# On Model-Based RIP-1 Matrices

Piotr Indyk[*]        Ilya Razenshteyn[†]

**Abstract**

The Restricted Isometry Property (RIP) is a fundamental property of a matrix enabling sparse recovery [CRT06]. Informally, an $m \times n$ matrix satisfies RIP of order $k$ in the $\ell_p$ norm if $\|Ax\|_p \approx \|x\|_p$ for any vector $x$ that is $k$-sparse, i.e., that has at most $k$ non-zeros. The minimal number of rows $m$ necessary for the property to hold has been extensively investigated, and tight bounds are known. Motivated by signal processing models, a recent work of Baraniuk et al [BCDH10] has generalized this notion to the case where the support of $x$ must belong to a given *model*, i.e., a given family of supports. This more general notion is much less understood, especially for norms other than $\ell_2$.

In this paper we present tight bounds for the model-based RIP property in the $\ell_1$ norm. Our bounds hold for the two most frequently investigated models: tree-sparsity and block-sparsity. We also show implications of our results to sparse recovery problems.

## 1   Introduction

In recent years, a new "linear" approach for obtaining a succinct approximate representation of $n$-dimensional vectors (or signals) has been discovered. For any signal $x$, the representation is equal to $Ax$, where $A$ is an $m \times n$ matrix, or possibly a random variable chosen from some distribution over such matrices. The vector $Ax$ is often referred to as the *measurement vector* or *linear sketch* of $x$. Although $m$ is typically much smaller than $n$, the sketch $Ax$ often contains plenty of useful information about the signal $x$.

A particularly useful and well-studied problem is that of *stable sparse recovery*. We say that a vector $x'$ is $k$-sparse if it has at most $k$ non-zero coordinates. The sparse recovery problem is typically defined as follows: for some norm parameters $p$ and $q$ and an approximation factor $C > 0$, given $Ax$, recover an "approximation" vector $x^*$ such that

$$\|x - x^*\|_p \le C \min_{k\text{-sparse } x'} \|x - x'\|_q \tag{1}$$

(this inequality is often referred to as $\ell_p/\ell_q$ *guarantee*). Sparse recovery has a tremendous number of applications in areas such as compressive sensing of signals [CRT06, Don06], genetic data acquisition and analysis and data stream algorithms [Mut05, GI10].

It is known [CRT06] that there exist matrices $A$ and associated recovery algorithms that produce approximations $x^*$ satisfying Equation (1) with $p = q = 1$[1] constant approximation factor $C$, and sketch length

$$m = O(k \log(n/k)) \tag{2}$$

---

[*]MIT CSAIL, `indyk@mit.edu`

[†]MIT CSAIL, `ilyaraz@mit.edu`

[1]In fact, one can prove a somewhat stronger guarantee, referred to as the $\ell_2/\ell_1$ guarantee.

This result was proved by showing that there exist matrices $A$ with $m = O(k \log(n/k))$ rows that satisfy the *Restricted Isometry Property (RIP)*. Formally, we say that $A$ is *a $(k, \varepsilon)$-RIP-p matrix*, if for every $x \in \mathbb{R}^n$ with at most $k$ non-zero coordinates we have

$$(1 - \varepsilon)\|x\|_p \le \|Ax\|_p \le (1 + \varepsilon)\|x\|_p.$$

The proof of [CRT06] proceeds by showing that (i) there exist matrices with $m = O(k \log(n/k))$ rows that satisfy $(k, \varepsilon)$-RIP-2 for some constant $\varepsilon > 0$ and (ii) for such matrices there exist a polynomial time recovery algorithm that given $Ax$ produces $x^*$ satisfying Equation 1. Similar results were obtained for RIP-1 matrices [BGI$^+$08]. The latter matrices are closely connected to hashing-based streaming algorithms for heavy-hitter problems, see [GI10] for an overview.

It is known that the bound on the number of measurements in Equation (2) is asymptotically optimal for some constant $C$ and $p = q = 1$, see [BIPW10] and [FPRU10] (building on [Don06, GG84, Glu84, Kas77]). The necessity of the "extra" logarithmic factor multiplying $k$ is quite unfortunate: the sketch length determines the "compression rate", and for large $n$ any logarithmic factor can worsen that rate tenfold. Fortunately, a more careful modeling offers a way to overcome the aforementioned limitation. In particular, after decades of research in signal modeling, signal processing researchers know that not all supports (i.e., sets of non-zero coordinates) are equally common . For example, if a signal is a function of time, large coefficients of the signal tend to occur consecutively. This phenomenon can be exploited by searching for the best $k$-sparse approximation $x^*$ whose support belongs to a given "model" family of supports $\mathcal{M}_k$ (i.e., $x^*$ is $\mathcal{M}_k$-*sparse*). Formally, we seek $x^*$ such that

$$\|x - x^*\|_p \le C \cdot \min_{\substack{\text{supp}(x') \subseteq T \\ T \in \mathcal{M}_k}} \|x - x'\|_q \tag{3}$$

for some family $\mathcal{M}_k$ of $k$-subsets of $[n]$. Clearly, the original $k$-sparse recovery problem corresponds to the case, when $\mathcal{M}_k$ is a family of all $k$-subsets of $[n]$.

A prototypical example of a sparsity model is *block sparsity* [EM09]. Here the signal is divided into blocks of size $b$, and the non-zero coefficients belong to at most $k/b$ blocks. This model is particularly useful for bursty time signals, where the "activity" occurs during a limited time period, and is therefore contained in a few blocks. Another example is *tree sparsity* [RCB01] which models the structure of wavelet coefficients. Here the non-zero coefficients form a rooted subtree in a full binary tree defined over the coordinates.[2]. For many such scenarios the size of the family $\mathcal{M}_k$ is much smaller than $\binom{n}{k}$, which in principle makes it possible to recover an approximation from fewer measurements.

