ECE 901 Spring 2014 Statistical Learning Theory

instructor: R. Nowak

A Note on Compressed Sensing

This is a short note on the compressed sensing analysis in the paper "Simple Bounds for Recovering Low-Complexity Models," by Candes and Recht. To simplify the analysis, the notation used here is slightly different than in the paper.

Suppose that $x^* \in \mathbb{R}^n$ has support on $S \subset \{1, \dots, n\}$. Consider a matrix $\Phi \in \mathbb{R}^{m \times n}$ and suppose that we observe $z = \Phi x^* \in \mathbb{R}^m$. Our goal is to recover x^* from z by solving the optimization

$$\min_x \|x\|_1 \text{ subject to } \Phi x = \Phi x^* \ .$$

Under certain conditions, x^* is the unique solution to this optimization. Note that every x that satisfies the constraint can be decomposed as $x = x^* + h$, where $\Phi h = 0$. We want to show that $\|x^* + h\|_1 > \|x^*\|_1$ for every non-zero h satisfying $\Phi h = 0$.

Before stating the key conditions, let us introduce a little notation. Let S^c denote the complement of S and let $|S|$ denote the size of S. For any vector $y \in \mathbb{R}^n$, let y_S denote the restriction of y to S (i.e., y_S is equal to y on S and zero on S^c) and define y_{S^c} analogously. Note that $y = y_S + y_{S^c}$. Let Φ_S denote the $m \times |S|$ submatrix of Φ obtained by discarding all columns in S^c. Finally, recall that $g \in R^n$ is a subgradient of a function $f : \mathbb{R}^n \to \mathbb{R}$ at a point x if for all $h \in \mathbb{R}^n$

$$f(x + h) \ \geq \ f(x) \ + \ \langle g, h \rangle \ .$$

We are interested in subgradients of the function $\|x\|_1$. Let x_i denote the i-th element of x. If $x_i > 0$, then the subgradient in that direction is $+1$. If $x_i < 0$, then the subgradient in that direction is -1. If $x_i = 0$, then the subdifferential (set of subgradients) in that direction is $[-1, +1]$.

Now the following conditions suffice to guarantee that x^* is the unique solution. Suppose that

1. Φ_S has full rank

and there exists a $q \in \mathbb{R}^m$ such that $y = \Phi' q$ satisfies

2. $y_S = \text{sign}(x^*_S)$

3. $\|y_{S^c}\|_\infty < 1$

Then x^* is the unique solution to the optimization above.

To see this, define $v \in \mathbb{R}^n$ such that $v_{S^c} = \text{sign}(h_{S^c})$ and 0 elsewhere. Observe that $\text{sign}(x^*_S) + v$ is a subgradient of $\|\cdot\|_1$ at x^*. Then

$$\begin{aligned}
\|x^* + h\|_1 \ &\geq \ \|x^*\|_1 \ + \ \langle \text{sign}(x^*_S) + v, h \rangle \ , \text{ by definition of subgradient} \\
&= \ \|x^*\|_1 \ + \ \langle \text{sign}(x^*_S) + v - y, h \rangle \ , \text{ since } \langle y, h \rangle = q' \Phi h = 0 \\
&= \ \|x^*\|_1 \ + \ \langle \text{sign}(x^*_S) + v_S + v_{S^c} - (y_S + y_{S^c}), h \rangle \\
&= \ \|x^*\|_1 \ + \ \langle v_S + v_{S^c} - y_{S^c}, h \rangle \ , \text{ since } y_S = \text{sign}(x^*_S) \\
&= \ \|x^*\|_1 \ + \ \langle v_{S^c} - y_{S^c}, h \rangle \ , \text{ since } v_S = 0 \\
&= \ \|x^*\|_1 \ + \ \langle v_{S^c} - y_{S^c}, h_{S^c} \rangle \\
&\geq \ \|x^*\|_1 \ + \ \|h_{S^c}\|_1 - \|y_{S^c}\|_\infty \|h_{S^c}\|_1 \ , \text{ since } v_{S^c} = \text{sign}(h_{S^c}) \\
&= \ \|x^*\|_1 \ + \ (1 - \|y_{S^c}\|_\infty)\|h_{S^c}\|_1 > \|x^*\|_1 \ , \text{ since } \|y_{S^c}\|_\infty < 1 \text{ and } \|h_{S^c}\|_1 > 0
\end{aligned}$$

which gives us the result. The fact that $\|h_{S^c}\|_1 > 0$ follows from the assumption that Φ_S has full rank. Note that $h = h_S + h_{S^c}$, so $\Phi h = 0$ implies that $\Phi h_S = -\Phi h_{S^c}$. Because Φ_S has full rank, $h_{S^c} = 0$ implies that $h_S = 0$. Therefore, if $h \neq 0$, then $h_{S^c} \neq 0$.

Recall that for the compressed sensing result, $\Phi \in \mathbb{R}^{m\times n}$ with iid $\mathcal{N}(0,1)$ entries and $m > |S|$. It follows immediately that Φ_S has full rank for every S with probability 1. The other key issue is finding a q such that $y = \Phi' q$ satisfies the other two conditions. To satisfy the second condition, it suffices to take $q = \Phi_S(\Phi'_S\Phi_S)^{-1}\text{sign}(x^*(S))$, where $x^*(S)$ is the $|S| \times 1$ subvector of x^* composed of the elements in S. Using the concentration inequalities, it is possible to show that this choice also satisfies the third condition with high probability (see "Simple Bounds for Recovering Low-Complexity Models," by Candes and Recht, for details).