Paper Reading Report: SAT Reduces to Minimum Circuit Size Problem with a Random Oracle

Tingqiang Xu, Yusi Chen, Jingyi Lyu

April 22, 2025

SAT Reduces to Minimum Circuit Size Problem with a Random Oracle [Ila23] is a paper written by Rahul Ilango, which won the **Best Student Paper Award** at FOCS '23. We will briefly summarize contribution, techniques and impact of this paper. We also give some of our thoughts.

1 Contribution

This paper provided a reduction from approximating τ-frequency set cover problem (proved NP-hard in [DGKR03]) to approximating minimum circuit size problem (with random oracle access, denoted $\mathsf{MCSP}^{\mathcal{O}}$) and time-bounded Kolmogorov complexity problem (with random oracle access, denoted $\mathsf{K}^{t,\mathcal{O}}$). The reduction is uniform and deterministic with access to random oracle $\mathcal{O}$, runs in polynomial time and fails with negligible probability over choice of $\mathcal{O}$.

Specifically, the author proved that if the reduction outputs $|x|$, then hardness result of [DGKR03] implies

- Promise problem of $\mathsf{MCSP}^{\mathcal{O}}$ with gap $\Omega(\frac{|x|}{\log|x|})$ is NP-hard;
- Promise problem of $\mathsf{K}^{t,\mathcal{O}}$ with gap $\Omega(|x|)$ and a polynomial slowdown on time bound is NP-hard.

However, notice that $\mathsf{K}^{t,\mathcal{O}}(x)$ must be bounded by $|x|+O(1)$ when $t \geq \mathsf{poly}(|x|)$, and any function $f:\{0,1\}^n \to \{0,1\}$ holds circuit complexity $O(\frac{2^n}{n})$, which is $O(\frac{|x|}{\log|x|})$ when x is truth table of f. Therefore, this paper proved additive hardness of approximation within a constant factor of optimal, under Random Oracle Model.

At the cost of introducing non-uniformity into reduction (which removes requirement of oracle access, by hard-coding randomness into advice), one can re-state the result as

$$\mathsf{NP} \subset \mathsf{P}^{\mathsf{MCSP}^{\mathcal{O}}}/\mathsf{poly} \quad \text{or} \quad \mathsf{NP} \subset \mathsf{P}^{\mathsf{K}^{t,\mathcal{O}}}/\mathsf{poly}.$$

The author informally claimed that this paper provides a strong evidence that $\mathsf{MCSP}^{\mathcal{O}}$ and K^t are NP-Complete, based on following observations:

- Previous counterexample on random oracle hypothesis (Omnipresent possibility of instantiating random oracle in complexity statement) mainly focuses on class equality, like $\mathsf{IP}^{\mathcal{O}} \neq \mathsf{PSPACE}^{\mathcal{O}}$, while this paper proves inclusion relationship;
- Sufficiently pseudorandom cryptographic functions exist in P, which are secure against computationally bounded adversaries. In the setting of this paper, the adversary is mainly bounded-size circuits, whose computational power is limited. In cryptographic works, random oracle can usually be heuristically instantiated as concrete hash functions.

2 Reduction

The reduction by [Ila23] is stated below.

Input: An instance $S_1, \cdots, S_m \subset [n]$ of Gap τ-Frequency Set Cover for some power-of-two n.

Parameter: Secret key length $\lambda(n)$ computable in time $O(\log \lambda(n))$.

