Analyzing Weighted ℓ_1 Minimization for Sparse Recovery with Nonuniform Sparse Models

M. Amin Khajehnejad[†], Weiyu Xu[†], A. Salman Avestimehr[*] and Babak Hassibi[†]

[†]California Institute of Technology, Pasadena CA 91125

[*]Cornell University, Ithaca NY 14853 [*†]

Abstract

In this paper we introduce a nonuniform sparsity model and analyze the performance of an optimized weighted ℓ_1 minimization over that sparsity model. In particular, we focus on a model where the entries of the unknown vector fall into two sets, with entries of each set having a specific probability of being nonzero. We propose a weighted ℓ_1 minimization recovery algorithm and analyze its performance using a Grassmann angle approach. We compute explicitly the relationship between the system parameters-the weights, the number of measurements, the size of the two sets, the probabilities of being nonzero- so that when i.i.d. random Gaussian measurement matrices are used, the weighted ℓ_1 minimization recovers a randomly selected signal drawn from the considered sparsity model with overwhelming probability as the problem dimension increases. This allows us to compute the optimal weights. We demonstrate through rigorous analysis and simulations that for the case when the support of the signal can be divided into two different subclasses with unequal sparsity fractions, the weighted ℓ_1 minimization outperforms the regular ℓ_1 minimization substantially. We also generalize our results to signal vectors with an arbitrary number of subclasses for sparsity.

1 Introduction

Compressed sensing is an emerging technique of joint sampling and compression that has been recently proposed as an alternative to Nyquist sampling (followed by compression) for scenarios where measurements can be costly [29]. The whole premise is that sparse signals (signals with many zero or negligible elements over a known basis) can be recovered with far fewer measurements than the ambient dimension of the signal itself. In fact, the major breakthrough in this area has been the demonstration that ℓ_1

[*]The results of this paper were presented in part at the International Symposium on Information Theory, ISIT 2009

[†]This work was supported in part by the National Science Foundation under grants CCF-0729203, CNS-0932428 and CCF-1018927, by the Office of Naval Research under the MURI grant N00014-08-1-0747, and by Caltech's Lee Center for Advanced Networking.

minimization can efficiently recover a sufficiently sparse vector from a system of underdetermined linear equations [2]. ℓ_1 minimization is usually posed as the convex relaxation of ℓ_0 minimization which solves for the sparsest solution of a system of linear equation and is NP hard.

The conventional approach to compressed sensing assumes no prior information on the unknown signal other than the fact that it is sufficiently sparse over a particular basis. In many applications, however, additional prior information is available. In fact, in many cases the signal recovery problem that compressed sensing addresses is a detection or estimation problem in some statistical setting. Some recent work along these lines can be found in [8], which considers compressed detection and estimation, [9], which studies Bayesian compressed sensing, and [10] which introduces model-based compressed sensing allowing for model-based recovery algorithms. In a more general setting, compressed sensing may be the inner loop of a larger estimation problem that feeds prior information on the sparse signal (e.g., its sparsity pattern) to the compressed sensing algorithm [12, 13].

In this paper we will consider a particular model for the sparse signal where the entries of the unknown vector fall into a number u of classes, with each class having a specific fraction of nonzero entries. The standard compressed sensing model is therefore a special case where there is only one class. As mentioned above, there are many situations where such prior information may be available, such as in natural images, medical imaging, or in DNA microarrays. In the DNA microarrays applications for instance, signals are often *block sparse*, i.e., the signal is more likely to be nonzero in certain blocks rather than in others [11]. While it is possible (albeit cumbersome) to study this model in full generality, in this paper we will focus on the case where the entries of the unknown signal fall into a fixed number u of categories; in the ith set K_i with cardinality n_i, the fraction of nonzero entries is p_i. This model is rich enough to capture many of the salient features regarding prior information. We refer to the signals generated based on this model as *nonuniform sparse* signals.

A signal generated based on this model could resemble the vector representation of a natural image in the domain of some linear transform (e.g. Discrete Fourier Transform, Discrete Cosine Transform, Discrete Wavelet Transform, ...) or the spatial representation of some biomedical image, e.g., a brain fMRI image (see [5] for an application of model based compressed sensing and modified sparsity models for MRI images). Although a brain fMRI image is not necessarily sparse, the subtraction of the brain image at any moment during an experiment from an initial background image of inactive brain mode is indeed a sparse signal which, demonstrates the additional brain activity during the specific course of experiment. Moreover, depending on the assigned task, the experimenter might have some prior information. For example it might be known that some regions of the brain are more likely to be entangled with the decision making process than the others. This can be captured in the above *nonuniform sparse* model by considering a higher value p_i for the more active region. Similarly, this model is applicable to other problems like network monitoring (see [20] for an application of compressed sensing and nonlinear estimation in compressed network monitoring), DNA microarrays [22, 23, 24], astronomy, satellite imaging and many more practical examples. In particular, we will give one example in satellite imaging with details, where such a model is applicable for real data. In general, achieving such a probabilistic prior in practice requires comprehensive knowledge of the model from which the sparse signal is generated, or detailed analysis of some post processing. We will elaborate on the latter later in this paper.

In this paper we first analyze this model for the case where there are $u \geq 2$ categories of entries, and demonstrate through rigorous analysis and simulations that the recovery performance can be significantly boosted by exploiting the additional information. We find a closed form expression for the recovery threshold for $u = 2$. We also generalize the results to the case of $u > 2$. A further interesting question to be addressed in future work would be to characterize the gain in recovery percentage as a function of the number of distinguishable classes u. It is worth mentioning that a somewhat similar model for prior information has been considered in [6]. There, it has been assumed that part of the support is completely known a priori or due to previous processing. A modification of the regular ℓ_1 minimization based on the given information is proven to achieve significantly better recovery guarantees. A variation of this algorithm for noisy recovery under partially known support is also considered in [7]. As will be discussed, this model can be cast as a special case of the nonuniform sparse model, where the sparsity fraction is equal to unity in one of the classes . Therefore, using the generalized tools of this work, we can explicitly find the recovery thresholds for the method proposed in [6]. This is in contrast to the recovery guarantees of [6] which are given in terms of the restricted isometry property (RIP).

The contributions of the paper are the following. We propose a weighted ℓ_1 minimization approach for sparse recovery where the ℓ_1 norms of different classes (K_i's) are assigned different weights w_{K_i} ($1 \leq i \leq u$). Intuitively, one would want to give a larger weight to the entries with a higher chance of being zero and thus further force them to be zero.[1] The second contribution is that we *explicitly* compute the relationship between p_i, w_{K_i},$\frac{n_i}{n}$, $1 \leq i \leq u$ and the number of measurements so that the unknown signal can be recovered with overwhelming probability as $n \to \infty$ (the so-called weak and strong thresholds) for measurement matrices drawn from an i.i.d. Gaussian ensemble. The analysis uses the high-dimensional geometry techniques first introduced by Donoho and Tanner [1, 3] (e.g., Grassmann angles) to obtain sharp thresholds for compressed sensing. However, rather than the *neighborliness* condition used in [1, 3], we find it more convenient to use the null space characterization of Xu and Hassibi [4, 19]. The resulting Grassmannian manifold approach is a general framework for incorporating additional factors into compressed sensing: in [4] it was used to incorporate approximately sparse signals; here it is used to incorporate prior information and weighted ℓ_1 optimization. The techniques for computing the probability decay exponents are adapted from the works of [1, 3]. Our analytic results allow us to precisely compute the optimal weights for any p_i,n_i, $1 \leq i \leq u$. We also provide certain robustness conditions for the recovery scheme for compressible signals or under model mismatch. We present simulation results to show the advantages of the weighted method over standard ℓ_1 minimization. Furthermore, the results of this paper for the case of two classes ($u = 2$) builds a rigid framework for analyzing certain classes of

[1] A somewhat related method that uses weighted ℓ_1 optimization is by Candès et al. [12]. The main difference is that there is no prior information. At each step, the ℓ_1 optimization is re-weighted using the estimate of the signal obtained in the last minimization step.

re-weighted ℓ_1 minimization algorithms. In a re-weighted ℓ_1 minimization algorithm, the post processing information from the estimate of the signal at each step can be viewed as additional prior information about the signal, and can be incorporated into the next step as appropriate weights. In a further work we have been able to analytically prove the threshold improvement in a reweighted ℓ_1 minimization using this framework [21]. It is worth mentioning that we have prepared a software package based on the results of this paper for threshold computation using weighted ℓ_1 minimization, and it is available in [27].

The paper is organized as follows. In the next section we briefly describe the notations that we use throughout the paper. In Section 3 we describe the model and state the principal assumptions of nonuniform sparsity that we are interested in. We also sketch the objectives that we are shooting for and, clarify what we mean by *recovery improvement* in the weighted ℓ_1 case. In Section 4, we skim through our critical theorems and try to present the big picture of the main results. Section 5 is dedicated to the concrete derivation of these results. In Section 6, we briefly introduce the reweighted ℓ_1 minimization algorithm, and provide some insights in how the derivations of this work can be used to analyze the improved recovery thresholds. In Section 7 some simulation results are presented and are compared to the analytical bounds of the previous sections. The paper ends with a conclusion and discussion of future work in Section 8.

2 Basic Definitions and Notations

Throughout the paper, vectors are denoted by small boldface letters $\mathbf{x}, \mathbf{w}, \mathbf{z}, \dots$, scalars are shown by small regular letters $a, b, \alpha, \dots$, and matrices are denoted by bold capital letters($\mathbf{A}, \mathbf{I}, \dots$). For referring to geometrical objects and subspaces, we use Calligraphic notation, e.g. $\mathcal{Z}, \mathcal{F}, \mathcal{G}, \mathcal{P}, \mathcal{C}, \dots$. This includes the notations that we use to indicate the faces of a high dimensional polytope, or the polytope itself. Sets and random variables are denoted by regular capital letters($K, S, \dots$). The normal distribution with mean μ and variance σ^2 is denoted by $\mathcal{N}(\mu, \sigma^2)$. For functions we use both little and capital letters and it should be generally clear from the context. We use the phrases RHS and LHS as abbreviations for Right Hand Side and Left Hand Side respectively throughout the paper.

Definition 1. *A random variable Y is said to have a Half Normal distribution $HN(0, \sigma^2)$ if $Y = |X|$ where X is a zero mean normal variable $X \sim \mathcal{N}(0, \sigma^2)$.*

3 Problem Description

We first define the signal model. For completeness, we present a general definition.

Definition 2. *Let $\mathcal{K} = \{K_1, K_2, ..., K_u\}$ be a partition of $\{1, 2, \cdots, n\}$, i.e. ($K_i \cap K_j = \emptyset$ for $i \neq j$, and $\bigcup_{i=1}^{u} K_i = \{1, 2, ..., n\}$), and $P = \{p_1, p_2, \cdots, p_u\}$ be a set of positive numbers in $[0, 1]$. A $n \times 1$ vector $\mathbf{x} = (x_1, x_2, \cdots, x_n)^T$ is said to be a **random nonuniformly sparse** vector with sparsity fraction p_i over the set K_i for $1 \leq i \leq u$, if $\mathbf{x}$ is generated from the following random procedure:*

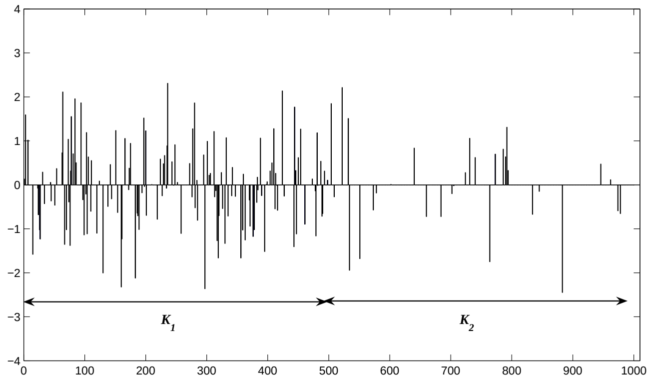


Figure 1: Illustration of a nonuniformly sparse signal.

- *Over each set K_i, $1 \leq i \leq u$, the set of nonzero entries of* $\mathbf{x}$ *is a random subset of size $p_i|K_i|$. In other words, a fraction p_i of the entries are nonzero in K_i. p_i is called the sparsity fraction over K_i. The values of the nonzero entries of* $\mathbf{x}$ *can arbitrarily be selected from any distribution. We can choose $\mathcal{N}(0,1)$ for simplicity.*

In Figure 1, a sample nonuniformly sparse signal with Gaussian distribution for nonzero entries is plotted. The number of sets is considered to be $u = 2$ and both classes have the same size $\frac{n}{2}$, with $n = 1000$. The sparsity fraction for the first class K_1 is $p_1 = 0.3$, and for the second class K_2 is $p_2 = 0.05$. In fact, the signal is much sparser in the second half than it is in the first half. The advantageous feature of this model is that all the resulting computations are independent of the actual distribution on the amplitude of the nonzero entries. However, as expected, it is not independent of the properties of the measurement matrix. We assume that the measurement matrix $\mathbf{A}$ is a $m \times n$ matrix with i.i.d. standard Gaussian distributed $\mathcal{N}(0,1)$ entries, with $\frac{m}{n} = \delta < 1$. The measurement vector is denoted by $\mathbf{y}$ and obeys the following:

$$\mathbf{y} = \mathbf{A}\mathbf{x}. \tag{1}$$

As mentioned in Section 1, ℓ_1 minimization can recover a randomly selected vector $\mathbf{x}$ with $k = \mu n$ nonzero entries with high probability, provided μ is less than a known function of δ. ℓ_1 minimization has the following form:

$$\min_{\mathbf{A}\mathbf{x}=\mathbf{y}} \|\mathbf{x}\|_1. \tag{2}$$

The reference [1] provides an explicit relationship between μ and the minimum δ that guarantees success of ℓ_1 minimization recovery in the case of Gaussian measurements and provides the corresponding numerical curve. The optimization in (2) is a linear program and can be solved polynomially fast ($O(n^3)$). However, it fails to encapsulate additional prior information of the signal nature, might there be any such information available. One can simply think of modifying (2) to a weighted ℓ_1 minimization as follows:

$$\min_{\mathbf{A}\mathbf{x}=\mathbf{y}} \|\mathbf{x}\|_{\mathbf{w},1} = \min_{\mathbf{A}\mathbf{x}=\mathbf{y}} \sum_{i=1}^{n} w_i |x_i| \tag{3}$$

The index, $\mathbf{w}$, on the norm is an indication of the $n \times 1$ positive weight vector. Now the questions are i) what is the optimal set of weights for a certain set of available prior information?, and ii) can one improve the recovery threshold using the weighted ℓ_1 minimization of (3) by choosing a set of optimal weights? We have to be more clear with our objective at this point and clarify what we mean by improving the recovery threshold. Generally speaking, if a recovery method can reconstruct all signals of a certain model with certainty, then that method is said to be ***strongly successful*** on that signal model. If we have a class of models that can be identified with a parameter θ, and if for all models corresponding to $\theta < \theta_0$ a recovery scheme is strongly successful, then the threshold θ_0 is called a ***strong recovery threshold*** for the parameter θ. For example, for fixed $\frac{m}{n}$, if $k < n$ is sufficiently small, then ℓ_1 minimization can provably recover all k-sparse signals, provided that appropriate linear measurements have been made from the signal. The maximum such k is called the strong recovery threshold of the sparsity for the success of ℓ_1 minimization. Likewise, for a fixed ratio $\mu = \frac{k}{n}$, the minimum ratio of measurements to ambient dimension $\frac{m}{n}$ for which, ℓ_1 minimization always recovers k-sparse signals from the given m linear measurements is called the strong recovery threshold for the number of measurements for ℓ_1 minimization. In contrast, one can also look into the ***weak recovery*** threshold, defined as the threshold below which, with very high probability a random vector generated from the model is recoverable. For the nonuniformly sparse model, the quantity of interest is the overall sparsity fraction of the model defined as $(\frac{\sum_{i=1}^{u} p_i n_i}{n})$. The question we ask is whether by adjusting w_i's according to p_i's one can extend the strong or weak recovery threshold for sparsity fraction to a value above the known threshold of ℓ_1 minimization. Equivalently, for given classes $K_1, \cdots, K_u$ and sparsity fractions p_i's, how much can the strong or weak threshold be improved for the minimum number of required measurements, as apposed to the case of uniform sparsity with the same overall sparsity fraction.

