1 RIP matrices

**Definition:** A matrix $A \in \mathbb{R}^{m \times n}$ is $(k, \varepsilon)$-RIP (satisfies the restricted isometry property) if for every $k$-sparse vector $x \in \mathbb{R}^n$,

$$(1 - \varepsilon)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \varepsilon)\|x\|_2^2.$$  

Another interpretation: Let $A_S$ denote the restriction of $A$ to columns in $S \subset [n]$. Then the above requires that for all $S$ of cardinality $k$, and all $x \in \mathbb{R}^S$, we have

$$(1 - \varepsilon)\|x\|_2^2 \leq \|x A_S^T A_S x\|_2 \leq (1 + \varepsilon)\|x\|_2^2,$$

which means that $A_S^T A_S \approx I_k$ in the sense that all its eigenvalues are close to 1. We can further write it as $|x^T (A_S^T A_S - I)x| \leq \varepsilon \|x\|_2^2$, which in matrix notation is just a bound on the operator norm (spectral radius):

$$\|A_S^T A_S - I\| \leq \varepsilon.$$

**Exer:** Show that that this implies $A_S$ is invertible.

**Exer:** Show that every $(\varepsilon/k)$-coherent matrix is $(k, \varepsilon)$-RIP.

Recall that a matrix $A \in \mathbb{R}^{m \times n}$ is called $\alpha$-coherent if its columns $A^i$ satisfy that every $\|A^i\|_2 = 1$ and every $|\langle A^i, A^j \rangle| \leq \varepsilon$ (for $i \neq j$).

By the homework exercise, this implies that for every $(n, k, \varepsilon)$, there exists a $(k, \varepsilon)$-RIP matrix with $m = O(\varepsilon^{-2} k^2 \log n)$ rows.

**Hint:** Given $A$ that is $\alpha$-coherent matrix for $\alpha = \varepsilon/k$, let $B = A_S^T A_S - I$, and bound $\|B\|$ which is the largest-magnitude eigenvalue of $B$. 

*These notes summarize the material covered in class, usually skipping proofs, details, examples and so forth, and possibly adding some remarks, or pointers. The exercises are for self-practice and need not be handed in. In the interest of brevity, most references and credits were omitted.*
2 Compressed Sensing via Basis Pursuit

**Theorem 1 [Candes, Romberg and Tao [2004], and Donoho [2004]:** There is a polynomial-time algorithm that given a matrix $A \in \mathbb{R}^{m \times n}$ which is $(2k, \epsilon)$-RIP for $1 + \epsilon < \sqrt{2}$, together with $y = Ax$ for some (unknown) $x \in \mathbb{R}^n$, computes $\tilde{x} \in \mathbb{R}^n$ satisfying

$$\|x - \tilde{x}\|_2 \leq O(1/\sqrt{k})\|x_{\text{tail}(k)}\|_1.$$ 

This condition is usually called an $\ell_2/\ell_1$ guarantee.

**Exer:** Show that the above implies the following $\ell_1/\ell_1$ guarantee for $x^* = \tilde{x}_{\text{top}(k)}$:

$$\|x - x^*\|_1 \leq O(1)\|x_{\text{tail}(k)}\|_1.$$ 

Hint: Let $T$ be the top $k$ coordinates of $x$, and $\hat{T}$ the top $k$ coordinates of $\tilde{x}$. Split the coordinates into $\hat{T}$, $T \setminus \hat{T}$, and the rest.

**Comparison with previously seen result:** We saw previously an algorithm of [Cormode and Muthukrishnan, 2006] achieving WHP $\ell_1/\ell_1$ guarantee

$$\|x - x'\|_1 \leq (1 + 3\epsilon)\|x_{\text{tail}(k)}\|_1.$$ 

* The current $\ell_2/\ell_1$ guarantee is stronger as it implies an $\ell_1/\ell_1$ guarantee, although with constant factor and not $1 + 3\epsilon$.

* The current result is deterministic and holds for all $x$ simultaneously, while the previous one holds WHP separately for every $x$.

