corresponding extractor construction in [11] can be tuned to handle low min-entropies without blowing up the seed length.

Proceedings of the 47th Annual IEEE Symposium
on Foundations of Computer Science (FOCS'06)
0-7695-2720-5/06 $20.00 © 2006 IEEE

The curve sampling problem arises quite often. We frequently need to sample a 0/1 function f defined on $\mathbb{F}^D$ such that:

- the sample space is t-wise independent so that we can apply t-wise independent tail bounds and make the sampling error small enough, and
- simultaneously, we need to exploit the special properties of sampling along low-degree curves – namely that the restriction of a low-degree function over $\mathbb{F}^D$ to a curve is low degree.

This combination of requirements arises in the above extractor construction [11], but also in PCP constructions, hardness amplification and decoding of Reed-Muller codes [14], algebraic PRG constructions [11, 19], and some pure complexity results (e.g. [10]).

The randomness required to sample the curve is an important parameter in these settings: it is the seed length for condensers, and it is related to the PCP length in the PCP setting, the list-size in the decoding setting, and the non-uniformity in hardness amplification.

A probabilistic argument shows that there exists a small subset of degree t curves, samplable using $O(\log n)$ randomness that samples f well; i.e., the error behaves as it would for t-wise independent samplers. The challenge is to describe such a subset explicitly, or "derandomize curve samplers." Motivated by the goal of constructing short PCPs, this problem was tackled before for the case of degree 1 curves, or lines, resulting in two beautiful papers. In [2] Ben-Sasson et al. show how to derandomize line samplers by picking one random point in $\mathbb{F}^D$ and a random direction for the line in an ϵ-biased set. Moshkovitz and Raz [6] use a different approach: they pick a direction for the line in a subfield. However, it is not at all clear how to generalize these results to higher degree curves.

In this paper we show how to get close to optimal curve samplers, and we do that using a new technique. Our idea is to use sub-sampling. We illustrate the idea with a toy example. A useful rule-of-thumb is that the sampling error when choosing N points t-wise independently is approximately $N^{-t/2}$. Assume we want to pick degree $t = O(\log n)$ curves in $\mathbb{F}_q^D$, with $q^D = poly(n)$. Such curves produce sampling error that is roughly $q^{-t/2}$, and they require $O(tD \log q) = O(\log^2 n)$ randomness to sample directly. Rather than pick curves immediately, we first sample a random $t^{1/2}$-dimensional subspace V of $\mathbb{F}_q^D$. The $q^{t^{1/2}}$ points in V are close to being $t^{1/2}$-wise independent[2], and so we expect a sampling error of about $|V|^{-t^{1/2}/2} = q^{-t/2}$. Now we pick a random degree t curve in the subspace V, which gives a sampling error of about $q^{-t/2}$ as before. Overall, the sampling error is about what it would have been for picking the curve directly. But we have gained in the randomness: we picked $t^{1/2}$ points in $\mathbb{F}_q^D$, and t points in V for a total of $t^{1/2}\log(q^D) + t\log(q^{t^{1/2}}) = O(\log^{3/2} n \log q)$ random bits, which is an improvement for typical settings of q.

A natural idea is to use more steps, implementing the above sub-sampling process gradually. At each step i the dimension d_i of the vector space we work with becomes smaller, and the independence t_i used must become larger (to keep the error small). At a certain stage the required independence t_i becomes larger than the dimension of the vector space. At this point we cannot get t_i-wise independence by choosing a linear subspace. So, when the dimension of the vector space we work with becomes too small, we achieve the required independence by picking a *low-degree manifold*. At the end of the process, the dimension is as small as possible – one – and the independence is t, so we are choosing a degree t one-dimensional manifold, otherwise known as a curve.

However, there is a basic bug in the above argument. Although "most" d-tuples of points in a random d dimensional subspace of $\mathbb{F}_q^D$ are d-wise independent, a $1/q^c$ fraction are only $(d-c)$-wise independent, and we are shooting for an error of about $q^{-d^2/2} \ll q^{-c}$. Indeed, David Zuckerman [21] showed us an example in which d dimensional subspaces suffer a huge sampling error, in fact the same as just using pair-wise independence! Our intended application makes critical use of the tail-bounds afforded by greater-than-pairwise independence, and so we cannot use subspaces as intermediate samplers.

Surprisingly, we bypass this problem in an easy way, as follows. We identify the vector space $\mathbb{F}_q^D$ with the field $\mathbb{F}_{q^D}$ and we choose a random degree t univariate polynomial over the field $\mathbb{F}_{q^D}$. The evaluations of this polynomial are t-wise independent points in $\mathbb{F}_{q^D}$, as are any subset of evaluations. A simple, but crucial, point is that the when we view this function as a function from the vector space $\mathbb{F}_q^D$ to the vector space $\mathbb{F}_q^D$, then each coordinate function is a D-variate low degree polynomial over $\mathbb{F}_q$ (for the simple proof see Section 3.2). We now identify the subspace $\mathbb{F}_q^d$ with a subspace in $\mathbb{F}_q^D$ via a linear map and we compose this map with the sampled function. The composition is a good sampler (because we get t-wise independence when evaluating the polynomial over the subset of points) and is low degree (because both mappings are low-degree). We give full details in Section 3.2.

Altogether, we obtain in Theorems 1 and 2:

- a curve sampler in $\mathbb{F}^D$ with error δ, that samples curves of degree $(\log D/\delta)^{\log D}$, using randomness $O(\log(|\mathbb{F}^D|) + \log(1/\delta)(\log D))$, and
- a curve sampler in $\mathbb{F}^D$ with error δ, that samples curves of degree $\log(1/\delta)^{O(1)}$, using randomness

[2] But not close enough, as we discuss below.

Proceedings of the 47th Annual IEEE Symposium
on Foundations of Computer Science (FOCS'06)
0-7695-2720-5/06 $20.00 © 2006 IEEE

$O(\log(|\mathbb{F}^D|) + \log(1/\delta)D^\alpha)$, for any constant $\alpha > 0$.

Note that an optimal curve sampler would have degree $O(\log(1/\delta)/\log q)$ and randomness $O(\log|\mathbb{F}^D| + \log(1/\delta))$. Our curve samplers immediately give rise to the two condensers mentioned earlier, by plugging them into the reconstructive extractor construction of [11].

