Model-based Sketching and Recovery with Expanders[*]

Bubacarr Bah, Luca Baldassarre, Volkan Cevher [†]

Abstract

Linear sketching and recovery of sparse vectors with randomly constructed sparse matrices has numerous applications in several areas, including compressive sensing, data stream computing, graph sketching, and combinatorial group testing. This paper considers the same problem with the added twist that the sparse coefficients of the unknown vector exhibit further correlations as determined by a known sparsity model. We prove that exploiting model-based sparsity in recovery provably reduces the sketch size without sacrificing recovery quality. In this context, we present the model-expander iterative hard thresholding algorithm for recovering model sparse signals from linear sketches obtained via sparse adjacency matrices of expander graphs with rigorous performance guarantees. The main computational cost of our algorithm depends on the difficulty of projecting onto the model-sparse set. For the tree and group-based sparsity models we describe in this paper, such projections can be obtained in linear time. Finally, we provide numerical experiments to illustrate the theoretical results in action.

1 Introduction

1.1 Problem overview and statement In recent years, we have witnessed the advent of a new paradigm of succinct approximate representations of signals. A signal $\mathbf{x} \in \mathbb{R}^N$ can be represented by the *measurements* or *sketch* $\mathbf{A}\mathbf{x}$, for $\mathbf{A} \in \mathbb{R}^{m \times N}$, 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 as data streaming [31, 24], compressing sensing [13, 10], graph sketching [1, 20] and even to breaking privacy of databases via aggregate queries [15].

The standard approach in linear sketching assumes that $\mathbf{x}$ is k-sparse, i.e., only k of its N components are nonzero. We then find conditions on the sketch matrix $\mathbf{A}$ and the recovery algorithm Δ to obtain an estimate

$\widehat{\mathbf{x}} = \Delta(\mathbf{A}\mathbf{x} + \mathbf{e})$, where $\mathbf{e}$ measures the perturbation of the linear model, that yields an error guarantee such as:

$$(1.1) \qquad \|\widehat{\mathbf{x}} - \mathbf{x}\|_p \leq C_1 \sigma_k(\mathbf{x})_q + C_2 \|\mathbf{e}\|_p,$$

where $\sigma_k(\mathbf{x})_q = \min_{k-\text{sparse } \mathbf{x}'} \|\mathbf{x} - \mathbf{x}'\|_q$, for some constants $C_1, C_2 > 0$ and $1 \leq q \leq p \leq 2$. We denote this type of error guarantees as ℓ_p/ℓ_q. It has been established in CS that a randomly generated $\mathbf{A}$ achieves the optimal sketch length of $\mathcal{O}\left(k \log(N/k)\right)$ and an approximate reconstruction, with error guarantee such as in (1.1), of the original signal is possible using an appropriate algorithm, such as Δ. This leads to two different probabilistic models for which (1.1) holds. The *for all* case where a randomly generated $\mathbf{A}$ and recovery algorithm Δ satisfy (1.1) *for all* $\mathbf{x} \in \mathbb{R}^N$ with high probability; and the *for each* case where a randomly generated $\mathbf{A}$, independent of $\mathbf{x}$, and recovery algorithm Δ satisfy (1.1) *for each* given $\mathbf{x} \in \mathbb{R}^N$ with high probability [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 signal's sparse coefficients. The acquisition and recovery of signals with such *a priori* known structures, like tree and block structures is referred to as model-based compressed sensing (CS) [4, 26]. Model-based CS has the advantage of reducing the minimum sketch length for which a variant of (1.1) holds. However, model-based CS with random *dense* matrices require large storage space and can lead to implementation difficulties. To overcome this computational drawback, this work, similarly to [26], focuses on sparse measurement matrices that have the so called ℓ_1-norm restricted isometry property (to be defined later), designed for model-based CS, and referred to as model-expanders.

The recent work of [26] is the first to address 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 they provided bounds for the sketch sizes. They also propose a recovery algorithm that has exponential time complexity with respect to the dimension of the signal.

1.2 Main results In this paper, we propose an efficient polynomial time algorithm for model-based CS when the measurement matrices are model-expanders for sparsity models endowed with a tractable projection operator. We show that the proposed algorithm has ℓ_1/ℓ_1 theoretical guarantees. Our theoretical results are supported with numerical experiments. In addition, this work builds on [26] to extend the class of models for which we can show existence of model-expanders that yield sampling bounds similar to those in [26]. Note that our results are for the *for all* case.

Formally, in order to recover k-model-sparse signals (to be defined in Section 2), we propose the algorithm: *Model-expander iterative hard thresholding (MEIHT)*, which is an adaptation of the *iterative hard thresholding for expanders (EIHT)* algorithm [18]. The runtime of MEIHT is determined by the complexity of the *projection* onto a given model. In this paper we consider two sparsity models endowed with linear time projections via dynamic programs, adapted from [3]. Therefore, for these models, MEIHT has linear time complexity.

MEIHT yields a k-model-sparse approximation $\hat{\mathbf{x}}$, with a *for all*, ℓ_1/ℓ_1 error guarantee as follows:

$$(1.2) \qquad \|\hat{\mathbf{x}} - \mathbf{x}_{\mathcal{S}}\|_1 \leq C_1 \sigma_{\mathcal{M}_k}(\mathbf{x})_1 + C_2 \|\mathbf{e}\|_1 ,$$

where $\sigma_{\mathcal{M}_k}(\mathbf{x})_1 := \min_{k-\text{model-sparse } \mathbf{x}'} \|\mathbf{x} - \mathbf{x}'\|_1$, for some constants $C_1, C_2 > 0$. This is a modification of (1.1) that ensures that the approximations belong to the model considered.

We consider two categories of structured signals: (a) k-rooted connected signals which have their k nonzero components lie in a rooted connected tree [5, 4, 26]; and (b) k-group sparse signals whose support is contained in the union of k overlapping groups [27, 34]. Note that k-block sparse signals whose support consists of k equal size non-overlapping blocks [36, 35, 4, 26] are a special case of (b). We show that our algorithm can efficiently recover signals from these models and our numerical results support this. In addition, we generalize the results of [26] to D-ary tree models for (a) and to overlapping groups, instead of only block sparse signals.

Our key contribution is introducing the first linear time algorithm for these models that resolves a key bottleneck in the application of model-expanders. Another contribution is the generalization of the results of [26] to cover a broader class of models. This is also significant for applications, since for instance we can consider wavelets quad-trees instead of just binary wavelet trees.

A hallmark of our approach is that the algorithm can be used also for sparsity models that do not necessary lead to a reduction in sketch size. However, the algorithm always returns a signal from the given sparsity model. In many applications, this may be more important than reducing the sketch size, because it allows to interpret the solution with respect to the chosen sparsity structure even in the presence of perturbations.

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 [14, 17, 16, 35], just to cite a few. Model-based compressed sensing [4, 26] 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 $\mathcal{O}\left(k\log\left(N/k\right)\right)$. For instance using dense sketching matrices [4] shows that we can have sketch lengths of $\mathcal{O}(k)$ and $\mathcal{O}\left(kg + k\log(N/(kg))\right)$ for the rooted-connected tree and the block structure models respectively - where g is the size of the blocks. Recently, [26] used sparse sketching matrices to obtain $\mathcal{O}\left(k\log\left(N/k\right)/\log\log\left(N/k\right)\right)$ and $\mathcal{O}\left(kg\log(N)/\log(kg)\right)$ for the same models considered in [4]. The former work provided ℓ_2/ℓ_1 guarantees in the for all case with recovery algorithms typically of time complexity $\mathcal{O}\left(N\log(N)\right)$; while the latter gave ℓ_1/ℓ_1 guarantees in the for all case with an exponential time recovery algorithm.

It is well known that sparse matrices have computational advantages over their dense counterparts. Berinde et al. in [7] introduced sparse binary, non-zero mean, sensing matrices that are adjacency matrices of expander graphs and they showed such matrices have the ℓ_1-norm restricted isometry property (RIP-1). A binary matrix $\mathbf{A}$ with d nonzeros per column has RIP-1 [7] if for all k-sparse vectors $\mathbf{x} \in \mathbb{R}^N$, it satisfies

$$(1.3) \qquad (1 - \delta_k)d\|\mathbf{x}\|_1 \leq \|\mathbf{A}\mathbf{x}\|_1 \leq d\|\mathbf{x}\|_1,$$

where δ_k is the restricted isometry constant (RIC) of order k. It is the smallest such constant over all sets of size k. 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 ℓ_1-minimization [7] and combinatorially greedy algorithms: Sparse Matching Pursuit (SMP) [8] and Expander Recovery Algorithm [28] (see also [19, 2]). However, [26] 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 improve on this work by proposing the first polynomial

time algorithm for model-sparse signals using model expander sensing matrices, when the sparsity model is endowed with a tractable projection operator.

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 [12]. 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 exist classes of group-based models that admit tractable projections that can be computed via dynamic programming in $\mathcal{O}(Nk)$ time, where k is the maximum number of allowed groups in the support of the solution [3]. In [12], another dynamic program for rooted connected subtree projections has been proposed that achieves a computational complexity of $\mathcal{O}(Nk)$. The projection step in our algorithm uses a modification of 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, $m = \mathcal{O}\left(k\log(N/k)/\epsilon\right)$ for $\epsilon > 0$. Deterministic construction of these objects with optimal parameters (graph degree d and m) is still an open challenge. The current best *sub-optimal* deterministic construction is due to [22] with $m = \mathcal{O}\left(d^2 k^{1+\alpha}\right)$ for $\alpha > 0$. The recent work of [26] on model-based RIP-1 matrices (model expanders) also used a probabilistic construction. As an extension of this work, we show that it is possible to probabilistically construct more general model expanders. It is conceivable that *model-optimal* deterministic construction of these model expanders is possible, but it is not the focus of this work. We remark that the first work on model-based sparse matrices [33] considered signal models similar to those in [26] and achieved ℓ_2/ℓ_2 error guarantees, but only for the *for each* case.

Outline In the next section we set our notation and give definitions needed in the sequel. In Section 3, we present our algorithm and analyze its error guarantees. In Section 4, we extend the existence of model-based expanders for D-ary rooted connected tree models and group-based models with overlaps. We provide numerical evidence in Section 5 and some concluding considerations in Section 6. Finally, Section 7 contains detailed proofs of the main results (Theorem 3.1 and Corollary 3.1).