4 Summary of Main Results

We state the two problems more formally using the notion of recovery thresholds that we defined in the previous section. We only consider the case of $u = 2$.

- **Problem 1** Consider the random nonuniformly sparse model with two classes K_1, K_2 of cardinalities $n_1 = \gamma_1 n$ and $n_2 = \gamma_2 n$ respectively, and given sparsity fractions p_1 and p_2. Let $\mathbf{w}$ be a given weight vector. As $n \to \infty$, what is the weak (strong) recovery threshold for $\delta = \frac{m}{n}$ so that a randomly chosen vector (all vectors) $\mathbf{x}_0$ selected from the nonuniformly sparse model is successfully recovered by the weighted ℓ_1 minimization of (3) with high probability?

Upon solving Problem.1, one can exhaustively search for the weight vector $\mathbf{w}$ that results in the minimum recovery threshold for δ. This is what we recognize as the optimum set of weights. So the second problem can be stated as:

- **Problem 2** Consider the random nonuniformly sparse model defined by classes K_1, K_2 of cardinalities n_1 and n_2 respectively, with $\gamma_1 = \frac{n_1}{n}$ and $\gamma_2 = \frac{n_2}{n}$, and given sparsity fractions p_1 and p_2.

What is the optimum weight vector $\mathbf{w}$ in (3) that results in the minimum number of measurements for almost sure recovery of signals generated from the given random nonuniformly sparse model?

We will fully solve these problems in this paper. We first connect the misdetection event to the properties of the measurement matrix. For the non-weighted case, this has been considered in [19] and is known as the null space property. We generalize this result to the case of weighted ℓ_1 minimization, and mention a necessary and sufficient condition for (3) to recover the original signal of interest. The theorem is as follows.

Theorem 4.1. *For all $n \times 1$ vectors $\mathbf{x}^*$ supported on the set $K \subseteq \{1, 2, ..., n\}$, $\mathbf{x}^*$ is the unique solution to the linear program $\min_{\mathbf{Ax}=\mathbf{y}} \sum_{i=1}^{n} w_i|x_i|$ with $\mathbf{y} = \mathbf{Ax}^*$, if and only if for every nonzero vector $\mathbf{z} = (z_1, z_2, \cdots, z_n)^T$ in the null space of $\mathbf{A}$, the following holds: $\sum_{i\in K} w_i|z_i| < \sum_{i\in \overline{K}} w_i|z_i|$.*

This theorem will be proved in Section 5. Note that if the null space condition given in the above theorem is weaker than the null space condition for regular ℓ_1 minimization, then it immediately implies that the weighted ℓ_1 minimization is better than ℓ_1 minimization. To clarify this, consider the following example. suppose that a subset K' of the support set K is known a priori. If we assign weights $w_i = 0$ to $i \in K'$ and $w_i = 1$ to $i \in K$, then whenever the null space condition $\sum_{i\in K} |z_i| < \sum_{i\in\overline{K}} |z_i|$ holds for regular ℓ_1 minimization, the condition $\sum_{i\in K} w_i|z_i| < \sum_{i\in\overline{K}} w_i|z_i|$ also holds (since the left hand size decreases and the right hand size is untouched). Therefore the weighted ℓ_1 scheme is at least as good as ℓ_1 minimization (in the strong recovery sense).

As will be explained in Section 5.1, Theorem 4.1 along with known facts about the null space of random Gaussian matrices, help us interpret the probability of recovery error in terms of a high dimensional geometrical object called the complementary Grassmann angle; namely the probability that a uniformly chosen $(n-m)$-dimensional subspace $\mathcal{Z}$ shifted by a point $\mathbf{x}$ of unity weighted ℓ_1-norm, $\sum_{i=1}^{n} w_i x_i = 1$, intersects the weighted ℓ_1-ball $\mathcal{P}_{\mathbf{w}} = \{\mathbf{y} \in \mathbb{R}^n \mid \sum_{i=1}^{n} w_i|y_i| \le 1\}$ nontrivially at some other point besides $\mathbf{x}$. The shifted subspace is denoted by $\mathcal{Z} + \mathbf{x}$. The fact that we can take for granted, without explicitly proving it, is that due to the identical marginal distribution of the entries of $\mathbf{x}$ in each of the sets K_1 and K_2, the entries of the optimal weight vector take at most two (or in the general case u) distinct values w_{K_1} and w_{K_2} depending on their index. In other words

$$\forall i \in \{1, 2, \cdots, n\} \quad w_i = \begin{cases} w_{K_1} \; if \; i \in K_1 \\ w_{K_2} \; if \; i \in K_2 \end{cases} \tag{4}$$

Leveraging on the existing techniques for computing the complementary Grassmann angle [17, 18], we will be able to state and prove the following theorem along the same lines, which upper bounds the probability that the weighted ℓ_1 minimization does not recover the signal. Please note that in the following theorem, the rigorous mathematical definitions to some of the terms (internal angle and external angle) is not presented, due to the extent of descriptions. They will however be defined rigorously later in the derivations of the main results in Section 5.

Theorem 4.2. *Let K_1 and K_2 be two disjoint subsets of $\{1,2,\cdots,n\}$ such that $|K_1| = n_1, |K_2| = n_2$, and p_1 and p_2 be real numbers in $[0,1]$. Also, let $k_1 = p_1 n_1$, $k_2 = p_2 n_2$, and E be the event that a random nonuniformly sparse vector $\mathbf{x}_0$ (Definition 2) with sparsity fractions p_1 and p_2 over the sets K_1 and K_2 respectively is recovered via the weighted ℓ_1 minimization of (3) with $\mathbf{y} = \mathbf{A}\mathbf{x}_0$. Also, let E^c denote the complement event of E. Then*

$$\mathbb{P}\{E^c\} \leq \sum_{\substack{0 \leq t_1 \leq n_1 - k_1 \\ 0 \leq t_2 \leq n_2 - k_2 \\ t_1 + t_2 > m - k_1 - k_2 + 1}} 2^{t_1+t_2+1}\binom{n_1-k_1}{t_1}\binom{n_2-k_2}{t_2}\beta(k_1,k_2|t_1,t_2)\zeta(t_1+k_1,t_2+k_2) \tag{5}$$

where $\beta(k_1,k_2|t_1,t_2)$ is the internal angle between a (k_1+k_2-1)-dimensional face $\mathcal{F}$ of the weighted ℓ_1-ball $\mathcal{P}_{\mathbf{w}} = \{y \in \mathbb{R}^n | \sum_{i=1}^n w_i|y_i| \leq 1\}$ with k_1 vertices supported on K_1 and k_2 vertices supported on K_2, and another $(k_1+k_2+t_1+t_2-1)$-dimensional face $\mathcal{G}$ that encompasses $\mathcal{F}$ and has t_1+k_1 vertices supported on K_1 and the remaining t_2+k_2 vertices supported on K_2. $\zeta(d_1,d_2)$ is the external angle between a face $\mathcal{G}$ supported on set L with $|L\cap K_1| = d_1$ and $|L \cap K_2| = d_2$ and the weighted ℓ_1-ball $\mathcal{P}_{\mathbf{w}}$. See Section 5.1 for the definitions of integral and external angles.

The proof of this theorem will be given in Section 5.2. We are interested in the regimes that make the above upper bound decay to zero as $n \to \infty$, which requires the cumulative exponent in (5) to be negative. We are able to calculate sharp upper bounds on the exponents of the terms in (5) by using large deviations of sums of normal and half normal variables. More precisely, if we assume that the sum of the terms corresponding to particular indices t_1 and t_2 in (5) is denoted by $F(t_1,t_2)$, and define $\tau_1 = \frac{t_1}{n}$ and $\tau_2 = \frac{t_2}{n}$, then using the angle exponent method from [3, 1], we are able to find and compute an exponent function $\psi_{tot}(\tau_1,\tau_2) = \psi_{com}(\tau_1,\tau_2) - \psi_{int}(\tau_1,\tau_2) - \psi_{ext}(\tau_1,\tau_2)$ so that $\frac{1}{n}\log F(t_1,t_2) \sim \psi_{tot}(\tau_1,\tau_2)$ as $n \to \infty$. The terms $\psi_{com}(\cdot,\cdot)$, $\psi_{int}(\cdot,\cdot)$ and $\psi_{ext}(\cdot,\cdot)$ are contributions to the cumulative exponent ψ_{tot} by the so called combinatorial, internal angle and external angle terms respectively, existing in the upper bound (5). The derivations of these terms will be elaborated in Section 5.2.2. Consequently, we state a key theorem that is the implicit answer to Problem 1.

Theorem 4.3. *Let $\delta = \frac{m}{n}$ be the ratio of the number of measurements to the signal dimension, $\gamma_1 = \frac{n_1}{n}$ and $\gamma_2 = \frac{n_1}{n}$. For fixed values of γ_1, γ_2, p_1, p_2, $\omega = \frac{w_{K_2}}{w_{K_1}}$, define E to be the event that a random nonuniformly sparse vector $\mathbf{x}_0$ (Definition 2) with sparsity fractions p_1 and p_2 over the sets K_1 and K_2 respectively with $|K_1| = \gamma_1 n$ and $|K_2| = \gamma_2 n$ is recovered via the weighted ℓ_1 minimization of (2) with $\mathbf{y} = \mathbf{A}\mathbf{x}_0$. There exists a critical threshold $\delta_c = \delta_c(\gamma_1,\gamma_2,p_1,p_2,\omega)$ such that if $\delta = \frac{m}{n} \geq \delta_c$, then $\mathbb{P}\{E^c\}$ decays exponentially to zero as $n \to \infty$. Furthermore, δ_c is given by*

$$\boxed{\begin{gathered}\delta_c = \min\{\delta \mid \psi_{com}(\tau_1,\tau_2) - \psi_{int}(\tau_1,\tau_2) - \psi_{ext}(\tau_1,\tau_2) < 0 \ \forall\ 0 \leq \tau_1 \leq \gamma_1(1-p_1), \\ 0 \leq \tau_2 \leq \gamma_2(1-p_2), \tau_1+\tau_2 > \delta - \gamma_1 p_1 - \gamma_2 p_2\}\end{gathered}}$$

where ψ_{com}, ψ_{int} and ψ_{ext} are obtained from the following expressions:

Define $g(x) = \frac{2}{\sqrt{\pi}} e^{-x^2}$, $G(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2} dy$ *and let* $\varphi(.)$ *and* $\Phi(.)$ *be the standard Gaussian pdf and cdf functions respectively.*

1. *(Combinatorial exponent)*

$$\psi_{com}(\tau_1, \tau_2) = \left(\gamma_1(1-p_1)H(\frac{\tau_1}{\gamma_1(1-p_1)}) + \gamma_2(1-p_2)H(\frac{\tau_2}{\gamma_2(1-p_2)}) + \tau_1 + \tau_2 \right) \log 2 \tag{6}$$

   where $H(\cdot)$ *is the entropy function defined by* $H(x) = -x\log x - (1-x)\log(1-x)$.

