Compressive Sensing for Sparse Approximations: Constructions, Algorithms, and Analysis

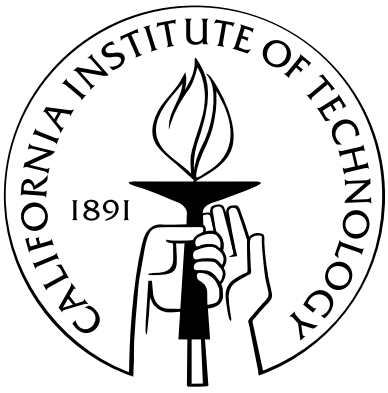
Thesis by

Weiyu Xu

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

2010

(Defended August 13th, 2009)

Acknowledgements

I thank my advisor Professor Babak Hassibi for his guidance and support throughout my PhD study at Caltech. His great passion and capability for fundamental theoretical research over a broad spectrum of fields have been, and will always be, a great inspiration to me. It is a great pleasure and privilege for me to have worked together with him over the past four years. He was always available to help work out the problems whenever I was stuck in research. His nice personality has made me feel at home at Caltech even though my motherland is thousands of miles away.

I thank Professor Emmanuel Candès for introducing me to his seminal work on the magic of ℓ_1 minimization in a homework set when I was a first-year Caltech graduate student taking his course ACM 103. It was hard to imagine then that this homework problem, together with a little bit curiosity of my own, would attract me to the beautiful topics in compressive sensing and would eventually lead to the completion of this thesis. Professor Candès's encouragement and support have been a very valuable part of my graduate study.

My special thanks go to my other collaborators besides my advisor, Mihailo Stojnic, Ben Recht, M. Amin Khajehnejad, Salman Avestimehr, Alex Dimakis, Sina Jafarpour, Morten Hansen, and Robert Calderbank, without whom some of the chapters in this thesis would not have been possible. I have been learning a lot through the collaborations. I would especially like to acknowledge that Professor Rolf Schneider very kindly mailed to me his paper on random projections of polytopes when I could not find it in the library.

My sincere thanks go to Professors Robert McEliece, Emmanuel Candès, P. P. Vaidy-anathan, Abu Mostafa, and Joel Tropp for being on my candidacy and thesis

committees and providing valuable feedback.

I am grateful to Shirley for her pleasant and efficient help over the last four years. It has always been fun to hang out in Moore 155 and collaborate on research with the labmates. The friendship and the enormous fun of working there are just hard to express in a single page. Amir Farajidana, Radhika Gowaikar, Vijay Gupta, Yindi Jing, Devdutt Marathe, Michela Munoz-Fernandez, Frédérique Oggier, Farzad Parvaresh, Chaitanya Rao, Masoud Sharif, Mihailo Stojnic, Sormeh Shadbakht and Haris Vikalo all have set up very good role models and guided me through the happy days at the Institute, and I enjoy interacting with them very much. I am especially thankful towards Chaitanya Rao, who, on my first day in Pasadena, spared his Friday evening and took me to a dinner at a Thai restaurant on the Colorado Boulevard. The delicious food there had me deeply attracted to this beautiful sunny place.

I would also like to thank many friends of mine outside Caltech, who have helped me and made my life more fun.

Finally, my special thanks go to my family. I thank them for their patience and unreserved support, which are the cornerstones of my PhD life.

Abstract

Compressive sensing is an emerging research field that has applications in signal processing, error correction, medical imaging, seismology, and many more other areas. It promises to efficiently recover a sparse signal vector via a much smaller number of linear measurements than its dimension. Naturally, how to design these linear measurements, how to construct the original high-dimensional signals efficiently and accurately, and how to analyze the sparse signal recovery algorithms are important issues in the developments of compressive sensing. This thesis is devoted to addressing these fundamental issues in the field of compressive sensing.

In compressive sensing, random measurement matrices are generally used and ℓ_1 minimization algorithms often use linear programming or other optimization methods to recover the sparse signal vectors. But explicitly constructible measurement matrices providing performance guarantees were elusive and ℓ_1 minimization algorithms are often very demanding in computational complexity for applications involving very large problem dimensions. In chapter 2, we propose and discuss a compressive sensing scheme with deterministic performance guarantees using deterministic explicitly constructible expander graph-based measurement matrices and show that the sparse signal recovery can be achieved with linear complexity. This is the first of such a kind of compressive sensing scheme with linear decoding complexity, deterministic performance guarantees of linear sparsity recovery, and deterministic explicitly constructible measurement matrices.

The popular and powerful ℓ_1 minimization algorithms generally give better sparsity recovery performances than known greedy decoding algorithms. In chapter 3, starting from a necessary and sufficient null-space condition for achieving a certain

signal recovery accuracy, using high-dimensional geometry, we give a unified *null-space Grassmann angle*-based analytical framework for compressive sensing. This new framework gives sharp quantitative trade-offs between the signal sparsity and the recovery accuracy of the ℓ_1 optimization for approximately sparse signal. Our results concern the fundamental "balancedness" properties of linear subspaces and so may be of independent mathematical interest.

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 chapter 4, we will consider a particular model for the sparse signal that assigns a probability of being zero or nonzero to each entry of the unknown vector. The standard compressed sensing model is therefore a special case where these probabilities are all equal. Following the introduction of the *null-space Grassmann angle*-based analytical framework in this thesis, we are able to characterize the optimal recoverable sparsity thresholds using weighted ℓ_1 minimization algorithms with the prior information.

The roles of ℓ_1 minimization algorithm in recovering sparse signals from incomplete measurements are now well understood, and sharp recoverable sparsity thresholds for ℓ_1 minimization have been obtained. The iterative reweighted ℓ_1 minimization algorithms or related algorithms have been empirically observed to boost the recoverable sparsity thresholds for certain types of signals, but no rigorous theoretical results have been established to prove this fact. In chapter 5, we try to provide a theoretical foundation for analyzing the iterative reweighted ℓ_1 algorithms. In particular, we show that for a nontrivial class of signals, the iterative reweighted ℓ_1 minimization can indeed deliver recoverable sparsity thresholds larger than the ℓ_1 minimization. Again, our results are based on the null-space Grassmann angle-based analytical framework.

Evolving from compressive sensing problems, where we are interested in recovering sparse vector signals from compressed linear measurements, we will turn our attention to recovering matrices of low rank from compressed linear measurements in chapter 6, which is a challenging problem that arises in many applications in machine learning,

control theory, and discrete geometry. This class of optimization problems is NP-HARD, and for most practical problems there are no efficient algorithms that yield exact solutions. A popular heuristic replaces the rank function with the nuclear norm of the decision variable and has been shown to provide the optimal low rank solution in a variety of scenarios. We analytically assess the practical performance of this heuristic for finding the minimum rank matrix subject to linear constraints. We start from the characterization of a necessary and sufficient condition that determines when this heuristic finds the minimum rank solution. We then obtain probabilistic bounds on the matrix dimensions and rank and the number of constraints, such that our conditions for success are satisfied for almost all linear constraint sets as the matrix dimensions tend to infinity. Empirical evidence shows that these probabilistic bounds provide accurate predictions of the heuristic's performance in non-asymptotic scenarios.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivations

Compressive sensing, also referred to as *compressed sensing* or *compressive sampling*, is an emerging area in signal processing and information theory which has attracted a lot of attention recently [CT06] [Don06b]. The motivation behind compressive sensing is to do "sampling" and "compression" at the same time. In conventional wisdom, in order to fully recover a signal, one has to sample the signal at a sampling rate equal or greater to the *Nyquist* sampling rate. However, in many applications such as imaging, sensor networks, astronomy, high-speed analog-to-digital compression and biological systems, the signals we are interested in are often "sparse" over a certain basis. For example, an image of a million pixels has a million degrees of freedom, however, a typical *interesting* image is very *sparse* or *compressible* over the wavelet basis, namely, very likely only a small fraction of wavelet coefficients, say, one hundred thousand out of a million wavelet coefficients, are significant in recovering the original images, while the rest of wavelet coefficients are "thrown away" in many compression algorithms. This process of "sampling at full rate" and then "throwing away in compression" can prove to be wasteful of sensing and sampling resources, especially in application scenarios where such resources as sensors, energy, and observation time etc. are limited.

Instead of thinking in the traditional way, compressive sensing promises to recover

the high-dimensional signals exactly or accurately, by using a much smaller number of *non-adaptive* linear samplings or measurements. In general, signals in this context are represented by vectors from linear spaces, many of which in the applications will represent images or other objects. The fundamental theorem of linear algebra, "as many equations as unknowns," tells us that it is not possible to reconstruct a unique signal from an incomplete set of linear measurements. However, as we said before, many signals such as real-world images or audio signals are often sparse or compressible over some basis, such as smooth signals or signals whose variations are bounded. This opens the room for recovering these signals accurately or even exactly from incomplete linear measurements. However, even though we know that the signal itself is sparse, it is a non-trivial job to recover the signals from the compressed measurements since we do not know the locations of the non-zero or significant components of that vector. One of the cornerstone techniques enabling compressive sensing is then about efficient and effective decoding algorithms to recover the sparse signals from the "compressed" measurements. One of the most important and popular decoding algorithms for compressive sensing is the Basis Pursuit algorithm [Che95, CDS98], namely the ℓ_1-minimization algorithm.

1.2 Mathematical Formulation for Sparse Recovery

Before we go into greater technical details in later chapters, we will first give the general signal models discussed in this thesis. We say that a n-dimensional signal $\mathbf{x}$ is k-sparse if it has k or fewer non-zero components:

$$\mathbf{x} \in \mathcal{R}^n, \quad \|\mathbf{x}\|_0 := |\mathrm{supp}(\mathbf{x})| \leq k \ll n,$$

where $|\mathrm{supp}(\mathbf{x})|$ denotes the cardinality of the support set of $\mathbf{x}$, and thus $\|\cdot\|_0$, namely the number of non-zero components, is a quasi-norm. For $1 \leq p < \infty$, we denote by

$\|\cdot\|_p$ the usual p-norm,

$$\|\mathbf{x}\|_p := \Big(\sum_{i=1}^{n} |x_i|^p\Big)^{1/p},$$

and $\|\mathbf{x}\|_\infty = \max|x_i|$, where x_i denotes the i-th component of the vector $\mathbf{x}$. In the case when signals are not exactly sparse, but their coefficients decay rapidly, we call these signals *approximately sparse* signals. In particular, compressible signals are those satisfying a power law decay:

$$|x_i^*| \leq Li^{-\frac{1}{q}}, \tag{1.2.1}$$

where $\mathbf{x}^*$ is a non-increasing rearrangement of $\mathbf{x}$, L is some positive constant, and $0 < q < 1$. Of course, sparse signals are special cases of compressible signals.

Compressive sensing only measures sparse signals from a small set of non-adaptive linear measurements. Each measurement is seen as an inner product between the signal $\mathbf{x} \in \mathcal{R}^n$ and a measurement vector $a_i \in \mathcal{R}^n$, where $i = 1, \ldots, m$. If we collect m measurements in this way, we may then consider the $m \times n$ measurement matrix A whose rows are the vectors a_i. We can then view the sparse recovery problem as the recovery of the k-sparse signal $\mathbf{x} \in \mathcal{R}^n$ from its measurement vector $\mathbf{y} = A\mathbf{x} \in \mathcal{R}^m$.

One may wonder how to reconstruct the signals from this incomplete set of measurements. With the prior information of the sigal being sparse, one of the theoretically simplest ways to recover such a vector from its measurements $\mathbf{y} = A\mathbf{x}$ is to solve the ℓ_0-minimization problem

$$\min_{\mathbf{x}\in\mathcal{R}^n} \|\mathbf{x}\|_0 \quad \text{subject to} \quad A\mathbf{x} = \mathbf{y}. \tag{1.2.2}$$

If $\mathbf{x}$ is k-sparse and the rank of A is larger than $2k$, then the solution to (1.2.2) must be the signal $\mathbf{x}$. Indeed, if the solution is $\mathbf{z}$, then since $\mathbf{x}$ is a feasible solution, $\mathbf{z}$ must be k-sparse as well. Since $A\mathbf{z} = \mathbf{y}$, $\mathbf{y} - \mathbf{x}$ must be in the null-space of A. But $\mathbf{z} - \mathbf{x}$ has at most $2k$ non-zero elements, and since the rank of A is larger than $2k$, we must have that $\mathbf{z} = \mathbf{x}$. Thus this ℓ_0-minimization problem works perfectly

theoretically. However, it is computationally NP-Hard in general [CT05]. Fortunately, in the framework of compressive sensing, there have been computationally efficient relaxation algorithms for this computationally NP-Hard problem.

1.2.1 ℓ_1 Minimization

One major approach, Basis Pursuit, relaxes the ℓ_0-minimization problem to an ℓ_1 minimization problem:

$$\min_{\mathbf{x}\in\mathcal{R}^n} \|\mathbf{x}\|_1 \quad \text{subject to} \quad A\mathbf{x} = \mathbf{y}. \tag{1.2.3}$$

Simply put, instead of trying to find the solution with the smallest ℓ_0-norm, ℓ_1 minimization tries to find the solution with the minimum ℓ_1 norm. Surprisingly, this relaxation often recovers $\mathbf{x}$ *exactly* when $\mathbf{x}$ is sparse or *accurately* when $\mathbf{x}$ is an approximately sparse signal or compressible signal. Please note that the measurement matrix A is given and fixed in advance, and does not depend on the signal, but as long as the signals are sufficiently sparse and the measurement matrix satisfies some conditions independent of the signals, the ℓ_1 minimization will succeed [CT05, Don06c]. That is, even though ℓ_1-norm is different from the quasi-norm ℓ_0, the solution of ℓ_1 often comes as the sparsest solution.

As mentioned in [CT08], this sparsity-promoting feature of ℓ_1 minimization was already observed in the 1960's by Logan [Log65], where he proved probably the first ℓ_1-uncertainty principle. Suppose we have the observation over time

$$y(t) = f(t) + n(t), t \in \mathcal{R}, \tag{1.2.4}$$

where $f(t)$ is bandlimited, namely

$$f \in B(\Omega) := \{f : \hat{f}(\omega) = 0 \quad \text{for} \quad |\omega| > \Omega\}, \tag{1.2.5}$$

and $n(t)$ is an impulsive noise term supported on a sparse set T. Logan observed that

if we recover f by the following program

$$\min \|y - \hat{f}\|_{L_1(\mathcal{R})} \quad \text{subject to} \quad \hat{f} \in B(\Omega), \tag{1.2.6}$$

then the recovery is exact provided that $|T||\Omega| \leq \pi/2$. This holds for whatever $f \in B(\Omega)$ and for whatever values of the noise. ℓ_1 minimization also appeared early in reflection seismology, where people tried to infer a sparse reflection function (indicating meaningful changes between subsurface layers) from bandlimited data. For example, Taylor, Banks and McCoy and others began proposing the application of ℓ_1 for deconvolving seismic traces [TBM79] and a refined idea better handling the observation noise was introduced in [SS86]. In the meanwhile, some rigorous theoretical results started appearing in the late 1980's, when Donoho and Stark [DS89] and Donoho and Logan [DL92] extended Logan's 1965 result and quantified the ability to recover sparse reflectivity functions from bandlimited data. With the LASSO algorithm [Tib96] proposed as a method in statistics for sparse model selection, the application areas for ℓ_1 minimization began to broaden. Basis Pursuit [CDS98] was proposed in computational harmonic analysis for extracting a sparse signal representation from highly overcomplete dictionaries, and a related technique known as total variation minimization was proposed in image processing [GGI$^+$02].

It then came as a breakthrough in [CT05, CT06] and [Don06c] that Basis Pursuit method was shown to be able to recover sparse signals with a linear fraction of non-zero elements. Certainly this requires some conditions on the measurement matrix A stronger than the simple rank conditions mentioned above. For example, the restricted isometry property (RIP) conditions were given in [CT05, CT06] to guarantee that ℓ_1 minimization accurately recovers sparse or compressible signals. It is now known that many kinds of matrices satisfy these conditions with the number of measurements $m = k\log(n)^{O(1)}$ [CT05, CT06], where $O(\cdot)$ is the standard notion for computational complexity in computer science.

1.2.2 Greedy Algorithms

The ℓ_1-minimization approach provides uniform guarantees over all sparse signals and also stability and robustness under measurement noises and approximately sparse signals, but relies on optimization which has relatively high complexity, for example, linear programming, the complexity of which grows cubic in the problem dimension n. In many applications which involve very large dimension processing, these approaches are not optimally fast. The other main approaches use greedy algorithms such as Orthogonal Matching Pursuit [MZ93, TG], Stagewise Orthogonal Matching Pursuit [DTDS07], Regularized Orthogonal Matching Pursuit [NV09], Iterative Thresholding [FR07, BD] and Compressive Sampling Matching Pursuit (CoSaMP) [NT08]. Most of these approaches calculate the support of the signal iteratively. With the support S of the signal calculated, the signal x is reconstructed from its measurements $\mathbf{y} = A\mathbf{x}$ as $\mathbf{x} = (A_S)^\dagger \mathbf{y}$, where A_S denotes the measurement matrix A restricted to the columns indexed by S and $\dagger$ denotes the pseudoinverse. Greedy approaches are relatively fast compared with the Basis Pursuit algorithm, both in theory and practice, but most of them deliver smaller recoverable sparsity compared to ℓ_1 minimization and most of them often come without provable uniform guarantees and stability, with the exception of [NV09, NT08].

1.2.3 High Dimensional Geometry for Compressive Sensing

The idea of compressive sensing and ℓ_1 minimization certainly did not come from nowhere. The theoretical foundation for compressed sensing is high dimensional geometry, which is deeply connected with the field of geometric functional analysis. For example, Kashin in the 1970's studied how many and what linear measurements (as described before) need to be taken so that we can recover a vector with a precision ϵ from the ℓ_1 ball,

$$\{\mathbf{x} \in \mathcal{R}^n : \quad \|\mathbf{x}\|_1 \leq 1\}, \tag{1.2.7}$$

in principle (regardless of the computational complexity of the decoding methods).

After we know the measurement matrix A, we know that all the solutions $\mathbf{x}$ to the underdetermined system lies in an affine space parallel to the null space of A. Then Kashin's problem clearly becomes a high dimensional geometrical problem, which is about how we should select this null space so that the affine space's intersection with the ℓ_1 ball has minimal radius. The answer to this question was given by Kashin [Kas77] and later refined by Garnaev and Gluskin [GG]. Their *existential* results rely on randomly choosing the linear projections (or measurements) and are optimal in order of the number of measurements, which is within a multiplicative factor of what the ℓ_1 minimization compressive sensing provides. However, their results were mostly *existential* while compressive sensing comes with at least one practical algorithm, the ℓ_1 minimization algorithm, which is nearly optimal over many classes of signals.

In addition, the deep probabilistic techniques from [Bou89, BT, BT87, BT91], especially the generic chaining technique developed in [Bou89, Tal96], which controls the suprema of random processes, were used as important technical tools in verifying that certain measurement matrix ensembles satisfy the conditions for recovering sparse signals [CT06].

1.3 Applications

Besides its deep roots in fundamental theories, compressive sensing has widespread applications, both in the past, for example, in seismology and promisingly in the future. It has already advanced the state of the arts in imaging, MRI (magnetic resonance imaging), NMR (Nuclear Magnetic Resonance) spectroscopy, radar design, function approximation, real-numbered error correction, communications and high-speed analog-to-digital conversion. Several examples of its applications are listed below.

1.3.1 Compressive Imaging

Naturally, one of the most prominent applications of compressive sensing is to acquire images efficiently. The images we are interested in are often sparse over some basis so that they fit just into the framework of compressive sensing. Today's digital cameras capture images with one sensor for each pixel and acquire every pixel in an image before compressing that captured data and storing the compressed image. Due to the use of silicon, everyday digital cameras today can operate in the megapixel range. Consistent with the motivation for compressive sensing, a natural question asks why we need to acquire this many data, just to throw most of it away immediately.

In the newly developed compressive imaging, the sensors directly acquire random linear measurements of an image while avoiding using sensors for each pixel. Compressive sensing provides a guideline framework for implementing such an idea, including designing the measurement methods and the decoding algorithms. Researchers have worked on the construction of such systems, for example, in a prototype "single-pixel" compressive sampling camera [WLD+06]. This camera consists of a digital micromirror device (DMD), two lenses, a *single* photon detector and an analog-to-digital (A/D) converter. The first lens focuses the light onto the DMD. Each mirror on the DMD is for a pixel in the image, and can be tilted toward or away from the second lens. This operation is analogous to creating inner products with random vectors. This light is then collected by the lens and focused onto the photon detector where the measurement is computed. This optical computer computes the random linear measurements of the image in this way and passes those to a digital computer that reconstructs the image.

In a very different style from the traditional camera, this camera utilizes only *one* photon detector. One key highlight is that the single-pixel camera can operate over a much broader range of the light spectrum than traditional cameras that use silicon. Over some light spectrum, the sensors can be very expensive. A digital camera made of silicon to capture infrared images would be very complicated and costly if it is built in a traditional way.

In medical imaging, in particular in magnetic resonance imaging (MRI) which sample Fourier coefficients of an image, compressive sensing finds another important application. MR images are implicitly sparse: some MR images such as angiograms are sparse in their actual pixel representation, whereas more complicated MR images are sparse over some other basis, such as the wavelet Fourier basis. As we all know, MRI in general is very time costly, as the speed of data collection is limited by physical and physiological constraints. Thus it is very helpful to reduce the number of measurements collected without sacrificing quality of the MR image, or said in another way, increasing the recovered image quality with the same number of measurements. In fact, compressive MRI is a very active topic in compressive sensing and has attracted a large number of researchers in this field. Many compressive sensing algorithms have been specifically designed for MRI application [LDP07].