Oracle: For all powers-of-two n, we have random oracle[1] $\mathcal{O}_n : [n] \times \{0,1\}^\lambda \times \{0,1\}^{2\lambda} \to \{0,1\}^\lambda$.
Algorithm:

1. Pick $sk_1, \cdots, sk_m \in \{0,1\}^\lambda$ uniformly at random.
2. For all $i \in [n]$, pick $v_i \in \{0,1\}^\lambda$ uniformly at random.
3. For all $i \in [n]$ and $k \in [\tau]$, let $c_{i,k}$ be a uniformly random element of the set $\{c \in \{0,1\}^{2\lambda} : \mathcal{O}_n(i, sk_j, c) = v_i\}$, where j is the index of the k-th set containing i.[2]
4. Output the $4\tau n\lambda$-bit truth table of function $f : [n] \times [\tau] \times \{0,1\} \times [2\lambda] \to \{0,1\}$ given by

$$f(i,k,b,d) = \begin{cases} d\text{-th bit of } c_{i,k} & b = 0 \\ d\text{-th bit of } v_i & b = 1 \end{cases}$$

This is a **deterministic** reduction since all randomness come from a general random oracle, invoked in some specific way.

Understand the variables. The author interpreted variables in reduction algorithms as follow, which could be essential to understand the intuition behind:

- sk_j: **secret key**;
- v_i: **message**;
- $c_{i,k}$: a random "**encoding**" of an "**encryption**" of message v_i, using the secret key sk_j.

One might hope the "optimal" way to compute f is to memorize all $c_{i,k}$ as well as all secret keys sk_j in an **optimal set cover**, then decrypt to find values of v_i.

3 Intuition

Symmetry of Information. The idea of symmetry of information arises from information theory: we have $H(XY) = H(Y) + H(X|Y)$. Also for Kolmogorov complexity we have $\mathsf{K}^\infty(xy) \approx \mathsf{K}^\infty(y) + \mathsf{K}^\infty(x|y)$. [ZL70]

Approximating conditional version of K^t is known **NP-hard** [HIR23], then we expect such symmetry also holds for K^t, i.e.

$$\mathsf{K}^t(xy) \approx \mathsf{K}^t(y) + \mathsf{K}^t(x|y).$$

Then we can estimate $\mathsf{K}^t(x|y)$ by unconditional K^t, which proves NP-hardness of approximating K^t. However intuitively, when one-way functions exist, consider the case $y = f(x)$ where f is OWF, we can approximately think

- $\mathsf{K}^t(y) \approx |y|$, which is most of the case (up to approximation of additive constant);
- $\mathsf{K}^t(x|y) \approx \mathsf{K}^t(x) \approx |x|$, since OWF cannot leak any information to compute x from $y = f(x)$.

However, to compute xy, we can hard-code f and x, then compute y from $f(x)$. This in general costs $|x|$. If $|x| = |y|$, then the symmetry of information in this case fails.

[1] This is easy to produce given a general random oracle $\{0,1\}^* \to \{0,1\}$.
[2] By rejection sampling.

Pseudo Symmetry of Information. [Ila23] stated a generalized idea. Assume $\tilde{y}$ is some encoding of y that one can easily decode to y. If there is some encoding scheme such that

$$\mathsf{K}^t(x\tilde{y}) \approx \mathsf{K}^t(x|y) + \mathsf{K}^t(\tilde{y}),$$

then this scheme holds **pseudo symmetry of information**.

Intuitively, we need the encoding to have enough **randomness** that diminishes the latent relationship between x, y (for example, an OWF), so $\tilde{y}$ cannot be obtained from x easily. In fact, we expect the optimal way to compute $\tilde{y}$ is to hard-code it. We would also like the encoding holds **incompressibility**, so we need to hard-code $\tilde{y}$ (nearly) completely, rather than in a compressed fashion.

Consider $\mathcal{O} : \{0,1\}^{2\lambda} \to \{0,1\}^{\lambda}$ as a random oracle. Divide y into $r = n/\lambda$ blocks $y_1, \cdots, y_r$, for each y_i find a random $\tilde{y}_i$ satisfying $\mathcal{O}(\tilde{y}_i) = y_i$. Then let $\tilde{y}$ be the concatenation of all $\tilde{y}_i$. Since we are limited to t oracle queries, we can only compress $O(\log t)$ bits for each block (in a sense that random oracle is information theoretical safe on each bit of its preimage information, so we need 2^k queries to compress k bits.). Therefore, one can (informally) conjecture

$$\mathsf{K}^t(x\tilde{y}) \geq n + \frac{n}{\lambda}(\lambda - O(\log t)).$$

When λ overwhelms $\log t$ (for example, $\lambda = \gamma \log t$ for sufficiently large γ), we arrive at $\mathsf{K}^t(x\tilde{y}) \geq (3-\epsilon)n$, so we can think pseudo symmetry is achieved by this encoding scheme.

