THESIS FOR THE DEGREE OF LICENTIATE OF ENGINEERING

Applications of Gaussian Noise Stability in Inapproximability and Social Choice Theory

MARCUS ISAKSSON

DEPARTMENT OF MATHEMATICAL SCIENCES
CHALMERS UNIVERSITY OF TECHNOLOGY AND GÖTEBORG UNIVERSITY
GÖTEBORG, SWEDEN 2009

Applications of Gaussian Noise Stability in Inapproximability and Social Choice Theory
Marcus Isaksson

ISSN 1652-9715/NO 2009:1
Department of Mathematical Sciences
Chalmers University of Technology and Göteborg University
412 96 GÖTEBORG, Sweden
Phone: +46 (0)31-772 10 00

Printed in Göteborg, Sweden, 2009

Applications of Gaussian Noise Stability in Inapproximability and Social Choice Theory

Marcus Isaksson
Department of Mathematical Sciences
Chalmers University of Technology and Göteborg University

ABSTRACT

Gaussian isoperimetric results have recently played an important role in proving fundamental results in hardness of approximation in computer science and in the study of voting schemes in social choice theory. In this thesis we prove a generalization of a Gaussian isoperimetric result by Borell and show that it implies that the majority function is optimal in Condorcet voting in the sense that it maximizes the probability that there is a single candidate which the society prefers over all other candidates. We also show that a different Gaussian isoperimetric conjecture which can be viewed as a generalization of the "Double Bubble" theorem implies the Plurality is Stablest conjecture and also that the Frieze-Jerrum semidefinite programming based algorithm for MAX-q-CUT achieves the optimal approximation factor assuming the Unique Games Conjecture. Both applications crucially depend on the invariance principle of Mossel, O'Donnell and Oleszkiewicz which lets us rephrase questions about noise stability of low-influential discrete functions in terms of noise stability of functions on $\mathbb{R}^n$ under Gaussian measure. We prove a generalization of this invariance principle needed for our applications.

Keywords: Gaussian noise stability, inapproximability theory, invariance principle, max-q-cut, condorcet voting.

Preface

This thesis contains the following papers:

- ▷ Marcus Isaksson and Elchanan Mossel, "Some Gaussian Noise Stability Conjectures and their Applications".
- ▷ Marcus Isaksson, "K-wise Gaussian Noise Stability".

Acknowledgments

I would like to thank my adviser Jeffrey Steif for inspiration, support and for many valuable comments on the draft of this thesis. I also want to thank my co-adviser Devdatt Dubhashi. Without you I would not be where I am. And of course the co-author of the first paper, Elchanan Mossel, who introduced me to the topics of this thesis.

I also want to thank all my colleagues at the department. Oscar, Janeli, Emilio, Alexandra, Marcus and many more for making my life here much more fun. Special thanks also to the floorball group and the less well-defined “lunch club”.

Finally, I want to thank my family for all your support and for always believing in me. And Eugenia for all your love and encouragement.

Marcus Isaksson
Göteborg, December 19, 2008

Contents

List of Figures

Part I

INTRODUCTION

1
Introduction

1.1 Gaussian Noise Stability

Gaussian noise stability measures the stability of partitions of Gaussian space under noise. In the simplest form we have two jointly standard Gaussian vectors X and Y in $\mathbb{R}^n$, with a covariance matrix $\mathbf{Cov}(X,Y) = \mathbf{E}[XY^T] = \rho I_n$, i.e. the coordinate pairs (X_i, Y_i) are i.i.d. $\mathrm{N}\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$. The stability of a subset A of $\mathbb{R}^n$ is defined to be the probability that both X and Y fall into A. Borell [3] proved that for sets of fixed Gaussian measure, half-spaces maximize this stability (it follows from his result that in an Ornstein-Uhlenbeck process the hitting time of sets of fixed measure is maximized by half-spaces). For simplicity, we will restrict attention to balanced partitions, i.e. sets of Gaussian measure $\frac{1}{2}$.

THEOREM 1. *[3] Fix $\rho \in [0,1]$. Suppose $X, Y \sim \mathrm{N}(0, I_n)$ are jointly normal and $\mathbf{Cov}(X,Y) = \rho I_n$. Let $A \subseteq \mathbb{R}^n$ with $\mathbf{P}(X \in A) = \frac{1}{2}$. Then*

$$\mathbf{P}(X \in A, Y \in A) \leq \mathbf{P}(X \in H, Y \in H)$$

where $H = \{x \in \mathbb{R}^n | x_1 \geq 0\}$.

In this thesis, several generalizations of this theorem are considered which are also motivated by applications.

- We may consider the situation with $k > 2$ correlated vectors.
- We may consider vectors that are negatively correlated, i.e. $\rho \in \left[-\frac{1}{k-1}, 0\right]$.
- Instead of selecting one set (and implicitly its complement) we may consider a partition of $\mathbb{R}^n$ into $q > 2$ subsets and ask for the probability that all k vectors fall into the same subset.
  We will still restrict attention to balanced partitions, i.e. into disjoint sets $A_1, \dots A_q \subseteq \mathbb{R}^n$ with equal Gaussian measure $\frac{1}{q}$.

It is conjectured that

- For fixed q, increasing k will not change the optimal partition. For instance, for $q = 2$ but $k \geq 3$ half-spaces would still be optimal.
- The most stable partition for positive ρ is the least stable partition for negative ρ. [1]
- If partitions with $q > 2$ subsets are considered then the stability, now defined as
  $$\mathbf{P}\left((X,Y) \in \bigcup_{j=1}^{q} A_j^2\right) \tag{1}$$
  (where A_j^2 denotes the Cartesian product $A_j \times A_j$) is maximized by a *standard simplex partition* (for $n \geq q - 1$).

A standard simplex partition divides $\mathbb{R}^n$ into q partitions depending on which of q maximally separated unit vectors are closest (ties may be broken arbitrarily):

DEFINITION 1. *For $n+1 \geq q \geq 2$, $A_1, \dots, A_q$ is a standard simplex partition of $\mathbb{R}^n$ if for all i*

$$A_i = \{x \in \mathbb{R}^n | x \cdot a_i > x \cdot a_j, \forall j \neq i\}$$

where $a_1, \dots a_q \in \mathbb{R}^n$ are q vectors satisfying $a_i \cdot a_j = \begin{cases} 1 & \text{if } i = j \\ -\frac{1}{q-1} & \text{if } i \neq j \end{cases}$

[1] This is known for $k = q = 2$

When $n \geq q$ a standard simplex partition can be formed by picking q orthonormal vectors $e_1, \ldots, e_q$, subtracting their mean and scaling appropriately, i.e.

$$a_i = \sqrt{\frac{q}{q-1}} \left(e_i - \frac{1}{q} \sum_{i=1}^{q} e_i \right)$$

and for $n = q-1$ it is enough to project these vectors onto the $q-1$-dimensional space which they span.

When $q = 3$ the standard simplex partition, also known as the *standard Y partition* or the *peace sign partition*, is described in $\mathbb{R}^2$ by three half-lines meeting at an 120 degree angle at the origin (Figure 1.1) and in $\mathbb{R}^n$, where $n > 2$, it can be exemplified by taking the Cartesian product of the peace sign partition and $\mathbb{R}^{n-2}$.

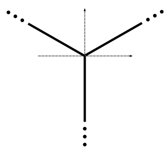

Figure 1.1: *The peace sign partition*

Paper I considers applications of two specific generalizations of Theorem 1. The first generalization was proved in Paper II: [2]

THEOREM 2. *Fix $\rho \in [0,1]$. Suppose $X_1, \ldots, X_k \sim \mathrm{N}(0, I_n)$ are jointly normal and $\mathbf{Cov}(X_i, X_j) = \rho I_n$ for $i \neq j$. Let $A \subseteq \mathbb{R}^n$ with $\mathbf{P}(X_i \in A) = \frac{1}{2}$. Then*

$$\mathbf{P}(\forall i : X_i \in A) \leq \mathbf{P}(\forall i : X_i \in H)$$

where $H = \{x \in \mathbb{R}^n | x_1 \geq 0\}$.

The second generalization is still open:

CONJECTURE 1. *Fix $\rho \in [0,1]$ and $3 \leq q \leq n+1$. Suppose $X, Y \sim \mathrm{N}(0, I_n)$ are jointly normal and $\mathbf{Cov}(X,Y) = \rho I_n$. Let $A_1, \ldots, A_q \subseteq \mathbb{R}^n$ be a balanced partition of $\mathbb{R}^n$. Then,*

$$\mathbf{P}((X,Y) \in A_1^2 \cup \cdots \cup A_q^2) \leq \mathbf{P}\left((X,Y) \in \left(S_1^2 \cup \cdots \cup S_q^2\right)\right) \qquad (2)$$

[2]This result has also been obtained independently by Guy Kindler and Elchanan Mossel.

where $S_1, \ldots, S_q$ *is a standard simplex partition of* $\mathbb{R}^n$. *Further, for* $\rho \in [-1, 0]$, (2) *holds in reverse:*

$$\mathbf{P}((X,Y) \in A_1^2 \cup \cdots \cup A_q^2) \geq \mathbf{P}\left((X,Y) \in \left(S_1^2 \cup \cdots \cup S_q^2\right)\right)$$

Since it is not known whether the second conjecture holds and the standard simplex partition is optimal, it should be pointed out that one of the main contribution of paper I is to show that the optimality of certain discrete problems can be reduced to the question of finding optimal partitions with respect to Gaussian noise stability.

1.2 The Invariance Principle

By the Fourier-Walsh transform, any Boolean function $f : \{-1, 1\}^n \to \{-1, 1\}$ can be written uniquely as a multilinear polynomial in the input variables

$$f(x) = \sum_{S \subseteq [n]} \hat{f}(S) \prod_{i \in S} x_i \tag{3}$$

The degree of f is

$$\deg f = \max_{S | \hat{f}(S) \neq 0} |S|$$

We will usually think of the input as being uniformly distributed over $\{-1, 1\}^n$ and denote it by X. For any coordinate $i \in [n]$ we may define its influence on $f(X)$ as the probability that changing the value of that coordinate will change the value of $f(X)$, i.e.

$$\mathrm{Inf}_i(f) = \mathbf{P}(f(X) \neq f(X^{(i)}))$$

where $X^{(i)}$ is obtained from X by flipping the i:th coordinate. Note that for a *dictator* function $\mathrm{DICT}_{n,i}(x) := x_i$ exactly one coordinate has influence 1 while the others have influence 0. For the *majority* function $\mathrm{MAJ}_n := 1_{\sum_{i=1}^n x_i > 0}$ one can show that each coordinate has influence $\Theta\left(\frac{1}{\sqrt{n}}\right)$. Thinking of the functions as social choice functions, that given n voters preferences between two candidates determines the winning candidate it is natural to ask which function minimizes the most influential voter. This was answered by the KKL theorem [4], THEOREM 3 (KKL). *For any $f : \{-1,1\}^n \to \{-1,1\}$ there exists an $i \in [n]$ such that*

$$\mathrm{Inf}_i(f) \geq \Omega\left(\mathbf{Var}(f)\frac{\log(n)}{n}\right)$$

In its simplest form, the invariance principle of [7], states that if f is of low degree and each coordinate has small influence on f, then the distribution of $f(X)$ will not change by much if we replace the X_i's in (3) by i.i.d. standard Gaussians $Z_i \sim \mathrm{N}(0,1)$. The change of the distribution is measured by an arbitrary $\mathcal{C}^3$ function Ψ having bounded third order derivatives.

THEOREM 4. *([7], special case of Theorem 3.18)*
Suppose $X_1,\ldots,X_n$ are i.i.d. uniform on $\{-1,1\}$, $f : \{-1,1\}^n \to \{-1,1\}$ has $\deg f \leq d$ and $\mathrm{Inf}_i f \leq \tau, \forall i$. Let $\Psi : \mathbb{R} \to \mathbb{R}$ be a $\mathcal{C}^3$ function with $|\Psi^{(\mathbf{r})}| \leq B$ for $|\mathbf{r}| = 3$. Then,

$$\left|\mathbf{E}\,\Psi(f(X)) - \mathbf{E}\,\Psi\left(\sum_{S\subseteq[n]}\hat{f}(S)\prod_{i\in S}Z_i\right)\right| \leq B10^d\tau$$

where $Z_1,\ldots Z_n$ are i.i.d $\mathrm{N}(0,1)$.

The theorems in [7] and [6] are much more general. For example

- The underlying probability space is generalized to an arbitrary finite product space $(\Omega,\mu) = (\prod_{i=1}^n \Omega_i, \prod_{i=1}^n \mu_i)$ where $|\Omega_i| < \infty, \forall i$. Functions $f : \Omega \to \mathbb{R}$ can still be written as a multilinear polynomial by constructing an orthonormal basis $\mathcal{X}_i = (\mathcal{X}_{i,0} = 1, \mathcal{X}_{i,1}, \ldots, \mathcal{X}_{i,|\Omega_i|-1})$ for the space of functions $\Omega_i \to \mathbb{R}$ and expressing f as

$$f(x) = \sum_{\sigma}\hat{f}(\sigma)\prod_{i=1}^{n}\mathcal{X}_{i,\sigma_i}(x)$$

  where the sum is over all tuples $\sigma = (\sigma_1,\ldots,\sigma_n)$ such that $0 \leq \sigma_i < |\Omega_i|$.

- Multidimensional functions $f : \Omega \to \mathbb{R}^k$ can be handled similarly using a test function $\Psi : \mathbb{R}^k \to \mathbb{R}$.

Paper I introduces a few more generalizations that are useful in applications.

- The $\mathcal{C}^3$ restriction on Ψ is removed and replaced with a Lipschitz continuity requirement.
- Non-orthonormal bases for the functions spaces $\Omega_i \to \mathbb{R}$ are handled (this was also discussed in [6]).

1.3 Plurality is Stablest

Consider an election with n voters choosing between q candidates. We call a function $f : [q]^n \to [q]$, which given the n votes determines the winning candidate, a social choice function. Letting $\Delta_q = \{x \in \mathbb{R}^q | x \geq 0, \sum_{i=1}^q x_i = 1\}$ denote the standard q-simplex, we can generalize this notion a bit and call a function $f : [q]^n \to \Delta_q$, which given the n votes assigns a probability distribution to the set of candidates, a "fuzzy" social choice function.

The noise stability of such functions measures the stability of the output when the votes are chosen independently and uniformly at random, and then re-randomized with probability ρ.

DEFINITION 2. *For $\rho \in [0, 1]$, the noise stability of $f : [q]^n \to \Delta_q$ is*

$$\mathbb{S}_\rho(f) = \mathbf{E} \sum_{j=1}^{q} f_j(\omega) f_j(\lambda)$$

where ω is uniformly selected from $[q]^n$ and each λ_i is independently selected using the conditional distribution

$$\mu(\lambda_i | \omega_i) = \rho 1_{\{\lambda_i = \omega_i\}} + (1 - \rho)\frac{1}{q}$$

We say that a social choice function f is balanced if $\mathbf{E}\, f_j(\omega) = \frac{1}{q}$ when $\omega \in [q]^n$ is chosen uniformly at random.

It is natural to require that a social choice function has low influence in each coordinate, so that a single voter has a very small chance of changing the outcome of the election. Another natural requirement is for the function to be as noise stable as possible, so that even if an ϵ fraction of the votes are miscounted the result is unlikely to change. One application considered in Paper I is to show that for balanced functions f having low influence in each coordinate, the most stable function is essentially determined by the most stable partition of

Gaussian space into q subsets as in (1). It is conjectured that the noise stability is maximized by the plurality function $\mathrm{PLUR}_{n,q}$, which assigns a mass 1 to the most popular candidate (ties broken arbitrarily).

CONJECTURE 2 (Plurality is Stablest).
For any $q \geq 2$*,* $\rho \in [0,1]$ *and* $\epsilon > 0$ *there exists a* $\tau > 0$ *such that if* $f : [q]^n \to \Delta_q$ *is a balanced function with* $\mathrm{Inf}_i(f_j) \leq \tau, \forall i, j$*, then*

$$\mathbb{S}_\rho(f) \leq \lim_{n\to\infty} \mathbb{S}_\rho(\mathrm{PLUR}_{n,q}) + \epsilon \quad \text{if } \rho \geq 0$$

This is already known [7] under the name *Majority is stablest* in the case $q = 2$. In paper I we show that the general Plurality is Stablest conjecture follows from Conjecture 1.

THEOREM 5. *Conjecture 1* $\Rightarrow$ *Conjecture 2*

1.4 Inapproximability Theory

1.4.1 Introduction to computational complexity theory

In computational complexity theory, one is interested in the asymptotics of the amount of time (or space) required to compute discrete functions. For simplicity we will assume that all combinatorial objects used (numbers, sets, graphs, formal mathematical proofs etc.) are represented as binary strings, i.e. elements in $\Sigma^* = \bigcup_{n\in\mathbb{N}}\{0,1\}^n$. The exact encoding used for different objects does not matter for our purposes (as long as it is a reasonable one). The length of a string $x \in \Sigma^*$ is denoted by $|x|$.

In general a *computational problem* is defined by a function $f : \Sigma^* \to \Sigma^*$. A *decision problem* is a problem which can be answered by *yes* or *no*. For instance,

- 3-COLOR: given a graph, can the vertices be colored using 3 colors such that no neighboring vertices have the same color?

- TRUE_Γ: given a proposition T in a formal mathematical theory Γ and an empty proof consisting of n zeroes [3] , does there exist a formal proof of

[3]The reason that we include an empty proof of length n in the instance and not just the number n is that the number n is encoded by a string of length $\Theta(log(n))$ but we later want a polynomial in the length of the instance to be polynomial in n.

T of length at most n?

DEFINITION 3. *A* decision problem *L is a subset of Σ^*.*

The complexity class **P** consists of all decision problems that can be computed in polynomial time (on any (and thus all) universal Turing machines, which the reader may think of as a regular computer equipped with unlimited amount of memory). If an algorithm's running time is bounded above by a polynomial in the length of the input (for some fixed universal Turing machine) we say that it is a polynomial time algorithm.

DEFINITION 4. *The complexity class **P** consists of all decision problems L for which there exists a polynomial time algorithm A such that*

$$\begin{cases} x \in L \Rightarrow A(x) = yes \\ x \notin L \Rightarrow A(x) = no \end{cases}$$

The complexity class **NP** consists of all decision problems for which *yes* instances have proofs that can be verified in polynomial time.

DEFINITION 5. *The complexity class **NP** consists of all decision problems L for which there exists a polynomial q and a polynomial time algorithm (verifier) V such that*

$$\begin{cases} x \in L \Rightarrow \exists \Pi \in \Sigma^* \textit{ such that } |\Pi| \leq q(|x|) \textit{ and } V(\Pi) = \textit{yes} \\ x \notin L \Rightarrow \forall \Pi \in \Sigma^* : V(\Pi) = \textit{no} \end{cases}$$

Note that both 3-COLOR and TRUE$_\Gamma$ are in **NP**. For instance, for 3-COLOR the verifier V can be taken to be an algorithm that simply checks that Π is a string that describes a coloring of all vertices in the graph in a way such that no neighboring vertices have the same color. Clearly, such a Π exists iff $x \in$ 3-COLOR.

Further, **P** $\subseteq$ **NP**, since for $L \in$ **P** we can simply ignore the proof Π and use the algorithm A as verifier. It remains an open problem whether **P** $=$ **NP**, although equality would be very surprising (implying e.g. that mathematical theorems can be proved in time polynomial in the statement and the length of the proof).

In inapproximability theory one is interested in showing non-existence of polynomial time algorithms for approximating combinatorial optimization problems (assuming **P** $\neq$ **NP**). Let us first define combinatorial optimizations problems.

