Communication Complexity: Lecture 7

Nathan Harms

February 27, 2026

In this lecture, we will see some relationships between communication complexity and *margin*, a complexity measure important in learning theory (and communication as well).

1 Margin in Learning Theory

In learning theory, we usually have a *hypothesis class* $\mathcal{H}$ of functions $\mathcal{X} \to \{\pm 1\}$. There is some unknown function $f \in \mathcal{H}$ that we want to learn. There are several different models that formalize this task.

We can think of hypothesis classes $\mathcal{H}$ over finite domains $\mathcal{X}$ as matrices. Let $\mathcal{H} = \{h_1, h_2, \dots, \}$; since $\mathcal{X}$ is finite this set is also finite. Then define the matrix $M \in \{\pm 1\}^{\mathcal{X} \times \mathcal{H}}$ as

$$M_{i,x} \coloneqq h_i(x).$$

One useful condition on the matrix M that allows for efficient learning is for it to have large *margin*.

Definition 1.1 (Margin)**.** Let $M \in \{\pm 1\}^{m \times n}$ be a sign matrix.

$$\mathrm{mar}(M) \coloneqq \max_{M = \mathrm{sign}(UV^\top)} \min_{i,j} \frac{|\langle u_i, v_j \rangle|}{\|u_i\|_2 \cdot \|v_i\|_2}$$

where the minimum is over $U \in \mathbb{R}^{m \times d}$, $V \in \mathbb{R}^{n \times d}$ in any dimension d, such that $M_{i,j} = \mathrm{sign}(\langle u_i, v_j \rangle)$ where u_i and v_j are the i^{th} and j^{th} row of U and V respectively.

Hypothesis classes (i.e. sign matrices) can always be represented as *halfspaces*: a halfspace $h \colon \mathbb{R}^d \to \{\pm 1\}$ is a function of the form

$$h(x) = \mathrm{sign}(\langle x, v \rangle)$$

where $v \in \mathbb{R}^d$. For example, by definition of the rank, any sign matrix M can be factored as

$$M = UV^\top$$

where $U \in \mathbb{R}^{m \times r}$ and $V \in \mathbb{R}^{n \times r}$, with $r = \mathrm{rank}(M)$. Writing $u_i \in \mathbb{R}^r$ and $v_j \in \mathbb{R}^r$ for the rows of U, V respectively, we have

$$M_{i,j} = \langle u_i, v_j \rangle = \mathrm{sign}(\langle u_i, v_j \rangle),$$

so we have written the matrix in a form where each row (or column) is a halfspace. If we can force the halfspaces to have large *margin*, it can help learning algorithms perform well.

1.1 Perceptron

Here is a famous example of how large margin helps a learning algorithm perform well. Let $\mathcal{X} \subset \mathbb{S}^d$ be a set of unit vectors and $\mathcal{H}$ a set of halfspaces of the form $h(x) = \text{sign}(\langle x, v\rangle)$ for unit vector v.

We are given an arbitrary sequence of points $x^{(1)}, x^{(2)}, x^{(3)}, \dots$ from $\mathcal{X}$, together with their labels $h(x^{(i)})$ according to some $h \in \mathcal{H}$. We don't know h. Our goal is to learn h.

The following algorithm is known as the *perceptron* algorithm. Initiate unit vector $v^{(0)}$ arbitrarily. Given a sequence of points $x^{(i)}$ with labels $h(x^{(i)})$, for each $x^{(t)}$ in sequence:

1. If $\text{sign}(\langle v^{(t-1)}, x^{(t)}\rangle) = h(x^{(t)})$, set $v^{(t)} = v^{(t-1)}$;
2. Otherwise, set $v^{(t)} = v^{(t-1)} + h(x^{(t)})\frac{x^{(t)}}{\|x^{(t)}\|_2}$.

We say the algorithm *makes a mistake* in round t if it runs step 2, i.e. $h(x^{(t)}) \neq \text{sign}(\langle x^{(t)}, v^{(t)}\rangle)$. Let

$$\lambda := \min_{x\in\mathcal{X}, h\in\mathcal{H}} |\langle x, v\rangle|$$

be the margin of this point-halfspace arrangment. We will show that the number of mistakes made by the perceptron in any sequence is at most $1/\lambda^2$

Theorem 1.2. *The Perceptron algorithm makes at most $\frac{2}{\lambda^2}$ on any input sequence $x^{(1)}, x^{(2)}, \dots$ in $\mathcal{X}$ labelled by any function $h \in \mathcal{H}$.*

