Proceedings of the International Congress of Mathematicians
Hyderabad, India, 2010

Inapproximability of NP-complete Problems, Discrete Fourier Analysis, and Geometry

Subhash Khot *

Abstract. This article gives a survey of recent results that connect three areas in computer science and mathematics: (1) (Hardness of) computing approximate solutions to NP-complete problems. (2) Fourier analysis of boolean functions on boolean hypercube. (3) Certain problems in geometry, especially related to isoperimetry and embeddings between metric spaces.

Mathematics Subject Classification (2000). Primary 68Q17.

Keywords. NP-completeness, Approximation algorithms, Inapproximability, Probabilistically Checkable Proofs, Discrete Fourier analysis.

1. Introduction

The well-known $\mathsf{P} \neq \mathsf{NP}$ hypothesis says that a large class of computational problems known as NP-complete problems do not have efficient algorithms. An algorithm is called efficient if it runs in time polynomial in the length of the input. A natural question is whether one can efficiently compute *approximate* solutions to NP-complete problems and how good an approximation one can achieve. We are interested in both upper and lower bounds: designing algorithms with a guarantee on the approximation (upper bounds) as well as results showing that no efficient algorithm exists that achieves an approximation guarantee beyond a certain threshold (lower bounds). It is the latter question, namely the lower bounds, that is the focus of this article. Such results are known as *inapproximability* or *hardness of approximation* results, proved under a standard hypothesis such as $\mathsf{P} \neq \mathsf{NP}$.

Let us consider the Max-3Lin problem as an illustration. We are given a system of linear equations over $GF(2)$ with three variables in each equation and the goal is to find an assignment that satisfies the maximum number of equations. This is known to be an NP-complete problem. There is a trivial approximation algorithm that achieves a multiplicative approximation guarantee of 2. The algorithm simply assigns a random value in $GF(2)$ to each variable and in expectation satisfies half

*Supported by NSF CAREER grant CCF-0833228, NSF Expeditions grant CCF-0832795, and BSF grant 2008059.

of the equations. The optimal assignment may satisfy all (or nearly all) equations, and thus the assignment produced by the algorithm is within factor 2 of the optimal assignment. On the other hand, a famous result of Håstad [25] shows that such a trivial algorithm is the best one can hope for! Specifically, let $\varepsilon > 0$ be an arbitrarily small constant. Then given an instance of $\mathsf{Max\text{-}3Lin}$ that has an assignment satisfying $1-\varepsilon$ fraction of the equations, no efficient algorithm can find an assignment that satisfies $\frac{1}{2}+\varepsilon$ fraction of the equations unless $\mathsf{P} = \mathsf{NP}$.

It turns out that such inapproximability results are closely related to Fourier analysis of boolean functions on a boolean hypercube and to certain problems in geometry, especially related to isoperimetry. This article aims to give a survey of these connections. We anticipate that the intended audience of this article is not necessarily familiar with the techniques in computer science. We therefore focus more on the Fourier analytic and geometric aspects and only give a brief overview of how such results are used in (and often arise from) the context of inapproximability. We describe an overall framework in Section 2 and then illustrate the framework through several examples in the succeeding sections.

2. Framework for Inapproximability Results

Approximation Algorithms and Reductions

Let $\mathcal{I}$ denote an NP-complete problem. For an instance I of the problem with input size N, let $\mathsf{OPT}(I)$ denote the value of the optimal solution. For a specific polynomial time approximation algorithm, let $\mathsf{ALG}(I)$ denote the value of the solution that the algorithm finds (or its expected value if the algorithm is randomized). Let $C > 1$ be a parameter that could be a function of N.

Definition 2.1. *An algorithm is said to achieve an approximation factor of C if on every instance I,*

$$\begin{aligned} \mathsf{ALG}(I) &\geq \mathsf{OPT}(I)/C & \quad & \text{if } \mathcal{I} \text{ is a maximization problem,} \\ \mathsf{ALG}(I) &\leq C \cdot \mathsf{OPT}(I) & & \text{if } \mathcal{I} \text{ is a minimization problem} \end{aligned}$$

A maximization problem $\mathcal{I}$ is proved to be inapproximable by giving a reduction from a canonical NP-complete problem such as $\mathsf{3SAT}$ [1] to a *gap version* of $\mathcal{I}$. Specifically, suppose there is a polynomial time reduction that maps a $\mathsf{3SAT}$ formula ϕ to an instance I of the problem $\mathcal{I}$, such that for constants $0 < s < c$, we have:

1. (Completeness): If ϕ has a satisfying assignment, then $\mathsf{OPT}(I) \geq c$.
2. (Soundness): If ϕ has no satisfying assignment, then $\mathsf{OPT}(I) \leq s$.

[1] A $\mathsf{3SAT}$ formula ϕ is a logical AND of a set of clauses, where each clause is a logical OR of three boolean variables, possibly negated. The goal is to decide whether the formula has a satisfying boolean assignment.

Such a reduction implies that if there were an algorithm with approximation factor strictly less than $\frac{c}{s}$ for the problem $\mathcal{I}$, then it would enable one to efficiently decide whether a 3SAT formula is satisfiable, and hence $\mathsf{P} = \mathsf{NP}$. Inapproximability results for minimization problems can be proved in a similar way.

The PCP Theorem

In practice, a reduction as described above is often a sequence of (potentially very involved) reductions. In fact, the first reduction in the sequence is the famous *PCP Theorem* [18, 4, 2] which can be phrased as a reduction from 3SAT to a gap version of 3SAT. For a 3SAT formula ϕ, let $\mathsf{OPT}(\phi)$ denote the maximum fraction of clauses that can be satisfied by any assignment. Thus $\mathsf{OPT}(\phi) = 1$ if and only if ϕ is satisfiable. The PCP Theorem states that there is a universal constant $\alpha < 1$ and a polynomial time reduction that maps a 3SAT instance ϕ to another 3SAT instance ψ such that:

1. (Completeness): If $\mathsf{OPT}(\phi) = 1$, then $\mathsf{OPT}(\psi) = 1$.
2. (Soundness): If $\mathsf{OPT}(\phi) < 1$, then $\mathsf{OPT}(\psi) \leq \alpha$.

We stated the PCP Theorem as a combinatorial reduction. There is an equivalent formulation of it in terms of *proof checking*. The theorem states that every NP statement has a polynomial size proof that can be checked by a probabilistic polynomial time verifier by reading only a constant number of bits in the proof! The verifier has the completeness and the soundness property: every correct statement has a proof that is accepted with probability 1 and every proof of an incorrect statement is accepted with only a small probability, say at most 1%. The equivalence between the two views, namely reduction versus proof checking, is simple but illuminating, and has influenced much of the work in this area.

Gadgets based on Hypercube

The core of a reduction often involves a combinatorial object called a *gadget* and the reduction itself consists of taking several copies of the gadget and then appropriately connecting them together. The class of gadgets that is relevant for this article is the class of hypercube based gadgets. A simple example is the hypercube $\{-1, 1\}^n$ itself thought of as a graph. The edges of the hypercube are all pairs of inputs that differ on exactly one co-ordinate. When the computational problem under consideration is the Graph Partitioning problem, we are interested in partitioning a graph into two equal parts so as to minimize the number of crossing edges. A cut in the hypercube is same as a function $f : \{-1, 1\}^n \mapsto \{-1, 1\}$. The number of edges cut divided by a normalizing factor of 2^n is known as *average sensitivity* of the function. It is well-known that the minimum average sensitivity of a balanced function is 1 and the minimizer is precisely the *dictatorship* function, i.e. the function $f(x) = x_{i_0}$ for some fixed co-ordinate $i_0 \in \{1, \ldots, n\}$. Note that the dictatorship function depends only on a single co-ordinate. On the other hand, a theorem of Friedgut [19] shows that any function whose average sensitivity is at most k, is very *close* to a function that depends only on $2^{O(k)}$ co-ordinates. In the contrapositive, if a function depends on too many co-ordinates and thus is *far from being a dictatorship*, then its average sensitivity must be large. Such "dictatorship is good; any function that is far from being a dictatorship is bad" kind of results are precisely the properties that we need from the gadget.

