29. Other Lower Bound Methods

Lecturer: Meena Mahajan *Scribe: Meena Mahajan*

In this last lecture of the course, we give an overview of some norm-based lower bound methods, and also see how they are very elegantly unified using the language of LP duality. Finally, we mention further topics that we have not considered in this course at all.

29.1 Duality-based Methods: the general strategy

Let A be a sign matrix $A \in \{-1,+1\}_{m\times n}$, and let $\mathrm{CC}(A)$ denote the communication complexity of the associated function in the model of interest (deterministic, randomized ...). The general framework in which norm-based lower bounds work has three major steps:

Step 1. Embed the problem into reals. Specifically, find a function $G : \mathbb{R}^{m\times n} \to \mathbb{R}$, such that for every sign matrix A, $G(A) \leq \mathrm{CC}(A)$. Now minimize G to get a lower bound on CC. Since G has a continuous domain as opposed to the discrete domain of CC, hopefully more mathematical tools will be available and will make minimization easier.

If G is suitably *nice*, for instance if it is a convex function, then optimizing G could be significantly easier.

Of course, there is a price paid: G is a lower bound on CC, but it could be much less than CC. We can't say much right away about the quality of the lower bound.

Step 2. Reformulate "Minimize G" as "Maximize G^*".

Obtaining a lower bound on $\min G$ may be hard since we need to show that for every A, $G(A)$ is at least as large as the target bound. Obtaining a lower bound on $\max G^*$ ought to be easier, since it suffices to show that there exists a single B with $G^*(B)$ as large as the target bound. That is, we just need to find a single *dual witness*.

Note that transforming $\min G$ to $\max G^*$ may not always be possible. It is possible if G is a "nice" function.

Step 3. Find a good witness B. Conclude that $\mathrm{CC}(A) \geq G^*(B)$.

Typically, the functions G used are some kind of norm defined on any matrices, not just sign matrices. We will see several examples of norm-based lower bound methods.

29.2 Lower bounds for deterministic communication $\mathrm{D}(A)$

Before moving on to the norm-based methods, it is instructive to see how some of the earlier lower bound techniques (roughly) fit this duality framework.

The fooling set technique in this framework

One of the earliest lower bound methods we saw was the fooling set method: if function f has a fooling set of size k, then $\mathrm{D}(A)$ is at least as large as $\log k$. This fits into the above framework:

Step 1. Let $C(f)$ denote the minimum number of monochromatic rectangles partitioning the input space, and define G as $G(f) = \log(C(f))$. Then $\mathrm{D}(f) \geq G(f)$.

Note that, we didn't embed the problem into reals, but stayed in the discrete world.

Step 2. If F is any fooling set for f, then $C(f) \geq |F|$. So let $G^*(f)$ be defined as the log of the size of the largest fooling set. Then $G(f) = \log(C(f)) \geq G^*(f)$.

Step 3. Find a large fooling set of size k, then $G^*(f) \geq \log k$.

The rank method in this framework

The rank lower bound $\mathrm{D}(A) \geq \log \mathrm{rank}(A)$ also can be stated in this format.

Step 1. If we can find r pairs of vectors x_i, y_i such that $M_f = \sum_{i=1}^{r} x_i y_i^t$, then M_f has rank no more than r. We can always find $C(f)$ pairs of vectors like this. So $r \leq C(f)$. Similarly, the positive rank $\mathrm{rank}^+(M_f)$ is the smallest r for which we can find r pairs of *non-negative* vectors as above. Clearly,

$$\mathrm{rank}(M_f) \leq \mathrm{rank}^+(M_f) \leq C(f)$$

We can use $\log \mathrm{rank}(M_f)$ or $\log \mathrm{rank}^+(M_f)$ as $G(f)$.

Note that here we moved to the continuous domain: G is defined for any matrix over reals, not just sign matrices.

Step 2. This step is easy: $\mathrm{rank}(M)$ is most naturally stated as a maximization, namely, the maximum number of linearly independent rows or columns in M. (Similarly for rank^+, with an appropriate notion of independence.)

Step 3. Any large linearly independent set of rows / columns is a dual witness.

