

# A New Proof of a Theorem of Green, Tao, and Ziegler

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

April 24, 2008

## Abstract

Green, Tao and Ziegler prove structure 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 abstractly, 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.”

We present a new proof of this result inspired by Nisan’s proof of Impagliazzo’s hardcore set theorem. A similar proof was discovered independently by Gowers [Gow]. This note is an extract, rewritten in a different notation, from our paper [?] which includes a more general result and applications to complexity theory.

## 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  $f, g : X \rightarrow \mathbb{R}$  we define their *inner product* as

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

A *measure* on  $X$  is a function  $g : X \rightarrow \mathbb{R}$  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  $f : X \rightarrow [-1, 1]$ . We say that two measures  $g, h$  are  $\epsilon$ -*indistinguishable* according to  $\mathcal{F}$  if

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

---

\*Faculty of Mathematics and Computer Science, Weizmann Institute of Science, Rehovot 76100, Israel. [omer.reingold@weizmann.ac.il](mailto: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](mailto: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 grant 2006060.

‡Computer Science Division, U.C. Berkeley. [madhurt@cs.berkeley.edu](mailto: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 grant 2006060.

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

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

If  $\mathcal{F}$  is a collection of bounded functions  $f : X \rightarrow [-1, 1]$ , 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  $\mathcal{F}_k = \mathcal{F}$ . [**Madhur's Note: Changed definition of  $\mathcal{F}_k$  slightly to have upto  $k$  functions instead of exactly  $k$ .**]

**Theorem 1.1 (Green, Tao, Ziegler)** *For every  $\epsilon, \delta$  there is a  $k = (\epsilon)^{O(1)}$  and an  $\epsilon' = \exp(-\epsilon^{O(1)})$  such that:*

*If  $\mathcal{F}$  is a collection of bounded functions  $f : X \rightarrow [-1, 1]$ ,  $\nu : X \rightarrow \mathbb{R}$  is an  $\epsilon'$ -pseudorandom measure according to  $\mathcal{F}_k$ , and  $g : X \rightarrow \mathbb{R}$  is a measure such that  $0 \leq g \leq \nu$  and*

$$\mathbb{E}_{x \in X} g(x) = \delta ;$$

*Then there is a bounded measure  $g_1 : X \rightarrow [0, 1]$  such that  $\mathbb{E}_{x \in X} g_1(x) = \mathbb{E}_{x \in X} g(x) = \delta$  and  $g_1$  and  $g$  are  $\epsilon$ -indistinguishable according to  $\mathcal{F}$ .*

Note that a way to restate the conclusion is that 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  $\mathcal{F}$ , meaning that

$$\forall f \in \mathcal{F}. |\langle g_2, f \rangle| \leq \epsilon$$

## 2 Our Proof

We prove the contrapositive: assuming that  $g_1$  as required does not exist, we prove that  $\nu$  cannot be pseudorandom.

Let us denote, for convenience, by  $G$  the set of functions  $g_1 : X \rightarrow [0, 1]$  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 have

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

Our first claim is that we can, essentially, reverse the order of quantifiers.

**Claim 2.1** *There is a function  $\bar{f}$  which is a convex combination of functions from  $\mathcal{F}'$ , and is such that*

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

**Proof:** We use the min-max theorem for 2-player zero-sum games.<sup>1</sup> 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”  $\alpha$  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

$$\forall g_1 \in G, \quad \langle g - g_1, \bar{f} \rangle \geq \alpha \quad (1)$$

and

$$\forall f \in \mathcal{F}', \quad \langle g - \bar{g}_1, f \rangle \leq \alpha \quad (2)$$

[**Madhur’s Note: Changed  $\mathbb{E}\langle g - \bar{g}_1, f \rangle$  to  $\langle g - \bar{g}_1, f \rangle$  in the above equation. It seems either both the equations should have the expectation or neither should, depending on how you interpret  $\bar{g}_1$ .**]

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 (2), we get that  $\alpha \geq \epsilon$ . The claim now follows from Equation (1). ■

Let now  $S \subseteq X$  be the set of  $\delta|X|$  elements of  $X$  that maximize  $\bar{f}$ . The function  $1_S$  is a bounded measure of expectation  $\delta$ , and hence an element of  $G$ , so we have:

$$\langle g - 1_S, \bar{f} \rangle \geq \epsilon .$$

or, equivalently,

$$\langle g, \bar{f} \rangle \geq \langle 1_S(x), \bar{f}(x) \rangle + \epsilon$$

For a threshold  $t$ , define  $\bar{f}_t : X \rightarrow \{0, 1\}$  to be the boolean function such that  $\bar{f}_t(x) = 1$  if and only if  $\bar{f}(x) \geq t$ .

