Approximate Testing with Error Relative to Input Size

Marcos Kiwi[1]
Dept. de Ing. Matemática, U. de Chile, Santiago 170–3, Chile

and

Frédéric Magniez[2]
CNRS–LRI, UMR 8623, Université Paris–Sud, 91405 Orsay, France

and

Miklos Santha[3]
CNRS–LRI, UMR 8623, Université Paris–Sud, 91405 Orsay, France

---

We formalize the notion and initiate the investigation of approximate testing for arbitrary forms of the error term. Until now only the case of absolute error had been addressed ignoring the fact that often only the most significant figures of a numerical calculation are valid. This work considers approximation errors whose magnitude grows with the size of the input to the program. We demonstrate the viability of this new concept by addressing the basic and benchmark problem of self–testing for the class of linear and polynomial functions. We obtain stronger versions of results of Ergün, Ravi Kumar, and Rubinfeld [11] by exploiting elegant techniques from Hyers–Ulam stability theory.

Key Words: program verification, approximation error, self–testing programs, robustness and stability of functional equations.

A preliminary version of this paper appeared in *Proceedings of the 31st ACM Symposium on Theory of Computing*, pp. 51–60, 1999, under the title "Approximate Testing with Relative Error".

[1]Gratefully acknowledges the hospitality of LRI, U. Paris Sud and the support of Conicyt, via Fondecyt No. 1981182 and a CNRS–Conicyt'98 Project.
[2]Partially supported by a CNRS–Conicyt'98 Project, Fondap'98 in Applied Mathematics, and ESPRIT Working Group RAND2, No. 21726.
[3]Partially supported by a CNRS–Conicyt'98 Project and ESPRIT Working Group RAND2, No. 21726.

1. INTRODUCTION

The following is a quote from Knuth [19, Ch. 4, § 2.2]: *Floating point computation is by nature inexact, and programmers can easily misuse it so that the computed answers consist almost entirely of "noise." One of the principal problems of numerical analysis is to determine how accurate the results of certain numerical methods will be. There's a credibility gap: we don't know how much of the computer's answers to believe.... Many serious mathematicians have attempted to analyze a sequence of floating point operations rigorously, but have found the task so formidable that they have tried to be content with plausibility arguments instead.* Then, Knuth goes on to say: *A rough (but reasonably useful) way to express the behavior of floating point arithmetic can be based on the concept of "significant figures" or relative error.*

If the exact real number x is represented inside the computer, an approximation $\widehat{x} = x(1+\theta)$ is often used. The quantity $\theta = (\widehat{x} - x)/x$ is called the relative error of approximation.

Consider now the task of writing a program P purported to compute a real valued function f. One of the difficulties of such an endeavor is that once P is implemented it is difficult to verify its correctness, i.e., that $P(x) = f(x)$ for all valid inputs x. Moreover, due to the inexact nature of digital computations, it might be impossible to compute f exactly. A more realistic requirement is that P compute f in such a way that $|P(x) - f(x)| \leq \beta(x)$ on every valid input x, where $\beta(x)$ is some appropriate error function.

The inaccuracies of many computational processes are made worst by a crucial and pervasive issue that arises in the programming practice; it is not easy to get a program right. To address the software correctness problem the notion of program checking [7, 5], self–testing programs [6], and self–correcting programs [6, 20] was pioneered by Blum et al. during the late 80's and early 90's. A *program checker* verifies whether the program gives the correct answer on a particular input, a *self–testing program* for f verifies whether the program P evaluates to f on most inputs, and a *self–correcting program* for f takes a program P that is correct on most inputs and uses it to compute f correctly on every input with high probability. Checkers and self–testers/correctors, testers for short, may call the program as a black box but are required to do something different and simpler than to actually compute the function f in a sense that is formalized in [5].

Initially, it was assumed in the testing literature, that programs performed exact computations and that the space of valid inputs was closed under the standard arithmetic operations, i.e., was an algebraically closed domain. Early on, it was recognized that these assumptions were too simplistic to capture the real nature of many computations, in particular the computation of real valued functions and of functions defined over rational domains (finite subsets of fixed point arithmetic of the form $\{i/s : |i| \leq n, i \in \mathbb{Z}\}$ for some $n, s > 0$). This led to the development of approximate testers [15, 2], testers over finite rational domains [20], and testers that consider both aspects simultaneously [11].

A key issue that arises throughout the testing literature is to verify whether a program belongs to a particular function class, i.e., the *property testing* problem. Once this problem has been resolved, testers for the members of the function class are often easier to derive. This justifies why we henceforth focus on this problem.

But, we concentrate on an aspect of the problem that has been ignored in the literature; the magnitude of the inaccuracies in many numerical computations depends on the size of the values involved in the calculations. This leads us to the following:

Problem: For a program P purportedly computing over a finite domain D a function in the class of real valued functions $\mathcal{F}$, and given error functions β and β', find a simple and efficient self–tester for P which, for some fiability parameters $0 \leq \delta < \delta' \leq 1$ and some subdomain $D' \subseteq D$,

- Outputs PASS with high probability if $\Pr_{x \in D}\left[|P(x) - f(x)| > \beta(x)\right]$ is at most δ for some function $f \in \mathcal{F}$.
- Outputs FAIL with high probability if $\Pr_{x \in D'}\left[|P(x) - f(x)| > \beta'(x)\right]$ is at least δ' for all functions $f \in \mathcal{F}$.

Exact self–testing is an instance of the above problem where β and β' are identically 0. Approximate self–testing with *absolute error* is the case where the error functions are constants. Testers have been built, in both of the latter scenarios, for different function classes and domains. But, they suffer from the following problem: when the error term is a small constant they fail good programs, e.g., those in which the error in the computation of $P(x)$ grows with the size of x. If on the contrary, the error term is a large constant, they might pass programs that make incorrectly large errors in the computation of $P(x)$ for small values of x. Self–testing with *relative error* refers to the case when the error can be proportional to the function value to be computed. This work addresses the intermediate case where the acceptable error terms are measured relative to some pre-specified function of the input x to the program P being tested — they do not depend on the function f purportedly being computed. In a subsequent work, for the case of multi-linear functions, Magniez [21] showed how to self–test with relative error. To derive our results we will require the error terms to satisfy certain conditions. But, before we describe our specific contributions let us discuss the context where they arise.

Previous Work: Here we will only discuss the literature concerned with testing over rational domains and/or approximate testing. For a thorough exposition of the motivations, applications, and work on exact testing, see the survey of Blum and Wasserman [8] and the thesis of Rubinfeld [25].

Self–testers/correctors for programs whose input values are from finite rational domains were first considered by Lipton [20] and further developed by Rubinfeld and Sudan [23]. In [20] a self–corrector for multivariate polynomials over a finite rational domain is given. In the same scenario [23] describes more efficient versions of this result as well as a self-tester for univariate polynomials.

The study of testing in the context of inexact computations was started by Gemmell et al. [15] who provided approximate self–testers/correctors for linear functions, logarithmic functions, and floating point exponentiation. Nevertheless, their work was limited to the context of algebraically closed domains. Program checking in the approximate setting was first considered by Ar et al. [2] who provided, among others, approximate checkers for some trigonometric functions and matrix operations.

The works discussed above left many open questions, several of which were settled by Ergün, Ravi Kumar, and Rubinfeld [11] who addressed the testing problem in the approximate context and over finite rational domains. Among other things, they showed how to perform approximate testing with absolute error for linear functions, polynomials, and for functions satisfying addition theorems. One of their significant contributions was to recognize the importance of stability theory in the context of testing. It is beyond the possibilities of this brief discussion to give a fair account of the achievements of this theory (the interested reader is referred to the surveys of Forti [12] and Hyers and Rassias [17]). But, a description of its goals is due. In order to do so, and also for concreteness sake, it will be convenient to recall a simple albeit fundamental testing problem that has played a key role in the development of the theory of testers; the Blum–Luby–Rubinfeld linearity test [6]. Given a program P purportedly computing a homomorphism from one finite Abelian group G into another such group, this test picks $u, v \in G$ at random and verifies whether $P(u) + P(v) = P(u{+}v)$. The analysis of this test described in [6] is due to Coppersmith [9] and goes as follows; define a function g whose value at u is the Majority of the multi-set $\{P(u{+}v) - P(v) : v \in G\}$ (here, the Majority of a multi-set is the most commonly occurring element in the multi-set, where ties are broken arbitrarily). Then, show that if the probability of the test failing is sufficiently small, three things happen. First, an overwhelming majority of the values $\{P(u{+}v) - P(v) : v \in G\}$ agree with $g(u)$, second, g is linear, and last, g is close to P. The analyzes of approximate tests with absolute error over algebraically non-closed domains follow a similar approach as the one described above. But, there are two significant differences. First, instead of taking Majority, the Median is used [11]. Second, both the closeness of g to P and the linearity of g can be ascertained only approximately. Therefore, to conclude that P is close to a linear function, a result showing that g is close to a linear function is needed. In general, this last step consisting in proving that a function g approximately satisfying a property is approximately close to a function that exactly satisfies the property is referred to as *proving stability.* The setting of the stage where a stability type result can be applied, i.e., showing that a function P that satisfies a condition, not necessarily everywhere, is close to a function g that satisfies the condition everywhere is referred to as *proving robustness.* The latter term was coined and formally defined in [24] and studied in [26].

Our Contributions: In Section 2 we address several instances of the so called *local stability problem.* Specifically, we consider functional equations, e.g., $f(x + y) = f(x) + f(y)$, and provide conditions under which a function that approximately satisfies the functional equation (over an algebraically non-closed domain) is well approximated by a function that satisfies it exactly (on some subset of the domain). We restrict our discussion to real valued functions whose domain is $D_n \stackrel{\text{def}}{=} \{i \in \mathbb{Z} : |i| \leq n\}$. But, our results can be directly extended to finite rational domains as those considered in [20, 23, 11].