In the following, we will sketch the overall framework for inapproximability results proved via hypercube based gadgets. We refrain from describing the components of a reduction other than the gadget itself, as these typically involve computer science techniques that the reader may not be familiar with. We then illustrate this framework through several examples.

The Framework

Let $\mathcal{F} := \{f \mid f : \{-1,1\}^n \mapsto \{-1,1\}, \mathbb{E}[f] = 0\}$ be the class of all balanced boolean functions on the hypercube. Let

$$\mathsf{DICT} := \{f \mid f \in \mathcal{F},\ \forall x \in \{-1,1\}^n, f(x) = x_{i_0} \text{ for some } i_0 \in \{1,\ldots,n\}\},$$

be the class of dictatorship functions. Note that a dictatorship function depends only on a single co-ordinate. We aim to define a class FFD of functions that are to be considered as functions far from being a dictatorship. This class should include functions such as $\mathsf{MAJORITY} := \mathsf{sign}(\sum_{i=1}^n x_i)$, $\mathsf{PARITY} := \prod_{i=1}^n x_i$, and random functions; these functions depend on all the co-ordinates in a non-trivial manner. Towards this end, let the influence of the i^{th} co-ordinate on a function f be defined as:

$$\mathrm{Infl}_i(f) := \mathrm{Pr}_x\left[f(x_1,\ldots,x_i,\ldots,x_n) \neq f(x_1,\ldots,-x_i,\ldots,x_n)\right].$$

For a dictatorship function, the relevant co-ordinate has influence 1 and all other influences are zero. Thus one may define FFD as the class of functions all of whose influences are small. This includes MAJORITY (all influences are $O(\frac{1}{\sqrt{n}})$), but excludes PARITY (all influences are 1) and random functions (all influences are very close to $\frac{1}{2}$). We therefore give a more refined definition that also turns out to be the most useful for the applications.

It is well-known that any function $f : \{-1,1\}^n \mapsto \mathbb{R}$ has a Fourier (or Fourier-Walsh) representation:

$$f(x) = \sum_{S \subseteq \{1,\ldots,n\}} \widehat{f}(S) \prod_{i \in S} x_i,$$

where the $\widehat{f}(S) \in \mathbb{R}$ are the Fourier coefficients. When f is a boolean function, by Parseval's identity, $\sum_S \widehat{f}(S)^2 = \mathbb{E}[f^2] = 1$. It is easily proved that:

$$\mathrm{Infl}_i(f) = \sum_{i \in S} \widehat{f}(S)^2.$$

For an integer d, we define the *degree d influence* as:

$$\mathrm{Infl}_i^d(f) = \sum_{i \in S, |S| \leq d} \widehat{f}(S)^2.$$

Finally, for an integer d and a parameter $\eta > 0$, let

$$\mathsf{FFD}_{d,\eta} := \{f \mid f \in \mathcal{F},\ \forall i \in \{1,\ldots,n\},\ \mathrm{Infl}_i^d(f) \le \eta\}.$$

In words, $\mathsf{FFD}_{d,\eta}$ is the class of all functions that are far from being a dictatorship, in the sense that all degree d-influences are at most η. We will think of d as a large and η as a small constant, and $n \to \infty$ as an independent parameter. Clearly, MAJORITY, PARITY, and random functions are in this class. For MAJORTIY, the influences are $O(\frac{1}{\sqrt{n}})$, and so are the degree d influences. For PARITY, the only non-zero Fourier coefficient $\widehat{f}(S)$ is for $S = \{1,\ldots,n\}$ and hence all degree d-influences are zero. For a random function, the Fourier mass is concentrated on sets $|S| = \Omega(n)$, and hence the degree d-influences are negligible. We are now ready to informally state the connection between inapproximability results and Fourier analytic results:

Theorem 2.2. *(Informal) Suppose $\mathcal{I}$ is a maximization problem and $\mathsf{Val} : \mathcal{F} \mapsto \mathbb{R}^+$ is a valuation on balanced boolean functions. Suppose there are constants $0 < s < c$ such that,*

1. *(Completeness): $\forall f \in \mathsf{DICT}, \quad \mathsf{Val}(f) \ge c$.*
2. *(Soundness): $\forall f \in \mathsf{FFD}_{d,\eta}, \quad \mathsf{Val}(f) \le s$.*

Assume a certain complexity theoretic hypothesis. Then given an instance of the problem $\mathcal{I}$ that has a solution with value at least c, no polynomial time algorithm can find a solution with value exceeding s. In particular, there is no polynomial time algorithm for the problem $\mathcal{I}$ with approximation factor strictly less than c/s.

The theorem is stated in a very informal manner and calls for several comments: (1) The choice of the valuation $\mathsf{Val}(\cdot)$ depends very much on the problem $\mathcal{I}$ and different problems lead to different interesting valuations. (2) We will be interested in the limiting case when $d \to \infty, \eta \to 0$. Often we will have $s = s' + \delta$ where s' is a specific constant and $\delta \to 0$ as $d \to \infty, \eta \to 0$. (3) The complexity theoretic hypothesis should ideally be $\mathsf{P} \neq \mathsf{NP}$, but often it will be the *Unique Games Conjecture* (see below). (4) An analogous theorem holds for minimization problems as well.

We apply the framework of Theorem 2.2 to several computational problems in the rest of the article. For each problem, we state the problem definition, the valuation $\mathsf{Val}(\cdot)$ that is used, how the soundness property follows from a Fourier analytic result, related geometric results, and then the inapproximability result that can be proved. Before we begin, we state several properties of the dictatorship functions that will be useful and state the Unique Games Conjecture for the sake of completeness.

The valuation $\mathsf{Val}(\cdot)$ is supposed to capture a certain property of dictatorship functions. Let us observe a few such properties:

1. Dictatorships are linear, i.e. $\forall x, y \in \{-1,1\}^n, \quad f(xy) = f(x)f(y)$, where xy denotes the string that is bitwise product of strings x and y.

2. Dictatorships are stable under noise, i.e. if input $x \in \{-1,1\}^n$ is chosen uniformly at random, and $y \in \{-1,1\}^n$ is obtained by flipping every bit of x with probability ε, then the probability that $f(x) \neq f(y)$ is ε. In contrast, MAJORITY is less stable and the probability is $\theta(\sqrt{\varepsilon})$, whereas PARITY is very unstable and the probability is very close to $\frac{1}{2}$.

3. If $C \subseteq \{-1,1\}^n$ is a random sub-cube with dimension εn, then with probability $1-\varepsilon$, a dictatorship function is constant on C. A sub-cube of dimension k is the set of all inputs that agree on a specific setting of input bits outside of T for some subset of co-ordinates $T \subseteq \{1,\ldots,n\}$, $|T| = k$.

4. The Fourier mass of a dictatorship function is concentrated at the first level, i.e. on sets $|S| = 1$. In contrast, the Fourier mass of MAJORITY at the first level is very close to $\frac{2}{\pi}$ and that of the PARITY function is zero.

The Unique Games Conjecture

Most of the inapproximability results presented in this article rely on the Unique Games Conjecture [28] stating that a certain computational problem called the Unique Game is very hard to approximate. We do state the conjecture here, but since we are focussing only on a certain component of a reduction, we will not have an occasion to use the statement. It is easier to understand the conjecture in terms of a special case: an instance of the Linear Unique Game is a system of linear equations over $\mathbb{Z}_n$ where every equation is of the form $x_i - x_j = c_{ij}$, $\{x_1,\ldots,x_N\}$ are variables, and $c_{ij} \in \mathbb{Z}_n$ are constants. The goal is to find an assignment to the variables that satisfies a *good* fraction of the equations.

