6. MSRI Graduate Summer School Notes
Jeff Blanchard, 2018
Epilogue: Asymmetric Restricted Isometry Constants and ℓ_1-minimization

In Chapter 1.2, we extensively studied singular values of matrices. Let's study the RIP in terms of singular values. We do this by introducing a more relaxed version of the RIP which does not use the same restricted isometry constant for both the lower and upper bounds. These are called the Asymmetric Restricted Isometry Constants (ARIC) and when we impose conditions on ARICs, it is an asymmetric restricted isometry property (ARIP).

In most of the literature on compressed sensing, the standard RIC are used. The nice symmetry of Def. 5.1 does provide for a more tidy analysis, but it also unintentionally establishes RIP conditions that are much more difficult to satisfy than Asymmetric RIP conditions obtained by the same general analysis. The reason is that the singular values of matrices do not deviate from 1 in a symmetric fashion.

Definition 6.1. Let $A \in \mathbb{R}^{m\times n}$. A has the asymmetric restricted isometry constants L_s and U_s if

$$L_s = \min_{c\geq 0} c \ \text{ subject to } \ (1-c)\|x\|_2^2 \leq \|Ax\|_2^2 \text{ for all } x \in \chi_n(s);$$
$$U_s = \min_{c\geq 0} c \ \text{ subject to } \ (1+c)\|x\|_2^2 \geq \|Ax\|_2^2 \text{ for all } x \in \chi_n(s).$$

L_s and U_s are the *lower* and *upper* RICs of A, respectively. (This is a generalization of the standard RICs from Def. 5.1.)

Exercise 6.1. First make a simple observation. Prove that $\delta_s = \max\{L_s, U_s\}$.

Now, we see that the ARIP are closely related to the singular values of all the submatrices of A of size s. In particular, we prove the following lemma.

Lemma 6.1. *Let $\Omega = \{1,\dots,n\}$ and let $\lambda^{min}(B), \lambda^{max}(B)$ denote the minimum and maximum eigenvalues of a matrix B. Then,*

$$1 - L_s = \min_{T\subset\Omega, |T|=s} \lambda^{min}(A_T^*A_T),$$
$$1 + U_s = \max_{T\subset\Omega, |T|=s} \lambda^{max}(A_T^*A_T).$$

Exercise 6.2. Prove the second part of this lemma for $1+U_s$. Recall all we learned in Chapter 1.2 about singular values of a matrix B and their relationship to eigenvalues of the matrix B^*B. Recall that the eigenvalues of $A_T^*A_T$ are the squares of the singular values of A_T and therefore nonnegative.

The definition of $1-L_s$ follows a very similar argument since

$$\lambda^{min}(A_T^*A_T) = (\sigma_{min}(A_T))^2 \leq \frac{\|A_Tx\|_2^2}{\|x\|_2^2}$$

for all $x \in \mathbb{R}^n$. The fact that singular values do not deviate from 1 symmetrically (the smallest singular values of a full rank matrix must be within 1 of 1 while the largest singular value could be much larger than 1) leaves room for misinterpretation of various statements in the compressed sensing literature. For example, the sufficient condition $\delta_{2k} < 1$ for unique k-sparse representations you proved in Exer. 5.2 is often stated as a necessary condition. The necessary condition is instead $L_{2k} < 1$.

Exercise 6.3. Let $A \in \mathbb{R}^{m\times n}$, let $x \in \chi_n(k)$, and suppose $y = Ax$. Prove that x is the unique, sparsest solution to $y = Ax$ if and only if A satisfies the ARIP $L_{2k} < 1$. What does the statement $L_{2k} < 1$ say about spark(A)?

Lemma 6.1 also gives us a nice way to bound an operator norm that will come up regularly in the analysis of greedy algorithms. We state it as the following lemma and prove it in the subsequent exercises.

Lemma 6.2. *Suppose A has the asymmetric restricted isometry constants L_s and U_s. Then, if $T \subset \Omega, |T| = s$,*

$$\|I - A_T^*A_T\|_2 \leq \max\{L_s, U_s\} = \delta_s.$$

Exercise 6.4. Use Lem. 6.1 to prove that if $\lambda(A_T^*A_T)$ is any eigenvalue of $A_T^*A_T$, then

$$|1 - \lambda(A_T^*A_T)| = |\lambda(A_T^*A_T) - 1| \leq \max\{L_s, U_s\}.$$

Exercise 6.5. Prove that if $\lambda(A_T^*A_T)$ is an eigenvalue of $A_T^*A_T$, then $1-\lambda(A_T^*A_T)$ is an eigenvalue of $I-A_T^*A_T$.

