Randomness-Efficient Curve Samplers

Zeyu Guo*

Computer Science Department
California Institute of Technology
Pasadena, CA 91125

Abstract. Curve samplers are sampling algorithms that proceed by viewing the domain as a vector space over a finite field, and randomly picking a low-degree curve in it as the sample. Curve samplers exhibit a nice property besides the sampling property: the restriction of low-degree polynomials over the domain to the sampled curve is still low-degree. This property is often used in combination with the sampling property and has found many applications, including PCP constructions, local decoding of codes, and algebraic PRG constructions.

The randomness complexity of curve samplers is a crucial parameter for its applications. It is known that (non-explicit) curve samplers using $O(\log N + \log(1/\delta))$ random bits exist, where N is the domain size and δ is the confidence error. The question of explicitly constructing randomness-efficient curve samplers was first raised in [TSU06] where they obtained curve samplers with near-optimal randomness complexity.

We present an explicit construction of low-degree curve samplers with *optimal* randomness complexity (up to a constant factor), sampling curves of degree $\left(m \log_q(1/\delta)\right)^{O(1)}$ in $\mathbb{F}_q^m$. Our construction is a delicate combination of several components, including extractor machinery, limited independence, iterated sampling, and list-recoverable codes.

1 Introduction

Randomness has numerous uses in computer science, and sampling is one of its most classical applications: Suppose we are interested in the size of a particular subset A lying in a large domain D. Instead of counting the size of A directly by enumeration, one can randomly draw a small sample from D and calculate the density of A in the sample. The approximated density is guaranteed to be close to the true density with probability $1 - \delta$ where δ is very small, known as the *confidence error*. This sampling technique is extremely useful both in practice and in theory.

One class of sampling algorithms, known as *curve samplers*, proceed by viewing the domain as a vector space over a finite field, and picking a random low-degree curve in it. Curve samplers exhibit the following nice property besides the sampling property: the restriction of low-degree polynomials over the domain to

* zguo@caltech.edu. Supported by NSF CCF-1116111, NSF CCF-1038578 and BSF 2010120.

the sampled curve is still low-degree. This special property, combined with the sampling property, turns out to be useful in many settings, e.g local decoding of Reed-Muller codes and hardness amplification [STV01], PCP constructions [AS98, ALM+98, MR08], algebraic constructions of pseudorandom-generators [SU05, Uma03], extractor constructions [SU05, TSU06], and some pure complexity results (e.g. [SU06]).

The problem of explicitly constructing low-degree curve samplers was raised in [TSU06]. Typically, we are looking for low-degree curve samplers with small sample complexity (polylogrithmic in the domain size) and confidence error (polynomially small in the domain size), and we focus on minimizing the randomness complexity. The simplest way is picking a completely random low-degree curve whose sampling properties are guaranteed by tail bounds for limited independence. The randomness complexity of this method, however, is far from being optimal. The probabilistic method guarantees the existence of (non-explicit) low-degree curve samplers using $O(\log N + \log(1/\delta))$ random bits where N is the domain size and δ is the confidence error. The real difficulty, however, is to find an explicit construction matching this bound.

1.1 Previous work

Randomness-efficient samplers (without the requirement that the sample points form a curve) are constructed in [CG89, Gil98, BR94, Zuc97]. In particular, [Zuc97] obtains explicit samplers with optimal randomness complexity (up to a $1+\gamma$ factor for arbitrary small $\gamma > 0$) using the connection between samplers and extractors. See [Gol11] for a survey of samplers.

Degree-1 curve samplers are also called *line samplers*. Explicit randomness-efficient line samplers are constructed in the PCP literature [BSSVW03, MR08], motivated by the goal of constructing almost linear sized PCPs. In [BSSVW03] line samplers are derandomized by picking a random point and a direction sampled from an ϵ-biased set, instead of two random points. An alternative way is suggested in [MR08] where directions are picked from a subfield. It is not clear, however, how to apply these techniques to higher degree curves.

In [TSU06] it was shown how to explicitly construct derandomized curve samplers with near-optimal parameters. Formally they obtained

- curve samplers picking curves of degree $(\log\log N + \log(1/\delta))^{O(\log\log N)}$ using randomness $O(\log N + \log(1/\delta)\log\log N)$, and
- curve samplers picking curves of degree $(\log(1/\delta))^{O(1)}$ using randomness $O(\log N + \log(1/\delta)(\log\log N)^{1+\gamma})$ for any constant $\gamma > 0$

for domain size N, field size $q \geq (\log N)^{\Theta(1)}$ and confidence error $\delta = N^{-\Theta(1)}$. Their work left the problem of explicitly constructing low-degree curve samplers (ideally picking curves of degree $O(\log_q(1/\delta))$) with essentially optimal $O(\log N + \log(1/\delta))$ random bits as a prominent open problem.

1.2 Main results

We present an explicit construction of low-degree curve samplers with optimal randomness complexity (up to a constant factor). In particular, we show how to sample degree-$\left(m \log_q(1/\delta)\right)^{O(1)}$ curves in $\mathbb{F}_q^m$ using $O(\log N + \log(1/\delta))$ random bits for domain size $N = |\mathbb{F}_q^m|$ and confidence error $\delta = N^{-\Theta(1)}$. Before stating our main theorem, we first present the formal definition of samplers and curve samplers.

Samplers. Given a finite set $\mathcal{M}$ as the domain, the *density* of a subset $A \subseteq \mathcal{M}$ is $\mu(A) \stackrel{\text{def}}{=} \frac{|A|}{|\mathcal{M}|}$. For a collection of elements $\mathcal{T} = \{t_i : i \in I\} \in \mathcal{M}^I$ indexed by set I, the density of A in $\mathcal{T}$ is $\mu_{\mathcal{T}}(A) \stackrel{\text{def}}{=} \frac{|A \cap \mathcal{T}|}{|\mathcal{T}|} = \Pr_{i \in I}[t_i \in A]$.

Definition 1 (sampler). *A* sampler *is a function $S : \mathcal{N} \times \mathcal{D} \to \mathcal{M}$ where $|\mathcal{D}|$ is its* sample complexity *and $\mathcal{M}$ is its* domain*. We say S samples $A \subseteq \mathcal{M}$ with* accuracy error ϵ *and* confidence error δ *if $\Pr_{x \in \mathcal{N}}[|\mu_{S(x)}(A) - \mu(A)| > \epsilon] \leq \delta$ where $S(x) \stackrel{def}{=} \{S(x, y) : y \in \mathcal{D}\}$. We say S is an (ϵ, δ) sampler if it samples all subsets $A \subseteq \mathcal{M}$ with accuracy error ϵ and confidence error δ. The* randomness complexity *of S is $\log(|\mathcal{N}|)$.*

Definition 2 (curve/line sampler). *Let $\mathcal{M} = \mathbb{F}_q^D$ and $\mathcal{D} = \mathbb{F}_q$. The sampler $S : \mathcal{N} \times \mathcal{D} \to \mathcal{M}$ is a* degree-t curve sampler *if for all $x \in \mathcal{N}$, the function $S(x, \cdot) : \mathcal{D} \to \mathcal{M}$ is a curve (see Definition 3) of degree at most t over $\mathbb{F}_q$. When $t = 1$, S is also called a* line sampler*.*

Theorem 1 (main). *For any $\epsilon, \delta > 0$, integer $m \geq 1$, and sufficiently large prime power $q \geq \left(\frac{m \log(1/\delta)}{\epsilon}\right)^{\Theta(1)}$, there exists an explicit degree-t curve sampler for the domain $\mathbb{F}_q^m$ with $t = \left(m \log_q(1/\delta)\right)^{O(1)}$, accuracy error ϵ, confidence error δ, sample complexity q, and randomness complexity $O\left(m \log q + \log(1/\delta)\right) = O(\log N + \log(1/\delta))$ where $N = q^m$ is the domain size. Moreover, the curve sampler itself has degree $\left(m \log_q(1/\delta)\right)^{O(1)}$ as a polynomial map.*

Theorem 1 has better degree bound and randomness complexity compared with the constructions in [TSU06]. We remark that the degree bound, being $\left(m \log_q(1/\delta)\right)^{O(1)}$, is still sub-optimal compared with the lower bound $\log_q(1/\delta)$ (see Appendix A for the proof of this lower bound). However in many cases it is satisfying to achieve such a degree bound.