Thus, we obtain our curve samplers by composing a sequence of low-degree manifolds, starting with high-dimension, low-degree manifolds and proceeding through lower and lower dimension manifolds with (moderately) growing degrees, until we finish with dimension-one, low-degree manifolds.

Outline. The next section contains relevant definitions and a tail bound for t-wise independence. Section 3 gives the basic manifold sampler based on Reed-Solomon codes, and Section 4 shows how to compose it with itself to obtain randomness-efficient curve samplers. Section 5 adapts to larger alphabets the proof from [16] that the advice function of reconstructive extractors is a lossless condenser and applies it to the reconstructive extractor construction of [11]. Finally Section 6 plugs in the new curve samplers to obtain improved lossless condensers.

2. Preliminaries

A *probability distribution* D on Λ is a function $D : \Lambda \to [0,1]$ such that $\sum_{x\in\Lambda} D(x) = 1$. U_n is the uniform distribution on $\{0,1\}^n$. The *variation distance* $|D_1 - D_2|$ between two probability distributions on Λ is $\frac{1}{2}\sum_{x\in\Lambda}|D_1(x) - D_2(x)| = \max_{S\subseteq X}|D_1(S) - D_2(S)|$. We say D_1 is ϵ-close to D_2 if $|D_1 - D_2| \le \epsilon$. The *support* of a distribution D is the set of all x for which $D(x) \neq 0$. A distribution D is *flat* over its support $A \subseteq \Lambda$ if $D(a) = \frac{1}{|A|}$ for all $a \in A$. If A is a set, we use A to also refer to the flat distribution with support A, when this meaning is clear from context.

If D is a distribution and f a function, then $f(D)$ denotes the distribution obtained by picking d according to the distribution D and evaluating $f(d)$. Thus, e.g., $E(X, U_t)$ denotes the distribution obtained by picking x according to the distribution X, picking y uniformly at random from $\{0,1\}^t$, and evaluating $E(x,y)$.

Distinguishers and predictors. A distinguisher is a test that distinguishes between a given distribution and the uniform distribution:

Definition 1 (distinguisher). *A function $D : \Sigma^m \to \{0,1\}$ ϵ–distinguishes a distribution X, if $|\Pr_{x\leftarrow X}[D(x) = 1] - \Pr_{u\leftarrow\Sigma^m}[D(u) = 1]| \ge \epsilon$.*

A next-element predictor is a special distinguisher that is able to predict well the i-th element of $x \in X$ given the first $i-1$ elements of x, i.e.,

Definition 2 (next-element predictor). *Let X be a distribution over Σ^m. A function $T : \Sigma^{<m} \to \Sigma$ is a next-element predictor for X with success p, if $\Pr_{i\in[m],x\leftarrow X}[T(x_1, x_2, \ldots, x_{i-1}) = x_i] \ge p$.*

Note that a next-element predictor (or a distinguisher) need not be efficient.

Extractors and condensers. We say a distribution X has *min-entropy* k, if no element x has probability mass larger than 2^{-k}. Formally:

Definition 3 (min-entropy). *The* min-entropy *of a distribution X is $H_\infty(X) = \min_a\{-\log_2 X(a)\}$.*

Definition 4 (condenser). *Let $C : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$ be a function.*

1. *We say C is a $(n,k_1) \to_\epsilon (m,k_2)$ condenser if for every distribution X with k_1 min-entropy, $C(X,U_t)$ is ϵ-close to a distribution with k_2 min-entropy.*
2. *We say C is a* strong *$(n,k_1) \to_\epsilon (m,k_2)$ condenser, if for every distribution X with k_1 min-entropy, $U_t \circ C(X,U_t)$ is ϵ-close to a distribution $U_t \circ D$ with $t + k_2$ min-entropy.*
3. *We say C is a (strong)* lossless *condenser if it is a (strong) $(n,k) \to_\epsilon (m,k)$ condenser.*

In this language we can define an extractor as a special case of a condenser.

Definition 5 (extractor). *The function $E : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$ is a (strong) (k,ϵ)-extractor if it is a (strong) $(n,k) \to_\epsilon (m,m)$ condenser.*

Both extractors and condensers are *explicit* if they can be computed in polynomial time. In the definitions above, we may equivalently take the source distribution X to be a flat distribution. This follows from two standard facts: (1) any distribution X with min-entropy k_1 can be written as a convex combination of flat distributions with min-entropy k_1; and (2) a convex combination of distributions that are each ϵ-close to distributions with min-entropy k_2 is ϵ-close to a single distribution with min-entropy k_2.

A tail bound for t-wise independence. The main tool for analyzing our new samplers is the following tail bound from [1]:

Proceedings of the 47th Annual IEEE Symposium
on Foundations of Computer Science (FOCS'06)
0-7695-2720-5/06 $20.00 © 2006 IEEE

Lemma 1 ([1]). *Let $t \geq 4$ be an even integer. Suppose $X_1, \ldots, X_m$ are t-wise independent random variables taking values in $[0,1]$ and denote $X = \sum_{i=1}^{m} X_i$ and $\mu = E(x)$. Then, for every $A > 0$ we have*

$$\Pr[|X - \mu| \geq A] \leq 8\left(\frac{t \cdot E(X) + t^2}{A^2}\right)^{t/2}.$$

Samplers. The *density* of a set $A \subseteq \mathbb{F}^D$ is $\rho(A) = \frac{|A|}{|\mathbb{F}^D|}$. The *density of A in a subset $S \subseteq \mathbb{F}^D$* is $\rho_S(A) = \Pr_{x \in S}[x \in A]$.

Definition 6 (sampler). *A sampler is a probabilistic procedure R that outputs a subset $S \subseteq \mathbb{F}^D$. We say R samples $A \subseteq \mathbb{F}^D$ with* accuracy error ϵ *and* confidence error δ *if*

$$\Pr\ [\ |\rho_S(A) - \rho(A)| \geq \epsilon\rho(A)\ ] \quad \leq \quad \delta,$$

where the probability is over the randomness of R. We say R is a (ρ, ϵ, δ) sampler if it samples all sets $A \subseteq \mathbb{F}^D$ of density at least ρ with accuracy error ϵ and confidence error δ. The randomness of the sampler is the number of random coins it uses.