DEFINITION 6. *A combinatorial maximization problem is defined by a function* $f : \Sigma^* \times \Sigma^* \to \mathbb{R} \cup \{-\infty\}$ *assigning a value* $f(x, l)$ *to any solution* l *of an instance* x *such that for each* x*, there are only a finite number of solutions* l *(called* feasible *for* x*) for which* $f(x, l) \neq -\infty$.
An instance x *is said to be valid if it has a feasible solution* l.
The value of an instance $x \in \Sigma$ *is*

$$\mathrm{VAL}(x) = \max_l f(x, l)$$

A minimization problem is defined similarly by replacing the $\max$ *by* $\min$ *and* $-\infty$ *by* $+\infty$.

We can now define the corresponding complexity classes **PO** and **NPO**.

DEFINITION 7. *The complexity class* ***NPO*** *consists of all combinatorial optimization problems* f *for which there exist*

i) a polynomial time algorithm that determines whether an instance x *is valid.*

ii) a polynomial q *such that for any instance* x*, all feasible solutions* l *satisfy* $|l| \leq q(|x|)$.

iii) a polynomial time algorithm that computes f.

PO *is the subset of* ***NPO*** *for which* $\mathrm{VAL}(x)$ *is computable by a polynomial time algorithm.*

There is a natural pre-ordering of computational problems given by polynomial time reducibility.

DEFINITION 8. *Given two computational problems* X *and* Y*, we say that* X *is polynomial time reducible to* Y*, denoted* $X \leq_P Y$*, if there exists a polynomial time algorithm* A *which computes the value of instances* $x \in X$ *in polynomial time, given access to an oracle for* Y *(i.e. a hypothetical algorithm that computes* Y *in constant time).*

From this we may define the complexity classes **NP-complete** consisting of the hardest problems in **NP** and **NP-hard** consisting of all problems that are at least as hard as **NP**. More generally, DEFINITION 9. *Let $\mathcal{C}$ be a complexity class. Then $\mathcal{C}-\mathbf{hard}$ consists of all computational problems Y such that $X \leq_P Y, \forall X \in \mathcal{C}$. Further, $\mathcal{C}-\mathbf{complete} = \mathcal{C}-\mathbf{hard} \cap \mathcal{C}$*

1.4.2 Approximation algorithms

Many NP-hard optimization problems (for which no polynomial time algorithm exists unless **P** $=$ **NP**) are possible to approximate within a constant factor in polynomial time. For instance, for the Euclidean Traveling Salesman Problem where one is given a set of points in Euclidean space, computing the shortest round-trip route visiting all points is NP-hard. However, for any $\epsilon > 0$ there exist a polynomial time approximation algorithm that computes a route no more than $1 + \epsilon$ times longer than the optimal route.

DEFINITION 10. *If $f : \Sigma^* \times \Sigma^* \to \mathbb{R} \cup \{-\infty\}$ is a maximization problem in **NPO**, A is an algorithm and $r \in [0, 1)$, we say that A is an r-approximation algorithm for f if for all valid instances x,*

$$f(x, A(x)) \geq r \,\mathrm{VAL}(x)$$

Similarly, if f is a minimization problem and $r > 1$ we say that A is an r-approximation algorithm for f if for all valid instances x,

$$f(x, A(x)) \leq r \,\mathrm{VAL}(x)$$

Thus the Euclidean Traveling Salesman Problem has a polynomial time $1+\epsilon$-approximation algorithm for any $\epsilon > 0$.

Other problems can only be efficiently approximated up to a certain approximation constant. For instance, consider MAX-3-SAT defined as

DEFINITION 11. *An instance of the MAX-3-SAT problem consists of m clauses, each being a disjunction (logical or) of at most three literals, where each literal is either a variable or the negation of a variable from a set of n Boolean variables $b_1, \ldots, b_n$. A feasible solution is an assignment $l : [n] \to \{0, 1\}$ to these variables. The value $f(x, l)$ of an assignment is the fraction of clauses that are satisfied by the assignment.*

For MAX-3-SAT there exist a $\frac{7}{8}$ approximation algorithm based on semidefinite programming [10]. For the restricted problem MAX-E3-SAT, where we require that each clause contains exactly three (different) variables, then this can be achieved by picking a random assignment which will satisfy a $\frac{7}{8}$ fraction of the clauses in expectation (this algorithm can be derandomized by repeatedly setting each variable to the value which maximizes the conditional expectation over the remaining variables). On the other hand it is known [9] that no $\frac{7}{8} + \epsilon$ polynomial time approximation can be achieved (unless $\mathbf{P} = \mathbf{NP}$), for any $\epsilon > 0$.

$$(b_1 \vee \neg b_2 \vee b_4) \wedge (\neg b_1 \vee \neg b_3 \vee b_2) \wedge (\neg b_2 \vee b_3 \vee b_5)$$

Figure 1.2: *A MAX-E3-SAT instance. All 3 clauses can be satisfied simultaneously so the value is 1.*

MAX-3-SAT is an example of class of optimization problems called Constraint Satisfaction Problems (CSP's).

DEFINITION 12. *A Constraint Satisfaction Problem (CSP) $\Lambda = (P, q)$ is specified by a set of predicates P over the finite domain $[q]$. The arity of Λ is the maximal arity of the predicates in P.*
An instance of Λ consists of a set of variables $x_1, \ldots, x_n$ and a set of predicates from P, each applied to a subset of the variables and their negations.

Thus, MAX-3-SAT is a ternary CSP over a Boolean domain.

1.4.3 The PCP Theorem and the Unique Games Conjecture

The $\frac{7}{8} + \epsilon$ inapproximability result for MAX-3-SAT (and similar results for other CSP's) is obtained by a reduction from a standard problem called the Label Cover problem for which arbitrarily good inapproximability results exist.

DEFINITION 13.
An instance of the Label Cover *problem, $\mathcal{L}(V, W, E, M, N, \{\sigma_{v,w}\}_{(v,w)\in E})$, consists of a bipartite graph $(V \cup W, E)$ with a function $\sigma_{v,w} : [M] \to [N]$ associated with every edge $(v, w) \in E \subseteq V \times W$. A labeling $l = (l_V, l_W)$, where $l_V : V \to [M]$ and $l_W : W \to [N]$, is said to satisfy an edge (v, w) if*

$$\sigma_{(v,w)}(l_W(w)) = l_V(v)$$

The value of a labeling l, $\mathrm{VAL}_l(\mathcal{L})$, is the fraction of edges satisfied by l and the value of $\mathcal{L}$ is the maximal fraction of edges satisfied by any labeling,

$$\mathrm{VAL}(\mathcal{L}) = \max_l \mathrm{VAL}_l(\mathcal{L})$$

The PCP (Probabilistically Checkable Proofs) theorem [1, 2] asserts that the Label Cover problem is NP-hard to approximate within any constant $\epsilon > 0$, for suitable choices of M and N.

THEOREM 6 (Label Cover version of the PCP Theorem). *For any $\epsilon > 0$ there exists a M and N such that it is NP-hard to distinguish between instances $\mathcal{L}$ of the Label Cover problem with label set sizes M and N having $\mathrm{VAL}(\mathcal{L}) = 1$ from those having $\mathrm{VAL}(\mathcal{L}) \leq \epsilon$.*

This implies that any problem in **NP** (for instance TRUE_Γ) has a probabilistically checkable proof, which can be verified by looking only at a constant (depending on ϵ, but not on the length of the instance $|x|$) number of bits in such a way that a false proof is accepted with probability ϵ while a correct proof is always accepted. The proof structure is given by the polynomial time reduction from the **NP** problem to a Label Cover problem for which a correct proof (assignment) satisfies all edges while any other (incorrect) proof satisfies at most an ϵ fraction of the edges.

However, the PCP theorem is not strong enough to give sharp inapproximability results for binary CSP's (2-CSP's). To this end Khot [5] introduced the Unique Games Conjecture.

DEFINITION 14. *A* Label Cover *problem $\mathcal{L}(V, W, E, M, N, \{\sigma_{v,w}\}_{(v,w)\in E})$ is called* unique *if $M = N$ and each $\sigma_{v,w} : M \to M$ is a permutation.*

CONJECTURE 3 (Unique Games Conjecture). *For any $\eta, \gamma > 0$ there exists a $M = M(\eta, \gamma)$ such that it is NP-hard to distinguish instances $\mathcal{L}$ of the Unique Label Cover problem with label set size M having $\mathrm{VAL}(\mathcal{L}) \geq 1-\eta$ from those having $\mathrm{VAL}(\mathcal{L}) \leq \gamma$.*

It was recently shown [8] how to obtain optimal approximation algorithms for any CSP including 2-CSP's assuming the Unique Games Conjecture. However, the optimal approximation constants in [8] are generally not very explicit but given as the optimum of certain optimization problems. It should be noted that it is still not known whether the Unique Games Conjecture holds.

1.4.4 MAX-q-CUT

In Paper I, we consider the MAX-q-CUT problem or Approximate q-Coloring, where given a (possible edge weighted) graph one seeks a coloring of the vertices using q colors that minimizes the number of edges between nodes of the same color (i.e. maximizes the number of edges between different colors).

DEFINITION 15.
An instance of the weighted MAX-q-CUT problem, $\mathcal{M}_q(V, E, w)$, consists of a graph (V, E) with a weight function $w : E \to [0, 1]$ assigning a weight to each edge. A q-cut $l : V \to [q]$ is a partition of the vertices into q parts. The value of a q-cut l is

$$\mathrm{VAL}_l(\mathcal{M}_q) = \sum_{(u,v) \in E : l(u) \neq l(v)} w_{(u,v)}$$

The value of $\mathcal{M}_q$ is

$$\mathrm{VAL}(\mathcal{M}_q) = \max_l \mathrm{VAL}_l(\mathcal{M}_q)$$

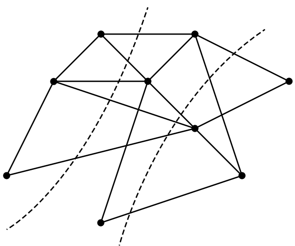

Figure 1.3: *In MAX-3-CUT we want to find a partition of the vertices into 3 sets so as to maximize the weight of edges between different sets.*

Note that MAX-q-CUT is a (weighted) binary CSP over the alphabet $[q]$.

In Paper I we find the optimal inapproximability constant for MAX-q-CUT assuming the unique games conjecture and Conjecture 1.

THEOREM 7. *Assume Conjecture 1 and the UGC. Then, for any $\epsilon > 0$ there exist a polynomial time algorithm that approximates MAX-q-CUT within $\alpha_q - \epsilon$ while no algorithm exists the approximates MAX-q-CUT within $\alpha_q + \epsilon$.*

Here,

$$\alpha_q = \inf_{-\frac{1}{q-1} \le \rho \le 1} \frac{q}{q-1} \frac{1 - qI(\rho)}{1-\rho}$$

where $qI(\rho)$ is the noise stability the standard simplex partition, i.e.

$$qI(\rho) = \mathbf{P}((X,Y) \in S_1^2 \cup \cdots \cup S_q^2)$$

where $X, Y \sim \mathrm{N}(0, I_{q-1})$ are jointly normal with $\mathbf{Cov}(X,Y) = \rho I_{q-1}$ and $S_1, \ldots S_q$ is a standard simplex partition of $\mathbb{R}^{q-1}$.

For instance, for $q = 3$ this value is given by

$$\alpha_3 = \inf_{-\frac{1}{2} \le \rho \le 1} \frac{1 - \frac{9}{8\pi^2}(\arccos(-\rho)^2 - \arccos(\rho/2)^2)}{1-\rho} \approx 0.83601$$

References

[1] Arora and Safra, *Probabilistic checking of proofs; A new characterization of NP (draft)*, FOCS: IEEE Symposium on Foundations of Computer Science (FOCS), 1992.

[2] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy, *Proof verification and intractability of approximation problems*, Proc. 33rd IEEE Symp. on Foundations of Computer Science, 1992.

[3] C. Borell, *Geometric bounds on the ornstein-uhlenbeck velocity process*, Probability Theory and Related Fields **70** (1985), no. 1, 1–13.

[4] J. Kahn, G. Kalai, and N. Linial, *The influence of variables on boolean functions*, Proceedings of the 29th Annual Symposium on Foundations of Computer Science, 1988, pp. 68–80.

[5] S. Khot, *On the power of unique 2-prover 1-round games*, STOC '02: Proceedings of the thirty-fourth annual ACM symposium on Theory of computing (New York, NY, USA), ACM, 2002, pp. 767–775.

[6] E. Mossel, *Gaussian bounds for noise correlation of functions*, Submitted, 2008.

[7] E. Mossel, R. O'Donnell, and K. Oleszkiewicz, *Noise stability of functions with low influences: invariance and optimality*, To appear in Ann. Math., 2008.

[8] Prasad Raghavendra, *Optimal algorithms and inapproximability results for every CSP?*, Proceedings of the 40th Annual ACM Symposium on Theory of Computing, ACM, 2008, pp. 245–254.

[9] J. Håstad, *Some optimal inapproximability results*, Journal of the ACM, 1997, pp. 1–10.

[10] U. Zwick, *Computer assisted proof of optimal approximability results*, Proc. of 13th SODA, 2002, pp. 496–505.

Part II

PAPERS

PAPER I

Some Gaussian Noise Stability Conjectures and their Applications

Marcus Isaksson and Elchanan Mossel

2
PAPER I

ABSTRACT
Gaussian isoperimetric results have recently played an important role in proving fundamental results in hardness of approximation in computer science and in the study of voting schemes in social choice. We propose two Gaussian isoperimetric conjectures and derive consequences of the conjectures in hardness of approximation and social choice. Both conjectures generalize isoperimetric results by Borell on the heat kernel. One of the conjectures may be also be viewed as a generalization of the "Double Bubble" theorem. The applications of the conjecture include an optimality result for majority in the context of Condorcet voting and a proof that the Frieze-Jerrum SDP for MAX-q-CUT achieves the optimal approximation factor assuming the Unique Games Conjecture.

2.1 Introduction

Recent results in hardness of approximation in computer science and in the study of voting schemes in social choice crucially rely on Gaussian isoperimetric results. The first result in hardness of approximation established a tight inapproximability result for MAX-CUT assuming unique games [11] while the latest results achieve optimal inapproximation factors for very general families of constraint satisfaction problems [16]. Results in social choice include optimality of the majority function among low influence functions in the context of Condorcet voting on 3 candidates [8] and near optimality for any number of candidates [13]. A common feature of these results is the use of "Invariance Principles" [13, 15] together with optimal Gaussian isoperimetric results [1].

In the current paper we propose two conjectures generalizing the results of Borell [1] and develop an extension of the invariance principle so that assuming the conjectures new results in hardness of approximation and in social choice are obtained. In the introduction we state the conjectures and their applications.

2.1.1 The Conjectures

We will be concerned with finding partitions of $\mathbb{R}^n$ that maximizes the probability that correlated Gaussian vectors remain within the same part. More specifically we would like to partition $\mathbb{R}^n$ into $q \geq 2$ disjoint sets of equal Gaussian measure.

Borell [1] proved that when $q = 2$ and we have two standard Gaussian vectors with covariance $\rho \geq 0$ in corresponding coordinates then half-spaces (e.g. $H := \{x \in \mathbb{R}^n | x_1 \geq 0\}$) are optimal. Let I_n be the $n \times n$ identity matrix. For two n-dimensional random variables $X = (X_1, \dots, X_n)$ and $Y = (Y_1, \dots, Y_n)$ write $\mathbf{Cov}(X, Y)$ for the $n \times n$ matrix whose (i, j)'th entry is given by $\mathbf{Cov}[X_i, Y_j] = \mathbf{E}[X_i Y_j] - \mathbf{E}[X_i]\,\mathbf{E}[Y_j]$. Borell's result states the following:

THEOREM 1. *[1] Fix $\rho \in [0, 1]$. Suppose $X, Y \sim \mathrm{N}(0, I_n)$ are jointly normal and $\mathbf{Cov}(X, Y) = \rho I_n$. Let $A \subseteq \mathbb{R}^n$ with $\mathbf{P}(X \in A) = \frac{1}{2}$. Then*

$$\mathbf{P}(X \in A, Y \in A) \leq \mathbf{P}(X \in H, Y \in H)$$

We conjecture that Theorem 1 can be generalized in two different directions . The first conjecture claims that half-spaces are still optimal if we have $k > 2$ correlated vectors and seek to maximize the probability that they all fall into the same part.

CONJECTURE 1. *Fix* $\rho \in [0,1]$. *Suppose* $X_1, \ldots, X_k \sim \mathrm{N}(0, I_n)$ *are jointly normal and* $\mathbf{Cov}(X_i, X_j) = \rho I_n$ *for* $i \neq j$. *Let* $A \subseteq \mathbb{R}^n$ *with* $\mathbf{P}(X_i \in A) = \frac{1}{2}$. *Then*

$$\mathbf{P}(\forall i : X_i \in A) \leq \mathbf{P}(\forall i : X_i \in H) \tag{1}$$

We call the conjecture above the *Exchangeable Gaussians Conjecture* (EGC). Recall that a collection of random variables is exchangeable if its distribution is invariant under any permutation.

The second conjecture generalizes Theorem 1 by asking for the optimal partition of $\mathbb{R}^n$ into $q > 2$ sets of equal measure. We conjecture that the optimal partition can be formed by splitting the standard $(q-1)$-simplex into q parts determined by the closest q-dimensional basis vector and further that this is the *least stable* partition for $\rho \leq 0$.

Let $S'_q = \{x \in \mathbb{R}^q | \sum_{j=1}^q x_i = 1\}$ be the affine hyperplane containing the standard $(q-1)$-simplex and take $M : \mathbb{R}^q \to \mathbb{R}^{q-1}$ to be a mapping from this hyperplane to $\mathbb{R}^{q-1}$ by letting $M = M_2 M_1$, where $M_1 = I_q - \frac{\mathbf{1_q}\mathbf{1_q}^t}{q}$ is the projection along the vector $\mathbf{1_q}$ and M_2 is any orthogonal linear mapping with nullspace $\{a\mathbf{1_q} | a \in \mathbb{R}\}$. For $1 \leq j \leq q$, let $S'_{q,j} = \{x \in S'_q | x_j > x_i, \forall i \neq j\}$ with mapping $S_{q,j} = M(S'_{q,j}) \subseteq \mathbb{R}^{q-1}$.

We call $A_1, \ldots, A_q$ a balanced partition of $\mathbb{R}^n$ if $A_1, \ldots, A_q$ are disjoint with $\mathbf{P}(X \in A_j) = \frac{1}{q}, \forall j$.

CONJECTURE 2. *Fix* $\rho \in [0,1]$ *and* $3 \leq q \leq n+1$. *Suppose* $X, Y \sim \mathrm{N}(0, I_n)$ *are jointly normal and* $\mathbf{Cov}(X,Y) = \rho I_n$. *Let* $A_1, \ldots, A_q \subseteq \mathbb{R}^n$ *be a balanced partition of* $\mathbb{R}^n$. *Then,*

$$\mathbf{P}((X,Y) \in A_1^2 \cup \cdots \cup A_q^2) \leq \mathbf{P}\left((X,Y) \in \left(S_{q,1}^2 \cup \cdots \cup S_{q,q}^2\right) \times \mathbb{R}^{n+1-q}\right) \tag{2}$$

Further, for $\rho \in [-1, 0]$, (2) *holds in reverse:*

$$\mathbf{P}((X,Y) \in A_1^2 \cup \cdots \cup A_q^2) \geq \mathbf{P}\left((X,Y) \in \left(S_{q,1}^2 \cup \cdots \cup S_{q,q}^2\right) \times \mathbb{R}^{n+1-q}\right)$$

The particular case of $q = 3$ is easier to visualize and we call this the "Peace Sign Partition". For this reason we call the conjecture above, the Piece Sign Conjecture (PSC).