In the literature, the local stability problem has been addressed only in the absolute error case, i.e., when the approximation error is constant. On the contrary, we consider more general forms of the approximation error. In particular, we allow error terms that grow with the size of the input on which a function is evaluated. As discussed earlier, we believe that this is a more realistic and interesting scenario. Moreover, our results generalize those previously obtained for the absolute error case. Nevertheless, our arguments are simpler than those previously used in the related literature. We illustrate them in Section 2.1. For the sake of precision, below we give an example of the kind of results we can derive.

First, we need to introduce the notion of *valid error terms of degree* $p \in \mathbb{R}$. These are nonnegative functions $\beta\colon \mathbb{Z}\times\mathbb{Z} \to \mathbb{R}_+$ which are, in each of their coordinates, even and nondecreasing for nonnegative integers, and such that $\beta(2s,2t) \le 2^p\beta(s,t)$ for all integers s,t. Examples of this type of functions are $\beta(s,t) = |s|^p + |t|^p$, and $\beta(s,t) = \max\{a, |s|^p, |t|^p\}$ for some nonnegative real number a. Whenever it is clear from context, we will abuse notation and will interpret a valid error term $\beta(\cdot,\cdot)$ as the function of one variable, denoted $\beta(z)$, that evaluates to $\beta(z,z)$ at z. Also, for every $p \in [0,1)$, we set $C_p = (1{+}2^p)/(2{-}2^p)$, and we will use this notation throughout the paper.

Theorem 1.1. *Let $\beta(\cdot,\cdot)$ be a valid error term of degree $p \in [0,1)$. Let $g\colon D_{2n} \to \mathbb{R}$ be such that for all $x, y \in D_n$,*

$$|g(x{+}y) - g(x) - g(y)| \;\le\; \beta(x,y)\,.$$

Then, the linear mapping $T\colon \mathbb{Z} \to \mathbb{R}$ defined by $T(n) = g(n)$ is such that for all $x \in D_n$,

$$|g(x) - T(x)| \;\le\; C_p\beta(x)\,.$$

Analogous results for multi-linear functions are derived in Section 2.2, and for univariate polynomials in Section 2.3.

Making $\beta(s,t) = |s|^p + |t|^p$ in Theorem 1.1 we immediately get the following:

Corollary 1.1. *Let $p \in [0,1)$ and $\theta > 0$. If $g\colon D_{2n} \to \mathbb{R}$ is such that for all $x, y \in D_n$,*

$$|g(x{+}y) - g(x) - g(y)| \;\le\; \theta\left(|x|^p + |y|^p\right)\,,$$

then there exists a linear function $T\colon D_n \to \mathbb{R}$ such that for all $x \in D_n$,

$$|g(x) - T(x)| \;\le\; 2C_p\theta|x|^p\,.$$

The condition that p be strictly less than 1 is necessary for Corollary 1.1 to hold. Too see this, it suffices to consider the counterexample of [18] to a similar statement where $\mathbb{R}$ takes the place of D_n and D_{2n}. Indeed, let f be the function which at x takes the value $x\log_2|x{+}1|$ if $x \ge 0$, and $x\log_2|x{-}1|$ if $x < 0$. Observe that f is nonlinear and that $|f(x{+}y) - f(x) - f(y)| \le |x| + |y|$ for all $x, y \in \mathbb{Z}$.

The particular case of Corollary 1.1 where $p = 0$ is implicit in [27] and reduces to the result of [11] concerning stability of the functional equation $f(x{+}y) - f(x) - f(y) = 0$. But, even in this special case, the analysis in [11] is rather technical and requires first approximating the function at hand by two additive functions (one defined over the negative elements and another for the positive elements of the domain) and then combining them to get the desired additive function. We bypass all of these technical problems.

In Section 3 we undertake the task of proving robustness in the scenario where non-constant error terms are allowed. These proofs of robustness follow those of [11] albeit with one major technical difference; the Median is taken over sets of non-fixed size. In Section 3.1 we give the first of our robustness results. It concerns linear functions. From it we prove the following:

THEOREM 1.2. *Let $\delta \in [0,1]$ and $\beta(\cdot,\cdot)$ be a valid error term of degree $p \in [0,1)$. If $P\colon D_{8n} \to \mathbb{R}$ is such that*

$$\Pr_{x,y\in D_{4n}} [|P(x+y) - P(x) - P(y)| > \beta(x,y)] \ \leq\ \delta/384\,,$$

then there exists a linear function $T\colon \mathbb{Z} \to \mathbb{R}$ such that

$$\Pr_{x\in D_n} [|P(x) - T(x)| > 17C_p\beta(x)] \ \leq\ 7\sqrt{\delta}/6\,.$$

(If $p = 0$, then $\beta(\cdot,\cdot)$ is a constant function and the latter inequality holds with $\delta/6$ in its RHS.)

We conclude Section 3.1 showing that without additional conditions on $\beta(\cdot,\cdot)$ the $\sqrt{\delta}$ in the conclusion of Theorem 1.2 is tight, up to constant factors.

In Section 3.2 we derive a similar result for univariate polynomials.

In Section 4, we extend the approximate self–testers definition of [11, 15] in order to capture the idea of approximate self–testing with relative error. We then show how the results of Section 2 and Section 3 yield approximate self–testers for more general forms of the error term than previously known. This is achieved through standard arguments when self–testing is done in the exact or in the absolute error case. In the relative error case, it is not as simple. Indeed, the issue is somewhat complicated by the fact that the error function might be too costly to compute. (It is interesting to note that the testing literature has so far implicitly assumed that the error term is efficiently computable.) In Section 4 we discuss this issue, and show that a good approximation of the error function suffices for self–testing. We conclude, in Section 5, by stating an open problem.

We restrict our discussion to approximate self–testing with relative error and will not address issues concerning approximate program checkers and self–correctors in a similar setting.

Relationship to other work: Although initially intended to address the problem of program correctness, the theory of self–testers/correctors had unanticipated consequences. Indeed, all known constructions of probabilistically checkable proofs [3] use in some way or another ideas and results concerning testers. Moreover, it has been shown that it has implications in learning theory and approximation theory [14], local stability theory [11], and coding theory [24].

2. STABILITY WITH RELATIVE ERROR

2.1. Approximate Linearity

In this section we prove Theorem 1.1 and illustrate an elegant technique for proving stability results in the context of approximate testing over finite rational domains. We bring together and strengthen two ideas developed in stability theory.

Our argument first relates a function g satisfying the hypothesis of Theorem 1.1 to a function h satisfying the same type of inequality but for all $x, y \in \mathbb{Z}$. Moreover, we will carefully choose h so that $h(x) = g(x)$ for all $x \in D_n$. In other words, h will be an *extension* of g restricted to D_n. Thus, in order to establish that the function g can be well approximated by a linear function it will suffice to show that h can be well approximated by a linear function T over all of $\mathbb{Z}$. This task is greatly simplified by the fact that the domain of h is a group. In fact, an elegant sequence of papers, starting with the 1941 paper of Hyers [16], addresses such a problem for functions whose domain have a semi–group structure. Hyers work was motivated by a question posed by Ulam who asked whether a function f that satisfies the functional equation $f(x+y) = f(x) + f(y)$ only approximately could always be approximated by a linear function. Hyers showed that if the equality was satisfied within a constant error term then f could be approximated, also within a constant error term, by a linear function. Many other positive answers to Ulam's question and variations of it are now known (see [17, 12] for a discussion of such results), e.g.,

LEMMA 2.1. [Rassias [22]] *Let E_1 be a normed semi–group, let E_2 be a Banach space, and let $h\colon E_1 \to E_2$ be a mapping for which there exists $\theta > 0$ and $p \in [0, 1)$ such that for all $x, y \in E_1$,*

$$\|h(x+y) - h(x) - h(y)\| \;\leq\; \theta(\|x\|^p + \|y\|^p)\,.$$

Then, the function $T\colon E_1 \to E_2$ defined by $T(x) = \lim_{m\to\infty} h(2^m x)/2^m$ is a well defined linear mapping such that for all $x \in E_1$,

$$\|h(x) - T(x)\| \;\leq\; \frac{2}{2-2^p}\theta\|x\|^p\,.$$

The main reason why we can not directly apply Rassias's Lemma to a function g such that $|g(x+y) - g(x) - g(y)| \leq \theta(|x|^p + |y|^p)$ for all $x, y \in D_n$, is that D_n is not a semi–group. It is in order to address this issue and to be able to exploit results like the one of Rassias that one would like to extend g into a function defined over all of $\mathbb{Z}$ (a group!) in such a way that the hypothesis of Rassias's lemma is satisfied. Although this is a rather natural approach, it requires more work than necessary. To explain this, let us advance that the extension h of g that we will propose is such that $h(x) = g(r_x) + q_x g(n)$, where $q_x \in \mathbb{Z}$ and $r_x \in D_n$ are the unique numbers such that $x = q_x n + r_x$ and $|q_x n| < |x|$ if $x \in \mathbb{Z} \setminus \{0\}$ and $q_0 = r_0 = 0$. Thus, the limit of $h(2^m x)/2^m$ when m goes to ∞ is $xg(n)/n$. Hence, we get for free that $T(x) = \lim_{t\to\infty} h(2^m x)/2^m$ is well defined and determines a linear mapping. Thus, when Lemma 2.1 is applied to a function like h the only nontrivial fact about $T(x) = \lim_{t\to\infty} h(2^m x)/2^m$ that we get is that T is close to h. The next lemma implies that to obtain this same conclusion a weaker hypothesis than that of Lemma 2.1 suffices.

