Lecture 15: Sparse Recovery - L1 minimization

Prof. Eric Price *Scribe: Anish Acharya, Rudrajit Das*

NOTE: THESE NOTES HAVE NOT BEEN EDITED OR CHECKED FOR CORRECTNESS

1 Sparse Recovery - L1 minimization

Recall the sparse-recovery problem. We are given noisy measurements $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$. We wish to recover $\mathbf{x} \in \mathbb{R}^n$ from the available noisy measurements $y \in \mathbb{R}^n$. $\mathbf{A} \in \mathbb{R}^{m \times n}$ is known. We are particularly interested in the vastly under-determined setting where $m << n$. As such, the problem is of course ill-posed but suppose $\mathbf{x}$ is k-sparse i.e. $\|x\|_0 \leq k$. This premise radically changes the problem, making the search for solutions feasible.
However, in general, finding sparse solutions to under-determined systems of linear equations is NP-hard. For example, we can formulate the problem of finding the sparsest solution as:

$$\begin{aligned} \min_x \quad & \|x\|_0 \\ \text{s.t.} \quad & \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon \end{aligned} \tag{P0}$$

To the best of our knowledge, solving P0 requires exhaustive searches over all subsets of columns of $\mathbf{A}$, a procedure which clearly is combinatorial in nature and has exponential complexity. This computational intractability has led to several different formulations. One widely studied formulation, known as basis pursuit (Chen et al. (2001); Candes and Tao (2005); Candes et al. (2006, 2008)), is to do a convex relaxation and solve the following problem instead:

$$\begin{aligned} \min_x \quad & \|x\|_1 \\ \text{s.t.} \quad & \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon \end{aligned} \tag{P1}$$

Unlike the ℓ_0 norm, ℓ_1 is convex and thus P1 reduces to a convex program.

Suppose $\mathbf{x}^*$ is the optimal solution to P1. Then, we have the following result relating $\mathbf{x}^*$ to $\mathbf{x}$.

Theorem 1. *Assume $\mathbf{A}$ satisfies RIP, $\|\mathbf{e}\|_2 \leq \epsilon$, $\mathbf{x}$ is k-sparse and $\mathbf{x}^*$ is the solution to P1 then:*

$$\|\mathbf{x}^* - \mathbf{x}\|_2 \leq \mathcal{O}(\epsilon) \tag{1}$$

Before we prove Theorem 1, we introduce a definition and a lemma which we use in the proof.

Definition 1. ***Restricted Eigenvalue Condition (REC)**: $\mathbf{A}$ satisfies the Restricted Eigenvalue Condition if for any $\mathbf{z}$, if $\exists\, S$ s.t. $\|\mathbf{z}_S\|_1 \geq \|\mathbf{z}_{\overline{S}}\|_1$, then $\|\mathbf{A}\mathbf{z}\|_2 \gtrsim \|\mathbf{z}\|_2$.*

Lemma 1. *(**RIP $\implies$ REC**). Suppose $\mathbf{A}$ satisfies RIP($\mathcal{O}(3k), \epsilon$). Also, let $\mathbf{z}$ be such that there exists a support of size k, say S, which contains more than half of its mass – i.e. $\|(\mathbf{z})_S\|_1 \geq \|(\mathbf{z})_{\overline{S}}\|_1$. Then, $\|\mathbf{z}\|_2 \lesssim \|\mathbf{A}\mathbf{z}\|_2$.*

Proof. We will prove this by a "shelling" argument. Suppose we sort the co-ordinates of $\mathbf{z}$ in decreasing order of magnitude. Then split the co-ordinates into groups of k i.e. $\mathbf{z}_1$ denote the top-k coordinates of $\mathbf{z}$, $\mathbf{z}_2$ denote the next top-k coordinates and so on i.e. $\mathbf{z} = \sum_{i=1}^{n/k} \mathbf{z}_i$. Note that if $\exists$ a support S of size k, such that $\|(\mathbf{z})_S\|_1 \geq \|(\mathbf{z})_{\overline{S}}\|_1$, then $\mathrm{Supp}(\mathbf{z}_1)$ must be one such S – so $\|\mathbf{z}_1\|_1 \geq \|\mathbf{z}_2 + \mathbf{z}_3 + \ldots\|_1$. Also, since each $\mathbf{z}_i$ is k-sparse by construction, then $(\mathbf{z}_i + \mathbf{z}_{i+1})$ is $2k$-sparse. Then:

$$\|\mathbf{z}_i + \mathbf{z}_{i+1}\|_2 \leq \sqrt{2k}\|\mathbf{z}_i + \mathbf{z}_{i+1}\|_\infty \leq \frac{1}{\sqrt{2k}}\|\mathbf{z}_{i-2} + \mathbf{z}_{i-1}\|_1 \tag{2}$$

The last inequality follows because the average value of the previous two shells is bigger than the maximum value in the current two shells.

