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

September 21, 2010

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 optimal weighted ℓ_1 minimization outperforms the regular ℓ_1 minimization substantially. We also generalize the results to an arbitrary number of classes.

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 [24]. The whole premise is that sparse signals (signals with many zero or negligible

---

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

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 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 [6], which considers compressed detection and estimation, [7], which studies Bayesian compressed sensing, and [8] 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 [10, 11].

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 [9]. 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. 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 [17] for an application of compressed sensing and nonlinear estimation in compressed network monitoring), DNA microarrays [19, 20, 21], astronomy, satellite imaging and many more practical examples.

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 [5]. 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. 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 [5]. This is in contrast to the recovery guarantees of [5] 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} ($1 \leq i \leq u$). Clearly, 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, 16]. 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. 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. [10]. 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 [18]. 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 [23].

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 $\mathbf{x}, \mathbf{w}, \mathbf{z}, \cdots$, scalars are shown by small regular letters $a, b, \alpha, \cdots$, and matrices are denoted by bold capital letters($\mathbf{A}, \mathbf{I}, \cdots$). For referring to geometrical objects and subspaces, we use Calligraphic notation, e.g. $\mathcal{Z}, \mathcal{F}, \mathcal{G}, \mathcal{P}, \mathcal{C}, \cdots$. 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($K, S, \cdots$). 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*
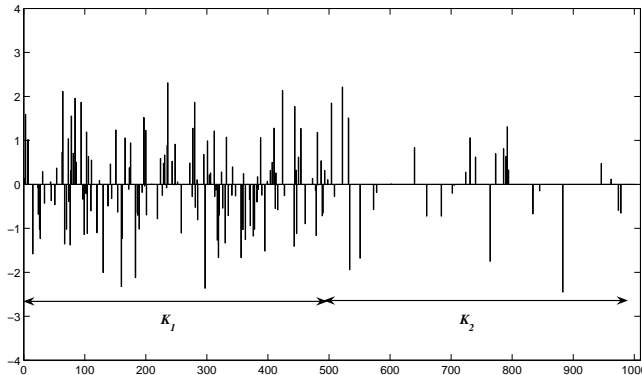
Figure 1: Illustration of a nonuniformly sparse signal.

$\mathbf{x} = (x_1, x_2, \cdots, x_n)^T$ *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:*

- *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 symmetric 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 ($O(n^3)$).

5

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:

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

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 [16] 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 $\min_{\mathbf{A}\mathbf{x}=\mathbf{y}} \sum_{i=1}^{n} w_i |x_i|$ with $\mathbf{y} = \mathbf{A}\mathbf{x}^*$, if and only if for every vector $\mathbf{z} = (z_1, z_2, \cdots, z_n)^T$ in the null space of $\mathbf{A}$, the following holds: $\sum_{i \in K} w_i |z_i| \geq \sum_{i \in \overline{K}} w_i |z_i|$.*

This theorem will be proved in Section 5. 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 $\mathcal{P}_{\mathbf{w}} = \{\mathbf{y} \in \mathbb{R}^n \mid \sum_{i=1}^{n} w_i |y_i| \leq 1\}$ 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 [14, 15], 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)

7

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*

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

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, for small enough ϵ, 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 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 5.2.3. 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

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

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)*

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

   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

$$\psi_{ext}(\tau_1, \tau_2) = cx_0^2 - \alpha_1 \log G(x_0) - \alpha_2 \log G(\omega x_0) \qquad (7)$$

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:*

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

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 ($u \ge 2$) will be given in Section 5.3.

9
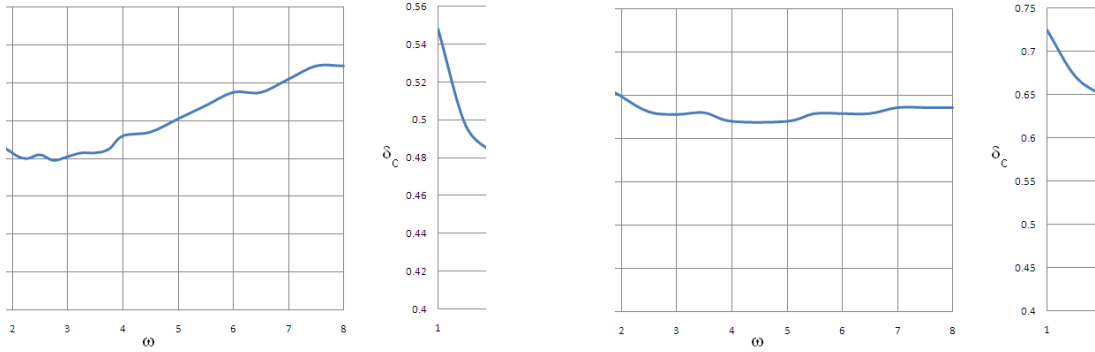
(a) $p_1 = 0.4$, $p_2 = 0.05$.
(b) $p_1 = 0.65$, $p_2 = 0.1$.

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

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.

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, 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 5.2.3.

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. [5], with the motivation that in some applications (or due to previous processing steps), part of the support set can be fully identified[2]. If the dense subclass is K_1 and $K_2 = K_1^c$, then [5] suggests solving the following minimization program:

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

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 $\omega \to \infty$

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

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 $\delta_c(\gamma_1, \gamma_2, 1, p_2, \omega) \to \gamma_1 + \gamma_2 \delta_c(0, 1, 0, p_2, 1)$. In other words, when a subset of entries of size $\gamma_1 n$ are known to be nonzero, the minimum number of measurements that is required for almost surely successful recovery using (11) is equal to the total number of measurements needed if we were allowed to independently make measurements from the two parts and recover each using ℓ_1 minimization.*

The proof of this theorem is given in Appendix E.

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*

---

[2]Thanks to anonymous reviewers for pointing this out to us!

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

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

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. 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 [16] 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 $\mathbf{W} = diag(w_1, w_2, \cdots, w_n)$. By triangular inequality, we have:

$$\|\mathbf{W\hat{x}}\|_1 = \|(\mathbf{W\hat{x}})_K\|_1 + \|(\mathbf{W\hat{x}})_{\overline{K}}\|_1 \quad = \quad \|(\mathbf{Wx^*} + \mathbf{W\hat{x}} - \mathbf{Wx^*})_K\|_1 + \|(\mathbf{W\hat{x}})_{\overline{K}}\|_1$$

$$\geq \quad \|(\mathbf{Wx^*})_K\|_1 - \|(\mathbf{W\hat{x}} - \mathbf{Wx^*})_K\|_1 + \|(\mathbf{W\hat{x}} - \mathbf{Wx^*})_{\overline{K}}\|_1$$

$$\geq \quad \|\mathbf{Wx^*}\|_1$$

Where the last inequality is a result of the fact that $\hat{\mathbf{x}} - \mathbf{x}^*$ is 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 $\|\mathbf{W\hat{x}}\|_1 \leq \|\mathbf{Wx^*}\|_1$. This implies that $\hat{\mathbf{x}} = \mathbf{x}^*$. Conversely, suppose there is some vector $\mathbf{z}$ in $\mathcal{N}(\mathbf{A})$ such that $\|(\mathbf{Wz})_K\|_1 > \|(\mathbf{Wz})_{\overline{K}}\|_1$. Taking define $\mathbf{x}^* = (\mathbf{z}_K \ 0)^T$ and $\hat{\mathbf{x}} = (0 \ \mathbf{z}_{\overline{K}})^T$ implies that $\mathbf{Ax^*} = \mathbf{A\hat{x}}$ and $\|\mathbf{Wx^*}\|_1 > \|\mathbf{W\hat{x}}\|_1$. 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 which will be proven in Appendix A.

