Dense Model Theorems and Their Applications

Luca Trevisan*

Department of Computer Science, Stanford University,
Stanford CA 94305

In 2004, Ben Green and Terry Tao [6] proved that, for every k, there are infinitely many length-k arithmetic progressions made entirely of prime numbers. This settled a very long-standing open question in number theory that had been open even for the $k = 4$ case.

A key innovation in their proof is a "transference" or "dense model" theorem that, described in a "computer science terminology," states that if R is a "pseudorandom" set of integers, and $P \subseteq R$ is a subset of R that has positive density within R, then there exists a set of integers M that has positive density within all of $\mathbb{N}$ and that is "indistinguishable" from P in the sense that if a statistical property of a certain type is true for M then it must also be true for P.

Green and Tao apply their theorem to the case in which: (i) R are the almost-primes, integers with few, large, prime factors, which were known to have strong pseudorandomness properties; and (ii) P are the primes, which are known to have positive density in R, when the proper quantitative definition of R is given. Because of Szemerédi's Theorem [11, 12], M contains infinitely many arithmetic progressions of any length and, in fact, the probability that a random length-k progression is entirely contained in M is within a constant factor of the probability that are random k-tuple of elements is entirely contained in M. The notion of "indistinguishability" between M and P is such that this statistical property must be true for P too, and so not only it follows that P contains infinitely many arithmetic progressions of any length, but even the stronger statement that there are $\Omega_k(n^2/(\log n)^k)$ length-k progressions entirely composed of primes less than n.

A later paper of Tao and Ziegler [13] presents a more abstract version of the Green-Tao dense model theorem that, *qualitatively*, can be given the following translation in computer science terminology: if R is a pseudorandom distribution over a universe X, and D is a distribution that is c-dominated by R, in the sense that $D(x) \leq c \cdot R(x)$ for a domination parameter c, then there is a model distribution M such that D and M are indistinguishable, and M is $2c$-dominated by the uniform distribution over X. In particular, M has very high entropy $\log_2 |X| - \log_2 c + 1$, and so D has very high *pseudoentropy*. Unfortunately, the proof in [6,13], which is based on the proof of the Szemerédi Regularity Lemma, does not give the right quantitative result, because the complexity of the "security reduction" is exponential in the accuracy of the indistinguishability that one wants to achieve.

* This material is based upon work supported by the National Science Foundation under Grant No. CCF-1017403.

Y. Ishai (Ed.): TCC 2011, LNCS 6597, pp. 55–57, 2011.

Reingold, Trevisan, Tulsiani and Vadhan [10] provide an alternative proof based on duality of linear programming, inspired by Nisan's proof of the Impagliazzo hard-core lemma [7], which has a security reduction of polynomial complexity and hence works with "cryptographically strong" notions of pseudorandomness and indistinguishability. This gave a new characterization of the notion of pseudoentropy, which adds to the study of Barak, Shaltiel and Wigderson [1], who had proved equivalences between various computational analogs of entropy. Gowers [3] independently discovered the same argument based on duality of linear programming, which he has applied to other problems in additive combinatorics in joint work with Wolfe [4, 5].

Dziembowski and Pietrzak [2] formulated the computational dense model theorem in independent work (not inspired by the number-theoretic analog), and proved it based on a result that is attributed to [1], although it is actually first proved in [10]. Dziembowski and Pietrzak apply the dense model theorem to the task of designing cryptosystems that are resilient to key leakage, introducing an approach that has been tremendously influential.

Mironov, Pandey, Reingold and Vadhan [9] give an application of the dense model theorem to the study of *computational differential privacy*.

Impagliazzo [8] showed that the dense model theorem can be proved from a weaker assumption, giving yet another new characterization of pseudoentropy; he also showed that one can derive the dense model theorem in a black box way from any proof of a sufficiently strong version of his hard-core set lemma. This implied that a computational dense model theorem can be derived from an "iterative" approach, which is different from both the original argument of Green and Tao and from the argument based on duality of linear programming. The work of Trevisan, Tulsiani and Vadhan [14] gives a different way of "tying together" the Szemerédi regularity lemma, the dense model theorem, the Impagliazzo hard-core set lemma, and the use of iterative or linear-programming based techniques.

In this talk we will tell the number-theoretic side of this story, and give a quick overview of the several different, but equivalent, ways in which this family of results can be thought about.

References

1. Barak, B., Shaltiel, R., Wigderson, A.: Computational analogues of entropy. In: Arora, S., Jansen, K., Rolim, J.D.P., Sahai, A. (eds.) RANDOM 2003 and APPROX 2003. LNCS, vol. 2764, pp. 200–215. Springer, Heidelberg (2003)
2. Dziembowski, S., Pietrzak, K.: Leakage-resilient cryptography. In: Proceedings of the 49th IEEE Symposium on Foundations of Computer Science, pp. 293–302 (2008)
3. Gowers, T.: Decompositions, approximate structure, transference, and the Hahn-Banach theorem, arXiv:0811.3103 (2008)
4. Gowers, T., Wolf, J.: Linear forms and higher-degree uniformity for functions on $\mathbb{F}_p^n$, arXiv:1002.2208 (2010)
5. Gowers, T., Wolf, J.: Linear forms and quadratic uniformity for functions on $\mathbb{F}_p^n$, arXiv:1002.2209 (2010)

6. Green, B., Tao, T.: The primes contain arbitrarily long arithmetic progressions. Annals of Mathematics 167, 481–547 (2008)
7. Impagliazzo, R.: Hard-core distributions for somewhat hard problems. In: Proceedings of the 36th IEEE Symposium on Foundations of Computer Science, pp. 538–545 (1995)
8. Impagliazzo, R.: Personal Communication (2008)
9. Mironov, I., Pandey, O., Reingold, O., Vadhan, S.P.: Computational differential privacy. In: Halevi, S. (ed.) CRYPTO 2009. LNCS, vol. 5677, pp. 126–142. Springer, Heidelberg (2009)
10. Reingold, O., Trevisan, L., Tulsiani, M., Vadhan, S.: Dense subsets of pseudorandom sets. In: Proceedings of the 49th IEEE Symposium on Foundations of Computer Science, pp. 76–85 (2008)
11. Szemerédi, E.: On sets of integers containing no four elements in arithmetic progression. Acta Math. Acad. Sci. Hung. 20, 89–104 (1969)
12. Szemerédi, E.: On sets of integers containing no k elements in arithmetic progression. Acta Arithmetica 27, 199–245 (1975)
13. Tao, T., Ziegler, T.: The primes contain arbitrarily long polynomial progressions. Acta Mathematica 201, 213–305 (2008)
14. Trevisan, L., Tulsiani, M., Vadhan, S.: Regularity, boosting, and efficiently simulating every high-entropy distribution. In: Proceedings of the 24th IEEE Conference on Computational Complexity (2009)