Now, the rest of the proof is as follows:

$$\begin{aligned}
\|\mathbf{A}\mathbf{z}\|_2 &\geq \|\mathbf{A}(\mathbf{z}_1 + \mathbf{z}_2 + \mathbf{z}_3)\|_2 - \|\mathbf{A}(\mathbf{z}_4 + \mathbf{z}_5)\|_2 - \|\mathbf{A}(\mathbf{z}_6 + \mathbf{z}_7)\|_2 - \ldots \text{ (by the triangle inequality)} \\
&\geq (1-\epsilon)\|\mathbf{z}_1 + \mathbf{z}_2 + \mathbf{z}_3\|_2 - (1+\epsilon)(\|\mathbf{z}_4 + \mathbf{z}_5\|_2 + \|\mathbf{z}_6 + \mathbf{z}_7\|_2 + \ldots) \text{ ($\mathbf{A}$ satisfies RIP of $\mathcal{O}(3k)$)} \\
&\geq (1-\epsilon)\|\mathbf{z}_1 + \mathbf{z}_2 + \mathbf{z}_3\|_2 - \frac{1+\epsilon}{\sqrt{2k}}(\|\mathbf{z}_2 + \mathbf{z}_3\|_1 + \|\mathbf{z}_4 + \mathbf{z}_5\|_1 + \ldots) \text{ (using Equation 2)} \\
&= (1-\epsilon)\|\mathbf{z}_1 + \mathbf{z}_2 + \mathbf{z}_3\|_2 - \frac{1+\epsilon}{\sqrt{2k}}\|\mathbf{z}_2 + \mathbf{z}_3 + \mathbf{z}_4 + \mathbf{z}_5 + \ldots\|_1 \text{($\mathbf{z}_i$ have disjoint supports)} \\
&\geq (1-\epsilon)\|\mathbf{z}_1\|_2 - \frac{1+\epsilon}{\sqrt{2k}}\|\mathbf{z}_2 + \mathbf{z}_3 + \mathbf{z}_4 + \mathbf{z}_5 + \ldots\|_1 \\
&\geq (1-\epsilon)\|\mathbf{z}_1\|_2 - \frac{1+\epsilon}{\sqrt{2k}}\|\mathbf{z}_1\|_1 \text{ (recall that $\|\mathbf{z}_1\|_1 \geq \|\mathbf{z}_2 + \mathbf{z}_3 + \ldots\|_1$)} \\
&\geq \|\mathbf{z}_1\|_2(1 - \epsilon - \frac{1+\epsilon}{\sqrt{2}}) \text{ (since $\|\mathbf{z}_1\|_1 \leq \|\mathbf{z}_1\|_2\sqrt{k}$)} \\
&\gtrsim \|\mathbf{z}_1\|_2
\end{aligned} \tag{3}$$

Hence, $\|\mathbf{z}_1\|_2 \leq \mathcal{O}(1)\|\mathbf{A}\mathbf{z}\|_2$. Also, note that:

$$\begin{aligned}
\|\mathbf{z} - \mathbf{z}_1\|_2 &= \|\mathbf{z}_2 + \mathbf{z}_3 + \ldots\|_2 \\
&\leq \sqrt{\|\mathbf{z}_2 + \mathbf{z}_3 + \ldots\|_\infty \|\mathbf{z}_2 + \mathbf{z}_3 + \ldots\|_1} \text{ (using Hölder's inequality)} \\
&\leq \sqrt{\left(\frac{\|\mathbf{z}_1\|_1}{k}\right)\|\mathbf{z}_1\|_1} \\
&\leq \frac{\|\mathbf{z}_1\|_1}{\sqrt{k}} \\
&\leq \|\mathbf{z}_1\|_2
\end{aligned} \tag{4}$$

Hence, using Equation (4), we get that $\|\mathbf{z}\|_2 \leq \|\mathbf{z}_1\|_2 + \|\mathbf{z} - \mathbf{z}_1\|_2 \leq 2\|\mathbf{z}_1\|_2 \lesssim \|\mathbf{A}\mathbf{z}\|_2$. This concludes the proof. ■

We are now ready to prove Theorem 1.