Exercise 6.6. Show that $I-A_T^*A_T$ is symmetric and therefore every singular value $\sigma(I-A_T^*A_T)$ is the equal to the absolute value of an eigenvalue of $I-A_T^*A_T$, i.e.

$$\sigma(I-A_T^*A_T)=|1-\lambda(A_T^*A_T)|\,.$$

Exercise 6.7. Finally, using the definition of the ℓ_2 norm of a matrix in terms of its singular values, employ Exers. 6.4 and 6.6 to establish Lem. 6.2.

So, Exer. 6.6 proves the lemma. Now, ponder this: how hard of a problem is it to determine the restricted isometry constants of a matrix for a given value of s? When you figure this out, you'll realize this is a bit unfortunate. However, using probability and tails on bounds of singular values, it can be proven that certain families of random matrices satisfy the RIP or ARIP with very high probability. The power of the RIP is that when we assume a matrix satisfies an RIP condition, it is now a linear algebraic property of the matrix which we can use to bound the approximations in an algorithm. Also, the probabilistic results about matrices satisfying the RIP only require $m \gtrsim Ck$. Therefore, the number of measurements required is proportional to the information in the signal. This is a vast improvement over the coherence based conditions of the previous chapter where the number of measurements was proportional to the square of the information, $m \gtrsim \tilde{C}k^2$. This linear scaling with k is the optimal order.

6.1. **RIP and ℓ_1 minimization.** In this section, we present our first RIC based proof that a tractable decoder, namely the ℓ_1 decoder, will correctly recover sparse solutions under the right conditions. Our method will be to show that under a sufficient condition on the ARIP constants of an encoder A, A satisfies an appropriate null space property. Candes and Tao were the first to utilize RICs to establish the equivalence of the ℓ_1 solution and the ℓ_0 solution. Many alternative proofs with improved sufficient conditions on the RICs were developed over the subsequent years. Here we present a relatively simple proof to show how the RICs are used in the analysis of a decoder. We will call upon the results of Chap. 5 and the following lemma.

Lemma 6.3. *Suppose $S,T\subset\{1,\dots,n\}$ with $|S|=|T|=k$. Let $v\in\mathbb{R}^n$.*

(i) $\|v_T\|_1\le\sqrt{k}\|v_T\|_2$*;*

(ii) If $\max_{j\in S}|v_j|<\min_{i\in T}|v_i|$*, then*

$$\|v_S\|_2\le\frac{1}{\sqrt{k}}\|v_T\|_1.$$

Proof. Observe that in either interpretation of the restriction, v_T and v_S each have at most k nonzero entries. We leave the proof of *(i)* as an exercise similar to those from Chap. 0 which is a clever application of the Cauchy-Schwartz inequality. To establish *(ii)*, we recall from Exercise 0.6 that

$$\|v_S\|_2\le\sqrt{k}\|v_S\|_\infty=\sqrt{k}\left(\max_{j\in S}|v_j|\right). \tag{7}$$

We now bound the minimum magnitude in v_T by the mean of the magnitudes and recognize the ℓ_1 norm:

$$\min_{i\in T}|v_i|\le\frac{1}{k}\sum_{i\in T}|v_i|=\frac{1}{k}\|v_T\|_1. \tag{8}$$

From the hypothesis of *(ii)*, $\|v_S\|_\infty\le\min_{i\in T}|v_i|$. Combining this fact with (7) and (8), we have

$$\|v_S\|_2\le\frac{\sqrt{k}}{k}\|v_T\|_1=\frac{1}{\sqrt{k}}\|v_T\|_1.$$

□

We are now ready to state and prove our main result of the chapter: when x is k-sparse and A has well behaved restricted isometry constants, the ℓ_1 decoder will find the unique sparsest solution to the linear measurements $y=Ax$.

Theorem 6.1.1. *Suppose $A \in \mathbb{R}^{m\times n}$ has asymmetric restricted isometry constants L_k, L_{2k}, U_{2k}. Let $x \in \chi_n(k)$ and let $y = Ax$. If*

$$2\max\{L_{2k}, U_{2k}\} + L_k < 1,$$

then x is the unique solution to the ℓ_1 decoder, namely $\hat{x} = x$. Therefore, every k-sparse vector is exactly recovered by the ℓ_1 decoder.

Proof. We prove the result by establishing that for any $T \subset \{1, \ldots, n\}$ with $|T| = k$ and any $v \in \mathrm{null}(A)-\{0\}$, we have

$$\|v_T\|_1 < \frac{1}{2}\|v\|_1.$$

By employing Exercise 4.16, this implies that A satisfies $NSP(k)$.