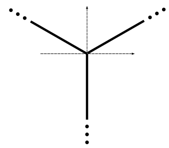

Figure 2.1: *The peace sign partition*

2.1.2 Applications

We show that the two conjectures have natural applications in Social Choice Theory. The conjectures imply

- The *Plurality is Stablest conjecture* as well as showing that the Frieze-Jerrum [4] SDP relaxation obtains the optimal approximation ratio for MAX-q-CUT assuming the Unique Games Conjecture.
- Certain optimality of majority in Condorcet Voting. More specifically, it asymptotically maximizes the probability of a unique winner in Condorcet voting with any number of candidates.

The main tool for proving these applications is the invariance principle of [13, 15] which we extend to handle general Lipschitz continuous functions. We note that previous work proved the invariance principle for $\mathcal{C}^3$ functions and some specific Lipschitz continuous functions. The generalization of the invariance principle may be of independent interest.

We proceed with formal statements of the applications.

2.1.2.1 Plurality is Stablest

Consider an election with n voters choosing between q candidates. We call a function $f : [q]^n \to [q]$, which given the n votes determines the winning candidate, *a social choice function*. Letting $\Delta_q = \{x \in \mathbb{R}^q | x \geq 0, \sum_{i=1}^q x_i = 1\}$ denote the standard q-simplex, we generalize this notion a bit and call a function $f : [q]^n \to \Delta_q$ assigning a probability distribution to the set of candidates a *"fuzzy" social choice function.* To be able to treat non-fuzzy social choice functions at the same time, we will usually embed their output into Δ_q and think of them as functions $f : [q]^n \to E_q$, where $E_q = \{(1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)\}$ are the q extreme points of Δ_q corresponding to assigning a probability mass 1 to one of the candidates.

The noise stability of such functions measures the stability of the output when the votes are chosen independently and uniformly at random, and then rerandomized with probability $1-\rho$.

DEFINITION 1. *For* $-\frac{1}{q-1} \le \rho \le 1$*, the noise stability of* $f : [q]^n \to \mathbb{R}^k$ *is*

$$\mathbb{S}_\rho(f) = \sum_{j=1}^{k} \mathbf{E}[f_j(\omega) f_j(\lambda)]$$

where ω *is uniformly selected from* $[q]^n$ *and each* λ_i *is independently selected using the conditional distribution*

$$\mu(\lambda_i|\omega_i) = \rho 1_{\{\lambda_i=\omega_i\}} + (1-\rho)\frac{1}{q} \tag{3}$$

Note that when $f : [q]^n \to E_q$ is a non-fuzzy social choice function $\mathbb{S}_\rho(f) = \mathbf{P}(f(\omega) = f(\lambda))$.

We say that $f : [q]^n \to \Delta_q$ is *balanced* if $\mathbf{E}[f(\omega)] = \frac{1}{q}\mathbf{1}$ where ω is uniformly selected from $[q]^n$ and say that the influence of the i:th coordinate on $f : [q]^n \to \mathbb{R}$ is

$$\mathrm{Inf}_i\, f(\omega) = \mathop{\mathbf{E}}_{\omega}[\mathbf{Var}_{\omega_i}\, f(\omega)]$$

Let $\mathrm{PLUR}_{n,q} : [q]^n \to \Delta_q$ denote the plurality function which assigns a probability mass 1 to the candidate with the most votes (ties can be broken arbitrarily, e.g. by splitting the mass equally among the tied candidates). The *Plurality is Stablest* conjecture claims that plurality is essentially the most stable of all low-influence functions under uniform measure:

CONJECTURE 3 (Plurality is Stablest). *For any* $q \ge 2$*,* $\rho \in [-\frac{1}{q-1}, 1]$ *and* $\epsilon > 0$ *there exists a* $\tau > 0$ *such that if* $f : [q]^n \to \Delta_q$ *is a balanced function with* $\mathrm{Inf}_i(f_j) \le \tau, \forall i, j$*, then*

$$\mathbb{S}_\rho(f) \le \lim_{n\to\infty} \mathbb{S}_\rho(\mathrm{PLUR}_{n,q}) + \epsilon \quad \textit{if } \rho \ge 0 \tag{4}$$

and

$$\mathbb{S}_\rho(f) \ge \lim_{n\to\infty} \mathbb{S}_\rho(\mathrm{PLUR}_{n,q}) - \epsilon \quad \textit{if } \rho \le 0$$

The case where $q = 2$, the *Majority is stablest theorem*, was proved in [15]. We show that the general case follows from PSC.

THEOREM 2. *PSC (Conj. 2) $\Rightarrow$ Plurality is Stablest (Conj. 3)*

It should be pointed out that our results imply a slightly stronger result where the low influence requirement is replaced by a low *low-degree influence* requirement. This strengthening turns out to be crucial to applications in hardness of approximation.

It is known [11] that the bound (4) in Conjecture 3 holds asymptotically as $q \to \infty$ up to a small multiplicative constant, i.e.

$$\mathbb{S}_\rho(f) \le \mathcal{O}_q(1) \cdot \lim_{n\to\infty} \mathbb{S}_\rho(\mathrm{PLUR}_{n,q}) + \epsilon \quad \text{if } \rho \ge 0$$

It may be helpful to think of the theorem in terms of a pure social choice function $f : [q]^n \to [q]$. In this case, there are n voters and each voter chooses one out of q possible candidates. Given individual choices $x_1, \ldots, x_n$, the winning candidate is defined to be $f(x_1, \ldots, x_n)$. In social choice theory it is natural to restrict attention to the class of low influence functions, where each individual voter has small effect on the outcome. We now consider the scenario where voters have independent and uniform preferences. Moreover, we assume that there is a problem with the voting machines so that each vote cast is rerandomized with probability $1 - \rho$. Denoting by $X_1, \ldots, X_n$ the intended votes and $Y_1, \ldots, Y_n$ the registered votes, it is natural to wonder how correlated are $f(X_1, \ldots, X_n)$ and $f(Y_1, \ldots, Y_n)$. The theorem above states that under PSC, the maximal amount of correlation is obtained for the plurality function if $\rho \ge 0$. The case where $\rho < 0$ corresponds to the situation where the voting machine's rerandomization mechanism favors votes that differ from the original vote. In this case the theorem states that plurality will have the least correlation between the intended outcome $f(X_1, \ldots, X_n)$ and the registered outcome $f(Y_1, \ldots, Y_n)$. In the next subsection we discuss applications of the result for hardness of approximation.

2.1.2.2 Hardness of approximating MAX-q-CUT

For NP-hard optimization problems it is natural to search for polynomial time approximation algorithms that are guaranteed to find a solution with value within a certain constant of the optimal value. Hardness of approximation results on the other hand bound the achievable approximation constants away from 1. For some problems, tight hardness results have been show where the bound matches the best known polynomial time approximation algorithm. For instance, Håstad [6] showed that for MAX-E3-SAT one cannot improved upon the simple randomized algorithm picking assignments at random thus achieving an approximation ratio of $\frac{7}{8}$.

In general, for constraint satisfaction problems (CSP's) where the object is to maximize the number of satisfied predicates selected from a set of allowed predicates and applied to a given set of variables, algorithms based on relaxations to semi-definite programming (SDP), first introduced by Goemans and Williamson [5] has proved very successful.

Still optimal hardness results are not known for many CSP's. To make progress on this Khot [10] introduced the Unique Games Conjecture (UGC), a strengthened form of the PCP Theorem. Recently Raghavendra [16] showed tight hardness results for any CSP assuming the UGC, albeit without giving explicit optimal approximation constants.

We consider the MAX-q-CUT or the Approximate q-Coloring problem where given a weighted graph on seeks a q-coloring of the vertices that maximizes the total weight of edges between differently colored vertices.

DEFINITION 2.
An instance of the weighted MAX-q-CUT problem, $\mathcal{M}_q(V, E, w)$, consists of a graph (V, E) with a weight function $w : E \to [0, 1]$ assigning a weight to each edge. A q-cut $l : V \to [q]$ is a partition of the vertices into q parts. The value of a q-cut l is

$$\mathrm{VAL}_l(\mathcal{M}_q) = \sum_{(u,v)\in E: l(u)\neq l(v)} w_{(u,v)}$$

The value of $\mathcal{M}_q$ is

$$\mathrm{VAL}(\mathcal{M}_q) = \max_l \mathrm{VAL}_l(\mathcal{M}_q)$$

Frieze-Jerrum gave an explicit SDP relaxation of MAX-q-CUT (see Section 2.6.3) which was rounded using the standard simplex partition of Conjecture 2. We show that Conjecture 2 implies that this is optimal.

THEOREM 3. *Assume Conjecture 2 and the UGC. Then, for any $\epsilon > 0$ there exist a polynomial time algorithm that approximates MAX-q-CUT within $\alpha_q - \epsilon$ while no algorithm exists the approximates MAX-q-CUT within $\alpha_q + \epsilon$.*

Here,

$$\alpha_q = \inf_{-\frac{1}{q-1} \le \rho \le 1} \frac{q}{q-1} \frac{1 - qI(\rho)}{1-\rho}$$

where $qI(\rho)$ *is the noise stability of the standard simplex partition of* $\mathbb{R}^{q-1}$*, i.e.*

$$qI(\rho) = \mathbf{P}((X,Y) \in S_{q,1}^2 \cup \cdots \cup S_{q,q}^2)$$

where $X, Y \sim \mathrm{N}(0, I_{q-1})$ *are jointly normal with* $\mathbf{Cov}(X,Y) = \rho I_{q-1}$.

2.1.2.3 Condorcet voting

Suppose n voters rank k candidates. It is assumed that each voter i has a linear order $\sigma_i \in S(k)$ on the candidates. In *Condorcet voting*, the rankings are aggregated by deciding for each pair of candidates which one is superior among the n voters.

More formally, the aggregation results in a tournament G_k on the set $[k]$. Recall that G_k is a *tournament* on $[k]$ if it is a directed graph on the vertex set $[k]$ such that for all $a, b \in [k]$ either $(a > b) \in G_k$ or $(b > a) \in G_k$. Given individual rankings $(\sigma_i)_{i=1}^n$ the tournament G_k is defined as follows. Let

$$x_i^{a>b} = \begin{cases} 1 & \text{if } \sigma_i(a) > \sigma_i(b) \\ -1 & \text{else} \end{cases}, \text{ for } i \in [n] \text{ and } a, b \in [k].$$

Note that $x^{b>a} = -x^{a>b}$. The binary decision between each pair of candidates is performed via a anti-symmetric function $f : \{-1,1\}^n \to \{0,1\}$ so that $f(-x) = 1 - f(x)$ for all $x \in \{-1,1\}^n$. The tournament $G_k = G_k(\sigma; f)$ is then defined by letting $(a > b) \in G_k$ if and only if $f(x^{a>b}) = 1$. A natural decision function is the majority function $\mathrm{MAJ}_n : \{-1,1\}^n \to \{0,1\}$ defined by $\mathrm{MAJ}_n(x) = 1_{\sum_{i=1}^n x_i \ge 0}$.

Note that there are $2^{\binom{k}{2}}$ tournaments while there are only $k! = 2^{\Theta(k \log k)}$ linear rankings. For the purposes of social choice, some tournaments make more sense than others.

Following [8, 9, 13], we consider the probability distribution over n voters, where the voters have independent preferences and each one chooses a ranking uniformly at random among all $k!$ orderings. Note that the marginal distributions on vectors $x^{a>b}$ is the uniform distribution over $\{-1,1\}^n$ and that if $f : \{-1,1\}^n \to \{0,1\}$ is anti-symmetric then $\mathbf{E}[f] = \frac{1}{2}$. The previous discussion and the following definition are essentially taken from [13].

DEFINITION 3. *For any anti-symmetric function* $f : \{-1,1\}^n \to \{0,1\}$ *let* $\mathrm{UniqueBest}_k(f)$ *denote the event that the Condorcet voting system described above results in a unique best candidate and* $\mathrm{UniqueBest}_k(f,i)$ *the event that the i:th candidate is unique best.*

The case that is now understood is $k = 3$. Note that in this case G_3 is unique max if and only if it is linear. Kalai [8] studied the *probability* of a rational outcome given that the n voters vote independently and at random from the 6 possible rational rankings. He showed that the probability of a rational outcome in this case may be expressed as $3\,\mathbb{S}_{1/3}(f)$.

It is natural to ask which function f with small influences is most likely to produce a rational outcome. Instead of considering small influences, Kalai considered the essentially stronger assumption that f is monotone and "transitive-symmetric"; i.e., that for all $1 \le i < j \le n$ there exists a permutation σ on $[n]$ with $\sigma(i) = j$ such that $f(x_1,\ldots,x_n) = f(x_{\sigma(1)},\ldots,x_{\sigma(n)})$ for all $(x_1,\ldots,x_n)$. Kalai conjectured that as $n \to \infty$ the maximum of $3\mathbb{S}_{1/3}(f)$ among all transitive-symmetric functions approaches $\lim_{n\to\infty} 3\,\mathbb{S}_{1/3}(\mathrm{MAJ}_n)$. This is a direct consequence of the Majority is Stablest Theorem proved in [14, 15]. In [13] similar, but sub-optimal results were obtained for any value of k. More specifically it was shown that if one considers Condorcet voting on k candidates, then for all $\epsilon > 0$ there exists $\tau = \tau(k,\epsilon) > 0$ such that if $f : \{-1,1\}^n \to \{0,1\}$ is anti-symmetric and $\mathrm{Inf}_i(f) \le \tau$ for all i, then

$$\mathbf{P}[\mathrm{UniqueBest}_k(f)] \le k^{-1+o_k(1)} + \epsilon.$$

Moreover for the majority function we have $\mathrm{Inf}_i(\mathrm{MAJ}_n) = O(n^{-1/2})$ and it holds that

$$\mathbf{P}[\mathrm{UniqueBest}_k(\mathrm{MAJ}_n)] \ge k^{-1-o_k(1)} - o_n(1).$$

Here we provide tight results for every value of k assuming EGC by showing that:

THEOREM 4. *Assume Conjecture 1. Then, for any* $k \ge 1$ *and* $\epsilon > 0$ *there exists a* $\tau(\epsilon,k) > 0$ *such that for any anti-symmetric* $f : \{-1,1\}^n \to \{0,1\}$ *satisfying* $\max_i \mathrm{Inf}_i f \le \tau$,

$$\mathbf{P}[\mathrm{UniqueBest}_k(f)] \le \lim_{n\to\infty} \mathbf{P}[\mathrm{UniqueBest}_k(\mathrm{MAJ}_n)] + \epsilon$$

.

2.1.3 The PSC and the Double Bubble Theorem

The famous Double Bubble Theorem [7] determines the minimal area that encloses and separates two fixed volumes in $\mathbb{R}^3$. The optimal partition is given by two spheres which intersect at an $120 \deg$ angle having a separating membrane in the plane of the intersection. The proof of this theorem is the culmination of a long line of work answering a conjecture which was open for more than a century.

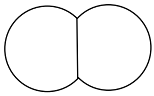

Figure 2.2: *A double bubble in* $\mathbb{R}^2$

An analogous question can be asked in Gaussian space, $\mathbb{R}^n$ equipped with a standard Gaussian density and the techniques and results used in the proof of the Double Bubble Theorem allow to find the partition of $\mathbb{R}^n (n \geq 2)$ into three volumes each having Gaussian volume $\frac{1}{3}$ minimizing the Gaussian surface area between the three volumes. Indeed, the results of [2] show that the optimal partition is the Peace Sign partition [1], which can be seen as the limit of the double bubble partition scaled up around one point on the intersection.

This indicates that the partition in Conjecture 2 is optimal (at least for $q = 3$ when $\rho \to 1$). Indeed Conjecture 2 is stronger than the results of [2]. It is easy to see that Conjecture 2 with $q = 3$ imply that the "standard Y" or "Peace Sign" are optimal by taking the limit $\rho \to 1$ (this is done similarly to the way in which Borell's result [1] implies the classical Gaussian isoperimetric result, see Ledoux's Saint-Flour lecture notes [3]).

2.2 Preliminaries

2.2.1 Multilinear polynomials

Consider a product probability space $(\Omega, \mu) = (\prod_{i=1}^n \Omega_i, \prod_{i=1}^n \mu_i)$. We will be interested in functions $f : \prod_{i=1}^n \Omega_i \to \mathbb{R}$ on such spaces. For simplicity, we will assume that each μ_i as full support, i.e. $\mu_i(\omega_i) > 0, \forall \omega_i \in \Omega_i$. Then

[1] Called the standard Y in that paper

clearly, for each coordinate i we can create a (possibly orthonormal) basis of the form

$$\mathcal{X}_i = (X_{i,0} = 1, X_{i,1}, \dots, X_{i,|\Omega_i|-1})$$

where $E[X_{i,j}] = 0$ for $j \geq 1$, for the space of functions $\Omega_i \to \mathbb{R}$.

DEFINITION 4. *We call a finite sequence of (orthonormal) real-valued random variables where the first variable is the constant* 1 *and the other variables have zero mean an (orthonormal) ensemble.*

Thus, $\mathcal{X} = (\mathcal{X}_1, \dots, \mathcal{X}_n)$ is an independent sequence of (possibly orthonormal) ensembles. We will only be concerned with independent sequences of ensembles, however we will not always require the ensembles to be orthonormal [2]. Another type of ensembles are the Gaussian ensembles, of which an independent sequence is typically denoted by $\mathcal{Z} = (\mathcal{Z}_1, \dots, \mathcal{Z}_n)$ where $\mathcal{Z}_i = (Z_{i,0} = 1, Z_{i,1}, \dots, Z_{i,m_i})$ and each $Z_{i,j}$ is a standard Gaussian variable.

DEFINITION 5. *A multi-index* σ *is a sequence of numbers* $(\sigma_1, \dots, \sigma_n)$ *such that* $\sigma_i \geq 0, \forall i$. *The degree* $|\sigma|$ *of* σ *is* $|\{i \in [n] : \sigma_i > 0\}|$. *Given a set of indeterminates* $\{x_{i,j}\}_{i\in[n],0\leq j\leq m_i}$, *let* $x_\sigma = \prod_{i=1}^n x_{i,\sigma_i}$. *A multilinear polynomial over such a set of indeterminates is an expression* $Q(x) = \sum_\sigma c_\sigma x_\sigma$ *where* $c_\sigma \in \mathbb{R}$ *are constants.*

Continuing from (2.2.1) and letting $X_\sigma = \prod_{i=1}^n X_{i,\sigma_i}$ it should be clear that $\{X_\sigma\}$ forms a basis for functions $\prod_{i=1}^n \Omega_i \to \mathbb{R}$, hence any function $f : \prod_{i=1}^n \Omega_i \to \mathbb{R}$ can be expressed as a multilinear polynomial Q over $\mathcal{X}$:

$$f(\omega_1, \dots, \omega_n) = Q(\mathcal{X}_1, \dots, \mathcal{X}_n) = \sum_\sigma c_\sigma X_\sigma \tag{5}$$

DEFINITION 6. *The degree of a multilinear polynomial* Q *is*

$$\deg Q = \max_{\sigma : c_\sigma \neq 0} |\sigma|$$

We will also use the notation $Q^{\leq d}$ *to denote the truncated multilinear polynomial*

$$Q^{\leq d}(x) = \sum_{\sigma : |\sigma| \leq d} c_\sigma x_\sigma$$

and the analogous for $Q^{=d}$ *and* $Q^{>d}$.

[2]Hence, we will deviate from the notation of [13, 15] where *sequences of ensembles* was used as an abbreviation for *sequences of orthonormal ensembles*.

