A very compressed primer on compressed sensing

Matthis Lehmkuehler

21/02/2024

1. Introduction

Compressed sensing is about selecting a solution $x \in \mathbb{R}^n$ to $y = Ax$ where $y \in \mathbb{R}^m$ and A a $m \times n$ matrix where $n \gg m$ (i.e. the problem is very is generically going to have a large solution space). For $p \in [0, \infty]$ we define the problem $\left(P_p\right)$ by

$$\left(P_p\right) \begin{cases} \text{minimize } \|x\|_p \text{ subject to} \\ Ax = y \end{cases}$$

where $\|x\|_p = \left(|x_1|^p + \dots + |x_n|^p\right)^{1/p}$ for $p \in (0, \infty)$, $\|x\|_\infty = \max\{|x_1|, \dots, |x_n|\}$ and where $\|x\|_0 = \#\{1 \leq i \leq n : x_i \neq 0\}$. If $p \geq 1$ (in particular $p = 1$), this is a convex problem and can be solved using convex optimization techniques, while the cases $p \in [0, 1)$ (in particular $p = 0$) do not enjoy this property. Compressed sensing is a theory on giving conditions under which the solutions to (P_0) and (P_1) agree, thus making the (P_0) problem amenable to convex optimization techniques as well.

Let us introduce some notation: We call $x \in \mathbb{R}^n$ s-sparse if $\|x\|_0 \leq s$. Moreover, we let $[n] = \{1, \dots, n\}$ and if $I = \{i_1 < \dots < i_a\} \subseteq [n]$ we define $x_I = \left(x_i 1(i \in I)\right)_i$ (i.e. we set all entries with index not in I to 0) and $A_I = \left(A_{j i_{a'}}\right)_{j a'}$ (i.e. we only keep the columns with index in I).

The proof in this short note follows the one in [1] fairly closely. For more in depth information, see any good textbook on the topic, like [2].

2. Recovery of a sparse solution

Suppose that x_0 is s-sparse, say supported on I with $\#I = s$, and satisfies $Ax_0 = y$. We can ask, under which condition do we have $\|x\|_1 > \|x_0\|_1$ for all $x \neq x_0$ satisfying $Ax = y$. If this holds, then clearly (P_1) with $y = Ax_0$ has a unique solution given by x_0.

Let $z = x - x_0$ which then satisfies $Az = 0$. If x_0 is supported on $I \subseteq [n]$ with $\#I = s$ then using that $\left(x_0\right)_I = x_0$ and using the triangle inequality, we get

$$\begin{aligned} \|x\|_1 - \|x_0\|_1 &= \|x_0 + z\|_1 - \|x_0\|_1 \\ &= \|x_0 + z_I\|_1 + \|z_{I^c}\|_1 - \|x_0\|_1 \\ &\geq \|z_{I^c}\|_1 - \|z_I\|_1 \ . \end{aligned}$$

So if the right hand side is either positive or $z = 0$, then the solutions to (P_0) and (P_1) agree. We will say that A has the null-space property on I, or $A \in \mathrm{NSP}(m, n, I)$, if each $z \in \ker(A) \setminus \{0\}$ satisfies $\|z_I\|_1 < \|z_{I^c}\|_1$.

Lemma 2.1. Suppose that x_0 is supported on I with $\#I = s$ and $A \in \mathrm{NSP}(m, n, I)$, then x_0 is the unique solution to (P_1) with $y = Ax_0$. Moreover, if $A \in \mathrm{NSP}(m, n, I')$ for all I' with $\#I' = s$ then x_0 is the unique solution to (P_0) with $y = Ax_0$.

Proof. The first part is clear from the discussion above. The second part follows since if x_0' is supported on I' with $\#I' = s$ and $Ax_0' = y$ then by the same argument, x_0' is the unique solution to (P_1) with $y = Ax_0 = Ax_0'$ and the uniqueness implies $x_0 = x_0'$. □

The question is now only, how do we check the condition in the lemma. For this, people have introduced the restricted isometry property. We write $A \in \mathrm{RIP}(m, n, s', \delta)$ for $\delta \in (0, 1)$ if

$$\sqrt{1-\delta}\ \|x\|_2 \leq \|Ax\|_2 \leq \sqrt{1+\delta}\ \|x\|_2$$

for all $x \in \mathbb{R}^n$ with $\|x\|_0 \leq s'$. The following proposition shows how the restricted isometry property implies the null-space property introduced above and that we can hence relate (P_0) and (P_1) via Lemma 2.1.

Proposition 2.2. If $A \in \mathrm{RIP}(m, n, s + s', \delta)$ for $s/s' < (1-\delta)/(1+\delta)$ and $s, s' \geq 1$, then $A \in \mathrm{NSP}(m, n, I)$ for all I with $\#I = s$.