The Unique Games Conjecture states that for every constant $\varepsilon > 0$, there is a large enough constant $n = n(\varepsilon)$, such that given an instance of Linear Unique Game over $\mathbb{Z}_n$ that has an assignment satisfying $1-\varepsilon$ fraction of the equations, no polynomial time algorithm can find an assignment that satisfies (even) an ε fraction of the equations. [2]

A comment about the term "Unique Game". The term "game" refers to the context of 2-Prover-1-Round games where the problem was studied initially. Given an instance of the Linear Unique Game, consider the following game between two provers and a verifier: the verifier picks an equation $x_i - x_j = c_{ij}$ at random, sends the variable x_i to prover P_1 and the variable x_j to prover P_2. Each prover is supposed to answer with a value in $\mathbb{Z}_n$, and the verifier accepts if and only if $a_1 - a_2 = c_{ij}$ where a_1 and a_2 are the answers of the two provers respectively. The strategies of the provers correspond to assignments $\sigma_1, \sigma_2 : \{x_1,\ldots,x_N\} \mapsto \mathbb{Z}_n$. The *value* of the game is the maximum over all prover strategies, the probability that the verifier accepts. It can be easily seen that this value is between β and

[2] The original conjecture is stated in terms of a more general problem, but it is shown in [29] that the conjecture is equivalent to the statement here in terms of linear unique games. Also, the "hardness" is conjectured to be NP-hardness rather than just saying that there is no polynomial time algorithm.

$\max\{1, 4\beta\}$ where β is the maximum fraction of equations that can be satisfied by any assignment. The term "unique" refers to the property of the equations $x_i - x_j = c_{ij}$ that for every value to one variable, there is a unique value to the other variable so that the equation is satisfied.

3. Max-3Lin and Linearity Test with Perturbation

Max-3Lin **Problem:** Given a system of linear equations over $GF(2)$ with each equation containing three variables. The goal is to find an assignment that satisfies a maximum fraction of equations.

Valuation: We define $\mathsf{Val}(f)$ as the probability that f passes the *linearity test* along with a small perturbation. Specifically, pick two inputs $x, y \in \{-1,1\}^n$ uniformly at random and let $w := xy$. Let z be a string obtained by flipping each bit of w with probability ε independently. Note that the correlation of every bit in z with the corresponding bit in w is $1-2\varepsilon$ and let $z \sim_{1-2\varepsilon} w$ denote this. Define

$$\mathsf{Val}(f) := \Pr_{x,y,z\sim_{1-2\varepsilon} w}\left[f(z) = f(x)f(y)\right].$$

The optimization problem concerns linear equations with three variables, and the valuation is defined in terms of a test that depends linearly on the values of f at three random (but correlated) inputs.

Completeness: If $f \in \mathsf{DICT}$, then it is easily seen that $\mathsf{Val}(f) = 1-\varepsilon$. Indeed, for some fixed co-ordinate $i_0 \in \{1,\ldots,n\}$, $f(x) = x_{i_0}, f(y) = y_{i_0}, f(z) = z_{i_0}$, and z_{i_0} is obtained by flipping the value of $x_{i_0}y_{i_0}$ with probability ε. Hence we have $f(z) = f(x)f(y)$ with probability $1-\varepsilon$.

Soundness: We will sketch a proof showing that if $f \in \mathsf{FFD}_{d,\eta}$, then $\mathsf{Val}(f) \le \frac{1}{2}+\delta$ where $\delta \to 0$ as $d \to \infty, \eta \to 0$. The key observation is that the probability of acceptance of the test can be written in terms of Fourier coefficients of f. It is a rather straightforward exercise (that we skip) to show that:

$$\begin{aligned}\mathsf{Val}(f) &= \frac{1}{2} + \frac{1}{2}\sum_{S\subseteq\{1,\ldots,n\}} \widehat{f}(S)^3(1-2\varepsilon)^{|S|} \\ &= \frac{1}{2} + \frac{1}{2}\sum_{S\subseteq\{1,\ldots,n\}} \widehat{f}(S)^2\left(\widehat{f}(S)\cdot(1-2\varepsilon)^{|S|}\right).\end{aligned}$$

Note that $\sum_S \widehat{f}(S)^2 = 1$ and since the function is balanced $\widehat{f}(\emptyset) = 0$. Thus it suffices to show that for every $S \neq \emptyset$, $\left|\widehat{f}(S)(1-2\varepsilon)^{|S|}\right| \le \delta$. Since the degree d-influence of each co-ordinate is at most η, it must be that for every set $S \neq \emptyset$, either $|S| > d$ or $\widehat{f}(S)^2 \le \eta$, as otherwise any co-ordinate in S will have degree d-influence at least η. Thus setting $\delta = \max\{(1-2\varepsilon)^d, \sqrt{\eta}\}$ proves the claim.

Inapproximability Result: Applying Theorem 2.2, gives the following inapproximability result proved by Håstad [25].

Theorem 3.1. *Assume* $\mathsf{P} \neq \mathsf{NP}$ *and let* $\varepsilon, \delta > 0$ *be arbitrarily small constants. Given an instance of the* Max-3Lin *problem that has an assignment satisfying* $1-\varepsilon$ *fraction of the equations, no polynomial time algorithm can find an assignment that satisfies* $\frac{1}{2}+\delta$ *fraction of the equations. In particular, there is no polynomial time algorithm for the* Max-3Lin *problem with approximation factor strictly less than* 2.

4. Max-kCSP and Gowers Uniformity

Max-kCSP **Problem:** Given a set of N boolean variables, and a system of *constraints* such that each constraint depends on k variables, find an assignment to the variables that satisfies a maximum fraction of constraints. For the ease of presentation, we assume that $k = 2^q - 1$ is a large constant.

Valuation: We define $\mathsf{Val}(f)$ to be the probability that f passes the *hypergraph linearity test* with perturbation. The test is a generalized and iterated version of the linearity test with perturbation in Section 3. Specifically, pick q inputs $x^1, \ldots, x^q \in \{-1,1\}^n$ at random. For every set $S \subseteq \{1,\ldots,q\}, |S| \geq 2$, let $w^S := \prod_{i \in S} x^i$ and z^S be obtained by flipping each bit of w^S with probability ε independently, i.e. $z^S \sim_{1-2\varepsilon} w^S$. The test passes if for every S, $f(z^S) = \prod_{i\in S} f(x^i)$, i.e.

$$\mathsf{Val}(f) := \Pr_{x^1,\ldots,x^q, z^S \sim_{1-2\varepsilon} w^S} \left[\forall |S| \geq 2,\ f(z^S) = \prod_{i \in S} f(x^i) \right].$$

Completeness: If $f \in \mathsf{DICT}$, it is easily seen that $\mathsf{Val}(f) \geq 1 - \varepsilon \cdot 2^q$, as there are $2^q - q - 1$ sets $|S| \geq 2$, and the test for each S could fail with probability ε due to the ε-noise/perturbation.

Soundness: It can be shown that if $f \in \mathsf{FFD}_{d,\eta}$, then $\mathsf{Val}(f) \leq \frac{1}{2^{2^q-q-1}} + \delta$ where $\delta \to 0$ as $d \to \infty$ and $\eta \to 0$. Note that there are $2^q - q - 1$ sub-tests, one for each $|S| \geq 2$. If f has all influences small, then these tests behave as if they were independent tests, each tests accepts with probability essentially $\frac{1}{2}$, and hence the probability that all tests accept simultaneously is essentially $\frac{1}{2^{2^q-q-1}}$.

Samorodnitsky and Trevisan [43] relate the acceptance probability of the test to the Gowers Uniformity norms [22] of a function, and then show that for a function with all influences small, the Gowers Uniformity norm is small as well.

Definition 4.1. Gowers Uniformity: *Let* $f : \{-1,1\}^n \mapsto \{-1,1\}$ *be a function, and* $\ell \geq 1$ *be an integer. The dimension-*ℓ *uniformity of* f *is defined as:*

$$U^\ell(f) := \mathbb{E}_{x,x^1,\ldots,x^\ell} \left[\prod_{S \subseteq \{1,\ldots,\ell\}} f\left(x \cdot \prod_{i \in S} x^i \right) \right].$$

Theorem 4.2. ([43]) *If* f *is a balanced function such that* $\forall i \in \{1,\ldots,n\}$, $\mathrm{Infl}_i(f) \leq \eta$, *then* $U^\ell(f) \leq \sqrt{\eta} \cdot 2^{O(\ell)}$.

