MTH 995-003: Intro to CS and Big Data Spring 2014

Lecture 25 – April 01, 2014

Inst. Mark Iwen *Scribe: Atreyee Majumder*

Contents

1 Coherence

Definition 1. *Let* $A \in \mathbb{C}^{m \times N}$ *be a matrix with* ℓ_2-*normalized columns (i.e. such that* $A = (\vec{a}_1, \ldots, \vec{a}_N)$ *has* $\|\vec{a}_j\| = 1 \;\; \forall j \in [\mathrm{N}]$*). 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* λ *be an eigenvalue of a square matrix* $A \in \mathbb{C}^{m \times m}$*. Then,* $\exists$ *an index* $j \in [m]$ *such that* $|\lambda - A_{j,j}| \leq \sum_{l \in [m]-\{j\}} |A_{j,l}|$*.*

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

$$\begin{aligned} |\lambda - A_{j,j}| \; |u_j| &\leq \sum_{l \in [m]-\{j\}} |A_{j,l}| \; |u_l| \\ &\leq \|\vec{u}\|_\infty \sum_{l \in [m]-\{j\}} |A_{j,l}| \end{aligned} \tag{1}$$

Dividing through by $|u_j| = \|\vec{u}\|_\infty$ yields the desired result. □

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 $\vec{x} \in \mathbb{C}^N$,*

$$\big(1-(k-1)\mu(A)\big)\ \|\vec{x}\|_2^2 \ \le\ \|A\vec{x}\|_2^2 \ \le\ \big(1+(k-1)\mu(A)\big)\ \|\vec{x}\|_2^2 \qquad (\dagger)$$

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

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

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

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

- Let λ_{max} = largest eigenvalue $\ge \lambda_{min}$ = smallest eigenvalue ≥ 0.
- If $\vec{x}$ has support $\subseteq$ S, then $\|A\vec{x}\| = \|A_S\vec{x}_S\| = \langle A_S^*A_S\vec{x}, \vec{x}\rangle \le \lambda_{max}\|\vec{x}\|_2^2$. Similarly, $\|A\vec{x}\| = \|A_S\vec{x}_S\| = \langle A_S^*A_S\vec{x}, \vec{x}\rangle \ge \lambda_{min}\|\vec{x}\|_2^2$. Thus, it suffices to bound $\lambda_{min}, \lambda_{max}$.
- Let $\lambda_{min} \le \lambda \le \lambda_{max}$ be an eigenvalue of $A_S^*A_S$.
- Note that $(A_S^*A_S)_{j,j} = 1\ \ \forall j \in [k]$. Theorem 1 now tells us that $|1-\lambda| \le \sum_{l\in[k]-\{j\}} |(A_S^*A_S)_{j,l}|$ for some j. So, $|1-\lambda| \le \sum_{l\in S-\{j'\}} |\langle \vec{a}_l, \vec{a}_{j'}\rangle| \le (k-1)\mu(A)$.

(†) follows by setting $\lambda = \lambda_{max}$ or λ_{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 $\mu(A) \ge \sqrt{\frac{N-m}{m(N-1)}}$.*

Proof : See Theorem 5.7 from [1]. □

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

  $\implies \left(\frac{1}{\mu(A)}\right)^2 \le \frac{m(N-1)}{N-m}$, $(m \ge 1)$

  $\implies m \gtrsim \left(\frac{1}{\mu(A)}\right)^2$, $(N \gg m)$
- In order to get $\epsilon_k(A) \le \epsilon$ by Theorem 2 we need $\mu(A)(k-1) \le \epsilon$

  $\implies m \gtrsim \frac{(k-1)^2}{\epsilon^2}$ – 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.

- 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

[1] Simon Foucart, Holger Rauhut. A Mathematical Introduction to Compressive Sensing. *Birkhauser Basel*, 2013.