Non Linear Invariance & Applications

Elchanan Mossel

U.C. Berkeley
http://stat.berkeley.edu/~mossel

Talk Plan

- Probability and Gaussian Geometry:
  - Non linear invariance for low influence functions
  - Gaussian Geometry and "Majority is Stablest".
- Quantitative Social choice
  - Qauntitative Arrow theorem.
- Approximate Optimization
  - Unique Games and hardness of Max-Cut
  - General Optimization
- More on Gaussian Geometry.

Lindeberg & Berry Esseen

- Let X_i = +1/-1 w.p $\frac{1}{2}$, $N_i \sim N(0,1)$ ind.
- $f(x) = \sum_{i=1}^n c_i x_i$ with $\sum c_i^2 = 1$.
- Thm: (Berry Esseen CLT):
- $\sup_t |P[f(X) \le t] - P[f(N) \le t]| \le 3 \max |c_i|$
- Note that $f(N) = f(N_1,\ldots,N_n) \sim N(0,1)$.
- Lindeberg pf idea: can replace X_i with N_i "one at a time" as long as all coefficients are small.
- Q: can this be done for other functions f? e.g. multi-linear polynomials?

Some Examples

- Q: Is it possible to apply Lindeberg principle to other functions f with small coefficients to show that $f(X) \sim f(N)$?

- Ex 1: $f(x) = (n^3/6)^{-1/2} \sum_{i<j<k} x_i x_j x_k$
- → Okay: Limit is $N^3 - 3N$

- Ex 2: $f(x) = (2n)^{-1/2} (x_1 - x_2) (x_1 + ... + x_n)$
- → Not OK
- For X: $P[f(X) = 0] \geq \frac{1}{2}$.

Invariance Principle

- Thm (MOO := M-O' Donnell-Oleszkiewicz):
- Let $f(x) = \sum_S c_s X_s$ be a multi-linear of degree k with $\sum c_S^2 = 1$ ($X_S = \prod_{i \in S} x_i$)
- $I_i(f) := \sum_{S : i \in S} c_S^2$, $\delta(f) = \max_i I_i(f)$
- Then:
- $\sup_t |P[f(X) \leq t] - P[f(N) \leq t]| \leq 3\,k\,\delta^{1/8d}$
- Works if X has $2+\varepsilon$ moments + other setups.

The Role of Hyper-Contraction

- Pf Ideas:
- Lindeberg trick (replace one variable at a time)
- Hyper-contraction allows to bound high moments in term of lower ones.
- Key fact: A degree d polynomial S of hyp. contract. variables satisfies $||S||_q \leq C(q)^d \, ||S||_2$

An Invariance Principle

- Invariance Principle [M+O'Donnell+Oleszkiewicz(05)]:
- Let $p(x) = \sum_{0 < |S| \le k} a_S \prod_{i \in S} x_i$ be a degree k multi-linear polynomial with $|p|_2 = 1$ and $I_i(p) \le \delta$ for all i.
- Let $X = (X_1,\dots,X_n)$ be i.i.d. $P[X_i = \pm 1] = 1/2$.
  $N = (N_1,\dots,N_n)$ be i.i.d. Normal(0,1).
- Then for all t:

$$|P[p(X) \le t] - P[p(N) \le t]| \le O(k\,\delta^{1/(4k)})$$

(Proof works for any hyper-contractive random vars).

Invariance Principle – Proof Sketch

- Suffices to show that $\forall$ smooth F (sup $|F^{(4)}| \leq C$), E[F(p($X_1,\ldots,X_n$)] is close to E[F(p($N_1,\ldots,N_n$))].

Main Lemma.

$$|\mathbf{E}[F(p(X_1,\ldots,X_{i-1},N_i,N_{i+1}\ldots,N_n)]- \mathbf{E}[F(p(X_1,\ldots,X_{i-1},X_i,N_{i+1},\ldots,N_n)]| \leq C9^k I_i^2 \leq C9^k \delta I_i.$$

Therefore

$$|\mathbf{E}[F(p(X_1,\ldots,X_n)] - \mathbf{E}[F(p(N_1,\ldots,N_n)]| \leq C9^k\delta \sum_i I_i \leq Ck9^k\delta.$$

