Tight Measurement Bounds for Exact Recovery of Structured Sparse Signals

Nikhil Rao[1]   Benjamin Recht[2]   Robert Nowak[1]

[1] Department of Electrical and Computer Engineering
[2] Department of Computer Sciences
University of Wisconsin-Madison

Abstract

Standard compressive sensing results state that to exactly recover an s sparse signal in $\mathbb{R}^p$, one requires $\mathcal{O}(s \cdot \log p)$ measurements. While this bound is extremely useful in practice, often real world signals are not only sparse, but also exhibit structure in the sparsity pattern. We focus on group-structured patterns in this paper. Under this model, groups of signal coefficients are active (or inactive) together. The groups are predefined, but the particular set of groups that are active (i.e., in the signal support) must be learned from measurements. We show that exploiting knowledge of groups can further reduce the number of measurements required for exact signal recovery, and derive universal bounds for the number of measurements needed. The bound is universal in the sense that it only depends on the number of groups under consideration, and not the particulars of the groups (e.g., compositions, sizes, extents, overlaps, etc.). Experiments show that our result holds for a variety of overlapping group configurations.

1   Introduction

In many fields such as genetics, image processing, and machine learning, one is faced with the task of recovering very high dimensional signals from relatively few measurements. In general this is not possible, but fortunately many real world signals are, or can be transformed to be, sparse, meaning that only a small fraction signal coefficients are non-zero. Compressed Sensing [3, 6] allows us to recover sparse, high dimensional signals with very few measurements. In fact, results indicate that one only needs $\mathcal{O}(s \cdot \log p)$ random measurements to exactly recover an s sparse signal of length p.

In many applications however, one not only has knowledge about the sparsity of the signal, but some additional information about the structure of the sparsity pattern as well:

- In genetics, the genes are arranged into pathways, and genes belonging to the same pathway are highly correlated with each other [22].

- In image processing, the wavelet transform coefficients can be modeled as belonging to a tree, with parent-child coefficients exhibiting similar properties [5, 21, 19].

- In wideband spectrum sensing applications, the spectrum typically displays clusters of non-zero frequency coefficients, each corresponding to a narrowband transmission [15]

In cases such as these, the sparsity pattern can be represented as a union of certain groups of coefficients (e.g., coefficients in certain pathways, tree branches, or clusters). This knowledge about the signal structure can help further reduce the number of measurements one needs to exactly recover the signal. Indeed, the authors in [10] derive information theoretic bounds for the number of measurements needed for a variety of

1

signal ensembles, including trees. In, [2, 7], the authors show that one needs far fewer measurements when the signal can be expressed as lying in a union of subspaces, and explicit bounds are derived when using a modified version of CoSaMP [17] to recover the signal. In this paper, we derive bounds on the number of random iid gaussian measurements needed to exactly recover a sparse signal when its pattern of sparsity lies in a union of groups, when solving the *convex* recovery algorithm introduced in [11].

We analyze the group-structured sparse recovery problem using a random Gaussian measurement model. We emphasize that although the derivation assumes the measurement matrix to be Gaussian, it can be extended to *any* subgaussian case, by paying a small constant penalty, as shown in [14]. We restrict ourselves to the Gaussian case here since it highlights the main ideas and keeps the analysis as simple as possible.

Note that in this work, variables can be grouped into arbitrary sets, and we make *no* assumptions about the nature of the groups, except that they are known in advance. In short, we derive bounds for any generic group structure of variables, whether the groups overlap or form a partition of the ambient high dimensional space.

To the best of our knowledge, these results are new and distinct from prior theoretical characterizations of group lasso methods. Asymptotic consistency results are derived for the group lasso when the groups partition the space of variables in [1]. Similarly, in [9], the authors consider the groups to partition the space, and derive conditions for recovery using the group lasso [25]. In [12, 13], the authors derive consistency results for the group lasso under arbitrary groupings of variables. In [18], the authors consider overlapping groups and derive sample bounds under the group lasso [25] setting. The authors in [11] derive consistency results in an asymptotic setting, for the group lasso with overlap, but do not provide exact recovery results. The general group lasso scenarios is different from what we consider, in that the group lasso yields vectors whose support can be expressed as a complement of a union of groups, while we consider cases where we require the support to lie in a union of groups, a distinction made in [11]. Note that in the case of non-overlapping groups, the complement of a union of groups is a union of (a different set of) groups. In this paper, we (a) derive sample complexity bounds in a compressive-sensing framework when the measurement matrix is *i.i.d.* gaussian. (b) We focus on non-asymptotic sample bounds, and in a case where the support is contained in a union of groups, and (c) make no assumptions about the nature of groups. To derive our results, we appeal to the notion of restricted minimum singular values of an operator.

We bound number of measurements needed for exact recovery with two terms. The first term grows linearly in the total number of non-zero coefficients (with a small constant of proportionality). This is close to the bare minimum of one measurement per non-zero component. The second term only depends on the number of groups under consideration, and not the particulars of the groups (e.g., compositions, sizes, extents, etc.). In particular, the groups need not be disjoint. The degree to which groups overlap, remarkably, has no effect on our bounds. In this regard, our bounds can be termed to be *universal*. This is somewhat surprising since overlapping groups are strongly coupled in the observations, tempting one to suppose that overlap may make recovery more challenging.

