Abstract. It is well known that sparse signals can be succinctly represented by certain low-dimensional linear sketches with applications in compressive sensing, data streaming and graph-sketching, among others. Recently, structured sparsity has emerged as a promising new tool for reducing sketch size and improving recovery. By structured sparsity, we mean that the sparse coefficients exhibit further correlations as determined by a model. Existing work on sketching structured sparse signals requires dense sketching matrices that satisfy the 2-norm restricted isometry property. On the other hand, sparse sketching matrices, usually from expanders, are computationally much more efficient, easier to store and apply in recovery. In this paper, we focus on model-based expanders, that is expanders that capture a given structure sparsity model, and show that they exist for a larger class of models than previously considered. We present the first polynomial time algorithm for recovering structured sparse signals from low-dimensional linear sketches obtained via sparse matrices. The algorithm is guaranteed to yield signals with bounded recovery error and is quite easy to implement and customize for structured sparse models that are endowed with a “projection” operator. As a result, we characterize a broad class of structured sparsity models that have polynomial time projection property. We also provide numerical experiments to illustrate the theoretical results in action.

*LIONS, EPFL, Lausanne, Switzerland ({bubacarr.bah, luca.baldassarre, volkan.cevher}@epfl.ch).
1. Introduction.

1.1. Problem overview and statement. In the recent years, we witnessed the advent of a new paradigm of succinct approximate representations of signals. A signal $x \in \mathbb{R}^N$ can be represented by the measurements or sketch $Ax$, for $A \in \mathbb{R}^{m \times N}$, where $m \ll N$. The sketching is linear and dimensionality reducing, however it can still retain most of the information about the signal. These representations have been successfully applied in such diverse domains such as data streaming \cite{22,29}, compressing sensing \cite{10,12}, graph sketching \cite{1,19} and even to breaking privacy of databases via aggregate queries \cite{14}.

The standard framework is to assume that $x$ is $k$-sparse, i.e., only $k$ of its $N$ components are nonzero. The aim is then to find conditions on the sketch matrix $A$ and the recovery algorithm $\Delta$ that yields $\hat{x} = \Delta(Ax + e)$, where $e$ measures the perturbation of the linear model, such that the following error guarantee holds

$$\|\hat{x} - x\|_p \leq C_1 \sigma_k(x)_q + C_2 \|e\|_p \quad \text{where} \quad \sigma_k(x)_q = \min_{k\text{-sparse } x'} \|x - x'\|_q$$

for some constants $C_1, C_2 > 0$ and $1 \leq q \leq p \leq 2$. We denote this type of error guarantees as $\ell_p/\ell_q$. It has been shown that the minimum sketch length can be obtained via a randomly generated $A$, leading to two different probabilistic models for which (1.1) holds. The for all case where a randomly generated $A$ and recovery algorithm $\Delta$ satisfy (1.1) for all $x \in \mathbb{R}^N$ with high probability; and the for each case where a randomly generated $A$, independent of $x$, and recovery algorithm $\Delta$ satisfy (1.1) for each given $x \in \mathbb{R}^N$ with high probability \cite{7}.

In many cases, additional prior knowledge on the signal is available and can be encoded in a structure that models the relationships between the components of $x$. The acquisition and recovery of signals with such a-priori known structures, like tree and block structures \cite{4,24} is referred to as model-based compressed sensing (CS). Model-based CS has the advantage of reducing the minimum sketch length for which a variant of (1.1) holds. This work focuses on sparse measurement matrices, that have the so called $\ell_1$-norm restricted isometry property (to be defined later), designed for model-based CS, and will be referred to a model-expanders.

The recent work of \cite{24} is the first to tackle model-based CS with sparse measurement matrices and in the for all case. They show that model-expanders exist when the structures (models) considered are binary trees and non-overlapping block structures and the number of rows are above a certain bound. They also propose a recovery algorithm that is exponential in the dimension of the signal. In this work, we propose instead a polynomial time algorithm.

1.2. Main results. In this manuscript we propose an efficient linear time algorithm for model-based CS when the measurement matrices are model-expanders with provable $\ell_1/\ell_1$ theoretical guarantees (to be made precise below). Our theoretical results are supported with numerical experiments. In addition, this work builds on \cite{24} to extend the class of models for which we can show existence of model-expanders that meet the bounds in \cite{24}. Note that our results are for the for all case.

Formally, in order to recover $k$-model sparse (to be defined in Section 2) signals, we propose the algorithm: Model-based iterative hard thresholding (MIHT). This is a combination of the iterative hard thresholding for expanders (EIHT) algorithm \cite{17} (which is a modification of the popular sparse matching pursuit (SMP) compressed sensing algorithm \cite{8}) and efficient algorithms for projecting onto our models \cite{3}. MIHT runs in linear time and yields an approximation $\hat{x}$ that is in our model, with the following $\ell_1/\ell_1$ error guarantee in the for all case

$$\|\hat{x} - x_S\|_1 \leq C_1 \sigma_{M_k}(x)_1 + C_2 \|e\|_1,$$

for some constants $C_1, C_2 > 0$, where $\sigma_{M_k}(x)_1 := \min_{k\text{-model sparse } x'} \|x - x'\|_1$. This is an adaptation of (1.1) that ensures that the approximations belong to the model considered. The algorithm runtime complexity crucially depends on the cost of projecting onto the model, but we show that the projections for the considered models can be computed exactly and rather efficiently in $O(Nk)$ time.

We consider two categories of structured signals: a) $k$-rooted connected signals which have their $k$ nonzero components lie in a rooted connected tree \cite{4,5,24}, and b) $k$-group sparse signals whose support is contained in the union of $k$ overlapping groups \cite{25,32}. Note that $k$-block sparse signals whose support consists of $k$ equal size non-overlapping blocks \cite{4,24,33,34} are a special case of b). We provably show that our algorithm can efficiently recover signals from these models and our numerical results support this. In addition, we generalize the results of \cite{24}. We show the existence of model-expanders based on model a) for $D$-ary tree models and model b), which satisfy their bounds.
Our key contribution is introducing the first linear time algorithm that resolves a huge bottleneck in the application of model-expanders. Another contribution is the generalization of the results of [24] to cover a broader class of models. This is also significant for applications, since for instance we can consider 4-ary wavelet trees instead of just binary wavelet trees.

