Information and Computation 261 (2018) 175–201

Contents lists available at ScienceDirect

Information and Computation

www.elsevier.com/locate/yinco

Proofs of proximity for context-free languages and read-once branching programs ☆

Oded Goldreich [a], Tom Gur [b,*], Ron D. Rothblum [c]

[a] *Weizmann Institute, Israel*
[b] *UC Berkeley, CA, USA*
[c] *MIT and Northeastern University, MA, USA*

ARTICLE INFO

Article history:
Received 27 May 2015
Available online 9 February 2018

Keywords:
Property testing
Probabilistic proof systems
Interactive proofs

ABSTRACT

Proofs of proximity are proof systems wherein the verifier queries a sublinear number of bits, and soundness only asserts that inputs that are far from valid will be rejected. In their minimal form, called *MA proofs of proximity* ($\mathcal{MAP}$), the verifier receives, in addition to query access to the input, also free access to a short (sublinear) proof. A more general notion is that of *interactive proofs of proximity* ($\mathcal{IPP}$), wherein the verifier is allowed to interact with an omniscient, yet untrusted prover.

We construct proofs of proximity for two natural classes of properties: (1) context-free languages, and (2) languages accepted by small read-once branching programs. Our main results are:

1. $\mathcal{MAP}$s for these two classes, in which, for inputs of length n, both the verifier's query complexity and the length of the $\mathcal{MAP}$ proof are $\widetilde{O}(\sqrt{n})$.
2. $\mathcal{IPP}$s for the same two classes with constant query complexity, poly-logarithmic communication complexity, and logarithmically many rounds of interaction.

1. Introduction

The field of property testing, initiated by Rubinfeld and Sudan [1] and Goldreich, Goldwasser and Ron [2], studies a computational model that consists of probabilistic algorithms, called *testers*, that need to decide whether a given object has a certain global property or is far (say, in Hamming distance) from all objects that have the property, based only on a local view of the object.

A line of work [3–9] has considered the question of designing *proof systems* within the property testing model. The minimal type of such a proof system, which was recently studied by Gur and Rothblum [7], augments the property testing framework by replacing the tester with a *verifier* that receives, in addition to oracle access to the input, also free access to

☆ This research was partially supported by the Israel Science Foundation (grant No. 671/13). The research was conducted while the second and third authors were students at the Weizmann Institute.

* Corresponding author.
E-mail addresses: oded.goldreich@weizmann.ac.il (O. Goldreich), tom.gur@berkeley.edu (T. Gur), ronr@csail.mit.edu (R.D. Rothblum).

https://doi.org/10.1016/j.ic.2018.02.003
0890-5401/

an explicitly given short (i.e., sub-linear length) proof. The guarantee is that for inputs that have the property there exists a proof that makes the verifier accept with high probability, whereas, for inputs that are far from the property, the verifier will reject *every* alleged proof with high probability. These proof systems can be thought of as the $\mathcal{NP}$ (or more accurately $\mathcal{MA}$) analogue of *property testing*, and are called *Merlin–Arthur proofs of proximity ($\mathcal{MAP}$)*.[1]

A more general notion was considered by Rothblum, Vadhan and Wigderson [6] (prior to [7]). Their proof system, which can be thought of as the $\mathcal{IP}$ analogue of property testing, consists of an all powerful (but untrusted) prover who interacts with a verifier that only has oracle access to the input x. The prover tries to convince the verifier that x has a particular property Π. Here, the guarantee is that for inputs in Π, there exists a prover strategy that will make the verifier accept with high probability, whereas for inputs that are far from Π, the verifier will reject with high probability no matter what prover strategy is employed. The latter proof systems are known as *interactive proofs of proximity ($\mathcal{IPP}$s)*.[2]

The focus of this paper is identifying natural classes of properties that are known to be hard to test, but become easy to *verify* using the power of a proof ($\mathcal{MAP}$) or interaction with a prover ($\mathcal{IPP}$).

1.1. Our results

One well-known class of properties that is hard to test is the class of *context-free languages*. Alon et al. [10] showed that there exists a context-free language that requires $\Omega\left(\sqrt{n}\right)$ queries to test (where here and throughout this work, n denotes the size of the input) and a context-free language that requires $\Omega(n)$ queries to test with *one-sided error*. Furthermore, there are no known (non-trivial) testers for general context-free languages.

Another interesting class is the class of languages that are accepted by small *read-once branching programs (ROBPs)*. Newman [11] showed that the set of strings accepted by any small width ROBP can be efficiently tested.[3] More specifically, Newman showed that width w ROBPs can be tested using $(2^w/\varepsilon)^{O(w)}$ queries, where ε is the proximity parameter. Bollig [12] showed that Newman's result cannot be extended to polynomial-sized ROBPs, by exhibiting an $O(n^2)$-sized ROBP that requires $\Omega(\sqrt{n})$ queries to test. No (non-trivial) testers for general ROBPs are known for width $\Omega(\sqrt{\log n})$.

