Tight Performance Bounds for Compressed Sensing With Conventional and Group Sparsity

Shashank Ranjan and Mathukumalli Vidyasagar, *Life Fellow, IEEE*

***Abstract*—In this paper, we study the problem of recovering a group sparse vector from a small number of linear measurements. In the past, the common approach has been to use various "group sparsity-inducing" norms such as the Group LASSO norm for this purpose. By using the theory of convex relaxations, we show that it is also possible to use ℓ_1-norm minimization for group sparse recovery. We introduce a new concept called group robust null space property (GRNSP), and show that, under suitable conditions, a group version of the restricted isometry property (GRIP) implies the GRNSP, and thus leads to group sparse recovery. When all groups are of equal size, our bounds are sometimes less conservative than known bounds. Moreover, our results apply even to situations where the groups have different sizes. When specialized to conventional sparsity, our bounds reduce to one of the well-known "best possible" conditions for sparse recovery. This relationship between GRNSP and GRIP is new even for conventional sparsity, and substantially streamlines the proofs of some known results. Using this relationship, we derive bounds on the ℓ_p-norm of the residual error vector for all $p \in [1, 2]$, and not just when $p = 2$. When the measurement matrix consists of random samples of a sub-Gaussian random variable, we present bounds on the number of measurements, which are sometimes less conservative than currently known bounds.**

***Index Terms*—Compressed sensing, convex functions, optimization.**

I. Introduction

Compressed sensing refers to the recovery of high-dimensional vectors with very few nonzero components from a limited number of linear measurements. This is referred to here as the "conventional" sparsity problem, and it has been the subject of a great deal of research. In recent years, attention has also been focused on the "group sparsity" problem, where there is additional information available about the locations of the nonzero components of the unknown vector. In this paper, we advance the status of knowledge in compressed sensing for both conventional as well as group sparsity. Precise details are given in subsequent sections, but in brief the contributions of the paper are the following:

- In conventional sparsity, the two most widely used techniques are RIP (restricted isometry property) and the RNSP (robust null space property); very few papers relate the two approaches.[1] One of the currently best available results [2] on the use of the RIP states that if the measurement matrix A satisfies the RIP of order tk for some $t \geq 4/3$, then it is possible to achieve robust k-sparse recovery via the basis pursuit formulation, that is, minimizing an ℓ_1-norm objective function.[2] Moreover, this bound is tight, as shown in [2]. In the present paper, we offer two improvements to these results. First, we show that the above sufficient condition continues to be sufficient whenever $t > 1$, and not just when $t \geq 4/3$. Second, we prove this by showing that in this case the RIP implies the RNSP. Ours is the best available relationship between RIP and RNSP. Moreover, the connection between RIP and RNSP allows us to prove bounds on the ℓ_p-norm of the residual error for all $p \in [1, 2]$, and not just the ℓ_2-norm. The papers based on the RIP alone are not able to prove such bounds.
- In group sparsity, until now researchers have replaced the ℓ_1-norm objective function by various "group sparsity-inducing" norms in order to achieve robust recovery. In the present paper, we show that the standard ℓ_1-norm can also be interpreted as the convex relaxation of two distinct group sparsity indices, so that ℓ_1-norm minimization also has the potential to achieve group sparse recovery. Then we proceed to derive conditions under which ℓ_1-norm minimization actually achieves group sparse recovery. These conditions reduce to those for conventional sparsity when all "groups" consist of one element each. Our method of proof is based on the group version of the RIP, but also a new (though very natural) group version of the RNSP. As with conventional sparsity, we show that GRIP implies the GRNSP. Thus, using our approach, we can derive bounds on the ℓ_p-norm of the residual error for all $p \in [1, 2]$, which are generally not available with group sparsity-inducing norms. We also derive bounds on the number of samples that suffice to achieve group sparse recovery when the measurement matrix consists of random sub-Gaussian samples. In some situations, these bounds are smaller than currently available bounds from other papers. Not surprisingly, it is also shown that group sparse recovery can be achieved with fewer samples than for

Manuscript received January 22, 2018; revised July 17, 2018 and January 29, 2019; accepted March 10, 2019. Date of publication March 27, 2019; date of current version April 26, 2019. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Mark A. Davenport. This work was supported by the National Science Foundation under Award #ECCS-1306630 and by the Department of Science and Technology, Government of India. A brief description of a part of these results was presented at the Indian Control Conference 2017 and is published in the Proceedings [1]. *(Corresponding author: Mathukumalli Vidyasagar.)*

The authors are with the Electrical Engineering Department, Indian Institute of Technology Hyderabad, Kandi 502285, India (e-mail: ee16resch11020@iith.ac.in; m.vidyasagar@iith.ac.in).

Digital Object Identifier 10.1109/TSP.2019.2907228

[1]All terms are defined in subsequent sections.

[2]Other complementary results from [3] and [4] are also discussed below.

conventional sparse recovery. Given that there are now very efficient methods for ℓ_1-norm minimization, our results suggest that ℓ_1-norm minimization is a viable alternative to the use of group sparsity-inducing norms, for problems of group sparse recovery.

II. Conventional Sparsity

A. Summary of Some Compressed Sensing Results

Let $\Sigma_k \subseteq \mathbb{R}^n$ denote the set of k-sparse vectors in $\mathbb{R}^n$; that is

$$\Sigma_k := \{x \in \mathbb{R}^n : \|x\|_0 = |\mathrm{supp}(x)| \le k\},$$

where, as is customary, $\|\cdot\|_0$ denotes the number of nonzero components of x. Given a norm $\|\cdot\|$ on $\mathbb{R}^n$, the k-sparsity index of x with respect to that norm is defined by

$$\sigma_k(x, \|\cdot\|) := \min_{z \in \Sigma_k} \|x - z\|.$$

Now we can define the conventional compressed sensing problem precisely.

Definition 1: Suppose $A \in \mathbb{R}^{m\times n}$ and $\Delta : \mathbb{R}^m \to \mathbb{R}^n$. The pair (A, Δ) is said to achieve robust sparse recovery of order k and indices p, q if there exist constants C and D such that, for all $\eta \in \mathbb{R}^m$ with $\|\eta\|_2 \le \epsilon$, it is the case that

$$\|\Delta(Ax + \eta) - x\|_p \le C\sigma_k(x, \|\cdot\|_q) + D\epsilon, \ \forall x \in \mathbb{R}^n. \quad (1)$$

Among the most popular decoder maps is ℓ_1-norm minimization, also known as basis pursuit. When $y = Ax + \eta$ with $\|\eta\|_2 \le \epsilon$, it is defined as follows:

$$\Delta_{\mathrm{BP}}(y) := \operatorname*{argmin}_{z\in\mathbb{R}^n} \|z\|_1 \text{ s.t. } \|y - Az\|_2 \le \epsilon, \quad (2)$$

There are two widely used sufficient conditions for basis pursuit to achieve robust sparse recovery, namely the restricted isometry property (RIP) and the robust null space property (RNSP). We begin by discussing the RIP.

Definition 2: A matrix $A \in \mathbb{R}^{m\times n}$ is said to satisfy the restricted isometry property (RIP) of order k with constant δ if

$$(1-\delta)\|u\|_2^2 \le \|Au\|_2^2 \le (1+\delta)\|u\|_2^2, \ \forall u \in \Sigma_k. \quad (3)$$

Starting with [5], it is shown in a series of papers that the RIP of A is sufficient for $(A, \Delta_{\mathrm{BP}})$ to achieve robust sparse recovery. In [6] it is shown that $\delta_{2k} < \sqrt{2} - 1$ is sufficient for robust k-sparse recovery. This bound has been subsequently improved in several papers, but to save space, we cite only the most recent "best possible" results relating RIP and robust recovery.

Theorem 1: If A satisfies the RIP of order tk with constant $\delta_{tk} < \sqrt{(t-1)/t}$ for some $t \ge 4/3$, or with constant $\delta_{tk} < t/(4-t)$ for some $t \in (0, 4/3)$,[3] then $(A, \Delta_{\mathrm{BP}})$ achieves robust sparse recovery with $q = 1$ and $p = 2$. Moreover, both bounds are tight.

Note that the first bound is proved in [2], while the second bound is proved in [4]. In [3], it is shown that $\delta_k < 0.307$ is sufficient, which is slightly worse than the bound $\delta_k < 1/3$ implied by [4].

An alternative to the RIP approach to compressed sensing is provided by the robust null space property; see [7] or [8, Definition 4.21].

Definition 3: A matrix $A \in \mathbb{R}^{m\times n}$ is said to satisfy the ℓ_2-robust null space property (RNSP) of order k with constants $\rho \in (0, 1)$ and $\tau > 0$ if, for every set $S \subseteq [n]$ with $|S| \le k$, we have that[4]

$$\|h_S\|_2 \le \frac{\rho}{\sqrt{k}}\|h_{S^c}\|_1 + \frac{\tau}{\sqrt{k}}\|Ah\|_2, \ \forall h \in \mathbb{R}^n. \quad (4)$$

Schwarz' inequality implies that if A satisfies the ℓ_2-RNSP, then for every set $S \subseteq [n]$ with $|S| \le k$ it also satisfies

$$\|h_S\|_1 \le \rho\|h_{S^c}\|_1 + \tau\|Ah\|_2, \ \forall h \in \mathbb{R}^n. \quad (5)$$

We refer to this property as just RNSP without the prefix "ℓ_2."

Theorem 2: (See [8, Ths. 4.19 and 4.22].) Suppose A satisfies (5) with constants ρ and τ. Then the pair $(A, \Delta_{\mathrm{BP}})$ achieves robust k-sparse recovery for $p = q = 1$, with

$$C = 2\frac{1+\rho}{1-\rho}, D = \frac{4\tau}{1-\rho}. \quad (6)$$

If A satisfies (4), then $(A, \Delta_{\mathrm{BP}})$ achieves robust k-sparse recovery for $p = 1$ and all $q \in [1, 2]$.

There are relatively few results relating the RIP and the RNSP. Currently the best available result is [7, Proposition 8], in which it is shown that if A satisfies the RIP of order $2k$ with constant $\delta_{2k} < 1/9$, then it also satisfies the RNSP. Note that $1/9$ is far smaller than $\sqrt{1/2}$ which is the bound on δ_{2k} from Theorem 1.

B. Our Contributions

Against this background, in this paper we show that, if A satisfies the RIP of order tk with constant $\delta_{tk} < \sqrt{(t-1)/t}$ for some $t > 1$, then A also satisfies the ℓ_2-RNSP for appropriate constants; see Theorem 9.[5] This has several consequences. First, this is by far the best result that relates RIP to RNSP. As mentioned in the previous paragraph, the bound on δ_{2k} to satisfy the RNSP is improved from $1/9$ to $1/\sqrt{2}$. Second, by establishing that the condition $\delta < \sqrt{(t-1)/t}$ implies the ℓ_2-RNSP, we can establish that for such a matrix A, basis pursuit achieves robust k-sparse recovery for all $p \in [1, 2]$ (and $q = 1$), and are able to prove bounds on the ℓ_p-norm of the residual for all $p \in [1, 2]$. This is in contrast to existing papers papers based on the RIP including [2], [4] where robust k-sparse recovery is established using the RIP, and thus error bounds are available only for $p = 2$. Moreover, our bounds on the ℓ_p-norm of the residual error are an improvement over those in [8, Th. 4.25].