Inapproximability Result:

Theorem 4.3. ([43]) *Assume the Unique Games Conjecture and let $\varepsilon, \delta > 0$ be arbitrarily small constants. Then given an instance of* Max-kCSP *problem, $k = 2^q - 1$, that has an assignment satisfying $1 - \varepsilon \cdot 2^q$ fraction of the constraints, no polynomial time algorithm can find an assignment that satisfies at least $\frac{1}{2^{2^q-q-1}} + \delta$ fraction of the constraints. In particular, there is no polynomial time algorithm for the* Max-kCSP *problem with approximation factor strictly less than $2^{2^q-q-1} = \theta(2^k/k)$.*

We note that an algorithm with approximation factor of $O(2^k/k)$ is known [8] and therefore the inapproximability result is nearly optimal.

5. Graph Partitioning and Bourgain's Noise Sensitivity Theorem

Graph Partitioning **Problem:** Given a graph $G(V, E)$, find a partition of the graph into two equal (or roughly equal) parts so as to minimize the fraction of edges cut. Note that this is a minimization problem.

Valuation: We define $\mathsf{Val}(f) = \mathsf{NS}_\varepsilon(f)$, the ε-noise sensitivity of f, i.e. the probability that f passes the perturbation test with ε-noise. Specifically, pick input $x \in \{-1, 1\}^n$ at random, and let y be a string obtained by flipping each bit of the string x with probability ε, i.e. $x \sim_{1-2\varepsilon} y$. Define

$$\mathsf{Val}(f) := \mathsf{NS}_\varepsilon(f) := \mathrm{Prob}_{x \sim_{1-2\varepsilon} y} \left[f(x) \neq f(y) \right].$$

The optimization problem concerns balanced cuts in graphs. Consider a complete graph with vertices $\{-1, 1\}^n$ and non-negative weights on edges where the weight of an edge (x, y) is exactly the probability that the pair (x, y) is picked by the perturbation test. View a balanced function $f : \{-1, 1\}^n \mapsto \{-1, 1\}$ as a cut in the graph. Thus $\mathsf{Val}(f)$ is exactly the total weight of edges cut by f.

Completeness: If $f \in \mathsf{DICT}$, then it is easily seen that $\mathsf{Val}(f) = \varepsilon$.

Soundness: It turns out that for large enough d and small enough η (depending on ε), if $f \in \mathsf{FFD}_{d,\eta}$, then $\mathsf{Val}(f) \geq \Omega(\sqrt{\varepsilon})$. This follows either from the Majority Is Stablest Theorem [38] that we will describe in the next section or essentially from the Bourgain's Theorem stated below. Bourgain's Theorem only gives a lower bound of $\Omega(\varepsilon^c)$ for any constant $c > \frac{1}{2}$, but its conclusion is stronger in the following sense: if the noise sensitivity of a balanced boolean function is $O(\varepsilon^c)$, then not only that f has a variable with significant influence, in fact f is close to a function that depends only on a bounded number of co-ordinates. The precise statement is:

Theorem 5.1. (Bourgain [7]) *Let $c > \frac{1}{2}$ be fixed. Then for all sufficiently small $\varepsilon > 0$, if f is a balanced function with ε-noise sensitivity $O(\varepsilon^c)$, then there is a* *boolean function g that agrees with f on 99% of the inputs and g depends only on $2^{O(1/\varepsilon^2)}$ co-ordinates.*

We would like to point out that Bourgain's Theorem came as an answer to a question posed by Håstad who was interested in such a theorem towards application to inapproximability.

Inapproximability Result: Applying Theorem 2.2 gives the following inapproximability result proved by Khot and Vishnoi [34]. The result applies to a generalization of the Graph Partitioning problem: one has so-called *demands*, i.e. a collection of pairs of vertices and we are interested in cuts that are balanced w.r.t. the demands, i.e. cuts that separate at least a constant fraction of the demands. The Graph Partitioning problem is a special case when all $\binom{|V|}{2}$ vertex pairs occur as demands.

Theorem 5.2. ([34]) *Assume the Unique Games Conjecture. Given a graph $G(V, E)$ along with demands that has a balanced partition that cuts at most ε fraction of the edges, no polynomial time algorithm can find a balanced partition that cuts at most $o(\sqrt{\varepsilon})$ fraction of the edges. In particular, there is no polynomial time algorithm for the* Graph Partitioning *problem with an approximation factor that is a universal constant.*

Connection to Metric Embeddings

The Graph Partitioning problem has a close connection to the theory of metric embeddings. We refer to Naor's article [39] for a detailed treatment of this connection and give a brief overview here. Theorem 5.2 rules out a constant factor approximation algorithm for the Graph Partitioning problem with demands; however the result is conditional on the Unique Games Conjecture. It is also interesting to have unconditional results that rule out a specific class of algorithms such as those based on *Semi-definite Programming relxation.* It turns out that the performance of an SDP algorithm for the Graph Partitioning problem is closely related to the question of embedding the *negative type* metrics into the class of ℓ_1 metrics. An N-point finite metric $d(\cdot, \cdot)$ is said to be of *negative type* if the metric $\sqrt{d}$ is isometrically embeddable in ℓ_2. Let $c_1(\mathsf{NEG}, N)$ be the least number such that every N-point negative type metric embeds into the class of ℓ_1 metrics with *distortion* $c_1(\mathsf{NEG}, N)$, i.e. preserving all distances up to a factor of $c_1(\mathsf{NEG}, N)$. It is known that $c_1(\mathsf{NEG}, N)$ is same up to a constant factor, the performance of the SDP algorithm for the Graph Partitioning problem on N-vertex graphs. Given an N-vertex graph that has a balanced partition that cuts ε fraction of the edges, the SDP algorithm finds a balanced partition that cuts $O(\varepsilon \cdot c_1(\mathsf{NEG}, N))$ fraction of the edges. Goemans and Linial [21, 37] conjectured that $c_1(\mathsf{NEG}, N)$ is a universal constant independent of N; this would be contrary to the statement of Theorem 5.2 since the theorem rules out every polynomial time algorithm that might achieve a constant factor approximation, and in particular an SDP-based one. In fact, using the techniques used to prove Theorem 5.2, Khot and Vishnoi [34] were able to disprove the Goemans and Linial conjecture:

Theorem 5.3. ([34]) $c_1(\mathsf{NEG}, N) \geq \Omega((\log\log N)^c)$ *for some constant* $c > 0$.

An interesting aspect of this theorem is that the construction of the negative type metric is inspired by the Unique Games Conjecture and the PCP reduction used to prove Theorem 5.2, but the construction itself is explicit and the lower bound unconditional. Regarding the upper bounds on $c_1(\mathsf{NEG}, N)$, in a breakthrough work, Arora, Rao, and Vazirani [3] showed that the SDP algorithm gives $O(\sqrt{\log N})$ approximation to the Graph Partitioning problem (without demands). This was extended to the demands version of the problem by Arora, Lee, and Naor [1], albeit with a slight loss in the approximation factor. As discussed, the latter result is equivalent to an upper bound on $c_1(\mathsf{NEG}, N)$.

Theorem 5.4. ([1]) $c_1(\mathsf{NEG}, N) \leq O(\sqrt{\log N} \cdot \log\log N)$.[3]

Using an alternate construction based on the geometry of Heisenberg group, a sequence of works by Lee and Naor [36], Cheeger and Kleiner [11, 12], Cheeger, Kleiner, and Naor [13, 14] obtained a stronger lower bound than Theorem 5.3:

Theorem 5.5. ([36, 11, 12, 13, 14]) $c_1(\mathsf{NEG}, N) \geq \Omega((\log N)^c)$ *for some constant* $c > 0$.