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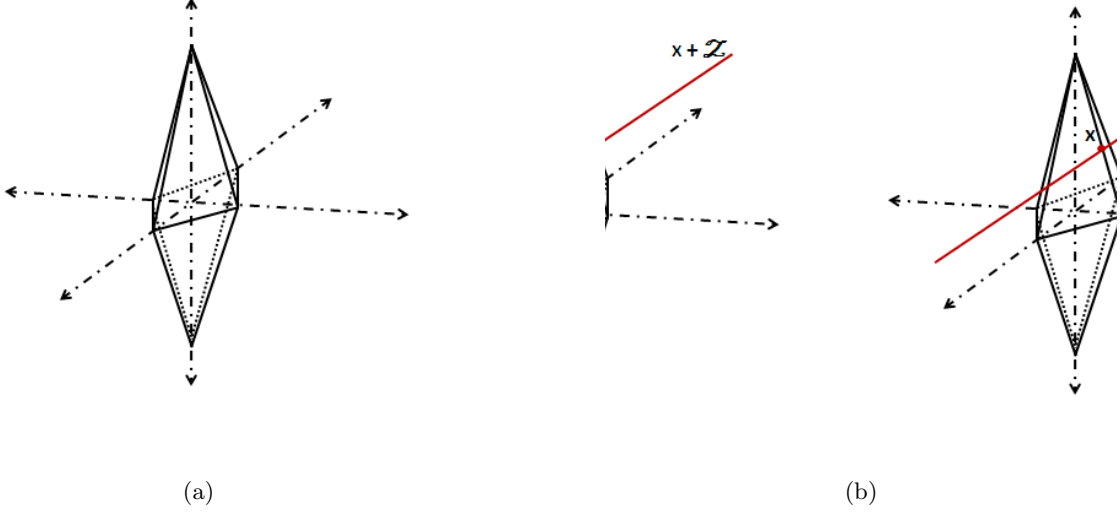
Figure 3: A weighted ℓ_1-ball, $\mathcal{P}_\mathbf{w}$, in $\mathbb{R}^3$ (a), and a linear hyperplane $\mathcal{Z}$ passing through a point $\mathbf{x}$ in the interior of a one dimensional face of $\mathcal{P}_\mathbf{w}$ (b).

any convex ball that surrounds the origin. Therefore we restrict our attention to those vectors $\mathbf{x}$ from 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. Since we are assuming that the distribution of the nonzero entries of $\mathbf{x}$ is symmetric, 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 $\mathcal{P}_\mathbf{w}$:

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

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 3a 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 3b). 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 [12, 13], 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** [12] 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. Santaló [14] and P. McMullen [15] 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 [13]:

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

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 [13, 15]:

- 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 $\mathcal{F}_1$. [3] 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 $\mathcal{F}_1 = \mathcal{F}_2$.

- 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 $\mathcal{F}_3 = \mathcal{F}_4$.

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 [4] if $|K \cap K_1| = |K' \cap K_1|$ and $|K \cap K_2| = |K' \cap K_2|$[5]. In other words the only thing that distinguishes the morphology of the faces of $\mathcal{P}_{\mathbf{w}}$ is the

---

[3]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.

[4]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. $\mathcal{F}' = \Theta \mathcal{F}$.

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

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 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 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

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

(24)

where 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 5.2.3, and derive thresholds for the negativity of the cumulative exponent.

5.2.1 Computation of Internal Angle

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

$$
Z = (k_1 w_{K_1}^2 + k_2 w_{K_2}^2) X_1 - w_{K_1}^2 \sum_{i=1}^{t_1} X_1' - w_{K_2}^2 \sum_{i=1}^{t_2} X_1'',
$$

where $X_1 \sim N(0, \frac{1}{2(k_1 w_{K_1}^2 + k_2 w_{K_2}^2)})$ 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

$$\beta(k_1, k_2 | t_1, t_2) = c_0 p_Z(0) \tag{25}$$

We now prove this Theorem. Suppose that $\mathcal{F}$ is a $(k-1)$-dimensional face of the weighted ℓ_1-ball

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

supported on the subset K with $|K| = k = k_1 + k_2$. Let $\mathcal{G}$ be a $l-1$ dimensional face of $\mathcal{P}_{\mathbf{w}}$ supported on the set L with $\mathcal{F} \subset \mathcal{G}$. Also, let $|L \cap K_1| = k_1 + t_1$ and $|L \cap K_2| = k_2 + t_2$.

We first state the following lemma the proof of which is given in Appendix B.

Lemma 5.2. *Let $\mathcal{F}$ be a $(k-1)$-dimensional face of $\mathcal{P}_{\mathbf{w}}$ supported on the set $K = \{1, 2, \cdots, k\}$, and $\mathcal{G}$ be a $l-1$-dimensional face of $\mathcal{P}_{\mathbf{w}}$ that contains $\mathcal{F}$ and is supported on the set $L = \{1, 2, \cdots, l\}$. Let $\mathcal{C}_{\mathcal{F}^\perp, \mathcal{G}}$ be the positive cone of all the vectors $\mathbf{x} \in \mathbb{R}^n$ that take the form:*

$$-\sum_{i=1}^{k} b_i e_i + \sum_{i=k+1}^{l} b_i e_i, \tag{26}$$

where $b_i, 1 \leq i \leq l$ are nonnegative real numbers and

$$\sum_{i=1}^{k} w_i b_i = \sum_{i=k+1}^{l} w_i b_i, \quad \frac{b_1}{w_1} = \frac{b_2}{w_2} = \cdots = \frac{b_k}{w_k}.$$

Then

$$\int_{\mathcal{C}_{\mathcal{F}^\perp, \mathcal{G}}} e^{-\|\mathbf{x}\|^2} d\mathbf{x} = \beta(\mathcal{F}, \mathcal{G}) \cdot \pi^{(l-k)/2}. \tag{27}$$

From (27) we can find the expression for the internal angle. Define $U \subseteq \mathbb{R}^{l-k+1}$ as the set of all nonnegative vectors $(x_1, x_2, \cdots, x_{l-k+1})$ satisfying:

$$(\textstyle\sum_{r=1}^{k} w_r^2) x_1 = \sum_{r=k+1}^{l} w_r^2 x_{r-k+1}$$

and define $f(x_1, \cdots, x_{l-k+1}) : U \to \mathcal{C}_{F^\perp, G}$ to be the following linear and bijective map:

$$f(x_1, \cdots, x_{l-k+1}) = -\sum_{r=1}^{k} x_1 w_r \mathbf{e}_r + \sum_{r=k+1}^{l} x_{r-k+1} w_r \mathbf{e}_r.$$

Then

$$\int_{\mathcal{C}_{\mathcal{F}^\perp,\mathcal{G}}} e^{-\|\mathbf{x}'\|^2} d\mathbf{x}' = \int_U e^{-\|f(\mathbf{x})\|^2} df(\mathbf{x}) = |J(\mathbf{M})| \int_\Gamma e^{-\|f(x)\|^2} dx_2 \cdots dx_{l-k+1}$$

$$= |J(\mathbf{M})| \int_\Gamma e^{-(\sum_{r=1}^k w_r^2)x_1^2 - \sum_{r=k+1}^l w_r^2 x_{r-k+1}^2} dx_2 \cdots dx_{l-k+1} \tag{28}$$

Γ is the region described by

$$(\sum_{r=1}^k w_r^2)x_1 = \sum_{r=k+1}^l w_r^2 x_{r-k+1}, \ x_r \geq 0 \ \ 2 \leq r \leq l - k + 1 \tag{29}$$

where $|J(\mathbf{M})|$ is due to the change of integral variables and is essentially the determinant of the Jacobian of the variable transform given by the $l \times (l-k)$ matrix $\mathbf{M}$ below:

$$\mathbf{M}_{i,j} = \begin{cases} -\frac{1}{\Omega} w_i w_{k+j}^2 & 1 \leq i \leq k, 1 \leq j \leq l - k \\ w_i & k + 1 \leq i \leq l, j = i - k \\ 0 & \text{Otherwise} \end{cases} \tag{30}$$

where $\Omega = \sum_{r=1}^k w_r^2$. The Jacobian is obtained by $|J(\mathbf{M})| = \det(\mathbf{M}^T\mathbf{M})^{1/2}$. By finding the eigenvalues of $\mathbf{M}^T\mathbf{M}$ we obtain:

$$|J(\mathbf{M})| = w_{K_1}^{t_1} w_{K_2}^{t_2} (\frac{\Omega + t_1 w_{K_1}^2 + t_2 w_{K_2}^2}{\Omega})^{1/2} \tag{31}$$