III. Group Sparsity

A. Literature Review

At about the same time that the problem of robust sparse recovery was being addressed via ℓ_1-norm minimization, the

[3]Here and elsewhere. when we write δ_α and α is not necessarily an integer, we mean $\delta_{\lceil \alpha \rceil}$.

[4]Note that, for the sake of consistency, we have introduced a factor of $\sqrt{k}$ to divide τ in (4).

[5]However, this results in an improvement only when $t \ge 4/3$.

research community began to propose that the number of nonzero components of a vecor might not be the only reasonable measure of the sparsity of a vector. Alternate notions under the broad umbrella of "group sparsity" and "group sparse recovery" began to appear, starting with [9]. In its simplest form, group sparsity refers to the case where the index set $[n]$ is partitioned into g disjoint sets $G_1, \ldots, G_g$. In the early papers such as [10]–[13], it is assumed that all groups G_i have the same size d, so that $n = gd$. However, starting with [14], the groups are not required to have a common size. In almost all current papers on group sparse recovery, the ℓ_1-norm objective function in (2) is changed to the so-called Group LASSO norm introduced in [9], defined as

$$\|x\|_{\mathrm{GL}} := \sum_{j \in [g]} \|x_{G_j}\|_2, \tag{7}$$

where x_{G_j} denotes the projection of x onto the set G_j.

For this formulation, a variety of recovery results are proved by several authors. In [12], a block RIP analogous to the RIP is introduced, as follows: A vector $x \in \mathbb{R}^n$ is said to be l-group sparse if there are no more than l groups G_j such that the projection x_{G_j} of x onto G_j is nonzero. A matrix $A \in \mathbb{R}^{m \times n}$ is said to satisfy the "group RIP" if there exists a constant δ_l such that

$$(1 - \delta_l)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_l)\|x\|_2^2 \tag{8}$$

whenever the vector x is l-group sparse. It is shown that recovery of group k-sparse vectors is achieved by minimizing the GL norm $\|x\|_{\mathrm{GL}}$ if $\delta_{B,2k} < \sqrt{2} - 1$, which is an extension of the conventional sparsity result in [6]. In [13], a notion of block-coherence is introduced; it is shown that, just as in conventional sparsity, block-coherence implies block-RIP, which in turn implies group sparse recovery. The well-known orthogonal matching pursuit algorithm [15] is modified to a block-OMP and it is shown that block-OMP recovers a block-sparse signal under suitable conditions. In [10], [11], and [14], the measurement matrix A is assumed to be a randomly generated Gaussian or sub-Gaussian matrix, and bounds are derived on the number of samples that sufficie for the (probabilistic) recovery of a group-sparse vector x by minimizing the GL norm. In the first two papers, it is assumed that all groups have the same size and it is shown that the required number of measurements with group sparsity is less than with conventional sparsity. In [11], the behavior of well-known algorithms such as CoSaMP and IHT (iterative hard thresholding) is analyzed with the Group LASSO norm. In [14] the authors dispense with the requirement of equal group sizes, and derive a sufficient condition (see [14, Assumption 4.3]) for group sparse recovery. Regarding another assumption, namely [14, Assumption 4.2] the authors state "Note that this assumption does not show the benefit of group Lasso over standard Lasso." However, this statement does not apply to [14, Assumption 4.3]. In [16], group sizes need not be equal, and the GL norm is modified by replacing $\|x_{G_j}\|_2$ with $\|x_{G_j}\|_q$ for any $q \geq 1$. Sufficient conditions for group sparse recovery are established in terms of the group RIP, and block coherence. This work is extended in [17] to incorporate subspace coherence, whose value is in general smaller than block coherence. In [18] and [19], Theorem 1 is extended to group sparsity as follows: It is shown in [18] that if the group RIP constant δ_l satisfies $\delta_{tl} < \sqrt{(t-1)/t}$ for some $t > 1$, then robust sparse recovery results. This is an improvement over [2] in that the limit $t \geq 4/3$ in [2] is reduced to $t > 1$. In [19] the above bound is replaced by something completely analogous that in [4], namely $\delta_{tl} < \sqrt{(t-1)/t}$ when $t \geq 4/3$, or $\delta_{tl} < t/(4-t)$ if $t \in (0, 4/3)$. Moreover, it is shown that both bounds are tight.

The above papers can be thought as representing the first phase of research into group sparse recovery. Subsequent papers follow several different and unrelated directions. Several papers in the statistics community analyze the *asymptotic* behavior of minimizing the Group LASSO norm as $n \to \infty$, with k either kept fixed, or increasing more slowly than n. Some of these papers also study the problem of "simultaneous" estimation of several unknown vectors that share a common sparsity pattern, using a common measurement matrix. In [20, Corollary 4.1, Th. 7.1], it is shown that *in the problem of simultaneous estimation*, the Group LASSO norm offers advantages over the standard ℓ_1-norm. In [21], the authors study the problem of *support* recovery, that is, recovering the set of nonzero components of the unknown vector, under group sparsity. They give a very tight bound on the rate at which the number of samples must grow in order to achieve support recovery. These results show that, when the unknown vector is supported over *a union* of unknown subspaces, the Group LASSO formuation is natural. Support recovery is also the subject of [22]. Unlike other papers, the results in this paper are not asymptotic. Note that it is possible to recover the support of an unknown vector while not recovering the vector itself. On the other hand, if the nondominant components of a (group) sparse vector are much smaller than the dominant ones, recovering a good approximation to the unknown vector also leads to support recovery.

In [23] and the references therein, the emphasis is on removing the assumption that the sets G_j are pairwise disjoint; thus the focus is on *overlapping* group decompositions. In [24] and [25], the authors study the case where there is uncertainty and/or error in implementing the measurement matrix A. Instead of the designed matrix A, the measurements equal $y = (A + BE)x$ for suitable models of B, E. One of the important innovations of these papers is the notion of "joint" sparsity. To illustrate, suppose $n = 2l$. Then for a given $k < n$, a vector x is said to be jointly k-sparse if its support is concentrated a set of the form $S \cup (l + S)$, where $|S| \leq k$, and $l + S$ denotes shifting every element of S by l. This model is apparently very natural in problems of detecting the Direction of Arrival (DoA). A joint RIP is defined for such vectors. It is clear that, for the same number m of measurements, the joint $2k$ RIP constant is smaller than not just the standard $2k$ RIP constant, but also the group $2k$ RIP constant, because of the restrictions on the support set. Therefore, in order to make the RIP constant smaller than a specified threshold, group sparse recovery would require fewer samples than conventional sparse recovery, while joint sparse recovery would require still fewer samples. Finally, in [26], the authors relax the requirement from recovering *every* group sparse vector to *average case* recovery. Naturally, the sufficient conditions for recovery in this case are weaker than for the recovery of *every* vector. The main drawback of this approach is that there is no way to know whether the *particular* group sparse vector that one is attempting to recover lies within the set of recoverable vectors.

The above discussion can be briefly summarized as follows: The Group LASSO formulation is better than the conventional LASSO formulation when it comes to simultaneous estimation, and in support recovery. Under suitable assumptions as specified in [14], Group LASSO is also superior to conventional LASSO for vector recovery as well. In the present paper, the focus is on estimating a single vector (therefore not simultaneous recovery, nor support recovery). In the opinion of the authors, currently available results such as [20] do not establish conclusively whether Group LASSO offers any unambiguous advantages in this situation.

B. The ℓ_1-Norm as a Group Sparsity-Inducing Norm

The subsection has two objectives. The first is to introduce the concept of the convex relaxation of a nonconvex function, and to give explicit formulas for the convex relaxations of group sparsity measures over product sets. The second objective is to show that commonly used objective functions such as the Group LASSO norm of (7), and the weighted ℓ_1-norm introduced in [27] can both be interpreted as convex relaxations of various group sparsity measures over suitably defined product sets. Rather surprisingly perhaps, it is shown that even the unweighted ℓ_1-norm is the convex relaxation of two sparsity measures over suitably defined product sets. The conclusion is that, in addition to the Group LASSO and the weighted ℓ_1-norms, the unweighted ℓ_1-norm can also be used as a "sparsity-inducing" norm, and can thus be used to recover *group sparse* vectors.

We begin with a discussion of group sparsity indices. In the conventional setting, the quantity $\|x\|_0$ which counts the number of nonzero components of x is taken as a measure of the sparsity of x. In the case of group sparsity, it is possible to think of two distinct-looking definitions.[6]

$$\|x\|_{\mathrm{UG},0} = \sum_{j\in[g]} \mathbf{1}_{\{x_{G_j}\neq \mathbf{0}\}}, \tag{9}$$

$$\|x\|_{\mathrm{G},0} = \sum_{j\in[g]} |G_j|\mathbf{1}_{\{x_{G_j}\neq \mathbf{0}\}}, \tag{10}$$

where x_{G_j} denotes the projection of $x \in \mathbb{R}^n$ onto the indices in G_j, and $\mathbf{1}$ denotes the indicator function. Thus $\|x\|_{\mathrm{UG},0}$ counts the number of groups on which x has a nonzero projection, whereas $\|x\|_{\mathrm{G},0}$ counts the cardinality of the union of groups over which x has a nonzero projection. It is obvious that if all groups have the same size d (as was assumed in many early papers), then both definitions differ only by a factor of d. However, when group sizes differ widely, the two quantities can be very different. While a majority of papers use the definition in (9), [14] uses a combination of both parameters.

Let us define a vector $x \in \mathbb{R}^n$ to be l-group sparse if $\|x\|_{\mathrm{UG},0} \leq l$, and group k-sparse if $\|x\|_{\mathrm{G},0} \leq k$. Further, define $d_{\max}$ and $d_{\min}$ denote the largest and smallest group sizes. Then an l-group sparse vector is also $ld_{\max}$-sparse in the conventional sense, but the converse is not true. Similarly, a group k-sparse vector is also k-sparse in the conventional sense, but the converse is not true. In the proofs of Theorems 3 and 4, we make use of the fact that there is a known prior bound on the sparsity count of the unknown vector, irrespective of the number of groups over which it is supported. Thus we prefer to work with group k-sparse vectors and not l-group sparse vectors. In principle our proofs could be adapted to l-group sparse vectors by treating them as group $ld_{\max}$-sparse vectors. Working with the latter would lead to more conservative bounds for recovery.

In conventional sparse recovery, one could attempt to recover a k-sparse vector x from a linear measurement vector $y = Ax$ by solving

$$\hat{x} = \operatorname*{argmin}_z \|z\|_0 \text{ s.t. } Az = y.$$

However, the function $\|\cdot\|_0$ is not convex. In conventional sparse recovery, replacing the nonconvex objective function $\|x\|_0$ by $\|x\|_1$ is justified using the concept of a convex relaxation. Therefore, for group sparse recovery it is desirable to determine the convex relaxations of the group sparsity indices $\|\cdot\|_{\mathrm{UG},0}$ and $\|\cdot\|_{\mathrm{G},0}$. Accordingly, we first formally define the concept of a convex relaxation, then a group decomposable norm, and then show that not just the Group LASSO and the weighted ℓ_1-norm, but also the unweighted ℓ_1-norm, are all convex relaxations of both $\|\cdot\|_{\mathrm{UG},0}$ and $\|\cdot\|_{\mathrm{G},0}$ over suitably defined convex sets.

Definition 4: Suppose $\Omega \subseteq \mathbb{R}^n$ is a convex set, and that $f : \Omega \to \mathbb{R}$. Then a function $g : \Omega \to \mathbb{R}$ is said to be the convex relaxation of f over Ω if: (i) $g(x) \leq f(x)\ \forall x \in \Omega$, and (ii) if $h : \Omega \to \mathbb{R}$ is convex and satisfies $h(x) \leq f(x)\ \forall x \in \Omega$, then $h(x) \leq g(x)\ \forall x \in \Omega$.