2 Preliminaries

2.1 Notation Throughout this paper, we denote scalars by lowercase letters (e.g. k), vectors by lowercase boldface letter (e.g., $\mathbf{x}$), sets by uppercase calligraphic letters (e.g., $\mathcal{S}$) and matrices by uppercase boldface let-

ter (e.g. $\mathbf{A}$). We use the shorthand $[N] := \{1, \dots, N\}$. Given a set $\mathcal{S} \subseteq [N]$, we denote its complement by $\bar{\mathcal{S}} := [N] \setminus \mathcal{S}$ and use $\mathbf{x}_{\mathcal{S}}$ for the restriction of $\mathbf{x} \in \mathbb{R}^N$ onto $\mathcal{S}$, i.e., $(\mathbf{x}_{\mathcal{S}})_i = \mathbf{x}_i$ if $i \in \mathcal{S}$ and 0 otherwise. We use $|\mathcal{S}|$ to denote the cardinality of a set $\mathcal{S}$. We denote the set of k-sparse vectors as Σ_k.

We denote bipartite graphs by $(\mathcal{U}, \mathcal{V}, \mathcal{E})$ where $\mathcal{U}$ and $\mathcal{V}$ are the set of left and right nodes, respectively, and $\mathcal{E}$ 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 $(\mathcal{U}, \mathcal{V}, \mathcal{E})$ and a set $\mathcal{K} \subseteq \mathcal{U}$, 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 ℓ_p norm of a vector $\mathbf{x} \in \mathbb{R}^N$ is defined as $\|\mathbf{x}\|_p := \left(\sum_{i=1}^N x_i^p\right)^{1/p}$.

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. We start by defining lossless expanders and describing models of structured signals before we define model-based expanders.

DEFINITION 2.1. (LOSSLESS EXPANDER [11]) *Let a graph $G = ([N], [m], \mathcal{E})$ 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 some $\epsilon \in (0, 1/2)$ and any $\mathcal{K} \subset [N]$ of size $|\mathcal{K}| \leq k$, we have $|\Gamma(\mathcal{K})| \geq (1-\epsilon)d|\mathcal{K}|$, then G is referred to as a (k, d, ϵ)-lossless expander graph.*

The property $|\Gamma(\mathcal{K})| \geq (1-\epsilon)d|\mathcal{K}|$ is termed the *expansion* (or ϵ-expansion) property and we refer to ϵ as the expansion coefficient. Typically, lossless expanders have $\epsilon \ll 1$, for instance $\epsilon < 1/4$ [28].

Informally, 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 hierarchical model, where the elements of $[N]$ are organized in a *tree*, and a group model, where the elements of $[N]$ are grouped in M *overlapping groups*.

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

DEFINITION 2.3. (k-TREE SPARSE MODEL AND SET) *Given a D-ary tree $\mathcal{T}$ ($D \geq 2$), we define the collection of all k-rooted connected subtrees of $\mathcal{T}$ as the k-tree*

sparse model $\mathcal{T}_k$. A set $\mathcal{S} \subseteq [N]$ is k-tree sparse (i.e., $\mathcal{T}_k$-sparse) if $\mathcal{S} \subseteq \mathcal{K}$ for some $\mathcal{K} \in \mathcal{T}_k$.

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

DEFINITION 2.4. (GROUP STRUCTURE) *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 $\bigcup_{\mathcal{G} \in \mathfrak{G}} \mathcal{G} = [N]$.*

We say 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 group* structure. We consider this structure in the sequel, because it turns out that projecting onto such group models can be done in linear time.

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 $\mathcal{S} \in [N]$ 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 use $\mathcal{M}_k$ to jointly refer to the $\mathcal{T}_k$ and $\mathfrak{G}_k$ models and $\mathcal{M} = \{\mathcal{M}_1, \mathcal{M}_2, \ldots\}$ to represent the *model class*. Next we define model-sparse vectors/signals.

DEFINITION 2.6. (MODEL-SPARSE VECTORS) *Given a model $\mathcal{M}_k$, a vector $\mathbf{x} \in \mathbb{R}^N$ is* model-sparse, *denoted as $\mathcal{M}_k$-sparse, if $supp(\mathbf{x}) \subseteq \mathcal{K}$ 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 a graph $G = ([N], [m], \mathcal{E})$ 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, ϵ)-model expander graph.*

We interchangeably refer to the adjacency matrices of model-expander graphs as model-expanders. As noted in [26], it is straightforward to see that model-expanders satisfy a variant of the RIP-1, [7]. Precisely, we say a model-expander, $\mathbf{A}$ satisfies *model* RIP-1 if for all $\mathcal{M}_k$-sparse vectors $\mathbf{x} \in \mathbb{R}^N$

$$(2.4) \qquad (1 - \delta_{\mathcal{M}_k})d\|\mathbf{x}\|_1 \leq \|A\mathbf{x}\|_1 \leq d\|\mathbf{x}\|_1.$$

where $\delta_{\mathcal{M}_k}$ is the model RIC of order k. Henceforth, we also refer to model-expander matrices as model RIP-1 (denoted as $\mathcal{M}_k^\epsilon$-RIP-1) matrices as in [26].

2.3 Motivation Block models have direct applications in graph sketching, group testing and micro-arrays data analysis. Furthermore, these models and overlapping groups models have been advocated for several applications that include detecting genetic pathways for diagnosis or prognosis of cancer [27], background-subtraction in videos [4], mental state decoding via neuroimaging [21], to mention just a few. Loopless pairwise overlapping groups 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 pairwisely 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.

On the other hand, rooted connected tree models arise in many applications ranging from 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, [29] for a more extensive list of applications and references). Another application of the rooted connected tree models is in Earth Mover Distance (EMD) sparse recovery [25], where the pyramid transform can be used to convert the EMD recovery to a rooted connected sparse recovery.

3 Model based recovery with expanders

3.1 Algorithm overview The model-expander iterative hard thresholding (MEIHT) algorithm proposed is a hybrid adaptation of the *expander iterative hard thresholding* (EIHT) algorithm proposed in [18] (which is a modification of the *sparse matching pursuit* (SMP) algorithm [8]) and projections onto the considered models using the results in [3] and [12]. Let the compressive measurements be given as $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ for a signal of interest $\mathbf{x} \in \mathbb{R}^N$ where the noise vector $\mathbf{e} \in \mathbb{R}^m$ characterizes the measurement error. Typically initializing $\mathbf{x}^0 = \mathbf{0} \in \mathbb{R}^N$, the SMP iterates the following

$$(3.5) \qquad \mathbf{x}^{n+1} = H_k\{\mathbf{x}^n + H_{2k}[\mathcal{M}(\mathbf{y} - \mathbf{A}\mathbf{x}^n)]\}$$

where $H_s(\mathbf{u})$ is the hard thresholding operator which retains the s largest in magnitude components of $\mathbf{u}$ and sets the rest to zero and $\mathcal{M}$ is the *median operator* which is nonlinear and defined componentwise on a vector, say $\mathbf{u} \in \mathbb{R}^m$, as follows

$$(3.6) \qquad [\mathcal{M}(\mathbf{u})]_i := \text{median}\,[u_j, j \in \Gamma(i)] \text{ for } i \in [N].$$

SMP was modified to become the *expander iterative hard thresholding* (EIHT) algorithm in [18]. Precisely, the EIHT usually initializes $\mathbf{x}^0 = \mathbf{0} \in \mathbb{R}^N$ and iterates

$$(3.7) \qquad \mathbf{x}^{n+1} = H_k \left[\mathbf{x}^n + \mathcal{M} \left(\mathbf{y} - \mathbf{A} \mathbf{x}^n \right) \right].$$

The stability and robustness to noise of EIHT was shown in [18]. We adapt EIHT and call it the *model expander iterative hard thresholding* (MEIHT) algorithm for the purpose of sparse recovery with model based expanders. In MEIHT, we replace the hard thresholding operator, $H_s(\mathbf{b})$, by the projector $\mathcal{P}_{\mathcal{M}_s}(\mathbf{b})$, which projects the vector $\mathbf{b}$ onto our model of size s, $\mathcal{M}_s$, with respect to the ℓ_1 norm (see Section 3.5 for more details). The pseudo-code of MEIHT is given in Algorithm 1 below. Note that the proposed algorithm always returns a signal from the given sparsity model.

Algorithm 1 Model-Expander IHT (MEIHT)

Input: Data: $\mathbf{A}$ and $\mathbf{y}$ and a bound of $\|\mathbf{e}\|_1$
Output: k-sparse approximation $\widehat{\mathbf{x}}$
Initialization:
    Set $\mathbf{x}^0 = 0$
Iteration: While some *stopping criterion* is not true
    Compute $\mathbf{x}^{n+1} = \mathcal{P}_{\mathcal{M}_k} \left[\mathbf{x}^n + \mathcal{M} \left(\mathbf{y} - \mathbf{A} \mathbf{x}^n \right) \right]$
Return $\widehat{\mathbf{x}} = \mathbf{x}^{n+1}$

REMARK 3.1. *SMP, EIHT and MEIHT could be regarded as projected gradient descent algorithms [32]. This is because the median operator behaves approximately like the adjoint of $\mathbf{A}$ [18, 8]. 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 Runtime complexity of MEIHT The complexity of MEIHT depends on the cost of the median operation and the cost of the projection onto a sparsity model. It turns out that in the settings with model expander measurement matrices the projection on the models is costlier than the median operation. Therefore, for the models we consider, the runtime complexity of MEIHT is linear. A detailed discussion on the projection step and its cost is given in Section 3.5. The following proposition formally states the complexity of MEIHT for the models considered.

PROPOSITION 3.1. *The runtime of MEIHT is $\mathcal{O}\left(knN\right)$ and $\mathcal{O}\left(M^2 kn + nN\right)$ for the models $\mathcal{T}_k$ and $\mathfrak{G}_k$ respectively, where k is the model order, n is the number of iterations, M is the number of groups and N is the dimension of the signal.*

Recall that k being the model order means that signals in $\mathcal{T}_k$ are k-sparse while signals in $\mathfrak{G}_k$ are k-group sparse.

