Green's Conjecture and Testing Linear Invariant Properties

Asaf Shapira *

Abstract

A system of ℓ linear equations in p unknowns $Mx = b$ is said to have the *removal property* if every set $S \subseteq \{1, \ldots, n\}$ which contains $o(n^{p-\ell})$ solutions of $Mx = b$ can be turned into a set S' containing no solution of $Mx = b$, by the removal of $o(n)$ elements. Green [GAFA 2005] proved that a single homogenous linear equation always has the removal property, and conjectured that every set of homogenous linear equations has the removal property. In this paper we confirm Green's conjecture by showing that every set of linear equations (even non-homogenous) has the removal property. We also discuss some applications of our result in theoretical computer science, and in particular, use it to resolve a conjecture of Bhattacharyya, Chen, Sudan and Xie [4] related to algorithms for testing properties of boolean functions.

1 Background on removal lemmas

The (triangle) removal lemma of Ruzsa and Szemerédi [18], which is by now a cornerstone result in combinatorics, states that a graph on n vertices that contains only $o(n^3)$ triangles can be made triangle free by the removal of only $o(n^2)$ edges. Or in other words, if a graph has asymptomatically few triangles then it is asymptotically close to being triangle free. While the lemma was proved in [18] for triangles, an analogous result for any fixed graph can be obtained using the same proof idea. Actually, the main tool for obtaining the removal lemma is Szemerédi's regularity lemma for graphs [20], another landmark result in combinatorics. The removal lemma has many applications in different areas like extremal graph theory, additive number theory and theoretical computer science. Perhaps its most well known application appears already in [18] where it is shown that an ingenious application of it gives a very short and elegant proof of Roth's Theorem [16], which states that every $S \subseteq [n] = \{1, \ldots, n\}$ of positive density contains a 3-term arithmetic progression.

Recall that an r-uniform hypergraph $H = (V, E)$ has a set of vertices V and a set of edges E, where each edge $e \in E$ contains r distinct vertices from V. So a graph is a 2-uniform hypergraph. Szemeredi's famous theorem [19] extends Roth's theorem by showing that every $S \subseteq [n]$ of positive density actually contains arbitrarily long arithmetic progressions (when n is large enough). Motivated by the fact that a removal lemma for graphs can be used to prove Roth's theorem, Frankl and Rödl [5] showed that a removal lemma for r-uniform hypergraphs could be used to prove Szemeredi's theorem on $(r + 1)$-term arithmetic progressions. They further developed a regularity lemma, as well as a corresponding removal lemma, for 3-uniform hypergraphs thus obtaining a new proof of Szemeredi's theorem for 4-term arithmetic progressions. In recent years there have been many exciting results in this area, in particular the results of Gowers [8] and of Nagle, Rödl Schacht and Skokan [14, 15],

*School of Mathematics and College of Computing, Georgia Institute of Technology, Atlanta, GA 30332

who independently obtained regularity lemmas and removal lemmas for r-uniform hypergraph, thus providing alternative combinatorial proofs of Szemeredi's Theorem [19] and some of it generalizations, notably those of Furstenberg and Katznelson [6]. Tao [21] and Ishigami [11] later obtained another proof of the hypergraph removal lemma and of its many corollaries mentioned above. For more details see [9].

2 Our main result

In this paper we will use the above mentioned hypergraph removal lemma in order to resolve a conjecture of Green [10] regarding the removal properties of sets of linear equations. Let $Mx = b$ be a set of linear equations, and let us say that a set of integers S is (M, b)-free if it contains no solution to $Mx = b$, that is, if there is no vector x, whose entries all belong to S, which satisfies $Mx = b$. Just like the removal lemma for graphs states that a graph that has few copies of H should be close to being H-free, a removal lemma for sets of linear equations $Mx = b$ should say that a subset of the integers $[n]$ that contains few solutions to $Mx = b$, should be close to being (M, b)-free. Let us start be defining this notion precisely.

Definition 2.1 (Removal Property) *Let M be an $\ell \times p$ matrix of integers and let $b \in \mathbb{N}^{\ell}$. The set of linear equations $Mx = b$ has the* removal property *if for every $\delta > 0$ there is an $\epsilon = \epsilon(\delta, M, b) > 0$ with the following property: if $S \subseteq [n]$ is such that there are at most $\epsilon n^{p-\ell}$ vectors $x \in S^p$ satisfying $Mx = b$, then one can remove from S at most δn elements to obtain an (M, b)-free set.*

Green [10] has initiated the study of the removal properties of sets of linear equations. His main result was the following:

Theorem 1 (Green [10]) *Any single homogenous linear equation has the removal property.*

The main result of Green actually holds over any abelian group. To prove this result, Green developed a regularity lemma for abelian groups, which is somewhat analogous to Szemerédi's regularity lemma for graphs [20]. Although the application of the group regularity lemma for proving Theorem 1 was similar to the derivation of the graph removal lemma from the graph regularity lemma, the proof of the group regularity lemma was far from trivial. One of the main conjectures raised in [10] is that a natural generalization of Theorem 1 should also hold (Conjecture 9.4 in [10]).

