

Limitations of the sum of squares algorithm

In the last lecture we have seen how the sum of squares algorithm can achieve non-trivial performance guarantees for several interesting problems. But it is not all powerful. In this lecture we will see some *negative results* for the sum of squares algorithm, showing lower bounds on the degree needed to certify certain problems. In many cases, we do not know of any algorithms that do better, but in some cases we do, and we will see examples of both types. In the standard parlance, these negative results are known as *integrality gaps*, since these are instances in which there is a gap between the value that the pseudo-distribution “pretends” and the true objective value.¹

The cycle as an integrality gap for Cheeger’s Inequality and Max-Cut

Recall that the discrete Cheeger’s inequality states that every d -regular graph G with adjacency matrix A satisfies $\lambda \geq \Omega(\varphi(G)^2)$, where λ is the second smallest eigenvalue of the normalized Laplacian $L_G = I - \frac{1}{d}A$. It turns out the humble cycle shows that this bound is tight.

1. Lemma (Second eigenvalue of the cycle). *Let C_n be the cycle on n vertices and let L_{C_n} be its normalized Laplacian. Then the second smallest eigenvalue of L_{C_n} is at most $O(1/n^2)$.*

The lemma shows that Cheeger’s inequality is tight for the C_n because every subset $S \subseteq V(C_n)$ with $1 \leq |S| \leq n - 1$ has at least one neighbor outside of S . Therefore, the expansion of C_n is at least $\varphi(C_n) \geq \Omega(1/n)$.

Proof. Let $\omega = e^{2\pi i/n} \in \mathbb{C}$ be the n -th root of unity. The vector $v = (\omega^0, \omega^1, \dots, \omega^{n-1})$ is orthogonal to the all-ones vector $\mathbf{1}$,

$$\langle v, \mathbf{1} \rangle = \sum_{\ell=0}^{n-1} \omega^\ell = \frac{\omega^0 - \omega^n}{1 - \omega} = 0 \quad (1)$$

using the formula for the sum of a geometric progression. At the same time, we can upper bound the quadratic form of L_{C_n} at v ,

$$\begin{aligned} \langle v, L_{C_n} v \rangle &= \sum_{\ell=0}^{n-1} |\omega^\ell - \omega^{\ell+1}|^2 \\ &= |1 - \omega|^2 \cdot n \\ &\leq O(1/n^2) \cdot \|v\|^2. \end{aligned} \quad (2)$$

¹ The name “integrality gap”, as well as the related notion of a “rounding algorithm” arise from the setting of using a linear program as a relaxation of an integer linear program, where the optimal value of the linear programming relaxation is known as the *fractional value*, and the optimal value of the integer linear program is known as the *integral value*.

The last step uses $|1 - \omega|^2 = O(1/n^2)$. It follows that the second smallest eigenvalue of L_{C_n} is $O(1/n^2)^2$ \square

² In fact, it turns out that the second eigenvalue of L_{C_n} is exactly $\langle v, L_{C_n} v \rangle / \|v\|^2$ and v is an eigenvector corresponding to this eigenvalue.

In one of the exercises in the [chapter about Max Cut](#) you showed that if a graph G satisfies $\text{maxcut}(G) \leq 1 - \varepsilon$ then every degree-2 pseudo-distribution μ satisfies that $\tilde{\mathbb{E}}_{\mu} f_G(x) \leq 1 - \Omega(\varepsilon^2)$. The following lemma shows that this tradeoff is tight for cycles with an odd number of vertices. Concretely, if n is odd then C_n is not bipartite. Therefore, $\text{maxcut}(C_n) \leq 1 - 1/n$. In contrast, there are degree-2 pseudo-distributions μ such that $\tilde{\mathbb{E}}_{\mu} f_{C_n}(x) \geq 1 - O(1/n^2)$.

2. Lemma (max cut pseudodistribution for the odd cycle). *Let $n \in \mathbb{N}$ be odd. Then, there exists a degree two pseudo-distribution μ such that $\tilde{\mathbb{E}}_{\mu} f_{C_n}(x) \geq 1 - O(1/n^2)$.*

Proof. Let $\omega \in \mathbb{C}$ be the n -th complex root of unity as before. Suppose $n = 2k + 1$. Let $u \in \mathbb{C}^n$ with

$$(0, \omega^k, \omega^{2k}, \dots, \omega^{(n-1)k}) \quad (3)$$

Let $v, w \in \mathbb{R}^n$ be the real and imaginary part of u so that $u = v + i \cdot w$. Let X be the n -by- n positive-semidefinite matrix,

$$X = vv^{\top} + ww^{\top}. \quad (4)$$

Since $X_{ii} = v_i^2 + w_i^2 = |u_i|^2 = 1$, the diagonal of X is the all-ones vector $\mathbf{1}$.