As an example, consider the following setting of parameters: domain size $N = q^m$, field size $q = (\log N)^{\Theta(1)}$, confidence error $\delta = N^{-\Theta(1)}$, and accuracy error $\epsilon = (\log N)^{-\Theta(1)}$. Note that this is the typical setting in PCP and other literature [ALM+98, AS98, STV01, SU05]. In this setting, we have the following corollary in which the randomness complexity is logarithmic and the degree is polylogarithmic.

Corollary 1. *Given domain size $N = |\mathbb{F}_q^m|$, accuracy error $\epsilon = (\log N)^{-\Theta(1)}$, confidence error $\delta = N^{-\Theta(1)}$, and large enough field size $q = (\log N)^{\Theta(1)}$, there exists an explicit degree-t curve sampler for the domain $\mathbb{F}_q^m$ with accuracy error ϵ, confidence error δ, randomness complexity $O(\log N)$, sample complexity q, and $t \leq (\log N)^c$ for some constant $c > 0$ independent of the field size q.*

It remains an open problem to explicitly construct curve samplers that have optimal randomness complexity $O(\log N + \log(1/\delta))$ (up to a constant factor), and sample curves with optimal degree bound $O(\log_q(1/\delta))$. It is also an interesting problem to achieve the optimal randomness complexity up to a $1+\gamma$ factor for any constant $\gamma > 0$ (rather than just an $O(1)$ factor), as achieved by [Zuc97] for general samplers. The standard techniques as in [Zuc97] are not directly applicable as they increase the dimension of samples and only yield $O(1)$-dimensional manifold samplers.

1.3 Techniques

Extractor machinery. It was shown in [Zuc97] that samplers are equivalent to *extractors*, objects that convert weakly random distributions into almost uniform distributions. Therefore the techniques of constructing extractors are extremely useful in constructing curve samplers. Our construction employs the technique of block source extraction [NZ96, Zuc97, SZ99]. In addition, we also use the techniques appeared in [GUV09], especially their constructions of *condensers*.

Limited independence. It is well known that points on a random degree-$(t-1)$ curve are t-wise independent. So we may simply pick a random curve and use tail inequalities to bound the confidence error. However, the sample complexity is too high, and hence we need to use the technique of *iterated sampling* to reduce the number of sample points.

Iterated sampling. Iterated sampling is a useful technique for explicitly constructing randomness-efficient samplers [BR94, TSU06]. The idea is first picking a large sample from the domain and then draw a sub-sample from the previous sample. The drawback of iterated sampling, however, is that it invests randomness twice while the confidence error does not shrink correspondingly. To remedy this problem, we add another ingredient into our construction, namely the technique of *error reduction*.

Error reduction via list-recoverable codes. We will use explicit list-recoverable codes (a strengthening of list-decodable codes [GI01]). More specifically, we will employ the list-recoverability from (folded) Reed-Solomon codes [GR08, GUV09]. List-recoverable codes provide a way of obtaining samplers with very small confidence error from those with mildly small confidence error. We refer to this transformation as *error reduction*, which plays a key role in our construction.

1.4 Sketch of the construction

Our curve sampler is the composition of two samplers which we call the *outer sampler* and the *inner sampler* respectively. The outer sampler picks manifolds (see Definition 3) of dimension $O(\log m)$ from the domain $\mathcal{M} = \mathbb{F}_q^m$. The outer sampler has near-optimal randomness complexity but the sample complexity is large. To fix this problem, we employ the idea of iterated sampling. Namely we regard the manifold picked by the outer sampler as the new domain $\mathcal{M}'$, and then construct an inner sampler picking a curve from $\mathcal{M}'$ with small sample complexity.

The outer sampler is obtained by constructing an extractor and then using the extractor-sampler connection [Zuc97]. We follow the approach in [NZ96, Zuc97, SZ99]: Given an arbitrary random source with enough min-entropy, we will first use a *block source converter* to convert it into a *block source*, and then feed it to a *block source extractor*. In addition, we need to construct these components carefully so as to maintain the low-degree-ness. The way we construct the block source converter is different from those in [NZ96, Zuc97, SZ99] (as they are not in the form of low-degree polynomial maps), and is based on the Reed-Solomon condenser proposed in [GUV09]: To obtain one block, we simply feed the random source and a fresh new seed into the condenser, and let the output be the block. We show that this indeed gives a block source.

The inner sampler is constructed using techniques of iterated sampling and error reduction. We start with the basic curve samplers picking totally random curves, and then apply the error reduction as well as iterated sampling techniques repeatedly to obtain the desired inner sampler. Either of the two operations improves one parameter while worsening some other one: Iterated sampling reduces sample complexity but increases the randomness complexity, whereas error reduction reduces the confidence error but increases the sample complexity. Our construction applies the two techniques alternately such that (1) we keep the invariant that the confidence error is always exponentially small in the randomness complexity, and (2) the sample complexity is finally brought down to q.

Outline. The next section contains relevant definitions and some basic facts. Section 3 gives the construction of the outer sampler using block source extraction. Section 4 introduces the techniques of error reduction and iterated sampling, and then uses them to construct the inner sampler. These components are finally put together in Section 5 and yield the curve sampler construction.

2 Preliminaries

We denote the set of numbers $\{1, 2, \ldots, n\}$ by $[n]$. Given a prime power q, write $\mathbb{F}_q$ for the finite field of size q. Write $U_{n,q}$ for the uniform distribution over $\mathbb{F}_q^n$. Logarithms are taken with base 2 unless the base is explicitly specified.

Random variables and distributions are represented by upper-case letters whereas their specific values are represented by lower-case letters. Write $x \leftarrow X$

if x is sampled according to distribution X. The support of a distribution X over set S is $\mathrm{supp}(X) \stackrel{\text{def}}{=} \{x \in S : \Pr[X = x] > 0\}$. The statistical distance between distributions X, Y over set S is defined as $\Delta(X,Y) = \max_{T \subseteq S} |\Pr[X \in T] - \Pr[Y \in T]|$. We say X is ϵ-*close* to Y if $\Delta(X,Y) \leq \epsilon$.

For an event A, let $\mathbf{I}[A]$ be the indicator variable that evaluates to 1 if A occurs and 0 otherwise. For a random variable X and an event A that occurs with nonzero probability, define the *conditional distribution* $X|_A$ by $\Pr[X|_A = x] = \frac{\Pr[(X=x) \wedge A]}{\Pr[A]}$.

Manifolds and curves. Let $f : \mathbb{F}_q^d \to \mathbb{F}_q^D$ be a polynomial map. We may view f as D individual polynomials $f_i : \mathbb{F}_q^d \to \mathbb{F}_q$ describing its operation on each output coordinate, i.e., $f(x) = (f_1(x), \ldots, f_D(x))$ for all $x \in \mathbb{F}_q^d$. Such maps are called *curves* or *manifolds*, depending on the dimension d.

