A Sub-Constant Error-Probability PCP Characterization of NP

PART II: The Consistency Test
(Preliminary Version)
DRAFT

Ran Raz
Weizmann Inst. of Science
ISRAEL

Shmuel Safra
Tel-Aviv University
ISRAEL

September 24, 1996

Abstract

This paper introduces a new consistency-test for a class of codes, referred to as geometric-codes, and proves the test to be of small error-probability.

This consistency-test enables us to conclude a strong characterization of NP in terms of PCP. Specifically, our theorem states that, for any given $\epsilon > 0$, membership in any NP language can be verified with $O(1)$ accesses, each reading logarithmic number of bits, and such that the error-probability is as low as $2^{-\log^{1-\epsilon} n}$. Our results are in fact stronger, as stated in Appendix 2.

Previous characterizations of NP in terms of PCP have managed to achieve, with constant number of accesses, error-probability of, at best, a constant.

Low-degree polynomials are a special case of geometric-codes, hence our consistency-test implies a low-degree–test, which is the first to exhibit sub-constant probability of error. The proof for the small error-probability consistency-test is, nevertheless, simpler (combinatorial and geometrical in nature) than previous proofs, which achieved only constant probability of error.

In Part I of this work, a complete exposition of the proof for the improved PCP theorem is given[1]. The exposition there starts off with the general scheme of [AS92], which is altered, using several new ideas as well as known ones, so as to take advantage of the test described herein to obtain the results claimed.

1 Introduction

Characterizing the class NP, by itself or with respect to other computational-complexity classes, is perhaps one of the most fundamental avenues of research in theory of computer-science.

[1]This draft version of PART II somehow appears before PART I

Since the early days, when the classes P and NP were defined, and the question was posed as to whether they are the same or do they differ, many problems were shown to be NP-complete, thereby increasing the weight on finding stricter characterization for the class NP.

NP has since been given a few alternative characterizations. The one most commonly applied being Cook's [Coo71], which characterizes NP in terms of efficient verification of proofs (or nondeterministic computations).

A new perspective, by which improved characterizations of NP can be obtained, has been recently proposed. The motivation for which stems from questions regarding the hardness of approximation versions, for problems whose exact computation is known to be NP-hard.

This avenue of research was initiated by [FGL$^+$91], which introduced a new methodology for proving hardness results for approximation problems. The method takes advantage of results in a seemingly unrelated area — that of interactive proofs [GMR89, Bab85, BGKW88, LFKN92, Sha92, BFL91] — however interprets those results with quite a different perspective in mind.

Much effort has been invested since towards a better understanding of this methodology, and the class NP has consequently gained stricter characterizations [AS92, ALM$^+$92, BGS95], which are referred to as *characterizations of NP in terms of PCP* (or, in short, PCP characterizations of NP).

The PCP characterization of NP — though has taken around 20 years to be formulated — seems now as the most natural extension of the old characterization of NP [Coo71], if one has in mind proving hardness results for approximation problems. This characterization has already been used to obtain quite a few hardness results for approximation problems [FGL$^+$91, AS92, ALM$^+$92, PY91, LY94, BGLR93, KLS93, BGS95, Hås96, Hås96].

Previous proofs for PCP characterizations of NP tend to be very complicated, imposing a major obstacle for any attempt of further research in the area.

The previous characterization of NP in terms of the PCP hierarchy [AS92, ALM$^+$92], seemed at first as the best possible up to constant factors. A stronger characterization was later conjectured in [BGLR93]; one that, as an immediate outcome, implies NP-hardness of approximating set-cover to within logarithmic factors [LY94, BGLR93]. The conjecture itself, nonetheless, seems to be of independent interest, allowing a deeper understanding into probabilistic checking of proofs (PCP) and the class NP.

1.1 Characterizations of NP

The known PCP characterization of NP and the one obtained herein are surveyed next. Let us first present (a slight variation of) Cook's characterization of NP.

Cook's characterization of NP

Theorem 1 ([Coo71]) *For every $L \in NP$ there exists a polynomial time procedure that, on input $I \in \{0,1\}^n$, constructs a set of Boolean functions*

$$\Phi_{L,I} = \{\varphi_1, \dots, \varphi_l\}$$

over Boolean variables $\mathcal{Y} = \{y_1, ..., y_l\}$ — representing the guess bits (or witness) — such that

1. *Each $\varphi_i \in \Phi_{L,I}$ depends on $O(1)$ variables in $\mathcal{Y}$*
2. *$I \in L$ if and only if there exists an assignment $A: \mathcal{Y} \to \{0, 1\}$ such that all $\varphi_i \in \Phi_{L,I}$ evaluate to* `true` *on A.*

PCP Characterization of NP

The characterization of NP in terms of PCP differs from Cook's characterization only in regards to how many of the Boolean functions can be simultaneously satisfied in case the input I is not in L (see condition 2):

Theorem 2 ([AS92, ALM+92]) *For every $L \in NP$ there exists a polynomial time procedure that, on input $I \in \{0, 1\}^n$, constructs a set of Boolean functions*

$$\Phi_{L,I} = \{\varphi_1, ..., \varphi_l\}$$

over Boolean variables $\mathcal{Y} = \{y_1, ..., y_l\}$ such that

1. *Each $\varphi_i \in \Phi_{L,I}$ depends on $O(1)$ variables in $\mathcal{Y}$*
2. *For every input instance I:*
   - *if $I \in L$ then there exists an assignment A to the variables $\mathcal{Y}$ such that all $\varphi_i \in \Phi_{L,I}$ evaluate to* `true` *on A.*
   - *if $I \notin L$ then for any assignment A to the variables $\mathcal{Y}$, more than $\frac{l}{2}$ of $\varphi_i \in \Phi_{L,I}$ evaluate to* `false` *on A.*

In contrast to Cook's characterization, here, if the input I is not in L, at least half of the Boolean functions are unsatisfied, no matter which assignment is chosen. One should note that this variation, of the new characterization of NP from the old one, however syntactically small, is in fact not at all minor. It causes both a great increase in how difficult it is to prove the theorem, and, in return, enables quite stronger implications deduced.

1.2 Stronger PCP Characterizations of NP

The characterization of NP described above seemed, at first, as the best possible. The PCP hierarchy [AS92] was defined so as to consider the number of random bits (which is transparent in the formalism used herein) and the total number of answer bits (which is, according to our formalism, the number of bits each Boolean function depends on). The possibility of letting variables range over sets containing more than two elements was not considered in [AS92].

Once a larger range is allowed, the number of variables each Boolean function depends on may be considerably smaller than the total number of bits the function takes as input. One may therefore contemplate stronger characterizations of NP in terms of PCP. Namely, NP can possibly be characterized as in the above except that, while the number of variables would remain constant, their range could become larger, thereby achieving smaller, hopefully sub-constant, error-probability. The error-probability is defined to be the fraction of functions that can be satisfied simultaneously in case the original input is not in L.

A stronger characterization of NP along these lines, taking these parameters into consideration, was conjectured in [BGLR93] (formulated there for a somewhat different model). These two parameters — the number of variables each function depends on and the probability of error — play a crucial role in some applications of PCP characterizations of NP. The number of variables preferably remains constant while the probability of error would hopefully become exponentially small in the number of bits required to represent the value for each variable.

Proving that conjecture would directly imply NP-hardness for the problem of approximating set-cover to within logarithmic factors (see [BGLR93]). Nevertheless, the conjecture has the potential of an impact quite larger in scope, deepening our understanding with regards to the class NP.

According to a different perspective of the PCP theorem the above conjecture suggests that any mathematical proof (presented in the right format) can be verified by reading only a *constant* number of its pages, each consisting of m bits, and such that the probability of not detecting an error in the proof would be *exponentially small in m*.

PCP Characterization of NP — Strong Version

As mentioned above, the two differences, between the current PCP characterization of NP and the conjectured one, are in regards to the range of values for the variables, and the probability of error. The conjecture proceeds as follows:

Conjecture 3 ([BGLR93]) *For every $L \in NP$, and at most logarithmic m, there exists a polynomial time procedure that, on input $I \in \{0,1\}^n$, constructs a set of Boolean functions*

$$\Phi_{L,I} = \{\varphi_1, ..., \varphi_l\}$$

over variables $\mathcal{Y} = \{y_1, ..., y_l\}$ which range over $[1, ..., 2^m]$ such that

1. *Each $\varphi_i \in \Phi_{L,I}$ depends on $O(1)$ variables in $\mathcal{Y}$*
2. *For every input instance I:*
   - *if $I \in L$ then there exists an assignment A to the variables $\mathcal{Y}$ such that all $\varphi_i \in \Phi_{L,I}$ evaluate to* `true` *on A.*
   - *if $I \notin L$ then for any assignment A to the variables $\mathcal{Y}$, at most $\frac{l}{2^m}$ of $\varphi_i \in \Phi_{L,I}$ evaluate to* `true` *on A.*

The conjecture takes an extreme view, expecting every bit the functions depend on to decrease the probability of error by a factor of 2.

