Electronic Colloquium on Computational Complexity, Report No. 32 (2016)

A Note on Tolerant Testing with One-Sided Error

Roei Tell

March 10, 2016

Abstract

A tolerant tester with *one-sided error* for a property is a tester that accepts every input that is close to the property, with probability 1, and rejects every input that is far from the property, with positive probability. In this note we show that such testers require a linear number of queries.

1 Introduction

This note deals with property testing, and assumes familiarity with the basic notions involved; for expository texts, see, e.g., [Ron09, Gol16]. We will specifically be interested in *tolerant testers*, introduced by Parnas, Ron, and Rubinfeld [PRR06]. These are testers that distinguish, with high probability, between inputs that are close to the property, and inputs that are far from the property.

We prove that it is impossible to test a property tolerantly with both a sub-linear number of queries and *one-sided error*. Specifically, for a property Π and two constants $\epsilon > 0$ and $\epsilon' < \epsilon$, an ϵ'-tolerant ϵ-tester with one-sided error for Π accepts inputs that are ϵ'-close to Π, *with probability* 1, and rejects inputs that are ϵ-far from Π, with positive probability. We show that for essentially any Π, and any pair of constants $\epsilon > 0$ and $\epsilon' < \epsilon$, any ϵ'-tolerant ϵ-tester with one-sided error for Π requires a linear number of queries. The proof is based on a simple claim that is implicit in the proofs of two similar previous results (see [Tel15, Prop. 5.6 and Cor. 5.7]). In Section 4 we reproduce the proofs of these two results, as corollaries of the said claim.

2 Preliminaries

We will be interested in properties of Boolean strings. The distance between two n-bit strings, denoted by δ, is the relative Hamming distance (i.e., $\delta(x,y) = \frac{|\{i \in [n] : x_i \neq y_i\}|}{n}$). The distance between a string $x \in \{0,1\}^n$ and a non-empty set $S \subseteq \{0,1\}^n$ is $\min_{s \in S} \delta(x,s)$. We say that a string is ϵ-close to a set (resp., ϵ-far from a set), if its distance from the set is at most ϵ (resp., at least ϵ). In the following definition we refer to algorithms that get oracle access to a string $x \in \{0,1\}^n$; by this we mean that for any $i \in [n]$, the algorithm can query for the value of the i^{th} bit of x.

ISSN 1433-8092

Definition 1 *(tolerant testers with one-sided error). Let $\Pi_n \subseteq \{0,1\}^n$, and let $\epsilon > 0$ and $\epsilon' < \epsilon$. An ϵ'-tolerant ϵ-tester with one-sided error for Π_n is a probabilistic algorithm T that satisfies the following conditions:*

1. *For every x that is ϵ'-close to Π_n it holds that $\Pr[T^x(1^n) = 1] = 1$.*

2. *For every x that is ϵ-far from Π_n it holds that $\Pr[T^x(1^n) = 0] > 0$.*

Note that in Condition (2) of Definition 1 we only require rejection of "far" inputs with *positive* probability, rather than with high probability. Indeed, the lower bound holds even for this relaxed definition.

3 The new result

The following claim is implicit in the proof of [Tel15, Claim 5.6.1].

Claim 2. *Let $n \in \mathbb{N}$, and let $S \subseteq \{0,1\}^n$. Assume that there exists a probabilistic algorithm A that queries any input string in q locations, and satisfies the following conditions:*

- *For every $s \in S$ it holds that A accepts s, with probability 1.*

- *There exists $w \notin S$ such that A rejects w, with positive probability.*

Then, every $x \in \{0,1\}^n$ is (q/n)-close to $\overline{S} = \{0,1\}^n \setminus S$.