More general norm-based methods take off from here; norms are used to lower bound the rank. Thus $C(f) \geq \log(\mathrm{rank}(M_f)) \geq \log(\text{some norm of } M_f)$. Thus lower bounding such a norm is weaker than lower bounding rank itself. But it is very useful in randomised/multiparty settings, where generalisations of rank may not be easy to compute. For now, we will see such bounds for the deterministic case.

29.2.1 The Trace Norm

The trace norm of a matrix is defined as follows: Let M be an $m \times n$ matrix of rank r. Let σ be the vector of its r non-zero singular values, $\sigma = \langle \sigma_1, \sigma_2, \ldots, \sigma_r \rangle$. The trace norm of A is the ℓ_1 norm of σ;

$$\|A\|_{tr} = \|\sigma\|_1 = \sum_{i=1}^{r} |\sigma_i|$$

Recall that in contrast, the Frobenius norm of A is the ℓ_2 norm of σ, and also satisfies

$$\|A\|_F = \|\sigma\|_2 = \sqrt{\sum_{i=1}^{r} |\sigma_i|^2} = \sqrt{\sum_{i\in[m],j\in[n]} A_{ij}^2}$$

Using the Cauchy-Schwartz inequality, we see that

$$\|A\|_{tr} = \sum_{i=1}^{r} |\sigma_i| \leq \sqrt{r}\sqrt{\sum_{i=1}^{r} |\sigma_i|^2} = \sqrt{r}\|A\|_F.$$

Hence

Proposition 29.1. *For every matrix A,*

$$\operatorname{rank}(A) = r \geq \left(\frac{\|A\|_{tr}}{\|A\|_F}\right)^2.$$

In particular, if A is a sign matrix, then $\|A\|_F^2 = mn$. Hence for any sign matrix A, $\operatorname{rank}(A) \geq \frac{\|A\|_{tr}^2}{mn}$. Along with the log rank bound, this yields

Theorem 29.2. *For any $m \times n$ sign matrix A, $\mathrm{D}(A) \geq \log\left(\frac{\|A\|_{tr}^2}{mn}\right)$.*

Example 29.3. *If A is the sign matrix for the the Inner Product function, then it can be verified that $\|A\|_{tr}^2 = 2^{3n}$. So $\mathrm{D}(A) \geq \log \frac{2^{3n}}{2^n 2^n} = n$, and hence $\mathrm{D}(IP) = n$.*

One drawback of the trace norm is that the lower bounds it yields are not "monotone". For instance, consider the sequences of matrices H_k, M_k defined as follows: $H_1 = \begin{pmatrix} +1 & +1 \\ +1 & -1 \end{pmatrix}$, $H_k = H_1^{\otimes k}$, $M_k = \begin{pmatrix} H_k & J_k \\ J_k & J_k \end{pmatrix}$, where J_k is the $2^k \times 2^k$ all-ones matrix. It can be verified that $\|H_k\|_{tr} = 2^{3k/2}$, and $\|M_k\|_{tr} = 2^{3k/2} + 3 \cdot 2^k$. Clearly, $\mathrm{D}(H_k) \leq \mathrm{D}(M_k) \leq 2 + \mathrm{D}(H_k) \leq 2 + k$. Using the trace norm bound, we get the lower bound $\mathrm{D}(H_k) \geq \log 2^{3k}/(2^k \times 2^k) = k$. But we only get

$$\mathrm{D}(M_k) \geq \log\left(\frac{(2^{3k/2} + 3 \cdot 2^k)^2}{2^{k+1} \times 2^{k+1}}\right) \sim k - 2.$$

Though the complexity has increased in going from H_k to M_k, the lower bound via the trace bound has decreased. Intuitively, this happens because the norm for M_k gives equal weightage to all parts of M_k, whereas we need to somehow focus on the "difficult" part H_k.