The barrier of constant error-probability (for a constant number of accesses whatever m is), however, seemed to have been unbreakable, so far.

1.3 Background and Our Results

Our results, concerning characterization of NP in terms of PCP, approaches the above conjecture, however, do not manage to go the entire way. For logarithmic m (recall that m denotes the number of bits required to represent each variable), the error-probability obtained is $2^{-m^{1-\epsilon}}$ for any $\epsilon > 0$. If a similar result for ϵ equal 0 were shown, the conjecture would follow. Our results in fact bring us closer to the above conjecture, as stated in appendix 2. The full proof for all these results can be found in [RS1].

[BGLR93] formulated the above conjecture, for a somewhat different model, in order to be applied to obtain NP-hardness result for approximating SET-COVER to within logarithmic factors. Indeed, combined with some observations, regarding the relation between the two models, and other properties the chracterization should comply with, NP-hardness for SET-COVER can be deduced. This is to be elaborated in the full version of the paper.

Our proof starts off with the general scheme of [AS92], as amended (following the technique of [LS91], improved in [FL92]) by [ALM+92]. This scheme, however, is insufficient when sub-constant error-probability is sought after. Further alterations are therefore required, in order to allow undisturbed flow of such small error-probability between various parts of the proof.

Of critical importance, in the development of the ideas leading so far, is the work of [GS96]. The work reported here is in fact an extension of the line of research pursued there. [GS96] shows how to avoid the all familiar low-degree test, and obtain instead a *consistency test*. Namely, a test that, assuming an arbitrary function f, enables consistent reading of tuples of several (say poly-logarithmic in number) values of f. A consistency-test of constant error-probability is presented there. This is in contrast to previous proofs for the old PCP theorem that assumed f to be a low-degree polynomial. The theorem of [GS96], although implying a low-degree test, is proven there using only combinatorial properties of the representation of f.

It follows that, for constant error-probability, one needs not assume any properties regarding the function which is to be read consistently. Our work here proceeds to show that in order to obtain sub-constant error-probability, one can incorporate some restrictions regarding f. Namely, requiring it to comply with some conditions akin to properties of error-correcting–codes.

Functions over geometric spaces, represented by their values over planes, which is the formalism used herein, are a special case of the format used in [GS96].

It seems hence to be the case that low-degree polynomials were previously taken advantage of for two different type of properties, which were however not distinguished. The first being the geometrical nature of their domain. The other being extended error-correcting properties. This distinction is now made explicit (and hopefully clear).

2 Low-Degree–Tests

A critical section of previous proofs for the old PCP theorem is the so called "low-degree–test". The analysis of the error-probability of that test is the most technical part of those proofs. This analysis, however, is not tight. An improved analysis (for possibly a different test) may imply a stronger PCP characterization of NP, and may at the same time simplify the entire proof.

Domains, Tables, Lines, and in between

Let $\mathcal{D}$ be a vector-space of dimension d over a field $\mathcal{F}$, that is,

$$\mathcal{D} = \mathcal{F}^d$$

and let $T: \mathcal{D} \rightarrow \mathcal{F}$ be an arbitrary function, which can be thought of as a table of values that has one entry for each element of $\mathcal{D}$. Such a table T can fall into one of the following categories:

1. T is a polynomial of low-degree, i.e. it can be described as a polynomial of d variables, and such that the total-degree of the polynomial is small — how small is determined by a parameter of the test, denoted r. In most cases the ratio $r/|\mathcal{F}|$ is taken to be very small (say $|\mathcal{F}|^{-1/2}$).

2. T is not a polynomial of low degree, however, there exists a low-degree polynomial $g: \mathcal{D} \rightarrow \mathcal{F}$, such that T and g agree on a large fraction of the domain $\mathcal{D}$ — how large a fraction is again given as a parameter of the test, denoted δ. T, in that case, is said to be *close* to g.

3. T is not a polynomial of low-degree, and is not close to any polynomial of low-degree.

A line in $\mathcal{D}$ is an affine subspace of $\mathcal{D}$ of dimension 1. Denote by $\mathcal{S}^1$ the set of all lines in $\mathcal{D}$. For every such line ℓ, let us consider the restriction of the function T to ℓ. If T is a degree-r polynomial[2] then, obviously, its restriction to each line is a uni-variate–polynomial of degree r.

Let now T be a table, as above, supposedly containing the values for each point of a certain low-degree polynomial. Let T' be a table, supposedly containing the restriction of T to every line of $\mathcal{D}$.

For $\ell \in \mathcal{S}^1$, denote by $T'(\ell)$ the ℓ^{th} element of T' (which is supposedly the restriction of T to ℓ). Given T and T', let us say that T' is *honest* if indeed, for every ℓ, $T'(\ell)$ is the restriction of T to ℓ.

The aim of the standard low-degree–test is to output `true` with probability 1 if T falls into the first category **and** T' is honest, however to output `false` with probability at least $1-\gamma$ (where γ is again a parameter of the test) if T falls into the third category (no matter what T' is in that case). If T falls into the second category, or alternatively, if T falls into the first one, however T' is not honest, then the output of the test may be arbitrary.

By a converse perspective, if the test outputs `false` with probability greater than 0, then either T' is not honest, or T is not a polynomial of low-degree. On the other hand, if the test outputs `true` with probability greater than γ, then T is either a polynomial of low-degree or is close to a certain polynomial of low-degree.

Note that the smaller δ and γ are, the stronger the test is.

[2] A polynomial of degree $< r$ is nonetheless of degree r as well

The Standard Test

The standard test proceeds as follows: Pick a random line $\ell \in \mathcal{S}^1$, hence ℓ contains $|\mathcal{F}|$ elements of $\mathcal{D}$. Now choose a random element $x \in \ell$, i.e., an element of $\mathcal{D}$ contained in ℓ. If $T'(\ell)$ is indeed a polynomial of low-degree, and $T'(\ell)$ agrees with $T(x)$, then the test outputs `true`. Otherwise, the test outputs `false`.

Clearly, if the test outputs `false` then either T' is not honest, or T is not a polynomial of low-degree. Therefore, to prove the correctness of the test, one just needs to show that if the test outputs `true` with probability greater than γ, then T is close to a polynomial of low-degree g (or, even better, T is exactly a polynomial of low-degree).

Such theorems were proven in [BLR90, Lip91, FSH94, GLR+91, RS92, She91, Rub90, Sud92] for some values of δ and γ. These proofs, however, do not generalize to the cases where γ is small, say $\gamma < 1/2$. In order to reduce the error-probability in the PCP theorem, one should prove the correctness of the test for much smaller values.

Assume now a low-degree test proven to satisfy the above requirements. Consider a two-phase process that first apply the test, choosing randomly one of its options and then, if the value returned is `true`, proceed under the assumption that T represents a certain polynomial g of low-degree. Such a process has an error-probability, i.e., the probability of accepting erroneous values, bounded from above by γ. In this case one can sample g on a random element x of $\mathcal{D}$ just by reading the value $T(x)$. By the correctness of the test, one can assume that $T(x) = g(x)$. This, of course, is not always true. The smaller γ and δ are, the safer it is to trust that assumption.

3 Consistency with Low Error-Probability

In the old low-degree test, just described, the table is first checked so that serious inconsistencies are detected with high probability. It is then assumed that the values obtained for the requested entries indeed correspond to a single global low-degree polynomial. The probability, over the random choice, of the test to be erroneously passed (i.e., when there is no global polynomial with which the values agree) is rather small.

When trying to achieve much smaller probabilities of error for the PCP theorem, however, such a two-phase process cannot do. E.g., it is possible to assign values to one subset of the entries to the table, that would cause the first phase (that of checking consistency) to succeed with small, however non-negligible, probability. At the same time, however, a disjoint subset of the entries can be assigned values that have nothing to do with those of the first subset. Values which the second phase would happily return with non-negligible probability. Hence, with tests that perform only a constant number of accesses, the barrier of constant error-probability cannot be broken.

Consistent Readers

Before describing the new test, let us shift to a more general paradigm by which to approach the issue. The procedure in quest is to be given a table, whose values supposedly comply with some global consistency condition (being of low degree is just one possible consistency condition). Given any entry to the table, the procedure is required to return, with very high probability, a value for the requested entry that would correspond to a function that conforms to the consistency condition. The procedure is to achieve that, while performing only a very small number of accesses to the table.

Let us dwell a little on what should be considered a faulty outcome for a random test of such a procedure. If the procedure just happens to step on values that cannot possibly agree with any globally consistent function, it might as well announce it. The table is supposedly everywhere consistent, hence any such inconsistency disqualifies it, and the procedure might as well halt.

Our attention therefore should be focused on the values the procedure returns in case no inconsistency is detected, and how safe it is to assume that the values returned correspond to a globally consistent function.