2. *(External angle exponent) Define* $c = (\tau_1 + \gamma_1 p_1) + \omega^2(\tau_2 + \gamma_2 p_2)$, $\alpha_1 = \gamma_1(1-p_1) - \tau_1$ *and* $\alpha_2 = \gamma_2(1-p_2) - \tau_2$. *Let* x_0 *be the unique solution to* x *of the following:*

$$2c - \frac{g(x)\alpha_1}{xG(x)} - \frac{\omega g(\omega x)\alpha_2}{xG(\omega x)} = 0$$

   Then

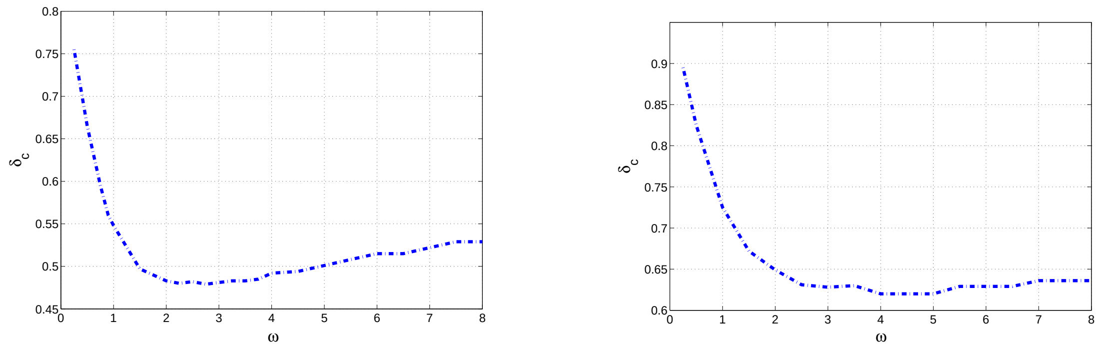
$$\psi_{ext}(\tau_1, \tau_2) = c x_0^2 - \alpha_1 \log G(x_0) - \alpha_2 \log G(\omega x_0) \tag{7}$$

3. *(Internal angle exponent) Let* $b = \frac{\tau_1 + \omega^2 \tau_2}{\tau_1 + \tau_2}$, $\Omega' = \gamma_1 p_1 + \omega^2 \gamma_2 p_2$ *and* $Q(s) = \frac{\tau_1 \varphi(s)}{(\tau_1+\tau_2)\Phi(s)} + \frac{\omega \tau_2 \varphi(\omega s)}{(\tau_1+\tau_2)\Phi(\omega s)}$. *Define the function* $\hat{M}(s) = -\frac{s}{Q(s)}$ *and solve for* s *in* $\hat{M}(s) = \frac{\tau_1+\tau_2}{(\tau_1+\tau_2)b + \Omega'}$. *Let the unique solution be* s^* *and set* $y = s^*(b - \frac{1}{\hat{M}(s^*)})$. *Compute the rate function* $\Lambda^*(y) = sy - \frac{\tau_1}{\tau_1+\tau_2}\Lambda_1(s) - \frac{\tau_2}{\tau_1+\tau_2}\Lambda_1(\omega s)$ *at the point* $s = s^*$, *where* $\Lambda_1(s) = \frac{s^2}{2} + \log(2\Phi(s))$. *The internal angle exponent is then given by:*

$$\psi_{int}(\tau_1, \tau_2) = (\Lambda^*(y) + \frac{\tau_1 + \tau_2}{2\Omega'} y^2 + \log 2)(\tau_1 + \tau_2) \tag{8}$$

Theorem 4.3 is a powerful result, since it allows us to find (numerically) the optimal set of weights for which the fewest possible measurements are needed to recover the signals almost surely. To this end, for fixed values of γ_1, γ_2, p_1 and p_2, one should find the ratio $\frac{w_{K_2}}{w_{K_1}}$ for which the critical threshold $\delta_c(\gamma_1, \gamma_2, p_1, p_2, \frac{w_{K_2}}{w_{K_1}})$ from Theorem 4.3 is minimum. We discuss this by some examples in Section 7. A generalization of Theorem 4.3 for a nonuniform model with an arbitrary number of classes ($u \geq 2$) will be given in Section 5.3.

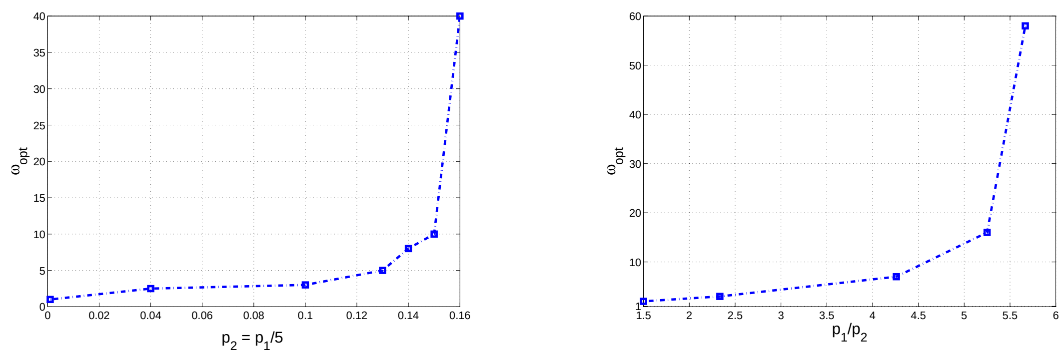
As mentioned earlier, using Theorem 4.3, it is possible to find the optimal ratio $\frac{w_{K_2}}{w_{K_1}}$. It however requires an exhaustive search over the δ_c threshold for all possible values of ω. For $\gamma_1 = \gamma_2 = 0.5$, $p_1 = 0.3$ and $p_2 = 0.05$, we have numerically computed $\delta_c(\gamma_1, \gamma_2, p_1, p_2, \frac{w_{K_2}}{w_{K_1}})$ as a function of $\frac{w_{K_2}}{w_{K_1}}$ and depicted the resulting curve in Figure 2a. This suggests that $\frac{w_{K_2}}{w_{K_1}} \approx 2.5$ is the optimal ratio that one can choose. Later we will confirm this using simulations. The value of δ_c for another choice of p_1, p_2 is shown in Figure 2b. Note that for given class sizes γ_1, γ_2 the optimal value of ω does not solely depend on $\frac{p_2}{P_1}$. To see this, we have obtained the numerical value of the optimal ω for a fixed ratio $\frac{p_1}{p_2} = 5$, for various values of p_2 using exhaustive search and the result of Theorem 4.3. The result is displayed in Figure 3a. As can be seen, although $\frac{p_2}{P_1}$ is fixed, as p_1 approaches 1, the optimal value of weight becomes very large (in fact for the case of $p_1 = 1$, we later prove that the optimal ω is indeed $\omega_{\text{opt}} = \infty$). In Figure 3b, the



(a) $p_1 = 0.4$, $p_2 = 0.05$.

(b) $p_1 = 0.65$, $p_2 = 0.1$.

Figure 2: δ_c as a function of $\omega = \frac{w_{K_2}}{w_{K_1}}$ for $\gamma_1 = \gamma_2 = 0.5$.



(a)

(b)

Figure 3: (a) Optimum value of weight $\omega = \frac{w_{K_2}}{w_{K_1}}$ vs. $p_2 = p_1/5$, $\gamma_1 = \gamma_2 = 0.5$. (b) Optimum value of weight $\omega = \frac{w_{K_2}}{w_{K_1}}$ vs. $\frac{p_1}{p_2}$, for $\gamma_1 = \gamma_2 = 0.5$, $p_1\gamma_1 + p_2\gamma_2 = 0.5$.

optimal weight is plotted as a function of $\frac{p_1}{p_2}$ for a situation where the overall sparsity, i.e. $\gamma_1 p_1 + \gamma_2 p_2$ is constant.

Note that δ_c given in Theorem 4.3 is a weak bound on the ratio $\delta = \frac{m}{n}$. In other words, it determines the minimum number of measurements so that for a random sparse signal from the nonuniform sparse model and a random support set, the recovery is successful with high probability. It is possible to obtain a strong bound for δ, using a union bound on all possible support sets in the model, and all possible sign patterns of the sparse vector. Similarly, a sectional bound can be defined which accounts for all possible support sets but almost all sign patterns. Therefore, the expressions for the strong and sectional thresholds(see [1, 3] for definitions), which we denote by $\delta_c^{(S)}$ and $\delta_c^{(T)}$ are very similar to δ_c in Theorem 4.3, except for a slight modification in the combinatorial exponent term ψ_{com}. This will be elaborated in Section 5.2.2.

It is worthwhile to consider some asymptotic cases of the presented nonuniform model and some of their implications. First of all, when one of the subclasses is empty, e.g. $\gamma_1 = 0$, then the obtained weak and strong thresholds are equal to the corresponding thresholds of ℓ_1 minimization for a sparsity fraction

$p = p_2$. Furthermore, if the sparsity fractions p_1 and p_2 over the two classes are equal, and a unitary weight $\omega = 1$ is used, then the weak threshold δ_c is equal to the threshold of ℓ_1 minimization for a sparsity fraction $p = p_1 = p_2$. In other words:

$$\delta_c(\gamma_1, \gamma_2, p, p, 1) = \delta_c(0, 1, 0, p, 1). \tag{9}$$

This follows immediately from the derivations of the exponents in Theorem 4.3. However, the latter is not necessarily true for the strong threshold. In fact the computation of the strong threshold for regular ℓ_1 minimization involves a union bound over a larger set of possible supports, and therefore the combinatorial exponent becomes larger. Therefore:

$$\delta_c^{(S)}(\gamma_1, \gamma_2, p, p, 1) \leq \delta_c^{(S)}(0, 1, 0, p, 1). \tag{10}$$

A very important asymptotic case is when the unknown signal is fully dense over one of the subclasses, e.g. $p_1 = 1$, which accounts for a *partially known* support. This model is considered in the work of Vaswani et al. [6], with the motivation that in some applications (or due to previous processing steps), part of the support set can be fully identified ($p_1 = 1$), or can be approximated very well which corresponds to $p_1 < 1$. If the dense subclass is K_1 and $K_2 = K_1^c$, then [6] suggests solving the following minimization program:

$$\min_{\mathbf{Ax}=\mathbf{y}} \|\mathbf{x}_{K_2}\|_1. \tag{11}$$

It is possible to find exact thresholds for the above problem using the weighted ℓ_1 minimization machinery presented in this paper. First, note that (11) is the asymptotic solution of the following weighted ℓ_1 minimization, when $\omega \to \infty$,

$$\min_{\mathbf{Ax}=\mathbf{y}} \|\mathbf{x}_{K_1}\|_1 + \omega\|\mathbf{x}_{K_2}\|_1. \tag{12}$$

Therefore the recovery threshold for (11) can be given by $\delta_c(\gamma_1, \gamma_2, 1, p_2, \omega)$ for $\omega \to \infty$. We prove the following theorem about the latter threshold:

Theorem 4.4. *If $\omega \to \infty$, then $\delta_c(\gamma_1, \gamma_2, 1, p_2, \omega) \to \gamma_1 + \gamma_2\delta_c(0, 1, 0, p_2, 1)$.*

The interpretation of the above theorem is that when a subset of entries of size $\gamma_1 n$ are known to be nonzero, the minimum number of measurements that is required for successful recovery with high probability using (11) is equal to the total number of measurements needed if we were allowed to independently measure everything in the first subclass (i.e. the known subset of the support) , plus the number of measurements we needed for recovering the remaining entries using ℓ_1 minimization. The proof of this theorem is given in Appendix A.

A very important factor regarding the performance of any recovery method is its robustness. In other words, it is important to understand how resilient the recovery is in the case of compressible signals or in the presence of noise or model mismatch(i.e. incorrect knowledge of the the sets or sparsity factors). We address this in the following theorem.

Theorem 4.5. *Let K_1 and K_2 be two disjoint subsets of $\{1,2,\cdots,n\}$, with $|K_1| = \gamma_1 n, |K_2| = \gamma_2 n$ and $\gamma_1 + \gamma_2 = 1$. Also suppose that the dimensions of the measurement matrix $\mathbf{A}$ satisfy $\delta = \frac{m}{n} \geq \delta_c^{(S)}(\gamma_1,\gamma_2,p_1,p_2,\omega)$ for positive real numbers p_1 and p_2 in $[0,1]$ and $\omega > 0$. For positive ϵ_1, ϵ_2, assume that L_1 and L_2 are arbitrary subsets of K_1 and K_2 with cardinalities $(1-\epsilon_1)\gamma_1 p_1 n$ and $(1-\epsilon_2)\gamma_2 p_2 n$ respectively. With high probability, for every vector $\mathbf{x}_0$, if $\hat{\mathbf{x}}$ is the solution to the following linear program:*

$$\min_{\mathbf{A}\mathbf{x}=\mathbf{A}\mathbf{x}_0} \|\mathbf{x}_{K_1}\|_1 + \omega\|\mathbf{x}_{K_2}\|_1, \tag{13}$$

then the following holds

$$\|(\mathbf{x}_0 - \hat{\mathbf{x}})_{K_1}\|_1 + \omega\|(\mathbf{x}_0 - \hat{\mathbf{x}})_{K_2}\|_1 \leq C_{\epsilon_1,\epsilon_2}\left(\|(\mathbf{x}_0)_{\overline{L_1}\cap K_1}\|_1 + \omega\|(\mathbf{x}_0)_{\overline{L_2}\cap K_2}\|_1\right), \tag{14}$$

where

$$C_{\epsilon_1,\epsilon_2} = \frac{1+\min(\frac{\epsilon_1 p_1}{1-p_1},\frac{\epsilon_2 p_2}{1-p_2})}{1-\min(\frac{\epsilon_1 p_1}{1-p_1},\frac{\epsilon_2 p_2}{1-p_2})}.$$

The above theorem has the following implications. First, if $\mathbf{x}_0$ is a (compressible) vector, such that its "significant" entries follow a nonuniform sparse model, then the recovery error of the corresponding weighted ℓ_1 minimization can be bounded in terms of the ℓ_1 norm of the "insignificant" part of $\mathbf{x}_0$ (i.e. the part where a negligible fraction of the energy of the signal is located or most entries have significantly small values, compared to the other part that has an overall large norm). Theorem 4.5 can also be interpreted as the robustness of weighted ℓ_1 scheme to the model mismatch. If K_1, K_2, p_1, p_2 are the estimates of an actual nonuniform decomposition for $\mathbf{x}_0$ (based on which the minimum number of required measurements have been estimated), then the recovery error can be relatively small if the model estimation error is slight. Theorem 4.5 will be proved in Section 5.4.

5 Derivation of the main results

In this section we provide detailed proofs to the claims of Section 4. We mention that, due to space limitations, and also due to the similarity of some proof techniques to those of [3, 1, 4], we have skipped proofs of some of the claims or the details of derivation. For completeness, we have provided a complete version of this paper available online [28]. Let $\mathbf{x}_0$ be a random nonuniformly sparse signal with sparsity fractions p_1 and p_2 over the index subsets K_1 and K_2 respectively (Definition 2), and let $|K_1| = n_1$ and $|K_2| = n_2$. Also let K be the support of $\mathbf{x}$. Let E be the event that $\mathbf{x}$ is recovered exactly by (3), and E^c be its complimentary event. In order to bound the conditional error probability $\mathbb{P}\{E^c\}$ we adopt the idea of [19] to interpret the failure recovery event (E^c) in terms of the null space of the measurement matrix $\mathbf{A}$. This is stated in Theorem 4.1, which we prove here.

Proof of Theorem 4.1.