29.2.2 The γ_2 Norm

This bound attempts to rectify the non-monotonicity of the trace norm bound. It allows us to focus on any chosen submatrix and take its trace norm, and selects the choice that maximises the value. A submatrix can be picked by taking the pointwise product (also known as Hadamard product or Schur product) of A with a suitably chosen 0-1 matrix

B. (The Hadamard product of two matrices X, Y, denoted $Z = X \circ Y$ is defined as $Z_{ij} = X_{ij}Y_{ij}$.) In the γ_2 norm, we further relax the condition that B is 0-1, but require it to be of normalised rank 1 (expressible as uv^t for vectors unit u, v). That is,

$$\gamma_2(A) = \max_{u,v:\|u\|=\|v\|=1} \|A \circ uv^t\|_{tr}$$

Generalizing the trace norm bound, we use Proposition 29.1 to see that for every matrix A, and every rank one matrix $W = uv^t$, $\text{rank}(A) \geq \text{rank}(A \circ uv^t) \geq \left(\frac{\|A \circ uv^t\|_{tr}}{\|A \circ uv^t\|_F}\right)^2$. Further, if A is a sign matrix and u, v are unit vectors, then $\|A \circ uv^t\|_F^2 = \|u\|_2^2 \cdot \|v\|_2 = 1$. Hence

$$\text{rank}(A) \geq (\gamma_2(A))^2.$$

This, with the log rank bound, yields

Theorem 29.4. *For any $m \times n$ sign matrix A, $\mathrm{D}(A) \geq 2 \log \gamma_2(A)$.*

Note that the γ norm of a matrix can be computed efficiently (Step 3 of the General Strategy) using semi-definite programming.

29.2.3 The μ Norm and the nuclear ν norm

Recall the monochromatic cover bound: $2^{\mathrm{D}(A)} \geq C(f) = \min\{r : A = \sum_{i=1}^{r} \alpha_i R_i\}$ where each α_i is in $\{\pm 1\}$ and each R_i is a rectangle, that is, a rank-one Boolean matrix. Relaxing the rectangle constraint by allowing R_i to be any rank-one matrix gives the log rank bound. Instead, we can relax the constraint on the weights α_i; then, instead of the number of α_i, the measure is the total weight. This gives the μ norm, formally defined below.

$$\mu(A) = \min\left\{\sum |\alpha_i| : \begin{array}{l} A = \sum_i \alpha_i R_i,\ \alpha \in \mathbb{R},\ R_i \text{ is Boolean rank-one} \\ \text{(a combinatorial rectangle)} \end{array}\right\}$$

Similarly, allowing arbitrary real weights and also relaxing the rectangle constraint to allow any rank-one sign matrix gives the nuclear norm $\nu(A)$ formally defined below.

$$\nu(A) = \min\left\{\sum |\alpha_i| : A = \textstyle\sum_i \alpha_i x_i y_i^t,\ \alpha \in \mathbb{R},\ x_i,\ y_i \text{ are sign vectors}\right\}$$

Clearly, $\text{rank}(A)$ is bounded from below by $\mu(A)$ and $\nu(A)$. It is known that for every matrix A,

$$\gamma_2(A) \leq \nu(A) \leq \mu(A) \leq 4\nu(A)$$

Hence

Theorem 29.5. *For any $m \times n$ sign matrix A, $\mathrm{D}(A) \geq \log \mu(A) \geq \log \nu(A)$.*

29.2.4 Bounding a norm via the dual norm

To use the γ, μ or ν (or any other) norms to bound $\mathrm{D}(A)$, we must bound the norm itself. This can be done via the dual norm defined below – this is Step 2 of the general strategy.

Definition 29.6. *For any norm Φ, the dual norm Φ^* is given by the expression*

$$\Phi^*(A) = \max_{Z:\Phi(Z)\leq 1} \langle A, Z\rangle = \max_{Z:\Phi(Z)\leq 1} tr(AZ^t)$$

Unfolding this definition, we can show that $\nu^*(A)$ is given by the *infinitiy-to-one norm* $\|A\|_{\infty\to 1}$; $\mu^*(A)$ is given by the *cut-norm* of A; and $\gamma^*(A)$ is given by a kind of SDP relaxation of μ^*.

$$\begin{aligned}\nu^*(A) &= \max_{y:\|y\|_\infty\leq 1} \|Ay\|_1 \\ \mu^*(A) &= \max_{x\in\{0,1\}^m, y\in\{0,1\}^n} \sum_{i,j} A_{ij}x_iy_j\end{aligned}$$

29.3 Approximate norms and randomized communication

Bounding randomized communication needs a further extension of the notion of norms, to approximate norms. This is defined as follows:

Definition 29.7 (Approximate Norm)**.** *Let Φ be any norm, $\alpha \geq 1$ any real value, and A a sign matrix. The α-approximate Φ norm of A, denoted $\Phi^\alpha(A)$, is given by the expression*

$$\Phi^\alpha(A) = \min_{B:\ \forall i,j,\ 1\leq A_{ij}B_{ij}\leq\alpha} \Phi(B)$$

The approximate norm bears a relation to randomized communication similar to the relation between the norm and deterministic communication.