Proof. Let T be the total number of mistakes. Since the algorithm makes no updates on any input where no mistake is made, we can consider the subsequence $x^{(1)}, \dots, x^{(T)}$ of inputs where the algorithm makes a mistake. By induction, for every $t \in [T]$:

$$\begin{aligned}
\langle v, v^{(t)}\rangle &= \langle v, v^{(t-1)} + h(x^{(t)})x^{(t)}\rangle \\
&= \langle v, v^{(t-1)}\rangle + h(x^{(t)})\langle v, x^{(t)}\rangle \\
&= \langle v, v^{(t-1)}\rangle + |\langle v, x^{(t)}\rangle| \\
&\geq \langle v, v^{(t-1)}\rangle + \lambda \\
&\geq \langle v, v^{(0)}\rangle + t\lambda \\
&\geq t\lambda - 1.
\end{aligned}$$

By induction, since $\|v^{(0)}\|_2 = 1$,

$$\begin{aligned}
\|v^{(t)}\|_2^2 &= \langle v^{(t)}, v^{(t)}\rangle \\
&= \langle v^{(t-1)} + h(x^{(t)})x^{(t)}, v^{(t-1)} + h(x^{(t)})x^{(t)}\rangle \\
&= \|v^{(t-1)}\|_2^2 + 2h(x^{(t)})\langle x^{(t)}, v^{(t-1)}\rangle + 1 \\
&= \|v^{(t-1)}\|_2^2 - 2|\langle x^{(t)}, v^{(t-1)}\rangle| + 1 \\
&\leq \|v^{(t-1)}\|_2^2 + 1 \leq t.
\end{aligned}$$

Now consider the normalized inner product $\langle v, \frac{v^{(T)}}{\|v^{(T)}\|_2}\rangle$:

$$1 \geq \langle v, \frac{v^{(T)}}{\|v^{(T)}\|_2}\rangle \geq \frac{1}{\sqrt{t}}(T\lambda - 1) = \sqrt{T}\lambda - 1/\sqrt{T}$$

Therefore $T \leq 2/\lambda^2$. ∎

2 Margin in Communication Complexity

Today, we will think of communication matrices as sign matrices, $M \in \{\pm 1\}^{m \times n}$. A boolean matrix $M \in \{0,1\}^{m \times n}$ can be replaced with a sign matrix $M \in \{\pm 1\}^{m \times n}$ either by replacing all 0s with -1s, or by replacing $b \in \{0,1\}$ with $(-1)^b$; neither change affects the communication complexity.

2.1 Upper Bounds via Margin

Theorem 2.1. *Let $M \in \{\pm 1\}^{m \times n}$. Then*

$$\mathsf{R}^{\rightarrow}_{1/4}(M) = O\left(\frac{1}{\mathrm{mar}(M)^2}\right).$$

To prove the theorem, we require the following lemma, which we will use again later.

Lemma 2.2. *Let $u, v \in \mathbb{S}^{d-1}$ be any unit vectors. Let $\boldsymbol{w}_1, \dots, \boldsymbol{w}_i \sim \mathbb{S}^{d-1}$ be uniformly random unit vectors. Define the random variables*

$$\boldsymbol{\phi}(u) := (\mathrm{sign}(\langle x, \boldsymbol{w}_1 \rangle), \dots, \mathrm{sign}(\langle x, \boldsymbol{w}_k \rangle)),$$

and $\boldsymbol{\phi}(v)$ similarly defined for v. Then

$$\mathbb{P}\left[\frac{1}{k}\left|\langle \boldsymbol{\phi}(u), \boldsymbol{\phi}(v) \rangle\right| < (1-\delta)\frac{2}{\pi}|\langle u, v \rangle| \ \vee\ \mathrm{sign}(\langle \boldsymbol{\phi}(u), \boldsymbol{\phi}(v) \rangle) \neq \mathrm{sign}(\langle u, v \rangle)\right] \le e^{-\frac{2\delta^2 k \langle u, v \rangle^2}{\pi^2}}.$$

Proof. Write

$$p_{u,v} := \mathbb{P}_{\boldsymbol{w}}\left[\mathrm{sign}(\langle u, \boldsymbol{w} \rangle) \neq \mathrm{sign}(\langle v, \boldsymbol{w} \rangle)\right].$$

where $\boldsymbol{w}$ is a random unit vector. Let α be the angle between u, v, which satisfies $\alpha = \cos^{-1}(\langle u, v \rangle)$. Then

$$p_{u,v} = \frac{2\alpha}{2\pi} = \frac{\alpha}{\pi}.$$

So

$$\mathbb{E}\left[\langle \boldsymbol{\phi}(u), \boldsymbol{\phi}(v) \rangle\right] = k((1 - p_{u,v}) - p_{u,v}) = k\left(1 - 2\frac{\alpha}{\pi}\right).$$