Suppose the mentioned null space condition holds and define $\hat{\mathbf{x}} = argmin_{\mathbf{A}\mathbf{x}=\mathbf{y}}\sum_{i=1}^n w_i|x_i|$. Let $\mathbf{W} = diag(w_1, w_2, \cdots, w_n)$. If $\hat{\mathbf{x}} \neq \mathbf{x}$, then by triangular inequality, we have:

$$
\begin{aligned}
\|\mathbf{W}\hat{\mathbf{x}}\|_1 = \|(\mathbf{W}\hat{\mathbf{x}})_K\|_1 + \|(\mathbf{W}\hat{\mathbf{x}})_{\overline{K}}\|_1 &= \|(\mathbf{W}\mathbf{x}^* + \mathbf{W}\hat{\mathbf{x}} - \mathbf{W}\mathbf{x}^*)_K\|_1 + \|(\mathbf{W}\hat{\mathbf{x}})_{\overline{K}}\|_1, \\
&\geq \|(\mathbf{W}\mathbf{x}^*)_K\|_1 - \|(\mathbf{W}\hat{\mathbf{x}} - \mathbf{W}\mathbf{x}^*)_K\|_1 + \|(\mathbf{W}\hat{\mathbf{x}} - \mathbf{W}\mathbf{x}^*)_{\overline{K}}\|_1, \\
&> \|\mathbf{W}\mathbf{x}^*\|_1,
\end{aligned}
$$

where the last inequality is a result of the fact that $\hat{\mathbf{x}} - \mathbf{x}^*$ is a nonzero vector in the null space of $\mathbf{A}$ and satisfies the mentioned null space condition. However, by assumption if $\hat{\mathbf{x}} \neq \mathbf{x}^*$ then $\|\mathbf{W}\hat{\mathbf{x}}\|_1 \leq \|\mathbf{W}\mathbf{x}^*\|_1$. This is a contradiction, and hence we should have $\hat{\mathbf{x}} = \mathbf{x}^*$. Conversely, suppose there is some vector $\mathbf{z}$ in $\mathcal{N}(\mathbf{A})$ such that $\|(\mathbf{W}\mathbf{z})_K\|_1 \geq \|(\mathbf{W}\mathbf{z})_{\overline{K}}\|_1$. Taking define $\mathbf{x}^* = (\mathbf{z}_K \ 0)^T$ and $\hat{\mathbf{x}} = (0 \ \mathbf{z}_{\overline{K}})^T$ implies that $\mathbf{A}\mathbf{x}^* = \mathbf{A}\hat{\mathbf{x}}$ and $\|\mathbf{W}\mathbf{x}^*\|_1 \geq \|\mathbf{W}\hat{\mathbf{x}}\|_1$. Therefore, $\mathbf{x}^*$ cannot be recovered from the weighted ℓ_1 minimization. ■

From this point on, we follow closely the steps towards calculating the upper bound on the failure probability from [4], but with appropriate modifications. The key to our derivations is the following lemma, the proof of which is skipped and can be found in the online version of this paper [28].

Lemma 5.1. *For a certain subset $K \subseteq \{1, 2, ..., n\}$ with $|K| = k$, the event that the null-space $\mathcal{N}(A)$ satisfies*

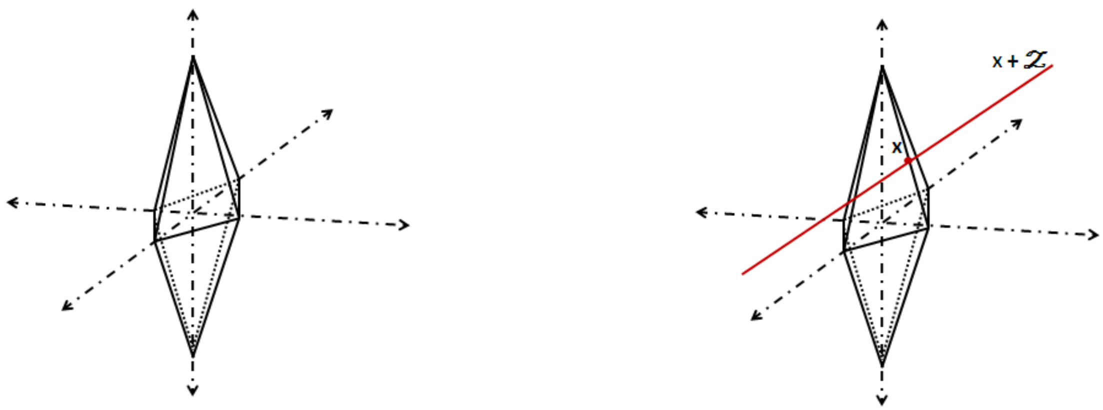
$$\sum_{i \in K} w_i |z_i| \leq \sum_{i \in \overline{K}} w_i |z_i|, \forall \mathbf{z} \in \mathcal{N}(A), \tag{15}$$

is equivalent to the event that for each $\mathbf{x}$ supported on the set K (or a subset of K)

$$\sum_{i \in K} w_i |x_i + z_i| + \sum_{i \in \overline{K}} w_i |z_i| \geq \sum_{i \in K} w_i |x_i|, \forall \mathbf{z} \in \mathcal{N}(A). \tag{16}$$

5.1 Upper Bound on the Failure Probability

Knowing Lemma 5.1, we are now in a position to derive the probability that condition (15) holds for a support set K with $|K| = k$, if we randomly choose an i.i.d. Gaussian matrix $\mathbf{A}$. In the case of a random i.i.d. Gaussian matrix, the distribution of null space of $\mathbf{A}$ is right-rotationally invariant, and sampling from this distribution is equivalent to uniformly sampling a random $(n-m)$-dimensional subspace $\mathcal{Z}$ from the Grassmann manifold $\mathrm{Gr}_{(n-m)}(n)$. The Grassmann manifold $\mathrm{Gr}_{(n-m)}(n)$ is defined as the set of all $(n-m)$-dimensional subspaces of $\mathbb{R}^n$. We need to upper bound the complementary probability $P = \mathbb{P}\{E^c\}$, namely the probability that the (random) support set K of $\mathbf{x}$ (of random sign pattern) fails the null space condition (16). We denote the null space of $\mathbf{A}$ by $\mathcal{Z}$. Because $\mathcal{Z}$ is a linear space, for every vector $\mathbf{z} \in \mathcal{Z}$, $\alpha\mathbf{z}$ is also in $\mathcal{Z}$ for all $\alpha \in \mathbb{R}$. Therefore, if for a $\mathbf{z} \in \mathcal{Z}$ and $\mathbf{x}$ condition (16) fails, by a simple re-scaling of the vectors, we may assume without loss of generality that $\mathbf{x}$ lies on the surface of any convex ball that surrounds the origin. Therefore we restrict our attention to those vectors $\mathbf{x}$ from



(a) (b)

Figure 4: A weighted ℓ_1-ball, $\mathcal{P}_{\mathbf{w}}$, in $\mathbb{R}^3$ (a), and a linear hyperplane $\mathcal{Z}$ passing through a point $\mathbf{x}$ in the interior of a one dimensional face of $\mathcal{P}_{\mathbf{w}}$ (b).

the weighted ℓ_1-sphere:

$$\{\mathbf{x} \in \mathbb{R}^n \mid \sum_{i=1}^{n} w_i|x_i| = 1\}$$

that are only supported on the set K , or a subset of it. Due to the similarity of the above weighted ℓ_1 balls in all orthants, we can write:

$$P = P_{K,-} \tag{17}$$

where $P_{K,-}$ is the probability that for a specific *support set* K , there exist a k-sparse vector $\mathbf{x}$ of a specific *sign pattern* which fails the condition (16). By symmetry, without loss of generality, we assume the signs of the elements of $\mathbf{x}$ to be non-positive. Now we can focus on deriving the probability $P_{K,-}$. Since $\mathbf{x}$ is a non-positive k-sparse vector supported on the set K and can be restricted to the weighted ℓ_1-sphere $\{\mathbf{x} \in \mathbb{R}^n \mid \sum_{i=1}^{n} w_i|x_i| = 1\}$, $\mathbf{x}$ is also on a $(k-1)$-dimensional face, denoted by $\mathcal{F}$, of the weighted ℓ_1-ball $\mathcal{P}_{\mathbf{w}}$:

$$\mathcal{P}_{\mathbf{w}} = \{\mathbf{y} \in \mathbb{R}^n \mid \sum_{i=1}^{n} w_i|y_i| \leq 1\} \tag{18}$$

The subscript $\mathbf{w}$ in $\mathcal{P}_{\mathbf{w}}$ is an indication of the weight vector $\mathbf{w} = (w_1, w_2, \cdots, w_n)^T$. Figure 4a shows $\mathcal{P}_{\mathbf{w}}$ in $\mathbb{R}^3$ for some nontrivial weight vector $\mathbf{w}$. Now the probability $P_{K,-}$ is equal to the probability that there exists an $\mathbf{x} \in \mathcal{F}$, and there exists a $\mathbf{z} \in \mathcal{Z}$ ($\mathbf{z} \neq 0$) such that

$$\sum_{i \in K} w_i|x_i + z_i| + \sum_{i \in \bar{K}} w_i|z_i| \leq \sum_{i \in K} w_i|x_i| = 1. \tag{19}$$

We start by studying the case for a specific point $\mathbf{x} \in \mathcal{F}$ and, without loss of generality, we assume $\mathbf{x}$ is in the relative interior of this $(k-1)$-dimensional face $\mathcal{F}$. For this particular $\mathbf{x}$ on $\mathcal{F}$, the probability, denoted by $P'_{\mathbf{x}}$, that there exists a $\mathbf{z} \in \mathcal{Z}$ ($\mathbf{z} \neq 0$) such that

$$\sum_{i \in K} w_i |x_i + z_i| + \sum_{i \in \bar{K}} w_i |z_i| \leq \sum_{i \in K} w_i |x_i| = 1. \tag{20}$$

is essentially the probability that a uniformly chosen $(n-m)$-dimensional subspace $\mathcal{Z}$ shifted by the point $\mathbf{x}$, namely $(\mathcal{Z}+\mathbf{x})$, intersects the weighted ℓ_1-ball $\mathcal{P}_{\mathbf{w}}$ *non-trivially*, namely, at some other point besides $\mathbf{x}$ (Figure 4b). From the fact that $\mathcal{Z}$ is a *linear* subspace, the event that $(\mathcal{Z}+\mathbf{x})$ intersects $\mathcal{P}_{\mathbf{w}}$ is equivalent to the event that $\mathcal{Z}$ intersects nontrivially with the cone $\mathcal{C}_{\mathbf{w}}(\mathbf{x})$ obtained by observing the weighted ℓ_1-ball $\mathcal{P}_{\mathbf{w}}$ from the point $\mathbf{x}$. (Namely, $\mathcal{C}_{\mathbf{w}}(\mathbf{x})$ is conic hull of the point set $(\mathcal{P}_{\mathbf{w}} - \mathbf{x})$ and of course $\mathcal{C}_{\mathbf{w}}(\mathbf{x})$ has the origin of the coordinate system as its apex.) However, as noticed in the geometry for convex polytopes [14, 15], the cones $\mathcal{C}_{\mathbf{w}}(\mathbf{x})$ are identical for any $\mathbf{x}$ lying in the relative interior of the face $\mathcal{F}$. This means that the probability $P_{K,-}$ is equal to $P'_{\mathbf{x}}$, regardless of the fact that $\mathbf{x}$ is only a single point in the relative interior of the face $\mathcal{F}$. There are some singularities here because $\mathbf{x} \in \mathcal{F}$ may not be in the relative interior of $\mathcal{F}$, but it turns out that the $\mathcal{C}_{\mathbf{w}}(\mathbf{x})$ in this case is only a subset of the cone we get when $\mathbf{x}$ is in the relative interior of $\mathcal{F}$. So we do not lose anything if we restrict $\mathbf{x}$ to be in the relative interior of the face $\mathcal{F}$, namely we have

$$P_{K,-} = P'_{\mathbf{x}}.$$

Now we only need to determine $P'_{\mathbf{x}}$. From its definition, $P'_{\mathbf{x}}$ is exactly the ***complementary Grassmann angle*** [14] for the face $\mathcal{F}$ with respect to the polytope $\mathcal{P}_{\mathbf{w}}$ under the Grassmann manifold $\mathrm{Gr}_{(n-m)}(n)$: a uniformly distributed $(n-m)$-dimensional subspace $\mathcal{Z}$ from the Grassmannian manifold $\mathrm{Gr}_{(n-m)}(n)$ intersecting non-trivially with the cone $\mathcal{C}_{\mathbf{w}}(\mathbf{x})$ formed by observing the weighted ℓ_1-ball $\mathcal{P}_{\mathbf{w}}$ from the relative interior point $\mathbf{x} \in \mathcal{F}$.

Building on the works by L.A. Santalö [17] and P. McMullen [18] in high dimensional geometry and convex polytopes, the complementary Grassmann angle for the $(k-1)$-dimensional face $\mathcal{F}$ can be explicitly expressed as the sum of products of internal angles and external angles [16, 18]:

$$2 \times \sum_{s \geq 0} \sum_{\mathcal{G} \in \Im_{m+1+2s}(\mathcal{P}_{\mathbf{w}})} \beta(\mathcal{F}, \mathcal{G}) \zeta(\mathcal{G}, \mathcal{P}_{\mathbf{w}}), \tag{21}$$

where s is any nonnegative integer, $\mathcal{G}$ is any $(m+1+2s)$-dimensional face of the $\mathcal{P}_{\mathbf{w}}$ ($\Im_{m+1+2s}(\mathcal{P}_{\mathbf{w}})$ is the set of all such faces), $\beta(\cdot,\cdot)$ stands for the internal angle and $\zeta(\cdot,\cdot)$ stands for the external angle, and are defined as follows [15, 18]:

- An internal angle $\beta(\mathcal{F}_1, \mathcal{F}_2)$ is the fraction of the hypersphere S covered by the cone obtained by observing the face $\mathcal{F}_2$ from the face $\mathcal{F}_1$. [2] The internal angle $\beta(\mathcal{F}_1, \mathcal{F}_2)$ is defined to be zero when $\mathcal{F}_1 \nsubseteq \mathcal{F}_2$ and is defined to be one if $\mathcal{F}_1 = \mathcal{F}_2$.

[2] Note the dimension of the hypersphere S here matches the dimension of the corresponding cone discussed. Also, the center of the hypersphere is the apex of the corresponding cone. All these defaults also apply to the definition of the external angles.

- An external angle $\zeta(\mathcal{F}_3, \mathcal{F}_4)$ is the fraction of the hypersphere S covered by the cone of outward normals to the hyperplanes supporting the face $\mathcal{F}_4$ at the face $\mathcal{F}_3$. The external angle $\zeta(\mathcal{F}_3, \mathcal{F}_4)$ is defined to be zero when $\mathcal{F}_3 \nsubseteq \mathcal{F}_4$ and is defined to be one if $\mathcal{F}_3 = \mathcal{F}_4$.