1.3.2 Radar Signal Processing

A traditional radar system transmits some kind of pulse form, and then uses a matched filter to correlate the signal received with that pulse. The receiver then uses a pulse compression system together with a high-rate analog-to-digital (A/D) converter for signal processing. This conventional approach is not only complicated and expensive, but also the resolution in traditional radar system is limited by the radar uncertainty principle. Compressive Radar Imaging discretizes the time-frequency plane into a grid and treats each possible target scene as a matrix. If the number of targets is small enough, then the occupations of the grids will be sparse, and compressive sensing techniques can be used to recover the target scene [HS07].

1.3.3 Biology

Compressive sensing can also be used for efficient and low-cost sensing in the area of biological applications. In fact, the applications of Group Testing, an idea closely related to compressive sensing, was used as early as in World War II to test soldiers for syphilis [WNB06]. Because the test (or sensing) for syphilis antigen in a blood

sample is expensive, the method was to group people and test the entire pool of blood samples for this group. Only if syphilis antigen was found in a pool of samples, further testings into the subgroups of that group would then take place.

A more modern example of compressive sensing idea in biology is for comparative DNA Microarray [VPMH, MBSR]. Microarrays (DNA, protein, etc.) are massively parallel affinity-based biosensors capable of detecting and quantifying a large number of different genomic particles simultaneously. Generally, DNA microarrays comprising tens of thousands of probe spots are being used to test a multitude of targets in a single experiment. In conventional microarrays, each spot contains a large number of copies of a single probe designed to capture a single target, and hence collects only a single data point. But in comparative DNA microarray experiments, only a fraction of the total number of genes represented by the reference sample and the test sample is differentially expressed. So we can use the compressive ideas to create the so-called *compressed microarrays* [VPMH], wherein each spot contains copies of several different probes and the total number of spots is potentially much smaller than the number of targets being tested.

Gene expression studies also provide examples of compressive sensing. For example, one would like to infer the gene expression level of thousands of genes from only a limited number of observations [Can06].

1.3.4 Error Correcting

Compressive sensing also has impacts on the coding theory and practices and can be seen as a dual to the error correction problem over the real number field. Error correction problem is a classic problem in coding theory: when signals are sent from the sender to the receiver in communications, the signal is usually corrupted by errors. It is thus interesting to see how to design system and decoding algorithms to correct the errors. Over the real field, because the errors usually occur in few places, sparse recovery can be used to reconstruct the signal from the corrupted encoded data [CT05]. What is novel here is that this error correction problem is over the real

field, while on the contrary, traditional coding theory usually assumes data values over the finite field. Indeed there are many practical applications for encoding over the continuous reals. In digital communications, for example, one wishes to protect results of onboard computations that are real-valued. These computations are performed by circuits that experience faults caused by effects of the outside world. This and many other examples are difficult real-world problems of error correction.

The error correction problem can be formulated as follows, from where we can also see the close relationship between coding theory and compressive sensing. Consider a m-dimensional input vector $\mathbf{f} \in \mathcal{R}^m$, the "plaintext," that we wish to transmit reliably to a remote receiver. We transmit the n-dimensional coded text, namely "ciphertext," $\mathbf{z} = B\mathbf{f}$ where B is the $n \times m$ coding matrix, or the *linear code*. In the case of no noise, it is clear that if the linear code B has full rank, we can recover the input vector $\mathbf{f}$ from the ciphertext $\mathbf{z}$. But as is often the case in practice, we consider the setting where the ciphertext $\mathbf{z}$ has been corrupted by sparse noises (similar to the finite field coding literature, a few bit errors). We then wish to reconstruct the input signal f from the corrupted received codeword $z' = Bf + e$ where $\varepsilon \in \mathcal{R}^n$ is the sparse error vector. To realize this in the usual compressed sensing setting, consider a matrix A whose null-space is the range of B. Apply A to both sides of the equation $\mathbf{z}' = B\mathbf{f} + \varepsilon$ to get $A\mathbf{z}' = A\varepsilon$. Set $\mathbf{y} = B\mathbf{z}'$ and the problem becomes reconstructing the sparse vector ε from its linear measurements $\mathbf{y}$. Once we have recovered the error vector ε, we have access to the actual measurements $A\mathbf{f}$ and, since A is full rank, can recover the input signal $\mathbf{f}$. For the details, please refer to [RIC, CT05].

1.4 Some Important Issues

There are many interesting theoretical and practical issues that merit attention in the field of compressive sensing.

1.4.1 Explicit Constructions of Sensing Matrices

There are several classes of random matrices used nowadays in compressive sensing, for example, random Gaussian matrices, random Bernoulli matrices, or random minors of a discrete Fourier transform matrix, but they are all probabilistic in nature; in particular, these randomly constructed matrices are not perfectly guaranteed to actually produce a "good" sensing matrix, although in many cases the failure rate can be proven to be exponentially small in the size of the matrix. Moreover, there were no fast algorithms known to test whether any given matrix is a good measurement matrix, for example, satisfying the RIP condition [Tao07]. It is thus interesting to find a deterministic construction which can give and test "good" sensing matrices efficiently. In analogy with error-correcting codes, it may be that algebraic or number-theoretic constructions may give such deterministic "good" matrices; some efforts have been made towards this end, for example, in [DeV07] where deterministic sensing matrices with *suboptimal* sparsity parameters have been given. However, deterministic explicit efficient constructions of sensing matrices, which offers stable recovery of sparse signals with sparsity scaling linearly with the problem dimension, were elusive.

1.4.2 Efficient Decoding Algorithms with Provable Performance Guarantees

In the compressed sensing literature, there have been many numerically feasible decoding algorithms to the sparse recovery problem from compressed observations. One major approach, Basis Pursuit, relaxes the ℓ_0-minimization problem to an ℓ_1-minimization problem. The ℓ_1-minimization approach provides uniform guarantees and stability, but relies on optimization methods for ℓ_1-minimization, for example, linear programming. These algorithms are not optimally fast (generally not linearly bounded in the problem dimension). One could also use other greedy algorithms such as Orthogonal Matching Pursuit [MZ93, TG], Regularized Orthogonal Match-

ing Pursuit [NV09], and Compressive Sampling Matching Pursuit (CoSaMP) [NT08]. These algorithms can provide similar uniform guarantees and stability results as the Basis Pursuit algorithm, but the complexity of these algorithms are growing super-linearly in the problem dimension. It would be interesting to design sensing matrices and decoding algorithms which are able to provide provable strong sparsity recovery performances while having low computational complexity, hopefully linear in the problem dimension.

1.4.3 Performance Analysis of Sparse Signal Recoveries

There are many algorithms for sparse signals recoveries, for example, the ℓ_1-minimization algorithm. It is very important to understand how well these algorithms perform in recovering sparse signal recoveries, which is often characterized by the recoverable sparsity threshold. Thus what is in particular interesting then is to characterize the recoverable sparsity thresholds or sharp performance bounds for different decoding algorithms and measurement matrices. When the measured signals are not exactly sparse and the measurement results are corrupted by noises, we are more interested in analyzing the stability and robustness of these sparse signal recovery algorithms and how they interact with the signal sparsity. The analysis made for sparse signal recovery is often tightly connected to fundamental probabilistic or geometric phenomena and problems in high dimensional geometry, thus often advancing the fruitful interactions between signal processing, optimization theory and high dimensional geometrical and probabilistic analysis.

1.4.4 More than Sparse Vectors

One may naturally wonder whether compressive sensing principle only applies to recovering sparse vectors. The answer is certainly no because sparse vectors are not the only information objects which have structures even though their ambient dimension might be large. For example, the matrices we run into in applications may have low rank. It is thus possible to recover the information objects (e.g., matrices

and information tables). Some results have been obtained in this direction [RFP, CR09, RXH08a]. All the important technical challenges with compressive sensing also appear in these more general problems.

1.5 Contributions

The contributions of this thesis are closely related to addressing the problems of how to design these linear measurements, how to construct the original high-dimensional signals efficiently and accurately, and how to analyze the sparse signal recovery algorithms.

1.5.1 Expander Graphs for Explicit Sensing Matrices Constructions

As we mentioned before, explicitly constructible measurement matrices providing performance guarantees and linearly scaling sparsity recoverability were elusive and the ℓ_1 minimization methods are very demanding in computational complexity for problems with very large dimension. In chapter 2, we proposed and discussed a compressive sensing scheme with deterministic performance guarantees using deterministic explicitly constructible expander-graphs-based measurement matrices. Moreover, we showed that sparse signal recoveries can be achieved with *linear* complexity. This is the first of such a kind of compressive sensing scheme with linear decoding complexity, deterministic performance guarantees and deterministic explicitly constructible measurement matrices.

1.5.2 Grassmann Angle Analytical Framework for Subspaces Balancedness

ℓ_1 minimization algorithms generally give better sparsity recovery performances than known greedy decoding algorithms in compressive sensing. In chapter 3, starting from

a necessary and sufficient null-space condition for achieving a certain signal recovery accuracy, we reduce the analysis of sparse signal recovery robustness to investigating a linear subspace balancedness property. Using high-dimensional geometry, we give a unified *null-space Grassmann angle*-based analytical framework for analyzing the linear subspace property. This new framework gives *sharp* quantitative tradeoffs between the signal sparsity and the recovery accuracy of the ℓ_1 optimization for approximately sparse signals.

1.5.3 Weighted ℓ_1 Minimization Algorithm

The conventional approach to compressed sensing assumes no prior information on the unknown signal and a plain ℓ_1 minimization was used. In chapter 4, we consider a particular model for the sparse signal that assigns a probability of being zero or nonzero to each entry of the unknown vector. The standard compressed sensing model is therefore a special case where these probabilities are all equal.

We proposed to use weighted ℓ_1 minimization algorithm for signal recovery under this model. Assuming that the Gaussian measurement matrix ensemble is used, using the *null-space Grassmann angle*-based analytical framework, we are able to characterize the optimal recoverable sparsity thresholds, the optimal weights or the smallest number of measurement using weighted ℓ_1 minimization algorithms under the prior information.

1.5.4 An Analysis for Iterative Reweighted ℓ_1 Minimization Algorithm

Even though iterative reweighted ℓ_1 minimization algorithms or related algorithms [CWB08] have been empirically observed to boost the recoverable sparsity thresholds for certain types of signals, no rigorous theoretical results have been established to prove this fact. In chapter 5, we try to provide a theoretical foundation for analyzing the iterative reweighted ℓ_1 algorithms. Our idea is that some of the support set

information was able to be obtained in the process of iterations and this information can be utilized in the weighted ℓ_1 minimization to boost the sparse signal recovery.

In particular, we showed that for a nontrivial class of signals, the iterative reweighted ℓ_1 minimization can indeed deliver recoverable sparsity thresholds larger than the ℓ_1 minimization. Again, our results are based on the *null-space Grassmann angle*-based analytical framework and the Gaussian measurement matrix ensemble.

1.5.5 Null Space Conditions and Thresholds for Rank Minimization

Evolving from compressive sensing problems, we turned our attention to recovering matrices of low rank from compressed linear measurements in chapter 6.

Minimizing the rank of a matrix subject to constraints is a challenging problem that arises in many applications in machine learning, control theory, and discrete geometry. This class of optimization problems, known as rank minimization, is NP-HARD, and for most practical problems there are no efficient algorithms that yield exact solutions. A popular heuristic replaces the rank function with the nuclear norm—equal to the sum of the singular values—of the decision variable and has been shown to provide the optimal low rank solution in a variety of scenarios.

In chapter 6, we analyze the practical performance of this heuristic for finding the minimum rank matrix subject to linear constraints. Our starting point is the characterization of a necessary and sufficient condition that determines when this heuristic finds the minimum rank solution. We then obtain conditions, as a function of the matrix dimensions and rank and the number of constraints, such that our conditions for success are satisfied for almost all linear constraint sets as the matrix dimensions tend to infinity. Finally, we provide empirical evidence that these probabilistic bounds provide accurate predictions of the heuristic's performance in non-asymptotic scenarios.

Chapter 2

Expander Graphs for Compressive Sensing

As discussed, compressive sensing is an emerging technology which can recover a sparse signal vector of dimension n via a much smaller number of measurements than n. However, on the encoding side, there were no explicit constructions of "good" measurement matrices which come with guaranteed performances under computationally feasible and robust decoding methods [Tao07]. On the decoding side, the known decoding algorithms have relatively high recovery complexity, such as $O(n^3)$, or can only work efficiently when the signal is super sparse, sometimes without deterministic performance guarantees. In this chapter, we propose a compressive sensing scheme using measurement matrices constructed from expander graphs. It is the first of its kind that comes with *explicit* measurement matrix constructions, *deterministic* decoding performance guarantees with the capability of recovering signals with *linear* sparsity, where the number of non-zero elements k grows linearly with n, and, at the same time, with *linear* ($O(n)$) decoding complexity. When the number of nonzero elements k does not grow linearly with the dimension n, similar to the ℓ_1 minimization using dense random matrices, this scheme can exactly recover any k-sparse signal using only $O(k \log(n/k))$ measurements.

Our main approach is to construct measurement matrices from the adjacency matrix of an unbalanced expander graph. We further show that the expander graph based scheme is applicable to approximately sparse signals and robust to measurement

noise. Simulation results are given to show the performance and complexity of the new method and we also compare our work with recent works on expander graph based compressive sensing schemes.

2.1 Introduction

Compressive sensing has recently received a great amount of attention in the applied mathematics and signal processing community. The theory of compressive sensing, as developed over the past few years, attempts to perform sampling and compression simultaneously, thus significantly reducing the sampling rate. What allows this theory is the fact that, in many applications, signals of interest have a "sparse" representation over an appropriate basis. In fact, compressive sampling is intimately related to solving underdetermined systems of linear equations with sparseness constraints. The work of Candès, Romberg and Tao [CRT06, CT06] and Donoho [Don06c] came as a major breakthrough in that they rigorously demonstrated, for the first time, that, under some very reasonable assumptions, the solution could be found using simple linear programming—thus rendering the solution practically feasible. The method is essentially constrained ℓ_1 minimization, which for a long time was empirically known to perform well for finding sparse solutions and has been known in the literature as "Basis Bursuit" [Che95, CDS98]. Interestingly, the area of compressive sensing is closely connected to the related areas of coding [CT05], high-dimensional geometry [DT05a], sparse approximation theory [Don06a], data streaming algorithms [CM06, GSTV05] and random sampling [GGI$^+$02]. Furthermore, promising applications of compressive sensing are emerging in compressive imaging, medical imaging, sensor networks and analog-to-digital conversion [Can06].

While solving the linear program resulting from ℓ_1 optimization can be done in polynomial-time (often $O(n^3)$, where n is the number of unknowns), this may still be infeasible in applications where n is quite large (e.g., in current digital cameras the number of pixels is of the order $n = 10^6$ or more) [CR]. Therefore there is a need for methods and algorithms that are more computationally efficient. Also, in many of

the previous works, random measurement matrices are used where a successful signal recovery can not be always guaranteed although it succeeds with a high probability. So it is also desirable to have an explicit construction of a measurement matrix for compressive sensing [Can06, DeV07].

2.1.1 Related Works

Recently, some significant progress has been made in addressing these two problems for compressive sensing. Orthogonal Matching Pursuit (OMP) algorithms can be used as alternative recovery algrithms which require $O(nk^2)$ computations [TG], where k is the number of non-zero entries in the unknown vector; however, this may also be too high a complexity. Stage-wise OMP [DTDS07] has recently been proposed that solves the problem in $O(n \log n)$ computations. In [CM06] a certain sparse coefficient matrix has been used, along with group testing, that yields an algorithm with $O(k \log^2 n)$ complexity; however, this comes at the expense of more measurements—$O(k \log^2 n)$ measurements, as opposed to the $O(k \log n)$ measurements required of the aforementioned methods. Chaining pursuit has been introduced in [GSTV05], which has complexity $O(k \log^2 n \log^2 k)$ and also requires $O(k \log^2 n)$ measurements. From the number of measurements needed, we can see that both the group testing methods [CM06] and the chaining pursuit methods only work in the "supersparse" case, i.e., when the ratio k/n is very small—when k/n is increased, an enormous number of measurements is required, as noted in [SBB06a]. Motivated by low-density parity-check codes (LDPCs) a method called sudocodes has been proposed in [SBB06b] to recover sparse signal with high probability, which requires $O(k \log n \log k)$ recovery complexity, yet only $O(k \log n)$ measurements. The Homotopy methods are able to recover the sparse solutions by reducing the computational complexity from $O(n^3)$ to $O(nk^2)$ [DT06a]. In [Fuc04], it was shown that by using the Vandermonde measurement matrix and linear programming, one can recover k nonzero elements using approximately $2k$ measurements when the nonzero elements are restricted to positive numbers. In [AT07], motivated by Reed-Solomon codes rather than LDPC codes as

in [SBB06b], a scheme of recovery complexity $O(k^2)$ is proposed to recover any signal vector with k nonzero elements using the Vandermonde measurement matrix. List decoding was proposed for similar schemes in [PH08].

With the exception of the method in [DeV07], the group testing methods in [CM06] and the Vandermonde measurement matrix-based methods in [Fuc04, AT07], all the results described above hold with "high probability" either over the random measurement matrix or over some assumptions on the input signals [SBB06b]. While the methods in [DeV07, CM06] can guarantee sparse signal recovery deterministically with explicit measurement matrices, they suffer from the fact that they only work in the supersparse case where k can not be kept as a constant fraction of n. But recovering a constant fraction of n non-zero elements via a small number of measurements is of great practical interests [CT05]. For this reason, in this chapter, we will allow k to grow linearly in n, i.e., $k = \Theta(n)$. In this sparsity regime, the complexity of the methods of [Fuc04, AT07] are of order $O(n^3)$ and $O(n^2)$ respectively, which will still be impractical for problems of large dimensions. Sometimes, it is also required that the recovery schemes are applicable to approximately sparse signals and robust to the noise in the measurements and numerical errors.

2.1.2 Contributions

In this chapter, we propose a new scheme for compressive sensing with deterministic performance guarantees based on bipartite expander graphs and show that even with $k = \Theta(n)$, the recovery complexity of our algorithm is $O(n)$ while saving a constant fraction of n measurements. The new scheme thus pushes the performance and complexity bounds of compressive sensing to be asymptotically linear in n at the same time. By the time that this result was published [XH07a, XH07b], this had not been achieved by other methods. Bipartite expander graphs [CRVW02, SS96] are a certain class of graphs whose existence has been known for quite some time and whose recent explicit constructions are considered to be a major feat [CRVW02]. In some sense our approach is closest to that of [SBB06b], which is inspired by LDPCs, certain classes of

which are related to expander graphs [SS96], but in our works we provide performance guarantees. Preliminary analytic results further show the feasibility of application of the new method to approximately sparse signals and the noisy measurement cases. When the number of nonzero elements k does not grow linearly with the dimension n, similar to the ℓ_1 minimization using dense random matrices, this scheme can exactly recover any k-sparse signal using only $O(k \log(n/k))$ measurements.

2.1.3 Recent Developments

After the publication of our works [XH07a, XH07b], an explicit construction for compressive sensing matrices was given in [Ind08] which used extractors. But the construction in [Ind08] only works for recovering sparse signals with sublinear sparsity. In a more recent very interesting work [BGI$^+$08], it was shown that the expander graph-based measurement matrices can work with performance guarantees under the ℓ_1 minimization methods. Indyk and Ruzic [IR08], and Berinde, Indyk and Ruzic [BIR08] proposed new compressed sensing algorithms based on the properties of the expander graphs. Those algorithms are similar to the CoSaMP algorithm [NT08], from the orthogonal matching framework, and are designed to be robust against more general noise and compressible signals; however, this comes with a cost on complexity of the algorithm and its analysis. The algorithm that we proposed in this chapter is much simpler, and also the analysis on why the algorithm works is only based on the unique neighborhood properties of the expander graphs. In contrast, the other algorithms require expander graphs with stronger expansion and a complicated preprocessing step, and the analysis is based on more involved properties of expander graphs with larger expansions.

The rest of this chapter is organized as follows. In the next section we review the background and give the problem formulation. We introduce expander graphs in Section 2.3 and show how they can be used to develop deterministic methods with $O(n)$ recovery complexity. The analysis of the new compressive sensing scheme for approximately sparse signals is given in Section 2.4. RIP-1 property and full recovery

property are described in Section 2.6. Optimized expander graphs are used in Section 2.7 to reduce the number of iterations to $O(k)$. Simulation results are given in the section 2.8.

2.2 Background and Problem Formulation

In compressive sensing the starting point is an n-dimensional signal vector which admits a sparse representation in some particular basis. Since the basis is not of primary concern to us, we may, without loss of generality, assume that it is the standard basis. In other words, we shall assume that we have an n-dimensional vector $x \in \mathcal{R}^n$, such that no more than k entries are non-zero. Clearly, $k < n$. Here we assume k can be up to a constant fraction of n, since this case is of great practical interest [CT05].