Intuition in the reduction. We can see how this idea is used in the reduction construction. $c_{i,k}$ is a random encoding of encryption on messages, which means there are two steps establishing the soundness of f:

- Encryption step: Use sk to encrypt v to ciphertext c'. Then $\mathsf{K}^t(v|c')$ is large, unless you hard-code sk. This step couples the size of the Set Cover into the length of hard-coding of sk, which establishes a relationship between two problems;
- Encoding step: Encode c' to c by random oracle preimage. Pseudo symmetry of information established, thus

$$\mathsf{K}^t(f) = \mathsf{K}^t(vc) \approx \mathsf{K}^t(v|c') + \mathsf{K}^t(c).$$

So the optimal way of computing f is to hard-code $sk_j, c_{i,k}$ over optimal set cover and decode to v_i.

4 Impact

This paper presents significant theoretical advancements with broad implications for complexity theory and cryptographic applications.

Theoretical Significance. By demonstrating an $\mathcal{O}$-oracle reduction from the τ-frequency set cover problem to $\mathsf{MCSP}^{\mathcal{O}}$ and $\mathsf{K}^{t,\mathcal{O}}$, the paper establishes that these problems are likely to be NP-hard to approximate. This adds a new layer to our understanding of the intrinsic difficulty of these computational problems.

The use of the random oracle model is crucial in this paper. The findings support the hypothesis that certain complexity classes and cryptographic assumptions hold true relative to random oracles, reinforcing the utility of random oracles in theoretical computer science.

Cryptographic Implications. The reduction techniques and the reliance on pseudorandom functions that behave similarly to random oracles indicate potential pathways for constructing cryptographic primitives that are secure against bounded adversaries. This has practical implications for designing secure encryption schemes, hash functions, and other cryptographic protocols.

The results contribute to the field of meta-complexity by providing evidence that MCSP and K^t might be NP-complete. This aligns with the broader goal of understanding the complexity of problems related to circuit minimization and information theory, which are fundamental in both theoretical and applied cryptography.

Practical Considerations. The reductions proposed in this paper can inform the design of algorithms for practical problems in coding theory, data compression, and secure computation. By understanding the limitations and hardness of these problems, researchers can better assess the feasibility and efficiency of various algorithmic approaches.

The techniques and results presented open up several avenues for future research, including exploring whether similar reductions can be applied to other complexity classes or cryptographic problems and investigating the practical instantiation of random oracles in real-world systems.

5 Discussion

The random oracle plays an important role both intuitively and technically. It follows some useful properties:

- **No partial information on preimage.** One cannot distinguish the distribution of suffix of r, conditioned on knowing $y = \mathcal{O}(r)$ and a prefix of r, or only knowing $y = \mathcal{O}(r)$. This mainly supports incompressibility of random oracle encoding in pseudo symmetry of information.
- **Pairwise independence.** This supports the query length argument in formal proof.
- **Collision-resistance.** So you cannot cheat by potentially producing some collisions to bypass the secret key requirement.

If we can utilize some real-world hash function with these properties (even just approximately), maybe instantiation of $\mathcal{O}$ is possible.

References

[DGKR03] Irit Dinur, Venkatesan Guruswami, Subhash Khot, and Oded Regev. A new multilayered pcp and the hardness of hypergraph vertex cover, 2003.

[HIR23] Yizhi Huang, Rahul Ilango, and Hanlin Ren. Np-hardness of approximating meta-complexity: A cryptographic approach, 2023.

[Ila23] Rahul Ilango. Sat reduces to the minimum circuit size problem with a random oracle, 2023.

[ZL70] Alexander K Zvonkin and Leonid A Levin. The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms, 1970.