Lecture 18 — Oct 31st, 2013

Prof. Jelani Nelson | *Scribe: Cyril Stark*

1 Overview

This lecture is still about compressed sensing. Before we approximately recovered a signal x from measurements Πx using l_1-minimization which is a linear program. Unfortunately, linear programming is slow (poly(n) time instead of linear time). In this lecture we are stating and analyzing an iterative recovery algorithm to replace the slow l_1-minimization.

There exist many different iterative algorithms for compressed sensing: OMP, StOMP, ROMP, IHT, etc. We are focusing on IHT, i.e., *iterative hard thresholding* [2]. In each iteration of IHT, the complexity is determined by multiplication with Π. Thus, IHT is fast if Π allows for fast matrix multiplication. In Section 3 we show how such sensing matrices can be constructed from expanders.

2 Iterative hard thresholding (IHT)

We aim at recovering a signal $x \in \mathbb{R}^n$ from measurement data $y \in \mathbb{R}^m$,

$$y = \Pi x + e,$$

where e captures noise.

2.1 IHT Algorithm

- Initialize $x^{[0]} = 0$
- for $i = 0$ to t

  $x^{[i+1]} \leftarrow H_k(x^{[i]} + \Pi^T(y - \Pi x^{[i]}))$
- output $x^{[t]}$

Here, H_k is the so called *hard thresholding operator*; $H_k(x)$ keeps the k largest components of x invariant but sets to zero all the remaining components. Note that each iteration can be computed quickly if the RIP matrices Π allow for fast matrix multiplication.

2.2 Analysis

Theorem 1. *Suppose* Π *satisfies* $3k$*-RIP with error* $\varepsilon_{3k} < \frac{1}{4\sqrt{2}}$. *Then, if* $\tilde{x} = x^{[t+1]}$, *we have*

$$\|x - \tilde{x}\|_2 \lesssim 2^{-t}\|H_k(x)\|_2 + \|e\|_2 + \|x - H_k(x)\|_2 + \frac{\|x - H_k(x)\|_1}{\sqrt{k}}.$$

The Theorem implies that after $t = \theta\big(\log\big(\frac{\|H_k(x)\|}{\varepsilon}\big)\big)$ iterations

$$\text{error } \lesssim \|x - H_k(x)\|_2 + \frac{1}{\sqrt{k}}\|x - H_k(x)\|_1 + \|e\|_2 \ .$$

In the remainder

$$x^k = H_k(x).$$

Note: In contrast to the error bounds from lecture 16, the bounds from Theorem 1 involve $\|x - x^k\|_2$ (2-norm instead of 1-norm). However,

$$\|x - x^{2k}\|_2 \le \frac{1}{\sqrt{k}}\|x - x^k\|_1.$$

This can be seen as follows: Let S_0 denote the indices of the k largest components of x, let S_1 denote the next k largest components of x, etc. Then, by shelling,

$$\begin{aligned}
\|x - x^{2k}\|_2 &= \Big\|\sum_{r=2}^{n/k} x_{S_r}\Big\|_2 \\
&\le \sum_{r=2}^{n/k} \|x_{S_r}\|_2 \\
&\le \sqrt{k}\sum_{r=2}^{n/k} \|x_{S_r}\|_\infty \\
&\le \frac{1}{\sqrt{k}}\sum_{r=1}^{n/k} \|x_{S_r}\|_1 \\
&= \frac{1}{\sqrt{k}}\|x - x^k\|_1
\end{aligned}$$

2.3 Proof

First we prove Theorem 1 for signals x satisfying $x = x^k$. To proceed we need the following notation:

- $y = \Pi x^k + e$
- $r^{[t]} = x^k - x^{[t]}$
- $a^{[t+1]} = x^{[t]} + \Pi^T(y - \Pi x^{[t]})$
- $x^{[t+1]} = H_k(a^{[t+1]})$
- Γ_k^* denotes support(x^k) (thus, $|\Gamma_k^*| \le k$)
- $\Gamma^{[t+1]}$ denotes support$(x^{[t+1]})$ (thus, $|\Gamma^{[t+1]}| \le k$)
- $B^{[t+1]} = \Gamma_k^* \cup \Gamma^{[t+1]}$