We claim that there is such a treshold function that distinguishes  $g$  and  $1_S$  in a “robust” way.

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

$$\langle g, \bar{f}_t \rangle \geq \left\langle 1_S(x), \bar{f}_{t-\frac{\epsilon}{3}}(x) \right\rangle + \frac{\epsilon}{3}$$

**Proof:** First, observe that

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

and also that, since  $\langle g, \mathbf{1} \rangle = \langle 1_S, \mathbf{1} \rangle = \delta$ , we have

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

which is equivalent to

$$\int_{-1}^1 \langle g, \bar{f}_t \rangle dt \geq \int_{-1}^1 \langle 1_S, \bar{f}_t \rangle dt + \epsilon \quad (3)$$

---

<sup>1</sup>We could also directly phrase the argument as a consequence of duality in linear programming.

Now if the claim were false, we would have

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

Contradicting Equation(3). ■

**[Madhur's Note: Changed  $\bar{f}_{t-\frac{\epsilon}{3}}$  to  $\bar{f}_t$  in the last step of the proof, since the limits of the integral are now different.]**

Now we shall show that the above threshold function distinguishes  $\nu$  from  $\mathbf{1}$ .

We have

$$\langle g(x), \bar{f}_t(x) \rangle \leq \langle \nu(x), \bar{f}_t(x) \rangle$$

because  $g(x) \leq \nu(x)$  pointwise and  $\bar{f}_t$  is non-negative.

Observe that one consequence of Claim 2.2 is that there are elements  $x \in S$  such that  $\bar{f}_t(x) < t - \frac{\epsilon}{3}$ . Otherwise, we would have

$$\langle 1_S, \bar{f}_{t-\frac{\epsilon}{3}} \rangle = \langle 1_S, \mathbf{1} \rangle = \delta = \langle g, \mathbf{1} \rangle \geq \langle g, \bar{f}_t \rangle$$

in contradiction to the Claim. Since  $S$  was chosen to maximize  $\bar{f}$ , we can conclude that for all elements  $x \in X - S$  we must have  $\bar{f}_t(x) < t - \frac{\epsilon}{3}$ , so that

$$\langle \mathbf{1}, \bar{f}_{t-\frac{\epsilon}{3}} \rangle = \langle 1_S, \bar{f}_{t-\frac{\epsilon}{3}} \rangle$$

Putting everything together, we have

$$\langle \nu, \bar{f}_t \rangle \geq \langle \mathbf{1}, \bar{f}_{t-\frac{\epsilon}{3}} \rangle + \frac{\epsilon}{3} \tag{4}$$

It remains to 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  $\mathcal{F}'$ .

**Claim 2.3** *For every  $\alpha, \beta \in [0, 1]$ ,  $t \in [\alpha, 1]$ , there exists a polynomial  $p$  of degree  $\text{poly}(1/\alpha, 1/\beta)$  and with coefficients bounded in absolute value by  $\exp(\text{poly}(1/\alpha, 1/\beta))$  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  $d = \text{poly}(1/\epsilon)$  with coefficients satisfying  $|c_i| \leq \exp(\text{poly}(1/\epsilon))$  and such that for every  $x$  we have

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

Combining the properties of the polynomial  $p$  with the above equations we get:

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

and

$$\langle \mathbf{1}, p(\bar{f}(\cdot)) \rangle \leq \langle \mathbf{1}, \bar{f}_t \rangle + \frac{\epsilon}{12}$$

giving

$$\langle \nu, p(\bar{f}(\cdot)) \rangle \geq \langle \mathbf{1}, p(\bar{f}(\cdot)) \rangle + \frac{\epsilon}{6} \tag{5}$$

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

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

Suppose that

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

(the reasoning will be analogous in the other case.) Recall that the function  $\bar{f}$  is a convex combination of functions from  $\mathcal{F}'$ , hence  $\bar{f}(x) = \sum_{f \in \mathcal{F}'} \lambda_f f(x)$ . We may think of  $\bar{f}(x)$  as being the expectation of the random variable  $f(x)$ , in the random process where we pick function  $f$  with probability  $\lambda_f$ . The value  $\bar{f}(x)^k$  is the expectation of the process where we sample independently  $k$  functions  $f_1, \dots, f_k$  as before, and then compute  $\prod_i f_i(x)$ . By linearity of expectation, we can write

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

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  $f_1, \dots, f_k \in \mathcal{F}'$  such that

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

Finally, replacing  $f_i$  with  $-f_i$  as appropriate, we have  $f_i \in \mathcal{F}$ .

## Acknowledgments

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

## References

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