Now we define a random variable

$$Z = (\sum_{r=1}^k w_r^2)X_1 - \sum_{r=k+1}^l w_r^2 X_{r-k+1}$$

where $X_1, X_2, \cdots, X_{l-k+1}$ are independent random variables, with $X_r \sim HN(0, \frac{1}{2w_{r+k-1}^2})$, $2 \leq r \leq (l-k+1)$, are half-normal distributed random variables and $X_1 \sim N(0, \frac{1}{2\sum_{r=1}^k w_r^2})$ is a normal distributed random variable. Then by inspection, (28) is equal to $c_1 p_Z(0)$, where $p_Z(\cdot)$ is the probability density function for the random variable Z and $p_Z(0)$ is the probability density function $p_Z(\cdot)$ evaluated at the point $Z = 0$, and

$$c_1 = \frac{\sqrt{\pi}^{l-k+1}}{2^{l-k}} \prod_{q=k+1}^l \frac{1}{w_q} (\sum_{r=1}^k w_r^2)^{1/2} |J(A)| = \frac{\sqrt{\pi}^{l-k+1}}{2^{l-k}} ((k_1 + t_1)w_{K_1}^2 + (k_2 + t_2)w_{K_2}^2)^{1/2} \tag{32}$$

Combining these results, the proof of Theorem 5.1 is complete.

5.2.2 Computation of External Angle

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 $\mathcal{G}$ is supported on the set L with $|L \cap K_1| = d_1$ and $|L \cap K_2| = d_2$ is given by:*

$$\zeta(d_1, d_2) = \pi^{-\frac{n-l+1}{2}} 2^{n-l} \int_0^\infty e^{-x^2} \left(\int_0^{\frac{w_{K_1} x}{\xi(d_1, d_2)}} e^{-y^2} \, dy \right)^{r_1} \left(\int_0^{\frac{w_{K_2} x}{\xi(d_1, d_2)}} e^{-y^2} \, dy \right)^{r_2} dx, \tag{33}$$

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 $r_2 = n_2 - d_2$.

Proof. Without loss of generality, assume that the support set of $\mathcal{G}$ is given by $L = \{n - l + 1, n - l + 2, \cdots, n\}$ and consider the $(l-1)$-dimensional face

$$\mathcal{G} = \mathrm{conv}\{\frac{\mathbf{e}_{n-l+1}}{w_{n-l+1}}, ..., \frac{\mathbf{e}_{n-k}}{w_{n-k}}, \frac{\mathbf{e}_{n-k+1}}{w_{n-k+1}}, ..., \frac{\mathbf{e}_n}{w_n}\}$$

of the weighted ℓ_1-ball $\mathcal{P}$. The 2^{n-l} outward normal vectors of the supporting hyperplanes of the facets containing $\mathcal{G}$ are given by

$$\{\sum_{i=1}^{n-l} j_i w_i \mathbf{e}_i + \sum_{p=n-l+1}^{n} w_i \mathbf{e}_i, j_i \in \{-1, 1\}\}.$$

Then the outward normal cone $\mathcal{C}_{\mathcal{G}, \mathcal{P}_\mathbf{w}}^\perp$ at the face $\mathcal{G}$ is the positive hull of these normal vectors. Thus

$$\int_{\mathcal{C}_{\mathcal{G}, \mathcal{P}_\mathbf{w}}^\perp} e^{-\|x\|^2} \, dx = \zeta(\mathcal{G}, \mathcal{P}_\mathbf{w}) V_{n-l}(S^{n-l}) \int_0^\infty e^{-r^2} r^{n-l} \, dx$$

$$= \zeta(\mathcal{G}, \mathcal{P}_\mathbf{w}).\pi^{(n-l+1)/2}, \tag{34}$$

where $V_{n-l}(S^{n-l})$ is the spherical volume of the $(n-l)$-dimensional unit sphere S^{n-l}. Now define U to be the set

$$\{x \in R^{n-l+1} \mid x_{n-l+1} \geq 0, |x_i/w_i| \leq x_{n-l+1}, 1 \leq i \leq (n-l)\}$$

and define $f(x_1, \cdots, x_{n-l+1}) : U \to \mathcal{C}_{\mathcal{G}, \mathcal{P}_\mathbf{w}}^\perp$ to be the linear and bijective map

$$f(x_1, \cdots, x_{n-l+1}) = \sum_{i=1}^{n-l} x_i \mathbf{e}_i + \sum_{i=n-l+1}^{n} w_i x_{n-l+1} \mathbf{e}_i.$$

Then

$$\int_{\mathcal{C}^{\perp}_{\mathcal{G},\mathcal{P}_{\mathbf{w}}}} e^{-\|x'\|^2}\, dx' = |J(\mathbf{M})| \int_U e^{-\|f(x)\|^2}\, dx$$

$$= |J(\mathbf{M})| \int_0^\infty \int_{-w_{K_1} x_{n-l+1}}^{w_{K_1} x_{n-l+1}} \cdots \int_{-w_{n-l} x_{n-l+1}}^{w_{n-l} x_{n-l+1}} e^{-x_1^2 - \cdots - x_{n-l}^2 - (\sum_{i=n-l+1}^n w_i^2) x_{n-l+1}^2}\, dx_1 \cdots dx_{n-l+1}$$

$$= |J(\mathbf{M})| \int_0^\infty e^{-(\sum_{i=n-l+1}^n w_i^2) x^2} \left(\int_{-w_{K_1} x}^{w_{K_1} x} e^{-y^2}\, dy \right)^{n_1 - d_1} \left(\int_{-w_{K_2} x}^{w_{K_2} x} e^{-y^2}\, dy \right)^{n_2 - d_2} dx \qquad (35)$$

$\mathbf{M}$ is the $n \times (n - l + 1)$ change of variable matrix given by $\mathbf{M} = \begin{pmatrix} \mathbf{I_{n-l}} & 0 \\ 0 & \mathbf{w}_L \end{pmatrix}$, where $\mathbf{w}_L = (w_{n-l+1}, w_{n-l+2}, \cdots, w_n)^T$. Therefore $J(\mathbf{M}) = \det(\mathbf{M}^T \mathbf{M})^{1/2} = (d_1 w_{K_1}^2 + d_2 w_{K_2}^2)^{1/2}$. Replacing this and a change of variable for x (replace ξx with x) in (35), along with (34), complete the proof. ∎

5.2.3 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

$$\mathbb{P}\{E^c\} \leq \sum_{\substack{0 \leq t_1 \leq n_1 - k_1 \\ 0 \leq t_2 \leq n_2 - k_2 \\ t_1 + t_2 > m - k_1 - k_2 + 1}} 2^{t_1 + t_2 + 1} \binom{n_1 - k_1}{t_1} \binom{n_2 - k_2}{t_2} \beta(k_1, k_2 | t_1, t_2) \zeta(t_1 + k_1, t_2 + k_2), \qquad (36)$$

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 Sections 5.2.1 and 5.2.2, respectively. Now our objective is to show that the R.H.S of (36) 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 do this end we bound the exponents of the combinatorial, internal angle and external angle terms in (36), 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 $\epsilon \to 0$

$$\frac{1}{n} \log \left(2^{t_1 + t_2 + 1} \binom{n_1 - k_1}{t_1} \binom{n_2 - k_2}{t_2} \right) \to \left(\gamma_1(1 - p_1) H\left(\frac{\tau_1}{\gamma_1(1 - p_1)} \right) + \gamma_2(1 - p_2) H\left(\frac{\tau_2}{\gamma_2(1 - p_2)} \right) + \tau_1 + \tau_2 \right) \log 2,$$
$$(37)$$

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:

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

21

Define

$$\psi_{ext}(\tau_1, \tau_2) = cx_0^2 - \alpha_1 \log G(x_0) - \alpha_2 \log G(\omega x_0) \tag{38}$$

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:

$$\psi_{int}(\tau_1, \tau_2) = \left(\Lambda^*(y) + \frac{\tau_1 + \tau_2}{2\Omega'}y^2 + \log 2\right)(\tau_1 + \tau_2). \tag{39}$$

We now state the following lemmas, which are proved in Appendix C and D.

