New Proofs of the Green–Tao–Ziegler Dense Model Theorem: An Exposition

Omer Reingold[*] Luca Trevisan[†] Madhur Tulsiani[‡] Salil Vadhan[§]

June 1, 2008

Abstract

Green, Tao and Ziegler [GT, TZ] prove "Dense Model Theorems" of the following form: if R is a (possibly very sparse) pseudorandom subset of set X, and D is a dense subset of R, then D may be modeled by a set M whose density inside X is approximately the same as the density of D in R. More generally, they show that a function that is majorized by a pseudorandom measure can be written as a sum of a bounded function having the same expectation plus a function that is "indistinguishable from zero." This theorem plays a key role in the proof of the Green–Tao Theorem [GT] that the primes contain arbitrarily long arithmetic progressions.

In this note, we present a new proof of the Green–Tao–Ziegler Dense Model Theorem, which was discovered independently by ourselves [RTTV] and Gowers [Gow]. Our presentation follows the argument in [RTTV] (which in turn was inspired by Nisan's proof of the Impagliazzo Hardcore Set Theorem [Imp]), but is translated to the original notation of Green, Tao, and Ziegler.

We refer to our full paper [RTTV] for variants of the result with connections and applications to computational complexity theory, and to Gowers' paper [Gow] for applications of the proof technique to "decomposition, "structure," and "transference" theorems in arithmetic and extremal combinatorics (as well as a broader survey of such theorems).

1 The Green-Tao-Ziegler Theorem

Let X be a finite universe. We use the notation $\mathbb{E}_{x\in X} f(x) := \frac{1}{|X|}\sum_{x\in X} f(x)$. For two functions $f, g : X \to \mathbb{R}$ we define their *inner product* as

$$\langle f, g\rangle := \mathop{\mathbb{E}}_{x\in X} f(x)g(x)$$

[*] Faculty of Mathematics and Computer Science, Weizmann Institute of Science, Rehovot 76100, Israel. omer.reingold@weizmann.ac.il. Research supported by US-Israel Binational Science Foundation grant 2006060.

[†] Computer Science Division, U.C. Berkeley. luca@cs.berkeley.edu. Work partly done while visiting Princeton University and the IAS. This material is based upon work supported by the National Science Foundation under grants CCF-0515231 and CCF-0729137 and by the US-Israel Binational Science Foundation under grant 2006060.

[‡] Computer Science Division, U.C. Berkeley. madhurt@cs.berkeley.edu Work partly done while visiting Princeton University. This material is based upon work supported by the National Science Foundation under grants CCF-0515231 and CCF-0729137 and by the US-Israel Binational Science Foundation under grant 2006060.

[§] School of Engineering and Applied Sciences, Harvard University. salil@eecs.harvard.edu. Work done during a visit to U.C. Berkeley, supported by the Miller Foundation for Basic Research in Science, a Guggenheim Fellowship, US-Israel Binational Science Foundation grant 2006060, and the Office of Naval Research grant N00014-04-1-0478.

A *measure* on X is a function $g : X \to \mathbb{R}$ such that $g \geq 0$ and $\mathbb{E}_{x \in X}\, g(x) \leq 1$. A measure g is *bounded* if $g \leq 1$.

Let $\mathcal{F}$ be a collection of bounded functions $f : X \to [-1, 1]$. We say that two measures g, h are *ϵ-indistinguishable* according to $\mathcal{F}$ if

$$\forall f \in \mathcal{F}. |\langle g - h, f\rangle| \leq \epsilon$$

(It can be noted, although this fact will not be used, that if define $\|g\|_{\mathcal{F}} = \max_{f\in\mathcal{F}} |\langle g, f\rangle|$, then $\|\cdot\|_{\mathcal{F}}$ is a semi-norm, and we have that g and h are ϵ-indistinguishable if and only if $\|g - h\|_{\mathcal{F}} \leq \epsilon$. Hence the notion of indistinguishability may be seen as a semi-metric imposed on the space of functions $X \to \mathbb{R}$. If $\mathcal{F}$ contains *all* bounded functions $f : X \to [-1, 1]$, then $\|\cdot\|_{\mathcal{F}}$ is the standard ℓ_1 norm.)

We say that a measure g is *ϵ-pseudorandom* according to $\mathcal{F}$ if g and 1_X are *ϵ-indistinguishable* according to $\mathcal{F}$, where 1_X is the function that is identically equal to 1.