DEFINITION 7. *Given a multilinear polynomial Q over an independent sequence of ensembles $\mathcal{X} = (\mathcal{X}_1, \ldots, \mathcal{X}_n)$, the influence of the i:th coordinate on $Q(\mathcal{X})$ is*

$$\mathrm{Inf}_i\, Q(\mathcal{X}) = \mathbf{E}\left[\mathbf{Var}[Q(\mathcal{X})|\mathcal{X}_1, \ldots, \mathcal{X}_{i-1}, \mathcal{X}_{i+1}, \ldots \mathcal{X}_n]\right]$$

We also define the d-degree influence of the i:th coordinate as

$$\mathrm{Inf}_i^{\leq d} Q(\mathcal{X}) = \mathrm{Inf}_i\, Q^{\leq d}(\mathcal{X})$$

Note that neither the degree nor influences of $Q(\mathcal{X})$ depends on the actual basis selected in (2.2.1), hence we can write $\deg f = \deg Q$, $\mathrm{Inf}_i\, f = \mathrm{Inf}_i\, Q(\mathcal{X})$ and $\mathrm{Inf}_i^{\leq d} f = \mathrm{Inf}_i\, Q^{\leq d}(\mathcal{X})$.

2.2.2 Bonami-Beckner noise

Let us first define the Bonami-Beckner noise operator.

DEFINITION 8. *Let $(\Omega, \mu) = (\prod_{i=1}^n \Omega_i, \prod_{i=1}^n \mu_i)$. be a finite product probability space and α the minimum probability of any atom in any Ω_i. For $-\frac{\alpha}{1-\alpha} \leq \rho \leq 1$ the Bonami-Beckner operator on functions $f : \prod_{i=1}^n \Omega_i \to \mathbb{R}^k$ is defined by*

$$T_\rho f(\omega_1, \ldots, \omega_n) = \mathbf{E}[f(\lambda_1, \ldots, \lambda_n)|\omega_1, \ldots \omega_n]$$

where each λ_i is independently selected from the conditional distribution

$$\mu(\lambda_i|\omega_i) = \rho 1_{\{\lambda_i = \omega_i\}} + (1-\rho)\mu(\lambda_i)$$

For $\rho \in [0, 1]$ this is equivalent to $T_\rho f$ being the expected value of f when each coordinate independently is rerandomized with probability $1 - \rho$.

2.2.3 Orthonormal ensembles

Most of the time we will work with *orthonormal* ensembles. Using independence and linearity of expectation it is easy to see that if $Q(\mathcal{X}) = \sum_\sigma c_\sigma \mathcal{X}_\sigma$ is a multilinear polynomial over an independent sequence of *orthonormal* ensembles, then

$$\mathbf{E}[Q(\mathcal{X})] = c_{\mathbf{0}}\,; \qquad \mathbf{Var}[Q(\mathcal{X})] = \sum_{\sigma:|\sigma|>0} c_\sigma^2\,; \quad T_\rho Q(\mathcal{X}) = \sum_\sigma \rho^{|\sigma|} c_\sigma X_\sigma$$

$$\mathbf{E}[Q(\mathcal{X})^2] = \sum_\sigma c_\sigma^2\,; \quad \mathrm{Inf}_i\, Q(\mathcal{X}) = \sum_{\sigma:\sigma_i>0} c_\sigma^2\,; \quad \mathrm{Inf}_i^{\le d}\, Q(\mathcal{X}) = \sum_{\sigma:\left\{\substack{\sigma_i>0\\ |\sigma|\le d}\right.} c_\sigma^2$$

Combining these expressions it is also easy to see that $\mathrm{Inf}_i^{\le d} f$ is convex in f and satisfy the following bound on the sum of low-degree influences:

$$\sum_{i=1}^n \mathrm{Inf}_i^{\le d} f \le d\,\mathbf{Var}\, f \tag{7}$$

2.2.4 Vector-valued functions

Since we will work extensively with vector-valued functions we make the following definitions:

DEFINITION 9. *For a vector-valued function* $f = (f_1, \ldots, f_k)$*, let*

$$\mathbf{Var}\, f = \sum_{j=1}^k \mathbf{Var}\, f_j\,, \quad \mathrm{Inf}_i\, f = \sum_{j=1}^k \mathrm{Inf}_i\, f_j$$

and similarly for $\mathrm{Inf}_i^{\le d}$.

Thus (7) holds even for vector-valued f. Also, all expressions in (6) hold for vector-valued multilinear polynomials $Q(\mathcal{X}) = \sum_\sigma c_\sigma \mathcal{X}_\sigma$, where $c_\sigma \in \mathbb{R}^k$ and $\mathcal{X}$ is an independent sequence of *orthonormal* ensembles, if we replace c_σ^2 with $||c_\sigma||_2^2$.

Finally, by expressing functions $f : [q]^n \to \mathbb{R}^k$ under the uniform measure on the input space $[q]^n$ as a multilinear polynomial

$$f(\omega) = \sum_\sigma c_\sigma \prod_{i=1}^n X_{i,\sigma_i}(\omega_i)$$

this lets us express the noise stability of Definition 1 as

$$\mathbb{S}_\rho(f) = \mathbf{E}[\langle f, T_\rho f\rangle] = \sum_\sigma \rho^{|\sigma|} ||c_\sigma||_2^2 \tag{8}$$

2.2.5 Correlated probability spaces

It will be important for us to bound the effect of the Bonami-Beckner noise operator on functions on correlated probability spaces.

DEFINITION 10. *Let $(\Omega_1 \times \Omega_2, \mu)$ be a correlated probability space. The correlation between Ω_1 and Ω_2 with respect to μ is then*

$$\rho(\Omega_1, \Omega_2; \mu) = \sup_{f_i:\Omega_i \to \mathbb{R}, \mathbf{Var}\, f_i = 1} \mathbf{Cov}(f_1(\omega_1), f_2(\omega_2))$$

For $(\Omega_1 \times \cdots \times \Omega_k, \mu)$ we let

$$\rho(\Omega_1, \ldots, \Omega_k; \mu) = \max_{1 \le i \le k} \rho \left(\prod_{j=1}^{i-1} \Omega_j \times \prod_{j=i+1}^{k} \Omega_j, \Omega_i; \mu \right)$$

The following theorem shows that the expected value of products of functions where corresponding coordinates form correlated probability spaces does not change by much when some small noise is applied to each coordinate:

LEMMA 1. *[13, Lemma 6.2] Let $(\prod_{i=1}^n \Omega_i, \prod_{i=1}^n \mu_i)$ be a finite product probability space where $\Omega_i = (\Omega_i^1, \ldots, \Omega_i^k)$ are correlated probability spaces with $\rho(\Omega_i^1, \ldots, \Omega_i^k; \mu_i) \le \rho < 1$. Further, let $\mathcal{X}^j = (\mathcal{X}_1^j, \ldots, \mathcal{X}_n^j)$ be independent sequences of orthonormal ensembles such that $\mathcal{X}_i^j$ forms a basis for functions $\Omega_i^j \to \mathbb{R}$ and $Q_1, \ldots, Q_k$ multilinear polynomials such that $\mathbf{Var}\, Q_j(\mathcal{X}^j) \le 1$. Then, for all $\epsilon > 0$ there exists a $\gamma = \gamma(\epsilon, \rho) > 0$ such that*

$$\left| \mathbf{E} \prod_{j=1}^{k} Q_j(\mathcal{X}^j) - \mathbf{E} \prod_{j=1}^{k} T_{1-\gamma} Q_j(\mathcal{X}^j) \right| \le \epsilon \cdot k$$

2.2.6 Gaussian noise

DEFINITION 11. *Let $X \sim \mathrm{N}(0, I_n)$. The Ornstein-Uhlenbeck operator U_ρ is defined on functions $f : \mathbb{R}^n \to \mathbb{R}$ such that $f(X) \in L^2$ by*

$$U_\rho f(X) = \mathbf{E} \left[f(\rho X + \sqrt{1-\rho^2}\xi) | X \right]$$

where $\xi \sim \mathrm{N}(0, I_n)$ is independent of X.

It is easy to see that if $\mathcal{Z} = (\mathcal{Z}_1, \dots, \mathcal{Z}_n)$ is a Gaussian sequence of independent ensembles and $Q(\mathcal{Z}) = \sum_\sigma c_\sigma \mathcal{Z}_\sigma$, then

$$U_\rho Q(\mathcal{Z}) = \sum_\sigma \rho^{|\sigma|} c_\sigma \mathcal{Z}_\sigma$$

Thus U_ρ and T_ρ acts identically on multi-linear polynomials over Gaussian sequences of independent ensembles.

Analogous to the discrete setting we say that $f : \mathbb{R}^n \to \Delta_q$ is balanced if $\mathbf{E}[f(X)] = \frac{1}{q}\mathbf{1}$ for $X \sim \mathrm{N}(0,1)$.

The following lemma shows for any fuzzy partition a non-fuzzy partition with almost the same expectation and noise stability (as measured in 1 and 2) can be created.

LEMMA 2. *Fix* $\rho \in \left[-\frac{1}{k-1}, 1\right]$ *and* $q_0 \le q$. *Suppose* $X_1, \dots, X_k \sim \mathrm{N}(0, I_n)$ *and* $\mathbf{Cov}(X_i, X_j) = \rho I_n$ *for* $i \neq j$. *Then, for any* $\epsilon > 0$ *and* $f : \mathbb{R}^n \to \Delta_q$, *there exists a* $g : \mathbb{R}^n \to E_q$ *such that*

$$\sum_{i=1}^{q} |\mathbf{E}\, g_i(X_1) - \mathbf{E}\, f_i(X_1)| \le k\epsilon \tag{9}$$

and

$$\left|\mathbf{E} \sum_{i=1}^{q_0} \prod_{j=1}^{k} g_i(X_j) - \mathbf{E} \sum_{i=1}^{q_0} \prod_{j=1}^{k} f_i(X_j)\right| \le \epsilon \tag{10}$$

The proof can be found in Appendix 2.A.

We also need a simple result that states that almost balanced functions cannot be much more stable than balanced functions:

LEMMA 3. *Fix* $\rho \in [-\frac{1}{k-1}, 1]$ *and* $q_0 \le q$. *Suppose* $X_1, \dots, X_k \sim \mathrm{N}(0, I_n)$ *are jointly normal with* $\mathbf{Cov}(X_i, X_j) = \rho I_n$ *for* $i \neq j$. *Let* $f : \mathbb{R}^n \to E_q$ *with* $\mathbf{E}\, f(X_1) = \mu$, *where*

$$\sum_{i=1}^{q} \left|\mu_i - \frac{1}{q}\right| = \delta$$

Then, there exists a balanced $g : \mathbb{R}^n \to E_q$ *such that*

$$\left|\mathbf{E} \sum_{i=1}^{q_0} \prod_{j=1}^{k} f_i(X_j) - \mathbf{E} \sum_{i=1}^{q_0} \prod_{j=1}^{k} g_i(X_j)\right| \le k\frac{\delta}{2}$$

Proof. Since the density function is continuous we can easily find a balanced g such that $P(f(X_j) = g(X_j)) = \frac{\delta}{2}$, hence the result follows by the union bound. □

2.3 Invariance Principle

Let $f : \prod_{i=1}^n \Omega_i \to \mathbb{R}$ be a function on a finite product probability space and express it as a multilinear polynomial $Q(\mathcal{X})$ over an independent sequence of orthonormal ensembles as in (5). The invariance principle of [15] shows that if Q has low degree and each coordinate has small influence then the distribution of $Q(\mathcal{X})$ does not change by much if we replace the variables $X_{i,j}$ with independent standard Gaussians $Z_{i,j}$.

In [13] the invariance principle was extended to the case of vector-valued functions $f = (f_1, \ldots, f_k)$ where $f_j : \prod_{i=1}^n \Omega_i \to \mathbb{R}$ for each j.

THEOREM 5. *([13], Theorem 4.1 and 3.16) Let $(\prod_{i=1}^n \Omega_i, \prod_{i=1}^n \mu_i)$ be a finite product probability space, $\alpha > 0$ the minimum probability of any atom in any μ_i and $\mathcal{X} = (\mathcal{X}_1, \ldots, \mathcal{X}_n)$ an independent sequence of orthonormal ensembles such that $\mathcal{X}_i$ is a basis for functions $\Omega_i \to \mathbb{R}$. Let Q be a k-dimensional multilinear polynomial such that $\mathbf{Var}\, Q_j(\mathcal{X}) \leq 1$, $\deg Q_j \leq d$ and $\mathrm{Inf}_i\, Q_j(\mathcal{X}) \leq \tau$. Finally, let $\Psi : \mathbb{R}^k \to \mathbb{R}$ be a $\mathcal{C}^3$ function with $|\Psi^{(\mathbf{r})}| \leq B$ for $|\mathbf{r}| = 3$. Then,*

$$|\mathbf{E}\, \Psi(Q(\mathcal{X})) - \mathbf{E}\, \Psi(Q(\mathcal{Z}))| \leq 2dBk^3 \left(8/\sqrt{\alpha}\right)^d \sqrt{\tau} = \mathcal{O}(\sqrt{\tau})$$

where $\mathcal{Z}$ is an independent sequence of standard Gaussian ensembles.

As suggested in [13, Corollary 4.3], since neither $\mathbf{Var}\, Q_j(\mathcal{X})$, $\deg Q_j$ or $\mathrm{Inf}_i\, Q_j$ depend on whether the ensembles are orthonormal, we can simply replace the orthonormal requirement by a matching covariance structure requirement.

DEFINITION 12. *We say that two independent sequences of ensembles $\mathcal{X} = (\mathcal{X}_1, \ldots, \mathcal{X}_n)$ and $\mathcal{Y} = (\mathcal{Y}_1, \ldots, \mathcal{Y}_n)$ have a matching covariance structure if for all i, $|\mathcal{X}_i| = |\mathcal{Y}_i|$ and $\mathbf{E}[\mathcal{X}_i^t \mathcal{X}_i] = \mathbf{E}[\mathcal{Y}_i^t \mathcal{Y}_i]$.*

THEOREM 6. *Let $\mathcal{X} = (\mathcal{X}_1, \ldots, \mathcal{X}_n)$ be an independent sequence of ensembles, such that $\mathbf{P}(\mathcal{X}_i = x) \geq \alpha > 0$. Let Q be a k-dimensional multilinear* *polynomial such that* $\mathbf{Var}\, Q_j(\mathcal{X}) \leq 1$, $\deg Q_j \leq d$ *and* $\mathrm{Inf}_i\, Q_j(\mathcal{X}) \leq \tau$. *Finally, let* $\Psi : \mathbb{R}^k \to \mathbb{R}$ *be a* $\mathcal{C}^3$ *function with* $|\Psi^{(\mathbf{r})}| \leq B$ *for* $|\mathbf{r}| = 3$. *Then,*

$$|\mathbf{E}\,\Psi(Q(\mathcal{X})) - \mathbf{E}\,\Psi(Q(\mathcal{Z}))| \leq 2dBk^3 \left(8/\sqrt{\alpha}\right)^d \sqrt{\tau} = \mathcal{O}(\sqrt{\tau})$$

where $\mathcal{Z}$ *is an independent sequence of Gaussian ensembles with the same covariance structure as* $\mathcal{X}$.

Proof. For each i, let Ω_i be the σ-algebra generated by the variables in $\mathcal{X}_i$. Since $\alpha > 0$, Ω_i is finite, hence we can find an orthonormal ensemble $\mathcal{X}'_i$ which is a basis for $\Omega_i \to \mathbb{R}$ and a linear transformation A_i such that $\mathcal{X}_i = \mathcal{X}'_i A_i$. Let $\mathcal{Z}'$ be any standard Gaussian ensemble and $\mathcal{Z}_i = \mathcal{Z}'_i A_i$. Then $\mathcal{Z}$ has the same covariance structure as $\mathcal{X}$. Let Q' be the multilinear polynomial defined by $Q'(\mathcal{X}') = Q(\mathcal{X}'_1 A_1, \ldots, \mathcal{X}'_n A_n)$. The result then follows by applying Theorem 5 to $Q'(\mathcal{X}')$ while noting that it has the same variances, degrees and influences as $Q(\mathcal{X})$. □

For our applications we will need a version of Theorem 6 for functions Ψ which are not $\mathcal{C}^3$ functions. Instead we will assume that Ψ is Lipschitz continuous with Lipschitz constant A, i.e. $|\Psi(x) - \Psi(y)| \leq A||x - y||_2$.

THEOREM 7. *Let* $\mathcal{X} = (\mathcal{X}_1, \ldots, \mathcal{X}_n)$ *be an independent sequence of ensembles, such that* $\mathbf{P}(\mathcal{X}_i = x) \geq \alpha > 0$. *Let* Q *be a k-dimensional multilinear polynomial such that* $\mathbf{Var}\, Q_j(\mathcal{X}) \leq 1$, $\deg Q_j \leq d$ *and* $\mathrm{Inf}_i\, Q_j(\mathcal{X}) \leq \tau$. *Finally, let* $\Psi : \mathbb{R}^k \to \mathbb{R}$ *be Lipschitz continuous with Lipschitz constant* A. *Then,*

$$|\mathbf{E}\,\Psi(Q(\mathcal{X})) - \mathbf{E}\,\Psi(Q(\mathcal{Z}))| \leq 4Ak \left(dB_{3,k} \left(8/\sqrt{\alpha}\right)^d \sqrt{\tau}\right)^{1/3} = \mathcal{O}(\tau^{1/6})$$

where $\mathcal{Z}$ *is an independent sequence of Gaussian ensembles with the same covariance structure as* $\mathcal{X}$ *and* $B_{3,k}$ *are universal constants.*

To prove Theorem 7 we need the following lemma which assures that Lipschitz continuous functions can be approximated well by $\mathbb{C}^3$ functions.

LEMMA 4. *Suppose* $\Psi : \mathbb{R}^k \to \mathbb{R}$ *is Lipschitz continuous, i.e.* $|\Psi(x) - \Psi(y)| \leq A||x - y||_2$ *for some constant* $A > 0$. *Then, for all* $\lambda > 0$ *there exists a* $\mathcal{C}^\infty$ *function* $\Psi_\lambda : \mathbb{R}^k \to \mathbb{R}$ *such that* $\forall x \in \mathbb{R}^k$ *and* $\forall \mathbf{r} : |\mathbf{r}| = r \geq 1$,

1. $|\Psi(x) - \Psi_\lambda(x)| \leq A\lambda$

2. $|\Psi_\lambda^{(\mathbf{r})}(x)| \leq \frac{AB_{r,k}}{\lambda^{r-1}}$

where $B_{r,k}$ are universal constants.

Proof. Let μ denote the Lebesgue measure on $\mathbb{R}^k$ and let $\phi : \mathbb{R}^k \to \mathbb{R}$ be the k-dimensional bump function defined by

$$\phi(x) = \begin{cases} Ce^{-\frac{1}{1-||x||_2^2}} & \text{if } ||x||_2 < 1 \\ 0 & \text{else} \end{cases}$$

where the constant C is chosen so that $\int_{x\in\mathbb{R}^k} \phi(x)\mu(dx) = 1$. It is well-known that $\phi(x)$ is $\mathcal{C}^\infty$ with bounded derivatives, hence there exist constants $B_r < \infty$ such that $|\phi^{(\mathbf{r})}(x)| \leq B_r$.

