Key points: The Restricted Isometry Property, Compression and Decompression, Linear Programming.

Definition 1 (Restricted Isometry Property (RIP)). \[ A \in \mathbb{R}^{k \times n} \text{ has } (s, \delta) - \text{RIP if } \forall x \in \mathbb{R}^n, \|x\|_2 = 1 \text{ and } x \text{ is } s\text{-sparse : } 1 - \delta \leq \|Ax\|_2^2 \leq 1 + \delta \]

This definition implies that if we let \( T \leftarrow [n] \) be defined by \( i \in T \iff x_i \neq 0 \) and we consider \( A_T \mathbb{R}^{k \times n} \) the submatrix of \( A \) obtained by keeping the columns indexed by \( T \) and assigning 0 everywhere else, then \( A_T \mathbb{R}^{k \times s} \) is almost an isometry:
\[ \|A_T x\|_2 \approx \|x\|_2 \]
Recall that we previously showed that if \( k = \Theta(s \log n) \exists A \text{ with RIP with parameter } (s, \delta = 0.2) \).

Compression and Decompression

Using a matrix \( A \) having the RIP property allows us to compress a signal, which we assume is represented by a sparse vector \( x \), during its acquisition (hardware property). We would then obtain \( y = A \cdot x \).

Problem: How do we decompress \( y \) ?

- Where are the \( s \) zeros of \( x \)? There is \( \binom{n}{s} \) (\( \approx n^s \) for \( s \) small) possibilities.
- If we know where these values are, what are they?

Observation 2. Images are not sparse but with some definition of derivative we have that the derivative of an image is sparse. Indeed the values of neighboring pixels are likely to be close.

A naive solution:

Definition 3 (zero-norm). \( \|x\|_0 \) is the number of coordinates of \( x \) which are different from 0.

We choose \( x \) such that \( x^* = \text{Argmin} \|x'\|_0 \text{ with } Ax' = y \). Unfortunately solving this minimization problem is NP-hard.

A better solution:

\( x^* = \text{Argmin} \|x'\|_1 \text{ with } Ax' = y \). Using the norm 1 allows us to regularize the minimization problem. Indeed it is equivalent to a linear programming problem:

\[
\begin{align*}
\min \sum_{i=1}^{n} z_i \\
\forall i, -z_i \leq x'_i \leq z_i \\
Ax' \leq y \\
\forall i, z_i \geq 0
\end{align*}
\]
Another solution in the presence of noise:

Assuming $y = A \cdot x + \text{noise} : x^* = \text{Argmin} \|x\|_1$ with $\|Ax' - y\|_2 \leq \varepsilon$ , $\varepsilon$ being the magnitude of the noise.

In this case we have the following theorem:

**Theorem 4.** If $A$ has $(s, \delta \leq 0.2) - \text{RIP}$ then $\|x^* - x\|_2 \leq C \cdot \varepsilon$

**Proof.** First the triangle inequality gives us:

$$\|Ax^* - Ax\|_2 \leq \|Ax^* - y\|_2 + \|Ax - y\|_2$$

$$\|Ax^* - Ax\| \leq 2\varepsilon$$  \hspace{1cm} (1)

We define $h = x^* - x$ then $\|x^*\|_1 = \|x + h\|_1 \leq \|x\|_1$

In a general scheme we define $x_T$ the vector whose first coordinates are $x_i, i \in T$ and the other are set to zero. It is a $s$ sparse vector.

Then:

$$\|x_T + h_T\|_1 + \|x_T^c + h_T^c\| = \|x + h\|_1$$ using the property of $L_1$ norm.

Recall that by definition of $T$ $x_T^c = 0$ and $x_T = x$.

Again by the triangle inequality:

$$\|x\|_1 \geq \|x + h_T\|_1 + \|h_T^c\| \geq \|x\|_1 - \|h_T\| + \|h_T^c\|$$

Which implies:

$$\|h_T^c\|_1 \leq \|h_T\|_1$$  \hspace{1cm} (2)

Some useful notations and considerations:

- Without loss of generality we can assume the $s$ non zeros coordinates of $x$ are its $s$ first. Then $T = T_0 = \{1, \ldots s\}$.