Let $v \in \mathrm{null}(A) - \{0\}$ and let $T_0 = \mathrm{PrincipalSupport}_k(v)$. Then iteratively define the index sets

$$S_i = \bigcup_{j=0}^{i} T_j \qquad \text{and} \qquad T_{i+1} = \mathrm{PrincipalSupport}_k(v_{S_i^c}).$$

By construction $T_i \cap T_j = \emptyset$ for all $i \neq j$, $|T_i| = k$ for $j < \frac{n}{k}$, and $|T_{\lceil \frac{n}{k} \rceil}| \leq k$. Also by construction,

$$\max_{j\in T_{i+1}} |v_j| \leq \min_{l\in T_i} |v_l|.$$

Therefore, from Lem. 6.3(ii),

$$\|v_{T_{i+1}}\|_2 \leq \frac{1}{\sqrt{k}}\|v_{T_i}\|_1. \tag{9}$$

Since $v \in \mathrm{null}(A) - \{0\}$, $Av_{T_0} = -Av_{T_0^c} = A(-v_{T_0^c})$. Therefore,

$$\begin{aligned}
\|Av_{T_0}\|_2^2 = \langle Av_{T_0}, Av_{T_0}\rangle &= \left|\left\langle Av_{T_0}, A\left(-v_{T_0^c}\right)\right\rangle\right| \\
&= \left|\left\langle Av_{T_0}, \sum_{i\geq 1} A\left(-v_{T_i}\right)\right\rangle\right| \leq \sum_{i\geq 1} \left|\left\langle Av_{T_0}, A\left(-v_{T_i}\right)\right\rangle\right|, \qquad (10)
\end{aligned}$$

where the second equality follows from our observation, the third equality follows from the construction of the sets $\{T_j\}_{j\geq 0}$ where $T_0^c = \cup_{j\geq 1} T_j$, and the final inequality is an application of the triangle inequality.

We now make use of the RICs. First, with $\delta_{2k} = \max\{L_{2k}, U_{2k}\}$, define

$$\rho = \frac{\delta_{2k}}{1 - L_k}.$$

For our first application of the RICs, note that the definition of the RICs implies

$$\|v_{T_0}\|_2^2 \leq \frac{1}{1-L_k}\|Av_{T_0}\|_2^2. \tag{11}$$

Our second application of the RICs is to employ Lem. 5.1(ii) so that

$$\left|\left\langle Av_{T_0}, A\left(-v_{T_i}\right)\right\rangle\right| \leq \delta_{2k}\|v_{T_0}\|_2\|v_{T_i}\|_2. \tag{12}$$

Then inserting (11) into (10), we have

$$\|v_{T_0}\|_2^2 \leq \frac{1}{1-L_k}\sum_{i\geq 1}\left|\left\langle Av_{T_0}, A\left(-v_{T_i}\right)\right\rangle\right|.$$

Using the bound from (12), we have

$$\begin{aligned}
\|v_{T_0}\|_2^2 &\leq \frac{1}{1-L_k}\sum_{i\geq 1}\delta_{2k}\|v_{T_0}\|_2\|v_{T_i}\|_2 \\
&= \rho\|v_{T_0}\|_2\sum_{i\geq 1}\|v_{T_i}\|_2. \qquad (13)
\end{aligned}$$

Diving by $\|v_{T_0}\|_2$ and shifting the index in the sum,

$$\|v_{T_0}\|_2 \leq \rho\sum_{j\geq 0}\|v_{T_{j+1}}\|_2.$$

From (9), we can bound the individual summands so that

$$\|v_{T_0}\|_2 \leq \frac{\rho}{\sqrt{k}} \sum_{j \geq 0} \|v_{T_j}\|_1.$$

We now invoke Lem. 6.3(i) to see that

$$\|v_{T_0}\|_1 \leq \sqrt{k}\|v_{T_0}\|_2 \leq \rho \sum_{j \geq 0} \|v_{T_j}\|_1 = \rho\|v\|_1,$$

where the final equality is due to the fact that the sets $\{T_j\}_{j \geq 0}$ are mutually disjoint. Finally, the hypothesis on the RIC constants of A ensures that $\rho < \frac{1}{2}$ and therefore

$$\|v_{T_0}\|_1 < \frac{1}{2}\|v\|_1.$$

Therefore, A satisfies $NSP(k)$ and Thm. 4.3.1 guarantees that every $x \in \chi_n(k)$ is the unique solution to (5). □

A more common formulation of a sufficient condition in terms of the standard restricted isometry constant is given in the following corollary.

Corollary 6.1. *Suppose $A \in \mathbb{R}^{m \times n}$ has the restricted isometry constant δ_{2k}. Let $x \in \chi_n(k)$ and let $y = Ax$. If*

$$\delta_{2k} < \frac{1}{3},$$

then x is the unique solution to the ℓ_1 decoder, namely $\hat{x} = x$. Therefore, every k-sparse vector is exactly recovered by the ℓ_1 decoder.

Proof. This is an exercise. Observe that $L_k \leq L_{2k}$ and show that the hypotheses of the corollary imply those of Thm. 6.1.1. □