The lower bound of Theorem 5.3 is also strengthened in a different direction by Raghavendra and Steurer [40] (also by Khot and Saket [33] with quantitatively weaker result):

Theorem 5.6. ([40, 33]) *There is an N-point negative type metric such that its submetric on any subset of t points is isometrically ℓ_1-embeddable, but the whole metric incurs distortion of at least t to embed into ℓ_1, and $t = (\log\log N)^c$ for some constant $c > 0$.*

The KKL Theorem

A result of Kahn, Kalai, and Linial [27] was used by Chawla *et al* [10] to prove a theorem analogous to Theorem 5.2, and also by Krauthagamer and Rabani [35] and Devanur *et al* [15] to improve the lower bound in Theorem 5.3 to $\Omega(\log\log N)$. The KKL result has many other applications and we state it below:

Theorem 5.7. ([27]) *Every balanced boolean function $f : \{-1,1\}^n \mapsto \{-1,1\}$ has a variable whose influence is $\Omega\left(\frac{\log n}{n}\right)$.*

[3] Arora, Lee, and Naor [1] in fact give an embedding of an N-point negative type metric into ℓ_2 (which is isometrically embeddable into ℓ_1) with distortion $O(\sqrt{\log N} \cdot \log\log N)$. Since ℓ_1 metrics are of negative type, this gives an embedding of an N-point ℓ_1 metric into ℓ_2 with the same distortion. The result essentially matches a decades long lower bound of Enflo [17] who showed that embedding N-point ℓ_1 metric into ℓ_2 incurs distortion $\Omega(\sqrt{\log N})$.

6. Majority Is Stablest and Borell's Theorem

In the last section, we studied the ε-noise sensitivity of balanced boolean functions. Bourgain's Theorem gives a lower bound of $\Omega(\varepsilon^c)$ on the noise sensitivity of a balanced function whose all influences are small and $c > \frac{1}{2}$. We also mentioned that the Majority Is Stablest Theorem gives a lower bound of $\Omega(\sqrt{\varepsilon})$. In fact it gives an exact lower bound, namely $\frac{1}{\pi}\arccos(1-2\varepsilon)$, which turns out to be useful for an inapproximability result for the Max-Cut problem presented in the next section. Indeed, the Majority Is Stablest Theorem was invented for this application!

Theorem 6.1. (Mossel, O'Donnell, Oleszkiewicz [38]) *Let $0 < \varepsilon < \frac{1}{2}$. If $f \in \mathsf{FFD}_{d,\eta}$, then*

$$\mathsf{NS}_\varepsilon(f) \geq \frac{1}{\pi}\arccos(1-2\varepsilon) - \delta$$

and $\delta \to 0$ as $d \to \infty, \eta \to 0$.

We present a sketch of the proof as it demonstrates the connection to an isoperimetric problem in geometry and its solution by Borell [6]. The proof involves an application of the *invariance principle* that has also been studied by Rotar [42] and Chatterjee [9]. Here is a rough statement of the invariance principle:

Invariance Principle [38, 42, 9]: *Suppose f is a low degree multi-linear polynomial in n variables and all its variables have small influence. Then the distribution of the values of f is nearly identical when the input is a uniform random point from $\{-1,1\}^n$ or a random point from $\mathbb{R}^n$ with standard Gaussian measure.*

The invariance principle allows us to translate the noise sensitivity problem on boolean hypercube to a similar problem in the Gaussian space and the latter problem has already been solved by Borell! Towards this end, let $f \in \mathsf{FFD}_{d,\eta}$ be a boolean function on n-dimensional hypercube. We intend to lower bound its ε-noise sensitivity. We know that f has a representation as a multi-linear polynomial, namely its Fourier expansion:

$$f(x) = \sum_S \widehat{f}(S) \prod_{i\in S} x_i \qquad \forall x \in \{-1,1\}^n.$$

Let $f^* : \mathbb{R}^n \mapsto \mathbb{R}$ be a function that has the same representation as a multi-linear polynomial as f:

$$f^*(x^*) = \sum_S \widehat{f}(S) \prod_{i\in S} x_i^* \qquad \forall x^* \in \mathbb{R}^n. \tag{1}$$

Since $f \in \mathsf{FFD}_{d,\eta}$, all its influences are small. Assume for the moment that f is also of *low* degree. By the invariance principle, the distributions of $f(x)$ and $f^*(x^*)$ are nearly identical, and let us assume them to be identical for the sake of simplicity. This implies that $\mathbb{E}[f^*] = \mathbb{E}[f] = 0$ and since f is boolean, so is f^*. In other words, f^* is a partition of $\mathbb{R}^n$ (with Gaussian measure) into two sets of equal measure. The next observation is that the ε-noise sensitivity of f is same as the ε-"Gaussian noise sensitivity" of $f^* : \mathbb{R}^n \mapsto \{-1,1\}$. To be precise, let (x^*, y^*) be a pair of $(1-2\varepsilon)$-correlated n-dimensional Gaussians, i.e. for every co-ordinate

i, (x_i^*, y_i^*) are $(1-2\varepsilon)$-correlated standard Gaussians. One way to generate such a pair is to pick two independent standard n-dimensional Gaussians x^* and z^*, and let $y^* = (1-2\varepsilon)x^* + \sqrt{1-(1-2\varepsilon)^2}z^*$, and thus one can think of y^* as a small perturbation of x^*. Let the ε-noise sensitivity of a function $f^* : \mathbb{R}^n \mapsto \{-1, 1\}$ be defined as:

$$\mathsf{NS}_\varepsilon(f^*) := \Pr_{x^* \sim_{1-2\varepsilon} y^*}\left[f^*(x^*) \neq f^*(y^*)\right].$$

When f^* is a multi-linear polynomial as in (1), it is easily observed that

$$\mathsf{NS}_\varepsilon(f^*) = \frac{1}{2} - \frac{1}{2}\sum_S \widehat{f}(S)^2(1-2\varepsilon)^{|S|}.$$

But this expression is same as the ε-noise sensitivity of the boolean function f and thus $\mathsf{NS}_\varepsilon(f) = \mathsf{NS}_\varepsilon(f^*)$ and Theorem 6.1 follows from Borell's result that lower bounds $\mathsf{NS}_\varepsilon(f^*)$.

Theorem 6.2. (Borell [6]) *If $g^* : \mathbb{R}^n \mapsto \{-1, 1\}$ is a measurable function with $\mathbb{E}[g^*] = 0$, then*

$$\mathsf{NS}_\varepsilon(g^*) \geq \mathsf{NS}_\varepsilon(\mathsf{HALF\ SPACE}) = \frac{1}{\pi}\arccos(1-2\varepsilon),$$

where HALF-SPACE *is the partition of $\mathbb{R}^n$ by a hyperplane through origin.*

We note that the parameter δ in the statement of Theorem 6.1 accounts for additive errors involved at multiple places during the argument: firstly, the distributions $f(x)$ and $f^*(x^*)$ are only nearly identical. Secondly, even though $f \in \mathsf{FFD}_{d,\eta}$, f is not necessarily of bounded degree, and the invariance principle is not directly applicable. One gets around this issue by *smoothening* f that *kills* the high order Fourier coefficients (which are then discarded) and only slightly affecting the noise sensitivity. The *truncated* version of f has bounded degree and the invariance principle can be applied. We also note that the statement of Borell's Theorem holds for g^* that is $[-1, 1]$-valued when the noise sensitivity is defined as $\frac{1}{2} - \frac{1}{2}\langle g^*, T_{1-2\varepsilon}g^*\rangle$ and $T_{1-2\varepsilon}$ is the Ornstein-Uhlenbeck operator.

7. Max-Cut problem

Max-Cut **Problem:** Given a graph $G(V, E)$, find a partition that maximizes the number of edges cut.

Valuation: We define $\mathsf{Val}(f)$ as the ε-noise sensitivity of f for an appropriately chosen constant $\varepsilon > \frac{1}{2}$. Specifically, pick input $x \in \{-1, 1\}^n$ at random, and let y be a string obtained by flipping each bit of the string x with probability ε, i.e. $x \sim_{1-2\varepsilon} y$. Define

$$\mathsf{Val}(f) := \mathrm{Prob}_{x \sim_{1-2\varepsilon} y}\left[f(x) \neq f(y)\right].$$

The optimization problem concerns cuts in graphs. As in Section 5, we consider the complete graph with vertices $\{-1, 1\}^n$ and non-negative weights on edges representing the probability that a pair (x, y) is picked, and $f : \{-1, 1\}^n \mapsto \{-1, 1\}$

as a cut in the graph. An important thing to note here is that for the Graph Partitioning problem, the goal is to minimize the noise sensitivity for a balanced cut, and $\varepsilon > 0$ is a small constant. On the other hand, for the Max-Cut problem, the goal is to maximize the noise sensitivity, and $\varepsilon > \frac{1}{2}$.