Invariance Principle – Proof Sketch

- Write: $p(X_1,\ldots,X_{i-1}, N_i, N_{i+1},\ldots,N_n) = R + N_i\, S$
- $p(X_1,\ldots,X_{i-1}, X_i, N_{i+1},\ldots,N_n) = R + X_i\, S$
- $F(R+N_i\, S) = F(R) + F'(R)\, S\, N_i + F''(R)\, (S^2/2)\, N_i^2 + F^{(3)}(R)\, (S^3/6)\, N_i^3 + F^{(4)}(*)\, N_i^4\, S^4/24$
- $E[F(R+ N_i\, S)] = E[F(R)] + E[F''(R)\, S^2]\, /2 + E[F^{(4)}(*) N_i^4 S^4]/24$
- $E[F(R + X_i\, S)] = E[F(R)] + E[F''(R)\, S^2]\, /2 + E[F^{(4)}(*) X_i^4\, S^4]/24$
- $|E[F(R + N_i\, S) - E[F(R + X_i\, S)| \leq C\, E[S^4]$
- But, $E[S^2] = I_i(p)$.
- And by Hyper-Contractivity, $E[S^4] \leq 9^{k-1}\, E[S^2]^2$
- So: $|E[F(R + N_i\, S) - E[F(R + X_i\, S) \leq C\, 9^k\, I_i^2$ ■

A direct proof of $E[S^4] \leq 9^{k-1} E[S^2]$

- Assuming: $E[X_i] = E[X_i^3] = 0$, $E[X_i^2] = 1$, $E[X_i^4] \leq 9$.
  Note: $deg(S) = k-1$.
- Pf by induction on number of variables.
- Write $S = R + X_n T$ so $deg(T) \leq k-2$.
- $E[S^4] = E[R^4] + 6\, E[R^2\, T^2] + E[X_n^4]\, E[T^4]$
- $\leq E[R^4] + 6\, E[R^2\, T^2] + 9\, E[T^4]$
- CS: $\leq (E[R^4]^{1/2} + 3\, E[T^4]^{1/2})^2$
- Induction: $\leq (3^{k-1}\, E[R^2] + 3*3^{k-2}\, E[T^2])^2$
- $= 9^{k-1}\, (E[R^2] + E[T^2])^2 = 9^{k-1}\, E[S^2]^2$

Related Work

- Many works generalizing Lindeberg ideaa.
- Rotar 79: Similar but no hyper-contraction, Berry-Esseen.
- Classical results for U,V statistics.
- M (FOCS 08, Geom. and Functional Analysis 10):
- Multi-function versions.
- General "noise".
- Bounds in terms of cross influences.
- Motivation: Proving "Majority is Stablest".

Majority is Stablest

- Let $(X_i, Y_i) \in \{-1,1\}^2$ & $E[X_i] = E[Y_i] = 0$; $E[X_i\, Y_i] = \rho$.
- Let $Maj(x) = sgn(\sum x_i)$.
- Thm (Sheffield 1899):
- $E[Maj(X)\, Maj(Y)] \rightarrow M(\rho) := (2 \arcsin \rho)/\pi$
- Pf Idea:
- Let $N, M \sim N(0,1)$ jointly Gaussian with $E[N\, M] = \rho$.
- Then:
- $\lim E[Maj(X)\, Maj(Y)] = E[sgn(N)\, sgn(M)] = M(\rho)$

Majority is Stablest

- Let $(X_i,Y_i) \in \{-1,1\}^2$ & $E[X_i] = E[Y_i] = 0$; $E[X_i Y_i] = \rho$.
- Let $Maj(x) = sgn(\sum x_i)$.
- Thm (Sheffield 1899):
- $E[Maj(X)\ Maj(Y)] \rightarrow M(\rho) := (2 \arcsin \rho)/\pi$
- Thm (Borell, 1985):
- Let N,M be two n-dim normal vectors
- where (N_i,M_i) i.i.d. & $E[N_i] = E[N_i] = 0$; $E[N_i M_i] = \rho$.
- Let $f : R^n \rightarrow [-1,1]$ with $E[f] = 0$.
- Then: $E[f(N)\ f(M)] \le E[sgn(N_1)\ sgn(M_1)] = M(\rho)$