Requiring, as in the old test, that the values returned for most random tests correspond to a single consistent function, cannot be guaranteed when such small probabilities of error are at aim. E.g., it is possible to partition the set of entries to the table into several subsets, where entries in each subset are assigned values corresponding to a distinct consistent function. This can be carried out while the probability of all accesses of a random test to fall within one subset is non-negligible. No inconsistency can be detected in those cases. The values returned, however, clearly do not all (even most) correspond to a single function.

Limited Pluralism

Our requirements regarding the values returned by the procedure are therefore relaxed so as to allow several permissible functions. The functions allowed are those whose values agree with a non-negligible fraction of the entries in the table. The procedure is required to return *impermissible* values — i.e., values that do not correspond to any permissible function — only with a very small probability.

If, as in the case with low-degree polynomials, two globally consistent functions can agree on at most a small fraction of the entries, there can be only a small number of functions that would be considered permissible.

A short summary of the requirements of our test is as follows. Given a table supposedly describing a globally consistent function, the outcome for a particular random choice of the test is considered non-faulty in one of the following cases:

1. The values the test returns correspond to one of the permissible functions.

2. An inconsistency is announced, and the table is indeed inconsistent.

It should be the case that all other possibilities, where the outcome can be labeled faulty, occur only with a very small probability.

Implementation Issues

There are several other, implementation, aspects by which the new test vary from the old one.

The new test assumes the table contains the values for the restriction of the globally consistent function to *planes* (i.e. affine spaces of dimension 2). This replaces the two tables, one for lines and one for points, of the old test. If one is interested in accessing a single element of the function, one may obtain it via the value for one of the planes containing that element. Consistency is tested by choosing two random planes that intersect by a line, and checking that the values given to them in the table agree on the common line. This is in contrast to checking the consistency of the value for a line vs. the value on one of its points.

The New Low-degree Consistent-Reader

In the next section, the class of codes, referred to as "geometric-codes", is defined. A new consistency-test for such codes is introduced as well, and the correctness of that test is proven. Low-degree polynomials are a special case of this class of codes, hence a consistent reader for low-degree polynomials is implied. Let us describe here the low-degree test that results by restricting the consistency-test described below (for geometric codes) to the special case of low-degree polynomials.

Consider a degree-r polynomial $g: \mathcal{D} \to \mathcal{F}$. A plane in $\mathcal{D}$ is an affine subspace of $\mathcal{D}$ of dimension 2. Denote by $\mathcal{S}^2$ the set of all planes in $\mathcal{D}$. Consider, for every such plane $\wp$, the restriction of g to $\wp$. Since g is of degree r then, obviously, its restriction to each plane is a two-variables–polynomial of degree r.

Consider a table T consisting of one entry, $T(\wp)$, for each $\wp \in \mathcal{S}^2$; assume that that entry contains an r-degree polynomial, which is supposedly the restriction of g to $\wp$.

T is said to be ***globally-consistent*** if indeed there exists a polynomial g of degree r, such that for every $\wp$, $T(\wp)$ is the restriction of g to $\wp$. A degree-r polynomial g' is said to be T-permissible if T agrees with g' on a non-negligible, determined by a parameter denoted δ, fraction of the planes. E.g., δ may be fixed to be $\sqrt{r/|\mathcal{F}|}$.

Denote by $\mathcal{P}$ the set of all pairs of planes of $\mathcal{D}$ that intersect by a line:

$$\mathcal{P} = \left\{ \langle \wp_1, \wp_2 \rangle \, | \wp_1, \wp_2 \in \mathcal{S}^2 \text{ and } \wp_1 \cap \wp_2 \in \mathcal{S}^1 \right\}.$$

Consider a pair $\langle \wp_1, \wp_2 \rangle \in \mathcal{P}$ of planes of $\mathcal{D}$ that intersect by a line $\ell = \wp_1 \cap \wp_2$. The pair $\langle \wp_1, \wp_2 \rangle$ is said to be *T-pairwise-consistent* if $T(\wp_1)$ agrees with $T(\wp_2)$ on ℓ.

One consequence of Theorem 4 proven below would be as follows:

Corollary 1 *Given any table T, the fraction of pairs of planes $\langle \wp_1, \wp_2 \rangle \in \mathcal{P}$ that are T-pairwise-consistent however do not agree with any T-permissible polynomial g' is at most $O\left(d \cdot \left(\frac{r}{|\mathcal{F}|}\right)^{1/c} + \delta\right)$, where c is a (large) universal constant.* $\square^{\rightarrow}$ include proof

$^{\rightarrow}$ Let us now try to ease the mind of the reader regarding the specific value of agreement chosen here for a function to be defined as T-permissible. Denote by γ the probability of a pair $\langle \wp_1, \wp_2 \rangle \in \mathcal{P}$ to be T-pairwise-consistent. It is then the case that the definition (of a T-permissible function) can be strengthened so as to require the fraction of points that agree with T to be not much smaller than γ (specifically $\frac{1}{c'}\gamma^{c'}$, for a small constant c'). The corollary would still hold, although maybe with a slightly larger error-probability, for the strengthened definition. This is to be made explicit in the full version of the paper. improve

It follows that, given a random point $e \in \mathcal{D}$, one may choose a random pair of planes $\langle \wp_1, \wp_2 \rangle \in \mathcal{P}$, one of which containing e, and verify that $\langle \wp_1, \wp_2 \rangle$ is T-pairwise-consistent. If successful, one proceeds while trusting the value thus obtained for e to agree with one of the T-permissible low-degree polynomials. Only a small fraction of the random tests would lead to trust impermissible values.

Furthermore, by a slight extension of this test it is possible to obtain a similar consistent-reader for an arbitrary point e (rather than just for a random one).

4 Geometric-Codes and Consistency

Time is ripe to proceed with the definition of the class of geometric-codes and the consistency test applied to it. Let us first try motivate our choice to proceed with the more general class:

1. Once the PCP theorem is proven assuming only properties of geometric-codes, any code belonging to that class may be plugged in, leaving the proof intact.

2. This class of codes identifies a small number of properties — combinatorial and geometrical in nature — sufficient for the consistency test. The reader can therefore be sure not to be surprised at some point by some additional properties, pulled off some hat, regarding low-degree polynomials.

3. Our belief is that the consistency-test defined below is of independent interest, not necessarily confined within the scope of low-degree polynomials or PCP.

Geometry

Let us assume a vector-space $\mathcal{D}$ of dimension d over a field $\mathcal{F}$

$$\mathcal{D} = \mathcal{F}^d$$

and consider all of $\mathcal{D}$'s affine sub-spaces[3]; denote by $\mathcal{S}^i$ the set of affine sub-spaces of $\mathcal{D}$ of dimension i. We are interested in functions whose domain is one of those sub-spaces and of an arbitrary range R; let us thus denote

$$\mathcal{M}^i[R] = \left\{ f \mid f \colon \mathcal{F}^i \to R \right\}$$

and define, for each i, a metric space over the set of functions $\mathcal{M}^i[R]$: for $f, g \in \mathcal{M}^i[R]$, denote by $\Delta(f, g)$ the fraction of $\mathcal{F}^i$ on which f and g disagree.

Codes

An α-code e over a vector-space $\mathcal{D}$ and range R is a subset of the functions from $\mathcal{D}$ to R

$$\mathsf{e} \subseteq \mathcal{M}^d[R]$$

[3]A set of i vectors in $\mathcal{D}$ induces a vector sub-space of dimension i in $\mathcal{D}$; adding a fixed vector to all vectors in such a sub-space forms *an affine sub-space* of $\mathcal{D}$.

such that, for distinct $f \neq g \in \mathbf{e}$, the fraction of entries which f and g agree on is at most α, i.e.,

$$\Delta(f,g) \geq 1-\alpha.$$

A member of $\mathbf{e}$ is referred to as an *encoding.*

The presentation proceeds next to define a special type of an α-code. Let us first discuss shortly some technicalities regarding the value that α may take. For technical ease, α is assumed to be such that α is an integer multiple of $1/|\mathcal{F}|$. For the specific manner in which these codes are to be later applied, the reader might want to think of α as being very small, however still larger than $|\mathcal{F}|^{-1}$, (say $|\mathcal{F}|^{-1/2}$). Let us denote by $\alpha^+ = \alpha + |\mathcal{F}|^{-1}$, that is, α^+ is a fraction of points two distinct encodings of an α-code cannot agree on. Let also $a = \alpha|\mathcal{F}| + 1$.