Conjecture 1 (Green [10]) *Any system of homogenous linear equations $Mx = 0$ has the removal property.*

Very recently, Král', Serra and Vena [12] gave a surprisingly simple proof of Theorem 1, which completely avoided the use of Green's regularity lemma for groups. In fact, their proof is an elegant and simple application the removal lemma for directed graphs [1], which is a simple variant of the graph removal lemma that we have previously discussed. The proof given in [12] actually extends Theorem 1 to any single non-homogenous linear equation over arbitrary groups. Král', Serra and Vena [12] also show that Conjecture 1 holds when M is a 0/1 matrix, which satisfies certain conditions. But these conditions are not satisfied even by all 0/1 matrices.

In this paper we confirm Green's conjecture for every homogenous set of linear equations. In fact, we prove the following more general result.

Theorem 2 (Main Result) *Any set of linear equations $Mx = b$ has the removal property.*

3 Applications to testing properties of boolean functions

Besides being a natural problem from the perspective of additive number theory, it turns out that Theorem 2 has some applications in Theoretical Computer Science, in the area of *Property Testing* [3, 17, 7]. Property testers are fast randomized algorithms that can distinguish between objects satisfying a certain property $\mathcal{P}$ and objects that are "far" from satisfying it. In an attempt to prove a general sufficient condition that would guarantee that certain properties of boolean functions have efficient testing algorithms, Bhattacharyya, Chen, Sudan and Xie [4] conjectured that certain properties of boolean functions (that are related to the notion of being (M, b)-free) can be efficiently tested. As we show in this paper, our main result gives a positive answer to their open problem.

After our paper appeared on the Arxiv we learned that independently of our work, Král', Serra and Vena managed to improve upon their results in [12, 13] and obtain a proof of Conjecture 1.

References

[1] N. Alon and A. Shapira, Testing Subgraphs in Directed Graphs, Journal of Computer and System Sciences, 69 (2004), 354-382.

[2] T. Austin and T. Tao, On the testability and repair of hereditary hypergraph properties, manuscript, 2008.

[3] M. Blum, M. Luby and R. Rubinfeld, Self-testing/correcting with applications to numerical problems, JCSS 47 (1993), 549-595.

[4] A. Bhattacharyya, V. Chen, M. Sudan and N. Xie, Testing linear-invariant non-linear properties, manuscript, 2008.

[5] P. Frankl and V. Rödl, Extremal problems on set systems, Random Structures and Algorithms 20 (2002), 131-164.

[6] H. Furstenberg and Y. Katznelson, An ergodic Szemerédi theorem for commuting transformations, J. Analyse Math. 34 (1978), 275-291.

[7] O. Goldreich, S. Goldwasser and D. Ron, Property testing and its connection to learning and approximation, JACM 45(4): 653-750 (1998).

[8] T. Gowers, Hypergraph regularity and the multidimensional Szemerédi theorem, Ann. of Math. Volume 166, Number 3 (2007), 897-946.

[9] T. Gowers, Quasirandomness, counting and regularity for 3-uniform hypergraphs, Combinatorics, Probability and Computing, 15 (2006), 143-184.

[10] B. Green, A Szemerédi-type regularity lemma in abelian groups, GAFA 15 (2005), 340-376.

[11] Y. Ishigami, A simple regularization of hypergraphs, http://arxiv.org/abs/math/0612838.

[12] D. Král', O. Serra and L. Vena, A combinatorial proof of the removal lemma for groups, arXiv:0804.4847v1.

[13] D. Král', O. Serra and L. Vena, A removal lemma for linear systems over finite fields, Jornadas de Matematica Discreta y algortimica 2008.

[14] B. Nagle, V. Rödl and M. Schacht, The counting lemma for regular k-uniform hypergraphs, Random Structures and Algorithms 28 (2006), 113-179.

[15] V. Rödl and J. Skokan, Regularity lemma for k-uniform hypergraphs, Random Structures and Algorithms 25 (2004), 1-42.

[16] K.F. Roth, On certain sets of integers, J. London Math. Soc. 28 (1953), 104-109.

[17] R. Rubinfeld and M. Sudan, Robust characterization of polynomials with applications to program testing, *SIAM J. on Computing* 25 (1996), 252–271.

[18] I. Ruzsa and E. Szemerédi, Triple systems with no six points carrying three triangles, in Combinatorics (Keszthely, 1976), Coll. Math. Soc. J. Bolyai 18, Volume II, 939-945.

[19] E. Szemerédi, Integer sets containing no k elements in arithmetic progression, Acta Arith. 27 (1975), 299-345.

[20] E. Szemerédi, Regular partitions of graphs, In: *Proc. Colloque Inter. CNRS* (J. C. Bermond, J. C. Fournier, M. Las Vergnas and D. Sotteau, eds.), 1978, 399–401.

[21] T. Tao, A variant of the hypergraph removal lemma, J. Combin. Theory, Ser. A 113 (2006), 1257-1280.