An elegant and very general model-based sparse recovery scheme was recently provided in a seminal work of Baraniuk et al [BCDH10]. The scheme has the property that, for any "computationally tractable" family of supports of "small" size, it guarantees a near-optimal sketch length $m = O(k)$, i.e., without any logarithmic factors. This is achieved by showing the existence of matrices $A$ satisfying the *model-based* variant of RIP. Formally, we say that $A$ satisfies $\varepsilon$-$\mathcal{M}_k$-RIP-$p$ if

$$(1 - \varepsilon)\|x\|_p \le \|Ax\|_p \le (1 + \varepsilon)\|x\|_p \tag{4}$$

for any $\mathcal{M}_k$-sparse vector $x \in \mathbb{R}^n$.

---

[2]See Section 2 for formal definitions of the two models.

In [BCDH10] it was shown that there exist matrices with $m = O(k)$ rows that satisfy $\varepsilon$-$\mathcal{M}_k$-RIP-2 as long as (i) either $\mathcal{M}_k$ is the block-sparse model and $b = \Omega(\log n)$ or (ii) $\mathcal{M}_k$ is the tree-sparse model. This property can be then used to give an efficient algorithm that, given $Ax$, finds $x^*$ satisfying a variant of the guarantee of Equation 3. However, the guarantees offered in [BCDH10], when phrased in the $\ell_1/\ell_1$ framework, results in a super-constant approximation factor $C = \Theta(\sqrt{\log n})$ [IP11]. The question of whether this bound can be improved has attracted considerable attention in signal processing and streaming communities. In particular, one of the problems[3] listed in the Bertinoro workshop open problem list [B1111] asks whether there exist matrices $A$ with $m = O(k)$ rows that provide the $\ell_1/\ell_1$ guarantee for the tree-sparse model with some constant approximation factor $C$.

**Our results**   In this paper we make a substantial progress on this question. In particular:

1. For both block-sparse and tree-sparse models, we show that there exist $m \times n$ matrices $A$ that provide the $\ell_1/\ell_1$ guarantee for some constant approximation factor $C$, such that the number of measurements improves over the bound of Equation 2 for a wide range of parameters $k$ and $b$. In particular we show that for the block-sparse model we can achieve $m = O(k \cdot (1 + \log_b(n/k)))$. This improves over the $O(k \log(n/k))$ bound of Equation 2 for any $b = \omega(1)$. In particular, if $b = n^{\Omega(1)}$, we obtain $m = O(k)$. For the tree-sparse model we achieve $m = O(k \log(n/k)/ \log \log(n/k))$ as long as $k = \omega(\log n)$. This also improves over the $O(k \log(n/k))$ bound of Equation 2.

   We note, however, that our results are not accompanied by efficient recovery algorithms. Instead, we show the existence of model-based RIP-1 matrices with the given number of rows. This implies that $Ax$ contains enough information to recover the desired approximation $x^*$ (see Appendix A for more details).

2. We complement the aforementioned results by showing that the measurement bounds achievable for a matrix satisfying block-sparse or tree-sparse RIP-1 property cannot be improved (i.e., our upper bounds are tight). This provides strong evidence that the number of measurements required for sparse recovery itself cannot be $O(k)$.

Our results show a significant difference between the model-based RIP-1 and RIP-2 matrices. For the $\ell_2$ norm, the original paper [BCDH10] shows that the number of measurements is fully determined by the cardinality of the model. Specifically, their proof proceeds by applying the union bound over all elements of $\mathcal{M}_k$ on top of the Johnson–Lindenstrauss-type concentration inequality. This leads to a measurement bound of $m = O(k + \log |\mathcal{M}_k|)$, which is $O(k)$ for the tree-sparse or block-sparse models. In contrast, in case of the $\ell_1$ norm we obtain an upper bound with a very different dependence on $|\mathcal{M}_k|$: namely, if $|\mathcal{M}_k| \geq n/k$, then we are able to achieve

$$m = O\left( \frac{k \cdot \log \frac{n \log(n/k)}{\log |\mathcal{M}_k|}}{\log \frac{k \log(n/k)}{\log |\mathcal{M}_k|}} \right).$$

Moreover, our lower bounds show that this bound is tight for the block- and the tree-sparse cases.

---

[3]See Question 15: Sparse Recovery for Tree Models. The question was posed by the first author.

**Our techniques** Our lower bounds are obtained by relating RIP-1 matrices to novel combinatorial/geometric structures we call *generalized expanders*. Specifically, it is known [BGI⁺08] that any *binary* 0-1 matrix $A$ that satisfies $(k, \varepsilon)$-RIP-1 is an adjacency matrix of an unbalanced $(k, \varepsilon)$-expander (see Section 2 for the formal definition). The notion of a generalized expander can be viewed as extending the notion of expansion to matrices that are not binary. Formally, we define it as follows.

**Definition 1** (Generalized expander)**.** *Let $A$ be an $m \times n$ real matrix. We say that $A$ is a generalized $(k, \varepsilon)$-expander, if all $A$'s columns have $\ell_1$-norm at most $1 + \varepsilon$, and for every $S \subseteq [n]$ with $|S| \leq k$ we have*

$$\sum_{i \in [m]} \max_{j \in S} |a_{ij}| \geq |S| \cdot (1 - \varepsilon).$$

Observe that the notion coincides with the standard notion of expansion for binary 0-1 matrices (after a proper scaling).

In this paper we show that any (not necessarily binary) RIP-1 matrix is also a generalized expander. We then use this fact to show that any RIP-1 matrix can be sparsified by replacing most of its entries by 0. This in turn lets us use counting arguments to lower bound the number of rows of such matrix.

Our upper bounds are obtained by constructing low-degree expander-like graphs. However, we only require that the expansion holds for the sets from the given model $\mathcal{M}_k$. This allows us to reduce the number of the right nodes of the graph, which corresponds to reducing the number of rows in its adjacency matrix.

## 2 Definitions

In this section we provide the definitions we will use throughout the text.