The vector x itself is not directly observable. What is observable are *measurements* of x that correspond to linear combinations of the form

$$\sum_{j=1}^{n} a_j x_j. \tag{2.2.1}$$

We often have control over what measurements to employ, and this may turn out to help us. In any event, assuming we have m $(k < m < n)$ measurements of this form, we may collect them in a $m \times n$ matrix A so that

$$y = Ax, \tag{2.2.2}$$

or, in other words,

$$y_i = \sum_{j=1}^{n} A_{ij} x_j, \qquad i = 1, \ldots, m. \tag{2.2.3}$$

The system of equations (2.2.2) is, of course, underdetermined. However, the fact that a sparse solution exists, allows us to be able to find the solution. It was a significant result when it was rigorously shown by Candès, Romberg and Tao [CRT06,

CT06] and Donoho [Don06c] that, under the sparsity assumption, the solution could be found via solving the ℓ_1 optimization problem

$$\min_{x, Ax=y} \|x\|_1, \tag{2.2.4}$$

where $\|x\|_1 = \sum_{i=1}^{n} |x_i|$ is the ℓ_1-norm of the vector x. The upshot is that something that appeared to be practically infeasible can now be potentially computed. For example, in [CT05], it was shown that if the measurement matrix A satisfies the restricted isometry conditions, then the ℓ_1 minimization can recover a vector with up to k nonzero elements, where k is a constant fraction of n.

In spite of the recent developments, one unanswered question is whether we can develop compressive sensing schemes and recovery algorithms with complexity $O(n)$ even when $k = \Theta(n)$? If yes, can one *explicitly* develop constructions of measurement matrices that *deterministically* guarantee finding the optimal solution for all signal instances in such schemes, provided the vector x is sparse enough? We shall presently answer both questions in the affirmative and discuss all these developments in the next section.

2.3 Expander Graphs and Efficient Algorithms

2.3.1 Expander Graphs

Expander graphs can be defined for arbitrary graphs, however, here we shall restrict ourselves to *bipartite* graphs. For a bipartite graph, we have two types of nodes. Following coding theory parlance, we will call one type left *variable nodes* of which there are n and which correspond to the entries of x, and right *parity check nodes* of which there are m and which correspond to the entries of y (or the measurements). We assume that $n \geq m$. In a bipartite graph, connections within the variable nodes and within the parity check nodes are not allowed. The existence of edges between the different variables and parity check nodes are represented by a $m \times n$ matrix A.

In particular,

$$A_{ij} = \begin{cases} 1 & \text{if right node } i \text{ connected to left node } j \\ 0 & \text{otherwise} \end{cases} \tag{2.3.1}$$

for $i = 1, \ldots, m$ and $j = 1, \ldots, n$. In what follows we shall use the matrix thus obtained from a suitably chosen bipartite graph as the measurement matrix for compressive sampling.

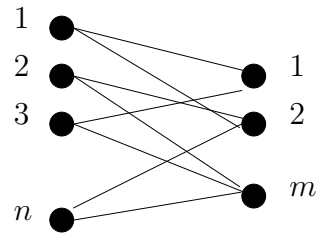


Figure 2.1: A bipartite graph

A bipartite graph will be said to have *regular left degree c* if the number of edges emanating from each variable node is c.

Definition 2.3.1 (Expander)**.** *A bipartite graph with n variable nodes, m parity check nodes and regular left degree c will be called a $(\alpha n, \beta c)$ expander, for some $0 < \alpha, \beta < 1$, if for every subset of variable nodes $\mathcal{V}$ with cardinality less than or equal to αn, i.e., $|\mathcal{V}| \leq \alpha n$, the number of neighbors connected to $\mathcal{V}$ is larger than $\beta c|\mathcal{V}|$, i.e., $|\mathcal{N}(\mathcal{V})| > \beta c|\mathcal{V}|$, where $\mathcal{N}(V)$ is the set of neighbors of $\mathcal{V}$.*

Here we assume that each righthand side node also has a regular degree d, where $cn = md$. The existence of expander graphs has been known for quite some time since the work of Pinsker and Bassylago [BP73], who used probabilistic arguments to prove their existence. Expander graphs arise in many applications: fast, distributed routing algorithms [PU89], LDPC codes [SS96], storage schemes [UW87], and proof systems [BSW99], to name a few. An explicit construction of constant regular left degree *lossless* (with β arbitrarily close to 1) expander graph is recently given in

[CRVW02]. An existence result, which holds for the setting we are interested in, is the following [BM01]:

Theorem 2.3.2. *Let $0 < \beta < 1$ and the ratio $r = \frac{m}{n}$ be given. Then for large enough n there exists a regular left degree c and a regular right degree d bipartite expander $(\alpha n, \beta c)$ for some $0 < \alpha < 1$ and some constant (not growing with n) c.*

2.3.2 The Main Algorithm

We are now in a position to describe our main algorithm. We begin with $\beta = \frac{3}{4}$ and some fixed $r = \frac{m}{n}$. (Thus, our number of measurements is $m = nr$. We can use the construction of [CRVW02], or any other recent one, to construct an expander with some $0 < \alpha < 1$ and constant c.) Denote the resulting measurement matrix by A. In particular, assuming $x \in \mathcal{R}^n$ is sparse with at most k nonzero entries, we perform the m measurements

$$y = Ax. \tag{2.3.2}$$

We will assume that

$$k \leq \frac{\alpha n}{2}. \tag{2.3.3}$$

We need one further notation: given an estimate $\hat{x}$ of x, we define as the *gap* in the i-th equation the quantity

$$g_i = y_i - \sum_{j=1}^{n} A_{ij}\hat{x}_j. \tag{2.3.4}$$

Algorithm 1 is incredibly simple. What is remarkable about it is that, in step 2 of the algorithm, if $y \neq A\hat{x}$ one can always find a variable node with the property that $c' > \frac{c}{2}$ among the measurement equations it participates in have *identical* nonzero gap g. Furthermore, the algorithm terminates in at most ck steps. We proceed to establish these two claims via a series of lemmas. At any step of the algorithm, let $\mathcal{S}$ denote the set

$$\mathcal{S} = \{j | \hat{x}_j \neq x_j\}. \tag{2.3.5}$$

Algorithm 1

1: Start with $\hat{x} = 0_{n\times 1}$.
2: **if** $y = A\hat{x}$ **then**
3: declare $\hat{x}$ the solution and exit.
4: **else**
5: find a variable node, say $\hat{x}_j$, such that of the c measurement equations it participates in $c' > \frac{c}{2}$ of them have an identical *nonzero* gap g.
6: Set $\hat{x}_j = \hat{x}_j + g$. Go to 2.
7: **end if**

Lemma 2.3.3 (Initialization)**.** *When $\hat{x} = 0$,$y \neq A\hat{x}$ and $k \leq \frac{\alpha n}{2}$, there always exists a variable node such that $c' > \frac{c}{2}$ of the measurement equations it participates in has identical nonzero gap g.*

Proof: Initially since $\hat{x}_i = 0$, the set $\mathcal{S}$ has cardinality $|\mathcal{S}| = k \leq \alpha n/2$. We can therefore apply the property of the expander with $\beta = \frac{3}{4}$ to $\mathcal{S}$ to conclude that

$$|\mathcal{N}(\mathcal{S})| > \frac{3}{4} c\,|\mathcal{S}|\,. \tag{2.3.6}$$

Let us now divide the set $\mathcal{N}(\mathcal{S})$ into two disjoint sets: $\mathcal{N}_{unique}(\mathcal{S})$ comprised of those elements of $\mathcal{N}(\mathcal{S})$ that are connected to only one edge emanating from $\mathcal{S}$ and $\mathcal{N}_{>1}(\mathcal{S})$ which are the remaining elements of $\mathcal{N}(\mathcal{S})$ that are connected to more than one edges emanating from $\mathcal{S}$. Clearly, (2.3.6) implies

$$|\mathcal{N}_{unique}(\mathcal{S})| + |\mathcal{N}_{>1}(\mathcal{S})| > \frac{3}{4} c\,|\mathcal{S}|\,. \tag{2.3.7}$$

Counting the edges emanating from $\mathcal{S}$ leads to

$$|\mathcal{N}_{unique}(\mathcal{S})| + 2\,|\mathcal{N}_{>1}(\mathcal{S})| \leq c\,|\mathcal{S}|\,, \tag{2.3.8}$$

since the total number of edges is $c\,|\mathcal{S}|$ and since some of the nodes in $\mathcal{N}_{>1}(\mathcal{S})$ may have more than 2 edges connecting to $\mathcal{S}$. Eliminating $|\mathcal{N}_{>1}(\mathcal{S})|$ from the inequalities (2.3.7) and (2.3.8) yields

$$|\mathcal{N}_{unique}(\mathcal{S})| > \frac{c}{2}\,|\mathcal{S}|\,. \tag{2.3.9}$$

The above inequality implies that there must be at least one element of $\mathcal{S}$ that is connected to $c' > \frac{c}{2}$ elements of $\mathcal{N}_{unique}(\mathcal{S})$. But since this is the only element of $\mathcal{S}$ connected to these c' measurements, and since the A_{ij}'s are all 1 for the edges connecting these nodes, they must all have the same nonzero gap g. ■

We now need another definition. At any step of the algorithm, let $\mathcal{T}$ denote the set

$$\mathcal{T} = \left\{ i | y_i \neq \sum_{j=1}^{n} A_{ij} x_j \right\}. \tag{2.3.10}$$

Lemma 2.3.4 (Decrease in $|\mathcal{T}|$)**.** *After the first step of the algorithm, the cardinality of the set $\mathcal{T}$ decreases at least by 1.*

Proof: According to the proof of Lemma 2.3.3, we have found a variable node with $c' > \frac{c}{2}$ measurements with identical nonzero gap g. Setting $\hat{x}_j = \hat{x}_j + g$ sets the gap on these c' equations to zero. However, it may make some zero gaps on the remaining $c - c'$ measurements nonzero. Nonetheless, since $c' - (c - c') = 2c' - c \geq 1$ (note that $c' - c/2 \geq \frac{1}{2}$) the cardinality of $\mathcal{T}$ decreases at least by one.

■

We can now proceed to the main induction argument.

Lemma 2.3.5 (Induction)**.** *Consider a regular left degree c bipartite graph with n variable nodes and m parity check nodes. Assume further that the graph is an $(\alpha n, \frac{3}{4}c)$ expander and consider Algorithm 1. If for all iterations of the algorithm up to step l:*

(1) $\left|\mathcal{S}^{(l')}\right| < \alpha n$, $l' = 1, \ldots, l$, where $\mathcal{S}^{(l')}$ is the same definition as in (2.3.5), except for at the l'-th iteration.

(2) There always exists a variable node such that $c' > \frac{c}{2}$ of the measurement equations it participates in have identical nonzero gap g.

(3) $\left|\mathcal{T}^{(l')}\right| \leq \left|\mathcal{T}^{(l'-1)}\right| - 1$, for $l' = 1, \ldots, l$, where $\mathcal{T}^{(l')}$ is the same as in the definition (2.3.10), except at the l'-th iteration.

Then at the $(l+1)$-th iteration we have

(i) $\left|\mathcal{S}^{(l+1)}\right| < \alpha n$

(ii) If $y \neq A\hat{x}$*, there always exists a variable node such that* $c' > \frac{c}{2}$ *of the measurement equations it participates in have identical nonzero gap* g*.*

(iii) $\left|\mathcal{T}^{(l+1)}\right| \leq \left|\mathcal{T}^{(l)}\right| - 1$

Proof: Let us begin with claim (ii). The argument is very similar to that of the proof of Lemma 2.3.3, which we essentially repeat here. Due to assumption (1) in the lemma, $\left|\mathcal{S}^{(l)}\right| < \alpha n$. Therefore we can apply the property of the expander with $\beta = \frac{3}{4}$ to $\mathcal{S}^{(l)}$ to conclude that

$$\left|\mathcal{N}(\mathcal{S}^{(l)})\right| > \frac{3}{4} c \left|\mathcal{S}^{(l)}\right|. \tag{2.3.11}$$

As before, we divide the set $\mathcal{N}(\mathcal{S}^{(l)})$ into two disjoint sets: $\mathcal{N}_{unique}(\mathcal{S}^{(l)})$ comprised of those elements of $\mathcal{N}(\mathcal{S}^{(l)})$ that are connected to only one edge of $\mathcal{S}^{(l)}$ and $\mathcal{N}_{>1}(\mathcal{S}^{(l)})$ which are the remaining elements of $\mathcal{N}(\mathcal{S}^{(l)})$ that are connected to more than one edges emanating from $\mathcal{S}^{(l)}$. Clearly, (2.3.11) implies

$$\left|\mathcal{N}_{unique}(\mathcal{S}^{(l)})\right| + \left|\mathcal{N}_{>1}(\mathcal{S}^{(l)})\right| > \frac{3}{4} c \left|\mathcal{S}^{(l)}\right|. \tag{2.3.12}$$

Counting the edges emanating from $\mathcal{N}(S^{(l)})$ leads to

$$\left|\mathcal{N}_{unique}(\mathcal{S}^{(l)})\right| + 2\left|\mathcal{N}_{>1}(\mathcal{S}^{(l)})\right| \leq c \left|\mathcal{S}^{(l)}\right|, \tag{2.3.13}$$

since the total number of edges is $c\left|\mathcal{S}^{(l)}\right|$ and since some of the nodes in $\mathcal{N}_{>1}(\mathcal{S}^{(l)})$ may have more than 2 nodes emanating from them. Eliminating $\left|\mathcal{N}_{>1}(\mathcal{S}^{(l)})\right|$ from the inequalities (2.3.12) and (2.3.13) yields

$$\left|\mathcal{N}_{unique}(\mathcal{S}^{(l)})\right| > \frac{c}{2} \left|\mathcal{S}^{(l)}\right|, \tag{2.3.14}$$

which implies that there must be at least one element of $\mathcal{S}^{(l)}$ that is connected to $c' > \frac{c}{2}$ elements of $\mathcal{N}_{unique}(\mathcal{S}^{(l)})$. But since this is the only element of $\mathcal{S}^{(l)}$ connected

to these c' nodes, and since the A_{ij}'s are all 1 for the edges connecting these nodes, they must all have the same nonzero gap g.

This establishes (ii). Establishing (iii) is similar to the proof of Lemma 2.3.4. We have already found a variable node with $c' > \frac{c}{2}$ measurements with identical nonzero gap g. Setting $\hat{x}_j^{(l+1)} = \hat{x}_j^{(l)} + g$ sets the gap on these c' equations to zero. However, it may make some zero gaps on the remaining $c-c'$ measurements nonzero. Nonetheless, since $c' - (c - c') = 2c' - c \geq 1$ (note that $c' - c/2 \geq \frac{1}{2}$), the cardinality of $\mathcal{T}^{(l+1)}$ decreases at least by one compared to $\mathcal{T}^{(l)}$.

This establishes (iii). We finally turn to (i). Note that, since in each iteration of Algorithm 1 we change the value of only one entry of $\hat{x}$, the cardinality of the set $\mathcal{S}^{(l')}$ can change at most by one. Since, due to assumption (1) of the lemma we have $\mathcal{S}^{(l)} < \alpha n$, (iii) can only be violated if $\mathcal{S}^{(l+1)} = \alpha n$. Let us assume this and arrive at a contradiction. Note that we can apply the property of the expander with $\beta = \frac{3}{4}$ to the set $\mathcal{S}^{(l+1)}$ to obtain

$$\left|\mathcal{N}(\mathcal{S}^{(l+1)})\right| > \frac{3}{4} c\alpha n. \tag{2.3.15}$$

Once again, we divide the set $\mathcal{N}(\mathcal{S}^{(l+1)})$ into two disjoint sets: $\mathcal{N}_{unique}(\mathcal{S}^{(l+1)})$ and $\mathcal{N}_{>1}(\mathcal{S}^{(l+1)})$. Clearly, (2.3.15) implies

$$\left|\mathcal{N}_{unique}(\mathcal{S}^{(l+1)})\right| + \left|\mathcal{N}_{>1}(\mathcal{S}^{(l+1)})\right| > \frac{3}{4} c\alpha n. \tag{2.3.16}$$

Counting the edges emanating from $\mathcal{N}(S^{(l+1)})$ leads to

$$\left|\mathcal{N}_{unique}(\mathcal{S}^{(l+1)})\right| + 2\left|\mathcal{N}_{>1}(\mathcal{S}^{(l+1)})\right| \leq c\alpha n \tag{2.3.17}$$

since the total number of edges is $c\alpha n$ and since some of the nodes in $\mathcal{N}_{>1}(\mathcal{S}^{(l)})$ may have more than 2 nodes emanating from them. (2.3.16) and (2.3.17) imply

$$\left|\mathcal{N}_{unique}(\mathcal{S}^{(l+1)})\right| > \frac{c}{2} \alpha n. \tag{2.3.18}$$

Since the nodes in $\mathcal{N}_{unique}(\mathcal{S}^{(l+1)})$ are connected to unique elements in $\mathcal{S}^{(l+1)}$, we

conclude that $\mathcal{N}_{unique}(\mathcal{S}^{(l+1)}) \subseteq \mathcal{T}^{(l+1)}$. This in turn implies that

$$\left|\mathcal{T}^{(l+1)}\right| > \frac{c}{2}\alpha n. \tag{2.3.19}$$

Note, however, that since $k \leq \alpha n/2$ and the left degree of the graph is c, at the beginning of the algorithm we have $\left|\mathcal{T}^{(0)}\right| \leq \frac{c}{2}\alpha n$. However, from assumption (3) and property (iii), which we just established, we know that $\left|\mathcal{T}^{(l')}\right|$ is a decreasing function for all $l' \leq l+1$. Therefore,

$$\left|\mathcal{T}^{(l+1)}\right| < \left|\mathcal{T}^{(0)}\right| \leq \frac{c}{2}\alpha n, \tag{2.3.20}$$

which contradicts (2.3.19). This establishes (i) and hence all claims of the lemma. ■

The above sequence of lemmas establishes the following main result regarding Algorithm 1.

Theorem 2.3.6 (Validity of Algorithm 1)**.** *Consider a regular left degree bipartite graph with n variable nodes and m parity check nodes. Assume further that the graph is an $(\alpha n, \frac{3}{4}c)$ expander and consider its corresponding A matrix. Let $x \in \mathcal{R}^n$ be an arbitrary vector with at most $k \leq \alpha n/2$ nonzero entries and consider the m measurements*

$$y = Ax. \tag{2.3.21}$$

Then Algorithm 1 finds the value of x in at most $kc \leq \frac{c}{2}\alpha n$ iterations. If we assume that the bipartite graph has a regular right degree, we will have a recovery algorithm with complexity linear in n.

Proof: The theorem has essentially been proven in Lemmas 2.3.3, 2.3.4 and 2.3.5. We essentially have shown that at each iteration the cardinality of the set $\mathcal{T}^{(l)}$ decreases by at least one. Since the initial cardinality is *at most* kc, $\mathcal{T}^{(l)}$ will be empty after *at most* kc steps. But, of course, an empty $\mathcal{T}^{(l)}$ implies that the algorithm has found x (This is because in this process $\mathcal{S}$ is always smaller than αn and we can see that a non-zero vector x' satisfying $Ax' = 0$ must have larger than αn nonzero elements following

essentially the same arguments as in the proof of Lemma 3). If the bipartite graph has a regular right degree, then in each iterative step of algorithm 1, we only need a fixed number of operations to update the variable nodes and its related measurements by keeping track of the list of variable nodes. ■

Remarks

Here we can allow for $k = \Theta(n)$ nonzero entries in x since α is a constant (not going to zero as n grows) which depends on the expander graph. The number of measurements is $m = rn$, where r can take any value from $(0, 1)$ and determines the value of α.

2.4 Expander Graphs for Approximately Sparse Signals

In this section, we will give preliminary analytic results on expander graph-based compressive sensing for approximately sparse signals. In an approximately sparse signal vector, only a few signal entries are significant and the remaining signal entries are near zero but possibly not exactly zero. In practice, the approximately sparse model is a more realistic model for signals. Here we use the same measurement matrix as in the previous section except that we apply it to approximately sparse signals. We also assume a two-level ("near-zero" and "significant") signal model for the approximately sparse signal vector. (Of course, this is a coarse signal model, but it captures the nature of approximately sparse signal vectors.) The entries of the "near-zero" level in the signal vector are near-zero elements taking values from the set $[-\lambda, +\lambda]$ while the "significant" level of entries take values from the set $\{x|(L-\Delta) \leq |x| \leq (L+\Delta)\}$, where $L > \Delta$ and $L > \lambda$. Let $\rho = max\{2\Delta, \lambda\}$ and d be the regular right check node degree. Now we apply the following signal recovery algorithm to y with the measurement matrix A.

The following theorem establishes the validity of Algorithm 2.

Theorem 2.4.1 (Validity of Algorithm 2)**.** *Consider a bipartite graph with n variable*

Algorithm 2

1. Start with $\hat{x} = 0_{n\times 1}$.

2. If $\|y - A\hat{x}\|_\infty \leq \rho d$, determine the positions and signs of the significant components in x as the positions and signs of the non-zero signal components in $\hat{x}$; exit.

   Else, find one variable node, say $\hat{x}_j$, such that of the c measurement equations it participates in $c' > \frac{c}{2}$ of them are in either of the following categories:

   (a) They have gaps which are of the same sign and have absolute values between $L-\Delta-\lambda-\rho(d-1)$ and $L+\Delta+\lambda+\rho(d-1)$. Moreover, there exists a number t in the set $\{x|x = 0, |x| = (L-\Delta), |x| = (L+\Delta)\}$ such that $|y - A\hat{x}|$ are all $\leq \rho d$ over these c' measurements if we change $\hat{x}_j$ to t.