Geometric-Codes

An encoding $g \in \mathbf{e}$ can be restricted onto any of the affine sub-spaces $S \in \mathcal{S}^i$, resulting in a function in $\mathcal{M}^i[R]$ denoted $g_{\downarrow S}$.[4]

Definition 1 *An α–*geometric-code $\mathbf{e}$ *is an α-code, with the additional following properties:*

1. **(Symmetry under) Restriction**[5]**:** *There exists a (1-dimensional) α-code $\mathbf{e}_1 \subset \mathcal{M}^1[R]$, such that $\mathbf{e}$ is exactly the set of functions whose restriction onto any line of $\mathcal{D}$ results in an encoding in $\mathbf{e}_1$, that is*

$$\mathbf{e} = \left\{ g \in \mathcal{M}^d[R] \,\middle|\, \forall \ell \in \mathcal{S}^1, g_{\downarrow \ell} \in \mathbf{e}_1 \right\}.$$

2. **Induction:** *There exist $\mathbf{e}_2, ..., \mathbf{e}_{d-1}$, each $\mathbf{e}_i$ being an i-dimensional α-code, so that the restriction of $\mathbf{e}$ onto any sub-space $S \in \mathcal{S}^i$, results in $\mathbf{e}_i$.*[6]

3. **Interpolation:** *Assume any $(i+1)$-dimensional subspace $S \in \mathcal{S}^{i+1}$, and some parallel i-dimensional sub-spaces of it $S_1, ..., S_a \subset S$. For any sub-encodings $g_1, ..., g_a \in \mathbf{e}_i$ over $S_1, ..., S_a$, there exists exactly one encoding $g \in \mathbf{e}_{i+1}$ over S such that for every j, $g_{\downarrow S_j} = g_j$.*

Let us specify here one clear, mentioned above, example for such codes:

Example: The set of all d-variables polynomials of total-degree r is an $\frac{r}{|\mathcal{F}|}$–geometric-code.

From now on, let us assume $\mathbf{e}$ to be an α–geometric-code, and $\mathbf{e}_1, ..., \mathbf{e}_d$ its restrictions. Note that each $\mathbf{e}_i$ can in fact be proven to be, by itself, an i-dimensional α-geometric-code.

[4]This is a bit of an abuse of notation, as the correspondence between $\mathcal{F}^i$ and $S \in \mathcal{S}^i$ is not determined. Any 1-to-1 linear transformation, however, will do.

[5]This property seems to be rather stronger than what is necessary for the proof. Weaker variants may be considered, thereby widening the set of codes to which these results apply.

[6]Property 2 seems to follow from property 1.

Plane Representation

Given an encoding $f \in \mathbf{e}$, one may consider its restriction to every plane $\wp \in \mathcal{S}^2$, so that each plane is assigned a value in $\mathbf{e}_2$. Let

$$F: \mathcal{S}^2 \to \mathbf{e}_2$$

be a function that assigns an arbitrary value in $\mathbf{e}_2$ to every plane of $\mathcal{D}$. The values assigned by F to the planes may be consistent — that is, every element of $\mathcal{D}$ is assigned the same value by F in all planes containing it — in which case F, by the restriction property (or, alternatively, by lemma 5 below) corresponds to a single encoding $f \in \mathbf{e}$.

The function F, however, may be assigning different values to different occurrences of some elements of $\mathcal{D}$, in which case there is no encoding f whose plane representation is F.

Of special interest to us herein is the case where F is locally consistent in a weak sense (as defined bellow). Our aim is to be able to conclude some sort of global consistency for F in that case.

Assume from now on F is as above.

Local Consistency

First, let us denote by $\mathcal{P}$ the set of all pairs of planes of $\mathcal{D}$ that intersect by a line:

$$\mathcal{P} = \left\{ \langle \wp_1, \wp_2 \rangle \, | \wp_1, \wp_2 \in \mathcal{S}^2 \text{ and } \wp_1 \cap \wp_2 \in \mathcal{S}^1 \right\}$$

and let $\mathrm{CONS}_F(\mathcal{P})$ be the set of all such pairs of planes $\langle \wp_1, \wp_2 \rangle \in \mathcal{P}$ that are given consistent values by F, i.e.,

$$F(\wp_1)_{\downarrow \ell} = F(\wp_2)_{\downarrow \ell}$$

where $\ell = \wp_1 \cap \wp_2 \in \mathcal{S}^1$. Local consistency is now defined as follows:

Definition 2 *F is* δ–pairwise–consistent *if*

$$\Pr_{\langle \wp_1, \wp_2 \rangle \in_{\mathrm{R}} \mathcal{P}} [\langle \wp_1, \wp_2 \rangle \in \mathrm{CONS}_F(\mathcal{P})] \geq \delta$$

where " $\in_{\mathrm{R}}$ " denotes a choice of an element according to the uniform distribution.

It is not hard to prove that if all pairs of planes in $\mathcal{P}$ are pairwise consistent (that is, F is 1-pairwise-consistent) then there exists an encoding f whose plane representation is F. We are interested however in the case where δ is much smaller, sub-constant in fact, (say $|\mathcal{F}|^{-1/100}$).

In that case, perfect global consistency does not necessarily hold, as the set of planes can be partitioned into a few subsets, each given values corresponding to a distinct encoding. This can be carried out while the probability of a pair of planes to belong to the same subset (hence the probability of a pair of planes to be in $\mathrm{CONS}_F(\mathcal{P})$) is $\geq \delta$.

Global Consistency

Nevertheless, a weaker type of global consistency — one however that still suffices for our purposes — is introduced herein.

The straightforward notion of global consistency would be that of a single encoding that F, on most planes, agree with. Assuming only weak local consistency, however, one cannot hope for such a strong global consistency, as mentioned above. Consequently, our notion of global consistency allows a small (say $\delta^{-O(1)}$) set of global encodings, and considers F's values on various planes to be permissible as long as they correspond to one of those encodings.

Still, F could be assigning impermissible values to most (say $1 - \sqrt{\delta}$) of the planes. It would suffice however, for our purposes, to be able to trust F's values on a random pair of planes in $\mathcal{P}$. Such a pair, if not in $\mathrm{CONS}_F(\mathcal{P})$, suits us fine as it supplies an immediate evidence to the fact that F is not the plane representation of a single global encoding — its values would therefore not be trusted. Our concern is only with pairs of planes in $\mathrm{CONS}_F(\mathcal{P})$, whose values however are not consistent with any of the permissible global encodings. The fraction of those (in $\mathcal{P}$) is referred to as the *probability of error* of F.

The *support* of an encoding $g \in \mathsf{e}$ is the set of all planes $\wp \in \mathcal{S}^2$ that are assigned by F values that agree with g, i.e., where $F(\wp) = g_{\downarrow \wp}$. An $[F, \mathsf{e}]$*-extension* is a function

$$A \colon \mathcal{S}^2 \to \mathsf{e}$$

labeling every plane $\wp$ by an encoding g so that $\wp$ is in the support of g, that is,

$$\forall \wp \in \mathcal{S}^2, A(\wp)_{\downarrow \wp} = F(\wp).$$

A is γ*-consistent* if at least γ fraction of the pairs of planes intersecting by a line are labeled the same by A, that is, if

$$\Pr_{\langle \wp_1, \wp_2 \rangle \in_R \mathcal{P}} [A(\wp_1) = A(\wp_2)] \geq \gamma.$$

Definition 3 *F is* γ-consistent *if there exists a γ-consistent $[F, \mathsf{e}]$-extension A.*

F being γ-consistent implies that F is γ–pairwise–consistent, as if two planes are assigned the same global encoding by A they must agree on their intersection. The other direction is nonetheless not necessarily true.

From Local to Global

The reader might note that γ-consistency, even for a rather small γ, implies that there is only a small number of global encodings whose support is not negligible. Our Ultimate Claim can now be stated:

Theorem 4 *There exists a global constant $c \geq 2$ so that for any δ, if F is δ–pairwise–consistent then F is $(\delta - \epsilon \cdot d)$-consistent, for $\epsilon = c \cdot \alpha^{1/c}$.*

In other words, theorem 4 implies that only a small ($\leq \epsilon \cdot d$) fraction of the pairs of planes in $\mathcal{P}$ can be given pairwise consistent values by F, however not belong to the support of one, out of a small number of, global encodings. That is, the probability of error of F is $\leq \epsilon \cdot d$.

The reader may be reminded that α should be thought of as very small, however still larger than $|\mathcal{F}|^{-1}$ (say $\alpha = |\mathcal{F}|^{-1/2}$). ϵ is clearly much larger than α, however still relatively small (say $\epsilon = O(|\mathcal{F}|^{-1/10})$). The dimension d, in our application of the theorem, would usually be logarithmic in $|\mathcal{F}|$. In addition, let us assume w.l.o.g that $\delta \geq \epsilon \cdot d$ (otherwise the theorem becomes trivial). δ, however, should still be thought of as much smaller than 1.

4.1 Proof of the Theorem

Proof: Theorem 4 is proven by induction on d, the dimension of $\mathcal{D}$: For $d = 2$ there is only one plane and nothing to prove.

Let us now consider the case where $d \geq 3$. For that purpose, some of the notations defined for $\mathcal{D}$ are now adopted to each $(d-1)$-dimensional sub-space $S \in \mathcal{S}^{d-1}$: denote by $\mathcal{S}^2[S]$ the set of planes in S, and by $\mathcal{P}[S]$ the set of pairs of planes in S that intersect by a line. In addition, let δ_S be the fraction of pairs of planes $\langle \wp_1, \wp_2 \rangle \in \mathcal{P}[S]$ that are assigned consistent values by F; and let A_S be an $[F_{\downarrow \mathcal{S}^2[S]}, \mathbf{e}_{d-1}]$-extension of maximal consistency for S.

