CS 2429 - Foundations of Communication Complexity

Lecture #11: 24 November 2009

Lecturer: Avner Magen

Scribe Notes by: Wesley George

1 Unbounded Error Probabilistic Communication Complexity and Sign-Rank

Definition Let $f : \{0,1\}^n \times \{0,1\}^m \to \{0,1\}$. An *unbounded error protocol* for f, is a protocol P such that

$$\Pr[P(x,y) = f(x,y)] > 1/2$$

the probability is over the coin tosses of the players. The (worst-case) cost of P is the maximum number of bits sent during an execution of P taken over choices of x and y, the players inputs. The *unbounded error communication complexity of* f, denoted $\tilde{C}_f$, is the minimum cost of an unbounded error protocol for f.

Most functions have linear unbounded error communication complexity. This lecture is an exposition of Jürgen Forster's 2002 paper 'A linear lower bound on the unbounded error probabilistic communication complexity' from the Journal of Computer and System Sciences, which was the first to give an explicit function with linear unbounded error communication complexity.

We denote the set of n by m real-valued matrices by $\mathbb{R}^{n\times m}$, and the set of n by m ± 1-valued matrices by $\{+1,-1\}^{n\times m}$.

Definition Let $A \in \{+1,-1\}^{n\times m}$. $B \in \mathbb{R}^{n\times m}$ *sign-approximates* A if

$$\text{sign}(B_{i,j}) = A_{i,j} \text{ for each entry } i,j$$

The *sign-rank of* A is the minimum rank of a matrix B that sign-approximates A. I.e.,

$$\text{sign-rank}(A) = \min_{B\in\mathbb{R}^{n\times m}} \{\text{rank}(B) : \forall i,j \ \text{sign}(B_{i,j}) = A_{i,j}\}$$

For a function $f : \{0,1\}^n \times \{0,1\}^m \to \{0,1\}$, M_f is the 2^n by 2^m matrix whose i,jth entry is $2f(i,j) - 1$, i.e.

$$M_f = [2f(i,j) - 1]_{i,j}$$

The following theorem motivates our interest in the sign-rank of M_f:

Theorem 1 *For* $f : \{0,1\}^n \times \{0,1\}^m \to \{0,1\}$, *if* $r = \text{sign-rank}(M_f)$ *then*

$$\lceil \log r \rceil \le \tilde{C}_f \le \lceil \log r \rceil + 1$$

Theorem 1 will be proven in a later lecture; we concern ourselves with deriving a lower bound for the sign-rank.

Definition The *spectral norm*, of a matrix $A \in \mathbb{R}^{n\times m}$ is

$$\|A\| = \sup_{x\neq 0} \frac{\|Ax\|}{\|x\|}$$

Where $\|v\|$ is the standard Euclidean norm if v is a vector.

Theorem 2 *Let* $A \in \{+1, -1\}^{n\times m}$

$$\text{sign-rank}(A) \geq \frac{\sqrt{nm}}{\|A\|}$$

Corollary 3 *For* $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$,

$$\tilde{C}_f \geq n - \log_2 \|M_f\|$$

Example The Hadamard matrices $\{H_n\}_{n\in\mathbb{N}}$ are a sequence of matrices. H_n is 2^n by 2^n

$$H_1 = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} \qquad H_{n+1} = \begin{vmatrix} H_n & H_n \\ H_n & -H_n \end{vmatrix}$$

It is easy to see that distinct columns, x, y of H_n match in exactly half their entries and so the columns of H_n are orthogonal and $H_n^T H_n = 2^n I_{2^n}$ and so $\|H_n^T H_n\| = 2^n$.

$$\|H_n\| = \sqrt{\|H_n^T H_n\|} = 2^{n/2}$$

Let $f = IP_n$, i.e. $f(x,y) = \sum_{i=1}^n x_i y_i \mod 2$. M_f is the nth Hadamard matrix. So

$$\tilde{C}_{IP} \geq n - \log_2 \|H_n\| = n/2$$

So we have given a function (the inner product) whose unbounded error probabilistic communication complexity is at least $n/2$. The rest of the lecture is concerned with the proof Theorem 2, the heart of which is is linear algebra.

2 Background in Matrix Theory

In this section we recall some terminology and basics facts concerning matrices. Readers needing no such reminders may safely skip this section.

