**Communication Complexity** 15 Nov, 2011 (@ IMSc)

# 25. Pattern Matrix Method

*Lecturer: Prahladh Harsha* *Scribe: Nitin Saurabh*

Last lecture, we saw a duality-based approach to bound the discrepancy via the threshold degree of a function. Today, we will see a more general approach called the pattern matrix method due to Sherstov [She11]. The main reference for today's lecture are the pattern matrix paper [She11] and Sherstov's survey on application of dual polynomials in communication complexity [She08].

## 25.1 Generalized discrepancy method

Recall the discrepancy method to lower bound the randomized complexity.

**Theorem 25.1.** *For every function $f : X \times Y \to \{0, 1\}$, every probability distribution $\mu$ on $X \times Y$ and every $\varepsilon \geq 0$,*

$$\mathrm{R}_{\frac{1}{2}-\varepsilon}(f) \geq \mathrm{D}^{\mu}_{\frac{1}{2}-\varepsilon}(f) \geq \log \frac{2\varepsilon}{\mathrm{disc}_{\mu}(f)}$$

Thus, an inverse exponential upper bound on the discrepancy gives polynomial upper bounds on the randomized communication cost. In fact, one of the strengths of this approach is that it gives bounds even for protocols with error very close to 1/2. But this also happens to be one of the reasons it fails to give good bounds for functions, such as $\mathsf{DISJ}$, which have constant bit protocols that achieve error at most $1/2 - 1/\mathrm{poly}(n)$. Thus, the discrepancy of $\mathsf{DISJ}$ is at least $1/\mathrm{poly}(n)$ and the discrepany method will not give any better than logarithmic lower bound on the randomized communication cost.

We will now discuss an extension of the discrepancy method, originally due to Razborov and formalized by Klauck, called the generalized discrepancy method, which will help get around this weakness of the discrepancy approach for certain functions.

Let $f : X \times Y \to \{1, -1\}$ be a function whose communication complexity is of interest. Suppose there exist $h : X \times Y \to \{1, -1\}$ and a distribution $\mu$ on $X \times Y$ such that the following holds.

1. $f$ and $h$ are well correlated with respect to $\mu$:

$$\mathbb{E}_{(x,y)\sim\mu}[f(x, y)h(x, y)] \geq \varepsilon$$

2. If $\Pi$ is a randomized protocol for $h$ with cost $c$, then

$$\mathbb{E}_{(x,y)\sim\mu,\Pi}[h(x, y)\Pi(x, y)] \leq 2^{O(c)}\nu$$

In other words, no low cost protocol $\Pi$ has large advantage in computing $h$ under $\mu$.

Let $\Pi'$ be the best cost protocol with cost $c$ that computes $f$ and

$$\mathsf{P}r[\Pi'(x,y) \neq f(x,y)] \leq \varepsilon/3$$

for all $(x,y)$. Hence we get,

$$2^{O(c)}\nu \geq \mathbb{E}_{(x,y)\sim\mu,\Pi'}[h(x,y)\Pi'(x,y)] \geq \mathbb{E}_{(x,y)\sim\mu}[f(x,y)h(x,y)] - 2\cdot\frac{\varepsilon}{3} \geq \frac{\varepsilon}{3}$$

This implies

$$\mathrm{R}_{\varepsilon/3}(f) = c = \Omega(\log\frac{\varepsilon}{\nu})$$

Thus, to show that $f$ has large communication cost it is sufficient to show that $f$ is well-correlated with another function $h$ (under some distribution) $\mu$ that has large communication cost. Specializing this to the discrepancy method, we get the following theorem.

**Theorem 25.2.** *Let $f : X \times Y \to \{1,-1\}$ and $\varepsilon \in [0,1/2)$. Then*[1] ,

$$2^{\mathrm{R}_\varepsilon(f)} \geq \max_{H,P}\left\{\frac{\langle H\circ P, F\rangle - 2\varepsilon}{\mathrm{disc}_P(H)}\right\}$$

*where $F$ is the sign matrix corresponding to $f$, and the maximization is over all $H$ that are $\pm1$ sign matrices and $P$ that are probability matrices (i.e, $P \geq 0$ and $\|P\|_1 = 1$).*

*Proof.* Let $H : X \times Y \to \{1,-1\}$ viewed as a sign matrix and $P$ be a distribution on $X \times Y$ such that

1. $f$ and $H$ are well correlated wrt. $P$, that is, $\langle F, H\circ P\rangle = \mathbb{E}_{(x,y)\sim P}[f(x,y)h(x,y)] \geq \delta$.
2. $\mathrm{disc}_P(H)$ is small.