Note that, if $\langle u, v \rangle \ge 0$, then $\alpha \le \frac{\pi}{2} - \langle u, v \rangle$, and if $\langle u, v \rangle \ge 0$, then $\alpha \ge \frac{\pi}{2} + |\langle u, v \rangle|$. So if $\langle u, v \rangle \ge 0$ then $\mathbb{E}\left[\langle \boldsymbol{\phi}(u), \boldsymbol{\phi}(v) \rangle\right] \ge k \cdot \frac{2}{\pi}|\langle u, v \rangle|$, while if $\langle u, v \rangle < 0$ then $\mathbb{E}\left[\langle \boldsymbol{\phi}(u), \boldsymbol{\phi}(v) \rangle\right] \le -k \cdot \frac{2}{\pi}|\langle u, v \rangle|$. Suppose $\langle u, v \rangle \ge 0$. Then, by the Hoeffding bound,

$$\mathbb{P}\left[\langle \boldsymbol{\phi}(u), \boldsymbol{\phi}(v) \rangle < (1-\delta)k \cdot \frac{2}{\pi}|\langle u, v \rangle|\right] \le \exp\left(-\frac{2\delta^2 k \cdot \langle u, v \rangle^2}{\pi^2}\right).$$

∎

We can now prove Theorem 2.1:

Proof of Theorem 2.1. Write $M = UV^\top$ where the rows u_i and v_j of U, V are unit vectors which satisfy $M_{i,j} = \mathrm{sign}\,\langle u_i, v_j \rangle$ and $|\langle u_i, v_j \rangle| \ge \mathrm{mar}(M)$ for all i, j. On inputs i, j, Alice and Bob use shared randomness to choose k uniformly random unit vectors $\boldsymbol{w}_1, \dots, \boldsymbol{w}_k$. Alice sends $\boldsymbol{\phi}(u_i) \in \{\pm 1\}^k$ and Bob outputs $\mathrm{sign}(\langle \boldsymbol{\phi}(u_i), \boldsymbol{\phi}(v_j) \rangle)$. By the lemma, the probability that this succeeds is at least

$$1 - \exp\left(-\Theta\left(k \cdot \mathrm{mar}(M)^2\right)\right) \ge 3/4$$

when we choose $k = \Theta(1/\mathrm{mar}(M)^2)$. ∎

2.2 Margin vs. Discrepancy

Recall the discrepancy of a matrix. For sign matrices $M \in \{\pm 1\}^{m\times n}$, we can write discrepancy as

$$\begin{aligned}\mathrm{disc}(M) &:= \min_{\mu} \max_{R=X\times Y} \left| \sum_{(i,j)\in R} \mu_{i,j} M_{i,j} \right| \\ &= \min_{\mu} \max_{s\in\{0,1\}^m, t\in\{0,1\}^n} \left| \sum_{i,j} \mu_{i,j} M_{i,j} s_i t_j \right|,\end{aligned}$$

where μ is a probability distribution over entries $(i,j) \in [m]\times[n]$. Below, we will always write μ for such a probability distribution.

We will prove the remarkable fact that discrepancy is equivalent to margin! This was originally shown by Linial and Shraibman; our proof will differ in some places and attain constant 2π instead of $4K_G \le 7.13$ where K_G is the Grothendieck constant.

Theorem 2.3 ([LS09])**.** *Let $M \in \{\pm 1\}^{m\times n}$. Then*

$$\mathrm{disc}(M) \le \mathrm{mar}(M) \le 2\pi \cdot \mathrm{disc}(M).$$

Corollary 2.4. *For all $M \in \{\pm 1\}^{m\times n}$,*

$$\Omega\left(\log \frac{1}{\mathrm{disc}(M)}\right) \le \mathsf{R}_{1/4}(M) \le \mathsf{R}^{\rightarrow}_{1/4}(M) = O\left(\frac{1}{\mathrm{disc}(M)^2}\right)$$

2.2.1 Sign-vector versions of discrepancy and margin.

To prove the theorem, we will introduce variations of $\mathrm{disc}(M)$ and $\mathrm{mar}(M)$. First some notation. For fixed m, n, we will write

$$\mathcal{B} := \{uv^\top \mid u \in \{0,1\}^m, v \in \{0,1\}^n\}$$

for the set of rank-1 boolean matrices, and

$$\mathcal{S} := \{st^\top \mid s \in \{\pm 1\}^m, t \in \{\pm 1\}^n\}$$

for the set of rank-1 sign matrices.

