Grothendieck's Inequality

Leqi Zhu

1 Introduction

Let $A = (A_{ij}) \in \mathbb{R}^{m \times n}$ be an $m \times n$ matrix. Then A defines a linear operator between normed spaces $(\mathbb{R}^m, \| \cdot \|_p)$ and $(\mathbb{R}^n, \| \cdot \|_q)$, for $1 \leq p, q \leq \infty$. The $(p \to q)$-*norm* of A is the quantity $\|A\|_{p \to q} = \max_{x \in \mathbb{R}^n : \|x\|_p = 1} \|Ax\|_q$. (Recall that, for a vector $x = (x_i) \in \mathbb{R}^d$, the p-*norm* of x is $\|x\|_p = (\sum_i |x_i|^p)^{1/p}$; the ∞-norm of x is $\|x\|_\infty = \max_i |x_i|$.) If $p = q$, then we denote the norm by $\|A\|_p$.

For what value of p and q is $\|A\|_{p \to q}$ maximized? Since A is linear, it suffices to consider p such that $\{x \in \mathbb{R}^n : \|x\|_p \leq 1\}$ contains as many points as possible. We also want $\|Ax\|_q$ as large as possible. Figure 1 gives an illustration of $\{x \in \mathbb{R}^2 : \|x\|_p = 1\}$, for $p \in \{1, 2, \infty\}$. Going by the figure, $\|A\|_{\infty \to 1} \geq \|A\|_{p \to q}$.

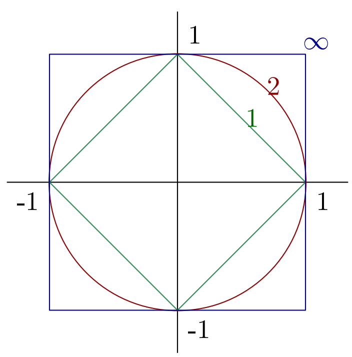

Figure 1: Depiction of $\{x \in \mathbb{R}^2 : \|x\|_p = 1\}$ for $p \in \{1, 2, \infty\}$.

Besides providing an upper bound on any $(p \to q)$-norm, it is known that the $(\infty \to 1)$-norm provides a constant approximation to the *cut norm* of a matrix, $\|A\|_C = \max_{S \subseteq [m], T \subseteq [n]} \max_{i \in S, j \in T} A_{ij}$, which is closely related to the MAX-CUT problem on a graph.

One way to compute $\|A\|_{\infty \to 1}$ is by solving a quadratic integer program:

$$\begin{aligned} \max \quad & \sum_{i,j} A_{ij} x_i y_j \\ \text{s.t.} \quad & (x, y) \in \{-1, 1\}^{m+n} \end{aligned}$$

To see this, note that $\sum_{i,j} A_{ij}x_iy_j = \sum_i (Ay)_i x_i$. Taking the maximum over $x \in \{-1,1\}^m$ gives us $\|Ay\|_1$. Then taking the maximum over $y \in \{-1,1\}^n$ gives us $\|A\|_{\infty\to 1}$ (this requires an argument, but it follows from the convexity of $\{x \in \mathbb{R}^m : \|x\|_\infty = 1\}$ and triangle inequality). This quadratic integer program may be relaxed to the following semidefinite program:

$$\begin{aligned} \max \quad & \sum_{i,j} A_{ij} \langle x^{(i)}, y^{(j)} \rangle \\ \text{s.t.} \quad & x^{(1)}, \dots x^{(m)}, y^{(1)}, \dots, y^{(n)} \\ & \text{are unit vectors in } (\mathbb{R}^d, \|\cdot\|_2) \end{aligned}$$

Notice that, if $d = 1$, then we have exactly the same optimization problem.

It is known that exactly computing $\|A\|_{p\to q}$, for $1 \le q < p \le \infty$, is NP-hard, while exactly computing $\|A\|_p$ is NP-hard for $p \notin \{1, 2, \infty\}$. (As far as I can tell, it is an open question whether or not $\|A\|_{p\to q}$ is computable in polynomial time when $1 \le p < q \le \infty$.) Hence, if $d = 1$, we cannot hope for the ellipsoid method to converge quickly on all instances of A. However, there are no hardness results for $d > 1$. Thus, in principle, the ellipsoid method could converge quickly.