1.3. Related work. It has been observed that many signals have an underlying structure that can be exploited in compressed sensing, see for example [13,15,16,33], just to cite a few. Model-based compressed sensing [4,24] was the terminology coined to refer to compressed sensing that exploits the structure of the signal in order to obtain approximations that satisfy (1.2) from fewer measurements than compressed sensing, see for example [13, 15, 16, 33], just to cite a few. Model-based compressed sensing [4, 24] was recently [24, 31] most of the works in this area (for the case). It is important to point out that the first work [24] on model-based RIP-1 matrices (in the case) also relied on a probabilistic construction. We also show that it is possible to probabilistically construct more general model expanders. It is important to point out that the first work to use model-based sparse matrices is [31]. Their signal models are similar to those considered in [24] and they showed such matrices have the \( \ell_1 \)-norm restricted isometry property (RIP-1). A binary matrix \( A \) with \( d \) nonzeros per column has RIP-1 [7] if it satisfies

\[
(1 - \delta_k)\|x\|_1 \leq \|Ax\| \leq d\|x\|_1, \quad \text{for all } k\text{-sparse vectors } x \in \mathbb{R}^N.
\]

These sparse matrices have low storage complexity since they have only \( d \), a small fraction, nonzero entries per column and they lead to much faster matrix products. Sparse recovery algorithms, for the generic sparse recovery problem with sparse non-zero mean matrices and for all guarantees, include \( \ell_1 \)-minimization [7] and combinatorially greedy algorithms: Sparse Matching Pursuit (SMP) [8] and Expander Recovery Algorithm [26] (see [2, 18] for more exhaustive lists). However, [24] is the first to propose an exponential time algorithm for model-sparse signals using model expanders (which they called model-based RIP-1 matrices) in the for all case. We significantly improve on this work by proposing the first polynomial time algorithm for model-sparse signals using model expanders.

The proposed algorithm involves projecting a signal onto the considered model. Projections onto group-based models and rooted connected subtrees have been recently considered in [3] and [11]. In the former paper, the authors showed that projections onto group-based models are equivalent to solving the combinatorial Weighted Maximum Coverage problem. Interestingly, there exists classes of group-based models that admit tractable projections that can be computed via dynamic programming in \( O(Nk) \) time, where \( k \) is the maximum number of allowed groups in the support of the solution [3]. In [11], another dynamic program for rooted connected subtree projections has been proposed that achieves a computational complexity of \( O(Nk) \). The projection step in our algorithm uses the results of [3].

Typically the adjacency matrices of expander graphs are constructed probabilistically [6] leading to sparse sketching matrices with the desired optimal number of rows, \( m = O(k \log(N/k)/\epsilon) \) for \( \epsilon > 0 \). It has also been shown in [21] that deterministic constructions require sub-optimal number of rows, \( m = O(d^2k^{1+\alpha}) \) for \( \alpha > 0 \). The recent work [24] on model-based RIP-1 matrices (in the for all case) also relied on a probabilistic construction. We also show that it is possible to probabilistically construct more general model expanders. It is important to point out that the first work to use model-based sparse matrices is [31]. Their signal models are similar to those considered in [24] and they achieved \( \ell_2/\ell_2 \) error guarantees, but only for the for each case.

Outline. In the next section we set our notation and give definitions of the objects under study. In Section 3 we present our algorithm and analyze its error guarantees. We provide numerical evidence in Section 4. In Section 5 we present the extensions of the existence of model-based expanders for \( D \)-ary rooted connected tree models and group-based models with overlaps, while Section 6 contains some conclusions and outlooks.

2. Preliminaries.

2.1. Notation. Throughout this paper, scalars will be denoted by lowercase letters (e.g. \( k \)), vectors by lowercase boldface letter (e.g., \( x \)), sets by uppercase calligraphic letters (e.g., \( \mathcal{S} \)) and matrices by uppercase boldface letter (e.g. \( A \)). The cardinality of a set \( S \) is denoted by \( |S| \). We use the shorthand \( [N] := \{1, \ldots, N\} \). Given a set \( S \subseteq [N] \), we denote its complement by \( \overline{S} := [N] \setminus S \) and use \( x_S \) for the restriction of \( x \in \mathbb{R}^N \) onto \( S \), i.e. \((x_S)_i = x_i \) if \( i \in S \) and 0 otherwise. We let the set of \( k \)-sparse vectors be \( \Sigma_k \). We denote bipartite graphs by \( G(U,V,E) \) where \( U \) and \( V \) are the set of left and right nodes, respectively, and \( E \) is the set of edges. A bipartite graph is left-regular if each left node has the same number of edges. We call this number the degree of the graph. Given a bipartite graph \( G(U,V,E) \) and a
set $\mathcal{K} \subset \mathcal{U}$, we use $\Gamma(\mathcal{K})$ to indicate the set of neighbors of $\mathcal{K}$, that is the right nodes that are connected to the nodes in $\mathcal{K}$. The $\ell_p$ norm of a vector $x \in \mathbb{R}^N$ is defined as $\|x\|_p := \left(\sum_{i=1}^N x_i^p\right)^{1/p}$.

2.2. Model-Based Expanders. In this paper, we consider sketching operators that are adjacency matrices of particular bipartite graphs (i.e. model-based expanders) that capture the structure of signals that belong to some pre-specified models. Following [24], we start by defining lossless expander, we then describe models of structured signals and define model-based expanders.

**Definition 2.1 (Lossless Expander).** Let $G ([N], [m], \mathcal{E})$ be a left-regular bipartite graph with $N$ left (variable) nodes, $m$ right (check) nodes, a set of edges $\mathcal{E}$ and left degree $d$. If, for any $\epsilon \in (0, 1/2)$ and any $\mathcal{K} \subset [N]$ of size $|\mathcal{K}| \leq k$, we have that $|\Gamma(\mathcal{K})| \geq (1 - \epsilon)d|\mathcal{K}|$, then $G$ is referred to as a $(k, d, \epsilon)$-lossless expander graph.

The property $|\Gamma(\mathcal{K})| \geq (1 - \epsilon)d|\mathcal{K}|$ is termed the expansion (or $\epsilon$-expansion) property and we refer to $\epsilon$ as the expansion coefficient. These expanders are lossless because $\epsilon \ll 1$.

By model-based expanders, we mean that the set $\mathcal{K}$ in Definition 2.1 has some structure, so that the expansion property must hold for a reduced number of index sets $\mathcal{K}$. Specifically, we restrict $\mathcal{K}$ to two models: a hierarchal model, where the elements of $[N]$ are organized in a tree, and a group model, where the elements of $[N]$ are grouped in $M$ groups that could have a fixed overlap.

**Definition 2.2 (k-rooted connected subtree).** Given a $D$-ary tree $T$ for an integer $D \geq 2$, an index set $\mathcal{K}$ is a $k$-rooted connected subtree of $T$, if it contains at most $k$ elements from $T$ and for each element in $\mathcal{K}$, all its ancestors with respect to $T$ are also in $\mathcal{K}$. This implies that the root of $T$ is always in $\mathcal{K}$.

**Definition 2.3 (k-tree sparse model and set).** Given a $D$-ary tree $T$ ($D \geq 2$), the $k$-tree sparse model $T_k$ is the collection of all possible $k$-rooted connected subtrees of $T$. A set $\mathcal{S} \subseteq [N]$ is $k$-tree sparse (i.e. $T_k$-sparse) if $\mathcal{S} \subseteq \mathcal{K}$ for some $\mathcal{K} \in T_k$.

The second model is based on groups of variables that should be selected together.