Proof of Theorem 1: In this proof, we consider the case of $\|\mathbf{e}\|_2 = \epsilon$. However, this proof

can be easily extended to the general case of $\|\mathbf{e}\|_2 \leq \epsilon$.
Let $\mathbf{z} = \mathbf{x}^* - \mathbf{x}$. Since $\mathbf{x}^*$ solves P1, it is also feasible and thus we have:

$$\begin{aligned}
&\|\mathbf{A}\mathbf{x}^* - \mathbf{y}\|_2 \leq \epsilon \\
\Rightarrow &\|\mathbf{A}\mathbf{x}^* - (\mathbf{A}\mathbf{x} + \mathbf{e})\|_2 \leq \epsilon \\
\Rightarrow &\|\mathbf{A}(\mathbf{x}^* - \mathbf{x}) - \mathbf{e}\|_2 \leq \epsilon \\
\Rightarrow &\|\mathbf{A}\mathbf{z} - \mathbf{e}\|_2^2 \leq \epsilon^2 = \|\mathbf{e}\|_2^2 \\
\Rightarrow &\|\mathbf{A}\mathbf{z}\|_2^2 + \|\mathbf{e}\|_2^2 - 2\mathbf{e}^T\mathbf{A}\mathbf{z} \leq \|\mathbf{e}\|_2^2 \\
\Rightarrow &\|\mathbf{A}\mathbf{z}\|_2^2 \leq 2\mathbf{e}^T\mathbf{A}\mathbf{z} \leq 2\|\mathbf{e}\|_2\|\mathbf{A}\mathbf{z}\|_2 \\
\Rightarrow &\|\mathbf{A}\mathbf{z}\|_2 \leq 2\|\mathbf{e}\|_2 \qquad (5)
\end{aligned}$$

Now, we can use the fact that since $\mathbf{x}^*$ is the minimizer of P1 : $\|\mathbf{x}^*\|_1 \leq \|\mathbf{x}\|_1 = \|\mathbf{x}_S\|_1$ where $S = \text{Supp}(\mathbf{x})$. Using this we have:

$$\begin{aligned}
\|\mathbf{x}_S\|_1 &\geq \|\mathbf{x}^*\|_1 \\
&= \|\mathbf{x}^*_S\|_1 + \|\mathbf{x}^*_{\overline{S}}\|_1 \\
&= \|(\mathbf{x} + \mathbf{z})_S\|_1 + \|(\mathbf{x} + \mathbf{z})_{\overline{S}}\|_1 \\
&= \|(\mathbf{x} + \mathbf{z})_S\|_1 + \|\mathbf{z}_{\overline{S}}\|_1 \text{ ... since } S = \text{Supp}(\mathbf{x}). \\
&\geq \|\mathbf{x}_S\|_1 - \|\mathbf{z}_S\|_1 + \|\mathbf{z}_{\overline{S}}\|_1 \text{ ... by the triangle inequality.} \\
\Rightarrow \|\mathbf{z}_S\|_1 &\geq \|\mathbf{z}_{\overline{S}}\|_1 \qquad (6)
\end{aligned}$$

From Equation 6 we see that most of the mass of $\mathbf{z}$ is concentrated in S (i.e. its top k coordinates). Now since Equation 6 holds, using Lemma 1, we have that $\|\mathbf{z}\|_2 \lesssim \|\mathbf{A}\mathbf{z}\|_2$. Plugging this in Equation 5, we get: $\|\mathbf{z}\|_2 \lesssim 2\|\mathbf{e}\|_2 \Rightarrow \|\mathbf{x}^* - \mathbf{x}\|_2 \leq \mathcal{O}(\epsilon)$.

This concludes the proof of Theorem 1.

2 Packing and Covering Numbers

Consider a metric space $(\mathcal{X}, d)$. Note that d must satisfy the following properties:

- $d(x, y) \geq 0 \; \forall \; x, y \in \mathcal{X}$.
- $d(x, y) = d(y, x) \; \forall \; x, y \in \mathcal{X}$.
- $d(x, y) \leq d(x, z) + d(z, y) \; \forall \; x, y, z \in \mathcal{X}$.
- $d(x, y) = 0 \iff x = y \; \forall \; x, y \in \mathcal{X}$.