A natural questions is: how well does an optimal solution to the semidefinite program approximate $\|A\|_{\infty\to 1}$? Grothendieck's Inequality provides an answer to this question:

Theorem (Grothendieck's Inequality)**.** *There exists a fixed constant $C > 0$ such that, for all $m, n \ge 1$, $A \in \mathbb{R}^{m\times n}$, and any Hilbert space H (vector space over $\mathbb{R}$ with an inner product),*

$$\max_{\substack{\text{unit vectors} \\ x^{(i)}, y^{(j)} \in H}} \sum_{i,j} A_{ij} \langle x^{(i)}, y^{(j)} \rangle_H \le C \|A\|_{\infty\to 1} .$$

Grothendieck's constant is the smallest C such that the above inequality holds. As far as I can tell, determining the exact value of Grothendieck's constant is an open problem. However, it is known that it lies between $\frac{\pi}{2} \approx 1.57$ and $K = \frac{\pi}{2\ln(1+\sqrt{2})} \approx 1.78$.

Hence, the value of an optimal solution to the semidefinite program provides a constant approximation of $\|A\|_{\infty\to 1}$. However, this is a bit unsatisfying because, given an optimal solution to the semidefinite program, we do not know how to round the solution to obtain an integer solution $(x, y) \in \{-1,1\}^{m+n}$ with a good approximation ratio.

Alon and Naor resolved this problem by rather nicely by adapting Krivine's proof of Grothendieck's Inequality, which obtains the upper bound K on Grothendieck's constant, to obtain a randomized rounding method:

Theorem (Alon and Naor)**.** *For $d = m + n$, given an optimal solution to the semidefinite program, $x^{(i)}, y^{(j)} \in \mathbb{R}^{m+n}$, it is possible to obtain $(x, y) \in \{-1,1\}^{m+n}$ (using randomized rounding) such that*

$$\mathbb{E}\left[\sum_{i,j} A_{ij}x_iy_j\right] = \frac{1}{K}\sum_{i,j} A_{ij}\langle x^{(i)}, y^{(j)}\rangle \ge \frac{1}{K}\|A\|_{\infty\to 1} \approx 0.56\|A\|_{\infty\to 1}.$$

It turns out that Krivine's proof can also be adapted to prove a theorem about degree-2 pseudo-distributions $\mu : \{-1,1\}^{m+n} \to \mathbb{R}$. Recall that μ has to satisfy two properties for the pseudo-expectation $\tilde{\mathbb{E}}_\mu$ that arises from μ: $\tilde{\mathbb{E}}_\mu[1] = \sum_{x\in\{-1,1\}^{m+n}} \mu(x) = 1$ and $\tilde{\mathbb{E}}_\mu[f^2] = \sum_{x\in\{-1,1\}^{m+n}} \mu(x)(f(x))^2 \geq 0$ for all degree-1 polynomials $f : \{-1,1\}^{m+n} \to \mathbb{R}$.

Theorem (SOS)**.** *For any degree-2 pseudo-distribution $\mu : \{-1,1\}^{m+n} \to \mathbb{R}$,*

$$\tilde{\mathbb{E}}_{\mu(x,y)}\left[\sum_{i,j} A_{ij}x_iy_j\right] \leq K\|A\|_{\infty\to 1}$$

In this note, we will prove Grothendieck's Inequality when $H = \mathbb{R}^{m+n}$. The proof is mainly due to Krivine. However, we use a nice simplification of a key lemma in Krivine's proof (which holds for general H), due to Alon and Naor. This will provide us with the tools to prove Alon and Naor's theorem. Finally, we will discuss the connection to SOS and how to prove the SOS theorem.