Let $\Pi$ be a cost $c$ protocol that computes $f$ with error at most $\varepsilon$ with respect to $\mu$. Then,

$$\mathsf{P}r_{\mu,\Pi}[\Pi(x,y) = h(x,y)] \geq \mathsf{P}r_{\mu,\Pi}[\Pi(x,y) = f(x,y)] - \mathsf{P}r_{\mu}[f(x,y) \neq h(x,y)] \geq 1-\varepsilon-\frac{1-\delta}{2} \geq \frac{1}{2}+\frac{\delta-2\varepsilon}{2}$$

Hence, by Theorem 25.1, we get $c \geq \log\frac{\delta-2\varepsilon}{\mathrm{disc}_\mu(h)}$. Therefore,

$$2^{\mathrm{R}_\varepsilon(f)} \geq \max_{H,P}\left\{\frac{\langle H\circ P, F\rangle - 2\varepsilon}{\mathrm{disc}_P(H)}\right\}$$

where $H = [h(x,y)]_{x\in X, y\in Y}$ and $P = [\mu(x,y)]_{x\in X, y\in Y}$. □

[1] Here, $\langle A, B\rangle$ denotes $\sum A_{ij}B_{ij}$ while $H\circ P$ denotes the matrix obtained by the Hadamard product of $H$ and $P$ i.e., $(H\circ P)_{ij} = H_{ij}P_{ij}$

**Spectral approach:** Recall that the discrepancy $\mathrm{disc}_P(H)$ can be bounded using the spectral norm of the matrix $K = H \circ P$ given by $\|K\| = \sigma_1(K) = \max_{\|x\|=1} \|Kx\|$ as follows.

$$\mathrm{disc}_P(H) = \max_{S,T} |1_S^T \cdot K \cdot 1_T| \leq \max_{S,T} \|1_S\| \cdot \|K\| \cdot \|1_T\| = \|K\|\sqrt{|X||Y|}.$$

Substituting this into Theorem 25.2, we get the following theorem.

**Theorem 25.3.** *Let $F : X \times Y \to \{\pm 1\}$ and $K = [K_{xy}]_{x \in X, y \in Y}$ be any real matrix with $\|K\|_1 = 1$. Then for any $\varepsilon > 0$,*

$$2^{\mathrm{R}_\varepsilon(f)} \geq \frac{\langle F, K\rangle - 2\varepsilon}{\|K\|\sqrt{|X||Y|}}.$$

What is a good choice for $K$? Does $F$ itself work? Indeed, when the function is parity function we can take $K = \frac{1}{2^{2n}}F$. But, in general, its clear from Theorem 25.3 that we want $K$ with small spectral norm such that $\langle F, K\rangle$ is large, that is, good correaltion with $F$. In the rest of today's lecture, we will see how to get hold of such a $K$ using the notion of approximate degree.

## 25.2 Approximate degree

Let $f : \{1, -1\}^n \to \{1, -1\}$. The $\varepsilon$-approximate degree of $f$, $\deg_\varepsilon(f)$, is the least degree of a real polynomial $p(x_1, x_2, \ldots, x_n)$ such that $|f(x) - p(x)| \leq \varepsilon$ for all $x \in \{1, -1\}^n$. The dual characterization of approximate degree is given by the following theorem.

**Theorem 25.4.** *let $f$ be a boolean function and $\varepsilon \geq 0$. Then, $\deg_\varepsilon(f) \geq d$ if and only if there exist a function $\psi : \{1, -1\}^n \to \mathbb{R}$ such that*

$$\hat{\psi}(S) = 0, \quad |S| < d \tag{25.2.1}$$

$$\sum_{x \in \{\pm 1\}^n} |\psi(x)| = 1 \tag{25.2.2}$$

$$\sum_{x \in \{\pm 1\}^n} \psi(x) f(x) > \varepsilon \tag{25.2.3}$$

**Comparison to threshold degree:** Consider the function $g : \{1, -1\}^n \to \{1, -1\}$ given by $g(x) = \mathrm{sign}(\psi(x))$ and the distribution $\mu(x) = |\psi(x)|$ where $\psi$ is as in the theorem above. Equations (25.2.1) and (25.2.2) imply that $\deg_\pm(g) \geq d$ and the distribution $\mu$ is the dual witness to this fact as defined in the last lecture. Inequality (25.2.3) is equivalent to stating that $\mathbb{E}_\mu [g(x)f(x)] > \varepsilon$. In other words, if $f$ has large approximate degree, then $f$ has large correlation with a large threshold degree function wrt. the distribution that witnesses the large threshold degree. Recall that in last lecture we showed that if $g$ has large threshold degree, then a related $G$ has large communication complexity via the discrepancy method. Similarly, we will show in today's lecture that if $f$ has large approximate degree, then a related function $F$ has large communication complexity via the generalized discrepancy method.