We have seen that we can specify a pseudo-distribution μ by specifying its pseudo-expectation operator. Specifically, we will fix the expectation operator such that

$$\begin{aligned} \tilde{\mathbb{E}}_{\mu(x)} x &= \frac{1}{2} \mathbf{1}, \\ \tilde{\mathbb{E}}_{\mu(x)} xx^{\top} &= \frac{1}{4} \cdot \mathbf{1}\mathbf{1}^{\top} + \frac{1}{4} \cdot X. \end{aligned} \quad (5)$$

We leave it as an exercise to the reader to verify that the above is indeed a valid degree two pseudo-expectation.³

It remains to estimate the pseudo-expectation of f_G with respect to μ ,

$$\begin{aligned} \tilde{\mathbb{E}}_{\mu(x)} f_G(x) &= \frac{1}{4} \sum_{\{i,j\} \in E(G)} (v_i - v_j)^2 + (w_i - w_j)^2 \\ &= \frac{1}{4} \sum_{\{i,j\} \in E(G)} |u_i - u_j|^2 \\ &= \frac{1}{4} n \cdot |1 - \omega^k|^2 \geq n \cdot (1 - O(1/n^2)) \end{aligned} \quad (6)$$

³ **Hint:** This holds because the diagonal of the second moment $\frac{1}{4} \cdot \mathbf{1}\mathbf{1}^{\top} + \frac{1}{4} \cdot X$ agrees with mean $\frac{1}{2} \mathbf{1}$ and the formal covariance $\tilde{\mathbb{E}} \left(x - \frac{1}{2} \mathbf{1} \right) \left(x - \frac{1}{2} \mathbf{1} \right)^{\top} = \frac{1}{4} \cdot X$ is positive semidefinite. See exercise below.

In the last step, we use that

$$\begin{aligned} |1 - \omega^k|^2 &= (1 - \omega^{-k})(1 - \omega^k) = 2 - \omega^k - \omega^{-k} \\ &= 2 + 2 \cos \frac{\pi}{n} \geq 4 - O\left(\frac{1}{n}\right)^2. \end{aligned} \tag{7}$$

□

3. Exercise (degree-2 pseudo-distributions). Show that for every vector $v \in \mathbb{R}^n$ and every matrix $M \in \mathbb{R}^{n \times n}$ with $\text{diag}(M) = v$ and $M - vv^\top \succeq 0$, there exists a degree-2 pseudo-distribution μ over the hypercube $\{0, 1\}^n$ with $\mathbb{E}_{\mu(x)} x = v$ and $\mathbb{E}_{\mu(x)} xx^\top = M$.

Eigenvalues of Cayley graphs

The above ad-hoc computations of eigenvalues are actually special cases of a more general theory. Let (Γ, \cdot) be a group and $S \subseteq \Gamma$. The *Cayley graph* corresponding to Γ and S , which we denote as $G(\Gamma, S)$, has vertices corresponding to Γ and edges a, b for every $a, b \in \Gamma$ such that $ab^{-1} \in S \cup S^{-1}$. The n -cycle is simply the Cayley graph corresponding to the group \mathbb{Z}_n (integers in $\{0, \dots, n-1\}$ with addition modulo n) and the set $S = \{1\}$.

It turns out that for every Abelian group Γ and Cayley graph $G(\Gamma, S)$, we can explicitly calculate the eigenvectors and eigenvalues of $G(\Gamma, S)$. The following series of exercises works this out, first for cyclic groups and then for every Abelian group:

4. Exercise. Let $n \in \mathbb{N}$ and $\omega = e^{2\pi i/n}$. For every $\alpha \in \mathbb{Z}_n$, we define $\chi^\alpha \in \mathbb{C}^n$ to be the vector $\chi_j^\alpha = \omega^{\alpha \cdot j}$. Prove that if $S \subseteq \mathbb{Z}_n$, and A is the adjacency matrix of $G = G(\mathbb{Z}_n, S)$, then $A\chi^\alpha = \lambda_\alpha \cdot \chi^\alpha$ where $\lambda_\alpha = \sum_{j \in S} \omega^{\alpha \cdot j}$. (In particular, χ^α is an eigenvector of A with eigenvalue λ_α .)