3. A manifold sampler

In this section we describe the Reed-Solomon code based sampler that underlies our later constructions.

3.1. Low-degree manifolds

Let $\mathbb{F} = \mathbb{F}_q$ be the finite field of size q. A *manifold* is a function $C : \mathbb{F}^d \to \mathbb{F}^D$. We call d the *dimension* of the manifold. We view C as D individual functions $C_i : \mathbb{F}^d \to \mathbb{F}$ describing its operation on each output coordinate, i.e., $C(a) = (C_1(a), \ldots, C_d(a))$. We are interested in low-degree manifolds, defined below:

Definition 7 (low-degree manifold). *A manifold $C : \mathbb{F}^d \to \mathbb{F}^D$ has degree t if for every $1 \leq i \leq D$ the function $C_i : \mathbb{F}^d \to \mathbb{F}$ is a d-variate polynomial of degree at most t.*

Note that a (parametric) degree t curve is just a one-dimensional manifold of degree t. In discussions below, we often identify a manifold $C : \mathbb{F}^d \to \mathbb{F}^D$ with its image in $\mathbb{F}^D$.

Let $A : \mathbb{F}^{d_1} \to \mathbb{F}^D$ be a d_1-dimensional manifold and $B : \mathbb{F}^{d_2} \to \mathbb{F}^{d_1}$ a d_2-dimensional manifold. Then their composition $A \circ B : \mathbb{F}^{d_2} \to \mathbb{F}^D$ is defined to be

$$(A \circ B)(a_1, \ldots, a_{d_2}) = A(B(a_1, \ldots, a_{d_2})).$$

The composition is a new manifold of dimension d_2. Its degree is $deg_1 \cdot deg_2$ where deg_1, deg_2 are the degrees of the manifolds A and B, respectively. To see that notice that each coordinate function $A_i(b_1, \ldots, b_{d_1})$ (for $1 \leq i \leq D$) is a degree deg_1 polynomial in $b_1, \ldots, b_{d_1}$, and we substitute for each b_j a degree deg_2 polynomial in $a_1, \ldots, a_{d_2}$. Notice also that the image of $A \circ B$ is a subset of the image of A.

3.2. The Reed-Solomon manifold sampler

Definition 8 (manifold sampler). *We say R is a* manifold sampler *of dimension d and degree t if R is a sampler that outputs a dimension d, degree t manifold C.*

We now present a simple low-degree manifold sampler based on Reed-Solomon codes.

Lemma 2. *Let q be a prime power, $d < D$ integers, $\epsilon > 0$, and let $\rho > 0$ be arbitrary. For every even $4 \leq t \leq \rho q^D$ there exists a degree t manifold sampler R that outputs a manifold $C : \mathbb{F}^d \to \mathbb{F}^D$, and for which R is a*

$$\left(\rho, \epsilon, \delta = O\left(\frac{2t}{\epsilon^2 \rho q^d}\right)^{t/2}\right)$$

sampler with randomness complexity $tD \log q$.

Proof. We pick $y = (y_1, \ldots, y_t)$ with each $y_i \in \mathbb{F}_{q^D}$ uniformly at random. We define $RS_y : \mathbb{F}_{q^D} \to \mathbb{F}_{q^D}$ by

$$RS_y(x) \quad = \quad \sum_{i=1}^{t} y_i \cdot x^i,$$

where additions and multiplications are in the field $\mathbb{F}_{q^D}$.

We identify $\mathbb{F}_q^D$ with the field F_{q^D} via an arbitrary basis $\{e_1, \ldots, e_D\}$ for $\mathbb{F}_{q^D}$ over $\mathbb{F}_q$. This allows us to view RS_y as a function from $\mathbb{F}_q^D$ to $\mathbb{F}_q^D$.

The following claim is simple, but crucial:

Claim 1. *Viewing RS_y as a function from $\mathbb{F}_q^D$ to $\mathbb{F}_q^D$, each coordinate function $(RS_y)_i$ is a D-variate degree t polynomial mapping from $\mathbb{F}_q^D$ to $\mathbb{F}_q$.*

Proof. Writing the variable x in $\mathbb{F}_{q^D}$ as $\sum_{j=1}^{D} x_j e_j$ with the x_j in $\mathbb{F}_q$, and each coefficient $y_i \in \mathbb{F}_{q^d}$ as $\sum_{j=1}^{D} y_{i,j} e_j$ with the $y_{i,j} \in \mathbb{F}_q$, we obtain

$$RS_y(x) = \sum_{i=1}^{t} \left(\sum_{j=1}^{D} y_{i,j} e_j\right) \left(\sum_{j=1}^{D} x_j e_j\right)^i.$$

After multiplying out, the monomials in the x_j all have total degree at most t, and their coefficients are polynomials in the $y_{i,j}$ and e_j elements. Rewriting each of these values in the basis $(e_1, \ldots, e_D)$ and gathering the coefficients on e_i, we obtain the i-th coordinate function, which is a D-variate, degree t polynomial in the x_j variables. □

Proceedings of the 47th Annual IEEE Symposium
on Foundations of Computer Science (FOCS'06)
0-7695-2720-5/06 $20.00 © 2006 IEEE

We define the random variable $RS(f)$, for $f \in \mathbb{F}_{q^D} \setminus \{0\}$, to be the value in $\mathbb{F}^D$ obtained by picking y at random and evaluating $RS_y(f)$. We know that the random variables $(RS(f))_{f \in \mathbb{F}_{q^D} \setminus \{0\}}$ are t-wise independent.

Finally, we define a linear map $\Phi : \mathbb{F}^d \to \mathbb{F}^D$ by $\Phi(b_1, \ldots, b_d) = \sum_{j=1}^{d} b_j e_j$ and we take our manifold $C_y : \mathbb{F}^d \to \mathbb{F}^D$ to be the function $RS_y \circ \Phi$; i.e.,

$$C_y(b_1, \ldots, b_d) \quad = \quad RS_y \left(\sum_j b_j e_j \right).$$

The q^d evaluations of C_y are a subset of the q^D evaluations of RS_y and so they give rise to a t-wise independent distribution. Applying Lemma 1 we get the desired accuracy and confidence error.