The formula (21) comes from the [18] based on the nonlinear angle sum of relations for polyhedral cones. In order to calculate the internal and external angles, it is important to use the symmetrical properties of the weighted cross-polytope $\mathcal{P}_{\mathbf{w}}$. First of all, $\mathcal{P}_{\mathbf{w}}$ is nothing but the convex hull of the following set of $2n$ vertices in $\mathbb{R}^n$

$$\mathcal{P}_{\mathbf{w}} = conv\{\pm\frac{\mathbf{e}_i}{w_i} \mid 1 \leq i \leq n\} \tag{22}$$

where $\mathbf{e}_i$ $1 \leq i \leq n$ is the standard unit vector in $\mathbb{R}^n$ with the ith entry equal to 1. Every $(k-1)$-dimensional face $\mathcal{F}$ of $\mathcal{P}_{\mathbf{w}}$ is simply the convex hull of k of the linearly independent vertices of $\mathcal{P}_{\mathbf{w}}$. In that case we say that $\mathcal{F}$ *is supported* on the index set K of the k indices corresponding to the nonzero coordinates of the vertices of $\mathcal{F}$ in $\mathbb{R}^n$. More precisely, if $\mathcal{F} = conv\{j_1\frac{\mathbf{e}_{i_1}}{w_{i_1}}, j_2\frac{\mathbf{e}_{i_2}}{w_{i_2}}, \cdots, j_k\frac{\mathbf{e}_{i_k}}{w_{i_k}}\}$ with $j_i \in \{-1,+1\}$ $\forall 1 \leq i \leq k$, then $\mathcal{F}$ is said to be supported on the set $K = \{i_1, i_2, \cdots, i_k\}$.

5.2 Special Case of $u = 2$

The derivations of the previous section were for a general weight vector $\mathbf{w}$. We now restrict ourselves to the case of two classes, i.e. $u = 2$, namely K_1 and K_2 with $|K_1| = n_1$ and $|K_2| = n_2$. For this case, we may assume that $w_i's$ have the following particular form

$$\forall i \in \{1,2,\cdots,n\} \quad w_i = \begin{cases} w_{K_1} \; if \; i \in K_1 \\ w_{K_2} \; if \; i \in K_2 \end{cases} \tag{23}$$

proof of Theorem 4.2. The choice of $\mathbf{w}$ as in (23) results in $\mathcal{P}_{\mathbf{w}}$ having two classes of geometrically identical vertices, and many of faces of $\mathcal{P}_{\mathbf{w}}$ being isomorphic. In fact, two faces $\mathcal{F}$ and $\mathcal{F}'$ of $\mathcal{P}_{\mathbf{w}}$ that are respectively supported on the sets K and K' are geometrically isomorphic [3] if $|K \cap K_1| = |K' \cap K_1|$ and $|K \cap K_2| = |K' \cap K_2|$[4]. In other words the only thing that distinguishes the morphology of the faces of $\mathcal{P}_{\mathbf{w}}$ is the proportion of their support sets that is located in K_1 or K_2. Therefore for two faces $\mathcal{F}$ and $\mathcal{G}$ with $\mathcal{F}$ supported on K and $\mathcal{G}$ supported on L $(K \subseteq L)$, $\beta(\mathcal{F}, \mathcal{G})$ is only a function of the parameters $k_1 = |K \cap K_1|$, $k_2 = |K \cap K_2|$, $k_1 + t_1 = |L \cap K_1|$ and $k_2 + t_1 = |K \cap K_2|$. So, instead of $\beta(\mathcal{F}, \mathcal{G})$ we may write $\beta(k_1, k_2|t_1, t_2)$ to indicate the internal angle between a $(k_1 + k_2 - 1)$-dimensional face $\mathcal{F}$ of $\mathcal{P}_{\mathbf{w}}$ with k_1 vertices supported on K_1 and k_2 vertices supported on K_2, and a $(k_1 + k_2 + t_1 + t_2 - 1)$-dimensional face $\mathcal{G}$ that encompasses $\mathcal{F}$ and has $t_1 + k_1$ vertices supported on K_1 and the remaining $t_2 + k_2$ vertices supported on K_2. Similarly instead of $\zeta(\mathcal{G}, \mathcal{P}_{\mathbf{w}})$ we write $\zeta(t_1 + k_1, t_2 + k_2)$ to denote the external angle

[3] This means that there exists a rotation matrix $\Theta \in \mathbb{R}^{n \times n}$ which is unitary i.e. $\Theta^T\Theta = I$, and maps $\mathcal{F}$ isometrically to $\mathcal{F}'$ i.e. $\mathcal{F}' = \Theta\mathcal{F}$.

[4] Remember that K_1 and K_2 are the same sets as defined in the model description of Section 3.

between a face $\mathcal{G}$ supported on set L with $|L \cap K_1| = d_1$ and $|L \cap K_2| = d_2$, and the weighted ℓ_1-ball $\mathcal{P}_{\mathbf{w}}$. Using this notation and recalling the formula (21) we can write

$$\begin{aligned} P_{K,-} &= 2\sum_{s\geq 0}\sum_{G\in\Im_{m+1+2s}(\mathcal{P}_{\mathbf{w}})} \beta(\mathcal{F},\mathcal{G})\zeta(\mathcal{G},\mathcal{P}_{\mathbf{w}}) \\ &= \sum_{\substack{0\leq t_1\leq n_1-k_1 \\ 0\leq t_2\leq n_2-k_2 \\ t_1+t_2+k_1+k_2-m\in\mathbb{E}}} 2^{t_1+t_2+1}\binom{n_1-k_1}{t_1}\binom{n_2-k_2}{t_2}\beta(k_1,k_2|t_1,t_2)\zeta(t_1+k_1,t_2+k_2), \end{aligned} \tag{24}$$

where $\mathbb{E}$ is the set of positive even integers, and in (24) we have used the fact that the number of faces $\mathcal{G}$ of $\mathcal{P}_{\mathbf{w}}$ of dimension $k_1+k_2+t_1+t_2-1$ that encompass $\mathcal{F}$ and have k_1+t_1 vertices supported on K_1 and its remaining k_2+t_2 are vertices supported on K_2 is $2^{t_1+t_2}\binom{n_1-k_1}{t_1}\binom{n_2-k_2}{t_2}$. In fact $\mathcal{G}$ has $k_1+k_2+t_1+t_2$ vertices including the k_1+k_2 vertices of $\mathcal{F}$. The remaining t_1+t_2 vertices can each be independently in the positive or negative orthant, therefore resulting in the term $2^{t_1+t_2}$. The two other combinatorial terms are the number of ways one can choose t_1 vertices supported on the set K_1-K and t_2 vertices supported on K_2-K. From (24) and (17) we can conclude theorem 4.2. ■

In the following sub-sections we will derive the internal and external angles for a face $\mathcal{F}$, and a face $\mathcal{G}$ containing $\mathcal{F}$, and will provide closed form upper bounds for them. We combine the terms together and compute the exponents using the Laplace method in Section 5.2.2, and derive thresholds for the negativity of the cumulative exponent.

5.2.1 Computation of Internal and External Angles

Theorem 5.1. *Let Z be a random variable defined as*

$$Z = (k_1w_{K_1}^2 + k_2w_{K_2}^2)X_1 - w_{K_1}^2\sum_{i=1}^{t_1}X_1' - w_{K_2}^2\sum_{i=1}^{t_2}X_1'',$$

where $X_1 \sim N(0, \frac{1}{2(k_1w_{K_1}^2+k_2w_{K_2}^2)})$ is a normal distributed random variable, $X_i' \sim HN(0, \frac{1}{2w_{K_1}^2})$ $1 \leq i \leq t_1$ and $X_i'' \sim HN(0, \frac{1}{2w_{K_2}^2})$ $1 \leq i \leq t_2$ are independent (from each other and from X_1) half normal distributed random variables. Let $p_Z(\cdot)$ denote the probability distribution function of Z and $c_0 = \frac{\sqrt{\pi}}{2^{l-k}}\left((k_1+t_1)w_{K_1}^2 + (k_2+t_2)w_{K_2}^2\right)^{1/2}$. Then

$$\beta(k_1,k_2|t_1,t_2) = c_0p_Z(0) \tag{25}$$

Theorem 5.2. *The external angle $\zeta(\mathcal{G},\mathcal{P}_{\mathbf{w}}) = \zeta(d_1,d_2)$ between the face $\mathcal{G}$ and $\mathcal{P}_{\mathbf{w}}$, where G is supported* *on the set L with $|L \cap K_1| = d_1$ and $|L \cap K_2| = d_2$ is given by:*

$$\zeta(d_1, d_2) = \pi^{-\frac{n-l+1}{2}} 2^{n-l} \int_0^\infty e^{-x^2} \left(\int_0^{\frac{w_{K_1} x}{\xi(d_1,d_2)}} e^{-y^2}\, dy \right)^{r_1} \left(\int_0^{\frac{w_{K_2} x}{\xi(d_1,d_2)}} e^{-y^2}\, dy \right)^{r_2} dx, \tag{26}$$

Where $\xi^2(d_1, d_2) = \sum_{i \in L} w_i^2 = d_1 w_{K_1}^2 + d_2 w_{K_2}^2$, $r_1 = n_1 - d_1$ and $r_2 = n_2 - d_2$.

The detailed proofs of these theorems which are generalizations of similar Theorems in [4] are given in the online version of this paper [28].

5.2.2 Derivation of the Critical Weak and Strong δ_c Threshold

So far we have proved that the probability of the failure event is bounded by the formula

$$\mathbb{P}\{E^c\} \leq \sum_{\substack{0 \leq t_1 \leq n_1 - k_1 \\ 0 \leq t_2 \leq n_2 - k_2 \\ t_1 + t_2 > m - k_1 - k_2 + 1}} 2^{t_1+t_2+1} \binom{n_1 - k_1}{t_1} \binom{n_2 - k_2}{t_2} \beta(k_1, k_2 | t_1, t_2) \zeta(t_1 + k_1, t_2 + k_2), \tag{27}$$

where we gave expressions for $\beta(t_1, t_2|k_1, k_2)$ and $\zeta(t_1 + k_1, t_2, k_2)$ in Section 5.2.1. Now our objective is to show that the R.H.S of (27) will exponentially decay to 0 as $n \to \infty$, provided that $\delta = \frac{m}{n}$ is greater than a critical threshold δ_c, which we are trying to evaluate. To this end we bound the exponents of the combinatorial, internal angle and external angle terms in (27), and find the values of δ for which the net exponent is strictly negative. The maximum such δ will give us δ_c. Starting with the combinatorial term, we use Stirling approximating on the binomial coefficients to achieve the following as $n \to \infty$ and $\epsilon \to 0$

$$\frac{1}{n} \log \left(2^{t_1+t_2+1} \binom{n_1 - k_1}{t_1} \binom{n_2 - k_2}{t_2} \right) \to \left(\gamma_1(1-p_1) H(\frac{\tau_1}{\gamma_1(1-p_1)}) + \gamma_2(1-p_2) H(\frac{\tau_2}{\gamma_2(1-p_2)}) + \tau_1 + \tau_2 \right) \log 2, \tag{28}$$

where $\tau_1 = \frac{t_1}{n}$ and $\tau_2 = \frac{t_2}{n}$.

For the external angle and internal angle terms we prove the following two exponents

1. Let $g(x) = \frac{2}{\sqrt{\pi}} e^{-x^2}$, $G(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2} dy$. Also define $c = (\tau_1 + \gamma_1 p_1) + \omega^2(\tau_2 + \gamma_2 p_2)$, $\alpha_1 = \gamma_1(1 - p_1) - \tau_1$ and $\alpha_2 = \gamma_2(1 - p_2) - \tau_2$. Let x_0 be the unique solution to x of the following:

$$2c - \frac{g(x)\alpha_1}{xG(x)} - \frac{\omega g(\omega x)\alpha_2}{xG(\omega x)} = 0$$

   Define

$$\psi_{ext}(\tau_1, \tau_2) = c x_0^2 - \alpha_1 \log G(x_0) - \alpha_2 \log G(\omega x_0) \tag{29}$$

2. Let $b = \frac{\tau_1 + \omega^2 \tau_2}{\tau_1 + \tau_2}$ and $\varphi(.)$ and $\Phi(.)$ be the standard Gaussian pdf and cdf functions respectively. Also let $\Omega' = \gamma_1 p_1 + \omega^2 \gamma_2 p_2$ and $Q(s) = \frac{\tau_1 \varphi(s)}{(\tau_1+\tau_2)\Phi(s)} + \frac{\omega \tau_2 \varphi(\omega s)}{(\tau_1+\tau_2)\Phi(\omega s)}$. Define the function $\hat{M}(s) = -\frac{s}{Q(s)}$ and solve for s in $\hat{M}(s) = \frac{\tau_1+\tau_2}{(\tau_1+\tau_2)b + \Omega'}$. Let the unique solution be s^* and set $y = s^*(b - \frac{1}{\hat{M}(s^*)})$. Compute the rate function $\Lambda^*(y) = sy - \frac{\tau_1}{\tau_1+\tau_2}\Lambda_1(s) - \frac{\tau_2}{\tau_1+\tau_2}\Lambda_1(\omega s)$ at the point $s = s^*$, where

$\Lambda_1(s) = \frac{s^2}{2} + \log(2\Phi(s))$. The internal angle exponent is then given by:

$$\psi_{int}(\tau_1, \tau_2) = (\Lambda^*(y) + \frac{\tau_1 + \tau_2}{2\Omega'} y^2 + \log 2)(\tau_1 + \tau_2). \tag{30}$$

We now state the following lemmas, the proof techniques of which are very similar to similar arguments in [3, 1]. The details of the proof can be found in the online version of this paper [28].

Lemma 5.2. *Fix δ, $\epsilon > 0$. There exists a finite number $n_0(\delta, \epsilon)$ such that*

$$\frac{1}{n} \log(\zeta(t_1 + k_1, t_2 + k_2)) < -\psi_{ext}(\tau_1, \tau_2) + \epsilon, \tag{31}$$

uniformly in $0 \le t_1 \le n_1 - k_1$, $0 \le t_2 \le n_2 - k_2$ and $t_1 + t_2 \ge m - k_1 - k_2 + 1$, $n \ge n_0(\delta, \epsilon)$.

Lemma 5.3. *Fix δ, $\epsilon > 0$. There exists a finite number $n_1(\delta, \epsilon)$ such that*

$$\frac{1}{n} \log(\beta(t_1, t_2 | k_1, k_2)) < -\psi_{int}(\tau_1, \tau_2) + \epsilon, \tag{32}$$

uniformly in $0 \le t_1 \le n_1 - k_1$, $0 \le t_2 \le n_2 - k_2$ and $t_1 + t_2 \ge m - k_1 - k_2 + 1$, $n \ge n_1(\delta, \epsilon)$.

Combining Lemmas 5.2 and 5.3, (28), and the bound in (27) we readily get the critical bound for δ_c as in the Theorem 4.3.

