1 Coherence

Definition 1. Let \( A \in \mathbb{C}^{m \times N} \) be a matrix with \( \ell_2 \)-normalized columns (i.e. such that \( A = (\vec{a}_1, \ldots, \vec{a}_N) \) has \( \| \vec{a}_j \| = 1 \) \( \forall j \in [N] \)). The coherence of \( A \), \( \mu(A) \) is

\[
\mu(A) := \max_{i \neq j} |\langle \vec{a}_j, \vec{a}_i \rangle|.
\]

- If \( A \) is an orthogonal matrix, then, \( \mu(A) = 0 \). This is the best case.
- If \( A \) has two identical columns, then, \( \mu(A) = 1 \). This is the worst case.
- In general, \( \mu(A) \in [0, 1] \). The smaller it is, the better.
- It turns out that matrices with low coherence have the R.I.P..
- In order to prove this we need Gershgorin’s Disk Theorem.

Theorem 1 (Gershgorin’s Disk Theorem). Let \( \lambda \) be an eigenvalue of a square matrix \( A \in \mathbb{C}^{m \times m} \). Then, \( \exists \) an index \( j \in [m] \) such that \( |\lambda - A_{j,j}| \leq \sum_{t \in [m] - \{j\}} |A_{j,t}|. \)

Proof: Let \( \vec{u} \in \mathbb{C}^m - \{0\} \) be an eigenvector for \( \lambda \). Let \( j \in [m] \) be such that \( |\vec{u}_j| = \|\vec{u}\|_\infty \). Then, \( \sum_{t \in [m]} A_{j,t} u_t = \lambda u_j \) such that \( \sum_{t \in [m] - \{j\}} A_{j,t} u_t = \lambda u_j - A_{j,j} u_j \). Therefore,

\[
|\lambda - A_{j,j}| |u_j| \leq \sum_{t \in [m] - \{j\}} |A_{j,t}| |u_t| \\
\leq \|\vec{u}\|_\infty \sum_{t \in [m] - \{j\}} |A_{j,t}| \quad (1)
\]

Dividing through by \( |u_j| = \|\vec{u}\|_\infty \) yields the desired result. \( \square \)

We can now show that a matrix with small coherence will also have reasonable Restricted Isometry Constants.
Theorem 2. Let $A \in \mathbb{C}^{m \times N}$ be a matrix with $l_2$-normalized columns. Let $k \in [N]$. For all $k$-sparse vectors $\vec{x} \in \mathbb{C}^N$,

$$(1 - (k - 1)\mu(A)) \|\vec{x}\|_2^2 \leq \|A\vec{x}\|_2^2 \leq (1 + (k - 1)\mu(A)) \|\vec{x}\|_2^2 \quad (\dagger)$$

Note: If $(\dagger)$ is true then it implies that $A_S^*A_S$ has all eigenvalues in

$$[1 - \mu(A)(k - 1), 1 + \mu(A)(k - 1)] \quad \forall S \subset [N] \text{ with } |S| \leq k$$

- If $(\dagger)$ is true then it also implies that the R.I.C. $\epsilon_k(A) \leq \mu(A)(k - 1)$

Proof of Theorem 2: Let $S \subseteq [N]$ have $|S| = k$. Then, $A_S^*A_S \in \mathbb{C}^{k \times k}$ is positive semi-definite and has $k$ orthonormal eigenvectors.

- Let $\lambda_{\max} = \text{largest eigenvalue} \geq \lambda_{\min} = \text{smallest eigenvalue} \geq 0$.

- If $\vec{x}$ has support $\subseteq S$, then $\|A\vec{x}\| = \|A_S\vec{x}_S\| = \langle A_S^*A_S\vec{x}, \vec{x} \rangle \leq \lambda_{\max}\|\vec{x}\|_2^2$. Similarly, $\|A\vec{x}\| = \|A_S\vec{x}_S\| = \langle A_S^*A_S\vec{x}, \vec{x} \rangle \geq \lambda_{\min}\|\vec{x}\|_2^2$. Thus, it suffices to bound $\lambda_{\min}, \lambda_{\max}$.

- Let $\lambda_{\min} \leq \lambda \leq \lambda_{\max}$ be an eigenvalue of $A_S^*A_S$.

- Note that $(A_S^*A_S)_{i,j} = 1 \quad \forall j \in [k]$. Theorem 1 now tells us that $|1 - \lambda| \leq \sum_{l \in [k]\setminus\{j\}} |\langle A_S^*A_S\rangle_{j,l}|$ for some $j$. So, $|1 - \lambda| \leq \sum_{l \in S\setminus\{j\}} |\langle \vec{a}_l, \vec{a}_{j'} \rangle| \leq (k - 1)\mu(A)$.

$(\dagger)$ follows by setting $\lambda = \lambda_{\max}$ or $\lambda_{\min}$ and doing some algebra.

- Although coherence implies R.I.P., we do not get the right scaling for sparsity in the necessary number of measurements.

Theorem 3 (Welch Bound). The coherence of a matrix $A \in \mathbb{C}^{m \times N}$ with $l_2$-normalized columns satisfies $\mu(A) \geq \sqrt{\frac{N - m}{m(2N - 1)}}$.

Proof: See Theorem 5.7 from [1].

- Theorem 3 implies that $(\mu(A))^2 \geq \frac{N - m}{m(2N - 1)}$

$$\implies \left(\frac{1}{\mu(A)}\right)^2 \leq \frac{m(N - 1)}{N - m}, \quad (m \geq 1)$$

$$\implies m \geq \left(\frac{1}{\mu(A)}\right)^2, \quad (N \gg m)$$

- In order to get $\epsilon_k(A) \leq \epsilon$ by Theorem 2 we need $\mu(A)(k - 1) \leq \epsilon$

$$\implies m \geq \frac{(k - 1)^2}{\epsilon^2} \quad - \text{We end up with quadratic dependence on sparsity!}$$

- Sub-gaussian random matrices and bounded orthonormal (BON) results give us R.I.P. of order $k$ with $m \sim C \cdot k \cdot \log^c(N) -$, which scales much better (linearly) in $k$. 

2
• Note that asymptotically, the lower bound for $\mu(A)$ approaches $\frac{1}{\sqrt{m}}$ as $N \to \infty$.

– There are constructions that match this asymptotic lower bound on coherence for small $m$ and $N$. An example follows:

**Proposition 1** (Proposition 5.13 from [1]). *For each prime number $m \geq 5$ there is an explicit $m \times m^2$ complex matrix $A$ with $\mu(A) = \frac{1}{\sqrt{m}}$.*

References