It is clear from the proposition that the runtime is linear in N for both models.

Proof. The algorithm has linear convergence (see below) and hence a fixed, say n, number of iterations are sufficient to achieve an approximate solution. At each iteration MEIHT performs N median operations of d elements which involve sorting with $\mathcal{O}\left(d \log d\right)$ complexity and a projection step that costs $\mathcal{O}\left(kN\right)$ for the $\mathcal{T}_k$ model and $\mathcal{O}\left(M^2 k + N\right)$ for the $\mathfrak{G}_k$ model. The projection step dominates since d is small, to be precise $d = \mathcal{O}\left(\log(N)/\log\log(N/k)\right)$ and $d = \mathcal{O}\left(\log(N)/\log\left(kg_{\max}\right)\right)$ for the former and latter models respectively, see Section 4. Therefore the runtime of MEIHT is a product of the number of iterations and the complexity of the projections.

3.3 Convergence of MEIHT The convergence analysis of MEIHT follows similar arguments as the proof of the linear time convergence of EIHT given in [18]. The key difference between the analysis of EIHT and MEIHT is the projections $H_k(\cdot)$ and $\mathcal{P}_{\mathcal{M}_k}(\cdot)$ respectively. We leverage and adapt recent results in [3] to perform the projection $\mathcal{P}_{\mathcal{M}_k}(\cdot)$ exactly and efficiently in linear time. The projections are analysed in Section 3.5. Theorem 3.1 and Corollary 3.1 below bound the error of the output of the MEIHT algorithm.

THEOREM 3.1. *Let $\mathcal{M}_k$ be a model and let $\mathcal{S}$ be $\mathcal{M}_k$-sparse. Consider $\mathbf{A} \in \{0,1\}^{m \times N}$ to be the adjacency matrix of a model-expander with $\epsilon_{\mathcal{M}_{3k}} < 1/12$. For any $\mathbf{x} \in \mathbb{R}^N$ and $\mathbf{e} \in \mathbb{R}^m$, the sequence of updates $(\mathbf{x}^n)$ of MEIHT with $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ satisfies, for any $n \geq 0$,*

$$(3.8) \quad \begin{aligned} &\|\mathbf{x}^n - \mathbf{x}_{\mathcal{S}}\|_1 \\ &\leq \alpha^n \|\mathbf{x}^0 - \mathbf{x}_{\mathcal{S}}\|_1 + (1 - \alpha^n) \beta \|\mathbf{A}\mathbf{x}_{\bar{\mathcal{S}}} + \mathbf{e}\|_1, \end{aligned}$$

where $\alpha < 1$ and β depends only on $\epsilon_{\mathcal{M}_{3k}}$.

COROLLARY 3.1. *If the sequence of updates $(\mathbf{x}^n)$ converges to $\mathbf{x}^\star$ as $n \to \infty$, then*

$$(3.9) \qquad \|\mathbf{x}^\star - \mathbf{x}_{\mathcal{S}}\|_1 \leq C_1 \sigma_{\mathcal{M}_k}(\mathbf{x})_1 + C_2 \|\mathbf{e}\|_1,$$

for some constants $C_1, C_2 > 0$ depending only on $\epsilon_{\mathcal{M}_{3k}}$.

Sketch proof of Theorem 3.1 & Corollary 3.1 The first step of the proof consists in bounding $\| \left[\mathcal{M} \left(\mathbf{A}\mathbf{x}_{\mathcal{S}} + \mathbf{e} \right) - \mathbf{x} \right]_{\mathcal{S}} \|_1$. This is done inductively over the set $\mathcal{S}$ and key ingredients to this proof are properties of the median operator and the expansion property of expander graphs. We obtain that

$$(3.10) \quad \begin{aligned} &\| \left[\mathcal{M} \left(\mathbf{A}\mathbf{x}_{\mathcal{S}} + \mathbf{e} \right) - \mathbf{x} \right]_{\mathcal{S}} \|_1 \\ &\leq \frac{4\epsilon_{\mathcal{M}_s}}{1 - 4\epsilon_{\mathcal{M}_s}} \|\mathbf{x}_{\mathcal{S}}\|_1 + \frac{2}{(1 - 4\epsilon_{\mathcal{M}_s})d} \|\mathbf{e}_{\Gamma(\mathcal{S})}\|_1, \end{aligned}$$

where $\mathcal{S} \in \mathcal{M}_s$-sparse for $|\mathcal{S}| \leq s$, $\epsilon_{\mathcal{M}_s}$ is the expansion coefficient of our model and $\Gamma(\mathcal{S})$ is the set of neighbours of $\mathcal{S}$. Now, assuming exact projections onto our model (in the ℓ_1 norm), we use the above bound with the triangle inequality to have the following upper bound

$$(3.11) \quad \begin{aligned} &\|\mathbf{x}^{n+1} - \mathbf{x}_{\mathcal{S}}\|_1 \\ &\leq \alpha \|\mathbf{x}^n - \mathbf{x}_{\mathcal{S}}\|_1 + (1-\alpha)\beta \|\mathbf{A}\mathbf{x}_{\bar{\mathcal{S}}} + \mathbf{e}\|_1, \end{aligned}$$

for some constants $\alpha < 1$ and $\beta > 0$ which depend only on $\epsilon_{\mathcal{M}_{3s}}$. This result relies on the fact that our models have the *nestedness property* discussed in Section 3.4. By induction over n, this bounds leads to (3.8).

Corollary 3.1 easily follows from Theorem 3.1 by taking the limits of (3.8) as $n \to \infty$ and using a property of model RIP-1 matrices, which is also a property of the underlying expander graphs. We provide the detailed proofs of Theorem 3.1 and Corollary 3.1 in Section 7.

3.4 Nestedness property of sparsity models The RIP-1 requirement in traditional CS for the basic k-sparse case holds for sums and differences of sparse signals (Minkowski sums of signal supports). In model-based recovery, however, the Minkowski sums of sparsity models from a given model class (e.g., rooted connected tree models or group models) may not necessarily belong to the same model class and, as a result, model RIP-1 may not hold for such Minkowski sums [4]. For instance, the Minkowski sum, $\mathcal{M}_k \oplus \mathcal{M}_k \oplus \mathcal{M}_k$, of model orders of size k may not be in the model $\mathcal{M}_{3k}$. Consequently, the model RIC of this Minkowski sum is not necessarily $\delta_{\mathcal{M}_{3k}}$ which is proportional to ϵ_{3k}. However, if a given model has what we refer to as the *nestedness property*, defined below, the Minkowski sum, $\mathcal{M}_k \oplus \mathcal{M}_k \oplus \mathcal{M}_k$ is in the model class and the model RIC of the Minkowski sum is $\delta_{\mathcal{M}_{3k}}$.

In particular, the convergence proof of our algorithm relies on the fact that the models we consider are *nested*, in other words we require that Minkowski sums of signals in a model remain in the model. This is equivalent to asking the model to have the *nestedness property* defined below which is similar to the *nested approximation property* (NAP) introduced in [4].

DEFINITION 3.1. (NESTEDNESS PROPERTY) *A model class* $\mathcal{M} = \{\mathcal{M}_1, \mathcal{M}_2, \ldots, \}$ *has the* nestedness property *if, for any* $\mathcal{S} \in \mathcal{M}_k$ *and* $\mathcal{S}' \in \mathcal{M}_{k'}$, *with* $\mathcal{M}_k, \mathcal{M}_{k'} \in \mathcal{M}$, *we have* $\mathcal{S} \cup \mathcal{S}' \in \mathcal{M}_{k+k'}$.

It is important to note that the models we consider in this work, rooted connected tree models and overlapping group models, have the nestedness property. In other words, the model-expanders expand on models of order $k, 2k, 3k$, and so on, simultaneously. For instance, the union of of the supports of two $\mathcal{T}_k$-sparse signals is the support of a $\mathcal{T}_{2k}$-sparse signal, similarly for $\mathfrak{G}_k$-group sparse signals.

Not all sparsity models have the nestedness property. An example is the dispersive model used to described neuronal spikes proposed in [23], where the model-sparse signal is defined as a sparse signal whose nonzero coefficients must be separated by a fixed number of zero coefficients. Summing two such signals can yield a signal whose nonzero coefficients are not separated enough and hence does not belong to the desired sparsity model.

3.5 Projections The projections for the rooted connected tree model and the loopless overlapping groups model can be computed in linear time via dynamic programming, leveraging the results in [3] and [12], where the authors considered projections in the ℓ_2 norm.

Given a general model $\mathcal{M}_k$, we define the projections in the ℓ_1 norm as

$$(3.12) \quad \mathcal{P}_{\mathcal{M}_k}(\mathbf{x}) \in \underset{\mathbf{z}:\mathrm{supp}(\mathbf{z}) \in \mathcal{M}_k}{\mathrm{argmin}} \|\mathbf{x} - \mathbf{z}\|_1.$$

We can reformulate (3.12) as a discrete optimization problem over sets in $\mathcal{M}_k$. Indeed, we have

$$\begin{aligned} \min_{\mathbf{z}:\mathrm{supp}(\mathbf{z}) \in \mathcal{M}_k} \|\mathbf{x} - \mathbf{z}\|_1 &= \min_{\mathcal{S} \in \mathcal{M}_k} \min_{\mathbf{z}:\mathrm{supp}(\mathbf{z}) \subseteq \mathcal{S}} \|\mathbf{x} - \mathbf{z}\|_1 \\ &= \max_{\mathcal{S} \in \mathcal{M}_k} \max_{\mathbf{z}:\mathrm{supp}(\mathbf{z}) \subseteq \mathcal{S}} \|\mathbf{x}\|_1 - \|\mathbf{x} - \mathbf{z}\|_1 \\ (3.13) \quad &= \max_{\mathcal{S} \in \mathcal{M}_k} \|\mathbf{x}_{\mathcal{S}}\|_1. \end{aligned}$$

Therefore, finding the projection in the ℓ_1 norm of a signal $\mathbf{x}$ onto a model $\mathcal{M}_k$ amounts to finding a set in $\mathcal{M}_k$ that maximizes the "cover" of $\mathbf{x}$ in absolute value. Tractable models $\mathcal{M}_k$ are models for which it is possible to solve (3.13) in polynomial time, leading to an overall polynomial time complexity for the proposed recovery algorithm, see Section 3.2.