**Definition 2** (Expander)**.** *Let $G = (U, V, E)$ with $|U| = n$, $|V| = m$, $E \subseteq U \times V$ be a bipartite graph such that all vertices from $U$ have the same degree $d$. Then we say that $G$ is a $(k, \varepsilon)$-expander, if for every $S \subseteq U$ with $|S| \leq k$ we have*

$$| \{v \in V \mid \exists u \in S \ (u, v) \in E\} | \geq (1 - \varepsilon)d|S|.$$

**Definition 3** (Model)**.** *Let us call any non-empty subset*

$$\mathcal{M}_k \subseteq \Sigma_k = \{A \subseteq [n] \mid |A| = k\}$$

*a model.*

In particular, $\Sigma_k$ is a model as well.

**Definition 4** (Block-sparse model)**.** *Suppose that $b, k \in [n]$. Moreover, $b$ divides both $k$ and $n$. Let us partition our universe $[n]$ into $n/b$ disjoint blocks $B_1$, $B_2$, ..., $B_{n/b}$ of size $b$. We consider the following* block-sparse model*: $\mathcal{B}_{k,b}$ consists of all unions of $k/b$ blocks.*

**Definition 5** (Tree-sparse model)**.** *Suppose that $k \in [n]$ and $n = 2^{h+1} - 1$, where $h$ is a non-negative integer. Let us identify the elements of $[n]$ with the vertices of a full binary tree of depth $h$. Then, tree-sparse model $\mathcal{T}_k$ consists of all subtrees of size $k$ that contain the root of the full binary tree.*

4

**Definition 6** (Model-sparse vector/set). *Let $\mathcal{M}_k \subseteq \Sigma_k$ be any model. We say that a set $S \subseteq [n]$ is $\mathcal{M}_k$-sparse, if $S$ lies within a set from $\mathcal{M}_k$. Moreover, let us call a vector $x \in \mathbb{R}^n$ $\mathcal{M}_k$-sparse, if its support is a $\mathcal{M}_k$-sparse set.*

It is straightforward to generalize the notions of RIP-$p$ matrix, expanders and generalized expanders to the case of $\mathcal{M}_k$-sparse vectors and sets. Let us call the corresponding objects $\varepsilon$-$\mathcal{M}_k$-*RIP-$p$ matrix*, $\varepsilon$-$\mathcal{M}_k$-*expander* and *generalized $\varepsilon$-$\mathcal{M}_k$-expander*, respectively. Clearly, the initial definitions correspond to the case of $\Sigma_k$-sparse vectors and sets.

Our two main objects of interest are $\mathcal{B}_{k,b}$- and $\mathcal{T}_k$-RIP-1 matrices.

## 3    Sparsification of RIP-$1$ matrices

In this section we show that any $n \times m$ matrix, which is $(k, \varepsilon)$-RIP-1, can be sparsified after removing $(1 - \Omega(1))n$ columns (Theorem 1). Then we state an obvious generalization of this fact (Theorem 2), which will be useful for proving lower bounds on the number of rows for $\mathcal{B}_{k,b}$- and $\mathcal{T}_k$-RIP-1 matrices.

**Theorem 1.** *Let $A$ be any $m \times n$ matrix, which is $(k, \varepsilon)$-RIP-1. Then there exists an $m \times \Omega(n)$ matrix $B$ which is $(k, O(\varepsilon))$-RIP-1, has at most $O(m/k)$ non-zero entries per column and can be obtained from $A$ by removing some columns and then setting some entries to zero.*

We prove this theorem via the sequence of lemmas. First we prove that for every matrix $A$ there exists a $\pm 1$-vector $x$ such that $\|Ax\|_1$ is small.

**Lemma 1.** *Let $A$ be any $m \times k$ matrix. Then there exists a vector $x \in \{-1, 1\}^k$ such that*

$$\|Ax\|_1 \leq \sum_{i \in [m]} \left( \sum_{j \in [k]} a_{ij}^2 \right)^{1/2}. \tag{5}$$

*Proof.* Let us use a probabilistic argument. Namely, let us sample all coordinates $x_i$ independently and uniformly at random from $\{-1, 1\}$. Then

$$\mathrm{E}\left[\|Ax\|_1\right] = \sum_{i \in [m]} \mathrm{E}\left[\left| \sum_{j \in [k]} a_{ij} x_j \right|\right] \leq \sum_{i \in [m]} \left( \mathrm{E}\left[ \left( \sum_{j \in [k]} a_{ij} x_j \right)^2 \right] \right)^{1/2} =$$

$$= \sum_{i \in [m]} \left( \mathrm{E}\left[ \sum_{j \in [k]} a_{ij}^2 x_j^2 \right] \right)^{1/2} = \sum_{i \in [m]} \left( \sum_{j \in [k]} a_{ij}^2 \right)^{1/2}.$$

Thus, there exists a vector $x \in \{-1, 1\}^k$ that satisfies (5). $\qquad\square$

As a trivial corollary we have the following statement.

**Corollary 1.** *Let $A$ be any $m \times k$ matrix that preserves (up to $1 \pm \varepsilon$) $\ell_1$-norms of all vectors. Then*

$$\sum_{i \in [m]} \left( \sum_{j \in [k]} a_{ij}^2 \right)^{1/2} \geq (1 - \varepsilon)k.$$

The next lemma shows that every $(k,\varepsilon)$-RIP-1 matrix is a generalized $(k, O(\varepsilon))$-expander. This is a generalization of a theorem from [BGI$^+$08].

**Lemma 2.** *Let $A$ be any $m \times n$ matrix, which is $(k, \varepsilon)$-RIP-1. Then, $A$ is a generalized $(k, 3\varepsilon)$-expander.*

*Proof.* For the proof we need the following lemma.