LEMMA 2.2. *Let E_1 be a normed semi–group and E_2 be a normed vector space over $\mathbb{R}$. Let $\beta(\cdot,\cdot)$ be a valid error term of degree $p \in [0, 1)$, and let $h\colon E_1 \to E_2$ be such that for all $x \in E_1$,*

$$\|h(2x) - 2h(x)\| \;\leq\; 2\beta(x)\,.$$

If $T\colon E_1 \to E_2$ is such that $T(x) = \lim_{m\to\infty} h(2^m x)/2^m$ is a well defined mapping, then for all $x \in E_1$,

$$\|h(x) - T(x)\| \;\leq\; \frac{2}{2{-}2^p}\beta(x)\,.$$

Proof. We follow the same argument used by Rassias [22] to prove Lemma 2.1. We claim that for any positive integer m,

$$\|h(2^m x)/2^m - h(x)\| \;\leq\; \beta(x)\sum_{t=0}^{m-1} 2^{t(p-1)}\,.$$

The verification of this claim follows by induction on m. The case $m=1$ is clear because of the hypothesis. Assume the claim holds for m. To prove it for $(m{+}1)$, note that

$$\begin{aligned}\left\|\frac{h(2^{m+1}x)}{2^{m+1}} - h(x)\right\| \;&\leq\; \left\|\frac{h(2x)}{2} - h(x)\right\| + \frac{1}{2}\left\|\frac{h(2^m\cdot 2x)}{2^m} - h(2x)\right\| \\ &\leq\; \beta(x) + \frac{1}{2}\beta(2x)\sum_{t=0}^{m-1} 2^{t(p-1)}\,.\end{aligned}$$

To conclude the induction observe that $\beta(2x) \leq 2^p\beta(x)$. Thus, for all $x \in E_1$ and any positive integer m

$$\|h(2^m x)/2^m - h(x)\| \;\leq\; \beta(x)\sum_{t=0}^{m} 2^{t(p-1)} \;\leq\; \frac{2}{2{-}2^p}\,\beta(x)\,.$$

Since T is well defined and $\|\cdot\|$ is continuous the desired conclusion follows letting $m \to \infty$. ∎

Remark. We have stated and proved Lemma 2.2 in its full generality in order to highlight the properties required of the domain and range of the functions we deal with. But, we will apply Lemma 2.2 to functions from $E_1 = \mathbb{Z}$ to $E_2 = \mathbb{R}$.

To prove Theorem 1.1 we now show that an appropriate extension $h\colon \mathbb{Z} \to \mathbb{R}$ of a function $g\colon D_{2n} \to \mathbb{R}$ such that $|g(x{+}y) - g(x) - g(y)| \leq \beta(x,y)$ for all $x, y \in D_n$ satisfies the hypothesis of Lemma 2.2.

Lemma 2.3. *Let $\beta(\cdot,\cdot)$ be a valid error term of degree $p \in [0,1)$. Let $g\colon D_{2n} \to \mathbb{R}$ be such that for all $x, y \in D_n$,*

$$|g(x{+}y) - g(x) - g(y)| \;\leq\; \beta(x,y)\,.$$

Then, the function $h\colon \mathbb{Z} \to \mathbb{R}$ such that $h(x) = g(r_x) + q_x g(n)$ satisfies that for all $x \in \mathbb{Z}$,

$$|h(2x) - 2h(x)| \;\leq\; (1{+}2^p)\beta(x)\,.$$

Proof. Let $x, y \in \mathbb{Z}$. We first mention a couple of inequalities which we will repeatedly use without explicitly stating it. By definition of D_n, we have that $x \in D_n$ implies that $|x| \leq n$. By definition of r_x we have that $|r_x| \leq |x|$. Finally, by definition of valid error term, $\beta(r_x, r_y) = \beta(|r_x|, |r_y|) \leq \beta(|x|, |y|) = \beta(x, y)$.

Now, let $x \in \mathbb{Z}$. By definition of h and since $r_{2x} = 2r_x - n(q_{2x} - 2q_x)$,

$$|h(2x) - 2h(x)| \;=\; |g(2r_x - n(q_{2x} - 2q_x)) - 2g(r_x) + (q_{2x} - 2q_x)g(n)|\,.$$

Our objective is to bound the RHS of this equality by $(1+2^p)\beta(x)$. Note that $q_{2x} - 2q_x \in \{-1, 0, 1\}$. We consider three cases depending on the value that this latter quantity takes. (The following case analysis is rather tedious and may be skipped without loss on a first reading.)

Case 1: Assume $q_{2x} - 2q_x = 0$. Then, since $r_x \in D_n$, the hypothesis implies that $|h(2x) - 2h(x)| = |g(2r_x) - 2g(r_x)| \leq \beta(r_x, r_x)$. To conclude recall that $\beta(r_x, r_x) \leq \beta(x, x)$.

Case 2: Assume now that $q_{2x} - 2q_x = 1$. Hence, $r_{2x} = 2r_x - n$ and

$$\begin{aligned}
|h(2x) - 2h(x)| &= |g(2r_x - n) - 2g(r_x) + g(n)| \\
&\leq |g(2r_x) - 2g(r_x)| + |g(2r_x - n) + g(n) - g(2r_x)| \\
&\leq \beta(r_x, r_x) + \beta(2r_x - n, n)\,,
\end{aligned}$$

where the first inequality is due to the triangle inequality and the second inequality follows from the hypothesis since $r_x, r_{2x} = 2r_x - n, n \in D_n$. Recall that $\beta(r_x, r_x) \leq \beta(x, x)$ so the only thing that remains to be shown is that $\beta(2r_x - n, n) \leq 2^p\beta(x, x)$. To see this, note that $r_{2x} = 2r_x - n$ is at least $-n$, thus r_x can not be negative implying that $x \geq 0$. Hence, since $2x \geq 0$, we get that $r_{2x} \geq 0$ implying that $2r_x = r_{2x} + n \geq n$. Moreover, $|2r_x - n| = |r_{2x}| \leq n$. Thus, $\beta(2r_x - n, n) \leq \beta(n, n) \leq \beta(2r_x, 2r_x) \leq 2^p\beta(r_x, r_x)$. Recalling that $\beta(r_x, r_x) \leq \beta(x, x)$ we obtain the desired conclusion.

Case 3: Assume $q_{2x} - 2q_x = -1$. Hence, $r_{2x} = 2r_x + n$ which is at most n. Thus, r_x can not be positive. This implies that $r_x + n \in D_n$ and

$$\begin{aligned}
|h(2x) - 2h(x)| &= |g(2r_x + n) - 2g(r_x) - g(n)| \\
&\leq |g(2r_x + n) - g(r_x + n) - g(r_x)| + |g(r_x + n) - g(r_x) - g(n)| \\
&\leq \beta(r_x + n, r_x) + \beta(r_x, n)\,,
\end{aligned}$$

where the first inequality is due to the triangle inequality and the second one follows from the hypothesis since $r_x + n, r_x, n \in D_n$. To obtain the desired bound we will show that $\beta(r_x + n, r_x) \leq \beta(x, x)$ and $\beta(r_x, n) \leq 2^p\beta(x, x)$. Since r_x is not positive, $x \leq 0$. Hence, since $2x \leq 0$, we get that $r_{2x} \leq 0$ implying that $2r_x = r_{2x} - n \leq -n$. It follows that $|r_x + n| \leq |r_x| \leq n \leq |2r_x|$. Thus, $\beta(r_x + n, r_x) \leq \beta(r_x, r_x)$ and $\beta(r_x, n) \leq \beta(2r_x, 2r_x) \leq 2^p\beta(r_x, r_x)$. Recalling that $\beta(r_x, r_x) \leq \beta(x, x)$ we obtain the desired conclusion. ■

An immediate consequence of the two previous results is Theorem 1.1.

In summary, in order to tackle the local stability problem we propose to follow a two step approach. First, extend the function in an appropriate way to a domain that is algebraically closed and then use a Rassias type result to obtain the desired conclusion.

2.2. Approximate Multi-linearity

In this section we consider functions of k-variables that satisfy in each of their k coordinates an approximately linear functional equation on a bounded hypercube of $\mathbb{Z}^k$. We again extend such a function, but now to a function defined over all of $\mathbb{Z}^k$. We then show that such an extension can be well approximated by a function which is linear in each of its coordinates. Thus, our approximate stability result for multi-linear functions is obtained using an extension technique similar to the one illustrated in Section 2.1. For clarity of exposition we limit our discussion to particular error terms. Specifically, we prove the following:

THEOREM 2.1. *Let $p \in [0,1)$, $\theta > 0$, and $a \geq 0$. Also, let $\vec{e}_i \in \mathbb{Z}^k$ be such that $(\vec{e}_i)_j = 1$ if $i = j$ and 0 otherwise. Let $g\colon (D_{2n})^k \to \mathbb{R}$ be such that for all $i \in \{1,\ldots,k\}$, for all $\vec{z} \in (D_n)^k$ where $z_i = 0$, and for all $x, x' \in D_n$,*

$$|g(\vec{z}+(x+x')\vec{e}_i) - g(\vec{z}+x\vec{e}_i) - g(\vec{z}+x'\vec{e}_i)| \;\leq\; \theta \max\{a, |z_1|^p, \ldots, |z_k|^p, |x|^p, |x'|^p\}\,.$$

Then, the multi-linear function $T\colon (D_n)^k \to \mathbb{R}$ defined by $T(n,\ldots,n) = g(n,\ldots,n)$ is such that for all $\vec{z} \in (D_n)^k$,

$$|g(\vec{z}) - T(\vec{z})| \;\leq\; \theta C_p(2k-1) \max\{a, |z_1|^p, \ldots, |z_k|^p\}\,.$$

When $p = 0$ and $a = 1$, Theorem 2.1 yields the result stated in [11, Theorem 9] up to a factor 2.

Proof. Let $\beta(\vec{z})$ denote the quantity $\theta \max\{a, |z_1|^p, \ldots, |z_k|^p\}$, for any $\vec{z} \in (D_n)^k$. The theorem will be proven by induction on k. Observe that the case $k = 1$ corresponds to Theorem 1.1.