Definition 2.5. Define

$$\mathrm{disc}^{\pm}(M) := \min_{\mu} \max_{s\in\{\pm 1\}^n, t\in\{\pm 1\}^m} \sum_{i,j} \mu_{i,j} M_{i,j} s_i t_j = \min_{\mu} \max_{S\in\mathcal{S}} \sum_{i,j} \mu_{i,j} M_{i,j} S_{i,j}$$

and

$$\mathrm{mar}^{\pm}(M) := \max_{X,Y} \min_{i,j} \frac{|\langle x_i, y_j\rangle|}{\|x_i\|_2 \|y_i\|_2},$$

where the max is over matrices $X \in \{\pm 1\}^{m\times d}$ and $Y \in \{\pm 1\}^{n\times d}$ with rows $x_i \in \{\pm 1\}^d$ and $y_j \in \{\pm 1\}^d$, in any dimension d, such that

$$\forall i, j: \qquad M_{i,j} = \mathrm{sign}(\langle x_i, y_i\rangle).$$

Proposition 2.6. *Let $M \in \{\pm 1\}^{m\times n}$. Then*

$$\mathrm{disc}(M) \le \mathrm{disc}^{\pm}(M) \le 4 \cdot \mathrm{disc}(M).$$

Proof. Start with the right hand side. Observe that any $S = st^\top \in \mathcal{S}$ can be written as $S = A + B - C - D$ where $A, B, C, D \in \mathcal{B}$ are rank-1 boolean matrices. Then

$$\begin{aligned}
\mathrm{disc}^{\pm}(M) &= \min_{\mu} \max_{S\in\mathcal{S}} \sum_{i,j} \mu_{i,j} M_{i,j} S_{i,j} \\
&\le \min_{\mu} \max_{S\in\mathcal{S}} \left(\left|\sum_{i,j} \mu_{i,j} M_{i,j} A_{i,j}\right| + \left|\sum_{i,j} \mu_{i,j} M_{i,j} B_{i,j}\right| + \left|\sum_{i,j} \mu_{i,j} M_{i,j} C_{i,j}\right| + \left|\sum_{i,j} \mu_{i,j} M_{i,j} D_{i,j}\right| \right) \\
&\le 4 \min_{\mu} \max_{B\in\mathcal{B}} \left|\sum_{i,j} \mu_{i,j} M_{i,j} B_{i,j}\right| \\
&= 4 \cdot \mathrm{disc}(M).
\end{aligned}$$

Now for the left hand side. Write $\mathrm{conv}(\mathcal{S})$ for the convex hull of rank-1 sign matrices. For any fixed distribution μ, the expression

$$\max_{S\in\mathcal{S}} \sum_{i,j} \mu_{i,j} M_{i,j} S_{i,j}$$

is maximizing function that is linear in S, and therefore

$$\max_{S\in\mathcal{S}} \sum_{i,j} \mu_{i,j} M_{i,j} S_{i,j} = \max_{S\in\mathrm{conv}(\mathcal{S})} \sum_{i,j} \mu_{i,j} M_{i,j} S_{i,j}.$$

Now,

$$\max_{S\in\mathrm{conv}(\mathcal{S})} \sum_{i,j} \mu_{i,j} M_{i,j} S_{i,j} = \max_{S\in\mathrm{conv}(\mathcal{S})} \left|\sum_{i,j} \mu_{i,j} M_{i,j} S_{i,j}\right|,$$

because, if the maximal absolute value is attained when $\sum_{i,j} \mu_{i,j} M_{i,j} S_{i,j} < 0$, we may replace S with $-S$. Finally, observe that any rank-1 boolean matrix $R = uv^\top$ for boolean u, v is the average of four rank-1 sign matrices (since $u = \frac{1}{2}(s_1 + s_2)$ for two sign vectors s_1, s_2, and similar for v), so

$$\max_{S\in\mathrm{conv}(\mathcal{S})} \left|\sum_{i,j} \mu_{i,j} M_{i,j} S_{i,j}\right| \ge \max_{R=uv^\top} \left|\sum_{i,j} \mu_{i,j} M_{i,j} u_i v_j\right|,$$

for $u \in \{0,1\}^m, v \in \{0,1\}^n$. Then

$$\mathrm{disc}^{\pm}(M) = \min_{\mu} \max_{S\in\mathcal{S}} \sum_{i,j} \mu_{i,j} M_{i,j} S_{i,j} \ge \min_{\mu} \max_{R=uv^\top} \left|\sum_{i,j} M_{i,j} u_i v_j\right| = \mathrm{disc}(M). \qquad \blacksquare$$

