A Hypergraph Dictatorship Test with Perfect Completeness

Victor Chen [*]

Linearity and dictatorship testing have been studied in the past decade both for their combinatorial interest and connection to complexity theory. These tests distinguish functions which are linear/dictator from those which are far from being a linear/dictator function. The tests do so by making queries to a function at certain points and receiving the function's values at these points. The parameters of interest are the number of queries a test makes and the completeness and soundness of a test.

In this work we deal with boolean functions of the form $f : \{0,1\}^n \to \{\text{-}1,1\}$. We say a function f is *linear* if $f = (-1)^{\sum_{i \in S} x_i}$ for some subset $S \subseteq [n]$. A *dictator* function is simply a linear function where $|S| = 1$, i.e., $f(x) = (-1)^{x_i}$ for some i. A dictator function is often called a *long code*, and it is first used in [3] for the constructions of probabilistic checkable proofs (PCPs), see e.g., [2, 1]. Since then, it has become standard to design a PCP system as the composition of two verifiers, an outer verifier and an inner verifier. In such case, a PCP system expects the proof to be written in such a way so that the outer verifier, typically based on the verifier obtained from Raz's Parallel Repetition Theorem [17], selects some tables of the proof according to some distribution and then passes the control to the inner verifier. The inner verifier, with oracle access to these tables, makes queries into these tables and ensures that the tables are the encoding of some error-correcting codes and satisfy some joint constraint. The long code encoding is usually employed in these proof constructions, and the inner verifier simply tests whether a collection of tables (functions) are long codes satisfying some constraints. Following this paradigm, constructing a PCP with certain parameters reduces to the problem of designing a long code test with similar parameters.

One question of interest is the tradeoff between the soundness and query complexity of a tester. If a tester queries the functions at every single value, then trivially the verifier can determine all the functions. One would like to construct a dictatorship test that has the lowest possible soundness while making as few queries as possible. One way to measure this tradeoff between the soundness s and the number of queries q is *amortized query complexity*, defined as $\frac{q}{\log s^{-1}.}$ This investigation, initiated in [25], has since spurred a long sequence of works [22, 20, 11, 6]. All the testers from these works run many iterations of a single dictatorship test by reusing queries from previous iterations. The techniques used are Fourier analytic, and the best amortized query complexity from this sequence of works has the form $1 + O\left(\frac{1}{\sqrt{q}}\right)$.

The next breakthrough occurs when Samorodnitsky [19] introduces the notion of a *relaxed* linearity test along with new ideas from additive combinatorics. In property testing, the goal is to distinguish objects that are very structured from those that are pseudorandom. In the case of linearity/dictatorship testing, the structured objects are the linear/dictator functions, and functions that are far from being linear/dictator are interpreted as pseudorandom. The recent paradigm in additive combinatorics is to find the right framework of structure and pseudorandomness and analyze combinatorial objects by dividing them into structured and pseudorandom components, see e.g. [24] for a survey. One success is the notion of Gowers norm [7], which has been fruitful in attacking many problems in additive combinatorics and computer science. In [19], the

[*] MIT CSAIL. victor@csail.mit.edu. Research supported in part by an NSF graduate fellowship and NSF Award CCR-0514915.

notion of pseudorandomness for linearity testing is relaxed; instead of designating the functions that are far from being linear as pseudorandom, the functions having small low degree Gowers norm are considered to be pseudorandom. By doing so, an optimal tradeoff between soundness and query complexity is obtained for the problem of relaxed linearity testing. (Here the tradeoff is stronger than the tradeoff for the traditional problem of linearity testing.)

In a similar fashion, in the PCP literature since [9], the pseudorandom objects in dictatorship tests are not functions that are far from being a dictator. The pseudorandom functions are typically defined to be either functions that are far from all "juntas" or functions whose "low-degree influences" are $o(1)$. Both considerations of a dictatorship test are sufficient to compose the test in a PCP construction. In [21], building on the analysis of the relaxed linearity test in [19], Samorodnitsky and Trevisan construct a dictatorship test (taking the view that functions with arbitrary small "low-degree influences are pseudorandom) with amortized query complexity $1+O\left(\frac{\log q}{q}\right)$. Furthermore, the test is used as the inner verifier in a conditional PCP construction (based on unique games [12]) with the same parameters. However, their dictatorship test suffers from an inherent loss of perfect completeness. Ideally one would like testers with one-sided errors. One, for aesthetic reasons, testers should always accept valid inputs. Two, for some hardness of approximation applications, in particular coloring problems (see e.g. [10] or [5]), it is important to construct PCP systems with one-sided errors.

In this paper, we prove the following theorem:

Theorem. *For every $q \geq 3$, there exists an (adaptive) dictatorship test that makes q queries, has completeness 1, and soundness $\frac{O(q^3)}{2^q}$; in particular it has amortized query complexity $1 + O\left(\frac{\log q}{q}\right)$.*

Our tester is a variant of the one given in [21]. Our tester is adaptive in the sense that it makes its queries in two stages. It first makes roughly $\log q$ nonadaptive queries into the function. Based on the values of these queries, the tester then selects the rest of the query points nonadaptively. Our analysis is based on techniques developed in [11, 21, 10, 8].

Related Works The problem of linearity testing was first introduced in [4]. The framework of property testing was formally set up in [18]. The PCP Theorems were first proved in [2, 1]; dictatorship tests first appeared in the PCP context in [3], and many dictatorship tests and variants appeared throughout the PCP literature. Dictatorship test was also considered as a standalone property testing in [16]. As mentioned, designing testers and PCPs focusing on amortized query complexity was first investigated in [25], and a long sequence of works [22, 20, 11, 6] followed. The first tester/PCP system focusing on this tradeoff while obtaining perfect completeness was achieved in [10].

