JMLR: Workshop and Conference Proceedings vol 35:1–20, 2014

Compressed Counting Meets Compressed Sensing

Ping Li PINGLI@STAT.RUTGERS.EDU
Department of Statistics and Biostatistics, Department of Computer Science,
Rutgers University, Piscataway, NJ 08854, USA

Cun-Hui Zhang CUNHUI@STAT.RUTGERS.EDU
Department of Statistics and Biostatistics, Rutgers University, Piscataway, NJ 08854, USA

Tong Zhang TZHANG@STAT.RUTGERS.EDU
Department of Statistics and Biostatistics, Rutgers University, Piscataway, NJ 08854, USA

Abstract

Compressed sensing (sparse signal recovery) has been a popular and important research topic in recent years. By observing that natural signals (e.g., images or network data) are often nonnegative, we propose a framework for nonnegative signal recovery using *Compressed Counting (CC)*. CC is a technique built on *maximally-skewed α-stable random projections* originally developed for data stream computations (e.g., entropy estimations). Our recovery procedure is computationally efficient in that it requires only one linear scan of the coordinates.

In our settings, the signal $\mathbf{x} \in \mathbb{R}^N$ is assumed to be nonnegative, i.e., $x_i \geq 0, \forall\, i$. We prove that, when $\alpha \in (0,\ 0.5]$, it suffices to use $M = (C_\alpha + o(1))\epsilon^{-\alpha}\left(\sum_{i=1}^N x_i^\alpha\right)\log N/\delta$ measurements so that, with probability $1-\delta$, all coordinates will be recovered within ϵ additive precision, in one scan of the coordinates. The constant $C_\alpha = 1$ when $\alpha \to 0$ and $C_\alpha = \pi/2$ when $\alpha = 0.5$. In particular, when $\alpha \to 0$, the required number of measurements is essentially $M = K \log N/\delta$, where $K = \sum_{i=1}^N 1\{x_i \neq 0\}$ is the number of nonzero coordinates of the signal.

1. Introduction

We develop a new framework for **compressed sensing** (sparse signal recovery) (Donoho and Stark, 1989; Donoho and Huo, 2001; Cormode and Muthukrishnan, 2005; Donoho, 2006; Candès et al., 2006). We focus on nonnegative sparse signals, i.e., $\mathbf{x} \in \mathbb{R}^N$ and $x_i \geq 0, \forall\, i$, by observing that real-world signals are often **nonnegative** (a phenomenon which we will comment more in the paper). We consider a typical scenario in which neither the magnitudes nor the locations of the nonzero entries of $\mathbf{x}$ are known. The task of compressed sensing reconstruction is to recover both the locations and the magnitudes of the nonzero entries. Our framework differs from mainstream work in compressed sensing in that we use maximally-skewed α-stable distributions for generating the design matrix, while classical compressed sensing algorithms typically adopt Gaussian or Gaussian-like distributions (e.g., distribution with finite variance). The use of skewed stable random projections was originally developed by Li (2009a,b); Li and Zhang (2011), named **Compressed Counting (CC)**, for data stream computations such as moment and entropy estimations.

In compressed sensing, the standard procedure collects M non-adaptive linear measurements

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

© 2014 P. Li, C.-H. Zhang & T. Zhang.

and reconstructs the signal $\mathbf{x}$ from the measurements, y_j, and the design matrix, s_{ij}. The design matrix is "designed" in that one can manually generate the entries to facilitate the recovery task. In fact, the design matrix can be integrated into sensing hardware (e.g., cameras or scanners). In classical settings, entries of the design matrix are sampled from Gaussian or Gaussian-like distributions. Well-known recovery algorithms are often based on linear programming (LP) (e.g., *basis pursuit* (Chen et al., 1998)) or greedy methods such as orthogonal matching pursuit (OMP) (Pati et al., 1993; Mallat and Zhang, 1993; Zhang, 2011; Tropp, 2004). In general, LP is computationally expensive. OMP is often faster but it requires scanning the coordinates (at least) K times.

It is desirable to develop a framework for sparse recovery which requires only one scan of the coordinates and does not require more measurements than LP or OMP. It is also desirable if the method is robust against measurement noises and is applicable to data streams. In this paper, our proposed method meets these requirements by sampling the entries of a design matrix from maximally-skewed α-stable distributions (Zolotarev, 1986).

1.1. Maximally-Skewed Stable Distributions

In our proposal, we sample entries of the design matrix s_{ij} from an α-stable maximally-skewed distribution, denoted by $S(\alpha, 1, 1)$, where the first "1" denotes maximal skewness and the second "1" denotes unit scale. If a random variable $Z \sim S(\alpha, 1, 1)$, then its characteristic function is

$$\mathscr{F}_Z(\lambda) = \text{E}\left(\exp\left(\sqrt{-1}Z\lambda\right)\right) = \exp\left(-|\lambda|^\alpha\left(1-\text{sign}(\lambda)\sqrt{-1}\tan\left(\frac{\pi\alpha}{2}\right)\right)\right), \quad \alpha \neq 1 \qquad (2)$$

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 $c_1 s_1 + c_2 s_2 \sim S(\alpha, 1, c_1^\alpha + c_2^\alpha)$. More generally, $\sum_{i=1}^N x_i s_i \sim S\left(\alpha, 1, \sum_{i=1}^N x_i^\alpha\right)$ if $s_i \sim S(\alpha, 1, 1)$ i.i.d.

To sample from $S(\alpha, 1, 1)$, we first generate an exponential random variable with mean 1, $w \sim \exp(1)$, and a uniform random variable $u \sim unif\ (0, \pi)$, and then compute (Chambers et al., 1976)

$$\frac{\sin(\alpha u)}{\left[\sin u \cos(\alpha\pi/2)\right]^{\frac{1}{\alpha}}}\left[\frac{\sin(u-\alpha u)}{w}\right]^{\frac{1-\alpha}{\alpha}} \sim S(\alpha, 1, 1) \qquad (3)$$

In practice, we can replace the stable distribution with a heavy-tailed distribution in the domain of attractions (Feller, 1971), for example, $[unif(0,1)]^{-1/\alpha}$ (at least for α not too close to 1).

1.2. Data Streams and Linear Projections

The use of maximally-skewed stable random projections for nonnegative data stream computations is known as *Compressed Counting (CC)* (Li, 2009a,b; Li and Zhang, 2011). Prior to the advent of CC, it was popular to use *symmetric stable random projections* (Indyk, 2006; Li, 2008). In the standard *turnstile* data stream model (Muthukrishnan, 2005), at time t, an arriving stream element (i_t, I_t) updates one entry of the data vector in a linear fashion: $x_{i_t}^{(t)} = x_{i_t}^{(t-1)} + I_t$. When the data streams arrive at high-speed (e.g., network traffic), the dynamic nature makes the task more challenging for computing summary statistics (e.g., $\sum_{i=1}^N |x_i|^2$) or for recovering nonzero entries.

Linear projections are naturally capable of handling data streams. To see this, suppose we denote the linear measurements as

$$y_j^{(t)} = \sum_{i=1}^N x_i^{(t)} s_{ij}, \quad j = 1, 2, ..., M \qquad (4)$$

When a new stream element (i_t, I_t) arrives, we only need to update the measurement as

$$y_j^{(t)} = y_j^{(t-1)} + I_t s_{i_t,j}, \quad j = 1, 2, ..., M \tag{5}$$

The entries $s_{i_t,j}$ are re-generated as needed by using pseudo-random numbers (Nisan, 1990), i.e., no need to materialize the entire design matrix. This is the standard practice in stream computations.

Note that when the data are nonnegative, i.e., $x_i^{(t)} \geq 0$ although the increment I_t can be either negative or positive, computing the first moment is trivial, because $\sum_{i=1}^N |x_i^{(t)}| = \sum_{i=1}^N x_i^{(t)} = \sum_{j=0}^t I_j$, i.e., merely one counter is needed. Based on this observation, (Li, 2009a,b; Li and Zhang, 2011) developed efficient estimators of $\sum_{i=1}^N |x_i^{(t)}|^\alpha$ for α close to 1. As the Shannon entropy is essentially the first derivative of $\sum_{i=1}^N |x_i^{(t)}|^\alpha$ as $\alpha \to 1$, this makes the known difficult problem of entropy estimation a trivial task. In a recent paper (Li and Zhang, 2011), merely 10 measurements seem to be sufficient for nonnegative data streams. Li and Zhang (2012) also developed *correlated symmetric stable random projections* for entropy estimation when the data streams can be negative.

For the rest of paper, we will drop the superscript (t) in $y_j^{(t)}$ and $x_i^{(t)}$, while readers should keep in mind that our results are naturally applicable to data streams.

1.3. Nonnegative Signals

Natural signals (such as images) are often nonnegative. For example, in database applications, a user can delete an existing entry but can not delete an entry which does not exist. Compressed sensing has been popular for estimating large entries in databases (Muthukrishnan, 2005). Readers probably have noticed that the turnstile data stream model is basically the **histogram model**. In network applications, monitoring traffic histograms is an important mechanism for (e.g.,) anomaly detections (Feinstein et al., 2003). Detecting (recovering) heavy components (e.g., so called "elephant detection" (Zhao et al., 2007)) using compressed sensing is an active research topic in networks; see (e.g.,) (Lin and Kung, 2012; Wang et al., 2012a,b) for some recent work in networks using compressed sensing. Of course, in machine learning applications (e.g.,vision and national language processing), it is a standard practice to generate features from histogram (e.g., bag-of-words model).