Following [3], we can show that projections in ℓ_1 norm onto general group sparse models, which include the rooted connected tree model, can be reformulated as the combinatorial Weighted Maximum Coverage (WMC) problem. Since the WMC is NP-hard in general, these projections are NP-hard too. However, [3] identified group structures that lead to linear time algorithms. These structures correspond to the models we consider: loopless overlapping group models and rooted connected tree models.

For projections onto a group model $\mathfrak{G}_k$, we can reformulate (3.13) as

$$(3.14) \quad \max_{\substack{\mathcal{S} \subseteq \mathfrak{G} \\ |\mathcal{S}| \leq k \\ \mathcal{I} = \bigcup_{\mathcal{G} \in \mathcal{S}} \mathcal{G}}} \|\mathbf{x}_{\mathcal{I}}\|_1 \, .$$

Let now define the binary matrix $\mathbf{A}^{\mathfrak{G}} \in \mathbb{B}^{N \times M}$,

$$\mathbf{A}^{\mathfrak{G}}_{ij} = \begin{cases} 1, & \text{if } i \in \mathcal{G}_j; \\ 0, & \text{otherwise.} \end{cases}$$

This matrix fully encodes the group structure and allows to rewrite (3.14) as the following Weighted Maximum Coverage problem
(3.15)

$$\max_{\boldsymbol{\omega} \in \mathbb{B}^M, \ \mathbf{y} \in \mathbb{B}^N} \left\{ \sum_{i=1}^{N} y_i |x_i| : \mathbf{A}^{\mathfrak{G}} \boldsymbol{\omega} \geq \mathbf{y}, \sum_{j=1}^{M} \omega_j \leq k \right\},$$

where $\boldsymbol{\omega}$ and $\mathbf{y}$ are binary variables that specify which groups and which variables are selected, respectively. The constraint $\mathcal{A}^{\mathfrak{G}} \boldsymbol{\omega} \geq \mathbf{y}$ makes sure that for every selected variable at least one group that contains it is selected, while $\sum_{j=1}^{M} \omega_j \leq k$ limits the number of selected groups to at most k. Note that the weights for the WMC are given by the absolute values of the coefficients of $\mathbf{x}$.

Two recently proposed dynamic programs [3, 12] compute the projections in the ℓ_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 ℓ_1 norm, it is sufficient to replace the weights of the nodes of the graphs by using the absolute values of the components of $\mathbf{x}$ instead of the square value.

For the k-tree model, the graph is the given tree where each node correspond to a variable and has weight equal to the variable absolute value. 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 is $\mathcal{O}(ND^2k)$, where D is the tree degree ($D = 2$ for binary trees).

For the case of loopless overlapping groups, [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. By defining an appropriate graph exploration rule, the dynamic program has time complexity equal to $\mathcal{O}(M^2k + N)$.

REMARK 3.2. (GENERALIZATION) *In [3], the overlapping groups model has been generalized to within-group sparsity with an overall sparsity budget K, together with the group budget k. A dynamic program solves the projection* (3.12) *exactly in* $\mathcal{O}(M^2K^2k + N \log K)$ *operations for this model.*

4 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 and m) as possible. Bounds, both lower and upper, on m for model-expanders, for binary tree-sparse and block-sparse models, are derived in [26]. 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 [26].

However, since the upper bounds involve the enumeration of the cardinality of the models which is different from [26], 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 introduced and proved [9].

LEMMA 4.1. *[9] There exist constants $C > 1$ and $\mu > 0$ such that, whenever $m \geq Cdt/\epsilon$, for any $T \subseteq [N]$ with $|T| = t$ one has*

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

4.1 D-ary tree model D-ary tree model-expanders or $\mathcal{T}_k^\epsilon$-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 4.1 states the existence and the sizes of the parameters of these matrices.

THEOREM 4.1. *For $\epsilon \in (0, 1/2)$ and $k = \omega (\log N)$ there*

exists a $\mathcal{T}_k^\epsilon$-RIP-1 matrix with

(4.16)
$$d = \mathcal{O}\left(\frac{\log{(N/k)}}{\epsilon \log\log{(N/k)}}\right) \quad and \quad m = \mathcal{O}\left(\frac{dk}{\epsilon}\right).$$

Proof. For a model $\mathcal{T}_k \subseteq \Sigma_k$ and $t \in [k]$ denote $\mathcal{T}_k$-sparse sets of size t as $\mathcal{T}_{k,t}$ 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 $\mathcal{T}_{k,t}$.

LEMMA 4.2. *For all $t \in [k]$, we have*

$$|\mathcal{T}_{k,t}| \leq \min\left[T_k \cdot \binom{k}{t}, \binom{N}{t}\right],$$
$$T_k = \frac{1}{(D-1)k+1}\binom{Dk}{k}.$$

Proof. It is sufficient to know that the T_k are the Pfaff-Fuss-Catalan numbers or k-Raney numbers [30, Note 12], which enumerates the total number of ordered, rooted D-ary trees of cardinality k.

Using the fact that $\binom{x}{y}$ is bounded above by $\left(\frac{ex}{y}\right)^y$, we have that for all $t \in [k]$, $|\mathcal{T}_{k,t}|$ is upper bounded by

(4.17)
$$\min\left[\frac{(eD)^k}{(D-1)k+1}\left(\frac{ek}{t}\right)^t, \left(\frac{eN}{t}\right)^t\right].$$

Taking a union bound over the number of $\mathcal{T}_{k,t}$ we see that the probability of Lemma 4.1 goes to zero with N if

$$\forall t \in [k] \quad |\mathcal{T}_{k,t}| \cdot \left(\mu \cdot \frac{\epsilon m}{dt}\right)^{-\epsilon dt} \leq \frac{1}{N},$$

where m satisfies the bound in the lemma. Let us define $t^* := \frac{k\log(eD) - \log(Dk-D+1)}{\log(N/k)}$. By simple algebra we see that at $t = t^*$ the two quantities in the square brackets of (4.17) 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 that $\forall t \in [1, t^*]$

(4.18)
$$\left(\frac{eN}{t}\right)^t\left(\mu \cdot \frac{\epsilon m}{dt}\right)^{-\epsilon dt} \leq \frac{1}{N},$$

and that $\forall t \in [t^*, k]$

(4.19)
$$\frac{(eD)^k}{(D-1)k+1}\left(\frac{ek}{t}\right)^t\left(\mu \cdot \frac{\epsilon m}{dt}\right)^{-\epsilon dt} \leq \frac{1}{N}.$$

In accordance with Theorem 4.1, we let the left degree $d = C'\log{(N/k)}/[\epsilon\log\log{(N/k)}]$ and $m = C''dk/\epsilon$. Note that the left hand side of Inequalities (4.18) and (4.19) 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 (4.18) becomes

$$eN\left(C''\mu k\right)^{-C'\log(N/k)/\log\log(N/k)} \leq 1/N.$$

This inequality holds for sufficiently large C' and C'' given that $k = \omega(\log N)$. Thus (4.18) holds for $t = 1$. When $t = t^*$, we have the left hand side of inequality (4.18) equal to the left hand side of inequality (4.19). Substituting the value of t^* and those of m and d in the left hand side of (4.18) gives the product of the following two terms.

$$\left(\frac{eN\log{(N/k)}}{\eta}\right)^{\frac{k\log(eD) - \log(Dk-D+1)}{\log(N/k)}}$$
$$\times \left(\frac{C''\mu k\log{(N/k)}}{\eta}\right)^{-C'\frac{k\log(eD) - \log(Dk-D+1)}{\log\log(N/k)}},$$

where $\eta = k\log(eD) - \log(Dk - D + 1)$. The above quantity is less than $1/N$ for sufficiently large C' and C'' given that $k = \omega(\log N)$. Thus (4.18) and (4.19) hold for $t = t^*$. For $t = k$, with the values of m and d, the left hand side of (4.18) becomes

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

This quantity is far less than $1/N$ for sufficiently large C' and C'' given that $k = \omega(\log N)$. Thus (4.19) hold for $t = k$ and this concludes the proof.

4.2 Overlapping group model Fixed overlapping group model-expanders or $\mathfrak{G}_k^\epsilon$-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. Theorem 4.2 states the existence and the sizes of the parameters of these matrices.

THEOREM 4.2. *Let the total number of groups of the model be M and the maximum size of groups be $g_{\max}$. For any $\epsilon \in (0, 1/2)$, if $M \geq 2$, $g_{\max} = \omega{(\log N)}$, and $N > kg_{\max}$, then there exist a $\mathfrak{G}_k^\epsilon$-RIP-1 matrix, $\mathbf{A} \in \mathbb{R}^{m \times N}$ with d nonzero entries per column, such that*

(4.20)
$$d = \mathcal{O}\left(\frac{\log(N)}{\epsilon \log{(kg_{\max})}}\right), \quad and$$
$$m = \mathcal{O}\left(\frac{dkg_{\max}}{\epsilon}\right).$$

Intuitively, we see that the condition $N > kg_{\max}$ is necessary in order to not have just one large group; while the condition $g_{\max} = \omega(\log N)$ means that $g_{\max}$, the maximum size of groups, cannot be too small.