Majority is Stablest

- Let $(X_i, Y_i) \in \{-1,1\}^2$ & $E[X_i] = E[Y_i] = 0$; $E[X_i Y_i] = \rho$.
- Let $Maj(x) = sgn(\sum x_i)$.
- Thm (Sheffield 1899):
- $E[Maj(X)\ Maj(Y)] \to M(\rho) := (2 \arcsin \rho)/\pi$
- Thm (MOO; "Majority is Stablest"):
- Let $f : \{-1,1\}^n \to [-1,1]$ with $E[f] = 0$.
- $I_i(f) := P[f(X_1,\ldots,X_i,\ldots,X_n) \neq f(X_1,\ldots,-X_i,\ldots,X_n)]$,
- $I = \max I_i(f)$
- Then: $E[f(X)\ f(Y)] \leq M(\rho) + C/\log^2(1/I)$

Majority is Stablest – Pf Idea

- Pf Sketch:

- $E[f(X)\, f(Y)] = E[g(X')\, g(Y')]$ where
- $g = T_{\eta'} f$, X' and Y' are η correlated and $\rho = \eta'^2 \eta$

- g is essentially a low-degree function.

- Since g is of low influence and "low degree":
- $E[g(X)\, g(Y)] \sim E[g(N)\, g(M)] \leq M(\rho)$

Majority is Stablest – Context

- Conext:
- Conjectured by Kalai in 2002 as it implies majority minimized Arrow paradox in a class of functions.
- Proves the conjecture of Khot-Kindler-M-O' Donnell 2005 in the context of approximate optimization.
- More general versions proved in M-10
- M-10 allows truncation in general "noise" structure.
- E.g: In M-10: Majority is most predictable:
- Among low influence functions majority outcome is most predictable give a random sample of inputs.

Quantitative Social Choice

- Quantitative social choice studies different voting methods in a quantitative way.

- Standard assumption is of uniform voting probability.

- A "stress-test" distribution.
- Renewed interest in the context of computational agents.
- Consider general voting rule
- $f: \{-1,1\}^n \rightarrow \{-1,1\}$ or $f : [q]^n \rightarrow [q]$ etc.

Errors in Voting

- Suppose each vote is re-randomized with probability ε (by voting machine):
- Majority is Stablest in voting language:
- Majority minimizes probability of error in outcome among low influence functions.
- Plurality is Stablest (IM) 11:
- Plurality minimizes probability of error in outcome among low influence functions (this is equivalent to the Peace-Sign conjecture)

Errors in Voting

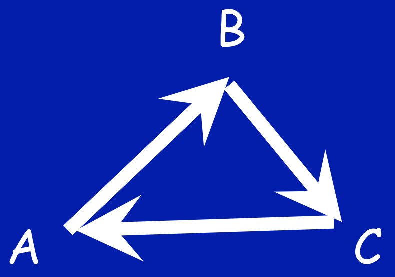
- Majority is Most Predictable (M 08; 10):
- Suppose each voter is in a poll with prob. p independently.
- Majority is most predictable from poll among all low influence functions.

- Next Example – Arrow theorem
- Fundamental theorem of modern social choice.

Condorcet Paradox

- n voters are to choose between 3 options / candidates.
- Voter i ranks the three candidates A, B & C via a permutation $\sigma_i \in S_3$
- Let $X^{AB}_i = +1$ if $\sigma_i(A) > \sigma_i(B)$

  $X^{AB}_i = -1$ if $\sigma_i(B) > \sigma_i(A)$
- Aggregate rankings via: $f,g,h : \{-1,1\}^n \to \{-1,1\}$.
- Thus: A is preferred over B if $f(x^{AB}) = 1$.
- A Condorcet Paradox occurs ("f irrational") if: $f(x^{AB}) = g(x^{BC}) = h(x^{CA})$.
- Defined by Marquis de Condorcet in 18'th century.

