Lecture 19. The Kadison-Singer problem

The Kadison-Singer problem originated in quantum mechanics in the 1950s. It has been formulated in several equivalent forms. Here we state it in a form known as Weaver’s conjecture.

**Conjecture 1** There exist $\epsilon, \delta > 0$ such that given any vectors $w_1, w_2, \ldots, w_m \in \mathbb{C}^n$ where

- $\|w_i\|_2 \leq \delta$ for all $i$
- $\sum_{i=1}^m w_i w_i^T = I$

there exists a partition of $[m]$ into $S_1$ and $S_2$ such that for all $j \in \{1, 2\}$,

$$\|\sum_{i \in S_j} w_i w_i^*\| \leq 1 - \epsilon$$

(This holds independently of $m, n \in \mathbb{N}$.)

Here $\|\sum_{i \in S_j} w_i w_i^*\|$ is the operator norm, for a Hermitian matrix simply the maximum eigenvalue, so our goal will be to prove an upper bound on the maximum eigenvalue. Note that since $w_i w_i^*$ are rank 1 matrices that add up to the identity, we must have $m \geq n$.

The setup here is similar to that of Ramanujan graphs. For Ramanujan graphs, we wanted all of the eigenvalues of the adjacency matrix to be in the set $\{d\} \cup \{-d\} \cup [-2\sqrt{d-1}, 2\sqrt{d-1}]$. Here we are again looking to bound the size of the eigenvalues, and we will use a similar approach.

### 19.1 Setup for the proof of Conjecture 1

The specific statement we will prove is the following, which is actually stronger than Conjecture 1.

**Theorem 19.1 (Marcus, Spielman, Srivastava ’13)** For all $\alpha > 0$ and vectors $w_1, w_2, \ldots, w_m \in \mathbb{C}^n$ where

- $\|w_i\|_2^2 \leq \alpha$ for all $i$
- $\sum_{i=1}^m w_i w_i^T = I$

there exists a partition of $[m]$ into $S_1$ and $S_2$ such that for all $j \in \{1, 2\}$,

$$\|\sum_{i \in S_j} w_i w_i^*\| \leq \frac{1}{2}(1 + \sqrt{2\alpha})^2$$

(This holds independently of $m, n \in \mathbb{N}$.)
Our strategy is as follows. We define random variables \( r_1 \ldots r_m \), each in \( \mathbb{C}^{2n} \), where either\\n\[
    r_j = (w_j \mid 0, 0\ldots)
\]
meaning that the first \( n \) entries are the \( n \) entries of \( w_j \), and the second \( n \) entries are all zeros, or\\n\[
    r_j = (0, 0\ldots \mid w_j)
\]
meaning that the first \( n \) entries are all zeros and the second \( n \) entries are the \( n \) entries of \( w_j \). We will use \( r^{(1)}_j \) to refer to \((w_j \mid 0, 0\ldots)\) and \( r^{(2)}_j \) to refer to \((0, 0\ldots \mid w_j)\). Each \( r_j \) is \( r^{(1)}_j \) with probability 1/2 and is \( r^{(2)}_j \) with probability 1/2.

Thus we have\\n\[
    \sum_{j=1}^{m} r_j r^*_j = \sum_{j \in S_1} \left( \begin{array}{c} w_jw_j^* \\ 0 \end{array} \right) + \sum_{j \in S_2} \left( \begin{array}{c} 0 \\ w_jw_j^* \end{array} \right)
\]
where the matrices on the right hand side are \( 2n \times 2n \) matrix where one \( n \times n \) quadrant is occupied by \( w_jw_j^* \), and the rest of the entries are zeros.

Thus\\n\[
    || \sum_{j=1}^{m} r_j r^*_j || = \max \left( || \sum_{i \in S_1} w_iw_i^* ||, || \sum_{i \in S_2} w_iw_i^* || \right)
\]
and this is the quantity that we should upper-bound.