A set of vectors in $\mathbb{R}^k$ are said to be in *general position* if no k or fewer of them are linearly dependent. The set of eigenvalues of a matrix A is called its spectrum and denoted spec(A). We will denote the eigenspace of A associated with eigenvalue λ as $\mathcal{E}_\lambda(A)$.

Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. The spectral theorem for symmetric matrices states that there exists an orthonormal basis $d_1, ..., d_n$ of $\mathbb{R}^n$ and values $\lambda_1, ..., \lambda_n \in \mathbb{R}$ such that

$$A = \sum_{i=1}^n \lambda_i d_i d_i^T$$

Equivalently, $A = D\Lambda D^T$ where $D = [d_1|...|d_n]$ and $\Lambda = \mathrm{diag}(\lambda_1, ..., \lambda_n)$. $\{d_1, ..., d_n\}$ are the eigenvectors of A and $\{\lambda_1, ..., \lambda_n\} = \mathrm{spec}(A)$. A is said to be *positive semi-definite* (PSD) if

$$\forall x \in \mathbb{R}^n \quad x^T A x \geq 0$$

If A is PSD, not only is its spectrum real, but it is non-negative. For $i \in \{1, ..., n\}$ we have

$$\begin{aligned}
0 &\leq d_i^T A d_i \\
&= d_i^T \left(\sum_j \lambda_j d_j d_j^T \right) d_i \\
&= d_i^T \left(\sum_j \lambda_j \langle d_i, d_j \rangle d_j \right) \\
&= d_i^T \lambda_i \langle d_i, d_i \rangle d_i && \text{(Orthogonality of D)} \\
&= \lambda_i && (\|d_i\| = 1)
\end{aligned}$$

Note that for any $b \in \mathbb{R}^n$, bb^T is PSD. Also, if $\{A_1, ..., A_m\}$ are PSD matrices and $\{\alpha_1, ..., \alpha_m\}$ are positive real numbers then $M = \sum \alpha_i A_i$ is PSD.

Fix a symmetric matrix A (so $\mathrm{spec}(A) \subset \mathbb{R}$) and a candidate spectrum lower-bound $\lambda_0 \in \mathbb{R}$. Let $B = A - \lambda_0 I_n$. Let λ_m be A's smallest eigenvalue and v an associated eigenvector. If $\lambda_m < \lambda_0$, then

$$v^T B v = v^T A v - v^T \lambda_0 v = (\lambda_m - \lambda_0)\|v\|^2 < 0$$

so B is not PSD.

The Spectral Norm

We recall the definition of the spectral norm of a matrix $A \in \mathbb{R}^{n \times m}$

$$\|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|} \tag{1}$$

This is just the operator norm for $\mathbb{R}^{n \times m}$ induced by the Euclidean norm on $\mathbb{R}^n$. It can be shown with little effort that the function defined in 1 is in fact a norm, and thus induces a topology on $\mathbb{R}^{n \times m}$ via the metric $p(A, B) = \|A - B\|$ makes $\mathbb{R}^{n \times m}$ so we can talk about convergence of sequences of matrices, and continuity of functions acting on a set of matrices.

First note that

$$\|A\| \stackrel{def}{=} \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \sup_{\|x\| \leq 1} \|Ax\| = \max_{\|x\|=1} \|Ax\|$$

The first equality is the definition, the second follows by the linearity of M. That the supremum is obtained follows from the continuity of the euclidean norm and the compactness of the unit ball in $\mathbb{R}^m$. This last implies that A is bounded (which is not true of all linear operators on infinite-dimensional vector spaces), so we can bound the length of images of vectors under the action of A, I.e.

$$\forall x \in \mathbb{R}^m \qquad \|Ax\| \leq \|A\| \|x\| \tag{2}$$

The LHS of (2) is the length of the vector Ax (i.e. the image of x under A), whereas the RHS is the product of the spectral norm of A and the length of x.

For all matrices A, A^TA is an n by n symmetric matrix. So $A^TA = \sum_i \sigma_i u_i u_i^T$ for orthonormal $\{u_1, ..., u_n\}$ and $\{\sigma_1, ..., \sigma_n\} = \text{spec}(A^TA)$. Without loss of generality, suppose $\sigma_1 > ... > \sigma_n$. Further, since

$$x^T(A^TA)x = (Ax)^T(Ax) = \|Ax\|^2 \geq 0$$

we have that A^TA is PSD, and so $\sigma_i \geq 0$ for all i. Now

$$\begin{aligned}\|Ax\|^2 &= \langle Ax, Ax\rangle \\ &= x^TA^TAx \\ &= x^T\left(\sum_i \sigma_i u_i u_i^T\right)x \\ &= \sum_i \sigma_i \langle x, u_i\rangle^2\end{aligned}$$

