COS 598D Lecture 10
Applications of Group Representation

Lecturer: Boaz Barak   Scribe: Moritz Hardt

April 18, 2008

In this lecture, we study fast matrix multiplication using techniques from the representation theory of non-Abelian groups. Secondly, we will see two explicit constructions of *dimension expanders*, a sort of generalization of expander graphs.

1 Fast Matrix Multiplication

We let ω denote the least exponent such that two $n \times n$ matrices can be multiplied with $O(n^{\omega+\epsilon})$ arithmetic operations for every $\epsilon > 0$. It is clear that $\omega \geq 2$, while Strassen showed that ω is strictly less than 3. Today it is widely believed that $\omega = 2$, although the best upper bound is roughly 2.34 due Coppersmith and Winograd. We will see a somewhat worse upper bound based on a group theoretic approach due to Cohen and Umans.

1.1 Strassen's main insight

Strassen showed that finding asymptotically fast matrix multiplication algorithms reduces to a finite problem. Namely, how many multiplications are necessary in order to multiply a $k \times k$ matrix for some constant k?

Lemma 1 (Strassen '69) *If there exists a $k \geq 2$ such that there is an algorithm which multiplies $k \times k$ matrices using k^ω multiplications, then we can multiply $n \times n$ matrices using $O(n^\omega)$ multiplications.*

Proof. The proof idea is to use recursion. Given two $n \times n$ matrices, we split each of them into $k \times k$ blocks. Now, we multiply the two matrices using the k^ω algorithm treating each block as a number. Whenever we have to multiply two blocks, we invoke a recursive call. Hence, on every fixed input size we invoke k^ω recursive calls.

Assuming $n = k^l$ for some positive integer l, we can compute the runtime of this algorithm using the recurrence equation

$$T(k^l) = k^\omega T(k^{l-1}) + f(k)k^{2l} = O(k^{l\omega}) = O(n^\omega).$$

□

Fact 2 (Strassen '69) *Two 2×2 matrices can be multiplied using $2^{\log 7} = 7$ multiplications.*

Using the previous fact, this gives us an $n^{\log 7}$ matrix multiplication algorithm where $\log 7 \approx 2.81 < 3$.

1.2 Bilinear Maps and Tensors

Before we proceed, we will mention a useful characterizations of the matrix multiplication exponent ω. The rank of a bilinear map $\phi\colon U \times V \to W$ is the least r such that

$$\phi(u, v) = \sum_{i=1}^{r} f_i(u) g_i(v) w_i, \tag{1}$$

where f_i (and g_i) are linear forms over U (and V), and $w_i \in W$.

Matrix multiplication is a bilinear map $\phi(A, B) = AB$ over the vector space $\mathbb{R}^{k\times k}$. Suppose the rank of ϕ is at most r for some k. Then, we can express $n \times n$ matrix multiplication for $n = k^{i+1}$ using (1) as $AB = \sum_{i=1}^{r} F_i(A)G_i(B)M_i$, where M_i is $k \times k$ and $F_i(A)$ is a $k \times k$ block decomposition of $k^i \times k^i$ matrices (likewise $G_i(B)$). Notice to compute AB we need precisely r multiplications of the form $F_i(A)G_i(A)$. Hence, this gives rises to the recursive algorithm of Lemma 1 and we obtain the following theorem.

Theorem 1 (Strassen) *If the rank of $k \times k$ matrix multiplication is at most r for some $k > 1$, then $\omega \leq \log_n r$.*

Often it is useful to think of bilinear maps as *tensors*. Every bilinear map $\phi\colon U \times V \to W$ corresponds uniquely to a tensor $t \in U^* \otimes V^* \otimes W$. This tensor is called the *structural tensor* of ϕ. In the case of $n \times n$ matrix multiplication we denote the structural tensor by $\langle n \rangle$.

1.3 The Group Representation Approach

The idea behind this approach is that matrix multiplication can be reduced to multiplication in the group algebra of suitable non-Abelian groups. The group algebra of a group G denoted $\mathbb{C}[G]$ is the set of formal sums $\sum_{g\in G} c_g g$ with the cyclic convolution as product between such sums. The group algebra is isomorphic to $\mathbb{C}^{d_1\times d_1} \times \cdots \times \mathbb{C}^{d_k\times d_k}$ where d_i denotes the dimension of the i-th irreducible group representation ρ_i. The isomorphism is given by $\sum c_g g \mapsto \bigoplus_i \sum_g c_g \rho_i(g)$. In particular, we can multiply two elements in the group algebra by multiplying k matrices of dimension $d_1 \times d_1, \ldots, d_k \times d_k$. The cost for this operation is $\sum_i d_i^{\omega}$. The specific criterion that G needs to satisfy is given in the next theorem.