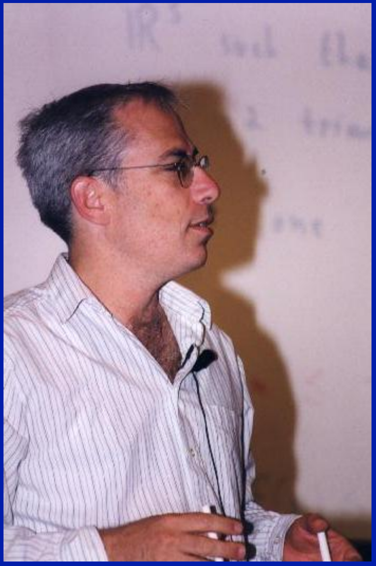


Arrow's Impossibility Thm

- Thm (Condorecet): If $n > 2$ and f is the majority function then there exists rankings $\sigma_1,\dots,\sigma_n$ resulting in a Paradox

- Thm (Arrow's Impossibility): For all $n > 1$, unless f is the dictator function, there exist rankings $\sigma_1,\dots,\sigma_n$ resulting in a paradox.
- Arrow received the Nobel prize (72)

Probability of a Paradox



- What is the probability of a paradox:
- $PDX(f) = P[f(x^{AB}) = f(x^{BC}) = f(x^{CA})]$?
- Arrow's: f = dictator iff PDX(f) = 0.
- Thm(Kalai 02): Majority is Stablest for $\rho=1/3$ → majority minimizes probability of paradox among low influences functions (7-8%).
- Thm(Isacsson-M 11): Majority maximizes probability of a unique winner for any number of alternatives.
- (Proof uses invariance + Exchangble Gaussian Theorem)

Probability of a Paradox

- Thm(Kalai 02): Majority is Stablest for $\rho=1/3 \rightarrow$ majority minimizes probability of paradox among low influences functions (7-8%).

- Pf Sketch:
- $PDX(f) = \frac{1}{4}\,(1+E[f(x^{AB})\, f(x^{BC})+f(x^{BC})\, f(x^{CA})+f(x^{CA})\, f(x^{AB})])$
- $|E[f(x^{AB})\, f(x^{BC})]| = |E[f(x^{AB})\, f^{-}(-x^{BC})]| = |E[f\, T_{1/3}\, f^{-}]|$
- $\leq E[f\, T_{1/3}\, f]^{1/2}\, E[f^{-}\, T_{1/3}\, f^{-}]^{1/2} \leq M(1/3)$
- $E[m\, T_{1/3}\, m^{-}] = E[-m\, T_{1/3}\, m] = -\, M(1/3).$

A quantitative Arrow Thm

- Arrow's: f = dictator iff PDX(f) = 0.
- Kalai 02: Is it true that $\forall \varepsilon \exists \delta$ such that
- if PDX(f) < δ
- then f is ε close to dictator?
- Kalai 02: Yes if there are 3 alternatives and E[f] = 0.

- M-11: True for any number of alternatives.
- Keller-11: Optimal dependency between δ ϵ.

- Pf uses Majority is stablest and inverse_hyper-contractive inequalities (including quantitative Barbera Thm we saw).

Approximate Computational Hardness and Fourier Analysis

- Fourier Analysis plays an important role in hardness of approximation since the beginning.

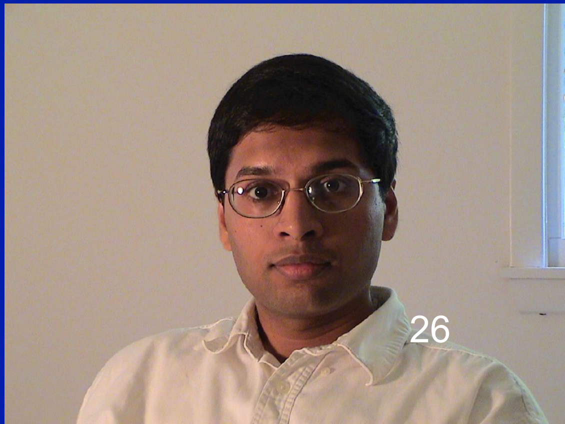
- We follow with a brief overview of the connection to Gaussian techniques.

- Optimist CS: Design efficient algorithms.
- Pessimist CS: Problem is NP-hard.
- Optimist CS: Design efficient approximation algs.
- Pessimist CS: Prove: computationally hard to approximate.

- New methodology: "UGC hardness".