1.4. The Proposed Algorithm and Main Result

For recovering a nonnegative signal $x_i \geq 0$, $i = 1$ to N, we collect linear measurements $y_j = \sum_{i=1}^N x_i s_{ij}$, $j = 1$ to M, where $s_{ij} \sim S(\alpha, 1, 1)$ i.i.d. In this paper, we focus on $\alpha \in (0, 0.5]$. At the decoding stage, we estimate the signal coordinate-wise using the following *minimum estimator*:

$$\hat{x}_{i,min} = \min_{1 \leq j \leq M} y_j / s_{ij} \tag{6}$$

The number of measurements M is chosen so that $\sum_{i=1}^N \mathbf{Pr}\left(\hat{x}_{i,min} - x_i \geq \epsilon\right) \leq \delta$ (e.g., $\delta = 0.05$).

Main Result: When $\alpha \in (0,\ 0.5]$, it suffices to use $M = (C_\alpha + o(1))\epsilon^{-\alpha}\left(\sum_{i=1}^N x_i^\alpha\right)\log N/\delta$ measurements so that, with probability $1-\delta$, all coordinates will be recovered within ϵ additive precision, in one scan of the coordinates. The constant $C_\alpha = 1$ when $\alpha \to 0$ and $C_\alpha = \pi/2$ when $\alpha = 0.5$. In particular, when $\alpha \to 0$, the required number of measurements is essentially $M = K \log N/\delta$, where $K = \sum_{i=1}^N 1\{x_i \neq 0\}$ is the number of nonzero coordinates of the signal.

In the literature, it is known that the sample complexity of compressed sensing using Gaussian design (i.e., $\alpha = 2$) is essentially $2K \log N/\delta$ (Donoho and Tanner, 2009). This means our work already achieves smaller complexity with an explicit constant, by requiring only one linear scan of the coordinates. Encouragingly, it is not surprising that our work in this paper is merely a tip of the iceberg and we expect many promising research directions can be explored along this line.

Organization. We first provide some relevant probability results on the ratio of two stable random variables in Section 2. Then we analyze the proposed recovery algorithm in Section 3 and present an experimental study in Section 4 to confirm the theoretical results. Section 5 is devoted to the discussions and possible future research problems. Finally, Section 6 concludes the paper.

2. Relevant Probability Results on the Ratio of Two Stable Random Variables

Our proposed algorithm utilizes only the ratio statistics y_j/s_{ij} for the recovery task, while the observed data appear to contain more information, i.e., (y_j, s_{ij}) for $i = 1, 2, ..., N$, and $j = 1, 2, ..., M$. Thus, we first provide an explanation why we restrict ourselves to the ratio statistics.

For convenience, we define

$$\theta = \left(\sum_{i=1}^{N} x_i^\alpha\right)^{1/\alpha}, \quad \theta_i = (\theta^\alpha - x_i^\alpha)^{1/\alpha} \tag{7}$$

and denote the probability density function of $s_{ij} \sim S(\alpha, 1, 1)$ by f_S. By a conditional probability argument, the joint density of (y_j, s_{ij}) can be shown to be $\frac{1}{\theta_i} f_S(s_{ij}) f_S\left(\frac{y_j - x_i s_{ij}}{\theta_i}\right) \propto \frac{1}{\theta_i} f_S\left(\frac{y_j - x_i s_{ij}}{\theta_i}\right)$. The MLE amounts to finding (x_i, θ_i) to maximize the joint likelihood

$$L(x_i, \theta_i) = \prod_{j=1}^{M} \frac{1}{\theta_i} f_S\left(\frac{y_j - x_i s_{ij}}{\theta_i}\right) \tag{8}$$

Interestingly, Lemma 1 shows that $L(x_i, \theta_i)$ approaches infinity at the poles $y_j - x_i s_{ij} = 0$.

Lemma 1 *The likelihood in (8) approaches infinity, i.e.,* $L(x_i, \theta_i) \to +\infty$, *if* $y_j - x_i s_{ij} \to 0$, *for any* j, $1 \le j \le M$.

Proof: *See Appendix A.* □

Lemma 1 suggests to use the ratio statistics y_j/s_{ij} to recover x_i. By property of stable distributions,

$$\frac{y_j}{s_{ij}} = \frac{\sum_{t=1}^N x_t s_{tj}}{s_{ij}} = x_i + \frac{\sum_{t\neq i}^N x_t s_{tj}}{s_{ij}} = x_i + \theta_i \frac{S_2}{S_1} \tag{9}$$

where $\theta_i = \left(\sum_{t\neq i} x_t^\alpha\right)^{1/\alpha}$ and $S_1, S_2 \sim S(\alpha, 1, 1)$ i.i.d. Recall $\sum_{t\neq i}^N x_t s_{tj} \sim S\left(\alpha, 1, \sum_{t\neq i}^N x_t^\alpha\right)$.

This result motivates us to study the probability distribution of two independent stable random variables, S_2/S_1. For convenience, we also define

$$F_\alpha(t) = \mathbf{Pr}\left((S_2/S_1)^{\alpha/(1-\alpha)} \le t\right), \quad t \ge 0 \tag{10}$$

Lemma 2 *For any $t \geq 0$, $S_1, S_2 \sim S(\alpha, 1, 1)$, i.i.d.,*

$$F_\alpha(t) = \mathbf{Pr}\left((S_2/S_1)^{\alpha/(1-\alpha)} \leq t\right) = \frac{1}{\pi^2}\int_0^\pi \int_0^\pi \frac{1}{1+Q_\alpha/t} du_1 du_2 \tag{11}$$

where

$$Q_\alpha = \left[\frac{\sin(\alpha u_2)}{\sin(\alpha u_1)}\right]^{\alpha/(1-\alpha)} \left[\frac{\sin u_1}{\sin u_2}\right]^{\frac{1}{1-\alpha}} \frac{\sin(u_2 - \alpha u_2)}{\sin(u_1 - \alpha u_1)} \tag{12}$$

In particular, closed-form expressions are available when $\alpha \to 0+$ or $\alpha = 0.5$:

$$\lim_{\alpha\to 0+} F_\alpha(t) = \frac{1}{1+1/t}, \qquad F_{0.5}(t) = \frac{2}{\pi}\tan^{-1}\sqrt{t} \tag{13}$$

Moreover, for any $t \in [0,\ 1]$, $0 < \alpha_1 \leq \alpha_2 \leq 0.5$, we have

$$\frac{1}{1+1/t} \leq F_{\alpha_1}(t) \leq F_{\alpha_2}(t) \leq \frac{2}{\pi}\tan^{-1}\sqrt{t} \tag{14}$$

Proof: *See Appendix B. Figure 1 (left panel) plots $F_\alpha(t)$ for selected α values.* □

Lemma 2 proves that, when $\alpha \to 0+$, $F_\alpha(t)$ is of order t, and when $\alpha = 0.5$, $F_\alpha(t)$ is of order $\sqrt{t}$. Lemma 3 provides a general result that $F_\alpha(t) = \Theta\left(t^{1-\alpha}\right)$.

Lemma 3 *For $0 \leq t < \alpha^{\alpha/(1-\alpha)}$ and $0 < \alpha \leq 0.5$,*

$$F_\alpha(t) = \frac{t^{1-\alpha}}{C_\alpha + o(1)} \tag{15}$$

Proof: *See Appendix C.* □

Remarks for Lemma 3:

- The result restricts $t < \alpha^{\alpha/(1-\alpha)}$. Here $\alpha^{\alpha/(1-\alpha)}$ is monotonically decreasing in α and $0.5 \leq \alpha^{\alpha/(1-\alpha)} \leq 1$ for $\alpha \in (0, 0.5]$. Later we will show that our method indeed uses small t.
- The constant C_α can be numerically evaluated as shown in Figure 1 (right panel).
- At $\alpha \to 0+$, we have $F_{0+}(t) = \frac{1}{1+1/t} = t - t^2 + t^3 ....$ Hence $C_{0+} = 1$.
- At $\alpha = 0.5$, we have $F_{0.5}(t) = \frac{2}{\pi}\tan^{-1}\sqrt{t} = \frac{2}{\pi}\left(t^{1/2} - t^{3/2}/3 + ...\right)$. Hence $C_{0.5} = \pi/2$.

To close this section, the next Lemma shows that the maximum likelihood estimator using the ratio statistics is actually the *minimum estimator*.

Lemma 4 *Use the ratio statistics, y_j/s_{ij}, $j = 1$ to M. When $\alpha \in (0,\ 0.5]$, the maximum likelihood estimator (MLE) of x_i is the sample minimum*

$$\hat{x}_{i,min} = \min_{1\leq j\leq M} \frac{y_j}{s_{ij}} \tag{16}$$

Proof: *See Appendix D.* □

Lemma 4 largely explains our proposed algorithm. In the next section, we analyze the error probability of $\hat{x}_{i,min}$ and the sample complexity bound.