Proposition 2.7. *Let $M \in \{\pm 1\}^{m\times n}$. Then*

$$\mathrm{mar}^{\pm}(M) \le \mathrm{mar}(M) \le \frac{\pi}{2} \cdot \mathrm{mar}^{\pm}(M).$$

Proof. By definition,

$$\mathrm{mar}^{\pm}(M) \leq \mathrm{mar}(M),$$

because $\mathrm{mar}^{\pm}(M)$ maximizes over a subset of unit vectors. In the other direction, fix any $U \in \mathbb{R}^{m\times d}$, $V \in \mathbb{R}^{n\times d}$ attaining the maximum in $\mathrm{mar}(M)$, so the rows $u_i \in \mathbb{R}^d$ of U and $v_j \in \mathbb{R}^d$ of V are unit vectors with $\mathrm{sign}(\langle u_i, v_j\rangle) = M_{i,j}$.

By applying Lemma 2.2, together with a union bound over all mn pairs of vectors u_i, v_j, using arbitrarily small $\delta > 0$ and parameter $k \gg \frac{1}{\delta^2 \mathrm{mar}(M)^2}\log(mn)$ we obtain vectors $\phi(u_i), \phi(v_j) \in \{\pm 1\}^k$ such that

$$\forall i,j: \qquad \mathrm{sign}(\langle \phi(u_i), \phi(v_j)\rangle) = \mathrm{sign}(\langle u_i, v_j\rangle) \text{ and } \frac{|\langle \phi(u_i), \phi(v_j)\rangle|}{k} \geq (1-\delta)\cdot \frac{2}{\pi}\mathrm{mar}(M).$$

Therefore $\mathrm{mar}^{\pm}(M) \geq (1-\delta)\frac{2}{\pi}\mathrm{mar}(M)$ for any choice of δ. ■

2.2.2 Equality of ± versions

Using the above propositions, Theorem 2.3 now follows from the next lemma:

Lemma 2.8. *Let $M \in \{\pm 1\}^{m\times n}$. Then*

$$\mathrm{mar}^{\pm}(M) = \mathrm{disc}^{\pm}(M)$$

Let's break this proof down into two parts: we will show that both quantities are equal to

$$\Phi(M) := \max_{S\in \mathrm{conv}(\mathcal{S})} \min_{i,j} M_{i,j}S_{i,j}$$

Proposition 2.9. *Let $M \in \{\pm 1\}^{m\times n}$. Then*

$$\mathrm{mar}^{\pm}(M) = \Phi(M).$$

Proof. Maximizing over $X \in \{\pm 1\}^{m\times d}$ and $Y \in \{\pm 1\}^{n\times d}$ of any dimension d with $M_{i,j} = \mathrm{sign}(\langle x_i, y_j\rangle)$, where x_i and y_j are rows of X and Y, we have

$$\mathrm{mar}^{\pm}(M) = \max_{X,Y}\min_{i,j}\frac{|\langle x_i, y_j\rangle|}{\|x_i\|_2\|y_i\|_2}. = \max_{X,Y}\min_{i,j}\frac{M_{i,j}\langle x_i, y_j\rangle}{\|x_i\|_2\|y_i\|_2}.$$

For fixed X, Y, we have $\|x_i\|_2 = \|y_j\|_2 = \sqrt{d}$ for all x_i, y_j. Write $X^{(\ell)} \in \{\pm 1\}^m$, $Y^{(\ell)} \in \{\pm 1\}^n$ for the ℓ^{th} columns of X and Y, respectively, and note that $x_i = (X_i^{(1)}, \ldots, X_i^{(d)})$ and $y_j = (Y_j^{(1)}, \ldots, Y_j^{(d)})$. Let S be the matrix

$$S := \frac{1}{d}\sum_{\ell=1}^{d} X^{(\ell)}(Y^{(\ell)})^{\top}, \tag{1}$$

so that

$$S_{i,j} = \frac{1}{d}\sum_{\ell=1}^{d} X_i^{(\ell)} Y_j^{(\ell)} = \frac{1}{d}\langle x_i, y_j\rangle.$$