   (b) They have gaps which are of the same sign and have absolute values between $2L - 2\Delta - \rho(d-1)$ and $2L + 2\Delta + \rho(d-1)$. Moreover, there exist a number t in the set $\{x|x = 0, |x| = (L-\Delta), |x| = (L+\Delta)\}$ such that $|y - A\hat{x}|$ are all $\leq \rho d$ over these c' measurements if we change $\hat{x}_j$ to t.

3. Reset $\hat{x}_j = t$. Go to 2.

nodes and m parity check nodes. Assume further that the graph is an $(\alpha n, \frac{3}{4}c)$ expander with regular right degree d and regular left degree c. Denote the corresponding measurement matrix as A. Let $x \in \mathcal{R}^n$ be an arbitrary vector with at most $k \leq \alpha n/2$ significant signal components and assume that $max\{\rho(2d-1) + \Delta + \lambda, \rho(2d-2) + 3\Delta + \lambda\} < L$. Consider the m measurements

$$y = Ax. \tag{2.4.1}$$

Then Algorithm 2 correctly finds the sign and positions of the significant components of x in at most $kc \leq \frac{c}{2}\alpha n$ iterations with complexity linear in n.

Proof: The arguments here basically follow the same reasoning as in the proof of Lemma 2.3.3, Lemma 2.3.4, Lemma 2.3.5 and Theorem 2.3.6. But now we define the set $\mathcal{S}$ as the set of variable nodes j's such that x_j and $\hat{x}_j$ are on different signal levels or have opposite signs while both being on the "significant" signal level. If a variable

node $j \in S$, then $L-\Delta-\lambda \leq |x_j - \hat{x}_j| \leq L+\Delta+\lambda$ or $2(L-\Delta) \leq |x_j - \hat{x}_j| \leq 2(L+\Delta)$. Also notice that $|x_j - \hat{x}_j| \leq \rho$ if x_j and $\hat{x}_j$ are both in the near-zero signal level or have the same sign while both being on the "significant" signal level. Define the set $\mathcal{T}$ as the set of measurements where $|y - A\hat{x}|$ have values larger than ρd. Notice that after each iteration, we can always decrease the cardinality of $\mathcal{T}$ by at least 1. ■

Now let us consider the case where the measurements themselves are not perfect and corrupted by additive noises. In this case, we have

$$y = Ax + w, \tag{2.4.2}$$

where w is a m-dimensional noise vector. We assume $|w|_\infty \leq \varepsilon$ and that x is generated according to the same approximately sparse signal model as stated previously. Then the previous algorithm and can be extended to the noisy measurements cases.

Algorithm 3

1. Start with $\hat{x} = 0_{n\times 1}$.
2. If $\|y - A\hat{x}\|_\infty \leq \rho d + \varepsilon$, determine the positions and signs of the significant components in x as the positions and signs of the non-zero signal components in $\hat{x}$; exit.

   Else, find one variable node, say $\hat{x}_j$, such that of the c measurement equations it participates in $c' > \frac{c}{2}$ of them are in either of the following categories:

   (a) They have gaps which are of the same sign and have absolute values between $L-\Delta-\lambda-\rho(d-1)-\varepsilon$ and $L+\Delta+\lambda+\rho(d-1)+\varepsilon$. Moreover, there exists a number t in the set $\{x|x=0, |x|=(L-\Delta), |x|=(L+\Delta)\}$ such that $|y - A\hat{x}|$ are all $\leq \rho d + \varepsilon$ over these c' measurements if we change $\hat{x}_j$ to t.

   (b) They have gaps which are of the same sign and have absolute values between $2L - 2\Delta - \rho(d-1) - \varepsilon$ and $2L + 2\Delta + \rho(d-1) + \varepsilon$. Moreover, there exist a number t in the set $\{x|x=0, |x|=(L-\Delta), |x|=(L+\Delta)\}$ such that $|y - A\hat{x}|$ are all $\leq \rho d + \varepsilon$ over these c' measurements if we change $\hat{x}_j$ to t.

3. Reset $\hat{x}_j = t$. Go to 2.

The following theorem establishes the validity of Algorithm 3 in the case of ap-

proximately sparse signals and noisy measurements.

Theorem 2.4.2 (Validity of Algorithm 2)**.** *Consider a bipartite graph with n variable nodes and m parity check nodes. Assume further that the graph is an $(\alpha n, \frac{3}{4}c)$ expander with regular right degree d and regular left degree c. Denote the corresponding measurement matrix as A. Let $x \in \mathcal{R}^n$ be an arbitrary vector with at most $k \leq \alpha n/2$ significant signal components and assume that $max\{\rho(2d-1)+\Delta+\lambda+2\varepsilon, \rho(2d-2)+3\Delta+\lambda+2\varepsilon\} < L$. Consider the m measurements*

$$y = Ax. \tag{2.4.3}$$

Then Algorithm 2 correctly finds the sign and positions of the significant components of x in at most $kc \leq \frac{c}{2}\alpha n$ iterations with complexity linear in n.

Proof: The arguments here basically follow the same reasoning as in the proof of Lemma 2.3.3, Lemma 2.3.4, Lemma 2.3.5 and Theorem 2.3.6. But now we define the set $\mathcal{S}$ as the set of variable nodes j's such that x_j and $\hat{x}_j$ are on different signal levels (one is on the "near-zero signal" level and the other is on the "significant signal" level) or have opposite signs while both being on the "significant" signal level. Suppose for a variable node $j \in S$, x_j and $\hat{x}_j$ are on different signal levels, then $L-\Delta-\lambda \leq |x_j - \hat{x}_j| \leq L+\Delta+\lambda$ or $2(L-\Delta) \leq |x_j - \hat{x}_j| \leq 2(L+\Delta)$. Also notice that $|x_j - \hat{x}_j| \leq \rho$ if x_j and $\hat{x}_j$ are both on the near-zero signal level or have the same sign while both being on the "significant" signal level. If $\rho(2d-1)+\Delta+\lambda+2\varepsilon < L$ and $\rho(2d-2)+3\Delta+\lambda+2\varepsilon < L$ respectively, we will respectively have

$$L-\Delta-\lambda-\rho(d-1)-\varepsilon > \rho d+\varepsilon; \tag{2.4.4}$$

and

$$2L-2\Delta-\rho(d-1)-\varepsilon > L+\Delta+\lambda+\rho(d-1)+\varepsilon. \tag{2.4.5}$$

Under these conditions, we can distinguish the case (a) and case (b) in Algorithm 3. Moreover, under these conditions, if $0 < |\mathcal{S}| < \alpha$n , there must be one variable

node lying in category (a) or category (b) (following the same arguments as in Lemma 2.3.3). Define the set $\mathcal{T}$ as the set of measurements where $\|y - A\hat{x}\|$ have values larger than $\rho d + \varepsilon$. Notice that after each iteration, we can always decrease the cardinality of $\mathcal{T}$ by at least 1. By similar arguments from Lemma 3, we will have each component of $\hat{x}$ and x belonging to the same signal level (if they are both on the "significant signal" level, they will have the same sign). ∎

Also, after knowing the signs and locations of significant components, the estimation for their amplitudes can be further refined by using other techniques like minimum mean square error estimations. Actually, simulation results in Section 2.8 we show that a slight modification of algorithm 1 is very effective even in the case of random Gaussian noises and approximately sparse signal.

2.5 Sufficiency of $O(k \log(\frac{n}{k}))$ Measurements

In the previous parts, we assume that the number of nonzero elements k in a sparse signal vector grows linearly with n and the number of measurements needed in compressive sensing need to grow linearly with n, too. However, in some cases, the number of nonzero elements k remains fixed while the dimension of the signal vector n can grow arbitrarily large. For the ℓ_1-minimization framework, it has been shown that $O(k \log(n/k))$ measurements suffice for perfectly recovering a sparse signal vector of dimension n with no more than k non-zero elements. In this part, we will show that only $O(k \log(n/k))$ measurements are needed in order to perfect recovering all k-sparse signal when n goes large while requiring much lower recovery complexity. Before going to the precise statement and formal proof, we should notice that in Section 2, the signal recovery mechanism still works as long as the parameters α β and c remain fixed for a fixed n even if they are a function of n as n grows. From the results of previous parts, to recover any k-sparse signal, we need a $(k, \frac{3}{4}c)$ bipartite expander graph with m measurements. So by showing the existence of such an expander graph with $m = O(k \log(n/k))$, we actually show that for any k, $O(k \log(n/k))$ measure-

ments are enough for recovering any k-sparse signal with deterministic guarantees even as n grows. Before showing this, in the following theorem, we will give a lower bound on the number of measurements, namely m, in order to make an expander graph possible. Please note that this lower bound is a general result in the sense that it is also true for expander graphs with irregular right degrees.

Theorem 2.5.1 (Lower Bound on the Number of Measurements to Make an Expander Graph)**.** *Consider a bipartite graph with n variable nodes and m measurement nodes. Assume further that the graph is a $(k, \frac{3}{4}c)$ expander graph with regular left degree c. Then m must satisfy $\binom{m}{\frac{3}{4}ck}/\binom{m-c}{\frac{3}{4}ck-c} > n/k$.*

Proof: We prove this theorem by 'double counting'. In order for a bipartite graph to be a $(k, \frac{3}{4}c)$ expander, every $\frac{3}{4}ck$ measurement vertices must 'dominate' less than k variable nodes. Here we say a measurement set Ω dominates a variable node v if v is not connected to measurement nodes outside Ω. We now double count the number of 2-tuple pairs (Ω, v), where Ω is any set of measurement nodes of cardinality $\frac{3}{4}ck$ and v is a variable vertex dominated by the set Ω.

Notice that there are in total $\binom{m}{\frac{3}{4}ck}$ measurement node set Ω with cardinality $\frac{3}{4}ck$ and for the j-th $(1 \leq j \leq \binom{m}{\frac{3}{4}ck})$ such set Ω_j, we denote the set of variable nodes that are dominated by Ω_j as V_j. So the total number of 2-tuple pairs (Ω, v) is $\sum_{j=1}^{\binom{m}{\frac{3}{4}ck}} |V_j|$. Now let us count the number of 2-tuple pairs (Ω, v) from the perspective of variable nodes. For the i-th variable node v_i, there are $\binom{m-l_i}{\frac{3}{4}ck-l_i}$ measurement node sets Ω of cardinality $\frac{3}{4}ck$ that dominate v_i, where l_i $(1 \leq l_i \leq c)$ is the number of measurement nodes that the variable node v_i is connected to. So the total number of 2-tuple pairs (Ω, v) is also equal to $\sum_{i=1}^{n} \binom{m-l_i}{\frac{3}{4}ck-l_i}$, which is no smaller than $\binom{m-c}{\frac{3}{4}ck-c}n$. For an $(k, \frac{3}{4}c)$ expander graph, $\sum_{j=1}^{\binom{m}{\frac{3}{4}ck}} |V_j| < k \times \binom{m}{\frac{3}{4}ck}$ because each set Ω dominates less than k variable nodes. By combining the results of double counting, we have$\binom{m-c}{\frac{3}{4}ck-c}n < k \times \binom{m}{\frac{3}{4}ck}$. This proves Theorem 2.5.1. ∎

Lemma 2.5.2 (Constant Left Degree not Achieving the $O(k\log(n))$ bound)**.** *Consider a bipartite graph with n variable nodes and m measurement nodes. Assume*

further that the graph is a $(k, \frac{3}{4}c)$ *expander graph with regular left degree* c*. If* $m = O(k \log(n/k))$*, then* c *can not be a constant independent of* n*.*

Proof: It is straightforward from Theorem 2.5.2 that $m \geq (\frac{n}{k})^{\frac{1}{c}}$, which is a polynomial over n. ■

In the following part, we will give the main result of this section.

Theorem 2.5.3 (The Sufficiency of $O(k \log(n/k))$ Measurements)**.** *Consider regular bipartite graphs with* n *variable nodes and* m *measurement nodes. Assume that they have regular left degree* c *and regular right degree* d*. For any* k*, if* n *is large enough, there exists a regular* $(k, \frac{3}{4}c)$ *expander bipartite graph with* $m = O(k \log(n/k))$ *for some number* c *(Note that the left-degree* c *depends on* n*). Let* $x \in \mathcal{R}^n$ *be an arbitrary vector with at most* $\frac{k}{2}$ *nonzero entries and consider the* m *measurements*

$$y = Ax. \tag{2.5.1}$$

Then Algorithm 1 finds the value of x *in at most* $\frac{kc}{2}$ *iterations.*

Proof: We show the existence of the expander graphs stated in Theorem 2.5.3. Then the signal recovery performance statement in Theorem 2.5.3 follows from the existence of such an expander graph and Theorem 2.3.6. In proving the existence of such an expander graph, we show that a regular bipartite graph randomly generated in a certain way will be a $(k, \frac{3}{4}c)$ expander graph with probability approaching 1 as n goes large.

Here we take $c = C \log(n/k)$ and $m = Dk \log(n/k)$, where C and D are constants independent of k and n and will be specified later. Consider the bipartite graph as shown in Figure 2.1. For the time being, we assume that $C \leq D$. So, in total, we have

$$T_E = (C \log(n/k)) \times n \tag{2.5.2}$$

edges emanating from the n variable nodes. We generate a random permutation of these $(C \log(n/k)) \times n$ emanating edges with a uniform distribution (over all

the possible permutations) and connect ('plug') these $(C\log(n/k)) \times n$ edges to the $(C\log(n/k)) \times n$ 'sockets' on the $Dk\log(n/k)$ parity check nodes according to the randomly generated permutation. So the number of edges each measurement node connects is

$$d = (C\log(n/k)) \times n/m = \frac{Cn}{Dk}. \tag{2.5.3}$$

Take an arbitrary variable node set S of cardinality k and consider the random variable Y, which is the number of check nodes connected to S in this randomly generated graph. Obviously,

$$Y = \sum_{i=1}^{kC\log(n/k)} I_i, \tag{2.5.4}$$

where I_i is the indicator function of whether the i-th edge is connected to a check node which is not connected to any of the previous $(i-1)$ edges. Suppose the previous $(i-1)$ edges are connected to L_{i-1} measurement nodes, then I_i takes the value '1' with probability

$$\frac{T_E - d \times L_{i-1}}{T_E - L_{i-1}}, \tag{2.5.5}$$

whatever measurement nodes the previous $(i-1)$ edges are connected to. Since any $(i-1)$ edges are connected to at most $(i-1)$ measurement nodes and $i \leq C\log(n)$, we have

$$\frac{T_E - d \times L_{i-1}}{T_E - L_{i-1}} \geq \frac{T_E - d \times (k \times C\log(n/k))}{T_E - (k \times C\log(n/k))}. \tag{2.5.6}$$

So the probability that $(1 - I_i)$ takes the value '1' is at most

$$1 - \frac{T_E - d \times (k \times C\log(n/k))}{T_E - (k \times C\log(n/k))} = \frac{\frac{Cn}{D} - k}{n - k} \leq \frac{C}{D}, \tag{2.5.7}$$

whatever $I_j, 1 \leq j \leq (i-1)$, are.

Define a new random variable

$$Z = kC\log(n/k) - Y = \sum_{i=1}^{kC\log(n/k)} (1 - I_i), \tag{2.5.8}$$

and consider another random variable

$$Z' = \sum_{i=1}^{kC\log(n)} b_i, \tag{2.5.9}$$

where b_i's are independent binary Bernoulli random variables of parameter $\frac{C}{C}$ (taking the value '1' with probability $\frac{C}{D}$ and taking the value '0' with probability $1-\frac{C}{D}$). Then the probability that $Z \geq \frac{1}{4}kC\log(n/k)$ is always no larger than the probability that $Z' \geq \frac{1}{4}kC\log(n/k)$. This is because *whatever* $I_j, 1 \leq j \leq (i-1)$ are, the probability of $(1-I_i)$ taking the value '1' is at most $\frac{C}{D}$ conditioned on $I_j, 1 \leq j \leq (i-1)$.

By the well-known Chernoff bound for the sum of independent Bernoulli random variables [DZ98], we know that if $\frac{C}{D} < \frac{1}{4}$,

$$P(Z \geq \frac{1}{4}kC\log(n/k)) \leq e^{-\mathbf{H}(\frac{1}{4}\|\frac{C}{D})kC\log(n/k)}. \tag{2.5.10}$$

Here $\mathbf{H}(a\|b)$ is the Kullback-Leibler divergence between two Bernoulli random variables with parameter a and b, namely,

$$\mathbf{H}(a\|b) = a\log\frac{a}{b} + (1-a)\log\frac{1-a}{1-b}. \tag{2.5.11}$$

In summary, with probability no larger than $e^{-\mathbf{H}(\frac{1}{4}\|\frac{C}{D})kC\log(n/k)}$, a variable node set S of cardinality k is connected to no more than $\frac{3}{4}kC\log(n/k)$ measurement nodes. Since there are at most $\binom{n}{k} \leq e^{k(\log(n/k)+1)}$ variable node sets of cardinality k, by a simple union bound, we have with probability at least

$$P_k = 1 - e^{k(log(n/k)+1)} \times e^{-\mathbf{H}(\frac{1}{4}\|\frac{C}{D})kC\log(n/k)}, \tag{2.5.12}$$

all variable node sets of cardinality of k are connected to more than $\frac{3}{4}kC\log(n/k)$ check nodes. If we take the constants C and D such that $\frac{C}{D}$ is sufficiently small, $e^{k(\log(n/k)+1)} \times e^{-\mathbf{H}(\frac{1}{4}\|\frac{C}{D})kC\log(n/k)}$ will go to zero exponentially in $kC\log n/k$. In fact, if $\frac{C}{D}$ is arbitrarily small, $\mathbf{H}(\frac{1}{4}\|\frac{C}{D})$ can go to infinity.

Similarly, for each integer $1 \leq l \leq k$, the probability that all variable sets of cardinality of l are connected to more than $\frac{3}{4}lC\log(n/k)$ check nodes is at least

$$P_l = 1 - e^{l(\log(n/l)+1)} \times e^{-\mathbf{H}(\frac{1}{4}\|\frac{C}{D})lC\log(n/k)}. \tag{2.5.13}$$

By union bound, the probability that any set of cardinality no larger than k has good expansion property satisfies

$$P \leq 1 - \sum_{l=1}^{k} e^{l(\log(n/l)+1)} \times e^{-\mathbf{H}(\frac{1}{4}\|\frac{C}{D})lC\log(n/k)}, \tag{2.5.14}$$

which is positive given that n is large enough and if we choose $\frac{C}{D}$ sufficiently small. This shows that we only need $O(k\log(n/k))$ check nodes to make a bipartite graph a $(k, \frac{3}{4}c)$ expander. ∎

2.6 RIP-1 Property and Full Recovery Property

2.6.1 Norm One Restricted Isometry Property

The standard Restricted Isometry Property [CT05] is an important sufficient condition that enables compressed sensing using random projections. Intuitively, it says that the measurement almost preserves the Euclidean distance between any two sufficiently sparse vectors. This property implies that recovery using ℓ_1 minimization is possible if a random projection is used for measurement. Berinde et al. in [BGI+08] showed that expander graphs satisfy a very similar property called "RIP-1" which states that if the adjacency matrix of an expander graph is used for measurement, then the Manhattan (ℓ_1) distance between two sufficiently sparse signals is preserved by measurement. They used this property to prove that ℓ_1-minimization is still possible in this case. However, we will show in this section how RIP-1 can guarantee that the algorithm described above will has full recovery.

Following [BI08, BGI+08], we show that the RIP-1 property can be derived from

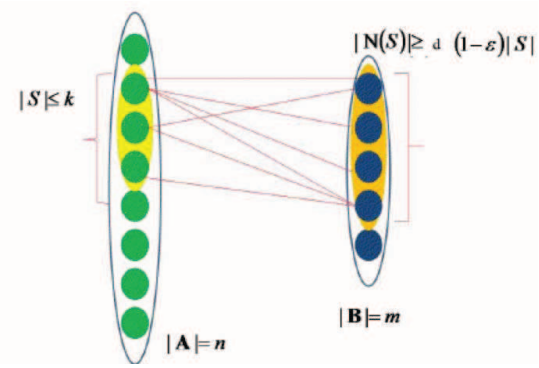


Figure 2.2: (k, ϵ) vertex expander graph

the expansion property and will guarantee the uniqueness of sparse representation.

We begin with the definition of the "unbalanced lossless vertex expander graphs" with expansion coefficient $1-\epsilon$, bearing in mind that we will be interested in $1-\epsilon > \frac{3}{4}$.

Definition 2.6.1 (Unbalanced Lossless Expander Graphs)**.** *A $(l, 1-\epsilon)$-unbalanced bipartite expander graph is a bipartite graph $V = (A, B), |A| = n,\ |B| = m$, where A is the set of variable nodes and B is the set of parity nodes, with regular left degree d such that for any $S \subset A$, if $|S| \leq l$ then the set of neighbors $N(S)$ of S has size $N(S) > (1-\epsilon)d|S|$.*

The following claim follows from the Chernoff bounds [BI08][1].

Claim 2.6.1. *for any $\frac{n}{2} \geq l \geq 1$, $\epsilon > 0$ there exists a $(l, 1-\epsilon)$ expander with left degree:*

$$d = O\left(\frac{\log(\frac{n}{l})}{\epsilon}\right)$$

and right set size:

$$m = O\left(\frac{l\log(\frac{n}{l})}{\epsilon^2}\right).$$

Lemma 2.6.2 (RIP-1 property of the expander graphs)**.** *Let $A_{m\times n}$ be the adjacency matrix of a $(k, 1-\epsilon)$ expander graph E, then for any k-sparse vector $x \in \mathbb{R}^n$ we have:*

$$(1-2\epsilon)d\|x\|_1 \leq \|Ax\|_1 \leq d\ \|x\|_1 \tag{2.6.1}$$

[1]This claim is also used in the expander codes construction.