Theorem 2 (Cohn, Umans '03) *Let G be a group of size n^{α} for some constant α with subsets S, T, U of cardinality n such that for all $s_1, s_2 \in S, t_1, t_2 \in T$ and $u_1, u_2 \in U$,*

$$s_1 s_2^{-1} t_1 t_2^{-1} u_1 u_2^{-1} = 1 \iff s_1 s_2^{-1} = t_1 t_2^{-1} = u_1 u_2^{-1} = 1. \tag{2}$$

Then,

$$n^{\omega} \leq \sum_i d_i^{\omega}.$$

where $d_1, \ldots, d_k$ are the dimensions of the irreducible representations of G.

It can be shown that if a group satisfies the assumption of the theorem, then α is between 2 and 3. Further, any Abelian group has $\alpha = 3$.

Proof. Let $|S| = k$ and suppose A, B are $k \times k$ matrices. Consider the product

$$\left(\sum_{s\in S, t\in T} A_{st}s^{-1}t\right)\left(\sum_{t'\in T, u\in U} B_{t'u}t'^{-1}u\right)$$

in the group algebra. By (2), we have

$$(s^{-1}t)(t'^{-1}u) = s'^{-1}u'$$

if and only if $s = s'$, $t = t'$ and $u = u'$. Hence, the coefficient of $s^{-1}u$ in the product is

$$\sum_{t\in T} A_{st}B_{tu} = (AB)_{su}.$$

This means we can multiply two $n \times n$ matrices at the cost of multiplication in the group algebra of G. By our previous discussion, this shows $n^\omega \leq \sum_i d_i^\omega$. □

The following corollary will be helpful in applying the theorem later.

Corollary 3 *Under the assumptions of the previous theorem, if* $\max d_i = |G|^{\frac{1}{\gamma}}$ *and* $2 \leq \alpha < \gamma$, *then* $\omega \leq \alpha\frac{\gamma-2}{\gamma-\alpha}$.

Proof.

$$n^\omega \leq \sum_i d_i^2 \cdot d_i^{\omega-2} \leq (\max d_i)^{\omega-2}\sum_i d_i^2 = n^{\frac{\alpha}{\gamma}(\omega-2)}n^\alpha.$$

Hence,

$$\omega \leq \frac{\alpha}{\gamma}(\omega-2) + \alpha.$$

□

It has been conjectured that using this approach one can show $\omega = 2$. We will next see an example of a group which achieves $\omega < 3$ even though the exact constant will be worse than in Strassen's algorithm. However, Cohn, Kleinberg, Szegedy and Umans '05 gave an example of a group that achieves $\omega < 2.41$.

1.4 Example for $\omega < 3$

For two groups G, H we define the semi-direct product $G \rtimes H$ to be the group induced by the group operation $(g, h) \times (g', h') = (g' \cdot h'(g), h \cdot h')$ where $g, g' \in G$ and $h, h' \in H$. Here we associated with every element $h \in H$ and automorphism on the group G.

To make this concrete, let $A = \mathbb{Z}_{17}$, the Abelian group of integers modulo 17 and let $G = (A^3)^2$. We think of elements in G as rectangular arrays, e.g., $\begin{array}{|c|c|c|}\hline 2 & 8 & 6 \\ \hline 3 & 0 & 1 \\ \hline\end{array}$. Let $H = S_2 = \{\mathrm{id}, f\}$, the symmetry group of two elements. Here, we think of f as an operation that flips the rows of an element in G, e.g, $f\left(\begin{array}{|c|c|c|}\hline 2 & 8 & 6 \\ \hline 3 & 0 & 1 \\ \hline\end{array}\right) = \begin{array}{|c|c|c|}\hline 3 & 0 & 1 \\ \hline 2 & 8 & 6 \\ \hline\end{array}$.