If $\mathcal{F}$ is a collection of bounded functions $f : X \to [-1, 1]$, we denote by $\mathcal{F}^k$ the collections of all functions of the form $\prod_{i=1}^{k'} f_i$, where $f_i \in \mathcal{F}$ and $k' \leq k$. In particular, if $\mathcal{F}$ is closed under multiplication, then $\mathcal{F}^k = \mathcal{F}$.

Theorem 1.1 (Green, Tao, Ziegler [GT, TZ]) *For every $\epsilon > 0$, there is a $k = (1/\epsilon)^{O(1)}$ and an $\epsilon' = \exp(-(1/\epsilon)^{O(1)})$ such that the following holds:*

Suppose that $\mathcal{F}$ is a finite collection of bounded functions $f : X \to [-1, 1]$ on a finite set X, $\nu : X \to \mathbb{R}$ is an ϵ'-pseudorandom measure according to $\mathcal{F}^k$, and $g : X \to \mathbb{R}$ is a measure such that $g \leq \nu$.

Then there is a bounded measure $g_1 : X \to [0, 1]$ such that

1. *$\mathbb{E}_{x\in X}\, g_1(x) = \mathbb{E}_{x\in X}\, g(x)$, and*
2. *g_1 and g are ϵ-indistinguishable according to $\mathcal{F}$.*

Green, Tao, and Ziegler [GT, TZ] state the conclusion in the following equivalent form: we can write $g = g_1 + g_2$, where g_1 is a bounded measure, g_1 and g have the same expectation, and g_2 is nearly orthogonal to $\mathcal{F}$ in the sense that $|\langle g_2, f\rangle| \leq \epsilon$ for all $f \in \mathcal{F}$.

We now describe how the theorem can be interpreted as saying that "every dense subset of a pseudorandom set has a dense model", as mentioned in the abstract. From any sets $D \subseteq R \subseteq X$, we can obtain measures $\nu \geq g$ by setting $\nu = 1_R \cdot |X|/|R|$ and $g = 1_D \cdot |X|/|R|$, where we write 1_S for the characteristic function of a set S. Then the condition that ν is ϵ'-pseudorandom according to $\mathcal{F}$ says that every function $f \in \mathcal{F}$ has the same average over R as it does over X, to within $\pm\epsilon'$, which is a natural pseudorandomness property of the set R. And the expectation of g is precisely the density of D in R, i.e. $|D|/|R|$. Now, assuming that R does indeed satisfy the foregoing pseudorandomness property, let g_1 be the bounded function given in the conclusion of the theorem. Suppose for starters that g_1 is the characteristic function of some set $M \subseteq X$. Then Item 1 says M has the same density in X as D has in R. And Item 2 says that D and M are indistinguishable from each other, in the sense that every function in $\mathcal{F}$ has the same average over both sets, to within $\pm\epsilon/\delta$, where $\delta = |D|/|R| = |M|/|X|$. So M is indeed a "dense model" of X.

The actual theorem above can be interpreted as simply allowing all of the sets, namely D and R in the hypothesis and M in the conclusion, to have their characteristic functions replaced with bounded measures of the same expectation. We note that, by an argument of Impagliazzo [Imp], allowing the function g_1 in the conclusion to be a measure rather than the characteristic function of some set M does not substantially weaken the theorem. Indeed, given g_1, we can construct a set M using the probabilistic method, including each element $x \in X$ in M independently with probability $g_1(x)$. Then, by Chernoff Bounds, M will have density at least $(1-\epsilon)\delta$ and its characteristic function will be 2ϵ-indistinguishable from g according to $\mathcal{F}$ with probability $1 - |\mathcal{F}| \cdot \exp(-\Omega(\delta\epsilon^2|X|))$.

2 Our Proof

We prove the contrapositive: assuming that g_1 is ϵ-distinguishable from all dense models g by functions in $\mathcal{F}$, we prove that ν cannot be pseudorandom, i.e. it is ϵ'-distinguishable from 1_X by some function in $\mathcal{F}^k$.