Assuming now, by the induction hypothesis, that every $(d-1)$-dimensional sub-space S is $(\delta_S - \epsilon \cdot (d-1))$-consistent, we prove that $\mathcal{D}$ is $(\delta - \epsilon \cdot d)$-consistent.[7]

Note that for $d = 3$, $S \in \mathcal{S}^{d-1}$ is a plane, and therefore $\mathcal{P}[S]$, and δ_S are undefined, and the set of planes $\mathcal{S}^2[S]$ contains only S itself. In this case we define $A_S : \mathcal{S}^2[S] \to \mathbf{e}_{d-1}$ by $A_S(S) = F(S)$, i.e. A_S is the obvious extension.

Need for Support

Let us begin with a simple claim concerning geometric-codes, namely that for any plane $\wp$, $F(\wp)$, unless being the restriction of an encoding g, must differ from $g_{\downarrow \wp}$ almost everywhere, and can therefore be pairwise consistent only with a small number of planes in the support of g:

Claim 1 *Let $g \in \mathbf{e}$, and T be its support; for every plane $\wp \notin T$, let $\mathcal{N}_\wp = \{\wp' \,|\langle \wp, \wp' \rangle \in \mathcal{P}\}$, then*

$$\Pr_{\wp' \in_{\mathrm{R}} \mathcal{N}_\wp} [\wp' \in T \text{ and } \langle \wp, \wp' \rangle \in \mathrm{CONS}_F(\mathcal{P})] \leq \alpha.$$

Proof: For any $\wp'$ in the support of g such that $\langle \wp, \wp' \rangle \in \mathrm{CONS}_F(\mathcal{P})$, $F(\wp)_{\downarrow \ell} = F(\wp')_{\downarrow \ell} = g_{\downarrow \ell}$, where $\ell = \wp \cap \wp'$. A fraction larger than α of planes $\wp' \in \mathcal{N}_\wp$ thus fixes $F(\wp)$ to agree with g on more than α fraction of the lines in $\wp$. Since $F(\wp) \in \mathbf{e}_2$, which is an α-code, it must be that $F(\wp) = g_{\downarrow \wp}$.[8] $\square$

[7] The reader might be advised to go over the proof first considering only the case where d is 3 — where much of the technical analysis becomes simpler — and then verify the proof for the rest of the cases.

[8] Note that this analysis would still work for a fraction much smaller than α.

It follows from claim 1 that for any encoding $g \in \mathbf{e}$, the fraction of pairs in $\mathcal{P}$ that are assigned pairwise consistent values by F and such that one is in the support of g and the other outside, is small. More precisely,

Corollary 2 *Given an encoding $g \in \mathbf{e}$, let T be the support of g. Then, $(T \times (\mathcal{S}^2 \setminus T)) \cap \mathrm{CONS}_F(\mathcal{P})$ is of fraction $\leq \alpha$ in $\mathcal{P}$.* $\square$

Hence, almost all of $\mathrm{CONS}_F(\mathcal{P})$ is divided between pairs of planes both in T and pairs both outside T.

Our proof now proceeds to show that, given any subset W of the planes, as long as $\mathrm{CONS}_F(\mathcal{P}) \cap \binom{W}{2}$ is of fraction $\geq \theta_d \equiv \epsilon \cdot (d-1) + \frac{1}{2}\epsilon$ in $\mathcal{P}$, there exists an encoding $g \in \mathbf{e}$ of large (i.e., $\geq \sqrt{\alpha}$) support within W.

Such an assertion is sufficient in order to prove theorem 4, due to the following argument: Let g_1 be the encoding of largest support; $g_2 \in \mathbf{e}$ of largest support in the set of remaining planes; and so on. Consider $g_1, ..., g_{\alpha^{-1/2}} \in \mathbf{e}$ and let W be the set of planes outside the support of any of these encodings. It is not hard to see that no encoding then can have a non-negligible (i.e., $\geq \sqrt{\alpha}$) support within W (otherwise the fraction of the union of the supports of $g_1, ..., g_{\alpha^{-1/2}}$ is > 1, which is obviously impossible). Assuming the above assertion, therefore, the pairwise consistency of F within W is small, that is,

$$\frac{\left|\mathrm{CONS}_F(\mathcal{P}) \cap \binom{W}{2}\right|}{|\mathcal{P}|} \leq \theta_d.$$

Let A be an $[F, \mathbf{e}]$-extension assigning g_1 to the support of g_1, g_2 to the support of g_2 (in the remaining plains), and so on. For plains in W, A assigns arbitrary value. Pairs of planes in $\mathrm{CONS}_F(\mathcal{P})$ that are not given the same values by A are either both in W, and therefore bounded to be of $\leq \theta_d$ fraction, or pairs one in the support of some g_i while the other not, which are bounded for each g_i to be of $\leq \alpha$ fraction. Hence, the total fraction of pairs in $\mathcal{P}$ given different values by A is $\leq \theta_d + \alpha \cdot \alpha^{-1/2} \leq \epsilon \cdot d$.

Our proof proceeds to identify an encoding of non-negligible support in $\mathcal{S}^2$, i.e., in the special case where W is the set of all planes. This assertion may be generalized to any W by defining F' which is the same as F on all planes in W, however assigns, for every plane $\wp \notin W$ a value that does not agree with any other plane $\wp'$ such that $\langle \wp, \wp' \rangle \in \mathcal{P}$ (extending R if necessary). An encoding of large support in F' thus corresponds to an encoding of large support within W in F.

Let us thus assume that $\delta \geq \theta_d (= \epsilon(d-1) + \frac{1}{2}\epsilon)$ and prove the existence of one encoding of large support.

Consistency Graph

In order to isolate and clearly identify the few places in which properties related to geometric-codes are used, the relationship between encodings on various sub-spaces of $\mathcal{D}$ are represented by a special type of graph G_F.

Define a *graph–with–distribution* $G = \langle V, E, \psi \rangle$ to be an undirected graph with a probability distribution ψ over the set of vertexes V.

Let $G_F = \langle V, E, \psi \rangle$ be a graph–with-distribution, whose vertex set has one vertex for every pair consisting of a $(d-1)$-dimensional sub-space S and an encoding over S, hence

$$V = \mathcal{S}^{d-1} \times \mathbf{e}_{d-1}.$$

The extensions A_S identify encodings that win a non-negligible support in S, which are used in order to show a global encoding of large support in $\mathcal{D}$. For every S, Let ψ_S be a probability distribution over $\mathbf{e}_{d-1}$, by which every encoding $f \in \mathbf{e}_{d-1}$ appears according to the probability of a random plane in S to be assigned f by A_S, i.e.,

$$\psi_S(f) = \Pr_{\wp \in_R \mathcal{S}^2[S]} [A_S(\wp) = f]$$

Let now ψ be the probability distribution, defined over V, normalized so that each S appears with equal probability: for every $S \in \mathcal{S}^{d-1}$ and $f \in \mathbf{e}_{d-1}$,

$$\psi(S, f) = \frac{\psi_S(f)}{|\mathcal{S}^{d-1}|}$$

Note that, for every S, most encodings appear with probability 0 according to ψ; those vertexes would be practically ignored in our analysis. For the special case $d = 3$, $\psi_S(f) = 1$ for $f = F(S)$, and $\psi_S(f) = 0$ otherwise. In this case, ignoring the probability-0-vertexes, the set of vertexes is equivalent to $\mathcal{S}^2$, and hence is very simple.

For two vertexes $\langle S_1, f_1 \rangle$ and $\langle S_2, f_2 \rangle$ in G_F, either S_1 and S_2 are parallel, or they intersect by a $(d-2)$-dimensional subspace. In the former case, the two vertexes are connected by an edge; in the latter, let S' be the $(d-2)$-dimensional sub-space in common, the two vertexes are connected if the encodings agree on S', i.e., if ${f_1}_{\downarrow S'} = {f_2}_{\downarrow S'}$.

If F corresponds to a single encoding $g \in \mathbf{e}$, then in each S there would be a single encoding $g_{\downarrow S} \in \mathbf{e}_{d-1}$ such that $\psi_S(g_{\downarrow S}) = 1$, and the set of vertexes

$$\left\{ \left\langle S, g_{\downarrow S} \right\rangle \right\}_{S \in \mathcal{S}^{d-1}}$$

would form a clique in G_F. In other words, two vertexes of G_F, chosen according to ψ, would then be connected with probability 1.

Note that for the case $d = 3$ the graph G_F is very simple: In this case, ignoring every vertex that appears with probability 0 (according to ψ), the set of vertexes of G_F is equivalent to $\mathcal{S}^2$. Two planes $S_1, S_2 \in \mathcal{S}^2$ are connected iff $F(S_1), F(S_2)$ agree on their intersection. That is, the set of edges is simply $\mathrm{CONS}_F(\mathcal{P})$.