By Claim 1, each coordinate function $(RS_y)_i$ is a D-variate degree t polynomial mapping from $\mathbb{F}_q^D$ to $\mathbb{F}_q$. Composing this with $\Phi : \mathbb{F}_q^d \to \mathbb{F}_q^D$, which is a degree 1, d-variate polynomial in each coordinate, gives that each coordinate function $(C_y)_i$ has degree at most t. Thus, the total degree of C_y is t. The randomness complexity is immediate. □

4. A randomness-efficient curve sampler

We save on randomness by using a small dimension, but large degree manifold sampler to *sub-sample* a larger dimension, small degree manifold. This sub-sampling corresponds to a composition of the manifold functions.

4.1. Sub-sampling

Definition 9 (composed manifold samplers). *Let R_1 be a manifold sampler outputting dimension d_1 manifolds in $\mathbb{F}^{d_0}$. Let R_2 be a manifold sampler outputting dimension d_2 manifolds in $\mathbb{F}^{d_1}$. Then, we define a new sampler $R_1 \circ R_2$ that does the following. It uses R_1 to sample a manifold $C_1 : \mathbb{F}^{d_1} \to \mathbb{F}^{d_0}$ and R_2 to sample a manifold $C_2 : \mathbb{F}^{d_2} \to \mathbb{F}^{d_1}$. It then outputs the manifold $C_1 \circ C_2 : \mathbb{F}^{d_2} \to \mathbb{F}^{d_0}$.*

We claim that if R_1 and R_2 are two good samplers then their composition $R_1 \circ R_2$ is also a good sampler.

Lemma 3. *Let R_1, R_2 be as above. If R_i (for $i \in \{1, 2\}$) is a $(\rho, \epsilon_i, \delta_i)$ sampler with randomness complexity r_i and degree t_i, then $R_1 \circ R_2$ is a $(\rho/(1-\epsilon_1), \epsilon_1 + \epsilon_2, \delta_1 + \delta_2)$ sampler with randomness complexity $r_1 + r_2$ and degree $t_1 t_2$.*

Proof. The randomness complexity and the degree are immediate. We now turn to analyze the error. Fix an arbitrary subset $A \subseteq \mathbb{F}^{d_0}$ of density at least $\rho/(1-\epsilon_1)$. As R_1 is a $(\rho, \epsilon_1, \delta_1)$ sampler, except for probability δ_1 the manifold C_1 samples A to within ϵ_1 accuracy. In other words, if A' is the set of points in $\mathbb{F}^{d_1}$ whose C_1-image is in A, then

$$|\rho(A') - \rho(A)| \leq \epsilon_1 \rho(A).$$

In particular, $\rho(A') \geq (1-\epsilon_1)\rho(A) \geq \rho$, and then because R_2 is a $(\rho, \epsilon_2, \delta_2)$ sampler, except for probability δ_2, the manifold C_2 samples A' to within ϵ_2 accuracy. This implies that the density of A in $(C_1 \circ C_2)(\mathbb{F}^{d_2})$ is within $\epsilon_1 + \epsilon_2$ accuracy of $\rho(A)$. □

4.2. Repeated sub-sampling

In the next two theorems, we give a construction of a curve sampler that minimizes the randomness complexity, and a second curve sampler that minimizes the degree (while still achieving relatively low randomness complexity). In both constructions we work over $\mathbb{F}_q^D$, and we repeatedly sub-sample with manifolds of decreasing dimension, and increasing degree, until we finally sample a dimension one manifold – a curve.

Our first construction reduces the dimension by $1/2$ in each stage, for a total of $\log_2 D$ stages.

Theorem 1. *Let ρ, ϵ, δ be arbitrary, D a power of 2, and $T = \log_2 D$. Let q be a prime power satisfying*

$$q = \Omega\left(\left(\frac{1}{\rho}\right)^2 \left(\frac{T}{\epsilon}\right)^4 \left(\log \frac{T}{\delta}\right)^2 \right).$$

There is a (ρ, ϵ, δ) manifold sampler that outputs a function from F_q to F_q^D with degree at most $\log(T/\delta)^T$, and it has randomness complexity

$$O(D \log q + \log(1/\delta) \log D).$$

Proof. Define $d_i = D/2^i$ and $t_i = 4 \left\lceil \frac{\log(T/\delta)}{d_i \log q} \right\rceil$ for $i = 0, 1, 2, \ldots, T$, and let R_i be the

$$\left(\rho' = \frac{\rho}{2}, \epsilon' = \frac{\epsilon}{2T}, \delta'\right)$$

manifold sampler of Lemma 2 that outputs a function from $F_q^{d_i}$ to $\mathbb{F}_q^{d_{i-1}}$. Our curve sampler is $R = R_1 \circ R_2 \circ \cdots \circ R_T$.

The randomness complexity of R is $\sum_{i=1}^{T} t_i d_{i-1} \log q$. This is bounded from above by $4 \sum_{i=1}^{T} d_{i-1} \log q + T \cdot O(\log(T/\delta))$ which is as claimed.

The total degree of R_i is at most $t_i \leq \log(T/\delta)$ and so the total degree of R is at most $deg_1 \deg_2 \ldots deg_T \leq \log(T/\delta)^T$.

We now analyze the error. Note that for all i, $t_i \leq \rho' q^{d_{i-1}}$ by our choice of q, and so Lemma 2 applies, and it shows that each R_i is a $(\rho', \epsilon', \delta')$ sampler with confidence error

$$\delta' = \left(\frac{c t_i T^2}{\epsilon^2 \rho q^{d_i}} \right)^{t_i/2}$$

Proceedings of the 47th Annual IEEE Symposium
on Foundations of Computer Science (FOCS'06)
0-7695-2720-5/06 $20.00 © 2006 IEEE

for some universal constant c. By our choice of q, we have $(ct_iT^2)/(\epsilon^2\rho q^{d_i}) \leq q^{-d_i/2}$. Thus, the confidence error of R_i is at most $q^{-d_i t_i/4} \leq \delta/T$.