Completeness: If $f \in \mathsf{DICT}$, then it is easily seen that $\mathsf{Val}(f) = \varepsilon$.

Soundness: The Majority Is Stablest Theorem states that MAJORITY is the most *stable* function among the class of low influence balanced boolean functions. It is implicit in this statement that the noise rate is strictly less than $\frac{1}{2}$. It turns out, essentially from the same theorem, that when the noise rate is above $\frac{1}{2}$, MAJORITY is the most *unstable* function among the class of low influence boolean functions (even including the unbalanced ones). This allows us to show that if $f \in \mathsf{FFD}_{d,\eta}$,

$$\mathsf{Val}(f) \le \frac{1}{\pi} \arccos(1 - 2\varepsilon) + \delta, \qquad \text{where } \delta \to 0 \text{ as } d \to \infty, \eta \to 0.$$

Inapproximability Result: Khot *et al* [29] proved the following inapproximability result for the Max-Cut problem and the Majority Is Stablest Theorem was conjectured therein.

Theorem 7.1. ([29]) *Assume the Unique Games Conjecture and let $\varepsilon > \frac{1}{2}$. Let $\delta > 0$ be an arbitrarily small constant. Given a graph $G(V, E)$ that has a partition that cuts at least ε fraction of the edges, no polynomial time algorithm can find a partition that cuts at least $\frac{1}{\pi}\arccos(1 - 2\varepsilon) + \delta$ fraction of the edges. In particular, there is no polynomial time algorithm for the* Max-Cut *problem with an approximation factor that is strictly less than $\frac{\varepsilon}{\frac{1}{\pi}\arccos(1-2\varepsilon)}$.*

In the above theorem, one can choose $\varepsilon > \frac{1}{2}$ so as to maximize the inapproximability factor. Let $\alpha_{GW} := \max_{\varepsilon \in [\frac{1}{2}, 1]} \frac{\varepsilon}{\frac{1}{\pi}\arccos(1-2\varepsilon)} \approx 1.13$. The theorem rules out an efficient algorithm with approximation factor strictly less than α_{GW}. On the other hand, the well-known SDP-based algorithm of Goemans and Williamson [20] achieves an approximation factor of exactly α_{GW} and thus is the optimal algorithm (modulo the Unique Games Conjecture).

8. Independent Set and the It Ain't Over Till It's Over Theorem

Independent Set **Problem:** Given a graph $G(V, E)$, find the largest independent set. A set $I \subseteq V$ is called independent if no edge of the graph has both endpoints in I. It is known from a result of Håstad [24], that given an N-vertex graph that has an independent set of size $N^{1-\varepsilon}$, no polynomial time algorithm can find an independent set of size N^{ε} unless $\mathsf{P} = \mathsf{NP}$. In this section, we are interested in the case when the graph is almost 2-colorable, i.e. has two disjoint independent sets of size $\left(\frac{1}{2} - \varepsilon\right) N$ each.

Valuation: We define $\mathsf{Val}(f)$ as probability that f is constant on a random εn dimensional sub-cube. For a set of co-ordinates $S \subseteq \{1, \ldots, n\}$ and a string $x \in \{-1,1\}^n$, a sub-cube $C_{S,x}$ corresponds to the set of all inputs that agree with x *outside* of S, i.e.

$$C_{S,x} := \{z \mid z \in \{-1,1\}^n,\ \forall i \notin S, z_i = x_i\}.$$

A random sub-cube $C_{S,x}$ of dimension εn is picked by selecting a random set $S \subseteq \{1, \ldots, n\}, |S| = \varepsilon n$ and a random string x. Define:

$$\mathsf{Val}(f) := \Pr_{|S|=\varepsilon n, x}\left[f \text{ is constant on } C_{S,x}\right].$$

The connection between this test and the Independent Set problem is rather subtle. One constructs a graph whose vertices are all pairs (C, b) where C is an εn-dimensional sub-cube and $b \in \{-1, 1\}$ is a bit. The intended purpose of this vertex is to capture the possibility that $f|_C \equiv b$. If two sub-cubes C, C' have non-empty intersection and $b \neq b'$, then we cannot have both $f_C = b$ and $f_{C'} = b'$, and we introduce an edge between vertices (C, b) and $C', b')$ to denote this conflict. This construction is known as the FGLSS construction, invented in [18]. It is not difficult to see that an independent set in this graph corresponds to a boolean function and the size of the independent set is proportional to the probability that the function passes the random sub-cube test.

Completeness: If $f \in \mathsf{DICT}$, then $f(x) = x_{i_0}$ for some fixed co-ordinate i_0. It is easily seen that for a random sub-cube $C_{S,x}$, unless $i_0 \in S$, f is constant on the sub-cube. Since $|S| = \varepsilon n$, we have $\mathsf{Val}(f) = 1 - \varepsilon$.

Soundness: If $f \in \mathsf{FFD}_{d,\eta}$, then it can be showed that $\mathsf{Val}(f) \leq \delta$ where $\delta \to 0$ as $d \to \infty, \eta \to 0$. It follows from the It Ain't Over Till It's Over Theorem of Mossel *et al* [38] which in fact says something stronger: if f has all influences small, then for almost all sub-cubes C, not only that f is non-constant on C, but f takes both the values $\{-1, 1\}$ on a constant fraction of points in C. A formal statement appears below:

Theorem 8.1. *For every $\varepsilon, \delta > 0$, there exist $\gamma, \eta > 0$ and integer d such that if $f \in \mathsf{FFD}_{d,\eta}$, and C is a random εn-dimensional sub-cube, then*

$$\Pr_C\left[\Big|\mathbb{E}[f(x)|x \in C]\Big| \geq 1 - \gamma\right] \ \leq\ \delta.$$

The theorem is proved using the invariance principle. Bansal and Khot [5] gave an alternate simple proof without using the invariance principle (the random sub-cube test is proposed therein), but the conclusion is only that f is non-constant on almost every sub-cube (which suffices for their application to Independent Set problem).

Inapproximability Result:

Theorem 8.2. ([5]) *Assume the Unique Games Conjecture and let $\varepsilon, \delta > 0$ be arbitrarily small constants. Then given an N-vertex graph $G(V, E)$ that is almost 2-colorable, i.e. has two disjoint independent sets of size $\left(\frac{1}{2} - \varepsilon\right) N$ each, no polynomial time algorithm can find an independent set of size δN.*

Friedgut's Theorem

Khot and Regev [32] proved a weaker result than Theorem 8.2: assuming the Unique Games Conjecture, given an N-vertex graph $G(V, E)$ that has an independent set of size $\left(\frac{1}{2} - \varepsilon\right) N$, no polynomial time algorithm can find an independent set of size δN. This gives $2 - \varepsilon$ inapproximability factor for the Vertex Cover problem.[4] The result is optimal since an algorithm that finds a maximal matching and takes all endpoints of the edges in the matching gives a 2-approximation for the Vertex Cover problem. Khot and Regev's paper (and its precursor Dinur and Safra [16]) use the following theorem of Friedgut [19]:

Theorem 8.3. ([19]) *Let $f : \{-1, 1\}^n \mapsto \{-1, 1\}$ be a function such that the average sensitivity (i.e. sum of all influences) is at most k. Then there exists a function g that agrees with f on $1 - \beta$ fraction of inputs and depends only on $2^{3k/\beta}$ co-ordinates.*

9. Kernel Clustering and the Propeller Problem

Kernel Clustering **Problem:** Given an $N \times N$ (symmetric) positive semidefinite matrix $A = (a_{ij})$ with $\sum_{i,j=1}^{N} a_{ij} = 0$, partition the index set $\{1, \ldots, N\}$ into k sets $T_1, \ldots, T_k$ so as to maximize $\sum_{\ell=1}^{k} \sum_{i,j \in T_\ell} a_{ij}$. In words, we seek to partition the matrix into $k \times k$ block diagonal form and then maximize the sum of entries of all diagonal blocks. Since the matrix is PSD, this sum is necessarily non-negative. The problem is actually a special case of the Kernel Clustering problem studied in [30, 31] and we don't state the more general problem here. We think of $k \geq 2$ as a small constant.