**Definition 2.4.** A group structure $\mathfrak{G} = \{\mathcal{G}_1, \ldots, \mathcal{G}_M\}$ is a collection of $M$ index sets, named groups, with $\mathcal{G}_j \subseteq [N]$ and $|\mathcal{G}_j| = g_j$ for $1 \leq j \leq M$ and $\cup_{\mathcal{G} \in \mathfrak{G}} \mathcal{G} = [N]$.

We say that two groups $\mathcal{G}_i$ and $\mathcal{G}_j$ overlap if $\mathcal{G}_i \cap \mathcal{G}_j \neq \emptyset$. Given a group structure $\mathfrak{G}$, it is possible to define a graph (the group-graph) whose nodes are the groups $\{\mathcal{G}_1, \ldots, \mathcal{G}_M\}$ and whose edges connect nodes $i$ and $j$ if $\mathcal{G}_i$ and $\mathcal{G}_j$ overlap. If the group-graph induced by $\mathfrak{G}$ does not contain loops, we say that $\mathfrak{G}$ is a loopless overlapping groups structure. We will consider this structure with fixed overlaps in the sequel, because it allows to exactly solve a certain signal approximation problem.

**Definition 2.5 (k-group sparse model and set).** Given a group structure $\mathfrak{G}$, the $k$-group sparse model $\mathfrak{G}_k$ is defined as the collection of all sets that are the union of at most $k$ groups from $\mathfrak{G}$. A set $\mathcal{S} \subseteq [N]$ is $k$-group sparse (i.e. $\mathfrak{G}_k$-sparse) if $\mathcal{S} \subseteq \mathcal{K}$ for some $\mathcal{K} \in \mathfrak{G}_k$.

We will use $\mathcal{M}_k$ to jointly refer to the $T_k$ and $\mathfrak{G}_k$ models. Next we define model-sparse vectors/signals.

**Definition 2.6 (model-sparse vectors).** Given a model $\mathcal{M}_k$, a vector $x \in \mathbb{R}^N$ is $\mathcal{M}_k$-sparse if $\text{supp}(x) \subseteq \mathcal{K}$ for some $\mathcal{K} \in \mathcal{M}_k$.

The following concept of model expander hinges on the previous definitions.

**Definition 2.7 (Model-expander).** Let $G ([N], [m], \mathcal{E})$ be a left-regular bipartite graph with $N$ left (variable) nodes, $m$ right (check) nodes, a set of edges $\mathcal{E}$ and left degree $d$. If, for any $\epsilon \in (0, 1/2)$ and any $\mathcal{S} \subseteq \mathcal{K}$ with $\mathcal{K} \in \mathcal{M}_k$, we have $|\Gamma(\mathcal{S})| \geq (1 - \epsilon)d|\mathcal{S}|$, then $G$ is called a $(k, d, \epsilon)$-model expander graph.

In a slight abuse of terminology, we refer to the adjacency matrices of model expander graphs as model-expanders. As noted in [24], it is straightforward to see that model-expanders do satisfy RIP-1, [7]. Precisely, we say a model-expander, $A$ satisfies model RIP-1 if for $\epsilon > 0$

$$
(1 - \delta_{\mathcal{M}_k})d\|x\|_1 \leq \|Ax\| \leq d\|x\|_1, \quad \text{for all } \mathcal{M}_k\text{-sparse vectors } x \in \mathbb{R}^N,
$$

(2.1)

where $A$ is the adjacency graph of a $(k, d, \epsilon)$-model expander graph. Since for the models we consider $\mathcal{M}_k \subseteq \mathcal{S}_k$, then by the monotonicity of RICs, $\delta_{\mathcal{M}_k} \leq \delta_k$. Since our model-expander matrices satisfy (2.1) we will also be referring to then as $\mathcal{M}'_k$-RIP-1 matrices as in [24].

2.3. Motivation. Rooted connected tree models arise in many applications ranging to image compression and denoising via wavelet coefficients to bioinformatics, where one observes hierarchical organization of gene networks.
to deep learning architectures with hierarchies of latent variables to topic models that naturally present a hierarchy (or hierarchies) of topics (see, for example, [27] for a more extensive list of applications and references). Another new application of RC-tree models is in Earth Mover Distance (EMD) sparse recovery [23].

Block model and overlapping groups models have been advocated for several applications that include detecting genetic pathways for diagnosis or prognosis of cancer [25], background-subtraction in videos [14], mental state decoding via neuroimaging [20], to mention just a few. Loopless pairwise overlapping groups (LOPG) models have been identified in [3] as a class of group models that allows to find group-sparse signal approximations in polynomial time. An example of such structure is given by defining groups over a wavelet tree containing a node and all its direct children. These groups will overlap pairwise on only one element and the corresponding group graph does not contain loops. They can be used to approximate signals which have significant features at different locations and at different scales.

3. Model-expander Iterative Hard Thresholding (MIHT).

3.1. Algorithm Overview. The model-based iterative hard thresholding (MIHT) algorithm is a hybrid of the expander iterative hard thresholding (EIHT) algorithm proposed in [17] (which is a modification of the sparse matching pursuit (SMP) algorithm [8]) and projections onto the considered models using the results in [3] and [11]. Let the compressive measurements be given as $y = Ax + e$ for a signal of interest $x \in \mathbb{R}^N$ where the noise vector $e \in \mathbb{R}^m$ characterizes the measurement error. Typically initializing $x^0 = 0 \in \mathbb{R}^N$, the SMP iterates the following

$$
\begin{align*}
    x^{n+1} &= H_k \left( x^n + H_{2k} \left[ M \left( y - Ax^n \right) \right] \right) \\
\end{align*}
$$

where $H_s(u)$ is the hard thresholding operator which retains the $s$ largest in magnitude components of $u$ and sets the rest to zero and $M$ is the median operator which is nonlinear and defined componentwise as follows

$$
\begin{align*}
    [M(u)]_i := \text{median} \left[ u_j, j \in \Gamma(i) \right] \quad \text{for} \quad u \in \mathbb{R}^m \quad \text{and} \quad i \in [N].
\end{align*}
$$

SMP was modified to become the expander iterative hard thresholding (EIHT) algorithm in [17]. Precisely, the EIHT usually initializes $x^0 = 0 \in \mathbb{R}^N$ and iterates

$$
\begin{align*}
    x^{n+1} &= H_k \left[ x^n + M \left( y - Ax^n \right) \right].
\end{align*}
$$

The stability and robustness to noise of EIHT was shown in [17]. We adapt EIHT and call it the model expander iterative hard thresholding (MIHT) algorithm for the purpose of sparse recovery with model based expanders. In MIHT we replace the hard thresholding operator, $H_s(b)$, by the projector $P_{M_s}(b)$, which projects $b$ onto our model of size $s$, $M_s$ (see Section 3.3 for more details). The pseudo-code of MIHT is given in Algorithm 1.