In other words, the convex relaxation of f is the largest convex function that is dominated by f on the set Ω. Observe that the same function f but on a different convex set Ω' could have a different convex relaxation g'. There is a conceptually simple way to determine the convex relaxation, namely through the use of convex duality. Theorem [28, Th. E.1.3.5] states that the second dual of f is its convex relaxation. Moreover, using the definition of the dual, it is easy to establish the following fact.

Lemma 1: Let $\{G_1, \ldots, G_j\}$ be a partition of $[n]$. Write $\mathbb{R}^n = \prod_{j=1}^g \mathbb{R}^{|G_j|}$, and suppose that $\Omega \subseteq \mathbb{R}^n = \prod_{j=1}^g \Omega_j$ where each $\Omega_j \subseteq \mathbb{R}^{|G_j|}$. Further, suppose $f : \Omega \to \mathbb{R}$ is decomposable as

$$f(x) = \sum_{j=1}^{g} f_j(x_{G_j}), \tag{11}$$

where x_{G_j} is the projection of x onto $\mathbb{R}^{|G_j|}$. Then the convex relaxation g of f equals

$$g(x) = \sum_{j=1}^{g} g_j(x_{G_j}), \tag{12}$$

where g_j is the convex relaxation of f_j over Ω_j.

The next lemma is also easy to prove.

[6]We thank one of the reviewers for suggesting this notation.

Lemma 2: Suppose $c > 0$ is some constant, and let $c \cdot \mathcal{B}$ denote the set of all $x \in \mathbb{R}^l$ with $\|x\| \leq c$. Then the convex relaxation of ϕ over $c \cdot \mathcal{B}$ is $(1/c)\| \cdot \|$.

Next, let us refer to a norm $\| \cdot \|$ on $\mathbb{R}^n$ as group decomposable if it is of the form

$$\|x\| = \sum_{j \in [g]} \|x_{G_j}\|_{G_j}, \qquad (13)$$

for suitably defined norms $\| \cdot \|_{G_j}$ on $\mathbb{R}^{|G_j|}$. The next result follows as a ready consequence of Lemmas 1 and 2.

Lemma 3: Let $\| \cdot \|_{G_j}, j \in [g]$ be arbitrary norms on $\mathbb{R}^{|G_j|}$. Let $\Omega_j \subseteq \mathbb{R}^{|G_j|}$ denote the unit ball of $\| \cdot \|_{G_j}$, and define $\Omega = \prod_{j \in [g]} \Omega_j$. Then the convex relaxation of $\| \cdot \|_{\mathrm{UG},0}$ over Ω is the norm defined in (13). More generally, define $\Omega'_j = |G_j| \cdot \Omega_j$ for all $j \in [g]$, and let $\Omega' = \prod_{j \in [g]} \Omega'_j$. Then the convex relaxation of $\| \cdot \|_{\mathrm{G},0}$ over Ω' is the norm defined in (13).

In short, *every* group decomposable norm is the convex relaxation of $\| \cdot \|_{\mathrm{UG},0}$ over a suitably defined product set Ω. Conversely, the convex relaxation of $\| \cdot \|_{\mathrm{UG},0}$ over every product set is a group decomposable norm. Moreover, every convex relaxation of $\| \cdot \|_{\mathrm{UG},0}$ over a product Ω is also a convex relaxation of $\| \cdot \|_{\mathrm{G},0}$ over the related set Ω', and vice versa. In particular, if we were to choose each of the sets Ω_j to be the unit balls in the ℓ_2-norm over the corresponding space, then the convex relaxation of $\| \cdot \|_{\mathrm{UG},0}$ over Ω would be the Group LASSO norm defined in (7). However, if we were to choose each of the sets Ω_j to be the unit balls in the ℓ_∞-norm over the corresponding space, then the convex relaxation of $\| \cdot \|_{\mathrm{UG},0}$ over Ω would be the ℓ_1-norm! If we were to choose the set Ω_j to be the ball of radius r_j in the ℓ_∞-norm over the corresponding space, then the convex relaxation of $\| \cdot \|_{\mathrm{UG},0}$ over Ω would be the *weighted* ℓ_1-norm.

To summarize, the point is that even ℓ_1-norm minimization can be used to achieve group sparse recovery, even though it is not obviously "group-sparsity inducing."

C. Our Contributions

In the present paper, we use ℓ_1-norm minimization, and establish that this approach can recover group k-sparse vectors under appropriately defined sufficient conditions. These results stand in contrast to earlier results of [12], [18], and [19] that pertain to the recovery of l-group sparse vectors. To reiterate, our results are for group k-sparse vectors, whereas earlier results are for l-group sparse vectors. Our results are based on defining group analogs of the RIP and the RNSP for group k-sparse vectors.[7] So far as we are able to determine, this is the first time that a group version of the RNSP is proposed and used to establish group sparse recovery. In some situations, the bounds derived here are less conservative than those proved earlier by others, based on the group RIP, both for the case where all groups are of equal size [11]–[13], and are of unequal size [16], [17]. When the measurement matrices consist of random samples of sub-Gaussian variables, and all groups are of equal size, our estimates are of the same order as in [11] and are smaller than for conventional sparse recovery. When group sizes are unequal and the measurement matrix is random, our bounds are less conservative than those in [14, Assumption 4.2, Th. 5.1], though not less conservative than those in [14, Assumption 4.3, Th. 5.1]. It is of course possible results similar to ours could be established using the Group LASSO norm instead of the ℓ_1-norm. That would be a topic for future research.

[7]Previous definitions are for l-group sparse vectors.

The rest of the paper is organised as follows: The main results of the paper concerning group sparse recovery and concerning conventional sparse recovery are stated in Sections IV and V respectively. These results are compared against known results in Section VI. Numerical examples are given in Section VII, and the proofs of the main results are given separately in Section VIII. Throughout the paper, we use the basis pursuit denoising approach. Therefore, given $y = Ax + \eta$ with $\|\eta\|_2 \leq \epsilon$, we define

$$\hat{x} = \operatorname*{argmin}_z \|z\|_1 \text{ s.t. } \|y - Az\|_2 \leq \epsilon. \qquad (14)$$

IV. Main Results-I: Group Sparse Recovery

We say that a vector u is group k-sparse if $\|u\|_{\mathrm{G},0} \leq k$, where $\| \cdot \|_{\mathrm{G},0}$ is defined in (10). We also require the notion of a group k-sparse subset of the index set $[n]$. Recall that $\{G_1, \ldots, G_g\}$ is a partition of $[n]$. If $L \subseteq [g]$, let G_L denote $\cup_{j \in L} G_j$. Then a set $S \subseteq [n]$ is said to be a group k-sparse subset of $[n]$ if $S = G_L$ for some subset $L \subseteq [g]$, and moreover, $|S| \leq k$. Note that a vector is a group k-sparse vector if and only if $\mathrm{supp}(x)$ is a group k-sparse set. We denote the set of all group k-sparse vectors by $\Sigma_{G,k}$, and the collection of all group k-sparse sets by GkS.

We begin by defining group analogs of the RIP and RNSP for group k-sparse vectors.

Definition 5: A matrix $A \in \mathbb{R}^{m \times n}$ is said to satisfy the group restricted isometry property (GRIP) of order k with constant $\delta_{G,k} \in (0, 1)$ if

$$(1 - \delta_{G,k})\|u\|_2 \leq \|Au\|_2 \leq (1 + \delta_{G,k})\|Au\|_2^2, \ \forall u \in \Sigma_{G,k}. \qquad (15)$$

Definition 6: A matrix $A \in \mathbb{R}^{m \times n}$ is said to satisfy the ℓ_2-group robust null space property (GRNSP) with constants $\rho_G \in (0, 1), \tau_G \in \mathbb{R}_+$, if, for all $h \in \mathbb{R}^n$ and all sets $S \in$ GkS, it is true that

$$\|h_S\|_2 \leq \frac{\rho_G}{\sqrt{k}} \|h_{S^c}\|_1 + \frac{\tau_G}{\sqrt{k}} \|Ah\|_2. \qquad (16)$$

As with RNSP, Schwarz' inequality implies that if A satisfies the ℓ_2-GRNSP, then for all $h \in \mathbb{R}^n$ and all sets $S \in$ GkS, it is true that

$$\|h_S\|_1 \leq \rho_G \|h_{S^c}\|_1 + \tau_G \|Ah\|_2. \qquad (17)$$

A. Group Robust Null Space Property

Now we present the first of our main results, which allows us to establish robust group k-sparse recovery. For notational convenience, define

$$d_{\max} = \max_{j \in [g]} |G_j|, d_{\min} = \min_{j \in [g]} |G_j|,$$

Given integers k, n and a real number $t > 1$, define

$$\bar{k} := [1 + (t-1)d_{\max}]k. \tag{18}$$

Also define

$$\nu := \sqrt{(t-1)t} - (t-1). \tag{19}$$

It is easy to verify via elementary calculus that $\nu = 0$ if $t = 1$, ν is an increasing function of t, and $\nu \to 0.5$ as $t \to \infty$.

Theorem 3: Suppose that the matrix A satisfies the GRIP of order $\bar{k}$ with constant $\delta_{G,\bar{k}} < \bar{\delta}_G$, where[8]

$$\bar{\delta}_G = \nu(1-\nu)\left(\frac{\nu^2 d_{\max}}{2(t-1)d_{\min}} + 0.5 - \nu + \nu^2\right)^{-1}. \tag{20}$$

Then A satisfies the ℓ_2 GRNSP with constants ρ_G, τ_G defined as follows:

$$\rho_G := c_G/a < 1, \tau_G := b\sqrt{k}/a^2, \tag{21}$$

where

$$\begin{aligned} a &:= [\nu(1-\nu) - \delta(0.5 - \nu(1-\nu))]^{1/2} \\ &= \frac{[(1-\delta) - (1+\delta)(1-2\nu)^2]^{1/2}}{2}, \end{aligned} \tag{22}$$

$$b := \nu(1-\nu)\sqrt{1+\delta}, \tag{23}$$

$$c_G := \left[\frac{\delta\nu^2 d_{\max}}{2(t-1)d_{\min}}\right]^{1/2}, \tag{24}$$

and δ is shorthand for $\delta_{G,\bar{k}}$.

A simplification is possible in the case where all groups have the same size, so that $d_{\max} = d_{\min}$.

Theorem 4: Suppose $n = dg$ for some integers d, g, and that all groups have size d. Suppose the matrix A satisfies the GRIP of order $\bar{k}$ with constant $\delta_{G,\bar{k}} < \bar{\delta} = \sqrt{(t-1)/t}$. Then A satisfies the ℓ_2 GRNSP with constants ρ, τ defined as follows:

$$\rho := c/a < 1, \tau := b\sqrt{k}/a^2, \tag{25}$$

where a, b are as in (22) and (23) respectively, and

$$c := \left[\frac{\delta\nu^2}{2(t-1)}\right]^{1/2}, \tag{26}$$

and δ is shorthand for $\delta_{G,\bar{k}}$.

B. Error Bounds on the Residual Vector

Suppose x is the unknown vector and $\hat{x}$ is the recovered vector, constructed according to (14). In this subsection we present bounds for $\|\hat{x} - x\|_p$ for $p \in [1, 2]$.