The orthogonal question of designing testers or PCPs with as few queries as possible was also considered. In a highly influential paper [9], Håstad constructed a PCP system making only three queries. Many variants also followed. In particular PCP systems with perfect completeness making three queries were also achieved in [8, 13]. Similar to our approach, O'Donnell and Wu [14] designed an optimal three bit dictatorship test with perfect completeness, and later the same authors constructed a conditional PCP system [15].

Future Direction Recently, Tamaki and Yoshida in their recent preprint [23] designed a dictatorship test that makes *non-adaptive* q queries, has completeness 1, and soundness $O(q \cdot 2^{-q})$. However, it is not clear how to extend either their test or ours to a PCP construction.

Analogous to the works in [14, 15], it would be interesting to extend these query-efficient dictator tests to PCPs using Khot's d-to-1 outer verifier [12]. In particular, we leave the following conjecture as a challenging open problem:

Conjecture. *For infinitely many q, there exists a* PCP *system that makes q queries, has completeness* 1*, and soundness* $\mathrm{poly}(q) \cdot 2^{-q}$.

References

[1] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *J. ACM*, 45(3):501–555, 1998.

[2] Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: a new characterization of NP. *J. ACM*, 45(1):70–122, 1998.

[3] Mihir Bellare, Oded Goldreich, and Madhu Sudan. Free bits, PCPs, and nonapproximability—towards tight results. *SIAM Journal on Computing*, 27(3):804–915, 1998.

[4] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of Computer and System Sciences*, 47(3):549–595, 1993.

[5] Irit Dinur, Elchanan Mossel, and Oded Regev. Conditional hardness for approximate coloring. In *STOC '06: Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 344–353, New York, NY, USA, 2006. ACM.

[6] Lars Engebretsen and Jonas Holmerin. More efficient queries in PCPs for NP and improved approximation hardness of maximum CSP. In *STACS*, pages 194–205, 2005.

[7] W. T. Gowers. A new proof of Szemerédi's theorem. *Geom. Funct. Anal.*, 11(3):465–588, 2001.

[8] Venkatesan Guruswami, Daniel Lewin, Madhu Sudan, and Luca Trevisan. A tight characterization of NP with 3 query PCPs. In *FOCS '98: Proceedings of the 39th Annual Symposium on Foundations of Computer Science*, page 8, Washington, DC, USA, 1998. IEEE Computer Society.

[9] Johan Håstad. Some optimal inapproximability results. *J. ACM*, 48(4):798–859, 2001.

[10] Johan Håstad and Subhash Khot. Query efficient PCPs with perfect completeness. *Theory of Computing*, 1(7):119–148, 2005.

[11] Johan Håstad and Avi Wigderson. Simple analysis of graph tests for linearity and PCP. *Random Struct. Algorithms*, 22(2):139–160, 2003.

[12] Subhash Khot. On the power of unique 2-prover 1-round games. In *STOC '02: Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 767–775, New York, NY, USA, 2002. ACM.

[13] Subhash Khot and Rishi Saket. A 3-query non-adaptive PCP with perfect completeness. In *CCC '06: Proceedings of the 21st Annual IEEE Conference on Computational Complexity*, pages 159–169, Washington, DC, USA, 2006. IEEE Computer Society.

[14] Ryan O'Donnell and Yi Wu. 3-bit dictator testing: 1 vs. 5/8. In *SODA '09: Proceedings of the Nineteenth Annual ACM -SIAM Symposium on Discrete Algorithms*, pages 365–373, Philadelphia, PA, USA, 2009. Society for Industrial and Applied Mathematics.

[15] Ryan O'Donnell and Yi Wu. Conditional hardness for satisfiable-3csps. In *STOC '09: Proceedings of the forty-first annual ACM symposium on Theory of computing*, page To appear, New York, NY, USA, 2009. ACM.

[16] Michal Parnas, Dana Ron, and Alex Samorodnitsky. Testing basic boolean formulae. *SIAM Journal on Discrete Mathematics*, 16(1):20–46, 2002.

[17] Ran Raz. A parallel repetition theorem. *SIAM J. Comput.*, 27(3):763–803, 1998.

[18] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM Journal on Computing*, 25(2):252–271, 1996.

[19] Alex Samorodnitsky. Low-degree tests at large distances. In *STOC '07: Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 506–515, New York, NY, USA, 2007. ACM.

[20] Alex Samorodnitsky and Luca Trevisan. A PCP characterization of NP with optimal amortized query complexity. In *STOC '00: Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 191–199, New York, NY, USA, 2000. ACM.

[21] Alex Samorodnitsky and Luca Trevisan. Gowers uniformity, influence of variables, and PCPs. In *STOC '06: Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 11–20, New York, NY, USA, 2006. ACM.

[22] Madhu Sudan and Luca Trevisan. Probabilistically checkable proofs with low amortized query complexity. In *FOCS '98: Proceedings of the 39th Annual Symposium on Foundations of Computer Science*, page 18, Washington, DC, USA, 1998. IEEE Computer Society.

[23] Suguru Tamaki and Yuichi Yoshida. A query efficient non-adaptive long code test with perfect completeness. *ECCC TR09-074*, 2009.

[24] Terence Tao. Structure and randomness in combinatorics. In *FOCS '07: Proceedings of the forty-eighth annual ACM symposium on Foundations of computer science*, pages 3–15, New York, NY, USA, 2007. ACM.

[25] Luca Trevisan. Recycling queries in PCPs and in linearity tests (extended abstract). In *STOC '98: Proceedings of the thirtieth annual ACM symposium on Theory of computing*, pages 299–308, New York, NY, USA, 1998. ACM.