Proof. Let $\mathfrak{G}_{k,t}$ denote $\mathfrak{G}_k$-sparse sets with sparsity t. We use the following estimate of the number of sets $\mathfrak{G}_{k,t}$. Then we have $\forall\, t \in [kg_{\max}]$

$$(4.21) \qquad |\mathfrak{G}_{k,t}| \le \min\left[\binom{M}{k}\binom{kg_{\max}}{t}, \binom{N}{t}\right].$$

Since $\binom{x}{y}$ is bounded above by $\left(\frac{ex}{y}\right)^y$, we have for all $t \in [kg_{\max}]$

$$(4.22) \quad |\mathfrak{G}_{k,t}| \le \min\left[\left(\frac{eM}{k}\right)^k\left(\frac{ekg_{\max}}{t}\right)^t, \left(\frac{eN}{t}\right)^t\right].$$

Using a union bound over the sets $\mathfrak{G}_{k,t}$, we see that the probability of Lemma 4.1 goes to zero with N if

$$\forall t \in [kg_{\max}], \quad |\mathfrak{G}_{k,t}| \cdot \left(\mu \cdot \frac{\epsilon m}{dt}\right)^{-\epsilon dt} \le \frac{1}{N},$$

where m satisfies the bound in the lemma. By simple algebra we see that at $t = t^* = \frac{k\log(eM/k)}{\log(N/(kg_{\max}))}$ the two quantities in the square brackets of (4.21) are equal; and $\forall t \in [1, t^*]$ the second quantity is smaller, while $\forall t \in [t^*, kg_{\max}]$ the first quantity is smaller. Therefore, we split the domain of t into two intervals and prove that $\forall t \in [1, t^*]$

$$(4.23) \qquad \left(\frac{eN}{t}\right)^t\left(\mu \cdot \frac{\epsilon m}{dt}\right)^{-\epsilon dt} \le \frac{1}{N},$$

and that $\forall t \in [t^*, kg_{\max}]$

$$(4.24) \quad \left(\frac{eM}{k}\right)^k\left(\frac{ekg_{\max}}{t}\right)^t\left(\mu \cdot \frac{\epsilon m}{dt}\right)^{-\epsilon dt} \le \frac{1}{N}.$$

In accordance with Theorem 4.2, we let the left degree $d = C'\log(N)/[\epsilon\log(kg_{\max})]$ and $m = C''dkg_{\max}/\epsilon$. The left hand side of Inequalities (4.23) and (4.24) 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 $t = 1, t^*, kg_{\max}$. For $t = 1$, Inequality (4.23) becomes

$$eN\left(C''\mu kg_{\max}\right)^{-C'\log(N)/\log(kg_{\max})} \le 1/N.$$

This is true for sufficiently large C' and C'' given that $g_{\max} = \omega(\log N)$. Thus (4.23) holds for $t = 1$. When $t = t^*$, we have the left hand side of inequality (4.23) equal to the left hand side of inequality (4.24). Substituting t^* for t and the values of m and d in the left hand side of (4.23) leads to

$$\left(\frac{eN}{t^*}\right)^{t^*}\left(\frac{C''\mu kg_{\max}}{t^*}\right)^{-C't^*\log(N)/\log(kg_{\max})} \le \frac{1}{N}.$$

The above inequality is true for sufficiently large C' and C'' given that $g_{\max} = \omega(\log N)$. Thus (4.23) and (4.24) hold for $t = t^*$. For $t = k$, with the values of m and d, the left hand side of (4.23) becomes

$$\left(\frac{e^2Mg_{\max}}{k}\right)^k (C''\mu g_{\max})^{-C'k\log(N)/\log(kg_{\max})} \le \frac{1}{N}.$$

This is also true for sufficiently large C' and C'' given that $g_{\max} = \omega(\log N)$. Thus (4.24) hold for $t = k$ and this concludes the proof.

The extra condition on N in the theorem can be explicitly derived if we impose that $t^* > 0$ which leads to $N > kg_{\max}$. But this condition is implicitly assumed in the above derivation.

REMARK 4.1. *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 straight forward way. The fixed overlapping group model is another special case of the more general overlapping groups model proven above.*

5 Experimental Results

5.1 Block sparsity In the first experiment, we consider recovering a block-sparse signal, where the M equal-sized blocks define a partition of the set $\{1, \ldots, N\}$, so that the block size is given by $g = N/M$.

We fix the number of active groups, $k = 5$, and let the group size g grow as $\log(N)$ to satisfy the assumptions of our bound for group-sparse recover (4.20). We also have that the number of groups M must grow as $N/\log(N)$. Specifically, we set $N \in \{2^7, 2^8, \ldots, 2^{13}\}$, $M = \lfloor N/\log_2(N)\rfloor$, and $g = \lfloor N/M\rfloor$. We vary the number of samples m and compute the relative recovery error $\|\widehat{\mathbf{x}} - \mathbf{x}\|_1/\|x\|_1$, where $\widehat{\mathbf{x}}$ is the signal estimated either with the EIHT or the MEIHT algorithms and $\mathbf{x}$ is the true signal. For each problem size N, we repeat the experiment 50 times. For each repetition, we randomly select k blocks from M and set the components of the true signal $\mathbf{x}$ 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 $\mathbf{A} \in \mathbb{R}^{m \times N}$ with $d = \lfloor 2\log(N)/\log(kg)\rfloor$ ones per column in order to satisfy the condition in (4.20). The columns of $\mathbf{A}$ are normalized to have unitary ℓ_1 norm. We then create the measurement vector as $\mathbf{y} = \mathbf{Ax}$, which we use to recover $\mathbf{x}$ using either the EIHT or the MEIHT algorithm. We record the minimum number of samples m^* that yields a median relative recovery error below 10^{-5}.

In the top panel of Figure 1, we plot m^* as a function of $\log_2(N)$, showing strong empirical evidence for our theoretical results. Indeed, in the setting of this experiment, our bound predicts that m^* for MEIHT should scale as $\log(N)^2/\log(C_1 \log(N))$, where C_1 is a positive constant. Instead the bound for standard expanders indicates that m^* for EIHT should scale as $\log(N)\log(N/C_2\log(N))$, where C_2 is a positive constant. The plot shows that empirically the minimum number of samples for good recovery grows faster for EIHT than for MEIHT.

5.2 Tree sparsity In the second experiment, we try to recover a signal $\mathbf{x} \in \mathbb{R}^N$ whose nonzero coefficients lie in a rooted connected subtree of a regular binary tree.

We set the number of nonzero coefficients $k = \lfloor 2\log_2(N)\rfloor$ as required by Theorem 4.1, while their values are randomly drawn from a standard Gaussian distribution. We randomly draw a sparse sketching matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ with $d = \lfloor 2.5 \frac{\log(N/k)}{\log\log(N/k)}\rfloor$ ones per column in order to satisfy the condition in (4.16). The columns of $\mathbf{A}$ are normalized to have unitary ℓ_1 norm. We use the EIHT and MEIHT algorithms to obtain an approximation $\widehat{\mathbf{x}}$ from the linear sketch $\mathbf{Ax}$. 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 $\widehat{\mathbf{x}}$ with both algorithms, recording the relative recovery error $\|\widehat{\mathbf{x}} - \mathbf{x}\|_1/\|\mathbf{x}\|_1$. We repeat the experiment 50 times with a different draw of the signal $\mathbf{x}$ and compute the median of the relative recovery errors for both algorithms. Finally, we record the minimum number of measurements, m^*, that yields a median relative recovery error less than 10^{-5}.

The results are presented in the bottom panel of Figure 1 and show a strong accordance to the theoretical results of the previous section. Indeed, in the setting of this experiment, our bound predicts that m^* for MEIHT should scale as $\frac{\log(N)\log(N/(C\log(N)))}{\log\log(N/C\log(N))}$, where C is a positive constant. Instead, the bound for standard expanders indicates that m^* for EIHT should scale as $\log(N)\log(\frac{N}{C\log(N)})$. The plot shows that in practice the minimum number of samples for good recovery grows faster for EIHT than for MEIHT, matching the theoretical bounds.

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 linear time re-
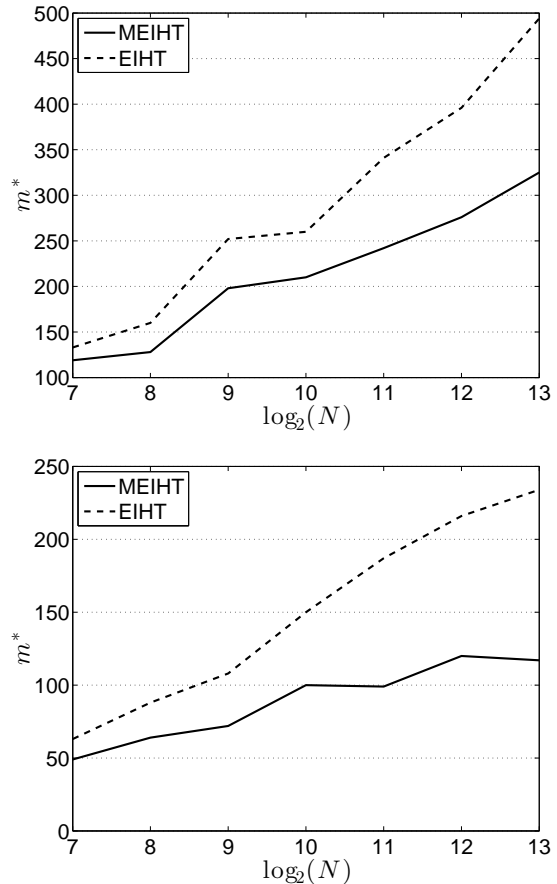

Figure 1: Model-based signal recovery example. *Top panel:* block-sparse model. The original signal is block-sparse (see text for details on how it is generated). The plot shows the minimum number of samples required to reach a median relative recovery error in the ℓ_1 norm below 10^{-5} for the EIHT and the MEIHT algorithms as the signal ambient dimension is increased. *Bottom panel:* rooted connected subtree model. The original signal is sparse and its support is a rooted connected subtree. The plot shows the minimum number of samples required to reach a median relative recovery error in the ℓ_1 norm below 10^{-5} for the EIHT and MEIHT algorithms as the signal ambient dimension is increased. Note that the x-axis in both plots reports the logarithm of the ambient dimension.

covery algorithm with ℓ_1/ℓ_1 recovery guarantees for the models we consider. Our results rely on: (i) identifying a special class of sparse matrices that are the adjacency matrices of model-based expanders and which have the model ℓ_1 norm restricted isometry property; (ii) performing exact projections onto the sparsity models by adapting the dynamic programs presented in [3]. We

show that by exploiting structured sparsity priors we are able to reduce the sketch length that guarantees recovery and our experimental results confirm these findings. In addition, we derive probabilistic constructions of a more general class of model expanders.