For $\lambda > 0$, let $\phi_\lambda(x) = \frac{1}{\lambda^k}\phi(\frac{x}{\lambda})$. Then $\int_{||x||_2\leq\lambda} \phi_\lambda(x)\mu(dx) = 1$ and $|\phi_\lambda^{(\mathbf{r})}(x)| \leq \frac{B_r}{\lambda^{k+r}}$. Let $\Psi_\lambda = \Psi * \phi_\lambda$, i.e.

$$\Psi_\lambda(x) = \int_{||x-t||_2\leq\lambda} \phi_\lambda(x-t)\Psi(t)\mu(dt)$$

By the mean value theorem, $\Psi_\lambda(x) = \Psi(t)$, for some $t : ||x-t||_2 \leq \lambda$. But $|\Psi(t) - \Psi(x)| \leq A||x-t||_2 \leq A\lambda$, which proves 1.
Without loss of generality we may assume that $\mathbf{r} = \mathbf{e}_1 + \mathbf{r}_2$, where $\mathbf{e}_1 = (1, 0, \ldots, 0)^t$ is the first unit vector. Since Ψ is bounded on $||x-t||_2 \leq \lambda$, Ψ_λ is $\mathcal{C}^\infty$ and for any s,

$$\Psi_\lambda^{(\mathbf{s})}(x) = \int_{||x-t||_2\leq\lambda} \phi_\lambda^{(\mathbf{s})}(x-t)\Psi(t)\mu(dt)$$

Thus we may write

$$
\begin{aligned}
\left|\Psi_\lambda^{(\mathbf{r})}(x)\right| &= \left|\frac{\delta}{\delta x_1}\int_{||x-t||_2\le\lambda}\phi_\lambda^{(\mathbf{r_2})}(x-t)\Psi(t)\mu(dt)\right| \\
&= \left|\frac{\delta}{\delta x_1}\int_{||t||_2\le\lambda}\phi_\lambda^{(\mathbf{r_2})}(t)\Psi(x-t)\mu(dt)\right| \\
&= \left|\lim_{h\to 0}\int_{||t||_2\le\lambda}\phi_\lambda^{(\mathbf{r_2})}(t)\frac{(\Psi(x+h\mathbf{e}_1-t)-\Psi(x-t))}{h}\mu(dt)\right| \\
&= \lim_{h\to 0}\left|\int_{||t||_2\le\lambda}\phi_\lambda^{(\mathbf{r_2})}(t)\frac{(\Psi(x+h\mathbf{e}_1-t)-\Psi(x-t))}{h}\mu(dt)\right| \\
&\le \lim_{h\to 0}\int_{||t||_2\le\lambda}\left|\phi_\lambda^{(\mathbf{r_2})}(t)\right|\left|\frac{(\Psi(x+h\mathbf{e}_1-t)-\Psi(x-t))}{h}\right|\mu(dt) \\
&\le \frac{B_{r-1}}{\lambda^{k+r-1}}A(2\lambda)^k = \frac{B_{r-1}}{\lambda^{r-1}}A2^k
\end{aligned}
$$

□

Proof of Theorem 7. Let Ψ_λ be the approximation given by Lemma 4. Then,

$$
\begin{aligned}
|\mathbf{E}\,\Psi(Q(\mathcal{X}))-\mathbf{E}\,\Psi(Q(\mathcal{Z}))| &\le |\mathbf{E}\,\Psi_\lambda(Q(\mathcal{X}))-\mathbf{E}\,\Psi_\lambda(Q(\mathcal{Z}))|+2A\lambda \le \\
&\le \frac{2A\epsilon}{\lambda^2}+2A\lambda\text{ , where }\epsilon = dB_{3,k}k^3\left(8/\sqrt{\alpha}\right)^d\sqrt{\tau}
\end{aligned}
$$

where we have used Theorem 6. Picking $\lambda=\epsilon^{1/3}$ gives the result. □

Our final version of the invariance principle replaces the bounded degree requirement with a smoothness requirement which can be achieved by applying the Bonami-Beckner operator $T_{1-\gamma}$ on $Q(\mathcal{X})$ for some small $\gamma>0$. Later we will use Lemma 1 to show that this smoothing is essentially harmless for our applications.

THEOREM 8. *Let $\mathcal{X}=(\mathcal{X}_1,\dots,\mathcal{X}_n)$ be an independent sequence of ensembles, such that $\mathbf{P}(\mathcal{X}_i = x)\ge\alpha>0$. Fix $\gamma,\tau\in(0,1)$ and let Q be a k-dimensional multilinear polynomial such that $\mathbf{Var}\,Q_j(\mathcal{X})\le 1$, $\mathbf{Var}\,Q_j^{>d}\le(1-\gamma)^{2d}$ and $\mathrm{Inf}_i\,Q_j^{\le d}(\mathcal{X})\le\tau$, where $d=\frac{1}{18}\log\frac{1}{\tau}/\log\frac{1}{\alpha}$. Finally, let $\Psi:\mathbb{R}^k\to\mathbb{R}$ be Lipschitz continuous with Lipschitz constant A. Then,*

$$
|\mathbf{E}\,\Psi(Q(\mathcal{X}))-\mathbf{E}\,\Psi(Q(\mathcal{Z}))|\le C_kA\tau^{\frac{\gamma}{18}/\log\frac{1}{\alpha}}
$$

where $\mathcal{Z}$ is an independent sequence of Gaussian ensembles with the same covariance structure as $\mathcal{X}$ and C_k is a constant depending only on k.

To prove Theorem 8 we need following easy lemma which bounds the effect of small deviations on Lipschitz continuous functions.

LEMMA 5. *Suppose $\Psi : \mathbb{R}^k \to \mathbb{R}$ is Lipschitz continuous, i.e. $|\Psi(x) - \Psi(y)| \le A||x-y||_2$ for some constant $A > 0$. Then,*

$$|\mathbf{E}\,\Psi(X+\xi) - \mathbf{E}\,\Psi(X)| \le A\left(\sum_{j=1}^{k} \mathbf{E}\,\xi_j^2\right)^{1/2}$$

Proof.

$$\begin{aligned} |\,\mathbf{E}\,\Psi(X+\xi) - \mathbf{E}\,\Psi(X)| &\le \mathbf{E}\,|\Psi(X+\xi) - \Psi(X)| \le \mathbf{E}\,A||\xi||_2 = \\ &= A\,\mathbf{E}\left(\sum_{j=1}^{k} \xi_j^2\right)^{1/2} \le A\left(\sum_{j=1}^{k} \mathbf{E}\,\xi_j^2\right)^{1/2} \end{aligned}$$

□

Proof of Theorem 8. The proof is by truncation of Q at degree $d = \frac{1}{18}\log\frac{1}{\tau}/\log\frac{1}{\alpha}$. Without loss of generality we may assume that $\alpha \le \frac{1}{2}$ (else, all random variables are constants and the result is trivial). By noting that Lemma 5 and Theorem 7 hold for all positive real values on d we have

$$\begin{aligned} &|\mathbf{E}\,\Psi(Q(\mathcal{X})) - \mathbf{E}\,\Psi(Q(\mathcal{Z}))| \le \\ &\le \left|\mathbf{E}\,\Psi(Q^{\le d}(\mathcal{X})) - \mathbf{E}\,\Psi(Q^{\le d}(\mathcal{Z}))\right| + A\sqrt{k}(1-\gamma)^d \le \\ &\le 4AkB_{3,k}^{1/3}\left(16/\sqrt{\alpha}\right)^{d/3}\tau^{1/6} + A\sqrt{k}e^{-\gamma d} \end{aligned}$$

The result now follows by noting that

$$e^{-\gamma d} = \tau^{\frac{\gamma}{18}/\log\frac{1}{\alpha}}$$

and

$$\begin{aligned} \left(16/\sqrt{\alpha}\right)^{d/3}\tau^{1/6} &= e^{\frac{d}{6}\log\frac{256}{\alpha}}\tau^{1/6} = \tau^{-\frac{1}{6\cdot 18}\log\frac{256}{\alpha}/\log\frac{1}{\alpha}}\tau^{1/6} \le \\ &\le \tau^{-\frac{1}{12}}\tau^{1/6} = \tau^{\frac{1}{12}} \le \tau^{\frac{\gamma}{18}/\log\frac{1}{\alpha}} \end{aligned}$$

where both inequalities uses that $\alpha \le \frac{1}{2}$ and the last also that $\gamma \le 1$. □

2.4 Application I: Plurality is Stablest

Here we show that Conjecture 2 implies the Plurality is Stablest conjecture (Theorem 2).

We start by showing an unconditional result that asserts that the most stable low low-degree influence functions are essentially determined by most stable partition of Gaussian space into q parts of equal measure.

DEFINITION 13. *For $\rho \in [-1, 1]$ and $q \geq 1$, let*

$$\Lambda_q^-(\rho) = \lim_{n\to\infty} \inf_{A_1,\ldots,A_q} \mathbf{P}((X, Y) \in A_1^2 \cup \ldots \cup A_q^2) \tag{11}$$

and

$$\Lambda_q^+(\rho) = \lim_{n\to\infty} \sup_{A_1,\ldots,A_q} \mathbf{P}((X, Y) \in A_1^2 \cup \ldots \cup A_q^2) \tag{12}$$

where $X, Y \in \mathrm{N}(0, I_n)$, $\mathbf{Cov}(X, Y) = \rho I_n$ and the inf and sup is over all balanced partitions $A_1, \ldots, A_q$ of $\mathbb{R}^n$.

Note that the limits in (11) and (12) exist since they are limits of bounded functions which are monotone in n (we can always ignore any number of dimensions while specifying the partitions).

THEOREM 9. *For any $q \geq 2$, $\rho \in [-\frac{1}{q-1}, 1]$ and $\epsilon > 0$ there exist d and $\tau > 0$ such that if $f : [q]^n \to \Delta_q$ is a balanced function with $\mathrm{Inf}_i^{\leq d}(f_j) \leq \tau$, $\forall i, j$, then*

$$\Lambda_q^-(\rho) - \epsilon \leq \mathbb{S}_\rho(f) \leq \Lambda_q^+(\rho) + \epsilon$$

DEFINITION 14. *For $q \geq 2$, let let $f_{\Delta_q} : \mathbb{R}^q \to \Delta_q$ denote the function which maps x to the point in Δ_q which is closest to x.*

Proof of Theorem 9. The result is trivial for $\rho = 1$ so assume $\rho \in [-\frac{1}{q-1}, 1)$. Let $(\Omega \times \Lambda, \mu)$, with the ρ-correlated measure $\mu(\omega, \lambda) = \rho 1_{\{\lambda=\omega\}} \frac{1}{q} + (1-\rho)\frac{1}{q^2}$ be our base space and let $(\omega, \lambda) \in [q]^n \times [q]^n$ be drawn from μ^n.

Fix an orthonormal basis $\mathcal{V}(x) = \{V_0(x) = 1, V_1(x), \ldots, V_{q-1}(x)\}$ for functions $[q] \to \mathbb{R}$ and construct two sequences of orthonormal ensembles $\mathcal{X} = \{\mathcal{X}_1, \ldots, \mathcal{X}_n\}$ and $\mathcal{Y} = \{\mathcal{Y}_1, \ldots, \mathcal{Y}_n\}$ for functions $\Omega \to \mathbb{R}$ and $\Lambda \to \mathbb{R}$ by letting $X_{i,j}(\omega_i) = V_j(\omega_i)$ and $Y_{i,j}(\lambda_i) = V_j(\lambda_i)$. Note that this means that

$$\mathbf{Cov}(X_{i_1,j_1}, Y_{i_2,j_2}) = \begin{cases} \rho & \text{if } i_1 = i_2 \text{ and } j_1 = j_2 > 0 \\ 0 & \text{else} \end{cases}$$

Expressing $f(x) = f(x_1, \ldots, x_n)$ as a q-dimensional multi-linear polynomial $Q(\mathcal{V}(x_1), \ldots, \mathcal{V}(x_n))$ we get

$$\mathbb{S}_\rho(f) = \sum_{i=1}^{q} \mathbf{E}[f_j(\omega)f_j(\lambda)] = \sum_{i=1}^{q} \mathbf{E}[Q_j(\mathcal{X})Q_j(\mathcal{Y})] \tag{13}$$

Let $\widetilde{Q} = T_{1-\gamma}Q$ be a slightly smoothed version of Q. By Lemma 1 we can find a $\gamma(\epsilon, \rho, q) > 0$ s.t.

$$\left|\mathbf{E}[Q_j(\mathcal{X})Q_j(\mathcal{Y})] - \mathbf{E}[\widetilde{Q}_j(\mathcal{X})\widetilde{Q}_j(\mathcal{Y})]\right| \leq \frac{\epsilon}{2q} \tag{14}$$

Since $Q(\mathcal{X})$ has range Δ_q, the same holds for $\widetilde{Q}(\mathcal{X})$. Hence,

$$f_{\Delta_q}\widetilde{Q}(\mathcal{X}) = \widetilde{Q}(\mathcal{X}) \tag{15}$$

(and similarly for $\mathcal{Y}$). We are now ready to apply the invariance principle (Theorem 8) using $\Psi(x, y) = \langle f_{\Delta_q}(x), f_{\Delta_q}(y)\rangle$. To see that $\Psi(x, y)$ is Lipschitz continuous, first note that the convexity of Δ_q implies

$$||f_{\Delta_q}(x) - f_{\Delta_q}(x')||_2 \leq ||x - x'||_2 \tag{16}$$

Also, for $u, v \in \mathbb{R}^q$,

$$\begin{aligned} |\langle u, v\rangle - \langle u', v'\rangle| &\leq |\langle u, v\rangle - \langle u', v\rangle| + |\langle u', v\rangle - \langle u', v'\rangle| \\ &\leq ||u - u'||_2||v||_2 + ||v - v'||_2||u'||_2 \end{aligned} \tag{17}$$

Combining (17) and (16) we get

$$\begin{aligned} |\Psi(x, y) - \Psi(x', y')| &\leq ||x - x'||_2||f_{\Delta_q}(y)||_2 + ||y - y'||_2||f_{\Delta_q}(x')||_2 \\ &\leq ||x - x'||_2 + ||y - y'||_2 \leq \sqrt{2}||(x, y) - (x', y')||_2 \end{aligned}$$

Hence Theorem 8 implies that for some $\tau > 0$ small enough,

$$\left|\mathbf{E}[\langle f_{\Delta_q}\widetilde{Q}(\mathcal{X}), f_{\Delta_q}\widetilde{Q}(\mathcal{Y})\rangle] - \mathbf{E}[\langle f_{\Delta_q}\widetilde{Q}(\mathcal{G}), f_{\Delta_q}\widetilde{Q}(\mathcal{H})\rangle]\right| \leq \frac{\epsilon}{4q} \tag{18}$$

where $\mathcal{G}$ and $\mathcal{H}$ are two Gaussian sequences of orthonormal ensembles with

$$\mathbf{Cov}(G_{i_1,j_1}, H_{i_2,j_2}) = \begin{cases} \rho & \text{if } i_1 = i_2, j_1 = j_2 > 0 \\ 0 & \text{else} \end{cases}$$

$f_{\Delta_q}\widetilde{Q}$ applied to $\mathcal{G}$ or $\mathcal{H}$ can be thought of as a function $\mathbb{R}^{n(q-1)} \to \Delta_q$ creating a fuzzy partition of the $n(q-1)$-dimensional Gaussian space. This partition is not balanced, but letting $\mu = \mathbf{E}\, f_{\Delta_q}\widetilde{Q}(\mathcal{G})$ and applying Theorem 8 again, using $\Psi(x) = f_{\Delta_q,j}(x)$ which by (16) is Lipschitz continuous with $A = 1 \leq \sqrt{2}$, we can bound the total variation distance by

$$\sum_{j=1}^{q} \left|\mu_j - \frac{1}{q}\right| = \sum_{j=1}^{q} |\,\mathbf{E}\, f_{\Delta_q,j}\widetilde{Q}(\mathcal{G}) - \mathbf{E}\, f_{\Delta_q,j}\widetilde{Q}(\mathcal{X})| \leq q\frac{\epsilon}{4q} = \frac{\epsilon}{4}$$

By Lemma 2 and 3 there exists a balanced function $g : \mathbb{R}^{n(q-1)} \to E_q$ such that

$$\left|\mathbf{E}[\langle f_{\Delta_q}\widetilde{Q}(\mathcal{G}), f_{\Delta_q}\widetilde{Q}(\mathcal{H})\rangle] - \mathbf{E}[\langle g(\mathcal{G}), g(\mathcal{H})\rangle]\right| \leq \frac{\epsilon}{4} \tag{19}$$

But any such g partitions $\mathbb{R}^{n(q-1)}$ into q parts of equal Gaussian measure $\frac{1}{q}$, hence

$$\Lambda_q^-(\rho) \leq \mathbf{E}[\langle g(\mathcal{G}), g(\mathcal{H})\rangle] \leq \Lambda_q^+(\rho) \tag{20}$$

Combining equations (13), (14), (15), (18), (19) and (20) gives the desired result. □

In order to prove Theorem 2 we first show that the limit of the noise stability of $\mathrm{PLUR}\, n, q$ corresponds to the right hand side of (2).

LEMMA 6. *Fix* $\rho \in [-\frac{1}{q-1}, 1]$ *and* $q \geq 3$. *Let* $X, Y \sim \mathrm{N}(0, I_{q-1})$ *and* $\mathbf{Cov}(X, Y) = \rho I_{q-1}$. *Then*

$$\lim_{n\to\infty} \mathbb{S}_\rho(\mathrm{PLUR}_{n,q}) = \mathbf{P}((X,Y) \in S_{q,1}^2 \cup \cdots \cup S_{q,q}^2)$$

Proof. By definition 1,

$$\mathbb{S}_\rho(\mathrm{PLUR}_{n,q}) = \mathbf{E}\langle \mathrm{PLUR}_{n,q}(\omega), \mathrm{PLUR}_{n,q}(\lambda)\rangle$$

where ω, λ are uniform on $[q]^n$ and satisfy (3). Represent each ω_i and λ_i by a q-dimensional unit vector $U_i = \mathbf{e}_{\omega_i}$ and $V_i = \mathbf{e}_{\lambda_i}$ and let $U = \frac{\sqrt{q}}{n}\sum_{i=1}^n U_i$ and $V = \frac{\sqrt{q}}{n}\sum_{i=1}^n V_i$. Then, conditioning on having no ties which will happen with probability 1 as $n \to \infty$, we have

$$\begin{aligned}\mathbb{S}_\rho(\mathrm{PLUR}_{n,q}) &= \mathbf{P}((U,V) \in (S'_{q,1})^2 \cup \cdots \cup (S'_{q,q})^2) \\ &= \mathbf{P}((MU, MV) \in (S_{q,1})^2 \cup \cdots \cup (S_{q,q})^2)\end{aligned}$$

The expectations and covariances of MU and MV are,

$$\begin{aligned}
\mathbf{E}[MV] = \mathbf{E}[MU] &= M\frac{1}{\sqrt{q}}\mathbf{1} = \mathbf{0} \\
\mathbf{E}[MV(MV)^T] = \mathbf{E}[MU(MU)^T] &= MI_qM^T = I_{q-1} \\
\mathbf{E}[MU(MV)^T] &= M\left(\rho I_q - (1-\rho)\frac{1}{q}\mathbf{1}_q\mathbf{1}_q^T\right)M^T = \rho I_{q-1}
\end{aligned}$$

Hence, by the central limit theorem, (MU, MV) converges to a normal distribution with the same parameters as (X, Y). Thus,

$$\lim_{n\to\infty} \mathbf{P}((MU, MV) \in (S_{q,1})^2 \cup \cdots \cup (S_{q,q})^2) = \mathbf{P}((X,Y) \in S_{q,1}^2 \cup \cdots \cup S_{q,q}^2)$$

□

Proof of Theorem 2. By Theorem 9 and Lemma 6 we only need to observe that Conjecture 2 is equivalent to

$$\Lambda_q^+(\rho) = \mathbf{P}((X,Y) \in S_{q,1}^2 \cup \cdots \cup S_{q,q}^2) \text{ for } \rho \in [0,1]$$

and

$$\Lambda_q^-(\rho) = \mathbf{P}((X,Y) \in S_{q,1}^2 \cup \cdots \cup S_{q,q}^2) \text{ for } \rho \in [-1,0]$$

where $X, Y \sim \mathrm{N}(0, I_n)$ are jointly normal with $\mathbf{Cov}(X,Y) = \rho I_n$. That we may replace the low low-degree influence requirement with the simpler low influence requirement follows by noting that

$$\mathrm{Inf}_i^{\leq d}(f_j) \leq \mathrm{Inf}_i(f_j)$$

□

2.5 Application II: Condorcet Voting

Here we show that Conjecture 1 implies that majority maximizes the probability of having a unique best candidate in Condorcet voting assuming (Theorem 4).