2 Grothendieck's Inequality

Krivine's proof of Grothendieck's Inequality relies on Grothendieck's Identity, which, as the name suggests, was first proved by Grothendieck:

Lemma (Grothendieck's Identity)**.** *Let x and y be unit vectors in $(\mathbb{R}^d, \|\cdot\|_2)$, where $d \geq 2$. If z is a unit vector picked uniformly at random from $(\mathbb{R}^d, \|\cdot\|_2)$, then*

$$\mathbb{E}[\mathrm{sign}(\langle x,z\rangle)\mathrm{sign}(\langle y,z\rangle)] = \frac{2}{\pi}\arcsin(\langle x,y\rangle)\,.$$

Here, $\mathrm{sign}(a) \in \{-1,1\}$ *is 1 if and only if $a \geq 0$.*

Proof. Consider $\mathrm{sign}(\langle x,z\rangle)\mathrm{sign}(\langle y,z\rangle)$. This has a nice geometric interpretation. First, we orient the sphere $\{x \in \mathbb{R}^d : \|x\|_2 = 1\}$ so that z is at the top. It can be verified that $\mathrm{sign}(\langle x,z\rangle)\mathrm{sign}(\langle y,z\rangle)$ is 1 if and only if both x and y lie in the same (upper or lower) half of the sphere when it is oriented this way. Equivalently, $\{x \in \mathbb{R}^d : \langle z,x\rangle = 0\}$ is a hyperplane passing through the origin (with normal z). A vector $x \in \mathbb{R}^d$ satisfies $\langle x,z\rangle > 0$ if and only if it lies above the hyperplane. Figure 2 contains a depiction of this.

Now, consider the expectation. Given the geometric interpretation, the expectation is $\Pr[x,y \text{ lie in same half}] - \Pr[x,y \text{ lie in different halves}] = 1 - 2\Pr[x,y \text{ lie in different halves}]$, when a random hyperplane passing through the origin is selected (with normal z). Then we note that the probability x and y lie in different halves of the circle is $\frac{2\theta}{2\pi}$, where θ is the angle between x and y (factor of 2 comes from z and $-z$ defining same hyperplane). Hence, the expectation is $1 - \frac{2\theta}{\pi}$. On the other hand, $\frac{2}{\pi}\arcsin(\langle x,y\rangle) = \frac{2}{\pi}\arcsin(\cos\theta) = \frac{2}{\pi}\arcsin(\sin(\frac{\pi}{2}-\theta)) = \frac{2}{\pi}(\frac{\pi}{2}-\theta) = 1 - \frac{2\theta}{\pi}$. □

This doesn't appear to help much as we don't know what $\arcsin(\langle x,y\rangle)$ is. The next lemma addresses this problem.

$\mathrm{sign}(\langle x, z\rangle)\mathrm{sign}(\langle y, z\rangle) = 1$ $\quad$ $\mathrm{sign}(\langle x, z\rangle)\mathrm{sign}(\langle y, z\rangle) = -1$

Figure 2: Geometric intepretation of $\mathrm{sign}(\langle x, z\rangle)\mathrm{sign}(\langle y, z\rangle)$.

Lemma (Krivine/Alon and Naor). *Suppose that $x^{(i)}, y^{(j)}$ are unit vectors in $\mathbb{R}^{m+n}$, for $i \in [m]$, $j \in [n]$. Then there are unit vectors $\hat{x}^{(i)}, \hat{y}^{(j)}$ in $\mathbb{R}^{m+n}$, for $i \in [m]$, $j \in [m]$, such that*

$$\arcsin(\langle \hat{x}^{(i)}, \hat{y}^{(j)}\rangle) = \ln(1+\sqrt{2})\langle x^{(i)}, y^{(j)}\rangle\,.$$

Proof. Let $c = \ln(1+\sqrt{2})$ and $d = m+n$. By Taylor's expansion,

$$\sin(c\langle x^{(i)}, y^{(j)}\rangle) = \sum_{k=0}^{\infty}(-1)^k \frac{c^{2k+1}}{(2k+1)!}(\langle x^{(i)}, y^{(j)}\rangle)^{2k+1}\,.$$