Now, define three sets of $S, T, U \subseteq G \rtimes H$ as follows:

$$S = \left\{\left(\begin{array}{|c|c|c|}\hline g_1 & 0 & 0 \\ \hline 0 & g_2 & 0 \\ \hline\end{array}, h\right) \mid g_1, g_2 \in G, h \in H\right\},$$

$$T = \left\{\left(\begin{array}{|c|c|c|}\hline 0 & g_1 & 0 \\ \hline 0 & 0 & g_2 \\ \hline\end{array}, h\right) \mid g_1, g_2 \in G, h \in H\right\},$$

$$U = \left\{\left(\begin{array}{|c|c|c|}\hline 0 & 0 & g_1 \\ \hline g_2 & 0 & 0 \\ \hline\end{array}, h\right) \mid g_1, g_2 \in G, h \in H\right\},$$

where neither g_1 nor g_2 may be zero. By case analysis, we can verify that these sets satisfy the requirement (2).

Also, we have $n = |S| = |T| = |U| = 2(|A|-1)^2$. On the other hand, $|G| = 2|A|^6$. Hence, $\alpha < 3$. On the other hand, $\max_i d_i = 2$. Computing γ and applying Corollary 3, this leads to the bound $\omega < 2.91$.

2 Dimension Expanders

We now come to our second application of group representation theory.

Definition 1 A set of matrices $A_1, \ldots, A_d \in \mathbb{F}^{n\times n}$ is called an ϵ-*dimension* expander if for every subspace $V \subseteq \mathbb{F}^n$ of dimension $\dim(V) < \frac{n}{2}$, we have

$$\dim\{V + A_1V + \cdots + A_dV\} \geq (1+\epsilon)\dim V.$$

Dimension expanders can be thought of as a stronger notion than expander graphs. To see this take $\mathbb{F} = \mathbb{F}_2$ (the binary field) and consider the graph on the set of vertices $\mathbb{F}_2^n$ with edges to $A_0v, A_1v, \ldots, A_dv$ from every vertex v. Fix some k-dimensional subspace V which we think of as a set of vertices in this graph of size 2^k. Then, we have the following different guarantees for expander graphs and dimension expanders:

$$|\Gamma(v)| \geq (1+\epsilon)2^k \qquad \text{(Expander Graphs)}$$

$$|\mathrm{span}(\Gamma(v))| \geq 2^{k(1+\epsilon)} \qquad \text{(Dimension Expanders)}$$

Random matrices give us good dimension expanders. We will demonstrate this argument over $\mathbb{F}_2$.

Lemma 4 *Let $A_1, \ldots, A_d$ be $n \times n$ matrices over $\mathbb{F}_2$ with i.i.d. 0/1 entries. Then, $A_1, \ldots A_d$ is a 1.1-dimension expander for $d \geq 10$.*

Proof. Fix subspaces V of dimension k and U of dimension $1.1k < n/2$. We have

$$\Pr_{A_i}(\forall i\colon A_iV \subseteq U) \leq 2^{-nkd/2}.$$

Since there are $2^{nk} \cdot 2^{1.1nk} = 2^{2.1nk}$ choices for U and V, the union bound finishes the proof. □

One original motivation to study dimension expanders came from the problem of explicitly constructing rigid matrices. The idea was that perhaps one could show (1) sparse matrices $B_0, \ldots, B_d$ cannot be dimension expanders in the sense that there is a subspace V of dimension $n/10$ such that $\dim\{B_0V + \cdots + B_dV\} \leq (1+o(1))n/10$, and (2) give an explicit construction of dimension expanders $A_0, \ldots, A_d$.

If these two statements were true, one would get rigid matrices as follows. Assuming (1), we cannot have that

$$\begin{pmatrix} A_0 \\ \vdots \\ A_d \end{pmatrix} = \begin{pmatrix} \text{low} \\ \text{rank} \end{pmatrix} + \begin{pmatrix} \text{sparse} \\ \ \end{pmatrix},$$

since neither term of the RHS would expand the dimension of subspace.

Unfortunately, this conjecture is false. There are now constructions of *sparse* dimension expanders.

2.1 Over the Complex Numbers

Lubotzky and Zelmanov give a construction of dimension expanders over the complex numbers based on the image of irreducible group representations on a generating set.