Adjacency matrices of (potentially weighted) Cayley graphs over the group \mathbb{Z}_n are known as *circulant matrices*.

5. Exercise. Let Γ be an Abelian group of the form $\Gamma = \mathbb{Z}_{n_1} \times \dots \times \mathbb{Z}_{n_\ell}$, and let $\omega_t = e^{2\pi i/n_t}$. For every $\alpha = (\alpha_1, \dots, \alpha_\ell) \in \Gamma$, we define a vector $\chi^\alpha \in \mathbb{C}^\Gamma$ (called *character*) such that for every $j = (j_1, \dots, j_\ell) \in \Gamma$

$$\chi_j^\alpha = \prod_{t=1}^{\ell} \omega_t^{\alpha_t \cdot j_t}. \tag{8}$$

Prove that if $S \subseteq \Gamma$ and A is the adjacency matrix of $G(\Gamma, S)$, then every vector χ^α is an eigenvector of A with eigenvalue $\lambda_\alpha = \sum_{j \in S} \prod_{t=1}^{\ell} \omega_t^{\alpha_t \cdot j_t}$ so that $A\chi^\alpha = \lambda_\alpha \cdot \chi^\alpha$.

One of the most common instantiations of this result in computer science is for the case that Γ is the Boolean cube $\{0, 1\}^\ell$ with the XOR operation. In this case we can think of Γ as \mathbb{Z}_2^ℓ , and then, since $e^{2\pi i/2} = -1$, we get that the eigenvectors of the adjacency matrix of a graph $G(\{0, 1\}^n, S)$ have the form $\chi^\alpha \in \mathbb{R}^{\{0, 1\}^n}$ where $\chi_\beta^\alpha = -1^{\langle \alpha, \beta \rangle}$ with $\alpha, \beta \in \{0, 1\}^n$. The corresponding eigenvalue is $\sum_{\beta \in S} -1^{\langle \alpha, \beta \rangle}$.

The map that transforms a vector $v \in \mathbb{C}^\Gamma$ from its representation in the standard basis into its representation in the basis of characters of Γ (which are the eigenvectors of $G(\Gamma, S)$) is known as the *Fourier transform*. For more on this topic, let us point again to the textbook (O'Donnell [2014]).

A sharper integrality gap for Max Cut

The odd cycle shows that, at least for degree 2 pseudo-distributions, our analysis was tight up to a constant factor, but it does not yield the optimal constant. However, there is a more sophisticated example, due to Feige and Schechtman [2002], that yields a tight bound. Recall that the approximation ratio was $\alpha_{\text{GW}} = \min_{0 \leq x \leq 1} \frac{\arccos(1-2x)}{\pi x} \approx 0.878$ and was achieved at $x_{\text{GW}} \approx 0.845$.

6. Theorem (Tight integrality gap for max cut). *For every $\varepsilon > 0$, there exists a graph $G = (V, E)$ such that $\text{maxcut}(G) \leq \alpha_{\text{GW}} \cdot x_{\text{GW}} + \varepsilon$ and there is a degree 2 pseudo-distribution μ with $\tilde{\mathbb{E}}_\mu f_G(x) \geq x_{\text{GW}}$.*

Proof. Let's think of ε as some small $o(1)$ value that we will fix later. Looking at the analysis of the GW rounding algorithm, we see that to prove the theorem we need to come up with a graph G on n vertices and a degree 2 pseudo-distribution μ on $\{0, 1\}^n$ such that $\text{maxcut}(G) \leq \alpha_{\text{GW}} x_{\text{GW}} + o(1)$ but for almost all edges $\{i, j\}$ of G , $\tilde{\mathbb{E}}_\mu (x_i - x_j)^2 \geq x_{\text{GW}} - o(1)$. By the same calculations we did before, if we assume $\tilde{\mathbb{E}}_\mu x_i = 1/2$, then this corresponds to the normalized covariance of x_i and x_j satisfying $\tilde{\mathbb{E}}_\mu (x_i - 1/2)(x_j - 1/2) \leq (\rho_{\text{GW}} + o(1))1/4$ where $\rho_{\text{GW}} = 1 - 2x_{\text{GW}}$ (note that ρ_{GW} is roughly -0.69). Typically, we think of the graph as fixed and then we come up with the pseudo-distribution, but for this proof we will do this the other way around. We will first come up with μ and then *define* the graph G to correspond to those pairs $\{i, j\}$ in which $\tilde{\mathbb{E}}_\mu x_i x_j$ is roughly equal to ρ_{GW} . Moreover, borrowing an idea from the rounding algorithm, we will let μ be an *actual distribution*, but one over \mathbb{R}^n instead of $\{0, 1\}^n$. In fact, μ will correspond to an actual multivariate Gaussian distribution over \mathbb{R}^n , with $\mathbb{E}_\mu x_i = 1/2$ for

