# Randomness extractors

Vsevolod Nagibin, Daniil Musatov

Moscow Institute of Physics and Technology

April 22, 2025

# Reminder

## Min entropy

Let $X$ be a random variable. The *min entropy* of $X$, denoted by $H_\infty(X)$ is $\sup\{k \in \mathbb{R} | \forall x \mathbb{P}[X = x] \leq 2^{-k}\} = \inf\{-\log_2 \mathbb{P}[X = x]\}$
If $X$ is a distribution over $\{0,1\}^n$ with $H_\infty(X) \geq k$ then it is called an $(n,k)$-*source*

## Statistical distance

Let $X$ and $Y$ be two distributions over domain $\Omega$. The *statistical distance* between $X$ and $Y$, denoted by $\Delta(X,Y)$, is equal to

$$\max_{S \subset \Omega} |\mathbb{P}[X \in S] - \mathbb{P}[Y \in S]| = \frac{1}{2} \sum_{\omega \in \Omega} |\mathbb{P}[X = \omega] - \mathbb{P}[Y = \omega]|$$

Two distributions $X$ and $Y$ are called $\varepsilon$-*close* if $\Delta(X,Y) \leq \varepsilon$.

# Reminder

## Randomness extractor

A function $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a $(k, \varepsilon)$-*extractor* if for any $(n, k)$-source $X$, the distribution $\mathrm{Ext}(X, U_d)$ is $\varepsilon$-close to $U_m$.

By probabilistic method, can be shown the existence of an almost optimal extractor.

## Theorem

$\forall k \leq n \in \mathbb{N}, \varepsilon > 0 \exists (k, \varepsilon)$-*extractor with* $m = k + d - 2log_2(\frac{1}{\varepsilon}) - O(1)$ *and* $d = \log_2(n - k) + 2log_2(\frac{1}{\varepsilon}) + O(1)$.

# Constructions

Important extractor construction is obtained by using hash functions.

**Leftover hash lemma**

If $\mathcal{H} = \{h : \{0,1\}^n \to \{0,1\}^{k-2log_2(\frac{1}{\varepsilon})}\}$ is a pairwise independent family of hash functions, then $\mathrm{Ext}(x,h) = (h, h(x))$ is a $(k,\varepsilon)$-extractor.

In this construction we get $m = k + d - 2\log_2(\frac{1}{\varepsilon})$, but the number of bits required to choose a hash function from a pairwise independent family is at least $n$. So, $d \geq n$ (and such families of hash functions with $d = n$ exist).

# Condenser

## Condenser

A function $\mathrm{Con} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a $k \to_\varepsilon k'$ *condenser* if for every $k$-source $X$ on $\{0,1\}^n$, $\mathrm{Con}(X, U_d)$ is $\varepsilon$-close to some $k'$-source. $\mathrm{Con}$ is *lossless* if $k' = k + d$.

Condensers can be viewed from graph perspective.

## Lemma

*Let $n, d, m \in \mathbb{N}$, $K = 2^k \in \mathbb{N}$ and $\varepsilon > 0$. A function $\mathrm{Con} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a $k \to_\varepsilon k + d$ lossless condenser if and only if the corresponding bipartite multigraph $G = ([2^n], [2^m], E)$ of left degree $D = 2^d$ is an $(= K, (1-\varepsilon)D)$ vertex expander.*

Here bipartite multigraph $G = ([N], [M], E)$ of left degree $D$ is an $(= K, \gamma)$ vertex expander if and only if $\forall S \subset [N]\ |S| = K \implies |\{u | \exists v \in S\ (v, u) \in E\}| \geq \gamma |S|$

# Condenser construction

A useful construction of lossless condenser is based on Parvaresh-Vardy codes [PV05]

**Theorem**

$\forall n \in \mathbb{N}, k_{\max} \leq n, \varepsilon > 0$ *and* $\alpha \in (0, \frac{\log_2(nk_{\max}/\varepsilon)}{\log_2(\log_2(nk_{\max}/\varepsilon)})$ *there is an explicit function* $\mathrm{Con} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *with* $d = (1 + \frac{1}{\alpha}) \cdot (\log_2 n + \log_2 k_{\max} + \log_2 \frac{1}{\varepsilon}) + O(1)$ *and* $m \leq 2d + (1 + \alpha)k_{\max}$ *such that for all* $k \leq k_{\max}$, $\mathrm{Con}$ *is a* $k \to_\varepsilon k + d$ *condenser.*

# Block sources

## Block source

A random variable $X = (X_1, X_2, \cdots X_t)$ is a $(k_1, k_2, \cdots k_t)$ *block source* if for every $x_1, \cdots x_{i-1}$, $X_i|_{X_1=x_1 \ldots X_{i-1}=x_{i-1}}$ is a $k_i$-source.

Block sources allow us to extract bits from each block basically independently.