*Proof of Theorem 25.4.* The proof is based on LP-duality. Consider the following primal LP:

$$\begin{array}{rcl}
\text{Minimize:} & \delta & \\
\text{subject to:} & |f(x) - \sum_{|S|<d} a_S \chi_S(x)| \leq \delta & \forall x \in \{1,-1\}^n \\
 & a_S \in \mathbb{R} & \forall S \\
 & \delta \geq 0 &
\end{array}$$

Its dual is given as follows:

$$\begin{array}{rcl}
\text{Maximize:} & \sum_{x \in \{\pm 1\}^n} \psi_x f(x) & \\
\text{subject to:} & \sum_x |\psi_x| = 1 & \\
 & \sum_x \psi_x \chi_S(x) = 0 & \forall S \\
 & \psi_x \in \mathbb{R} & \forall x \in \{1,-1\}^n
\end{array}$$

Since both LPs are feasible, they have same finite optimum and Note that $\deg_\varepsilon(f) \geq d$ iff opt(primal-LP) $> \varepsilon$. Hence, by duality $\deg_\varepsilon(f) \geq d$ iff opt(dual-LP) $> \varepsilon$, which proves the theorem. □

### 25.2.1 Large approximation degree to large communication complexity

Let $N$ and $n$ be positive integers such that $n$ divides $N$. Let $f : \{1,-1\}^n \to \{1,-1\}$. Define $F : \{1,-1\}^N \times \Gamma(N,n) \to \{1,-1\}$, where $\Gamma(N,n) \subseteq \binom{[N]}{n}$ and

$$F(y,V) = f(y|_V)$$

where $y|_V$ denotes the projection of $y$ onto the indices in $V$. Let us also assume that $\Gamma$ contains only the following types of sets $V$, $V = \{i_1, i_2, \dots, i_n\}$ such that

$$i_1 \in \left\{1, \dots, \frac{N}{n}\right\}, i_2 \in \left\{\frac{N}{n} + 1, \dots, \frac{2N}{n}\right\}, \dots, i_n \in \left\{\frac{(n-1)N}{N} + 1, \dots, N\right\}$$

Clearly, $|\Gamma(N,n)| = (N/n)^n$.

Our main theorem relating approximate degree to communication complexity is as follows.

**Theorem 25.5** (approximate degree implies large communication)**.** *Let $F$ be defined as above. Then,*

$$\mathrm{R}_{1/5}(F) \geq \frac{1}{2} \deg_{2/3}(f) \log \left\lfloor \frac{N}{2n} \right\rfloor - 2.$$

Setting $N = 4n$, we have following corllary.

**Corollary 25.6.** *Let $f : \{0,1\}^n \to \{0,1\}$. Define $F : \{0,1\}^{4n} \times \{0,1\}^{4n} \to \{0,1\}$ where*

$$\begin{aligned} F(x,y) = f\Big( & x_1y_1 \vee x_2y_2 \vee x_3y_3 \vee x_4y_4, \\ & x_5y_5 \vee x_6y_6 \vee x_7y_7 \vee x_8y_8, \\ & \vdots \\ & x_{4n-3}y_{4n-3} \vee \ldots \vee x_{4n}y_{4n}\Big) \end{aligned}$$

*where $x_iy_i = (x_i \wedge y_i)$. Then,*

$$\mathrm{R}_{1/5}(F) \geq \frac{1}{2}\deg_{1/3}(f) - 2.$$

Instead of applying the generalized discrepancy method to $F$ (as in Theorem 25.5), we will find it more convenient (for reasons that will become evident shortly) to apply the generalized discrepancy method to a related function, called the pattern matrix function.

## 25.3 The pattern matrix method

**Definition 25.7** (pattern matrix)**.** *Let $f : \{1,-1\}^n \to \mathbb{R}$. The $(N,n,f)$-pattern matrix is the real matrix $A_{N,n,f} \in \mathbb{R}^{\{\pm 1\}^N \times (\Gamma(N,n) \times \{\pm 1\}^n)}$, given by*

$$A_{N,n,f}(y,(V,w)) = f(y|_V \oplus w).$$

Observe that it is almost similar to the definition of $F$ from above, except that Bob gets an additional input $w$ and the function is $f(y|_V \oplus w)$ instead of just $f(y|_w)$. The matrix $A_{N,n,f}$ is a matrix consisting of several permuted copies of the matrix $F$ (one for each $w$) and hence, the name, pattern matrix.