Our main result shows that for signals with support on k of M possible groups, exact recovery is possible from $(\sqrt{2\log(M-k)} + \sqrt{B})^2 k + kB$ measurements using an overlapping group lasso algorithm, B being the maximum group size. Note that the bound depends on the sparsity s of the signal via the kB term, which is a loose upper bound for s when the groups highly overlap. This arises as an artifact of the general approach we use to bound the number of measurements, and in specific cases, this can be made much tighter.

Our proof derives from the techniques developed in [4]. The rest of this paper is organized as follows: in Section 2, we lay the groundwork for the main contribution of the paper, *viz.* applying the techniques from [4] tot he specific setting of group lasso with overlapping groups. We describe the theory and reasoning behind this approach. In Section 3 we derive bounds on the number of random *i.i.d.* gaussian measurements needed to be taken for exact recovery of group sparse signals. We further derive bounds for the number of measurements required for robust recovery of signals as well. Section 4 outlines the experiments we performed and the corresponding results obtained. We conclude our paper in Section 5.

1.1 Notations

We first introduce notations that we will use for the rest of the paper. Consider a signal of length p, that is s sparse. Note here that in case of multidimensional signals like images, we assume they are vectorized to have length p. The coefficients of the signal are grouped into sets $\{G_i\}_{i=1}^M$, such that $\forall i \in \{1, 2, \cdots, M\}, G_i \subset \{1, 2, \cdots, p\}$. We denote the set of groups by $\mathcal{G} = \{G_i\}_{i=1..M}$, and $|\cdot|$ denotes the cardinality of a set. We let $x^\star$ be the (sparse) signal to be recovered, whose non zero coefficients lie in k of the M groups $\mathcal{G}^\star \subset \mathcal{G}$. Formally,

$$\mathcal{G}^\star = \{G_i \in \mathcal{G}^\star : \text{supp}(x^\star) \cap G_i \neq 0\}$$

We assume $|\mathcal{G}^\star| = k \leq M = |\mathcal{G}|$. We let $\Phi_{n \times p}$ be a measurement matrix consisting of *i.i.d.* gaussian entries of mean 0 and unit variance so that every column is a realization of an *i.i.d.* gaussian length n vector with covariance I. For any vector $x \in \mathbb{R}^p$, we denote by x_G a vector in $\mathbb{R}^p$ such that $(x_G)_i = x_i$ if $i \in G$, and 0 otherwise. We denote the observed vector by $y \in \mathbb{R}^n$: $y = \Phi x^\star$. The absence of a subscript following a norm $\| \cdot \|$ implies the $\ell 2$ norm. The dual norm of $\| \cdot \|_p$ is denoted by $\| \cdot \|_p^\ast$. The convex hull of a set of points S is denoted by $\text{conv}(S)$.

2 Preliminaries

In this Section, we will set up the problem that we wish to solve in this paper. We will argue as to why exact recovery of the signal corresponds to the minimization of the atomic norm of the signal, with the atoms obeying certain properties governed by the signal structure.

2.1 Atoms and the atomic set

To begin with, let us formalize the notion of atoms and the atomic norm of a signal (or vector). We will restrict our attention to group-sparse signals in $\mathbb{R}^p$, though the same concepts can be extended to other spaces as well. We assume that $x \in \mathbb{R}^p$ can be decomposed as :

$$x = \sum_{i=1}^k c_i a_i, \quad c_i \geq 0$$

The vectors a_i are called *atoms*, and form the basic building blocks of any signal, which can be represented as a conic combination of the atoms. We denote $\mathcal{A} = \{a_i\}$ to be the *atomic set*. Given a vector $x \in \mathbb{R}^p$ and an atomic set, we define the *atomic norm* as

$$||x||_\mathcal{A} = \inf \left\{ \sum_{a \in \mathcal{A}} c_a : x = \sum_{a \in \mathcal{A}} c_a a, \quad c_a \geq 0 \ \ \forall a \in \mathcal{A} \right\} \tag{1}$$

The atomic decomposition of the signal has been known to be the simplest representation of the signal in some sense. Hence, to obtain a "simple" representation of a vector, we look to minimize the atomic norm subject to constraints (equation (2)):

$$\hat{x} = \underset{x \in \mathbb{R}^p}{\text{argmin}} ||x||_\mathcal{A} \ \ \textbf{s.t.} \ \ y = \Phi x \tag{2}$$

Indeed, when the atoms are merely the canonical basis in $\mathbb{R}^p$, the atomic norm reduces to the standard ℓ_1 norm, and minimization of the atomic norm yields the well known *lasso* procedure [23].

Assuming we are aware of the group structure $\mathcal{G}$, we now proceed to define the atomic set and the corresponding atomic norm for our framework:

$$\forall G \in \mathcal{G}, \ \text{let}$$
$$A_G = \{a^G \in \mathbb{R}^p : ||(a^G)_G||_2 = 1, (a^G)_{G^c} = 0\}$$

$$\mathcal{A} = \{A_G\}_{G \in \mathcal{G}} \tag{3}$$

We now show that the atomic norm of a vector $x \in \mathbb{R}^p$ under the atomic set defined in equation (3) is equivalent to the overlapping group lasso norm defined in [11], a special case of which is the standard group lasso norm [25]. Thus, minimizing the atomic norm in this case is exactly the same as the group lasso with overlapping groups.

