Compressed Counting Meets Compressed Sensing

Ping Li, Cun-Hui Zhang, Tong Zhang

Department of Statistics and Biostatistics
Department of Computer Science
Rutgers, the State University of New Jersey
Piscataway, NJ 08854, USA

June 15, 2014
New Direction on Compressed Sensing (Sparse Recovery)

The goal of **Compressed Sensing** (sparse signal recovery) is to recover the signal $x$ from non-adaptive measurements $y$:

$$y_j = \sum_{i=1}^{N} x_i s_{ij}, \quad j = 1, 2, ..., M$$

$s$ is the **design matrix** (or sensing matrix), which might be part of the sensing hardware (e.g., camera, scanner, etc).

Classical compressed sensing used **Gaussian design**, i.e., $s_{ij} \sim N(0, 1)$.

We propose to use **heavy-tailed design**.
A Simple Interesting Story about Compressed Sensing

Sparse Signal: \( x_1 = x_2 = 1, \quad x_i = 0, \quad 3 \leq i \leq N \)

We know neither the locations nor the magnitudes of the nonzero coordinates.

Task: Recover \( \mathbf{x} \) from a small number of linear nonadaptive Measurements:

\[
y_j = \sum_{i=1}^{N} x_i s_{ij} = x_1 s_{1j} + x_2 s_{2j} = s_{1j} + s_{2j}, \quad j = 1, 2, ..., M.
\]

for this particular example.
A 3-Iteration 3-Measurement Scheme

\(\{y_j\}\) is the measurement vector and \(\{s_{ij}\}\) is the design matrix.

For this example, \(y_j = s_{1j} + s_{2j}\), \(j = 1, 2, \ldots, M\).

**Ratio Statistics:**

\[
\begin{align*}
    z_{1,j} &= y_j / s_{1j} = 1 + \frac{s_{2j}}{s_{1j}} \\
    z_{2,j} &= y_j / s_{2j} = 1 + \frac{s_{1j}}{s_{2j}} \\
    z_{i,j} &= y_j / s_{ij} = \frac{s_{1j}}{s_{ij}} + \frac{s_{2j}}{s_{ij}}, \ i \geq 3
\end{align*}
\]

**Ideal Design:** \(s_{2j} / s_{1j}\) is either 0 or \(\pm\infty\), i.e., \(z_{1,j} = 1\) or \(\pm\infty\).
Suppose we use $M = 3$ measurements.

First coordinate

$$z_{1,j} = \frac{y_j}{s_{1j}} = 1 + \frac{s_{2j}}{s_{1j}}$$

Second coordinate

$$z_{2,j} = \frac{y_j}{s_{2j}} = 1 + \frac{s_{1j}}{s_{2j}}$$

Suppose $s_{2j}/s_{1j}$ is either 0 or $\pm \infty$:

- $j = 1$: If $\frac{s_{2j}}{s_{1j}} = 0$, then $z_{1,1} = 1$ (truth), $z_{2,1} = \pm \infty$ (useless)
- $j = 2$: If $\frac{s_{2j}}{s_{1j}} = \pm \infty$, then $z_{1,2} = \pm \infty$, $z_{2,2} = 1$
- $j = 3$: If $\frac{s_{2j}}{s_{1j}} = 0$, then $z_{1,3} = 1$, $z_{2,3} = \pm \infty$

With 3 measurements, we see $z_{1,j} = 1$ twice and we safely estimate $\hat{x}_1 = 1$. 
In the second iteration, we compute the residuals and update the ratio statistics:

\[ r_j = y_j - \hat{x}_1 s_{1j} = s_{2j} \]

\[ z_{2,j} = r_j / s_{2j} = 1, \quad j = 1, 2, 3 \]

Therefore, we can correctly estimate \( \hat{x}_2 = 1 \).
In the third iteration, we update the residual and ratio statistics:

\[ r_j = 0 \]

\[ z_{i,j} = \frac{r_j}{s_{ij}} = 0, \ i \geq 3 \]

This means, all zeros are identified.

An important (and perhaps surprising) consequence:

\[ M = 3 \] measurements suffice for \( K = 2 \), regardless of \( N \).
Realization of the Ideal Design

It suffices to sample \( s_{ij} \) from \( \alpha \)-stable distribution: \( s_{ij} \sim S(\alpha, 1) \) with \( \alpha \to 0 \).