For ease of reading, let us extend the notation ψ to sets of vertexes of G and to sets of pairs of vertexes, in the natural way, i.e., for $U \subseteq V$, $\psi(U) = \Pr_{v \in_\psi V}[v \in U]$ and for $M \subseteq V^2$, $\psi(M) = \Pr_{v_1, v_2 \in_\psi V}[\langle v_1, v_2 \rangle \in M]$ (where $\in_\psi$ corresponds to a choice of an element according to ψ).

The *edge-probability* of a graph–with-distribution $G = \langle V, E, \psi \rangle$ — denoted $e(G)$ — is defined to be

$$e(G) = \psi(E).$$

The rest of the proof of theorem 4 breaks into three parts. At first, the edge probability of G_F is shown to be non-negligible. The amount of inconsistencies in G_F is then quantified in terms of a combinatorial parameter β. That amount $\beta(G_F)$ is, on one hand, shown, using simple properties of geometric codes, to be rather small. On the other hand, our proof resorts to combinatorial arguments in order to show that any graph-with-distribution with small β has a large clique. Lemma 5 (below) implies the existence of an encoding of large support. These parts combined imply the theorem.

First, let us show that the edge probability of G_F is non-negligible. The edge-probability of G_F can be bounded from below as follows:

Claim 2 *There exist (large) global constants c_1 and c_2, such that* $e(G_F) \geq \frac{1}{c_1}(\delta - \epsilon \cdot (d-1))^{c_2}$.

A sketch of the proof for claim 2 appears in the Appendix. A complete proof to appear in the full version.

Transitivity

Let us proceed, assuming Claim 2, to prove the existence of at least one clique of large weight in G_F.

Let $G = \langle V, E \rangle$ be a graph; a *non-transitive triplet* in G are three vertexes, $v_1, v_2, v_3 \in V$ such that out of the three inner pairs (v_1, v_2), (v_1, v_3) and (v_2, v_3), exactly two of the edges exist in G (i.e., belong to E). G is *transitive* if it has no non-transitive triplets.

G_F is not necessarily transitive; nevertheless, it turns out to be "almost transitive" in the sense that, it is possible to remove edges from G_F such that the edge-probability of G_F remains almost the same, and so that G_F becomes transitive. For that purpose, let us introduce the following parameter:

Given any graph–with-distribution $G = \langle V, E, \psi \rangle$, consider a non-edge $\langle v_1, v_2 \rangle \in \binom{V}{2} \setminus E$ in G; let $\beta(v_1, v_2)$ be the probability, over $v_3 \in_\psi V$, that the triplet v_1, v_2, v_3 is non-transitive (due to the non-edge $\langle v_1, v_2 \rangle$), that is

$$\beta(v_1, v_2) = \Pr_{v_3 \in_\psi V} \left[\langle v_1, v_3 \rangle, \langle v_3, v_2 \rangle \in E \right].$$

Let $\beta(G)$ be the maximum, over $\langle v_1, v_2 \rangle \notin E$, of $\beta(v_1, v_2)$.

The consistency graph G_F is in fact such that $\beta(G_F)$ is rather small — as proven below. Let us, at this point, show a general upper bound — for any graph–with-distribution G — of the weight of edges to be removed from G, as a function of $\beta(G)$, in order for G to become transitive:

Lemma 3 *Any graph–with-distribution $G = \langle V, E, \psi \rangle$ has a transitive sub-graph $G' = \langle V, E', \psi \rangle$ (where $E' \subseteq E$) such that $\psi(E') \geq \psi(E) - 3\sqrt{\beta(G)}$.*

Proof: The proof of lemma 3 identifies a sequence of graphs $G = G_0, G_1, ..., G_l = G'$, where G' satisfies the necessary conditions. First, for a vertex $v \in V$, denote by $\mathcal{N}_i(v)$ the neighbor set of v in G_i. For $i = 0, 1, 2, ...,$ if G_i is transitive then $i = l$ and $G_i = G'$. Otherwise, $G_{i+1} = \langle V, E_{i+1}, \psi \rangle$ is constructed from $G_i = \langle V, E_i, \psi \rangle$ by applying one of the following two steps:

1. If there exists a vertex $v_i \in V$ such that

$$\psi(\mathcal{N}_i(v_i)) \leq \sqrt{\beta(G)}$$

then E_{i+1} is E_i except all edges containing v_i are removed.

2. Otherwise, unless G_i is transitive there must exist a vertex $v_i \in V$ that is disconnected from at least one vertex in its connected component C_i in G_i, and such that

$$\psi(\mathcal{N}_i(v_i)) > \sqrt{\beta(G)}.$$

In that case, all edges between $\mathcal{N}_i(v_i)$ and $C_i \setminus \mathcal{N}_i(v_i)$ are removed from E_i in order to obtain E_{i+1}.

Let us now show that the total edge-probability removed going from G to G' is small, that is,

$$\psi(E \setminus E') \leq 3\sqrt{\beta(G)}.$$

For that purpose, define I_1 to be the set of indexes in which step 1 is applied, and I_2 be the rest; we show that the two totals of edge probability removed in steps of I_1 and of I_2 are both small.

As far as for I_1, for each i, at most $2\sqrt{\beta(G)} \cdot \psi(v_i)$ edge probability is removed (the factor 2 is due to the fact that each edge contributes twice to the edge-probability). Since for distinct $i \neq i'$, $v_i \neq v_{i'}$, the total edge-probability removed due to enforcement of step 1 is $\leq 2\sqrt{\beta(G)}$.

Now let us turn to I_2, and show that

$$\sum_{i \in I_2} \psi(E_i \setminus E_{i+1}) \leq \sqrt{\beta(G)}$$

Denote

$$T_i = (C_i \setminus \mathcal{N}_i(v_i)) \times \mathcal{N}_i(v_i)$$

and note that it is clearly the case that for any distinct $i \neq i' \in I_2$, T_i and $T_{i'}$ are disjoint. Therefore, it is enough to prove that, for every $i \in I_2$,

$$\frac{\psi(E_i \setminus E_{i+1})}{\psi(T_i)} < \sqrt{\beta(G)}$$

That is, the weight of edges removed in each application of step 2 is bounded from above by a value smaller than $\sqrt{\beta(G)}$ time the weight of T_i. Since the total weight of all T_i's is at most 1, the total edge-probability removed in steps of I_2 must be $\leq \sqrt{\beta(G)}$. The remainder of the proof of lemma 3 is devoted to the proof of the above statement.

Let $v \in C_i \setminus \mathcal{N}_i(v_i)$ and $v' \in \mathcal{N}_i(v_i)$, such that $\langle v, v' \rangle \in E_i$, the triplet v_i, v, v' is non-transitive due to the non-edge $\langle v_i, v \rangle$. The weight of those non-transitive triangles is bounded from above by $\beta(G)$, hence for every $v \in C_i \setminus \mathcal{N}_i(v_i)$,

$$\Pr_{v' \in_\psi V}[v' \in \mathcal{N}_i(v_i) \text{ and } \langle v, v' \rangle \in E_i] \leq \beta(G)$$

and therefore

$$\psi(E_i \setminus E_{i+1}) \leq \beta(G) \cdot \psi(C_i \setminus \mathcal{N}_i(v_i)).$$

On the other hand, $\psi(\mathcal{N}_i(v_i)) > \sqrt{\beta(G)}$ hence

$$\psi(T_i) = \psi(\mathcal{N}_i(v_i)) \cdot \psi(C_i \setminus \mathcal{N}_i(v_i)) > \sqrt{\beta(G)} \cdot \psi(C_i \setminus \mathcal{N}_i(v_i))$$

Lemma 3 follows. □

Non-Transitivity Limited

Let us now consider, for every non-edge in G_F, the weight of non-transitive triangle due to that non-edge:

Claim 4

$$\beta(G_F) \leq \alpha^+.$$

Proof: [9] Let $\langle S_1, f_1\rangle$ and $\langle S_2, f_2\rangle$ be disconnected vertexes of G_F. It must then be the case that S_1 and S_2 intersect by a $(d-2)$-dimensional sub-space $S' = S_1 \cap S_2 \in \mathcal{S}^{d-2}$, and f_1 disagrees with f_2 on S'. f_1 and f_2 therefore may agree on at most an α fraction of S'. Denote by S'' the set of disagreement between f_1 and f_2. S'' is of $\geq 1-\alpha$ fraction in S'. Given a vertex $\langle S_3, f_3\rangle$ of G_F, it is either the case that S_3 does not intersect S', which seems to occur with probability $\leq |\mathcal{F}|^{-1}$. Otherwise, S_3 does intersects S', and $\langle S_3, f_3\rangle$ would then be connected to both $\langle S_1, f_1\rangle$ and $\langle S_2, f_2\rangle$ only if S_3 does not intersect S''. Since only an $\leq \alpha$ fraction of $S_3 \in \mathcal{S}^{d-1}$ does intersect S' however does not intersect S'', the claim follows. □

