STAT 206A: Polynomials of Random Variables | Grothendieck's Inequality

Lecture 19

Lecture date: Nov 22 | *Scribe: Ben Hough*

The scribes are based on a guest lecture by Ryan O'Donnell. In this lecture we prove Grothendieck's inequality, which states that for all real $n \times n$ matrices there is a universal constant $K < \infty$ (independent of n) such that

$$\sup_{\substack{x_1,\ldots,x_n\in[-1,1]\\ y_1,\ldots,y_n\in[-1,1]}} \sum_{i,j=1}^{n} A_{i,j}x_iy_j \geq \frac{1}{K} \sup_{\substack{\vec{x}_1,\ldots,\vec{x}_n\in B_d\\ \vec{y}_1,\ldots,\vec{y}_n\in B_d}} \sum_{i,j=1}^{n} A_{i,j}\vec{x}_i \cdot \vec{y}_j, \tag{1}$$

where $B_d = \{z \in \mathbb{R}^d : \|z\| \leq 1\}$ is the ball of radius 1 in $\mathbb{R}^d$. The optimal constant K for which equation (1) holds is called the *Grothendieck constant*, K_G. We observe that the suprema in (1) are obtained when $x_i, y_j \in \{-1, 1\}$, and $\|\vec{x}_i\|, \|\vec{y}_j\| = 1$, so it is equivalent to replace $[-1, 1]$ with $\{-1, 1\}$ and B_d with $S_d = \{z \in \mathbb{R}^d : \|z\| = 1\}$ in (1). Moreover, by taking the vectors $\vec{x}_i, \vec{y}_j$ to all lie along a common line in $\mathbb{R}^d$, we see that the reverse inequality to equation (1) holds with $K = 1$.

1 Why do we care?

One motivation for studying Grothendieck's inequality is that it provides a method for rapidly approximating the cut norm of a matrix.

Definition 1 *The cut norm of a matrix A is defined to be*

$$\|A\|_c = \max_{\substack{I\subset[n]\\ J\subset[n]}} \left|\sum_{\substack{i\in I\\ j\in J}} A_{i,j}\right| = \max_{x_i,y_j\in\{0,1\}} \left|\sum_{i,j=1}^{n} A_{i,j}x_iy_j\right|. \tag{2}$$

Here $[n]$ denotes $\{1, 2, \ldots, n\}$.

Let us denote the left hand side of (1) by $\|A\|_{\infty\to 1}$. An easy computation verifies that

Exercise 2 $\|A\|_c \leq \|A\|_{\infty\to 1} \leq 4\|A\|_c$.

The right hand side of (1) can be computed in polynomial time using semi-definite programming methods. On the other hand, computing the cut norm of a matrix is NP-hard.

Grothendieck's constant is known not to be too big, so this approach provides a reasonable method for estimating the cut norm.

Grothendieck's inequality also has important consequences in Banach spaces. Indeed, it implies the following:

Corollary 3 *Every bounded linear operator T from an L^1 space X to a Hilbert space is absolutely summable. That is, if $x_1, x_2, \cdots \in X$ have the property that $\sum_i x_i$ converges regardless of re-arrangement, then $\sum_i \|Tx_i\| < \infty$.*

2 Known bounds on K_G

When Grothendieck originally proved the inequality in 1956, he showed that $1.57 \sim \frac{\pi}{2} \leq K_G \leq \sinh\left(\frac{\pi}{2}\right) \sim 2.3$. Riesz ('74) improved the upper bound slightly to 2.26, and Krivine ('79) showed that

$$K_G \leq \frac{\pi}{2} \frac{1}{\sinh^{-1}} \sim 1.78, \tag{3}$$

and conjectured that this value is correct. A.M. Davie ('84, unpublished) improved the lower bound. Independently, Jim Reeds (c'91, unpublished) improved the lower bound to approximately 1.677. We shall discuss Reeds' proof below.