Let $\delta := \mathbb{E}_{x \in X}\, g(x)$ and let us denote, for convenience, by G the set of "dense measures" $g_1 : X \to [0,1]$ such that $\mathbb{E}\, g_1 = \delta$. Our assumption can be written as

$$\forall g_1 \in G. \exists f \in \mathcal{F}. |\langle g - g_1, f \rangle| > \epsilon$$

If we denote by $\mathcal{F}'$ the closure of $\mathcal{F}$ under negation, that is $\mathcal{F}' := \mathcal{F} \cup \{-f : f \in \mathcal{F}\}$, we can remove the absolute values:

$$\forall g_1 \in G. \exists f \in \mathcal{F}'. \langle g - g_1, f \rangle > \epsilon. \tag{1}$$

Proof outline. Suppose that we can manage to find a g_1, f pair such that the above holds *and* for which $g_1(x) = 1$ on every point in the support of f. Then it turns out that f must also distinguish ν from 1_X. Indeed, $\langle f, \nu \rangle \geq \langle f, g \rangle$, because $\nu \geq g$ pointwise, and $\langle f, 1_X \rangle = \langle f, g_1 \rangle$.

We will not be able to find such a g_1, f pair with $f \in \mathcal{F}'$, but we will be able to do so with a function f that is a *convex combination* of functions in $\mathcal{F}'$ composed with a *threshold function*. Then we show how to convert a distinguisher of such a form into a distinguisher that is a product of at most k functions from $\mathcal{F}$.

In more detail, the proof will proceed in the following steps:

1. By replacing $\mathcal{F}'$ with its convex hull, we reverse the order of quantifiers in (1), and obtain a single $\bar{f}$ that ϵ-distinguishes g from *every* bounded measure g_1 of expectation δ.

2. With an appropriate choice of g_1 (namely, the characteristic function of the $\delta|X|$ inputs on which $\bar{f}$ is largest), we argue that a thresholded version of $\bar{f}$, denoted $\bar{f}_t$, continues to $\Omega(\epsilon)$-distinguish g from g_1, and has support contained in $g_1^{-1}(1)$. By the above argument, $\bar{f}_t$ $\Omega(\epsilon)$-distinguishes ν from 1_X.

3. By approximating the threshold function with a low-degree polynomial that has relatively small coefficients, we deduce that there are at most k functions from $\mathcal{F}$ whose product ϵ'-distinguishes ν from 1_X.

Proof Details. We now proceed with Step 1, where we reverse the order of quantifiers in (1).

Claim 2.1 *There is a function $\overline{f}$ that is a convex combination of functions from $\mathcal{F}'$ and satisfies:*

$$\forall g_1 \in G. \langle g - g_1, \overline{f} \rangle > \epsilon$$

Proof of claim: We use the min-max theorem for 2-player zero-sum games (which is a consequence of the Hahn-Banach Theorem, as used in Gowers' version of the proof [Gow]). We think of a zero-sum game where the first player picks a function $f \in \mathcal{F}'$, the second player picks a function $g_1 \in G$, and the payoff is $\langle g - g_1, f \rangle$ for the first player, and $-\langle g - g_1, f \rangle$ for the second player.

By the min-max theorem, the game has a "value" α for which the first player has an optimal mixed strategy (a convex combination of strategies) $\bar{f}$, and the second player has an optimal mixed strategy $\bar{g}_1$, such that

$$\forall g_1 \in G, \quad \langle g - g_1, \bar{f} \rangle \geq \alpha \tag{2}$$

and

$$\forall f \in \mathcal{F}', \quad \langle g - \bar{g}_1, f \rangle \leq \alpha \tag{3}$$

Since G is convex, $\bar{g}_1 \in G$, and our hypothesis tells us that there exists a function f such that

$$\langle g - \bar{g}_1, f \rangle > \epsilon$$

Taking this f in Inequality (3), we get that $\alpha \geq \epsilon$. The claim now follows from Equation (2). □

We now proceed with Step 2 of the proof. Let $S \subseteq X$ be the set of $\delta|X|$ elements of X that maximize $\bar{f}$, and let g_1 be the characteristic function of S.[1] Then g_1 is a bounded measure of expectation δ, i.e. an element of G, so we have:

$$\langle g - g_1, \bar{f} \rangle \geq \epsilon \ .$$

or, equivalently,

$$\langle g, \bar{f} \rangle \geq \langle g_1, \bar{f} \rangle + \epsilon \tag{4}$$

Now, we argue that by applying a threshold function to $\bar{f}$, we can ensure that $g_1 = 1$ at every point in the support, while preserving the fact that we distinguish g from g_1. Specifically, for a threshold t, define $\bar{f}_t : X \to \{0, 1\}$ to be the boolean function such that $\bar{f}_t(x) = 1$ if and only if $\bar{f}(x) \geq t$. We will show that for some value of t, $\bar{f}_t$ has the properties we desire. Moreover, it will be important for the final step to argue that the threshold is "robust" in the sense that it does not matter what happens in a small interval around t, where the discontinuity of the threshold function could cause problems. (Gowers [Gow] handles this issue differently, by instead showing that there is a distinguisher of the form $\max\{0, \bar{f}(x) - t\}$, which has the advantage of being continuous everywhere as a function $\bar{f}(x)$.)