**The standard procedure:** \( w \sim exp(1), u \sim unif(-\pi/2, \pi/2) \), \( w \) and \( u \) are independent. Then

\[
\frac{\sin(\alpha u)}{(\cos u)^{1/\alpha}} \left[ \frac{\cos(u - \alpha u)}{w} \right]^{(1-\alpha)/\alpha} \sim S(\alpha, 1)
\]

which can be practically replaced by \( \pm \frac{1}{[unif(0,1)]^{1/\alpha}} \).

**Stability:** If \( S_1, S_2 \sim S(\alpha, 1) \) i.i.d., then for any constants \( C_1, C_2 \),

\[
C_1 S_1 + C_2 S_2 = S \times (|C_1|^\alpha + |C_2|^\alpha)^{1/\alpha}, \quad S \sim S(\alpha, 1)
\]
Ratio of Two Independent $\alpha$-Stable Variables

$S_1, S_2 \sim S(\alpha, 1)$ independent. Ratio $|S_2/S_1|$ is either very small or very large.

Recall, when $K = 2$, the ratio statistics are

\[
\begin{align*}
z_{1,j} &= y_j / s_{1j} = 1 + \frac{s_{2j}}{s_{1j}} \\
z_{2,j} &= y_j / s_{2j} = 1 + \frac{s_{1j}}{s_{2j}}
\end{align*}
\]
Advantages of the New Compressed Sensing Framework

Our proposal uses $\alpha$-stable distributions for small $\alpha$ (for this paper).

- **Computationally very efficient**, with the main cost being one linear scan.
- **Very robust to measurement noise**, unlike traditional methods.
- **Fewer (or at most the same) measurements** compared to LP.
- **The design matrix can be made very sparse**
  (e.g., Ping Li, *Very Sparse Stable Random Projections*, KDD’07)
- **Special treat for nonnegative signals**, i.e., this paper
Most natural signals (e.g., images) are nonnegative. Instead of using symmetric stable projections, skewed stable projections have significant advantages.

Ref: Li Compressed Counting, SODA’09.
Ref: Li Improving Compressed Counting, UAI’09.
Ref: Li and Zhang A New Algorithm for Compressed Counting ..., COLT’11.

With Compressed Counting, a very simple sparse recovery algorithm can be developed and precisely analyzed.
Maximally-Skewed $\alpha$-Stable Distribution

Denote $Z \sim S(\alpha, 1, 1)$, where the first “1” denotes maximal skewness and the second “1” denotes unit scale. Its characteristic function is

$$E \left( \exp \left( \sqrt{-1} Z \lambda \right) \right) = \exp \left( -|\lambda|^\alpha \left( 1 - \text{sign}(\lambda) \sqrt{-1} \tan \left( \frac{\pi \alpha}{2} \right) \right) \right)$$

Suppose $s_1, s_2 \sim S(\alpha, 1, 1)$ i.i.d. For any constants $c_1 \geq 0, c_2 \geq 0$, we have

$$c_1 s_1 + c_2 s_2 \sim S(\alpha, 1, c_1^\alpha + c_2^\alpha)$$

To sample from $S(\alpha, 1, 1)$, we sample $w \sim \exp(1), u \sim \text{unif} (0, \pi)$. Then

$$\frac{\sin (\alpha u)}{[\sin u \cos (\alpha \pi/2)]^{\frac{1}{\alpha}}} \left[ \frac{\sin (u - \alpha u)}{w} \right]^{\frac{1-\alpha}{\alpha}} \sim S(\alpha, 1, 1)$$
The Recovery Algorithm

Linear measurements

\[ y_j = \sum_{i=1}^{N} x_i s_{ij}, \quad j = 1, 2, \ldots, M \]

where \( s_{ij} \) is sampled i.i.d. from \( S(\alpha, 1, 1) \).