Lemma 5.3. *Fix $\delta, \epsilon > 0$. There exists a finite number $n_0(\delta, \epsilon)$ such that*

$$\frac{1}{n}\log(\zeta(t_1 + k_1, t_2 + k_2)) < -\psi_{ext}(\tau_1, \tau_2) + \epsilon, \tag{40}$$

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

Lemma 5.4. *Fix $\delta, \epsilon > 0$. There exists a finite number $n_1(\delta, \epsilon)$ such that*

$$\frac{1}{n}\log(\beta(t_1, t_2|k_1, k_2)) < -\psi_{int}(\tau_1, \tau_2) + \epsilon, \tag{41}$$

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

Combining Lemmas 5.3 and 5.4, (37), and the bound in (36) 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:

$$\binom{n_1}{k_1}\binom{n_2}{k_2}2^k P_{K,-} \tag{42}$$

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

$$(\gamma_1 p_1 + \gamma_2 p_2 + \gamma_1 H(p_1) + \gamma_2 H(p_2))\log 2, \tag{43}$$

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:

$$(\gamma_1 H(p_1) + \gamma_2 H(p_2)) \log 2. \tag{44}$$

5.3 Generalizations

Except for some subtlety in the large deviation calculations, the generalization of the results of the previous section to an arbitrary $u \geq 2$ 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:

$$\mathbb{P}\{E^c\} \leq 2 \sum_{\substack{0 \,\leq\, \mathbf{t} \,\leq\, \mathbf{n} - \mathbf{k} \\ \mathbf{1}^T \mathbf{t} \,>\, m - \mathbf{1}^T \mathbf{k} + 1}} \Pi_{1 \leq i \leq u} 2^{t_i} \binom{n_i - k_i}{t_i} \beta(\mathbf{k}|\mathbf{t}) \zeta(\mathbf{t} + \mathbf{k}) \tag{45}$$

Where $\mathbf{t} = (t_1, \cdots, t_u)^T$, $\mathbf{k} = (k_1, \cdots, k_u)^T$ 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 $K_1, \cdots, K_u$ with $|K_i| = n_i = \gamma_1 n$, and sparsity fractions $p_1, p_2, \cdots, p_u$, 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:*

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

is given by the following expression:

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

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 $y = s^*(b - \frac{1}{M(s^*)})$. 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 5.5. *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*

$$\|\mathbf{W}(\mathbf{x}_0 - \hat{\mathbf{x}})\|_1 \leq 2\frac{C+1}{C-1} \sum_{i \in K} w_i |(x_0)_i|, \tag{46}$$

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

$$C \sum_{i \in K} w_i |z_i| \leq \sum_{i \in \overline{K}} w_i |z_i|. \tag{47}$$

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

$$C' \sum_{i \in L_1 \cup L_2} w_i |z_i| = \sum_{i \in \overline{L_1} \cap \overline{L_2}} w_i |z_i|. \tag{48}$$

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

$$K_{\epsilon_1} : \max_{K_{\epsilon_1} \subset K_1 \cap \overline{L_1}, |K_{\epsilon_1}| = \epsilon_1 \gamma_1 p_1 n} \sum_{i \in K_{\epsilon_1}} w_i |z_i|, \tag{49}$$

$$K_{\epsilon_2} : \max_{K_{\epsilon_2} \subset K_2 \cap \overline{L_2}, |K_{\epsilon_2}| = \epsilon_2 \gamma_2 p_2 n} \sum_{i \in K_{\epsilon_2}} w_i |z_i|. \tag{50}$$

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

$$\sum_{i \in K_{\epsilon_1}} w_i |z_i| \geq \frac{\epsilon_1 p_1}{1 - p_1} \sum_{i \in \overline{L_1'} \cap K_1} w_i |z_i|, \tag{51}$$

$$\sum_{K_{i \in \epsilon_2}} w_i |z_i| \geq \frac{\epsilon_2 p_2}{1 - p_2} \sum_{i \in \overline{L_2'} \cap K_2} w_i |z_i|. \tag{52}$$

24

Adding $C'\left(\sum_{K_{\epsilon_1}} w_i|z_i| + \sum_{K_{\epsilon_2}} w_i|z_i|\right)$ to both sides of (48) and using (51) and (52), we can write:

$$C'\sum_{i\in L_1'\cup L_2'} w_i|z_i| \geq \sum_{i\in\overline{L_1}\cap\overline{L_2}} w_i|z_i| + C'\left(\frac{\epsilon_1 p_1}{1-p_1}\sum_{i\in\overline{L_1'}\cap K_1} w_i|z_i| + \frac{\epsilon_2 p_2}{1-p_2}\sum_{i\in\overline{L_2'}\cap K_2} w_i|z_i|\right) \quad (53)$$

$$\geq \left(1 + (C'+1)\min(\frac{\epsilon_1 p_1}{1-p_1}, \frac{\epsilon_2 p_2}{1-p_2})\right)\sum_{i\in\overline{L_1'}\cap\overline{L_2'}} w_i|z_i|. \quad (54)$$

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 (54) we conclude that

$$C' \geq \left(1 + (C'+1)\min(\frac{\epsilon_1 p_1}{1-p_1}, \frac{\epsilon_2 p_2}{1-p_2})\right), \quad (55)$$

or equivalently

$$C' \geq \frac{1 + \min(\frac{\epsilon_1 p_1}{1-p_1}, \frac{\epsilon_2 p_2}{1-p_2})}{1 - \min(\frac{\epsilon_1 p_1}{1-p_1}, \frac{\epsilon_2 p_2}{1-p_2})}. \quad (56)$$

Using Lemma 5.5 and the above inequality, we conclude (14). ∎

6   Approximate Support Recovery and Reweighted ℓ_1

Using the analytical tools of this paper, it is possible to prove that a class of reweighted ℓ_1 minimization algorithms 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 is not brought here, since it stands beyond the scope of this paper. However, we briefly mention how a simple post processing on the output of ℓ_1 minimization results in a nonuniform sparsity model with $u = 2$ classes close to the one we introduced for the unknown signal. A more comprehensive study on this can be found in [11]. The reweighted ℓ_1 recovery algorithm proposed in [11] 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. The algorithm is as follows:

Algorithm 1.

1. *Solve the ℓ_1 minimization problem:*

$$\hat{\mathbf{x}} = \arg\min \|\mathbf{z}\|_1 \quad \text{subject to} \quad \mathbf{Az} = \mathbf{Ax}. \tag{57}$$

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:*

$$\mathbf{x}^* = \arg\min \|\mathbf{z}_L\|_1 + \omega\|\mathbf{z}_{\overline{L}}\|_1 \quad \text{subject to} \quad \mathbf{Az} = \mathbf{Ax}. \tag{58}$$

For a given number of measurements, if the support size of $\mathbf{x}$, namely $k = |K|$, is slightly larger than the sparsity threshold of ℓ_1 minimization, then a so-called robustness of ℓ_1 minimization helps find a lower bound f_1 for $\frac{|L \cap K|}{|L|}$, i.e. the sparsity fraction of $\mathbf{x}$ over the set L. If f_1 is sufficiently close to 1, the number of measurements could satisfy:

$$\delta \geq \max_{f_1' \geq f_1, f_1'k + f_2'(n-k)=k} \delta_c^{(T)}(\frac{k}{n}, 1 - \frac{k}{n}, f_1', f_2', \omega). \tag{59}$$

Then the recovery is successful in the second step with high probability. Recall that $\delta_c^{(T)}$ is the sectional threshold, which accounts for all possible support sets. Therefore, the condition for strict improvement in the reweighted ℓ_1 minimization is that:

$$\delta(0, 1, 0, \frac{k}{n}, 1) \geq \max_{f_1' \geq f_1, f_1'k + f_2'(n-k)=k} \delta_c^{(T)}(\frac{k}{n}, 1 - \frac{k}{n}, f_1', f_2', \omega). \tag{60}$$

7 Simulation Results

We demonstrate by some examples that appropriate weights can boost the recovery percentage. In Figure 4 we have shown the empirical recovery threshold of weighted ℓ_1 minimization for different values of the 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 5a 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 5b 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 5a for each single value of p_1. Figure 6 shows the result of simulations in another setting where $p_2 = 0.1$ and $m = 0.75n$ (similar to the setting of Section
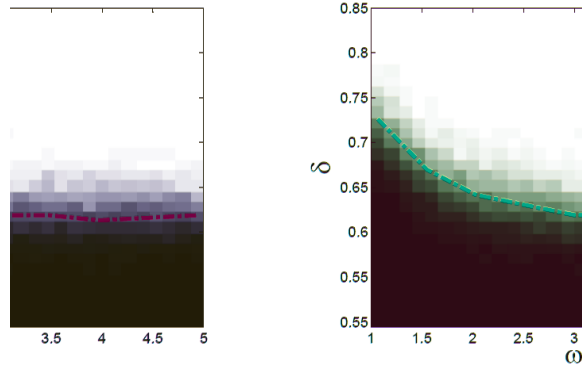
(a) $\gamma_1 = \gamma_2 = 0.5$, $p_1 = 0.4$ and $p_2 = 0.05$.   (b) $\gamma_1 = \gamma_2 = 0.5$, $p_1 = 0.65$ and $p_2 = 0.1$.