Derivation of the strong and sectional threshold can be easily done using union bounds to account for all possible support sets and/or all sign patterns. The corresponding upper bound on the failure probability for the strong threshold is given by:

$$\binom{n_1}{k_1} \binom{n_2}{k_2} 2^k P_{K,-} \tag{33}$$

It then follows that the strong threshold of δ is given by δ_c in Theorem 4.3, except that the combinatorial exponent $\psi_{com}(\cdot, \cdot)$ must be corrected by adding a term

$$(\gamma_1 p_1 + \gamma_2 p_2 + \gamma_1 H(p_1) + \gamma_2 H(p_2)) \log 2, \tag{34}$$

to the RHS of (6). Similarly, for the sectional threshold, which deals with all possible support sets but almost all sign patterns, the modification in the combinatorial exponent term is as follows:

$$(\gamma_1 H(p_1) + \gamma_2 H(p_2)) \log 2. \tag{35}$$

5.3 Generalizations

Except for some subtlety in the large deviation calculations, the generalization of the results of the previous section to an arbitrary $u \ge 2$ classes of entries is straightforward. Consider a nonuniform sparse model with u classes $K_1, \cdots, K_u$ where $|K_i| = n_i = \gamma_i n$, and the sparsity fraction over the set K_i is p_i, and a recovery scheme based on weighted ℓ_1 minimization with weight ω_i for the set K_i. The bound in (21) is general and can always be used. Due to isomorphism, the internal and external angles $\beta(\mathcal{F},\mathcal{G})$ and $\zeta(\mathcal{G},\mathcal{P}_{\mathbf{w}})$ only depend on the number of vertices that the supports of $\mathcal{F}$ and $\mathcal{G}$ have in common with each K_i. Therefore, a generalization to (5) would be:

$$\mathbb{P}\{E^c\} \leq 2 \sum_{\substack{0 \leq \mathbf{t} \leq \mathbf{n}-\mathbf{k} \\ \mathbf{1}^T\mathbf{t} > m - \mathbf{1}^T\mathbf{k}+1}} \Pi_{1\leq i\leq u} 2^{t_i}\binom{n_i-k_i}{t_i}\beta(\mathbf{k}|\mathbf{t})\zeta(\mathbf{t}+\mathbf{k}), \tag{36}$$

where $\mathbf{t} = (t_1,\dots,t_u)^T$, $\mathbf{k} = (k_1,\dots,k_u)^T$ and $\mathbf{1}$ is a vector of all ones. Invoking generalized forms of Theorems 5.2 and 5.1 to approximate the terms $\beta(\mathbf{k}|\mathbf{t})$ and $\zeta(\mathbf{k}+\mathbf{t})$, we conclude the following Theorem.

Theorem 5.3. *Consider a nonuniform sparse model with u classes $K_1,\dots,K_u$ with $|K_i| = n_i = \gamma_1 n$, and sparsity fractions $p_1,p_2,\dots,p_u$, where n is the signal dimension. Also, let the functions $g(.),G(.),\psi(.),\Psi(.)$ be as defined in Theorem 4.3. For positive values $\{\omega_i\}_{i=1}^u$, the recovery thresholds (weak, sectional and strong) of the weighted ℓ_1 minimization program:*

$$\min_{\mathbf{A}\mathbf{x}=\mathbf{y}} \sum_{i=1}^{u}\omega_i\|\mathbf{x}_{K_i}\|_1,$$

is given by the following expression:

$$\boxed{\begin{array}{c}\delta_c = \min\{\delta \mid \psi_{com}(\tau)-\psi_{int}(\tau)-\psi_{ext}(\tau) < 0 \ \forall \tau=(\tau_1,\cdots,\tau_u)^T : \\ 0\leq \tau_i \leq \gamma_i(1-p_i) \ \forall 1\leq i\leq u, \ \sum_{i=1}^u \tau_i > \delta - \sum_{i=1}^u \gamma_i p_i\}\end{array}}$$

where ψ_{com}, ψ_{int} and ψ_{ext} are obtained from the following expressions:

1. *$\psi_{com}(\tau) = \log 2\sum_{i=1}^u \gamma_i(1-p_i)H(\frac{\tau_i}{\gamma_i(1-p_i)}) + \tau_i$, for the weak threshold. For sectional threshold this must be modified by adding a term $\log 2\sum_{i=1}^u \gamma_i H(p_i)$. For strong threshold, it must be also added with $\sum_{i=1}^u \gamma_i p_i$.*
2. *$\psi_{ext}(\tau) = cx_0^2 - \sum_{i=1}^u \alpha_i \log G(\omega_i x_0)$, where $c = \sum_{i=1}^u \omega_i^2(\tau_i+\gamma_i p_i)$, $\alpha_i = \gamma_i(1-p_i)-\tau_i$ and x_0 is the unique solution of $2c = \sum_{i=1}^u \omega_i \frac{g(\omega_i x_0)\alpha_i}{x_0 G(\omega x_0)}$,*
3. *$\psi_{int}(\tau) = \lambda(\Lambda^*(y) + \frac{\lambda y^2}{2\sum_{i=1}^u \omega_i^2\gamma_i p_i} + \log 2)$, where $\lambda = \sum_{i=1}^u \tau_i$, and y and $\Lambda^*(y)$ are obtained as follows. Let $b = \frac{\sum_{i=1}^u \omega_i^2\tau_i}{\lambda}$, $Q(s) = \sum_{i=1}^u \frac{\tau_i\varphi(s)}{\lambda\Phi(s)}$. Let s^* be the solution to s in $-\frac{Q(s)}{s} = b + \frac{\sum_{i=1}^u \omega_i^2\gamma_i p_i}{\lambda}$, and $y = s^*(b - \frac{1}{\hat{M}(s^*)})$. Then $\Lambda^*(y) = s^* y - 1/\lambda\sum_{i=1}^u \tau_i\left(\frac{\omega_i^2 s^{*2}}{2} + \log(2\Phi(\omega_i s^*))\right)$.*

5.4 Robustness

Proof of Theorem 4.5. We first state the following lemma, which is very similar to Theorem 2 of [4]. We skip its proof for brevity.

Lemma 5.4. *Let $K \subset \{1, 2\cdots, n\}$ and the weight vector $\mathbf{w} = (w_1, w_2, \cdots, w_n)^T$ be fixed. Define $\mathbf{W} = diag(w_1, w_2, \cdots, w_n)$ and suppose $C > 1$ is given. For every vector $\mathbf{x}_0 \in \mathbb{R}^{n\times 1}$, the solution $\hat{\mathbf{x}}$ of (3) satisfies*

$$\|\mathbf{W}(\mathbf{x}_0 - \hat{\mathbf{x}})\|_1 \leq 2\frac{C+1}{C-1}\sum_{i\in\overline{K}} w_i|(x_0)_i|, \tag{37}$$

if and only if for every $\mathbf{z} \in \mathcal{N}(\mathbf{A})$ the following holds:

$$C\sum_{i\in K} w_i|z_i| \leq \sum_{i\in\overline{K}} w_i|z_i|. \tag{38}$$

Let $\mathbf{z} = (z_1, \cdots, z_n)^T$ be a vector in the null space of $\mathbf{A}$, and assume that

$$C'\sum_{i\in L_1\cup L_2} w_i|z_i| = \sum_{i\in\overline{L_1}\cap\overline{L_2}} w_i|z_i|. \tag{39}$$

Let K_{ϵ_1} and K_{ϵ_2} be the solutions of the following problems

$$K_{\epsilon_1} \;:\; \max_{K_{\epsilon_1}\subset K_1\cap\overline{L_1}, |K_{\epsilon_1}|=\epsilon_1\gamma_1 p_1 n} \sum_{i\in K_{\epsilon_1}} w_i|z_i|, \tag{40}$$

$$K_{\epsilon_2} \;:\; \max_{K_{\epsilon_2}\subset K_2\cap\overline{L_2}, |K_{\epsilon_2}|=\epsilon_2\gamma_2 p_2 n} \sum_{i\in K_{\epsilon_2}} w_i|z_i|. \tag{41}$$

Let $L_1' = L_1 \cup K_{\epsilon_1}$ and $L_2' = L_2 \cup K_{\epsilon_2}$. From the definition of K_{ϵ_1} and K_{ϵ_2}, it follows that

$$\sum_{i\in K_{\epsilon_1}} w_i|z_i| \;\geq\; \frac{\epsilon_1 p_1}{1-p_1}\sum_{i\in\overline{L_1'}\cap K_1} w_i|z_i|, \tag{42}$$

$$\sum_{j\in K_{\epsilon_2}} w_j|z_j| \;\geq\; \frac{\epsilon_2 p_2}{1-p_2}\sum_{j\in\overline{L_2'}\cap K_2} w_j|z_j|. \tag{43}$$

Adding $C'\left(\sum_{i\in K_{\epsilon_1}} w_i|z_i| + \sum_{j\in K_{\epsilon_2}} w_j|z_j|\right)$ to both sides of (39) and using (42) and (43), we can write:

$$C'\sum_{i\in L_1'\cup L_2'} w_i|z_i| \;\geq\; \sum_{i\in\overline{L_1}\cap\overline{L_2}} w_i|z_i| + C'\left(\frac{\epsilon_1 p_1}{1-p_1}\sum_{i\in\overline{L_1'}\cap K_1} w_i|z_i| + \frac{\epsilon_2 p_2}{1-p_2}\sum_{i\in\overline{L_2'}\cap K_2} w_i|z_i|\right) \tag{44}$$

$$\geq\; \left(1 + (C'+1)\min\left(\frac{\epsilon_1 p_1}{1-p_1}, \frac{\epsilon_2 p_2}{1-p_2}\right)\right)\sum_{i\in\overline{L_1'}\cap\overline{L_2'}} w_i|z_i|. \tag{45}$$

Note that $|L_1'| = \gamma_1 p_1 n$ and $|L_2'| = \gamma_2 p_2 n$. Therefore, since $\delta = \frac{m}{n} \geq \delta_c^{(S)}(\gamma_1, \gamma_2, p_1, p_2, \omega)$, we know that

$\sum_{i\in L_1'\cup L_2'} w_i|z_i| \leq \sum_{i\in \overline{L_1'}\cap\overline{L_2'}} w_i|z_i|$. From this and (45) we conclude that

$$C' \geq \left(1+(C'+1)\min(\frac{\epsilon_1 p_1}{1-p_1},\frac{\epsilon_2 p_2}{1-p_2})\right), \tag{46}$$

or equivalently

$$C' \geq \frac{1+\min(\frac{\epsilon_1 p_1}{1-p_1},\frac{\epsilon_2 p_2}{1-p_2})}{1-\min(\frac{\epsilon_1 p_1}{1-p_1},\frac{\epsilon_2 p_2}{1-p_2})}. \tag{47}$$

Using Lemma 5.4 and the above inequality, we conclude (14). ■

6 Approximate Support Recovery and Reweighted ℓ_1

So far, we have considered the case where some prior information about the support of the sparse signal is available. When no such information exists, we ask if there is a polynomial time algorithm that outperforms ℓ_1 minimization? Using the analytical tools of this paper, it is possible to prove that a class of reweighted ℓ_1 minimization algorithms, with complexity only twice as ℓ_1 minimization, have a strictly higher recovery thresholds for sparse signals whose nonzero entries follow certain classes of distributions (e.g. Gaussian). The technical details of this claim are not brought here, since they stand beyond the scope of this paper. However, we briefly mention the algorithm and provide insight into its functionality. The proposed reweighted ℓ_1 recovery algorithm (studied in more details in [13]) is composed of two steps. In the first step a standard ℓ_1 minimization is done, and based on the output, a set of entries where the signal is likely to reside (the so-called approximate support) is identified. The unknown signal can thus be thought of as two classes, one with a relatively high fraction of nonzero entries, and one with a small fraction. The second step is a weighted ℓ_1 minimization step where entries outside the approximate support set are penalized with a constant weight larger than 1. Although no prior information is available at the beginning, the post processing on the output of ℓ_1 minimization results in an approximate support inference, which is similar to the nonuniform sparsity model we introduced in this paper. A more comprehensive study on this can be found in [13]. The algorithm is described as follows:

Algorithm 1.

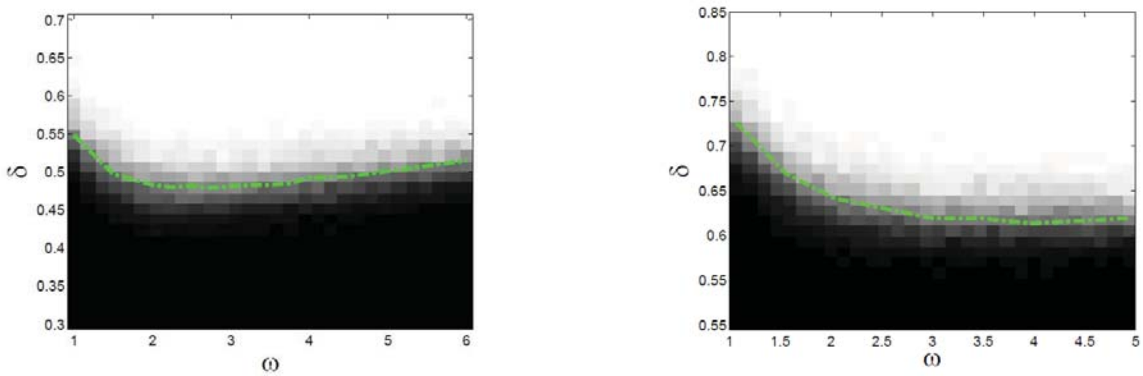
1. *Solve the ℓ_1 minimization problem:*

$$\hat{\mathbf{x}} = \arg\min \|\mathbf{z}\|_1 \quad \textit{subject to} \quad \mathbf{Az} = \mathbf{Ax}. \tag{48}$$

2. *Obtain an approximation for the support set of $\mathbf{x}$: find the index set $L \subset \{1, 2, ..., n\}$ which corresponds to the largest k elements of $\hat{\mathbf{x}}$ in magnitude.*

3. *Solve the following weighted ℓ_1 minimization problem and declare the solution as output:*

$$\mathbf{x}^* = \arg\min \|\mathbf{z}_L\|_1 + \omega\|\mathbf{z}_{\overline{L}}\|_1 \quad \textit{subject to} \quad \mathbf{Az} = \mathbf{Ax}. \tag{49}$$



(a) $\gamma_1 = \gamma_2 = 0.5$, $p_1 = 0.4$ and $p_2 = 0.05$.

(b) $\gamma_1 = \gamma_2 = 0.5$, $p_1 = 0.65$ and $p_2 = 0.1$.

Figure 5: Empirical recovery percentage of weighed ℓ_1 minimization for different weight values ω, and different number of measurements $\delta = \frac{m}{n}$ and $n = 200$. Signals have been selected from a nonuniform sparse models. White indicates perfect recovery..