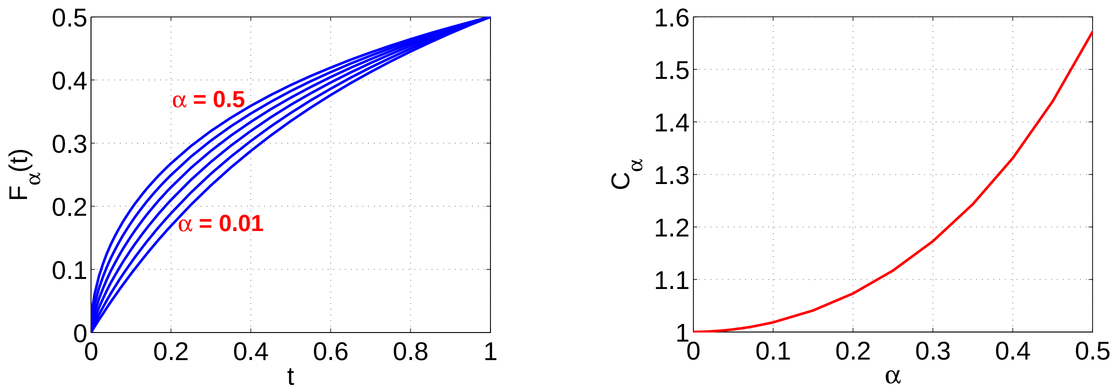


Figure 1: **Left panel**: $F_\alpha(t)$ for $t \in [0,\ 1]$, $\alpha = 0.01, 0.1, 0.2, 0.3, 0.4, 0.5$ (from bottom to top). **Right panel**: The constant C_α as in Lemma 3. Numerically, it varies between 1 and $\pi/2$.

3. Analysis of the Proposed Algorithm: Error Probability and Sample Complexity

Our proposed algorithm uses the *minimum estimator* $\hat{x}_{i,min} = \min_{1\leq j\leq M} \frac{y_j}{s_{ij}}$. The following Lemma concerns the tail probability of $\hat{x}_{i,min}$. Because $\hat{x}_{i,min}$ always over-estimates x_i, we only need to provide a one-sided error probability bound.

Lemma 5

$$\mathbf{Pr}\left(\hat{x}_{i,min} - x_i \geq \epsilon\right) = \left[1 - F_\alpha\left(\left(\epsilon/\theta_i\right)^{\alpha/(1-\alpha)}\right)\right]^M \tag{17}$$

$$\leq \left[\frac{1}{1 + \left(\epsilon/\theta_i\right)^{\alpha/(1-\alpha)}}\right]^M \tag{18}$$

For $0 < \alpha \leq 0.5$ and $\epsilon/\theta_i < \alpha$, we have

$$\mathbf{Pr}\left(\hat{x}_{i,min} - x_i \geq \epsilon\right) = \left[1 - \Theta\left(\epsilon^\alpha/\theta_i^\alpha\right)\right]^M \tag{19}$$

In particular, when $\alpha = 0.5$,

$$\mathbf{Pr}\left(\hat{x}_{i,min} - x_i \geq \epsilon, \alpha = 0.5\right) = \left[1 - \frac{2}{\pi}\tan^{-1}\sqrt{\frac{\epsilon}{\theta_i}}\right]^M \tag{20}$$

Proof: *Recall* $\frac{y_j}{s_{ij}} = x_i + \theta_i \frac{S_2}{S_1}$ *and* $\hat{x}_{i,min} = \min_{1\leq j\leq M} \frac{y_j}{s_{ij}}$. *We have*

$$\mathbf{Pr}\left(\hat{x}_{i,min} > x_i + \epsilon\right) = \mathbf{Pr}\left(\frac{y_j}{s_{ij}} > x_i + \epsilon,\ 1 \leq j \leq M\right)$$

$$= \left[\mathbf{Pr}\left(\frac{S_2}{S_1} > \frac{\epsilon}{\theta_i}\right)\right]^M = \left[1 - F_\alpha\left(\left(\epsilon/\theta_i\right)^{\alpha/(1-\alpha)}\right)\right]^M$$

The rest of the proof follows from Lemma 2 and Lemma 3. Recall we have defined in (7) that $\theta = \left(\sum_{i=1}^N x_i^\alpha\right)^{1/\alpha}$ *and* $\theta_i = \left(\theta^\alpha - x_i^\alpha\right)^{1/\alpha}$. □

Remark for Lemma 5: The probability bound (18) is convenient to use. However, it is conservative in that it does not give the right order unless α is small (i.e., when $\alpha/(1-\alpha) \approx \alpha$). In comparison, (19) provides the exact order, which will be useful for analyzing the precise sample complexity of our proposed algorithm. As shown in Lemma 3, $F_\alpha(t) = \Theta(t^{1-\alpha})$ holds for relatively small $t < \alpha^{\alpha/(1-\alpha)}$. In our case, $t = (\epsilon/\theta_i)^{\alpha/(1-\alpha)}$, i.e., the result requires $\epsilon/\theta_i < \alpha$, or $\epsilon^\alpha/\theta_i^\alpha = \epsilon^\alpha/(\sum_{l\neq i}^N x_l^\alpha) < \alpha^\alpha$. When $\alpha \to 0$, this means we need $1/K < 1$, which is virtually always true. For larger α, the relation $\epsilon^\alpha/(\sum_{l\neq i}^N x_l^\alpha) < \alpha^\alpha$ should hold in reasonable settings.

Theorem 6 *To ensure* $\sum_{i=1}^N \mathbf{Pr}\left(\hat{x}_{i,min} - x_i \geq \epsilon\right) \leq \delta$, *it suffices to choose* M *by*

$$M \geq \frac{\log N/\delta}{-\log\left[1 - F_\alpha\left((\epsilon/\theta)^{\alpha/(1-\alpha)}\right)\right]} \tag{21}$$

where F_α *is defined in Lemma 2. If* $\epsilon/\theta < 1$, *then it suffices to use*

$$M \geq \frac{\log N/\delta}{\log\left[1 + (\epsilon/\theta)^{\alpha/(1-\alpha)}\right]} \tag{22}$$

which is only sharp at $\alpha \to 0$. *In general, for* $\alpha \in (0,\ 0.5]$ *and* $\epsilon/\theta < \alpha$, *the sharp bound is*

$$M \geq (C_\alpha + o(1))\left(\frac{\theta}{\epsilon}\right)^\alpha \log N/\delta \tag{23}$$

where the constant C_α *is the same in Lemma 3. When* $\alpha = 0.5$ *and* $\epsilon/\theta < 1$, *the explicit bound is*

$$M \geq \frac{\pi}{2}\sqrt{\frac{\theta}{\epsilon}}\log N/\delta \tag{24}$$

Proof: *The result (21) follows from Lemma 5, (22) from Lemma 2, and (23) from Lemma 3.*

We provide more details for the proof of the more precise bound (24). When $\alpha = 0.5$,

$$M \geq \frac{\log N/\delta}{-\log\left[1 - \frac{2}{\pi}\tan^{-1}\sqrt{\frac{\epsilon}{\theta}}\right]}$$

which can be simplified to be $M \geq \frac{\pi}{2}\sqrt{\frac{\theta}{\epsilon}}\log N/\delta$, *using the fact that* $-\log\left(1 - \frac{2}{\pi}\tan^{-1}(z)\right) \geq \frac{2}{\pi}z, \forall z \in [0,\ 1]$. *To see this inequality, we can check*

$$\frac{\partial}{\partial z}\left(-\log(1 - \frac{2}{\pi}\tan^{-1}(z)) - \frac{2}{\pi}z\right) = \frac{\frac{2}{\pi}}{\left(1 - \frac{2}{\pi}\tan^{-1}z\right)(1+z^2)} - \frac{2}{\pi}$$

It suffices to show

$$z^2 - \frac{2}{\pi}\tan^{-1}z - \frac{2}{\pi}z^2\tan^{-1}z \leq 0$$

which is true because the equality holds when $z = 0$ *or* $z = 1$, *and*

$$\frac{\partial^2}{\partial z^2}\left(z^2 - \frac{2}{\pi}\tan^{-1}z - \frac{2}{\pi}z^2\tan^{-1}z\right) = 2 - \frac{2}{\pi}\left(2z\tan^{-1}z + \frac{2z}{1+z^2}\right) > 0$$

This completes the proof. □

Remarks for Theorem 6: The convenient bound (22) is only sharp for $\alpha \to 0$. For example, when $\alpha = 0.5$, $\alpha/(1-\alpha) = 1$, but the true order should be in terms of $\sqrt{\epsilon}$ instead of ϵ. The other bound (23) provides the precise order, where the constant C_α is the same as in Lemma 3. Note that, if we let $\alpha \to 0$, then $\left(\frac{\theta}{\epsilon}\right)^\alpha \to K$. In other words, the complexity for exact K-sparse recovery is essentially $K \log N/\delta$ with constant 1. This is a very sharp bound.

It is interesting to compare our complexity bound $C_\alpha \epsilon^{-\alpha} \sum_{i=1}^N x_i^\alpha \log N/\delta$ with the complexity bound of Count-Min sketch (Cormode and Muthukrishnan, 2005) $O\left(\epsilon^{-1} \sum_{i=1}^N x_i \log N/\delta\right)$. First of all, the bound we prove has the explicit constant C_α, which is just 1 when $\alpha \to 0$. Secondly, the order $\epsilon^{-\alpha}$ is an improvement over the order ϵ^{-1}. However, the interesting part is $\sum_{i=1}^N x_i^\alpha$ versus $\sum_{i=1}^N x_i$. If $x_i > 1$, then $x_i^\alpha < x_i$, and vice versa. This means whether $\sum_{i=1}^N x_i^\alpha$ is smaller than $\sum_{i=1}^N x_i$ will depend on the signal (and whether there are small components in the signal).

4. Experiments

Our proposed algorithm for sparse recovery is simple and requires merely one scan of the coordinates. Our theoretical analysis provides the sharp sample complexity bound with the constant (i.e., C_α) specified in Figure 1 (right panel). It is nevertheless still interesting to include an experimental study. All experiments presented were conducted in Matlab on a workstation with 256GB memory. We did not make special effort to optimize our code for improving efficiency.