Proof. Let r be random coins such that when A queries w with coins r it holds that A rejects w. Denote by $(i_1, ..., i_q)$ the corresponding q locations in w that A queries, given coins r. Now, let $x \in \{0,1\}^n$. Let x' be the string obtained by modifying the q locations $(i_1, ..., i_q)$ in x to the values $(w_{i_1}, ..., w_{i_q})$. Observe that when A queries x' with coins r, it queries locations $(i_1, ..., i_q)$, sees the values $(w_{i_1}, ..., w_{i_q})$, and thus rejects x'. Hence, it cannot be that $x' \in S$ (otherwise A would have to accept x' with probability 1). It follows that $\delta(x, \overline{S}) \leq \delta(x, x') = q/n$. ■

Note that Claim 2 also holds if we switch the roles of "accept" and "reject" in it (i.e., if we assume that A rejects every $s \in S$ with probability 1, and accepts some $w \in S$ with positive probability); we will use this fact in Section 4. Our lower bound on tolerant testers with one-sided error follows easily from Claim 2:

Theorem 3 *(tolerant testing with one-sided error). Let $\Pi_n \subseteq \{0,1\}^n$, and let $\epsilon > 0$. Assume that there exists $p \in \Pi_n$ and $z \in \{0,1\}^n$ such that $\delta(z, \Pi_n) \geq \epsilon$. Then, for every $\epsilon' < \epsilon$, every ϵ'-tolerant ϵ-tester with one-sided error for Π uses more than $\epsilon' \cdot n$ queries.*

Proof. For $\epsilon' < \epsilon$, let T be an ϵ'-tolerant ϵ-tester with one-sided error for Π_n, and denote the query complexity of T by $q = q(n)$. Let S be the set of strings that are ϵ'-close to Π_n, and let w be the string z such that $\delta(z, \Pi_n) \geq \epsilon$. Note that T accepts every $s \in S$, with probability 1, and rejects w, with positive probability. Invoking Claim 2 with the tester T, and with these S and w, we deduce that every $x \in \{0,1\}^n$ is (q/n)-close to being at distance more than ϵ' from Π_n (i.e., x is (q/n)-close to $\overline{S}$). This applies in particular to the string $p \in \Pi_n$. However, the distance of p from any string $y \in \overline{S}$ is more than ϵ', because $\epsilon' < \delta(y, \Pi_n) \leq \delta(y, p)$. It follows that $q/n \geq \delta(p, \overline{S}) > \epsilon'$. ■

Note that the two requirements in Theorem 3 (about the existence of p and of z) only exclude "degenerate" cases: If either of the two requirements does not hold, then the testing problem is trivial to begin with.

4 Previous results as corollaries of Claim 2

As mentioned in Section 1, Claim 2 is implicit in the proofs of two similar results, which we now reproduce.

4.1 Testers for dual problems with one-sided error

Dual problems were introduced in [Tel15], and involve the testing of properties of the form "all inputs that are far from the set" (e.g., testing the property of graphs that are far from being connected). Specifically, a tester for the dual problem of a set Π accepts, with high probability, every input that is far from Π, and rejects, with high probability, every input that is far from being so; that is, it rejects every input that is far from the set of inputs that are far from Π.

While dual testing problems turn out to be very interesting in general, solving dual problems with *one-sided error* (i.e., accepting inputs that are far from Π with probability 1) requires a linear number of queries. Let us formally state and prove this.

Definition 4 *(testers with one-sided error for dual problems). Let $\Pi_n \subseteq \{0,1\}^n$, and let $\epsilon > 0$ and $\epsilon' \leq \epsilon$. An ϵ'-*tester with one-sided error for the ϵ-dual problem of Π_n *is a probabilistic algorithm T that satisfies the following conditions:*

1. *For every x that is ϵ-far from Π_n it holds that $\Pr[T^x(1^n) = 1] = 1$.*
2. *For every x that is ϵ'-far from the set of strings that are ϵ-far from Π_n, it holds that $\Pr[T^x(1^n) = 0] > 0$.*

Theorem 5 *(dual problems with one-sided error; see [Tel15, Cor. 5.7]). Let $\Pi_n \subseteq \{0,1\}^n$, and let $\epsilon > 0$. Assume that there exists $p \in \Pi_n$ and $z \in \{0,1\}^n$ such that $\delta(z, \Pi_n) \geq 2 \cdot \epsilon$. Then, for every $\epsilon' \leq \epsilon$, every ϵ'-tester with one-sided error for the ϵ-dual problem of Π uses more than $\epsilon \cdot n$ queries.*