So $\|Ax\|^2$ is maximized over unit vectors when x is in the direction of u_1 (since $\sigma_1 \geq \sigma_i$ for all i).

$$\|A\| = \max_{\|x\|=1} \|Ax\| = \sqrt{\max_{\|x\|=1} \|Ax\|^2} = \sqrt{\sigma_1} = \sqrt{\|A^TA\|}$$

Now if A is symmetric, $A = D\Lambda D^T$ and $A^T = A$, so

$$A^TA = A^2 = (D\Lambda D^T)(D\Lambda D^T) = D\Lambda^2 D^T$$

So the squares of the eigenvalues of A equal the eigenvalues of A^TA, and

$$\|A\| = \sqrt{\sigma_1} = \max\{|\lambda| : \lambda \in \text{spec}(A)\}$$

Finally, we want to show that the function $m_\lambda : \mathbb{R}^{n\times n} \to \mathbb{R}$ with returns the smallest modulus of its eigenvalues, i.e.

$$m_\lambda(M) = \min\{|\lambda| : \lambda \in \text{spec}(M)\}$$

is a continuous function (with respect to the topology induced by the spectral norm). To do this, it is convenient to introduce the max norm, $\|\cdot\| : \mathbb{R}^{n\times m}$, which is the maximum magnitude of any of a matrix's entries, i.e.

$$\|M\|_\infty = \max_{i,j} |M_{i,j}|$$

It is straightforward to see that $\|M\| \leq \sqrt{n}\|M\|_\infty$, and so the topology induced by the spectral norm $\|\cdot\|$ is a refinement of the topology induced by the max norm $\|\cdot\|_\infty$. So to show that m_λ is continuous with respect to the spectral norm, it suffices to show that m_λ is a continuous function with respect to the max norm.

For a square matrix A, the eigenvalues of A are the roots of the characteristic polynomial $\chi_A(\lambda) = \det(M - \lambda I)$. The roots of a polynomial depend continuously on its coefficients, thus the eigenvalues of A depend continuously on its entries. Thus the function $m_\lambda(A)$ is continuous with respect to the max norm.

The Trace

The *trace of* A, trace(A), is the sum of the diagonal entries of A.

$$\text{trace}(A) = \sum_i A_{i,i}$$

Clearly trace is linear. Additionally, for a pair matrices A, B where the products AB and BA are defined, $\text{trace}(AB) = \text{trace}(BA)$ (apply the definition of trace, then of matrix multiplication and reorder the sum). Since $A = D\Lambda D^T$, and D is unitary, we have that

$$\text{trace}(A) = \text{trace}(D(\Lambda D^T)) = \text{trace}((\Lambda D^T)D) = \text{trace}(\Lambda)$$

So if A is a symmetric matrix, it is easily seen that $\text{trace}(A)$ is the sum of the eigenvalues of A (since every matrix A is similar to a matrix in Jordan Normal Form this is actually true for all matrices, but we will not need this more general fact).

3 A Result in Matrix Theory

In this section we state and prove the main technical result from Forster's paper which is the heart of the proof of Theorem 2. We define the operator $\tilde{M}$ on a set of vectors $\mathcal{X}$ as follows:

$$\tilde{M}(\mathcal{X}) = \sum_{x \in \mathcal{X}} \frac{1}{\|x\|^2} x x^T$$

If $\mathcal{X} \subset \mathbb{R}^r$, then $\tilde{M}(\mathcal{X})$ is an r by r matrix.

Theorem 4 *Let $\mathcal{X} = \{x_1, ..., x_n\}$ be a set of vectors in general position in $\mathbb{R}^r$ such that $n \geq r$. Then there exists a nonsingular $A \in \mathbb{R}^{r \times r}$ such that*

$$\tilde{M}(A\mathcal{X}) = \frac{n}{r} I_r$$

where $A\mathcal{X} = \{Ax : x \in \mathcal{X}\}$

$\tilde{M}(\mathcal{X}$ is a sum of outer-products of vectors, so it is PSD, so its trace gives us the sum of its eigenvalues.