Approximate Optimization

- Many optimization problems are NP-hard.
- Instead: Approximation algorithms
- These are algorithms that guarantee to give a solution which is at least
- α OPT or OPT - ε.
- S. Khot (2002) invented a new paradigm for analyzing approximation algorithms – called UGC (Unqiue Games Conjecture)



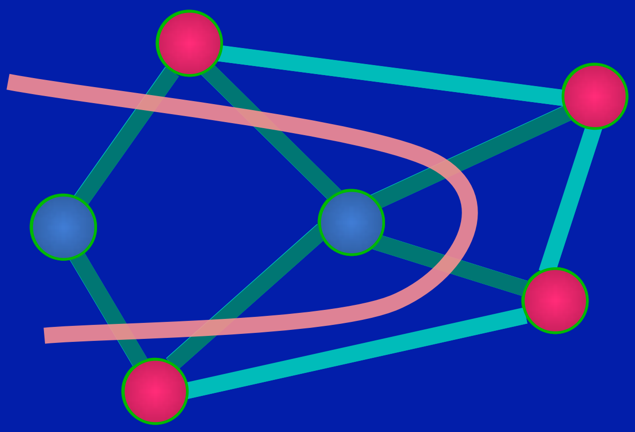
THE UGC

- UGC: For all $\epsilon > 0$ $\exists$ q s.t. given
- n equations of the form $x_i + x_j = c_{i,j} \bmod q$
- It is computationally hard to distinguish between the following two scenarios:
- 1. It is possible to satisfy at most ϵ fraction of the Equations simultaneously.
- It is possible to satisfy

at least $1-\epsilon$ of the equations.

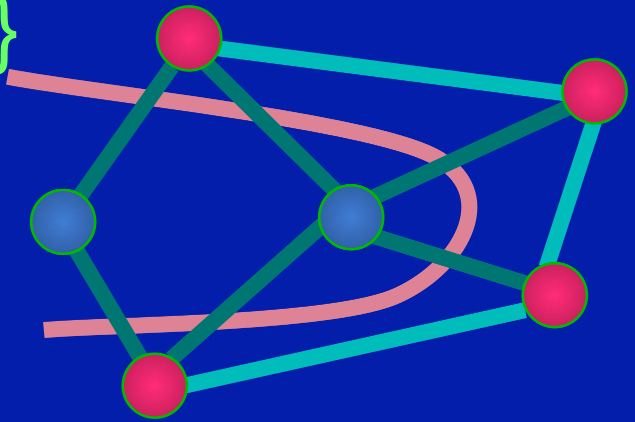
Example 1: The MAX-CUT Problem

- $G = (V,E)$
- $C = (S^c,S)$, partition of V
- $w(C) = |(S \times S^c) \cap E|$
- $w : E \longrightarrow R^+$
- $w(C) = \sum_{e \in E \cap S \times S^c} w(e)$



Example: The Max-Cut Problem

- OPT = OPT(G) = $\max_C \{|C|\}$



- **MAX-CUT problem:**
  find C with w(C)= OPT

- **α-approximation:**
  find C with w(C) ≥ α·OPT

- Goemans-Williamson-95:
- Rounding of
- Semi-Definite Program gives an

α = .878567 approximation algorithm.

MAX-Cut Approximation

- Thm (KKMO = Khot-Kindler-M-O' Donnell, 2007):
- Under UGC, the problem of finding an
- $\alpha > a_{GW} = \min\{2\,\theta / \pi\,(1-\cos\theta) : 0 < \theta < \pi\} = 0.87\ldots$ approximation for MAX-CUT is NP-hard.
- Moral: Semi-definite program does the best!

- Thm (IM-2011): Same result for MAX-q-CUT assuming the Peace-Sign Conjecture.

MAX-Cut Approximation

- Thm (KKMO):
- High level proof idea:
- Approximation factor is L/M where
- M = Opt E[f(x) f(y) : E[f] = 0]
- L =   lim Opt E[f(x) f(y) : E[f]= 0, I(f) < ε}
- (x,y) have some "noise structure"
- Second quantity studied via invariance + Majority is Stablest.

Other Approximation problems

- A second result using Invariance of M 08;10
- Raghavendra 08: Duality between Algorithms and Hardness for Constraint Satisfaction Problems.
- Thm: Every instance with gap β' < β can be used to prove UGC-based β'- hardness result !