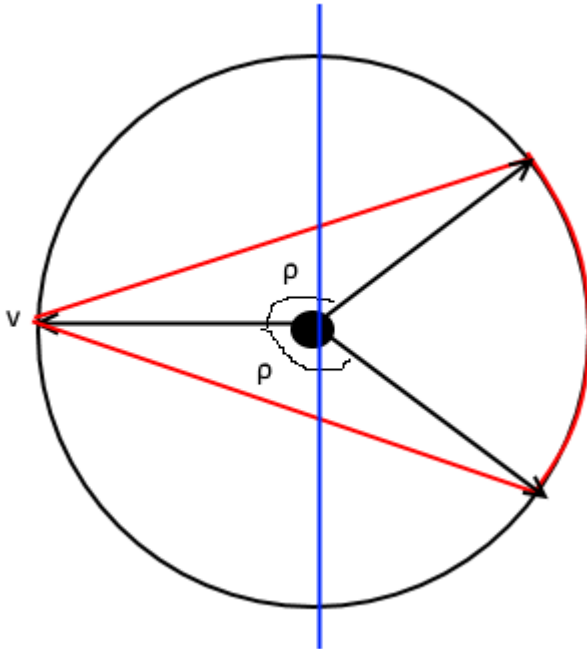


Figure 1: In the Feige-Schechtman graph, the vertices correspond to unit vectors in \mathbb{R}^d , and the neighbors of a vector v consists of all vectors w such that $\langle v, w \rangle \leq \rho$ for $\rho = \rho_{\text{GW}} + \epsilon$. The figure shows the case for $d = 2$, where the neighbors of v correspond to the red arc of vectors that have at most ρ correlation with v . The maximum cut in such a graph is obtained by partitioning the vertex set according to a hyperplane going through the origin.

all i (and hence $\mathbb{E}_\mu(x_i - 1/2)^2 = 1/4$ for all i). How do we come up with such a distribution? First note that we can make the number of vertices n as large as we like as a function of our desired accuracy ϵ , and hence we will think of n as very large. In fact, we will think of n as *very very* large: so large that it is practically infinite or even continuous! Concretely, we will identify the vertices of the graph G with the $d - 1$ dimensional unit sphere \mathcal{S}^{d-1} in \mathbb{R}^d (for some dimension parameter d depending on the desired accuracy) - that is the vertex set is the set of all $v \in \mathbb{R}^d$ with $\|v\| = 1$. (The one dimensional sphere is a circle, the two dimensional sphere is the boundary of a 3 dimensional ball, and so on..)

The set E of edges will be the set of pairs of vectors $(u, v) \in \mathcal{S}^{d-1} \times \mathcal{S}^{d-1}$ such that $\langle u, v \rangle \leq \rho_{\text{GW}} + \epsilon$. We can think of the max cut value of G as the maximum over all measurable subsets S of \mathcal{S}^{d-1} , of the measure of $E \cap S \times (\mathcal{S}^{d-1} \setminus S)$. Ultimately, we will obtain a finite graph by sampling n such vectors, but as long as n is large enough ($n \gg 2^d$ will do) then this finite graph will inherit both the max-cut value, as well as the pseudo-distribution value. However, the heart of the argument happens in the continuous setting, so you can ignore for the moment that final sampling stage.