**Algorithm 1** Model-expander Iterative Hard Thresholding (MIHT)

Input: Measurement matrix $A$, observations $y$ and noise vector $e$

Output: $k$-sparse approximation $\hat{x}$ of the original signal $x$

Initialization:
- Set $x^0 = 0$

Iteration: While some stopping criterion is not true
- Compute $x^{n+1} = P_{M_s} [x^n + M \left( y - Ax^n \right)]$

Return $\hat{x} = x^{n+1}$

Remark 1. SMP, EIHT and MIHT could be regarded as projected gradient descent algorithms, this is because the median operator behaves approximately like the adjoint of $A$. Essentially, these algorithms do a gradient descent by applying the median and then project onto the feasible set using the hard threshold operator or the projector onto the model.

3.2. Complexity and Convergence. The complexity of MIHT depends on the cost of the median operation and the projection onto our model sets. The following proposition gives the complexity of MIHT.

**Proposition 3.1.** The runtime of MIHT is $O(kN)$ where $k$ is the model size and $N$ is the dimension of the signal.
Proof. At each iteration MIHT performs $N$ median operations of $d$ elements which involve sorting ($O(d \log d)$ complexity) and a projection step that costs $O(k N)$. A detailed discussion on the projection step and its cost is given in Section 3.3. We see that the projection step dominates since $k = \omega(\log N)$ and $d$ may be fixed to a constant or assume the lower bound of $d = \log_k N$ which depends on some assumptions on $m$ \cite{t}. In either case $O(d \log d) = O(k)$. \qed

The convergence analysis of MIHT follows similar arguments as the proof of the linear time convergence of EIHT given in \cite{17}. The key difference between the analysis of EIHT and MIHT is the projections \cite{6} and $P_{M_k}(\cdot)$ respectively. We leverage recent results in \cite{3} to perform the projection $P_{M_k}(\cdot)$ exactly and efficiently in linear time. The projections are analysed in Section 3.3. Theorem 3.2 and Corollary 3.3 below bound the error of the output of the MIHT algorithm.

**Theorem 3.2.** Let $M_k$ be a model and let $S \in M_k$-sparse. Consider $A \in \{0, 1\}^{m \times n}$ to be the adjacency matrix of a model-expander for $M_k$ with $\epsilon_{M_k} < 1/12$. For any $x \in \mathbb{R}^N$ with $\text{supp}(x) = S$ and $e \in \mathbb{R}^n$, the sequence of updates $(x^n)$ of MIHT with $y = Ax + e$ satisfies, for any $n \geq 0$,

$$
\|x^n - x_S\|_1 \leq \alpha^n \|x^0 - x_S\|_1 + \beta \|Ax_S + e\|_1,
$$

(3.4)

where $\alpha < 1$ and $\beta$ depends only on $\epsilon_{M_k}$.

**Corollary 3.3.** If the sequence of updates $(x^n)$ for any $n$ converges to $x^*$ as $n \to \infty$, then

$$
\|x^* - x_S\|_1 \leq C_1 \sigma_{M_k}(x)_1 + C_2 \|e\|_1,
$$

(3.5)

for some constants $C_1, C_2 > 0$ depending only on $\epsilon_{M_k}$.

We discuss the projections onto our models first (in the next subsection) before we provide the proofs of Theorem 3.2 and Corollary 3.3 in the appendix.

### 3.3. Projections

The projections for the rooted connected tree model and the loopless overlapping groups model can be computed in polynomial time via dynamic programming, leveraging the results in \cite{3} and \cite{11}. We define the projections in the $\ell_1$ norm as

$$
P_{M_k}(x) \in \text{argmin}_{z \in M_k} \{\|x - z\|_1\},
$$

(3.6)

where $M_k$ is either the $k$-sparse RC-tree model or the $k$-group sparse model, see definitions in Section 2.

In \cite{3}, it was shown that projections in $\ell_2$ norm onto group-based models, which include the rooted connected tree model, can be reformulated as solving a discrete support selection problem that corresponds to the combinatorial Weighted Maximum Coverage (WMC) problem. Since the WMC is NP-hard in general, these projections are NP-hard too. However, \cite{3} identified group-structures that lead to tractable solutions. These structures correspond to the models we consider in this paper.

Two dynamic programs were recently proposed \cite{3,11} for computing the projections in the $\ell_2$ norm for the loopless overlapping groups model and the rooted connected tree model. The dynamic programs gradually explore a graph defined by the considered model and recursively compute the optimal solution. In order to adapt the dynamic programs to projections in the $\ell_1$ norm, it is sufficient to replace the weights of the nodes of the graphs.

For the RC-tree model, the graph is the given tree where each node correspond to a variable. For the projection in $\ell_2$ norm, the weight of each node is the square of the corresponding component of the signal to be projected, while for the $\ell_1$ norm the weights are given by the absolute values. The tree is explored from the leaves upward, updating at each node a table containing the optimal selection of $1$ to $k$ nodes from its subtrees. More specifically, consider being at a particular level in the tree. For each node in this level, we store a table that contains the optimal values for choosing $1$ to $k$ connected elements from its subtree. When we move one level up, the table for a node in the new level is computed by comparing the tables of its subtrees, considering one subtree at a time, from right to left. The computational complexity has been determined \cite{11} to be $O(ND^2k)$, where $D$ is the tree degree ($D = 2$ for binary trees).

For the case of loopless overlapping groups, \cite{3} considered the group-graph induced by the group structure, for which a node corresponds to an entire group of variables. The weight associated to each node is evaluated dynamically during the exploration of the group-graph: it is the sum of the weights of the variables included in that node that do not...
belong to an already selected group. For the $\ell_2$ norm, the weight associated to each variable correspond to its square value, while for the $\ell_1$ norm, the weight is its absolute value. By defining an appropriate graph exploration rule, it has been shown that the dynamic program requires at most $O(M^2k)$ operations to solve (3.6) for the generalized loopless overlapping groups model.

**Remark 2 (Generalization).** In [3], the overlapping groups model has been generalized allowing for within-group sparsity and by introducing an overall sparsity budget $K$ together with the group budget $k$. A dynamic program solves the projection exactly in $O(M^2K^2k)$ operations.

**Remark 3 (Relaxations).** Performing the exact projections might not always be computationally feasible. In these cases, it is possible to relax the constraints on the sparsity or the number of groups into regularization terms. In [3], it has been shown that these relaxations, for the RC-tree model and the loopless overlapping groups model, lead to binary linear programs with totally unimodular constraints that can be efficiently solved with standard linear program solvers. Furthermore, if the computed projection has model parameter $k$, it will be the exact projection. However, it may not be possible to always find a $k$-model sparse solution via relaxation for any value of $k$ [3].