Theorem 5: Suppose that the measurement matrix A satisfies the conditions of Theorem 3. Then the formulation (14) achieves robust group sparse recovery of order k. Specifically, let $\hat{x}$ be defined as in (14), and let $h = \hat{x} - x$ denote the residual error vector. Then

$$\|h\|_1 \leq \frac{2}{1-\rho_G}[(1+\rho_G)\sigma_{G,k}(x, \|\cdot\|_1) + 2\tau_G\epsilon], \tag{27}$$

and for all $p \in [1, 2]$,

$$\begin{aligned} \|h\|_p \leq{}& \frac{2}{1-\rho_G}\left(1 + \frac{\rho_G}{k^{1-1/p}}\right)\sigma_{G,k}(x, \|\cdot\|_1) \\ &+ \left[\frac{2}{1-\rho_G}\left(1 + \frac{\rho_G}{k^{1-1/p}}\right) + \frac{2}{k^{1-1/p}}\right]\tau_G\epsilon, \end{aligned} \tag{28}$$

where both ρ_G and τ_G are defined in (21), and $\sigma_{G,k}(x, \|\cdot\|_1)$ denotes the group k-sparsity index of x defined by

$$\sigma_{G,k}(x, \|\cdot\|_1) = \inf_{S \in \mathrm{GkS}} \|x - x_S\|_1.$$

C. Sample Complexity Estimates

In this subsection we study the case where the measurement matrix A equals

$$A = (1/\sqrt{m})\Phi, \tag{29}$$

where Φ consists of mn independent samples of a zero-mean, unit-variance random variable X that satisfies

$$E[\exp(\theta X)] \leq \exp(\bar{c}\theta^2), \ \forall \theta \in \mathbb{R}. \tag{30}$$

for some constant $\bar{c}$. Such a random variable is said to be sub-Gaussian. In such a case, it can be shown that there exists a constant $\tilde{c}$ such that

$$\Pr\{|\|Au\|_2^2 - \|u\|_2^2| \geq t\|u\|_2^2\} \leq 2\exp(\tilde{c}mt^2), \ \forall t \in \mathbb{R}. \tag{31}$$

The relationship between the sub-Gaussian parameter $\bar{c}$ in (30) and the constant $\tilde{c}$ can be derived by combining various arguments in [8]. See in particular [8, Lemma 9.8].

Lemma 4: Suppose X is a zero-mean, unit variance random variable, and satisfies (30) for some constant $\bar{c}$. Define

$$\gamma = 2, \zeta = 1/(4\bar{c}), \alpha = \gamma e^{-\zeta} + e^{\zeta}. \tag{32}$$

Then (31) is satisfied with

$$\tilde{c} = \frac{\zeta^2}{2(2\alpha + \zeta)}. \tag{33}$$

By adapting [8, Th. 9.11] via replacing k by tk throughout, we can give a bound on the number of measurements m that suffice to ensure that A defined in (29) satisfies the RIP or order tk with constant $\delta_{tk} < \delta$, with probability $\geq 1 - \xi$.

Theorem 6: Given integers $n, k < n$ and a small number $\xi \in (0, 1)$, choose any $t > 1$ and any $\delta < \sqrt{(t-1)/t}$. Let X be a sub-Gaussian random variable that satisfies (30) for some $\bar{c} > 0$, and define A as in (29). Choose an integer m_C such that

$$m_C \geq \frac{1}{\tilde{c}\delta^2}\left(\frac{4}{3}tk\ln\frac{en}{tk} + \frac{14tk}{3} + \frac{4}{3}\ln\frac{2}{\xi}\right). \tag{34}$$

Then A satisfies the RIP of order tk with constant $\delta_{tk} < \delta$ with probability $\geq 1 - \xi$. Consequently the pair $(A, \Delta_{\mathrm{BP}})$ achieves robust sparse recovery of order k with probability at least $1 - \xi$.

Now we present an extension of Theorem 6 to group sparse recovery.

Theorem 7: Given integers n, k, choose any $\delta < \bar{\delta}_G$ where $\bar{\delta}_G$ is defined in (20). Choose any $t > 1$, define $\bar{k}$ as in (18), and

[8] As before, when we write $\delta_{G,\alpha}$ and α is not necessarily an integer, we mean $\delta_{G,\lceil\alpha\rceil}$.

define

$$\phi = \left\lceil \frac{\bar{k}}{d_{\min}} \right\rceil. \tag{35}$$

Let X be a sub-Gaussian random variable that satisfies (30) for some $\bar{c} > 0$, and define A as in (29). Choose an integer m_G such that

$$m_G \geq \frac{1}{\tilde{c}\delta^2}\left(\frac{4}{3}\phi \ln \frac{eg}{\phi} + \frac{14\phi}{3} + \frac{4}{3}\ln\frac{2}{\xi}\right). \tag{36}$$

Then A satisfies the GRIP of order $\bar{k}$ with constant $\delta_{\bar{k}} < \delta$ with probability $\geq 1 - \xi$. Consequently the pair $(A, \Delta_{\mathrm{BP}})$ achieves robust group sparse recovery of order k with probability at least $1 - \xi$.

Now let us specialize Theorem 7 to the case where all groups have the same size d (so that $n = gd$), and in addition, $k = ld$ for some integer l. In this case (35) becomes

$$\phi = [1 + (t-1)d]l,$$

while the bound for m_G in (36) becomes

$$m_G \geq \frac{1}{\tilde{c}\delta^2}\left(\frac{4}{3}\phi \ln \frac{eg}{\phi} + \frac{14\phi}{3} + \frac{4}{3}\ln\frac{2}{\xi}\right). \tag{37}$$

The above bound m_G for the number of samples that suffices for robust *group* sparse recovery should be compared against the number m_C for conventional sparsity in (34). In this case the estimate for m_C from (34) becomes

$$m_C \geq \frac{1}{\tilde{c}\delta^2}\left(\frac{4}{3}tld \ln \frac{eg}{tl} + \frac{14tld}{3} + \frac{4}{3}\ln\frac{2}{\xi}\right), \tag{38}$$

after noting that $n/k = g/l$.

Theorem 8: If $d > 1$ and $tld = tk < g$, then $m_G < m_C$, where m_G is defined in (37), and m_C is defined in (38).

Thus, in the case where all groups are of equal size, achieving robust group k-sparse recovery using Theorem 8 requires *fewer measurements* than for conventional sparsity using Theorem 6, whenever tk is smaller than the number of groups, which is a very reasonable assumption. On the other hand, if there is a very large disparity between group sizes, the estimate given by (36) could be larger than the estimate for conventional sparsity given in (34); however, it can also be smaller. This is illustrated in the numerical example in Section VII.

V. Main Results–II: Conventional Sparse Recovery

In this section we present our results regarding conventional sparsity. The sufficient condition for sparse recovery is an immediate special case of Theorem 3. However, the bounds for the ℓ_p-norm of the residual error require a separate proof.

In the case of conventional sparsity, all groups have cardinality one, GRIP becomes RIP, and GRNSP becomes RNSP. Thus Theorem 3 immediately implies the following.

Theorem 9: Given integers k, n and a real number $t > 1$, suppose that the matrix A satisfies the RIP of order tk with constant $\delta_{tk} = \delta < \bar{\delta} := \sqrt{(t-1)/t}$. Then A satisfies the RNSP with constants

$$\rho = c/a < 1, \tau = b\sqrt{k}/a^2,$$

where a, b, c are as in (22), (23), and (26) respectively.

Because conventional sparsity is a special case of group sparsity where each group has cardinality one, it is possible to obtain bounds from Theorem 5 to generate bounds on the residual error for conventional sparsity. However, we can do better than this.

Theorem 10: Suppose that $A \in \mathbb{R}^{m \times n}$ satisfies the ℓ_2-robust null space property of order k as defined in Definition 3, and let $\hat{x}$ denote the solution of (14). Then

$$\|\hat{x} - x\|_1 \leq 2\frac{1+\rho}{1-\rho}\sigma_k(x, \|\cdot\|_1) + \frac{4\tau}{1-\rho}\epsilon. \tag{39}$$

Moreover, for all $p \in [1, 2]$, we have that

$$\|\hat{x} - x\|_p \leq \frac{1}{k^{1-1/p}} \cdot \frac{2}{1-\rho}[(1+2\rho)\sigma_k(x, \|\cdot\|_1) + 3\tau\epsilon]. \tag{40}$$

VI. Discussion of Our Contributions

A. Group Sparsity

When all groups have the same size, the GRIP defined here is essentially the same as the group- or block-RIP defined in earlier papers. However, the sufficient conditions we derive are weaker. When all groups have equal size d, the sufficient condition proved here in Theorem 4 is that, for some $t > 1$, we have

$$\delta_{G,[1+(t-1)d]k} < \sqrt{\frac{t-1}{t}}.$$

In particular, if we set $t = 2$, we get the bound

$$\delta_{G,(1+d)k} < \sqrt{1/2} \approx 0.707.$$

This can be compared to the bound derived in various papers including [12, Th. 1] or [16, Definition 2 and Th. 1], namely

$$\delta_{G,2dk} < \sqrt{2} - 1 \approx 0.414.$$

Obviously $\delta_{G,(1+d)k} \leq \delta_{G,2dk}$, and $\sqrt{2} - 1 < \sqrt{1/2}$. Thus the bound derived here is less conservative. Compared to the bound in [18], the bound is the same, namely $\sqrt{(t-1)/t}$. However, in [18], this is a bound on $\delta_{G,2dk}$, whereas here it is a bound on $\delta_{G,(1+d)k}$. Finally, because we prove our results by establishing the ℓ_2-GRNSP, we can derive bounds on the ℓ_p-norm of the residual error for all $p \in [1, 2]$, as opposed to just the Euclidean norm in existing papers.

Next we discuss the case where the measurement matrix A consists of random samples of sub-Gaussian variables. When all groups have the same size d, and $n = gd, k = ld$ for some integers g, l, the number of samples becomes $O(l \log(g/l))$ as opposed to $O(k \log(n/k))$ for conventional sparsity. This is not a novel observation, and is contained in practically every paper in the area, e.g. [10], [11], and [14] and others. For the case of unequal group sizes the condition in [14, Assumption 4.3] is (in the present notation) a bound on

$$\frac{\delta_{G,k+d_{\max}} + \delta_{G,2k+2d_{\max}}}{1 - \delta_{G,k+d_{\max}}}.$$

See [14, Th. 5.1]. Other papers on the topic cannot handle the case where group sizes are unequal. Thus replacing the "sparsity-inducing" Group LASSO norm with the ℓ_1-norm can sometimes lead to lower bounds for the number of measurements.

B. Conventional Sparsity

Note that the bound in Theorem 9 is precisely the bound given by [2] and stated here as Theorem 1, but with the restriction that $t \geq 4/3$. Here the bound is lowered to $t > 1$. However, the bound $\delta_{tk} < t/(4-t)$ for some $t \in (0, 4/3)$ proved in [4] is not covered by our approach. Moreover, it is clear that if $t \in (1, 4/3]$, then the bound $t/(4-t)$ proved in [4] is superior to $\sqrt{(t-1)/t}$. On the other hand, the method of proof given in [2] and [4] does not establish the robust null space property. Rather, the proof is based on directly manipulating various inequalities. As a result, the results in both [2] and [4] lead only to a bound on the Euclidean norm of the residual error $\hat{x} - x$ when $\hat{x}$ is computed via (14). In contrast, by first establishing that the RIP implies the RNSP, we are able to treat the cases of noise-free and noisy measurements in a common framework, and also to obtain bounds on $\|\hat{x} - x\|_p$ for all $p \in [1, 2]$, and not just for $p = 2$.