* Previously, the number of measurements was $m = O(\epsilon^{-1}k \log n)$. Here it depends on having an RIP matrix; the incoherent matrix from homework has worse (quadratic) dependence on $k$, but other constructions of RIP matrices are linear in $k$.

**Basis Pursuit Algorithm:** We will prove Theorem 1 using an algorithm called Basis Pursuit, which simply solves the linear program (LP)

$$\tilde{x} = \min\{\|z\|_1 : z \in \mathbb{R}^n, Az = y\}.$$ 

It is known that linear programs can be solved in polynomial time.

**Exer:** Show that $\tilde{x}$ above can indeed be solved using LP.

**Proof of Theorem 1 (based on [Candes’08]):**

As before, let $z_S$ denote a vector $z$ after zeroing all coordinates outside $S \subset [n]$.

Let $T_0 \subset [n]$ be the indices of the $k$ heaviest coordinates (largest in absolute value) in $x$. Thus $x_{T_0} = x_{\text{tail}(k)}$.

We now partition the rest ($T_0^c$) according to the heaviness in $h = x - \tilde{x}$ (not in $x$). Let $T_1 \subset T_0^c$ be the $k$ heaviest coordinates in $h_{T_0^c}$ largest ones (i.e., largest in $T_0^c$), and so forth.
To bound the error of $h = x - \tilde{x}$, we use the triangle inequality
\[
\|x - \tilde{x}\|_2 = \|h\|_2 = \|h_{T_0 \cup T_1} + h_{(T_0 \cup T_1)^c}\|_2 \\
\leq \|h_{T_0 \cup T_1}\|_2 + \|h_{(T_0 \cup T_1)^c}\|_2.
\]

The proof will be completed by the following two lemmas.

QED

**Lemma 1a:** \[\|h_{(T_0 \cup T_1)^c}\|_2 \leq O(1/\sqrt{k})\|x_{\text{tail}(k)}\|_1 + \|h_{T_0 \cup T_1}\|_2.\]

**Lemma 1b:** \[\|h_{T_0 \cup T_1}\|_2 \leq O(1/\sqrt{k})\|x_{\text{tail}(k)}\|_1.\]

We prove these two lemmas using another lemma.

**Lemma 1c:** \[\sum_{j \geq 2} \|h_{T_j}\|_2 \leq \frac{2}{\sqrt{k}} \cdot \|x_{T_0^c}\|_1 + \|h_{T_0 \cup T_1}\|_2.\]

**Proof of Lemma 1c:** Was seen in class using the so-called “shelling argument” and the fact that now $\tilde{x} = x - h$ is a minimizer of the LP, and $x$ is feasible.

**Proof of Lemma 1a:** was seen in class, follows almost immediate from Lemma 1a.

To prove Lemma 1b we need another lemma.

**Lemma 1d:** Suppose $h', h''$ are supported on disjoint sets $T', T'' \subset [n]$ respectively, and $A$ is $(|T'| + |T''|, \varepsilon_0)$-RIP. Then
\[|\langle Ah', Ah'' \rangle| \leq \varepsilon_0 \|h'\|_2 \|h''\|_2.\]

**Exer:** Prove this lemma.

Hint: First assume WLOG that $h'$, $h''$ are unit vectors. Then apply the formula \[\|u + v\|_2^2 - \|u - v\|_2^2 = 4\langle u, v \rangle\] to $u = Ah'$ and $v = Ah''$.

**Proof of Lemma 1b:** Was seen in class. The idea is to analyze the norm of $Ah_{T_0 \cup T_1}$ (instead of that of $h_{T_0 \cup T_1}$) to show
\[2(1 - \varepsilon)\|h_{T_0 \cup T_1}\|_2^2 \leq \|Ah_{T_0 \cup T_1}\|_2^2 \leq \varepsilon \sqrt{2} \|h_{T_0 \cup T_1}\|_2 \sum_{j \geq 2} \|h_{T_j}\|_2,\]

then plug in Lemma 1c, and rearrange.