Lemma 2.1 *Given any arbitrary set of groups $\mathcal{G}$, we have*

$$||x||_\mathcal{A} = \Omega^\mathcal{G}_{overlap}(x)$$

where $\Omega^\mathcal{G}_{overlap}(x)$ is the overlapping group lasso norm defined in [11].

Proof In (1), we can substitute $v_G = c_G a$, giving us $c_G = |c_G| \cdot ||a|| = ||c_G a|| = ||v_G||$. Hence,

$$||x||_\mathcal{A} = inf \left\{ \sum_{a \in \mathcal{A}} c_a : x = \sum_{a \in \mathcal{A}} c_a a \ \ c_a \geq 0 \ \ \forall a \in \mathcal{A} \right\}$$

$$= inf \left\{ \sum_{G \in \mathcal{G}} ||v_G|| \ \ : x = \sum_{G \in \mathcal{G}} v_G \right\}$$

$$= \Omega^\mathcal{G}_{overlap}(x) \quad \blacksquare$$

Corollary 2.2 *Under the atomic set defined in 3 , when $\mathcal{G}$ partitions $\mathbb{R}^p$,*

$$||x||_\mathcal{A} = \sum_{G \in \mathcal{G}} ||x_G||$$

Proof $\Omega^\mathcal{G}_{overlap} = \sum_{G \in \mathcal{G}} ||x_G||$ in the non overlapping case. $\quad \blacksquare$

Thus, (2) yields:

$$\hat{x} = \underset{x \in \mathbb{R}^p}{\operatorname{argmin}} \ \ \Omega^\mathcal{G}_{overlap}(x) \ \ \textbf{s.t.} \ \ y = \Phi x \tag{4}$$

which can be solved using [11].

Also note that we can directly compute the dual of the atomic norm from the set of atoms

$$||u||^*_\mathcal{A} = \sup_{a \in \mathcal{A}} \langle a, u \rangle = \max_{G \in \mathcal{G}} ||u_G|| \tag{5}$$

The dual norm will be useful in our derivations below.

2.2 Gaussian Widths and Exact Recovery

Following [4], we define the *tangent cone* and *normal cone* at $x^\star$ with respect to $conv(\mathcal{A})$ under $||x||_\mathcal{A}$ as [20]:

$$\mathcal{T}_\mathcal{A}(x^\star) = \operatorname{cone}\{z - x^\star \ : ||z||_\mathcal{A} \leq ||x^\star||_\mathcal{A}\} \tag{6}$$

$$\mathcal{N}_\mathcal{A}(x^\star) = \{u \ : \ \langle u, z \rangle \leq 0, \ \ \forall z \in \mathcal{T}_\mathcal{A}(x^\star)\} \tag{7}$$

$$= \{u \ : \ \langle u, x^\star \rangle = t||x||_\mathcal{A}$$

$$\text{and } ||u||^*_\mathcal{A} \leq t \text{ for some } t \geq 0\}$$

We note that, from [4] (Prop. 2.1), $\hat{x} = x^\star$ (2) is unique *iff*

$$null(\Phi) \cap \mathcal{T}_\mathcal{A}(x^\star) = \{0\} \tag{8}$$

Hence, we require that the tangent cone at $x^\star$ intersects the nullspace of Φ only at the origin, to guarantee exact recovery.

Before we state the main recovery result from [4], we define the *gaussian width* of a set:

Definition Let $\mathbb{S}^{p-1}$ denote the unit sphere in $\mathbb{R}^p$. The Gaussian width $\omega(S)$ of a set $S \in \mathbb{S}^{p-1}$ is

$$\omega(S) = \mathbb{E}_g \left[\sup_{z \in S} g^T z \right]$$

where $g \sim \mathcal{N}(0, I)$

Gordon uses the Gaussian width to provide bounds on the probability that a random subspace of a certain dimension misses a subset of the sphere [8]. In [4], these results are specialized to the case of atomic norm recovery. In particular, we will make use of the following:

Proposition 2.3 *[[4], Corollary 3.2] Let $\Phi : \mathbb{R}^p \to \mathbb{R}^n$ be a random map with i.i.d. zero-mean Gaussian entries having variance $1/n$. Further let $\Omega = T_{\mathcal{A}}(x^*) \cap \mathbb{S}^{p-1}$ denote the spherical part of the tangent cone $T_{\mathcal{A}}(x^*)$. Suppose that we have measurements $y = \Phi x^\star$, and we solve the convex program (2). Then $x^\star$ is the unique optimum of (2) with high probability provided that*

$$n \geq \omega(\Omega)^2 + \mathcal{O}(1).$$

To complete our problem setup we will also restate Proposition 3.6 in [4] :

Proposition 2.4 *(Proposition 3.6 in [4]) Let C be any non-empty convex cone in $\mathbb{R}^p$, and let $g \sim \mathcal{N}(0, I)$ be a Gaussian vector. Then:*

$$\omega(C \cap \mathbb{S}^{p-1}) \leq \mathbb{E}_g[\text{dist}(g, C^*)] \tag{9}$$

where $\text{dist}(.,.)$ *denotes the Euclidean distance between a point and a set, and C^* is the dual cone of C*

We can then square (9) use Jensen's inequality to obtain

$$\omega(C \cap \mathbb{S}^{p-1})^2 \leq \mathbb{E}_g[\text{dist}(g, C^*)^2] \tag{10}$$

We note here that the dual cone of the Tangent cone is the Normal cone, and vice-versa.