3 Proof of lower bound for K_G

Observe that

$$\sup_{\vec{x}_i, \vec{y}_j \in B_d} \sum_{i,j=1}^{n} A_{i,j} \vec{x}_i \cdot \vec{y}_j = \sup_{\substack{f:[n]\to B_d \\ g:[n]\to B_d}} \sum_{i=1}^{n} \left\langle f(i), \sum_{j=1}^{n} A_{i,j} g(j) \right\rangle. \tag{4}$$

This quantity is clearly increasing in d, we are interested in understanding by how much it can increase as d increases from 1 to ∞. Scaling the matrix A (which has no influence on the Grothendieck's inequality), the above may be expressed as

$$\sup_{f,g:[n]\to B_d} \int_{[n]} \langle f, Ag \rangle \, d\mu \tag{5}$$

where μ is the uniform measure on $[n]$. For any value of n, this quantity equals

$$\sup_{f,g:\mathbb{R}^n \to B_d} \int_{\mathbb{R}^n} \langle f, Ag \rangle \, d\gamma = \sup_{f:\mathbb{R}^n \to B_d} \int_{\mathbb{R}^n} \|Ag(x)\| d\gamma(x). \tag{6}$$

where γ denotes the n-dimensional Gaussian distribution, and A is now an appropriate linear operator on functions in Gaussian space. Conversely, as $n \to \infty$, any linear operator on Gaussian space can be approximated arbitrarily well by discrete operators, hence it suffices to consider by how much (6) can increase as d increases from 1 to ∞. In particular, to obtain a good lower bound for K_G, we shall exhibit an operator A for which (6) is large as $d \to \infty$, and small for $d = 1$.

To define this operator A, we first recall that any $g : \mathbb{R}^n \to B_d$ has a Hermite expansion. That is:

$$g(x) = \sum_{S \in \mathbb{N}^n} \hat{g}(S) H_S(x) \tag{7}$$

where $\hat{g}(S) \in B_d$ and $H_S(x) = \prod_{i=1}^n h_{S_i}(x_i)$ where h_j are the 1-dimensional Hermite polynomials. It is easy to show that the H_S's are orthonormal:

$$\int_{\mathbb{R}^n} H_S(x) H_T(x) = \begin{cases} 1 & \text{if} \quad S = T \\ 0 & \text{if} \quad S \neq T \end{cases} \tag{8}$$

Moreover, $H_{\vec{0}} = 1$ and if e_i denotes the i^{th} standard basis vector in $\mathbb{R}^n$, then $H_{e_i} = x_i$. Let us refer to $|S| = \sum S_i$ as the "level of S", and define P_1 to be the "projection to level 1" operator. That is, $P_1 g(x) = \sum_{i=1}^n \hat{g}(e_i) x_i$. Define $A = P_1 - \lambda \cdot \mathbf{1}$, where $\mathbf{1}$ denotes the identity operator, and we shall choose λ later. Our first claim gives a lower bound for (6) for large d (note we take $n = d$ here for convenience).

Proposition 4

$$\sup_{g:\mathbb{R}^n \to B_n} \int \|Ag\| \geq 1 - \lambda - O\left(\frac{1}{n}\right). \tag{9}$$

Proof: To prove this lower bound, it suffices to exhibit a function g for which the claimed lower bound holds. Set $g(x) = \frac{x}{\|x\|}$ and compute

$$(P_1 g)(x) = \sum_{i=1}^n x_i \hat{g}(e_i). \tag{10}$$

Now we have that

$$\begin{aligned} \hat{g}(e_i) &= \int_{y \in \mathbb{R}^n} y_i g(y) d\gamma(y) & (11) \\ &= \int_{y \in \mathbb{R}^n} y_i \frac{y}{\|y\|} d\gamma(y), & (12) \end{aligned}$$

and a straightforward computation shows that

$$\int_{\mathbb{R}^n} \frac{y_i y_j}{\|y\|} d\gamma(y) = \begin{cases} 0 & \text{if} \quad i \neq j \\ a_n := \frac{1}{n} \int_{\mathbb{R}^n} \|x\| d\gamma(x) & \text{if} \quad i = j \end{cases}. \tag{13}$$

Therefore, $(P_1 g)(\vec{x}) = a_n \vec{x}$ and we compute

$$\begin{aligned}
\int_{\mathbb{R}^n} \|Ag\| d\gamma(\vec{x}) &= \int_{\mathbb{R}^n} \|(P_1 g)(\vec{x}) - \lambda g(x)\| d\gamma(\vec{x}) && (14)\\
&= \int_{\mathbb{R}^n} \|a_n \vec{x} - \frac{\lambda \vec{x}}{\|\vec{x}\|}\| d\gamma(\vec{x}) && (15)\\
&= \int_{\mathbb{R}^n} (a_n - \frac{\lambda}{\|x\|})\|\vec{x}\| d\gamma(\vec{x}) && (16)\\
&= a_n \int_{\mathbb{R}^n} \|\vec{x}\| - \lambda && (17)\\
&= \frac{1}{n}\left(\int_{\mathbb{R}^n} \|\vec{x}\|\right)^2 - \lambda && (18)
\end{aligned}$$