We will prove now the existence of $g \in \mathbf{e}$ of large support (i.e. of fraction $\geq \sqrt{\alpha}$). As described above, this completes the proof of theorem 4:

By lemma 3, G_F has a transitive sub-graph G'_F such that

$$e(G'_F) \geq e(G_F) - 3\sqrt{\beta(G_F)}$$

hence, by claim 4, and claim 2, we have

$$e(G'_F) \geq \frac{1}{c_1}(\delta - \epsilon \cdot (d-1))^{c_2} - 3\sqrt{\alpha^+} \geq \frac{1}{c_1}\left(\frac{\epsilon}{2}\right)^{c_2} - 3\sqrt{\alpha + |\mathcal{F}|^{-1}}.$$

Recall, that in the statement of the theorem $\epsilon = c \cdot \alpha^{1/c}$. Therefore, a big enough constant c ensures

$$e(G'_F) \geq \sqrt{\alpha}.$$

Thus the graph G'_F, and hence also G_F, contains (at least one) large clique (of fraction more than $\sqrt{\alpha}$). I.e. there exists a set Y of vertexes of G_F, such that $\psi(Y) \geq \sqrt{\alpha}$, and every two vertexes of Y are connected in G_F. By lemma 5 (below), there exists $g \in \mathbf{e}$, such that for every $\langle S, f\rangle \in Y$, g agrees with f on S. Hence, for every $\langle S, f\rangle \in Y$, and every plane

[9] A better bound for $\beta(F_G)$, one that depends on d can be proved, this will be made explicit in future versions of this write-up.

$\wp \in \mathcal{S}^2[S]$ with $A_S(\wp) = f$, g agrees with $F(\wp)$ on $\wp$ (by the definition of A_S). In other words; such a plane $\wp$ is in the support of g.

The fact that $\psi(Y) \geq \sqrt{\alpha}$ ensures that the fraction of these planes in $\mathcal{S}^2$ is at least $\sqrt{\alpha}$. Hence g is of large support.

$\square$

Let us finally attend to "lemma 5 below".

Lemma 5 (Extended-Interpolation) *Consider $T \subset \mathcal{S}^{d-1}$ such that $|T| \geq \alpha^+ \cdot \left|\mathcal{S}^{d-1}\right|$. Let $H: T \to \mathbf{e}_{d-1}$ be assigning consistent values in $\mathbf{e}_{d-1}$ to every sub-space in T (i.e. for every two sub-spaces in T, with non-empty intersection, the values assigned by H agree on their intersection). Then, there exists exactly one encoding $h \in \mathbf{e}_d$ such that for every $S \in T$, $H(S) = h_{\downarrow S}$.*

Proof: Since $|T| \geq \alpha^+ \cdot \left|\mathcal{S}^{d-1}\right|$, there exist some parallel $(d-1)$-dimensional sub-spaces $S_1, ..., S_a \in T$, and some other parallel $(d-1)$-dimensional sub-spaces $S'_1, ..., S'_a \in T$ such that $S_1, ..., S_a$ are not parallel to $S'_1, ..., S'_a$.

Let us first consider $S_1, ..., S_a$. By the interpolation property, there exists exactly one encoding $h \in \mathbf{e}_d$ such that for every j, $h_{\downarrow S_j} = H(S_j)$. Any $(d-1)$-dimensional sub-space $S \in T$ which is not parallel to $S_1, ..., S_a$, intersects each of these sub-spaces in a $(d-2)$-dimensional sub-space. Since all of these $(d-2)$-dimensional sub-spaces are parallel, and $H(S)$ and $h_{\downarrow S}$ agree on those sub-spaces (as they both agree with $H(S_j)$ for every j), it follows, by the interpolation property for the sub-space S, that $H(S) = h_{\downarrow S}$.

Therefore, in particular, for every j, $h_{\downarrow S'_j} = H(S'_j)$. By the same argument, for any $(d-1)$-dimensional sub-space $S \in T$ which is not parallel to $S'_1, ..., S'_a$, $H(S) = h_{\downarrow S}$.

Since any $S \in T$ cannot be parallel to both $S_1, ..., S_a$, and $S'_1, ..., S'_a$, the lemma follows.

$\square$

Appendix 1: Proof of claim 2

(sketch only) Note that for $d = 3$ the proof is trivial, because by the definitions $e(G_F)$ is almost the same as δ (or greater). In particular, we certainly have $e(G_F) > \frac{1}{2}\delta$. We will therefore concentrate on the case $d > 3$.

Denote $\Delta = \delta - \epsilon \cdot (d-1)$. Note that $\Delta \geq \frac{1}{2}\epsilon$. Since F is δ–pairwise–consistent, the average value of δ_S (taken over $S \in \mathcal{S}^{d-1}$) is at least δ. Therefore, the average value of $(\delta_S - \epsilon \cdot (d-1))$ is at least Δ. Since, by the induction hypothesis, every $S \in \mathcal{S}^{d-1}$ is $(\delta_S - \epsilon \cdot (d-1))$-consistent, we get by a standard argument that a fraction of at least $\frac{1}{2}\Delta$ sub-spaces $S \in \mathcal{S}^{d-1}$ are $\frac{\Delta}{2}$-consistent. We call these sub-spaces *strong sub-spaces*.

Let $S \in \mathcal{S}^{d-1}$ be a strong sub-space. Since A_S is an extension of maximal consistency, we have

$$\Pr_{\langle \wp_1, \wp_2 \rangle \in_R \mathcal{P}[S]} [A_S(\wp_1) = A_S(\wp_2)] \geq \frac{1}{2}\Delta.$$

First, we would like to conclude the existence of (at least one) encoding $f \in \mathbf{e}_{d-1}$ such that the probability of a random plane in S to be assigned f by A_S is greater than, say, $\frac{1}{4}\Delta$, i.e. we would like to conclude the existence of $f \in \mathbf{e}_{d-1}$ with

$$\psi_S(f) = \Pr_{\wp \in_R \mathcal{S}^2[S]} [A_S(\wp) = f] \geq \frac{1}{4}\Delta.$$

To prove this, assume by the way of contradiction that no such encodings exist. It is enough to show that for any encoding $f \in \mathbf{e}_{d-1}$ with $\psi_S(f) = \Pr_{\wp \in_R \mathcal{S}^2[S]} [A_S(\wp) = f] < \frac{1}{4}\Delta$, we have $\Pr_{\langle \wp_1, \wp_2 \rangle \in_R \mathcal{P}[S]} [A_S(\wp_1) = A_S(\wp_2) = f] \leq \psi_S(f) \cdot \frac{1}{2}\Delta$. (If this is true, one gets a contradiction by summing over all the encodings $f \in \mathbf{e}_{d-1}$). The last fact, however, can be proved by a standard argument, using the fact that in any set of planes of non-negligible size, the number of pairs of planes intersecting in a line, is almost the same as in a random set of planes of the same size.

Thus for every strong sub-space S, there exists an encoding $f \in \mathbf{e}_{d-1}$, with $\psi_S(f) \geq \frac{1}{4}\Delta$. We call such an encoding: *a strong encoding for the sub-space* S. Denote by $SUP_f[S]$ the set of all planes $\wp$ in $\mathcal{S}^2[S]$, with $A_S(\wp) = f$.

A vertex $\langle S, f \rangle \in V$ of the graph G_F is called *strong* if S is strong, and f is a strong encoding for the sub-space S. Let $W \subset V$ be the set of strong vertexes, then by the above argument

$$\psi(W) \geq \frac{1}{2} \cdot \Delta \cdot \frac{1}{4} \cdot \Delta = \frac{1}{8}\Delta^2.$$

Let $(\langle S_1, f_1 \rangle, \langle S_2, f_2 \rangle)$ be a pair of strong vertexes of G_F. If S_1, S_2 are parallel then by the definition $(\langle S_1, f_1 \rangle, \langle S_2, f_2 \rangle)$ is an edge of G_F. Assume that S_1, S_2 are not parallel. Then their intersection is a sub-space of dimension $d-2$. Denote the intersection of S_1, and S_2 by Z. Then $\mathcal{S}^2[S_1] \cap \mathcal{S}^2[S_2] = \mathcal{S}^2[Z]$. In this case $(\langle S_1, f_1 \rangle, \langle S_2, f_2 \rangle)$ is an edge of G_F iff f_1, f_2 agree on Z.

Now, if $SUP_{f_1}[S_1] \cap SUP_{f_2}[S_2]$ is a non-negligible fraction of $\mathcal{S}^2[Z]$ (say a fraction of more than α — in reality, however, much smaller fractions are enough) then by the fact that $\mathbf{e}_{d-1}$ is an α-code, f_1, f_2 must agree on Z. Hence, in this case, $(\langle S_1, f_1 \rangle, \langle S_2, f_2 \rangle)$ is an edge of G_F.