We conclude with a surprising result that reveals an interesting research question. We repeat the experiment in Section 5.2, using sparse rooted connected tree signals with fixed sparsity, $k = 16$, and sparse sketching matrices with a fixed number of ones per column (corresponding to $d = 6$) in contrast to what required by the theory. Figure 2 shows that in this setting, the proposed algorithm seems to yield a constant sample complexity, comparable to the one predicted for dense sketching matrices. It may be possible that the better bound for dense sketching matrices may be achievable in the *for each* case also by sparse ones, bringing even greater efficiency advantages.
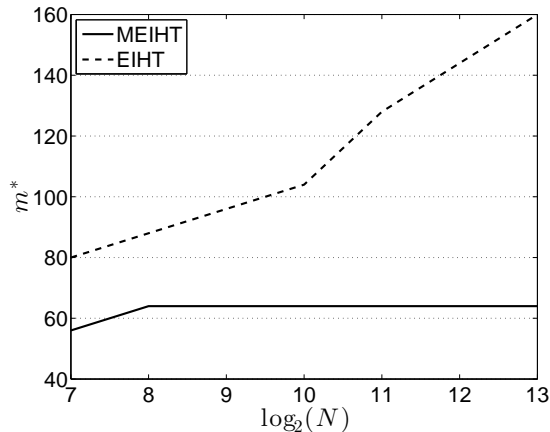

Figure 2: A surprising result. Model-based signal recovery for the rooted connected tree model. The original signal is $k = 16$ sparse and its support is a rooted connected subtree. The model expander is generated with $d = 6$ ones per column. The plot shows the minimum number of samples required to reach a median relative recovery error in the ℓ_1 norm below 10^{-5} for the EIHT and MEIHT algorithms as the signal ambient dimension is increased. Note that the x-axis reports the logarithm of the ambient dimension.

7 Proofs

Before we proceed with the proof of Theorem 3.1, we present an essential lemma. This lemma is similar to Lemma 13.16 in [18] which holds only for the basic k-sparse model. Here, we generalize this lemma to cover the models we consider in this paper using similar proof techniques to that of [18].

We first start with additional notations and definitions that are required in this section. Recall that $\mathbf{A}$ is the adjacency matrix of a (k, d, ϵ)-model expander, $G = ([N], [m], \mathcal{E})$. Let $\mathcal{S} \subseteq [N]$, then $\Gamma'(\mathcal{S})$ refers to the set of vertices connected to exactly one vertex in $\mathcal{S}$, while $\Gamma''(\mathcal{S})$ refers to the set of vertices connected to more than one vertex in $\mathcal{S}$. Note that $\Gamma''(\mathcal{S})$ is referred to as the *collision set* in the RIP-1 literature. The variable e_{ij} indicates the edge from vertex i to vertex j while e_i denotes the ith component of the vector $\mathbf{e}$.

7.1 Key lemma for the proof of Theorem 3.1

Now, we state the lemma which is the key ingredient to the proof of Theorem 3.1. This lemma is about the sub-problem of our algorithm involving the median operation and it shows that the median operator approximately inverts the action of $\mathbf{A}$ on sparse vectors, i.e., it behaves more-or-less like the adjoint of $\mathbf{A}$. For clarity, we provide the proof only for the k-tree sparse model, $\mathcal{T}_k$. The proof for the k-group sparse model, $\mathfrak{G}_k$, easily follows *mutatis mutandis*.

LEMMA 7.1. *Let* $\mathbf{A} \in \{0, 1\}^{m \times N}$ *be the adjacency matrix of a model expander with an expansion coefficient of order k, $\epsilon_{\mathcal{M}_k}$, satisfying $4\epsilon_{\mathcal{M}_k}d < d + 1$. If $\mathcal{S} \subseteq [N]$ is $\mathcal{M}_k$-sparse, then for all $\mathbf{x} \in \mathbb{R}^N$ and $\mathbf{e} \in \mathbb{R}^m$*

$$
\begin{aligned}
& \| \left[\mathcal{M} \left(\mathbf{A}\mathbf{x}_{\mathcal{S}} + \mathbf{e} \right) - \mathbf{x} \right]_{\mathcal{S}} \|_1 \\
(7.25) \quad & \leq \frac{4\epsilon_{\mathcal{M}_k}}{1 - 4\epsilon_{\mathcal{M}_k}} \|\mathbf{x}_{\mathcal{S}}\|_1 + \frac{2}{(1 - 4\epsilon_{\mathcal{M}_k})d} \|\mathbf{e}_{\Gamma(\mathcal{S})}\|_1.
\end{aligned}
$$

The proof of this lemma requires the following proposition which is a property of d-left regular bipartite graphs like (k, d, ϵ)-lossless and (k, d, ϵ)-model expander graphs, which is equivalent to the RIP-1 condition [7]. Remember that a d-left regular bipartite graph is a bipartite graph with each left vertex having degree d.

PROPOSITION 7.1. *Given a d-left regular bipartite graph with expansion coefficient ϵ, if $\mathcal{S}$ is a set of s left indices, then the set, $\Gamma'(\mathcal{S})$ of right vertices connected to exactly one left vertex in $\mathcal{S}$ has size*

$$
|\Gamma'(\mathcal{S})| \geq (1 - 2\epsilon)ds.
$$

Proof. **Proof of Proposition 7.1** By the expansion property of the graph we have $|\Gamma(\mathcal{S})| \geq (1 - \epsilon)d|\mathcal{S}|$ and by the definition of the sets $\Gamma'(\cdot)$ and $\Gamma''(\cdot)$ we have $|\Gamma(\mathcal{S})| = |\Gamma'(\mathcal{S})| + |\Gamma''(\mathcal{S})|$ which therefore means

$$
(7.26) \qquad |\Gamma'(\mathcal{S})| + |\Gamma''(\mathcal{S})| \geq (1 - \epsilon)d|\mathcal{S}|.
$$

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

$$
(7.27) \qquad |\Gamma'(\mathcal{S})| + 2|\Gamma''(\mathcal{S})| \leq |\mathcal{E}(\mathcal{S}, \Gamma(\mathcal{S}))| = d|\mathcal{S}|.
$$

Solving (7.26) and (7.27) simultaneously gives the bound $|\Gamma''(\mathcal{S})| \leq \epsilon d |\mathcal{S}|$ and thus $|\Gamma'(\mathcal{S})| \geq (1 - 2\epsilon)d|\mathcal{S}|$. Substituting $|\mathcal{S}|$ by s completes the proof.

Proof. **Proof of Lemma 7.1** 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 (3.6) as

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

where q_2 is a shorthand for the *second quantile (median) operator* that is $q_2 [b_1, \ldots, b_m] = b_{\pi(\lceil m/2 \rceil)}$.

Using this shorthand for the median, the left hand side of (7.25) can be written in the following way:

$$\| [\mathcal{M}(\mathbf{A}\mathbf{x}_{\mathcal{S}} + \mathbf{e}) - \mathbf{x}]_{\mathcal{S}} \|_1$$
$$= \sum_{i \in \mathcal{S}} \left| q_2 \left[(\mathbf{A}\mathbf{x}_{\mathcal{S}} + \mathbf{e})_j, \ j \in \Gamma(i) \right] - x_i \right|$$
$$= \sum_{i \in \mathcal{S}} \left| q_2 \left[(\mathbf{A}\mathbf{x}_{\mathcal{S}})_j + e_j - x_i, \ j \in \Gamma(i) \right] \right|.$$

The above can be bound as in (7.28) 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

$$(7.28) \quad \begin{aligned} & \| [\mathcal{M}(\mathbf{A}\mathbf{x}_{\mathcal{S}} + \mathbf{e}) - \mathbf{x}]_{\mathcal{S}} \|_1 \\ & \leq \sum_{i \in \mathcal{S}} q_2 \left[\left| (\mathbf{A}\mathbf{x}_{\mathcal{S}})_j + e_j - x_i \right|, \ j \in \Gamma(i) \right] \\ & = \sum_{i \in \mathcal{S}} q_2 \left[\left| \sum_{\substack{l \in \mathcal{S} \setminus \{i\} \\ e_{lj} \in \mathcal{E}}} x_l + e_j \right|, \ j \in \Gamma(i) \right], \end{aligned}$$

where the last equality is due to the property of the underlying expander graph of $\mathbf{A}$, say G, and remember $e_{lj} \in \mathcal{E}$ denotes an edge from node l to node j being in $\mathcal{E}$, the set of edges of G. Now, we proceed by induction on the cardinality of $\mathcal{S}$ i.e., $s = |\mathcal{S}|$ to show that the right hand side of (7.25) upper bounds the right hand side of (7.28).

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

$$(7.29) \quad q_2 [b_1, \ldots, b_m] \leq \frac{2(b_1 + \cdots + b_m)}{m}, \text{ if } \forall j \ b_j \geq 0,$$

we have the following stronger bound.

$$(7.30) \quad \begin{aligned} & \sum_{i \in \mathcal{S}} q_2 \left[\left| \sum_{\substack{l \in \mathcal{S} \setminus \{i\} \\ e_{lj} \in \mathcal{E}}} x_l + e_j \right|, \ j \in \Gamma(i) \right] \\ & = q_2 [|e_j|, j \in \Gamma(i)] \leq \frac{2}{d} \|\mathbf{e}_{\Gamma(i)}\|_1. \end{aligned}$$

Inductive step: We assume that the induction statement holds for $\mathcal{S}$ with $|\mathcal{S}| = s - 1$ ($s \geq 2$) and proceed to show that it also holds for $\mathcal{S}$ with $|\mathcal{S}| = s$. By the pigeonhole principle, given a $(s, d, \epsilon_{\mathcal{T}_s})$-model expander, Proposition 7.1 implies that there exists a variable node $i^* \in \mathcal{S}$ which is connected uniquely to at least $(1 - 2\epsilon_{\mathcal{T}_s})d$ check nodes. We will refer to this node as the *unique neighbor* index. Any set $\mathcal{S} \in \mathcal{T}_s$-sparse (with $|\mathcal{S}| = s$) can be partitioned into $\{i^*\}$ and $\mathcal{S} \setminus \{i^*\}$. We will bound (7.28) for (a) when $i \in \{i^*\}$, that is $i = i^*$, and (b) when $i \in \mathcal{R} := \mathcal{S} \setminus \{i^*\}$. Summing the bounds from (a) and (b) will result in an upper bound on (7.28) for the case $|\mathcal{S}| = s$ and this completes the proof.