The result in Theorem 9 is the best available to date showing that RIP implies the RNSP. Previously the best available result was [7, Proposition 8], in which it is shown that if A satisfies the RIP of order $2k$ with constant $\delta_{2k} < 1/9$, then it also satisfies the ℓ_2-RNSP. The bound $1/9$ is far smaller than the bound $\sqrt{1/2} \approx 0.7071$ that results from Theorem 9.

Equation (39) is the same as [8, Th. 4.19]. However, our method of proof is different, and this leads to an improvement in the bounds for the ℓ_p-norm of the residual error when $p > 1$, when compared to [8, Th. 4.25]. The bound in (40) is an improvement over that in [8, Th. 4.25]. If one were to substitute

$$\|z\|_1 - \|x\|_1 \leq 0$$

into the bound given in that theorem, the result would be

$$\|\hat{x} - x\|_p \leq \frac{1}{k^{1-1/p}} \cdot \frac{2}{1-\rho}[(1+\rho)^2 \sigma_k(x, \|\cdot\|_1) + (3+\rho)\tau\epsilon].$$

The bound in (40) is better in that $(1+\rho)^2$ is replaced by $1 + 2\rho$, and $3 + \rho$ is replaced by 3.

VII. Numerical Example

In this section we illustrate the application of the bounds in (36) and (37). Specifically, we compare the number of measurements for group sparse recovery given in (35) and (37) with the number for conventional sparse recovery using Theorem 6 as given in (34). As shown in [29], unless n is larger than about 10^5, the bound m_C in (34) often exceeds n, which makes "compressed" sensing meaningless. We study four different cases to illustrate the fact that even with small groups of equal size, robust group sparse recovery can require fewer samples than conventional sparse recovery. Moreover, as the minimum size of the groups increases, the advantage is even more on the side of group sparse recovery. The reason for this phenomenon is that, as the minimum group size increases, the total number of groups decreases. Consequently, the cardinality of the number of group sparse sets decreases fairly rapidly. This is the quantity referred to as $C(g, \phi)$ in the proof of Theorem 7.

Specifically, we choose $n = 10^6$ and $k = 60$. We use a sub-Gaussian random variable that satisfies the same rate of decay as a standard normal variable, namely $\bar{c} = 1/2$ in (30); see [8, Lemma 7.6]. To compute the RIP constant δ, we choose $t = 1.5$, which gives $1/\sqrt{3} \approx 0.577$ as the upper limit for conventional sparse recovery, as given in Theorem 1. We choose $\delta_C = 0.5$, or about 85% of the limit, as the RIP constant for conventional sparsity. In the case of GRIP, we compute the limit δ_G as 85% of the limit $\bar{\delta}_G$ given by (20). Note that for different choices of group sizes, this threshold would also be different. Finally, for the failure probability ξ we choose 10^{-9} for both conventional and group sparse recovery. Table I shows the number of samples needed by conventional sparsity and group sparsity for various values of $d_{\max}, d_{\min}, g$. The choice of the GRIP constant δ_G for each choice of $d_{\max}, d_{\min}, g$ is also shown in the table. Note that, from Theorem 4, when all groups have the same size d, and both n and k are multiples of d, then the GRIP bound δ_G and RIP bound δ_C are the same.

TABLE I
Comparison of Number of Measurements Required in Conventional and Group Sparsity For Various Group Sizes, With $n = 10^6$ and $k = 60$

$d_{\max}$	$d_{\min}$	g	δ_C	δ_G	m_C	m_G
4	2	300,000	0.5	0.3750	335,862	545,935
10	6	100,000	0.5	0.4091	335,862	296,499
4	4	250,000	0.5	0.5000	335,862	162,492
10	10	100,000	0.5	0.5000	335,862	124,505

It is noteworthy that, even when the GRIP constant δ_G that the matrix A is required to satisfy is smaller than the RIP constant δ_C, the number of samples can be smaller in the case of group sparsity, as happens in row 2 of the table. This is because the number of group sparse sets is substantially smaller than the number of sparse sets. As a final example, we increased k to 300, and chose the group sizes to be uniform at 50 (thus leading to 20,000 groups). With this choice, $m_C = 1,464,244$, that is, *more than* the size of the vector, whereas $m_G = 393,153$. Thus group sparse recovery is feasible when conventional sparse recovery is not feasible.

VIII. Proofs of Main Results

A. Polytope Decomposition Lemma

The key to the results in [2] is Lemma 1.1 of that paper, which the authors call the "polytope decomposition lemma." In this subsection we generalize this lemma to the case of group sparsity. Before presenting the lemma, we introduce a couple of terms. Given a vector $v \in \mathbb{R}^n$, we define the group support set of v, denoted by $\mathrm{Gsupp}(v)$, as

$$\mathrm{Gsupp}(v) := \{j \in [g] : v_{G_j} \neq 0\}. \quad (41)$$

Thus $\mathrm{Gsupp}(v)$ denotes the subset of the groups on which v has a nonzero support. Obviously $|\mathrm{Gsupp}(v)|$ is the number of distinct groups on which v is supported.

Lemma 5: Given a vector $v \in R^n$ such that,

$$\|v_{G_j}\|_1 \leq \alpha, \ \forall j \in [g], \text{ and } \|v\|_1 \leq s\alpha \quad (42)$$

for some integer s, there exist an integer N and vectors $u_i, i \in [N]$ such that

- $\mathrm{supp}(u_i) \subseteq \mathrm{supp}(v), \ \forall i \in [N]$.
- $\|u_i\|_1 = \|v\|_1, \ \forall i \in [N]$.

- u_i is group $sd_{\max}$-sparse for each i, and finally
- v is a convex combination of $u_i, i \in [N]$.

Remarks: In the case of conventional sparsity, each group G_j consists of the singleton $\{j\}$. In this case the condition $\|v_{G_j}\|_1 \leq \alpha$, $\forall j \in [g]$ reduces to $|v_j| \leq \alpha\ \forall j \in [n]$, or equivalently, $\|v\|_\infty \leq \alpha$. Moreover, $d_{\max} = 1$, in which case all vectors u_i are s-sparse. This is precisely [2, Lemma 1.1].

Proof: The proof is by induction. Define a subset of $\mathbb{R}^n$ as follows:

$$X := \{v \in \mathbb{R}^n : \|v_{G_j}\|_1 \leq \alpha\ \forall j \in [g], \|v\|_1 \leq s\alpha\}.$$

To begin the inductive process, suppose $|\mathrm{Gsupp}(v)| \leq s$. Then v is itself $sd_{\max}$-sparse. So we can take $N = 1$ and $u_1 = v$. Now suppose that the lemma is true for all $v \in X$ such that $|\mathrm{Gsupp}(v)| = r - 1$ where $r - 1 \geq s$. It is shown that the lemma is also true for all $v \in X$ satisfying $|\mathrm{Gsupp}(v)| = r$.

Let $Q \subseteq [g]$ denote the index set $\{j \in [g] : v_{G_j} \neq 0\}$, and observe that $|Q| = |\mathrm{Gsupp}(v)| = r$ by assumption. Then v can be expressed as $v = \sum_{j \in Q} v_{G_j}$. Now arrange the vectors v_{G_j} in decreasing order of their ℓ_1-norm. Denote the permuted vectors as p_1 through p_r. Define $a_i := \|p_i\|_1$, and $\hat{p}_i = (1/a_i)p_i$. Then each $\hat{p}_i$ has unit ℓ_1-norm. Moreover $a_i \geq a_{i+1}$ for all i, and $v = \sum_{i=1}^r p_i = \sum_{i=1}^r a_i \hat{p}_i$. Also, because the ℓ_1-norm is decomposable and the p_i have nonoverlapping support sets, it follows that $\|v\|_1 = \sum_{i=1}^r a_i$.

Now define a set

$$D := \left\{ \beta \in [r-1] : \sum_{i=\beta}^r a_\beta \leq (r-\beta)\alpha \right\}.$$

Then $1 \in D$ because

$$\sum_{i=1}^r a_i = \|v\|_1 \leq s\alpha \leq (r-1)\alpha.$$

Therefore D is nonempty. Now, by a slight abuse of notation, let β again denote the largest element of the set D. This implies that

$$\sum_{i=\beta}^r a_i \leq (r-\beta)\alpha, \sum_{i=\beta+1}^r a_i > (r-\beta-1)\alpha. \tag{43}$$

Define the constants

$$b_t := \frac{1}{r-\beta} \sum_{i=\beta}^r a_i - a_t, \beta \leq t \leq r.$$

Since the first term on the right side is independent of t, and $a_{t+1} \leq a_t$, it follows that $b_{t+1} \geq b_t$. Also

$$\begin{aligned} b_\beta &= \frac{1}{r-\beta} \sum_{i=\beta}^r a_i - a_\beta \\ &= \frac{1}{r-\beta} \sum_{i=\beta+1}^r a_i - \frac{r-\beta-1}{r-\beta} a_\beta \\ &\geq \frac{1}{r-\beta} \left[\sum_{i=\beta+1}^r a_i - (r-\beta-1)\alpha \right] > 0, \end{aligned}$$

where the last two steps follow from $a_i \leq \alpha$ for all i, and from the second inequality in (43). Also, it is easy to verify that

$$\sum_{i=\beta}^r a_i = (r-\beta) \sum_{i=\beta}^r b_i \tag{44}$$

Next, for $t = \beta, \ldots, r$, define

$$w_t := \sum_{i=1}^{\beta-1} a_i \hat{p}_i + \left(\sum_{i=\beta}^r b_i \right) \sum_{i=\beta, i \neq t}^r \hat{p}_i, \lambda_t := \frac{b_t}{\sum_{i=\beta}^r b_i}. \tag{45}$$

Now observe that

$$0 < \lambda_t < 1, \sum_{t=\beta}^r \lambda_t = 1, \text{ and } v = \sum_{t=\beta}^r \lambda_t w_t.$$

Next, $\mathrm{supp}(w_t) \subseteq \mathrm{supp}(v)$ for all t. Moreover, $|\mathrm{Gsupp}(w_t)| \leq r - 1$ for all t, because the corresponding term $\hat{p}_t$ is missing from the summation in (45). Also, note that each $\hat{p}_i$ has unit ℓ_1-norm. Therefore, for each t between β and r, we have that

$$\begin{aligned} \|w_t\|_1 &= \sum_{i=1}^{\beta-1} a_i + (r-\beta) \sum_{i=\beta}^r b_i \\ &= \sum_{i=1}^{\beta-1} a_i + \sum_{i=\beta}^r a_i = \sum_{i=1}^r a_i = \|v\|_1. \end{aligned}$$

Therefore each $w_t \in X$. By the inductive assumption, each w_t has a convex decomposition as in the statement of the lemma. It follows that v is also a convex combination as in the statement of the lemma. This completes the inductive step. ∎

Lemma 6: Let $u_i, i \in [N]$ be the vectors in the convex combination of Lemma 5. Then

$$\|u_i\|_2^2 \leq \frac{sd_{\max}}{d_{\min}} \alpha^2, \ \forall i \in [N]. \tag{46}$$

Proof: Fix the index $i \in [N]$. Define the index set

$$B_i := \{j \in [g] : (u_i)_{G_j} \neq 0\}.$$

Let $c_i = |B_i|$. Because u_i is $sd_{\max}$-sparse, it follows that $c_i \leq \frac{sd_{\max}}{d_{\min}}$. Moreover, for each index $j \in B_i$, we have that

$$\|(u_i)_{G_j}\|_2 \leq \|(u_i)_{G_j}\|_1 \leq \|u_i\|_1 = \alpha.$$

Now observe that

$$u_i = \sum_{j \in B_i} (u_i)_{G_j}.$$

Next, note that the various vectors $(u_i)_{G_j}$ are supported on disjoint sets. Therefore

$$\|u_i\|_2^2 = \sum_{j \in B_i} \|(u_i)_{G_j}\|_2^2.$$

Since there are c_i terms in the above summation, and each term is no larger than α^2, it follows that

$$\|u_i\|_2^2 \leq c_i \alpha^2 \leq \frac{sd_{\max}}{d_{\min}} \alpha^2,$$

which is the desired conclusion (46). ∎

B. Group Robust Null Space Property

Proof of Theorem 3: Recall the constants $\nu, a, b, c_G, \rho_G, \tau_G$ defined in the statement of Theorem 3. We will make use of these constants in the proof. Let $h \in \mathbb{R}^n$ be arbitrary. The objective is to establish that the inequality (16) is satisfied with ρ_G, τ_G defined as above.