We compare our proposed method with two popular packages: *L1Magic* (Candès and Romberg, 2005) and *SPGL1* (van den Berg and Friedlander, 2008)[1]. Although it is not our intension to compare these two solvers, we will present the results of both. While it is known that SPGL1 can often be faster than L1Magic, we observe that in some cases SPGL1 could not achieve the desired accuracy. On the other hand, SPGL1 better uses memory and can handle larger problems than L1Magic.

In each simulation, we randomly select K out of N coordinates and set their values (x_i) to be 1. The other $N-K$ coordinates are set to be 0. To simulate the design matrix $\mathbf{S}$, we generate two random matrices: $\{u_{ij}\}$ and $\{w_{ij}\}$, $i = 1$ to N, $j = 1$ to M, where $u_{ij} \sim unif(0,\pi)$ and $w_{ij} \sim exp(1)$, i.i.d. Then we apply the formula (3) to generate $s_{ij} \sim (\alpha, 1, 1)$, for $\alpha = 0.04$ to 0.5, spaced at 0.01. We also use the same u_{ij} and w_{ij} to generate standard Gaussian $N(0,1)$ variables for the design matrix used by L1Magic and SPGL1, based on the fact: $-\sqrt{2}\cos(u_{ij})\sqrt{w_{ij}} \sim N(0,1)$.

In this experimental setting, since $K = \sum_{i=1}^N x_i^\alpha$, the sample complexity of our algorithm is essentially $M = C_\alpha K/\epsilon^\alpha \log N/\delta$, where $C_{0+} = 1$ and $C_{0.5} = \pi/2 \approx 1.6$. In our simulations, we choose M by two options: (i) $M = K \log N/\delta$; (ii) $M = 1.6K \log N/\delta$, where $\delta = 0.01$.

1. We must specify some parameters in order to achieve good accuracies. For L1Magic, we use the following script:

```
l1eq_pd(x0, Afun, Atfun, y,1e-3,100,1e-8,1000);
```

For SPGL1, after consulting the author of (van den Berg and Friedlander, 2008), we used the following script:

```
opts = spgSetParms('verbosity',0);
opts.optTol = 1e-6;   opts.decTol = 1e-6;   spg_bp(A, y, opts);
```

However, it looks for $N = 10,000,000$ we probably should reduce the tolerance further.

We compare our method with L1Magic and SPGL1 in terms of their decoding times and recovery errors. The (normalized) recovery error is defined as

$$error = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \text{estimated } x_i)^2}{\sum_{i=1}^{N} x_i^2}} \tag{25}$$

4.1. $M = K \log N/\delta$

Figure 2 presents the recovery errors (left panel) and ratios of the decoding times (right panel), for $N = 1,000,000$, $K = 10$, and $M = K \log N/\delta$ (where $\delta = 0.01$). The results confirm that our proposed method is computationally efficient and produces accurate recovery for $\alpha < 0.38$.

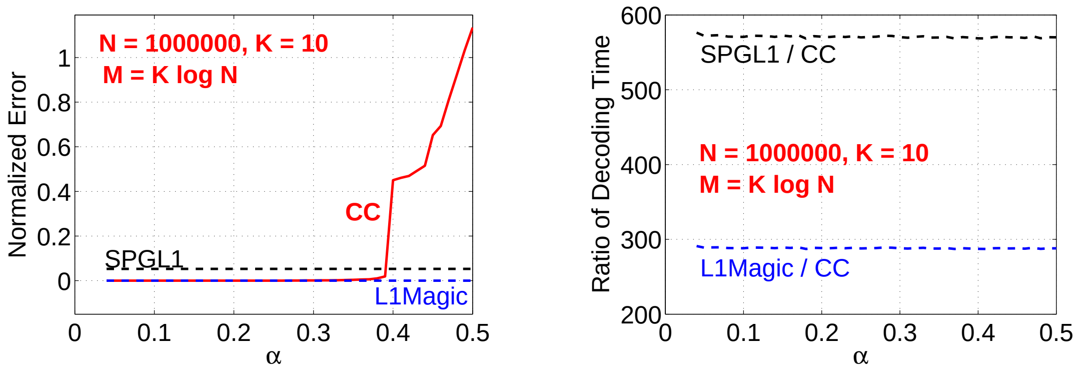


Figure 2: Experiments for comparing our proposed algorithm (labeled "CC") with *SPGL1* and *L1Magic*, for $N = 1,000,000$, $K = 10$, and $M = K \log N/\delta$ (where $\delta = 0.01$). For each α (from 0.04 to 0.5 spaced at 0.01), we conduct simulations 100 times and report the median results. In the **left panel**, our proposed method (solid curve) produces very accurate recovery results for $\alpha < 0.38$. For larger α values, however, the errors become large. This is expected because when $\alpha = 0.5$, the required number of samples should be $(\pi/2)K \log N/\delta$ instead of $K \log N/\delta$. In this case, L1Magic also produces accuracy recovery results. Note that for all methods, we report the top-K entries of the recovered signal as the estimated nonzero entries. In the **right panel**, we plot the ratios of the decoding times. SPGL1 uses about 580 times more time than our proposed method (which requires only one scan), and L1Magic needs about 290 times more time than ours.

Figure 3 presents the results for a larger problem, with $N = 10,000,000$ and $K = 10$. Because we can not run L1Magic in this case, we only present the comparisons with SPGL1. Again, our method is computationally very efficient and produces accurate recovery for about $\alpha < 0.38$.

For α close to 0.5, we need to increase the number of measurements, as shown in the analysis.

4.2. $M = 1.6K \log N/\delta$

To study the behavior as α approaches 0.5, we increase the number of measurements to $M = 1.6K \log N/\delta$. Figure 4 and Figure 5 present the experimental results for $N = 1,000,000$ and $N = 10,000,000$, respectively. Our algorithm still produces accurate recovery results (with the normalized errors around 0.007), although the results at smaller α values are even more accurate.

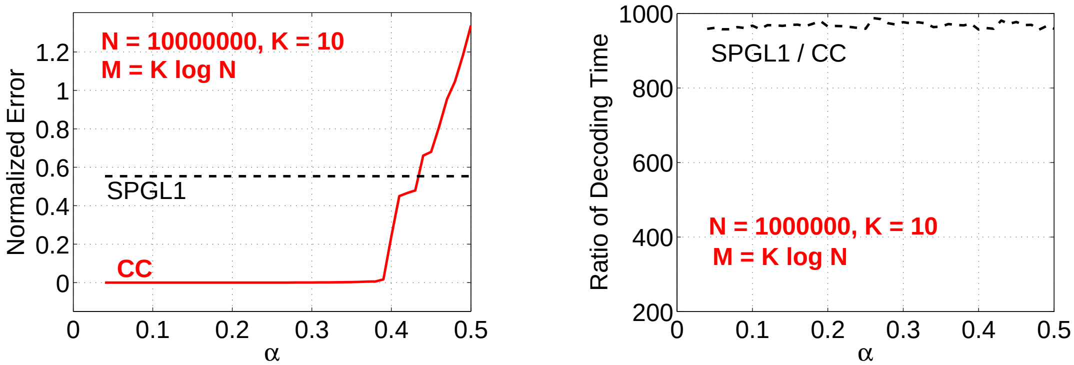


Figure 3: Experiments for comparing CC with SPGL1, for $N = 10,000,000$, $K = 10$, $M = K \log N/\delta$ (where $\delta = 0.01$). See the caption of Figure 2 for more details. For this larger problem, we can not run L1Magic as it simply halts without making progress.

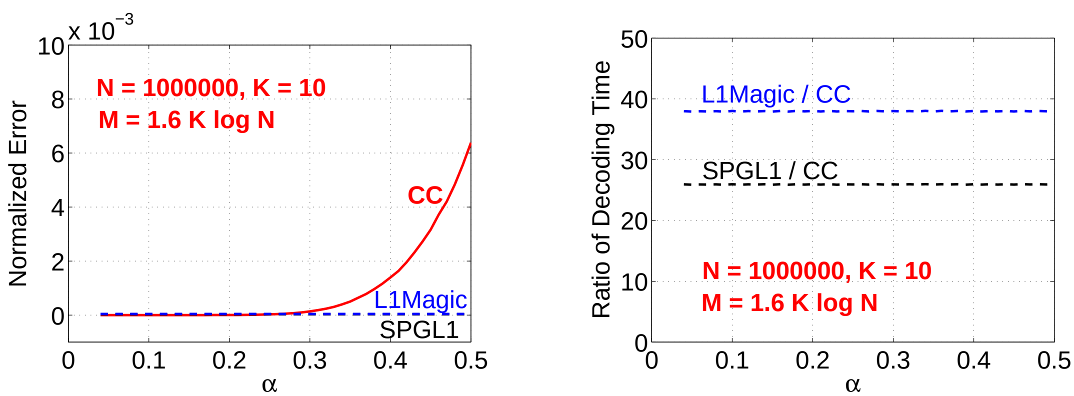
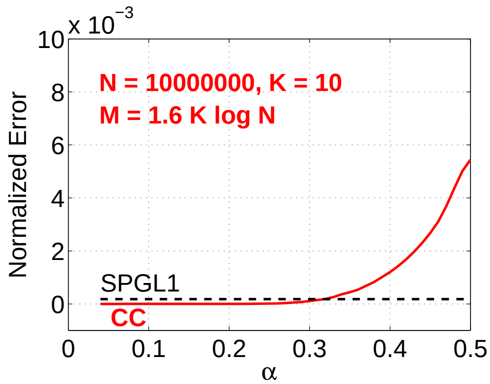


Figure 4: Experiments for comparing CC with SPGL1 and L1Magic, for $N = 1,000,000$, $K = 10$, $M = 1.6K \log N/\delta$ (where $\delta = 0.01$). In the **left panel**, our proposed method (solid curve) produces accurate recovery results, although the errors increase with increasing α (the maximum error is around 0.007). In the **right panel**, we can see that our method is, respectively, 27 times and 39 times faster than SPGL1 and L1Magic.