Now we use Lemma 3 to analyze the composition. It applies because the ρ' associated with each R_i is small enough so that even $\rho'/(1-\epsilon')^T \leq \rho$, as can be seen from the following calculation: $(1-\epsilon')^T\rho \geq (1-\epsilon'T)\rho \geq \rho/2 = \rho'$. We conclude, by Lemma 3, that $R = R_1 \circ R_2 \circ \cdots \circ R_T$ is a $(\rho, \epsilon'T, \delta'T)$ sampler, and note that $\epsilon'T < \epsilon$ and $\delta'T \leq \delta$. □

Doing the same thing but with only a constant number of rounds of sub-sampling, minimizes the degree (and the lower bound on the field size q) at the cost of using more randomness (but still much less than $O(\log^2 n)$).

To get down to dimension one after a constant number of rounds, we must reduce the dimension by a constant root in each stage.

Theorem 2. *Let ρ, ϵ, δ be arbitrary, T be a positive constant, and D a T-th power. Let q be a prime power satisfying*

$$q = \Omega\left(\left(\frac{1}{\rho}\right)^2\left(\frac{T}{\epsilon}\right)^4\left(\log\frac{T}{\delta}\right)^2\right).$$

There is a (ρ, ϵ, δ) manifold sampler that outputs a function from F_q to F_q^D with degree at most $\log(1/\delta)^T$, and it has randomness complexity

$$O(D\log q + \log(1/\delta)D^{1/T}).$$

Proof. Define $d_i = D^{1-i/T}$ and $t_i = 4\left\lceil \frac{\log(T/\delta)}{d_i \log q}\right\rceil$ for $i = 0, 1, 2, \ldots, T$, and let R_i be the

$$\left(\rho' = \frac{\rho}{2}, \epsilon' = \frac{\epsilon}{2T}, \delta'\right)$$

manifold sampler of Lemma 2 that outputs a function from $F_q^{d_i}$ to $\mathbb{F}_q^{d_{i-1}}$. Our curve sampler is $R = R_1 \circ R_2 \circ \cdots \circ R_T$.

The randomness complexity of R is

$$\sum_{i=1}^{T} t_i d_{i-1}\log q \leq 4\sum_{i=1}^{T} d_{i-1}\log q + O(D^{1/T}\log(1/\delta))$$

which is as claimed.

The total degree of R_i is at most $t_i \leq \log(1/\delta)$ and so the total degree of R is at most $deg_1 \deg_2 \ldots deg_T \leq \log(1/\delta)^T$.

Applying Lemma 2 exactly as in the proof of Theorem 1, we find that the confidence error δ' of R_i is at most $q^{-d_i t_i/4} \leq \delta/T$. We use Lemma 3 to analyze the composition just as in the proof of Theorem 1. We conclude that $R = R_1 \circ R_2 \circ \cdots \circ R_T$ is a (ρ, ϵ, δ) sampler, as desired. □

4.3. A remark

When we have a sequence of dimensions $d_0 = D, d_1, \ldots, d_T = 1$ such that $d_i | d_{i-1}$, the composition has a particularly pleasant structure. Indeed this is the case for both Theorem 1 and Theorem 2. In such a case $\mathbb{F}_{q^{d_i}}$ is a subfield of $\mathbb{F}_{q^{d_{i-1}}}$. As a consequence, we do not need to move between vector spaces and fields as we do in the proof of Lemma 2. Instead, the entire construction amounts to picking univariate polynomials (Reed-Solomon codewords) $p_i : \mathbb{F}_{q^{d_i}} \to \mathbb{F}_{q^{d_i}}$ with degree t_i, and composing them to obtain $R = p_1 \circ p_2 \circ \cdots p_T$. This is a function from $\mathbb{F}_q = \mathbb{F}_{q^{d_T}}$ to $\mathbb{F}_{q^D} = \mathbb{F}_{q^{d_0}}$, and notice that the composition is well defined because $\mathbb{F}_{q^{d_i}}$ is a subfield of $\mathbb{F}_{q^{d_{i-1}}}$.

5. Lossless condensers

The following framework is adapted from [20] to apply to larger alphabets.

5.1. Reconstructive extractors yield lossless condensers

Certain extractor constructions ([18], [17], [11]) implicitly define the following object. The original purpose was to prove that E is indeed an extractor, by arguing that if it is not, then there exists a next-element predictor, and then many strings in the supposedly high-entropy source X can be reconstructed from short advice, a contradiction.

Definition 10 (reconstructive extractor). *A triple (E, A, R) of functions where:*

- *$E : \{0,1\}^n \times \{0,1\}^{r_E} \to \Sigma^m$ is called the extractor function,*
- *$A : \{0,1\}^n \times \{0,1\}^{r_A} \to \Sigma^a$ is called the advice function, and,*
- *$R : \Sigma^a \times \{0,1\}^{r_A} \times \{0,1\}^{r_R} \to \{0,1\}^n$ is called the reconstruction function*

is a (p, q) reconstructive extractor *if for every $X \subseteq \{0,1\}^n$ and every next-element predictor $T : \Sigma^{<m} \to \Sigma$ for $E(X, U_{r_E})$ with success p, we have*

$$\Pr_{x \leftarrow X, y, z}[R^T(A(x,y), y, z) = x] \geq q.$$

If R is deterministic, then we drop the third argument and just write $R : \Sigma^a \times \{0,1\}^{r_A} \to \{0,1\}^n$.

We now have two claims. First, we claim that when E's output is long enough, then a good next-element predictor exists, and second that whenever such a predictor exists, A is a lossless condenser. We begin with:

Proceedings of the 47th Annual IEEE Symposium
on Foundations of Computer Science (FOCS'06)
0-7695-2720-5/06 $20.00 © 2006 IEEE

Lemma 4. *Let $E : \{0,1\}^n \times \{0,1\}^{r_E} \to \Sigma^m$ be a function, and let $X \subseteq \{0,1\}^n$ be a subset of cardinality at most 2^k. Then, there exists a next-element predictor $T : \{0,1\}^{<m} \to \{0,1\}$ for $E(X, U_{r_E})$ with success*

$$p \geq 1 - (\ln 2)(k + r_E)/m.$$

The proof idea is that if m is much larger than the entropy of X, then E encodes an input x from X with much redundancy, and hence a good predictor exists. A similar proof was given in [16] for next-*bit* predictors.