Proof. Fix I with $\#I = s$ and $z \in \ker(A) \setminus \{0\}$. Take an enumeration $I^c \cup (\mathbb{N} \setminus [n]) = \{i_j : j \geq 1\}$ such that the sequence $\left(|z_{i_j}|\right)_j$ is non-increasing where we make the convention $z_i = 0$ if $i > n$. Let

$$I_k = \left\{i_{s'(k-1)+1}, ..., i_{s'k}\right\} \quad \text{for } k \geq 1.$$

Then since $0 = Az = Az_{I \cup I_1} + \sum_{k \geq 1} Az_{I_{k+1}}$ and using the RIP property, we get

$$\begin{aligned}
\|z_I\|_1 &\leq \sqrt{s}\ \|z_I\|_2 \leq \sqrt{s}\ \|z_{I \cup I_1}\|_2 \\
&\leq \frac{\sqrt{s}}{\sqrt{1-\delta}}\ \|Az_{I \cup I_1}\|_2 \\
&\leq \frac{\sqrt{s}}{\sqrt{1-\delta}} \sum_{k \geq 1} \|Az_{I_{k+1}}\|_2 \\
&\leq \frac{\sqrt{s}\sqrt{1+\delta}}{\sqrt{1-\delta}} \sum_{k \geq 1} \|z_{I_{k+1}}\|_2\ .
\end{aligned}$$

Clearly $\|z_{I_{k+1}}\|_2 \leq \sqrt{s'}\ \|z_{I_{k+1}}\|_\infty$ and by the definition of the (I_k), we have $s'\|z_{I_{k+1}}\|_\infty \leq \|z_{I_k}\|_1$. Putting this together yields

$$\|z_I\|_1 \leq \frac{\sqrt{s/s'}\sqrt{1+\delta}}{\sqrt{1-\delta}} \sum_{k \geq 1} \|z_{I_k}\|_1 = \frac{\sqrt{s/s'}\sqrt{1+\delta}}{\sqrt{1-\delta}}\ \|z_{I^c}\|_1\ .$$

The assumption $s/s' < (1-\delta)/(1+\delta)$ implies $\|z_I\|_1 < \|z_{I^c}\|_1$ since $z \neq 0$. □

3. Random Gaussian matrices allow compressed sensing

In this section, we will see how to apply this framework in the case when the matrix A is given by i.i.d. Gaussian entries.

The key here will be to use a tailbound on $\mathbb{P}(\|M\|_{\mathrm{op}} > \delta)$ for a random $s \times s$ symmetric matrix M, so let us recall how to do this using the idea of ε-nets. Fix $\varepsilon \in (0, 1/2)$ and let N_ε be a maximal subset of points in $\{x \in \mathbb{R}^s : \|x\|_2 = 1\}$ which are all distance $> \varepsilon$ away from each other. Then by a crude volume bound, we have that

$$\#N_\varepsilon \leq (1 + \varepsilon/2)^s (\varepsilon/2)^{-s} \leq (3/\varepsilon)^s.$$

Since M is symmetric (by considering its eigendecomposition),

$$\|M\|_{\mathrm{op}} = \max_{x \in \mathbb{R}^s:\ \|x\|_2 = 1} |(x, Mx)| \ .$$

The point is that by maximality, whenever $x \in \mathbb{R}^s$ satisfies $\|x\|_2 = 1$, there exists $x' \in N_\varepsilon$ such that $\|x - x'\| \leq \varepsilon$ and hence

$$|(x, Mx)| \leq |(x', Mx')| + |(x - x', Mx)| + |(x', M(x - x'))| \leq |(x', Mx')| + 2\varepsilon \|M\|_{\mathrm{op}} \ .$$

By taking the maximum over all $x \in \mathbb{R}^s$ with $\|x\|_2 = 1$ and upper bounding the first term on the right side by the maximum over all $x' \in N_\varepsilon$, we get

$$\|M\|_{\mathrm{op}} \leq (1 - 2\varepsilon)^{-1} \max_{x' \in N_\varepsilon} |(x', Mx')|$$

which readily implies that for $\delta > 0$,

$$\begin{aligned} \mathbb{P}\big(\|M\|_{\mathrm{op}} > \delta\big) &\leq \mathbb{P}\left(\max_{x' \in N_\varepsilon} |(x', Mx')| > (1 - 2\varepsilon)\delta\right) \leq \sum_{x' \in N_\varepsilon} \mathbb{P}(|(x', Mx')| > (1 - 2\varepsilon)\delta) \\ &\leq \left(\frac{3}{\varepsilon}\right)^s \max_{x' \in \mathbb{R}^s:\ \|x'\|_2 = 1} \mathbb{P}(|(x', Mx')| > (1 - 2\varepsilon)\delta). \end{aligned} \tag{1}$$

The point is that we can essentially interchange the maximum defining the operator norm with the probability expression and only incur an exponential error term which we will be able to suppress.

The following lemma shows that a matrix A with i.i.d. Gaussian entries satisfies the RIP property. Here and below, $\sigma_1(M), \ldots, \sigma_s(M)$ denote the singular values of a $m \times s$ matrix (in increasing order) and $\lambda_1(M), \ldots, \lambda_s(M)$ denote the eigenvalues of a $s \times s$ matrix (also in increasing order).

