Improved Analysis of Sequential Sparse Matching Pursuit

Eric Price

November 12, 2015

Abstract

Sequential Sparse Matching Pursuit (SSMP) is a compressed sensing algorithm introduced in [BI] for nearly linear $\ell_1$ recovery with optimal measurements. We present a simpler and more general proof of convergence than in the original paper. Our proof shows that SSMP works for all matrices satisfying an appropriate RIP-1.

1 Introduction

We give a new proof that SSMP converges in nearly linear time. The key result is a new proof that the innermost loop of SSMP always makes significant progress. The rest of the proof proceeds as in [BI], or indeed any other matching pursuit algorithm. Nevertheless, we include the complete proof for completeness.

2 Preliminaries

Here we define things.

$H_k(x)$, $k$-sparse, RIP-1.

3 Algorithm

For any vector $x \in \mathbb{R}^n$, define $H_k(x)$ to be the $k$ largest coefficients of $x$. That is, $H_k(x)$ equals $x$ over its $k$ largest coefficients (in magnitude) and 0 otherwise.

```
1: procedure SSMP(A, b, k)    \triangleright Recover $k$-sparse approximation $x'$ to $x$ from $b = Ax + \mu$
2:     $x^0 = 0$
3:   for $j \leftarrow 1, \ldots, T = O(\log \|x\|_1 / \|\mu\|_1)$ do
4:       $x^{j,0} \leftarrow x^{j-1}$
5:         for $a \leftarrow 1, \ldots, (c-1)k$ do
6:             $(i, z) \leftarrow \text{arg min}_{(i,z)} \|b - A(x^j + ze_i)\|_1$
7:             $x^{j,a} \leftarrow x^{j,a-1} + ze_i$
8:       end for
9:     $x^j \leftarrow H_k(x^{j,(c-1)k})$ \hspace{1cm} \triangleright Extract the $k$ largest coefficients
10:   end for
11: return $x' = x^T$
12: end procedure
```

Algorithm 3.1: SSMP.
Define the remainders \( y^j = x - x^j \), \( y^{j,a} = x^j - x^{j,a} \). To show that SSMP works, we show that as long as \( \|y^{j,a}\|_1 = \omega(\|\mu\|_1) \), the successive \( y \) will decrease significantly. Hence the algorithm quickly converges to \( \|y^{j,a}\|_1 = O(\|\mu\|_1) \). In slightly more detail, we will show that when \( \|y^{j,a}\|_1 = \omega(\|\mu\|_1) \),

1. Each inner loop decreases the norm of the remainder (under \( A \)) by at least a \( (1 - \frac{1}{O(k)}) \) fraction.
   In particular \( \|Ax^{j,a+1} - b\|_1 \leq (1 - \frac{1}{2k+a})^{1/2} \|Ax^{j,a} - b\|_1 \), for \( A \) satisfying an appropriate RIP.
2. Hence \( \|Ax^{j+1,(c-1)k} - b\|_1 \leq \frac{1}{8} \|Ax^{j} - b\|_1 \) for \( c = 127 \).
3. Hence \( \|x^{j+1,(c-1)k} - x\|_1 \leq \frac{1}{4} \|x^{j} - x\|_1 \).
4. Hence \( \|x^{j+1} - x\|_1 \leq \frac{1}{2} \|x^{j} - x\|_1 \).

The tricky part, and the novel part of this paper, is step 1. We discuss it in Section 4. The other parts are straightforward, and covered in Section 5.

## 4 Proof of Sequential Progress

Before proving Lemma 2, we will prove a lemma about nearly orthogonal vectors in the \( \ell_1 \) norm:

**Lemma 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

\[
(1 - \delta)(\sum \|x_i\|_1) \leq \|\sum x_i\|_1,
\]

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

**Proof.** 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.

More formally, 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, \( \|b - p(a, b)\|_1 = \|b\|_1 - \|p(a, b)\|_1 \) for all \( a, b \).

For simplicity of notation, let \( v_i = x_i \) for \( i \geq 1 \) and \( v_0 = \mu \), so \( z = \sum v_i \). 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,

\[
\|z - v_i\|_1 = \|z - u_i\|_1 + \|u_i - v_i\|_1 = \|z\|_1 - \|u_i\|_1 + w_i
\]  \hspace{1cm} (1)

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, which will show that the difference is large on average, and hence large for at least one \( i \).
We claim
\[ \sum \| u_i \|_1 \geq \| z \|_1. \tag{2} \]

We will show that for any coordinate \( j \), \( \sum |(u_i)_j| \geq |z_j| \). 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, \( \sum (u_i)_j \geq \max (u_i)_j = z_j\); otherwise, \( \sum (u_i)_j \geq \sum (v_i)_j = z_j\).

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 \( \sum w_i \) must be small. In equations,
\[
\sum \| x_i \|_1 = \sum (w_i + \| u_i \|_1) \geq \| z \|_1 - \| u_0 \|_1 + \sum w_i \\
\frac{1}{1-\delta} \| \sum x_i \|_1 \geq \| \sum x_i \|_1 - \| \mu \|_1 - \| u_0 \|_1 + \sum w_i \\
\frac{\delta}{1-\delta} \| \sum x_i \|_1 \geq -2 \| \mu \|_1 + \sum w_i \\
2 \| \mu \|_1 + \frac{\delta}{1-\delta} (\| z \|_1 + \| \mu \|_1) \geq \sum w_i
\]
hence
\[
\sum w_i \leq \left( 2 + \frac{\delta}{1-\delta} \right) \| \mu \|_1 + \frac{\delta}{1-\delta} \| z \|_1. \tag{3}
\]