For the induction step, suppose that it holds for 1 and $(k-1)$. Let $g\colon (D_{2n})^k \to \mathbb{R}$ satisfy the hypothesis of theorem. Thus, there exist two real valued functions T_1 and T_2, which are defined over $(D_n)^k$, such that T_1 is linear in its $(k-1)$ first variables, T_2 is linear in its last variable, and for all $\vec{z} \in (D_n)^k$:

$$|T_1(\vec{z}) - g(\vec{z})| \;\leq\; (2k-3)C_p\beta(\vec{z})\,, \qquad \text{and} \qquad |T_2(\vec{z}) - g(\vec{z})| \;\leq\; C_p\beta(\vec{z})\,. \quad (1)$$

Moreover,

$$T_1(n,\ldots,n,z_k) = g(n,\ldots,n,z_k)$$

and

$$T_2(z_1,\ldots,z_{k-1},n) = g(z_1,\ldots,z_{k-1},n).$$

Let $T\colon (D_n)^k \to \mathbb{R}$ be the multi-linear function defined by

$$T(n,\ldots,n) = g(n,\ldots,n).$$

The function T also satisfies $T(z_1, \ldots, z_{k-1}, n) = T_1(z_1, \ldots, z_{k-1}, n)$. We will prove that T is close to g over $(D_n)^k$. Fix $\vec{z} \in (D_n)^k$, and denote by $\vec{z'}$ the vector which is obtained by replacing the last coordinate in $\vec{z}$ by n. Note that $|g(\vec{z}) - T(\vec{z})|$ is upper bounded by $|g(\vec{z}) - T_2(\vec{z})| + |T_2(\vec{z}) - T(\vec{z})|$. Since T and T_2 are linear in their last variable, the second term can be rewritten as $(|z_k|/n)|T_2(\vec{z'}) - T(\vec{z'})|$. The definition of T implies that $T(\vec{z'}) = T_1(\vec{z'})$, thus $|T_2(\vec{z'}) - T(\vec{z'})|$ is at most $|T_2(\vec{z'}) - g(\vec{z'})| + |g(\vec{z'}) - T_1(\vec{z'})|$. It follows that,

$$|g(\vec{z}) - T(\vec{z})| \ \leq \ |g(\vec{z}) - T_2(\vec{z})| + \frac{|z_k|}{n} \left(|T_2(\vec{z'}) - g(\vec{z'})| + |g(\vec{z'}) - T_1(\vec{z'})| \right) .$$

Hence, (1) yields,

$$|g(\vec{z}) - T(\vec{z})| \ \leq \ C_p \beta(\vec{z}) + C_p \frac{|z_k|}{n} \left(\beta(\vec{z'}) + (2k{-}3)\beta(\vec{z'}) \right) .$$

To conclude the proof, observe that $(|z_k|/n)\beta(\vec{z'}) \leq \beta(\vec{z})$. ■

2.3. Polynomials

The main purpose of this section is to prove a stability result similar to the one of Section 2.1 applicable to univariate polynomials. We require such a stability result in order to provide an approximate relative error test for univariate polynomials. Our stability result for polynomials, as well as its proof, is an extension of an argument in [1] generalized in [11] to the absolute error case over finite rational domains.

We adopt the standard convention of denoting the *forward difference operator* by ∇_t. Hence, by definition, $\nabla_t g(x) = g(x{+}t) - g(x)$ if g is a real valued function such that $x{+}t$ and x belong to g's domain. If we let ∇_t^d denote the operator corresponding to d applications of ∇_t and for $\vec{t} \in \mathbb{R}^d$ denote by $\nabla_{\vec{t}}$ the operator corresponding to the applications of $\nabla_{t_1}, \ldots, \nabla_{t_d}$, then it is easy to check that:

1. ∇_t is linear,
2. ∇_{t_1} and ∇_{t_2} commute,
3. $\nabla_{t_1,t_2} = \nabla_{t_1+t_2} - \nabla_{t_1} - \nabla_{t_2}$, and
4. $\nabla_t^d g(x) = \sum_{k=0}^{d} (-1)^{d-k} \binom{d}{k} g(x{+}kt)$.

The usefulness of the difference operator in testing was recognized by Rubinfeld and Sudan [23]. Using it, they were able to give more efficient self–correctors for polynomials, over finite fields and rational domains, than the one proposed by Lipton [20]. Its utility is mostly based on two facts; (1) $\nabla_t g(x)$ can be computed efficiently, (2) g is a degree $d{-}1$ polynomial over a vector space V, if and only if, $\nabla_t^d g(x) = 0$ for all $x \in V$ and $t \in \mathbb{Z}$ (see [4] for a more general form of this fact). In [11] it was shown that if the $\nabla_t^d g(x) = 0$ interpolation identity was approximately true, modulo an absolute error term, over a large bounded subset of the integers, then g was close, again modulo an absolute error term, to a degree $d{-}1$ polynomial over a smaller and coarser sub-domain. The reason that the closeness could only be obtained on a coarser sub-domain is that whereas the interpolation identity uses evenly spaced points, in order to prove stability, a more general condition with arbitrarily spaced points is needed. The lemmas stated below yield the same conclusion, but now modulo a relative error term. Both of the following statements are similar to results derived in [11] based on arguments found in [1, 13]. For completeness sake we include the proofs.

LEMMA 2.4. *Let $a \geq 0$, $\mu_d = \mathrm{lcm}\{1, \ldots, d\}$, and g be a real valued function over $D_{\mu_d \cdot d(d+1)n}$. Let $f\colon D_{dn} \to \mathbb{R}$ be such that $f(x) = g(\mu_d \cdot x)$. If for all $x, t \in D_{\mu_d \cdot dn}$,*

$$|\nabla_t^d g(x)| \;\leq\; \frac{\theta}{(\mu_d d)^p 2^d} \max\{a, |x|^p, |t|^p\},$$

then for all $\vec{t} \in (D_n)^d$,

$$|\nabla_{\vec{t}} f(0)| \;\leq\; \theta \max\{a, |t_1|^p, \ldots, |t_d|^p\}.$$

Proof. For any $\vec{\lambda} \in \{0,1\}^d$ and $\vec{t} \in (D_n)^d$, let $t'_{\vec{\lambda}} = -\sum_{i=1}^d \lambda_i t_i / i$ and $t''_{\vec{\lambda}} = \sum_{i=1}^d \lambda_i t_i$. Also, let $(-1)^{\vec{\lambda}} = (-1)^{\lambda_1 + \ldots + \lambda_d}$. Then, as pointed out in [11, Fact 17],

$$\nabla_{\vec{t}} f(0) = \sum_{\vec{\lambda} \in \{0,1\}^d} (-1)^{\vec{\lambda}} \nabla^d_{t'_{\vec{\lambda}}} f(t''_{\vec{\lambda}}).$$

Moreover, $\mu_d \cdot t'_{\vec{\lambda}}, \mu_d \cdot t''_{\vec{\lambda}} \in D_{\mu_d \cdot d \cdot \max\{|t_1|, \ldots, |t_d|\}} \subseteq D_{\mu_d \cdot dn}$. Thus,

$$\begin{aligned}
|\nabla^d_{t'_{\vec{\lambda}}} f(t''_{\vec{\lambda}})| &= |\nabla^d_{\mu_d \cdot t'_{\vec{\lambda}}} g(\mu_d \cdot t''_{\vec{\lambda}})| \\
&\leq \frac{\theta}{(\mu_d d)^p 2^d} \max\{a, |\mu_d \cdot t'_{\vec{\lambda}}|^p, |\mu_d \cdot t''_{\vec{\lambda}}|^p\} \\
&\leq \frac{\theta}{2^d} \max\{a, |t_1|^p, \ldots, |t_d|^p\}.
\end{aligned}$$

The desired conclusion follows by triangle inequality. ∎

LEMMA 2.5. *Let d be a positive integer, $p \in [0,1)$ and $\theta > 0$. Let $a \geq 0$ and $f\colon D_{dn} \to \mathbb{R}$ be such for all $\vec{t} \in (D_n)^d$,*

$$|\nabla_{\vec{t}} f(0)| \;\leq\; \theta \max\{a, |t_1|^p, \ldots, |t_d|^p\}.$$

Then, there exists a polynomial $h_{d-1}\colon D_n \to \mathbb{R}$ of degree at most $d-1$ such that for all $x \in D_n$,

$$|f(x) - h_{d-1}(x)| \;\leq\; \theta \prod_{i=1}^{d-1} ((2i-1) C_p) \max\{a, |x|^p\}.$$

Proof. The proof is by induction on d. When $d = 1$, simply let $h_0(x) = f(0)$. Thus, by hypothesis, $|f(x) - h_0(x)| = |\nabla_x f(0)| \leq \theta \max\{a, |x|^p\}$ for all

$x \in D_n$. For the induction step, suppose that the lemma holds for d. We define a function $G\colon (D_{2n})^d \to \mathbb{R}$ by $G(\vec{t}) = \nabla_{\vec{t}} f(0)$. Then, from the stated properties of the difference operator and our hypothesis, we get that for all $i \in \{1,\ldots,d\}$ and $\vec{t} \in (D_n)^d$ such that $(\vec{t})_i = 0$ the following holds for all $t_i, t'_i \in D_n$,

$$\begin{aligned} |G(\vec{t}+(t_i+t'_i)\vec{e}_i) - G(\vec{t}+t_i\vec{e}_i) - G(\vec{t}+t'_i\vec{e}_i)| &= |\nabla_{t_1,\ldots,t_d,t'_i} f(0)| \\ &\leq \theta \max\{a, |t_1|^p, \ldots, |t_d|^p, |t'_i|^p\}, \end{aligned}$$

where the $\vec{e}_i$'s are the elements of the canonical base of $\mathbb{Z}^d$ mentioned in the statement of Theorem 2.1. Therefore, Theorem 2.1 implies that there exists a multi-linear function $H\colon (D_n)^d \to \mathbb{R}$ such that for all $\vec{x} \in (D_n)^d$, $|G(\vec{x}) - H(\vec{x})| \leq \theta(2d-1)C_p \max\{a, |x_1|^p, \ldots, |x_d|^p\}$. Extend H multi-linearly to $\mathbb{Z}^d$. Let $\vec{1}$ be the d-dimensional vector all of whose components are 1 and define the function f' on $D_{(d+1)n}$ by $f'(x) = f(x) - H(x\vec{1})/d!$. From [10, Lemma 2] we know that for all $x \in D_n$ and $\vec{t} \in (D_n)^d$,

$$\nabla_{\vec{t}} H(x \cdot \vec{1}) = d!H(\vec{t}).$$

Then, using the linearity of ∇, the triangular inequality, the above equality, the hypothesis, and the bound on $|G(\vec{x}) - H(\vec{x})|$, we get that for all $\vec{t} \in (D_n)^d$,

$$\begin{aligned} |\nabla_{\vec{t}} f'(0)| &= |\nabla_{\vec{t}} f(0) - \nabla_{\vec{t}} H(0 \cdot \vec{1})/d!| \\ &= |G(\vec{t}) - H(\vec{t})| \\ &\leq \theta(2d-1)C_p \max\{a, |t_1|^p, \ldots, |t_d|^p\}. \end{aligned}$$