## Lemma

Let $\mathrm{Ext}_1 : \{0,1\}^{n_1} \times \{0,1\}^{d_1} \to \{0,1\}^{m_1}$ be $(k_1, \varepsilon_1)$-extractor, and $\mathrm{Ext}_2 : \{0,1\}^{n_2} \times \{0,1\}^{d_2} \to \{0,1\}^{m_2}$ be $(k_2, \varepsilon_2)$-extractor with $m_2 \geq d_1$. Let $\mathrm{Ext}'((x_1, x_2), y_2) = (\mathrm{Ext}_1(x_1, y_1), z_2)$ where $(y_1, z_2) := Ext(x_2, y_2)$.
Then for every $(k_1, k_2)$ block source $X = (X_1, X_2)$ in $\{0,1\}^{n_1} \times \{0,1\}^{n_2}$ it holds than $\mathrm{Ext}'(X, U_{d_2})$ is $(\varepsilon_1 + \varepsilon_2)$-close to $U_{m_1} \times U_{m_2 - d_1}$.

The lemma can be extended to extracting from many blocks.

# Block sources

Actually, high min-entropy sources are block sources.

**Lemma**

*If $X$ is a $(n, n-\Delta)$-source and $X = (X_1, X_2)$ is a partition of $X$ into blocks of lengths $n_1$ and $n_2$ then $(X_1, X_2)$ is $\varepsilon$-close to some $(n_1 - \Delta, n_2 - \Delta - \log_2 \frac{1}{\varepsilon})$ block sources.*

Another lemma, which helps extract the remaining min-entropy.

**Lemma**

*Let $\mathrm{Ext}_1 : \{0,1\}^n \times \{0,1\}^{d_1} \to \{0,1\}^{m_1}$ be $(k_1, \varepsilon_1)$-extractor, and $\mathrm{Ext}_2 : \{0,1\}^n \times \{0,1\}^{d_2} \to \{0,1\}^{m_2}$ be $(k_2, \varepsilon_2)$-extractor with $k_2 \leq k_1 - m_1 - s$. Then $E'\{0,1\}^n \times \{0,1\}^{d_1+d_2} \to \{0,1\}^{m_1+m_2}$ defined by $E'(x,(y_1,y_2)) = (E_1(x,y_1), E_2(x,y_2))$ is a $(k_1, \frac{1}{1-2^{-s}} \cdot \varepsilon_1 + \varepsilon_2)$-extractor*

# Ideas of best construction

To construct the optimal up to constant factor construction ([GUV09]), we will firstly construct the extractors with output length $m \geq \frac{k}{2}$ in the following steps:

1. Applies the condenser to get a source $X'$ which is $\varepsilon_0$-close to a $k$-source of length $\frac{9}{8}k + O(\log n + \log \frac{1}{\varepsilon_0})$.
2. Divides $X'$ into two halves $(X_1, X_2)$. Now the source $(X_1, X_2)$ is $2\varepsilon_0$-close to a $w \times k'$ block source with $k' = \frac{k}{2} - \frac{k}{8} - O(\log \frac{n}{\varepsilon_0})$.
3. Apply block source extraction with $\mathrm{Ext}_1$ (which is recursively constructed, uses $d$ random bits and extracts $\geq \frac{k'}{2} \geq \frac{k}{6}$ bits) and $\mathrm{Ext}_2$ (which uses only $\frac{d}{8}$ random bits and gives $d$ output bits).
4. The steps described above extracted only $\frac{k}{6}$ bits. Applying the last lemma with this construction, repeated 4 times, we can extract $\geq \frac{k}{2}$ bits.

**Theorem**

*Let constant $\alpha > 0$. $\forall n \in \mathbb{N}, k \in [0, n]$ and $\varepsilon > 0$ there is an explicit $(k, \varepsilon)$-extractor $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ with $m \geq (1 - \alpha)k$ and $d = O(\log \frac{n}{\varepsilon})$*

# Open problems

## Open problem 1

Give an explicit construction of $(k, 0.01)$-extractor $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ with seed length $d = O(\log n)$ and output length $m = k + d - O(1)$.

## Open problem 2

Give an explicit construction of $(k, 0.01)$-extractor $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ with seed length $d = \log n + O(1)$ and output length $m = \Omega(k)$.

Based on survey by S. Vadhan ([Vad12]).

# Questions

# References



Venkatesan Guruswami, Christopher Umans, and Salil Vadhan.
Unbalanced expanders and randomness extractors from parvaresh–vardy codes.
*J. ACM*, 56(4), July 2009.


F. Parvaresh and A. Vardy.
Correcting errors beyond the guruswami-sudan radius in polynomial time.
In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05)*, pages 285–294, 2005.


Salil P. Vadhan.
Pseudorandomness.
*Foundations and Trends® in Theoretical Computer Science*, 7(1–3):1–336, 2012.