Proof. (Of Lemma 4) We know that X has k min-entropy, but because X is flat it also has k entropy. Thus, $E(X, U_{r_E})$ has at most $k + r_E$ entropy. It follows that

$$\begin{aligned} E_i\left[H(Y_i \mid Y_{[1..i-1]})\right] &= \frac{1}{m}\sum_{i=1}^{m} H(Y_i \mid Y_{[1..i-1]}) \\ &= \frac{1}{m} H(E(X, U_{r_E})) \leq \frac{k + r_E}{m}. \end{aligned}$$

Now, assume there is no next-element predictor with success p. Then for every $1 \leq i \leq m$,

$$\Pr_{y_1,\ldots,y_{i-1}}[\text{optimal predictor succeeds}] \leq p.$$

But, the optimal predictor always guesses the element with the highest probability. Thus its success probability is exactly $2^{-H_\infty(Y_i \mid Y_{[1..i-1]}=y_{[1..i-1]})}$. Also, the min-entropy is bounded from above by the entropy. It follows that $E_i\left[2^{-H(Y_i \mid Y_{[1,i-1]})}\right] \leq p$. The function $g(z) = 2^z$ is convex, and so by Jensen's Inequality, $E_i\left[H(Y_i \mid Y_{[1,i-1]})\right] \geq \log(1/p)$.

We conclude that

$$\frac{k + r_E}{m} \geq \log\left(\frac{1}{p}\right) \geq (\log_2 e)(1 - p)$$

(using the fact that $p < e^{-(1-p)}$). This is a contradiction when $p < 1 - (\ln 2)(k + r_E)/m$. □

Our second claim is that if (E, A, R) is a reconstructive extractor, and if a good next-element predictor for $E(X, U_{r_E})$ exists, then $A(X, U_{r_A})$ retains the entropy of X. This argument is identical to [16, 20]. We state it here and give a proof for completeness.

Lemma 5. *Let (E, A, R) be a $(p, q = 1 - \epsilon)$ reconstructive extractor and $X \subseteq \{0,1\}^n$ a subset such that there exists a next-element predictor $T : \Sigma^{<m} \to \Sigma$ for $E(X, U_{r_E})$ with success p. Then the distribution $U_{r_A} \circ A(X, U_{r_A})$ is $O(\epsilon)$-close to a distribution $U_{r_A} \circ D$ with $r_A + \log_2 |X|$ min-entropy.*

Proof. Let us call a pair (x, y) with $x \in X$ and $y \in \{0,1\}^{r_A}$ *good* if

$$\Pr_z[R^T(A(x, y), y, z) = x] \quad > \quad 1/2 \qquad (1)$$

Let G be the set of good pairs (x, y). Since we know $\Pr_{x \leftarrow X, y, z}[R^T(A(x, y), y, z) = x] \geq 1 - \epsilon$, we obtain, by an averaging argument, that $\Pr_{x \leftarrow X, y}[(x, y) \in G] \geq 1 - 2\epsilon$.

Now notice that Equation (1) implies that if (x_1, y) and (x_2, y) are both good, then $A(x_1, y) \neq A(x_2, y)$. This holds because if $A(x_1, y) = A(x_2, y)$ then $\Pr_z[R^T(A(x_1, y), y, z) = x_2] > 1/2$, contradicting Equation (1). In particular, if we define $A'(x, y) = y \circ A(x, y)$, then A' is one-to-one on the set of good pairs G.

However, as argued above, almost every element of $X \times \{0,1\}^{r_A}$ is good, and so the flat distribution on the set G is $O(\epsilon)$-close to the distribution $X \circ U_{r_A}$. In particular, the probability mass on elements of $A'(X, U_{r_A})$ with multiple preimages is at most $O(\epsilon)$ (since A' is one-to-one on G). By redistributing this mass, we obtain a distribution $D \circ U_{r_A}$ with min-entropy $\log_2 |X| + r_A$ that is $O(\epsilon)$-close to $A'(X, U_{r_A})$, which proves the lemma. □

Combining Lemmas 5 and 4 we see that the advice function of a reconstructive extractor (with long enough output length m) is a lossless condenser:

Theorem 3. *Assume the triple of functions (E, A, R) as in Definition 10 is a $(p = \frac{7}{8}, 1 - \epsilon)$ reconstructive extractor. Then A is a strong $(n, k) \to_{O(\epsilon)} (a, k)$ condenser, provided $m \geq (8 \ln 2)(k + r_E)$.*

Proof. Let $X \subseteq \{0,1\}^n$ be an arbitrary subset of cardinality 2^k. By Lemma 4 there exists a next-element predictor T for $E(X, U_{r_E})$ with success $p \geq \frac{7}{8}$. By Lemma 5, $U_{r_A} \circ A(X, U_{r_A})$ is $O(\epsilon)$-close to a distribution with min-entropy $k + r_A$. Using the observation regarding flat distributions from Section 2, we find that A is the desired lossless condenser. □

5.2. The SU reconstructive extractor

We now present the Shaltiel-Umans (SU) construction using the "reconstructive extractor" terminology, and with the curve-samplers abstraction.

Theorem 4. *Let q be a prime power and let D be an integer for which $q^D = n^2$. Assume there exists a family of efficient $(\rho = 0.5, \epsilon = 0.01, \delta = 1/q^{2D})$ one-dimensional manifold samplers, sampling degree $t(n)$ curves in $\mathbb{F}_q^D$ with randomness $r(n)$.*

Then, provided that $q \geq (100Dt(n))^2$, for every $m = m(n)$, there exists a triple of functions

$$\begin{aligned} E &: \{0,1\}^n \times \{0,1\}^{r_E} \to \Sigma^m \\ A &: \{0,1\}^n \times \{0,1\}^{r_A} \to \Sigma^a \\ R &: \Sigma^a \times \{0,1\}^{r_A} \to \{0,1\}^n \end{aligned}$$

Proceedings of the 47th Annual IEEE Symposium
on Foundations of Computer Science (FOCS'06)
0-7695-2720-5/06 $20.00 © 2006 IEEE

that is a $(p = \frac{7}{8}, 1 - (1/q^D))$ *reconstructive extractor with* $r_E = D \log q$, $r_A = r(n)$ *and* $a \leq mq \log q$.