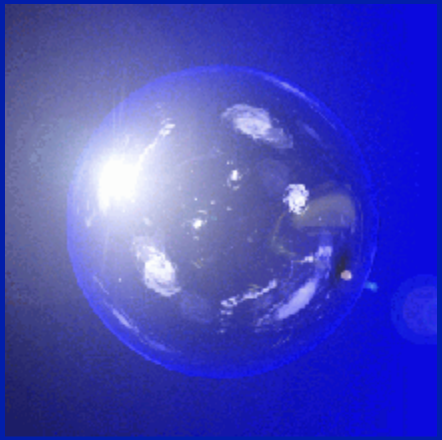
- Implies Semi-definite programs with "optimal rounding" are optimal algorithms for optimization of Constraint Satisfaction Problems.

Other Approximation problems

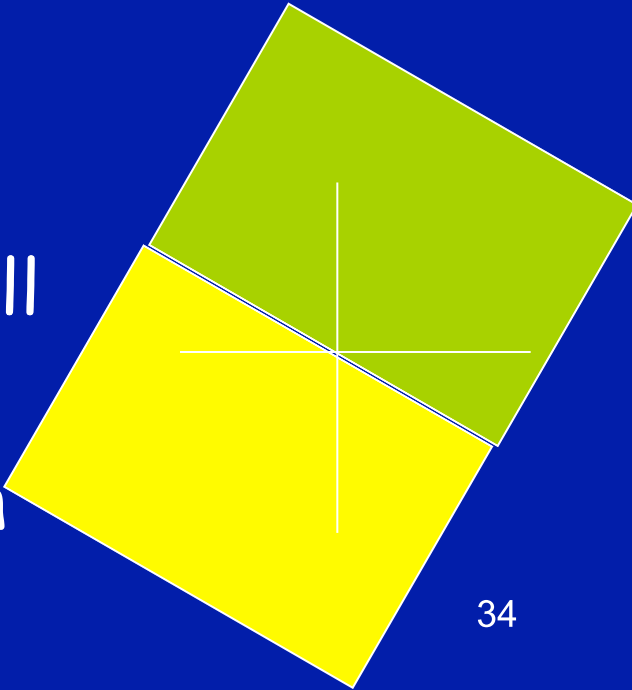
- After KKMO+MOO
- Dozens of papers use the same recipe
- → obtain optimal approximation ratio for many optimization problems.
- Best results use "general" invariance M-08;10.
- Ex :Thm: (Austrin-M):
- Predicates that are pairwise independent cannot be approximated better than random.

Geometry behind Borell's results

- I. Ancient: Among all sets with $\nu_n(A) = 1$ the minimizer of $\nu_{n-1}(\partial A)$ is A = Ball.

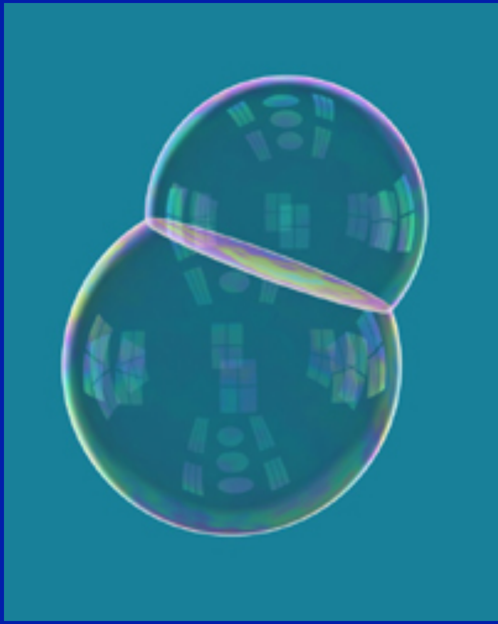


- II. Recent (Borell, Sudakov-Tsierlson 70's) Among all sets with $\gamma_n(A) = a$ the minimizer of $\gamma_{n-1}(\partial A)$ is A = Half-Space.
- III. More recent (Borell 85): For all ρ, among all sets with $\gamma(A) = a$ the maximizer of $E[A(N)A(M)]$ is given by A = Half-Space.