$$\begin{aligned}
\text{trace}(\tilde{M}(\mathcal{X})) &= \text{trace}\left(\sum_i 1/\|x_i\|^2 x_i x_i^T\right) \\
&= \sum_i 1/\|x_i\|^2 \,\text{trace}(x_i x_i^T) && \text{(linearity of trace)} \\
&= \sum_i (1/\|x_i\|^2)(\|x_i\|^2) \\
&= n
\end{aligned}$$

So the average eigenvalue of $\tilde{M}(\mathcal{X})$ is n/r for any $\mathcal{X}$. By the spectral theorem, $\tilde{M} = U\Lambda U^T$ for unitary U and $\Lambda = \text{diag}(\text{spec}\,\tilde{M}(\mathcal{X}))$. So if we can find an A such that

$$\min(\text{spec}(\tilde{M}(A\mathcal{X}))) = n/r \tag{3}$$

i.e. n/r is the only eigenvalue of $\tilde{M}(A\mathcal{X})$, then

$$\begin{aligned}
&\tilde{M}(A\mathcal{X}) = U(n/r I_r)U^T \\
\Longrightarrow\ &\tilde{M}(U^T A\mathcal{X}) = n/r I_r
\end{aligned}$$

and we are done. So (3) is our goal.

Lemma 5 *Let $\mathcal{X}$ be a set of unit vectors in general position in $\mathbb{R}^k$. Either*

$$\tilde{M}(\mathcal{X}) = \frac{n}{r} I_r$$

or there exists a nonsingular A such that $\tilde{M}(A\mathcal{X})$ whose smallest eigenvalue is strictly larger than the smallest eigenvalue of $\tilde{M}(\mathcal{X})$.

Proof Without loss of generality we suppose that

$$\tilde{M}(\mathcal{X}) = \mathrm{diag}(\lambda_1, ..., \lambda_r)$$

with $\lambda_1 \geq ... \geq \lambda_r$. Since λ_r is the smallest eigenvalue, the average of which is n/r, $\lambda_r \leq n/r$. If $\lambda_r = n/r$, then all the eigenvalues are n/r and we are done. So we suppose that $\lambda_r < n/r$ and that its multiplicity is less than r.

Since $\tilde{M}(\mathcal{X})$ is PSD, its spectrum is non-negative, so let

$$A = \mathrm{diag}(\lambda_1^{-1/2}, ..., \lambda_r^{-1/2})$$

Since $\mathcal{X}$ is a set of unit vectors, $\tilde{M}(\mathcal{X}) = \sum xx^T$ and so

$$I_r = A\tilde{M}(\mathcal{X})A^T = \sum_x (Ax)(Ax)^T \tag{4}$$

Let

$$B = \tilde{M}(A\mathcal{X}) - \lambda_r I_r$$

If B is PSD, then the smallest eigenvalue of $\tilde{M}(A\mathcal{X})$ is at least λ_r.

$$\begin{aligned} B &= \tilde{M}(A\mathcal{X}) - \lambda_r I_r \\ &= \left(\sum_x 1/\|Ax\|^2 (Ax)(Ax)^2 \right) - \lambda_r \sum_x (Ax)(Ax)^T & \text{(by (4))} \\ &= \sum_x (1/\|Ax\|^2 - \lambda_r)(Ax)(Ax)^T \end{aligned}$$

So B is PSD if and only if $\alpha_x = (1/\|Ax\|^2 - \lambda_r) \geq 0$ for all $x \in \mathcal{X}$. Now

$$\|Ax\|^2 = \sum_i \frac{x_i^2}{\lambda_i} \leq 1/\lambda_r \|x_i\|^2$$

$\mathcal{X}$ is a set of unit vectors so $\|Ax\|^2 \leq 1/\lambda_r$, $\alpha_x \geq 0$ for all x, B is PSD, and

$$\min(\mathrm{spec}(\tilde{M}(A\mathcal{X}))) \geq \lambda_r \tag{5}$$

If the inequality of (5) is strict, we're done. So suppose that the inequality of (5) is in fact an equality. Showing that

$$\dim \mathcal{E}_{\lambda_r}(\tilde{M}(A\mathcal{X})) < \dim \mathcal{E}_{\lambda_r}(\tilde{M}(\mathcal{X})) \tag{6}$$

will complete the proof of the lemma since though we have not increased the smallest eigenvalue of $\tilde{M}(A\mathcal{S})$, the multiplicity of λ_r is an integer which we can reduce to zero (thereby increasing the minimum eigenvalue) by iterating the above process.