4. Experimental Results.

4.1. Block sparsity. In the first experiment, we consider recovering a block-sparse signal, where the $M$ equal-size blocks define a partition of the set $\{1, \ldots, N\}$, so that the block size $g$ is given by $g = N/M$. We want to show that the number of measurements $m$ required for exact recovery in the noiseless case does not depend upon the ambient dimension $N$. Recall that our bound for group-sparse recovery (5.6) gives $m \geq ed k (g + \log (M/k))/\epsilon$. We therefore set $kg = C_1$ and $k \log (M/k) = C_2$, which combined with $g = N/M$, implies $k = O(1/\log (N))$, $g = O(\log (N))$ and $M = O(N/\log (N))$. We set $N = 2^7, \ldots, 2^{13}$, $M = \lfloor N/\log_2 (N)\rfloor$, $k = \lfloor 42/\log_2 (N)\rfloor$ and $g = \lfloor N/M\rfloor$. We vary the number of samples $m$ and compute the recovery error $\|\hat{x} - x\|_1$, where $\hat{x}$ is the signal estimated either with the EIHT or the MIHT algorithms and $x$ is the true signal. For each problem size $N$, we repeat the experiment 100 times. For each repetition, we randomly select $k$ blocks from $M$ and set the components of the true signal $x$ as identical and independent draws from a standard Gaussian distribution. The other components are set to zero. For each number of samples, we randomly draw a sparse sketching matrix $A \in \mathbb{R}^{m \times N}$ with $d = 6$ ones per column. The columns of $A$ are normalized to have unitary $\ell_1$ norm. We then create the measurement vector $y$ as $y = Ax$, from which we use the EIHT or the MIHT algorithm to recover $x$. We record the minimum number of sample $m^*$ that yields a median recovery error below $10^{-4}$. In the left panel of Figure 4.1, we plot $m^*$ as a function of $\log(N)$. As expected, the line for MIHT is almost constant with respect to $N$, while the line for EIHT grows linearly.

4.2. Tree sparsity. In the second experiment, we try to recover a tree-sparse signal $x \in \mathbb{R}^N$ with $k = 16$ nonzero components that lie in a rooted connected subtree of a regular binary tree. The nonzero components are randomly drawn from a standard Gaussian distribution. We randomly draw a sparse sketching matrix $A \in \mathbb{R}^{m \times N}$ with $d = 6$ ones per column. The columns of $A$ are normalized to have unitary $\ell_1$ norm. We use the EIHT and MIHT
algorithms to obtain an approximation \( \hat{x} \) from the linear sketch \( Ax \). For each value of \( N \in \{2^7, 2^8, \ldots, 2^{13}\} \), we vary the sketch length, \( m \in [2k, 10k \log_2 N] \) and compute an approximation \( \hat{x} \) with both algorithms, recording the recovery error \( \| x - \hat{x} \|_1 \). We repeat the experiment 10 times with a different draw of the signal \( x \) and compute the median of the recovery errors for both algorithms. Finally, we record the minimum number of measurements that yield a median recovery error less than \( 10^{-4} \). The results are present in the right panel of Figure 4.1. The plot clearly shows that the minimum sketch length for MIHT is approximately constant with respect to \( N \) as opposed to the increasing one for the standard EIH.

5. Existence of model-expanders. For practical applications of model expanders we need to be able to construct these objects. The key goal is to construct them with as small parameters \( (d, m) \) as possible. Bounds, both lower and upper, on \( m \) for model-expanders, for binary tree-sparse and block-sparse models, were derived in [24]. The derivation of the lower bounds depended mainly on RIP-1, since the sparsification technique they employed relies on the RIP-1 of these matrices. Therefore, the extension of these two models to \( D \)-ary tree-sparse and fixed overlapping tree-sparse models respectively, which we consider in this manuscript, also have these lower bounds. We skip the explicit derivation because this will be identical to the derivation in [24].

However, since the upper bounds involve the enumeration of the cardinality of the models which is different from [24], we explicitly show the derivations below. In essence, we will show the existence of model-expanders by random (probabilistic) construction. Precisely, for every vertex, \( i \), in the set of left vertices \( [N] \), i.e. \( i \in [N] \), we sample with replacement \( d \) vertices in the set of right vertices, \([m]\), and then we connect these \( d \) nodes to \( i \). Each \( d \)-subset of \([m]\) are sampled uniformly and independently of other sets. This leads to a standard tail bound whose proof uses a Chernoff bound argument. However, in our derivation we will use the following variant of this standard tail bound proved and used by [9].

**Lemma 5.1.** [9] There exist constants \( C > 1 \) and \( \mu > 0 \) such that, whenever \( m \geq C \frac{dt}{e} \), one has for any \( T \subseteq [N] \) with \( |T| = t \)

\[
\text{Prob}\left[ |\{ j \in [m] \mid \exists i \in T \text{ and } e_{ij} \in \mathcal{E}\} | < (1 - \epsilon)dt \right] \leq \left( \mu \cdot \frac{em}{dt} \right)^{-\epsilon dt} \tag{5.1}
\]

5.1. \( D \)-ary tree model. \( D \)-ary tree model-expanders or \( T_k^\epsilon \)-RIP-1 matrices are \( m \times N \) sparse binary matrices with \( d \) ones per column. The relation between the \([N]\) indices of the columns of these matrices is modeled by a \( D \)-ary tree. Theorem 5.2 states the existence and the sizes of the parameters of these matrices.

**Theorem 5.2.** For \( \epsilon \in (0, 1/2) \) and \( k = \omega (\log N) \) there exists a \( T_k^\epsilon \)-RIP-1 matrix with

\[
d = O\left( \frac{\log (N/k)}{\epsilon \log \log (N/k)} \right) \quad \text{and} \quad m = O\left( \frac{dk}{\epsilon} \right). \tag{5.2}
\]

**Proof.** For a model \( T_k \subseteq \Sigma_k \) and \( t \in [k] \) denote \( T_k \)-sparse sets of size \( t \) as \( T_{k,t} \) and denote the number of \( D \)-ary rooted connected trees with \( k \) nodes as \( T_k \). We use the following estimate of the number of sets \( T_{k,t} \).

**Lemma 5.3.** For all \( t \in [k] \), \( |T_{k,t}| \leq \min \left( T_k \cdot \left( \frac{k}{t}\right) \cdot \left( \frac{N}{t}\right) \right) \) and \( T_k = \frac{1}{(D - 1)k + 1} \binom{Dk}{k} \).

**Proof.** It is sufficient to know that the \( T_k \) are the Pfaff-Fuss-Catalan numbers or \( k \)-Raney numbers [28, Note 12], which enumerates the total number of ordered, rooted \( D \)-ary trees of cardinality \( k \).