[1] In case $\delta|X|$ is not an integer, we define g_1 to be 1 on the $\lfloor \delta|X| \rfloor$ inputs that maximize $\bar{f}$, to be 0 on the $|X| - \lceil \delta|X| \rceil$ inputs that minimize $\bar{f}$, and to be an appropriate fractional value on the remaining element in order to make the expectation of g_1 equal to δ.

Claim 2.2 *There is a threshold $t \in [-1+\epsilon/3, 1]$ such that*

$$\langle g, \bar{f}_t \rangle \geq \langle g_1, \bar{f}_{t-\epsilon/3} \rangle + \frac{\epsilon}{3}$$

Proof of claim: First, observe that

$$\bar{f}(x) = \int_{-1}^{1} \bar{f}_t(x) dt - 1.$$

From (4) and the fact that $\langle g, 1_X \rangle = \langle g_1, \mathbf{1} \rangle = \delta$, we have

$$\langle g, \bar{f} + 1 \rangle \geq \langle g_1, \bar{f} + 1 \rangle + \epsilon,$$

which is equivalent to

$$\int_{-1}^{1} \langle g, \bar{f}_t \rangle dt \geq \int_{-1}^{1} \langle g_1, \bar{f}_t \rangle dt + \epsilon \tag{5}$$

Now if the claim were false, we would have

$$\begin{aligned}
\int_{-1}^{1} \langle g, \bar{f}_t \rangle dt &= \int_{-1}^{-1+\epsilon/3} \langle g, \bar{f}_t \rangle dt + \int_{-1+\epsilon/3}^{1} \langle g, \bar{f}_t \rangle dt \\
&< \int_{-1}^{-1+\epsilon/3} \langle g, \mathbf{1} \rangle dt + \int_{-1+\epsilon/3}^{1} \left(\langle g_1, \bar{f}_{t-\epsilon/3} \rangle + \frac{\epsilon}{3} \right) dt \\
&\leq \frac{\epsilon}{3} \cdot \delta + \int_{-1+\epsilon/3}^{1} \langle g_1, \bar{f}_{t-\epsilon/3} \rangle dt + \left(2 - \frac{\epsilon}{3} \right) \cdot \frac{\epsilon}{3} \\
&< \int_{-1}^{1} \langle g_1, \bar{f}_t \rangle dt + \epsilon,
\end{aligned}$$

contradicting Equation(5). □

We now argue that g_1 is identically equal to 1 on the support of $\bar{f}_{t-\epsilon/3}$. Recall g_1 is the characteristic function of the set of $\delta|X|$ inputs maximizing $\bar{f}$. So if $g_1(x) < 1$ for some x in the support of $\bar{f}_{t-\epsilon/3}$, then $g_1(x) = 0$ everywhere outside the support of $\bar{f}_{t-\epsilon/3}$. But then

$$\langle g_1, \bar{f}_{t-\epsilon/3} \rangle = \langle g_1, 1_X \rangle = \delta = \langle g, 1_X \rangle \geq \langle g, \bar{f}_t \rangle$$

in contradiction to Claim 2.2.

Putting everything together, we have

$$\begin{aligned}
\langle \nu, \bar{f}_t \rangle &\geq \langle g, \bar{f}_t \rangle \\
&\geq \langle g_1, \bar{f}_{t-\epsilon/3} \rangle + \epsilon/3 \\
&\geq \langle 1_X, \bar{f}_{t-\epsilon/3} \rangle + \epsilon/3.
\end{aligned}$$

Finally, we proceed with Step 3, where we find a distinguisher that is defined as a product of functions from $\mathcal{F}$, rather than being a threshold function applied to a convex combination of elements of $\mathcal{F}'$. We do this by approximating the threshold function by a polynomial, using the following special case of the Weierstrass Approximation Theorem.