We have

$$\|x^k - x^{[t+1]}\|_2 = \|x^k_{B^{[t+1]}} - x^{[t+1]}_{B^{[t+1]}}\|_2 \leq \|x^k_{B^{[t+1]}} - a^{[t+1]}_{B^{[t+1]}}\|_2 + \|a^{[t+1]}_{B^{[t+1]}} - x^{[t+1]}_{B^{[t+1]}}\|_2.$$

Since $x^{[t+1]}_{B^{[t+1]}}$ is the best k-sparse approximation of $a^{[t+1]}_{B^{[t+1]}}$, it follows that

$$\begin{aligned}\|x^k - x^{[t+1]}\|_2 &\leq 2\|x^k_{B^{[t+1]}} - a^{[t+1]}_{B^{[t+1]}}\|_2 \\ &= 2\left\|x^k_{B^{[t+1]}} - x^{[t]}_{B^{[t+1]}} - \left[\Pi^T(y - \Pi x^{[t]})\right]_{B^{[t+1]}}\right\|_2 \\ &= 2\left\|r^{[t]}_{B^{[t+1]}} - \left[\Pi^T\Pi r^{[t]}\right]_{B^{[t+1]}} - \left[\Pi^T e\right]_{B^{[t+1]}}\right\|_2,\end{aligned}$$

where we have used $y - \Pi x^{[t]} = \Pi(x^k - x^{[t]}) + e = \Pi r^{[t]} + e$ (by definition of y and $r^{[t]}$). It follows that

$$\begin{aligned}\|x^k - x^{[t+1]}\|_2 &\leq 2\left\|r^{[t]}_{B^{[t+1]}} - \Pi^T_{B^{[t+1]}}\Pi r^{[t]} - \Pi^T_{B^{[t+1]}} e\right\|_2 \\ &= 2\left\|r^{[t]}_{B^{[t+1]}} - \Pi^T_{B^{[t+1]}}\Pi\left(r^{[t]}_{B^{[t+1]}} + r^{[t]}_{B^{[t]}\setminus B^{[t+1]}}\right) - \Pi^T_{B^{[t+1]}} e\right\|_2 \\ &\leq 2\left\|(\mathbb{I} - \Pi^T_{B^{[t+1]}}\Pi_{B^{[t+1]}})r^{[t]}_{B^{[t+1]}}\right\|_2 + 2\left\|\Pi^T_{B^{[t+1]}}\Pi_{B^{[t]}\setminus B^{[t+1]}} r^{[t]}_{B^{[t]}\setminus B^{[t+1]}}\right\|_2 + 2\left\|\Pi^T_{B^{[t+1]}} e\right\|_2 \\ &\leq 2\left\|\mathbb{I} - \Pi^T_{B^{[t+1]}}\Pi_{B^{[t+1]}}\right\|\left\|r^{[t]}_{B^{[t+1]}}\right\|_2 + 2\left\|\Pi^T_{B^{[t+1]}}\Pi_{B^{[t]}\setminus B^{[t+1]}}\right\|\left\|r^{[t]}_{B^{[t]}\setminus B^{[t+1]}}\right\|_2 + 2\left\|\Pi^T_{B^{[t+1]}}\right\|\|e\|_2.\end{aligned}$$

Recall that Π is $3k$-RIP with error ε_{3k} means that for any S with $|S| \leq 3k$ all eigenvalues of $\Pi_S^T\Pi_S$ are in $[1-\varepsilon_{3k}, 1+\varepsilon_{3k}]$. Thus, $\left\|\mathbb{I} - \Pi^T_{B^{[t+1]}}\Pi_{B^{[t+1]}}\right\| \leq \varepsilon_{3k}$ and $\left\|\Pi^T_{B^{[t+1]}}\right\| \leq \sqrt{1+\varepsilon_{3k}}$ because $|B^{[t+1]}| \leq 2k$. Moreover, by considerations from lecture 16, $\left\|\Pi^T_{B^{[t+1]}}\Pi_{B^{[t]}\setminus B^{[t+1]}}\right\| \leq \varepsilon_{3k}$. We conclude that