Using the fact that \( (\frac{k}{t})^y \) is bounded above by \( (\frac{ek}{t})^y \), we have for all \( t \in [k] \)

\[
|T_{k,t}| \leq \min \left[ \frac{(eD)^k}{(D - 1)k + 1} \left( \frac{ek}{t}\right)^t \left( \frac{eN}{t}\right)^t \right]. \tag{5.3}
\]

Using a union bound over the number of \( T_{k,t} \) we see that the probability of Lemma 5.1 goes to zero with \( N \) if

\[
\forall t \in [k] \quad \left| T_{k,t} \right| \cdot \left( \mu \cdot \frac{em}{dt} \right)^{-\epsilon dt} \leq \frac{1}{N},
\]
where \( m \) satisfies the bound in the lemma. Let \( t^* = \frac{k \log(e D) - \log(D k - D + 1)}{\log(N/k)} \). By simple algebra we see that at \( t = t^* \)
the two quantities in the square brackets of (5.3) are equal; and \( \forall t \in [1, t^*] \) the second quantity is smaller while \( \forall t \in [t^*, k] \) the first quantity is smaller. Therefore, we split the domain of \( t \) into two intervals and prove the following.

\[
\forall t \in [1, t^*] \quad \left(\frac{e N}{t}\right)^t \left(\frac{e k}{t}\right)^{e d t} \leq \frac{1}{N},
\]

\[
\forall t \in [t^*, k] \quad \left(\frac{(e D)^k}{(D-1)k + 1}\right) \left(\frac{e k}{t}\right)^t \left(\frac{e m}{d t}\right)^{-e d t} \leq \frac{1}{N}.
\]

In accordance with Theorem 5.2, we let \( \delta = C' \frac{\log(N/k)}{\log \log(N/k)} \) and \( m = C'' \frac{dk}{\epsilon} \). Note that the left hand side of inequalities (5.4) and (5.5) are log-convex whenever \( d > 1/\epsilon \). We therefore need to check the conditions only at the end points of the intervals, i.e. for \( t = 1, t^*, k \). For \( t = 1 \), inequality (5.4) becomes

\[
e N \left(\frac{C'' \mu k}{2}\right)^{\frac{C' \log(N/k)}{\log \log(N/k)}} \leq \frac{1}{N}.
\]

This inequality holds for sufficiently large \( C' \) and \( C'' \) given that \( k = \omega(\log N) \). Thus (5.4) holds for \( t = 1 \). When \( t = t^* \), we have the left hand side of inequality (5.4) equal to the left hand side of inequality (5.5). Substituting the value of \( t^* \) and those of \( m \) and \( d \) in the left hand side of (5.4) gives

\[
\left(\frac{e N \log(N/k)}{k \log(e D) - \log(D k - D + 1)}\right)^{\frac{k \log(e D) - \log(D k - D + 1)}{\log(N/k)}} \left(\frac{C'' \mu k}{2}\right)^{\frac{C' \log(N/k)}{\log \log(N/k)}} \leq \frac{1}{N}.
\]

The above quantity is less than \( 1/N \) for sufficiently large \( C' \) and \( C'' \) given that \( k = \omega(\log N) \). Thus (5.4) and (5.5) hold for \( t = t^* \). For \( t = k \), with the values of \( m \) and \( d \), the left hand side of (5.4) becomes

\[
\left(\frac{(e^2 D)^k}{(D-1)k + 1}\right) \left(\frac{C'' \mu k}{2}\right)^{\frac{C' \log(N/k)}{\log \log(N/k)}} \leq \frac{1}{N}
\]

This quantity is far less than \( 1/N \) for sufficiently large \( C' \) and \( C'' \) given that \( k = \omega(\log N) \). Thus (5.5) hold for \( t = k \) and this concludes the proof. \( \square \)

**5.2. Overlapping group model.** Fixed overlapping group model-expanders or \( G_k \)-RIP-1 matrices are also \( m \times N \) sparse binary matrices with \( d \) ones per column and the relationship between the indices of their columns is modeled by an overlapping group structure. Let the number of groups be \( M \) and the size of the overlap be \( L \). Theorem 5.4 states the existence and the sizes of the parameters of these matrices.

**Theorem 5.4.** For \( \epsilon \in (0, 1/2) \), the number of groups, \( M \geq 2 \), the maximum size of groups \( g_{\max} = \omega(\log N) \), and the ambient dimension \( N > kg_{\max} \), there exist a \( G_k \)-RIP-1 matrix with

\[
d = \mathcal{O} \left(\frac{\log k N}{\epsilon}\right) \quad \text{and} \quad m = \mathcal{O} \left(\frac{dk}{\epsilon}\right).
\]

**Proof.** We begin by bounding the number of subsets of cardinality \( t \) of a \( k \)-group-sparse set.

\[
\forall t \in [kg_{\max}], \quad |\Phi_{k,t}| \leq \min \left[\binom{M}{k}, \binom{kg_{\max}}{t}, \binom{N}{t}\right]
\]

Since \( \binom{a}{b} \) is bounded above by \( \left(\frac{e a}{b}\right)^b \), we have for all \( t \in [kg_{\max}] \)

\[
|\Phi_{k,t}| \leq \min \left[\left(\frac{e M}{k}\right)^k \frac{e k g_{\max}}{t}, \frac{e N}{t}\right]
\]

Using a union bound over the sets \( \Phi_{k,t} \), we see that the probability of Lemma 5.1 goes to zero with \( N \) if

\[
\forall t \in [kg_{\max}], \quad |\Phi_{k,t}| \cdot \left(\frac{e m}{d t}\right)^{-e d t} \leq \frac{1}{N},
\]
where $m$ satisfies the bound in the lemma. By simple algebra we see that at $t = t^* = \frac{k \log(e M/k)}{\log(N/(k g_{\text{max}}))}$, the two quantities in the square brackets of (5.7) are equal; and $\forall t \in [1, t^*]$ the second quantity is smaller, while $\forall t \in [t^*, k g_{\text{max}}]$ the first quantity is smaller. Therefore, we split the domain of $t$ into two intervals and prove the following.

$$
\forall t \in [1, t^*] \quad \left( \frac{eN}{t} \right)^t \left( \mu \cdot \frac{em}{dt} \right)^{-\epsilon dt} \leq \frac{1}{N},
$$

(5.9)

$$
\forall t \in [t^*, k g_{\text{max}}] \quad \left( \frac{eM}{k} \right)^{k} \left( \frac{ek g_{\text{max}}}{t} \right)^t \left( \mu \cdot \frac{em}{dt} \right)^{-\epsilon dt} \leq \frac{1}{N}.
$$

(5.10)

It is easy to check that, if $d > 1/\epsilon$, the left hand-sides of (5.9) and (5.10) are log-convex. Therefore, it is sufficient to check (5.9) and (5.10) for $t = 1, t^*, k g_{\text{max}}$. The rest of the proof follows the above proof for $D$-ary tree models if we choose $d = C'' \log_k(N)$, $m = C'' dk/\epsilon$, $k = \omega(\log N)$ and sufficiently large $C''$ and $C''$. We skip the rest of the details for space purposes.

For the extra conditions, firstly we need $t^* > 0$ which leads to a condition that $N > g_{\text{max}}$. This is necessary in order not have just one large group. Secondly, we need $t^* \leq k g_{\text{max}}$ which implies that $\log(e M/k) \leq g_{\text{max}} \log(N/(k g_{\text{max}}))$ and this inequality holds true if $g_{\text{max}} = \omega(\log N)$. This means $g_{\text{max}}$, the maximum size of groups, cannot be too small. 