Definition 3 (manifold). *A manifold in $\mathbb{F}_q^D$ is a polynomial map $M : \mathbb{F}_q^d \to \mathbb{F}_q^D$ where $M_1, \ldots, M_D$ are d-variate polynomials over $\mathbb{F}_q$. We call d the* dimension *of M. An 1-dimensional manifold is also called a curve. A curve of degree 1 is also called a* line*. The degree of M is $\deg(M) \stackrel{def}{=} \max\{\deg(M_1), \ldots, \deg(M_D))\}$.*

We need the following lemma, generalizing the one in [TSU06]. The proof is the same as in [TSU06] and we omit it.

Lemma 1. *A manifold $f : \left(\mathbb{F}_{q^D}\right)^n \to \left(\mathbb{F}_{q^D}\right)^m$ of degree t, when viewed as a manifold $f : \left(\mathbb{F}_q^D\right)^n \to \left(\mathbb{F}_q^D\right)^m$, also has degree at most t.*

Basic line/curve samplers The simplest line (resp. curve) samplers are those picking completely random lines (resp. curves), as defined below. We call them *basic line* (resp. *curve*) *samplers*.

Definition 4 (basic line sampler). *Given $m \geq 1$ and prime power q, let $\mathsf{Line}_{m,q} : \mathbb{F}_q^{2m} \times \mathbb{F}_q \to \mathbb{F}_q^m$ be the line sampler that picks a completely random line in $\mathbb{F}_q^m$. Formally,*

$$\mathsf{Line}_{m,q}((a,b),y) \stackrel{def}{=} (a_1 y + b_1, \ldots, a_m y + b_m)$$

for $a = (a_1, \ldots, a_m), b = (b_1, \ldots, b_m) \in \mathbb{F}_q^m$ and $y \in \mathbb{F}_q$.

Definition 5 (basic curve sampler). *Given $m \geq 1$, $t \geq 4$ and prime power q, let $\mathsf{Curve}_{m,t,q} : \mathbb{F}_q^{tm} \times \mathbb{F}_q \to \mathbb{F}_q^m$ be the curve sampler that picks a completely random curve of degree $t-1$ in $\mathbb{F}_q^m$. Formally,*

$$\mathsf{Curve}_{m,t,q}((c_0, \ldots, c_{t-1}), y) \stackrel{def}{=} \left(\sum_{i=0}^{t-1} c_{i,1} y^i, \ldots, \sum_{i=0}^{t-1} c_{i,m} y^i \right)$$

for each $c_0 = (c_{0,1}, \ldots, c_{0,m}), \ldots, c_{t-1} = (c_{t-1,1}, \ldots, c_{t-1,m}) \in \mathbb{F}_q^m$ and $y \in \mathbb{F}_q$.

Remark 1. Note that $\mathsf{Line}_{m,q}$ has degree 2 and $\mathsf{Curve}_{m,t,q}$ has degree t as polynomial maps.

Lemma 2. *For $\epsilon > 0$, $m \geq 1$ and prime power q, $\mathsf{Line}_{m,q}$ is an $\left(\epsilon, \frac{1}{\epsilon^2 q}\right)$ line sampler.*

Lemma 3. *For $\epsilon > 0$, $m \geq 1$, $t \geq 4$ and sufficiently large prime power $q = (t/\epsilon)^{O(1)}$, $\mathsf{Curve}_{m,t,q}$ is an $\left(\epsilon, q^{-t/4}\right)$ sampler.*

Lemma 2 follows from Chebyshev's inequality and pairwise independence of points on a random line. Similarly, Lemma 3 follows from the tail inequalities for t-wise independence. See [TSU06, Lemma 2] for more details.

Extractors and condensers. A *(seeded) extractor* is an object that takes an imperfect random variable called the *(weakly) random source*, invests a small amount of randomness called the *seed*, and produces an output whose distribution is very close to the uniform distribution.

Definition 6 (q-ary min-entropy). *We say X has* q-ary min-entropy k *if for any $x \in S$, it holds that $\Pr[X = x] \leq q^{-k}$ (or equivalently, X has min-entropy $k \log q$).*

Definition 7 (condenser/extractor). *Given a function $f : \mathbb{F}_q^n \times \mathbb{F}_q^d \to \mathbb{F}_q^m$, we say f is an $k_1 \to_{\epsilon,q} k_2$* condenser *if for every distribution X with q-ary min-entropy k_1, $f(X, U_{d,q})$ is ϵ-close to a distribution with q-ary min-entropy k_2. We say f is a (k, ϵ, q)* extractor *if it is a $k \to_{\epsilon,q} m$ condenser.*

Remark 2. We are interested in extractors and samplers that are polynomial maps. For such an object f, we denote by $\deg(f)$ its degree as a polynomial map.

The following connection between extractors and samplers was observed in [Zuc97].

Theorem 2 ([Zuc97], restated). *Given a map $f : \mathbb{F}_q^n \times \mathbb{F}_q^d \to \mathbb{F}_q^m$, we have the following:*

1. *If f is a (k, ϵ, q) extractor, then it is also an (ϵ, δ) sampler where $\delta = 2q^{k-n}$.*
2. *If f is an $(\epsilon/2, \delta)$ sampler where $\delta = \epsilon q^{k-n}$, then it is also a (k, ϵ, q) extractor.*

3 Outer sampler

In this section we construct a sampler whose randomness complexity is optimal up to a constant factor. We refer to it as the "outer sampler".

We need the machinery of block source extraction.

Definition 8 (block source [CG88]). *A random source $X = (X_1, \ldots, X_s)$ over $\mathbb{F}_q^{n_1} \times \cdots \times \mathbb{F}_q^{n_s}$ is a $(k_1, \ldots, k_s)$* q-ary block source *if for any $i \in [s]$ and $(x_1, \ldots, x_{i-1}) \in \mathrm{supp}(X_1, \ldots, X_{i-1})$, the distribution $X_i|_{X_1 = x_1, \ldots, X_{i-1} = x_{i-1}}$ has q-ary min-entropy k_i. Each X_i is called a* block.

Definition 9 (block source extractor). *A function* $E : (\mathbb{F}_q^{n_1} \times \cdots \times \mathbb{F}_q^{n_s}) \times \mathbb{F}_q^d \to \mathbb{F}_q^m$ *is a* $((k_1, \ldots, k_s), \epsilon, q)$ block source extractor *if for any* $(k_1, \ldots, k_s)$ *q-ary block source* $(X_1, \ldots, X_s)$ *over* $\mathbb{F}_q^{n_1} \times \cdots \times \mathbb{F}_q^{n_s}$, $E((X_1, \ldots, X_s), U_{d,q})$ *is* ϵ*-close to* $U_{m,q}$.

The special structure of block sources allows us to compose several extractors and get a block source extractor, with a small amount of randomness invested.

Definition 10 (block source extraction via composition). *Let* $s \geq 1$ *be an integer and* $E_i : \mathbb{F}_q^{n_i} \times \mathbb{F}_q^{d_i} \to \mathbb{F}_q^{m_i}$ *be a map for each* $i \in [s]$. *Suppose that* $m_i \geq d_{i-1}$ *for all* $i \in [s]$, *where we set* $d_0 = 0$. *Define* $E = \mathsf{BlkExt}(E_1, \ldots, E_s)$ *as follows:*

$$E : (\mathbb{F}_q^{n_1} \times \cdots \times \mathbb{F}_q^{n_s}) \times \mathbb{F}_q^{d_s} \to (\mathbb{F}_q^{m_1 - d_0} \times \cdots \times \mathbb{F}_q^{m_s - d_{s-1}})$$
$$((x_1, \ldots, x_s), y_s) \mapsto (z_1, \ldots, z_s)$$

where for $i = s, \ldots, 1$, *we iteratively define* (y_{i-1}, z_i) *to be a partition of* $E_i(x_i, y_i)$ *into the prefix* $y_{i-1} \in \mathbb{F}_q^{d_{i-1}}$ *and the suffix* $z_i \in \mathbb{F}_q^{m_i - d_{i-1}}$.