Lemma 3.1. There exists $c, C > 0$ such that the following is true: Let A be a $m \times n$ matrix with i.i.d. $N(0, 1/m)$ entries where $s \leq m \leq n$ with $n > 1$, and $\delta \in (0, 1)$. Then $A \in \mathrm{RIP}(m, n, s, \delta)$ with probability $\geq 1 - \exp(Cs \log(n) - cm\delta^2)$.

Proof. Note that

$$\{A \in \mathrm{RIP}(m, n, s, \delta)\} = \left\{\forall I \subseteq [n] \text{ with } \#I = s : \sigma_i(A_I) \in \left[\sqrt{1 - \delta}, \sqrt{1 + \delta}\right] \forall i \leq s\right\}.$$

Then by a union bound we get

$$\begin{aligned} \mathbb{P}(A \notin \mathrm{RIP}(m, n, s, \delta)) &\leq \binom{n}{s} \max_{I \subseteq [n]: \#I = s} \mathbb{P}\Big(\left\{\sigma_i(A_I) \in \left[\sqrt{1 - \delta}, \sqrt{1 + \delta}\right] \forall i \leq s\right\}^c\Big) \\ &= \binom{n}{s} \max_{I \subseteq [n]: \#I = s} \mathbb{P}\Big(\left\{\lambda_i\left(A_I^T A_I\right) \in [1 - \delta, 1 + \delta] \ \forall i \leq s\right\}^c\Big) \\ &= \binom{n}{s} \max_{I \subseteq [n]: \#I = s} \mathbb{P}\big(\|A_I^T A_I - I_s\|_{\mathrm{op}} > \delta\big) \\ &= \binom{n}{s} \mathbb{P}\Big(\|A_{[s]}^T A_{[s]} - I_s\|_{\mathrm{op}} > \delta\Big). \end{aligned}$$

The last equality followed since all expressions in the maximum are the same by the i.i.d. property. Let $B = A_{[s]}$ which is a $m \times s$ matrix with i.i.d. $N(0, 1/m)$ entries. Then by (1) we get for $\varepsilon \in (0, 1/2)$,

$$
\begin{aligned}
\mathbb{P}\big(\|B^T B - I_s\|_{\mathrm{op}} > \delta\big) &\leq \left(\frac{3}{\varepsilon}\right)^s \max_{x\in\mathbb{R}^s:\ \|x\|_2=1} \mathbb{P}(|(x, (B^T B - I)x)| > (1-2\varepsilon)\delta) \\
&\leq \left(\frac{3}{\varepsilon}\right)^s \max_{x\in\mathbb{R}^s:\ \|x\|_2=1} \mathbb{P}(|\,\|Bx\|_2^2 - 1| > (1-2\varepsilon)\delta) \\
&= \left(\frac{3}{\varepsilon}\right)^s \mathbb{P}\left(\left|\frac{N_1^2 + \cdots + N_m^2}{m} - 1\right| > (1-2\varepsilon)\delta\right)
\end{aligned}
$$

where $N_1, \ldots, N_m \sim N(0,1)$ are i.i.d.; this follows since for any $x \in \mathbb{R}^s$ with $\|x\|_2 = 1$, the vector Bx has i.i.d. $N(0, 1/m)$ entries. For the last term, we can use a standard exponential tail bound. Thus putting everything together, we see that there exist universal $c, C > 0$ such that

$$
\begin{aligned}
\mathbb{P}(A \notin \mathrm{RIP}(m,n,s,\delta)) &\leq \binom{n}{s}\left(\frac{3}{\varepsilon}\right)^s Ce^{-cm(1-2\varepsilon)^2\delta^2} \\
&\leq C \cdot \exp\big(s\log(3n/\varepsilon) - cm(1-2\varepsilon)^2\delta^2\big).
\end{aligned}
$$

By taking $\varepsilon = 1/3$, we see that after redefining the constants $c, C > 0$, we have that $\mathbb{P}(A \notin \mathrm{RIP}(m,n,s,\delta)) \leq \exp(Cs\log(n) - cm\delta^2)$ as required. □

In practice, one can simply apply this result with $\delta = 1/4$ to get a control on the probability of the event $A \in \mathrm{RIP}(m,n,s+s',\delta)$ holding where we consider $s' = 2s$, and we are in a setting where we can apply Proposition 2.2.

Remark 3.2. What's remarkable (no pun intended) about this result is that it suffices to take $m = \Omega(s\log(n))$ to get the RIP property and hence the sparse recovery result. The s prefactor is not surprising since in order to recover a vector with s entries, we certainly need at least data of dimension s but it is incredibly useful that we can take n (the dimension out of which we need to select the s sparse entries) exponential in m.

References

[1] E. J. Candès, "The restricted isometry property and its implications for compressed sensing", *C. R. Math.*, vol. 346, no. 9, pp. 589–592, 2008, doi: 10.1016/j.crma.2008.03.014

[2] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Birkhäuser New York, 2013. doi: 10.1007/978-0-8176-4948-7