Thus, to derive measurement bounds, we only need to calculate the square of the gaussian width of the interSection of the tangent cone at $x^\star$ with respect to the atomic norm and the unit sphere. This value can be bounded by the distance of a gaussian random vector to the Normal cone at the same point, as implied by (10). In the next Section, we derive bounds on this quantity.

3   Gaussian Width of the Normal Cone of the Group Sparsity Norm

For generic groups $\mathcal{G}$, we have

$$v \in \mathcal{N}_{\mathcal{A}}(x^\star) \Leftrightarrow \exists \gamma \geq 0 \quad : \langle v, x^\star \rangle = \gamma \|x^\star\|_{\mathcal{A}}, \|v_G\| = \gamma \text{ if } G \in \mathcal{G}^\star, \quad \|v_G\| \leq \gamma \text{ if } G \notin \mathcal{G}^\star. \tag{11}$$

It is not hard to see that, in the case of disjoint groups,

$$\mathcal{N}_{\mathcal{A}}(x^\star) = \{z \in \mathbb{R}^p : z_i = \gamma \frac{(x^\star)_i}{\|x^\star\|} \quad \forall G \in \mathcal{G}^\star, \tag{12}$$

$$\|z_G\| \leq \gamma \quad \forall G \notin \mathcal{G}^\star, \gamma \geq 0\}$$

However, in the case of overlapping groups, we do not know how to obtain such a closed from.

We now prove the main result of this paper, a sufficient number of gaussian measurements needed to recover a group-sparse signal:

Theorem 3.1 *To exactly recover a k- group sparse signal decomposed into M groups in $\mathbb{R}^p$, $(\sqrt{2\log(M-k)} + \sqrt{B})^2 k + kB$ iid gaussian measurements are sufficient.*

To prove this result, we need two lemmas:

Lemma 3.2 *Let $q_1, \ldots, q_L$ be L, χ-squared random variables with d-degrees of freedom. Then*

$$\mathbb{E}[\max_{1 \leq i \leq L} q_i] \leq (\sqrt{2\log(L)} + \sqrt{d})^2.$$

Proof Let $M_L := \max_{1 \leq i \leq L} q_i$. For $t > 0$, we have that

$$\mathbb{E}[M_L] = \frac{\log[\exp(t \cdot \mathbb{E}[M_L])]}{t}$$
$$\overset{(i)}{\leq} \frac{\log[\mathbb{E}[\exp(t \cdot M_L)]]}{t}$$
$$\overset{(ii)}{=} \frac{\log[\mathbb{E}[\max_{1 \leq j \leq L} \exp(t \cdot q_j)]]}{t}$$
$$\overset{(iii)}{\leq} \frac{\log[L\mathbb{E}[\exp(t \cdot q_1)]]}{t}$$
$$= \frac{\log(L) - \frac{d}{2}\log(1 - 2t)}{t}$$

Where (i) follows from Jensen's inequality , (ii) follows from the monotonicity of the exponential function, and (iii) merely bounds the maximum by the sum over all the elements. Now, setting $t = (2 + 2\epsilon)^{-1}$ with $\epsilon = \sqrt{\frac{d}{2\log(L)}}$ yeilds $\mathbb{E}[M_L] \leq (\sqrt{2\log(L)} + \sqrt{d})^2$ ∎

Note that t can be optimized depending on the application. We use this particular choice because it makes no assumptions about the relative magnitudes of $(M - k)$ and B.

Lemma 3.3 *Suppose $v \in \mathbb{R}^p$ is supported on some set of groups $\mathcal{G}^\star \subset \mathcal{G}$*

$$\|v\| \leq \sqrt{|\mathcal{G}^\star|} \ \|v\|_{\mathcal{A}}^*.$$

Proof By duality, it suffices to show that $\|z\|_{\mathcal{A}} \leq \sqrt{|\mathcal{G}^\star|} \ \|z\|$ for all z. For any z, there exists a representation $z = \sum_{G \in \mathcal{G}^\star} b_G$ where none of the supports of b_G overlap. It then follows that

$$\|z\|_{\mathcal{A}} \overset{(i)}{\leq} \sum_{G \in \mathcal{G}^\star} \|b_G\|$$
$$\overset{(ii)}{\leq} \sqrt{|\mathcal{G}^\star|} \left(\sum_{G \in \mathcal{G}^\star} \|b_G\|^2 \right)^{1/2}$$
$$= \sqrt{|\mathcal{G}^\star|} \ \|z\|$$

Where (i) follows from the definition of the norm $\|\cdot\|_{\mathcal{A}}$ and (ii) is a consequence of the relation $\|\beta\|_1 \leq \sqrt{k}\|\beta\|_2$ for k dimensional vectors β ∎

Proof of Theorem 3.1 *Intuition*: Note that, from (10), we need to bound the distance between $\mathcal{N}_{\mathcal{A}}(x^\star)$ and a random gaussian vector. In the following proof, we carefully construct a specific vector $r \in \mathcal{N}_{\mathcal{A}}(x^\star)$ and bound the distance from r to the gaussian vector. Naturally, this will be an upper bound to the distance desired.