The idea behind Definition 10 is to composes a chain of extractors E_i with decreasing output length and seed length. Then we use each E_i's output as the seed of the previous extractor E_{i-1}, so that the only seed the whole object actually needs is the (typically very short) one of E_s.

Lemma 4. *Let* $s \geq 1$ *be an integer and for each* $i \in [s]$, *let* $E_i : \mathbb{F}_q^{n_i} \times \mathbb{F}_q^{d_i} \to \mathbb{F}_q^{m_i}$ *be a* (k_i, ϵ_i, q) *extractor of degree* $t_i \geq 1$. *Then* $\mathsf{BlkExt}(E_1, \ldots, E_s)$ *is a* $((k_1, \ldots, k_s), \epsilon, q)$ *block source extractor of degree* t *where* $\epsilon = \sum_{i=1}^s \epsilon_i$ *and* $t = \prod_{i=1}^s t_i$.

The proof is standard and we defer it to the full version of this paper.

3.1 Block source conversion

Definition 11 (block source converter [NZ96]). *A function* $C : \mathbb{F}_q^n \times \mathbb{F}_q^d \to \mathbb{F}_q^{m_1} \times \cdots \times \mathbb{F}_q^{m_s}$ *is a* $(k, (k_1, \ldots, k_s), \epsilon, q)$ *block source converter if for any random source* X *over* $\mathbb{F}_q^n$ *with q-ary min-entropy* k, *the output* $C(X, U_{d,q})$ *is* ϵ*-close to a* $(k_1, \ldots, k_s)$ *q-ary block source.*

It was shown in [NZ96] that one can obtain a block by choosing a pseudorandom subset of bits of the random source. Yet the analysis is pretty delicate and cumbersome. Furthermore the resulting extractor does not have a nice algebraic structure. We observe that the following condenser from Reed-Solomon codes in [GUV09] can be used to obtain blocks and is a low-degree manifold.

Definition 12 (condenser from Reed-Solomon codes [GUV09]). *Let* $\zeta \in \mathbb{F}_q$ *be a generator of the multiplicative group* $\mathbb{F}_q^\times$. *Define* $\mathsf{RSCon}_{n,m,q} : \mathbb{F}_q^n \times \mathbb{F}_q \to \mathbb{F}_q^m$ *for* $n, m \geq 1$ *and prime power q:*

$$\mathsf{RSCon}_{n,m,q}(x, y) = \big(y, f_x(y), f_x(\zeta y), \ldots, f_x(\zeta^{m-2} y)\big)$$

where $f_x(Y) = \sum_{i=0}^{n-1} x_i Y^i$ *for* $x = (x_0, x_1, \ldots, x_{n-1}) \in \mathbb{F}_q^n$.

Theorem 3 ([GUV09]). $\mathsf{RSCon}_{n,m,q}$ *is a* $m \to_{\epsilon,q} 0.99m$ *condenser for large enough* $q \geq (n/\epsilon)^{O(1)}$.

Remark 3. The condenser $\mathsf{RSCon}_{n,m,q}(x, y)$ is a degree-n manifold, since each monomial in any of its coordinate is of the form y or $x_i(\zeta^j y)^i$ for some $0 \leq i \leq n-1$.

We apply the above condenser to the source with independent seeds to obtain a block source.

Definition 13 (block source converter via condensing). *Given integers* $n, m_1, \ldots, m_s \geq 1$ *and prime power* q*, define the function* $\mathsf{BlkCnvt}_{n,(m_1,\ldots,m_s),q} : \mathbb{F}_q^n \times \mathbb{F}_q^s \to \mathbb{F}_q^{m_1+\cdots+m_s}$ *by*

$$\mathsf{BlkCnvt}_{n,(m_1,\ldots,m_s),q}(x, y) = (\mathsf{RSCon}_{n,m_1,q}(x, y_1), \ldots, \mathsf{RSCon}_{n,m_s,q}(x, y_s))$$

for $x \in \mathbb{F}_q^n$ *and* $y = (y_1, \ldots, y_s) \in \mathbb{F}_q^s$.

The function $\mathsf{BlkCnvt}_{n,(m_1,\ldots,m_s),q}$ is indeed a block source converter. The intuition is that conditioning on the values of the previous blocks, the random source X still has enough min-entropy, and hence we may apply the condenser to get the next block. Formally, we have the following statement whose proof is deferred to the full version of this paper.

Theorem 4. *For* $\epsilon > 0$*, integers* $s, n, m_1, \ldots, m_s \geq 1$ *and sufficiently large prime power* $q = (n/\epsilon)^{O(1)}$*,* $\mathsf{BlkCnvt}_{n,(m_1,\ldots,m_s),q}$ *is a* $(k, (k_1, \ldots, k_s), 3s\epsilon, q)$ *block source converter of degree* n *where* $k = \sum_{i=1}^{s} m_i + \log_q(1/\epsilon)$ *and each* $k_i = 0.99m_i$.

3.2 Construction of the outer sampler

By Lemma 2 and Theorem 2, the basic line samplers are also extractors.

Lemma 5. *For* $\epsilon > 0$*,* $m \geq 1$ *and prime power* q*,* $\mathsf{Line}_{m,q}$ *is a* (k, ϵ, q) *extractor of degree* 2 *where* $k = 2m - 1 + 3\log_q(1/\epsilon)$.

We employ Lemma 4 and compose the basic line samplers to get a block source extractor. It is then applied to a block source obtained from the block source converter.

Definition 14 (Outer Sampler). *For* $\delta > 0$*,* $m = 2^s$ *and prime power* q*, let* $n = 4m + \lceil \log_q(2/\delta) \rceil$*,* $d = s + 1$*, and* $d_i = 2^{s-i}$ *for* $i \in [s]$*. For* $i \in [s]$*, view* $\mathsf{Line}_{2,q^{d_i}} : \mathbb{F}_{q^{d_i}}^4 \times \mathbb{F}_{q^{d_i}} \to \mathbb{F}_{q^{d_i}}^2$ *as a manifold over* $\mathbb{F}_q$*:* $\mathsf{Line}_{2,q^{d_i}} : \mathbb{F}_q^{4d_i} \times \mathbb{F}_q^{d_i} \to \mathbb{F}_q^{2d_i}$*. Composing these line samplers* $\mathsf{Line}_{2,q^{d_i}}$ *for* $i \in [s]$ *gives the function* $\mathsf{BlkExt}(\mathsf{Line}_{2,q^{d_1}}, \ldots, \mathsf{Line}_{2,q^{d_s}}) : \mathbb{F}_q^{4d_1+\cdots+4d_s} \times \mathbb{F}_q \to \mathbb{F}_q^m$*. Finally, define function* $\mathsf{OuterSamp}_{m,\delta,q} : \mathbb{F}_q^n \times \mathbb{F}_q^d \to \mathbb{F}_q^m$ *such that for* $x \in \mathbb{F}_q^n$*,* $y \in \mathbb{F}_q^s$ *and* $y' \in \mathbb{F}_q$*,* $\mathsf{OuterSamp}_{m,\delta,q}(x, (y, y'))$ *equals*

$$\mathsf{BlkExt}(\mathsf{Line}_{2,q^{d_1}}, \ldots, \mathsf{Line}_{2,q^{d_s}})\big(\mathsf{BlkCnvt}_{n,(4d_1,\ldots,4d_s),q}(x, y), y'\big).$$