$$\begin{aligned}\|x^k - x^{[t+1]}\|_2 &\leq 2\varepsilon_{3k}\left\|r^{[t]}_{B^{[t+1]}}\right\|_2 + 2\varepsilon_{3k}\left\|r^{[t]}_{B^{[t]}\setminus B^{[t+1]}}\right\|_2 + 2\sqrt{1+\varepsilon_{3k}}\,\|e\|_2 \\ &= 2\varepsilon_{3k}\left(\left\|r^{[t]}_{B^{[t+1]}}\right\|_2 + \left\|r^{[t]}_{B^{[t]}\setminus B^{[t+1]}}\right\|_2\right) + 2\sqrt{1+\varepsilon_{3k}}\,\|e\|_2 \\ &\leq 2\sqrt{2}\varepsilon_{3k}\|r^{[t]}\|_2 + 2\sqrt{1+\varepsilon_{3k}}\|e\|_2\end{aligned}$$

where we interpreted $\left\|r^{[t]}_{B^{[t+1]}}\right\|_2 + \left\|r^{[t]}_{B^{[t]}\setminus B^{[t+1]}}\right\|_2$ as an inner product which we bounded by Cauchy-Schwarz inequality. So we need

$$\varepsilon_{3k} < \frac{1}{2}\frac{1}{2\sqrt{2}} = \frac{1}{4\sqrt{2}}$$

so that

$$\|x^k - x^{[t+1]}\|_2 < \frac{1}{2}\|r^{[t]}\|_2 + 3\|e\|_2. \tag{1}$$

All of the above applies if the signal x satisfies $x = x^k$. For these k-sparse signals we have

$$y = \Pi x + e = \Pi x^k + e.$$

To deal with general (i.e., non-k-sparse) signals we now set $\tilde{e} = e + \Pi(x - x^k)$, so that the measurement of general signals x is captured by $y = \Pi x^k + \tilde{e}$. Plugging this into the iterated version of (1) yields

$$\|x^k - x^{[t]}\|_2 \lesssim 2^{-t}\|x^k\|_2 + \|\tilde{e}\|_2 \tag{2}$$

(recall that $x^{[0]} = 0$). Grouping the components of x according to their magnitude ($\rightarrow$ partitioning $(S_r)_r$), applying the triangle inequality twice and using that Π is $3k$-RIP we get

$$\begin{aligned}\|\tilde{e}\|_2 &\leq \|e\|_2 + \|\Pi(x - x^k)\|_2 \\ &\leq \|e\|_2 + \left\|\sum_{r=1}^{n/k} \Pi(x - x^k)_{S_r}\right\|_2 \\ &\leq \|e\|_2 + \sum_{r=2}^{n/k} \|\Pi(x - x^k)_{S_r}\|_2 \\ &\leq \|e\|_2 + \sqrt{1 + \varepsilon_{3k}} \sum_{r=2}^{n/k} \|(x - x^k)_{S_r}\|_2.\end{aligned}$$

By a shelling argument (see lecture 16),

$$\|\tilde{e}\|_2 \lesssim \|e\|_2 + \|x - x^k\|_2 + \frac{1}{\sqrt{k}}\|x - x^k\|_1.$$

Thus (see (2)),

$$\|x^k - x^{[t]}\|_2 \lesssim 2^{-t}\|x^k\|_2 + \|e\|_2 + \|x - x^k\|_2 + \frac{1}{\sqrt{k}}\|x - x^k\|_1.$$

This concludes the proof of the Theorem.

3 RIP$_1$ matrices for signal recovery

Recall that the standard RIP property is formulated in terms of the 2-norm.