5. Discussions and Future Work

The preprint available at *arXiv:1310.1076* includes more detailed proofs and additional technical results: (i) the proposed algorithm is very robust against measurement noise; and (ii) the minimum estimator can be further improved by reducing the estimation bias.

While our proposed algorithm based on compressed counting is simple, fast, and accurate for the task of exact sparse recovery, it is clear that the work as presented is merely a tip of the iceberg. We expect many interesting research problems will arise along this line for future work.



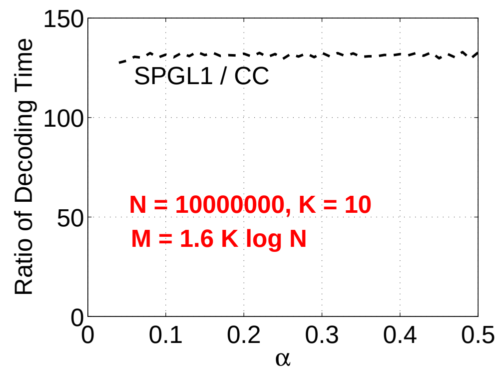


Figure 5: Experiments for comparing our proposed algorithm (CC) with SPGL1, for $N = 10,000,000$, $K = 10$, and $M = K \log N/\delta$ (where $\delta = 0.01$).

Choice of α. One important issue is the choice of α. In this paper, our analysis focuses on $\alpha \in (0,\ 0.5]$ and our theoretical results show that smaller α values lead to better performance. However, there are numerical issues which prevent us from using a too small α.

For the convenience of discussions, let us consider the approximate mechanism for generating $S(\alpha, 1, 1)$ from $U^{-1/\alpha}$, where $U \sim unif(0, 1)$ (based on the theory of domain of attractions (Feller, 1971)). If $\alpha = 0.04$, we need to compute U^{-25}, which may potentially create numerical problems. In our Matlab simulations in Section 4, we use $\alpha \in [0.04,\ 0.5]$ and we do not notice obvious numerical issues even with $\alpha = 0.04$. However, if a device (e.g., camera, cell phone, or inexpensive sensor) has limited precision and memory, then we might have to use a larger α. Fortunately, as shown in Section 4, the performance is not too sensitive to α. For example, in our experiments, the recovery accuracies are good for $\alpha < 0.38$ even when we choose $M = K \log N/\delta$ based on $\alpha \to 0$.

What about $\alpha > 0.5$? The same algorithm (i.e., the ratio statistic and the minimum estimator) can still be used but it will require more measurements. Lemma 4 has proved that when $\alpha \in (0, 0.5]$, the minimum estimator is the MLE. We no longer has such a proof when $\alpha > 0.5$ and most likely the estimator is not the MLE if $\alpha > 0.5$. In this study, for simplicity, we focus on $\alpha \in (0, 0.5]$.

Dense versus sparse design matrix. In this paper, we focus on dense design matrix. The prior work on "very sparse stable random projections" (Li, 2007) showed that one can significantly sparsify the design matrix without hurting the performance in estimating summary statistics. We have a separate tech report on *sparse recovery with very sparse compressed counting* (*arXiv:1401.0201*) (Li et al., 2013), which connects this line of work with the well-known "sparse matrix" algorithm (Gilbert and Indyk, 2010). In (Li et al., 2013), we proved a general worst-case bound $eK \log N/\delta$, which is an improvement over the result in this paper because the complexity bound does not contain the $\sum_{i=1}^N x_i^\alpha$ term (although the constant, i.e., e, might be larger).

Symmetric stable projections for symmetric signals. In a concurrent work (Li and Zhang, 2013), we developed an algorithm for using *symmetric α-stable projections* with very small α for exact sparse recovery. The algorithm used a *minimum* estimator for detection (of nonzero locations) and a *gap* estimator for estimation (of nonzero magnitudes), as well as an iterative procedure for further improving the estimates. For the task of exact sparse recovery, that method is extremely effective. However, it has a serious disadvantage in that the performance is sensitive to α (unlike the proposed method in this paper). In fact, the performance would drop substantially even if α is not too far away from $\alpha = 0$. More robust algorithms are needed to overcome this drawback.

Coding for stable random projections. Due to the use of heavy-tailed design matrix, our measurements are also heavy-tailed. This may raise concerns for storage and transmission, because it is necessary to use a large number of bits to store each measurement. Consequently, developing a good coding scheme for the measurements might be an urgent task, at least from the practical perspective. In the past years, we have worked on coding for random projections for the task of estimating summary statistics and approximate near neighbor search; examples include *Coding for Random Projections* (ICML 2014) and *Sign Cauchy Projections and Chi-square Kernels* (NIPS 2013).

Correlated or dependent stable random projections. So far we have only used independent α-stable random projections for sparse recovery. It appears to be an interesting idea to take advantage of stable projections simultaneously at multiple α values. Recall an α-stable random variable can be generated from an uniform random variable and an exponential random variable. Thus, we can generate multiple α-stable random variables for different α values from the same uniform and exponential variables. The hope is that such a *correlated* (or more appropriate, *dependent*) stable projection scheme might bring in unexpected improvements, just like our prior work on using correlated stable random projections to solve entropy estimation problem (Li and Zhang, 2012).

6. Conclusion

We develop a new compressed sensing algorithm for nonnegative (possibly streaming) signals, using *Compressed Counting (CC)* which is based on *maximally-skewed α-stable random projections*. Our method produces accurate recovery of nonnegative sparse signals and our procedure is computationally very efficient. The decoding cost is just one linear scan of the coordinates. Our theoretical analysis provides the sharp complexity bound and our experimental study confirms the theory.

Acknowledgement

The work of Ping Li is partially supported by ONR-N00014-13-1-0764, NSF-III-1360971, NSF-Bigdata-1419210, and AFOSR-FA9550-13-1-0137. The work of Cun-Hui Zhang is partially supported by ONR-N00014-13-1-0764, NSF-DMS-1106753, and NSF-DMS-1209014.

References

Emmanuel Candès and Justin Romberg. l_1-magic: Reocvery of sparse signals via convex programming. Technical report, Calinfornia Institute of Technology, 2005.

Emmanuel Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52 (2):489–509, 2006.

John M. Chambers, C. L. Mallows, and B. W. Stuck. A method for simulating stable random variables. *Journal of the American Statistical Association*, 71(354):340–344, 1976.

Scott Shaobing Chen, David L. Donoho, Michael, and A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20:33–61, 1998.

Graham Cormode and S. Muthukrishnan. An improved data stream summary: the count-min sketch and its applications. *Journal of Algorithm*, 55(1):58–75, 2005.

Daivd L. Donoho and Xiaoming Huo. Uncertainty principles and ideal atomic decomposition. *Information Theory, IEEE Transactions on*, 40(7):2845–2862, nov. 2001.

David L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

David L. Donoho and Philip B. Stark. Uncertainty principles and signal recovery. *SIAM Journal of Applied Mathematics*, 49(3):906–931, 1989.

David L. Donoho and Jared Tanner. Counting faces of randomly projected polytopes when the projection radically lowers dimension. *Journal of the American Mathematical Society*, 22(1), jan. 2009.

Laura Feinstein, Dan Schnackenberg, Ravindra Balupari, and Darrell Kindred. Statistical approaches to DDoS attack detection and response. In *DARPA Information Survivability Conference and Exposition*, pages 303–314, 2003.

William Feller. *An Introduction to Probability Theory and Its Applications (Volume II)*. John Wiley & Sons, New York, NY, second edition, 1971.

A. Gilbert and P. Indyk. Sparse recovery using sparse matrices. *Proc. of the IEEE*, 98(6):937 –947, june 2010.

Piotr Indyk. Stable distributions, pseudorandom generators, embeddings, and data stream computation. *Journal of ACM*, 53(3):307–323, 2006.

Ping Li. Very sparse stable random projections for dimension reduction in l_α $(0 < \alpha \leq 2)$ norm. In *KDD*, San Jose, CA, 2007.

Ping Li. Estimators and tail bounds for dimension reduction in l_α $(0 < \alpha \leq 2)$ using stable random projections. In *SODA*, pages 10 – 19, San Francisco, CA, 2008.

Ping Li. Compressed counting. In *SODA*, New York, NY, 2009a.

Ping Li. Improving compressed counting. In *UAI*, Montreal, CA, 2009b.

Ping Li and Cun-Hui Zhang. A new algorithm for compressed counting with applications in shannon entropy estimation in dynamic data. In *COLT*, 2011.

Ping Li and Cun-Hui Zhang. Entropy estimations using correlated symmetric stable random projections. In *NIPS*, Lake Tahoe, NV, 2012.

Ping Li and Cun-Hui Zhang. Exact sparse recovery with L0 projections. In *KDD*, pages 302–310, 2013.

Ping Li, Cun-Hui Zhang, and Tong Zhang. Sparse recovery with very sparse compressed counting. Technical report, 2013.

Tsung-Han Lin and H. T. Kung. Compressive sensing medium access control for wireless lans. In *Globecom*, 2012.

S.G. Mallat and Zhifeng Zhang. Matching pursuits with time-frequency dictionaries. *Signal Processing, IEEE Transactions on*, 41(12):3397 –3415, 1993.

S. Muthukrishnan. Data streams: Algorithms and applications. *Foundations and Trends in Theoretical Computer Science*, 1:117–236, 2 2005.