Proof. The upper bound is trivial using the triangle inequality, so we only prove the lower bound.

The left side inequality is not influenced by changing the position of the coordinates of x, so we can assume that they are in a non-increasing order: $|x_1| \geq |x_2| \geq \cdots \geq |x_n|$. Let E be the set of edges of G and $e_{ij} = (x_i, y_j)$ be the edge that connects x_i to y_j. Define

$$E_2 = \{e_{ij} : \exists\ k < i \text{ such that } e_{kj} \in E\}.$$

Intuitively E_2 is the set of the collision edges. Let

$$T_i = \{e_{i'j} \in E_2 \text{ such that } i' \leq i\},$$

and $a_i = |T_i|$. Clearly $a_1 = 0$; moreover by the expansion property of the graph for any k' less than or equal to k $a_{k'}$ is less than or equal to $\epsilon d k'$. Finally since the graph is k-sparse we know that for each k'' greater than k, $x_{k''}$ is zero. Therefore

$$\begin{aligned} rcl \sum_{e_{ij} \in E_2} |x_i| &= \sum_{i=1}^{n} |x_i|(a_i - a_{i-1}) \\ &= \sum_{i \leq k} a_i(|x_i| - |x_{i+1}|) \\ &\leq \sum_{i \leq k} \epsilon d i(|x_i| - |x_{i+1}|) \\ &\leq \sum_{i \leq k} |x_i| \epsilon d \\ &= \epsilon d \left\|x\right\|_1 . \end{aligned}$$

Now the triangle inequality, and the definition of E_2 imply

$$\begin{aligned} rcl\, \|Ax\|_1 &= \sum_{j=1}^{m} \left| \sum_{e_{ij} \in E} x_i \right| = \sum_{j=1}^{m} \left| \sum_{e_{ij} \in E_2} x_i + \sum_{e_{ij} \notin E_2} x_i \right| \\ &\geq \sum_{j=1}^{m} \left(|\sum_{e_{ij} \notin E_2} x_i| - |\sum_{e_{ij} \in E_2} x_i| \right) \\ &= \sum_{j=1}^{m} \left(|\sum_{e_{ij} \notin E_2} x_i| + |\sum_{e_{ij} \in E_2} x_i| - 2|\sum_{e_{ij} \in E_2} x_i| \right) \\ &= \sum_{e_{ij} \notin E_2} |x_i| + \sum_{e_{ij} \in E_2} |x_i| - 2 \sum_{e_{ij} \in E_2} |x_i| \\ &\geq d\, \|x\|_1 - 2\epsilon d\, \|x\|_1 \\ &= (1 - 2\epsilon)\ d\ \|x\|_1. \end{aligned}$$

■

2.6.2 Full Recovery Property

The full recovery property now follows immediately from Lemma 2.6.2.

Theorem 2.6.3 (Full recovery)**.** *Suppose $A_{m \times n}$ is the adjacency matrix of a $(3k, 1-\epsilon)$ expander graph, and suppose x_1 is a k-sparse and x_2 is a $2k$-sparse vector, such that $Ax_1 = Ax_2$. Then $x_1 = x_2$.*

Proof. Let $z = x_1 - x_2$. Since x_1 is $k - sparse$ and x_2 is $2k$-sparse, z is $3k$-sparse.[2] By Lemma 2.6.2 we have:

$$\|x_1 - x_2\|_1 \leq \frac{1}{(1 - 2\epsilon)\ d} \|Ax_1 - Ax_2\|_1 = 0,$$

hence $x_1 = x_2$. ■

Note that the proof of the above theorem essentially says that the adjacency matrix of a $(3k, 1 - \epsilon)$ expander graph does not have a null vector that is $3k$ sparse.

[2] $\|z\|_0 \leq \|x_1\|_0 + \|x_2\|_0 = 3k$.

We will also give a direct proof of this result (which does not appeal to RIP-1) since it gives a flavor of the arguments to come.

Lemma 2.6.4 (Null space of A)**.** *Suppose $A_{m\times n}$ is the adjacency matrix of a $(3k, 1-\epsilon)$ expander graph with $\epsilon \leq \frac{1}{2}$. Then any nonzero vector in the null space of A, i.e., any $z \neq 0$ such that $Az = 0$, has more than $3k$ nonzero entries.*

Proof. Define $\mathcal{S}$ to be the support set of z. Suppose that z has at most $3k$ nonzero entries, i.e., that $|\mathcal{S}| \leq 3k$. Then from the expansion property we have that $|N(\mathcal{S})| > (1-\epsilon)d|\mathcal{S}|$. Partitioning the set $N(\mathcal{S})$ into the two disjoint sets $N_1(\mathcal{S})$, consisting of those nodes in $N(\mathcal{S})$ that are connected to a single node in $\mathcal{S}$, and $N_{>1}(\mathcal{S})$, consisting of those nodes in $N(\mathcal{S})$ that are connected to more than a single node in $\mathcal{S}$, we may write $N_1(\mathcal{S}) + N_{>1}(\mathcal{S}) > (1-\epsilon)d|\mathcal{S}|$. Furthermore, counting the edges connecting $\mathcal{S}$ and $N(\mathcal{S})$, we have $|N_1(\mathcal{S})| + 2|N_{>1}(\mathcal{S})| \leq d|\mathcal{S}|$. Combining these latter two inequalities yields $|N_1(\mathcal{S})| > (1-2\epsilon)d|\mathcal{S}| \geq 0$. This implies that there is at least one nonzero element in z that participates in only one equation of $y = Az$. However, this contradicts the fact that $Az = 0$ and so z must have more than $3k$ nonzero entries. ■

2.7 Recovering Signals with Optimized Expanders

In Section 2.5, we showed the sufficiency of $k\log(n/k)$ measurements in recovering a k-sparse signal, but it seems that we would need $k\log(n/k)$ iterations in our algorithm. In this section, we generalized our result to more general expander graph with expansion factor as $1-\epsilon$, where $\epsilon > \frac{1}{4}$ and showed that we actually only need $O(k)$ iterations in recovering the k-sparse signal.

2.7.1 $O(k\log(\frac{n}{k}))$ Sensing with $O\left(n\log\left(\frac{n}{k}\right)\right)$ Complexity

Before proving the result, we introduce some notations used in the recovery algorithm and in the proof.

Definition 2.7.1 (gap)**.** *Recall the definition of the gap. At each iteration t, let G_t be the support*[3] *of the gaps vector at that iteration:*

$$G_t = support\ (\vec{g_t}) = \{i | y_i \neq \sum_{j=1}^{n} A_{ij} x_j\}.$$

Definition 2.7.2. *At each iteration t, we define S_t an indicator of the difference between the estimate $\hat{x}$ and x:*

$$S_t = support\ (\hat{x} - x) = \{j : \hat{x_j} \neq x_j\}.$$

Now we are ready to state the main result:

Theorem 2.7.3 (Expander Recovery Algorithm)**.** *Let $A_{m\times n}$ be the adjacency matrix of a $(2k, 1-\epsilon)$ expander graph, where $\epsilon \leq 1/4$, and $m = O\left(k \log\left(\frac{n}{k}\right)\right)$. Then, for any k-sparse signal $\hat{x}$, given $y = A\hat{x}$, the expander recovery algorithm (Algorithm 4 below) recovers $\hat{x}$ successfully in at most $2k$ iterations.*

Algorithm 4 Expander Recovery Algorithm

1: Initialize $x = 0_{n\times 1}$.
2: **if** $y = Ax$ **then**
3: output x and exit.
4: **else**
5: find a variable node say x_j such that at least $(1-2\epsilon)$ d of the measurements it participate in, have identical gap g.
6: set $x_j \leftarrow x_j + g$, and go to 2.
7: **end if**

The proof is virtually identical to that of [XH07a], except that we consider a general $(1-\epsilon)$ expander, rather than a $\frac{3}{4}$-expander, and it consists of the following lemmas.

- The algorithm never gets stuck, and one can always find a coordinate j such that x_j is connected to at least $(1-2\epsilon)d$ parity nodes with identical gaps.

[3] set of nonzero elements

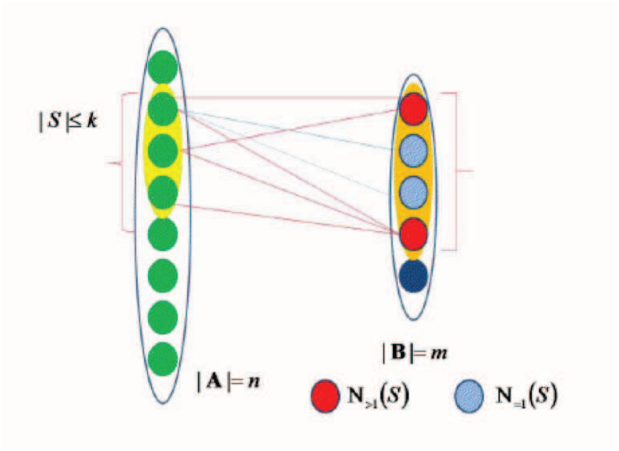


Figure 2.3: Progress lemma

- With certainty the algorithm will stop after at most $2k$ rounds. Furthermore, by choosing ϵ small enough the number of iterations can be made arbitrarily close to k.

Lemma 2.7.4 (progress)**.** *Suppose at each iteration t, $S_t = \{j : \hat{x_j} \neq x_j\}$. If $|S_t| < 2k$ then always there exists a variable node x_j such that at least $(1-2\epsilon)d$ of its neighbor check nodes have the same gap g.*

Proof. We will prove that there exists a coordinate j, such that x_j is uniquely connected to at least $(1-2\epsilon)d$ check nodes, in other words no other non-zero variable node is connected to these nodes. This immediately implies the lemma.
Since $|S_t| < 2k$ by the expansion property of the graph it follows that $N(S_t) \geq (1-\epsilon)d|S_t|$. Now we are going to count the neighbors of S_t in two ways. Figure 2.3 shows the notations in the progress lemma.
We partition the set $N(S_t)$ into two disjoint sets:

- $N_1(S_t)$: The vertices in $N(S_t)$ that are connected only to one vertex in S_t.
- $N_{>1}(S_t)$: The other vertices (that are connected to more than one vertex in S_t).

By double counting the number of edges between variable nodes and check nodes we have:

$$|N_1(S_t)| + |N_{>1}(S_t)| = |N(S_t)| > (1-\epsilon)d|S_t|$$

$$|N_1(S_t)| + 2|N_{>1}(S_t)| \leq \#\text{edges between } S_t, N(S_t) = d|S_t|$$

This gives

$$|N_{>1}(S_t)| < \epsilon d|S_t|,$$

hence

$$|N_1(S_t)| > (1 - 2\epsilon)d|S_t|, \tag{2.7.1}$$

so by the pigeonhole principle, at least one of the variable nodes in S_t must be connected uniquely to at least $(1 - 2\epsilon)d$ check nodes. ∎

Lemma 2.7.5 (gap elimination)**.** *At each step t if $|S_t| < 2k$ then $|G_{t+1}| < |G_t| - (1 - 4\epsilon)d$*

Proof. By the previous lemma, if $|S_t| < 2k$, there always exists a node x_j that is connected to at least $(1 - 2\epsilon)d$ nodes with identical nonzero gap, and hence to at most $2\epsilon d$ nodes possibly with zero gaps. Setting the value of this variable node to zero, sets the gaps on these uniquely connected neighbors of x_j to zero, but it may make some zero gaps on the remaining $2\epsilon d$ neighbors non-zero. So at least $(1 - 2\epsilon)d$ coordinates of G_t will become zero, and at most $2\epsilon d$ its zero coordinates may become non-zero. Hence

$$|G_{t+1}| < |G_t| - (1 - 2\epsilon)d + 2\epsilon d = |G_t| - (1 - 4\epsilon)d. \tag{2.7.2}$$

Figure 2.4 shows the gap elimination. ∎

Remark: The key to accelerating the algorithm is the above Lemma. For a $\frac{3}{4}$ expander, $\epsilon = \frac{1}{4}$ and so $|G_{t+1}| < |G_t|$, which only guarantees that $|G_{t+1}|$ is reduced by a constant number. However, when $\epsilon < \frac{1}{4}$, we have $|G_{t+1}| < |G_t| - (1 - 4\epsilon)d$, which means that $|G_{t+1}|$ is guaranteed to decrease proportionally to d. Since $d = \Omega(\log n)$, we save a factor of $\log n$.

The following lemma provides a direct connection between the size of G_t and the size of S_t.

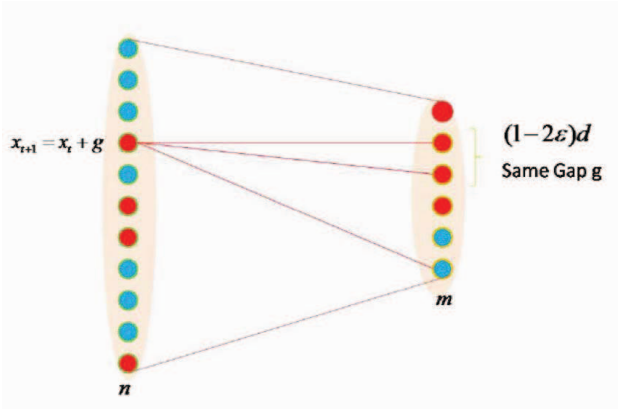


Figure 2.4: Gap elimination lemma

Lemma 2.7.6 (connection)**.** *If at iteration t, $|S_t| < 2k$, then $(1-2\epsilon)d|S_t| \leq |G_t|$.*

Proof. By Equation (2.7.1), $|N_1(S_t)| > (1-2\epsilon)d|S_t|$, also each node in $N_1(S_t)$ has non-zero gap and so is a member of G_t. ■

Lemma 2.7.7 (preservation)**.** *At each step t if $|S_t| < 2k$, after running the algorithm we have $|S_{t+1}| < 2k$.*

Proof. Since at each step we are only changing one coordinate of x, we have $|S_{t+1}| = |S_t| + 1$, so we only need to prove that $S_{t+1} \neq 2k$.

Suppose for a contradiction that $|S_{t+1}| = 2k$, and partition $N(S_{t+1})$ into two disjoint sets:

1. $N_1(S_{t+1})$: The vertices in $N(S_{t+1})$ that are connected only to one vertex in S_{t+1}.

2. $N_{>1}(S_{t+1})$: The other vertices (that are connected to more than one vertex in S_{t+1}).

The argument is similar to that given above; by double counting the number of vertices in $N_1(S_{t+1}), N_{>1}(S_{t+1})$ one can show that

$$|N_1(S_{t+1})| \geq (1-2\epsilon)\ d\ 2k.$$

Now we have the following facts:

- $|N_1(S_{t+1})| > |G_{t+1}|$: Coordinates in $N_1(S_{t+1})$ are connected uniquely to coordinates in S_{t+1}, hence each coordinate in $N_1(S_{t+1})$ has non-zero gap.
- $|G_{t+1}| > |G_1|$: gap elimination from Lemma 2.7.5.
- $|G_1| \leq kd$: x, $\hat{x}$ differ in at most k coordinates, so $Ax, A\hat{x}$ can differ in at most kd coordinates.

As a result we have

$$(1-2\epsilon)2\ dk \leq |N_1(S_{t+1})| \leq |G_{t+1}| \leq |G_1| \leq kd. \tag{2.7.3}$$

This implies $\epsilon \geq \frac{1}{4}$ which contradicts the assumption $\epsilon < \frac{1}{4}$. ∎

Proof of the Theorem 2.7.3. Preservation (Lemma 2.7.7) and progress (Lemma 2.7.4) together immediately imply that the algorithm will never get stuck. Also by Lemma 2.7.5 we had shown that $|G_1| \leq kd$ and $|G_{t+1}| < |G_t| - (1-4\epsilon)d$. Hence after at most $T = \frac{k}{1-4\epsilon}$ steps we will have $|G_T| = 0$ and this together with the connection lemma implies that $|S_T| = 0$, which is the exact recovery of the original signal.

Note that we have to choose $\epsilon < \frac{1}{4}$, and as an example, by setting $\epsilon = \frac{1}{8}$ the recovery needs at most $2k$ iterations. ∎

Remark: The condition $\epsilon < \frac{1}{4}$ in the theorem is necessary. Even $\epsilon = \frac{1}{4}$ leads to a $\frac{3}{4}$ expander graph, which needs $O(k \log n)$ iterations.

2.7.2 Explicit Constructions of Optimized Expander Graphs

In the definition of the expander graphs (Definition 2.6.1), we noted that probabilistic methods prove that such expander graphs exist and furthermore, that any random graph, with high probability, is an expander graph. Hence, in practice it may be sufficient to use random graphs instead of expander graphs.

Though, there is no efficient explicit construction for the expander graphs of Definition 2.6.1, there exists explicit construction for a class of expander graphs which are

very close to the optimum expanders of Definition 2.6.1. Recently Guruswami et al. [GUV07], based on the Parvaresh-Vardy codes [PV05], proved the following theorem:

Theorem 2.7.8 (Explicit Construction of expander graphs)**.** *For any constant $\alpha > 0$, and any $n, k, \epsilon > 0$, there exists a $(k, 1-\epsilon)$ expander graph with left degree*

$$d = O\left(\left(\frac{\log n}{\epsilon}\right)^{1+\frac{1}{\alpha}}\right)$$

and number of right side vertices

$$m = O(d^2 k^{1+\alpha})$$

, which has an efficient deterministic explicit construction.

Since our previous analysis was only based on the expansion property, which does not change in this case, a similar result holds if we use these expanders.

2.7.3 Efficient Implementations and Comparisons

We now compare our approach with recent analysis by Berinde et al [BGI+08]. This paper integrates Indyk's previous work which was based on randomness extractors [Ind08] and a combinatorial algorithm (employing an alternative approach to the RIP-1 results of Berinde-Indyk [BI08]) based on geometric convex optimization methods and suggests a recursive recovery algorithm which takes $m' = O(m \log m)$ sketch measurements and needs a recovery time $O(m \log^2 n)$. The recovery algorithm exploit the hashing properties of the expander graphs, and is sublinear. However, it has difficulties for practical implementation.

By comparison, our recovery algorithm is a simple iterative algorithm, that needs $O(k \log n)$ sketch measurements, and our decoding algorithm consists of at most $2k$ very simple iterations. Each iteration can be implemented very efficiently (see [XH07a]) since the adjacency matrix of the expander graph is sparse with all entries 0 or 1. Even the very naive implementation of the algorithm as suggested in this

chapter works efficiently in practice. The reason is that the unique neighborhood property of the expander graphs is much stronger than what we needed to prove the accuracy of our algorithm. Indeed, it can be shown [HLW06, IR08] that most of the variable nodes have $(1 - \epsilon/2)d$ unique neighbors, and hence at each of the $O(k)$ iterations, the algorithm can find one desired node efficiently. The efficiency of the algorithm can also be improved by using a priority queue data structure. The idea is to use preprocessing as follows: For each variable node v_i compute the median of its neighbors $m_i = Med\left(N(v_i)\right)$ and also compute n_i the number of neighbors with the same value m_i (Note that if a node has $(1 - 2\epsilon)d$ unique neighbors, their median should also be among them.) Then construct the priority queue based on the values n_i, and at each iteration extract the root node from the queue, perform the gap elimination on it, and then, if required, make the correction on corresponding dD variable nodes. The main computational cost of this variation of the algorithm will be the cost of building the priority queue which is $O\left(n \log\left(\frac{n}{k}\right)\right)$; finding the median of d elements can be done in $O\left(\log \frac{n}{k}\right)$ and building a priority queue requires linear computational time.

In this section we show how the analysis using optimized expander graphs that we proposed in the previous section can be used to illustrate that the robust recovery algorithm in [XH07a] can be done more efficiently in terms of the sketch size and recovery time for a family of almost k-sparse signals. With this analysis we will show that the algorithm will only need $O(k \log O\left(n \log\left(\frac{n}{k}\right)\right))$ measurements. Explicit constructions for the sketch matrix exist and the recovery consists of two simple steps. First, the combinatorial iterative algorithm in [XH07a], which is now empowered with the optimized expander sketches, can be used to find the position and the sign of the k largest elements of the signal x. Using an analysis similar to the analysis in section 2.7 we will show that the algorithm needs only $O(k)$ iterations, and similar to the previous section, each iteration can be done efficiently using a priority queue. Then restricting to the position of the k largest elements, we will use a robust theorem in expander graphs to show that simple optimization methods that are now restricted

on k dimensional vectors can be used to recover a k sparse signal that approximates the original signal with very high precision.

Before presenting the algorithm we will define precisely what we mean for a signal to be almost k sparse.

Definition 2.7.9 (almost k-sparse signal)**.** *A signal $x \in R^n$ is said to be almost k-sparse iff it has at most k large elements and the remaining elements are very close to zero and have very small magnitude. In other words, the entries of the near-zero level in the signal take values from the set $[-\lambda, \lambda]$ while the significant level of entries take values from the set $S = \{x : |L - \Delta| \leq |x| \leq |L + \Delta\}$. By the definition of the almost sparsity we have $|S| \leq k$. The general assumption for almost sparsity is intuitively the fact that the total magnitude of the almost sparse terms should be small enough that so that it does not disturb the overall structure of the signal which may make the recovery impossible or very erroneous. Since $\sum_{x \notin S} |x| \leq n\lambda$ and the total contribution of the 'near-zero' elements is small we can assume that $n\lambda$ is small enough. We will use this assumption throughout this section.*

In order to make the analysis for almost k-sparse signals simpler we will use a optimized expander graph which is right-regular as well[4]. The following lemma which appears as Lemma 2.3 in [GLR08] gives us a way to construct right-regular expanders from any expander graph without disturbing its characteristics.