Let $h_{\Lambda_0}, h_{\Lambda_1}, h_{\Lambda_2}, \ldots, h_{\Lambda_s}$ be an optimal group-k-sparse decomposition of h. This means the following: First,

$$h_{\Lambda_0} = \underset{\text{supp}(z) \in \text{GkS}}{\text{argmin}} \|x - z\|.$$

Next, for $i \geq 1$,

$$h_{\Lambda_i} = \underset{\text{supp}(z) \in \text{GkS}}{\text{argmin}} \left\| x - \sum_{j=0}^{i-1} h_{\Lambda_j} - z \right\|.$$

Now denote $h_{\Lambda_0^c} = h^*$. Define sets S_1 and S_2 as follows:

$$S_1 = \left\{ j : \|h^*_{G_j}\|_1 > \frac{\|h_{\Lambda_0^c}\|_1}{k(t-1)}, \ \forall j \in [g] \right\},$$

$$S_2 = \left\{ j : \|h^*_{G_j}\|_1 \leq \frac{\|h_{\Lambda_0^c}\|_1}{k(t-1)}, \ \forall j \in [g] \right\}.$$

Let $GS_1 = \cup_{j \in S_1} G_j$ and $GS_2 = \cup_{j \in S_2} G_j$. Now define

$$h^{(0)} = h_{\Lambda_0}, h^{(1)} = h^*_{GS_1}, h^{(2)} = h^*_{GS_2}.$$

Then we have

$$h_{\Lambda_0^c} = h^* = h^*_{GS_1} + h^*_{GS_2} = h^{(1)} + h^{(2)}.$$

Let $r = |S_1|$, and note that $r \leq k(t-1)$. This is because, by the manner in which we defined the set S_1, it follows that

$$\|h_{\Lambda_0^c}\|_1 \geq \|h^{(1)}\|_1 > r \frac{\|h_{\Lambda_0^c}\|_1}{k(t-1)}.$$

Next we establish upper bound on $\|h^{(2)}\|_1$. Because of the definition of set S_1, it follows that

$$\|h^{(1)}\|_1 \geq r \frac{\|h_{\Lambda_0^c}\|_1}{k(t-1)}. \tag{47}$$

Therefore

$$\begin{aligned} \|h^{(2)}\|_1 &= \|h_{\Lambda_0^c}\|_1 - \|h^{(1)}\|_1 \\ &\leq \|h_{\Lambda_0^c}\|_1 - r \frac{\|h_{\Lambda_0^c}\|_1}{k(t-1)} \\ &= [k(t-1) - r] \frac{\|h_{\Lambda_0^c}\|_1}{k(t-1)}. \end{aligned} \tag{48}$$

By the definition of set S_2

$$\|h^{(2)}_{G_j}\|_1 \leq \frac{\|h_{\Lambda_0^c}\|_1}{k(t-1)}, \ \forall j \in [g]. \tag{49}$$

From (48) and (49), we see that the vector $h^{(2)}$ satisfies the hypotheses of Lemma 5 with

$$\alpha = \frac{\|h_{\Lambda_0^c}\|_1}{k(t-1)}, s = k(t-1) - r.$$

Therefore we can apply Lemma 5 to $h^{(2)}$. So $h^{(2)}$ can be represented as

$$h^{(2)} = \sum_{i=1}^{N} \lambda_i u_i, \tag{50}$$

where each u_i is group $(k(t-1) - r)d_{\max}$-sparse, $h^{(1)}$ is group $(rd_{\max})$-sparse, and $h^{(0)}$ is group k-sparse. Therefore $u_i + h^{(1)} + h^{(0)}$ has group sparsity no larger than

$$\begin{aligned} k + rd_{\max} + (k(t-1) - r)d_{\max} &= k[1 + (t-1)d_{\max}] \\ &= \bar{k} \end{aligned}$$

for each $i \in [N]$. Now let, for all $i \in [N]$,

$$\begin{aligned} x_i &= \frac{1}{2}\left(h^{(0)} + h^{(1)}\right) + \frac{\nu}{2} u_i, \\ z_i &= \frac{1 - 2\nu}{2}\left(h^{(0)} + h^{(1)}\right) - \frac{\nu}{2} u_i, \\ \gamma &= x_i + z_i = (1 - \nu)\left(h^{(0)} + h^{(1)}\right), \\ \beta_i &= x_i - z_i = \nu\left(h^{(0)} + h^{(1)} + u_i\right). \end{aligned}$$

Then

$$\begin{aligned} \sum_{i=1}^{N} \lambda_i \langle A\gamma, A\beta_i \rangle &= \left\langle A\gamma, A \sum_{i=1}^{N} \lambda_i \beta_i \right\rangle \\ &= \nu(1-\nu)\langle A(h^{(0)} + h^{(1)}), Ah \rangle, \end{aligned} \tag{51}$$

where we make use of (50) and the fact that $h^{(0)} + h^{(1)} + h^{(2)} = h$. However, for each index set i, we have that

$$\begin{aligned} \langle A\gamma, A\beta_i \rangle &= \langle Ax_i + Az_i, Ax_i - Az_i \rangle \\ &= \|Ax_i\|_2^2 - \|Az_i\|_2^2. \end{aligned}$$

Therefore it follows that

$$\sum_{i=1}^{N} \lambda_i (\|Ax_i\|_2^2 - \|Az_i\|_2^2) = \nu(1-\nu)\left\langle A(h^{(0)} + h^{(1)}), Ah \right\rangle,$$

$$\sum_{i=1}^{N} \lambda_i \|Ax_i\|_2^2 = \sum_{i=1}^{N} \lambda_i \|Az_i\|_2^2 + \nu(1-\nu)\left\langle A(h^{(0)} + h^{(1)}), \ Ah \right\rangle.$$

Since $x_i, \ z_i, \ (h^{(0)} + h^{(1)})$ are all group $\bar{k}$-sparse, it follows from the GRIP and Schwarz' inequality that

$$(1-\delta)\sum_{i=1}^{N} \lambda_i \|x_i\|_2^2 \leq (1+\delta)\sum_{i=1}^{N} \lambda_i \|z_i\|_2^2 + \nu(1-\nu)\|A(h^{(0)} + h^{(1)})\|_2 \cdot \|Ah\|_2.$$

Since $h^{(0)}, \ h^{(1)}$ and u_i have disjoint support sets, it follows that, for all $i \in [N]$, we have

$$\|x_i\|_2^2 = 0.25\left(\|(h^{(0)} + h^{(1)})\|_2^2 + \nu^2\|u_i\|_2^2\right),$$

$$\|z_i\|_2^2 = 0.25\left[(1-2\nu)^2\|(h^{(0)} + h^{(1)})\|_2^2 + \nu^2\|u_i\|_2^2\right],$$

Substituting these relationships, multiplying both sides by 4, and noting that $\sum_{i=1}^N \lambda_i = 1$, leads to

$$(1-\delta)\cdot \left[\|(h^{(0)}+h^{(1)})\|_2^2 + \nu^2 \sum_{i=1}^N \lambda_i \|u_i\|_2^2 \right]$$
$$\leq (1+\delta) \left[(1-2\nu)^2 \|(h^{(0)}+h^{(1)})\|_2^2 + \nu^2 \sum_{i=1}^N \lambda_i \|u_i\|_2^2 \right] + 4\nu(1-\nu)\|A(h^{(0)}+h^{(1)})\|_2 \cdot \|Ah\|_2,$$

or upon rearranging,

$$\|(h^{(0)}+h^{(1)})\|_2^2 \cdot [(1-\delta)-(1+\delta)(1-2\nu)^2]$$
$$\leq 2\delta\nu^2 \sum_{i=1}^N \lambda_i \|u_i\|_2^2 + 4\nu(1-\nu)\|A(h^{(0)}+h^{(1)})\|_2 \cdot \|Ah\|_2.$$

Recall that

$$\alpha = \frac{\|h_{\Lambda_0^c}\|_1}{k(t-1)}, s = k(t-1)-r.$$

Substituting these values into (46), we get that

$$\|u_i\|_2^2 \leq [k(t-1)-r] \frac{d_{\max}}{d_{\min}} \frac{\|h_{\Lambda_0^c}\|_1^2}{k^2(t-1)^2}$$
$$\leq k(t-1) \frac{d_{\max}}{d_{\min}} \frac{\|h_{\Lambda_0^c}\|_1^2}{k^2(t-1)^2}$$
$$= \frac{d_{\max}}{d_{\min}} \frac{\|h_{\Lambda_0^c}\|_1^2}{k(t-1)}.$$

Substituting this bound, which is independent of i, into the above inequality, and noting that $\sum_{i=1}^N \lambda_i = 1$, we get

$$\|(h^{(0)}+h^{(1)})\|_2^2 [(1-\delta)-(1+\delta)(1-2\nu)^2]$$
$$\leq \frac{2\delta\nu^2 d_{\max}}{d_{\min}} \frac{\|h_{\Lambda_0^c}\|_1^2}{k(t-1)} + 4\nu(1-\nu)\sqrt{1+\delta}\|(h^{(0)}+h^{(1)})\|_2 \cdot \|Ah\|_2.$$

Denote $\|(h^{(0)}+h^{(1)})\|_2$ by f and invoke the definition of the constants a, b, c from (22) and (23). This gives

$$4f^2a^2 \leq 4c^2 \frac{\|h_{\Lambda_0^c}\|_1^2}{k} + 4bf\|Ah\|_2,$$

or after dividing both the sides by 4 and rearranging,

$$f^2a^2 - bf\|Ah\|_2 \leq c^2 \frac{\|h_{\Lambda_0^c}\|_1^2}{k}.$$

The next step is to complete the square on left side of the above inequality. This gives

$$f^2a^2 - bf\|Ah\|_2 + \frac{b^2}{4a^2}\|Ah\|_2^2 \leq \frac{b^2}{4a^2}\|Ah\|_2^2 + c^2 \frac{\|h_{\Lambda_0^c}\|_1^2}{k},$$

or equivalently,

$$\left[af - \frac{b}{2a}\|Ah\|_2 \right]^2 \leq \frac{b^2}{4a^2}\|Ah\|_2^2 + c^2 \frac{\|h_{\Lambda_0^c}\|_1^2}{k}.$$

Taking the square root on both sides, and using the obvious inequality that $\sqrt{x^2+y^2} \leq x+y$ whenever $x, y \geq 0$, leads to

$$af - (b/2a)\|Ah\|_2 \leq (b/2a)\|Ah\|_2 + c\frac{\|h_{\Lambda_0^c}\|_1}{\sqrt{k}},$$

or upon rearranging and replacing f by $\|(h^{(0)}+h^{(1)})\|_2$,

$$a\|(h^{(0)}+h^{(1)})\|_2 \leq (b/a)\|Ah\|_2 + c\frac{\|h_{\Lambda_0^c}\|_1}{\sqrt{k}}.$$

Dividing both the sides by a and observing that $h_{\Lambda_0} = h^{(0)}$ and

$$\|h^{(0)}\|_2 \leq \|(h^{(0)}+h^{(1)})\|_2,$$

we get

$$\|h_{\Lambda_0}\|_2 \leq \|(h^{(0)}+h^{(1)})\|_2 \leq \frac{b}{a^2}\|Ah\|_2 + \frac{c}{a}\frac{\|h_{\Lambda_0^c}\|_1}{\sqrt{k}}$$
$$= \frac{b\sqrt{k}}{a^2\sqrt{k}}\|Ah\|_2 + \frac{c}{a}\frac{\|h_{\Lambda_0^c}\|_1}{\sqrt{k}}.$$

This inequality is of the form (16) with ρ_G, τ_G given as in (21).