Note that $SUP_{f_1}[S_1] \cap SUP_{f_2}[S_2]$ is always a subset of $\mathcal{S}^2[Z]$. Call such a pair $(\langle S_1, f_1 \rangle, \langle S_2, f_2 \rangle)$ (of strong vertexes of G_F, such that $SUP_{f_1}[S_1] \cap SUP_{f_2}[S_2]$ is a non-negligible fraction of $\mathcal{S}^2[Z]$) :*a good pair of vertexes.*

Denote by X the set of all good pairs of vertexes of G_F. The argument is completed by a standard density argument showing that in the graph G_F there are many good pairs of vertexes. In particular, for some constants c_1 and c_2, we have

$$\psi(X) \geq \frac{1}{c_1}\Delta^{c_2}.$$

□

Appendix 2: The New Characterization of NP in terms of PCP

In Part I of this work, a complete exposition of the proof for the new PCP theorem is presented; let us state the theorem proven there.

Our results are best summarized by the following theorem:

Theorem 5 *Let $M(n)$ be any function such that $M(n) \leq (\log n)^{\beta}$ for $\beta < 1/4$, and no smaller than $\log^{(i)} n$ for a constant i (where $\log^{(i)}$ denotes i applications of the $\log$ function). Then for every $L \in NP$, there exists a polynomial time procedure that, on input $I \in \{0,1\}^n$, constructs a set of Boolean functions*

$$\Phi_{L,I} = \{\varphi_1, ..., \varphi_l\}$$

over variables $\mathcal{Y} = \{y_1, ..., y_l\}$ which range over $[1, ..., 2^{O(M(n))}]$, such that

1. *Each $\varphi_i \in \Phi_{L,I}$ depends on a constant C_M variables in $\mathcal{Y}$.*
2. *For every input instance I:*
   - *if $I \in L$ then there exists an assignment A to the variables $\mathcal{Y}$ such that all $\varphi_i \in \Phi_{L,I}$ evaluate to* `true` *on A.*
   - *if $I \notin L$ then for any assignment A to the variables $\mathcal{Y}$, at most $\frac{l}{2^{M(n)}}$ of $\varphi_i \in \Phi_{L,I}$ evaluate to* `true` *on A.*

In addition, if one is interested in $M(n)$ closer to $\log n$, the folowing holds as well:

Theorem 6 *Let $M(n)$ be any function which is at most $(\log n)^{1-\beta}$ for any constant $\beta > 0$, and no smaller than $\log^{(i)} n$ for a constant i (where $\log^{(i)}$ denotes i applications of the $\log$ function). Then for every $L \in NP$, there exists a polynomial time procedure that, on input $I \in \{0,1\}^n$, constructs a set of Boolean functions*

$$\Phi_{L,I} = \{\varphi_1, ..., \varphi_l\}$$

over variables $\mathcal{Y} = \{y_1, ..., y_l\}$ which range over $[1, ..., 2^{M(n)\cdot(\log M(n))^c}]$, for some constant c, such that

1. *Each $\varphi_i \in \Phi_{L,I}$ depends on a constant C_M variables in $\mathcal{Y}$.*
2. *For every input instance I:*
   - *if $I \in L$ then there exists an assignment A to the variables $\mathcal{Y}$ such that all $\varphi_i \in \Phi_{L,I}$ evaluate to* `true` *on A.*
   - *if $I \notin L$ then for any assignment A to the variables $\mathcal{Y}$, at most $\frac{l}{2^{M(n)}}$ of $\varphi_i \in \Phi_{L,I}$ evaluate to* `true` *on A.*

Acknowledgment

Johan Håstad had a small nevertheless critical contribution to this work, however persistently refuses to co-author the paper. Oded Goldreich deserves as good as grade for his contributions and support.

References

[ALM+92] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and intractability of approximation problems. In *Proc. 33rd IEEE Symp. on Foundations of Computer Science*, pages 13–22, 1992.

[AS92] S. Arora and S. Safra. Probabilistic checking of proofs: A new characterization of NP. In *Proc. 33rd IEEE Symp. on Foundations of Computer Science*, pages 2–13, 1992.

[Bab85] L. Babai. Trading group theory for randomness. In *Proc. 17th ACM Symp. on Theory of Computing*, pages 421–429, 1985.

[BFL91] L. Babai, L. Fortnow, and C. Lund. Non-deterministic exponential time has two-prover interactive protocols. *Computational Complexity*, 1:3–40, 1991.

[BGKW88] M. Ben-or, S. Goldwasser, J. Kilian, and A. Wigderson. Multi prover interactive proofs: How to remove intractability assumptions. In *Proc. 20th ACM Symp. on Theory of Computing*, pages 113–121, 1988.

[BGLR93] M. Bellare, S. Goldwasser, C. Lund, and A. Russell. Efficient multi-prover interactive proofs with applications to approximation problems. In *Proc. 25th ACM Symp. on Theory of Computing*, pages 113–131, 1993.

[BGS95] M. Bellare, O. Goldreich, and M. Sudan. Free bits and nonapproximability. In *Proc. 27th ACM Symp. on Theory of Computing*, 1995. to appear.

[BLR90] M. Blum, M. Luby, and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. In *Proc. 22nd ACM Symp. on Theory of Computing*, pages 73–83, 1990.

[Coo71] S. Cook. The complexity of theorem-proving procedures. In *Proc. 3rd ACM Symp. on Theory of Computing*, pages 151–158, 1971.

[FGL+91] U. Feige, S. Goldwasser, L. Lovász, S. Safra, and M. Szegedy. Approximating clique is almost NP-complete. In *Proc. 32nd IEEE Symp. on Foundations of Computer Science*, pages 2–12, 1991.

[FL92] U. Feige and L. Lovász. Two-prover one-round proof systems: Their power and their problems. In *Proc. 24th ACM Symp. on Theory of Computing*, pages 733–741, 1992.

[FSH94] K. Friedl, A. Shen, and Z. Hatsagi. The low-degree test. In *Proc. 5th SIAM Symposium on Discrete Algorithms*, 1994.

[GLR+91] P. Gemmell, R. Lipton, R. Rubinfeld, M. Sudan, and A. Wigderson. Self-testing/correcting for polynomials and for approximate functions. In *Proc. 23rd ACM Symp. on Theory of Computing*, pages 32–42, 1991.

[GMR89] S. Goldwasser, S. Micali, and C. Rackoff. The knowledge complexity of interactive proofs. *SIAM J. of Computation*, 18:186–208, 1989.

[GS96] O. Goldreich and S. Safra. A Combinatorial Consistency Lemma with application to proving the PCP Theorem. ECCC TR96-047.

[RS1] R. Raz and S .Safra. A Sub-Constant Error-Probability
PCP Characterization of NP; PART I

[Hås96] Johan Håstad. Testing of the long code and hardness for clique. In *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*, pages 11–19, Philadelphia, Pennsylvania, 22–24 May 1996.

[Hås96] Johan Håstad. ?? ??

[KLS93] S. Khanna, N. Linial, and S. Safra. On the hardness of approximating the chromatic number. In *Proceedings of the 2nd Israel Symposium on Theory and Computing Systems, ISTCS*, pages 250–260. IEEE Computer Society Press, 1993.

[LFKN92] C. Lund, L. Fortnow, H. Karloff, and N. Nisan. Algebraic methods for interactive proof systems. *Journal of the ACM*, 39(4):859–868, October 1992.

[Lip91] R. Lipton. New directions in testing. In J. Feigenbaum and M. Merritt, editors, *Distributed Computing and Cryptography*. American Mathematical Society, 1991. Dimacs Series in Discrete Mathematics and Theoretical Computer Science, volume 2.

[LS91] D. Lapidot and A. Shamir. Fully parallelized multi prover protocols for NEXPTIME. In *Proc. 32nd IEEE Symp. on Foundations of Computer Science*, pages 13–18, 1991.

[LY94] Carsten Lund and Mihalis Yannakakis. On the hardness of approximating minimization problems. *Journal of the ACM*, 41(5):960–981, 1994.

[PY91] C. Papadimitriou and M. Yannakakis. Optimization, approximation and complexity classes. *Journal of Computer and System Sciences*, 43:425–440, 1991.

[RS92] R. Rubinfeld and M. Sudan. Testing polynomial functions efficiently and over rational domains. In *Proc. 3rd Annual ACM-SIAM Symp. on Discrete Algorithms*, pages 23–32, 1992.

[Rub90] R. Rubinfeld. *A mathematical theory of self-checking, self-testing and self-correcting Programs*. PhD thesis, U.C. Berkeley, 1990.

[Sha92] A. Shamir. IP = PSPACE. *Journal of the ACM*, 39(4):869–877, October 1992. Prelim. version in 1990 FOCS, pages 11–15.

[She91] A. Shen. Multilinearity test made easy. Manuscript, 1991.

[Sud92] M. Sudan. *Efficient checking of polynomials and proofs and the hardness of approximation problems*. PhD thesis, U.C. Berkeley, 1992.