Noam Nisan. Pseudorandom generators for space-bounded computations. In *Proceedings of the twenty-second annual ACM symposium on Theory of computing*, STOC, pages 204–212, 1990.

Y.C. Pati, R. Rezaiifar, and P. S. Krishnaprasad. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*, pages 40–44 vol.1, Nov 1993.

J.A. Tropp. Greed is good: algorithmic results for sparse approximation. *Information Theory, IEEE Transactions on*, 50(10):2231 – 2242, oct. 2004.

Ewout van den Berg and Michael P. Friedlander. Probing the pareto frontier for basis pursuit solutions. *SIAM J. Sci. Comput.*, 31(2):890–912, 2008.

Jun Wang, Haitham Hassanieh, Dina Katabi, and Piotr Indyk. Efficient and reliable low-power backscatter networks. In *SIGCOMM*, pages 61–72, Helsinki, Finland, 2012a.

Meng Wang, Weiyu Xu, Enrique Mallada, and Ao Tang. Sparse recovery with graph constraints: Fundamental limits and measurement construction. In *Infomcom*, 2012b.

Tong Zhang. Sparse recovery with orthogonal matching pursuit under rip. *Information Theory, IEEE Transactions on*, 57(9):6215 –6221, sept. 2011.

Haiquan (Chuck) Zhao, Ashwin Lall, Mitsunori Ogihara, Oliver Spatscheck, Jia Wang, and Jun Xu. A data streaming algorithm for estimating entropies of od flows. In *IMC*, San Diego, CA, 2007.

Vladimir M. Zolotarev. *One-dimensional Stable Distributions*. American Mathematical Society, Providence, RI, 1986.

Appendix A. Proof of Lemma 1

For $S \sim S(\alpha, 1, 1)$, the sampling approach in (3) provides a method to compute its CDF

$$\begin{aligned}
F_S(s) =& \mathbf{Pr}\left(\frac{\sin(\alpha u)}{[\sin u \cos(\alpha\pi/2)]^{\frac{1}{\alpha}}}\left[\frac{\sin(u-\alpha u)}{w}\right]^{\frac{1-\alpha}{\alpha}} \leq s\right)\\
=& \mathbf{Pr}\left(\frac{[\sin(\alpha u)]^{\alpha/(1-\alpha)}}{[\sin u \cos(\alpha\pi/2)]^{\frac{1}{1-\alpha}}}\left[\frac{\sin(u-\alpha u)}{w}\right] \leq s^{\alpha/(1-\alpha)}\right)\\
=& \frac{1}{\pi}\int_0^\pi \exp\left\{-\frac{[\sin(\alpha u)]^{\alpha/(1-\alpha)}}{[\sin u \cos(\alpha\pi/2)]^{\frac{1}{1-\alpha}}}\left[\frac{\sin(u-\alpha u)}{s^{\alpha/(1-\alpha)}}\right]\right\} du\\
=& \frac{1}{\pi}\int_0^\pi \exp\left\{-q_\alpha(u) s^{-\alpha/(1-\alpha)}\right\} du
\end{aligned}$$

and the PDF

$$f_S(s) = \frac{1}{\pi}\int_0^\pi \exp\left\{-q_\alpha(u) s^{-\alpha/(1-\alpha)}\right\} q_\alpha(u)\alpha/(1-\alpha) s^{-\alpha/(1-\alpha)-1} du$$

Hence,

$$\begin{aligned}
&\frac{1}{\theta_i} f_S\left(\frac{y_j - x_i s_{ij}}{\theta_i}\right)\\
=& \frac{\alpha/(1-\alpha)}{\pi}\int_0^\pi q_\alpha(u) \exp\left\{-q_\alpha(u)\left(\frac{\theta_i}{y_j - x_i s_{ij}}\right)^{\alpha/(1-\alpha)}\right\}\left(\frac{\theta_i}{y_j - x_i s_{ij}}\right)^{\alpha/(1-\alpha)} \frac{1}{(y_j - x_i s_{ij})} du
\end{aligned}$$

Therefore, the likelihood $L(x_i, \theta_i) \to +\infty$ if $y_j - x_i s_{ij} \to 0$, provided $\theta_i/(y_j - x_i s_{ij}) \to const$. Note that here we can choose θ_i and x_i to maximize the likelihood.

Appendix B. Proof of Lemma 2

Since $S_1, S_2 \sim S(\alpha, 1, 1)$, i.i.d., we know that

$$\begin{aligned}
S_1 =& \frac{\sin(\alpha u_1)}{[\sin u_1 \cos(\alpha\pi/2)]^{\frac{1}{\alpha}}}\left[\frac{\sin(u_1 - \alpha u_1)}{w_1}\right]^{\frac{1-\alpha}{\alpha}},\\
S_2 =& \frac{\sin(\alpha u_2)}{[\sin u_2 \cos(\alpha\pi/2)]^{\frac{1}{\alpha}}}\left[\frac{\sin(u_2 - \alpha u)}{w_2}\right]^{\frac{1-\alpha}{\alpha}}
\end{aligned}$$

where $u_1, u_2 \sim uniform(0, \pi)$, $w_1, w_2 \sim \exp(1)$, u_1, u_2, w_1, w_2 are independent. Thus, we can write

$$\begin{aligned}
&(S_2/S_1)^{\alpha/(1-\alpha)} = Q_\alpha \frac{w_1}{w_2},\\
&Q_\alpha = \left[\frac{\sin(\alpha u_2)}{\sin(\alpha u_1)}\right]^{\alpha/(1-\alpha)}\left[\frac{\sin u_1}{\sin u_2}\right]^{\frac{1}{1-\alpha}} \frac{\sin(u_2 - \alpha u_2)}{\sin(u_1 - \alpha u_1)}
\end{aligned}$$

Using properties of exponential distributions, for any $t \geq 0$,

$$\begin{aligned}F_\alpha(t) =&\mathbf{Pr}\left((S_2/S_1)^{\alpha/(1-\alpha)} \leq t\right) = \mathbf{Pr}\left(Q_\alpha w_1/w_2 \leq t\right)\\ =&E\left(\frac{1}{1+Q_\alpha/t}\right) = \frac{1}{\pi^2}\int_0^\pi\int_0^\pi \frac{1}{1+Q_\alpha/t}du_1du_2\end{aligned}$$

When $\alpha \to 0+$, $Q_\alpha \to 1$ point-wise. By dominated convergence, $F_{0+}(t) = \frac{1}{1+1/t}$.
When $\alpha = 0.5$, Q_α can be simplified to be

$$Q_{0.5} = \left[\frac{\sin(u_2/2)}{\sin(u_1/2)}\right]\left[\frac{\sin u_1}{\sin u_2}\right]^2 \frac{\sin(u_2/2)}{\sin(u_1/2)} = \frac{\cos^2(u_1/2)}{\cos^2(u_2/2)}$$

which can be used to obtain the closed-form expression for $F_{0.5}(t)$:

$$\begin{aligned}F_{0.5}(t) =&\frac{1}{\pi^2}\int_0^\pi\int_0^\pi \frac{1}{1+Q_{0.5}/t}du_1du_2\\ =&\frac{1}{\pi^2}\int_0^\pi\int_0^\pi \frac{1}{1+\frac{\cos^2(u_1/2)}{t\cos^2(u_2/2)}}du_1du_2\\ =&\frac{4}{\pi^2}\int_0^{\pi/2}\int_0^{\pi/2}\frac{1}{1+b\cos^2(u_1)}du_1du_2, \qquad b = \frac{1}{t\cos^2(u_2)}\\ =&\frac{4}{\pi^2}\int_0^{\pi/2}\frac{-1}{\sqrt{1+b}}\tan^{-1}\left(\sqrt{1+b}\frac{\cos u_1}{\sin u_1}\right)\Big|_0^{\pi/2} du_2\\ =&\frac{2}{\pi}\int_0^{\pi/2}\frac{1}{\sqrt{1+\frac{1}{t}\sec^2 u_2}}du_2\\ =&\frac{2}{\pi}\int_0^1 \frac{1}{\sqrt{1+\frac{1}{t}-z^2}}dz = \frac{2}{\pi}\int_0^{1/\sqrt{1+1/t}}\frac{1}{\sqrt{1-z^2}}dz\\ =&\frac{2}{\pi}\sin^{-1}\left(1/\sqrt{1+1/t}\right) = \frac{2}{\pi}\tan^{-1}\sqrt{t}\end{aligned}$$

To show $F_\alpha(t) \geq 1/(1+1/t)$ for any $t \in [0,\ 1]$, we first note that the equality holds when $t = 0$ and $t = 1$. To see the latter case, we write $Q_\alpha = q_2/q_1$, where q_1 and q_2 are i.i.d. When $t = 1$, $F_\alpha(t) = E\left(1/(1+q_2/q_1)\right) = E\left(\frac{q_1}{q_1+q_2}\right) = \frac{1}{2}$ by symmetry.