The main technical theorem is the following:

**Theorem 19.2 (Marcus, Spielman and Srivastava)** For independently random vectors \( r_1, \ldots, r_m \) in \( \mathbb{C}^n \), if \( \sum_{i=1}^{m} \mathbb{E}[r_i r_i^*] = I \) and \( \mathbb{E} [||r_i||^2] \leq \alpha \), then with probability greater than 0,

\[
    || \sum_{j=1}^{m} r_j r^*_j || \leq (1 + \sqrt{\alpha})^2
\]

From here, Theorem 19.1 follows by setting \( r_j = \sqrt{2} (w_j \mid 0) \) or \( \sqrt{2} (0 \mid w_j) \) with probability \( \frac{1}{2} - \frac{1}{2} \). Let us denote the first option by \( w^{(1)}_j \) and the second option by \( w^{(2)}_j \). In the following, we prove Theorem 19.2.

To bound the eigenvalues of \( \sum_{j=1}^{m} r_j r^*_j \), we consider the expected characteristic polynomial \( \chi \):

\[
    \chi(x) = \mathbb{E}_{r_1 \ldots r_m} \left[ \det(xI - \sum_{j=1}^{m} r_j r^*_j) \right]
\]

As before, we fix the values of the first \( k \) vectors \( r_1, \ldots, r_k \) to be \( w^{(\sigma_1)}_1, \ldots, w^{(\sigma_k)}_k \), for some values of \( \sigma_1, \ldots, \sigma_k \in \{1, 2\} \). The remaining \( m - k + 1 \) vectors remain random. This yields the following polynomial:

\[
    \chi_{\sigma_1, \ldots, \sigma_k}(x) = \mathbb{E}_{r_{k+1} \ldots r_m} \left[ \det(xI - \sum_{j=1}^{k} w^{(\sigma_j)}_j w^{(\sigma_j)*}_j - \sum_{j=k+1}^{m} r_j r^*_j) \right]
\]
We saw when studying Ramanujan graphs that this is an interlacing family of stable polynomials. We know from before that if the maximum root of $\chi$ is at most $\lambda_{max}$, then there exists a choice of $\sigma_1, \ldots, \sigma_m$ such that the maximum root of $\chi_{\sigma_1, \ldots, \sigma_m}$ is also at most $\lambda_{max}$.

In the setting of Ramanujan graphs, the characteristic polynomial was the matching polynomial, so we already knew its maximum root. Here, we do not know the maximum root of $\chi$ (if we did, we would be done), so we will have to do some more work to bound the maximum root.

19.2 Bounding the maximum root of $\chi$

Recall how we prove that $\chi$ is in fact stable. We will do this by showing that $\chi$ can be obtained applying a set of differential operators to a polynomial we already known to be stable:

$$\chi(x) = (1 - \frac{\partial}{\partial t_1}) (1 - \frac{\partial}{\partial t_2}) \ldots (1 - \frac{\partial}{\partial t_m}) \det \left( xI + \sum_{j=1}^m t_j E[r_i r_j^*] \right)$$

The matrix $E[r_i r_j^*]$ is positive semi-definite, so the polynomial $\det \left( xI + \sum_{j=1}^m t_j E[r_i r_j^*] \right)$ is stable. We show next that these differential operators preserve stability.

**Lemma 19.3** Let $z$ refer to the vector of variables $(z_1 \ldots z_m)$. If a function $f(z_1, z_2 \ldots z_m) = f(z)$ is stable, then so is $(1 - \frac{\partial}{\partial z_1}) f(z) = f(z) - \frac{\partial f}{\partial z_1}(z)$.