We now need to come up with a collection of correlated random variables $\{X_v\}_{v \in \mathbb{R}^d}$ such that for every $v \in \mathbb{R}^d$, $\mathbb{E} X_v = 1/2$ and every edge (u, v) , the covariance of X_u and X_v is at most $-\rho_{\text{GW}} + O(\epsilon)$. This

collection will be very simple: we choose a random standard Gaussian $g \in \mathbb{R}^d$ (i.e., $g \in \mathbb{R}^n$ is chosen with $g_i \in N(0, 1)$ independently for all i), and for every $v \in \mathbb{R}^d$, we define $X_v = 1/2 + \langle v, g \rangle / 2$. Note that $\mathbb{E} X_v = 1/2$ and that $\mathbb{E}(X_v - 1/2)^2 = \mathbb{E}\langle v, g \rangle^2 / 4 = 1/4$.⁴

Now the normalized covariance (subtracting the expectation and dividing by the standard deviation) of u and v corresponds to $\mathbb{E}\langle u, g \rangle \langle v, g \rangle$ which by standard manipulations is the same as $\mathbb{E} \text{Tr}(uv^\top gg^\top)$ (thinking of u, v, g as column vectors and hence uv^\top and gg^\top are $n \times n$ matrices. But by linearity of trace and expectation this is the same as $\text{Tr}(uv^\top \mathbb{E} gg^\top)$ and since for a standard Gaussian g , $\mathbb{E} gg^\top = \text{Id}$, we get that this normalized covariance is equal to $\text{Tr}(uv^\top) = \langle u, v \rangle$ which equals at most $\rho_{\text{GW}} + \varepsilon$ by our definition of the edge set.

The above shows that we have a degree two pseudo-distribution μ satisfying that with high probability over $\{u, v\} \in E$, $\mathbb{E}_\mu (X_u - X_v)^2 \geq x_{\text{GW}} - \varepsilon$. But we still need to show that the true maximum cut value is at most $\alpha_{\text{GW}} x_{\text{GW}} + o(1)$. Luckily, here we can “stand on the shoulders of giants” and use previously known results. Specifically, by the geometric nature of this graph, intuitively the maximal cuts would be obtained via a geometric partition. Indeed, [Borell \[1975\]](#) (and, independently [Sudakov and Cirel'son \[1974\]](#)), proved that over the unit sphere, when we define the edge sets in such geometric terms, then the maximum cuts that optimize this will always be *spherical caps*. That is, the set S would be of the form $\{v \in \mathbb{R}^d : \langle v, a_0 \rangle \geq b_0\}$ for some $a_0 \in \mathbb{R}^d$ and $b_0 \in \mathbb{R}$. Specifically, in this case, one can show that the bipartition that would maximize the number of cut edges would be a *balanced one* (where S and $S^{d-1} \setminus S$ have the same measure) and hence $b_0 = 0$. By the rotational symmetry of the sphere (and appropriately scaling b_0), we can assume without loss of generality that a_0 is simply the first standard basis vector $(1, 0, \dots, 0)$. Now, since a random edge $(u, v) \in E$ is chosen by letting v be a random vector with correlation roughly ρ_{GW} with u , and since the first coordinate of a random unit vector has (essentially) the Gaussian distribution with mean zero and variance $1/n$, computing the value of the cut reduces to computing the probability that two ρ_{GW} correlated Gaussians disagree in their sign. This latter quantity is exactly what we computed in the last lecture as $1 - \arccos(\rho_{\text{GW}}) / \pi = \alpha_{\text{GW}}$. \square

Based on our construction, we can see that if the final graph has n vertices, then there is a unit vector v_i associated with each vertex i , and the pseudo-expectation operator is defined as $\tilde{\mathbb{E}} x_i x_j = 1/2 + 1/2 \langle v_i, v_j \rangle$. The resulting matrix is the sum of the psd all-1/2 matrix plus the Gram matrix (i.e., matrix of dot products) of the

⁴ Both equations follow from the rotational symmetry of a standard Gaussian, which means that without loss of generality $v = e_1$, in which case $\langle g, v \rangle = g_1$ is simply a one dimensional standard Gaussian.

vectors $\{v_1, \dots, v_m\}$. It is not hard to verify that such a matrix is psd—see also the exercises below:

7. Exercise. Prove that an $n \times n$ matrix M is a psd matrix of rank d if and only if there exist $v_1, \dots, v_n \in \mathbb{R}^d$ such that $M_{i,j} = \langle v_i, v_j \rangle$.

8. Exercise. Suppose that M, N are two $n \times n$ psd matrices such that $M_{i,j} = \langle v_i, v_j \rangle$ and $N_{i,j} = \langle u_i, u_j \rangle$. Show explicitly a tuple of vectors (w_1, \dots, w_n) such that the psd matrix $L = M + N$ satisfies $L_{i,j} = \langle w_i, w_j \rangle$.