Theorem 3 (Lubotzky, Zelmanov) *Let G be a finite group, and let S denote a generating set of G so that $\lambda(C(G,S)) \leq 1-\epsilon$. Here, $C(G,S)$ denotes the Cayley graph and λ is its second largest eigenvalue. Further let $\rho\colon G \to U_n$ denote an irreducible representation of G. Then, $\{\rho(s) \mid s \in S\}$ is an $\frac{\epsilon}{100|S|}$-dimension expander over $\mathbb{C}^n$.*

Proof of Theorem 3

Fix G and S. For every representation ρ we let $A_\rho = \frac{1}{|S|}\sum_{s\in S}\rho(s)$. Notice, A_{REG} is just the normalized adjacency matrix of $C(G,S)$. Every eigenvalue of A_ρ is also an eigenvalue of A_{REG} and also every eigenvalue of A_{REG} is an eigenvalue of A_ρ for some irreducible representation ρ. Indeed, if $\rho = \rho_1 \oplus \rho_2$, then every eigenvalue of ρ is either also an eigenvalue of ρ_1 or ρ_2. More precisely, every eigenvector v of ρ_1 with corresponding eigenvalue λ extends to an eigenvector of ρ as $(v,0)$ with the same eigenvalue. Since

$$\lambda(G,S) = \max_{0\neq v\perp \mathbf{1}} \frac{\langle v, A_{\mathrm{REG}}\rangle}{\langle v, v^*\rangle} = \frac{1}{|S|}\sum_s \frac{\langle v, \mathrm{REG}(s)v\rangle}{\langle v, v^*\rangle},$$

we have the following fact.

Fact 5 *If $\lambda(C(G,S)) \leq 1-\epsilon$, then for every vector v there exists an element $s \in S$ such that $\|v - \mathrm{REG}(s)v\|^2 \geq \frac{\epsilon'}{|S|}\|v\|^2$ for some $\epsilon' > 0$. Here, REG denotes the regular representation over some complex Hilbert space $\mathcal{H}$ and $v \in \mathcal{H}$.*

So, let us consider the following constant (called Kazhdan constant)

$$\begin{aligned} K(G,S) &= \max_{0\neq v\perp \mathbf{1}} \max_{s\in S} \frac{\|\mathrm{REG}(s)v - v\|^2}{\|v\|^2} \\ &= \min_\rho \min_{v\neq 0} \max_{s\in S} \frac{\|\rho(s)v - v\|^2}{\|v\|^2}, \end{aligned}$$

where the minimum in the second line is taken over all vectors v that are not fixed vectors of ρ. We will apply Fact 5 to the *adjoint representation* $\operatorname{adj}\rho$ defined as

$$\operatorname{adj}\rho(\gamma)A = \rho(\gamma)A\rho(\gamma^{-1}).$$

where $A \in \mathbb{C}^{n\times n}$. We think of $\operatorname{adj}\rho$ as a representation over the Hilbert space $\mathbb{C}^{n\times n}$ where we have the inner product $\langle A, B\rangle = \mathrm{tr}(AB^*)$. We remark that $\operatorname{adj}\rho$ is invariant on the n^2-1 dimensional subspace $\{A \mid \mathrm{tr}(A) = 0\}$. If ρ be an irreducible representation. It turns out, $\operatorname{adj}\rho$ has no fixed nonzero vector. To see this, suppose $\operatorname{adj}\rho(g)A = A$. Then $A = \rho(g)A\rho(g^{-1})$. This means that A is from the invariant subspace of $\operatorname{adj}\rho$ and hence has $\mathrm{tr}(A) = 0$. But we assumed ρ was irreducible. Therefore, by Schur's Lemma, A is either the identity matrix or the zero matrix. But, the identity matrix does not have trace zero. Hence, A must be the zero matrix.

Now, fix a subspace $V \subseteq \mathbb{C}^n$ of dimension $k < n/2$ and let P denote the linear projection onto V. Consider the matrix

$$A = P - \frac{k}{n}I.$$

We have $\mathrm{tr}(A) = \mathrm{tr}(P) - \frac{k}{n}\mathrm{tr}I = 0$. By the assumption of our theorem and Fact 5, we have that there exists an s such that

$$\|\mathrm{adj}\,\rho(\gamma)A - A\|^2 \geq \epsilon\|A\|^2,$$

where

$$\|A\|^2 = \mathrm{tr}\Big((P - \frac{k}{n}I)(P - \frac{k}{n}I)^*\Big) = \mathrm{tr}(P^2) - \frac{k}{n^2}\mathrm{tr}I = k - \frac{k^2}{n} \geq k/2.$$

On the other hand,

$$\mathrm{adj}\,\rho(\gamma)A = \rho(\gamma)P\rho(\gamma^{-1}) - \frac{k}{n}I = P' - \frac{\mathrm{tr}P'}{n}I,$$

where $P' = \rho(\gamma)P\rho(\gamma^{-1})$ is the projection onto the subspace $V' = \rho(\gamma)V$.