Remember that we have n voters selecting a linear order $\sigma_i \in S(k)$ uniformly at random and let

$$X_i^{a>b} = \begin{cases} 1 & \text{if } \sigma_i(a) > \sigma_i(b) \\ -1 & \text{else} \end{cases} \text{, for } i \in [n] \text{ and } a, b \in [k].$$

By considering the 6 possible linear orders of three candidates its easy to see that for any distinct $a, b, c \in [k]$ we have

$$\mathbf{E}[X_i^{a>b}] = 0\ , \ \mathbf{Var}\, X_i^{a>b} = 1 \text{ and } \mathbf{Cov}[X_i^{a>b}, X_i^{a>c}] = \frac{1}{3}$$

First we will show that the limit of the probability of having a unique best candidate using the majority function corresponds to the right hand side of (1).

LEMMA 7. *Let $X_2, \ldots, X_k \sim \mathrm{N}(0, I_n)$ be jointly normal with $\mathbf{Cov}(X_i, X_j) = \frac{1}{3} I_n$ for $i \neq j$. Then*

$$\lim_{n\to\infty} \mathbf{P}[\mathrm{UniqueBest}_k(\mathrm{MAJ}_n)] = \mathbf{P}(\forall i : X_i \in H)$$

Proof. Let $Y_j = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i^{1>j}$. By definition 3,

$$\mathbf{P}[\mathrm{UniqueBest}_k(\mathrm{MAJ}_n)] = \mathbf{P}(Y_2 \geq 0, \ldots, Y_k \geq 0)$$

But, $\mathbf{E}[Y_j] = 0$, $\mathbf{E}[Y_j^2] = 1$ and $\mathbf{Cov}[Y_i, Y_j] = \frac{1}{3}$ for $i \neq j$. Thus, by the central limit theorem, $(Y_2, \ldots, Y_k) \xrightarrow{\mathcal{D}} (X_2, \ldots, X_k)$ and the result follows. $\square$

Proof of Theorem 4. Clearly, any candidate has the same probability of being the unique best candidate. So it's enough to show that the probability that the first candidate is the unique best is maximized by majority,

$$\mathbf{P}[\mathrm{UniqueBest}_k(f, 1)] \leq \lim_{n\to\infty} \mathbf{P}[\mathrm{UniqueBest}_k(\mathrm{MAJ}_n, 1)] + \frac{\epsilon}{k} \tag{21}$$

Form orthonormal ensembles

$$\mathcal{X}_i^{1>j} = (1, X_i^{1>j})\ \text{, for } i \in [n] \text{ and } 2 \leq j \leq k$$

and independent sequences of orthonormal ensembles

$$\mathcal{X}^j = (\mathcal{X}_1^{1>j}, \ldots, \mathcal{X}_n^{1>j})$$

Thus $\mathcal{X}^j$ is a basis for real-valued functions on all voters' preferences between candidates 1 and j and we can compute the (unique) multilinear polynomial Q such that

$$f(X_1^{1>j}, \ldots, X_n^{1>j}) = Q(\mathcal{X}^j)$$

Hence we may write,

$$\mathbf{P}[\mathrm{UniqueBest}_k(f,1)] = \mathbf{E}\prod_{j=2}^{k} Q(\mathcal{X}^j) \tag{22}$$

Let $\widetilde{Q} = T_{1-\gamma}Q$ be a slightly smoothed version of Q, and let

$$\rho(k) = \rho(\Sigma\left(\mathcal{X}_i^{1>2}\right), \ldots, \Sigma\left(\mathcal{X}_i^{1>k}\right); \mathbf{P})$$

where $\Sigma(X)$ denotes the σ-algebra generated by X. Clearly, $\rho(k) < 1$, so by Lemma 1 we can find a $\gamma(\epsilon, k, n) > 0$ such that

$$|\,\mathbf{E}\prod_{j=2}^{k} Q(\mathcal{X}^j) - \mathbf{E}\prod_{j=2}^{k} \widetilde{Q}(\mathcal{X}^j)| \le \frac{\epsilon}{2k} \tag{23}$$

Let $f_{[0,1]}(x) = \max(0, \min(1,x))$. Since Q has range $[0,1]$, the same holds for $\widetilde{Q}$. Hence, for all j,

$$\widetilde{Q}(\mathcal{X}^j) = f_{[0,1]}\widetilde{Q}(\mathcal{X}^j) \tag{24}$$

We now apply the invariance principle (Theorem 8) using $\Psi(x_2, \ldots, x_k) = \prod_{j=2}^{k} f_{[0,1]}(x_j)$ which by convexity of $[0,1]^{k-1}$ is Lipschitz continuous with Lipschitz constant $A = 1$. Thus, by Theorem 8, there exist some $\gamma > 0$ such that,

$$\left|\mathbf{E}\prod_{j=2}^{k} f_{[0,1]}\widetilde{Q}(\mathcal{X}^j) - \mathbf{E}\prod_{j=2}^{k} f_{[0,1]}\widetilde{Q}(\mathcal{G}^j)\right| \le \frac{\epsilon}{4k^2} \tag{25}$$

where $\mathcal{G}^j = (\mathcal{G}_1^{1>j}, \ldots, \mathcal{G}_n^{1>j})$, and $\mathcal{G}_i^{1>j} = (1, G_i^{1>j})$ are Gaussian sequences of orthonormal ensembles with

$$\mathbf{Cov}[G_{i_1}^{1>j_1}, G_{i_2}^{1>j_2}] = \mathbf{Cov}[X_{i_1}^{1>j_1}, X_{i_2}^{1>j_2}] = \begin{cases} 1 & \text{if } i_1 = i_2, j_1 = j_2 \\ \frac{1}{3} & \text{if } i_1 = i_2, j_1 \neq j_2 \\ 0 & \text{else} \end{cases}$$

Now $(f_{[0,1]}\widetilde{Q}, 1 - f_{[0,1]}\widetilde{Q})$ applied to $\mathcal{G}^j$ can be thought of as a function $\mathbb{R}^n \to \Delta_2$ creating a fuzzy partition of the n-dimensional Gaussian space which is almost balanced. Let $\mu = \mathbf{E}\, f_{[0,1]}\widetilde{Q}(\mathcal{G}^j)$. Then a second application of Theorem 8 with $\Psi(x) = f_{[0,1]}(x)$ gives

$$\left|\mu - \frac{1}{2}\right| \leq \frac{\epsilon}{4k^2}$$

By Lemma 2 and 3, there exists a balanced function $g : \mathbb{R}^n \to E_2$.

$$\mathbf{E}\prod_{j=2}^{k} f_{[0,1]}\widetilde{Q}(\mathcal{G}^j) \leq \mathbf{E}\prod_{j=2}^{k} g_1(\mathcal{G}^j) + \frac{\epsilon}{4k} \tag{26}$$

But any such g partitions $\mathbb{R}^n$ into 2 parts of equal Gaussian measure $\frac{1}{2}$, so Conjecture 1 and Lemma 7 implies

$$\mathbf{E}\prod_{j=2}^{k} g_1(\mathcal{G}^j) \leq \lim_{n\to\infty} \mathbf{P}[\mathrm{UniqueBest}_k(\mathrm{MAJ}_n)] \tag{27}$$

Combining equations (22), (23), (24), (25), (26) and (27) gives (21) as needed. □

2.6 Approximability of MAX-q-CUT

In this section we will show that if we assume the Unique Games Conjecture, then the optimal approximability of MAX-q-CUT is directly related to the most stable partition of Gaussian space into q parts of equal measure as described in Conjecture 2.

2.6.1 The Unique Games Conjecture

The Unique Games Conjecture (UGC) was introduced by Khot in [10] as a possible way of proving inapproximability results for 2-CSPs and has since been used to prove optimal inapproximability results for many important problems, such as ....

The conjecture asserts the hardness of approximating the *Unique Label Cover* problem within any constant.

DEFINITION 15.
An instance of the Unique Label Cover *problem,* $\mathcal{L}(V, W, E, M, \{\sigma_{v,w}\}_{(v,w)\in E})$, *consists of a bipartite graph* $(V \cup W, E)$ *with a permutation* $\sigma_{v,w} : [M] \to [M]$ *associated with every edge* $(v, w) \in E \subseteq V \times W$. *A labeling* $l : V \cup W \to [M]$ *is said to satisfy an edge* (v, w) *if*

$$\sigma_{(v,w)}(l(w)) = l(v)$$

The value of a labeling l, $\mathrm{VAL}_l(\mathcal{L})$, *is the fraction of edges satisfied by* l *and the value of* $\mathcal{L}$ *is the maximal fraction of edges satisfied by any labeling,*

$$\mathrm{VAL}(\mathcal{L}) = \max_l \mathrm{VAL}_l(\mathcal{L})$$

CONJECTURE 4. ***The Unique Games Conjecture.*** *For any* $\eta, \gamma > 0$ *there exists a* $M = M(\eta, \gamma)$ *such that it is NP-hard to distinguish instances* $\mathcal{L}$ *of the Unique Label Cover problem with label set size* M *having* $\mathrm{VAL}(\mathcal{L}) \geq 1 - \eta$ *from those having* $\mathrm{VAL}(\mathcal{L}) \leq \gamma$.

Next, we will show that for any $\epsilon > 0$, MAX-q-CUT can be approximated within $\alpha_q - \epsilon$ in polynomial time while it is UG-hard to approximate it within $\beta_q + \epsilon$.

2.6.2 Optimal approximability constants

DEFINITION 16. *For* $q \geq 1$, *let*

$$\alpha_q = \lim_{n\to\infty} \sup_{A_1,\ldots,A_q} \inf_{-\frac{1}{q-1}\leq\rho\leq 1} \frac{q}{q-1} \frac{1 - \mathbf{P}((X, Y) \in A_1^2 \cup \cdots \cup A_q^2)}{1 - \rho} \tag{28}$$

and

$$\beta_q = \lim_{n\to\infty} \inf_{-\frac{1}{q-1}\leq\rho\leq 1} \sup_{A_1,\ldots,A_q} \frac{q}{q-1} \frac{1 - \mathbf{P}((X, Y) \in A_1^2 \cup \cdots \cup A_q^2)}{1 - \rho} \tag{29}$$

where $X, Y \in \mathrm{N}(0, I_n)$, $\mathbf{Cov}(X, Y) = \rho I_n$ *and the supremum is over all disjoint* $A_1, \ldots, A_q \subseteq \mathbb{R}^n$ *with* $\mathbf{P}(X \in A_j) = \frac{1}{q}$, $\forall j$.

Note that the limit in (28) and (29) exist since they are limits of bounded functions increasing with n (we can always ignore any number of dimensions while specifying the partition).

We now show that $\alpha_q = \beta_q$ assuming Conjecture 2. To do this, we first show that we can restrict attention to non-positive values of ρ and for all such values the standard simplex partition is optimal.

LEMMA 8. *Assume Conjecture 2. Then, with the notation of Definition 16, we have for all* $\rho \in [0,1]$,

$$\frac{q}{q-1}\frac{1-\mathbf{P}((X,Y)\in A_1^2\cup\cdots\cup A_q^2)}{1-\rho}\geq 1$$

with equality for $\rho = 0$.

Proof. By Conjecture 2 and Lemma 6,

$$\mathbf{P}((X,Y)\in A_1^2\cup\cdots\cup A_q^2)\leq \lim_{n\to\infty}\mathbb{S}_\rho(\mathrm{PLUR}_{n,q})$$

On the other hand, by (8) and (6)

$$\mathbb{S}_\rho(\mathrm{PLUR}_{n,q}) = \sum_\sigma \rho^{|\sigma|}||c_\sigma||_2^2 \leq ||\,\mathbf{E}[\mathrm{PLUR}_{n,q}]||_2^2 + \rho\,\mathbf{Var}[\mathrm{PLUR}_{n,q}] =$$
$$= \frac{1}{q}+\frac{q-1}{q}\rho$$

Hence,

$$\mathbf{P}((X,Y)\in A_1^2\cup\cdots\cup A_q^2)\leq \frac{1}{q}+\frac{q-1}{q}\rho$$

which holds with equality for $\rho = 0$. □

THEOREM 10. *Assume Conjecture 2. Then* $\alpha_q = \beta_q$.

Proof. By Lemma 8 the infimums in the definition of α_q and β_q are obtained for $-\frac{1}{q-1} \leq \rho \leq 0$. The result now follows from the fact that for ρ in this range, the least stable partition in Conjecture 2 does not depend on ρ. □

We now proceed to present the approximation algorithm and the inapproximability argument which together implies Theorem 3.

2.6.3 An approximation algorithm

The approximation algorithm presented here is a generalization of the algorithm presented in [4] allowing for an arbitrary partition to be used when rounding the relaxed solution. The algorithm in [4] corresponds exactly to using the simplex partition of Conjecture 2, which (as we will see) is optimal if Conjecture 2 is true.

Let $\widetilde{E}_q = \sqrt{\frac{q}{q-1}} M E_q = \left\{ \sqrt{\frac{q}{q-1}} M x | x \in E_q \right\}$ be the extreme points of the projected simplex scaled so that each point has unit norm:

LEMMA 9. *For* $\widetilde{x}, \widetilde{y} \in \widetilde{E}_q$,

$$\widetilde{x} \cdot \widetilde{y} = \begin{cases} 1 & \textit{if } \widetilde{x} = \widetilde{y} \\ -\frac{1}{q-1} & \textit{if } \widetilde{x} \neq \widetilde{y} \end{cases} \tag{30}$$

Proof. Let x and y be the preimages of $\widetilde{x}$ and $\widetilde{y}$, i.e. $\widetilde{x} = \sqrt{\frac{q}{q-1}} M x$ and similarly for y. Then,

$$\widetilde{x} \cdot \widetilde{y} = \frac{q}{q-1} \left(x - \frac{\mathbf{1}}{q} \right) \cdot \left(y - \frac{\mathbf{1}}{q} \right) = \frac{q}{q-1} \left(x \cdot y - \frac{\mathbf{1}}{q} \right) = \begin{cases} 1 & \text{if } \widetilde{x} = \widetilde{y} \\ -\frac{1}{q-1} & \text{if } \widetilde{x} \neq \widetilde{y} \end{cases}$$

□

Labeling the vertices with vectors from $\widetilde{E}_q$ instead of numbers from $[q]$, we can write the value of a MAX-q-CUT instance $\mathcal{M}_q(V, E, w)$ as the following discrete optimization problem:

$$\begin{aligned} \text{VAL}(\mathcal{M}_q) = \quad & \max \quad \tfrac{q-1}{q} \textstyle\sum_{(u,v) \in E} w_{(u,v)} (1 - l_u \cdot l_v) \\ & \text{subject to} \quad l_u \in \widetilde{E}_q \text{ , } \forall u \in V \end{aligned}$$

To obtain the SDP relaxation we allow the vectors to be arbitrary points on the unit sphere while adding the constraint $z_u \cdot z_v \geq -\frac{1}{q-1}$ which by (30) holds for vectors in $\widetilde{E}_q$,

$$\begin{aligned} \text{SDP-VAL}(\mathcal{M}_q) := \quad & \max \quad \tfrac{q-1}{q} \textstyle\sum_{(u,v) \in E} w_{(u,v)} (1 - z_u \cdot z_v) \\ & \text{subject to} \quad z_u \in \mathbb{R}^n, \forall u \in V \\ & \qquad\qquad\quad z_u \cdot z_u = 1, \forall u \in V \\ & \qquad\qquad\quad z_u \cdot z_v \geq -\tfrac{1}{q-1}, \forall u, v \in V \end{aligned}$$

where $n = |V|$ denotes the number of vertices.

The rounding applied to the solution of SDP-VAL is parametrized by an integer m, a partition $\mathcal{A} = \{A_1, \dots, A_q\}$ of $\mathbb{R}^m$ and an error constant $\delta > 0$,

Approximation algorithm $\mathcal{R}(m, \mathcal{A}, \delta)$**.**

1. *Compute an almost optimal solution* $(z_u)_{u \in V}$ *to SDP-VAL*$(\mathcal{M}_q)$ *using semidefinite programming. This will achieve a value of SDP-VAL*$(\mathcal{M}_q) - \delta$*.*

2. *Pick a projection matrix* $T : \mathbb{R}^{m \times n}$*, by letting* T_{ij} *be i.i.d.* $\mathrm{N}(0, 1)$.

3. *For each* $u \in V$*, let* $l(u) = i$ *iff* $Tz_u \in A_i$*.*

Let R-VAL$(\mathcal{M}_q) = \mathrm{VAL}_l(\mathcal{M}_q)$ be the value of the rounded labeling. Then, the expected approximation ratio is:

$$\begin{aligned}
\frac{\mathbf{E}[\text{R-VAL}(\mathcal{M}_q)]}{\mathrm{VAL}(\mathcal{M}_q)} \geq \frac{\mathbf{E}[\text{R-VAL}(\mathcal{M}_q)]}{\text{SDP-VAL}(\mathcal{M}_q) + \delta} &= \\
&= \frac{\sum_{(u,v) \in E} w_{(u,v)} \, \mathbf{P}\left(l(u) \neq l(v)\right)}{\frac{q-1}{q} \sum_{(u,v) \in E} w_{(u,v)} (1 - z_u \cdot z_v) + \delta} \geq \\
&\geq \frac{q}{q-1} \inf_{\substack{z_u, z_v \in S^{n-1} \\ z_u \cdot z_v \geq -\frac{1}{q-1}}} \frac{1 - \mathbf{P}((Tz_u, Tz_v) \in A_1^2 \cup \dots \cup A_q^2)}{1 - z_u \cdot z_v + \delta}
\end{aligned}$$

But, $Tz_u, Tz_v \in \mathrm{N}(0, I_m)$ and $\mathbf{Cov}(Tz_u, Tz_v) = (z_u \cdot z_v) I_m$, so by picking m large enough and $A_1, \dots, A_q$ so that the limit in (28) is almost achieved (bar, say δ), and then picking $\delta = \delta(\epsilon)$ small enough, we get an approximation ratio of $\alpha_q - \epsilon$, for any $\epsilon > 0$. We have proved the following result

THEOREM 11. *For any* $\epsilon > 0$ *there exists a polynomial time algorithm that approximates MAX-q-CUT within* $\alpha_q - \epsilon$*.*

2.6.4 Inapproximability results

We will now prove that MAX-q-CUT is UG-hard to approximate within any factor greater than β_q. To do so, we present a reduction from the Unique Label

Cover problem to MAX-q-CUT following the same outline as the corresponding reduction for MAX-CUT given in [11]. The reduction is based on a *Probabilistically Checkable Proof* (PCP) whose proof Π consists of the function tables of $\{f_w\}_{w \in W}$, where $f_w : [q]^M \to [q]$ is expected to be the *long code* of w's label $l(w)$, i.e. $f_w(x) = x_{l(w)}$. In order to be able to reduce the PCP to MAX-q-CUT, the PCP verifier $\mathcal{V}_\rho$ is designed to use an acceptance predicate which reads two random function values from the proof and accepts iff they differ. Thus, a MAX-q-CUT instance $\mathcal{M}_q$ can be created from the PCP by letting the vertices be the function values that can be read by $\mathcal{V}_\rho$, the edges the pairs of function values that are compared, and the weights the probability of that comparison being made by $\mathcal{V}_\rho$. The verifier is parametrized by $\rho \in [-\frac{1}{q-1}, 1]$.