Our goal is to write the above as the inner product of two vectors in some vector space. This suggests that we need an infinite dimensional vector space. Towards this end, consider the infinite-dimensional vector space H obtained by taking the direct product of $2k+1$ tensor powers of $\mathbb{R}^d$, i.e. $H = \oplus_{k=0}^{\infty}(\mathbb{R}^d)^{\otimes 2k+1}$. (As a bit of an aside, the direct sum of two vector spaces A and B of dimension α and β, respectively, is a vector space $A \oplus B$ of dimension $\alpha+\beta$; given vectors $a \in A$, $b \in B$, we get the vector $a \oplus b = (a_1, \ldots, a_\alpha, b_1, \ldots, b_\beta)$. Similarly, the tensor product of A and B, $A \otimes B$, gives a vector space of dimension $\alpha\beta$; given vectors $a \in A$ and $b \in B$, we get the vector $a \otimes b = (a_i b_j)_{i,j}$.)

Let $X^{(i)}$ and $Y^{(j)}$ be vectors in H with the following "coordinates" for the k'th part in the direct sum:

$$X_k^{(i)} = (-1)^k\sqrt{\frac{c^{2k+1}}{(2k+1)!}}(x^{(i)})^{\otimes(2k+1)}$$

$$Y_k^{(j)} = (-1)^k\sqrt{\frac{c^{2k+1}}{(2k+1)!}}(y^{(j)})^{\otimes(2k+1)}$$

It is a fact that $\langle a^{\otimes(2k+1)}, b^{\otimes(2k+1)}\rangle = (\langle a, b\rangle)^{2k+1}$. Hence, $\langle X^{(i)}, Y^{(j)}\rangle = \sin(c\langle x^{(i)}, y^{(j)}\rangle)$, as required. Moreover, it can be verified that $\langle X^{(i)}, X^{(i)}\rangle = \sinh(c\langle x^{(i)}, x^{(i)}\rangle) = \sinh(c) = 1$, by appealing to the Taylor's expansion of $\sinh(x) = \frac{1}{2}(e^x - e^{-x})$ and using the preceding fact. Similarly, $\langle Y^{(j)}, Y^{(j)}\rangle = 1$. It follows that $X^{(i)}$ and $Y^{(j)}$ are unit vectors in H.

Consider the span, S, of $\{X^{(i)}, Y^{(j)}\}$. As there are only $d = m + n$ vectors, S is isomorphic a subspace in $\mathbb{R}^d$. By finding an orthonormal basis for S (for example, using Gram-Schmidt) and mapping the basis to the standard basis for $\mathbb{R}^{m+n}$, we can preserve inner products. Thus, $X^{(i)}, Y^{(j)}$ correspond to unit vectors $\hat{x}^{(i)}, \hat{y}^{(j)}$ in $\mathbb{R}^d$ with the same inner product (in H and $\mathbb{R}^d$, respectively). It follows that $\arcsin(\langle \hat{x}^{(i)}, \hat{y}^{(j)} \rangle) = \arcsin(\langle X^{(i)}, Y^{(j)} \rangle) = c\langle x^{(i)}, y^{(j)} \rangle = \ln(1 + \sqrt{2})\langle x^{(i)}, y^{(j)} \rangle$, as required. $\square$

Proof of Grothendieck's Inequality. We may now prove Grothendieck's Inequality when $H = \mathbb{R}^{m+n}$. For $x^{(i)}, y^{(j)}$ that maximizes $\sum_{i,j} A_{ij} \langle x^{(i)}, y^{(j)} \rangle$, we first apply Krivine/Alon and Naor's lemma to obtain $\hat{x}^{(i)}, \hat{y}^{(j)}$. Define random variables $\hat{x}_i = \mathrm{sign}(\langle \hat{x}^{(i)}, z \rangle)$ and $\hat{y}_j = \mathrm{sign}(\langle \hat{y}^{(j)}, z \rangle)$, where z is a unit vector in $\mathbb{R}^{m+n}$ chosen uniformly at random. We may then compute:

$$\begin{aligned}
\mathbb{E}\left[\sum_{i,j} A_{ij} \hat{x}_i \hat{y}_j\right] &= \sum_{i,j} A_{ij} \mathbb{E}\left[\mathrm{sign}(\langle \hat{x}^{(i)}, z \rangle) \mathrm{sign}(\langle \hat{y}^{(j)}, z \rangle)\right] \\
&= \frac{2}{\pi} \sum_{i,j} A_{ij} \arcsin(\langle \hat{x}^{(i)}, \hat{y}^{(j)} \rangle) \\
&= \frac{2 \ln(1 + \sqrt{2})}{\pi} \sum_{i,j} A_{ij} \langle x^{(i)}, y^{(j)} \rangle \\
&= \frac{1}{K} \max_{\text{unit } x^{(i)}, y^{(j)} \in \mathbb{R}^{m+n}} \sum_{i,j} A_{i,j} \langle x^{(i)}, y^{(j)} \rangle
\end{aligned}$$