**Theorem 25.8** (pattern matrix theorem)**.** *Let $A_{N,n,f}$ be the $(N,n,f)$-pattern matrix. Then,*

$$\mathrm{R}_{1/5}(A_{N,n,f}) \geq \frac{1}{2}\deg_{2/3}(f)\log\frac{N}{n} - 2$$

Before proving the theorem, let us see how it implies Theorem 25.5.

*Proof of Theorem 25.5.* The $(\frac{N}{2},n,f)$-pattern matrix is actually a submatrix of $F = [f(y|_V)]_{(y,V)}$ in the following sense.

$$A_{\frac{N}{2},n,f}(y,(V,w)) = f(y'|_{V'}),$$

where $y'$ and $V'$ are defined as follows: if $y = (y_1, y_2, \ldots, y_{N/2}) \in \{\pm 1\}^{N/2}$, then $y' = (y_1, -y_1, \ldots, y_{N/2}, -y_{N/2})) \in \{\pm 1\}^N$. If $V = \{i_1, \ldots, i_n\}$ where $i_j \in \{(j-1)N/2n + 1, \cdots, jN/2n\}$ and $w = (w_1, \ldots, w_n)$, then $V' = \{i'_1, \ldots, i'_n\}$ where $i'_j \in \{(j-1)N/n + 1, \cdots, jN/n\}$ is defined as

$$i'_j = \begin{cases} 2i_j - 1 & \text{if } w_j = 1 \\ 2i_j & \text{if } w_j = -1 \end{cases}.$$

Hence, Theorem 25.5 follows from Theorem 25.8. □

Now, onto the proof of Theorem 25.8.

*Proof of Theorem 25.8.* By Theorem 25.4, there exist a function $\psi : \{\pm 1\}^n \to \mathbb{R}$ such that:

$$\begin{aligned} \hat{\psi}(S) &= 0, \ |S| < d \\ \|\psi(x)\|_1 &= 1 \\ \sum_{x \in \{\pm 1\}^n} \psi(x) f(x) &> \varepsilon \end{aligned}$$

A natural choice for the matrix $K$ in the generalized discrepancy approach (see Theorem 25.3) is the pattern matrix corresponding to $\psi$ (suitably scaled). More precisely, let $K$ be the $\left(N, n, \frac{\psi}{(N/n)^n 2^N}\right)$-pattern matrix. From the second and third equations above, it follows that

$$\|K\|_1 = 1, \quad \langle K, A_{N,n,f} \rangle > 2/3$$

Now all we need to show is that the spectral norm $\|K\|$ is small. For this, we will analyse the singular values of $K$.

Let $\psi = \sum_S \hat{\psi}(S)\chi_S$ be the Fourier decomposition of $\psi$. It follows that $K = \sum_S \hat{\psi}(S) A_S$, where $A_S$ is the $\left(N, n, \frac{\chi_S}{(N/n)^n 2^N}\right)$-pattern matrix. Since $\langle \chi_S, \chi_T \rangle = 0$ if $S \neq T$, we have $A_S^T A_T = A_S A_T^T = 0$ if $S \neq T$.[2] Suppose if we could obtain the singular values of $K$ from $A_S$'s and show that $\|A_S\|$ is upper bounded, then we shall be through. This is exactly what we do using the following two lemmata which we will prove later.

**Lemma 25.9.** *Let $A$, $B$ be real matrices such that $A^T B = AB^T = 0$. Then the non-zero singular values of $A + B$, even respecting multiplicities, are $\{$non-zero singular values of $A\} \cup \{$non-zero singular values of $B\}$ as a multiset.*

**Lemma 25.10.** *Let $g : \{1, -1\}^n \to \mathbb{R}$ and $A_{N,n,g}$ be the $(N, n, g)$-pattern matrix. Then the non-zero singular values of $A_{N,n,g}$, counting multiplicities, are:*

$$\bigcup_{S : \hat{g}(S) \neq 0} \left\{ \sqrt{2^{N+n} \left(\frac{N}{n}\right)^n} \cdot |\hat{g}(S)| \left(\frac{n}{N}\right)^{|S|/2} \right\}.$$