Claim 2.3 *For every $\alpha, \beta \in [0,1]$, $t \in [\alpha, 1]$, there exists a polynomial p of degree $\mathrm{poly}(1/\alpha, 1/\beta)$ and with coefficients bounded in absolute value by $\exp(\mathrm{poly}(1/\alpha, 1/\beta))$ such that*

1. *For all $z \in [-1,1]$, we have $p(z) \in [0,1]$.*
2. *For all $z \in [-1, t-\alpha]$, we have $p(z) \in [0, \beta]$.*
3. *For all $z \in [t, 1]$, we have $p(z) \in [1-\beta, 1]$.*

We set $\alpha = \epsilon/3$ and $\beta = \epsilon/12$ in the claim to obtain a polynomial $p(z) = \sum_{i=0}^{d} c_i z^i$ of degree $d = \mathrm{poly}(1/\epsilon)$ with coefficients satisfying $|c_i| \leq \exp(\mathrm{poly}(1/\epsilon))$ and such that for every x we have

$$\bar{f}_t(x) - \frac{\epsilon}{12} \leq (p \circ \bar{f})(x)) \leq \bar{f}_{t-\epsilon/3}(x) + \frac{\epsilon}{12},$$

where $\circ$ denotes composition. From the properties of the polynomial p, we get

$$\langle \nu, p \circ \bar{f} \rangle \geq \langle \nu, \bar{f}_t \rangle - \frac{\epsilon}{12}$$

and

$$\langle 1_X, p \circ \bar{f} \rangle \leq \langle 1_X, \bar{f}_t \rangle + \frac{\epsilon}{12},$$

giving

$$\langle \nu, p \circ \bar{f} \geq \langle 1_X, p \circ \bar{f} \rangle + \frac{\epsilon}{6}. \tag{6}$$

If the polynomial $p \circ \bar{f} = \sum_i c_i \bar{f}^i$ has inner product at least $\epsilon/6$ with $\nu - 1$, there must exist a single term $c_k \bar{f}^k$ whose inner product with $\nu - 1$ is at least $\epsilon/(6(d+1))$, which in turn implies that $\bar{f}^k$ has inner product of absolute value at least $\epsilon' := \epsilon/(4(d+1)|c_k|) = \exp(-\mathrm{poly}(1/\epsilon))$ with $\nu - 1$:

$$|\langle \nu - 1, \bar{f}^k \rangle| \geq \epsilon'$$

Suppose that $\langle \nu - 1, \bar{f}^k \rangle \geq \epsilon'$. (The reasoning is analogous in the case of a negative inner product.) Recall that the function $\bar{f}$ is a convex combination of functions from $\mathcal{F}'$. This means that we may think of $\bar{f}(x)$ as being the expectation of a random variable $f(x)$, in which the function f is picked according to some probability measure on $\mathcal{F}'$. Then the value $\bar{f}(x)^k$ is the expectation of the process where we sample independently k functions $f_1, \ldots, f_k$ as before, and then compute $\prod_i f_i(x)$. By linearity of expectation, we can write

$$\epsilon' \leq \langle \nu - 1, \bar{f}(x)^k \rangle = \mathbb{E} \left\langle \nu - 1, \prod_i f_i(\cdot) \right\rangle,$$

where the expectation is over the choices of the functions f_i as described above. We can now conclude that there is point in the sample space where a random variable takes values at least as large as its expectation, and so there are functions $f_1, \ldots, f_k \in \mathcal{F}'$ such that

$$\left\langle \nu - 1, \prod_i f_i(\cdot) \right\rangle \geq \epsilon'$$

Finally, replacing f_i with $-f_i$ as appropriate, we can have all the functions f_i be in $\mathcal{F}$ itself (rather than $\mathcal{F}'$).

Acknowledgments

We thank Terence Tao, Avi Wigderson, Noga Alon, Russell Impagliazzo, Yishay Mansour, and Timothy Gowers for comments, suggestions and references.

References

[Gow] Timothy Gowers. Decompositions, approximate structure, transference, and the Hahn-Banach theorem. Preprint, 2008. 1, 3, 4

[GT] Ben Green and Terence Tao. The primes contain arbitrarily long arithmetic progressions. To appear in Annals of Mathematics. math.NT/0404188, 2004. 1, 2

[Imp] Russell Impagliazzo. Hard-core distributions for somewhat hard problems. In *Proceedings of the 36th IEEE Symposium on Foundations of Computer Science*, pages 538–545, 1995. 1, 2

[RTTV] Omer Reingold, Luca Trevisan, Madhur Tulsiani, and Salil Vadhan. Dense subsets of pseudorandom sets. Technical Report TR08-045, ECCC, 2008. 1

[TZ] Terence Tao and Tamar Ziegler. The primes contain arbitrarily long polynomial progressions. arXiv:math/0610050, 2006. 1, 2