Valuation: We define $\mathsf{Val}(f)$ as the Fourier mass of f at the first level. We need to consider k-ary functions on k-ary hypercube, i.e. functions $f : \{1, \ldots, k\}^n \mapsto \{1, \ldots, k\}$. There is a natural generalization for the notions of dictatorship functions, Fourier representation, influences, and functions that are far from dictatorship. We don't formally state these notions here and directly state the definition of $\mathsf{Val}(f)$:

$$\mathsf{Val}(f) := \sum_{S \in \{0,1,\ldots,k-1\}^n, |S|=1} \widehat{f}(S)^2,$$

where $\widehat{f}(S)$ is the Fourier coefficient corresponding to a *multi-index* $S \in \{0, 1, \ldots, k-1\}^n$ and $|S|$ denotes the number of its non-zero co-ordinates. The connection between the Kernel Clustering problem and the specific valuation is that the (squared) Fourier mass is a PSD function of the values of f.

Completeness:[5] If $f \in \mathsf{DICT}$, then $\mathsf{Val}(f) = 1 - \frac{1}{k}$.

[4] A vertex cover in a graph is complement of an independent set. The Vertex Cover problem seeks to find a vertex cover of minimum size.

[5] When $k = 2$, we have boolean functions on boolean hypercube, and one would expect that for a dictatorship function, the Fourier mass at the first level equals 1. We instead get $\frac{1}{2}$ due to a slightly different (but equivalent) representation of functions.

Soundness: If $f \in \mathsf{FFD}_{d,\eta}$, then $\mathsf{Val}(f) \leq C(k) + \delta$ where $\delta \to 0$ as $d \to \infty$ and $\eta \to 0$. We would like to know functions that maximize the Fourier mass at the first level among the class of functions that are far from dictatorships. Since f has all its influences small, one can apply the invariance principle, and reduce this question to a certain geometric question, and the constant $C(k)$ is the solution to this geometric question. We state the geometric question below:

Definition 9.1. *Let $A_1, \ldots A_k$ be a partition of $\mathbb{R}^{k-1}$ into k measurable sets and for $1 \leq \ell \leq k$, let z_ℓ be the Gaussian moment vector over A_ℓ, i.e.*

$$z_\ell := \int_{A_\ell} x \; d\gamma \qquad \text{where } \gamma \text{ is standard Gaussian measure on } \mathbb{R}^{k-1}.$$

Then $C(k)$ is the supremum (it is achieved) of the sum of squared lengths of z_ℓ's over all possible partitions, i.e.

$$C(k) := \sup_{\mathbb{R}^{k-1} = A_1 \cup \ldots \cup A_k} \sum_{\ell=1}^{k} \|z_\ell\|^2. \tag{2}$$

It seems challenging to characterize an optimal partition for $k \geq 4$. For $k = 2$, the optimal partition of $\mathbb{R}$ into two sets is the partition into positive and negative real line, and $C(2) = \frac{1}{\pi}$. For $k = 3$, the optimal partition of $\mathbb{R}^2$ into three sets is the "propeller", i.e. partition into three cones with angle 120^o each, and $C(3) = \frac{9}{8\pi}$. One would expect that for $k = 4$, the optimal partition of $\mathbb{R}^3$ into four sets is the partition into four cones given by a regular tetrahedron. This turns out to be false as numerical computation shows that the value of this partition is worse than $C(3) = \frac{9}{8\pi}$ that can be achieved by letting $\mathbb{R}^3 = \mathbb{R}^2 \times \mathbb{R}$ and then partitioning $\mathbb{R}^2$ as a propeller. In fact Khot and Naor [30] conjecture that the propeller partition is the optimal one for any $k \geq 3$:

Conjecture 9.2. Propeller Conjecture: *For every $k \geq 3$, $C(k) = C(3)$. In words, the optimal partition of $\mathbb{R}^{k-1}$ into k sets in the sense of* (2) *is achieved by letting $\mathbb{R}^{k-1} = \mathbb{R}^2 \times \mathbb{R}^{k-3}$ and partitioning $\mathbb{R}^2$ as a propeller.*

Inapproximability Result:

Theorem 9.3. ([30, 31]) *Assume the Unique Games Conjecture and let $\varepsilon, \delta > 0$ be arbitrarily small constants. Then given an instance $A = (a_{ij})$ with value $1 - \frac{1}{k} - \varepsilon$, no polynomial time algorithm can find a solution with value at least $C(k) + \delta$. In particular, there is no polynomial time algorithm for the* Kernel Clustering *problem with approximation factor strictly less than $\frac{1-1/k}{C(k)}$.*

10. Conclusion

We have presented several examples to demonstrate the connections between inapproximability, discrete Fourier analysis, and geometry. There are many more examples and we conclude with pointing out a few:

- **Plurality is Stablest Conjecture:** In Section 6 and 7, we presented the connections between the Max-Cut problem, the Majority Is Stablest Theorem, and Borell's Theorem stating that a halfspace through origin is the most noise-stable balanced partition of $\mathbb{R}^n$. The Max-Cut problem can be generalized to the Max-kCut problem where one seeks to partition a graph into $k \geq 3$ sets so as to maximize the number of edges cut. An optimal inapproximability result for this problem is implied by the Plurality Is Stablest Conjecture stating that the Plurality function from $\{1, \ldots, k\}^n$ to $\{1, \ldots, k\}$ is the most stable under noise among the class of functions that are balanced and whose all influences are small. This conjecture in turn is implied by the Standard Simplex Conjecture stating that the standard k-simplex partition is the most noise-stable balanced partition of $\mathbb{R}^n$ with $n \geq k - 1$ (see [26]).

- **Sub-cube Test:** Consider a variant of the test discussed in Section 8: Assume that f is balanced, and one tests whether f is constant -1 on a random sub-cube of linear dimension. We know that if a function f passes the test with constant probability, say α, then it must have an influential variable. However f need not be close to a junta (i.e. a function depending on a bounded number of co-ordinates). Is it necessarily true that there is a function g that is close to a junta, monotonically above f, and passes the test with probability close to α? We say that g is monotonically above f if $\forall x, f(x) = 1 \implies g(x) = 1$. Such a result, though interesting on its own, might be useful towards inapproximability of graph coloring problem.

- **Lasserre Gaps:** Theorem 5.6 states that there is an N-point negative type metric that is locally ℓ_1-embeddable, but not globally ℓ_1-embeddable. In computer science, this result can be thought of as an *integrality gap* result for the so-called Sherali-Adams linear programming relaxation. An integrality gap result is an explicit construction showing that there is a gap between the true optimum and the optimum of the linear or semidefinite programming relaxation. Such results are taken as evidence that LP/SDP relaxation would not lead to a good approximation algorithm. There is a SDP relaxation known as Lasserre relaxation that is at least as powerful as the Sherali-Adams relaxation. It is a challenging open problem to prove integrality gap results for the Lasserre relaxation (for any problem of interest such as Max-Cut, Vertex Cover, or Unique Game). This could lead to interesting questions in Fourier analysis and/or geometry.

- **Small Set Expansion Problem:** Raghavendra and Steurer [41] give a connection between the small set expansion problem and the Unique Games Conjecture. Given an N-vertex graph, the goal is to find a set of vertices S of size δN that is nearly non-expanding, i.e. only a tiny fraction of edges incident on S leave S. One could conjecture that finding such sets is computationally intractable. Such a conjecture (see [41] for a formal statement) implies the Unique Games Conjecture as shown in [41].