From the inductive hypothesis we get that a polynomial $h_{d-1}\colon D_n \to \mathbb{R}$ of degree at most $d-1$ exists such that for all $x \in D_n$,

$$\begin{aligned} |f'(x) - h_{d-1}(x)| &\leq \theta(2d-1)C_p \prod_{i=1}^{d-1}((2i-1)C_p) \max\{a, |x|^p\} \\ &= \theta \prod_{i=1}^{d}((2i-1)C_p) \max\{a, |x|^p\}. \end{aligned}$$

To finish the proof, we let $h_d(x) = h_{d-1}(x) + H(x\vec{1})/d!$ for all $x \in D_n$. Then h_d is a polynomial of degree at most d, and for $x \in D_n$ we have that $|f(x) - h_d(x)| = |f'(x) - h_{d-1}(x)| \leq \theta \prod_{i=1}^{d}((2i-1)C_p) \max\{a, |x|^p\}$. ■

3. ROBUSTNESS WITH RELATIVE ERROR

3.1. Linearity

In this section we first prove approximate robustness in the relative error sense for the functional equation $f(x+y) - f(x) - f(y) = 0$. We then prove Theorem 1.2. Before proceeding we need to introduce some notation. Recall that the median of a set $S \subseteq \mathbb{R}$ is the smallest value of a such that $\Pr_{x \in S}[x \geq a]$ is at most 1/2. For $f\colon X \to \mathbb{R}$ we denote by $\mathrm{Med}_{x \in X}(f(x))$ the median of the values taken by f when x varies in X, i.e.,

$$\operatorname*{Med}_{x \in X}(f(x)) = \operatorname{Inf}\left\{ a \in \mathbb{R} : \Pr_{x \in X}[f(x) \geq a] \leq 1/2 \right\}.$$

To prove Theorem 1.2, we associate to P a function $g\colon D_{2n} \to \mathbb{R}$ such that

$$g(x) \;=\; \begin{cases} \operatorname*{Med}_{y\in D_{|x|}} \left(P(x+y) - P(y)\right), & \text{if } |x| \geq \sqrt{\delta}n\,, \\ \operatorname*{Med}_{y\in D_{\sqrt{\delta}n}} \left(P(x+y) - P(y)\right), & \text{otherwise}\,. \end{cases}$$

Note that the size of the set over which the median is taken is proportional to the size of x except when x is less than $\sqrt{\delta}n$, in which case the median is taken over $D_{\sqrt{\delta}n}$. The latter is the main departure of our proof technique from traditional analyzes of absolute error approximate testers. Below we prove two lemmas and derive from them Theorem 1.2. First, we state a fact that we will repeatedly use.

FACT 3.1. [Halving principle] *Let Ω and S denote finite sets such that $S \subseteq \Omega$, and let ψ be a boolean function defined over Ω. Then,*

$$\Pr_{x\in S}\left[\psi(x)\right] \;\leq\; \frac{|\Omega|}{|S|}\Pr_{x\in\Omega}\left[\psi(x)\right]\,.$$

If Ω is twice the size of S, then $\Pr_{x\in\Omega}[\psi(x)]$ is at least one half of $\Pr_{x\in S}[\psi(x)]$. This motivates the choice of name for Fact 3.1.

LEMMA 3.1. *Let $\phi(z) = \max\{\sqrt{\delta}n, |z|\}$. Under the hypothesis of Theorem 1.2, if $x \in D_n$ is randomly chosen, then $|P(x) - g(x)| > \widetilde{\beta}(x)$ with probability at most $\delta/6$, where $\widetilde{\beta}(s) = \beta(\phi(s))$.*

Proof. Let $P_{x,y} = P(x+y) - P(x) - P(y)$. Observe that by definition of g and Markov's inequality,

$$\begin{aligned} \Pr_{x\in D_n}\left[|g(x) - P(x)| > \widetilde{\beta}(x)\right] &= \Pr_{x\in D_n}\left[|\operatorname*{Med}_{y\in D_{\phi(x)}}\left(P_{x,y}\right)| > \widetilde{\beta}(x)\right] \\ &\leq 2\Pr_{x\in D_n, y\in D_{\phi(x)}}\left[|P_{x,y}| > \widetilde{\beta}(x)\right]\,. \end{aligned}$$

But, $\phi(x) \geq \phi(y) \geq |y|$ and $\phi(x) \geq |x|$ for $x \in D_n$ and $y \in D_{\phi(x)}$ together with the halving principle yield,

$$\begin{aligned} &\Pr_{x\in D_n, y\in D_{\phi(x)}}\left[|P_{x,y}| > \widetilde{\beta}(x)\right] \\ \leq\;& \frac{|D_{4n}|^2}{|\{(x,y) : x \in D_n, y \in D_{\phi(x)}\}|}\Pr_{x,y\in D_{4n}}\left[|P_{x,y}| > \beta(x,y)\right] \\ \leq\;& 32\Pr_{x,y\in D_{4n}}\left[|P_{x,y}| > \beta(x,y)\right]\,. \end{aligned}$$

The hypothesis implies the desired conclusion. ■

LEMMA 3.2. *Let $\phi(z) = \max\{\sqrt{\delta}n, |z|\}$. Under the hypothesis of Theorem 1.2, if $x, y \in D_n$, then $|g(x+y) - g(x) - g(y)| \leq 16\max\{\widetilde{\beta}(x), \widetilde{\beta}(y)\}$, where $\widetilde{\beta}(s) = \beta(\phi(s))$.*

Proof. First we show that for all $c \in D_{2n}$ and $I \subseteq D_{\phi(c)}$ such that $|I| \geq \sqrt{\delta}n{+}1$,

$$\Pr_{y\in I}\left[|g(c) - (P(c{+}y) - P(y))| > 4\widetilde{\beta}(c)\right] < 1/3\,. \tag{2}$$

Let $P_{x,y} = P(x{+}y) - P(x) - P(y)$. Note that Markov's inequality implies that

$$\Pr_{y\in I}\left[|g(c) - (P(c{+}y) - P(y))| > 4\widetilde{\beta}(c)\right] \leq 2 \Pr_{y\in I, z\in D_{\phi(c)}}\left[|P_{c+y,z} - P_{c+z,y}| > 4\widetilde{\beta}(c)\right].$$

Observe now that if y and z are randomly chosen in I and $D_{\phi(c)}$ respectively, then from the union bound we conclude that

$$\Pr_{y,z}\left[|P_{c+y,z} - P_{c+z,y}| > 4\widetilde{\beta}(c)\right] \;\leq\; \Pr_{y,z}\left[|P_{c+z,y}| > 2\widetilde{\beta}(c)\right] + \Pr_{y,z}\left[|P_{c+y,z}| > 2\widetilde{\beta}(c)\right].$$

But, $\phi(c) \geq \max\{|y|,|z|,|c|\}$ so $2\widetilde{\beta}(c) \geq \beta(c{+}y,z)$ and $2\widetilde{\beta}(c) \geq \beta(c{+}z,y)$. Hence, the halving principle implies that

$$\begin{aligned}
&\Pr_{y,z}\left[|P_{c+z,y}| > 2\widetilde{\beta}(c)\right] + \Pr_{y,z}\left[|P_{c+y,z}| > 2\widetilde{\beta}(c)\right] \\
\leq\;& \Pr_{y,z}\left[|P_{c+z,y}| > \beta(c{+}z,y)\right] + \Pr_{y,z}\left[|P_{c+y,z}| > \beta(c{+}y,z)\right] \\
\leq\;& 2\frac{|D_{4n}|^2}{|I|\cdot|D_{\phi(c)}|}\Pr_{u,v\in D_{4n}}\left[|P_{u,v}| > \beta(u,v)\right].
\end{aligned}$$

Observing that $|D_{4n}|^2/(|I|\cdot|D_{\phi(c)}|) < 32/\delta$ and recalling that

$$\Pr_{u,v\in D_{4n}}\left[|P_{u,v}| > \beta(u,v)\right] \leq \delta/384$$

we obtain (2).