It remains to show $F_\alpha(t)$ is monotonically increasing in α for fixed $t \in [0,\ 1]$. For convenience, we define $q_\alpha(u)$ and $g_\alpha(u)$, where

$$Q_\alpha = q_\alpha(u_2)/q_\alpha(u_1), \qquad q_\alpha(u) = \left[\sin(\alpha u)\right]^{\alpha/(1-\alpha)}\left[\sin u\right]^{\frac{-1}{1-\alpha}}\sin(u-\alpha u)$$

$$g_\alpha(u) = \frac{\partial \log q_\alpha(u)}{\partial\alpha} = \frac{\cos\alpha u}{\sin\alpha u}\frac{\alpha u}{1-\alpha} + \frac{1}{(1-\alpha)^2}\log\sin\alpha u - \frac{1}{(1-\alpha)^2}\log\sin u - u\frac{\cos(u-\alpha u)}{\sin(u-\alpha u)}$$

We can check that both $q_\alpha(u)$ and $g_\alpha(u)$ are monotonically increasing in $u \in [0, \pi]$.

$$\begin{aligned}\frac{\partial g_\alpha(u)}{\partial u} =&\frac{-\alpha}{\sin^2\alpha u}\frac{\alpha u}{1-\alpha}+\frac{\cos\alpha u}{\sin\alpha u}\frac{\alpha}{1-\alpha}+\frac{\alpha}{(1-\alpha)^2}\frac{\cos\alpha u}{\sin\alpha u}-\frac{1}{(1-\alpha)^2}\frac{\cos u}{\sin u}\\ &-\frac{\cos(u-\alpha u)}{\sin(u-\alpha u)}+\frac{(1-\alpha)u}{\sin^2(u-\alpha u)}\\ =&\left\{\frac{(1-\alpha)u}{\sin^2(u-\alpha u)}-\frac{\alpha}{\sin^2\alpha u}\frac{\alpha u}{1-\alpha}\right\}+\left\{\frac{\cos\alpha u}{\sin\alpha u}\frac{\alpha}{1-\alpha}-\frac{\cos(u-\alpha u)}{\sin(u-\alpha u)}\right\}\\ &+\left\{\frac{\alpha}{(1-\alpha)^2}\frac{\cos\alpha u}{\sin\alpha u}-\frac{1}{(1-\alpha)^2}\frac{\cos u}{\sin u}\right\}\end{aligned}$$

We consider three terms (in curly brackets) separately and show they are all ≥ 0 when $\alpha \in [0,\ 0.5]$.

For the first term,

$$\begin{aligned}&\frac{(1-\alpha)u}{\sin^2(u-\alpha u)}-\frac{\alpha}{\sin^2\alpha u}\frac{\alpha u}{1-\alpha}\geq 0\\ \Longleftrightarrow&\frac{1-\alpha}{\sin((1-\alpha)u)}\geq\frac{\alpha}{\sin\alpha u}\\ \Longleftrightarrow&(1-\alpha)\sin\alpha u-\alpha\sin((1-\alpha)u)\geq 0\end{aligned}$$

where the last inequality holds because the derivative (w.r.t. u) is $(1-\alpha)\alpha\cos\alpha u-(1-\alpha)\alpha\cos((1-\alpha)u)\geq 0$. For the second term, it suffices to show

$$\begin{aligned}&\frac{\partial}{\partial u}\left\{\alpha\cos\alpha u\sin(u-\alpha u)-(1-\alpha)\sin\alpha u\cos(u-\alpha u)\right\}\geq 0\\ \Longleftrightarrow&-\alpha^2\sin\alpha u\sin(u-\alpha u)+(1-\alpha)^2\sin\alpha u\sin(u-\alpha u)\geq 0\end{aligned}$$

For the third term, it suffices to show

$$\alpha\sin u\cos\alpha u-\cos u\sin\alpha u\geq 0\Longleftrightarrow\alpha\sin(u-\alpha u)+(1-\alpha)\cos u\sin\alpha u\geq 0$$

Thus, we have proved the monotonicity of $g_\alpha(u)$ in $u\in[0,\ \pi]$, when $\alpha\in[0,\ 0.5]$.

To prove the monotonicity of $q_\alpha(u)$ in u, it suffices to check if its logarithm is monotonic, i.e.

$$\frac{\partial}{\partial u}\log q_a\alpha(u)=\frac{1}{1-\alpha}\left(\alpha^2\frac{\cos\alpha u}{\sin\alpha u}+(1-\alpha)^2\frac{\cos(u-\alpha u)}{\sin(u-\alpha u)}-\frac{\cos u}{\sin u}\right)\geq 0$$

for which it suffices to show

$$\begin{aligned}&\alpha^2\cos\alpha u\sin(u-\alpha u)\sin u+(1-\alpha)^2\cos(u-\alpha u)\sin\alpha u\sin u-\cos u\sin\alpha u\sin(u-\alpha u)\geq 0\\ \Longleftrightarrow&\alpha^2\sin^2(u-\alpha u)+(1-\alpha)^2\sin^2\alpha u-2\alpha(1-\alpha)\cos u\sin\alpha u\sin(u-\alpha u)\geq 0\\ \Longleftrightarrow&(\alpha\sin(u-\alpha u)-(1-\alpha)\sin\alpha u)^2+2\alpha(1-\alpha)(1-\cos u)\sin\alpha u\sin(u-\alpha u)\geq 0\end{aligned}$$

At this point, we have proved that both $q_\alpha(u)$ and $g_\alpha(u)$ are monotonically increasing in $u\in[0,\ \pi]$ at least for $\alpha\in[0,\ 0.5]$.

$$\frac{\partial F_\alpha(t)}{\partial\alpha}=E\left(\frac{-\frac{1}{t}\frac{g_\alpha(u_2)q_\alpha(u_2)q_\alpha(u_1)-g_\alpha(u_1)q_\alpha(u_1)q_\alpha(u_2)}{q_\alpha^2(u_1)}}{\left(1+\frac{q_\alpha(u_2)}{tq_\alpha(u_1)}\right)^2}\right)=\frac{1}{t}E\left(\frac{q_\alpha(u_1)q_\alpha(u_2)\left(g_\alpha(u_1)-g_\alpha(u_2)\right)}{\left(q_\alpha(u_1)+q_\alpha(u_2)/t\right)^2}\right)$$

By symmetry

$$\frac{\partial F_\alpha(t)}{\partial \alpha} = \frac{1}{t} E\left(\frac{q_\alpha(u_1) q_\alpha(u_2) \left(g_\alpha(u_2) - g_\alpha(u_1)\right)}{\left(q_\alpha(u_2) + q_\alpha(u_1)/t\right)^2} \right)$$

Thus, to show $\frac{\partial F_\alpha(t)}{\partial \alpha} \geq 0$, it suffices to show

$$E\left(\frac{q_\alpha(u_1) q_\alpha(u_2) \left(g_\alpha(u_1) - g_\alpha(u_2)\right)}{\left(q_\alpha(u_1) + q_\alpha(u_2)/t\right)^2} \right) + E\left(\frac{q_\alpha(u_1) q_\alpha(u_2) \left(g_\alpha(u_2) - g_\alpha(u_1)\right)}{\left(q_\alpha(u_2) + q_\alpha(u_1)/t\right)^2} \right) \geq 0$$
$$\Longleftrightarrow E\left(\frac{q_\alpha(u_1) q_\alpha(u_2) \left(g_\alpha(u_1) - g_\alpha(u_2)\right) \left(q_\alpha^2(u_1) - q_\alpha^2(u_2)\right) \left(1/t^2 - 1\right)}{\left(q_\alpha(u_1) + q_\alpha(u_2)/t\right)^2 \left(q_\alpha(u_2) + q_\alpha(u_1)/t\right)^2} \right) \geq 0$$

which holds because $1/t^2 - 1 \geq 0$ and $\left(g_\alpha(u_1) - g_\alpha(u_2)\right) \left(q_\alpha(u_1) - q_\alpha(u_2)\right) \geq 0$ as both $g_\alpha(u)$ and $q_\alpha(u)$ are monotonically increasing functions of $u \in [0,\ \pi]$. This completes the proof.

Appendix C. Proof of Lemma 3

The goal is to show that $F_\alpha(t) = \Theta\left(t^{1-\alpha}\right)$. By our definition,

$$F_\alpha(t) = E\left(\frac{1}{1 + Q_\alpha/t}\right) = E\left(\frac{1}{1 + \frac{1}{t}\frac{q_\alpha(u_2)}{q_\alpha(u_1)}}\right)$$

where

$$q_\alpha(u) = \left[\sin\left(\alpha u\right)\right]^{\alpha/(1-\alpha)} \left[\frac{1}{\sin u}\right]^{\frac{1}{1-\alpha}} \sin\left(u - \alpha u\right)$$

We can write the integral as

$$F_\alpha(t) = E\left(\frac{1}{1 + \frac{1}{t}\frac{q_\alpha(u_2)}{q_\alpha(u_1)}}\right)$$
$$= \frac{1}{\pi^2}\int_0^{\pi/2}\int_0^{\pi/2} \frac{1}{1 + t^{-1} q_\alpha(u_2)/q_\alpha(u_1)} du_1 du_2 + \frac{1}{\pi^2}\int_0^{\pi/2}\int_0^{\pi/2} \frac{1}{1 + t^{-1} q'_\alpha(u_2)/q_\alpha(u_1)} du_1 du_2$$
$$+ \frac{1}{\pi^2}\int_0^{\pi/2}\int_0^{\pi/2} \frac{1}{1 + t^{-1} q_\alpha(u_2)/q'_\alpha(u_1)} du_1 du_2 + \frac{1}{\pi^2}\int_0^{\pi/2}\int_0^{\pi/2} \frac{1}{1 + t^{-1} q'_\alpha(u_2)/q'_\alpha(u_1)} du_1 du_2$$

where

$$q'_\alpha(u) = \left[\sin\left(\alpha(\pi - u)\right)\right]^{\alpha/(1-\alpha)} \left[\frac{1}{\sin(\pi - u)}\right]^{\frac{1}{1-\alpha}} \sin\left(\pi - u - \alpha(\pi - u)\right)$$
$$= \left[\sin\left(\alpha(\pi - u)\right)\right]^{\alpha/(1-\alpha)} \left[\frac{1}{\sin u}\right]^{\frac{1}{1-\alpha}} \sin\left(u + \alpha(\pi - u)\right)$$

First, using the fact that $\alpha \sin u \leq \sin(\alpha u) \leq \alpha u$, we obtain

$$q_\alpha(u) \geq \left[\alpha \sin\left(u\right)\right]^{\alpha/(1-\alpha)} \left[\frac{1}{\sin u}\right]^{\frac{1}{1-\alpha}} (1-\alpha)\sin\left(u\right) = \alpha^{\alpha/(1-\alpha)}(1-\alpha)$$

We have proved in the proof of Lemma 2 that $q_\alpha(u)$ is a monotonically increasing function of $u \in [0,\ \pi]$. Since $q_\alpha(\pi/2) = [\sin(\alpha\pi/2)]^{\alpha/(1-\alpha)}\cos(\alpha\pi/2)$, we have

$$1/4 \leq \alpha^{\alpha/(1-\alpha)}(1-\alpha) \leq q_\alpha(u) \leq [\sin(\alpha\pi/2)]^{\alpha/(1-\alpha)}\cos(\alpha\pi/2) \leq 1, \qquad u \in [0,\ \pi/2]$$

In other words, we can view $q_\alpha(u)$ as a constant (i.e., $q_\alpha(u) \asymp 1$) when $u \in [0,\ \pi/2]$.