As $(\hat{x}, \hat{y}) \in \{-1, 1\}^{m+n}$, this is at most $\max_{x_i, y_j \in \{-1,1\}} \sum_{i,j} A_{ij} x_i y_j = \|A\|_{\infty \to 1}$. Grothendieck's Inequality immediately follows.

3 Alon and Naor's Theorem

Alon and Naor's rounding algorithm is as follows:

1. Compute the optimal solution $x^{(i)}, y^{(j)}$ of the semidefinite program.
2. Find $\hat{x}^{(i)}, \hat{y}^{(j)}$ such that $\arcsin(\langle \hat{x}^{(i)}, \hat{y}^{(j)} \rangle) = \ln(1 + \sqrt{2})\langle x^{(i)}, y^{(j)} \rangle$.
3. Pick a unit vector z uniformly at random from $\mathbb{R}^{m+n}$.
4. Set $\hat{x}_i = \mathrm{sign}(\langle \hat{x}^{(i)}, z \rangle)$ and $\hat{y}_j = \mathrm{sign}(\langle \hat{y}^{(j)}, z \rangle)$.

Then the same calculation as in the preceding section gives us that:

$$\mathbb{E}\left[\sum_{i,j} A_{ij} \hat{x}_i \hat{y}_j\right] = \frac{1}{K} \max_{\text{unit } x^{(i)}, y^{(j)} \in \mathbb{R}^{m+n}} \sum_{i,j} A_{i,j} \langle x^{(i)}, y^{(j)} \rangle \rangle \,.$$

Alon and Naor's theorem follows.

Perhaps the final detail to address is how to find $\hat{x}^{(i)}, \hat{y}^{(j)}$. We may do so with the following semidefinite program:

$$\begin{aligned} \min \quad & \sum_{i,j} A_{ij} \langle \hat{x}^{(i)}, \hat{y}^{(j)} \rangle - \ln(1+\sqrt{2}) \sum_{i,j} A_{ij} \langle x^{(i)}, y^{(j)} \rangle \\ \text{s.t.} \quad & \hat{x}^{(1)}, \dots \hat{x}^{(m)}, \hat{y}^{(1)}, \dots, \hat{y}^{(n)} \\ & \text{are unit vectors in } (\mathbb{R}^d, \|\cdot\|_2) \end{aligned}$$

By Krivine's lemma, the optimal value is 0.

4 Connection to SOS

Let $\mu : \{-1,1\}^{m+n} \to \mathbb{R}$ be an arbitrary degree-2 pseudo-distribution. Recall that we aim to prove that $\tilde{\mathbb{E}}_{\mu(x,y)}[\sum_{i,j} A_{ij} x_i y_j] \le K\|A\|_{\infty\to 1}$, for all such μ. To see the similarity between this statement and Grothendieck's Inequality, we appeal to the fact that, by considering $\mu'(w) = \frac{1}{2}(\mu(w) + \mu(-w))$, which has the same pseudo-expectation as μ on $\sum_{i,j} A_{ij} x_i y_j$, we may assume that $\tilde{\mathbb{E}}_{\mu(x,y)}[x_i] = \tilde{\mathbb{E}}_{\mu(x,y)}[y_j] = 0$. Moreover, as $(x,y) \in \{-1,1\}^{m+n}$, $\tilde{\mathbb{E}}_{\mu(x,y)}[x_i^2] = \tilde{\mathbb{E}}_{\mu(x,y)}[y_j^2] = \tilde{\mathbb{E}}_{\mu(x,y)}[1] = 1$. By the Quadratic Sampling Lemma (see lecture notes or Boaz's notes), there exists a joint normal probability distribution $\rho : \mathbb{R}^{m+n} \to \mathbb{R}$ that has the same first two moments as μ; i.e. the same mean (entry-wise) and covariance matrix (here we consider the "formal" covariance matrix under pseudo-expectation of μ). Geometrically, sampling a point in $\mathbb{R}^{m+n}$ according to ρ is essentially sampling a point from $\{x \in \mathbb{R}^{m+n} : \|x\|_2 \le 1\}$, the unit ball in $\mathbb{R}^{m+n}$. Hence, roughly speaking, we have:

$$\begin{aligned} \mathbb{E}_{(u,v)\sim\rho}\left[\sum_{i,j} A_{ij} u_i v_j\right] &\le \mathbb{E}_{x^{(i)},y^{(j)}\sim\rho}\left[\sum_{i,j} A_{ij} \langle x^{(i)}, u^{(j)} \rangle\right] \\ &\approx \mathbb{E}_{\substack{x^{(i)},y^{(j)}\in\mathbb{R}^{m+n}: \\ \|x^{(i)}\|_2, \|y^{(j)}\|_2 \le 1}}\left[\sum_{i,j} A_{ij} \langle u^{(i)}, v^{(j)} \rangle\right] \\ &\le \max_{\substack{x^{(i)},y^{(j)}\in\mathbb{R}^{m+n}: \\ \|x^{(i)}\|_2, \|y^{(j)}\|_2 \le 1}} \sum_{i,j} A_{ij} \langle x^{(i)}, y^{(j)} \rangle \\ &= \max_{\substack{x^{(i)},y^{(j)}\in\mathbb{R}^{m+n}: \\ \|x^{(i)}\|_2, \|y^{(j)}\|_2 = 1}} \sum_{i,j} A_{ij} \langle x^{(i)}, y^{(j)} \rangle \\ &\le K\|A\|_{\infty\to 1} \end{aligned}$$

The fourth line follows since the maximum value of the semidefinite program is achieved on the boundary (i.e. we could have relaxed the constraint on $x^{(i)}$, $y^{(j)}$ to $\|x^{(i)}\|_2, \|y^{(j)}\|_2 \le 1$). As μ and ρ have the same first two moments, the pseudo-expectation of $\sum_{i,j} A_{ij} x_i y_j$ under μ is the same as the expectation under ρ, so we have the desired bound. Of-course, this is rather imprecise as sampling from ρ does not exactly correspond to sampling from $\{x \in \mathbb{R}^{m+n} : \|x\|_2 \leq 1\}$ (although it is true with high probability), hence the second line is only "$\approx$". But, hopefully this lends some intuition on why it could be true. In the remainder of this section, we formally prove this.

We first prove a modified version of Grothendieck's Identity.

Lemma. *Let $(u, v) \in \mathbb{R}^2$ be a joint normal distribution such that:*

$$(u, v) \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right).$$

The latter matrix is the covariance matrix. Notice that, by Cauchy-Schwartz, $-1 \leq \rho \leq 1$. Then

$$\mathbb{E}[\operatorname{sign}(u)\operatorname{sign}(v)] = \frac{2}{\pi}\arcsin\left(\mathbb{E}[uv]\right) = \frac{2}{\pi}\arcsin(\rho).$$

Proof. We may write $v = \rho u + \sqrt{1-\rho^2}w$, where $w \sim \mathcal{N}(0,1)$ is independent of u. (This is because $\mathbb{E}[u(\rho u+\sqrt{1-\rho^2}w)] = \mathbb{E}[\rho u^2+\sqrt{1-\rho^2}uw] = \rho$ as $\mathbb{E}[u^2] = 1$ and u and w are independent, hence $\mathbb{E}[uw] = \mathbb{E}[u]\mathbb{E}[w] = 0$.) It follows that $\mathbb{E}[\operatorname{sign}(u)\operatorname{sign}(v)] = 1 - 2\Pr_{u,w}[\operatorname{sign}(u) \neq \operatorname{sign}(\rho u + \sqrt{1-\rho^2}w)]$. We observe that $\operatorname{sign}(u) \neq \operatorname{sign}(v)$ if and only if $\operatorname{sign}(u) \neq \operatorname{sign}(w)$ and $|\rho u| < \sqrt{1-\rho^2}|w|$ (equivalently, $\rho^2 < \frac{w^2}{u^2+w^2}$, except when $u = w = 0$).