Now, to prove the lemma, let $a, b \in D_n$ and let $G_{c,y} = g(c) - (P(c{+}y) - P(y))$. Without loss of generality, assume $|a| \leq |b|$. If $a \geq 0$ (respectively $a < 0$), by (2), with nonzero probability there is a $y \in \{-\sqrt{\delta}n, \ldots, 0\}$ (respectively $y \in \{0, \ldots, \sqrt{\delta}n\}$) for which we have that $|G_{a,y}| \leq 4\widetilde{\beta}(a)$, $|G_{b,a+y}| \leq 4\widetilde{\beta}(b)$, and $|G_{a+b,y}| \leq 4\widetilde{\beta}(a{+}b)$. Hence, since $\widetilde{\beta}(a{+}b) \leq 2\max\{\widetilde{\beta}(a), \widetilde{\beta}(b)\}$, we conclude that $|g(a{+}b) - g(a) - g(b)| \leq |G_{a+b,y} - G_{a,y} - G_{b,a+y}|$ is upper bounded by $16\max\{\widetilde{\beta}(a), \widetilde{\beta}(b)\}$. ■

To prove Theorem 1.2 observe that under its hypothesis $16\max\{\widetilde{\beta}(\cdot), \widetilde{\beta}(\cdot)\}$ is a valid error term of degree p, and Lemma 3.2 and Theorem 1.1 imply that there is a linear map $T\colon \mathbb{Z} \to \mathbb{R}$ such that $|g(x) - T(x)| \leq 16C_p\widetilde{\beta}(x)$ for all $x \in D_n$. Since $1 \leq C_p$, Lemma 3.1 implies that if $x \in D_n$ is randomly chosen, then $|P(x){-}T(x)| > 17\theta C_p\widetilde{\beta}(x)$ with probability at most $\delta/6 \leq \sqrt{\delta}/6$. To conclude the proof observe that $\Pr_{x\in D_n}\left[\widetilde{\beta}(x) = \beta(x)\right] \geq 1 - \sqrt{\delta}$.

TIGHTNESS: Let n be a positive integer, $0 < p < 1$, $0 < \delta < 1/4$, $\theta, c > 0$, let $\beta(x,y) = \theta\max\{|x|^p, |y|^p\}$, and consider the function $P\colon \mathbb{Z} \to \mathbb{R}$ such that

$$P(x) \;=\; \begin{cases} -\theta(\sqrt{\delta}n)^p & \text{if } -\sqrt{\delta}n \leq x < 0\,, \\ \theta(\sqrt{\delta}n)^p & \text{if } 0 < x \leq \sqrt{\delta}n\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

Observe that if $|x|$ or $|y|$ is greater than $\sqrt{\delta}n$ then $|P(x+y)-P(x)-P(y)| \le 2\beta(x,y)$. Hence, if $n' \ge n$, with probability at most δ it holds that $|P(x+y)-P(x)-P(y)| > 2\beta(x,y)$ when x, y are randomly chosen in $D_{n'}$.

We claim that for every linear function T, when $x \in D_n$ is randomly chosen, $|P(x)-T(x)| > c\beta(x)$ with probability greater than $\sqrt{\delta}/(2(\max\{1,2c\})^{1/p})$. Indeed, denote by $\mathrm{Dist}_{D_n}(P,T,\beta)$ the probability when $x \in D_n$ is randomly chosen that $|P(x)-T(x)| > \beta(x)$. We want to prove that if T is a linear function, then $\mathrm{Dist}_{D_n}(P,T,c\beta) \ge \sqrt{\delta}/(2(\max\{1,2c\})^{1/p})$. Let $d = 1/(\max\{1,2c\})^{1/p}$. Observe that $d \le 1$. Suppose that $T(x) = ax$ for some a, and $\mathrm{Dist}_{D_n}(P,T,c\beta) < d\sqrt{\delta}/2$. It follows that $|a|n^{1-p} \le c\theta/(1-d\sqrt{\delta})^{1-p}$, since otherwise for all x such that $|x| \ge n(1-d\sqrt{\delta}) > \sqrt{\delta}n$ we would have $P(x) = 0$ so $|P(x)-ax| = |a||x|^{1-p}|x|^p > c\beta(x)$, contradicting the fact that $\mathrm{Dist}_{D_n}(P,T,c\beta) < d\sqrt{\delta}/2$. Thus, $|a|(\sqrt{\delta}n)^{1-p} \le c\theta(\sqrt{\delta}/(1-d\sqrt{\delta}))^{1-p}$. Since, $d \le 1$ and $\delta < 1/4$, we get that $|a|(\sqrt{\delta}n)^{1-p} < c\theta$. But, for $x \neq 0$ such that $|x| \le d\sqrt{\delta}n$ we have that $|x| \le \sqrt{\delta}n$ so $|P(x)-T(x)| = |\theta(\sqrt{\delta}n)^p - a|x|| \ge \theta(\sqrt{\delta}n)^p - |ax| \ge (\theta(d)^{-p} - |a||x|^{1-p})|x|^p \ge (2c\theta - |a|(\sqrt{\delta}n)^{1-p})|x|^p > c\beta(x)$ contradicting the fact that $\mathrm{Dist}_{D_n}(P,T,c\beta) < d\sqrt{\delta}/2$. Hence, it must be that $\mathrm{Dist}_{D_n}(P,T,c\beta) \ge d\sqrt{\delta}/2$.

3.2. Polynomials

In this section we prove approximate robustness, in the relative error sense, for the interpolation identity $\nabla_t^d f(x) = 0$. As explained in Section 2.3, we establish such a result on a coarser domain than the one where this identity approximately holds. From this, if kD_n denotes $\{kx \in \mathbb{Z} : x \in D_n\}$, we get:

THEOREM 3.1. *Let $\theta > 0$, $p \in [0,1)$, $\delta \in [0,1]$, and d be a positive integer. Furthermore, let $\mu_d = \mathrm{lcm}\{1,\ldots,d\}$, and $P\colon D_{2(d+1)^3\mu_d n} \to \mathbb{R}$ be such that*

$$\Pr_{x,t}\left[|\nabla_t^d P(x)| > \theta\max\{|x|^p,|t|^p\}\right] \;\le\; \delta/(16(d+1)^5)\,,$$

where x and t are randomly chosen in $D_{d(d+1)^2\mu_d n}$ and $D_{d(d+1)\mu_d n}$ respectively. Then, there exists a constant $C = 2^{\Theta(d\log(dC_p))}$ and a polynomial $h_{d-1}\colon \mu_d D_n \to \mathbb{R}$ of degree at most $d{-}1$ such that

$$\Pr_{x\in\mu_d D_n}\left[|P(x)-h_{d-1}(x)| > C\theta|x|^p\right] \;\le\; \mu_{d-1}\delta + d\sqrt{\delta}\,.$$

To prove Theorem 3.1 we proceed as in the proof of Theorem 1.2 but now based on the analogues of Lemma 3.1 and Lemma 3.2 stated below. Indeed, let $m = \mu_d\cdot dn$, $\alpha_k = (-1)^{k+1}\binom{d}{k}$, and associate to P a function $g\colon D_{(d+1)m} \to \mathbb{R}$ such that

$$g(x) \;=\; \begin{cases} \underset{t\in D_{|x|}}{\mathrm{Med}}\left(\displaystyle\sum_{k=1}^{d}\alpha_k P(x+kt)\right), & \text{if } |x| \ge \sqrt{\delta}m\,,\\[2ex] \underset{t\in D_{\sqrt{\delta}m}}{\mathrm{Med}}\left(\displaystyle\sum_{k=1}^{d}\alpha_k P(x+kt)\right), & \text{otherwise}\,.\end{cases}$$

(Observe that $\nabla_t^d P(x) = 0$, if and only if, $P(x) = \sum_{k=1}^{d}\alpha_k P(x+kt)$. This motivates g's definition.)

LEMMA 3.3. *Let* $\phi(z) = \max\{\sqrt{\delta}m, |z|\}$, *where* $m = \mu_d{\cdot}dn$. *Under the hypothesis of Theorem 3.1, if* $x \in \mu_d D_n$ *is randomly chosen, then* $|P(x) - g(x)| > \theta\phi(x)^p$ *with probability at most* $\mu_d\delta/(4(d{+}1))$.

Proof. Let $P_{x,t} = \sum_{i=0}^{d} \alpha_i P(x{+}it)$. By definition of g and Markov's inequality,

$$\Pr_{x\in D_m}[|g(x){-}P(x)| > \theta\phi(x)^p] \;=\; \Pr_{x\in D_m}\left[|\operatorname*{Med}_{t\in D_{\phi(x)}} P_{x,t}| > \theta\phi(x)^p\right] \;\le\; 2\Pr_{x\in D_m, t\in D_{\phi(x)}}[|P_{x,t}| > \theta\phi(x)^p] \,.$$

But, $\phi(x) \ge \phi(t) \ge |t|$ and $\phi(x) \ge |x|$ for $x \in D_m$ and $t \in D_{\phi(x)}$ together with the halving principle imply that

$$\Pr_{x\in D_m, t\in D_{\phi(x)}}[|P_{x,t}| > \theta\phi(x)^p] \le \frac{|D_{(d+1)^2 m}| \cdot |D_{(d+1)m}|}{|\{(x,t) : x\in D_m, t\in D_{\phi(x)}\}|} \Pr_{u,s}[|P_{u,s}| > \theta\max\{|u|^p, |s|^p\}] \,,$$

where u and s are randomly chosen in $D_{(d+1)^2 m}$ and $D_{(d+1)m}$ respectively. Since $|P_{u,s}| = |\nabla_s^d P(u)|$, the hypothesis implies that $\Pr_{x\in D_m, t\in D_{\phi(x)}}[|P_{x,t}| > \theta\phi(x)^p] \le 2(d{+}1)^3\delta/(16(d{+}1)^5) = \delta/(8(d{+}1)^2)$. Thus,

$$\Pr_{x\in D_m}[|g(x){-}P(x)| > \theta\phi(x)^p] \;\le\; \delta/(4(d{+}1)^2) \,.$$

By the halving principle,

$$\Pr_{x\in \mu_d D_n}[|P(x){-}g(x)| > \theta\phi(x)^p] \le (m/n)\Pr_{x\in D_m}[|P(x) - g(x)| > \theta\phi(x)^p] \,.$$

Recalling that $m = \mu_d \cdot dn$, we obtain the desired conclusion. ■

LEMMA 3.4. *Let* $\phi(z) = \max\{\sqrt{\delta}m, |z|\}$, *where* $m = \mu_d{\cdot}dn$. *Under the hypothesis of Theorem 3.1, if* $x, t \in D_m$, *then* $|\nabla_t^d g(x)| \le 2^{2d} O(d^4)\theta\max\{\phi(x)^p, \phi(t)^p\}$.