On the other hand, note that $q'_\alpha(u) \to \infty$ as $u \to 0$. Moreover, when $u \in [0,\ \pi/2]$, we have $\alpha u \leq \pi - u$ and $u - \alpha u \leq u + \alpha(\pi - u)$. Thus, $q'_\alpha(u)$ dominates $q_\alpha(u)$. Therefore, the order of $F_\alpha(t)$ is determined by one term:

$$F_\alpha(t) \asymp \int_0^{\pi/2}\int_0^{\pi/2} \frac{1}{1 + t^{-1}q_\alpha(u_2)/q'_\alpha(u_1)} du_1 du_2 \asymp \int_0^{\pi/2} \frac{1}{1 + t^{-1}/q'_\alpha(u)} du$$

Since

$$q'_\alpha(u) \asymp \frac{\alpha^{\alpha/(1-\alpha)}\max\{u,\alpha\}}{u^{1/(1-\alpha)}} \asymp \max\left\{u^{-\alpha/(1-\alpha)}, \alpha u^{-1/(1-\alpha)}\right\}$$

we have, for $\alpha \in [0,\ 1/2]$,

$$\begin{aligned} F_\alpha(t) &\asymp \int_0^{\alpha} \frac{1}{1 + t^{-1}/q'_\alpha(u)} du + \int_\alpha^{\pi/2} \frac{1}{1 + t^{-1}/q'_\alpha(u)} du \\ &\asymp \int_0^{\alpha} \frac{1}{1 + (\alpha t)^{-1}u^{1/(1-\alpha)}} du + \int_\alpha^{\pi/2} \frac{1}{1 + t^{-1}u^{\alpha/(1-\alpha)}} du \end{aligned}$$

Consider $t < \alpha^{\alpha/(1-\alpha)}$. Because $t^{-1}u^{\alpha/(1-\alpha)} > (u/\alpha)^{\alpha/(1-\alpha)} \geq 1$ for $u \geq \alpha$, we have

$$\int_\alpha^{\pi/2} \frac{1}{1 + t^{-1}u^{\alpha/(1-\alpha)}} du \asymp \int_\alpha^{\pi/2} \frac{1}{t^{-1}u^{\alpha/(1-\alpha)}} du = t\ \frac{1-\alpha}{1-2\alpha} u^{(1-2\alpha)/(1-\alpha)}\Big|_\alpha^{\pi/2} \asymp t$$

uniformly for $\alpha < 1/2$. When $\alpha = 1/2$ (i.e., $t < 1/2$), we also have

$$\int_\alpha^{\pi/2} \frac{1}{1 + t^{-1}u^{\alpha/(1-\alpha)}} du = \int_{1/2}^{\pi/2} \frac{1}{1 + t^{-1}u} du = t\ \log(u+t)|_{1/2}^{\pi/2} \asymp t$$

For the other term with $u \in [0, \alpha]$, we have

$$\begin{aligned} \int_0^{\alpha} \frac{1}{1 + (\alpha t)^{-1}u^{1/(1-\alpha)}} du &= \int_0^{(\alpha t)^{1-\alpha}} \frac{1}{1 + (\alpha t)^{-1}u^{1/(1-\alpha)}} du + \int_{(\alpha t)^{1-\alpha}}^{\alpha} \frac{1}{1 + (\alpha t)^{-1}u^{1/(1-\alpha)}} du \\ &= \int_0^{(\alpha t)^{1-\alpha}} \frac{1}{1 + (\alpha t)^{-1}u^{1/(1-\alpha)}} du + \int_{(\alpha t)^{1-\alpha}}^{\alpha} \frac{1}{1 + (\alpha t)^{-1}u^{1/(1-\alpha)}} du \\ &\asymp (\alpha t)^{1-\alpha} - (\alpha t)\frac{1-\alpha}{\alpha}\ u^{(-\alpha)/(1-\alpha)}\Big|_{(\alpha t)^{1/(1-\alpha)}}^{\alpha} \\ &= (\alpha t)^{1-\alpha} - t(1-\alpha)\alpha^{(-\alpha)/(1-\alpha)} + t(1-\alpha)(\alpha t)^{-\alpha} \\ &= t^{1-\alpha}\alpha^{-\alpha} - t(1-\alpha)\alpha^{(-\alpha)/(1-\alpha)} \end{aligned}$$

Combining the results, we obtain

$$F_\alpha(t) \asymp t\left(1 - \alpha^{(-\alpha)/(1-\alpha)} + \alpha^{(1-2\alpha)/(1-\alpha)}\right) + t^{1-\alpha}\alpha^{-\alpha} \asymp t^{1-\alpha}$$

This completes the proof.

Appendix D. Proof of Lemma 4

Define $F_Z(t) = \mathbf{Pr}\left(\frac{y_j}{s_{ij}} \leq t\right)$ and $f_Z(t) = F_Z'(t)$. To find the MLE of x_i, we need to maximize $\prod_{j=1}^{M} f_Z(z_{i,j})$. Using the result in Lemma 2, for $S_1, S_2 \sim S(\alpha, 1, 1)$, we have

$$F_Z(t) = \mathbf{Pr}\left(\frac{y_j}{s_{ij}} \leq t\right) = \mathbf{Pr}\left(S_2/S_1 \leq \frac{t-x_i}{\theta_i}\right) = E\left(\frac{1}{1+\left(\frac{\theta_i}{t-x_i}\right)^{\alpha/(1-\alpha)} Q_\alpha}\right)$$

$$f_Z(t) = E\left(\frac{\theta_i^{\alpha/(1-\alpha)} Q_\alpha \alpha/(1-\alpha)(t-x_i)^{-1/(1-\alpha)}}{\left(1+\left(\frac{\theta_i}{t-x_i}\right)^{\alpha/(1-\alpha)} Q_\alpha\right)^2}\right)$$

$$f_Z'(t) = E\left(\frac{A}{\left(1+\left(\frac{\theta_i}{t-x_i}\right)^{\alpha/(1-\alpha)} Q_\alpha\right)^4}\right)$$

where Q_α is defined in Lemma 2 and

$$\begin{aligned}
A =& \theta_i^{\alpha/(1-\alpha)} Q_\alpha \alpha/(1-\alpha)(-1/(1-\alpha))(t-x_i)^{-1/(1-\alpha)-1}\left(1+\left(\frac{\theta_i}{t-x_i}\right)^{\alpha/(1-\alpha)} Q_\alpha\right)^2 \\
&+ 2\left(1+\left(\frac{\theta_i}{t-x_i}\right)^{\alpha/(1-\alpha)} Q_\alpha\right)\left(\theta_i^{\alpha/(1-\alpha)} Q_\alpha \alpha/(1-\alpha)(t-x_i)^{-1/(1-\alpha)}\right)^2 \\
=& \left(1+\left(\frac{\theta_i}{t-x_i}\right)^{\alpha/(1-\alpha)} Q_\alpha\right)\theta_i^{\alpha/(1-\alpha)} Q_\alpha \alpha/(1-\alpha)^2 (t-x_i)^{-1/(1-\alpha)-1} \\
&\times\left(-\left(1+\left(\frac{\theta_i}{t-x_i}\right)^{\alpha/(1-\alpha)} Q_\alpha\right) + 2\theta_i^{\alpha/(1-\alpha)} Q_\alpha \alpha (t-x_i)^{-\alpha/(1-\alpha)}\right) \\
=& \left(1+\left(\frac{\theta_i}{t-x_i}\right)^{\alpha/(1-\alpha)} Q_\alpha\right)\theta_i^{\alpha/(1-\alpha)} Q_\alpha \alpha/(1-\alpha)^2 (t-x_i)^{-1/(1-\alpha)-1} \\
&\times\left(-1-\left(\frac{\theta_i}{t-x_i}\right)^{\alpha/(1-\alpha)}(1-2\alpha)\right)
\end{aligned}$$

$A \leq 0$ if $\alpha \leq 0.5$. This means, $f_Z(t) \to \infty$ when $t \to x_i$ and $f_Z(t)$ is nondecreasing in $t \geq x_i$ if $\alpha \leq 0.5$. Therefore, given M observations, $z_{i,j} = y_j/s_{ij}$, the MLE is the sample minimum. This completes the proof.