Now, let $S = \cup_{G \in \mathcal{G}^*} G$, *i.e.* S is the indices corresponding to the union of groups that support x^*. Note that $S \subset \{1, 2, \ldots, p\}$. Since the normal cone is nonempty, let $v \in \mathcal{N}_{\mathcal{A}}(x^*)$ and let

$$\|v\|_{\mathcal{A}}^* = 1 \tag{13}$$

then we must have that $\langle v, x^* \rangle = \|x^*\|_{\mathcal{A}}$. Moreover, for each G that intersects S, $\|v_G\| = 1$. This follows from the definition in (11). Also, let $v_{S^c} = 0$. It can be verified that v satisfies all the properties in (11).

Let $w \sim \mathcal{N}(0, I_p)$ be a vector with iid gaussian entries. We can write $w = [w_S \quad w_{S^c}]^T$. Let $t(w) = \max_{G \notin \mathcal{G}^*} \|w_G\|$.

Let us now construct a vector $r \in \mathcal{N}_{\mathcal{A}}(x^*)$. We can decompose r as $r = [r_S \quad r_{S^c}]^T$. Let $r_S = t(w) \cdot v_S$, and $r_{S^c} = w_{S^c}$.

From (11), and from our definition of $t(w)$, we have $r \in \mathcal{N}_{\mathcal{A}}(x^*)$. Referring to (10), we now consider the expected squared distance between $\mathcal{N}_{\mathcal{A}}(x^*)$ and w:

$$
\begin{aligned}
\mathbb{E}[\text{dist}(w, C^*)] &\leq \mathbb{E}[\|r - w\|^2] \\
&= \mathbb{E}[\|r_S - w_S\|^2 + \|r_{S^c} - w_{S^c}\|^2] \\
&\overset{(i)}{=} \mathbb{E}[\|r_S - w_S\|^2] \\
&\overset{(ii)}{=} \mathbb{E}[\|r_S\|^2] + \mathbb{E}[\|w_S\|^2] \\
&= \mathbb{E}[\|t(w) \cdot v_S\|^2] + \mathbb{E}[\|w_S\|^2] \\
&\overset{(iii)}{=} \mathbb{E}[t(w)^2] \cdot \|v_S\|^2 + \mathbb{E}[\|w_S\|^2] \\
&\overset{(iv)}{=} \mathbb{E}[t(w)^2] \cdot \|v_S\|^2 + |S| \\
&\overset{(v)}{\leq} (\sqrt{2\log(M - k)} + \sqrt{B})^2 \cdot \|v_S\|^2 + kB \\
&\overset{(vi)}{\leq} (\sqrt{2\log(M - k)} + \sqrt{B})^2 \cdot k + kB
\end{aligned}
$$

Where (i) follows because S and S^c are disjoint, (ii) follows from the fact that r_S and w_S are independent, (iii) follows from the fact that v is deterministic. We obtain (iv) since $\|w_S\|^2$ is a χ^2 random variable with $|S|$ degrees of freedom. (v) follows from Lemma 3.2, and from the fact that kB is a upper bound on the signal sparsity. Finally, (vi) follows from Lemma 3.3, noting that $|\mathcal{G}^*| \leq k$, and $\|v\|_{\mathcal{A}}^* = 1$ from (13). ∎

If the groups are disjoint to begin with , the the normal cone will be given by (12), and $\|v_S\|^2 = k$. Also, in this case, we have $|S| = kB$. We see that we do not pay an additional penalty when the groups overlap, except that the bound for the signal sparsity becomes loose. This fact is surprising, since one would expect that one would need more measurements to effectively capture the dependencies among the overlapping groups.

3.1   Remarks

The kB term in the bound is an upper-bound on the signal sparsity. In the case of highly overlapping groups, this value may be much larger than the signal sparsity. This is an unfortunate artifact of the general approach we take to derive a bound on the number of measurements. If the specific structure of groups is known (trees, hierarchies, etc.), one can refine the bound accordingly. Of course, the bound will be tightest when there is a block-sparse structure, i.e. there is no overlap between groups.

It can be seen from Theorem 3.1 that the number of measurements is linear in k and B. Hence, the number of measurements that are sufficient for signal recovery grows linearly with the number of active groups in the signal, and also the maximum group size.

We note that although we pay no extra price to measure the signal when the groups overlap, there is an additional cost in the recovery process of the signal, in that the groups need to first be separated by replication of the coefficients [11], or resort to a primal-dual method to solve the problem [16].