Theorem 5. *For any $\epsilon, \delta > 0$, integer $m \geq 1$, and sufficiently large prime power $q \geq (n/\epsilon)^{O(1)}$, $\mathsf{OuterSamp}_{m,\delta,q}$ is an (ϵ, δ) sampler of degree t where $d = O(\log m)$, $n = O\left(m + \log_q(1/\delta)\right)$ and $t = O\left(m^2 + m \log_q(1/\delta)\right)$.*

Proof. We first show that $\mathsf{OuterSamp}_{m,\delta,q}$ is a $(4m, \epsilon, q)$ extractor. Consider any random source X over $\mathbb{F}_q^n$ with q-ary min-entropy $4m$. Let s, d_i be as in Definition 14. Let $k_i = 4 \cdot 0.99 \cdot d_i$ for $i \in [s]$. Let $\epsilon_0 = \frac{\epsilon}{4s}$.

We have $(\sum_{i=1}^{s} 4d_i) + \log_q(1/\epsilon_0) \leq 4m$ for sufficiently large $q \geq (n/\epsilon)^{O(1)}$. So by Theorem 4, $\mathsf{BlkCnvt}_{n,(4d_1,\ldots,4d_s),q}$ is a $(4m, (k_1, \ldots, k_s), 3s\epsilon_0, q)$ block source converter. Therefore the distribution $\mathsf{BlkCnvt}_{n,(4d_1,\ldots,4d_s),q}(X, U_{s,q})$ is $3s\epsilon_0$-close to a $(k_1, \ldots, k_s)$ q-ary block source X'. Then $\mathsf{OuterSamp}_{m,\delta,q}(X, U_{d,q})$ is $3s\epsilon_0$-close to $\mathsf{BlkExt}(\mathsf{Line}_{2,q^{d_1}}, \ldots, \mathsf{Line}_{2,q^{d_s}})(X', U_{1,q})$.

By Lemma 5, $\mathsf{Line}_{2,q^{d_i}}$ is a $\left(k_i/d_i, \epsilon_0, q^{d_i}\right)$ extractor for $i \in [s]$ since $3 + 3\log_{q^{d_i}}(1/\epsilon_0) \leq 4 \cdot 0.99 = k_i/d_i$. Equivalently it is a (k_i, ϵ_0, q) extractor. By Lemma 4, $\mathsf{BlkExt}(\mathsf{Line}_{2,q^{d_1}}, \ldots, \mathsf{Line}_{2,q^{d_s}})$ is a $((k_1, \ldots, k_s), s\epsilon_0, q)$ block source extractor. Therefore $\mathsf{BlkExt}(\mathsf{Line}_{2,q^{d_1}}, \ldots, \mathsf{Line}_{2,q^{d_s}})(X', U_{1,q})$ is $s\epsilon_0$-close to $U_{m,q}$, which by the previous paragraph, implies that $\mathsf{OuterSamp}_{m,\delta,q}(X, U_{d,q})$ is $4s\epsilon_0$-close to $U_{m,q}$. By definition, $\mathsf{OuterSamp}_{m,\delta,q}$ is a $(4m, \epsilon, q)$ extractor. By Theorem 2, it is also an (ϵ, δ) sampler.

Finally, we have $d = s + 1 = O(\log m)$ and $n = O\left(m + \log_q(1/\delta)\right)$. By Lemma 1, each $\mathsf{Line}_{2,q^{d_i}}$ has degree 2 as a manifold over $\mathbb{F}_q$. Therefore by Lemma 4, the map $\mathsf{BlkExt}(\mathsf{Line}_{2,q^{d_1}}, \ldots, \mathsf{Line}_{2,q^{d_s}})$ has degree 2^s. By Theorem 4, $\mathsf{BlkCnvt}_{n,(4d_1,\ldots,4d_s),q}$ has degree n. Therefore $\mathsf{OuterSamp}_{m,\delta,q}$ has degree $n2^s = O\left(m^2 + m \log_q(1/\delta)\right)$. □

Remark 4. We assume m is a power of 2 above. For general m, simply pick $m' = 2^{\lceil \log m \rceil}$ and let $\mathsf{OuterSamp}_{m,\delta,q}$ be the composition of $\mathsf{OuterSamp}_{m',\delta,q}$ with the projection $\pi : \mathbb{F}_q^{m'} \to \mathbb{F}_q^m$ onto the first m coordinates. It yields an (ϵ, δ) sampler of degree t for $\mathbb{F}_q^m$ since π is linear, and approximating the density of a subset A in $\mathbb{F}_q^m$ is equivalent to approximating the density of $\pi^{-1}(A)$ in $\mathbb{F}_q^{m'}$.

4 Inner sampler

The sampler $\mathsf{OuterSamp}_{m,\delta,q}$ has randomness complexity $O\left(m \log q + \log(1/\delta)\right)$ which is optimal up to a constant factor. Yet the sample complexity is large, being $q^{O(\log m)}$. We remedy this problem by composing it with an "inner sampler" with small sample complexity. Its construction is based on two techniques called error reduction and iterated sampling.

4.1 Error reduction

Given $f : \mathbb{F}_q^n \times \mathbb{F}_q^d \to \mathbb{F}_q^m$, define $\mathrm{LIST}_f(T, \epsilon) \stackrel{\text{def}}{=} \left\{x \in \mathbb{F}_q^n : \Pr_y[f(x, y) \in T] > \epsilon\right\}$ for any $T \subseteq \mathbb{F}_q^m$ and $\epsilon > 0$. We are interesting in functions f exhibiting a "list-recoverability" property that the size of $\mathrm{LIST}_f(T, \epsilon)$ is kept small when T is not too large.

Definition 15. *A function $f : \mathbb{F}_q^n \times \mathbb{F}_q \to \mathbb{F}_q^m$ is (ϵ, L, H) list-recoverable if $|\mathrm{LIST}_f(T, \epsilon)| \leq H$ for all $T \subseteq \mathbb{F}_q^m$ of size at most L.*

We then define an operation $\star$ as follows.

Definition 16. *For functions $f : \mathbb{F}_q^n \times \mathbb{F}_q^d \to \mathbb{F}_q^m$ and $S : \mathbb{F}_q^m \times \mathbb{F}_q^{d'} \to \mathbb{F}_q^{m'}$, define $S \star f : \mathbb{F}_q^n \times (\mathbb{F}_q^d \times \mathbb{F}_q^{d'}) \to \mathbb{F}_q^{m'}$ such that $(S \star f)(x, (y, y')) \stackrel{def}{=} S(f(x, y), y')$.*

See Figure 1 for an illustration. The following lemma states that a sampler with mildly small confidence error, when composed with a list-recoverable function via the $\star$ operation, gives a sampler with very small confidence error.