Lemma 2.7.10 (right-regular expanders)**.** *From any left-regular $(k, 1-\epsilon)$ unbalanced expander graph G with left size n, right size m, and left degree d it is possible to efficiently construct a left-right-regular $(k, 1-\epsilon)$ unbalanced expander graph H with left size n, right size $m' \leq 2m$, left side degree $d' \leq 2d$, and right side degree $D = [\frac{nd}{m}]$*

Corollary 2.7.11. *There exists a $(k, 1-\epsilon)$ left-right unbalanced expander graph with left side size n, right side size $m = O(k \log \frac{n}{k})$, left side degree $d = O(\log \frac{n}{k})$, right side degree $D = O(\frac{n \log \frac{n}{k}}{k \log \frac{n}{k}}) = O(\frac{n}{k})$. Also based on the explicit constructions of expander graphs, explicit construction for right-regular expander graphs exists.*

[4]The right-regularity assumption is just for the simplicity of the analysis and as we will discuss it is not mandatory.

We will use the above right-regular optimized expander graphs in order to perform robust signal recovery efficiently. The following algorithm generalizes the k-sparse recovery algorithm and can be used to find the position and sign of the k largest elements of an almost k-sparse signal x from $y = Ax$. At each iteration t in the algorithm, let $\rho_t = 2t\Delta + (D - t - 1)\lambda$ and $\phi_t = 2t\Delta + (D - t)\lambda$. where $D = O(n)$ is the right side degree of the expander graph. Throughout the algorithm we will assume that $L > 2k\Delta + D\lambda$. Hence the algorithm is appropriate for a family of almost k-sparse signals for which the magnitude of the significant elements is large enough. We will assume that k is a small constant; when k is large with respect to n, $(k = \theta(n))$, the $(\alpha n, \frac{3}{4})$ constant degree expander sketch proposed in [XH07a] works well.

Algorithm 5 Expander Recovery Algorithm for Almost k-sparse Signals

1: Initialize $x = 0_{n\times 1}$.
2: **if** $|y - Ax|_\infty \leq \phi_t$ **then**
3: determine the positions and signs of the significant components in $\hat{x}$ as the positions and signs of the non-zero signal components in x; go to 8.
4: **else**
5: find a variable node say x_j such that at least $(1 - 2\epsilon)d$ of the measurements it participate in are in either of the following categories:

(a) They have gaps which are of the same sign and have absolute values between $L - \Delta - \lambda - \rho_t$ and $L + \Delta + \lambda + \rho_t$. Moreover, there exists a number $G \in \{0, L + \Delta, L - \Delta\}$ such that $|y - Ax|$ are all $\leq \phi_t$ over these $(1 - 2\epsilon)$ d measurements if we change x_j to G.

(b) They have gaps which are of the same sign and have absolute values between $2L - 2\Delta - \rho_t$ and $2L + 2\Delta + \rho_t$. Moreover, there exists a number $G \in \{0, L + \Delta, L - \Delta\}$ such that $|y - Ax|$ are all $\leq \phi_t$ over these $(1 - 2\epsilon)$ d measurements if we change x_j to G.

6: set $x_j \leftarrow G$, and go to 2 for next iteration.
7: **end if**
8: pick the set of k significant elements of the candidate signal x_T. Let A' be the sensing matrix A restricted to these entries, output $A'^{\dagger}y$.

In order to prove the algorithm we need the following definitions which are the generalization of the similar definitions in the exactly k-sparse case.

Definition 2.7.12. *At each iteration t, we define S_t an indicator of the difference between $\hat{x}$ and the estimate x:*

$$S_t = \{j | \hat{x_j} \text{ and } x_j \text{in different levels or large with different signs.}\}.$$

Definition 2.7.13 (gap)**.** *At each iteration t, let G_t be the set of measurement elements in which at least one "significant" elements from x contributes:*

$$G_t = \{i || y_i - \sum_{j=1}^{n} A_{ij} x_j |_\infty > \lambda D\}.$$

Theorem 2.7.14 (Validity of the algorithm 5)**.** *The first part of the algorithm will find the position and sign of the k significant elements of the signal x (or more discussion see [XH07a]).*

Proof. This is very similar to the proof of the validity of the exactly k-sparse recovery algorithm. We will exploit the following facts.

- $\hat{x}$ is almost k-sparse so it has at most k significant elements. Initially $S_0 = k$ and $G_0 \leq kd$.
- Since at each iteration only one element x_j is selected, at each iteration t there are at most t elements x_j such that both x_j and $\hat{x}_j$ are in the significant level with the same sign.
- If $|S_t| < 2k$ then $|S_{t+1}| < 2k$ (Preservation Lemma), and by the neighborhood theorem at each round $(1 - 2\epsilon)|S_t|d \leq |G_t|$.
- If $S_t < 2k$ by the neighborhood theorem there exists a node $x_j \in S_t$ which is the unique node in S_t that is connected to at least $(1-2\epsilon)d$ parity check nodes. This node is in S_t. It differs from its actual value in the significance level or at sign. In the first case the part (a) of the recovery algorithm will detect and fix it and in the second case the part (b) of the algorithm will detect and fix it. For further discussion please refer to [XH07a].

- As a direct result, $|G_{t+1}| \leq |G_t| - (1-4\epsilon)d$. So after $T = \frac{kd}{(1-4\epsilon)d}$ iterations we have $|G_T| = 0$. Consequently $|S_T| = 0$ after at most $2k$ iterations.

This means that after at most $2k$ iterations the set

$$S_T = \{j | \hat{x}_j \text{ and } x_j \quad \text{in different levels or with different signs}\} \tag{2.7.4}$$

will be empty and hence the position of the k largest elements in $\hat{x}_T$ will be the position of the k largest elements in x. ■

Knowing the position of the k largest elements of $\hat{x}$ it is easier to recover a good k-sparse approximation. If k is large, parallel version of the Algorithm 4 may be applicable. If k is small, analytical solutions are achievable. Based on the RIP-1 property of the expander graph we propose a way to recover a good approximation for x efficiently and analytically. We need the following lemma which is a direct result of the RIP-1 property of the expander graphs and is proved in [BI08, BGI$^+$08]

Lemma 2.7.15. *Consider any $u \in R^n$ such that $\|Au\|_1 = b$, and let S be any set of k coordinates of u. Then we have*

$$\|u_S\|_1 \leq \frac{b}{d(1-2\epsilon)} + \frac{2\epsilon}{1-2\epsilon}\|u\|_1,$$

and

$$\frac{1-4\epsilon}{1-2\epsilon}\|u_S\|_1 \leq \frac{b}{d(1-2\epsilon)} + \frac{2\epsilon}{1-2\epsilon}\|u_{\bar{S}}\|_1.$$

Using Lemma 2.7.15 we prove that the following minimization recovers a k-sparse signal very close to the original signal:

Theorem 2.7.16 (Final recovery)**.** *Suppose x is an almost k-sparse signal and $y = Ax$ is given where $y \in R^m$ and $m = O\left(k \log\left(\frac{n}{k}\right)\right)$. Also suppose S is the set of the k largest elements of x. Now let A' be a submatrix of A restricted to S. Hence the following minimization problem can be solved analytically with solution $v = A'^{\dagger}y$ (where $A'^{\dagger}$ is the pseudoinverse of A'), and recovers a k-sparse signal v with close*

distance to the original x in the ℓ_1 metric:

$$min\ \|A'v - y\|_2$$

Proof. Suppose v is the recovered signal. Since v is k-sparse we have $Av = A'v$ and hence:

$$\begin{aligned}
\|Av - Ax\|_1 &= \|Av - y\|_1 \\
&= \|A'v - y\|_1 \\
&\leq \sqrt{m}\|A'v - y\|_2 \\
&\leq \sqrt{m}\|A'x_S - y\|_2 \\
&= \sqrt{m}\|Ax_S - Ax\|_2 \\
&\leq \sqrt{m}\lambda D\sqrt{m} \\
&\leq \sqrt{m^2}\lambda D \\
&= mD\lambda = nd\lambda. \qquad (2.7.5)
\end{aligned}$$

The first two equations are only definitions. The third one is the Cauchy-Schwartz inequality. The fourth one is from the definition of v and the last one is due to the almost k-sparsity of x. Since v is k-sparse and x is almost k-sparse with the same support, we may set $u = x - v$ in Lemma 2.7.15 to obtain

$$\begin{aligned}
\frac{1-4\epsilon}{1-2\epsilon}\|u_S\|_1 &\leq \frac{\|Ax - Av\|_1}{d(1-2\epsilon)} + \frac{2\epsilon}{1-2\epsilon}\|u_{\bar{S}}\|_1 \\
&\leq \frac{n\lambda}{(1-2\epsilon)} + \frac{2\epsilon}{1-2\epsilon}\|u_{\bar{S}}\|_1 \\
&\leq \frac{n\lambda}{(1-2\epsilon)} + \frac{2\epsilon}{1-2\epsilon}n\lambda \\
&= O(n\lambda).
\end{aligned}$$

As a result, since the signal is almost k-sparse, the value of $n\lambda$ is small, and hence

the recovered k-sparse signal is close to the best k-term approximation of the original signal.

Remark: Recall that the right-regularity assumption is just to make the analysis simpler. As we mentioned before, it is not necessary for the first part of the algorithm. For the second part, it is used in the inequality $|Av - Ax| \leq \sqrt{m}|Ax_S - Ax|_2$.

However, denoting the i-th row of A by A_i, we have

$$\|Ax_S - Ax\|_2 = \sqrt{m}\sqrt{\sum_{i=1}^{m}(A_i(x_S - x))^2} \leq \sqrt{m}\sqrt{\sum_{i=1}^{m}(\lambda D_i)^2}$$

where D_i denotes the number of ones in the i-th row of A. (In the right regular case, $D_i = D$, for all i.)

Therefore

$$\|Ax_S - Ax\|_2 \leq \sqrt{m}\lambda \sum_{i=1}^{m} D_i = \sqrt{m}\lambda nd$$

The only difference with the constant D_i case is the extra $\sqrt{m}$ but this does not affect the end result.

2.8 Simulation Results

In this section, we give simulation results of the proposed schemes for different n, m and sparsity levels. Although there exist explicit constructions of the required bipartite expander graphs as given in [CRVW02], we will simulate the proposed schemes using randomly generated bipartite graphs since it is easier to implement for evaluation at the current stage, and a randomly generated bipartite graph is expected to be an expander graph with high probability. Also, we can enhance the compressive sensing performance by making some refinements of the randomly generated bipartite graphs. Compressive sensing using explicitly constructed expander graphs for large n is an important topic of future study [CRVW02].

In the simulation, we set the regular left degree c as 5 and generate the random bipartite graphs using a uniformly random permutation of size $n \times c$. In randomly generating the bipartite graphs, there is a chance of getting a small set of variable nodes that have a few common neighbors. For example, sometimes, one variable node is connected to a measurement nodes via two or even more edges. When this occurs, we simply exchange those edges connected to those common measurement nodes with some other randomly chosen edges. After randomly generating bipartite graphs and doing the refinements (thus we get the measurement matrix A), we uniformly select the support set for the k non-zero elements of the signal vector x. The nonzero entries for x are sampled as i.i.d. Gaussian random variables with zero mean and unit variance. We repeat the experiments for 100 independent trials for each k in the experiment. In comparison, we present here the simulation result for linear programming decoding method with the same measurement matrix and the same sparse signal vectors. Here we use the CVX software [boy] to perform the linear programming decoding.

As shown in Figure 2.5, when $n = 1024$ and $m = 512$ we can recover up to the sparsity level of $k = 70$, 7 percent of the signal vector length using Algorithm 1. Although the performance of the proposed scheme is not comparable with ℓ_1 minimization method in [CT05] using Gaussian measurement matrices, we should notice that the randomly generated bi-partite graphs are not optimized expander graphs. Also, the signal recovery for each instance works instantly, taking much less time than the various linear programming solvers which will usually take more than one second and the average time for solving one problem instance is shown in Figure 2.6. The experiment is done using Matlab 7.4.0 on a Windows Platform with 3.00G Hz Intel Pentium CPU and 2.00 GB memory. Surprisingly, under linear programming decoding method, similar recovery performance is achieved by using the measurement matrices constructed from the random graphs and Gaussian ensemble matrix used in [CT05]. In the numerical experiments, the linear programming decoding using the Gaussian ensemble matrix runs much slower than the linear programming decoding

method for the measurement matrix generated from the random bipartite graphs. Similar numerical results are also observed for various n and m's, as shown in Figure 2.7,2.8, 2.9, and 2.10.

In Figure 2.11, we give the simulation results for the performance of recovering approximately sparse signal using Algorithm 2. The significant parts of the signal vector x will take $+1, -1$ with equal probability and the near-zero elements in x are taken uniformly over the interval $[-\lambda, \lambda]$. We can see that the proposed algorithm works well in these cases when λ is reasonably small.

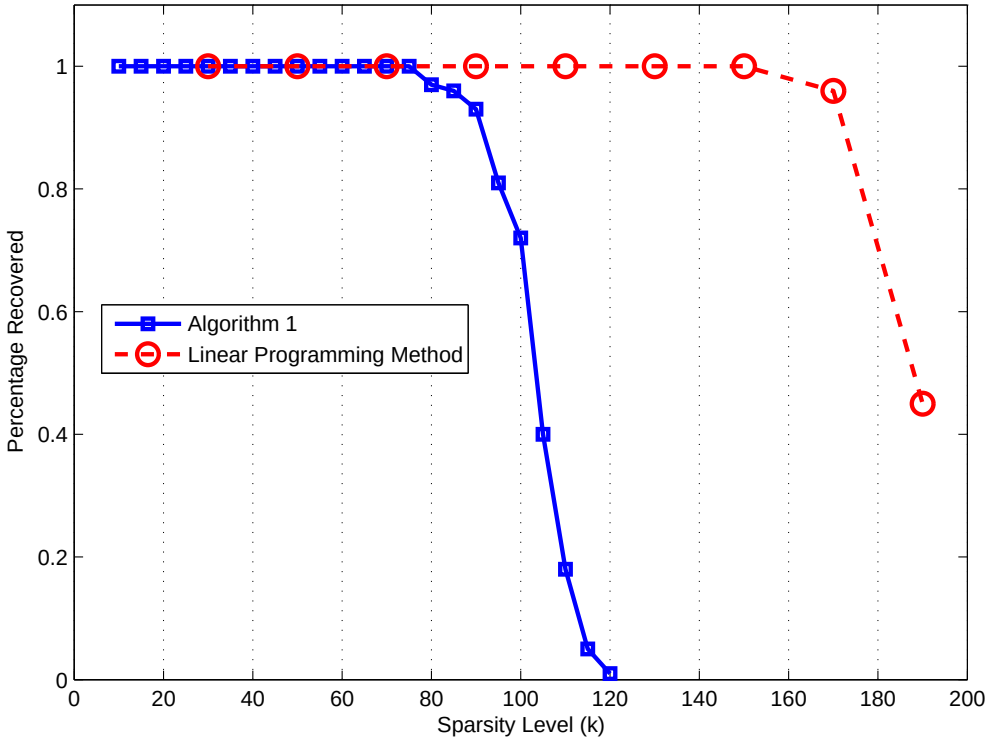


Figure 2.5: The probability of recovering a k-sparse signal with $n = 1024$ and $m = 512$

2.9 Conclusion

We propose to use bipartite expander graphs for compressive sensing of sparse signals and show that we can perform compressive sensing with deterministic performance

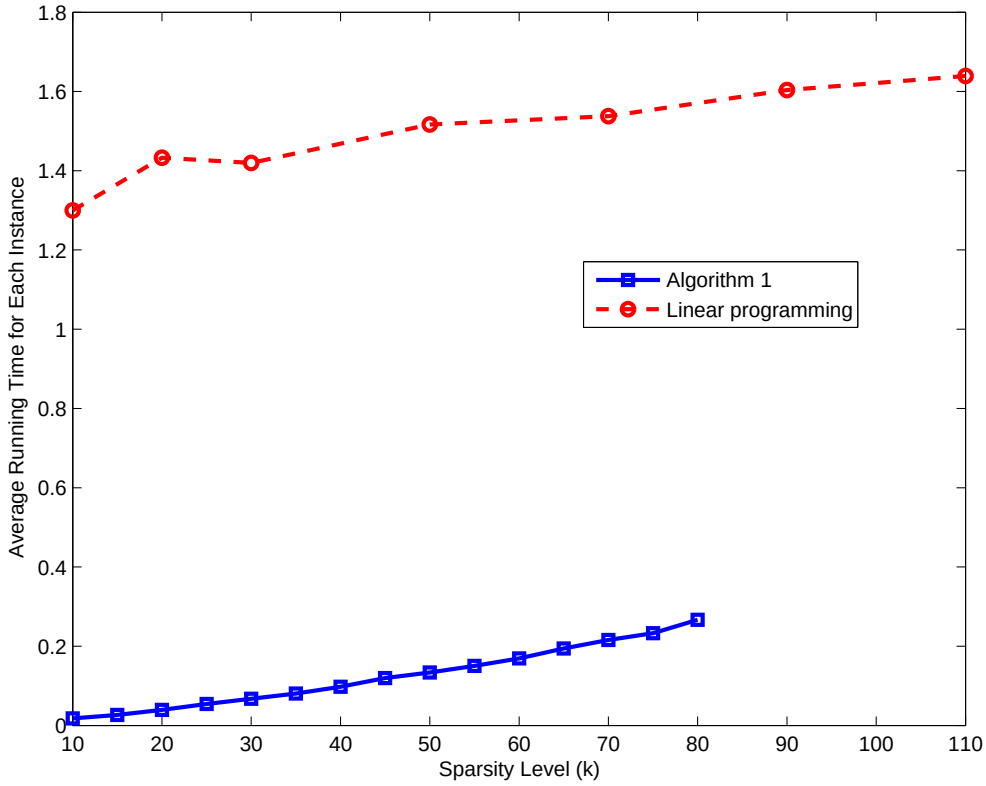


Figure 2.6: The average running time (seconds) of recovering a k-sparse signal with $n = 1024$ and $m = 512$

guarantees at a cost of $O(n)$ signal recovery complexity when the number of non-zero elements k grows linearly with n. At the same time, this expander graph-based scheme offer explicit constructions of the measurement matrices [CRVW02]. When the number of non-zero elements k does not grow linearly with n, we show that we need $O(k \log(n/k))$ measurements, $O(k)$ decoding iterations and total decoding time complexity $n \log(n/k)$. Also we showed how the expansion property of the expander graphs guarantees the full recovery of the original signal. Since random graphs are with high probability expander graphs and it is very easy to generate random graphs, in many cases we might use random graphs instead. When k grows linearly with n, we have an explicit construction of the measurement matrix where the number of measurements optimally scaling with n. When k does not grow linearly with n, just

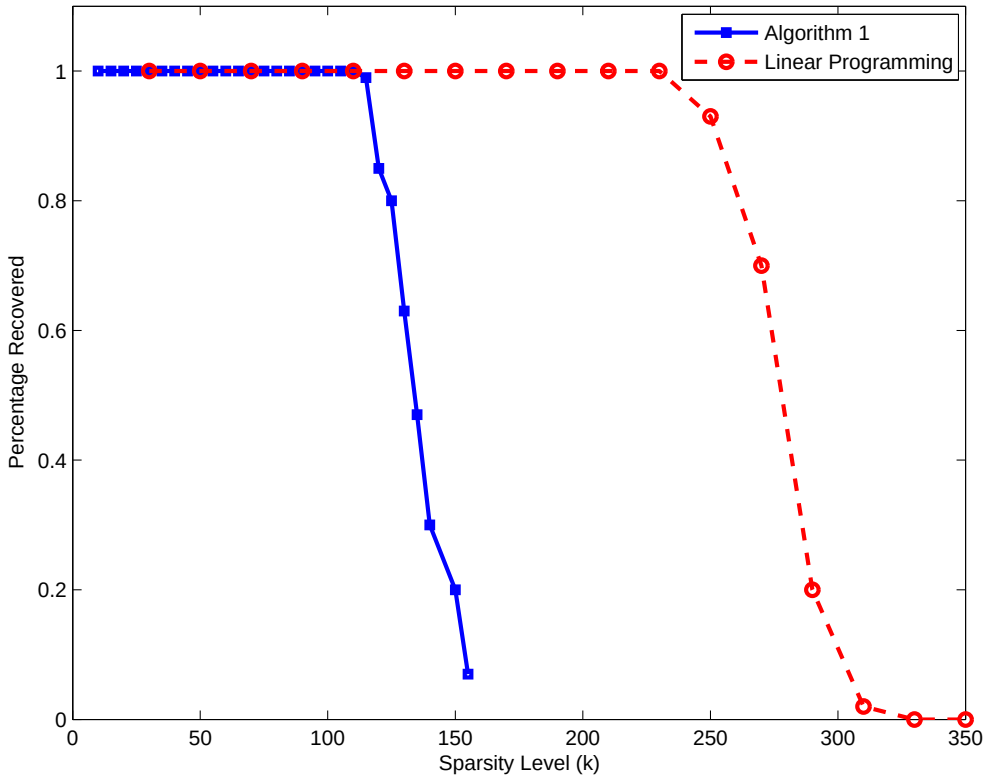


Figure 2.7: The probability of recovering a k-sparse signal with $n = 1024$ and $m = 640$

with a little penalty on the number of measurements and without affecting the number of iterations needed for recovery, one can construct a family of expander graphs for which explicit constructions exist. We also compared our results with a recent result by Berinde et al. [BGI$^+$08], and showed that our algorithm has advantages in terms of the number of required measurements, and the simplicity of the algorithm for practical use. Finally, we showed how the algorithm can be modified to be robust and handle almost k-sparse signals. In order to do this we slightly modified the algorithm by using right-regular optimized expander graphs to find the position of the k largest elements of an almost k-sparse signal. Then exploiting the robustness of the RIP-1 property of the expander graphs we showed how this information can be combined with efficient optimization methods to find a k-sparse approximation for x very efficiently. However, in the almost k-sparsity model that we used non-sparse components should have