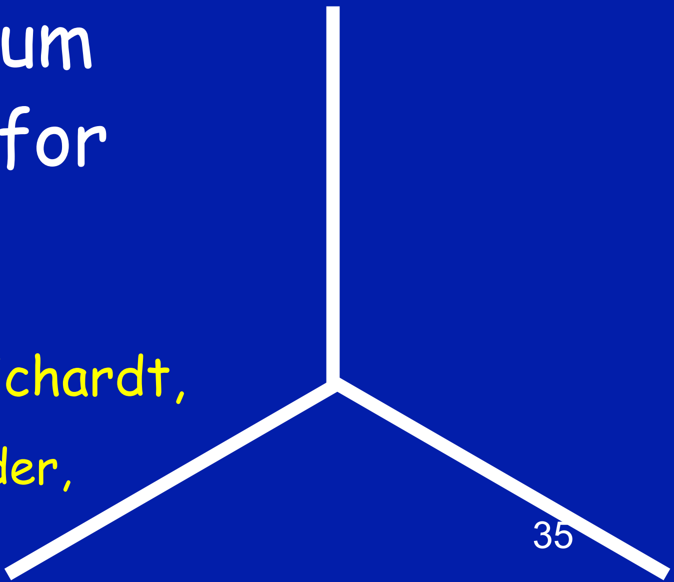


Double bubbles

- **<u>Thm1</u>** ("Double-Bubble"):
- Among all pairs of disjoint sets A,B with $\nu_n(A) = a$ $\nu_n(B) = b$, the minimizer of $\nu_{\nu-1}(\partial A \cup \partial B)$ is a "Double Bubble"



- **<u>Thm2</u>** ("Peace Sign"):
- Among all partitions A,B,C of R^n with $\gamma(A) = \gamma(B) = \gamma(C) = 1/3$, the minimum of $\gamma(\partial A \cup \partial B \cup \partial C)$ is obtained for the "Peace Sign"



- 1. Hutchings, Morgan, Ritore, Ros. + Reichardt, Heilmann, Lai, Spielman 2. Corneli, Corwin, Hurder, Sesum, Xu, Adams, Dvais, Lee, Vissochi

Newer Isoperimetric Results

- **Conj (Isaksson-M, Israel J. Math 2011):**
  For all $0 \le \rho \le 1$:

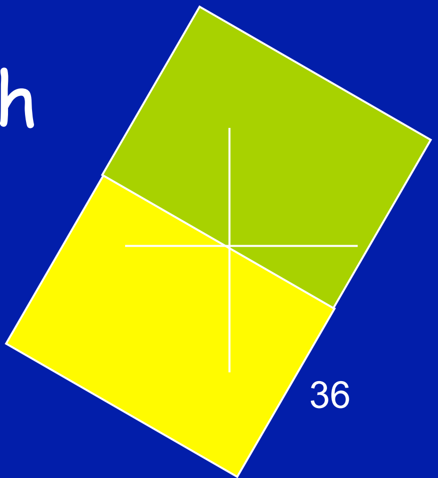
  $\mathrm{argmax}\ E[A(X)A(Y) + B(X)B(Y) + C(X)C(Y)]$
  = "Peace Sign"

  where max is over all partitions (A,B,C) of R^n with $\gamma_n(A) = \gamma_n(B) = \gamma_n(C) = 1/3$ is

Peace sign

Later we'll see applications

- Thm (Exchangble Guass. Thm, IM-11):
- Let X,Y,Z be Gaussian vectors each with pairwise $\rho \times Id$ covariance then
- $\mathrm{argmax}\{ E[A(X)A(Y)A(Z)] : \gamma_n(A) = \frac{1}{2}\}$ = half space.



A proof of Borell's result

- **<u>Cute proof</u>** <u>(Kinlder O'Donnell 2012)</u>**:**
- Let $P(A) = \frac{1}{2}$. Let M,N be $\rho = \cos\theta$ correlated $N(0,I)$
- $q(\theta) := P[N \in A,\ M \in A^c] =$
- $=P[N \in A, \cos\theta\, N + \sin\theta\, Z \in A^c]$
- $\leq kq(\theta/k)$.
- For $\theta = \pi/2$, $p(\theta) = \frac{1}{4}$.
- So $q(\pi/2k) \geq 1/(4k)$.
- For majority we get equality!
- $P[N_1 \in A, \cos\theta\, N_1 + \sin\theta\, Z_1 \in A^c] = \theta/(2\pi)$.

Summary

- Prove the "Peace Sign Conjecture" (Isoperimetry)
- $\Rightarrow$ "Plurality is Stablest" (Low Inf Bounds)
- $\Rightarrow$ MAX-3-CUT hardness (CS) and voting.

+ $\Rightarrow$ Results in Geometry.


Thank you!
ARTIE.COM