Figure 4: 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..

4). Note that these results very well match the theoretical results of Figures 2a and 2b.

In Figure 7, 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 $\gamma_1 = \gamma_2 = 0.5$. However, a white Gaussian noise vector is added before compression. Figure 7 shows a scatter plot of all output signal to recovery error ratios as a function of the input SNR, for all simulations. In Figure 8 the average curves are compared together for different values of weight ω.

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 9) taken at two different years, 1989 (left) and 2000 (right), from the New Britain rainforest in Papua Guina. 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 two subframes, identified by green and red rectangles in Figure 9. In addition, a weighted ℓ_1 minimization is also applied where the frame pixels are divided 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. For the right frame (red), the two classes are bottom half and top half of the frame, and for the left frame (green), they are left half and right half. We casually assign the weight value $\omega = 2$ for the sparser region for weighted ℓ_1 recovery, and unitary weight to the denser region. The recovery errors for the two methods are displayed in Figure 10. The error is averaged over 50 realizations of i.i.d. Gaussian measurement matrix for each δ. As can
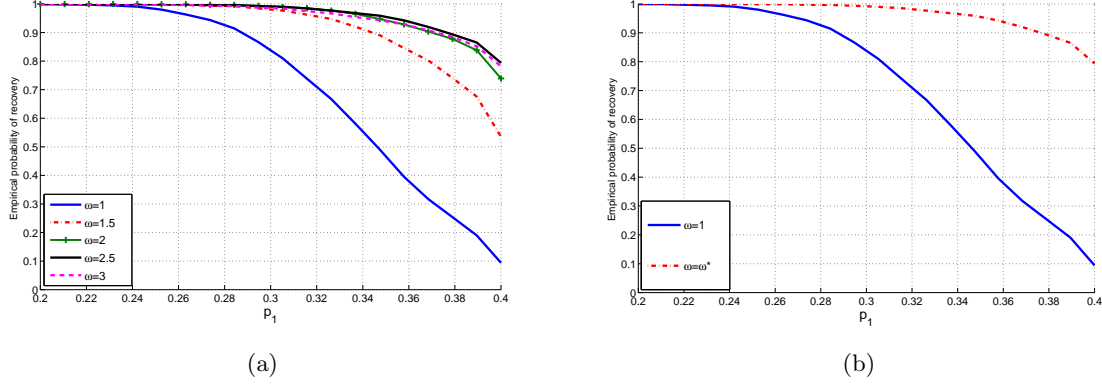
27

Figure 5: 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).
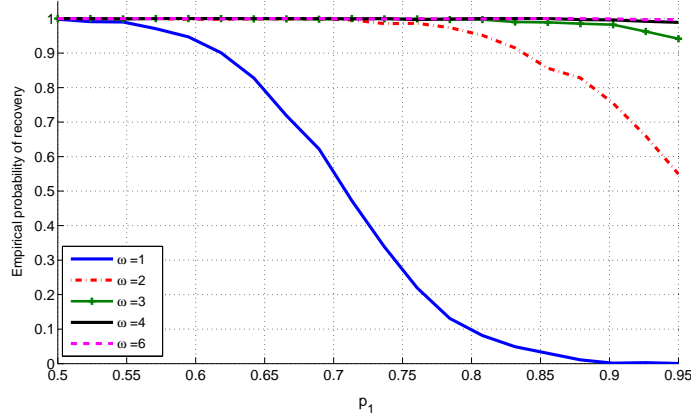

Figure 6: Empirical probability of successful recovery for different weights. $p_2 = 0.1, \gamma_1 = \gamma_2 = 0.5$ and $m = 0.75n = 150$.

be seen, even with this value of weight chosen intuitively, the recovery improvement is significant in the weighted ℓ_1 minimization.

In figure 11, we have compared the recovery performance for the regular ℓ_1 minimization and the reweighted ℓ_1 minimization of Algorithm 1, for different sparsity levels and different distributions for the nonzero entries. Here the signal dimension is $n = 200$, and the number of measurements is $m = 112$, which corresponds to a value of $\delta = 0.5555$. We generated random sparse signals with iid entries coming from certain distributions; Gaussian, uniform, Rayleigh , square root of χ-square with 4 degrees of freedom and, square root of χ-square with 6 degrees of freedom. Solid lines represent the simulation results for ordinary ℓ_1 minimization, and different colors indicate different distributions. Dashed lines are used to show the results for Algorithm 1. The reason why these distributions are selected and compared is elaborated in [11], as they demonstrate various levels of improvement. Note that for Gaussian and uniform distributions
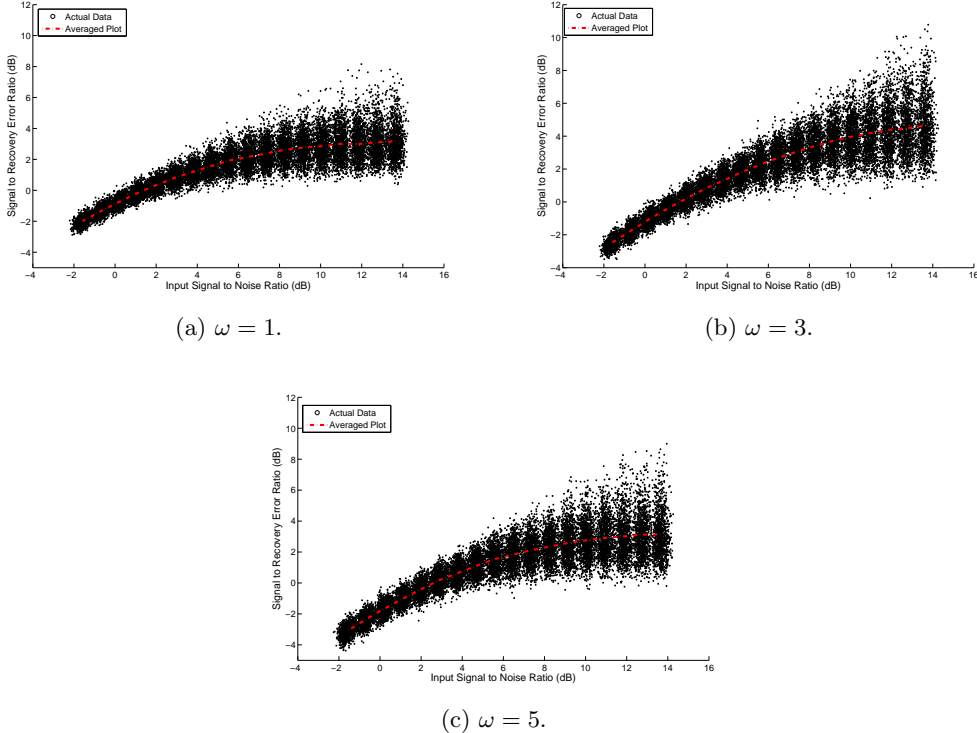
28

Figure 7: 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.

that are flat and nonzero at the origin, the reweighted algorithm shows an impressive more than 20% improvement in the weak threshold (from 45 to 55).