Proof. First we show that for all $c \in D_{(d+1)m}$ and $I \subseteq D_{(d+1)m}$ such that $|I| \ge \sqrt{\delta}m + 1$,

$$\Pr_{t'\in I}\left[|g(c) - \sum_{j=1}^{d} \alpha_j P(c{+}jt')| > (d{+}1)^2 2^{d+1}\theta\max\{\phi(c)^p, \phi(t')^p\}\right] \;<\; \frac{1}{2(d{+}1)} \,. \quad (3)$$

Let $P_{x,t} = \sum_{i=0}^{d} \alpha_i P(x{+}it)$ and $\Phi_{x,t} = \max\{\phi(x)^p, \phi(t)^p\}$. Note that, by Markov's inequality, the LHS of (3) is upper bounded by

$$2\Pr_{t\in D_{\phi(c)}, t'\in I}\left[|P_{c,t}{-}P_{c,t'}| > (d{+}1)^2 2^{d+1}\theta\Phi_{c,t'}\right] \,.$$

But, if t and t' are randomly chosen in $D_{\phi(c)}$ and I respectively, then

$$\Pr_{t,t'}\left[|P_{c,t}{-}P_{c,t'}| > (d{+}1)^2 2^{d+1}\theta\Phi_{c,t'}\right]$$

$$\leq \sum_{i=1}^{d} \Pr_{t,t'}\left[|\alpha_i P_{c+it,t'}| > (d+1)2^d\theta\Phi_{c,t'}\right] + \sum_{j=1}^{d} \Pr_{t,t'}\left[|\alpha_j P_{c+jt',t}| > (d+1)2^d\theta\Phi_{c,t'}\right]$$

$$\leq \sum_{i=1}^{d} \Pr_{t,t'}\left[|P_{c+it,t'}| > \theta \max\{|c+it|^p, |t'|^p\}\right]$$

$$+ \sum_{j=1}^{d} \Pr_{t,t'}\left[|P_{c+jt',t}| > \theta \max\{|c+jt'|^p, |t|^p\}\right]$$

$$\leq 2d\frac{|D_{(d+1)^2 m}| \cdot |D_{(d+1)m}|}{|D_{\phi(c)}| \cdot |I|} \Pr_{x\in D_{(d+1)^2 m}, t\in D_{(d+1)m}}\left[|P_{x,t}| > \theta \max\{|x|^p, |t|^p\}\right] ,$$

where the first inequality follows from the union bound, the second one is due to the fact that $2^d \geq |\alpha_k|$, and that $(d+1)|c| \geq |c+kt|$ and $(d+1)\max\{|c|, |t'|\} \geq \max\{|c+kt'|, |t|\}$ for $k \in \{0, \ldots, d\}$, and the third one follows from the halving principle. Note that $|P_{x,t}| = |\nabla_t^d P(x)|$ and $|D_{(d+1)^2 m}| \cdot |D_{(d+1)m}|/(|D_{\phi(c)}| \cdot |I|) < 2(d+1)^3/\delta$. Putting everything together and using the hypothesis yields (3).

Now, to prove the lemma, let $G_{z,s} = g(z) - \sum_{j=1}^{d} \alpha_j P(z+js)$. Fix $x, t \in D_m$. Consider, in (3), $c = x$ and $I = \left\{t + jt' : t' \in D_{\sqrt{\delta m}}\right\}$ and also $c = x + it$ and $I = \left\{it' : t' \in D_{\sqrt{\delta m}}\right\}$ for $i, j \in \{1, \ldots, d\}$. We get that with nonzero probability there exists $t' \in D_{\sqrt{\delta m}}$ such that $|G_{x,t+jt'}| \leq (d+1)^2 2^{d+1}\theta\Phi_{x,t+jt'}$ and $|G_{x+it,it'}| \leq (d+1)^2 2^{d+1}\theta\Phi_{x+it,it'}$ for $i, j \in \{0, \ldots, d\}$. Since $\phi(t') = \sqrt{\delta}m \leq \Phi_{x,t}$, we have that $\Phi_{x,t+jt'}$ and $\Phi_{x+it,it'}$ are upper bounded by $(d+1)\Phi_{x,t}$. Thus, since $\sum_{j=0}^{d} \alpha_j = 0$ and $2^d \geq |\alpha_k|$,

$$|\nabla_t^d g(x)| = |\sum_{i=0}^{d} \alpha_i g(x+it)| = |\sum_{i=1}^{d} \alpha_i G_{x+it,it'} - \sum_{j=1}^{d} \alpha_j G_{x,t+jt'}|$$

$$\leq \sum_{i=1}^{d} |\alpha_i G_{x+it,it'}| + \sum_{j=1}^{d} |\alpha_j G_{x,t+jt'}| \leq 2^{2d} O(d^4)\theta \max\{\phi(x)^p, \phi(t)^p\} .$$

■

Theorem 3.1 now follows by observing that under its hypothesis Lemmas 3.4, 2.4, and 2.5 imply that there is a degree $d-1$ polynomial $h_{d-1}\colon \mu_d D_n \to \mathbb{R}$ such that for all $x \in D_n$,

$$|g(\mu_d \cdot x) - h_{d-1}(\mu_d \cdot x)| \leq \theta 2^{3d}(\mu_d d)^p O(d^4) \prod_{i=1}^{d-1}(1+iC_p)\phi(x)^p$$

$$\leq \theta 2^{3d} O(d^5) C_p^d (d!)^2 \phi(x)^p$$

$$\leq \theta 2^{\Theta(d\log(dC_p))}\phi(x)^p .$$

Lemma 3.3 now implies that if $x \in \mu_d D_n$ is randomly chosen, then $|P(x) - h_{d-1}(x)| > C\theta\phi(x)^p$ with probability at most $\mu_d\delta/(4(d+1)) \leq \mu_{d-1}\delta$. To conclude the proof observe that $\Pr_{x\in\mu_d D_n}[\phi(x) = |x|] \geq 1 - d\sqrt{\delta}$.

4. TESTING WITH RELATIVE ERROR

In this section we show how we can obtain approximate relative error testers from the results derived so far. First, we introduce some basic terminology. Let $\mathcal{F}$ be a collection of real valued functions over D. For functions $P, f, \beta\colon D \to \mathbb{R}$ let

- $\mathrm{Dist}_D(P, f, \beta) = \mathrm{Pr}_{x\in D}\left[|P(x) - f(x)| > \beta(x)\right]$ — the *β-relative distance of P from f on D.*
- $\mathrm{Dist}_D(P, \mathcal{F}, \beta) = \mathrm{Inf}_{f\in\mathcal{F}}\,\mathrm{Dist}_D(P, f, \beta)$ — the *β-relative distance of P from $\mathcal{F}$ on D.*

The definition of a self–tester requires the notion of a probabilistic oracle program T, i.e., a program that can call as a subroutine another program P. As usual, we let T^P denote T making calls to the program P.

Definition 4.1. Let $0 \le \delta < \delta' \le 1$, $D' \subseteq D$, and $\mathcal{F}$ be a collection of real valued functions defined over D. Let β and β' be real valued functions also defined over D. A *$(D, \beta, \delta; D', \beta', \delta')$–self–tester for $\mathcal{F}$* is a probabilistic oracle program T which is allowed to make calls to a program $P\colon D \to \mathbb{R}$ and is such that on input $\gamma > 0$ (the confidence parameter),

- if the β–relative distance of P from $\mathcal{F}$ on D is at most δ, then the tester T on input γ outputs $PASS$ with probability at least $1-\gamma$, i.e.,

$$\mathrm{Dist}_D(P, \mathcal{F}, \beta) \le \delta \implies \Pr\left[T^P(\gamma) \text{ outputs PASS }\right] \ge 1-\gamma\,;$$

- if the β'–relative distance of P from $\mathcal{F}$ on D is at least δ', then the tester T on input γ outputs $FAIL$ with probability at least $1-\gamma$, i.e.,

$$\mathrm{Dist}_D(P, \mathcal{F}, \beta') \ge \delta' \implies \Pr\left[T^P(\gamma) \text{ outputs FAIL }\right] \ge 1-\gamma\,.$$

(Both probabilities are taken over the coin tosses of T.)

Usually one requires that a tester be different and simpler than any correct program for the function purportedly computed by P. A convenient, although sometimes too restrictive, way of enforcing this is to have the tester comply with the *little–oh property* [5], i.e., have its running time be asymptotically less than any correct program, where each call to the program counts as one time step in the tester's computation. Also, it is commonly assumed that the tester is faultless and performs exact computations. Nevertheless, our results remain valid in a less restrictive model similar to the one described in [2].

In what follows it will sometimes be convenient to allow a tester to have access to another oracle function ψ. In such a case we say that the tester has *ψ–help.* Initially, we build testers which receive as help a *c–testable error term of degree $p \in \mathbb{R}$*, i.e., such valid error terms $\beta(\cdot,\cdot)$ for which $\beta(s,s) + \beta(t,t) + \beta(s{+}t, s{+}t) \le c\beta(s,t)$ for some constant c. For the sake of clarity of exposition, from now on we restrict our discussion to the 4–testable error terms such as $\theta\max\{|s|^p, |t|^p\}$ or $\theta(|s|^p + |t|^p)$, where $\theta > 0$. It is a simple matter to re–derive the results that follow for the more general case of c–testable error terms.

The results presented in previous sections concerning linear functions allow us to prove the following

THEOREM 4.1. *Let $c, c' > 0$ be such that $6c \leq 1/2$ and $(6c'/7)^2 \geq 2$. Let $0 \leq \delta \leq 1/c'$, $\beta(\cdot,\cdot)$ be a 4–testable error term of degree $0 < p < 1$. Then, there is a $(D_{8n}, \beta/4, c\delta/384; D_n, 17C_p\beta, c'\sqrt{\delta})$–self–tester with $\beta(\cdot,\cdot)$–help for the class of real valued linear functions over D_{8n}. Moreover, the tester satisfies the little–oh property.*

Proof. Let N be a fixed positive integer whose value will be determined later. The tester T performs N independent rounds of the following experiment: randomly choose $x, y \in D_{4n}$ and verify whether $|P(x+y) - P(x) - P(y)| > \beta(x, y)$. If the inequality is satisfied we say that the round fails. If more than a $\delta/384$ fraction of the rounds fail, then T outputs FAIL, otherwise T outputs PASS. Given that T has oracle access to both P and β, that it can add/subtract and perform comparisons exactly and efficiently, T satisfies the little–oh property.

