FAST SPARSE RECOVERY FOR ANY RIP-1 MATRIX

Eric Price

UT Austin
ecprice@cs.utexas.edu

ABSTRACT

The Restricted Isometry Property (RIP) is a useful measure of which measurement matrices will work for sparse recovery. The RIP-1 is an L1 variant of the RIP that can be satisfied by sparse matrices, allowing for faster embedding and recovery. While L1 minimization is guaranteed to work for all matrices satisfying the RIP-1, faster iterative techniques were only known to work when the matrix is the adjacency of an expander graph. We show that Sequential Sparse Matching Pursuit (SSMP) works on all matrices satisfying the RIP-1, giving the first demonstration of near-linear recovery time for arbitrary RIP-1 matrices.

Index Terms**—** Sparse recovery, compressed sensing

1. INTRODUCTION

Compressed sensing is a popular framework for signal recovery developed over the past decade. The aim is to recover a k-sparse vector $x \in \mathbb{R}^n$ from a noisy measurement $b = Ax + \mu$, where $A \in \mathbb{R}^{m \times n}$ is a well-chosen "measurement matrix" with $m \ll n$ and μ is arbitrary noise. From this observation b, the goal is to recover an estimate $\widehat{x}$ of x with

$$\|\widehat{x} - x\| = O(\|\mu\|).$$

Much of the compressed sensing literature is based on the Restricted Isometry Property (RIP). Unfortunately, in the parameter regime of interest the RIP only holds for dense matrices [1], which are inefficient to store and manipulate. This led to the introduction of the RIP-1, an ℓ_1 variant of the RIP, which is achievable with sparse matrices and is sufficient for L1 minimization to achieve robust sparse recovery [2]:

Definition 1.1. *A matrix $A \in \mathbb{R}^{m \times n}$ satisfies the (k, ϵ)* RIP-1 *if, for all k-sparse $x \in \mathbb{R}^n$,*

$$(1 - \epsilon) \|Ax\|_1 \leq \|x\|_1 \leq \|Ax\|_1 .$$

For binary matrices, the RIP-1 is essentially equivalent to A being the adjacency matrix of an unbalanced bipartite expander [2]. As an example, a random binary matrix with $O(\frac{1}{\epsilon} \log(n/k))$ ones per column and $m = O(\frac{1}{\epsilon^2} k \log(n/k))$ will usually satisfy the (k, ϵ) RIP-1. However, the definition encompasses nonbinary matrices.

One interesting example of nonbinary RIP-1 matrices comes from randomly flipping the sign of each entry of an expander adjacency matrix. The result, which behaves like a random sparse $\{0, \pm 1\}$ matrix, is quite similar to a COUNTSKETCH matrix [3]. The random signs cause the noise to largely cancel itself out, leading to better performance than binary matrices like COUNT-MIN [4, 5].

In this paper, we show that Sequential Sparse Matching Pursuit (SSMP), an iterative algorithm introduced in [6], works on arbitrary RIP-1 matrices. Previously, SSMP was only known to work for expander adjacency matrices, as did all other fast iterative methods [7, 8, 9, 10, 11].

Theorem 1.2. *Let $A \in \mathbb{R}^{m \times n}$ satisfy the $(ck, \frac{1}{10})$ RIP-1 for some (sufficiently large) constant c. Let $x \in \mathbb{R}^n$ be k-sparse, $\mu \in \mathbb{R}^m$, and let $\widehat{x}$ be the result of running SSMP on $(A, Ax + \mu, k)$. Then*

$$\|\widehat{x} - x\|_1 = O(\|\mu\|_1). \tag{1}$$

Alternatively, one could give a guarantee for nonsparse inputs, e.g. for all $x \in \mathbb{R}^n$ SSMP gives $\widehat{x}$ with

$$\|\widehat{x} - x\|_1 = O(\min_{k\text{-sparse } x'} \|x' - x\|_1).$$

This follows from Theorem 1.2 and the triangle inequality, since we can set $\mu = A(x' - x)$ and then $\|\mu\|_1 \leq \sum_j \|A(x' - x)_j\|_1 \leq \frac{1}{1-\epsilon} \|x' - x\|_1$.