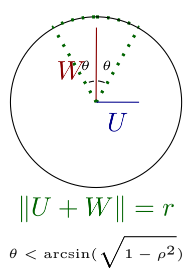

Figure 3: Geometric depiction of U, $U + W$, θ.

To compute the probability of this, we interpret this geometrically. First, as u and w are independent, u and w may be viewed as vectors $U = (u, 0)$, $W = (0, w)$ in $\mathbb{R}^2$, respectively. In this view, $\frac{w^2}{u^2+w^2} = (\cos\theta)^2$, where θ is the angle between W and $U + W$. As $u, w \sim \mathcal{N}(0,1)$ are independent, for any $r > 0$, if we do rejection sampling for (u, w) such that $u^2 + w^2 = r^2$, then this is the same as sampling a point uniformly at random from the circle of radius r in $\mathbb{R}^2$. Hence, $\Pr[\operatorname{sign}(u) \neq \operatorname{sign}(v)|u^2 + w^2 = r^2] = \Pr[(\cos(\theta))^2 > \rho^2|u^2 + w^2 = r^2] = \Pr[(\sin(\theta))^2 < 1 - \rho^2|u^2 + w^2 = r^2]$ is proportion of the arc $\|U + W\| = r$ such that $(\sin(\theta))^2 < 1 - \rho^2$, where θ is the angle between W and $U + W$. This is summarized in Figure 3. We may compute this to be $\frac{2\arcsin(\sqrt{1-\rho^2})r}{2\pi r} = \frac{1}{\pi}\arcsin(\sqrt{1-\rho^2})$. As $\sum_{r\geq 0}\Pr[u^2 + w^2 = r^2] = 1$, this gives us that $\Pr[\text{sign}(u) \neq \text{sign}(v)] = \frac{1}{\pi}\arcsin(\sqrt{1-\rho^2})$. It follows that $\mathbb{E}[\text{sign}(u)\text{sign}(v)] = 1 - \frac{2}{\pi}\arcsin(\sqrt{1-\rho^2}) = 1 - \frac{2}{\pi}\arccos(\rho) = 1 - \frac{2}{\pi}(\frac{\pi}{2} - \arcsin(\rho)) = \frac{2}{\pi}\arcsin(\rho)$. □

The preceding lemma implies that if $(u_1, \ldots, u_m, v_1, \ldots, v_m) \in \mathbb{R}^{m+n}$ is a joint normal distribution that satisfies $\mathbb{E}[u_i] = \mathbb{E}[v_j] = 0$ and $\mathbb{E}[u_i^2] = \mathbb{E}[v_j^2] = 1$, then, for all i, j,

$$\mathbb{E}[\text{sign}(u_i)\text{sign}(v_j)] = \frac{2}{\pi}\arcsin(\mathbb{E}[u_i v_j])\,.$$

To carry out the same calculation we did in the preceding section to prove Grothendieck's Inequality, we need

$$\arcsin(\mathbb{E}[u_i v_j]) = \ln(1+\sqrt{2})\tilde{\mathbb{E}}_{\mu(x,y)}[x_i y_j]\,,$$

which should look familiar from Krivine's lemma in the previous section. Put in another way, we need to pick the covariance of $u_i v_j$ according to $\tilde{\mathbb{E}}_{\mu(x,y)}[x_i y_j]$.

The next lemma proves exactly this:

Lemma. *There exists a joint normal distribution* $(u_1, \ldots, u_m, v_1, \ldots, v_n) \in \mathbb{R}^{m+n}$ *such that, for all* $i \in [m]$, $j \in [n]$,

- $\mathbb{E}[u_i] = \mathbb{E}[v_j] = 0$,
- $\mathbb{E}[u_i^2] = \mathbb{E}[v_j^2] = 1$, *and*
- $\arcsin(\mathbb{E}[u_i v_j]) = \ln(1+\sqrt{2})\tilde{\mathbb{E}}_{\mu(x,y)}[x_i y_j]$.