By Lemma 25.10 and using the fact the $\|K\| = \sigma_1(K)$, we have,

$$\|K\| = \max_{S \subseteq [n]} \sqrt{2^{N+n} \left(\frac{N}{n}\right)^n} \cdot \frac{|\hat{\psi}(S)|}{(N/n)^n 2^N} \cdot \left(\frac{n}{N}\right)^{|S|/2}.$$

But $\|\psi(x)\|_1 = 1$, implies $|\hat{f}(S)| \leq 2^{-n}$, and $\hat{\psi}(S) = 0$, for $|S| < d$. Hence,

$$\|K\| \leq \left(\frac{n}{N}\right)^{d/2} \left(2^{N+n} \left(\frac{N}{n}\right)^n\right)^{-1/2}$$

[2] Note that this is not true if we had worked with the matrix $F = [f(y|_V)]_{(y,V)}$ as opposed to $[f(y|_V \oplus w)]_{(y,(V,w))}$. This is the reason for working with the pattern matrix instead of the matrix $F$.

Since $\langle K, A_{N,n,f}\rangle > 2/3$ and $\varepsilon = 1/5$, using Theorem 25.3 we get,

$$\mathrm{R}_{1/5}(A_{N,n,f}) \geq \frac{1}{2}\deg_{2/3}(f)\log\frac{N}{n} - 2$$

□

*Proof of Lemma 25.9.* The singular values of $(A+B)$ are exactly the square roots of the eigenvalues of $(A+B)(A+B)^T$.

$$\begin{aligned}(A+B)(A+B)^T &= AA^T + BB^T + A^TB + AB^T \\ &= AA^T + BB^T\end{aligned}$$

Let $AA^T = \sum_{i=1}^{\operatorname{rank} A}\sigma_i^2(A)u_iu_i^T$ and $BB^T = \sum_{j=1}^{\operatorname{rank} B}\sigma_j^2(B)v_iv_i^T$ be the singular value decomposition of $AA^T$ and $BB^T$ respectively. Then,

$$\begin{aligned}\sum_{i,j}\sigma_i^2(A)\sigma_j^2(B)\langle u_i, v_j\rangle^2 &= \langle AA^T, BB^T\rangle \\ &= tr(AA^TBB^T) = 0\end{aligned}$$

This implies that $u_i's$ and $v_j's$ are orthonormal. Hence, the singular values of $(A+B)$ are infact, $\sigma_1(A), \ldots, \sigma_{\operatorname{rank} A}(A), \sigma_1(B), \ldots, \sigma_{\operatorname{rank} B}(B)$, with multiplicities. □

*Proof of Lemma 25.10.* Let $g = \sum_{S\subseteq[n]}\hat{g}(S)\chi_S$. Hence, $A_{N,n,g} = \sum_{S\subseteq[n]}\hat{g}(S)A_S$, where $A_S$ is $(N, n, \chi_S)$-pattern matrix. We have seen earlier that $A_SA_T^T = A_S^TA_S = 0$, if $S \neq T$. By Lemma 25.9, the non-zero singular values of $A$ are the union of the non-zero singular values of $\hat{g}(S)A_S$, counting multiplicities. So lets look at $A_S^TA_S$:

$$A_S^TA_s = [\chi_S(w)\chi_S(w')]_{w,w'} \otimes \left[\sum_y \chi_S(y|_V)\chi_S(y|_{V'})\right]_{V,V'}$$

The first term in above equation is a rank 1 matrix with entries ±1. Therefore, its singular value is $2^n$ with multiplicity 1. The second term is of the form

$$2^N\begin{bmatrix} J & & & \\ & J & & \\ & & \ddots & \\ & & & J\end{bmatrix}$$

where $J$ is the all ones matrix of order $\left(\frac{N}{n}\right)^{n-|S|}$. This implies that its singular value is $2^N(N/n)^{n-|S|}$ with multiplicity $(N/n)^{|S|}$. Hence, the non-zero singular value of $A_S^TA_S$ is $2^{N+n}(N/n)^{n-|S|}$, with multiplicity $(N/n)^{|S|}$ and the lemma follows. □

## References

[She08] Alexander A. Sherstov. *Communication lower bounds using dual polynomials*. Bulletin of the EATCS, 95:59–93, 2008. arXiv:0805.2135.

[She11] ———. *The pattern matrix method*. SIAM J. Computing, 40(6):1969–2000, 2011. (Preliminary version in *40th STOC*, 2008). arXiv:0906.4291, doi:10.1137/080733644.