1.1. Related work.

The RIP-1 was introduced in [2], where it was shown to imply robust recovery using L1 minimization. This gives fast embedding times, since the measurement matrix can be sparse, but not fast recovery. Subsequently, a number of fast iterative methods have been proposed. Expander Matching Pursuit (EMP) [7], Sparse Matching Pursuit (SMP) [8], and Sequential Sparse Matching Pursuit (SSMP) [6] all use $\widetilde{O}(n)$ recovery time and get the robustness guarantee (1) (see also [12] for a survey). Another set of work applies to "exact" recovery with $\mu = 0$ [9, 10] or very small [11]. The phase transition of some of these methods is explored in [13]. However, all of these iterative methods use more than just the RIP-1: they expect the matrix to be the adjacency of an expander.

2. SSMP ALGORITHM

The SSMP algorithm is shown as Algorithm 2.1. It uses the following definition: for any vector $x \in \mathbb{R}^n$, define $H_k(x) \in \mathbb{R}^n$ to be the restriction of x to its k largest coefficients. In [6] it was shown how to maintain a data structure to implement the algorithm in $O(nd(d + \log n))$ time per inner loop, where d is the matrix column sparsity (i.e. typically $d = O(\log \frac{n}{k})$). The method works for arbitrary RIP-1 matrices, not just expander adjacency matrices.

procedure SSMP(A, b, k)
  $\widehat{x}^0 = 0$
  for $j \leftarrow 1, \ldots, T = O(\log \|x\|_1 / \|\mu\|_1)$ **do**
    $\widehat{x}^{j,0} \leftarrow \widehat{x}^{j-1}$
    for $a \leftarrow 1, \ldots, r = (c-1)k + 1$ **do**
      $(i, z) \leftarrow \arg\min_{(i,z)} \left\|b - A(\widehat{x}^{j,a-1} + ze_i)\right\|_1$
      $\widehat{x}^{j,a} \leftarrow \widehat{x}^{j,a-1} + ze_i$
    end for
    $\widehat{x}^j \leftarrow H_k(\widehat{x}^{j,r})$ ▷ Restrict to k terms
  end for
  return $x' = \widehat{x}^T$
end procedure

Algorithm 2.1: SSMP.

The SSMP algorithm iteratively refines its estimate $\widehat{x}^{j,a}$ of the signal x. The inner loop adds a single coordinate to the estimate, which we show decreases the residual error $\left\|A(x - x^{j,a}) + \mu\right\|_1$ by at least a $1 - \frac{1}{O(k+a)}$ factor, unless we have already converged to $\left\|x - \widehat{x}^{j,a}\right\|_1 = O(\|\mu\|_1)$.

After $r = O(k)$ rounds, we will have that

$$\left\|A(x - \widehat{x}^{j,r}) + \mu\right\|_1 < \frac{1}{8} \left\|A(x - \widehat{x}^{j,0}) + \mu\right\|.$$

At this point, having updated our estimate many times, it is starting to lose the sparsity we need to apply the RIP-1. Therefore in our outer loop, we resparsify $\widehat{x}^{j,a}$ back down to k terms. This increases $\left\|A(x - \widehat{x}^{j,a}) + \mu\right\|_1$ by at most a $2 + O(\epsilon)$ factor, so we get

$$\left\|A(x - \widehat{x}^{j+1,0}) + \mu\right\|_1 \leq \frac{1}{2} \left\|A(x - \widehat{x}^{j,0}) + \mu\right\|_1$$

in every round, until it converges to $O(\|\mu\|_1)$.

The tricky part, and the novel part of this paper, is showing progress in each step of the inner loop. The original argument relied on properties of expander matrices; here, we solely consider geometry and the RIP-1. We discuss it in Section 3. The other parts are straightforward, and covered in Section 4.

