Randomness extractors

Vsevolod Nagibin, Daniil Musatov

Moscow Institute of Physics and Technology

March 18, 2025

Introduction

Generally, probabilistic algorithms require uniform and independent random bits. But real-world sources of randomness very rarely, if ever, have such qualities. That is why arises the question whether we can use real-world "weakly random" to run probabilistic algorithms.
The randomness extractor is a function that transforms a "weak" random source into an almost uniform distribution. In the 1950s, the problem of randomness extraction was first considered by von Neumann, who wanted to extract randomness from biased random coins. Later, the problem was generalized. In 1985 Chor and Goldreich introduced the notion of min entropy [CG85].

Definitions

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 Ω. 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 ε-*close* if $\Delta(X,Y) \leq \varepsilon$.

Definitions

Randomness extractor

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

Why do we need extra randomness?

Lemma

Let $k \leq n - 1$. $\forall \mathrm{Ext} : \{0,1\}^n \to \{0,1\}^m \exists (n, k)$*-source* X *such that first bit of* $\mathrm{Ext}(X)$ *is constant.*

From this lemma, the statistical distance between $\mathrm{Ext}(X)$ and U_m is at least $\frac{1}{2}$ when Ext is deterministic.

Bounds

The goal in constructing a good extractor is to maximize the output length m, minimize the seed length d and minimize the error ε at the same time.
Informally speaking, (n, k)-source contains k hidden bits of information, so the extractor, taking extra d real random bits with output length $m = k + d$ is optimal. Existence of an almost optimal (in this sense) extractor can be shown by probabilistic method.

Theorem

Consider ε fixed constant. $\forall k \leq n \exists (k, \varepsilon)$-extractor with $m = k + d - O(1)$ and $d = \log_2 n + O(1)$.

Bounds

In 1997 the following bounds on the seed length and output length were shown [RTS00].

Theorem

If $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *is a* (k, ε)*-extractor, then* $d \geq \log_2(n-k) + 2\log_2(\frac{1}{\varepsilon}) - O(1)$ *and* $d + k - m \geq 2\log_2 \frac{1}{\varepsilon} - O(1)$.

For the purpose of applications, of course, explicit constructions are needed. One of the known constructions has both seed length and output length within a constant seed factor of the optimal implicit extractor [CJLW03].

Theorem

$\forall \alpha, \varepsilon > 0 \forall n, k \leq n \exists$ *explicit* (k, ε)*-extractor* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *with* $d = O(\log n)$ *and* $m = (1-\alpha) \cdot k$.

Work Plan

There are many different constructions of extractors that have appeared in the recent years. But still there is a gap not only between known lower bounds and explicit constructions, but also a gap between known explicit and implicit constructions. Advanced extractors are complicated and their constructions based on other constructions. So the current goals are to study the best known constructions, try to simplify and improve them.
Also, extractors are tied with pseudo random generators [Tre01] and list-decodable error-correcting codes [Vad07]. So, the construction of any of these objects is often helpful in constructing the other two.

Questions

References

B. Chor and O. Goldreich.
Unbiased bits from sources of weak randomness and probabilistic communication complexity.
SIAM Journal on Computing, 17, 1985.

S. Vadhan C.-J. Lu, O. Reingold and W. Wigderson.
Extractors: optimal up to constant factors.
STOC, pages 602–611, 2003.

J. Radhakrishnan and A. Ta-Shma.
Bounds for dispersers, extractors, and depth-two superconcentrators.
SIAM J. Discrete Mathematics, 13:2–24, 2000.

L. Trevisan.
Extractors and pseudorandom generators.
Journal of the ACM, 48(4):860–879, 2001.

S. Vadhan.
The unified theory of pseudorandomness.
SIGACT News, 38, 2007.