Let $v = \dim \mathcal{E}_{\lambda_r}(\tilde{M}(\mathcal{X}))$. Now

$$\mathcal{E}_{\lambda_r}(\tilde{M}(A\mathcal{X})) = \ker B$$

$\operatorname{im} B$ is the span of vectors x such that $\alpha_x > 0$. On the other hand, $\alpha_x = 0$ if and only if x is an eigenvector for λ_r. Since the vectors in $\mathcal{X}$ are in general position, the span of B is k, unless the eigenspace of $\tilde{M}(\mathcal{X})$ associated with λ_r is large. I.e.

$$\dim \operatorname{im} B = \min\{k, n - v\}$$

By the rank theorem, we have that

$$\begin{aligned} \dim \ker B &= k - \min\{k, n - v\} \\ &\leq k - (n - v) \\ &= v - (n - k) \end{aligned}$$

Since $n > k$, we get that $\dim \ker B < v$ and so (6) holds and the proof of the lemma is complete.

To finish the proof of the theorem, we need a lemma which we cite but do not prove (see the Forster's paper for the proof).

Lemma 6 *Let $\mathcal{X}$ be a set of $n \geq k$ vectors in general position in $\mathbb{R}^k$. For all $\epsilon > 0$, let*

$$S_\epsilon = \{A \in \mathbb{R}^{k \times k} : \det A \neq 0, \|A\| = 1 \text{ and } \min(\operatorname{spec}(\tilde{M}(A\mathcal{X}))) \geq 1 + \epsilon\}$$

Then S_ϵ is compact for all ϵ.

We now prove Theorem 4. We are given $\mathcal{X}$. First normalize $\mathcal{X}$. Partition $\mathcal{X}$ arbitrarily into a set of r vectors, $\mathcal{X}_1$ and the remaining $n - r$ vectors $\mathcal{X}_2$. Let A be the (nonsingular) linear transformation mapping $\mathcal{X}_1$ to the canonical unit vectors in $\mathbb{R}^r$. Thus

$$\tilde{M}(A\mathcal{X}) = I_r + M$$

where $M = \tilde{M}(A\mathcal{X}_2)$. If $n = r$, M is nothing and we're finished, so suppose $n > r$. Now M is PSD, so $M = U\Lambda U^T$ where all diagonal entries of Λ are non-negative, so

$$\begin{aligned} \tilde{M}(A\mathcal{X} &= I_r + U\Lambda U^T \\ &= U I_r U^T + U\Lambda U^T && (U \text{ is unitary}) \\ &= U(I_r + \Lambda)U^T \end{aligned}$$

We have immediately that the eigenvalues of $\tilde{M}(A\mathcal{X})$ are all at least 1.

If $\tilde{M}(\mathcal{X}) \neq n/rI_r$, Lemma 5 guarantees the existence of nonsingular A such that the smallest eigenvalue of $\tilde{M}(A\mathcal{X})$ is larger than that of $\tilde{M}(\mathcal{X})$. In particular, the smallest eigenvalue of $\tilde{M}(A\mathcal{X})$ is strictly bigger than 1, so there exists an $\epsilon > 0$ such that all eigenvalues of $\tilde{M}(A\mathcal{X})$ are greater than $1 + \epsilon$.

The minimum eigenvalue of $\tilde{M}(A\mathcal{X})$ is a continuous function of A and by Lemma 6, S_ϵ is a compact set, so there exists a nonsingular $A \in S_\epsilon$ such that the smallest eigenvalue of $\tilde{M}(A\mathcal{X})$ is maximal. Again by Lemma 5, if $\tilde{M}(A\mathcal{X}) \neq n/rI_r$, then we could increase its minimum eigenvalue, but this contradicts the choice of A, so $\tilde{M}(A\mathcal{X} = n/rI_r$ as required.

4 Proof of the Main Theorem

We need one more lemma for the proof of the main theorem.