- We define $T_k = \{s + (k-1) \cdot M, \ldots, s + k \cdot M\}$ when it is relevant and $T_{\frac{n}{2}} = \{n - M + 1, \ldots n\}$

- Again, without loss of generality we can assume $h_{T_1}$ contains the $s$ largest coordinates of $h_{T^c}$, $h_{T_2}$ the second $s$ largest coordinates and so on.

With $M = 3s$

The third consideration implies that $h_{s+1}, \ldots h_n$ are such that $\forall j \geq 1, \|h_{T_{j+1}}\|_\infty \leq \frac{\|h_{T_j}\|_1}{M}$, it means that the largest element among the coordinates of $h$ indexed by $T_{j+1}$ is at most the average of the elements of $h$ indexed by $T_j$.

Then:

$$\|h_{T_{j+1}}\|_2^2 \leq M \cdot \|h_{T_{j+1}}\|_1^2 \leq \frac{\|h_{T_j}\|_1^2}{M}$$

If we take the square root and sum for $j \geq 2$:
\[
\sum_{j \geq 2} \| h_{T_{j+1}} \|_2 \leq \frac{\sum_{j \geq 1} \| h_{T_{j-1}} \|_1}{\sqrt{M}} = \frac{1}{\sqrt{M}} \| h_T \|
\]
\[
\leq \frac{1}{\sqrt{M}} \| h_{T_0} \|_1
\]

Where the last inequality comes from (2).

Then by using Cauchy-Schwarz inequality we get:
\[
\sum_{j \geq 2} \| h_{T_{j+1}} \|_2 \leq \sqrt{\frac{\kappa}{M}} \| h_{T_0} \|_2 \leq \sqrt{\frac{1}{3}} \| h_{T_0} \|_2
\]

(3)

Let’s wrap all this together:
\[
\| Ah \|_2 = \left\| A_{T_0 \cup T_1} h_{T_0 \cup T_1} + \sum_{j \geq 2} A_{T_j} h_{T_j} \right\|_2
\]
\[
\geq \left\| A_{T_0} h_{T_0} \right\|_2 - \left\| \sum_{j \geq 2} A_{T_j} h_{T_j} \right\|_2
\]
\[
\geq \left\| A_{T_0} h_{T_0} \right\|_2 - \sum_{j \geq 2} \left\| A_{T_j} h_{T_j} \right\|_2
\]

Using the triangle inequality twice.

Then we use the fact that \( A \) has RIP with \( \delta = 0.2 \)
\[
\| Ah \|_2 \geq \sqrt{0.8} \| h_{T_0} \|_2 - \sqrt{1.2} \sum_{j \geq 2} \| h_{T_j} \|
\]
\[
\geq \sqrt{0.8} \| h_{T_0} \|_2 - \sqrt{1.2} \sqrt{\frac{1}{3}} \| h_{T_0} \|_2
\]

Where we used (3) and \( \| h_{T_0} \|_2 \leq \| h_{T_0} \|_2 \).

Then we roughly have:
\[
\| Ah \|_2 \geq \approx 0.26 \| h_{T_0} \|_2
\]

(4)

Finally:
\[
\| x^* - x \|_2 = \| h_{T_0} \|_2^2 + \sum_{j \geq 2} \| h_{T_j} \|_2^2
\]
\[
\leq \| h_{T_0} \|_2^2 + \left( \sum_{j \geq 2} \| h_{T_j} \|_2 \right)^2
\]
\[
\leq \|h_{T_{01}}\|_2^2 + \frac{1}{3}\|h_{T_{02}}\|_2^2 = \frac{4}{3}\|h_{T_{01}}\|_2^2
\]
\[
\leq \frac{3}{4}\cdot 15 \cdot \|Ah\|^2 \leq 45c^2
\]

Where we successively used (3) (4) and (1).

Orthogonal Matching Pursuit[2]

In statistics L_1 Norm minimization is done using orthogonal matching pursuit:

Algorithm 1 OMP

1: for \( \forall t \in 1..l \) do
2: \[ h_t = y - A \cdot x_t \]
3: \[ i_t = \operatorname{argmax}_{j=1..n} |\langle A_j, h_t \rangle| \]
4: \[ S_t \leftarrow S_t \cup \{i_t\} \]
5: \[ x_t = \operatorname{argmin}_{x \in \mathbb{R}^n} \|A_{S_t}x\|_2 \]
6: end for

References


4