Then

$$\mathrm{mar}^{\pm}(M) = \sup_{S\in\mathcal{R}}\min_{i,j} M_{i,j}S_{i,j},$$

where the supremum is over matrices $\mathcal{R}$ defined as those of the form in Equation (1), which also satisfy $\operatorname{sign}(S_{i,j}) = M_{i,j}$ for all i,j. Observe that any such S is in $\operatorname{conv}(\mathcal{S})$, which gives $\operatorname{mar}^{\pm}(M) \le \Phi(M)$. In the other direction, let $S \in \operatorname{conv}(\mathcal{S})$ satisfy

$$\Phi(M) = \min_{i,j} M_{i,j} S_{i,j}.$$

By considering $S = 0$, we can see that $\min_{i,j} M_{i,j} S_{i,j} \ge 0$, which means that $\operatorname{sign}(S_{i,j}) = M_{i,j}$ for all i,j. Now we have

$$S = \sum_{\ell=1}^{r} \lambda_\ell X^{(\ell)} (Y^{(\ell)})^\top$$

for some values $\lambda_\ell \ge 0$ with $\sum_\ell \lambda_\ell = 1$. Pick an arbitrarily large d and choose k_ℓ such that $|k_\ell / d\lambda_\ell| < 1/2d$. Replacing S with S' defined by repeating $X^{(\ell)}(Y^{(\ell)})^\top$ k_ℓ times, each time with coefficient $1/d$, we get

$$S' = \sum_{\ell=1}^{r} \frac{k_\ell}{d} X^{(\ell)} (Y^{(\ell)})^\top$$

of the form in Equation (1) (i.e. in $\mathcal{R}$), so that $S'_{i,j} = S_{i,j} \pm 1/2d$ for all i,j, so, maximizing over S' of this form,

$$\operatorname{mar}^{\pm}(M) = \sup_{S' \in \mathcal{R}} \min_{i,j} M_{i,j} S'_{i,j} \ge M_{i,j} S_{i,j} - \frac{1}{2d}.$$

Since this holds for any d, it defines a sequence of S' converging to $\Phi(M)$ in the closed and bounded set $\operatorname{conv}(\mathcal{S})$, so $\operatorname{mar}^{\pm}(M) \ge \Phi(M)$, as desired. ■

Proposition 2.10. *Let $M \in \{\pm 1\}^{m \times n}$. Then*

$$\operatorname{disc}^{\pm}(M) = \Phi(M).$$

Proof. Observe that, for any fixed $S \in \operatorname{conv}(\mathcal{S})$,

$$\Phi(M) = \min_{i,j} M_{i,j} S_{i,j} = \min_{\mu} \sum_{i,j} \mu_{i,j} M_{i,j} S_{i,j}$$

because, in the right hand side, we are minimizing a function linear in μ over the convex hull of distributions with the form $\mu_{i,j} = 1$ for some fixed i,j. So

$$\Phi(M) = \max_{S \in \operatorname{conv}(\mathcal{S})} \min_{\mu} \sum_{i,j} \mu_{i,j} M_{i,j} S_{i,j},$$

Now, $\sum_{i,j} \mu_{i,j} M_{i,j} S_{i,j}$ is linear in both μ and S, and S and μ are both chosen from convex, compact sets. Therefore we may apply the von Neumann minimax theorem to obtain

$$\Phi(M) = \min_{\mu} \max_{S \in \operatorname{conv}(\mathcal{S})} \sum_{i,j} \mu_{i,j} M_{i,j} S_{i,j}.$$

Now, $\max_{S \in \operatorname{conv}(\mathcal{S})} \sum_{i,j} \mu_{i,j} M_{i,j} S_{i,j}$ is maximizing a linear function over the convex hull of $\mathcal{S}$, so for any fixed μ,

$$\max_{S \in \operatorname{conv}(\mathcal{S})} \sum_{i,j} \mu_{i,j} M_{i,j} S_{i,j} = \max_{S \in \mathcal{S}} \sum_{i,j} \mu_{i,j} M_{i,j} S_{i,j}.$$

Therefore

$$\Phi(M) = \min_{\mu} \max_{S \in \mathcal{S}} \sum_{i,j} \mu_{i,j} M_{i,j} S_{i,j}. = \operatorname{disc}^{\pm}(M).$$ ■

References

[LS09] Nati Linial and Adi Shraibman. Learning complexity vs communication complexity. *Combinatorics, Probability and Computing*, 18(1-2):227–245, 2009.