3. PROOF OF SEQUENTIAL PROGRESS

We start with a geometric lemma about the ℓ_1 norm:

Lemma 3.1. *Let $x_1, \ldots, x_s, \mu \in \mathbb{R}^m$, and $z = \mu + \sum x_i$. Suppose that $\|\mu\|_1 < c\|z\|_1$ and*

$$(1 - \epsilon)(\sum_{i=1}^{s} \|x_i\|_1) \leq \left\|\sum_{i=1}^{s} x_i\right\|_1,$$

for some constants $0 \leq c, \epsilon < 1/2$. Then there exists an i such that $\|z - x_i\|_1 \leq (1 - \frac{1}{s}(1 - 2\epsilon - 5c))\|z\|_1$.

Intuitively, the condition means the x_i form a chain that is nearly at its maximal length; it is nearly "taut." Almost all the mass needs to be oriented toward the final vector z; very little is "slack" that can be "wasted" by moving in superfluous directions. On average, the x_i are pointed in the right direction and fairly large; hence at least one x_i is both of these.

Proof. Define the "projection" operator $p(a, b)$ of a onto b to be the coordinatewise nearest neighbor of a to the intervals $[0, b_i]$ for each coordinate i. That is, for positive coordinates $b_i \geq 0$, we define

$$p(a,b)_i = \begin{cases} 0 & \text{if } a_i < 0 \\ a_i & \text{if } 0 \leq a_i \leq b_i \\ b_i & \text{if } a_i > b_i \end{cases}$$

and analogously for negative coordinates (so $p(a, b)_i = -p(-a, -b)_i$). As a property of this operator, for all a, b we have $\|b - p(a, b)\|_1 = \|b\|_1 - \|p(a, b)\|_1$.

For simplicity of notation, let $v_i = x_i$ for $i \geq 1$ and $v_0 = \mu$, so $z = \sum_{i=0}^{s} v_i$. Let $u_i = p(v_i, z)$, and $w_i = \|v_i - u_i\|_1 = \|v_i\|_1 - \|u_i\|_1$. Then u_i is the part of v_i moving in the right direction, and w_i is the amount of mass "wasted" in the wrong direction. In particular,

$$\begin{aligned} \|z - v_i\|_1 &= \|z - u_i\|_1 + \|u_i - v_i\|_1 \\ &= \|z\|_1 - \|u_i\|_1 + w_i \end{aligned} \tag{2}$$

So we just want to show that some i has large $\|u_i\|_1 - w_i$. First we will show that $\|u_i\|_1$ is large on average, then that w_i is small on average, and hence the difference is large for at least one i. First, we claim

$$\sum_{i=0}^{s} \|u_i\|_1 \geq \|z\|_1 . \tag{3}$$

We do this by showing that for any coordinate j, $\sum_{i=0}^{s} |(u_i)_j| \geq |z_j|$. WLOG suppose $z_j \geq 0$, so $(u_i)_j \geq 0$ for all i. Then by the definition of projection, for each i either $(u_i)_j \geq (v_i)_j$ or $(u_i)_j = z_j$. If the latter ever happens, $\sum_{i=0}^{s}(u_i)_j \geq \max\,(u_i)_j = z_j$; otherwise, $\sum_{i=0}^{s}(u_i)_j \geq \sum_{i=0}^{s}(v_i)_j = z_j$.