**Proof:** Recall that $\mathcal{H} = \{ z \in \mathbb{C} : \text{Im}(z) > 0 \}$. Suppose for sake of contradiction that $f(z) - \frac{\partial f}{\partial z_1}(z)$ were not stable: then there exists a root of $f(z) - \frac{\partial f}{\partial z_1}(z)$ in $\mathcal{H}^m$: that is, there exists $(\alpha_1 \ldots \alpha_m) \in \mathcal{H}^m$ such that $f(\alpha_1 \ldots \alpha_m) - \frac{\partial f}{\partial z_1}(\alpha_1 \ldots \alpha_m) = 0$. Then $f(\alpha_1 \ldots \alpha_m) = \frac{\partial f}{\partial z_1}(\alpha_1 \ldots \alpha_m)$.

Let $g(z)$ be the univariate polynomial $f(z_1, \alpha_2 \ldots \alpha_m)$. Then $g(\alpha_1) = f(\alpha_1 \ldots \alpha_m)$ and $g'(\alpha_1) = \frac{\partial f}{\partial z_1}(\alpha_1 \ldots \alpha_m)$, so we have $g(\alpha_1) = g'(\alpha_1)$.

If $g(z)$ had a root $\alpha' \in \mathcal{H}$, then $(\alpha' \ldots \alpha_m) \in \mathcal{H}^m$ would be a root of $f(z_1, \alpha_2 \ldots \alpha_m)$, which contradicts the stability of $f(z_1, \alpha_2 \ldots \alpha_m)$. Thus $g(z)$ is stable. Thus we can write

$$g(z) = C \prod_{i=1}^n (z - \lambda_i)$$

where $\lambda_1, \ldots, \lambda_n$ are the roots of $g(z)$. Therefore

$$\frac{g'(z)}{g(z)} = \sum_{i=1}^n \frac{1}{z - \lambda_i}$$

Examining the ratio of a function’s derivative divided by the function will be a useful tool at multiple points in this lecture.

Since $g(z)$ is stable, we know that $\lambda_i \notin \mathcal{H}$ for all $i$. Thus for all $z \in \mathcal{H}$, and all $i$, $z - \lambda_i \notin \mathcal{H}$, so $\frac{1}{z - \lambda_i} \in -\mathcal{H}$, where $-\mathcal{H}$ is the lower half-plane. Therefore $\frac{g'(z)}{g(z)} \in -\mathcal{H}$, so $\frac{g'(z)}{g(z)} \neq 1$ for all $z$.

But we know that $g'(\alpha_1) = g(\alpha_1)$, so $\frac{g'(\alpha_1)}{g(\alpha_1)} = 1$, which is a contradiction. Therefore $f(z) - \frac{\partial f}{\partial z_1}(z)$ must be stable.

Applying Lemma 19.3 repeatedly implies that $\chi$ is in fact stable. When $t_i = 0$ for all $i$, we have $\det \left( xI + \sum_{j=1}^m t_j E[r_i r_j^*] \right) = \det(xI) = x^n$, which has all of its roots at $x = 0$. If we could show that applying the $(1 - \frac{\partial}{\partial z_1})$ operators doesn’t increase the roots by much, that would give us an upper bound on the roots of $\chi$, and we would be done.
19.3 A toy example

Next, we examine a toy example: that of the univariate polynomial \( p(x) = x^n \). This polynomial initially has all of its roots at \( x = 0 \). Applying \((1 - \frac{d}{dx})\) once gives us

\[
(1 - \frac{d}{dx}) p(x) = x^n - nx^{n-1} = x^{n-1}(x - n)
\]

which has one root at \( x = n \). This is concerning, since applying the operator \((1 - \frac{d}{dx})\) just once caused the maximum root to jump from 0 to \( n \). The rest of the roots remain at \( x = 0 \). Applying the \((1 - \frac{d}{dx})\) operator again gives us

\[
(1 - \frac{d}{dx})^2 p(x) = x^n - nx^{n-1} - nx^{n-1} + n(n-1)x^{n-2}
= x^{n-2}(x^2 - 2nx + n(n-1))
\]

To compute the roots of this polynomial, we need to know the roots of \( x^2 - 2nx + n(n-1) \). This can be done using the quadratic formula, and the resulting roots are \( n + \sqrt{n} \) and one at \( n - \sqrt{n} \) (in addition to the \( n - 2 \) roots at \( x = 0 \) from \( x^{n-2} \)). This time, the maximum root increased only from \( n \) to \( n + \sqrt{n} \), which is a smaller jump than from 0 to \( n \).