3.2 Noisy Observations

The results we obtain can be easily extended to the case where we obtain noisy observations, assuming that the noise is bounded. In the noisy case, we observe

$$y = \Phi x^\star + \theta, \quad \|\theta\| \leq \delta$$

We then solve the atomic norm minimization problem, with a relaxed constraint to take into account the bounded noise:

$$\hat{x} = \underset{x \in \mathbb{R}^p}{\operatorname{argmin}} \|x\|_{\mathcal{A}} \quad \textbf{s.t.} \quad \|y - \Phi x\| \leq \delta \tag{14}$$

We restate corollary 3.3 from [4]:

Proposition 3.4 *[[4], Corollary 3.3] Let $\Phi : \mathbb{R}^p \to \mathbb{R}^n$ be a random map with i.i.d. zero-mean Gaussian entries having variance $1/n$. Further let $\Omega = T_{\mathcal{A}}(x^*) \cap \mathbb{S}^{p-1}$ denote the spherical part of the tangent cone $T_{\mathcal{A}}(x^\star)$. Suppose that we have measurements $y = \Phi x^\star + \theta$, and $\|\theta\| \leq \delta$. Suppose we solve the convex program (14). Let $\hat{x}$ denote the optimum of (14). Also, suppose $\|\Phi z\| \geq \epsilon \|z\| \quad \forall z \in T_{\mathcal{A}}(x^\star)$. Then $\|x^\star - \hat{x}\| \leq \frac{2\delta}{\epsilon}$ with high probability provided that*

$$n \geq \frac{\omega(\Omega)^2}{(1-\epsilon)^2} + \mathcal{O}(1).$$

Substituting the result in Theorem 3.1 in Proposition 3.4, we have the following corollary yielding a sufficient condition to accurately recover a signal when the measurements are corrupted with bounded noise:

Corollary 3.5 *Suppose we wish to recover a $k-$ group sparse signal having M groups, such that the maximum group size is B. Let $\hat{x}$ be the optimum of the convex program (14). To have $\|\hat{x} - x^\star\| \leq \frac{2\delta}{\epsilon}$ with high probability,*

$$\frac{(\sqrt{2\log(M-k)} + \sqrt{B})^2 k + kB}{(1-\epsilon)^2}$$

iid Gaussian measurements are sufficient.

4 Experiments and Results

We extensively tested our method against the standard lasso procedure. In the case where the groups overlap, we use the replication method outlined in [11], to reduce the problem to that of non overlapping groups. We compare the number of measurements needed for our method with that needed for the lasso. For the lasso, we use the bound derived in [4] , *viz.* $(2s + 1) \log(p - s)$. We generate length $p = 2000$ signals, made up of $M = 100$ non-overlapping groups of size $B = 20$. We set $k = 5$ groups to be "active", and the values within the groups are drawn from a uniform $[0, 1]$ distribution. The active groups are assigned uniformly at random. The sparsity of the signal will be $s = 100$

We use SpaRSA [24] for the lasso and the group lasso with overlap, learning λ over a grid. Figure 1 displays the mean reconstruction error $\|\hat{x} - x^*\|_2^2/p$ as a function of the number of random measurements taken. The errors have been averaged over 100 tests, and each time a new random signal was generated with the above mentioned parameters.

From the parameters considered, we conclude that ≈ 380 measurements are sufficient to recover the signal. Note that, when we have 380 measurements, the lasso does not recover the signal exactly, as seen in Figure 1.
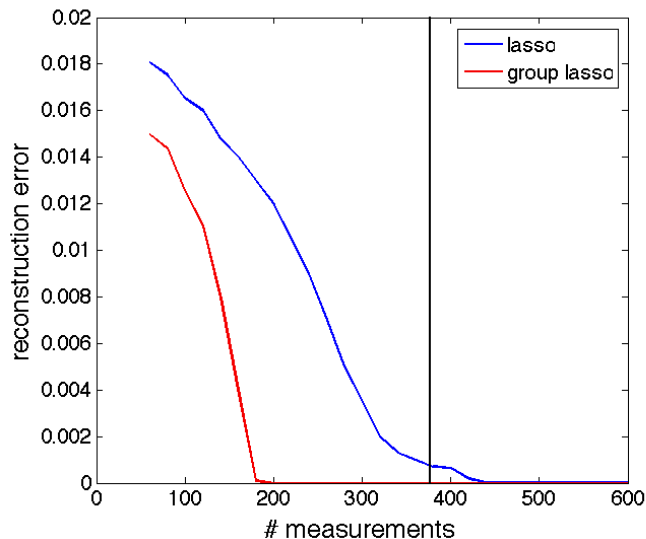
Figure 1: Comparison with the lasso. The vertical line indicates our bound. Note that our bound (380) predicts exact recovery of the signal, while at the same value, the lasso does not recover the signal

To show that the bound we compute holds regardless of the complexity of groupings, we consider the following scenario: Suppose we have $M = 100$ groups, each of size $B = 40$. $k = 5$ of those groups are active, and the values within each group are assigned from a uniform $[-1, 1]$ distribution. We arrange these groups in three configurations:

(i) The groups do not overlap, yielding a signal of length $p = 4000$, and signal sparsity $s = 200$.

(ii) A partial overlapping scenario, where apart from the first and last group, every group has 20 elements in common with a group above it, and 20 common with the group below, giving $p = 2020$, $s \in [120, \quad 200]$ depending on which of the 100 groups are active.