**Lemma 3.** *For any $y \in \mathbb{R}^k$*

$$\|y\|_1 - \|y\|_\infty \leq \left(1 + \frac{1}{\sqrt{2}}\right)(\|y\|_1 - \|y\|_2). \tag{6}$$

*Proof.* Clearly, if $y = 0$, then the desired inequality is trivial. Otherwise, by homogenity we can assume that $\|y\|_1 = 1$. If $\|y\|_\infty = 1$, then $\|y\|_2 = 1$, and both sides of (6) are equal to zero. So, we can assume that $\|y\|_\infty < 1$. Suppose that $\|y\|_\infty = t$ for some $t \in (0; 1)$. If $1/n > t \geq 1/(n+1)$ (thus, $n = \lceil 1/t - 1\rceil$) for some positive integer $n$, then, clearly, $\|y\|_2 \leq \sqrt{nt^2 + (1-nt)^2}$. One can check using elementary analysis that for every $t \in (0; 1)$

$$\frac{1 - \|y\|_\infty}{1 - \|y\|_2} \leq \frac{1 - t}{1 - \sqrt{\lceil\frac{1}{t} - 1\rceil t^2 + \left(1 - \lceil\frac{1}{t} - 1\rceil t\right)^2}} \leq 1 + \frac{1}{\sqrt{2}}$$

(equality is attained on $t = 1/2$). This concludes the proof. $\qquad\square$

Let $S \subseteq [n]$ be any subset of size at most $k$. For any $i \in [m]$ let us denote $y_i = (a_{ij})_{j \in S} \in \mathbb{R}^S$. We have

$$
\begin{aligned}
\sum_{i \in [m]} \|y_i\|_\infty &\geq \left(1 + \frac{1}{\sqrt{2}}\right) \sum_{i \in [m]} \|y_i\|_2 - \frac{1}{\sqrt{2}} \cdot \sum_{i \in [m]} \|y_i\|_1 \quad \text{(by Lemma 3)} \\
&\geq \left(1 + \frac{1}{\sqrt{2}}\right)(1 - \varepsilon)|S| - \frac{1}{\sqrt{2}} \cdot (1 + \varepsilon)|S| \quad \text{(by Corollary 1 and RIP-1)} \\
&= (1 - (1 + \sqrt{2})\varepsilon)|S|.
\end{aligned}
$$

So, $A$ is a generalized $(k, (1 + \sqrt{2})\varepsilon)$-expander. Since $1 + \sqrt{2} < 3$, this concludes the proof.
$\qquad\square$

Finally, we prove Theorem 1.

*Proof of Theorem 1.* By Lemma 2 $A$ is a generalized $(k, 3\varepsilon)$-expander. Let us partition $[n]$ into $n/k$ disjoint sets of size $k$ arbitrarily: $[n] = S_1 \cup S_2 \cup \ldots \cup S_{n/k}$. Now for every $i \in [m]$ and every $S_t$ let us zero out all the entries $a_{ij}$ for $j \in S_t$ except one with the largest absolute value. Let $A'$ be the resulting matrix. Since $A$ is a generalized $(k, 3\varepsilon)$-expander, we know that the (vector) $\ell_1$ norm of the difference $A - A'$ is at most $3\varepsilon n$. Thus, each column of $A - A'$ has the $\ell_1$ norm of at most $3\varepsilon$ *on the average*. The number of non-zero entries in $A'$ is at most $mn/k$, so a column has at most $m/k$ non-zero entries *on the average*. Thus, by Markov inequality there is a set of $n/3$ columns such that we have moved each of them by at most $9\varepsilon$ and each of them contains at most $3m/k$ non-zero entries. We define a matrix $B$ that consists of these columns. Since we have modified each of these columns by at most $9\varepsilon$ and $A$ is $(k, \varepsilon)$-RIP-1 we have that $B$ is $(k, 10\varepsilon)$-RIP-1. $\qquad\square$

The following theorem is a straightforward generalization of Theorem 1. It can be proved via literally the same argument.

**Theorem 2.** *Suppose that a model $\mathcal{M}_k \subseteq \Sigma_k$ has the following properties:*

- *for some $l \leq k$ all sets from $\Sigma_l$ are $\mathcal{M}_k$-sparse;*

- *there exists a partition of an $\Omega(1)$-fraction of $[n]$ into disjoint subsets of size $\Omega(k)$ such that each of these subsets is $\mathcal{M}_k$-sparse.*

*Then if $A$ is an $m \times n$ matrix which is $\varepsilon$-$\mathcal{M}_k$-RIP-1 for some sufficiently small $\varepsilon > 0$, there exists an $m \times \Omega(n)$ matrix $B$ which is $(l, O(\varepsilon))$-RIP-1, has at most $O(m/k)$ non-zero entries per column and can be obtained from $A$ by removing some columns and then setting some entries to zero.*

## 4 Lower bounds for model-based RIP-1 matrices

In this section we prove lower bounds on the number of rows for $\mathcal{B}_{k,b}$- and $\mathcal{T}_k$-RIP-1 matrices.

This is done using the following general theorem.

**Theorem 3.** *If a model $\mathcal{M}_k \subseteq \Sigma_k$ satisfies the statement of Theorem 2 and $A$ is an $m \times n$ matrix which is $\varepsilon$-$\mathcal{M}_k$-RIP-1 for some sufficiently small $\varepsilon > 0$, then*

$$m = \Omega\left(k \cdot \frac{\log(n/k)}{\log(k/l)}\right).$$

The proof is a combination of Theorem 2 and a counting argument similar to one used in [Nac10]. First, we need the following standard geometric fact.

**Theorem 4.** *Let $v_1, v_2, \ldots, v_n \in \mathbb{R}^d$ be a set of $d$-dimensional vectors such that*

- *for every $i \in [n]$ we have $\|v_i\|_1 \leq 1.1$;*

- *for every $i \neq j \in [n]$ we have $\|v_i - v_j\|_1 \geq 0.9$.*

*Then, $n \leq 4^d$.*

*Proof.* Denote $B(x,r)$ the ball in $\ell_1$-metric with center $x$ and radius $r$. Consider the balls $B_i = B(v_i, 0.45)$. On the one hand, these balls are disjoint, on the other hand, they lie within $B(0, 1.55)$. Thus, if we consider the balls' volumes we see that

$$n \leq \left(\frac{1.55}{0.45}\right)^d < 4^d.$$

$\square$

The next theorem shows a tradeoff between $m$ and column sparsity for any RIP-1 matrix. Its variant was proved in [Nac10], but we present here the proof for the sake of completeness.