Lemma 6. *Suppose $f : \mathbb{F}_q^n \times \mathbb{F}_q^d \to \mathbb{F}_q^m$ is (ϵ_1, L, H) list-recoverable, and $S : \mathbb{F}_q^m \times \mathbb{F}_q^{d'} \to \mathbb{F}_q^{m'}$ is an $(\epsilon_2, L/q^m)$ sampler. Then $S \star f$ is an $(\epsilon_1 + \epsilon_2, H/q^n)$ sampler.*

Proof. Let A be an arbitrary subset of $\mathbb{F}_q^{m'}$. Let $B = \{y \in \mathbb{F}_q^m : |\mu_{S(y)} - \mu(A)| > \epsilon_2\}$. By the sampling property of S, we have $|B| \leq (L/q^m) \cdot q^m = L$ and hence $|\mathrm{LIST}_f(B, \epsilon_1)| \leq H$. Therefore it suffices to show that for any $x \in \mathbb{F}_q^n \setminus \mathrm{LIST}_f(B, \epsilon_1)$, it holds that $|\mu_{(S \star f)(x)}(A) - \mu(A)| \leq \epsilon_1 + \epsilon_2$.

Fix $x \in \mathbb{F}_q^n \setminus \mathrm{LIST}_f(B, \epsilon_1)$. We have

$$\begin{aligned}\mu_{(S\star f)(x)}(A) = \Pr_{y,y'}[(S \star f)(x, (y, y')) \in A] &= \Pr_{y,y'}[S(f(x, y), y') \in A] \\ &= \mathbb{E}_y \left[\mu_{S(f(x,y))}(A)\right].\end{aligned}$$

Therefore

$$\begin{aligned}|\mu_{(S\star f)(x)}(A) - \mu(A)| = \left|\mathbb{E}_y \left[\mu_{S(f(x,y))}(A)\right] - \mu(A)\right| &\leq \mathbb{E}_y |\mu_{S(f(x,y))}(A) - \mu(A)| \\ &\leq \Pr_y[f(x, y) \in B] + \epsilon_2 \Pr_y[f(x, y) \notin B] \leq \epsilon_1 + \epsilon_2.\end{aligned}$$

To see the last two steps, note that $|\mu_{S(y)}(A) - \mu(A)| \leq \epsilon_2$ for $y \notin B$ by definition, and $\Pr_y[f(x, y) \in B] \leq \epsilon_1$ since $x \notin \mathrm{LIST}_f(B, \epsilon_1)$. □

It was shown in [GUV09] that the condenser $\mathsf{RSCon}_{n,m,q}$ (see Definition 12) enjoys the following list-recoverability property:

Theorem 6 ([GUV09]). *For sufficiently large $q \geq (n/\epsilon)^{O(1)}$, the function $\mathsf{RSCon}_{n,m,q}$ is $\left(\epsilon, q^{0.99m}, q^m\right)$ list-recoverable.*

Corollary 2. *For any $n \geq m \geq 1$, $\epsilon, \epsilon' > 0$ and sufficiently large prime power $q = (n/\epsilon)^{O(1)}$, suppose $S : \mathbb{F}_q^m \times \mathbb{F}_q^d \to \mathbb{F}_q^{m'}$ is an $(\epsilon', q^{-0.01m})$ sampler of degree t, then $S \star \mathsf{RSCon}_{n,m,q}$ is an $(\epsilon + \epsilon', q^{m-n})$ sampler of degree nt.*

Proof. Apply Lemma 6 and Theorem 6. Note that $\mathsf{RSCon}_{n,m,q}$ has degree n. Therefore $(S \star \mathsf{RSCon}_{n,m,q})(X, (Y, Y')) = S(\mathsf{RSCon}_{n,m,q}(X, Y), Y')$ has degree nt in its variables X, Y, Y'. □

4.2 Iterated sampling

We introduce the operation $\circ$ denoting the composition of two samplers.

Definition 17. *(composed sampler). Given functions* $S_1 : \mathbb{F}_q^{n_1} \times \mathbb{F}_q^{d_1} \to \mathbb{F}_q^{d_0}$ *and* $S_2 : \mathbb{F}_q^{n_2} \times \mathbb{F}_q^{d_2} \to \mathbb{F}_q^{d_1}$*, define* $S_1 \circ S_2 : (\mathbb{F}_q^{n_1} \times \mathbb{F}_q^{n_2}) \times \mathbb{F}_q^{d_2} \to \mathbb{F}_q^{d_0}$ *such that* $(S_1 \circ S_2)((x_1, x_2), y) \stackrel{def}{=} S_1(x_1, S_2(x_2, y))$.

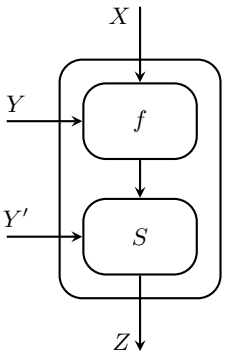

Fig. 1. The operation $S \star f$

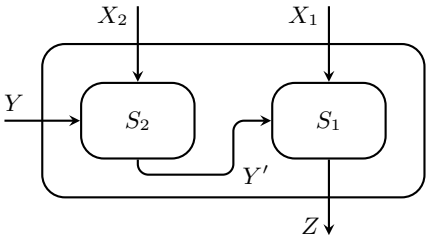

Fig. 2. The operation $S_1 \circ S_2$

See Figure 2 for an illustration. The composed sampler $S_1 \circ S_2$ first uses its randomness x_1 to get the sample $S_1(x_1) = \{S_1(x_1, y) : y \in \mathbb{F}_q\}$, and then uses its randomness x_2 to get the subsample $\{S_1(x_1, S_2(x_2, y)) : y \in \mathbb{F}_q\} \subseteq S_1(x_1)$. Intuitively, if S_1 and S_2 are good samplers then so is $S_1 \circ S_2$. This is indeed shown by [BR94, TSU06] and we formalize it as follows:

Lemma 7 ([BR94, TSU06]). *Let* $S_1 : \mathbb{F}_q^{n_1} \times \mathbb{F}_q^{d_1} \to \mathbb{F}_q^{d_0}$ *be an* (ϵ_1, δ_1) *sampler of degree* t_1 *and* $S_2 : \mathbb{F}_q^{n_2} \times \mathbb{F}_q^{d_2} \to \mathbb{F}_q^{d_1}$ *be an* (ϵ_2, δ_2) *sampler of degree* t_2*. Then* $S_1 \circ S_2 : (\mathbb{F}_q^{n_1} \times \mathbb{F}_q^{n_2}) \times \mathbb{F}_q^{d_2} \to \mathbb{F}_q^{d_0}$ *is an* $(\epsilon_1 + \epsilon_2, \delta_1 + \delta_2)$ *sampler of degree* $t_1 t_2$*.*

Proof. Fix an arbitrary subset $A \subseteq \mathbb{F}_q^{d_0}$. Define $B(x) = \left\{z \in \mathbb{F}_q^{d_2} : S_1(x, z) \in A\right\}$ for $x \in \mathbb{F}_q^{n_1}$. Pick $x_1 \leftarrow U_{n_1,q}$ and $x_2 \leftarrow U_{n_2,q}$. If $\left|\mu_{S_1 \circ S_2((x_1,x_2))}(A) - \mu(A)\right| > \epsilon_1 + \epsilon_2$ occurs, then either $|\mu_{S_1(x_1)}(A) - \mu(A)| > \epsilon_1$, or $|\mu_{S_1 \circ S_2((x_1,x_2))}(A) - \mu_{S_1(x_1)}(A)| > \epsilon_2$ occurs. Call the two events E_1 and E_2 respectively.