Isoperimetry, extremal questions, and sum of squares

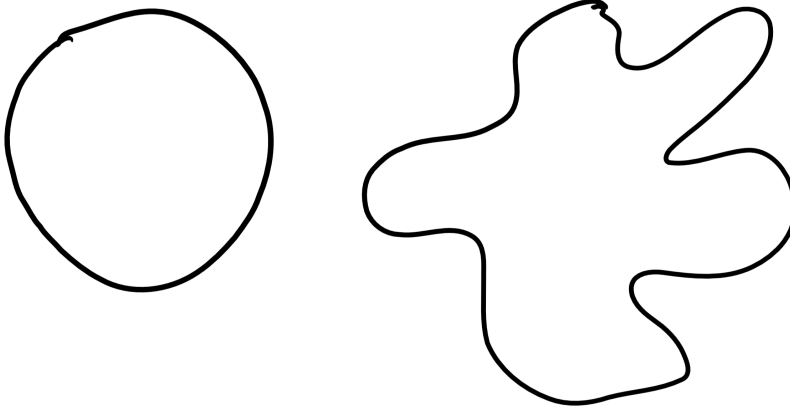


Figure 2: The classical isoperimetric inequality is that the circle minimizes the boundary among all closed curves with the same volume

The result of (Borell [1975], Sudakov and Cirel'son [1974]) we used above is one in a long line of work on **isoperimetric inequalities** and their many generalizations. The classical isoperimetric problem is to prove that among all shapes in the plane, the circle is the one that minimizes the ratio of its boundary to its volume. This question has been generalized to many other geometric spaces and notions of volume and boundaries. Indeed, it corresponds to the question we have already seen of finding the least expanding set in the graph: if we think of a very fine grid graph that discretizes the plane, then the isoperimetric problem corresponds to proving that the circle is the set of vertices that minimizes the *expansion*.

Isoperimetric questions themselves are just a special case of more general questions of finding and characterizing *extremal objects*. The general setting can be thought of as follows. We have:

- Some “universe” \mathcal{U} of possible objects (e.g., all subsets of a graph,

all closed curves in the plane, all functions or vectors in some space).

- Some “objective” function $F : \mathcal{U} \rightarrow \mathbb{R}$ (e.g., the ratio of boundary to volume, the sum of violations of some constraints)
- Some “nice” family $\mathcal{S} \subseteq \mathcal{U}$ (e.g., shifts of circles, spherical caps, codewords)

The type of theorems we might want to prove (in increasing order of difficulty, and, often, usefulness) would be:

- **An optimality theorem:** If $S \in \mathcal{U}$ is a global minimum of $F(\cdot)$ then $S \in \mathcal{S}$. Like in the isoperimetric case, such optimality theorems are often phrased as *inequalities* of the form $F(S) \geq \alpha^*$ for every $S \in \mathcal{U}$, where α^* is the (typically easily computable) minimum of $F(S)$ among all the “nice” $S \in \mathcal{S}$.
- **A stability theorem:** If $S \in \mathcal{U}$ is a “near global minimum” (i.e., $F(S)$ is close to $\min_{S' \in \mathcal{U}} f(S')$) then S is “close” to some “nice” $S^* \in \mathcal{S}$. The notion of “close” here of course needs to be defined and depends on the context.
- **An inverse theorem:** If $S \in \mathcal{U}$ has “non trivial $F(\cdot)$ value” (i.e., $F(S)$ is significantly smaller than the expected value of $F(S')$ for a random S') then S is “somewhat correlated” with some “nice” $S^* \in \mathcal{S}$. Again, the notion of “somewhat correlated” is context-dependent. Sometimes inverse theorems come with a “list decoding” variant in which the condition is that the non-trivial value of $F(S)$ can be explained by expressing S as some combination of a small number of “nice” objects in \mathcal{S} where again what is “small” and what combinations is one allowed to take.

A related question is the notion of **structure vs. randomness** as discussed by Tao (see for example (Tao [2007b]; Tao [2007a]; Tao [2008])). Given some object $S \in \mathcal{U}$ and some family of tests/objectives \mathcal{F} , we want to decompose S into the “structured part” that is some combination of objects from the “nice” family \mathcal{S} and the “random part” which we can think of as some “noise” object N such that $F(N)$ is close to the expectation of $F(S)$ over a random $S \in \mathcal{U}$ for all $F \in \mathcal{F}$.