**Theorem 5** ([Nac10]). *Let $A$ be an $m \times n$ matrix, which is $(k, \varepsilon)$-RIP-1 for some sufficiently small $\varepsilon > 0$. Moreover, suppose that every column of $A$ has at most $s$ non-zero entries. Then*

$$s \log\left(\frac{m}{sk}\right) = \Omega\left(\log\left(\frac{n}{k}\right)\right).$$

*Proof.* We need a lemma from [Nac10], which is proved by a standard probabilistic argument.

**Lemma 4.** *There exists a set $X \subseteq \mathbb{R}^n$ of $k/2$-sparse vectors such that*

- $\log |X| = \Omega(k \log(n/k))$;

- *every vector from $X$ has a unit $\ell_1$-norm;*

- *all pairwise $\ell_1$-distances between the elements of $X$ are at least $1$.*

Now let us see how $A$ acts on the elements of $X$. Clearly, for every $x \in X$ the vector $Ax$ is $sk$-sparse. By pigeonhole principle we have that for some $S \subseteq [m]$ with $|S| \leq sk$ there exists a subset $X' \subseteq X$ with

$$|X'| \geq \frac{|X|}{\binom{m}{sk}} \tag{7}$$

such that for every $x \in X'$ the support of $Ax$ lies within $S$.

On the other hand, since $A$ is $(k, \varepsilon)$-RIP-1 one can easily see that the set $\{Ax\}_{x \in X'}$ (which lies in the $sk$-dimensional subspace) has the following properties:

- every vector from the set has $\ell_1$-norm at most $1 + \varepsilon$;

- all pairwise distances are at least $1 - \varepsilon$.

Since this set lies in the $sk$-dimensional subspace by Theorem 4 its cardinality is bounded by $4^{sk}$ (provided that $\varepsilon$ is sufficiently small). Thus, we have by plugging this bound into (7)

$$\frac{2^{\Omega(k \log(n/k))}}{\binom{m}{sk}} \leq 4^{sk}.$$

Now by using a standard estimate $\binom{m}{sk} \leq 2^{O(sk \log(m/sk))}$ we have the desired statement. $\square$

Now we can finish the proof of Theorem 3.

*Proof of Theorem 3.* By Theorem 2 we can get an $m \times \Omega(n)$ matrix $A$ with column sparsity $s = O(m/k)$ and which is $(l, O(\varepsilon))$-RIP-1. Then applying Theorem 5 we have $s \log(m/sl) = \Omega(\log(n/l))$. Since, $s = O(m/k)$ we get the desired bound

$$m = \Omega \left( k \cdot \frac{\log(n/k)}{\log(k/l)} \right).$$

$\square$

Next we apply Theorem 3 to $\mathcal{B}_{k,b}$- and $\mathcal{T}_k$-RIP-1 matrices.

**Theorem 6.** *For any $k \geq 2b$ and sufficiently small $\varepsilon > 0$ if $A$ is an $m \times n$ matrix which is $\varepsilon$-$\mathcal{B}_{k,b}$-RIP-1, then $m = \Omega(k \cdot (1 + \log_b(n/k)))$.*

*Proof.* Clearly, $\mathcal{B}_{k,b}$ satisfies the conditions of Theorem 2 for $l = k/b$. Thus, by Theorem 3 we have

$$m = \Omega \left( k \cdot \log_b(n/k) \right).$$

At the same time, the lower bound $\Omega(k)$ is obvious. Combining them together we get the desired result. $\square$

**Theorem 7.** *Let $A$ be an $m \times n$ matrix which is $\varepsilon$-$\mathcal{T}_k$-RIP-1. Then, if $\varepsilon$ is sufficiently small and $k = \omega(\log n)$,*

$$m = \Omega\left(k \cdot \frac{\log(n/k)}{\log\log(n/k)}\right).$$

*Proof.* The next Lemma shows that for any $k = \omega(\log n)$ the model $\mathcal{T}_k$ satisfies the first condition of Theorem 2 with $l = \Omega(k/\log(n/k))$.

**Lemma 5.** *Let $S \subseteq [n]$ be a subset of the full binary tree. Then there exists a subtree that contains both $S$ and the root with at most $O(|S|\log(n/|S|))$ vertices.*

*Proof.* Let $T$ be a subtree that consists of $\log|S|$ levels of the full binary tree that are closest to the root. Let $T'$ be a subtree that is a union of $T$ and paths from the root to all the elements of $|S|$. It is not hard to see that $|T' \setminus T| \leq |S|\log(n/|S|)$. As a result we get

$$|T'| \leq |T| + |S|\log(n/|S|) \leq O(|S|\log(n/|S|)).$$

$\square$

The second condition of Theorem 2 is satisfied as well (here we use that $k = \omega(\log n)$). Thus, applying Theorem 3 we have

$$m = \Omega\left(k \cdot \frac{\log(n/k)}{\log\log(n/k)}\right).$$

$\square$

# 5   Upper bounds for model-based RIP-1 matrices

In this section we complement the lower bounds by upper bounds.

We use the following obvious modification of a theorem from [BGI+08].