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 3 from Equation 2,
\[
\sum_{i \geq 1} \| u_i \|_1 - w_i \geq (\| z \|_1 - \| u_0 \|_1) - ((2 + \frac{\delta}{1-\delta}) \| \mu \|_1 + \frac{\delta}{1-\delta} \| z \|_1) \\
\geq (1 - \frac{\delta}{1-\delta}) \| z \|_1 - (3 + \frac{\delta}{1-\delta}) \| \mu \|_1 \\
\geq (1 - 3c - \frac{\delta(1+c)}{1-\delta}) \| z \|_1.
\]

So for \( \delta \leq 1/2 \),
\[
\sum_{i \geq 1} \| u_i \|_1 - w_i \geq (1 - 2\delta - 5c) \| z \|_1. \tag{4}
\]

Let \( j \) be such that \( \| u_j \|_1 - w_j \) is above the mean. Then by Equation 4 and Equation 1,
\[
\| u_j \|_1 - w_j \geq \frac{1}{s} (1 - 2\delta - 5c) \| z \|_1 \\
\| z - x_j \|_1 = \| z \|_1 - \| u_j \|_1 + w_j \\
\geq (1 - \frac{1}{s} (1 - 2\delta - 5c)) \| z \|_1
\]
as desired.
Now we can apply Lemma 1 to matrices satisfying the RIP:

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

**Proof.** First, note that $\|w\|_1 \leq \frac{1}{30(1-\delta)} \|Ay\|_1 \leq \frac{1}{29} \|Ay\|_1 \leq \frac{1}{26} \|Ay + w\|_1$.

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

$$
\left\| \sum v_i \right\|_1 = \|Ay\|_1 \geq (1 - \delta) \sum \|y_i\|_1 \geq (1 - \delta) \sum \|v_i\|_1
$$

Hence

$$(1 - \delta) \sum \|v_i\|_1 \leq \left\| \sum v_i \right\|_1 \leq \sum \|v_i\|_1.
$$

So we can apply Lemma 1: for any noise vector $w$ with $\|w\|_1 \leq c \|Ay + w\|_1$, there exists a $j$ with

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

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

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

for $s \geq 2$. So one of the $y_j$ is an acceptable $z$ for the result. □

A corollary will apply this to SSMP:

**Corollary 3.** In SSMP, if $A$ satisfies an RIP-1 of order $((c+1)k, 1/10)$, and $\|\mu\|_1 \leq \frac{1}{30} \|x^{j,a} - x\|_1$, then

$$
\|Ax^{j,a+1} - b\|_1 \leq (1 - \frac{1}{2k+a})^{1/2} \|Ax^{j,a} - b\|_1
$$

for all $j$ and $a$.

**Proof.** Note that $x^{j,a}$ is $k+a$-sparse, so $x^{j,a} - x$ is $2k+a \leq (c+1)k$-sparse. Hence Lemma 2 applies, and there exists some update $z$ such that

$$
\|A(x^{j,a} + z) - b\|_1 = \|A(x^{j,a} - x + z) - \mu\|_1 \leq (1 - \frac{1}{2k+a})^{1/2} \|A(x^{j,a} - x) - \mu\|_1
$$

Because $x^{j,a+1}$ chooses an update that minimizes this quantity,

$$
\|Ax^{j,a+1} - b\|_1 \leq \|Ax^{j,a+1} - b\|_1 \leq (1 - \frac{1}{2k+a})^{1/2} \|Ax^{j,a} - b\|_1
$$

as desired. □
5 Full Proof of SSMP

From Corollary 3, we have

\[
\|Ax^{j+1, t} - b\|_1 \leq \left(\prod_{a=0}^{t-1} \left(1 - \frac{1}{2k + a}\right)^{1/2}\right) \|Ax^{j+1,0} - b\|_1
\]

\[
= \left(\prod_{a=0}^{t-1} \frac{2k + a - 1}{2k + a}\right)^{1/2} \|Ax^{j} - b\|_1
\]

\[
= \left(\frac{2k - 1}{2k + t - 1}\right)^{1/2} \|Ax^{j} - b\|_1
\]

\[
\leq \left(\frac{2k}{2k + t}\right)^{1/2} \|Ax^{j} - b\|_1
\]

so for \(t = 126k\), corresponding to the last iteration if \(c = 127\),

\[
\|Ax^{j+1, t} - b\|_1 \leq \frac{1}{8} \|Ax^{j} - b\|_1
\]

Because \(A\) satisfies an RIP we know

\[
\|A(x^{j+1, t} - x) - \mu\|_1 \geq \|A(x^{j+1, t} - x)\|_1 - \|\mu\|_1 \geq (1 - \delta) \|x^{j+1, t} - x\|_1 - \|\mu\|_1
\]

so since \(\delta < 1/2\),

\[
\|x^{j+1, t} - x\|_1 \leq 2 \|Ax^{j+1, t} - b\|_1 + 2 \|\mu\|_1
\]

\[
\leq \frac{1}{4} \|Ax^{j} - b\|_1 + 2 \|\mu\|_1
\]

\[
\leq \frac{1}{4} \|A(x^{j} - x)\|_1 + \frac{9}{4} \|\mu\|_1
\]

\[
\leq \frac{1}{4} \|x^{j} - x\|_1 + \frac{9}{4} \|\mu\|_1
\]

and by a well-known property of \(H_k\),

\[
\|x^{j+1} - x\|_1 \leq 2 \|x^{j+1, t} - x\|_1 \leq \frac{1}{2} \|x^{j} - x\|_1 + \frac{9}{2} \|\mu\|_1
\]

Now, if \(\|\mu\|_1 \leq \frac{1}{18} \|x^{j} - x\|_1\), we have \(\|x^{j+1} - x\|_1 \leq \frac{3}{4} \|x^{j} - x\|_1\). So the error decreases quickly until it is \(O(\|\mu\|_1)\).

References