Now, consider showing that the w_i are small. Intuitively, this is "wasted" mass that doesn't help reach the goal: it's in the wrong direction, or overshooting the mark. We don't have enough slack to waste much mass, so $\sum_{i=1}^{s} w_i$ must be small. In equations,

$$\begin{aligned} \frac{1}{1-\epsilon}\left\|\sum_{i=1}^{s} x_i\right\|_1 &\geq \sum_{i=1}^{s} \|x_i\|_1 \\ &= \sum_{i=1}^{s}(w_i + \|u_i\|_1) \\ &\geq \|z\|_1 - \|u_0\|_1 + \sum_{i=1}^{s} w_i \\ &\geq \left(\left\|\sum_{i=1}^{s} x_i\right\|_1 - \|\mu\|_1\right) - \|u_0\|_1 + \sum_{i=1}^{s} w_i \end{aligned}$$

and hence

$$\begin{aligned} \frac{\epsilon}{1-\epsilon}\left\|\sum_{i=1}^{s} x_i\right\|_1 &\geq -2\|\mu\|_1 + \sum_{i=1}^{s} w_i \\ \frac{\epsilon}{1-\epsilon}(\|z\|_1 + \|\mu\|_1) &\geq -2\|\mu\|_1 + \sum_{i=1}^{s} w_i \end{aligned}$$

so

$$\sum_{i=1}^{s} w_i \leq (2 + \frac{\epsilon}{1-\epsilon})\|\mu\|_1 + \frac{\epsilon}{1-\epsilon}\|z\|_1 . \tag{4}$$

Hence we have that the "non-wasted" mass u_i is large, and the "wasted" mass w_i is small. We just need to show that some particular i has large $\|u_i\|_1 - w_i$, but this will be true on average.

Subtracting Equation 4 from Equation 3,

$$\begin{aligned} &\sum_{i=1}^{s} \|u_i\|_1 - w_i \\ &\geq (\|z\|_1 - \|u_0\|_1) - ((2 + \frac{\epsilon}{1-\epsilon})\|\mu\|_1 + \frac{\epsilon}{1-\epsilon}\|z\|_1) \\ &\geq (1 - \frac{\epsilon}{1-\epsilon})\|z\|_1 - (3 + \frac{\epsilon}{1-\epsilon})\|\mu\|_1 \\ &\geq (1 - 3c - \frac{\epsilon(1+c)}{1-\epsilon})\|z\|_1 . \end{aligned}$$

So for $\epsilon \leq 1/2$,

$$\sum_{i=1}^{s} \|u_i\|_1 - w_i \geq (1 - 2\epsilon - 5c)\|z\|_1 . \tag{5}$$

Let j be such that $\|u_j\|_1 - w_j$ is above the mean. Then by Equation 5 and Equation 2,

$$\begin{aligned} \|u_j\|_1 - w_j &\geq \frac{1}{s}(1 - 2\epsilon - 5c)\|z\|_1 \\ \|z - x_j\|_1 &= \|z\|_1 - \|u_j\|_1 + w_j \\ &\geq (1 - \frac{1}{s}(1 - 2\epsilon - 5c))\|z\|_1 \end{aligned}$$

as desired. □

Now we can apply Lemma 3.1 to matrices satisfying the RIP:

Lemma 3.2. *Suppose A satisfies an RIP-1 of order $(s, 1/10)$, $s > 1$. If y is s-sparse, and $\|w\|_1 \leq \frac{1}{30}\|y\|_1$, then there exists a 1-sparse z such that*

$$\|A(y - z) + w\|_1 \leq e^{-\frac{1}{2s}}\|Ay + w\|_1 .$$

Proof. First, note that

$$\|w\|_1 \leq \frac{1}{30(1-\epsilon)} \|Ay\|_1 \leq \frac{1}{27} \|Ay\|_1 \\ \leq \frac{1}{26} \|Ay + w\|_1 .$$

Split y as $y = y_1 + y_2 + \ldots + y_s$, for orthogonal 1-sparse y_i. Let $v_i = Ay_i$. Let $\epsilon = 1/10$, so we have by the RIP-1 of order (s, ϵ) that

$$(1-\epsilon)\|v_i\|_1 \leq \|y_i\|_1 \leq \|v_i\|_1$$

and

$$\left\|\sum_{i=1}^{s} v_i\right\|_1 = \|Ay\|_1 \geq \|y\|_1 \geq (1-\epsilon)\sum_{i=1}^{s}\|v_i\|_1 .$$