The proof is therefore complete once it is shown that $\rho_G = c_G/a < 1$ if and only if $\delta_{G,\bar{k}} < \bar{\delta}_G$. Towards this end, define

$$\alpha = \frac{\nu^2 d_{\max}}{2(t-1)d_{\min}}.$$

Then $c_G^2 = \alpha\delta$. Next, observe that $c_G < a$ if and only if $c_G^2 < a^2$. Now we can invoke the definitions of a and c_G from (22) and (24), which leads to

$$\begin{aligned} c_G^2 < a^2 &\iff \alpha\delta < \nu(1-\nu) - \delta(0.5-\nu(1-\nu)) \\ &\iff \delta[\alpha + 0.5 - \nu(1-\nu)] < \nu(1-\nu) \\ &\iff \delta < \bar{\delta}_G, \end{aligned}$$

where δ is shorthand for $\delta_{G,\bar{k}}$. ∎

Proof of Theorem 4: If all groups have the same size, then $d_{\max} = d_{\min} = d$, and the bound (20) on the restricted isometry constant $\delta_{G,\bar{k}}$ becomes

$$\bar{\delta} = \nu(1-\nu)\left(\frac{\nu^2}{2(t-1)} + 0.5 - \nu + \nu^2 \right)^{-1}. \quad (52)$$

The objective is to show that $\bar{\delta} = \sqrt{(t-1)/t} =: \psi$ say. From (52), the statement that $\bar{\delta} = \psi$ is equivalent to

$$\nu(1-\nu) = \frac{\psi}{2}\left(\frac{\nu^2}{(t-1)} + 1 \right) - \psi\nu(1-\nu),$$

which in turn is equivalent to

$$2(1+\psi)\nu(1-\nu) = \psi\left(\frac{\nu^2}{(t-1)} + 1 \right). \quad (53)$$

Now note that, from the definition (19) of the constant ν, it follows that

$$\nu = t\sqrt{\frac{t-1}{t}} - (t-1) = t\psi - t + 1 = 1 - t(1-\psi),$$

and

$$1-\nu = t(1-\psi).$$

Therefore the left side of (53) becomes

$$2(1+\psi)\nu(1-\nu) = 2t(1-\psi^2)\nu.$$

However

$$t(1-\psi^2) = t\left(1-\frac{t-1}{t}\right) = 1.$$

Therefore the left side of (53) equals 2ν. The proof is therefore complete if it can be shown that the right side of (53) also equals 2ν. Towards this end, note that

$$\nu^2 = t(t-1) - 2(t-1)\sqrt{t(t-1)} + (t-1)^2,$$

$$\frac{\nu^2}{t-1} = t - 2\sqrt{t(t-1)} + t - 1,$$

$$\frac{\nu^2}{t-1} + 1 = 2(t - \sqrt{t(t-1)}),$$

and finally

$$\psi\left(\frac{\nu^2}{t-1}+1\right) = 2\sqrt{\frac{t-1}{t}}\left(t-\sqrt{t(t-1)}\right) = 2\left[\sqrt{t(t-1)}-(t-1)\right] = 2\nu.$$

Proof of Theorem 9: This consists of the observation that if $d_{\max} = d_{\min} = 1$, then $\bar{k} = [1+(t-1)d_{\max}]k = tk$. ■

C. Error Bounds on the Recovered Vector

Proof of Theorem 5: Define $\hat{x}$ as in (14), and let $h = \hat{x} - x$ denote the residual error. Then by definition we have that $\|\hat{x}\|_1 \le \|x\|_1$. Let $x_{S_0}, x_{S_1}, \ldots, x_{S_b}$ be an optimal group k-sparse decompostion of x. Then

$$\|x_{S_0^c} + h_{S_0^c}\|_1 + \|x_{S_0} + h_{S_0}\|_1 \le \|x_{S_0^c}\|_1 + \|x_{S_0}\|_1.$$

Applying triangle inequality twice to the left hand side of the above inequality, we get

$$\|x_{S_0}\|_1 - \|h_{S_0}\|_1 - \|x_{S_0^c}\|_1 + \|h_{S_0^c}\|_1 \le \|x_{S_0^c}\|_1 + \|x_{S_0}\|_1.$$

Cancelling the common term $\|x_{S_0}\|_1$ and denoting $\|x_{S_0^c}\|$ by $\sigma_{k,G}(x, \|\cdot\|_1) = \sigma_{k,G}$, we get

$$\|h_{S_0^c}\|_1 - \|h_{S_0}\|_1 \le 2\sigma_{k,G} \tag{54}$$

Now let $h_{\Lambda_0}, h_{\Lambda_1}, \ldots, h_{\Lambda_s}$ be an optimal group k-sparse decomposition of h. Then

$$\|h_{\Lambda_0}\|_1 \ge \|h_{S_0}\|_1, \text{ and } \|h_{\Lambda_0^c}\|_1 \le \|h_{S_0^c}\|_1.$$

Using the above facts in (54), we get

$$\|h_{\Lambda_0^c}\|_1 - \|h_{\Lambda_0}\|_1 \le 2\sigma_{k,G}. \tag{55}$$

Next, because both x and $\hat{x}$ are feasible for the optimization problem in (14), we get

$$\|Ah\|_2 = \|(A\hat{x}-y)-(Ax-y)\|_2 \le 2\epsilon.$$

Using the inequality (17) and the above fact, we have that

$$\|h_{\Lambda_0}\|_1 \le \rho_G \|h_{\Lambda_0^c}\|_1 + 2\tau_G\epsilon. \tag{56}$$

Now the two inequalities (55) and (56) can be neatly expressed in the form

$$\begin{bmatrix} 1 & -1 \\ -\rho_G & 1 \end{bmatrix} \begin{bmatrix} \|h_{\Lambda_0^c}\|_1 \\ \|h_{\Lambda_0}\|_1 \end{bmatrix} \le \begin{bmatrix} 2\sigma_{k,G} \\ 2\tau_G\epsilon \end{bmatrix}. \tag{57}$$

Let the M denote the coefficient matrix on the left hand side. Then, because $\rho_G < 1$, it follows that all elements of

$$M^{-1} = \frac{1}{1-\rho_G}\begin{bmatrix} 1 & 1 \\ \rho_G & 1 \end{bmatrix}$$

are positive. Therefore we can multiply both the sides of (57) by M^{-1}, which gives

$$\begin{bmatrix} \|h_{\Lambda_0^c}\|_1 \\ \|h_{\Lambda_0}\|_1 \end{bmatrix} \le \frac{1}{1-\rho_G}\begin{bmatrix} 1 & 1 \\ \rho_G & 1 \end{bmatrix}\begin{bmatrix} 2\sigma_{k,G} \\ 2\tau_G\epsilon \end{bmatrix} = \frac{2}{1-\rho_G}\begin{bmatrix} (\sigma_{k,G}+\tau_G\epsilon) \\ (\rho_G\sigma_{k,G}+\tau_G\epsilon) \end{bmatrix} \tag{58}$$

Finally using the triangle inequality, we get

$$\|h\|_1 \le \|h_{\Lambda_0^c}\|_1 + \|h_{\Lambda_0}\|_1 = \begin{bmatrix} 1 & 1 \end{bmatrix}\begin{bmatrix} \|h_{\Lambda_0^c}\|_1 \\ \|h_{\Lambda_0}\|_1 \end{bmatrix} \le \frac{2}{1-\rho_G}[(1+\rho_G)\sigma_{k,G} + 2\tau_G\epsilon]$$

This is the same as (27).

Next we derive bounds on $\|h\|_p$ for $p \in [1,2]$. From the triangle inequality,

$$\|h\|_p \le \|h_{\Lambda_0}\|_p + \|h_{\Lambda_0^c}\|_p. \tag{59}$$

Now we will obtain the upper bound for both of the terms in right hand side of (59). It is easy to show that

$$\|h_{\Lambda_0^c}\|_p \le \|h_{\Lambda_0^c}\|_1. \tag{60}$$

Next, it is a ready consequence of Hölder's inequality that

$$\|h_{\Lambda_0}\|_p \le k^{1/p-1/2}\, \|h_{\Lambda_0}\|_2.$$

Using the above fact and the ℓ_2-GRNS property (16), together with $\|Ah\|_2 \le 2\epsilon$, we get

$$\|h_{\Lambda_0}\|_2 \le \frac{\rho_G}{\sqrt{k}}\|h_{\Lambda_0^c}\|_1 + \frac{2\tau_G\epsilon}{\sqrt{k}}. \tag{61}$$

Therefore

$$\|h_{\Lambda_0}\|_p \le \frac{1}{k^{1-1/p}}[\rho_G\|h_{\Lambda_0^c}\|_1 + 2\tau_G\epsilon]. \tag{62}$$

Combining (60) and (62) leads to

$$\|h\|_p \le \left(1+\frac{\rho_G}{k^{1-1/p}}\right)\|h_{\Lambda_0^c}\|_1 + \frac{2\tau_G\epsilon}{k^{1-1/p}}. \tag{63}$$

Now we can substitute the upper bound for $\|h_{\Lambda_0^c}\|_1$ obtained from (58), namely

$$\|h_{\Lambda_0^c}\|_1 \le \frac{2}{1-\rho_G}(\sigma_{k,G}+\tau_G\epsilon).$$

Substituting this bound into (63) leads finally to the bound

$$\|h\|_p \leq \frac{2}{1-\rho_G}\left(1+\frac{\rho_G}{k^{1-1/p}}\right)\sigma_{k,G} + \left[\frac{2}{1-\rho_G}\left(1+\frac{\rho_G}{k^{1-1/p}}\right)+\frac{2}{k^{1-1/p}}\right]\tau_G\epsilon.$$

This is precisely (40). ■

Proof of Theorem 10: In this case the optimal group k-sparse decomposition becomes just the conventional optimal k-sparse decomposition. To prove (40), we proceed as above. However, instead of (60), we use the inequality from [8, Th. 2.5], namely

$$\|h_{\Lambda_0^c}\|_p = \sigma_k(h, \|\cdot\|_p) \leq \frac{1}{k^{1-1/p}}\|h\|_1. \quad (64)$$

Note that an analogous inequality does not exist for *groupk*-sparse decompositions. Now we merely substitute the bound from (64) instead of the bound from (58) into (63); this leads to (40). ■

D. Sample Complexity Estimates

Proof of Theorem 7: The proof is a fairly straight-forward adaptation of that of [8, Th. 9.9, p. 276], and [8, Th. 9.11, p. 278]. By assumption, (31) holds for *every fixed* $u \in \mathbb{R}^n$. By applying compactness arguments, it is shown in the cited proofs that, for *every fixed* subset S of cardinality s in $[n]$, the corresponding $m \times s$ submatrix A_S satisfies the bound

$$\sigma_{\min}(A_S^T A_S) \leq 1-\delta, \text{ and } \sigma_{\max}(A_S^T A_S) \geq 1+\delta, \quad (65)$$

with probability $\geq 1-\theta$, where

$$\theta = 2\left(1+\frac{2}{\rho}\right)^s \exp(-\tilde{c}(1-2\rho)^2\delta^2 m). \quad (66)$$

The above bound holds for *all* constants ρ. See [8, (9.12)]. Now by enumerating *all possible subsets* of $[n]$ of cardinality s, we get that the quantity

$$\xi = \binom{n}{s}\theta =: C(n,s)\theta \quad (67)$$

is an upper bound on the probability that A *fails* to satisfy the RIP of order s with constant δ. By choosing $\rho = 2/(e^{7/2}-1)$, we get the bound in the last (unnumbered) equation in the proof of [8, Th. 9.11, p. 278]. Substituting $s = tk$ gives the bound in (34).