(a) By the definition of q_2, there exist $\lceil d/2 \rceil$ distinct $j_1, \ldots j_{\lceil d/2 \rceil} \in \Gamma(i^*)$ (i^* is the unique neighbor node) such that, for all $h \in \{1, \ldots, \lceil d/2 \rceil\}$,

$$(7.31) \quad \begin{aligned} & q_2 \left[\left| \sum_{\substack{l \in \mathcal{S} \setminus \{i^*\} \\ e_{lj} \in \mathcal{E}}} x_l + e_j \right|, \ j \in \Gamma(i^*) \right] \\ & \leq \left| \sum_{\substack{l \in \mathcal{S} \setminus \{i^*\} \\ e_{lj_h} \in \mathcal{E}}} x_l + e_{j_h} \right|, \ j_h \in \Gamma(i^*). \end{aligned}$$

Note that there are no $l \in \mathcal{S} \setminus \{i^*\}$ with $e_{lj_h} \in \mathcal{E}$ since i^* is the unique neighbor index. Therefore, we can also bound (7.31) by the mean over the j_h. Let us define $\gamma := |\Gamma'(i^*)| \geq (1 - 2\epsilon_{\mathcal{T}_s})d$. There are at most $d - \gamma \leq 2\epsilon_{\mathcal{T}_s}d$ vertices in the collision set $\Gamma''(i^*)$ and there are at least $\lceil d/2 \rceil - (d - \gamma) \leq \lceil d/2 \rceil - 2\epsilon_{\mathcal{T}_s}d$ elements among the j_h that are in $\Gamma'(i^*)$. We use this fact to upper bound the mean over the j_h elements which in turn upper bounds (7.31) as follows

$$(7.32) \quad \begin{aligned} & q_2 \left[\left| \sum_{\substack{l \in \mathcal{S} \setminus \{i^*\} \\ e_{lj} \in \mathcal{E}}} x_l + e_j \right|, \ j \in \Gamma(i^*) \right] \\ & \leq \frac{1}{\lceil d/2 \rceil - (d - \gamma)} \|\mathbf{e}_{\Gamma'(i^*)}\|_1 \\ & \leq \frac{2}{(1 - 4\epsilon_{\mathcal{T}_s})d} \|\mathbf{e}_{\Gamma'(i^*)}\|_1 \end{aligned}$$

(b) If on the other hand $i \in \mathcal{R} := \mathcal{S} \setminus \{i^*\}$, firstly we rewrite

$$\left| \sum_{\substack{l \in \mathcal{S} \setminus \{i\} \\ e_{lj} \in \mathcal{E}}} x_l + e_j \right|, \ j \in \Gamma(i^*) \quad \text{as}$$

$$(7.33) \quad \left| \sum_{\substack{l \in \mathcal{R} \setminus \{i\} \\ e_{lj} \in \mathcal{E}}} x_l + \mathbb{I}_{\{e_{i^*j} \in \mathcal{E}\}} x_{i^*} + e_j \right|, \ j \in \Gamma(i^*),$$

where $\mathbb{I}_{\{\cdot\}}$ is an indicator function defined as follows

$$\mathbb{I}_{\{i\in\Omega\}} = \begin{cases} 1 & \text{if } i \in \Omega, \\ 0 & \text{if } i \notin \Omega. \end{cases}$$

Since we assume the induction statement to be true for $\mathcal{S}$ with $|\mathcal{S}| = s-1$, we replace $\mathcal{S}$ by $\mathcal{R}$ and e_j by $e_j' = \mathbb{I}_{\{e_{i^*j}\in\mathcal{E}\}}x_{i^*} + e_j$ and we use the fact that $\epsilon_{\mathcal{T}_{s-1}} \le \epsilon_{\mathcal{T}_s}$ to get an upper bound from (7.25) as follows:

$$(7.34) \quad \begin{aligned} &\sum_{i\in\mathcal{R}} q_2\Big[\Big|\sum_{\substack{l\in\mathcal{R}\setminus\{i\}\\e_{lj}\in\mathcal{E}}} x_l + e_j\Big|, \ j \in \Gamma(i)\Big] \\ &\le \frac{4\epsilon_{\mathcal{T}_s}d}{d-4\epsilon_{\mathcal{T}_s}d}\|\mathbf{x}_\mathcal{R}\|_1 + \frac{2}{(1-4\epsilon_{\mathcal{T}_s})d}\|\mathbf{e}'_{\Gamma(\mathcal{R})}\|_1, \end{aligned}$$

where $\mathbf{e}'$ is a vector with components e_j' for $j = 1,\ldots,m$. Next we upper bound $\|\mathbf{e}'_{\Gamma(\mathcal{R})}\|_1$ as follows

$$(7.35) \quad \begin{aligned} \|\mathbf{e}'_{\Gamma(\mathcal{R})}\|_1 &\le \sum_{j\in\Gamma(\mathcal{R})} \mathbb{I}_{\{e_{i^*j}\in\mathcal{E}\}} |x_{i^*}| + \|\mathbf{e}_{\Gamma(\mathcal{R})}\|_1 \\ &\le 2\epsilon_{\mathcal{T}_s}d\,|x_{i^*}| + \|\mathbf{e}_{\Gamma(\mathcal{R})}\|_1, \end{aligned}$$

where we used the fact that

$$\begin{aligned} \sum_{j\in\Gamma(\mathcal{R})} \mathbb{I}_{\{e_{i^*j}\in\mathcal{E}\}} &= \sum_{j\in\Gamma''(i^*)} \mathbb{I}_{\{e_{i^*j}\in\mathcal{E}\}} \\ &= |\Gamma''(i^*)| \le 2\epsilon_{\mathcal{T}_s}d, \end{aligned}$$

which is the cardinality of the collision set deducible from Proposition 7.1.

We now bound the last expression in (7.28), i.e., $\sum_{i\in\mathcal{S}} q_2\Big[\Big|\sum_{l\in\mathcal{S}\setminus\{i\}_{e_{lj}\in\mathcal{E}}} x_l + e_j\Big|, \ j \in \Gamma(i)\Big]$, which is the sum of the two cases (a) and (b) leading respectively to the bounds in (7.32) and (7.34). This sum can thus be rewritten as follows

$$(7.36) \quad \begin{aligned} &q_2\Big[\Big|\sum_{\substack{l\in\mathcal{S}\setminus\{i^*\}\\e_{lj}\in\mathcal{E}}} x_l + e_j\Big|, \ j \in \Gamma(i^*)\Big] \\ &+ \sum_{i\in\mathcal{R}} q_2\Big[\Big|\sum_{\substack{l\in\mathcal{S}\setminus\{i\}\\e_{lj}\in\mathcal{E}}} x_l + e_j\Big|, \ j \in \Gamma(i)\Big]. \end{aligned}$$

Now using the bounds in (7.32) and (7.34) with (7.35), we upper bound (7.36) by

$$\begin{aligned} &\frac{2}{(1-4\epsilon_{\mathcal{T}_s})d}\|\mathbf{e}_{\Gamma'(i^*)}\|_1 + \frac{4\epsilon_{\mathcal{T}_s}d}{d-4\epsilon_{\mathcal{T}_s}d}\|\mathbf{x}_\mathcal{R}\|_1 \\ &+ \frac{2}{(1-4\epsilon_{\mathcal{T}_s})d}\left(2\epsilon_{\mathcal{T}_s}d\,|x_{i^*}| + \|\mathbf{e}_{\Gamma(\mathcal{R})}\|_1\right), \end{aligned}$$

which in turn can be bounded by

$$(7.37) \quad \frac{4\epsilon_{\mathcal{T}_s}}{1-4\epsilon_{\mathcal{T}_s}}\|\mathbf{x}_\mathcal{S}\|_1 + \frac{2}{(1-4\epsilon_{\mathcal{T}_s})d}\|\mathbf{e}_{\Gamma(\mathcal{S})}\|_1.$$

The upper bound (7.37) is due to the fact that $\Gamma(\mathcal{R})$ and $\Gamma'(i^*)$ are two disjoint subsets of $\Gamma(\mathcal{S})$ and this completes the proof.

7.2 Proof of Theorem 3.1

Proof. We use Lemma 7.1 to prove Theorem 3.1. In order to show that (3.8) holds, we show that (7.38) below holds and by induction this leads to (3.8).

$$(7.38) \quad \begin{aligned} &\|\mathbf{x}^{n+1} - \mathbf{x}_\mathcal{S}\|_1 \\ &\le \alpha\|\mathbf{x}^n - \mathbf{x}_\mathcal{S}\|_1 + (1-\alpha)\beta\|\mathbf{A}\mathbf{x}_{\bar{\mathcal{S}}} + \mathbf{e}\|_1. \end{aligned}$$

Let us define $\mathcal{Q}^{n+1} := \mathcal{S} \cup \mathrm{supp}(\mathbf{x}^n) \cup \mathrm{supp}(\mathbf{x}^{n+1})$ and $\mathbf{u}^{n+1} := (\mathbf{x}^n + \mathcal{M}(\mathbf{y} - \mathbf{A}\mathbf{x}^n))_{\mathcal{Q}^{n+1}}$. Since we are able to do exact projections (in the ℓ_1 norm) onto the given model, then $\mathbf{x}^{n+1}$ is a better $\mathcal{M}_s$-sparse approximation to $\mathbf{u}^{n+1}$ than is $\mathbf{x}_\mathcal{S}$, i.e.,

$$\|\mathbf{u}^{n+1} - \mathbf{x}^{n+1}\|_1 \le \|\mathbf{u}^{n+1} - \mathbf{x}_\mathcal{S}\|_1.$$

With this fact we use the triangle inequality to have

$$(7.39) \quad \begin{aligned} &\|\mathbf{x}^{n+1} - \mathbf{x}_\mathcal{S}\|_1 \\ &\le \|\mathbf{x}^{n+1} - \mathbf{u}^{n+1}\|_1 + \|\mathbf{u}^{n+1} - \mathbf{x}_\mathcal{S}\|_1 \\ &\le 2\|\mathbf{u}^{n+1} - \mathbf{x}_\mathcal{S}\|_1. \end{aligned}$$

We substitute the value of $\mathbf{u}^{n+1}$ and use the fact that $\mathcal{S} \subseteq \mathcal{Q}^{n+1}$ to upper bound (7.39) by the following

$$(7.40) \quad 2\|\left[\mathbf{x}_\mathcal{S} - \mathbf{x}^n - \mathcal{M}\left(\mathbf{A}\left(\mathbf{x}^n - \mathbf{x}_\mathcal{S}\right) + \bar{\mathbf{e}}\right)\right]_{\mathcal{Q}^{n+1}}\|_1,$$

where $\bar{\mathbf{e}} := \mathbf{A}\mathbf{x}_{\bar{\mathcal{S}}} + \mathbf{e}$. Next we apply Lemma 7.1 to bound (7.40) by

$$(7.41) \quad \frac{8\epsilon_{\mathcal{M}_{3s}}}{1-4\epsilon_{\mathcal{M}_{3s}}}\|\mathbf{x}_\mathcal{S} - \mathbf{x}^n\|_1 + \frac{4}{(1-4\epsilon_{\mathcal{M}_{3s}})d}\|\bar{\mathbf{e}}\|_1,$$

where we use $\mathcal{M}_{3s}$ because $\mathcal{Q}^{n+1} \in \mathcal{M}_{3s}$ due to the fact that our models have the *nestedness property* (see Section 3.4). This bound is the desired upper bound in (7.38), with $\alpha := \frac{8\epsilon_{\mathcal{M}_{3s}}}{(1-4\epsilon_{\mathcal{M}_{3s}})}$ and $\beta := \frac{4}{(1-12\epsilon_{\mathcal{M}_{3s}})d}$.