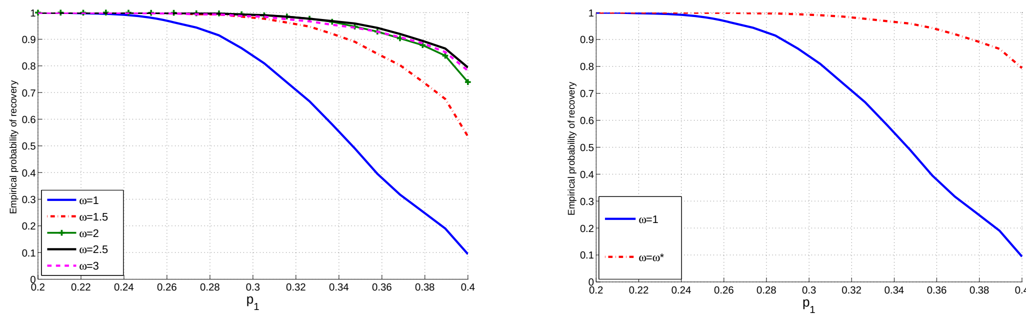
In general, for Algorithm 1, k is chosen based on the signal model in different applications. A general guideline is to choose k such that the k positions chosen intersect greatly with the support set of the signal vector. In other words, if K is the support set of $\mathbf{x}$, then $\frac{|L \cap K|}{|L|}$ is close to 1. In this way, the k positions chosen will correspond to a block in the weighted basis pursuit algorithm considered in this paper, where a higher portion of this block will correspond to the signal support. So we have extracted some support prior information from the first iteration, even though we did not have any prior information on the support of the signal at the beginning. When we do the next iteration of decoding, we can use this information to improve the decoding result. The analytical tools we developed in this paper for studying the weighted ℓ_1 minimization in the presence of prior information help analyze the performance of the last stage of Algorithm 1. More details can be found in [13].

Other variations of reweighted ℓ_1 minimization are given in the literature. For example the algorithm in [12] assigns a different weight to every entry of the signal based on the inverse absolute values of the entries of ℓ_1 approximation, $\hat{\mathbf{x}}$. To the best of our knowledge, there is no theoretical justification for strict threshold improvement in the approach of [12], or any similar proposed algorithm. However, we have been able to show that the proposed iterative reweighted algorithm will provably improve the recovery threshold[5] for certain types of signals even when no prior information is available [13].

7 Simulation Results

We demonstrate by some examples that appropriate weights can boost the recovery percentage. In Figure 5 we have shown the empirical recovery threshold of weighted ℓ_1 minimization for different values of the

[5]Here, by recovery threshold we mean the maximum number of nonzero entries of a random sparse signal that can be recovered by ℓ_1 minimization for a fixed number of measurements.



(a) (b)

Figure 6: Empirical probability of successful recovery for weighted ℓ_1 minimization with different weights (unitary weight for the first subclass and ω for the other one) and suboptimal weights in a nonuniform sparse setting. $p_2 = 0.05$, $\gamma_1 = \gamma_2 = 0.5$ and $m = 0.5n = 100$. ω^* is (b) is the optimum value of ω for each p_1 among the values shown in (a).

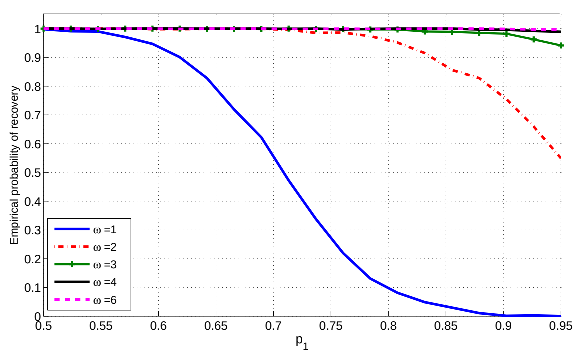


Figure 7: Empirical probability of successful recovery for different weights. $p_2 = 0.1$,$\gamma_1 = \gamma_2 = 0.5$ and $m = 0.75n = 150$.

weight $\omega = \frac{w_{K_1}}{w_{K_2}}$, for two particular nonuniform sparse models. Note that the empirical threshold is somewhat identifiable with naked eye, and is very similar to the theoretical curve of Figure 2 for similar settings. In another experiment, we fix p_2 and $n = 2m = 200$, and try ℓ_1 and weighted ℓ_1 minimization for various values of p_1. We choose $n_1 = n_2 = \frac{n}{2}$. Figure 6a shows one such comparison for $p_2 = 0.05$ and different values of w_{K_2}. Note that the optimal value of w_{K_2} varies as p_1 changes. Figure 6b illustrates how the optimal weighted ℓ_1 minimization surpasses the ordinary ℓ_1 minimization. The optimal curve is basically achieved by selecting the best weight of Figure 6a for each single value of p_1. Figure 7 shows the result of simulations in another setting where $p_2 = 0.1$ and $m = 0.75n$ (similar to the setting of Section 4). Note that these results very well match the theoretical results of Figures 2a and 2b.

Another nonuniform model with $\gamma_1 = 0.25$, $\gamma_2 = 0.75$ was considered for simulations. We fixed $\delta = 0.45$, and the overall sparsity fraction $\gamma_1 p_1 + \gamma_2 p_2 = 0.3$. For random vectors of size $n = 200$, the probability of successful recovery as a function of the sparsity of the second subclass, i.e. p_2, for various weights ω was obtained, and is depicted in Figure 8. As displayed, although ℓ_1 minimization fails in all of the cases to recover the sparse vectors, various weighted ℓ_1 approaches have higher chances of success, especially when the first subclass is very dense and a high weight is used ($p_1 \approx 1$, $\omega >> 1$).

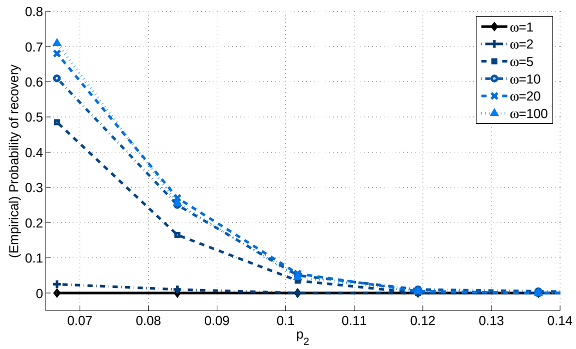
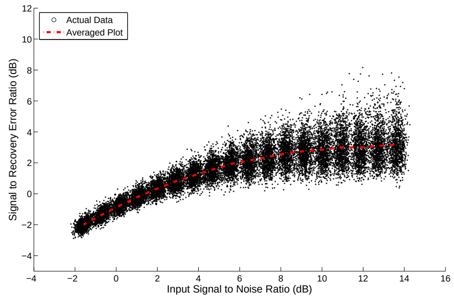


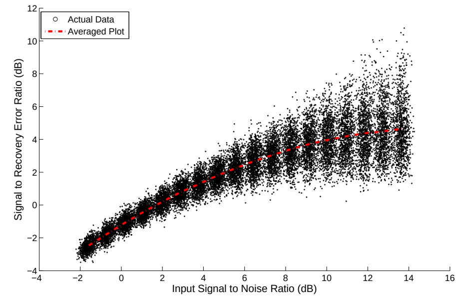
Figure 8: Probability of successful recovery (empirical) of nonuniform sparse signals with $\gamma_1 = 0.25, \gamma_2 = 0.75, p_1\gamma_1 + p_2\gamma_2 = 0.3$ vs. the sparsity of the second subclass p_2. $\delta = 0.45$.

In Figure 9, we have displayed the performance of weighted ℓ_1 minimization in the presence of noise. The original signal is a nonuniformly sparse vector with sparsity fractions $p_1 = 0.4, p_2 = 0.05$ over two subclasses $\gamma_1 = \gamma_2 = 0.5$, with $n = 200$. However, a white Gaussian noise vector is added before compression, i.e. $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$, where $\mathbf{v}$ (0, is an i.i.d. Gaussian vector. Figure 9 shows a scatter plot of all output signal to recovery error ratios as a function of the input SNR, for all simulations. The input SNR in dB is defined as $10\log_{10}\frac{\|\mathbf{x}\|_2^2}{\|\mathbf{v}\|_2^2}$, and output signal to error recovery in dB is defined as $10\log_{10}\frac{\|\mathbf{x}\|_2^2}{\|\hat{\mathbf{x}}-\mathbf{x}\|_2^2}$. The facts that the signal to recovery error does not drop drastically in small SNR regimes, and is mostly concentrated around the average values indicate the robustness of the weighted ℓ_1 algorithm in the presence of a moderate level noise. In Figure 10 the average curves are compared together for different values of weight ω. We can see that in the high input SNR regime, a non-unit weight $\omega = 3$ is advantageous over the regular ℓ_1 minimization. However, for stronger noise variances, ℓ_1 minimization seems to be more robust and yields better performance.

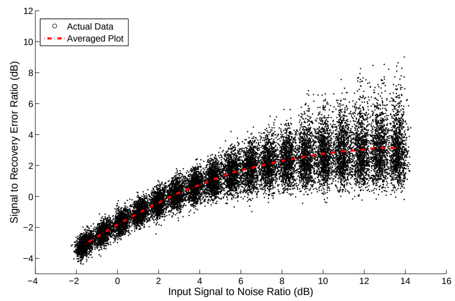
We have done some experiments with regular ℓ_1 and weighted ℓ_1 minimization recovery on some real world data. We have chosen a pair of satellite images (Figure 11) taken in two different years, 1989 (left) and 2000 (right), from the New Britain rainforest in Papua New Guinea. Image originally belongs to Royal Society for the Protection of Birds and was taken from the Guardian archive, an article on deforestation. These images are generally recorded to evaluate environmental effects such as *deforestation*. The difference of images taken at different times is generally not very significant, and thus can be thought of as compressible. In addition, the difference is usually more substantial over certain areas, e.g. forests. Therefore, it can be cast in a nonuniform sparse model. We have applied ℓ_1 minimization to recover the difference image over the subframe (subset of the original images), identified by the red rectangles in Figure 11. In addition, recovery by weighted ℓ_1 minimization was also implemented. To assign weights, we divided the pixels of each frame into two classes of equal sizes, where the concentration of the forestal area is larger over one of the classes, and hence the difference image is less sparse. This class is identified by the union of two rectangles with green frames on the left and bottom right of the image. The justification for such a decomposition is the concentration of green area. In other words, with



(a) $\omega = 1$.



(b) $\omega = 3$.



(c) $\omega = 5$.

Figure 9: Signal to recovery error ratio for weighted ℓ_1 minimization with weight ω vs. input SNR for nonuniform sparse signals with $\gamma_1 = \gamma_2 = 0.5$, $p_1 = 0.4, p_2 = 0.05$ superimposed with Gaussian noise.

the knowledge of the fact that the environmental changes are more apparent over forests in large scale, the difference image is expected to be denser over those regions. We also implemented the reweighted ℓ_1 minimization of Algorithm 1, with $k = 0.1n$ (n being the total number of frame pixels), which assumed no prior knowledge about the structural sparsity of the signal. This value of k was chosen heuristically, and is close to the actual support size of the signal. The original size of the image is 275×227. We reduced the resolution by roughly a factor of 0.05 for more tractability of ℓ_1 solver in MATLAB. In addition, only the gray scale version of the difference image was taken into account, and was normalized so that the maximum intensity is 1. Furthermore, prior to compression, the difference image was further sparsified by rounding the intensities less than 0.1 to zero. We pick the weight value $\omega = 2$ for both the weighted ℓ_1 and the reweighted ℓ_1 recovery. We define the normalized recovery error to be the sum square of the intensity differences in the recovered and the original image, divided by the sum square of the original image intensity, i.e. $\sum_{i\in\text{frame}}(I_i - \hat{I}_i)^2 / \sum_{i\in\text{frame}} I_i^2$. The average normalized error for different recovery methods, namely ℓ_1, weighted ℓ_1 and reweighted ℓ_1 minimization, is displayed in Figure 13a as a function of δ. The average is taken over 50 realizations of i.i.d. Gaussian measurement matrices for each δ. As can be seen, the recovery improvement is significant in the weighted ℓ_1 minimization. The reweighted ℓ_1 algorithm is superior to the regular ℓ_1 algorithm, but is worse than the weighted ℓ_1 method, which is due to the availability of support prior information and the proper choice of the subclasses in the weighted ℓ_1 algorithms.

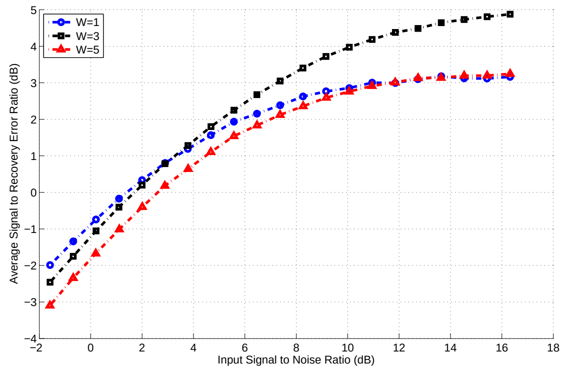


Figure 10: Average signal to recovery error ratio for weighted ℓ_1 minimization with weight ω vs. input SNR for nonuniform sparse signals with $\gamma_1 = \gamma_2 = 0.5$, $p_1 = 0.4, p_2 = 0.05$ superimposed with Gaussian noise.

Figure 11: Satellite images taken from the New Britain rainforest in Papua Guina at 1989 (left) and 2000 (right). Red boxes identify the subframe used for the experiment, and green boxes identify the regions with higher associated weight in the weighted ℓ_1 recovery. Image originally belongs to Royal Society for the Protection of Birds and was taken from the Guardian archive, an article on deforestation `http://www.guardian.co.uk/environment/2008/jan/09/endangeredspecies.endangeredhabitats`.

Another experiment was done on a pair of brain fMRI images taken at two different instances of time, shown in Figure 12. Similar to the satellite images, the objective is recover the difference image from a set of compressed measurements. The significant portion of the difference image in fMRI lies on the regions where the brain is identified as most active. Depending on the particular task that the patient undertakes, these regions can be (roughly) known a priori. The original image size is 271×271, and similar preprocessing steps as for the satellite images were done before compression. We used ℓ_1 minimization and weighted ℓ_1 minimization with a higher weight $\omega = 1.3$ on the regions identified by the green boxes. This choice of ω resulted in a slightly better performance in weighted ℓ_1 algorithm than $\omega = 2$. We also tried algorithm 1 with no presumed prior information, with $k = 0.1n$ and $\omega = 1.3$. The average normalized recovery errors are displayed in Figure 13b, from which we can infer similar conclusions as in the case of satellite images.

8 Conclusion and Future Work

We analyzed the performance of the weighted ℓ_1 minimization for nonuniform sparse models. We computed explicitly the phase transition curves for the weighted ℓ_1 minimization, and showed that with proper weighting, the recovery threshold for weighted ℓ_1 minimization can be higher than that of regular ℓ_1 minimization. We provided simulation results to verify this both in the noiseless and noisy situation.