- **Bounded Spectral Norm:** A result of Green and Sanders [23] states that every function $f : GF(2)^n \mapsto \{0, 1\}$ that has bounded spectral norm (defined as the sum of absolute values of its Fourier coefficients) can be expressed as a sum of a bounded number of functions each of which is an indicator function of an affine subspace of $GF(2)^n$. This result has the same flavor as "dictatorships are good; functions far from dictatorships are bad", except that now indicators of affine subspaces are considered as the "good" functions. Since there is such a close connection between such theorems and inapproximability results, it would be interesting to find an application to inapproximability, if there is one.

References

[1] S. Arora, J. Lee, and A. Naor. Euclidean distortion and the sparsest cut. In *Proc. 37th ACM Symposium on Theory of Computing*, pages 553–562, 2005.

[2] S. Arora, C. Lund, R. Motawani, M. Sudan, and M. Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, 1998.

[3] S. Arora, S. Rao, and U. Vazirani. Expander flows, geometric embeddings and graph partitioning. In *Proc. 36th ACM Symposium on Theory of Computing*, pages 222–231, 2004.

[4] S. Arora and S. Safra. Probabilistic checking of proofs : A new characterization of NP. *Journal of the ACM*, 45(1):70–122, 1998.

[5] N. Bansal and S. Khot. Optimal long code test with one free bit. In *Proc. 50th IEEE Symposium on Foundations of Computer Science*, 2009.

[6] C. Borell. Geometric bounds on the Ornstein-Uhlenbeck velocity process. *Z. Wahrsch. Verw. Gebiete*, 70(1):1–13, 1985.

[7] J. Bourgain. On the distribution of the Fourier spectrum of boolean functions. *Israel J. of Math.*, (131):269–276, 2002.

[8] M. Charikar, K. Makarychev, and Y. Makarychev. Near-optimal algorithms for maximum constraint satisfaction problems. In *SODA*, 2007.

[9] S. Chatterjee. A simple invariance theorem. *arXiv:math/0508213v1*, 2005.

[10] S. Chawla, R. Krauthgamer, R. Kumar, Y. Rabani, and D. Sivakumar. On the hardness of approximating multicut and sparsest-cut. In *Proc. 20th IEEE Conference on Computational Complexity*, pages 144–153, 2005.

[11] J. Cheeger and B. Kleiner. Differentiating maps into L^1 and the geometry of BV functions. *To appear in Ann. Math., preprint available at http://arxiv.org/abs/math/0611954*, 2006.

[12] J. Cheeger and B. Kleiner. On the differentiation of Lipschitz maps from metric measure spaces to Banach spaces. *Inspired by S.S. Chern, Volume 11 of Nankai Tracts. Math.*, pages 129–152, 2006.

[13] J. Cheeger, B. Kleiner, and A. Naor. Compression bounds for Lipschitz maps from the Heisenberg group to l_1. *Preprint*, 2009.

[14] J. Cheeger, B. Kleiner, and A. Naor. A $(\log n)^{\Omega(1)}$ integrality gap for the sparsest cut sdp. In *Proc. 50th IEEE Symposium on Foundations of Computer Science*, 2009.

[15] N. Devanur, S. Khot, R. Saket, and N. Vishnoi. Integrality gaps for sparsest cut and minimum linear arrangement problems. In *Proc. 38th ACM Symposium on Theory of Computing*, 2006.

[16] I. Dinur and S. Safra. The importance of being biased. In *Proc. 34th Annual ACM Symposium on Theory of Computing*, 2002.

[17] P. Enflo. On the nonexistence of uniform homeomorphisms between lp-spaces. *Ark. Mat.*, 8:103–105, 1969.

[18] U. Feige, S. Goldwasser, L. Lovász, S. Safra, and M. Szegedy. Interactive proofs and the hardness of approximating cliques. *Journal of the ACM*, 43(2):268–292, 1996.

[19] E. Friedgut. Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18(1):27–35, 1998.

[20] M. Goemans and D. Williamson. 0.878 approximation algorithms for MAX-CUT and MAX-2SAT. In *Proc. 26th ACM Symposium on Theory of Computing*, pages 422–431, 1994.

[21] M. X. Goemans. Semidefinite programming in combinatorial optimization. *Math. Program.*, 79:143–161, 1997.

[22] T. Gowers. A new proof of Szemeredi's theorem for progressions of length four. *Geometric and Functional Analysis*, 8(3):529–551, 1998.

[23] B. Green and T. Sanders. Boolean functions with small spectral norm. *Geom. Funct. Anal.*, 18(1):144–162, 2008.

[24] J. Hastad. Clique is hard to approximate within $n^{1-\epsilon}$. *Acta Mathematica*, 182:105–142, 1999.

[25] J. Hastad. Some optimal inapproximability results. *Journal of ACM*, 48:798–859, 2001.

[26] M. Isaksson and E. Mossel. New maximally stable gaussian partitions with discrete applications. *arXiv:0903.3362v3*, 2010.

[27] J. Kahn, G. Kalai, and N. Linial. The influence of variables on boolean functions. In *Proc. 29th Symposium on the Foundations of Computer Science*, pages 68–80, 1988.

[28] S. Khot. On the power of unique 2-prover 1-round games. In *Proc. 34th ACM Symposium on Theory of Computing*, 2002.

[29] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell. Optimal inapproximability results for max-cut and other 2-variable csps? In *Proc. 45th IEEE Symposium on Foundations of Computer Science*, pages 146–154, 2004.

[30] S. Khot and A. Naor. Approximate kernel clustering. In *Proc. 49th IEEE Symposium on Foundations of Computer Science*, 2008.

[31] S. Khot and A. Naor. Sharp kernel clustering algorithms and their associated Grothendieck inequalities. In *SODA*, 2010.

[32] S. Khot and O. Regev. Vertex cover might be hard to approximate to within $2-\epsilon$. In *Proc. 18th IEEE Conference on Computational Complexity*, 2003.

[33] S. Khot and R. Saket. Sdp integrality gaps with local ℓ_1-embeddability. In *Proc. 50th IEEE Symposium on Foundations of Computer Science*, pages 565–574, 2009.

[34] S. Khot and N. Vishnoi. The unique games conjecture, integrality gap for cut problems and embeddability of negative type metrics into ℓ_1. In *Proc. 46th IEEE Symposium on Foundations of Computer Science*, 2005.

[35] R. Krauthgamer and Y. Rabani. Improved lower bounds for embeddings into l_1. In *Proc. ACM-SIAM Symposium on Discrete Algorithms*, 2006.

[36] J. R. Lee and A. Naor. l_p metrics on the Heisenberg group and the Goemans-Linial conjecture. In *Proc. 47th IEEE Symposium on Foundations of Computer Science*, pages 99–108, 2006.

[37] N. Linial. Finite metric spaces-combinatorics, geometry and algorithms. In *Proc. International Congress of Mathematicians*, volume 3, pages 573–586, 2002.

[38] E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. In *Proc. 46th IEEE Symposium on Foundations of Computer Science*, pages 21–30, 2005.

[39] A. Naor. L1 embeddings of the Heisenberg group and fast estimation of graph isoperimetry. In *Proc. the International Congress of Mathematicians*, 2010.

[40] P. Raghavendra and D. Steurer. Integrality gaps for strong sdp relaxations of unique games. In *Proc. 50th IEEE Symposium on Foundations of Computer Science*, pages 575–585, 2009.

[41] P. Raghavendra and D. Steurer. Graph expansion and the unique games conjecture. In *Proc. 42nd ACM Symposium on Theory of Computing*, 2010.

[42] V. Rotar'. Limit theorems for polylinear forms. *J. Multivariate Anal.*, 9(4):511–530, 1979.

[43] A. Samorodnitsky and L. Trevisan. Gowers uniformity, influence of variables, and PCPs. In *Proc. 38th ACM Symposium on Theory of Computing*, 2006.

251 Mercer Street, Courant Institute of Mathematical Sciences, New York University,
New York, NY-10012, USA.
E-mail: khot@cs.nyu.edu