(iii) An almost complete overlap, where apart from one element in each group, the remaining elements are common to each group. This leads to $p = 139$ and $s = 44$

(iv) We also considered cases intermediate to the ones listed above. Specifically, we considered (a) a highly overlapping scenario which is identical to the previous case, but with odd and even groups disjoint. We also consider (b) a random overlap case where the first 50 groups are non overlapping and the remaining 50 are assigned uniformly at random from the existing $p = 2000$ indices.

The scenarios we consider are depicted in Figure 2(a). In each of the cases, we compute the bound to be ≈ 630. We can see from Figure 2(b) that the bound holds for all cases. The bound becomes looser as the complexity of the groupings increases. This, as argued before, is a result of the bound for the signal sparsity becoming looser. From the values of p and s computed for the three cases, we have the corresponding bounds for the lasso to be 3305 for the no overlap case (i), [1819, 3010] for the partial overlap case (ii) and 405 for the almost complete overlap case (iii)respectively. The lasso bounds for case (iv) will lie between those for case (ii) and (iii).

Finally, we consider the wavelet transform coefficients of the "blocks" signal (Figure 4(a)). It was shown in [19] that the coefficients can be grouped, to account for parent child dependencies across scales of the wavelet transform, as shown in Figure 3. In this case, for a $p = 16384$ length signal, we have $M = 16382$ groups, and the maximum group size is $B = 2$. We use the Haar wavelet bases to decompose the image. Figure 4(b) shows the reconstruction obtained from 1690 measurements, corresponding to the bound computed for

9

(a) types of groupings considered. Each set of coefficients encompassed by one color belongs to one group
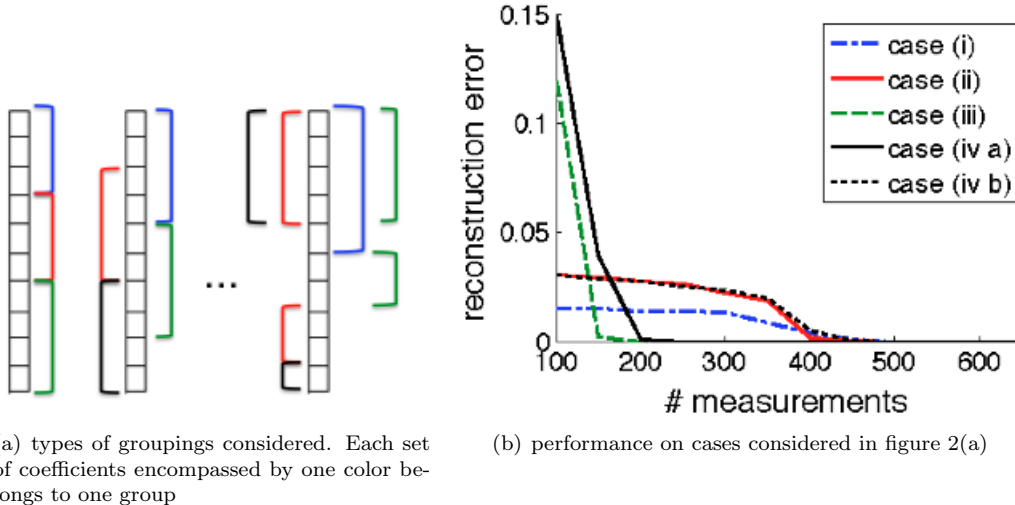
(b) performance on cases considered in figure 2(a)

Figure 2: (Best seen in color) Performance on various grouping schemes. Note that our bound evaluates to 630, clearly sufficient measurements to recover the signal. The corresponding bounds for the lasso (for cases (i), (ii) and (iii)) are 3305, [1819, 3010] (depending on s) and 405 respectively. We can see that, as the amount of overlap increases, our bound loosens, and for pathological cases the lasso bound is tighter

$k = 47$. We see that our bound yields a sufficient number of measurements for exact recovery.
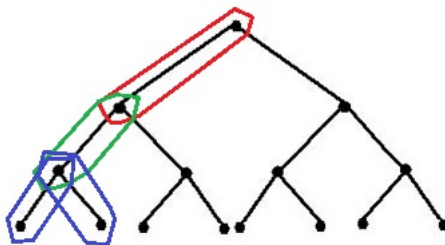