In the case of group sparsity, define ϕ as in (35), and observe that any group $\bar{k}$-subset of $[n]$ can be the union of no more than ϕ sets from the collection $G_1, \ldots, G_g$. Therefore the number of group $\bar{k}$-sparse subsets of $[n]$ is bounded by the combinatorial parameter $C(g,\phi)$. Therefore, replacing $C(n,tk)$ by $C(g,\phi)$, or what is the same, changing n to g and tk to ϕ in (34), gives the desired sample complexity estimate (36). ■

Proof of Theorem 8: The bound in (37) follows readily from that in (36) by substituting $n = gd$, $k = ld$, and $\phi = l[1+(t-1)d]$. The following fact can be easily proved using undergraduate calculus: The function $x \mapsto x\ln(eg/x)$ is strictly increasing for $x < g$. With $n = gd, k = ld$ where $d > 1$, we get

$$\bar{k} = [1+(t-1)d]k = [1+(t-1)d]ld,$$

$$\phi = \frac{\bar{k}}{d} = [1+(t-1)d]l < [d+(t-1)d]l = tld.$$

Now compare the right sides of (37) and (38). First, because $\phi < tld$, we infer that

$$\phi\ln\frac{eg}{\phi} < tld\ln\frac{eg}{tld} < tld\ln\frac{eg}{tl}$$

because $d > 1$. So the first term in (37) is smaller than the corresponding term in (38) if $tld < g$. The second term in (37) is smaller than the corresponding term in (38) because $\phi < tld$, and the third terms are the same in both equations. Therefore $m_G < m_C$ if $tld < g$.

IX. Conclusion

In this paper we have shown that the ℓ_1-norm is the convex relaxation of two commonly used group sparsity indices. Therefore ℓ_1-norm minimization can be used for recovering group sparse vectors, and not just for recovering conventionally sparse vectors. We have presented sufficient conditions for ℓ_1-norm minimization to achieve robust group sparse recovery, which are sometimes less conservative than currently available results, based on minimizing a group LASSO type of norm. We achieved this by introducing a group version of the robust null space property, and showing that GRNSP implies a group restricted isometry property. This relationship is new even for conventional sparsity. When specialized to conventional sparsity, our conditions for group sparse recovery reduce to some known "best possible" bounds proved earlier. We have also derived bounds for the ℓ_p-norm of the residual error between the true vector x and its approximation $\hat{x}$, for all $p \in [1,2]$. These bounds are new even for conventional sparsity and of course also for group sparsity. For the case where the measurement matrix consists of random sub-Gaussian samples, we have derived bounds for the number of samples that suffice for group sparse recovery. When all groups have the same size, our bounds are the same as known bounds, while our bounds are less conservative when group sizes are not all equal. We have illustrated our approach through numerical examples.

There are two interesting avenues of research that are worth pursuing. First, our results extend those in [2] to conventional and group sparsity in terms of the RIP (or group RIP) coefficient of order δ_{tk} when $t > 1$. It appears worthwhile to see whether the approach presented here can also be applied to the case $t \in (0, 4/3)$ studied in [4]. Second, there is yet another model of sparsity referred to as "joint" sparsity in [16] and [17]. It would be worthwhile to study whether the problem of recovering jointly sparse vector is amenable to the approach presented here.

Acknowledgment

The authors would like to thank three anonymous reviewers for very helpful suggestions on previous versions of the paper that have greatly improved the contents and readability of the paper.

REFERENCES

[1] S. Ranjan and M. Vidyasagar, "Tight performance bounds for compressed sensing with group sparsity," in *Proc. Indian Control Conf.*, 2017, pp. 206–208.

[2] T. Cai and A. Zhang, "Sparse representation of a polytope and recovery of sparse signals and low-rank matrices," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 122–132, Jan. 2014.

[3] T. Cai, L. Wang, and G. Xu, "New bounds for restricted isometry constants," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4388–4394, Sep. 2010.

[4] R. Zhang and S. Li, "A proof of conjecture on restricted isometry property constants $\delta_{tk} (0 < t < \frac{4}{3})$," *IEEE Trans. Inf. Theory*, vol. 64, no. 3, pp. 1699–1705, Mar. 2018.

[5] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[6] E. Candès, "The restricted isometry property and its implications for compresed sensing," *Comptes rendus de l'Académie des Sciences, Série I*, vol. 346, pp. 589–592, 2008.

[7] S. Foucart, "Stability and robustness of ℓ_1-minimizations with Weibull matrices and redundant dictionaries," *Linear Algebra Appl.*, vol. 441, pp. 4–21, 2014.

[8] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. New York, NY, USA: Springer-Verlag, 2013.

[9] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Roy. Statist. Soc., Ser. B*, vol. 68, pp. 49–67, 2006.

[10] M. Stojnic, F. Parvaresh, and B. Hassibi, "On the reconstruction of block-sparse signals with an optimal number of measurements," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3075–3085, Aug. 2009.

[11] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.

[12] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.

[13] Y. C. Eldar, P. Kuppinger, and H. Blcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.

[14] J. Huang and T. Zhang, "The benefits of group sparsity," *Ann. Statist.*, vol. 38, no. 4, pp. 1978–2004, 2010.

[15] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with application to wavelet decomposition," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, 1993, pp. 40–44.

[16] E. Elhamifar and R. Vidal, "Robust classification using structured sparse representation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2011, pp. 1873–1879.

[17] E. Elhamifar and R. Vidal, "Block-sparse recovery via convex optimization," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4094–4107, Aug. 2012.

[18] W. Chen and Y. Li, "The high order block RIP condition for signal recovery," 2016, arXiv: 1610.06294.

[19] J. Huang, J. Wang, W. Wang, and F. Zhang, "A sharp sufficient condition of block signal recovery via ℓ_2/ℓ_1-minimization," 2017, arXiv:1710.11125.

[20] K. Lounici, M. Pontil, S. van de Geer, and A. B. Tsybakov, "Oracle inequalities and optimal inference under group sparsity," *Ann. Statist.*, vol. 39, no. 4, pp. 2164–2204, Aug. 2011.

[21] G. Obozinski, M. J. Wainwright, and M. I. Jordan, "Support union recovery in high-dimensional multivariate regression," *Ann. Statist.*, vol. 39, no. 1, pp. 1–47, Feb. 2011.

[22] M. K. Elyaderani, S. Jain, J. Druce, S. Gonella, and J. Haupt, "Group-level support recovery guarantees for group LASSO estimator," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2017, pp. 4366–4370.

[23] N. Rao, B. Recht, and R. Nowak, "Universal measurement bounds for structured sparse signal recovery," in *Proc. 15th Int. Conf. Artif. Intell. Statist.*, 2012, pp. 942–950.

[24] Z. Yang, C. Zhang, and L. Xie, "Robustly stable signal recovery in compressed sensing with structured matrix perturbation," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4658–4671, Sep. 2012.

[25] Z. Tan, P. Yang, and A. Nehorai, "Joint sparse recovery method for compressed sensing with structured dictionary mismatches," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 4997–5008, Oct. 2014.

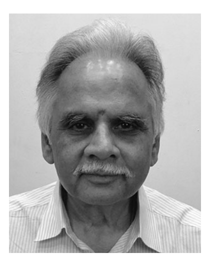
[26] W. U. Bajwa, M. F. Duarte, and R. Calderbank, "Conditioning of random block subdictionaries with applications to block-sparse recovery and regression," *IEEE Trans. Inf. Theory*, vol. 61, no. 7, pp. 4060–4079, Jul. 2015.

[27] B. Vinzamuri, K. K. Padthe, and C. K. Reddy, "Feature grouping using weighted ℓ_1 norm for high-dimensional data," in *Proc. IEEE Conf. Data Mining*, 2016, pp. 1233–1238.

[28] J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of Convex Analysis*. Berlin, Germany: Springer-Verlag, 2001.

[29] M. Lotfi and M. Vidyasagar, "A fast noniterative algorithm for compressive sensing using binary measurement matrices," *IEEE Trans. Signal Process.*, vol. 67, no. 15, pp. 4079–4089, Aug. 2018.

Shashank Ranjan was born in Allahabad, India, in 1991. He received the B.Tech. degree in electronics and instrumentation engineering from the Uttar Pradesh Technical University, Lucknow, India, in 2013 and the M.Tech. degree in systems and control from the Indian Institute of Technology, Hyderabad, Hyderabad, India, in 2016, where he is currently working toward the Ph.D. degree in systems and control.



Mathukumalli Vidyasagar (LF'12) was born in Guntur, India, on September 29, 1947. He received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of Wisconsin in Madison, Madison, WI, USA, in 1965, 1967, and 1969, respectively. Between 1969 and 1989, he was a Professor of electrical engineering with the Marquette University, Concordia University, and the University of Waterloo. In 1989, he returned to India as the Director of the newly created Centre for Artificial Intelligence and Robotics, Bangalore, India. In 2000, he moved to the Indian private sector, and joined India's largest software company, Tata Consultancy Services, as an Executive Vice President, and was located in Hyderabad. In 2009, he retired from TCS and joined the Erik Jonsson School of Engineering and Computer Science, University of Texas at Dallas, as a Cecil & Ida Green Chair in Systems Biology Science. He retired from UT Dallas at the end of 2017, and joined the Indian Institute of Technology Hyderabad, Hyderabad, India, where he is a Science and Engineering Research Board Distinguished Fellow and is a Distinguished Professor. His current research interests include the areas of machine learning, and compressed sensing. On the applications front, he is interested in applying ideas from machine learning to problems in computational biology with emphasis on cancer.

Dr. Vidyasagar has received a number of awards in recognition of his research contributions, including Fellowship in The Royal Society, the world's oldest scientific academy in continuous existence, the IEEE Control Systems (Technical Field) Award, the Rufus Oldenburger Medal of ASME, the John R. Ragazzini Education Award from AACC, and others. He is the author of 12 books and about 150 papers in peer-reviewed journals.