Taking a wild guess, maybe the multiplicity of the roots influences the effect of the \((1 - \frac{d}{dx})\) on the maximum root? In the following, we design a way of relaxing the notion of maximum root in a way that takes multiplicity and proximity of other roots into account.

19.4 The barrier function

Given a univariate polynomial \( p(x) \), we define the “barrier function” \( \phi_p(x) \) for \( x \in \mathbb{R} \) by

\[
\phi_p(x) = \frac{d}{dx} \left( \log p(x) \right)
\]

Thus

\[
\phi_p(x) = \frac{p'(x)}{p(x)} = \sum_{i=1}^{n} \frac{1}{z - \lambda_i}
\]

where \( \lambda_1...\lambda_n \) are the roots of \( p \). (Note the re-appearance of the ratio of a function’s derivative divided by the function, or equivalently the derivative of the logarithm.)

For each \( \alpha \in (0, 1) \), we define a set \( \alpha_{\text{max}}(p) \) by

\[
\alpha_{\text{max}}(p) = \left\{ x > \text{maximum root of } p : \frac{p'(x)}{p(x)} < \alpha \right\}
\]

The intuition for Lemma 19.4 is that \( \alpha_{\text{max}}(p) \) doesn’t change too much when we apply \((1 - \frac{d}{dx})\).

**Lemma 19.4** If \( x \in \alpha_{\text{max}}(p) \), then for \( \delta = \frac{1}{1-\alpha} \), we have \( x + \delta \in \alpha_{\text{max}}(p - p') \).
**Proof:** We have

\[
\phi_{p-p'}(x) = \frac{d}{dx} \left( \log((p-p')(x)) \right) \\
= \frac{d}{dx} \left( \log(p(x) \cdot (1 - \phi_p(x))) \right) \\
= \frac{d}{dx} \left( \log p(x) + \log(1 - \phi_p(x)) \right) \\
= \phi_p(x) - \frac{\phi_{p'}(x)}{1 - \phi_p(x)}
\]

Using our above computation of \( \phi_{p-p'}(x) \), we have

\[
\phi_{p-p'}(x + \delta) = \phi_p(x + \delta) - \frac{\phi_{p'}(x + \delta)}{1 - \phi_p(x + \delta)}
\]

When \( x \) is larger than the maximum root, \( \phi_p(x) \) is decreasing and convex. Also, since \( x \in \alpha_{\text{max}}(p) \), we have \( \phi_p(x) = \frac{p'(x)}{p(x)} < \alpha \) by assumption. Therefore

\[
\phi_p(x + \delta) - \frac{\phi_{p'}(x + \delta)}{1 - \phi_p(x + \delta)} \leq \phi_p(x + \delta) - \frac{\phi_{p'}(x + \delta)}{1 - \phi_p(x)} \leq \phi_p(x + \delta) - \frac{\phi_{p'}(x + \delta)}{1 - \alpha} \\
\leq \phi_p(x + \delta) - \delta \phi_{p'}(x + \delta) \leq \phi_p(x) < \alpha
\]

(19.1) uses the fact that \( \phi_p(x) \) is decreasing when \( x \) is larger than the maximum root. (19.1) to (19.2) is because \( \phi_p(x) < \alpha \). (19.2) to (19.3) is by definition of \( \delta \). (19.3) to (19.4) is by convexity, and (19.4) to (19.5) is again because \( \phi_p(x) < \alpha \).

Putting everything together gives us \( \phi_{p-p'}(x + \delta) < \alpha \), so \( (x + \delta) \in \alpha_{\text{max}}(p-p') \), as required. \( \Box \)

Recall our toy example of \( p(x) = x^n \). Here \( \phi_p(x) = n/x \), so \( n/\alpha \in \alpha_{\text{max}}(p) \). So by applying Lemma 19.4 \( m \) times, we know that the polynomial \( (1 - \frac{d}{dx})^m p \) has maximum root less than \( \frac{n}{\alpha} + \frac{m}{1-\alpha} \).