There are often algorithmic questions associated with such theorems. One such question is the “decoding” task of finding the “nice” S^* that is close to an S with small $F(\cdot)$ value. Another is the task of, given some description of \mathcal{U} and $F(\cdot)$, *certifying* an optimality theorem. Indeed, we will see that a very interesting question is often whether such an inequality has a low degree sum of squares proof. Often, an *algorithmic* proof of an optimality theorem will imply at

least a stability theorem if not stronger results. Indeed such algorithmic proofs often give an explicit process for optimizing F that given any starting point S ends up in an optimum point. Typically, if the starting point S already had a pretty good $F(\cdot)$ value then the algorithm would presumably not take too many steps and hence its final output will be “close” to the initial point.

At this point, when we’ve seen only one concrete example, this discussion might feel somewhat abstract, but many important results, including hypercontractivity, the “invariance principle”, Brascamb Lieb inequalities, results on list decoding, the Gowers norm, and others can be thought of as falling into this general framework. We will see several other examples of such results in this course.

Beyond degree 2

The examples above show that the rounding algorithms of the previous lecture are tight with respect to *degree two* pseudodistributions. But of course, we can run the sum of squares algorithm for larger degrees. We do pay a price in the running time, but it remains polynomial time for every constant degree, and as long as the degree is significantly smaller than n it would still be significantly faster than brute force.

Could it be that using larger, but still small degree, we can beat the guarantees for max cut, graph expansion, or boolean quadratic forms that are achieved respectively by Goemans-Williamson, Cheeger, or Grothendieck? The short answer that we do not know. It is known that if Khot’s *Unique Games Conjecture* is true (or the closely related *Small Set Expansion Hypothesis*) then no polynomial (or even $2^{n^{o(1)}}$ time) algorithms can beat those guarantees. In particular, for every $d = n^{o(1)}$ these conjectures predict that we can obtain instances with the same gaps as we showed in this lecture but with respect not to merely degree 2 but to the value achieved by degree d pseudo-distributions. However, even for $d = O(1)$ (even $d = 4$) this is still wide open. What we do know is that the same examples that we saw in this lecture do *not* yield such gaps. Here is one example:

9. Exercise. Let n be odd and C_n be the n -length cycle. Prove that for every degree 6 pseudo-distribution μ over $\{0, 1\}^n$, $\mathbb{E}_\mu f_{C_n} \leq (1 - 1/n)|E|$.⁵

⁵ **Hint:** Start by showing the *square triangle inequality* for degree 6 pseudo-distributions. That is prove that for every degree 6 pseudo-distribution μ over $\{0, 1\}^n$ and every $i, j, k \in [n]$, $\mathbb{E}_\mu (x_i - x_k)^2 \leq \mathbb{E}_\mu (x_i - x_j)^2 + (x_j - x_k)^2$.

References

- Christer Borell. The Brunn-Minkowski inequality in Gauss space. *Invent. Math.*, 30(2):207–216, 1975. ISSN 0020-9910.
- Uriel Feige and Gideon Schechtman. On the optimality of the random hyperplane rounding technique for MAX CUT. *Random Struct. Algorithms*, 20(3):403–440, 2002.
- Ryan O’Donnell. *Analysis of Boolean functions*. Cambridge University Press, New York, 2014. ISBN 978-1-107-03832-5. doi: 10.1017/CBO9781139814782. URL <http://dx.doi.org/10.1017/CBO9781139814782>.
- V. N. Sudakov and B. S. Cirel’son. Extremal properties of half-spaces for spherically invariant measures. *Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)*, 41:14–24, 165, 1974. Problems in the theory of probability distributions, II.
- Terence Tao. Structure and randomness in combinatorics. In *FOCS*, pages 3–15. IEEE Computer Society, 2007a.
- Terence Tao. The dichotomy between structure and randomness, arithmetic progressions, and the primes. In *International Congress of Mathematicians. Vol. I*, pages 581–608. Eur. Math. Soc., Zürich, 2007b. doi: 10.4171/022-1/22. URL <http://dx.doi.org/10.4171/022-1/22>.
- Terence Tao. *Structure and randomness*. American Mathematical Society, Providence, RI, 2008. ISBN 978-0-8218-4695-7. doi: 10.1090/mbk/059. URL <http://dx.doi.org/10.1090/mbk/059>. Pages from year one of a mathematical blog.