Proof. For $\epsilon' \leq \epsilon$, let T be an ϵ'-tester with one-sided error for the ϵ-dual problem of Π, and denote the query complexity of T by $q = q(n)$. Let S be the set of strings that are ϵ-far from Π_n, and let w be the string $p \in \Pi_n$. Note that T accepts every $s \in S$, with probability 1. Also observe that p is at distance at least $\epsilon \geq \epsilon'$ from any $z' \in S$ (because $\epsilon \leq \delta(z', \Pi_n) \leq \delta(z', p)$), and thus T rejects p with positive probability.

We can thus invoke Claim 2 with the tester T, and with S and $w = p$ as above, and deduce that every $x \in \{0,1\}^n$ is (q/n)-close to being at distance less than ϵ from Π_n; that is, $\delta(x, \Pi_n) < (q/n) + \epsilon$. However, by our hypothesis, there exists z such that $\delta(z, \Pi_n) \geq 2 \cdot \epsilon$. It follows that $2 \cdot \epsilon \leq \delta(z, \Pi_n) < (q/n) + \epsilon$, which implies that $q(n) > \epsilon \cdot n$. ■

4.2 Testers with perfect soundness

Testers with *perfect soundness* are testers that accept every input in Π, with positive probability, and reject every input that is far from Π, with probability 1. It turns out that this task also requires a linear number of queries. The proof of this result, which we now detail, is very similar to the proof of Theorem 5.

Definition 6 *(testers with perfect soundness). Let $\Pi_n \subseteq \{0,1\}^n$, and let $\epsilon > 0$. An ϵ-*tester with perfect soundness for Π_n *is a probabilistic algorithm T that satisfies the following conditions:*

1. *For every $x \in \Pi_n$ it holds that $\Pr[T^x(1^n) = 1] > 0$.*
2. *For every x that is ϵ-far from Π_n it holds that $\Pr[T^x(1^n) = 0] = 1$.*

Theorem 7 *(testing with perfect soundness; see [Tel15, Prop. 5.6]). Let $\Pi_n \subseteq \{0,1\}^n$, and let $\epsilon > 0$. Assume that there exists $p \in \Pi_n$ and $z \in \{0,1\}^n$ such that $\delta(z, \Pi_n) \geq 2 \cdot \epsilon$. Then, every ϵ-tester with perfect soundness for Π uses more than $\epsilon \cdot n$ queries.*

Proof. Let T be an ϵ-tester with perfect soundness for Π, and denote its query complexity by $q = q(n)$. Let S be the set of strings that are ϵ-far from Π_n, and let w be the string $p \in \Pi_n$. Note that T rejects every $s \in S$, with probability 1, and accepts w, with positive probability. Invoking Claim 2 with the tester T, and with these S and w, we deduce that every $x \in \{0,1\}^n$ is (q/n)-close to being at distance less than ϵ from Π_n. However, by our hypothesis, there exists z such that $\delta(z, \Pi_n) \geq 2 \cdot \epsilon$, which implies (as in the proof of Theorem 5) that $q(n) > \epsilon \cdot n$. ■

Acknowledgements

The author thanks his advisor, Oded Goldreich, for suggesting the question that this note answers, and for encouraging the author to write this note. This research was partially supported by the Israel Science Foundation (grant No. 671/13).

References

[Gol16] Oded Goldreich. Lecture notes on property testing, 2016. Available at http://www.wisdom.weizmann.ac.il/~oded/pt-ln.html.

[PRR06] Michal Parnas, Dana Ron, and Ronitt Rubinfeld. Tolerant property testing and distance approximation. *Journal of Computer and System Sciences*, 72(6):1012–1042, 2006.

[Ron09] Dana Ron. Algorithmic and analysis techniques in property testing. *Foundations and Trends in Theoretical Computer Science*, 5(2):73–205, 2009.

[Tel15] Roei Tell. On being far from far and on dual problems in property testing. *Electronic Colloquium on Computational Complexity: ECCC*, 22:72, 2015.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de