**Theorem 8** ([BGI+08])**.** *If a graph $G = (U, V, E)$ is an $\varepsilon$-$\mathcal{M}_k$-expander for some model $\mathcal{M}_k \subseteq \Sigma_k$, then the normalized (by a factor of $d$, where $d$ is the degree of all vertices from $U$) adjacency matrix of $G$ (which size is $|V| \times |U|$) is an $O(\varepsilon)$-$\mathcal{M}_k$-RIP-1 matrix.*

Thus, it is sufficient to build $\mathcal{M}_k$-expanders with as small $m$ as possible. We use the standard probabilistic argument to show the existence of such graphs. Namely, for every vertex $u \in U$ we sample a subset of $[m]$ of size $d$ ($d$ and $m$ have to be carefully chosen). Then, we connect $u$ and all the vertices from this subset. All sets we sample are uniform (among all $d$-subsets of $[m]$) and independent.

We use the following tail inequality, which can be proved using Chernoff bound (and whose slight variant is proved and used in [BMRV02]).

**Lemma 6** ([BMRV02])**.** *There exist constants $C > 1$ and $\delta > 0$ such that, whenever $m \geq Cdt/\varepsilon$, one has for any $T \subseteq U$ with $|T| = t$*

$$\Pr\left[|\{v \in V \mid \exists u \in T \; (u, v) \in E\}| < (1 - \varepsilon)dt\right] \leq \left(\delta \cdot \frac{\varepsilon m}{dt}\right)^{-\varepsilon dt}$$

For the proof see Appendix (Section B).

For a model $\mathcal{M}_k \subseteq \Sigma_k$ denote $\#(\mathcal{M}_k, t)$ the number of $\mathcal{M}_k$-sparse sets of size $t$ (for $t \in [k]$). We use the following simple estimate.

**Lemma 7.**
$$\forall t \in [k] \quad \#(\mathcal{M}_k, t) \le \min\left\{ |\mathcal{M}_k| \cdot \binom{k}{t}, \binom{n}{t} \right\}.$$

Now if we combine this Lemma with the standard estimate $\binom{u}{v} \le (eu/v)^v$, we get the following bound.

$$\forall t \in [k] \quad \#(\mathcal{M}_k, t) \le \min\left\{ |\mathcal{M}_k| \cdot \left(\frac{ek}{t}\right)^t, \left(\frac{en}{t}\right)^t \right\} \tag{8}$$

Now we combine these estimates, Lemma 6, Theorem 8 and the union bound to get upper bounds for $m$ for $\mathcal{M}_k$-RIP-1 matrices.

**Theorem 9.** *For any $0 < \varepsilon < 1/2$ and any model $\mathcal{M}_k$ that is of size at least $n/k$ there exists an $\varepsilon$-$\mathcal{M}_k$-RIP-1 matrix with*
$$m = O\left( \frac{k}{\varepsilon^2} \cdot \frac{\log(n/l)}{\log(k/l)} \right),$$
*where*
$$l = \frac{\log |\mathcal{M}_k|}{\log(n/k)}.$$

The proof is in the appendix. We note that the condition $|\mathcal{M}_k| \ge n/k$ is needed to make sure that $l \ge 1$.

Now we derive corollaries for the block- and tree-sparse cases.

**Corollary 2.** *For any $0 < \varepsilon < 1/2$ and any $1 \le b \le k \le n$ there exists an $\varepsilon$-$\mathcal{B}_{k,b}$-RIP-1 matrix with*
$$m = O\left( \frac{k \cdot (1 + \log_b(n/k))}{\varepsilon^2} \right)$$
*rows.*

*Proof.* It is easy to see that
$$|\mathcal{M}_k| = \binom{n/b}{k/b} \ge n/k.$$
Thus, we can apply Theorem 9 with $l = k/b$ and get the required bound. $\qquad\square$

**Corollary 3.** *For any $0 < \varepsilon < 1/2$ and any $k = \omega(\log n)$ there exists an $\varepsilon$-$\mathcal{T}_k$-RIP-1 matrix with*
$$m = O\left( \frac{k}{\varepsilon^2} \cdot \frac{\log(n/k)}{\log\log(n/k)} \right)$$
*rows.*

*Proof.* Using a simple estimate on Catalan numbers we can see that $|\mathcal{T}_k| \le 4^k$. Since $k = \omega(\log n)$, we are in position to apply Theorem 9 and get the required bound. $\qquad\square$

# 6 Acknowledgments

# References

[B1111] Open problems in data streams, property testing, AND related topics. http://people.cs.umass.edu/~mcgregor/papers/11-openproblems.pdf, 2011. 3

[BCDH10] Richard G. Baraniuk, Volkan Cevher, Marco F. Duarte, and Chinmay Hegde. Model-based compressive sensing. *IEEE Transactions on Information Theory*, 56(4):1982–2001, 2010. 1, 2, 3

[BGI$^+$08] Radu Berinde, Anna C. Gilbert, Piotr Indyk, Howard J. Karloff, and Martin J. Strauss. Combining geometry and combinatorics: A unified approach to sparse signal recovery. *Allerton*, 2008. 2, 3, 5, 9

[BIPW10] Khanh Do Ba, Piotr Indyk, Eric Price, and David P. Woodruff. Lower Bounds for Sparse Recovery. In *SODA*, pages 1190–1197, 2010. 2

[BMRV02] Harry Buhrman, Peter Bro Miltersen, Jaikumar Radhakrishnan, and Srinivasan Venkatesh. Are Bitvectors Optimal? *SIAM J. Comput.*, 31(6):1723–1744, 2002. 9, 13

[CRT06] Emmanuel J. Candès, Justin K. Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, August 2006. 1, 2

[Don06] David L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006. 1, 2

[EM09] Y. Eldar and M. Mishali. Robust recovery of signals from a structured union of subspaces. *IEEE Trans. Inform. Theory*, 55(11):5302–5316, 2009. 2

[FPRU10] Simon Foucart, Alain Pajor, Holger Rauhut, and Tino Ullrich. The Gelfand widths of lp-balls for 0p ≤ 1. *J. Complex.*, 26(6):629–640, December 2010. 2

[GG84] A.Yu. Garnaev and E.D. Gluskin. On widths of the Euclidean ball. *Sov. Math., Dokl.*, 30:200–204, 1984. 2