Then, we use induction on the bound in (7.38) to get the following bound

$$\begin{aligned} &\|\mathbf{x}^n - \mathbf{x}_\mathcal{S}\|_1 \\ &\le \alpha^n\|\mathbf{x}^0 - \mathbf{x}_\mathcal{S}\|_1 + \sum_{k=0}^{n-1} \alpha^k(1-\alpha^n)\beta\|\mathbf{A}\mathbf{x}_{\bar{\mathcal{S}}} + \mathbf{e}\|_1. \end{aligned}$$

Simplifying the geometric series leads to (3.8). To have $\alpha < 1$ and $\beta > 0$ we need $\epsilon_{\mathcal{M}_{3s}} < 1/12$. This concludes the proof.

7.3 Proof of Corollary 3.1

Proof. Let the sequence $(\mathbf{x}^n)$ for $n \geq 0$ converges to $\mathbf{x}^\star$. But, as $n \to \infty$, $\alpha^n \to 0$, the first term in (3.8) goes to zero, and hence, the term multiplying β in the second terms becomes 1. Therefore, by the triangle inequality we have the following

$$\|\mathbf{x}^\star - \mathbf{x}_\mathcal{S}\|_1 \leq \beta \|\mathbf{A}\mathbf{x}_{\bar{\mathcal{S}}}\|_1 + \beta \|\mathbf{e}\|_1.$$

Note that $\|\mathbf{A}\mathbf{x}_{\bar{\mathcal{S}}}\|_1 \leq d\|\mathbf{x}_{\bar{\mathcal{S}}}\|_1$ holds irrespective of $\mathbf{x}_{\bar{\mathcal{S}}}$ being sparse or not. This is a direct consequence of expanders and an indirect one of model RIP-1 matrices. Therefore, we bound our error as thus

$$\|\mathbf{x}^\star - \mathbf{x}\|_1 \leq \|\mathbf{x}_{\bar{\mathcal{S}}}\|_1 + \|\mathbf{x}^\star - \mathbf{x}_\mathcal{S}\|_1$$
$$\leq (1 + \beta d)\sigma_{\mathcal{M}_s}(\mathbf{x})_1 + \beta \|\mathbf{e}\|_1.$$

Taking $C_1 = 1 + \beta d$ and $C_2 = \beta$ completes the proof.

References

[1] K.J. Ahn, S. Guha, and A. McGregor, *Graph sketches: sparsification, spanners, and subgraphs*, in Proceedings of the 31st symposium on Principles of Database Systems, ACM, 2012, pp. 5–14.

[2] B. Bah and J. Tanner, *Vanishingly sparse matrices and expander graphs, with application to compressed sensing*, arXiv preprint arXiv:1207.3094, (2012).

[3] L. Baldassarre, N. Bhan, V. Cevher, and A. Kyrillidis, *Group-sparse model selection: Hardness and relaxations*, arXiv preprint arXiv:1303.3207v3, (2013).

[4] R.G. Baraniuk, V. Cevher, M.F. Duarte, and C. Hegde, *Model-based compressive sensing*, Information Theory, IEEE Transactions on, 56 (2010), pp. 1982–2001.

[5] R.G. Baraniuk and D.L. Jones, *A signal-dependent time-frequency representation: fast algorithm for optimal kernel design*, Signal Processing, IEEE Transactions on, 42 (1994), pp. 134–146.

[6] L.A. Bassalygo and M.S. Pinsker, *Complexity of an optimum nonblocking switching network without reconnections*, Problemy Peredachi Informatsii, 9 (1973), pp. 84–87.

[7] R. Berinde, A.C. Gilbert, P. Indyk, H. Karloff, and M.J. Strauss, *Combining geometry and combinatorics: A unified approach to sparse signal recovery*, in Communication, Control, and Computing, 2008 46th Annual Allerton Conference on, IEEE, 2008, pp. 798–805.

[8] R. Berinde, P. Indyk, and M. Ruzic, *Practical near-optimal sparse recovery in the ℓ_1-norm*, in 46th Annual Allerton Conference on Communication, Control, and Computing, IEEE, 2008, pp. 198–205.

[9] H. Buhrman, P. B. Miltersen, J. Radhakrishnan, and S. Venkatesh, *Are bitvectors optimal?*, SIAM Journal on Computing, 31 (2002), pp. 1723–1744.

[10] E.J. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, Information Theory, IEEE Transactions on, 52 (2006), pp. 489–509.

[11] M. Capalbo, O. Reingold, S. Vadhan, and A. Wigderson, *Randomness conductors and constant-degree lossless expanders*, in Proceedings of the thirty-fourth annual ACM symposium on Theory of computing, ACM, 2002, pp. 659–668.

[12] C. Cartis and A. Thompson, *An exact tree projection algorithm for wavelets*, submitted.

[13] D.L. Donoho, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306.

[14] M.F. Duarte, S. Sarvotham, M.B. Wakin, D. Baron, and R.G. Baraniuk, *Joint sparsity models for distributed compressed sensing*, in Proceedings of the Workshop on Signal Processing with Adaptive Sparse Structured Representations, 2005.

[15] C. Dwork, F. McSherry, and K. Talwar, *The price of privacy and the limits of lp decoding*, in Proceedings of the thirty-ninth annual ACM symposium on Theory of computing, ACM, 2007, pp. 85–94.

[16] Y.C. Eldar and M. Mishali, *Robust recovery of signals from a structured union of subspaces*, Information Theory, IEEE Transactions on, 55 (2009), pp. 5302–5316.

[17] M. Fornasier and H. Rauhut, *Recovery algorithms for vector-valued data with joint sparsity constraints*, SIAM Journal on Numerical Analysis, 46 (2008), pp. 577–613.

[18] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*, Applied Numerical Harmonic Analysis Birkhäuser, Boston, (2013).

[19] A.C. Gilbert and P. Indyk, *Sparse recovery using sparse matrices*, Proceedings of the IEEE, 98 (2010), pp. 937–947.

[20] A.C. Gilbert and K. Levchenko, *Compressing network graphs*, in Proceedings of the LinkKDD workshop at the 10th ACM Conference on KDD, 2004.

[21] A. Gramfort and M. Kowalski, *Improving m/eeg source localizationwith an inter-condition sparse prior*, in Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on, IEEE, 2009, pp. 141–144.

[22] V. Guruswami, C. Umans, and S. Vadhan, *Unbalanced expanders and randomness extractors from parvaresh–vardy codes*, Journal of the ACM (JACM), 56 (2009), p. 20.

[23] C. Hegde, M.F. Duarte, V. Cevher, et al., *Compressive sensing recovery of spike trains using a structured sparsity model*, in SPARS'09-Signal Processing with Adaptive Sparse Structured Representations, 2009.

[24] P. Indyk, *Sketching, streaming and sublinear-space algorithms*, Graduate course notest, (2007).

[25] P. Indyk and E. Price, *K-median clustering, model-based compressive sensing, and sparse recovery for earth mover distance*, in Proceedings of the 43rd annual ACM symposium on Theory of computing, ACM, 2011, pp. 627–636.

[26] P. Indyk and I. Razenshteyn, *On model-based RIP-1 matrices*, arXiv preprint arXiv:1304.3604, (2013).

[27] L. Jacob, G. Obozinski, and J.-P. Vert, *Group lasso with overlap and graph lasso*, in Proceedings of the 26th Annual International Conference on Machine Learning, ACM, 2009, pp. 433–440.

[28] S. Jafarpour, W. Xu, B. Hassibi, and R. Calderbank, *Efficient and robust compressed sensing using optimized expander graphs*, Information Theory, IEEE Transactions on, 55 (2009), pp. 4299–4308.

[29] R. Jenatton, J. Mairal, G. Obozinski, and F. Bach, *Proximal methods for hierarchical sparse coding*, Journal of Machine Learning Reasearch, 12 (2011), pp. 2297–2334.

[30] W. Lang, *Combinatorial interpretation of generalized stirling numbers*, J. Integer Seq, 12 (2009), p. 3.

[31] S. Muthukrishnan, *Data streams: Algorithms and applications*, Now Publishers Inc, 2005.

[32] J. Nocedal and S.J. Wright, *Numerical optimization*, vol. 2, Springer New York, 1999.

[33] E. Price, *Efficient sketches for the set query problem*, in Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, 2011, pp. 41–56.

[34] N.S. Rao, B. Recht, and R.D. Nowak, *Signal recovery in unions of subspaces with applications to compressive imaging*, arXiv preprint arXiv:1209.3079, (2012).

[35] M. Stojnic, F. Parvaresh, and B. Hassibi, *On the reconstruction of block-sparse signals with an optimal number of measurements*, Signal Processing, IEEE Transactions on, 57 (2009), pp. 3075–3085.

[36] M. Yuan and Y. Lin, *Model selection and estimation in regression with grouped variables*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68 (2006), pp. 49–67.