Minimum estimator

\[ \hat{x}_{i,min} = \min_{1 \leq j \leq M} \frac{y_j}{s_{ij}}, \quad i = 1, 2, \ldots, N \]
Sample Complexity of One-Scan Technique

**Theorem:** Suppose signal $\mathbf{x} \in \mathbb{R}^N$ is nonnegative, i.e., $x_i \geq 0, \forall i$. When $\alpha \in (0, 0.5]$, with $\alpha$-stable maximally-skewed stable projections, it suffices to use $M = C_{\alpha} \epsilon^{-\alpha} \left( \sum_{i=1}^{N} x_i^{\alpha} \right) \log N/\delta$ measurements, so that all coordinates will be recovered in one-scan within $\epsilon$ additive precision, with probability $1 - \delta$.

The constant $C_{0+} = 1$ and $C_{0.5} = \pi/2$. In particular, when $\alpha \rightarrow 0$ (exact sparse recovery), $M = K \log N/\delta$, where $K = \sum_{i=1}^{N} 1\{x_i \neq 0\}$. 
The Constant $C_\alpha$

$C_{0+} = 1$ and $C_{0.5} = \pi/2$. 
Comparison with Count-Min Sketch

**Complexity of Count-min sketch:** \( O \left( \epsilon^{-1} \sum_{i=1}^{N} x_i \log N/\delta \right) \).

**Our complexity bound:** \( C_{\alpha} \epsilon^{-\alpha} \sum_{i=1}^{N} x_i^{\alpha} \log N/\delta \)

- We know the exact constant \( C_{\alpha} \).
- Our \( \epsilon^{-\alpha} \) is an improvement of \( \epsilon^{-1} \).
- Whether or not \( \sum_{i=1}^{N} x_i \) is larger than \( \sum_{i=1}^{N} x_i^{\alpha} \) depends on the data.
- In fact, we can remove \( \sum_{i=1}^{N} x_i^{\alpha} \) by using very sparse compressed counting.
Experiments

Recovery Error

Decoding Time (Ratio)

N = 1000000, K = 10
M = K log N

Normalized Error

Ratio of Decoding Time

$\alpha$
For this case, we can not run L1Magic and only present comparison with SPGL1.

For $\alpha$ close to 0.5, we need to increase measurements, as in the analysis.
Recovery Error

Decoding Time (Ratio)

Normalized Error

Ratio of Decoding Time

L1Magic

SPGL1

CC

$N = 1000000, K = 10$

$M = 1.6K \log N$

$N = 1000000, K = 10$

$M = 1.6K \log N$
Recovery Error

Decoding Time (Ratio)

Normalized Error

Ratio of Decoding Time

\( N = 10000000, K = 10 \)
\( M = 1.6 K \log N \)

\( N = 10000000, K = 10 \)
\( M = 1.6 K \log N \)
Extensions


   **Main results**: The design matrix can be significantly sparsified (i.e., very sparse stable random projections as in KDD’07). The complexity is $eK \log N$ if the right sparsity is chosen.

2. **One Scan 1-Bit Compressed Sensing**, to be posted soon

   **Main Results**: Using heavy-tailed design and only the signs (1-bit) of the measurements, a one-scan algorithm can recover the support and the signs of the signals with $12.3K \log N$ measurements (a conservative version) or about $6K \log N$ (a more practical version).
Sign Recovery by One Scan 1-Bit Compressed Sensing

\[ \text{Error} = \frac{\sum_i |\text{sgn}(x_i) - \text{sgn}(\hat{x}_i)|}{K} \]

- **N = 1000, K = 10**
  - Sign Signal
  - Gaussian Signal

- **N = 10000, K = 20**
  - Sign Signal
  - Gaussian Signal
Summary of Contributions on Compressed Sensing

- Sparse recovery is a very active area of research in many disciplines: Mathematics, EE, CS, and perhaps Statistics.

- In classical settings, the design matrix for sparse recovery is sampled from Gaussian distribution, which is \( \alpha = 2 \)-stable distribution.

- Using \( \alpha \)-stable distribution with \( \alpha \approx 0 \) leads to simple, fast, robust, accurate exact sparse recovery. Cost is one linear scan, with no catastrophic failures.

- The design matrix can be made very sparse without hurting the performance. This connects to the influential work on sparse recovery with sparse matrices.

- This is just very preliminary work. There are numerous research problems and applications which we will study in the next a few years.