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. Some of our simulations were performed on real world data of satellite images, where the nonuniform sparse model is a valid assumption. 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 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) [11, 18]. 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
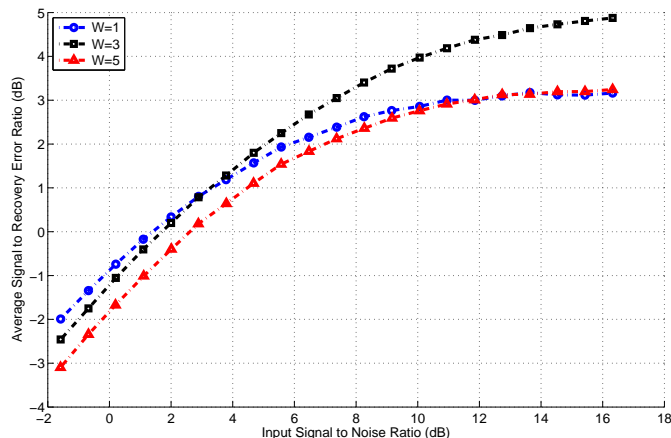
29

Figure 8: 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 9: Satellite images taken from the New Britain rainforest in Papua Guina at 1989 (left) and 2000 (right). Image originally belongs to Royal Society for the Protection of Birds and was taken from the Guardian archive, an article on deforestation..

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] N. Vaswani and W. Lu, "Modified-CS: Modifying Compressive Sensing for Problems with Partially Known Support," To Appear in IEEE Trans. Signal Processing.

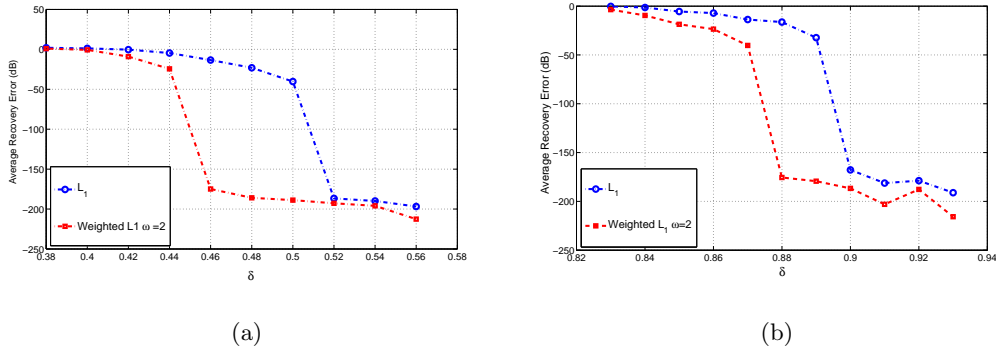(a)   (b)

Figure 10: Average Recovery Error for ℓ_1 and weighted ℓ_1 minimization recovery of the difference between the subframes of a pair of satellite images shown in Figure 9. (a) corresponds to the right(red) subframe in Figure 9, and (b) corresponds to the left(green) frame. Data is averaged over 50 realizations of measurement matrices for each δ.
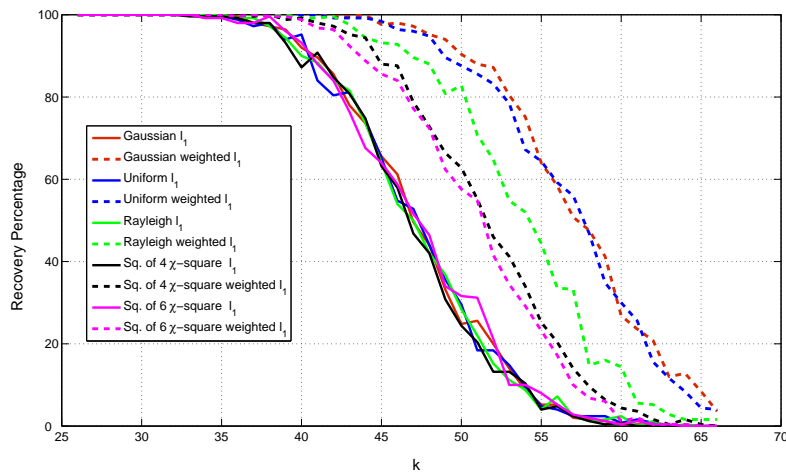

Figure 11: Empirical Recovery Percentage for Algorithm 1 with $n = 200$ and $\delta = 0.5555$.

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

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

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

[9] M. Stojnic, F. Parvaresh and B. Hassibi, "On the reconstruction of block-sparse signals with an optimal number of measurements," Preprint 2008.

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

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

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

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

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

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

[16] 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.

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

[18] 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.

[19] 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.

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

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

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

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

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

Appendix. Proof of Important Lemmas

A   Proof of Lemma 5.1

First, let us assume that $\sum_{i \in K} w_i |z_i| \leq \sum_{i \in \overline{K}} w_i |z_i|, \forall \mathbf{z} \in \mathcal{N}(A)$. Note that by assumption w_is are all nonnegative. Using the triangular inequality for the weighted ℓ_1 norm (or for each absolute value term

on the LHS) we obtain

$$\sum_{i \in K} w_i |x_i + z_i| + \sum_{i \in \overline{K}} w_i |z_i| \geq \sum_{i \in K} w_i |x_i| - \sum_{i \in K} w_i |z_i| + \sum_{i \in \overline{K}} w_i |z_i|$$

$$\geq \sum_{i \in K} w_i |x_i|.$$

thus proving the forward part of this lemma. Now let us assume instead that $\exists \mathbf{z} \in \mathcal{N}(A)$, such that $\sum_{i \in K} w_i |z_i| > \sum_{i \in \overline{K}} w_i |z_i|$. Then we can construct a vector $\mathbf{x}$ supported on the set K (or a subset of K), with $\mathbf{x}_K = -\mathbf{z}_K$ (i.e. $x_i = -z_i \; \forall i \in K$). Then we have

$$\sum_{i \in K} w_i |x_i + z_i| + \sum_{i \in \overline{K}} w_i |z_i| = 0 + \sum_{i \in \overline{K}} w_i |z_i| < \sum_{i \in K} w_i |x_i|$$

proving the reverse part of this lemma.

B   Proof of Lemma 5.2

Without loss of generality, assume that $\mathcal{F}$ has the following k vertices: $\{\frac{\mathbf{e}_r}{w_r}, 1 \leq r \leq k\}$, where $\mathbf{e}_r$ is the n-dimensional standard unit vector with the r-th element equal to 1. Also assume that the $(l-1)$-dimensional face G is the convex hull of the following l vertices: $\frac{\mathbf{e}_r}{w_r}, 1 \leq r \leq l$. Then the cone $\mathcal{C}_{\mathcal{F},\mathcal{G}}$ formed by observing the $(l-1)$-dimensional face $\mathcal{G}$ of the weighted ℓ_1-ball $\mathcal{P}_{\mathbf{w}}$ from an interior point $x^{\mathcal{F}}$ of the face $\mathcal{F}$ is the positive cone of the vectors:

$$\frac{\mathbf{e}_j}{w_j} - \frac{\mathbf{e}_i}{w_i}, \quad \text{for all } j \in J \backslash K, \; i \in K, \tag{61}$$

and also the vectors

$$\frac{\mathbf{e}_{i_1}}{w_{i_1}} - \frac{\mathbf{e}_{i_2}}{w_{i_2}}, \quad \text{for all } i_1 \in K, \; i_2 \in K, \tag{62}$$

where $L = \{1, 2, ..., l\}$ is the support set for the face $\mathcal{G}$. So the cone $\mathcal{C}_{\mathcal{F},\mathcal{G}}$ is the direct sum of the linear hull $\mathcal{L}_{\mathcal{F}} = \lin\{\mathcal{F} - x^{\mathcal{F}}\}$ formed by the vectors in (62) and the cone $\mathcal{C}_{\mathcal{F}^{\perp},G} = \mathcal{C}_{F,G} \bigcap \mathcal{L}_{\mathcal{F}}^{\perp}$, where $\mathcal{L}_F^{\perp}$ is the orthogonal complement to the linear subspace $\mathcal{L}_{\mathcal{F}}$. Then $\mathcal{C}_{\mathcal{F}^{\perp},\mathcal{G}}$ has the same (relative) spherical volume as $\mathcal{C}_{\mathcal{F},\mathcal{G}}$, and by definition the internal angle $\beta(\mathcal{F},\mathcal{G})$ is the relative spherical volume of the cone $\mathcal{C}_{\mathcal{F},\mathcal{G}}$. Now let us analyze the structure of $\mathcal{C}_{\mathcal{F}^{\perp},\mathcal{G}}$. We notice that the vector $\mathbf{e}_0 = \sum_{r=1}^{k} w_r \mathbf{e}_r$ is in the linear space $\mathcal{L}_{\mathcal{F}}^{\perp}$ and is also the only such a vector (up to linear scaling) supported on K. Thus a vector $\mathbf{x}$ in the positive cone $\mathcal{C}_{\mathcal{F}^{\perp},\mathcal{G}}$ must take the form