Lemma 7 *Let $M \in \{+1, -1\}^{n \times m}$, and unit vectors $\mathcal{X} = \{x_1, ..., x_n\}, \mathcal{Y} = \{y_1, ..., y_m\} \subset \mathbb{R}^d$ such that $M_{i,j} = \text{sign}\langle x_i, y_j\rangle$. Then*

$$\sum_{i,j} |x_i \cdot y_j| \leq \|M\|\sqrt{nm}$$

Proof Let $x_i^{(k)}$ denote the kth component of the vector x_i. If $X = [x_1|...|x_n]$, let $\mathcal{W} = \{w_1, ..., w_d\} \subset \mathbb{R}^n$ be the set of rows of X. Likewise, if $Y = [y_1|...|y_m]$, let $\mathcal{Z} = \{z_1, ..., z_d\} \subset \mathbb{R}^m$ be the set of rows of Y. (So $w_i^{(k)} = w_k^{(i)}$ and likewise for y and z).

$$\text{sign}(x_i \cdot y_j) = M_{i,j} \implies |x_i \cdot y_j| = M_{i,j}(x_i \cdot y_j)$$

so

$$\begin{aligned}
\sum_{i,j} |x_i \cdot y_j| &= \sum_{i,j} M_{i,j}(x_i \cdot y_j) \\
&= \sum_{i,j} M_{i,j} \sum_k x_i^{(k)} y_i^{(k)} \\
&= \sum_k \sum_{i,j} M_{i,j}\, x_i^{(k)} y_i^{(k)} \\
&= \sum_k w_k^T M z_k
\end{aligned}$$

Since $\sum_{i,j} |x_i \cdot y_j|$ is positive, so is $\sum_k w_k^T M z_k$, and so

$$\begin{aligned}
\sum_k w_k^T M z_k &= \left|\sum_k w_k^T M z_k\right| \\
&\leq \sum_k |w_k^T M z_k| \\
&\leq \sum_k \|w_k\|\|M z_k\| \\
&\leq \sum_k \|w_k\|\|M\|\|z_k\| \\
&= \|M\| \sum_k \|w_k\|\|z_k\|
\end{aligned}$$

so

$$\sum_{i,j} |x_i \cdot y_j| \leq \|M\| \sum_k \|w_k\|\|z_k\| \tag{7}$$

Now

$$\left(\sum_k \|w_k\|\|z_k\|\right)^2 \leq \left(\sum_k \|w_k\|^2\right)\left(\sum_k \|z_k\|^2\right) \qquad \text{(Cauchy-Schwartz)}$$
$$= \left(\sum_i \|x_i\|^2\right)\left(\sum_j \|y_j\|^2\right)$$
$$= nm$$

Thus

$$\sum_k \|w_k\|\|z_k\| \leq \sqrt{nm} \tag{8}$$

The result follows from (7) and (8).

Now we can prove the main theorem.

Proof [Of Theorem 2] We have $M \in \{+1,-1\}^{n\times n}$. Let $C \in \mathbb{R}^{n\times n}$ be a rank r matrix that sign-approximates M. So there exists $\mathcal{X} = \{x_1, ..., x_n\}, \mathcal{Y} = \{y_1, ..., y_m\} \subset \mathbb{R}^r$ such that

$$x_i \cdot y_j = C_{i,j}$$

We can assume that $\mathcal{X}$ and $\mathcal{Y}$ are in general position; otherwise perturbing all entries by a value in $(-\epsilon, \epsilon)$ for small enough ϵ gives us a set of vectors that still sign-approximates M (for small enough ϵ) but are almost surely in general position.

By Theorem 4 there exists a nonsingular A such that $\tilde{M}(A\mathcal{X}) = n/rI_r$. Let $B = (A^T)^{-1}$. Note that

$$Ax_i \cdot By_j = (Ax_i)^T By_j = x_i^T A^T By_j = x_i^T y_j = x_i \cdot y_j$$

So replace $\mathcal{X}$ with $x_i = Ax_i/\|Ax_i\|$ and $\mathcal{Y}$ by $v_j = By_j/\|By_j\|$. $\mathcal{X}$ and $\mathcal{Y}$ still sign-approximate M but now $\mathcal{X}, \mathcal{Y}$ are unit vectors and so by lemma 7

$$\sum_{i,j} |x_i \cdot y_j| \leq \|M\|\sqrt{nm} \tag{9}$$

and by choice of A, $\sum_{x\in\mathcal{X}} xx^T = \frac{n}{r}I_r$.

Fix $y_j \in \mathcal{Y}$.

$$\sum_i |x_i \cdot y_j| \geq \sum_i (x_i \cdot y_j)^2$$
$$= y^T\left(\sum_i x_i x_i^T\right) y$$
$$= y^T(n/rI_r)y$$
$$= n/r$$

Summing over j gives

$$\sum_{i,j} |x_i \cdot y_j| \geq nm/r \tag{10}$$

Combining (9) and (10) we get that

$$r \geq \sqrt{nm}/\|M\|$$

The result follows.