This holds for any \( \alpha \). If we choose \( \alpha \) to be \( (1 + \sqrt{\frac{m}{n}})^{-1} \), this implies that the maximum root of \( (1 - \frac{d}{dx}) \) is at most \( (\sqrt{n} + \sqrt{m})^2 \). When \( m \) is smaller than \( n \), this is at most \( 4n \).

### 19.5 Back to the multivariate setting

Recall that

\[
\chi(x) = \left(1 - \frac{\partial}{\partial t_1}\right) \left(1 - \frac{\partial}{\partial t_2}\right) \ldots \left(1 - \frac{\partial}{\partial t_m}\right) \det(xI + \sum_{j=1}^{m} t_j \mathbb{E}[r_{i_1}])
\]

Let \( A_i = \mathbb{E}[r_{i_1}] \). The matrix \( A_i \) is positive semi-definite.

**Theorem 19.5** If \( \sum_{i=1}^{m} A_i = I \), \( A_i \) is positive semi-definite for all \( i \), and \( Tr(A_i) \leq \alpha \) for all \( i \), then the maximum root of \( \chi \) when evaluated at \( t = 0 \) is at most \( (1 + \sqrt{\alpha})^2 \).
First, we will do a variable substitution by letting \( z_i = x + t_i \). Then
\[
\det(xI + \sum_{i=1}^{m} t_i A_i) = \det(\sum_{i=1}^{m} (x + t_i) A_i) = \det(\sum_{i=1}^{m} z_i A_i)
\]
Since \( z_i \) is linear in \( t_i \), the operator \((1 - \frac{\partial}{\partial z_i})\) is equivalent to \((1 - \frac{\partial}{\partial t_i})\). Therefore
\[
\chi(x) = (1 - \frac{\partial}{\partial z_1}) \cdots (1 - \frac{\partial}{\partial z_m}) \det(\sum_{i=1}^{m} z_i A_i)
\]
We will evaluate this at \( z_i = x \), which corresponds to \( t = 0 \). Let \( Q(z_1 \ldots z_m) \) be defined by
\[
Q(z_1 \ldots z_m) = (1 - \frac{\partial}{\partial z_1}) \cdots (1 - \frac{\partial}{\partial z_m}) \bigg|_{z_i=x} \det(\sum_{i=1}^{m} z_i A_i)
\]
Also, for each \( k \in [m] \), let
\[
Q_k(z_1 \ldots z_m) = (1 - \frac{\partial}{\partial z_m-k+1}) \cdots (1 - \frac{\partial}{\partial z_m}) \bigg|_{z_i=x} \det(\sum_{i=1}^{m} z_i A_i)
\]
Thus \( Q_k(z_1 \ldots z_m) \) contains \( k \) differentiation operations.

We will also use a multivariate version of the barrier function. We define the “barrier function in direction \( j \)” \( \phi^j_p(z) \) by
\[
\phi^j_p(z) = \frac{\partial}{\partial z_j} (\log p(z)) = \frac{\partial p}{p(z)} \cdot \frac{\partial p}{\partial z_j}
\]
We say that \( w \in \mathbb{R}^m \) is “above” the roots of \( p \) (real stable) if \( \forall t \in \mathbb{R}^m \) where \( t \geq 0 \), we have \( p(w + t) > 0 \).

Fixing \( z_i \in \mathbb{R} \) for each \( i \neq j \), \( p(z_1 \ldots z_j \ldots z_m) \) has roots \( \lambda_1 \ldots \lambda_n \in \mathbb{R} \). We can write
\[
\phi^j_p(z_j) = \sum_{t=1}^{n} \frac{1}{z_j - \lambda_t}
\]
To be continued...