Suppose the linear function $l\colon D_{8n} \to \mathbb{R}$ is such that $\mathrm{Dist}_{D_{8n}}(P, l, \beta/4) \leq c\delta/384$. Then, by the halving principle, we have that

$$\begin{aligned} \Pr_{x \in D_{4n}}\left[|P(x) - l(x)| > \tfrac{\beta(x,x)}{4}\right] &\leq \frac{2c\delta}{384}, \\ \Pr_{y \in D_{4n}}\left[|P(y) - l(y)| > \tfrac{\beta(y,y)}{4}\right] &\leq \frac{2c\delta}{384}, \\ \Pr_{x,y \in D_{4n}}\left[|P(x+y) - l(x+y)| > \tfrac{\beta(x+y,x+y)}{4}\right] &\leq \frac{2c\delta}{384}. \end{aligned}$$

Hence, since $\beta(s,s) + \beta(t,t) + \beta(s+t,s+t) \leq 4\beta(s,t)$, the union bound implies that a round fails with a probability of at most $6c\delta/384 \leq \frac{1}{2}(\delta/384)$. A standard Chernoff bound argument yields that if $N = \Omega(\frac{1}{\delta}\log\frac{1}{\gamma})$ the probability that T outputs FAIL is at most γ. Suppose now that $\mathrm{Dist}_{D_n}(P, l, 17C_p\beta) > c'\sqrt{\delta}$ for all real valued linear functions over D_n. Then, Theorem 1.2 implies that the probability that a round fails is at least $(6c'/7)^2(\delta/384) \geq 2(\delta/384)$. Again, if $N = \Omega(\frac{1}{\delta}\log\frac{1}{\gamma})$ the desired conclusion follows from a Chernoff bound. ∎

The self–tester of Theorem 4.1 needs oracle access to the error function. In the context of this work, this is an unrealistic assumption. Also, the tester will not comply with the little–oh property if it has to evaluate a hard to compute help function. We would like to have testers that achieve their goals, comply with the little–oh property, and do not have oracle access to the error function. Surprisingly, this is feasible, as our next stated result shows, provided the testable error function $\beta(\cdot,\cdot)$ is such that for some positive constants λ and λ' there is a function $\varphi(\cdot,\cdot)$ that is (λ, λ')–*equivalent* to $\beta(\cdot,\cdot)$, i.e., $\lambda\varphi(s,t) \geq \beta(s,t) \geq \lambda'\varphi(s,t)$ for all integers s, t. In addition, evaluating φ should be asymptotically faster than executing the program being tested. For example, let k be a positive integer, let k' be an integer, and let $\lg(n)$ denote the length of an integer n in binary. (Note that $\lg(n) = \lceil\log_2(|n|+1)\rceil$ or equivalently $\lg(0) = 0$ and $\lg(n) = \lfloor\log_2(|n|)\rfloor + 1$ if $n \neq 0$.) Then, $\beta(s,t) = 2^{k'}(|s|^{1/2^k} + |t|^{1/2^k})$ or $\beta(s,t) = 2^{k'}\max\{|s|^{1/2^k}, |t|^{1/2^k}\}$ are testable error terms of degree $1/2^k$ which are $(1, 1/2)$–equivalent to $\varphi(s,t) = 2^{k'}(2^{\lceil\lg(s)/2^k\rceil} + 2^{\lceil\lg(t)/2^k\rceil})$ and $\varphi(s,t) = 2^{k'+\max\{\lceil\lg(s)/2^k\rceil, \lceil\lg(t)/2^k\rceil\}}$ respectively. The computation of these latter functions requires only counting and shifting bits.

THEOREM 4.2. *Under the same hypothesis of Theorem 4.1, if* $\varphi(\cdot,\cdot)$ *is* (λ,λ')*–equivalent to* $\beta(\cdot,\cdot)$*, then there is a* $(D_{8n}, \beta/(4\lambda), c\delta/384\ ;\ D_n, 17C_p\beta/\lambda', c'\sqrt{\delta})$*–self–tester (without help) for the class of real valued linear functions over* D_{8n}*. Moreover, the tester satisfies the little–oh property provided* φ *is easy to compute relative to the cost of executing the program being tested.*

Proof. The proof is almost identical to that of Theorem 4.1 except that now the tester T performs N independent rounds of the following experiment; randomly choose $x, y \in D_{4n}$ and verify whether $|P(x+y) - P(x) - P(y)| > \varphi(x, y)$. ■

Similar results follow for the class of polynomials based on the results presented in Section 2.3 and Section 3.2.

5. FINAL COMMENTS

The error terms considered in this work do not depend on the function f purportedly computed by the program P which we wish to test. We leave open the following

Problem: For a class of real valued functions $\mathcal{F}$, and given constants $c > c' > 0$, find a simple and efficient self–tester which, with high probability, for every program P,

- Outputs PASS if $\Pr_{x\in D}\left[|P(x) - f(x)| > c|f(x)|\right]$ is at most δ for some function $f \in \mathcal{F}$.
- Outputs FAIL if $\Pr_{x\in D'}\left[|P(x) - f(x)| > c'|f(x)|\right]$ is at least δ' for all functions $f \in \mathcal{F}$.

In particular, what can be said when $\mathcal{F}$ is; (1) the class of real valued linear functions, (2) the class of real valued polynomials of degree at most d, and (3) the class whose only member is the map $x \in D \subseteq \mathbb{R} \to x^{-1}$.

ACKNOWLEDGMENT

We would like to thank Stéphane Boucheron for useful discussions.

REFERENCES

1. M. Albert and J. Baker. Functions with bounded nth difference. *Annales Polonici Mathematici*, 43:93–103, 1983.
2. S. Ar, M. Blum, B. Codenotti, and P. Gemmell. Checking approximate computations over the reals. In *Proc. 25th STOC*, pp. 786–795, 1993.
3. S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and intractability of approximation problems. In *Proc. 33rd FOCS*, pp. 14–23, 1992.
4. D. Beaver and J. Feigenbaum. Hiding instances in multioracle queries. In *Proc. 7th STACS*, LNCS 415, pp. 37–48, 1990.
5. M. Blum and S. Kannan. Designing programs that check their work. In *Proc. 21st STOC*, pp. 86–97, 1989.
6. M. Blum, M. Luby, and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. *J. of Comp. and Syst. Sci.*, 47:549–595, 1993.
7. M. Blum. Designing programs to check their work. Technical Report TR-88-009, ICSI, 1988.

8. M. Blum and H. Wasserman. Program result-checking: A theory of testing meets a test of theory. In *Proc. 35th FOCS*, pp. 382–392, 1994.
9. D. Coppersmith. Manuscript. See discussion in [6], December 1989.
10. D. Djoković. A representation theorem for $(X_1 - 1)(X_2 - 1) \dots (X_n - 1)$ and its applications. *Annales Polonici Mathematici* , 22:189–198, 1969.
11. F. Ergün, S. Ravi Kumar, and R. Rubinfeld. Approximate checking of polynomials and functional equations. In *Proc. 37th FOCS*, pp. 592–601, 1996.
12. G. L. Forti. Hyers–Ulam stability of functional equations in several variables. *Aeq. Mathematicae*, 50:143–190, 1995.
13. Z. Gajda. Local stability of the functional equation characterizing polynomial functions. *Ann. Polonici Mathemtici*, 42:119–137, 1990.
14. O. Goldreich, S. Goldwasser, and D. Ron. Property testing and its connection to learning and approximation. In *Proc. 37th FOCS*, pp. 339–348, 1996.
15. P. Gemmell, R. Lipton, R. Rubinfeld, M. Sudan, and A. Wigderson. Self-testing/correcting for polynomials and for approximate functions. In *Proc. 23rd STOC*, pp. 32–42, 1991.
16. D. H. Hyers. On the stability of the linear functional equation. *Proc. of the National Academy of Science, U.S.A.*, 27:222–224, 1941.
17. D. H. Hyers and T. M. Rassias. Approximate homomorphisms. *Aeq. Mathematicae*, 44:125–153, 1992.
18. D. H. Hyers and P. Šemrl. On the behaviour of mappings which do not satisfy Hyers–Ulam stability. *Proc. American Mathematical Society*, 114(4):989–993, April 1992.
19. D. E. Knuth. *The Art of Computer Programming*, Vol. 2, Seminumerical Algorithms. Addison-Wesley, 3rd edition, 1998.
20. R. Lipton. *New directions in testing*, Vol. 2 of *Series in Discrete Mathematics and Theoretical Computer Science*, pp. 191–202. ACM/AMS, 1991.
21. F. Magniez. Multi-linearity self-testing with relative error. In *Proc. 17th STACS*, LNCS 1770, pp. 302–313, 2000.
22. T. M. Rassias. On the stability of the linear mapping in Banach spaces. *Proc. of the American Mathematical Society*, 72(2):297–300, November 1978.
23. R. Rubinfeld and M. Sudan. Testing polynomial functions efficiently and over rational domains. In *Proc. 3rd SODA*, pp. 23–32, 1992.
24. R. Rubinfeld and M. Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM J. on Computing*, 25(2):252–271, April 1996.
25. R. Rubinfeld. *A mathematical theory of self-checking, self-testing and self-correcting programs.* PhD thesis, University of California, Berkeley, 1990.
26. R. Rubinfeld. On the robustness of functional equations. *SIAM J. on Computing*, 28(6):1972–1997, 1999.
27. F. Skof. Sull'approssimazione delle applicazioni localmente δ-additive. *Atti della Accademia delle Sciencze di Torino*, 117:377–389, 1983. (In Italian.)