[GI10] A. Gilbert and P. Indyk. Sparse recovery using sparse matrices. *Proceedings of IEEE*, 2010. 1, 2

[Glu84] E.D. Gluskin. Norms of random matrices and widths of finite-dimensional sets. *Math. USSR, Sb.*, 48:173–182, 1984. 2

[IP11]    Piotr Indyk and Eric Price. K-median clustering, model-based compressive sensing, and sparse recovery for earth mover distance. In *STOC*, pages 627–636, 2011. 3

[Kas77]   B.S. Kasin. Diameters of some finite-dimensional sets and classes of smooth functions. *Math. USSR, Izv.*, 11:317–333, 1977. 2

[MR95]    Rajeev Motwani and Prabhakar Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995. 13

[Mut05]   S. Muthukrishnan. Data streams: algorithms and applications. *Found. Trends Theor. Comput. Sci.*, 1(2):117–236, August 2005. 1

[Nac10]   Mergen Nachin. Lower Bounds on the Column Sparsity of Sparse Recovery Matrices. MIT Undergraduate Thesis, 2010. 7

[RCB01]   J. Romberg, H. Choi, and R. Baraniuk. Bayesian tree-structured image modeling using wavelet-domain Hidden Markov Models. 10(7):1056–1068, 2001. 2

# A    RIP-$1$ yields sparse recovery

In this section we show improved upper bounds on the number of measurements needed to recover good block- or tree-sparse approximations with $\ell_1/\ell_1$ guarantee and constant approximation factor. This result is folklore, but we include it for completeness.

Suppose that $\mathcal{M}_k \subseteq \Sigma_k$ is some model. We say that an $m \times n$ matrix $A$ is $\varepsilon$-$\mathcal{M}_k^{(2)}$-RIP-1, if for every $x \in \mathbb{R}^n$ such that $\operatorname{supp} x \subseteq S_1 \cup S_2$ for some $\mathcal{M}_k$-sparse sets $S_1$ and $S_2$ one has

$$(1 - \varepsilon)\|x\|_1 \le \|Ax\|_1 \le (1 + \varepsilon)\|x\|_1.$$

Let $A$ be *any* $\varepsilon$-$\mathcal{M}_k^{(2)}$-RIP-1 matrix for a sufficiently small $\varepsilon$ such that $\|A\|_1 \le 1 + \varepsilon$. Algorithm 1 (whose running time is exponential in $n$) given $y = Ax$ for some $x \in \mathbb{R}^n$ recovers a vector $x^* \in \mathbb{R}^n$ such that

$$\|x - x^*\|_1 \le (3 + O(\varepsilon)) \cdot \min_{x' \text{ is } \mathcal{M}_k\text{-sparse}} \|x - x'\|_1. \tag{9}$$

Note that the optimization problem within the for-loop can be easily reduced to a linear program.

---

**Algorithm 1** Model-based sparse recovery

---
**Input:** $y = Ax$ for some $x \in \mathbb{R}^n$
**Output:** a good $\mathcal{M}_k$-approximation $x^*$ of $x$
  $x^* \leftarrow 0$
  **for** $S \subseteq [n]$ is an $\mathcal{M}_k$-sparse set **do**
    $\tilde{x} \leftarrow \operatorname{argmin}_{\operatorname{supp} x' \subseteq S} \|y - Ax'\|_1$
    **if** $\|y - A\tilde{x}\|_1 \le \|y - Ax^*\|_1$ **then**
      $x^* \leftarrow \tilde{x}$
    **end if**
  **end for**

---

Now let us prove that the resulting vector $x^*$ satisfies (9). Denote $x_{\mathcal{M}_k}$ an $\mathcal{M}_k$-sparse vector that minimizes $\|x - x_{\mathcal{M}_k}\|_1$. Now we have

$$\|x - x^*\|_1 \leq \|x - x_{\mathcal{M}_k}\|_1 + \|x_{\mathcal{M}_k} - x^*\|_1 \leq \|x - x_{\mathcal{M}_k}\|_1 + (1 + O(\varepsilon))\|A(x_{\mathcal{M}_k} - x^*)\|_1 \leq$$
$$\leq \|x - x_{\mathcal{M}_k}\|_1 + (1 + O(\varepsilon))(\|A(x - x_{\mathcal{M}_k})\|_1 + \|A(x - x^*)\|_1) \leq$$
$$\leq (2 + O(\varepsilon))\|x - x_{\mathcal{M}_k}\|_1 + (1 + O(\varepsilon))\|A(x - x^*)\|_1 \leq$$
$$\leq (2 + O(\varepsilon))\|x - x_{\mathcal{M}_k}\|_1 + (1 + O(\varepsilon))\|A(x - x_{\mathcal{M}_k})\|_1 \leq (3 + O(\varepsilon))\|x - x_{\mathcal{M}_k}\|_1.$$

The second inequality is true, since $A$ is $\varepsilon\text{-}\mathcal{M}_k^{(2)}$-RIP-1 and both $x_{\mathcal{M}_k}$ and $x^*$ are $\mathcal{M}_k$-sparse. The fourth inequality is true, because $\|A\|_1 \leq 1 + \varepsilon$. The fith inequality is due to the construction of the algorithm: clearly, $\|A(x - x^*)\|_1 \leq \|A(x - x_{\mathcal{M}_k})\|_1$.

It is immediate to see that any $\varepsilon\text{-}\mathcal{B}_{2k,b}$-RIP-1 matrix is $\varepsilon\text{-}\mathcal{B}_{k,b}^{(2)}$-RIP-1. Similarly, any $\varepsilon\text{-}\mathcal{T}_{2k}$-RIP-1 matrix is $\varepsilon\text{-}\mathcal{T}_k^{(2)}$-RIP-1. Moreover, since all the singletons are both block- and tree-sparse, we have that these matrices have $\ell_1$-norm at most $1 + \varepsilon$. Thus, plugging Corollaries 2 and 3 we get the following result.