An example of a metric space is $(\mathbb{R}^m, \|.\|_p)$ which is the m-dimensional real space with the ℓ_p norm.

Definition 2. *An ϵ-cover of $\mathcal{X}$ with respect to d is a set of points $\{x_1, \ldots, x_n\} \in \mathcal{X}$ such that $\forall$ $y \in \mathcal{X}$, $\exists \; i \in [n]$ satisfying $d(y, x_i) \leq \epsilon$. Further, the covering number, $N(\epsilon, \mathcal{X}, d)$, of $\mathcal{X}$ with respect to d is defined as follows:*

$$N(\epsilon, \mathcal{X}, d) := \textit{minimum } n \textit{ such that } \exists \textit{ an } \epsilon\textit{-cover of } \mathcal{X} \textit{ with respect to } d.$$

Also:

$$\log(N(\epsilon, \mathcal{X}, d)) := \text{“metric entropy”, i.e. the information to describe } \mathcal{X} \text{ to within } \epsilon \text{ precision.}$$

As an example, consider $\mathcal{X} = [-1, 1]$ and $d(x, y) = |x - y|$, i.e. the unit line. It is easy to see that $\{0, \pm 2\epsilon, \pm 4\epsilon, \ldots\}$ forms an ϵ-cover of $\mathcal{X}$ w.r.t d – so in this case, $N(\epsilon, \mathcal{X}, d) \leq \frac{1}{\epsilon} + 1$.
Generalizing the above to m dimensions, we get that:

$$N(\epsilon, [-1, 1]^m, \|.\|_\infty) \leq \left(1 + \frac{1}{\epsilon}\right)^m. \tag{7}$$

Definition 3. *An ϵ-packing of $\mathcal{X}$ with respect to d is a set of points $\{x_1, \ldots, x_n\} \in \mathcal{X}$ such that $d(x_i, x_j) \geq \epsilon \ \forall \ i \neq j$. Further, the packing number, $M(\epsilon, \mathcal{X}, d)$, of $\mathcal{X}$ with respect to d is defined as follows:*

$$M(\epsilon, \mathcal{X}, d) := \text{maximum } n \text{ such that } \exists \text{ an } \epsilon\text{-packing of } \mathcal{X} \text{ with respect to } d.$$

Lemma 2. $M(2\epsilon, \mathcal{X}, d) \leq N(\epsilon, \mathcal{X}, d) \leq M(\epsilon, \mathcal{X}, d)$.

The second inequality in Lemma 2, i.e. $N(\epsilon, \mathcal{X}, d) \leq M(\epsilon, \mathcal{X}, d)$ can be explained as follows. Let us make a greedy construction for the cover, such that at each step, we increment our cover to include all points which are more than ϵ away from the current cover that we have. Eventually, when we stop, there would be no point that is more than ϵ away from the final cover (and so this an ϵ-cover). But by this greedy construction, every point is ϵ away from the previous chosen points and hence this gives us a packing too! Hence, there exists a packing with size at least the maximum cover.

Also recall the volume argument that we made in the lecture on 10/06 for ℓ_2 balls. This actually works for any ℓ_p ball, and gives us the following bound:

Lemma 3. *Suppose $\mathcal{X}$ is the unit ℓ_p ball in $\mathbb{R}^m$, say $\mathcal{B}_p$, and $d = \|.\|_p$. Then:*

$$\frac{1}{\epsilon^m} \leq N(\epsilon, \mathcal{B}_p, \|.\|_p) \leq \left(1 + \frac{2}{\epsilon}\right)^m.$$

Compare the result in Lemma 3 for the ℓ_∞ case to the actual bound that we got above (eq. (7)). Notice that the result of Lemma 3 is tight upto constant factors.

References

Candes, E. J. et al. (2008). The restricted isometry property and its implications for compressed sensing. *Comptes rendus mathematique*, 346(9-10):589–592.

Candes, E. J., Romberg, J. K., and Tao, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 59(8):1207–1223.

Candes, E. J. and Tao, T. (2005). Decoding by linear programming. *IEEE transactions on information theory*, 51(12):4203–4215.

Chen, S. S., Donoho, D. L., and Saunders, M. A. (2001). Atomic decomposition by basis pursuit. *SIAM review*, 43(1):129–159.