PCP Verifier $\mathcal{V}_\rho$.

1. *Pick $v \in V$ at random and two of its neighbors w, w' at random.*

2. *Pick $x \in [q]^M$ at random.*

3. *Pick $y \in [q]^M$ to be a ρ-correlated copy of x, i.e. each y_i is independently selected using the conditional distribution*

$$\mu(y_i|x_i) = \rho 1_{\{y_i = x_i\}} + (1-\rho)\frac{1}{q}$$

4. *Accept if $f_w P_{\sigma_{v,w}}(x) \neq f_{w'} P_{\sigma_{v,w'}}(y)$, where $P_\sigma : [q]^M \to [q]^M$ denotes the function*

$$P_\sigma(x_1, \ldots, x_M) = (x_{\sigma(1)}, \ldots, x_{\sigma(M)})$$

Using a result from [12] we can assume that the graph is regular on the V side so that (v, w), and similarly (v, w'), picked by $\mathcal{V}_\rho$ corresponds to a an edge selected uniformly at random. By folding, we may also assume that the functions f_w are balanced, i.e. by using the functions $f'_w(x_1, \ldots, x_M) := f_w(0, x_2 - x_1, \ldots, x_M - x_1) + x_1$ (where addition and subtraction in $[q]$ is performed modulo q) instead of the original functions f_w. Note that folding does not change any function which is a long code, but still forces any function to become balanced.

LEMMA 10. **(Completeness).** *Fix $\rho \in [-\frac{1}{q-1}, 1)$. Then, for any Unique Label Cover problem $\mathcal{L}$ with $\mathrm{VAL}(\mathcal{L}) \geq 1-\eta$ there exists a proof Π such that*

$$\mathbf{P}[\mathcal{V}_\rho \textit{ accepts } \Pi] \geq (1-2\eta)\frac{q-1}{q}(1-\rho)$$

Proof. Let l be the optimal assignment for $\mathcal{L}$ and f_w be the long code of $l(w)$, i.e.

$$f_w(x) = x_{l(w)}$$

With probability at least $1-2\eta$, both edges (v,w) and (v,w') are satisfied by l. In this case,

$$f_w P_{\sigma_{v,w}}(x) = x_{\sigma_{v,w}(l(w))} = x_{l(v)} \text{ and } f_{w'} P_{\sigma_{v,w'}}(y) = y_{l(v)}$$

and $\mathcal{V}_\rho$ accepts with probability

$$\mathbf{P}[x_{l(v)} \neq y_{l(v)}] = 1 - \left(\rho + \frac{1-\rho}{q}\right) = \frac{q-1}{q}(1-\rho)$$

□

LEMMA 11. **(Soundness).** *Fix $\rho \in [-\frac{1}{q-1}, 1]$ and $\epsilon > 0$. Then, there exists a $\gamma = \gamma(q,\rho,\epsilon) > 0$ such that for any Unique Label Cover problem $\mathcal{L}$ with $\mathrm{VAL}(\mathcal{L}) \leq \gamma$ and any proof Π,*

$$\mathbf{P}[\mathcal{V}_\rho \textit{ accepts } \Pi] \leq 1 - \Lambda_q^-(\rho) + \epsilon \tag{31}$$

Proof. For $w \in W$, let $\tilde{f}_w : [q]^M \to E_q$ defined by

$$\tilde{f}_w(x) = \mathbf{e}_{f_w(x)}$$

map the value of f_w onto one of q unit vectors, and for $v \in V$, let $g_v : [q]^M \to \Delta_q$ be defined by

$$g_v(x) = \mathop{\mathbf{E}}_w[\tilde{f}_w P_{\sigma_{v,w}}(x)]$$

where the expectation is over a random neighbor w of v. Then,

$$\begin{aligned}\mathbf{P}[\mathcal{V}_\rho \text{ accepts } \Pi] &= \mathop{\mathbf{E}}_{v,w,w',x,y}[1 - \langle \tilde{f}_w P_{\sigma_{v,w}}(x), \tilde{f}_{w'} P_{\sigma_{v,w'}}(y)\rangle] = \\ &= 1 - \mathop{\mathbf{E}}_{v,x,y}[\langle g_v(x), g_v(y)\rangle] = 1 - \mathop{\mathbf{E}}_v \mathbb{S}_\rho(g_v)\end{aligned}$$

Now suppose Π is a proof such that (31) is not satisfied, i.e,

$$\mathop{\mathbf{E}}_{v} \mathbb{S}_\rho(g_v) < \Lambda_q^-(\rho) - \epsilon \tag{32}$$

We need to show that this implies $\mathrm{VAL}(\mathcal{L}) > \gamma$. To do so it is enough to create a random labeling l such that

$$\mathop{\mathbf{E}}_{l}[\mathrm{VAL}_l(\mathcal{L})] > \gamma$$

Let $V_{\text{good}} = \{v \in V | \mathbb{S}_\rho(g_v) \leq \Lambda_q^-(\rho) - \frac{\epsilon}{2}\}$. Since $\mathbb{S}_\rho(g_v) \geq 0$, (32) implies that $|V_{\text{good}}| \geq \frac{\epsilon}{2}|V|$. Further, for $v \in V_{\text{good}}$, Theorem 9 implies that $\max_i \mathrm{Inf}_i^{\leq d} g_v \geq \tau$, for some d and $\tau > 0$ depending only on q,ρ and ϵ.

The assignment l is created as follows:

1. For $v \in V$, let $l(v) = i$, where i maximizes $\mathrm{Inf}_i^{\leq d} g_v$ (ties broken arbitrarily)

2. For $w \in W$, let $l(w) = i$ with probability proportional to $\mathrm{Inf}_i^{\leq d} \tilde{f}_w$.

Since (7) holds for vector-valued functions, this means that

$$\mathop{\mathbf{P}}_{l}(l(w) = i) \geq \frac{\mathrm{Inf}_i^{\leq d} \tilde{f}_w}{qd}$$

For $v \in V_{\text{good}}$,

$$\begin{aligned}
\tau &\leq \mathrm{Inf}_{l(v)}^{\leq d} g_v = \mathrm{Inf}_{l(v)}^{\leq d} \mathop{\mathbf{E}}_{w}[\tilde{f}_w P_{\sigma_{v,w}}(x)] \leq \mathop{\mathbf{E}}_{w} \mathrm{Inf}_{l(v)}^{\leq d} \tilde{f}_w P_{\sigma_{v,w}}(x) = \\
&= \mathop{\mathbf{E}}_{w} \mathrm{Inf}_{\sigma_{v,w}^{-1}(l(v))}^{\leq d} \tilde{f}_w(x) \leq qd \mathop{\mathbf{P}}_{w,l}[l(w) = \sigma_{v,w}^{-1}(l(v))] = \\
&= qd \mathop{\mathbf{P}}_{w,l}[l \text{ satisfies } (v,w)]
\end{aligned}$$

where the second inequality follows from convexity of $\mathrm{Inf}_i^{\leq d}$. Hence,

$$\mathop{\mathbf{E}}_{l}[\mathrm{VAL}_l(\mathcal{L})] = \mathop{\mathbf{P}}_{l,v,w}(l \text{ satisfies } (v,w)) \geq \frac{\epsilon}{2} \cdot \frac{\tau}{qd}$$

Picking $\gamma = \frac{\epsilon}{4} \cdot \frac{\tau}{qd} > 0$ finishes the proof. $\square$

Together, the soundness and completeness lemmas implies the following inapproximability result for MAX-q-CUT:

THEOREM 12. *For any $\epsilon > 0$ it is UG-hard to approximate MAX-q-CUT within $\beta_q + \epsilon$.*

Proof. By Lemma 10 and 11 it is UG-hard to distinguish instances of MAX-q-CUT with value at least $(1-2\eta)\frac{q-1}{q}(1-\rho)$ from instances with value at most $1 - \Lambda_q^-(\rho) + \epsilon$ for any $\gamma, \epsilon > 0$. Thus, it is UG-hard to approximate MAX-q-CUT within

$$\frac{1 - \Lambda_q^-(\rho) + \epsilon}{(1-2\eta)\frac{q-1}{q}(1-\rho)} = \frac{q}{q-1}\frac{1-\Lambda_q^-(\rho)}{1-\rho} + \epsilon'$$

where $\epsilon' > 0$ can be made arbitrarily small by picking γ and ϵ small enough. Since this holds for any $\rho \in [-\frac{1}{q-1}, 1]$ the result follows. □

References

[1] C. Borell, *Geometric bounds on the ornstein-uhlenbeck velocity process*, Probability Theory and Related Fields **70** (1985), no. 1, 1–13.

[2] J. Corneli, I. Corwin, Y. Xu, S. Hurder, V. Sesum, E. Adams, D. Davis, M. Lee, R. Pettit, and N. Hoffman, *Double bubbles in gauss space and spheres*, Houston journal of mathematics **34** (2008), no. 1, 181–204.

[3] R. Dobrushin, P. Groeneboom, and M. Ledoux, *Lectures on probability theory and statistics*, Lecture Notes in Mathematics, vol. 1648, Springer-Verlag, Berlin, 1996, Lectures from the 24th Saint-Flour Summer School held July 7–23, 1994, edited by P. Bernard. MR MR1600892 (98g:60002)

[4] A. Frieze and M. Jerrum, *Improved approximation algorithms for MAX-k-CUT and MAX-BISECTION*, Integer Programming and Combinatorial Optimization (Egon Balas and Jens Clausen, eds.), vol. 920, Springer, 1995, pp. 1–13.

[5] M. Goemans and D. Williamson, *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming*, JACM **42** (1995), 1115–1145.

[6] J. Håstad, *Some optimal inapproximability results*, Proceedings of the 39th Annual ACM Symposium on Theory of Computing (STOC), May 1997, pp. 1–10.

[7] M. Hutchings, F. Morgan, M. RitorÃ©, and A. Ros, *Proof of the double bubble conjecture*, Annals of Mathematics **155** (2002), no. 2, 459–489.

[8] G. Kalai, *A Fourier-theoretic perspective on the Concordet paradox and Arrow's theorem*, Adv. in Appl. Math. **29** (2002), no. 3, 412–426.

[9] ———, *Social Indeterminacy*, Econometrica **72** (2004), 1565–1581.

[10] S. Khot, *On the power of unique 2-prover 1-round games*, STOC '02: Proceedings of the thirty-fourth annual ACM symposium on Theory of computing (New York, NY, USA), ACM, 2002, pp. 767–775.

[11] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell, *Optimal inapproximability results for max-cut and other 2-variable csps?*, SIAM J. Comput. **37** (2007), 319–357.

[12] S. Khot and O. Regev, *Vertex cover might be hard to approximate to within* $2-\varepsilon$, Proceedings of the 18th Annual IEEE Conference on Computational Complexity, IEEE, 2003, pp. 379–388.

[13] E. Mossel, *Gaussian bounds for noise correlation of functions*, Submitted, 2008.

[14] E. Mossel, R. O'Donnell, and K. Oleszkiewicz, *Noise stability of functions with low influences: invariance and optimality (extended abstract)*, 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2005), 23-25 October 2005, Pittsburgh, PA, USA, Proceedings, IEEE Computer Society, 2005, pp. 21–30.

[15] ———, *Noise stability of functions with low influences: invariance and optimality*, To appear in Ann. Math., 2008.

[16] Prasad Raghavendra, *Optimal Algorithms and Inapproximability Results For Every CSP?*, To appear in STOC, 2008.

Appendices

2.A Proof of Lemma 2

Proof. Assume first that $\rho \in \left(-\frac{1}{k-1}, 1\right)$ so that the normal distribution is non-degenerate. Discretize $\mathbb{R}^n$ with cubes $[0,\delta)^n$, i.e. write $\mathbb{R}^n = \delta\mathbb{Z}^n \times [0,\delta)^n$. where $\delta\mathbb{Z}^n$ denotes the n-dimensional integer lattice scaled by a factor δ.

Let $Z_{i,j} = \delta \left\lfloor \frac{X_{i,j}}{\delta} \right\rfloor$ so that Z_i denotes the cube X_i is in, and let $U_{i,j}$ be i.i.d. uniform on $[0,\delta]$, independent of $X_1, \dots X_k$.

Further let η be the density of $(X_1, \dots, X_k)$ and $\tilde{\eta}$ the density of $(Z_1 + U_1, \dots, Z_k + U_k)$. It is easy to see that

$$\tilde{\eta}(x) = \frac{1}{\delta^{nk}} \int_{[0,\delta)^{nk}} \eta(z_1 + u_1, \dots z_q + u_q) d(u_1, \dots u_q) \to \eta(x) \text{ as } \delta \to 0$$

since η is Lipschitz continuous. By dominated convergence, this implies that we can choose δ so that

$$\int_{\mathbb{R}^{nk}} |\eta(x) - \tilde{\eta}(x)|\, dx \leq \frac{\epsilon}{2}$$

Similar to Scheffés Lemma we have for any $h : \mathbb{R}^{nk} \to [0,1]$,

$$\left| \int_{\mathbb{R}^{nk}} h(x)\eta(x)dx - \int_{\mathbb{R}^{nk}} h(x)\tilde{\eta}(x)dx \right| \leq \int_{\mathbb{R}^{nk}} h(x) \left|\eta(x) - \tilde{\eta}(x)\right| dx \leq \frac{\epsilon}{2} \tag{33}$$

The non-fuzzy function g is constructed from f by transferring masses internally in each cube. More specifically, g is defined arbitrarily on each cube with the only restriction that

$$\mathbf{E}[g(Z_1 + U_1)|Z_1] = \mathbf{E}[f(Z_1 + U_1)|Z_1]$$

(For instance, if $\mathbf{E}[g(Z_1 + U_1)|Z_1 = z_1] = \mu$, then we may divide the cube $z_1 + [0, \delta)^n$ into q parts of conditional measure $\mu_1, \dots \mu_q$ and assign the value $e_1, \dots, e_q$ respectively to each part.) Thus,

$$\mathbf{E} \sum_{i=1}^{q_0} \prod_{j=1}^{k} f_i(Z_j + U_j) = \mathbf{E} \sum_{i=1}^{q_0} \prod_{j=1}^{k} \mathbf{E}[f_i(Z_j + U_j)|Z_j] =$$
$$= \mathbf{E} \sum_{i=1}^{q_0} \prod_{j=1}^{k} \mathbf{E}[g_i(Z_j + U_j)|Z_j] = \mathbf{E} \sum_{i=1}^{q_0} \prod_{j=1}^{k} g_i(Z_j + U_j)$$

Applying (33) twice gives (10). Similarly

$$\mathbf{E}\, f_i(Z_1 + U_1) = \mathbf{E}[\mathbf{E}[f_i(Z_1 + U_1)|Z_1]] = \mathbf{E}[\mathbf{E}[g_i(Z_1 + U_1)|Z_1]] = \mathbf{E}\, g_i(Z_1 + U_1)$$

and two more applications of (33) gives $|\,\mathbf{E}\, f_i(X_1) - \mathbf{E}\, g_i(X_1)| \leq \epsilon$ and (9) follows.

The two degenerate cases can be handled in a similar way by using a density with respect to a lower dimensional Lebesgue measure. □

PAPER II

K-wise Gaussian Noise Stability

Marcus Isaksson

3
PAPER II

ABSTRACT
We introduce k-wise Gaussian noise stability and show that among subsets of $\mathbb{R}^n$ of fixed measure, half-spaces maximizes this stability. This extends a Gaussian isoperimetric inequality by Borell which proved the result for $k = 2$.

3.1 Introduction

DEFINITION 1. *For $k \geq 1$, $\rho \in [-\frac{1}{k-1}, 1]$, and $A \in \mathbb{B}(\mathbb{R}^n)$, the k-wise Gaussian noise stability of A at ρ is*

$$\mathbb{S}_\rho^{(k)}(A) = \mathbf{P}(X_1 \in A, \ldots, X_k \in A)$$

where $X_1, \ldots, X_k \sim \mathrm{N}(0, I_n)$ are jointly normal with $\mathbf{Cov}(X_i, X_j) = \rho I_n$ for $i \neq j$.
We also let $\mu = \mathbb{S}_\rho^{(1)}$ denote the standard Gaussian measure on $\mathbb{R}^n$.

We prove that among sets of fixed measure, half spaces are most stable under k-wise Gaussian noise for $\rho \geq 0$.

THEOREM 1. *For any* $k \geq 1$, $\rho \in [0,1]$ *and* $A \in \mathbb{B}(\mathbb{R}^n)$,

$$\mathbb{S}_\rho^{(k)}(A) \leq \mathbb{S}_\rho^{(k)}(H)$$

where $H = \{x \in \mathbb{R}^n | x_1 \leq a\}$ *for* a *chosen so that* $\mu(H) = \mu(A)$.

Note that the case $k = 1$ is of course trivial. Further, the case $k = 2$ was proved by Borell [1].

3.2 Spherical Case

We start by defining the corresponding problem on $\mathrm{S}^{m-1}(\sqrt{m})$, the $m-1$-dimensional sphere in $\mathbb{R}^m$ with radius $\sqrt{m}$.

DEFINITION 2. *For* $k \geq 1$, $\rho \in [-\frac{1}{k-1}, 1]$, *and* $A \in \mathbb{B}(\mathbb{R}^m)$, *the k-wise spherical noise stability of* A *at* ρ *is*

$$\widetilde{\mathbb{S}}_\rho^{(k)}(A) = \mathbf{P}(\widetilde{X}_1 \in A, \ldots, \widetilde{X}_k \in A)$$

where $X_1, \ldots, X_k \sim \mathrm{N}(0, I_m)$ *are jointly normal with* $\mathbf{Cov}(X_i, X_j) = \rho I_m$ *for* $i \neq j$ *and* $\widetilde{X}_i = \frac{\sqrt{m}}{||X_i||_2} X_i$.
We also let $\tilde{\mu} = \widetilde{\mathbb{S}}_\rho^{(1)}$ *denote the uniform measure on the sphere* $\mathrm{S}^{m-1}(\sqrt{m})$.

THEOREM 2. *For any* $k \geq 1$, $\rho \in [0,1]$ *and* $A \in \mathbb{B}(\mathbb{R}^m)$,

$$\widetilde{\mathbb{S}}_\rho^{(k)}(A) \leq \widetilde{\mathbb{S}}_\rho^{(k)}(H)$$

where $H = \{x \in \mathbb{R}^m | x_1 \leq a\}$ *for* a *chosen so that* $\tilde{\mu}(H) = \tilde{\mu}(A)$.

Our reduction from the spherical result to the Gaussian result is based on Poincarés observation that Gaussian measure on $\mathbb{R}^n$ is obtained by projection of the uniform measure on $\mathrm{S}^{m-1}(\sqrt{m})$ onto $\mathbb{R}^n$, as $m \to \infty$. The convergence is strong enough for the measure of any Borel set to converge:

LEMMA 1. *For any* $A \in \mathbb{B}(\mathbb{R}^n)$,

$$\tilde{\mu}(A \times \mathbb{R}^{m-n}) \to \mu(A) \text{ as } m \to \infty$$

Proof. This is mentioned with references in [4]. See also [2] □

LEMMA 2. *For any $k \geq 1$, $\rho \in [0,1]$ and $A \in \mathbb{B}(\mathbb{R}^n)$,*

$$\widetilde{\mathbb{S}}_\rho^{(k)}(A \times \mathbb{R}^{m-n}) \to \mathbb{S}_\rho^{(k)}(A) \text{ as } m \to \infty$$

Proof. To do.

Suppose $X_1, \ldots, X_k \sim \mathrm{N}(0, I_m)$ and $\mathbf{Cov}(X_i, X_j) = \rho I_m$ for $i \neq j$. Let $\widetilde{X}_i = \frac{\sqrt{m}}{||X_i||_2} X_i$.