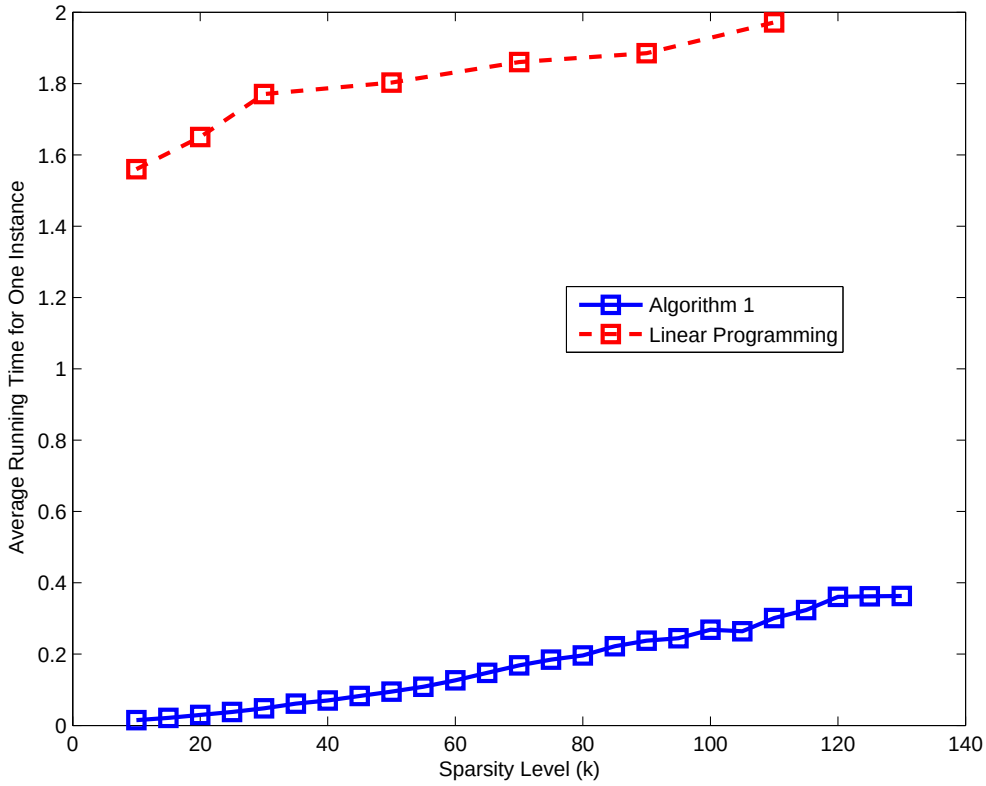


Figure 2.8: The average running time (seconds) of recovering a k-sparse signal with $n = 1024$ and $m = 640$

"almost equal" magnitudes. This is because of the assumption that $L > k\Delta$ which restricts the degree of deviation for significant components. As a result, one important future work will be finding robust algorithms based on more general assumptions, or investigating alternative noise models in which the expander graphs are beneficial. Table 2.1 compares our results with other algorithms. Simulation results verified the effectiveness and efficiency of our methods.

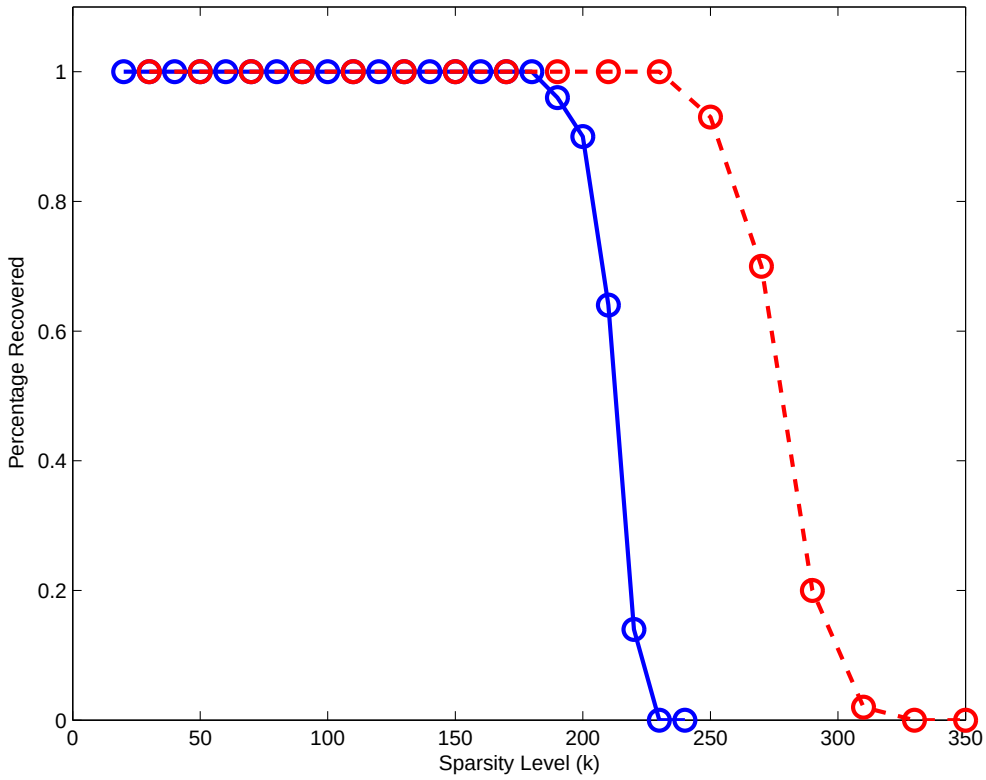


Figure 2.9: The probability of recovering a k-sparse signal with $n = 2048$ and $m = 1024$, with solid line for "Linear Programming" and dashed line for "Algorithm 1"

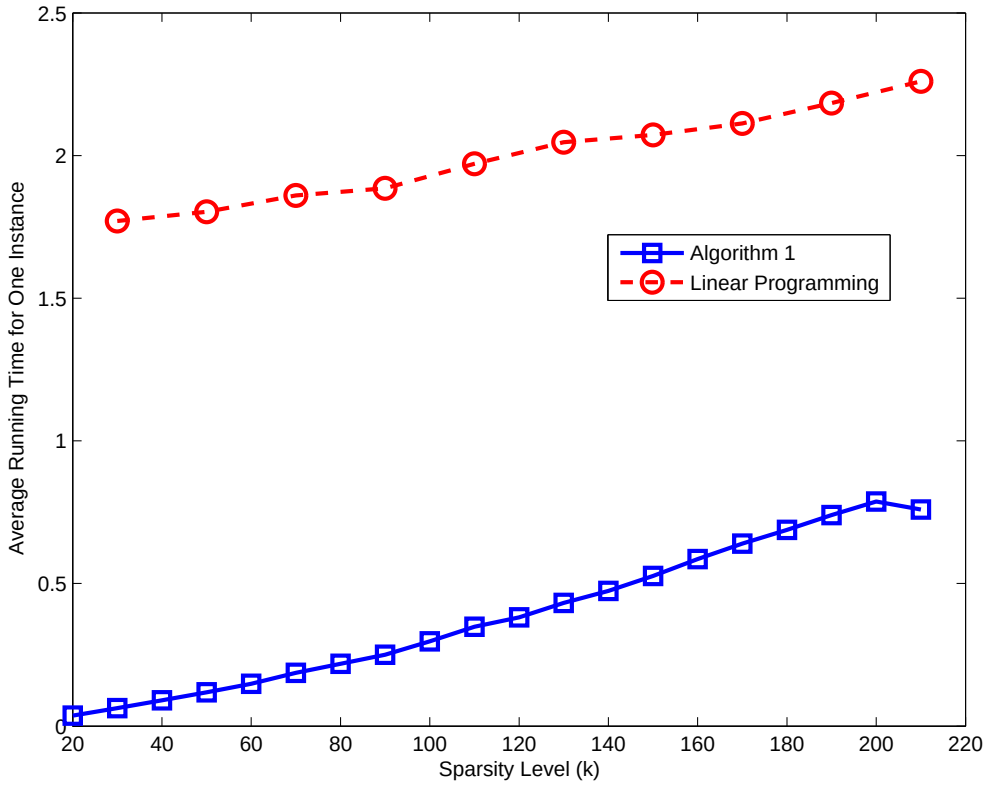


Figure 2.10: The average running time (seconds) of recovering a k-sparse signal with $n = 2048$ and $m = 1024$

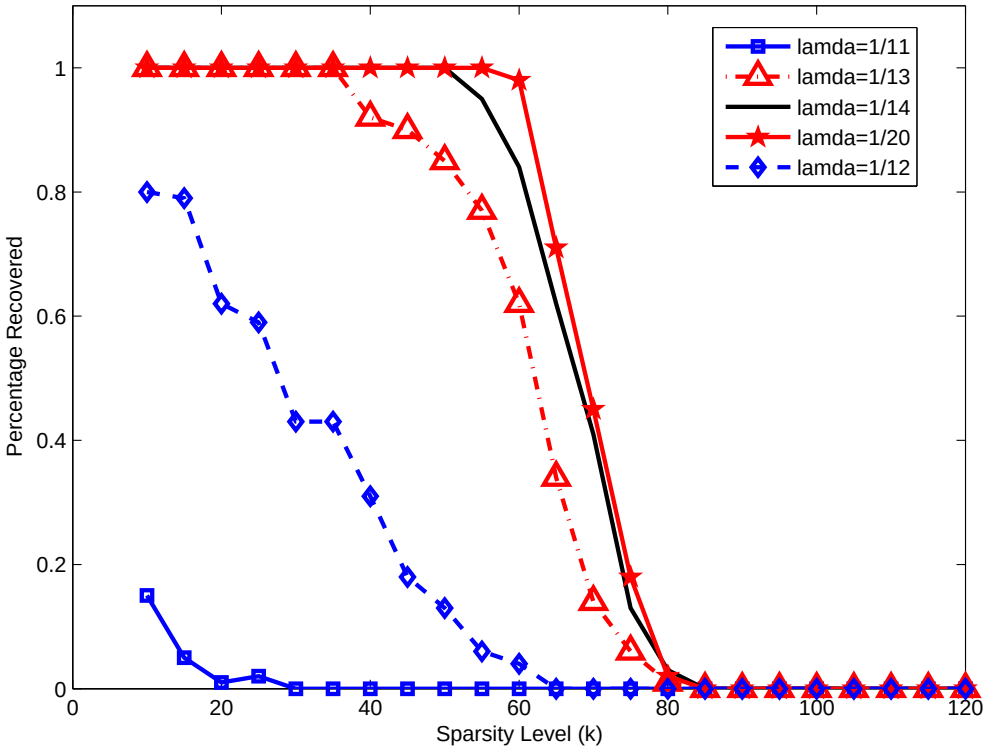


Figure 2.11: The probability of recovering a k-approximately-sparse signal with $n = 1024$ and $m = 512$

Table 2.1: Properties of k-sparse reconstruction algorithms that employ expander matrices with m rows and n columns to reconstruct a vector x from its noisy sketch $Ax + e$

Paper	Approach	Geometric/ /Combinatorial	Number of Measurements m	Number of Iterations	Worst Case Time Complexity	k-term Approximation	Noise Resilience	Explicit Construction
[GLR08]	$\ell 1$ Embedding of Expander Codes	Combinatorial	$O\left(k(\log n)^{\log n}\right)$	-	$O(n^3)$	Compressible	No	[GLR08]
[XH07a]	Unique Neighborhood	Combinatorial	$O(n)$	O(n)	$O(n)$	Almost k-sparse	Sparse Noise	[CRVW02]
[XH07b]	Unique Neighborhood	Combinatorial	$O(k \log n)$	$O(k \log n)$	$O(n \log n)$	Almost k-sparse	Sparse Noise	[GUV07]
[BGI+08]	RIP-1	Geometric	$O\left(k \log\left(\frac{n}{k}\right)\right)$	-	$O(n^3)$	Compressible	Sparse Noise	[GUV07]
[BGI+08]	Hash Functions	Combinatorial	$O(k \log \frac{n}{k} \log n)$	$O(k \log \frac{n}{k} \log n)$	$O(k \log^3 n)$	Compressible	Sparse Noise	[GUV07]
[IR08]	All Unique Neighborhoods	Combinatorial	$O\left(k \log\left(\frac{n}{k}\right)\right)$	$O(n)$	$O\left(n \log\left(\frac{n}{k}\right)\right)$	Compressible	Sparse Noise	[GUV07]
[BIR08]	RIP-1	Geometric	$O\left(k \log\left(\frac{n}{k}\right)\right)$	$R = O\left(\log\left(\frac{d\|x\|_1}{\|e\|_1}\right)\right)$	$O\left(n \log\left(\frac{n}{k}\right) \log R\right)$	Compressible	Sparse Noise	[GUV07]
[JXHC]	Unique Neighborhood	Combinatorial	$O\left(k \log\left(\frac{n}{k}\right)\right)$	$O(k)$	$O\left(n \log\left(\frac{n}{k}\right)\right)$	Almost k-sparse	Sparse Noise	[GUV07]

Chapter 3

Grassmann Angle Analytical Framework for Subspaces Balancedness

It is well known that compressed sensing problems reduce to finding the sparse solutions of a large under-determined system of equations. Although finding the sparse solution in general may be computationally difficult, starting with the seminal work of [CT05], it has been shown that linear programming techniques, obtained from an ℓ_1-norm relaxation of the original non-convex problem, can provably find the unknown vector in certain instances. In particular, using a certain restricted isometry property, [CT05] shows that for measurement matrices chosen from a random Gaussian ensemble, ℓ_1 optimization can find the correct solution with overwhelming probability even when the support size of the unknown vector is proportional to its dimension. The paper [Don06c] uses results on neighborly polytopes from [VS92] to give a "sharp" bound on what this proportionality should be in the Gaussian measurement ensemble.

In this chapter we shall focus on finding sharp bounds on the recovery of "approximately sparse" signals and also under noisy measurements. While the restricted isometry property can be used to study the recovery of approximately sparse signals in the presence of noisy measurements, the obtained bounds on achievable sparsity level can be quite loose. On the other hand, the neighborly polytope technique which yields sharp bounds for ideally sparse signals cannot be generalized to approximately sparse signals. In this chapter, starting from a necessary and sufficient condition, the

"balancedness" property of linear subspaces, for achieving a certain signal recovery accuracy, using high-dimensional geometry, we give a unified *null-space Grassmann angle*-based analytical framework for analyzing ℓ_1 minimization in compressive sensing. This new framework gives sharp quantitative tradeoffs between the signal sparsity and the recovery accuracy of the ℓ_1 optimization for approximately sparse signals.

As a consequence, the neighborly polytope result of [Don06c] for ideally sparse signals can be viewed as a special case of ours. We give the asymptotic analytical results of the sparsity level satisfying the new null-space necessary and sufficient conditions for. In addition to the "strong" notion of robustness, we also discuss the notion of "weak" and "sectional" robustness in sparsity recovery. Our results concern fundamental properties of linear subspaces and so may be of independent mathematical interest.

3.1 Introduction

Compressive sensing is an emerging area in signal processing and information theory which has attracted a lot of attention recently [Can06] [Don06b]. The motivation behind compressive sensing is to do "sampling" and "compression" at the same time. In conventional wisdom, in order to fully recover a signal, one has to sample the signal at a sampling rate equal or greater to the Nyquist sampling rate. However, in many applications such as imaging, sensor networks, astronomy, biological systems [RIC], the signals we are interested in are often "sparse" over a certain basis. This process of "sampling at full rate" and then "throwing away in compression" can prove to be wasteful of sensing and sampling resources, especially in application scenarios where resources like sensors, energy, and observation time are limited. In these cases, compressive sensing promises to use a much smaller number of samplings or measurements while still being able to recover the original sparse signal exactly or accurately. The cornerstone techniques enabling practical compressive sensing are the effective decoding algorithms to recover the sparse signals from the "compressed" measurement results. One of the most important and popular decoding algorithms

for compressive sensing is the Basis Pursuit algorithm, namely the ℓ_1 minimization algorithm.

In this chapter we are interested in the general principles behind the ℓ_1 minimization decoding algorithm for compressed sensing of approximately sparse signals under noisy measurements. Mathematically, in compressive sensing problems, we would like to find a $n \times 1$ vector $\mathbf{x}$, such that

$$A\mathbf{x} = \mathbf{y}, \tag{3.1.1}$$

where A is an $m \times n$ measurement matrix and $\mathbf{y}$ is $m \times 1$ measurement vector. In the usual compressed sensing context $\mathbf{x}$ is $n \times 1$ unknown k-sparse vector. This assumes that $\mathbf{x}$ has only k nonzero components. In this chapter we will consider a more general version of the k-sparse vector $\mathbf{x}$. Namely, we will assume that k components of the vector $\mathbf{x}$ have large magnitudes and that the vector comprised of the remaining $n - k$ components has an ℓ_1-norm less than Δ. We will refer to this type of signal as approximately k-sparse signal, or for brevity only approximately sparse signal. Possibly the $\mathbf{y}$ can be further corrupted with measurement noise. The interested readers can find More on similar type of problems in [CDD08] and other references. This problem setup is more realistic of practical applications than the standard compressed sensing of ideally k-sparse signals (see, e.g., [TWD$^+$06, Can06, CRT06] and the references therein).

In the rest of the chapter we will further assume that the number of the measurements is $m = \delta n$ and the number of the "large" components of $\mathbf{x}$ is $k = \rho\delta n = \zeta n$, where $0 < \rho < 1$ and $0 < \delta < 1$ are constants independent of n (clearly, $\delta > \zeta$).

3.1.1 ℓ_1 Minimization for Exactly Sparse Signal

A particular way of solving (3.1.1) which recently generated a large amount of research is called ℓ_1-optimization (basis pursuit) [CT05]. It proposes solving the following

problem:

$$\begin{aligned} \min \quad & \|\mathbf{x}\|_1 \\ \text{subject to} \quad & A\mathbf{x} = \mathbf{y}. \end{aligned} \tag{3.1.2}$$

Quite remarkably in [CT05] the authors were able to show that if the number of the measurements is $m = \delta n$ and if the matrix A satisfies a special property called the restricted isometry property (RIP), then any unknown vector $\mathbf{x}$ with no more than $k = \zeta n$ (where ζ is an absolute constant which is a function of δ, but independent of n, and explicitly bounded in [CT05]) non-zero elements can be recovered by solving (3.1.2). As expected, this assumes that $\mathbf{y}$ was in fact generated by that $\mathbf{x}$ and given to us (more on the case when the available measurements are noisy versions of $\mathbf{y}$ can be found in, e.g., [HN06, Wai06]).

As can be immediately seen, the previous results heavily rely on the assumption that the measurement matrix A satisfies the RIP condition. It turns out that for several specific classes of matrices, such as matrices with independent zero-mean Gaussian entries or independent Bernoulli entries, the RIP holds with overwhelming probability [CT05, BDDW08, RV05]. However, it should be noted that the RIP is only a sufficient condition for ℓ_1-optimization to produce a solution of (3.1.1).

Instead of characterizing the $m \times n$ matrix A through the RIP condition, in [Don06c, DT05a] the authors assume that A constitutes a k-neighborly poly-tope. It turns out (as shown in [Don06c]) that this characterization of the matrix A is in fact a necessary and sufficient condition for (3.1.2) to produce the solution of (3.1.1). Furthermore, using the results of [VS92], it can be shown that if the matrix A has i.i.d. zero-mean Gaussian entries with overwhelming probability it also constitutes a k-neighborly poly-tope. The precise relation between m and k in order for this to happen is characterized in [Don06c] as well. It should also be noted that for a given value m, i.e., for a given value of the constant δ, the value of the constant ζ is significantly better in [Don06c, DT05a] than in [CT05]. Furthermore, the values of constants ζ obtained for different values of δ in [Don06c] approach the ones obtained

by simulation as $n \longrightarrow \infty$.