In this work we consider the question of constructing *efficient* $\mathcal{MAP}$s and $\mathcal{IPP}$s for these two classes.[4] Here, by "efficient", we mean that *both* the *query complexity* (i.e., the number of queries performed by the verifier to the input) and the *proof complexity* (i.e., the length of the $\mathcal{MAP}$ proof) or *communication complexity* (i.e., the amount of communication with the $\mathcal{IPP}$ prover) are small and, in particular, sub-linear.[5]

Our first pair of results are efficient $\mathcal{MAP}$s for context-free languages and for ROBPs. These $\mathcal{MAP}$s offer a multiplicative trade-off between the query and proof complexities. Here and throughout this work, $n \in \mathbb{N}$ specifies the length of the main input and $\varepsilon \in (0, 1)$ denotes the proximity parameter.

Theorem 1.1. *For every context-free language $\mathcal{L}$ and every $k = k(n)$ such that $2 \le k \le n$, there exists an $\mathcal{MAP}$ for $\mathcal{L}$ that uses a proof of length $O(k \cdot \log n)$ and has query complexity $O\left(\frac{n}{k} \cdot \varepsilon^{-1}\right)$. Furthermore, the $\mathcal{MAP}$ has one-sided error.*

Theorem 1.2. *If a language $\mathcal{L}$ is recognized by a size $s = s(n)$ ROBP, then for every $k = k(n)$ such that $2 \le k \le n$, there exists an $\mathcal{MAP}$ for $\mathcal{L}$ that uses a proof of length $O(k \cdot \log s)$ and has query complexity $O\left(\frac{n}{k} \cdot \varepsilon^{-1}\right)$. Furthermore, the $\mathcal{MAP}$ has one-sided error.*

Hence, by setting $k = \sqrt{n}$, every context-free language and every language accepted by an ROBP of size at most $2^{\mathsf{polylog}(n)}$, has an $\mathcal{MAP}$ in which both the proof and query complexity are $\widetilde{O}\left(\sqrt{n}\right)$ (w.r.t. constant proximity parameter).

Next, we ask whether the query and proof complexity in Theorems 1.1 and 1.2 can be significantly reduced by allowing more extensive *interaction* between the verifier and the prover (i.e., arbitrary interactive communication rather than just a fixed non-interactive proof). Very relevant to this question is a recent result of [6] by which, loosely speaking, every language in $\mathcal{NC}$ (which contains all context-free languages [15] and languages accepted by small ROBPs[6]) has an $\mathcal{IPP}$ with $\widetilde{O}(\sqrt{n})$ query and communication complexities. While the [6] result is more general, for context-free languages and ROBPs it achieves roughly the same query and communication complexities as the $\mathcal{MAP}$s in Theorems 1.1 and 1.2, but uses much more interaction (i.e., at least logarithmically many rounds of interaction compared to just a single message in our $\mathcal{MAP}$s).

---

[1] A related notion is that of a *probabilistically checkable proof of proximity ($\mathcal{PCPP}$)* [4,5]. $\mathcal{PCPP}$s differ from $\mathcal{MAP}$s in that the verifier is only given *query* (i.e., oracle) access to the proof, whereas in $\mathcal{MAP}$s, the verifier has free (*explicit*) access to the proof. Hence, $\mathcal{PCPP}$s are a $\mathcal{PCP}$ analogue of property testing.

[2] Indeed, $\mathcal{MAP}$s can be thought of as a restricted case of $\mathcal{IPP}$s, in which the interaction is limited to a single message sent from the prover to the verifier.

[3] The result in [11] is stated only for *oblivious* ROBPs but in [12, Section 1.3] it is stated that Newman's result holds also for general *non-oblivious* ROBPs.

[4] To see that these two classes do not contain each other, observe that the language $\{0^i1^j2^i3^j : i, j \ge 1\}$, which is *not* a context-free language [13, Example 7.20], has a $\mathsf{poly}(n)$-width ROBP (which simply counts the number of repeated occurrences of 0, 1, 2 and 3). On the other hand, Kriegal and Waack [14] showed that every ROBP for the Dyck_2 language, which is a context-free language, has size $2^{\Omega(n)}$.

[5] As pointed out in [7], if we do not restrict the length of the proof, then *every* property Π can be verified trivially using only a constant amount of queries, by considering an $\mathcal{MAP}$ proof that contains a full description of the input.

[6] See Appendix B for a discussion on why languages accepted by ROBPs can be computed in small depth.