We now give a sketch of the proof. For a full correctness proof the reader should consult [11]. We note however that one of the main complications in that work is avoided in the present setting when our goal is a condenser construction rather than an extractor construction. When using the SU reconstructive extractor as an extractor construction, one needs to work with a next-element predictor that has only slightly better success than random guessing. A basic step in the reconstruction procedure is to use the predictor to learn the restriction of a low-degree polynomial to a curve. When the predictor has such a low success rate, it makes many errors, and one must use list-decoding to recover from these. Picking the "correct" decoding out of the list entails a complicated analysis of two separate "interleaved" curves, where the first curve is used to disambiguate along the second, and vice versa.

In the present setting, the existence of a next-element predictor will be guaranteed by Lemma 4, and we can set parameters so that the predictor has success rate close to one. When using this predictor to learn the restriction of a low-degree polynomial to a curve, we suffer few errors, and we can use unique decoding. Consequently there is no need for interleaved curves, and the construction and its analysis become more straightforward.

Proof. (of Theorem 4, sketch) Pick a subset $H \subseteq \mathbb{F}_q$ with $h = |H| = q^{1/2}$, and identify $[n]$ with the set H^D. We think of $x \in \{0,1\}^n$ as a mapping $x : H^D \to \mathbb{F}_q$ and we extend it to the unique polynomial $\widehat{x} : \mathbb{F}_q^D \to \mathbb{F}_q$ with individual variable degree at most $h - 1$. We let α be a generator of the multiplicative group $\mathbb{F}^*_{q^D}$.

- For $x \in \{0,1\}^n$ and $y \in \mathbb{F}_q^D$ we define $E(x, y)$ to be $(\widehat{x}(y), \widehat{x}(\alpha y), \ldots, \widehat{x}(\alpha^{m-1} y))$.

- For $x \in \{0,1\}^n$ and $y \in \{0,1\}^{r_A}$ we define $A(x, y)$ as follows. Use y as randomness to sample a curve $C : \mathbb{F}_q \to \mathbb{F}_q^D$ using the curve sampler. Then, $A(x, y)$ outputs the evaluations of $\widehat{x}$ at the mq evaluation points $\alpha^{-1}C(\mathbb{F}_q), \alpha^{-2}C(\mathbb{F}_q), \ldots, \alpha^{-m}C(\mathbb{F}_q)$.

- We now define the reconstruction procedure. It receives as input a random y which describes a curve $C : \mathbb{F}_q \to \mathbb{F}_q^D$, together with the evaluations of $\hat{x}$ at $\alpha^{-1}C(\mathbb{F}_q)$, $\alpha^{-2}C(\mathbb{F}_q), \ldots, \alpha^{-m}C(\mathbb{F}_q)$. Using this data, we can apply the next-element predictor that succeeds with probability 7/8 to predict the evaluations of $\widehat{x}$ at the points $C(\mathbb{F}_q)$. The points on which the predictor succeeds form the set we are sampling using the curve sampler. We know that except for probability $\frac{1}{q^{2D}}$, the sampled curve is good, in which case our predictions are correct on at least $0.8q$ elements. We also know that the true evaluations form a low-degree univariate polynomial of degree at most $Dht(n) \leq q/100$ and so the distance is at least $.99q$ and we can uniquely correct more than $0.45q$ errors. We error correct our predictions and obtain, with probability at least $1 - 1/q^{2D}$, the restriction of $\widehat{x}$ to the curve C.

  At this point, we would like to repeat the argument to predict the "next" curve αC. For this we need to argue that a random *shifted* curve from a curve sampler is good with high probability. This holds by the following simple, and general, observation:

  Claim 2. *Let* R *be a sampler that outputs subsets of* $\mathbb{F}_q^D$. *For* $\alpha \in \mathbb{F}_{q^D}$ *define* αR *to be the sampler that samples a subset* $S \subseteq \mathbb{F}_q^D$ *according to* R *and outputs the subset* $\alpha S = \{\alpha s \mid s \in S\}$. *Then* αR *is a* (ρ, ϵ, δ) *sampler iff* R *is.*

  Proof. The number of times a subset $S \subseteq \mathbb{F}_q^D$ hits $A \subseteq \mathbb{F}_q^D$ is the same as the number of times αS hits αA. In particular, αR samples A with accuracy error ϵ and confidence error δ iff $R = \alpha^{-1}(\alpha R)$ samples $\alpha^{-1}A$ with accuracy error ϵ and confidence error δ. But $\alpha^{-1}A$ is of the same density as A, and so αR is a (ρ, ϵ, δ) sampler iff R is. □

  If R is our curve sampler, then we can view the curve αC as coming from the curve sampler αR. Thus, except for probability $\frac{1}{q^{2D}}$, the curve αC is good, and after decoding, we learn $\widehat{x}$ restricted to αC. Repeating the argument, we learn $\widehat{x}$ restricted to $\alpha^2 C$, $\alpha^3 C$, and so on (for q^D successive shifts), until we learn the evaluations of $\widehat{x}$ on the whole space $\mathbb{F}_q^D$ and then we recover x. Each step succeeds with probability $1 - 1/q^{2D}$, and so by a union bound, we succeed in recovering x with probability at least $1 - (1/q^D)$.

  Finally, it is easy to check that r_A and r_E are as stated. □

6. Lossless condenser constructions

Plugging in parameters we get:

Theorem 5. *Let* $k = k(n) = \Omega(\log n \log\log n)$. *Then* $A : \{0,1\}^n \times \{0,1\}^{r(n)} \to \{0,1\}^{a(n)}$ *as described in Theorem 4, using the sampler in Theorem 1, is an explicit strong*

$$(n, k(n)) \to_{O(1/n^2)} (a(n), k(n))$$

lossless condenser, with output length $a(n) = k \cdot O(\log n)^{2 \log\log n}$ *and randomness* $r(n) = O(\log n \log\log n)$.