Definition 2. *A matrix Π satisfies the (ε, k)-RIP$_1$ property if for all k-sparse vectors x*

$$(1 - \varepsilon)\|x\|_1 \leq \|\Pi x\|_1 \leq (1 + \varepsilon)\|x\|_1$$

We already noted before that the iterations of the IHT Algorithm can be computed quickly if Π supports fast matrix multiplication. This motivates the consideration of signal recovery for RIP$_1$-measurement matrices because there exist RIP$_1$-matrices which are sparse (see below). Even though not as good as for RIP matrices, we will still be able to prove (in the next lecture) recovery guarantees of the form

$$\|x - \tilde{x}\| \leq C\|x - x^k\|_1 \tag{3}$$

where $C = 1 + \delta$ for $\delta > 0$ arbitrarily small. Our main tool to construct RIP$_1$-matrices are expanders, and the approach described in the remaining part of today's notes are given in [1].

Definition 3. *Let $G = (U, V, E)$ be a d-regular bipartite graph with left vertices U, right vertices V and edges E. The graph G is a (k, ε)-expander if for all $S \subseteq U$ ($|S| \leq k$)*

$$|\Gamma(S)| \geq (1 - \varepsilon)d|S|.$$

Here, $\Gamma(S)$ denotes the neighborhood of S.

Claim 4. *There exist d-regular (k, ε)-expanders satisfying*

- $n = |U|$
- $m = |V| = \mathcal{O}\big(\frac{k}{\varepsilon^2} \log\big(\frac{n}{k}\big)\big)$
- $d = \mathcal{O}\big(\frac{1}{\varepsilon^2} \log\big(\frac{n}{k}\big)\big)$

This claim can be proven by picking G at random and showing that G happens to satisfy all this properties with positive probability. Thus, this approach is not constructive. As shown in [3], you can explicitly construct d-regular (k, ε)-expanders satisfying

- $n = |U|$
- $m = |V| = \mathcal{O}\big(d^2 k^{1+\alpha}\big)$
- $d = \mathcal{O}\big(\frac{1}{\varepsilon} \log(k) \log(n)\big)^{1+\frac{1}{\varepsilon}}$

Given a d-regular (k, ε)-expander G, we make the following ansatz for Π_G:

$$\Pi_G = \frac{1}{d}\big(\text{adjacency matrix of } G\big)$$

Then, $\Pi_G \in \mathbb{R}^{|V| \times |U|}$ with exactly d non-zero entries (equal to $1/d$) in each column.

Claim 5 ([1, Theorem 1])**.** *If G is a d-regular (k, ε)-expander, then Π_G is $(2\varepsilon, k)$-RIP$_1$.*

Definition 6. *The matrix Π satisfies the "C-restricted nullspace property of order k" if for all $\eta \in \mathrm{Ker}(\Pi)$ and for all $S \subseteq [n]$ of size k*

$$\|\eta\|_1 \leq C \|\eta_{\bar{S}}\|_1.$$

Details of the proof of the following claims can be found in the lecture notes [4].

Claim 7. *If Π satisfies $(10k, \varepsilon)$-RIP$_1$ property then Π satisfies the C-restricted nullspace property of order $2k$ with $C = 1 + (1 - \varepsilon^2)/2$.*

Punchline. If Π satisfies $C/2$-restricted nullspace property of order $2k$ and $\tilde{x}$ is the result of l_1-minimization, then

$$\|x - \tilde{x}\|_1 \leq \frac{2C}{2 - C} \|x - x^k\|_1$$

(recall (3)).

References

[1] Radu Berinde, Anna Gilbert, Piotr Indyk, Howard Karloff, and Martin Strauss. Combining geometry and combinatorics: a unified approach to sparse signal recovery. *Allerton*, 2008.

[2] Thomas Blumensath, Mike E. Davies. Iterative hard thresholding for compressed sensing. *Appl. Comput. Harmon. Anal.*, 27:265–274, 2009.

[3] Venkatesan Guruswami, Christopher Umans, and Salil P. Vadhan. Unbalanced expanders and randomness extractors from Parvaresh-Vardy codes. In *IEEE Conference on Computational Complexity*, pages 96–108. IEEE Computer Society, 2007.

[4] Piotr Indyk, Ronitt Rubinfeld. Sublinear algorithms. `http://stellar.mit.edu/S/course/6/sp13/6.893/courseMaterial/topics/topic2/lectureNotes/riplp/riplp.pdf`.