Theorem 29.8. *For any norm Φ, if there is an absolute constant c such that for every sign matrix A, $2^{c\,\mathrm{D}(A)} \geq \Phi(A)$, then for every sign matrix A, for every $\varepsilon < 1/2$, and for $\alpha = \frac{1}{1-2\varepsilon}$ the following is true:*

$$c\,\mathrm{R}_\varepsilon(A) \geq \log\Phi^\alpha(A) - \log\alpha$$

Proof. Recall the mainstay of randomized communication lower bounds, Yao's lemma:

$$\mathrm{R}_\varepsilon(A) = \max_\mu \mathrm{D}^\mu_\varepsilon(A)$$

The hard direction of this lemma is $\mathrm{R}_\varepsilon(A) \leq \max_\mu \mathrm{D}^\mu_\varepsilon(A)$. Viewing an optimal randomized protocol as a distribution on deterministic protocols, we see that there are sign matrices $B_1, B_2, \ldots, B_m$, and a probability distribution $(p_1, \ldots, p_m)$, such that for each $i \in [m]$, $\mathrm{D}(B_i) \leq \mathrm{R}_\varepsilon(A)$, and $\|A - \sum_{i\in[m]} p_iB_i\| \leq 2\varepsilon$.

Consider the matrix $B = \frac{1}{1-2\varepsilon}\sum_i p_iB_i = \alpha\sum_i p_iB_i$. Then, from the error bound of the protocol, we have for each x, y, $1 \leq A(x,y)B(x,y) \leq \alpha = \frac{1}{1-2\varepsilon}$. Hence

$$\Phi^\alpha(A) \leq \Phi(B) \leq \alpha\sum_i p_i\Phi(B_i) \leq \alpha\max_i \Phi(B_i) \leq \alpha\max_i 2^{c\,\mathrm{D}(B_i)} \leq \alpha 2^{c\,\mathrm{R}_\varepsilon(A)},$$

yielding the claimed bound. □

29.4 Unifying these techniques via LP Duality: the Jain-Klauck partition bound

For randomized communication, we have seen several kinds of lower bounds:

- Rectangle bounds – discrepancy, corruption bounds,
- Norm-based bounds - $\gamma_2, \ldots,$
- Information-theoretic bounds.

All of these can be stated in the language of linear programming (LP) duality. This gives the *partition bound* and a weaker version, the *smooth rectangle bound*. We will see now how to do this.

The first task is to formulate the search for an optimal randomzied protocol as a linear programming instance. Let $f : X \times Y \to Z$ be a partial function, and imagine that we have an ε-error randomized protocol Π for f. Construct (conceptually) a matrix with rows indexed by all combinatorial rectangles and columns by random coin choices of Π. (The number of rows is $(2^{|X|}-1)(2^{|Y|}-1)$; the number of columns can be infinite.) Fill up the entries as follows:

$$\text{Entry}(R,r) \quad = \quad \begin{cases} z & \text{if in deterministic protocol } \Pi_r, \ R \text{ is a leaf rectangle labeled } z \\ \perp & \text{otherwise} \end{cases}$$

Let $q(r)$ denote the probability that protocol Π_r is chosen. We want to search for the best q with acceptable error. We write the first set of constraints to bound the error. For each x, y such that $f(x,y)$ is defined, we require

$$1-\varepsilon \leq \Pr_r[\Pi(x,y) = f(x,y)] = \sum_{R \ni (x,y)} \left[\sum_{r: R \text{ is a leaf in } \Pi_r, \text{ labeled } f(x,y)} q(r) \right] = \sum_{R \ni (x,y)} w'_{f(x,y),R}$$

where we define variable $w'_{z,R}$ as the inner sum

$$w'_{z,R} = \Pr[R \text{ is a leaf labeled } z] = \sum_{r: R \text{ is a leaf in } \Pi_r, \text{ labeled } z} q(r).$$