**Remark 4.** The block structured model is a special case of the overlapping group model, when the groups do not overlap. Usually in this setup the blocks are assumed to be of the same length which is equivalent to setting $g_1 = g_2 = \cdots = g_M = N/M =: g$. The above analysis on overlapping groups carries through to this particular case in a very straightforward way. The fixed overlapping group model is another special case of the more general overlapping groups model proven above.

### 6. Conclusions

In this paper, we consider signals that belong to some general structured sparsity models, such as rooted connected trees or overlapping groups models. We focus on the problem of recovering such signals from noisy low-dimensional linear sketches obtained from sparse measurement matrices. We present the first polynomial time recovery algorithm with noisy low-dimensional linear sketches obtained from sparse measurement matrices. We present the first polynomial time recovery algorithm with noisy low-dimensional linear sketches obtained from sparse measurement matrices.

### REFERENCES


Now let the set of edges between $S$ such that the action of $A$ problem of our algorithm involving the median operation and it shows that the median operator approximately inverts the lemma to be used in this proof. But the proof of this lemma requires the following proposition that states a property of left regular bipartite graphs, which can be said to be equivalent to the RIP-1 condition [7].

**Proof.** By the expansion property of the graph $|\Gamma(S)| \geq (1 - \epsilon_{M_2})d|S|$ and remember we defined $\Gamma'(\cdot)$ and $\Gamma''(\cdot)$ such that $|\Gamma(S)| = |\Gamma'(S)| + |\Gamma''(S)|$ which therefore means

$$|\Gamma'(S)| + |\Gamma''(S)| \geq (1 - \epsilon_{M_2})d|S|. \quad (7.1)$$

Now let the set of edges between $S$ and $\Gamma(S)$ be denoted by $E(S, \Gamma(S))$. Hence $|E(S, \Gamma(S))|$ is the number of edges between $S$ and $\Gamma(S)$ and by enumerating these edges we have

$$|\Gamma'(S)| + 2|\Gamma''(S)| \leq |E(S, \Gamma(S))| = d|S|. \quad (7.2)$$

Solving (7.1) and (7.2) simultaneously gives $|\Gamma''(S)| \leq \epsilon_{M_2}d|S|$ and thus $|\Gamma'(S)| \geq (1 - 2\epsilon_{M_2})d|S|$. Substituting $|S|$ by $s$ completes the proof. \(\square\)

Now we state the lemma which is the key ingredient to the proof of Theorem 3.2. This lemma is about the subproblem of our algorithm involving the median operation and it shows that the median operator approximately inverts the action of $A$ on sparse vectors.
Lemma 7.2 ([17 Lemma 14.16]). Let \( A \in \{0, 1\}^{m \times N} \) be the adjacency matrix of a model expander with an expansion coefficient of order \( k \), \( \epsilon_{M_k}, \) satisfying \( 4\epsilon_{M_k}d < d + 1 \). If \( S \subset [N] \) such that \( |S| \leq k \), then

\[
\| [M(Ax_S + e) - x]_S \|_1 \leq \frac{4\epsilon_{M_k}}{1 - 4\epsilon_{M_k}} \| x_S \|_1 + \frac{2}{(1 - 4\epsilon_{M_k})d} \| e_{\Gamma(S)} \|_1,
\]

for all \( x \in \mathbb{R}^N \) and \( e \in \mathbb{R}^m \).

Proof. Note that for a permutation \( \pi : [m] \to [m] \) such that \( b_{\pi(1)} \geq b_{\pi(2)} \geq \ldots \geq b_{\pi(m)} \), we can rewrite the median operation of (3.2) as

\[
[M(u)]_i := q_2 [u_j, j \in \Gamma(i)] \quad \text{for } u \in \mathbb{R}^m \text{ and } i \in [N]
\]

where \( q_2 \) is a shorthand for the quantile operator that is true of more general quantile operators. This can be bound as in (7.4) below using the following property of the median operator, which is true of more general quantile operators:

\[
|q_2 [b_1, \ldots, b_m]| \leq q_2 [|b_1|, \ldots, |b_m|] .
\]

Therefore, we have

\[
\| [M(Ax_S + e) - x]_S \|_1 \leq \sum_{i \in S} q_2 \left( \| (Ax_S)_j + e_j - x_i \|, j \in \Gamma(i) \right) = \sum_{i \in S} q_2 \left( \sum_{l \in S \setminus \{i\}} x_l + e_j, j \in \Gamma(i) \right)
\]

where the last equality due to the property of the underlying expander graph of \( A \), say \( G \), and remember \( e_{l,j} \in E \) signifies an edge from node \( l \) to node \( j \) being in \( E \), the set of edges of \( G \). Now we proceed by induction on the cardinality of \( S \) i.e. \( s = |S| \) to show that the right hand side of (7.3) upper bounds (7.4).

Base case: If \( s = 1 \) that is \( S = \{i\} \) for some \( i \in S \) then there is no \( l \in S \setminus \{i\} \). Using this other property of the median operator, which is also true for more general quantile operators:

\[
q_2 [b_1, \ldots, b_m] \leq \frac{2 (b_1 + \ldots + b_m)}{m}, \quad \text{if } b_j \geq 0 \quad \text{for all } j,
\]

we have the following stronger bound.

\[
\sum_{i \in S} q_2 \left( \sum_{l \in S \setminus \{i\}} x_l + e_j, j \in \Gamma(i) \right) = q_2 \left( |e_j|, j \in \Gamma(i) \right) \leq \frac{2}{d} \| e_{\Gamma(i)} \|_1.
\]

Inductive step: We assume that the induction statement holds for \( |S| = s - 1 \) for \( s \geq 2 \) and proceed to show that it also holds for \( |S| = s \). By the pigeonhole principle, Proposition 7.4 implies that there exists a variable node \( \gamma \in S \) which is connected uniquely to at least \((1 - 2\epsilon_{M_s})d\) check nodes. This holds for the model-expanders since \( \epsilon_{M_s} \geq \epsilon_{M_s} \). Let \( \gamma = |\Gamma'(|\gamma|)| \geq (1 - 2\epsilon_{M_s})d \). This means that there are at most \( d - \gamma \leq 2\epsilon_{M_s}d \) in \( \Gamma'(|\gamma|) \). By the definition of \( q_2 \), there exist \([d/2]\) distinct \( j_1, \ldots, j_{[d/2]} \in \Gamma'(|\gamma|) \) such that, for all \( h \in \{1, \ldots, [d/2]\} \),

\[
q_2 \left[ \sum_{l \in S \setminus \{i\}} x_l + e_j, j \in \Gamma(i) \right] \leq \sum_{l \in S \setminus \{i\}} x_l + e_{j_h}.
\]