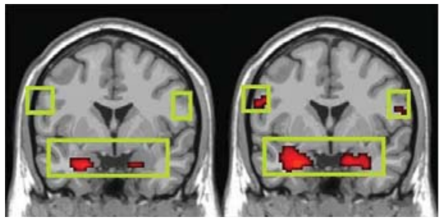
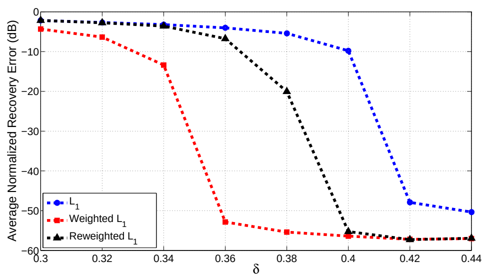
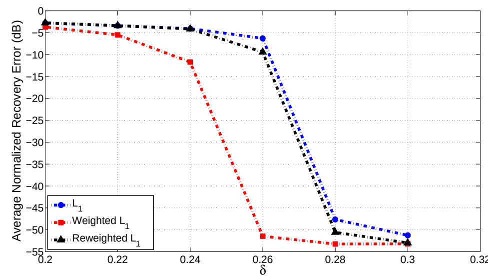


Figure 12: Functional MRI images of the brain at two different instances illustrating the brain activity. Green boxes identify the region with higher associated weight in the weighted ℓ_1 recovery. Image is adopted from `https://sites.google.com/site/psychopharmacology2010/student-wiki-for-quiz-9`.



(a)



(b)

Figure 13: Average normalized recovery error for ℓ_1, weighted ℓ_1 and reweighted ℓ_1 minimization recovery of the difference between the subframes of (a) a pair of satellite images shown in Figure 11, and (b) the pair of brain fMRI images shown in Figure 12. Data is averaged over different realizations of measurement matrices for each δ.

Some of our simulations were performed on real world data of satellite images, where the nonuniform sparse model is a valid assumption. Future work shall address generalizing the results of this paper to other measurement matrices with different distributions than i.i.d Gaussian. In particular, by using a similar idea as [26], one might be able to assert that a for class of distributions of measurement matrices including random Fourier ensembles, random Bernoulli, etc. similar sharp thresholds can be obtained for the weighted ℓ_1 minimization. A further interesting research topic to be addressed in future work would be to characterize the gain in recovery percentage as a function of the number of distinguishable classes u in the nonuniform model. In addition, we have used the results of this paper to build iterative reweighted ℓ_1 minimization algorithms that are provably strictly better than ℓ_1 minimization, when the nonzero entries of the sparse signals are known to come from certain distributions (in particular Gaussian distributions) [13, 21]. The basic idea there is that a simple post processing procedure on the output of ℓ_1 minimization results, with high probability, in a hypothetical nonuniform sparsity model for the unknown signal, which can be exploited for improved recovery.

References

[1] D. Donoho, "High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension,", Discrete and Computational Geometry , 102(27), pp. 617-652, 2006, Springer.

[2] E. Candès and T. Tao, "Decoding by linear programming,", IEEE Trans. on Information Theory, 51(12), pp. 4203 - 4215, December 2005.

[3] D. Donoho and J. Tanner, "Neighborliness of randomly-projected simplices in high dimensions," Proc. National Academy of Sciences, 102(27), pp. 9452-9457, 2005.

[4] W. Xu and B. Hassibi, "Compressed Sensing Over the Grassmann Manifold: A Unified Analytical Framework," Allerton Conference 2008.

[5] W. Lu and N. Vaswani, "Modified Compressive Sensing for Real-time Dynamic MR Imaging," IEEE Intl. Conf. Image Proc (ICIP), 2009.

[6] N. Vaswani and W. Lu, "Modified-CS: Modifying Compressive Sensing for Problems with Partially Known Support," IEEE Trans. Signal Processing, September 2010.

[7] W. Lu and N. Vaswani, "Regularized Modified-BPDN for compressive sensing with partially known support," Preprint 2010.

[8] M. Davenport, M. Wakin, and R. Baraniuk, "Detection and estimation with compressive measurements", Rice ECE Department Technical Report TREE 0610, November 2006

[9] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," IEEE Trans. on Signal Processing, 56(6) pp. 2346 - 2356, June 2008

[10] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," submitted to IEEE Transactions on Information Theory.

[11] M. Stojnic, F. Parvaresh and B. Hassibi, "On the reconstruction of block-sparse signals with an optimal number of measurements," IEEE Trans. on Signal Proc., 2009.

[12] E. J. Candès, M. Wakin and S. Boyd. "Enhancing sparsity by reweighted l1 minimization," J. Fourier Anal. Appl., 14 877-905.

[13] A. Khajehnejad, W. Xu, S. Avestimher and B. Hassibi, "Improved Sparse Recovery Thresholds with Two-Step Reweighted ℓ_1 Minimization," ISIT 2010.

[14] B. Grünbaum, "Grassmann angles of convex polytopes," Acta Math., 121:pp.293-302, 1968.

[15] B. Grünbaum, "Convex polytopes," volume 221 of Graduate Texts in Mathematics, Springer-Verlag, New York, second edition, 2003.

[16] F. Affentranger and R. Schneider, "Random projections of regular simplices," Discrete and Computational Geometry, Vol. 7, 1992, pp. 219-226.

[17] L.A.Santaló, "Geometría integral en espacios de curvatura constante," Rep. Argetina Publ. Com. Nac. Energí Atómica, Ser.Mat 1,No.1(1952)

[18] P. McMullen, "Non-linear angle-sum relations for polyhedral cones and polytopes," Math. Proc. Cambridge Philos. Soc., 78(2):pp.247-261, 1975.

[19] M. Stojnic, W. Xu, and B. Hassibi, "Compressed sensing - probabilistic analysis of a null-space characterization," IEEE International Conference on Acoustics, Speech and Signal Processing, Pages:3377-3380, March 31 2008-April 4 2008.

[20] J. Coates, Y. Pointurier and M. Rabbat, "Compressed network monitoring," Proceedings of IEEE Workshop on Statistical Signal Processing, Madison, WI, Aug. 2007.

[21] B. Hassibi, A. Khajehnejad, W. Xu, S. Avestimehr, "Breaking the ℓ_1 recovery thresholds with reweighted ℓ_1 optimization," in proceedings of Allerton Conference 2009.

[22] O.Milenkovic, R. Baraniuk, and T. Simunic-Rosing, "Compressed sensing meets bionformatics: a new DNA microarray architecture," Information Theory and Applications Workshop, San Diego, 2007.

[23] S. Erickson, and C. Sabatti, "Empirical Bayes estimation of a sparse vector of gene expression," Statistical Applications in Genetics and Molecular Biology, 2005.

[24] H.Vikalo, F. Parvaresh, S. Misra and B. Hassibi,"Recovering sparse signals using sparse measurement matrices in compressed DNA microarrays," Asilomor conference, November 2007.

[25] H. Hadwiger, "Gitterpunktanzahl im Simplex und Willssche Vermutung," Math. Ann. 239 (1979), 271-288.

[26] D. Donoho, "For most large underdetermined systems of linear equations the minimal l1-norm Solution is also the sparsest solution (2004) ," Comm. Pure Appl. Math 2004.

[27] `http://www.its.caltech.edu/~amin/weighted_l1_codes/`.

[28] A. Khajehnejad, W. Yu, A. S. Avestimehr and B. Hassibi, "Analyzing weighted ℓ_1 minimization for sparse recovery with nonuniform sparse models, full report.," arXiv:1009.3525.

[29] `http://www.dsp.ece.rice.edu/cs/`.

A Proof of Theorem 4.4

Let $\delta' = \delta_c(\gamma_1, \gamma_2, 1, p_2, \omega)$ and $\delta'' = \delta_c(0, 1, 0, p_2, 1)$. From Theorem 4.3 we know that:

$$\begin{aligned}
\delta' &= \min\{\delta \mid \psi'_{com}(0,\tau_2) - \psi'_{int}(0,\tau_2) - \psi'_{ext}(0,\tau_2) < 0 \ \forall\ 0 \le \tau_2 \le \gamma_2(1-p_2), \tau_2 > \delta - \gamma_1 - \gamma_2 p_2\}, \\
&= \gamma_2 \min\{\delta \mid \psi'_{com}(0,\gamma_2\tau_2) - \psi'_{int}(0,\gamma_2\tau_2) - \psi'_{ext}(0,\gamma_2\tau_2) < 0 \ \forall\ 0 \le \tau_2 \le 1 - p_2 \\
&\qquad , \tau_2 > \delta - p_2\} + \gamma_1, \qquad (50) \\
\delta'' &= \min\{\delta \mid \psi''_{com}(0,\tau_2) - \psi''_{int}(0,\tau_2) - \psi''_{ext}(0,\tau_2) < 0 \ \forall\ 0 \le \tau_2 \le 1 - p_2, \tau_2 > \delta - p_2\}, \qquad (51)
\end{aligned}$$

where the exponents ψ'_{com},ψ'_{int},ψ'_{ext},ψ''_{com},ψ''_{int} and ψ''_{ext} can be found using Theorem 4.3. From the definition of ψ_{com} in (6) it immediately follows that $\psi'_{com}(0, \gamma_2\tau_2) = \gamma_2\psi''_{com}(0,\tau_2)$. From (7), for $\omega \to \infty$ we know that $\psi'_{ext}(0,\gamma_2\tau_2) = c'{x'_0}^2 - \alpha'_2 \log G(\omega x'_0)$, and $\psi''_{ext}(0,\tau_2) = c'' {x''_0}^2 - \alpha''_2 \log G(x''_0)$. Following the details of derivations as in Theorem 4.3, we realize that: $c' = \gamma_2\omega^2 c''$, $\omega x'_0 = x''_0$, $\alpha'_2 = \gamma_2\alpha''_2$,, which implies $\psi'_{ext}(0,\gamma_2\tau_2) = \gamma_2\psi''_{ext}(0,\tau_2)$. Finally, from (8), we know that $\psi'_{int}(0,\gamma_2\tau_2) = (\Lambda^*(y') + \frac{\gamma_2\tau_2}{2\Omega'}y'^2 + \log 2)\gamma_2\tau_2$, $\psi''_{int}(0,\tau_2) = (\Lambda^*(y'') + \frac{\tau_2}{2\Omega''}y''^2 + \log 2)\tau_2$. Following the details of derivations as in Theorem 4.3, we realize that if $\omega \to \infty$, then $y' = y''$, $\Omega' = \gamma_2\Omega''$, which implies that $\psi'_{int}(0,\gamma_2\tau_2) = \gamma_2\psi''_{int}(0,\tau_2)$. From (50), (51) and the above conclusions, it follows that $\delta' = \gamma_2\delta'' + \gamma_1$.

Amin Khajehnejad received his Bachelors degree in Electrical Engineering from University of Tehran in 2007, and his Masters degree in Electrical Engineering from California Institute of Technology in 2009. He is currently a PhD candidate at Caltech. His general interests are in signal processing, coding, information theory and optimization. He has been with Lyric semiconductors Inc., and NEC laboratories America, Inc. summers of 2009 and 2010 respectively.

Weiyu Xu received the Bachelor degree in Information Engineering from Beijing University of Posts and Telecommunications in 2002. After getting a Master degree in Electronic Engineering from Tsinghua University, Beijing, he went on to pursue his graduate study in the Electrical Engineering Department at California Institute of Technology, Pasadena, California, beginning in September 2005. In August 2009, he obtained his Ph.D. degree in Electrical Engineering at California Institute of Technology, with a minor in Applied and Computational Mathematics. Since September 2009, he has been a postdoctoral associate at Cornell University, Ithaca, New York. His research interests are in networking science and systems, communication, control and signal processing, with an emphasis on detection and estimation, coding and information Theory He received the Charles and Ellen Wilts best dissertation prize in 2010.

Salman Avestimehr A. Salman Avestimehr is currently an assistant Professor at the School of Electrical and Computer Engineering at Cornell University. He received his Ph.D. in 2008 and M.S. degree in 2005 in Electrical Engineering and Computer Science, both from the University of California, Berkeley. Prior to that, he obtained his B.S. in Electrical Engineering from Sharif University of Technology in 2003. He was also a postdoctoral scholar at the Center for the Mathematics of Information (CMI) at Caltech in 2008. He has received a number of awards including the 2011 Young Investigator Program (YIP) award from the U. S. Air Force Office of Scientific Research (AFOSR), the NSF CAREER award (2010), the David J. Sakrison Memorial Prize from the U.C. Berkeley EECS Department (2008), and the Vodafone U.S. Foundation Fellows Initiative Research Merit Award (2005). His research interests include information theory, communications, and networking."

Babak Hassibi was born in Tehran, Iran, in 1967. He received the B.S. degree from the University of Tehran in 1989, and the M.S. and Ph.D. degrees from Stanford University in 1993 and

1996, respectively, all in electrical engineering.

He has been with the California Institute of Technology since January 2001, where he is currently Professor and Executive Officer of Elecrical Engineering. From October 1996 to October 1998 he was a research associate at the Information Systems Laboratory, Stanford University, and from November 1998 to December 2000 he was a Member of the Technical Staff in the Mathematical Sciences Research Center at Bell Laboratories, Murray Hill, NJ. He has also held short-term appointments at Ricoh California Research Center, the Indian Institute of Science, and Linkoping University, Sweden. His research interests include wireless communications and networks, robust estimation and control, adaptive signal processing and linear algebra. He is the coauthor of the books (both with A.H. Sayed and T. Kailath) *Indefinite Quadratic Estimation and Control: A Unified Approach to* H^2 *and* H^∞ *Theories* (New York: SIAM, 1999) and *Linear Estimation* (Englewood Cliffs, NJ: Prentice Hall, 2000). He is a recipient of an Alborz Foundation Fellowship, the 1999 O. Hugo Schuck best paper award of the American Automatic Control Council (with H. Hindi and S.P. Boyd), the 2002 National Science Foundation Career Award, the 2002 Okawa Foundation Research Grant for Information and Telecommunications, the 2003 David and Lucille Packard Fellowship for Science and Engineering and the 2003 Presidential Early Career Award for Scientists and Engineers (PECASE), and was a participant in the 2004 National Academy of Engineering "Frontiers in Engineering" program.

He has been a Guest Editor for the IEEE Transactions on Information Theory special issue on "space-time transmission, reception, coding and signal processing" was an Associate Editor for Communications of the IEEE Transactions on Information Theory during 2004-2006, and is currently an Editor for the Journal "Foundations and Trends in Information and Communication".