**Theorem 10.** *The problem of model-based stable sparse recovery with $\ell_1/\ell_1$ guarantee and a constant approximation factor can be solved*

- *with*
$$m = O(k \cdot (1 + \log_b(n/k)))$$
*measurements for $\mathcal{B}_{k,b}$,*

- *with*
$$m = O\left(k \cdot \frac{\log(n/k)}{\log\log(n/k)}\right)$$
*measurements for $\mathcal{T}_k$, provided that $k = \omega(\log n)$.*

# B  Proof of Lemma 6

To prove the lemma we need the following version of Chernoff bound [MR95].

**Theorem 11** ([MR95]). *Suppose that $X_1$, $X_2$, ..., $X_n$ are independent binary random variables. Denote $\mu = \mathrm{E}\left[X_1 + X_2 + \ldots + X_n\right]$. Then for any $\tau > 0$*

$$\Pr\left[X_1 + X_2 + \ldots + X_n \geq (1 + \tau)\mu\right] \leq \left(\frac{e^\tau}{(1 + \tau)^{(1+\tau)}}\right)^\mu.$$

We can enumerate all $dt$ outgoing from $T$ edges arbitrarily. Then let us denote $C_u$ the event "$u$-th of these edges collides with $v$-th edge on the right side, where $v < u$".

We would like to upper bound the probability of the event "at least $\varepsilon dt$ of events $C_u$ happen". As shown in [BMRV02] this probability is at most

$$\Pr\left[B\left(dt, \frac{dt}{m}\right) \geq \varepsilon dt\right].$$

Thus we can apply Theorem 11 with $\tau = \varepsilon m/dt - 1$ and $\mu = (dt)^2/m$. Then, the statement of the Lemma can be verified routinely.

# C  Proof of Theorem 9

If we combine Lemma 6 with the union bound we see that it is sufficient to prove that

$$\forall 1 \leq t \leq l \quad \#(\mathcal{M}_k, t) \cdot \left(\delta \cdot \frac{\varepsilon m}{dt}\right)^{-\varepsilon dt} < \frac{1}{2l},$$

and

$$\forall l \leq t \leq k \quad \#(\mathcal{M}_k, t) \cdot \left(\delta \cdot \frac{\varepsilon m}{dt}\right)^{-\varepsilon dt} < \frac{1}{2k},$$

provided that $m \geq Cdk/\varepsilon$ (here $C$ and $\delta$ are the constants from the statement of Lemma 6 and $l$ is the quantity from the statement of Theorem 9). Now let us plug (8). Namely, for $1 \leq t \leq l$ we use the second estimate, for $l \leq t \leq k$ we use the first estimate. Thus, it is left to choose $d$ such that we have

$$\forall 1 \leq t \leq l \qquad \left(\frac{en}{t}\right)^t \left(\frac{C'k}{t}\right)^{-\varepsilon dt} < \frac{1}{2l} \tag{10}$$

$$\forall l \leq t \leq k \qquad |\mathcal{M}_k| \cdot \left(\frac{ek}{t}\right)^t \left(\frac{C'k}{t}\right)^{-\varepsilon dt} < \frac{1}{2k}, \tag{11}$$

for $C'$ being sufficiently large constant. It would allow us to set $m = O(dk/\varepsilon)$.

It is straightforward to check that, whenever $d > 1/\varepsilon$, the left-hand sides of (10) and (11) are log-convex. Thus, to check (10) and (11), it sufficient to check them for $t = 1, l, k$.

Let us first handle (10). For it to hold for a given $1 \leq t \leq l$ it is enough to set

$$d = O\left(\frac{1}{\varepsilon \cdot \log(k/t)} \cdot \left(\frac{\log l}{t} + \log(n/t)\right)\right).$$

For $t = 1$ this expression is $O(\varepsilon^{-1} \cdot \log_k n)$, for $t = l$ it is

$$O\left(\frac{\log(n/l)}{\varepsilon \cdot \log(k/l)}\right). \tag{12}$$

Clearly, the latter bound is always larger than the former one.

Now let us handle (11). Similarly, for $l \leq t \leq k$ it is enough to set

$$d = O\left(\frac{1}{\varepsilon} \cdot \left(1 + \frac{\log|\mathcal{M}_k| + \log k}{t \cdot \log(k/t)}\right)\right) = O\left(\frac{1}{\varepsilon} \cdot \left(1 + \frac{l \cdot \log(n/k) + \log k}{t \cdot \log(k/t)}\right)\right), \tag{13}$$

where the second step is due to the definition of $l$ ($l = \log|\mathcal{M}_k|/\log(n/k)$).

Now we will prove that for $t = l, k$ the right-hand side of (13) is at most (12).

For $t = l$ we obviously have

$$\frac{l \cdot \log(n/k)}{t \cdot \log(k/t)} = O\left(\frac{\log(n/l)}{\log(k/l)}\right).$$

At the same time

$$\frac{\log k}{t \cdot \log(k/t)} = O\left(\frac{\log(n/l)}{\log(k/l)}\right),$$

14

since $\log k = O(l \cdot \log(n/l))$. Indeed, if $l \geq \log n$, then we are done. Otherwise, $\log(n/l) = \Omega(\log n) = \Omega(\log k)$, and we are again done.

Now let us handle $t = k$. Indeed,

$$\frac{\log k}{t \cdot \log(k/t)} = O(1)$$

in this case. At the same time,

$$\frac{l \cdot \log(n/k)}{t \cdot \log(k/t)} = O\left(\frac{\log(n/l)}{\log(k/l)}\right),$$

since $l \cdot \log(k/l) = O(k)$.

Overall, this argument shows that we can take

$$d = O\left(\frac{\log(n/l)}{\varepsilon \cdot \log(k/l)}\right)$$

and

$$m = O\left(\frac{k}{\varepsilon^2} \cdot \frac{\log(n/l)}{\log(k/l)}\right).$$