$$-\sum_{i=1}^{k} b_i e_i + \sum_{i=k+1}^{l} b_i e_i, \tag{63}$$

where $b_i, 1 \leq i \leq l$ are nonnegative real numbers and

$$\sum_{i=1}^{k} w_i b_i = \sum_{i=k+1}^{l} w_i b_i \quad \frac{b_1}{w_{K_1}} = \frac{b_2}{w_{K_2}} = \cdots = \frac{b_k}{w_k}.$$

Now that we have identified $\mathcal{C}_{\mathcal{F}^\perp, \mathcal{G}}$ we try to calculate its relative spherical volume with respect to the sphere surface S^{l-k-1} to derive $\beta(\mathcal{F}, \mathcal{G})$. First, we notice that $\mathcal{C}_{\mathcal{F}^\perp, \mathcal{G}}$ is a $(l-k)$-dimensional cone. Also, all the vectors $(x_1, \cdots, x_n)$ in the cone $\mathcal{C}_{\mathcal{F}^\perp, \mathcal{G}}$ take the form in (63). From [22],

$$\int_{\mathcal{C}_{\mathcal{F}^\perp, \mathcal{G}}} e^{-\|\mathbf{x}\|^2} d\mathbf{x} = \beta(\mathcal{F}, \mathcal{G}) V_{l-k-1}(S^{l-k-1}) \int_0^\infty e^{-r^2} r^{l-k-1} dx = \beta(\mathcal{F}, \mathcal{G}) \cdot \pi^{(l-k)/2},$$

where $V_{l-k-1}(S^{l-k-1})$ is the spherical volume of the $(l-k-1)$-dimensional sphere S^{l-k-1} and is given by the well-known formula

$$V_{i-1}(S^{i-1}) = \frac{i\pi^{\frac{i}{2}}}{\Gamma(\frac{i}{2}+1)},$$

where $\Gamma(\cdot)$ is the usual Gamma function. This completes the proof.

C Proof of Lemma 5.3

Let G denote the cumulative distribution function of a half-normal $HN(0, 1/2)$ random variable, i.e. a random variable $X = |Z|$ where $Z \sim N(0, 1/2)$, and $G(x) = \mathbb{P}\{X \leq x\}$. Since X has density function $g(x) = \frac{2}{\sqrt{\pi}} \exp(-x^2)$, we know that

$$G(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2} dy; \tag{64}$$

and so G is just the classical error function $erf(\cdot)$. We now justify the external angle exponent computations in Theorem 4.3 and Lemma 5.3 using Laplace methods [3]. Using the same set of notations as in Theorem 4.3, let $t_1 = \tau_1 n$, $t_2 = \tau_2 n$. 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:

$$2c - \frac{g(x)\alpha_1}{xG(x)} - \frac{\omega g(\omega x)\alpha_2}{xG(\omega x)} = 0 \tag{65}$$

Since $xG(x)$ is a smooth strictly increasing function (~ 0 as $x \to 0$ and $\sim x$ as $x \to \infty$), and $g(x)$ is strictly decreasing, the function $\frac{g(x)\alpha_1}{xG(x)} + \frac{\omega g(\omega x)\alpha_2}{xG(\omega x)}$ is one-one on the positive axis, and x_0 is a well-defined function of τ_1 and τ_2. Hence, we denote it as $x_0(\tau_1, \tau_2)$. Then

$$\psi_{ext}(\tau_1, \tau_2) = cx_0^2 - \alpha_1 \log G(x_0) - \alpha_2 \log G(\omega x_0). \tag{66}$$

To prove Lemma 5.3, we start from the explicit integral formula

$$\zeta(d_1, d_2) = \pi^{-\frac{n-l+1}{2}} 2^{n-l} \int_0^\infty e^{-x^2} \left(\int_0^{\frac{w_{K_1} x}{\xi(d_1,d_2)}} e^{-y^2} \, dy \right)^{r_1} \left(\int_0^{\frac{w_{K_2} x}{\xi(d_1,d_2)}} e^{-y^2} \, dy \right)^{r_2} dx, \tag{67}$$

After a changing of integral variables (Noticing that $w_{K_1} = 1$, $w_{K_2} = \omega$, $\frac{n_1 - d_1}{n} = \alpha_1$, and $\frac{n_2 - d_2}{n} = \alpha_2$), we have

$$\zeta(t_1 + k_1, t_2 + k_2) = \sqrt{cn/\pi} \int_0^\infty e^{-n(cx^2 - \alpha_1 \log(G(x)) - \alpha_2 \log(G(\omega x)))} \, dx. \tag{68}$$

This suggests that we should use Laplace's method; we define

$$f_{\tau_1,\tau_2,n} = e^{-n\psi_{t_1',t_2'}(y)} \cdot \sqrt{cn/\pi} \tag{69}$$

with

$$\psi_{t_1',t_2'}(y) = cy^2 - c_1 \log G(y) - \alpha_2 \log G(\omega y)$$

We note that the function $\psi_{t_1',t_2'}$ is smooth and convex. Applying Laplace's method to $\psi_{t_1',t_2'}$, but taking care about regularity conditions and remainders as in [3], gives a result with the uniformity in (t_1', t_2').

Lemma C.1. *For t_1', t_2', let $x_0(\tau_1, \tau_2)$ denote the minimizer of $\psi_{t_1',t_2'}$. Then*

$$\int_0^\infty f_{t_1',t_2',n}(x) \, dx \le e^{-n\psi_{t_1',t_2'}(x_0(t_1',t_2'))}(1 + R_n(t_1', t_2')),$$

where for any $\delta, \eta > 0$,

$$\sup_{0 \le t_1' \le \gamma_1 - \rho_1, 0 \le \tau_2 \le (\gamma_2 - \rho_2), \delta - \rho_1 - \rho_2 \le \tau_1 + \tau_2 \le (1 - \rho_1 - \rho_2 - \eta)} R_n(t_1', t_2') = o(1) \quad as \quad n \to \infty.$$

where $\rho_1 = k_1/n$, $\rho_2 = k_2/n$, $n_1/n = \gamma_1$ and $n_2/n = \gamma_2$.

In fact, in this lemma, the minimizer $x_0(t_1', t_2')$ is exactly the same $x_0(t_1', t_2')$ defined earlier in (65) and the corresponding minimum value is the same as the defined exponent ψ_{ext}:

$$\psi_{ext}(t_1', t_2') = \psi_{t_1',t_2'}(x_{t_1'}, x_{t_2'}). \tag{70}$$

We can derive Lemma 5.3 from Lemma C.1. We note that as $t_1' + t_2' + \gamma_1 + \gamma_2 \to 1$, $x_0(t_1', t_2') \to 0$ and $\psi_{ext}(t_1', t_2') \to 0$. For given $\epsilon > 0$ in the statement of Lemma 5.3, there is a largest $\eta_\epsilon < 1$ such that as

long as $\tau_1 + \tau_2 + \rho_1 + \rho_2 > \eta_\epsilon$, $\psi_{ext}(t_1', t_2') < \epsilon$. Note that $\zeta(\mathcal{G}, \mathcal{P}_\mathbf{w}) \leq 1$, so that for $\tau_1 + \tau_2 + \rho_1 + \rho_2 > \eta_\epsilon$,

$$n^{-1} \log(\zeta(t_1 + k_1, t_2 + k_2)) \leq 0 < -\psi_{ext}(t_1', t_2') + \epsilon,$$

for $n \geq 1$. Applying the uniformity in t_1', t_2' given in Lemma C.1, we have as $n \to \infty$, uniformly over the feasible region for t_1', t_2',

$$n^{-1} \log(\zeta(t_1 + k_1, t_2 + k_2)) \leq -\psi_{ext}(t_1', t_2') + o(1). \tag{71}$$

Then Lemma 5.3 follows.