Hence we can apply Lemma 3.1: for any noise vector w with $\|w\|_1 \leq c\|Ay+w\|_1$, there exists a j with

$$\|A(y-y_j)+w\|_1 \leq (1-\frac{1}{s}(1-2\epsilon-5c))\|Ay+w\|_1 .$$

For $\epsilon \leq 1/10$ and $c \leq 1/25$, this gives

$$\|A(y-y_j)+w\|_1 \leq (1-\frac{3}{5s})\|Ay+w\|_1 \\ \leq e^{-\frac{1}{2s}}\|Ay+w\|_1$$

So $z = y_j$ satisfies the desired result. □

4. ANALYSIS OF SSMP

We now apply our results to analyze SSMP. Since $\widehat{x}^{j,a}$ is $(k+a)$-sparse, $\widehat{x}^{j,a} - x$ is $(2k+a)$-sparse. Hence we have the following corollary of Lemma 3.2:

Corollary 4.1. *In SSMP, if A satisfies an RIP-1 of order $((c+1)k, 1/10)$, and $\|\mu\|_1 \leq \frac{1}{30}\left\|\widehat{x}^{j,a} - x\right\|_1$, then*

$$\left\|A\widehat{x}^{j,a+1} - b\right\|_1 \leq e^{-\frac{1}{2(2k+a)}}\left\|A\widehat{x}^{j,a} - b\right\|_1$$

for all j and a.

If we telescope this, with $H_n \approx \log n$ denoting the nth harmonic number $\sum_{i=1}^{n}\frac{1}{i}$, we have

$$\left\|A\widehat{x}^{j,t} - b\right\|_1 \leq e^{-\frac{1}{2}(H_{2k+t-1}-H_{2k-1})}\left\|A\widehat{x}^{j,0} - b\right\|_1 .$$

Setting $t = ck+1$, since $H_{(2+c)k} - H_{2k-1} > \log\frac{2+c}{2}$, we have

$$\left\|A\widehat{x}^{j,ck+1} - b\right\|_1 \leq \sqrt{\frac{2}{2+c}}\left\|A\widehat{x}^{j,0} - b\right\|_1 .$$

For $c = 128$, this gives

$$\left\|A\widehat{x}^{j,ck+1} - b\right\|_1 \leq \frac{1}{8}\left\|A\widehat{x}^{j,0} - b\right\|_1 .$$

Because A satisfies the RIP-1 we know

$$\left\|A(\widehat{x}^{j+1,t} - x) - \mu\right\|_1 \geq \left\|A(\widehat{x}^{j+1,t} - x)\right\|_1 - \|\mu\|_1 \\ \geq (1-\epsilon)\left\|\widehat{x}^{j+1,t} - x\right\|_1 - \|\mu\|_1$$

so since $\epsilon < 1/2$,

$$\left\|\widehat{x}^{j+1,t} - x\right\|_1 \leq 2\left\|A\widehat{x}^{j+1,t} - b\right\|_1 + 2\|\mu\|_1 \\ \leq \frac{1}{4}\left\|A\widehat{x}^{j} - b\right\|_1 + 2\|\mu\|_1 \\ \leq \frac{1}{4}\left\|A(\widehat{x}^{j} - x)\right\|_1 + \frac{9}{4}\|\mu\|_1 \\ \leq \frac{1}{4}\left\|\widehat{x}^{j} - x\right\|_1 + \frac{9}{4}\|\mu\|_1 .$$

Since x is k-sparse and $\widehat{x}^{j+1} = H_k(\widehat{x}^{j+1,t})$ is the nearest k-sparse vector to $\widehat{x}^{j+1,t}$, we have by the triangle inequality that

$$\left\|\widehat{x}^{j+1} - x\right\|_1 \leq 2\left\|\widehat{x}^{j+1,t} - x\right\|_1 \leq \frac{1}{2}\left\|\widehat{x}^{j} - x\right\|_1 + \frac{9}{2}\|\mu\|_1$$