Let $Y_i = (X_{i,1}, \ldots, X_{i,n})$ denote the restriction of X_i to the first n coordinates (think of $m \geq n$), and similarly $\widetilde{Y}_i = (\widetilde{X}_{i,1}, \ldots, \widetilde{X}_{i,n})$. Then it's easy to see that

$$(\widetilde{Y}_1, \ldots, \widetilde{Y}_k) \xrightarrow{\mathcal{D}} (Y_1, \ldots, Y_k) \text{ as } m \to \infty$$

But to show that

$$\mathbf{P}(\widetilde{Y}_1 \in A, \ldots \widetilde{Y}_k \in A) \to \mathbf{P}(Y_1 \in A, \ldots Y_k \in A)$$

for all A, we need a stronger convergence (convergence of densities is enough). □

LEMMA 3. *Theorem 2 ⇒ Theorem 1.*

Proof. Fix $A \in \mathbb{B}(\mathbb{R}^n)$ and let $H = \{x \in \mathbb{R}^n | x_1 \leq a\}$ where $\mu(H) = \mu(A)$. We need to show that

$$\mathbb{S}_\rho^{(k)}(A) \leq \mathbb{S}_\rho^{(k)}(H)$$

For each $m \geq n$, let $H_m = \{x \in \mathbb{R}^m | x_1 \leq a_m\}$ where a_m is chosen so that $\tilde{\mu}(H_m) = \tilde{\mu}(A \times \mathbb{R}^{m-n})$. Note that, by Lemma 1, as $m \to \infty$,

$$\tilde{\mu}(H_m) = \tilde{\mu}(A \times \mathbb{R}^{m-n}) \to \mu(A) = \mu(H) \leftarrow \tilde{\mu}(H \times \mathbb{R}^{m-n})$$

Since both H and H_m are half-spaces defined by the first coordinate, this implies

$$\tilde{\mu}(H_m \setminus H \times \mathbb{R}^{m-n}) \to 0$$

which by the union bound implies

$$\widetilde{\mathbb{S}}_\rho^{(k)}(H_m) - \widetilde{\mathbb{S}}_\rho^{(k)}(H \times \mathbb{R}^{m-n}) \leq k\tilde{\mu}(H_m \setminus H \times \mathbb{R}^{m-n}) \to 0 \qquad (1)$$

Now, by Theorem 2,

$$\widetilde{\mathbb{S}}_{\rho}^{(k)}(A \times \mathbb{R}^{m-n}) \leq \widetilde{\mathbb{S}}_{\rho}^{(k)}(H_m)$$

Taking limits (as $m \to \infty$) and using Lemma 2 and (1) we have

$$\mathbb{S}_{\rho}^{(k)}(A) \leq \mathbb{S}_{\rho}^{(k)}(H)$$

as needed. $\square$

3.3 Symmetrization

The main tool in the proof is the following symmetrization operation which given a hyperplane tries to push every point in A from one pre-determined side of the hyperplane to it's reflection point on the other side of the hyperplane as long as that point is not already in A. Defining the symmetrization process in terms of set operations we have,

DEFINITION 3. *For any $A \in \mathbb{B}(\mathbb{R}^n)$ and $h \in \mathbb{R}^n \setminus \{0\}$, we define the* two-point symmetrization *of A with respect to h by*

$$R_h(A) = \left([A \cap \sigma(A)] \cap H_+^C\right) \ \cup \ \left([A \cup \sigma(A)] \cap H_+\right) \tag{2}$$

where $H_+ = \{x \in \mathbb{R}^n | x \cdot h > 0\}$ and $\sigma(A)$ denotes the reflection of A with respect to the hyperplane $H_0 = \{x \in \mathbb{R}^n | x \cdot h = 0\}$.

As we will show, both Gaussian and spherical k-wise noise stability increases under this symmetrization for $\rho \geq 0$.

LEMMA 4. *For any $k \geq 1$, $\rho \in [0, 1)$, $A \in \mathbb{B}(\mathbb{R}^n)$ and $h \in \mathbb{R}^n \setminus \{0\}$,*

$$\mathbb{S}_{\rho}^{(k)}(R_h(A)) \geq \mathbb{S}_{\rho}^{(k)}(A)$$

Proof. By spherical symmetry it is enough to prove the result for $h = e_1$, the first unit vector.

Let $X_1, \ldots, X_k$ be as in Definition 1 and let X be the matrix of random variables with row vectors $X_{i.} := X_i$. Then the column vectors $X_{.j} = (X_{1,j}, \ldots, X_{k,j})$ are independent $\mathrm{N}(0, \Sigma)$ vectors where $\Sigma_{i,j} = \rho + (1 - \rho)\delta_{ij}$.

It is easy to verify that the inverse of Σ is given by $\left(\Sigma^{-1}\right)_{i,j} = -a + b\delta_{ij}$, where $b = \frac{1}{1-\rho}$ and $a = \frac{\rho}{(1-\rho)(1+\rho(k-1))} \geq 0$ for $\rho \geq 0$. Hence,

$$\mathbb{S}_\rho^{(k)}(A) = \int_{\mathbb{R}^{n\times k}} \prod_{i=1}^{k} 1_{\{x_{i.}\in A\}} \prod_{j=1}^{n} f(x_{.j})dx$$

where $f : \mathbb{R}^k \to \mathbb{R}$ is the density of a $\mathrm{N}(0,\Sigma)$ variable, i.e.

$$f(y) = \frac{1}{\sqrt{(2\pi)^k|\Sigma|}} e^{-\frac{1}{2}(b\sum_{i=1}^k y_i^2 - a\sum_{i,j} y_i y_j)}$$

Splitting the integral depending on the signs $s_1,\ldots,s_k$ of $x_{.1}$, we may write

$$\mathbb{S}_\rho^{(k)}(A) = \int_{(\mathbb{R}^+\times\mathbb{R}^{n-1})^k} f_A(x) \prod_{j=2}^{n} f(x_{.j})dx$$

where

$$f_A(x) = \sum_{s\in\{-1,1\}^k} \prod_{i=1}^{k} 1_{\{(s_i x_{i,1}, x_{i,2},\ldots,x_{i,n})\in A\}} f(s_1 x_{1,1},\ldots,s_k x_{k,1})$$

Clearly, it is enough to show that $f_A(x)$ does not decrease under symmetrization of A, for any $x \in (\mathbb{R}^+ \times \mathbb{R}^{n-1})^k$. Fix such an x. By reordering the vectors $x_1,\ldots,x_k$, we may assume without loss of generality that for the first l vectors both x_i and $\sigma(x_i)$ are in A, while for the rest exactly one is (we can ignore cases where for some i neither x_i nor $\sigma(x_i)$ are in A since such cases do not contribute to $f_A(x)$ nor $f_{R_h}(A)$). Thus we can assume that

$$\{x_i, \sigma(x_i)\} \subseteq A, 1 \leq i \leq l$$

while

$$\begin{cases} x_i \in A \text{ and } \sigma(x_i) \notin A \text{ if } t_i = 1 \\ x_i \notin A \text{ and } \sigma(x_i) \in A \text{ if } t_i = -1 \end{cases}, \; l < i \leq k$$

for some $l \in [k]$ and $t_{l+1},\ldots,t_k \in \{-1,1\}$. Note that symmetrization of A corresponds to setting all t_i's to 1. Now,

$$\begin{aligned} f_A(x) &= \sum_{s\in\{-1,1\}^l} f(s_1 x_{1,1},\ldots,s_l x_{l,1}, t_{l+1}x_{l+1,1},\ldots,t_k x_{k,1}) = \\ &= \frac{1}{\sqrt{(2\pi)^k|\Sigma|}} \sum_{s\in\{-1,1\}^l} e^{-\frac{1}{2}(c_s+d_s)} \end{aligned}$$

where

$$c_s = b\sum_{i=1}^{k} x_{i,1}^2 - a\sum_{1\le i,j\le l} s_i s_j x_{i,1}x_{j,1} - a\sum_{l<i,j\le k} t_i t_j x_{i,1}x_{j,1}$$

and

$$d_s = -2a\sum_{1\le i\le l<j\le k} s_i t_j x_{i,1}x_{j,1} = -2a\sum_{1\le i\le l} s_i x_{i,1}\sum_{l<j\le k} t_j x_{j,1}$$

Pairing each s with $-s$ in (3) and noting that c_s is even in s while d_s is odd, we may write

$$\begin{aligned} f_A(x)\sqrt{(2\pi)^k|\Sigma|} &= \frac{1}{2}\sum_{s\in\{-1,1\}^l} e^{-\frac{1}{2}(c_s+d_s)} + e^{-\frac{1}{2}(c_{-s}+d_{-s})} \\ &= \sum_{s\in\{-1,1\}^l} e^{-\frac{1}{2}c_s}\cosh(-\frac{1}{2}d_s) \end{aligned}$$

The result now follows by noting that since $x_{.1}\ge 0$, setting all t_i's to 1 will decrease each c_s and increase the absolute value of each d_s, hence $f_A(x)$ will increase (unless all t_i's already are 1). $\square$

This symmetrization works just as well in the spherical case,

COROLLARY 1. *For any* $k\ge 1$, $\rho\in[0,1)$, $A\in\mathbb{B}(\mathbb{R}^m)$ *and* $h\in\mathbb{R}^m\setminus\{0\}$,

$$\widetilde{\mathbb{S}}_\rho^{(k)}(R_h(A)) \ge \widetilde{\mathbb{S}}_\rho^{(k)}(A)$$

Proof. Define $T:\mathbb{R}^m\to\mathbb{R}^m$ by $T(A) = \left\{x\in\mathbb{R}^m \,\middle|\, \frac{\sqrt{m}}{||x||_2}x\in A\right\}$. Then, using Lemma 4 and noting that T and R_h commute, we have

$$\widetilde{\mathbb{S}}_\rho^{(k)}(R_h(A)) = \mathbb{S}_\rho^{(k)}(T(R_h(A))) = \mathbb{S}_\rho^{(k)}(R_h(T(A))) \ge \mathbb{S}_\rho^{(k)}(T(A)) = \widetilde{\mathbb{S}}_\rho^{(k)}(A)$$

$\square$

3.4 Proof of Theorem 2

The proof of Theorem 2 given Lemma 4 is inspired by [3].

DEFINITION 4. *For $x, y \in \mathbb{R}^m$ and $A, B \subseteq \mathbb{R}^m$, let $d(x,y)$ denote the Euclidean distance between x and y, $d(x,A) = \inf_{y\in A} d(x,y)$ denote the distance from x to A and*

$$d_H(A,B) = \max\{\sup_{x\in A} d(x,B), \sup_{y\in B} d(y,A)\}$$

denote the Hausdorff distance between A and B.
Also, for $\epsilon > 0$, let

$$A_\epsilon = \{x \in \mathbb{R}^m | d(x,A) \leq \epsilon\}$$

DEFINITION 5. *Let $(\mathcal{C}^m, d_H)$ denote the metric space*

$$\mathcal{C}^m = \{C \in \mathbb{B}(\mathrm{S}^{m-1}(\sqrt{m})) | \, C \textit{ is closed } \}$$

equipped with the Hausdorff measure d_H.

Note that since $(\mathrm{S}^{m-1}(\sqrt{m}), d)$ is compact so is $(\mathcal{C}^m, d_H)$.

LEMMA 5. *For $B \in \mathcal{C}^m$, $\widetilde{\mathbb{S}}^{(k)}_\rho(B_\epsilon) \to \widetilde{\mathbb{S}}^{(k)}_\rho(B)$ as $\epsilon \to 0$.*

Proof. For $k = 1$, we only need to note that since B is closed, $\bigcap_{\epsilon>0}(B_\epsilon \setminus B) = \emptyset$, hence $\mu(B_\epsilon \setminus B) \to 0$ as $\epsilon \to 0$. By the union bound,

$$\widetilde{\mathbb{S}}^{(k)}_\rho(B_\epsilon) \geq \widetilde{\mathbb{S}}^{(k)}_\rho(B) \geq \widetilde{\mathbb{S}}^{(k)}_\rho(B_\epsilon) - k\mu(B_\epsilon \setminus B)$$

hence the result follows by letting $\epsilon \to 0$. □

LEMMA 6. *$\widetilde{\mathbb{S}}^{(k)}_\rho$ is upper semi-continuous on $(\mathcal{C}^m, d_H)$.*

Proof. Suppose B_n is a sequence in $\mathcal{C}^m$ such that $d_H(B_n, B) \to 0$. We need to show that $\widetilde{\mathbb{S}}^{(k)}_\rho(B) \geq \limsup \widetilde{\mathbb{S}}^{(k)}_\rho(B_n)$. But, for any $\epsilon > 0$, $B_\epsilon \supseteq \limsup B_n$, hence

$$\widetilde{\mathbb{S}}^{(k)}_\rho(B_\epsilon) \geq \widetilde{\mathbb{S}}^{(k)}_\rho(\limsup B_n) \geq \limsup \widetilde{\mathbb{S}}^{(k)}_\rho(B_n)$$

where the second inequality follows from the reverse Fatou Lemma. The result now follows from Lemma 5 by letting $\epsilon \to 0$. □

Proof of Theorem 2. Since $\tilde{\mu}$ is supported on $\mathrm{S}^{m-1}(\sqrt{m})$, we may assume $A \in \mathbb{B}(\mathrm{S}^{m-1}(\sqrt{m}))$ and let $H = \{x \in \mathrm{S}^{m-1}(\sqrt{m}) | x_1 \leq a\}$ where a is chosen so that $\tilde{\mu}(H) = \tilde{\mu}(A)$. We need to show that

$$\widetilde{\mathbb{S}}_\rho^{(k)}(A) \leq \widetilde{\mathbb{S}}_\rho^{(k)}(H)$$

Without loss of generality we may also assume that A is closed (else, by regularity of the uniform measure $\tilde{\mu}$, $\forall \epsilon > 0 : \exists$ closed $A' \subseteq A$ such that $\tilde{\mu}(A') \geq \tilde{\mu}(A) - \epsilon$, and hence $\widetilde{\mathbb{S}}_\rho^{(k)}(A') \geq \widetilde{\mathbb{S}}_\rho^{(k)}(A) - k\epsilon$, and the result follows from the result for closed sets by letting $\epsilon \to 0$).

Let $\mathcal{B} \subseteq \mathcal{C}^m$ be the set of all $B \in \mathcal{C}^m$ such that

$$\begin{array}{ll} i) & \tilde{\mu}(B) = \tilde{\mu}(A)\ (= \tilde{\mu}(H)) \\ ii) & \forall \epsilon > 0 : \tilde{\mu}(B_\epsilon) \leq \tilde{\mu}(A_\epsilon) \\ iii) & \widetilde{\mathbb{S}}_\rho^{(k)}(B) \geq \widetilde{\mathbb{S}}_\rho^{(k)}(A) \end{array}$$

Claim 1: $\mathcal{B}$ is closed in $(\mathcal{C}^m, d_H)$.

Proof: Suppose B_n is a sequence in $\mathcal{B}$ such that $d_H(B_n, B) \to 0$. We need to show that $B \in \mathcal{B}$. From Lemma 6, it follows that $\widetilde{\mathbb{S}}_\rho^{(k)}(B) \geq \widetilde{\mathbb{S}}_\rho^{(k)}(A)$ and $\tilde{\mu}(B) \geq \tilde{\mu}(A)$. Now fix $\epsilon > 0$. For all $\delta > 0$ we can pick $n = n(\delta)$ such that $(B_n)_{\epsilon+\delta} \supseteq B_\epsilon$. Hence,

$$\tilde{\mu}(B_\epsilon) \leq \tilde{\mu}((B_n)_{\epsilon+\delta}) \overset{B_n \in \mathcal{B}}{\leq} \tilde{\mu}(A_{\epsilon+\delta}) \xrightarrow{\text{Lem. 5}} \tilde{\mu}(A_\epsilon) \text{ as } \delta \to 0$$

Thus, $\tilde{\mu}(B_\epsilon) \leq \tilde{\mu}(A_\epsilon)$. Letting $\epsilon \to 0$ and using Lemma 5 we also get $\tilde{\mu}(B) \leq \tilde{\mu}(A)$.

Claim 2: $\mathcal{B}$ is closed under R_h.

Proof: Condition iii) was shown in Corollary 1. Condition i) follows from (2) by noting that

$$\tilde{\mu}(R_h(B)) = \tilde{\mu}(B \cap \sigma(B))\frac{1}{2} + \tilde{\mu}(B \cup \sigma(B))\frac{1}{2} = \frac{\tilde{\mu}(B) + \tilde{\mu}(\sigma(B))}{2} = \tilde{\mu}(B)$$

For condition ii) it is enough to see that $[R_h(A)]_\epsilon \subseteq R_h(A_\epsilon)$ for all $\epsilon > 0$. This can be seen by a simple case analysis.

Now, upper semi-continuity of $\tilde{\mu}$ implies upper semi-continuity of $B \to \tilde{\mu}(B \cap H)$ on $(\mathcal{C}^m, d_H)$. Hence, since $(\mathcal{C}^m, d_H)$ is compact and $\mathcal{B}$ is a non-empty (since $A \in \mathcal{B}$) closed subset, $\sup_{B \in \mathcal{B}} \tilde{\mu}(B \cap H)$ is achieved by some $B^* \in \mathcal{B}$.

Suppose first that, $\tilde{\mu}(B^* \cap H) < \tilde{\mu}(H)$. Then, by i), we must have

$$\tilde{\mu}(B^* \setminus H) = \tilde{\mu}(H \setminus B^*) > 0$$

Lebesgue's density theorem asserts that there are points $x \in B^* \setminus H$ and $y \in H \setminus B^*$ and a $\epsilon > 0$ such that say,

$$\begin{cases} \tilde{\mu}(\{x\}_\epsilon \cap B^* \setminus H) > 0.9\tilde{\mu}(\{x\}_\epsilon) \\ \tilde{\mu}(\{y\}_\epsilon \cap H \setminus B^*) > 0.9\tilde{\mu}(\{y\}_\epsilon) \end{cases}$$

Let $h = y - x$. Then, applying the symmetrization operator R_h to B^* will transfer a subset of measure at least $0.8\tilde{\mu}(\{x\}_\epsilon)$ of B^* from H^C to H, while no point of B^* in H will be transferred to a point outside H (since H is a half-space and points will be transferred in the direction $h = y - x$ where $y \in H$ and $x \notin H$). Thus,

$$\tilde{\mu}(R_h(B^*) \cap H) \geq \tilde{\mu}(B^* \cap H) + 0.8\tilde{\mu}(\{x\}_\epsilon) > \tilde{\mu}(B^* \cap H)$$

contradicting the optimality of B^*. Hence, we must have $\tilde{\mu}(B^* \cap H) = \tilde{\mu}(H)$. But then $B^* = H$ (a.s. $\tilde{\mu}$) and

$$\widetilde{\mathbb{S}}_\rho^{(k)}(H) = \widetilde{\mathbb{S}}_\rho^{(k)}(B^*) \geq \widetilde{\mathbb{S}}_\rho^{(k)}(A)$$

as needed. $\square$

References

[1] C. Borell, *Geometric bounds on the ornstein-uhlenbeck velocity process*, Probability Theory and Related Fields **70** (1985), no. 1, 1–13.

[2] J. Corneli, I. Corwin, Y. Xu, S. Hurder, V. Sesum, E. Adams, D. Davis, M. Lee, R. Pettit, and N. Hoffman, *Double bubbles in gauss space and spheres*, Houston journal of mathematics **34** (2008), no. 1, 181–204.

[3] Uriel Feige and Gideon Schechtman, *On the optimality of the random hyperplane rounding technique for max cut*, Tech. report, Algorithms, 2002.

[4] Michel Ledoux and Michel Talagrand, *Probability in banach spaces*, Springer, May 1991.