Note that E_1 occurs with probability at most δ_1 by the sampling property of S_1. Also note that

$$\begin{aligned}\mu_{S_1 \circ S_2((x_1,x_2))}(A) = \Pr_y[S_1(x_1, S_2(x_2, y)) \in A] &= \Pr_y[S_2(x_2, y) \in B(x_1)] \\ &= \mu_{S_2(x_2)}(B(x_1))\end{aligned}$$

whereas

$$\mu_{S_1(x_1)}(A) = \Pr_y[S_1(x_1, y) \in A] = \Pr_y[y \in B(x_1)] = \mu(B(x_1)).$$

So the probability that E_2 occurs is $\Pr_{x_1,x_2}[|\mu_{S_2(x_2)}(B(x_1)) - \mu(B(x_1))| > \epsilon_2]$ which is bounded by δ_2 by the sampling property of S_2. By the union bound, the event $\left|\mu_{S_1 \circ S_2((x_1,x_2))}(A) - \mu(A)\right| > \epsilon_1 + \epsilon_2$ occurs with probability at most $\delta_1 + \delta_2$, as desired.

Finally, we have $S_1 \circ S_2((X_1, X_2), Y)) = S_1(X_1, S_2(X_2, Y))$ which has degree $t_1 t_2$ in its variables X_1, X_2, Y since S_1 and S_2 have degree t_1 and t_2 respectively. □

4.3 Construction of the inner sampler

We use the basic curve samplers as the building blocks and apply the error reduction as well as iterated sampling repeatedly to obtain the inner sampler. The formal construction is as follows.

Definition 18 (inner sampler). *For $m \geq 1$, $\delta > 0$ and prime power q, pick $s = \lceil \log m \rceil$ and let $d_i = 2^{s-i}$ for $0 \leq i \leq s$. Let $n_i = 16^i$ for $0 \leq i \leq s-1$ and $n_s = 16^s + 20\left\lceil \log_q(1/\delta) \right\rceil$. Define $S_i : \mathbb{F}_q^{n_i d_i} \times \mathbb{F}_q^{d_i} \to \mathbb{F}_q^m$ for $0 \leq i \leq s$ as follows:*

- *$S_0 : \mathbb{F}_q \times \mathbb{F}_q^{d_0} \to \mathbb{F}_q^m$ projects (x, y) onto the first m coordinates of y.*
- *$S_i \stackrel{def}{=} \left(S_{i-1} \star \mathsf{RSCon}_{\frac{n_i}{4}, 2n_{i-1}, q^{d_i}}\right) \circ \mathsf{Curve}_{3, \frac{n_i}{4}, q^{d_i}}$ for $i = 1, \ldots, s$.*

Finally, let $\mathsf{InnerSamp}_{m,\delta,q} \stackrel{def}{=} S_s$.

Theorem 7. *For any $\epsilon, \delta > 0$, integer $m \geq 1$ and large enough prime power $q \geq \left(\frac{m \log(1/\delta)}{\epsilon}\right)^{O(1)}$, $\mathsf{InnerSamp}_{m,\delta,q} : \mathbb{F}_q^n \times \mathbb{F}_q \to \mathbb{F}_q^m$ is an (ϵ, δ) sampler of degree t where $n = O\left(m^{O(1)} + \log_q(1/\delta)\right)$ and $t = O\left(m^{O(\log m)} \log_q^2(1/\delta)\right)$.*

Proof. Let $\epsilon' = \frac{\epsilon}{2s}$ and d_i, n_i, S_i be as in Definition 18 for $0 \leq i \leq s$. We will prove that each S_i is an (ϵ_i, δ_i) sampler of degree t_i where $\epsilon_i = 2i\epsilon'$, $\delta_i = q^{-n_i d_i/20}$, and $t_i = \prod_{j=1}^{i} \left(\frac{n_j}{4}\right)^2$. The theorem follows by noting that $\mathsf{InnerSamp}_{m,\delta,q} = S_s$, $\epsilon = \epsilon_s$, $\delta \geq \delta_s$, and $t = t_s$.

Induct on i. The case $i = 0$ is trivial. Consider the case $i > 0$ and assume the claim holds for all $i' < i$. By the induction hypothesis, S_{i-1} is an $(\epsilon_{i-1}, \delta_{i-1})$ sampler of degree t_{i-1}.

Note that $\delta_{i-1} = q^{-n_{i-1}d_{i-1}/20} \leq q^{-0.01 n_{i-1} d_{i-1}}$. By Corollary 2, $S_{i-1} \star \mathsf{RSCon}_{\frac{n_i}{4}, 2n_{i-1}, q^{d_i}}$ is an $\left(\epsilon_{i-1} + \epsilon', q^{n_{i-1}d_{i-1} - (n_i/4)d_i}\right)$ sampler of degree $\frac{n_i}{4} \cdot t_{i-1}$. By Lemma 3, $\mathsf{Curve}_{3, \frac{n_i}{4}, q^{d_i}}$ is an $\left(\epsilon', q^{-n_i d_i/16}\right)$ sampler of degree $\frac{n_i}{4}$. Finally by Lemma 7, the function

$$S_i = \left(S_{i-1} \star \mathsf{RSCon}_{\frac{n_i}{4}, 2n_{i-1}, q^{d_i}}\right) \circ \mathsf{Curve}_{3, \frac{n_i}{4}, q^{d_i}}$$

is an $\left(\epsilon_{i-1} + 2\epsilon', q^{n_{i-1}d_{i-1} - (n_i/4)d_i} + q^{-n_i d_i/16}\right)$ sampler of degree $\left(\frac{n_i}{4}\right)^2 \cdot t_{i-1}$. It remains to check that

$$\epsilon_i = \epsilon_{i-1} + 2\epsilon', \quad \delta_i \geq q^{n_{i-1}d_{i-1} - (n_i/4)d_i} + q^{-n_i d_i/16} \quad \text{and} \quad t_i = \left(\frac{n_i}{4}\right)^2 \cdot t_{i-1}.$$

which hold by the choices of parameters. □

5 Putting it together

We compose the outer sampler and the inner sampler to get the desired curve sampler.

Definition 19. *For $m \geq 1$, $\delta > 0$ and prime power q, define*

$$\mathsf{Samp}_{m,\delta,q} \stackrel{def}{=} \mathsf{OuterSamp}_{m,\delta/2,q} \circ \mathsf{InnerSamp}_{d,\delta/2,q}.$$

Theorem 8 (Theorem 1 restated). *For any $\epsilon, \delta > 0$, integer $m \geq 1$ and sufficiently large prime power $q \geq \left(\frac{m \log(1/\delta)}{\epsilon}\right)^{O(1)}$, the function $\mathsf{Samp}_{m,\delta,q} : \mathbb{F}_q^n \times \mathbb{F}_q \to \mathbb{F}_q^m$ is an (ϵ, δ) sampler of degree t where $n = O\left(m + \log_q(1/\delta)\right)$ and $t = \left(m \log_q(1/\delta)\right)^{O(1)}$. In particular, $\mathsf{Samp}_{m,\delta,q}$ is an (ϵ, δ) degree-t curve sampler.*

Proof. By Theorem 5, $\mathsf{OuterSamp}_{m,\delta/2,q} : \mathbb{F}_q^{n_1} \times \mathbb{F}_q^d \to \mathbb{F}_q^m$ is an $(\epsilon/2, \delta/2)$ sampler of degree t_1 where $d = O(\log m)$, $n_1 = O\left(m + \log_q(1/\delta)\right)$ and $t_1 = O\left(m^2 + m \log_q(1/\delta)\right)$.

By Theorem 7, $\mathsf{InnerSamp}_{d,\delta/2,q} : \mathbb{F}_q^{n_2} \times \mathbb{F}_q \to \mathbb{F}_q^d$ is an $(\epsilon/2, \delta/2)$ sampler of degree t_2 with parameters $n_2 = O\left((\log m)^{O(1)} + \log_q(1/\delta)\right)$ and $t_2 = O\left((\log m)^{O(\log\log m)} \log_q^2(1/\delta)\right)$.