D   Proof of Lemma 5.4

Recall Theorem 5.1. By applying the large deviation techniques as in [3], we have

$$p_Z(0) \leq \frac{2}{\sqrt{\pi\Omega}} \left(\int_0^{\mu_{m'}} v e^{-v^2 - m'\Lambda^*(\frac{\sqrt{2\Omega}}{m'}v)} \, dv + e^{-\mu_{m'}^2} \right), \tag{72}$$

where Ω is the same as defined in Section 5.2.1, $w_{K_1} = 1$, $w_{K_2} = \omega$, $m' = t_1 + t_2$, $\mu_{m'} = (t_1 + t_2\omega)\sqrt{\frac{1}{\pi\Omega}}$ is the expectation of $\frac{1}{\sqrt{\Omega'}}(w_{K_1}^2 \sum_{i=1}^{t_1} X_i' - w_{K_2}^2 \sum_{i=1}^{t_2} X_i'')$, ($X_i'$ and X_i'' are defined as in Theorem 5.1), and

$$\Lambda^*(y) = \max_s sy - \frac{t_1}{t_1 + t_2}\Lambda_1(s) - \frac{t_2}{t_1 + t_2}\Lambda_2(s),$$

with

$$\Lambda_1(s) = \frac{s^2}{2} + \log(2\Phi(s)), \quad \Lambda_2(s) = \Lambda_1(\omega s).$$

In fact, the second term in the sum can be argued to be negligible [3]. After a changing of variables $y = \frac{\sqrt{2\Omega}}{m'}v$, we know that the first term of (72) is upper-bounded by

$$\frac{2}{\sqrt{\pi}} \cdot \frac{1}{\sqrt{\Omega}} \cdot \frac{m'^2}{2\Omega} \cdot \int_0^{\frac{t_1 + t_2\omega}{t_1 + t_2}\sqrt{2/\pi}} y e^{-m'(\frac{m'}{2\Omega})y^2 - m'\Lambda^*(y)} \, dy. \tag{73}$$

As we know, m' in the exponent of (73) is $t_1 + t_2$. Similar to evaluating the external angle decay exponent, we will resort to the Laplace's method in evaluating the internal angle decay exponent. Define the function

$$f_{t_1, t_2}(y) = y e^{-m'(\frac{m'}{2\Omega})y^2 - m'\Lambda^*(y)}.$$

If we apply similar arguments as in proving Lemma C.1 and take care of the uniformity, we have the following lemma:

Lemma D.1. *Let $y_{t_1,t_2}*$ denotes the minimizer of $(\frac{m'}{2\Omega})y^2 + \Lambda^*(y)$. Then*

$$\int_0^\infty f_{t_1,t_2}(x)\,dx \le e^{-m'\left((\frac{m'}{2\Omega})y_{t_1,t_2}*^2 + \Lambda^*(y_{t_1,t_2}*)\right)} \cdot R_{m'}(t_1,t_2)$$

where for $\eta > 0$

$$m'^{-1}\sup_{t_1,t_2}\log(R_{m'}(t_1,t_2)) = o(1) \quad as \quad m' \to \infty.$$

This means that

$$p_Z(0) \le e^{-m'\left((\frac{m'}{2\Omega})y_{t_1,t_2}*^2 + \Lambda^*(y_{t_1,t_2}*)\right)} \cdot R_{m'}(t_1,t_2),$$

where $m'^{-1}\sup_{\frac{(t_1+t_2)}{n}\in[\delta-\rho_1-\rho_2,1]}\log(R_{m'}(t_1,t_2)) = o(1)$ *as* $m' \to \infty$.

Now in order to find a lower bound on the decay exponent for $p_Z(0)$,(ultimately the decay exponent $\psi_{int}(\tau_1,\tau_2)$), we need to focus on finding the minimizer $y_{t_1,t_2}*$ for $(\frac{m'}{2\Omega})y^2 + \Lambda^*(y)$. On this way, by setting the derivative of $(\frac{m'}{2\Omega})y^2 + \Lambda^*(y)$ with respect to y to 0, and also noting the derivative $\Lambda^{*\prime}(y) = s$, we have

$$s = -\frac{m'}{\Omega}y. \tag{74}$$

At the same time, the s maximizing $\Lambda^*(y)$ must satisfy

$$y = \frac{t_1}{t_1+t_2}\Lambda_1'(s) + \frac{t_2}{t_1+t_2}\Lambda_2'(s), \tag{75}$$

namely, (by writing out (75)),

$$y = \frac{t_1 + \omega^2 t_2}{t_1 + t_2}s + Q(s), \tag{76}$$

where $Q(s)$ is defined as in Theorem 5.1. By combining (74) and (75), we can solve for the s and y, thus resulting in the decay exponent for $\psi_{int}(\tau_1,\tau_2)$ as calculated in Theorem 5.1

E 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:

$$\begin{aligned}
\delta' &= \min\{\delta \mid \psi'_{com}(0,\tau_2) - \psi'_{int}(0,\tau_2) - \psi'_{ext}(0,\tau_2) < 0 \; \forall \; 0 \le \tau_2 \le \gamma_2(1-p_2) \\
&\quad , \tau_2 > \delta - \gamma_1 - \gamma_2 p_2\}, \\
&= \gamma_2\min\{\delta \mid \psi'_{com}(0,\gamma_2\tau_2) - \psi'_{int}(0,\gamma_2\tau_2) - \psi'_{ext}(0,\gamma_2\tau_2) < 0 \; \forall \; 0 \le \tau_2 \le 1 - p_2 \\
&\quad , \tau_2 > \delta - p_2\} + \gamma_1,
\end{aligned} \tag{77}$$

and

$$\delta'' = \min\{\delta \mid \psi''_{com}(0,\tau_2) - \psi''_{int}(0,\tau_2) - \psi''_{ext}(0,\tau_2) < 0 \; \forall \; 0 \le \tau_2 \le 1 - p_2$$

$$, \tau_2 > \delta - p_2\}, \tag{78}$$

where the exponents $\psi'_{com}, \psi'_{int}, \psi'_{ext}, \psi''_{com}, \psi''_{int}$ and ψ''_{ext} can be found using Theorem 4.3. Here, we basically show that when $\omega \to \infty$:

$$\psi'_{com}(0,\gamma_2\tau_2) = \gamma_2\psi''_{com}(0,\tau_2), \tag{79}$$

$$\psi'_{int}(0,\gamma_2\tau_2) = \gamma_2\psi''_{int}(0,\tau_2), \tag{80}$$

$$\psi'_{ext}(0,\gamma_2\tau_2) = \gamma_2\psi''_{ext}(0,\tau_2). \tag{81}$$

(79) follows immediately from the definition of ψ_{com} in (6). On the other hand, from (7), for $\omega \to$ we know that

$$\psi'_{ext}(0,\gamma_2\tau_2) = c'{x_0'}^2 - \alpha_2' \log G(\omega x_0'),$$

$$\psi''_{ext}(0,\tau_2) = c''{x_0''}^2 - \alpha_2'' \log G(x_0'').$$

Following the details of derivations as in Theorem 4.3, we realize that:

$$c' = \gamma_2\omega^2 c'', \; \omega x_0' = x_0'', \; \alpha_2' = \gamma_2\alpha_2'', \tag{82}$$

which implies that $\psi'_{ext}(0,\gamma_2\tau_2) = \gamma_2\psi''_{ext}(0,\tau_2)$. 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 for $\omega \to \infty$:

$$y' = y'', \; \Omega' = \gamma_2\Omega''. \tag{83}$$

which implies that $\psi'_{int}(0,\gamma_2\tau_2) = \gamma_2\psi''_{int}(0,\tau_2)$. From (77), (78) and (79)-(81) it follows that

$$\delta' = \gamma_2\delta'' + \gamma_1. \tag{84}$$