Hence,

$$\epsilon k/2 \leq \|\mathrm{adj}\,\rho(\gamma)A - A\|^2 = \|P' - P\|^2,$$

and the following lemma finishes the proof.

Lemma 6 *If P, P' are projection matrices of k-dimensional subspaces V and V', respectively, such that $\|P - P'\|^2 \geq \epsilon k$, then* $\dim(V + V') \geq (1 + \epsilon')k$ *for $\epsilon' > 0$.*

Proof.

$$\begin{aligned}\|P - P'\|^2 &= \langle P' - P, P' - P\rangle \\ &= \langle P', P'\rangle + \langle P, P\rangle - \langle P, P'\rangle - \langle P', P\rangle \\ &= 2k - 2\mathrm{Re}(\mathrm{tr}PP').\end{aligned}$$

We claim that $\mathrm{Re}(\mathrm{tr}PP') \geq 4k - 3\dim(V + V')$. Notice, the operator PP' is the identity on $V \cap V'$ (its trace being $\dim(V \cap V')$), and it is zero on $(V + V')^{\perp}$. Also, the trace is at least -1 on $(V + V')\backslash(V \cap V')$. Hence,

$$\mathrm{Re}(\mathrm{tr}(PP')) = 2\dim(V \cap V') - \dim(V + V') = 4k - 3\dim(V + V'),$$

using the fact that

$$\dim(V \cap V') = \dim(V) + \dim(V') - \dim(V + V') = 2k - \dim(V + V').$$

□

2.2 Over Finite Fields

Let $\mathbb{F}$ denote a finite field and consider the vector space $\mathbb{F}^n$ for some integer $n = 2m$. For an index $j \in \{0, \ldots, n-1\}$, we define the cyclic right shift Π_j by putting

$$\Pi_j(v_1, v_2, \ldots, v_n) = (v_{1-j}, v_{2-j}, \ldots, v_{n-j})$$

where we identify $v_0, v_{-1}, \ldots, v_{-j+1}$ with $v_n, v_{n-1}, \ldots, v_{n-j+1}$ as usual.

We also define the projections $P_L(v', v'') = (v'', 0)$ and $P_R(v', v'') = (0, v')$ where v', v'' denote vectors of length m each.

Theorem 4 (Dvir, Shpilka) *Let $J \subseteq \{1, \ldots, m\}$ of order $|J| = O(\log m)$ such that the Cayley graph of $\mathbb{Z}_m$ with respect to J is an expander, i.e., for every set $S \in \mathbb{Z}_m$ of size $|S| < m/2$ we have*

$$|\{s + j \mod m : s \in S, j \in J\}| \geq 1.1|S|.$$

Then, the family $\{\Pi_j \mid j \in J\} \cup \{P_L, P_R\}$ is an ϵ-dimension expander for some positive constant ϵ.

We remark that a construction of dimension expanders over finite fields for a constant number of matrices is currently not known.

Proof. For a vector v we define the degree of v, denoted $\deg(v)$, to be the largest coordinate i such that $v_i \neq 0$. For a subspace V, we let $D_V = \{\deg(v) \mid v \in V\}$. Clearly, $\dim(V) = |D_V|$, since vectors with distinct degrees are linearly independent.

Now, suppose V is a subspace of dimension $k < n/10$. We split the set of degrees into a left side $D_L = D_V \cap [m]$ and a right side $D_R = D_V \cap [m+1, 2m]$.

The set $D_R \backslash (D_L + m)$ contains all the new distinct degrees that we get when projecting the left side into the right side using P_L. Likewise, $D_L \backslash (D_R - m)$ counts the new degrees we get from applying P_R. If either of these sets is of size ϵk, we are done. So, suppose both sets are smaller than ϵk.

Consider the set $D_L + J$. Since both D_L and J are subsets of $[m]$ we have $\deg(\Pi_j(v)) = \deg(v) + j$ for every $v \in V$. Hence, the set of $D_L + J$ is contained in the set of degrees of the subspace $\sum_j \Pi_j(V)$. To show that we get many new distinct degrees in this set, consider $R = D_L + J \mod m$. This is the neighborhood of D_L in the Cayley graph. From our previous discussion, it follows that $D_L \cup (D_R - m)$ is less than $(1+\epsilon)k$. On the other hand $|R| > 1.1|D_L|$. Hence, for small enough ϵ, we have that $|R| \backslash (D_L \cup (D_R - m))| > \epsilon' k$ for some positive constant ϵ'. □