Now, if $\|\mu\|_1 \leq \frac{1}{18}\left\|\widehat{x}^{j} - x\right\|_1$, we have

$$\left\|\widehat{x}^{j+1} - x\right\|_1 \leq \frac{3}{4}\left\|\widehat{x}^{j} - x\right\|_1 .$$

This means the error decreases to $O(\|\mu\|_1)$ in $T = O(\log\frac{\|x\|_1}{\|\mu\|_1})$ iterations, after which it never grows larger then $O(\|\mu\|_1)$. This gives Theorem 1.2.

5. CONCLUSION

We have shown that the SSMP algorithm gives deterministic ℓ_1 sparse recovery for arbitrary RIP-1 matrices, not just those binary ones. This allows us to use matrices, such as the COUNTSKETCH matrix, that are also suitable for high-probability ℓ_2 recovery. One interesting question is whether SSMP on such a matrix will have high-probability ℓ_2 recovery. One can show that one of the early estimates $x^{0,a}$ will be close to the Count-Sketch estimate, and hence a good ℓ_2 estimate with high probability; however, we do not know how to show that the later adjustments do not combine into a bad ℓ_2 result.

6. REFERENCES

[1] Venkat Bala Chandar, *Sparse graph codes for compression, sensing, and secrecy*, Ph.D. thesis, Massachusetts Institute of Technology, 2010.

[2] Radu Berinde, Anna C Gilbert, Piotr Indyk, Howard Karloff, and Martin J Strauss, "Combining geometry and combinatorics: A unified approach to sparse signal recovery," in *Communication, Control, and Computing, 2008 46th Annual Allerton Conference on*. IEEE, 2008, pp. 798–805.

[3] M. Charikar, K. Chen, and M. Farach-Colton, "Finding frequent items in data streams," *ICALP*, 2002.

[4] G. Cormode and S. Muthukrishnan, "Improved data stream summaries: The count-min sketch and its applications," *LATIN*, 2004.

[5] Gregory T Minton and Eric Price, "Improved concentration bounds for count-sketch," in *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 2014, pp. 669–686.

[6] Radu Berinde and Piotr Indyk, "Sequential sparse matching pursuit," in *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*. IEEE, 2009, pp. 36–43.

[7] Piotr Indyk and Milan Ruzic, "Near-optimal sparse recovery in the l1 norm," in *Foundations of Computer Science, 2008. FOCS'08. IEEE 49th Annual IEEE Symposium on*. IEEE, 2008, pp. 199–207.

[8] Radu Berinde, Piotr Indyk, and Milan Ruzic, "Practical near-optimal sparse recovery in the l1 norm," in *Communication, Control, and Computing, 2008 46th Annual Allerton Conference on*. IEEE, 2008, pp. 198–205.

[9] Weiyu Xu and Babak Hassibi, "Efficient compressive sensing with deterministic guarantees using expander graphs," in *Information Theory Workshop, 2007. ITW'07. IEEE*. IEEE, 2007, pp. 414–419.

[10] Shriram Sarvotham, Dror Baron, and Richard G Baraniuk, "Sudocodes fast measurement and reconstruction of sparse signals," in *2006 IEEE International Symposium on Information Theory*. IEEE, 2006, pp. 2804–2808.

[11] Sina Jafarpour, Weiyu Xu, Babak Hassibi, and Robert Calderbank, "Efficient and robust compressed sensing using optimized expander graphs," *IEEE Transactions on Information Theory*, vol. 55, no. 9, pp. 4299–4308, 2009.

[12] Anna Gilbert and Piotr Indyk, "Sparse recovery using sparse matrices," Institute of Electrical and Electronics Engineers, 2010.

[13] Bubacarr Bah and Jared Tanner, "Vanishingly sparse matrices and expander graphs, with application to compressed sensing," *IEEE transactions on information theory*, vol. 59, no. 11, pp. 7491–7508, 2013.