Thus for each $x, y \in f^{-1}$ we have a constraint:

$$\sum_{R \ni (x,y)} w'_{z,R} \geq 1-\varepsilon$$

Also, for each $z \in Z$ and each rectangle R we have a constraint:

$$0 \leq w'_{z,R} \leq 1$$

Now we fill in the remaining constraints. For every x, y, the protocol Π must output **some** value. That is, $\sum_{z \in Z} \Pr[\Pi(x,y) = z]$ should be 1. Using the new variables we have defined, we express this constraint as:

$$\sum_{z \in Z} \sum_{R \ni (x,y)} w'_{z,R} = 1.$$

Finally, we need to relate the variables w' to the cost c of the protocol.

$$\begin{aligned} 2^c = 2^c \sum_r q(r) &\geq \sum_r q(r) \,[\#\text{ of leaf rectangles in } \Pi_r] \\ &= \sum_r q(r) \left[\sum_z \sum_{R:\text{leaf in } \Pi_r \text{ labeled } z} 1 \right] \\ &= \sum_z \sum_R \left[\sum_{r:R \text{ is a leaf in } \Pi_r \text{ labeled } z} q(r) \right] \\ &= \sum_z \sum_R w'_{z,R} \end{aligned}$$

Putting all this together, we can now define the primal LP whose optimum is $2^{\mathrm{R}_\varepsilon(f)}$:

$$\begin{aligned} &\min \sum_z \sum_R w_{z,R} \\ &\text{subject to} \\ \forall (x,y) \in f^{-1} : \sum_{R \ni (x,y)} w_{f(x,y),R} &\geq 1-\varepsilon \\ \forall (x,y) : \sum_z \sum_{R \ni (x,y)} w_{z,R} &= 1 \\ \forall z, R : w_{z,R} &\geq 0 \end{aligned}$$

Note that for any function f, this LP is feasible: a naive deterministic protocol for f gives a feasible point. The LP is also bounded, because the objective function is non-negative. Hence by strong duality, the optimum of this primal LP equals the optimum of the dual LP, and any feasible solution for the dual LP lower bounds the opt of the primal. Let us write the dual LP. Use multipliers μ_{xy} for the error constraints, and multipliers ψ_{xy} for the output-some-value constraints. Then the dual LP can be formulated as:

$$\begin{aligned} &\max \sum_{(x,y) \in f^{-1}} (1-\varepsilon)\mu_{xy} + \sum_{(x,y)} \psi_{xy} \\ &\text{subject to} \\ \forall (x,y) \in f^{-1} : \mu_{xy} &\geq 0 \\ \forall z, R : \sum_{(x,y) \in f^{-1} \cap R} \mu_{xy} + \sum_{(x,y) \in R} \psi_{xy} &\leq 1 \end{aligned}$$

The preceding discussion directly leads to the partition bound as stated below:

Theorem 29.9 (Partition Bound; Jain & Klauck [JK10])**.** *For any function f, let $\mathrm{prt}_\varepsilon(f)$ denote the optimum value of the objective function of the primal/dual LPs formulated above. Then*

$$\mathrm{R}^{pub}_\varepsilon(f) \geq \log \mathrm{prt}_\varepsilon(f)$$

For each of the methods discrepancy, generalised discrepancy, norm-based bounds, Jain and Klauck showed that the method yields a dual-feasible point of the dual LP. Thus this partition bound subsumes **all** earlier duality-based bounds!

One may ask why bother to study other methods now that we have the stronger partition bound. The point is that the partition bound is a unifying framework, but it does not tell us how to get the dual witness. For that step we still need a more specific method.

29.5 Topics not covered

- Norm-based methods: we just skimmed the surface today. There's a wealth of stuff here ...
- Quantum communication complexity. In the quantum world, we need different lower bound techniques.
- The polynomial hierarchy in communication complexity ...
- Applications of multiparty (NOF) communication complexity:
  - Time-space tradeoffs
  - Hardness amplification in proof complexity
  - Simulation of AC^0 circuits by majority circuits
  - Pseudo randomness
- Applications of Direct Sum results.

References

[JK10] Rahul Jain and Hartmut Klauck. *The partition bound for classical communication complexity and query complexity.* In *Proc. 25th IEEE Conference on Computational Complexity*, pages 247–258. 2010. arXiv:0910.4266, doi:10.1109/CCC.2010.31.