□

Theorem 5 *The quantity*

$$\sup_{g:\mathbb{R}^n \to \{-1,1\}} \int_{\mathbb{R}^n} \|Ag\| d\gamma(\vec{x}) \qquad (19)$$

is maximized by a function g' of the following form:

1. *$g'(x)$ depends only on x_1.*
2. *$g'(x)$ is an odd function of x_1, and there exists a constant a so that $g(\vec{x}) = 1$ for $x_1 > a$ and $g(\vec{x}) = -1$ for $0 \leq x_1 \leq a$.*

Remark 6 *Granting Theorem 5, one can solve for the optimal constants a and λ and derive a lower bound for K_G. The resulting expression is ugly, so we won't repeat the computation here.*

Proof: If $g : \mathbb{R}^n \to \{-1, 1\}$, then $P_1 g$ is linear so $(P_1 g)(x) = \sum_{i=1}^n b_i x_i$ for some constant b_i. Rotating space appropriately, we may assume that $\vec{b} = (b, 0, \ldots, 0)$ for some $0 \leq b \leq 1$. This rotation will not influence the integral $\int_{\mathbb{R}^n} \|Ag\| d\gamma(\vec{x})$, since γ is rotationally symetric. Define $\alpha : \mathbb{R} \to [-1, 1]$ by $\alpha(z) = \mathbf{E}[g|P_1 g = z]$, and observe that $-1 \leq \alpha(z) \leq 1$ for all z, and

$$\mathbf{P}[g = 1|P_1 g = z] = \frac{1}{2} + \frac{1}{2}\alpha(z) \qquad (20)$$

$$\mathbf{P}[g = -1|P_1 g = z] = \frac{1}{2} - \frac{1}{2}\alpha(z). \qquad (21)$$

Define

$$\sigma = \hat{g}(e_1) = \int_{\mathbb{R}^n} x_1 g(\vec{x}) d\gamma(\vec{x}) = \int_{\mathbb{R}} z\alpha(z)\varphi(z) dz \qquad (22)$$

where $\varphi(z)$ is the density of a 1-dimensional Gaussian random variable. We compute:

$$\begin{aligned}
\int_{\mathbb{R}^n} |Ag| d\gamma(\vec{x}) &= \int_{\mathbb{R}^n} |(P_1 g)(\vec{x}) - \lambda g(\vec{x})| d\gamma(\vec{x}) && (23)\\
&= \int_{\mathbb{R}} [(\frac{1}{2} + \frac{1}{2}\alpha(z))|\sigma z - \lambda| + (\frac{1}{2} - \frac{1}{2}\alpha(z))|\sigma z + \lambda|]\varphi(z) dz && (24)\\
&= \int_{\mathbb{R}} \frac{1}{2}(|\sigma z - \lambda| + |\sigma z + \lambda|)\varphi(z) dz + \int_{\mathbb{R}} \alpha(z)\psi(z)\varphi(z) && (25)
\end{aligned}$$

where $\psi(z) = \frac{1}{2}(|\sigma z - \lambda| - |\sigma z + \lambda|)$.

To prove our theorem, we assume that σ is fixed, and consider the following linear program:

LP_σ: Maximize:

$$\text{Const.} + \int_{-\infty}^{\infty} \psi(z)\varphi(z)\alpha(z) dz \qquad (26)$$

subject to the constraints:

$$-1 \le \alpha(z) \le 1 \text{ for all } z \ge 0 \quad \text{and} \quad \sigma = \int_{-\infty}^{\infty} z\varphi(z)\alpha(z) dz. \qquad (27)$$

Since $\varphi(z)$ is even and $\psi(z)$ is odd, we see that there will be an optimizing α which is odd (replace $\alpha(z)$ with $(\frac{1}{2}(\alpha(z) - \alpha(-z))$ if necessary). Now, since $\frac{\psi(z)}{z}$ is strictly increasing on $[0, \infty]$ we see that the optimal α will satisfy $\alpha(z) = 1$ for $z > a$ and $\alpha(z) = -1$ for $0 \le z \le a$. (If $\alpha(z)$ is not of this form, then we can perturb α slightly to increase the value of $\int_{-\infty}^{\infty} \psi(z)\varphi(z)\alpha(z) dz$ while still satisfying the constraints). It now follows from (20) and (21) that g must have the same form. $\square$