At least \([d/2] - (d - \gamma) \leq [d/2] - 2\epsilon_{M_s}d\) elements among \( j_1, \ldots, j_{[d/2]} \) are in \( \Gamma'(|\gamma|) \). Noting that there are no \( l \in S \setminus \{i\} \) with \( e_{l,j_h} \), we take the mean over these \( j_h \) elements and obtain

\[
q_2 \left[ \sum_{l \in S \setminus \{i\}} x_l + e_j, j \in \Gamma(i) \right] \leq \frac{1}{[d/2] - (d - \gamma)} \| e_{\Gamma'(|\gamma|)} \|_1 \leq \frac{2}{(1 - 4\epsilon_{M_s})d} \| e_{\Gamma'(|\gamma|)} \|_1.
\]
If on the other hand we define \( R := S \setminus \{i^*\} \) and if \( i \in R \), we have
\[
\left| \sum_{l \in S \setminus \{i\}} x_l + e_j \right| = \sum_{l \in R \setminus \{i\}} x_l + 1_{\{e_{i^*} \in E\}} x_{i^*} + e_j.
\]
(7.9)

We proceed inductively by replacing \( S \) by \( R \) and \( e_j \) by \( e_j' = 1_{\{e_{i^*} \in E\}} x_{i^*} + e_j \) (using the fact that \( \epsilon_{s-1} \leq \epsilon_{M^s} \)) to get the following upper bound.
\[
\sum_{i \in R} q_2 \left[ \sum_{l \in S \setminus \{i\}} x_l + e_j \right], \quad j \in \Gamma(i) \leq \frac{4\epsilon_{M^s} d}{d - 4\epsilon_{M^s} d} \|x_S\|_1 + \frac{2}{(1 - 4\epsilon_{M^s})} \|e'_{\Gamma(R)}\|_1,
\]
(7.10)

where \( e' \) is a vector with components \( e'_j \) for \( j = 1, \ldots, m \). Next we upper bound \( \|e'_{\Gamma(R)}\|_1 \) as follows
\[
\|e'_{\Gamma(R)}\|_1 \leq \sum 1_{\{e_{i^*} \in E\}} |x_{i^*}| + \|e_{\Gamma(R)}\|_1 \leq 2\epsilon_{M^s} d |x_{i^*}| + \|e_{\Gamma(R)}\|_1,
\]
(7.11)

where we used the fact that
\[
\sum_{j \in \Gamma(i)} 1_{\{e_{i^*} \in E\}} = \sum_{j=1}^m 1_{\{e_{i^*} \in E\}} \text{ and } e_{i^*} \in E \text{ for some } i \in R
\]
(7.12)

and
\[
\sum_{j \in \Gamma(i^*)} 1_{\{e_{i^*} \in E\}} \text{ for some } i \in R = |\Gamma''(i^*)| \leq 2\epsilon_{M^s} d.
\]
(7.13)

Remember we are trying to bound \( \sum_{i \in S} q_2 \left[ \sum_{l \in S \setminus \{i\}} x_l + e_j \right], \quad j \in \Gamma(i) \) of (7.8) which is the sum of the two cases leading to (7.8) and (7.10), that is the following sum:
\[
q_2 \left[ \sum_{l \in S \setminus \{i^*\}} x_l + e_j \right] + \sum_{i \in R} q_2 \left[ \sum_{l \in S \setminus \{i\}} x_l + e_j \right], \quad j \in \Gamma(i).
\]
(7.14)

Now using the bounds in (7.8) and (7.10) with (7.11) we upper bound (7.14) by
\[
\frac{2}{(1 - 4\epsilon_{M^s})} \|e_{\Gamma''(i^*)}\|_1 + \frac{4\epsilon_{M^s} d}{d - 4\epsilon_{M^s} d} \|x_S\|_1 + \frac{2}{(1 - 4\epsilon_{M^s})} \|e_{\Gamma(R)}\|_1 \leq \frac{4\epsilon_{M^s}}{1 - 4\epsilon_{M^s}} \|x_S\|_1 + \frac{2}{(1 - 4\epsilon_{M^s})} \|e_{\Gamma(S)}\|_1.
\]
(7.15)

The upper bound (7.15) was in view of the fact that \( \Gamma(R) \) and \( \Gamma''(i^*) \) are two disjoint subsets of \( \Gamma(S) \) and this completes the proof. \( \Box \)

### 7.2. Proof of Theorem 3.2

**Proof.** Now we use Lemma 7.2 to prove Theorem 3.2. In order to show that (3.4) holds we show that the (7.16) below holds and by induction this leads to (3.4).
\[
\|x^{n+1} - x_S\|_1 \leq \alpha \|x^n - x_S\|_1 + (1 - \alpha)\beta \|Ax_S + e\|_1.
\]
(7.16)

Since we are able to do exact projections (in the \( \ell_1 \) norm) onto the given model, we have that \( x^{n+1} \) is a better \( s \)-term approximation than \( x_S \) to \( u^{n+1} := \left(x^n + M(y - Ax^n)\right)_{Q_{n+1}}, \) where \( Q^{n+1} := S \cup \text{supp}(x^n) \cup \text{supp}(x^{n+1}). \) With this fact we use the triangle inequality to have
\[
\|x^{n+1} - x_S\|_1 \leq \|x^{n+1} - u^{n+1}\|_1 + \|u^{n+1} - x_S\|_1
\]
(7.17)
\[
\leq 2\|u^{n+1} - x_S\|_1 \leq 2\|x_S - x^n - M(A(x^n - x_S) + e')|_{Q^{n+1}}\|_1
\]
(7.18)
where \( e' := Ax_S + e \) and therefore we know that \( y = Ax + e = Ax_S + e' \). Next we apply Lemma 7.2 to the above to get the desired bound of (7.16),

\[
\|x^{n+1} - x_S\|_1 \leq \frac{8\epsilon_{M_3s}}{1 - 4\epsilon_{M_3s}} \|x_S - x^n\|_1 + \frac{4}{1 - 4\epsilon_{M_3s}} d \|e'\|_1,
\]

(7.19)

with \( \alpha := \frac{8\epsilon_{M_3s}}{(1 - 4\epsilon_{M_3s})} \) and \( \beta := \frac{4}{(1 - 12\epsilon_{M_3s})} d \). Then inductively (7.19) implies (3.4) with \( \alpha < 1 \) and \( \beta > 0 \) resulting into \( \epsilon_{M_3s} < 1/12 \). This therefore concludes the proof. \( \square \)

**7.3. Proof of Corollary 3.3**

*Proof.* If the sequence \( (x^n) \) for \( n \geq 0 \) converges to \( x^* \) as \( n \to \infty \). But as \( n \to \infty \) the first term in (3.4) goes to zero and we have

\[
\|x^* - x_S\|_1 \leq \beta \|Ax_S\|_1 + \beta \|e\|_1.
\]

Note that \( \|Ax_S\|_1 \leq d \|x_S\|_1 \) as a consequence of RIP-1 (2.1) which is true for \( A \) being an \( M_k \)-RIP-1 matrix. Therefore, we bound our error as thus

\[
\|x^* - x\|_1 \leq \|x_S\|_1 + \|x^* - x_S\|_1 \leq (1 + \beta d) \sigma_{M_1}(x)_1 + \beta \|e\|_1.
\]

With \( C_1 = 1 + \beta d \) and \( C_2 = \beta \), which completes the proof. \( \square \)