Figure 3: Groups on the 1d wavelet transform

5   Conclusion

We showed that, when additional structure about the support of the signal to be estimated is known, we can recover the signal in much fewer measurements that what would be needed in the standard compressed sensing framework. Also, we showed that we surprisingly do not pay an extra penalty when the groups overlap each other. Moreover, the bound holds for arbitrary group structures, and can be used in a variety of applications. The bounds we derive are tight, and can be extended to subgaussian measurement matrices by incurring a constant penalty. Experimental results on both toy and real data agree with the bounds we obtained.
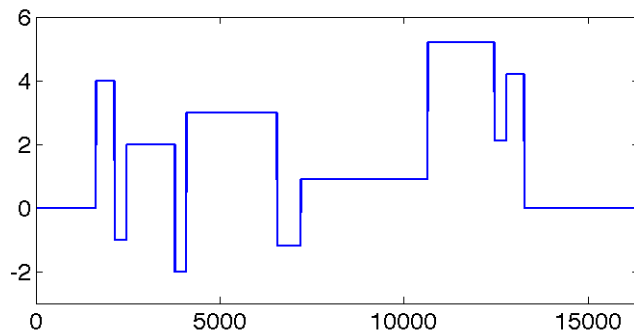
(a) original signal
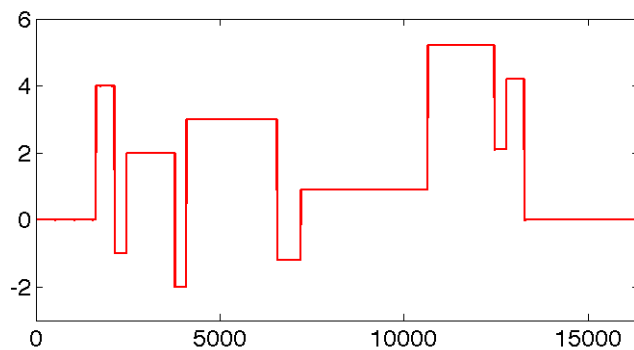
(b) reconstruction

Figure 4: Exact reconstruction of a length 16384 signal from 1690 measurements in the wavelet domain

References

[1] F. Bach. Consistency of the group lasso and multiple kernel learning. *Journal of Machine Learning Research*, 9:1179–1225, June 2008.

[2] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde. Model-based compressive sensing. *IEEE Transactions on Information Theory*, 2010.

[3] E. J. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Information Theory*, 52:489–509, 2006.

[4] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A Willsky. The convex geometry of linear inverse problems. *preprint arXiv:1012.0621v1*, 2010.

[5] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk. Wavelet based statistical signal processing using hidden markov models. *Transactions on Signal Processing*, 46(4):886–902, 1998.

[6] D. L. Donoho. Compressed sensing. *IEEE Trans. Information Theory*, 52:1289–1306, 2006.

[7] M. F. Duarte, V. Cevher, and R. G. Baraniuk. Model-based compressive sensing for signal ensembles. *Allerton*, 2009.

[8] Y. Gordon. On Milman's inequality and random subspaces which escape through a mesh in $\mathbb{R}^n$. *Geometric aspects of functional analysis, Isr. Semin.*, 1317:84–106, 1986 - 87.

[9] J. Huang and T Zhang. The benefit of group sparsity. *Technical report, arXiv:0901.2962. Preprint available at http://arxiv.org/pdf/0903.2962v2*, May 2009.

[10] J. Huang, T. Zhang, and D. Metaxas. Learning with structured sparsity. *Technical report, arXiv:0903.3002. Preprint available at http://arxiv.org/pdf/0903.3002v2*, May 2009.

[11] L. Jacob, G. Obozinski, and J. P. Vert. Group lasso with overlap and graph lasso. *Proceedings of the 26th International Conference on machine Learning*, 2009.

[12] R. Jenatton, J.Y. Audibert, and F. Bach. Structured variable selection with sparsity inducing norms. *Technical report, arXiv:0904.3523. Preprint available at http://arxiv.org/pdf/0904.3523v3*, Sep 2009.

[13] R. Jenatton, J. Mairal, G. Obozinski, , and F. Bach. Proximal methods for hierarchical sparse coding. *Technical report, arXiv:1009.3139. submitted*, 2010.

[14] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann. Reconstruction and subgaussian operators in asymptotic geometric analysis. *Geometric and Functional Analysis*, 17(4):1248 – 1282, 2006.

[15] M. Mishali and Y. Eldar. Blind multi-band signal reconstruction: compressed sensing for analog signals. *IEEE Trans. Signal Processing*, 57(30):993–1009, March 2009.

[16] S. Mosci, S. Villa, A. Verri, and L. Rosasco. A primal-dual algorithm for group sparse regularization with overlapping groups. *Neural Information Processing Systems*, 2010.

[17] D. Needell and J. Tropp. Cosamp: Iterative signal recovery from incomplete and inaccurate samples. *Appl. Comput. Harmon. Anal.*, 26:301–321, 2008.

[18] S. Negahban, P. Ravikumar, M. Wainwright, and B. Yu. A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers. *Preprint ArXiv :1010:2731v1*, October 2010.

[19] N. Rao, R. Nowak, S. Wright, and N. Kingsbury. Convex approaches to model wavelet sparsity patterns. *IEEE International Conference on Image Processing*, 2011.

[20] T. Rockafellar and J. B. Wets. Variational analysis. *Springer Series of Comprehensive Studies in Mathematics*, 317, 1997.

[21] J. K Romberg, H. Choi, and R. G Baraniuk. Bayesian tree structured image modeling using wavelet domain hidden markov models. *Transactions on Image Processing*, March 2000.

[22] A. Subramanian et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression proles. *National Academy of Sciences*, 102:1554515550, 2005.

[23] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B*, pages 267–288, 1996.

[24] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo. Sparse reconstruction by separable approximation. *Transactions on Signal Processing*, 57:2479–2493, 2009.

[25] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the royal statistical society. Series B*, 68:49–67, 2006.