Proceedings of the 47th Annual IEEE Symposium
on Foundations of Computer Science (FOCS'06)
0-7695-2720-5/06 $20.00 © 2006 IEEE

Proof. We set the parameters in the SU construction as follows: $q = \Omega\left((\log n)^{2\log\log n}\right)$, $D = 2\log n/\log q$ and $m = 100k$. We need a $(0.5, 0.01, 1/n^4)$ curve sampler. Theorem 1 gives us this as long as $q = \Omega(\log^4 D \log^2(\log(Dn^4)))$. We have chosen $q = \Omega((\log n)^{2\log\log n})$ and observe that for some constant c,

$$\log^4 D \log^2(\log(Dn^4)) \le c \cdot (\log\log n)^4 \log^2 n,$$

which is at most $(\log n)^{2\log\log n}$ for sufficiently large n. Therefore, by Theorem 1, the required curve sampler exists, and it samples degree $t(n) = (\log(n^4 \log D))^{\log D}$ curves, and uses randomness $r(n) = O(\log D \log n)$.

Now, we verify that $q > (100Dt(n))^2$. This holds because $(100Dt(n))^2 = O((\log n)^{2\log D + O(1)}) = O((\log n)^{2\log\log n})$ for sufficiently large n. Therefore Theorem 4 applies and gives us a reconstructive extractor (E, A, R) whose advice function A has randomness $r(n) = O(\log n \log\log n)$ and output length $mq \log q \le k \cdot O(\log n)^{2\log\log n}$, as claimed. We can also verify by our choice of m that Theorem 3 applies, and it shows that A is the desired lossless condenser. □

Plugging our second sampler minimizes the output length at the expense of the randomness:

Theorem 6. *Fix a constant $\alpha > 0$. Let $k = k(n) = \Omega(\log^{1+\alpha} n)$. Then $A : \{0,1\}^n \times \{0,1\}^{r(n)} \to \{0,1\}^{a(n)}$ as described in Theorem 4, using the sampler in Theorem 2 with $T = 1/\alpha$, is an explicit strong*

$$(n, k(n)) \to_{O(1/n^2)} (a(n), k(n))$$

lossless condenser, with output length $a(n) = k \cdot (\log n)^{O(1)}$ and randomness $r(n) = O((\log n)^{1+\alpha})$.

Proof. We set the parameters in the SU construction as follows : $q = (\log n)^{\Theta(1)}$, $D = 2\log n/\log q$ and $m = 100k$. We need a $(0.5, 0.01, 1/n^4)$ curve sampler. Theorem 2 gives us this as long as $q = \Omega(\log^4 D \log^2(\log(Dn^4)))$, which indeed holds by nearly the same verification as in the proof of Theorem 5. Therefore, by Theorem 2, the required curve sampler exists, and it samples degree $t(n) = O(\log n)^T$ curves, and uses randomness $r(n) = O(\log^{1+\alpha} n)$.

Now, we observe that $q > (100Dt(n))^2$ because $(100Dt(n))^2 = O(\log n)^{2(1+T)}$. Therefore Theorem 4 applies and gives us the reconstructive extractor (E, A, R) whose advice function A has the claimed randomness and output length. As in the proof of Theorem 5, by our choice of m Theorem 3 applies, and it shows that A is a lossless condenser. □

Acknowledgments. We thank Ronen Shaltiel and Eli Ben-Sasson for sharing their insights on the problem with us.

References

[1] M. Bellare and J. Rompel. Randomness-efficient oblivious sampling. In *FOCS*, pages 276–287, 1994.

[2] E. Ben-Sasson, M. Sudan, S. Vadhan, and A. Wigderson. Randomness-efficient low degree tests and short PCPs via epsilon-biased sets. In *STOC*, pages 612–621, 2003.

[3] M. Capalbo, O. Reingold, S. Vadhan, and A. Wigderson. Randomness conductors and constant-degree expansion beyond the degree/2 barrier. In *STOC*, pages 659–668, 2002.

[4] R. Impagliazzo, R. Shaltiel, and A. Wigderson. Extractors and pseudo-random generators with optimal seed length. In *STOC*, pages 1–10, 2000.

[5] C.-J. Lu, O. Reingold, S. Vadhan, and A. Wigderson. Extractors: Optimal up to constant factors. In *STOC*, pages 602–611, 2003.

[6] D. Moshkovitz and R. Raz. Sub-constant error low degree test of almost linear size. In *STOC*, 2006.

[7] N. Nisan and D. Zuckerman. Randomness is linear in space. *Journal of Computer and System Sciences*, 52(1):43–52, 1996.

[8] R. Raz and O. Reingold. On recycling the randomness of states in space bounded computation. In *STOC*, pages 159–168, 1999.

[9] O. Reingold, R. Shaltiel, and A. Wigderson. Extracting randomness via repeated condensing. In *FOCS*, pages 22–31, 2000.

[10] R. Shaltiel and C. Umans. Pseudorandomness for approximate counting and sampling. In *Computational Complexity*, pages 212–226, 2005.

[11] R. Shaltiel and C. Umans. Simple extractors for all min-entropies and a new pseudorandom generator. *Journal of the ACM*, 52(2):172–216, 2005.

[12] M. Sipser and D. A. Spielman. Expander codes. *IEEE Transactions on Information Theory*, 42(6):1710–1722, 1996.

[13] A. Srinivasan and D. Zuckerman. Computing with very weak random sources. *SIAM Journal on Computing*, 28:1433–1459, 1999.

[14] M. Sudan, L. Trevisan, and S. Vadhan. Pseudorandom generators without the xor lemma. *Journal of Computer and System Sciences*, 62(2):236–266, 2001.

[15] A. Ta-Shma. On extracting randomness from weak random sources. In *STOC*, pages 276–285, 1996.

[16] A. Ta-Shma, C. Umans, and D. Zuckerman. Loss-less condensers, unbalanced expanders, and extractors. In *STOC*, pages 143–152, 2001.

[17] A. Ta-Shma, D. Zuckerman, and S. Safra. Extractors from Reed-Muller codes. *Journal of Computer and System Sciences*, 72(5):786–812, 2006.

[18] L. Trevisan. Extractors and pseudorandom generators. *Journal of the ACM*, pages 860–879, 2001.

[19] C. Umans. Pseudo-random generators for all hardnesses. *Journal of Computer and System Sciences*, 67(2):419–440, 2003.

[20] C. Umans. Reconstructive dispersers and hitting set generators. In *RANDOM*, pages 460–471, 2005.

[21] D. Zuckerman. Personal communication, June 2006.