Finally, Lemma 7 implies that $\mathsf{Samp}_{m,\delta,q}$ is an (ϵ, δ) sampler of degree t with $n = n_1 + n_2 = O\left(m + \log_q(1/\delta)\right)$ and $t = t_1 t_2 = \left(m \log_q(1/\delta)\right)^{O(1)}$. A fortiori, it is a degree-t curve sampler since the degree of $\mathsf{Samp}_{m,\delta,q}(x, \cdot)$ is bounded by the degree of $\mathsf{Samp}_{m,\delta,q}$ for all $x \in \mathbb{F}_q^n$. □

Acknowledgements

The author is grateful to Chris Umans for his support and many helpful discussions.

References

[ALM+98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *J. ACM*, 45(3):501–555, May 1998.

[AS98] Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: a new characterization of NP. *J. ACM*, 45(1):70–122, January 1998.

[BR94] M. Bellare and J. Rompel. Randomness-efficient oblivious sampling. In *Proceedings of the 35th Annual Symposium on Foundations of Computer Science*, SFCS '94, pages 276–287, Washington, DC, USA, 1994. IEEE Computer Society.

[BSSVW03] Eli Ben-Sasson, Madhu Sudan, Salil Vadhan, and Avi Wigderson. Randomness-efficient low degree tests and short PCPs via epsilon-biased sets. In *Proceedings of the 35th Annual ACM Symposium on Theory of Computing*, STOC '03, pages 612–621, New York, NY, USA, 2003. ACM.

[CG88] Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM J. Comput.*, 17:230–261, April 1988.

[CG89] B. Chor and O. Goldreich. On the power of two-point based sampling. *J. Complex.*, 5(1):96–106, April 1989.

[Din07] Irit Dinur. The PCP theorem by gap amplification. *J. ACM*, 54(3), June 2007.

[GI01] V. Guruswami and P. Indyk. Expander-based constructions of efficiently decodable codes. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science*, FOCS '01, pages 658–, Washington, DC, USA, 2001. IEEE Computer Society.

[Gil98] David Gillman. A Chernoff bound for random walks on expander graphs. *SIAM J. Comput.*, 27(4):1203–1220, August 1998.

[Gol11] Oded Goldreich. A sample of samplers: A computational perspective on sampling. In *Studies in Complexity and Cryptography*, volume 6650 of *Lecture Notes in Computer Science*, pages 302–332. Springer Berlin Heidelberg, 2011.

[GR08] V. Guruswami and A. Rudra. Explicit codes achieving list decoding capacity: Error-correction with optimal redundancy. *IEEE Trans. Inf. Theor.*, 54(1):135–150, January 2008.

[GUV09] Venkatesan Guruswami, Christopher Umans, and Salil Vadhan. Unbalanced expanders and randomness extractors from Parvaresh–Vardy codes. *J. ACM*, 56:20:1–20:34, July 2009.

[MR08] Dana Moshkovitz and Ran Raz. Sub-constant error low degree test of almost-linear size. *SIAM J. Comput.*, 38(1):140–180, March 2008.

[NZ96] Noam Nisan and David Zuckerman. Randomness is linear in space. *J. Comput. Syst. Sci.*, 52:43–52, February 1996.

[Rei08] Omer Reingold. Undirected connectivity in log-space. *J. ACM*, 55(4):17:1–17:24, September 2008.

[RTS00] Jaikumar Radhakrishnan and Amnon Ta-Shma. Bounds for dispersers, extractors, and depth-two superconcentrators. *SIAM J. Discret. Math.*, 13(1):2–24, January 2000.

[STV01] Madhu Sudan, Luca Trevisan, and Salil Vadhan. Pseudorandom generators without the XOR lemma. *J. Comput. Syst. Sci.*, 62(2):236–266, March 2001.

[SU05] Ronen Shaltiel and Christopher Umans. Simple extractors for all min-entropies and a new pseudorandom generator. *J. ACM*, 52(2):172–216, March 2005.

[SU06] Ronen Shaltiel and Christopher Umans. Pseudorandomness for approximate counting and sampling. *Comput. Complex.*, 15(4):298–341, December 2006.

[SZ99] Aravind Srinivasan and David Zuckerman. Computing with very weak random sources. *SIAM J. Comput.*, 28:1433–1459, March 1999.

[TSU06] Amnon Ta-Shma and Christopher Umans. Better lossless condensers through derandomized curve samplers. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, pages 177–186, Washington, DC, USA, 2006. IEEE Computer Society.

[Uma03] Christopher Umans. Pseudo-random generators for all hardnesses. *J. Comput. Syst. Sci.*, 67(2):419–440, September 2003.

[Zuc97] David Zuckerman. Randomness-optimal oblivious sampling. In *Proceedings of the Workshop on Randomized Algorithms and Computation*, pages 345–367, New York, NY, USA, 1997. John Wiley & Sons, Inc.

A Lower bounds on the degree of sampled curves

We present the following lower bound on the degree of curves sampled by a curve sampler:

Theorem 9. *Let $S : \mathcal{N} \times \mathbb{F}_q \to \mathbb{F}_q^m$ be an (ϵ, δ) degree-t curve sampler where $m \geq 2$, $\epsilon < 1/2$ and $\delta < 1$. Then $t = \Omega\left(\log_q(1/\delta) + 1\right)$.*

Proof. Clearly $t \geq 1$. Suppose $S = (S_1, \ldots, S_m)$ and define $S' = (S_1, S_2)$. Let $\mathcal{C}$ be the set of curves of degree at most t in $\mathbb{F}_q^2$. Then $|\mathcal{C}| = q^{2(t+1)}$. Consider the map $\tau : \mathcal{N} \to \mathcal{C}$ that sends x to $S'(x, \cdot)$. We can pick $k = \lfloor q/2 \rfloor$ curves $C_1, \ldots, C_k \in \mathcal{C}$ such that the union of their preimages

$$B \stackrel{\text{def}}{=} \bigcup_{i=1}^{k} \tau^{-1}(C_i) = \bigcup_{i=1}^{k} \{x : S'(x, \cdot) = C_i\}$$

has size at least $\frac{k|\mathcal{N}|}{|\mathcal{C}|} = \frac{k|\mathcal{N}|}{q^{2(t+1)}}$.

Define $A \subseteq \mathbb{F}_q^m$ by

$$A \stackrel{\text{def}}{=} \{C_i(y) : i \in [k],\ y \in \mathbb{F}_q\} \times \mathbb{F}_q^{m-2},$$

i.e., let A be the set of points in $\mathbb{F}_q^m$ whose first two coordinates are on at least one curve C_i. We have $|A| \leq kq^{m-1}$ and hence $\mu(A) \leq k/q \leq 1/2 < 1 - \epsilon$. On the other hand, it follows from the definition of A that we have $S(x, y) \in A$ for all $x \in B$ and $y \in \mathbb{F}_q$. So $\mu_{S(x)}(A) = 1$ for all $x \in B$. Then

$$\delta \geq \Pr\left[|\mu_{S(x)}(A) - \mu(A)| > \epsilon\right] \geq \frac{|B|}{|\mathcal{N}|} \geq \frac{k}{q^{2(t+1)}}$$

and hence $t \geq \max\left\{1, \frac{1}{2}\log_q(k/\delta) - 1\right\} = \Omega\left(\log_q(1/\delta) + 1\right)$. □

We remark that the condition $m \geq 2$ is necessary in Theorem 9 since when $m = 1$, the sampler S with $S(x, y) = y$ for all $x \in \mathcal{N}$ and $y \in \mathbb{F}_q$ is a $(0, 0)$ degree-1 curve sampler.