Proof. Let $\Sigma \in \mathbb{R}^{(m+n)\times(m+n)}$ be the matrix defined by $\Sigma_{ij} = \mathbb{E}_{\mu(w)}[w_i w_j]$, where $w = (x, y)$. In other words, Σ is the "formal" covariance matrix of (x, y) (here is where we use the assumption that $\Sigma_{ij} = \tilde{\mathbb{E}}_{\mu(x,y)}[x_i] = \tilde{\mathbb{E}}_{\mu(x,y)}[y_j] = 0$). By the Quadratic Sampling Lemma, Σ is positive semidefinite and symmetric, i.e. it actually defines a covariance matrix. Moreover, since $\tilde{\mathbb{E}}_{\mu(x,y)}[1] = 1$, $\tilde{\mathbb{E}}_{\mu(x,y)}[x_i^2] = \tilde{\mathbb{E}}_{\mu(x,y)}[1] = 1$ and, similarly, $\tilde{\mathbb{E}}_{\mu(x,y)}[y_j^2] = 1$.

If we extend the pseudo-expectation function so that it applies to matrices (in terms of x, y) entry-wise, i.e. , $(\tilde{\mathbb{E}}_{\mu(x,y)}[B])_{ij} = \tilde{\mathbb{E}}_{\mu(x,y)}[B_{ij}]$, then we may view Σ more compactly as:

$$\begin{bmatrix} \tilde{\mathbb{E}}_{\mu(x,y)}[xx^t] & \tilde{\mathbb{E}}_{\mu(x,y)}[xy^t] \\ \tilde{\mathbb{E}}_{\mu(x,y)}[yx^t] & \tilde{\mathbb{E}}_{\mu(x,y)}[yy^t] \end{bmatrix}.$$

Consider the matrix $\Sigma' \in \mathbb{R}^{(m+n)\times(m+n)}$ defined as follows:

$$\begin{bmatrix} \sinh\circ(\ln(1+\sqrt{2})\tilde{\mathbb{E}}_{\mu(x,y)}[xx^t]) & \sin\circ(\ln(1+\sqrt{2})\tilde{\mathbb{E}}_{\mu(x,y)}[xy^t]) \\ \sin\circ(\ln(1+\sqrt{2})\tilde{\mathbb{E}}_{\mu(x,y)}[yx^t]) & \sinh\circ(\ln(1+\sqrt{2})\tilde{\mathbb{E}}_{\mu(x,y)}[yy^t]) \end{bmatrix}$$

where sinh and sin is applied entry-wise to each submatrix. Since Σ is positive semidefinite and symmetric, it can be shown that the same holds for Σ', i.e. Σ' defines a covariance matrix. Moreover, since $\sinh(\ln(1+\sqrt{2})) = 1$ and $\tilde{\mathbb{E}}_{\mu(x,y)}[x_i^2] = \tilde{\mathbb{E}}_{\mu(x,y)}[y_j^2] = 1$, we have that the diagonal of Σ' are all 1's. It follows that we can pick (u, v) to be a joint normal distribution with $\mathbb{E}[u_i] = \mathbb{E}[v_j] = 0$ and covariance matrix Σ'. $\square$

Putting this all together, we have that:

$$\begin{aligned}
\|A\|_{\infty\to 1} &\geq \mathbb{E}_{(u,v)}\left[\sum_{i,j} A_{ij}\mathrm{sign}(u_i)\mathrm{sign}(v_j)\right] \\
&= \sum_{i,j} A_{ij}\mathbb{E}_{(u,v)}[\mathrm{sign}(u_i)\mathrm{sign}(v_j)] \\
&= \frac{2}{\pi}\sum_{i,j} A_{ij}\arcsin(\mathbb{E}_{(u,v)}[u_i v_j]) \\
&= \frac{1}{K}\sum_{i,j} A_{ij}\tilde{\mathbb{E}}_{\mu(x,y)}[x_i y_j]\,.
\end{aligned}$$

The SOS theorem follows.