3.1.2 ℓ_1 Minimization for Approximately Sparse Signal

As mentioned earlier, in this chapter we will be interested in recovering approximately k-sparse signals from compressed observations. Since in this case the unknown vector $\mathbf{x}$ in general has no zeros, an exact recovery from a reduced number of measurements is not possible normally. Instead, we will prove that, if we denote the unknown approximately k-sparse vector is $\mathbf{x}$ and $\hat{\mathbf{x}}$ is the solution of (3.1.2), then for any given constant $0 \leq \delta \leq 1$, there exist a constant $\zeta > 0$ and a sequence of measurement matrices $A \in R^{m \times n}$ as $n \to \infty$ such that

$$||\hat{\mathbf{x}} - \mathbf{x}||_1 \leq \frac{2(C+1)\Delta}{C-1}, \tag{3.1.3}$$

holds for *all* $\mathbf{x} \in R^n$, where $C > 1$ is a given constant (saying how close in ℓ_1 norm the recovered vector $\hat{\mathbf{x}}$ should be to $\mathbf{x}$). Here ζ will be a function of C and δ, but independent of the problem dimension n. In particular, we have the following theorem

Theorem 3.1.1. *Let n, m, k, $\mathbf{x}$, $\hat{\mathbf{x}}$ and Δ be defined as above. Let K denotes a subset of $\{1, 2, \ldots, n\}$ such that $|K| = k$, where $|K|$ is the cardinality of K, and let K_i denote the i-th element of K and $\bar{K} = \{1, 2, \ldots, n\} \setminus K$.*

For any constant $C > 1$ and any $\delta = \frac{m}{n} > 0$, then there exists a $\zeta(\delta, C) > 0$ such that if the measurement matrix A is the basis for a uniformly distributed subspace, then with overwhelming probability as $n \to \infty$, for all vectors $\mathbf{w} \in R^n$ in the null-space of A, and for all K such that $|K| = k \leq \zeta(\delta, C)$, we have

$$C \sum_{i=1}^{k} |\mathbf{w}_{K_i}| \leq \sum_{i=1}^{n-k} |\mathbf{w}_{\bar{K}_i}|, \tag{3.1.4}$$

where $\mathbf{x}_K$ denotes the part of $\mathbf{x}$ over the subset K; and at the same time the solution $\hat{\mathbf{x}}$ produced by (3.1.2) will satisfy

$$||\hat{\mathbf{x}} - \mathbf{x}||_1 \leq \frac{2(C+1)\Delta}{C-1}. \tag{3.1.5}$$

for all $\mathbf{x} \in R^n$.

The main focus of this chapter is to establish a sharp relationship between ζ and C (when δ is fixed). For example, when $\delta = 0.5555$, we have the following figure showing the tradeoff between ζ and C:

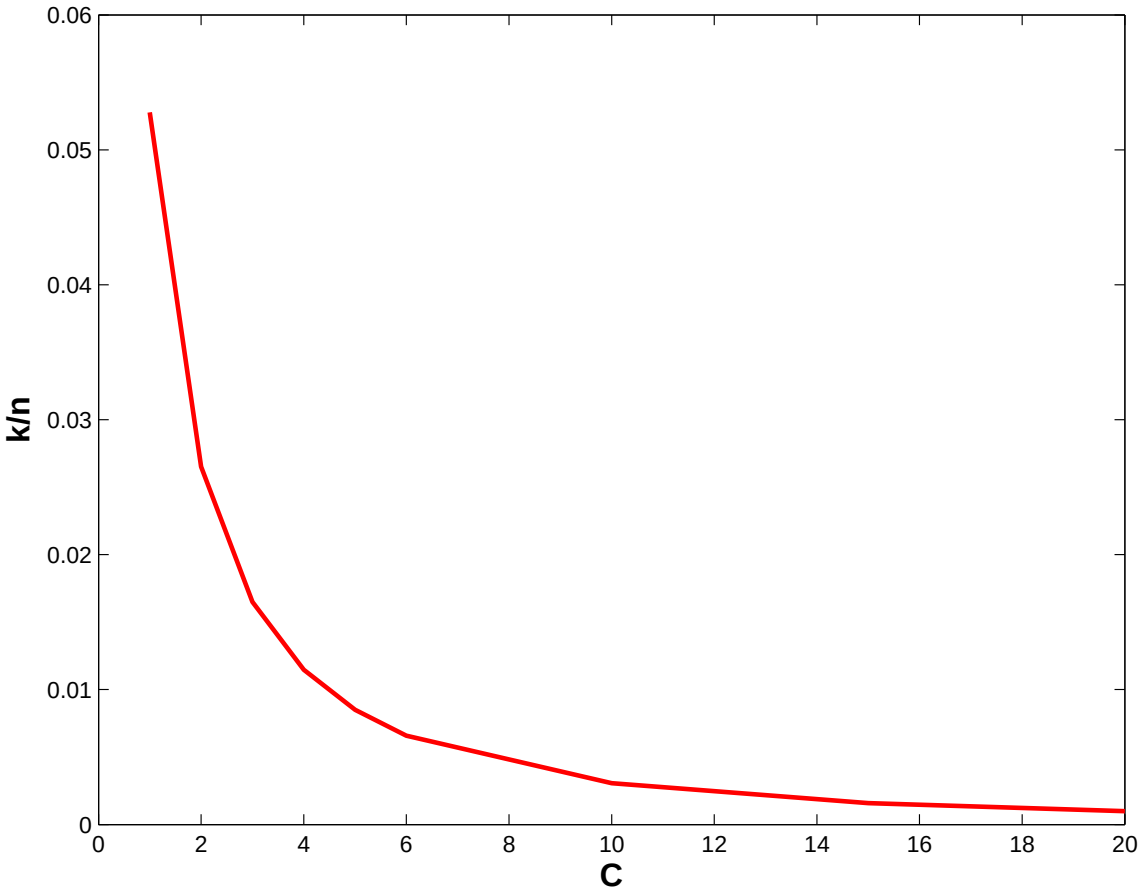


Figure 3.1: Allowable sparsity as a function of C (allowable imperfection of the recovered signal is $\frac{2(C+1)\Delta}{C-1}$)

To obtain the stated results, we will make use of a characterization that constitutes both necessary and sufficient conditions on the matrix A such that the solution of (3.1.2) approximates the original signal accurately enough such that (3.1.3) holds. This characterization will be equivalent to the neighborly polytope characterization

from [Don06c] in the "ideally sparse" case. Furthermore, as we will see later in the chapter, in the perfectly sparse signal case (which allows $C \longrightarrow 1$), our result for allowable ζ will match the result of [Don06c]. Our analysis will be directly based on the null-space Grassmann angle result in high dimensional integral geometry, which gives a unified analytic framework for ℓ_1 minimization.

A similar problem as discussed in this paper was considered with different proof techniques in [CDD08] based on the restricted isometry property from [CT05], where no explicit values of ζ were given. Since the RIP condition is a sufficient condition, it generally gives rather loose bounds on the explicit values of ζ even in the ideally sparse case. In this chapter we will provide sharp bounds on the explicit values of the allowable constants ζ for the general cases $C \geq 1$ based on high-dimensional geometry. Certainly there were also discussions of compressive sensing under different definitions of non-ideally sparse signals in the literature, for example, [Don06b] discussed compressive sensing for signals from a l_p ball with $0 < p \leq 1$ using sufficient conditions based on results of the Gelfand n-widths. However, the results in this chapter are dealing directly with approximately sparse signals defined in terms of the concentration of ℓ_1 norm, and furthermore, we give a neat *necessary and sufficient* condition for ℓ_1 optimization to work and we are also able to explicitly give much sharper compressive sensing performance bounds.

This rest of the chapter is organized as follows. In Section 3.2, we will introduce a null-space characterization of linear subspaces for guaranteeing the signal recovery robustness using the ℓ_1 minimizations. Section 3.3 presents a Grassmann angle based high dimensional geometrical framework for analyzing the null-space characterization. In Section 3.4, 3.6, and 3.7, analytical performance bounds are given for the null-space characterization. Section 3.8 shows how the Grassmann angle analytical framework can be extended to analyzing the "weak", "sectional" and "strong" notations of signal recovery robustness. In Section 3.9, we present the robustness analysis of the ℓ_1 minimization under noisy measurements using the null-space characterization. In Section 3.10, the numerical evaluations of the performance bounds for signal recovery

robustness are given. Section 3.11 concludes the chapter.

3.2 The Null Space characterization

In this section we introduce a useful characterization of the matrix A. The characterization will establish a necessary and sufficient condition on the matrix A so that solution of (3.1.2) approximates the solution of (3.1.1) such that (3.1.3) holds. (See [FN03, LN06, Zha06, CDD08, SXH08a, SXH08b, KT07] for variations of this result).

Theorem 3.2.1. *Assume that an $m \times n$ measurement matrix A is given. Further, assume that $\mathbf{y} = A\mathbf{x}$ and that $\mathbf{w}$ is an $n \times 1$ vector. Let K be any subset of $\{1, 2, \ldots, n\}$ such that $|K| = k$, where $|K|$ is the cardinality of K and let K_i denote the i-th element of K. Further, let $\bar{K} = \{1, 2, \ldots, n\} \setminus K$. Then the solution $\hat{\mathbf{x}}$ produced by (3.1.2) will satisfy*

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq \frac{2(C+1)}{C-1} \|\mathbf{x}_{\bar{K}}\|_1,$$

with $C > 1$, if and only if $\forall\ \mathbf{w} \in \boldsymbol{R}^n$ such that

$$A\mathbf{w} = 0$$

and $\forall\ K$ such that $|K| = k$, we have

$$C \sum_{i=1}^{k} |\mathbf{w}_{K_i}| \leq \sum_{i=1}^{n-k} |\mathbf{w}_{\bar{K}_i}|. \tag{3.2.1}$$

Proof. Sufficiency: Suppose the matrix A has the claimed null-space property. Now the solution $\hat{\mathbf{x}}$ of (3.1.2) satisfies

$$\|\hat{\mathbf{x}}\|_1 \leq \|\mathbf{x}\|_1,$$

where $\mathbf{x}$ is the original signal. Since $A\hat{\mathbf{x}} = \mathbf{y}$, it easily follows that $\mathbf{w} = \hat{\mathbf{x}} - \mathbf{x}$ is in the null space of A. Therefore we can further write $\|\mathbf{x}\|_1 \geq \|\mathbf{x} + \mathbf{w}\|_1$. Using the

triangular inequality for the ℓ_1 norm we obtain

$$\begin{aligned}
\|\mathbf{x}_K\|_1 + \|\mathbf{x}_{\bar{K}}\|_1 &= \|\mathbf{x}\|_1 \\
&\geq \|\hat{\mathbf{x}}\|_1 = \|\mathbf{x} + \mathbf{w}\|_1 \\
&\geq \|\mathbf{x}_K\|_1 - \|\mathbf{w}_K\|_1 + \|\mathbf{w}_{\bar{K}}\|_1 - \|\mathbf{x}_{\bar{K}}\|_1 \\
&\geq \|\mathbf{x}_K\|_1 + \frac{C-1}{C}\|\mathbf{w}_K\|_1 - \|\mathbf{x}_{\bar{K}}\|_1 \\
&\geq \|\mathbf{x}_K\|_1 - \|\mathbf{x}_{\bar{K}}\|_1 + \frac{C-1}{C+1}\|\mathbf{w}\|_1,
\end{aligned}$$

where the last two inequalities are from the claimed null-space property. Relating the first equality and the last inequality above, we finally get

$$2\|\mathbf{x}_{\bar{K}}\|_1 \geq \frac{(C-1)}{C+1}\|\mathbf{w}\|_1,$$

as desired.

Necessity: Since every step in the proof of the sufficiency can be reversed if equality is achieved in the triangular equality, the condition

$$C\sum_{i=1}^{k}|\mathbf{w}_{K_i}| \leq \sum_{i=1}^{n-k}|\mathbf{w}_{\bar{K}_i}|$$

is also a necessary condition for $\|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq \frac{2(C+1)}{C-1}\|\mathbf{x}_{\bar{K}}\|_1$ to hold for every $\mathbf{x}$. ∎

It should be noted that if the condition (3.2.1) is satisfied, then

$$2\|\mathbf{x}_{\bar{K}}\|_1 \geq \frac{(C-1)}{C+1}\|\mathbf{w}\|_1 = \frac{(C-1)}{C+1}\|\hat{\mathbf{x}} - \mathbf{x}\|_1,$$

for any K or $\bar{K}$. Hence it is also true for the set K which corresponds to the k largest components of the vector $\mathbf{x}$. In that case we can write

$$2\Delta \geq \frac{(C-1)}{C+1}\|\hat{\mathbf{x}} - \mathbf{x}\|_1$$

which exactly corresponds to (3.1.3). In fact, the condition (3.2.1) is also a sufficient

and necessary condition for unique exact recovery of ideally k-sparse signals after we take $C = 1$ and let (3.2.1) take strict inequality for all $\mathbf{w} \neq 0$ in the null space of $\mathcal{A}$. To see this, suppose the ideally k-sparse signal $\mathbf{x}$ is supported over the set K, namely, $\|\mathbf{x}_{\bar{K}}\|_1 = 0$. Then from the same triangular inequality derivation of Theorem 1, we know that $\|\hat{\mathbf{x}} - \mathbf{x}\|_1 = 0$, namely $\hat{\mathbf{x}} = \mathbf{x}$. Or we can just let C be arbitrarily close to 1 and since

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq \frac{2(C+1)}{C-1}\|\mathbf{x}_{\bar{K}}\|_1 = 0,$$

we also get $\hat{\mathbf{x}} = \mathbf{x}$. In this sense, when $C = 1$, the null-space condition is equivalent to the neighborly polytope condition [Don06c] for unique exact recovery of ideally sparse signals.

Remark: Clearly, we need not check (3.2.1) for all subsets K; checking the subset with the k largest (in absolute value) elements of $\mathbf{w}$ is sufficient. However, Theorem 1 will be more convenient for our subsequent analysis.

In the following section, for a given value $\delta = \frac{m}{n}$ and any value $C \geq 1$, we will determine the value of feasible $\zeta = \rho\delta = \frac{k}{n}$ for which there exists a sequence of A such that (3.2.1) is satisfied when n goes to infinity and $\frac{m}{n} = \delta$. It turns out that for a specific A, it is very hard to check whether the condition (3.2.1) is satisfied or not. Instead, we consider randomly choosing A from a certain distribution, and analyze for what ζ, the condition (3.2.1) for its null-space is satisfied with overwhelming probability as n goes to infinity. When we consider $C = 1$ corresponding to the ℓ_1 minimization for purely k-sparse signals, coarse bounds on k were established in [Zha06][SXH08a] using different analysis techniques for high-dimensional linear subspaces. However, no sharp bounds are available for under the general case $C \geq 1$.

The standard results on compressed sensing assume that the matrix A has i.i.d. $\mathcal{N}(0, 1)$ entries. In this case, the following lemma gives a characterization of the resulting null-space of A.

Lemma 3.2.2. *Let $A \in R^{m\times n}$ be a random matrix with i.i.d. $\mathcal{N}(0, 1)$ entries. Then the following statements hold:*

- *The distribution of A is left-rotationally invariant, $P_A(A) = P_A(A\Theta)$, $\Theta\Theta^* = \Theta^*\Theta = I$.*

- *The distribution of Z, any basis of the null-space of A is right-rotationally invariant. $P_Z(Z) = P_Z(\Theta^* Z)$, $\Theta\Theta^* = \Theta^*\Theta = I$.*

- *It is always possible to choose a basis for the null-space such that $Z \in R^{n\times(n-m)}$ has i.i.d. $\mathcal{N}(0,1)$ entries.*

In view of Theorem 1 and Lemma 1 what matters is that the null-space of A be rotationally invariantly. Sampling from this rotationally invariant distribution is equivalent to uniformly sampling a random $(n-m)$-dimensional subspace from the Grassmann manifold $\mathrm{Gr}_{(n-m)}(n)$. Here the Grassmannian manifold $\mathrm{Gr}_{(n-m)}(n)$ is the set of $(n-m)$-dimensional subspaces in the n-dimensional Euclidean space R^n [Boo86]. For any such A and ideally sparse signals, the sharp bounds of [Don06c], for example, apply. However, we shall see that the neighborly polytope condition for ideally sparse signals does not apply to the proposed null-space condition analysis for approximately sparse signals, since the null-space condition can not be transformed to the k-neighborly property in a high-dimensional polytope [Don06c]. Instead, in this chapter, we shall give a unified Grassmannian angle framework to analyze the proposed null-space property with applications to compressive sensing for approximately sparse signals.

3.3 The Grassmannian Angle Framework for the Null Space Characterization

In this section we derive and detail the Grassmannian angle-based framework for analyzing the bounds on $\zeta = \frac{k}{n}$ such that the condition (3.2.1) holds for the null-space of the measurement matrix A. Before proceeding further, let us make clear the problem that we are trying to solve: Let Z be the null-space of the randomly sampled

measurement matrix A. Given a certain constant $C > 1$ (or $C \geq 1$), which corresponds to a certain level of recovery accuracy for the approximately sparse signals, we are interested in how large the sparsity level k can be while satisfying the following condition on Z ($|K| = k$):

$$\forall K \subseteq \{1, 2, ..., n\}, C\|\mathbf{w}_K\|_1 \leq \|\mathbf{w}_{\bar{K}}\|_1, \forall \mathbf{w} \in Z, \tag{3.3.1}$$

with overwhelming probability as $n \to \infty$. From the definition of the condition (3.3.1), there is a tradeoff between the largest sparsity level k and the parameter C, which in turn is related to the allowable signal recovery imperfection. As C grows, clearly the largest k satisfying (3.3.1) will decrease, and at the same time, ℓ_1 minimization will be more robust in recovering approximately sparse signals. The key in our derivation is the following lemma

Lemma 3.3.1. *For a certain subset $K \subseteq \{1, 2, ..., n\}$ with $|K| = k$, the event that the null-space Z satisfies*

$$C\|\mathbf{w}_K\|_1 \leq \|\mathbf{w}_{\bar{K}}\|_1, \forall \mathbf{w} \in Z$$

is equivalent to the event that $\forall \mathbf{x}$ supported on the k-set K (or supported on a subset of K):

$$\|\mathbf{x}_K + \mathbf{w}_K\|_1 + \|\frac{\mathbf{w}_{\bar{K}}}{C}\|_1 \geq \|\mathbf{x}_K\|_1, \forall \mathbf{w} \in Z. \tag{3.3.2}$$

Proof. First, let us assume that $C\|\mathbf{w}_K\|_1 \leq \|\mathbf{w}_{\bar{K}}\|_1, \forall \mathbf{w} \in Z$. Using the triangular inequality for the ℓ_1 norm we obtain

$$\begin{aligned} & \|\mathbf{x}_K + \mathbf{w}_K\|_1 + \|\frac{\mathbf{w}_{\bar{K}}}{C}\| \\ \geq \; & \|\mathbf{x}_K\|_1 - \|\mathbf{w}_K\|_1 + \|\frac{\mathbf{w}_{\bar{K}}}{C}\| \\ \geq \; & \|\mathbf{x}_K\|_1, \end{aligned}$$

thus proving the forward part of this lemma. Now let us assume instead that $\exists \mathbf{w} \in Z$, such that $C\|\mathbf{w}_K\|_1 > \|\mathbf{w}_{\bar{K}}\|_1$. Then we can construct a vector $\mathbf{x}$ supported on the set K (or a subset of K), with $\mathbf{x}_K = -\mathbf{w}_K$. Then we have

$$\begin{aligned} & \|\mathbf{x}_K + \mathbf{w}_K\|_1 + \|\frac{\mathbf{w}_{\bar{K}}}{C}\| \\ = \; & 0 + \|\frac{\mathbf{w}_{\bar{K}}}{C}\| \\ < \; & \|\mathbf{x}_K\|_1, \end{aligned}$$

proving the inverse part of this lemma. ■

So the event that the condition in (3.3.1) on the null-space Z holds if and only if $\forall \;\; K \subseteq \{1, 2, ..., n\}$ with $|K| = k$, and $\forall \;\; \mathbf{x}$ supported on the set K (or on a subset of K),

$$\|\mathbf{x}_K + \mathbf{w}_K\|_1 + \|\frac{\mathbf{w}_{\bar{K}}}{C}\|_1 \geq \|\mathbf{x}_K\|_1, \forall \mathbf{w} \in Z. \tag{3.3.3}$$

Based on Lemma 3.3.1, we are now in a position to derive the probability that condition (3.3.1) holds for the sparsity $|K| = k$ if we uniformly sample a random $(n-m)$-dimensional subspace Z from the Grassmann manifold $\mathrm{Gr}_{(n-m)}(n)$. From the previous discussions, we can equivalently consider the complementary probability P, namely the probability there exists a subset $K \subset \{1, 2, ..., n\}$ with $|K| = k$, and a vector $\mathbf{x} \in R^n$ supported on the set K (or a subset of K) failing the condition (3.3.2). Due to the vector linear proportionality in the linear subspace Z, we can restrict our attention to those vectors $\mathbf{x}$ from the crosspolytope

$$\{\mathbf{x} \in R^n \mid \|\mathbf{x}\|_1 = 1\}$$

that are only supported on the set K (or a subset of K).

First, we upper bound the probability P by a union bound over all the possible support sets $K \subset \{1, 2, ..., n\}$ and all the sign patterns of the k-sparse vector $\mathbf{x}$. Since the k-sparse vector $\mathbf{x}$ has $\binom{n}{k}$ possible support sets of cardinality k and 2^k possible sign patterns (nonnegative or non-positive), we have

$$P \leq \binom{n}{k} \times 2^k \times P_{K,-} \tag{3.3.4}$$

,where $P_{K,-}$ is the probability that for a specific *support set*, there exist a k-sparse vector $\mathbf{x}$ of a specific *sign pattern* which fails the condition (3.3.2). By symmetry, without loss of generality, we assume the signs of the elements of $\mathbf{x}$ to be non-positive.

So 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 (or a subset of K) and can be restricted to the crosspolytope $\{\mathbf{x} \in R^n \mid \|\mathbf{x}\|_1 = 1\}$, $\mathbf{x}$ is also on a $(k-1)$-dimensional face, denoted by F, of the skewed crosspolytope SP:

$$\text{SP} = \{\mathbf{y} \in R^n \mid \|\mathbf{y}_K\|_1 + \|\frac{\mathbf{y}_{\bar{K}}}{C}\|_1 \leq 1\}. \tag{3.3.5}$$

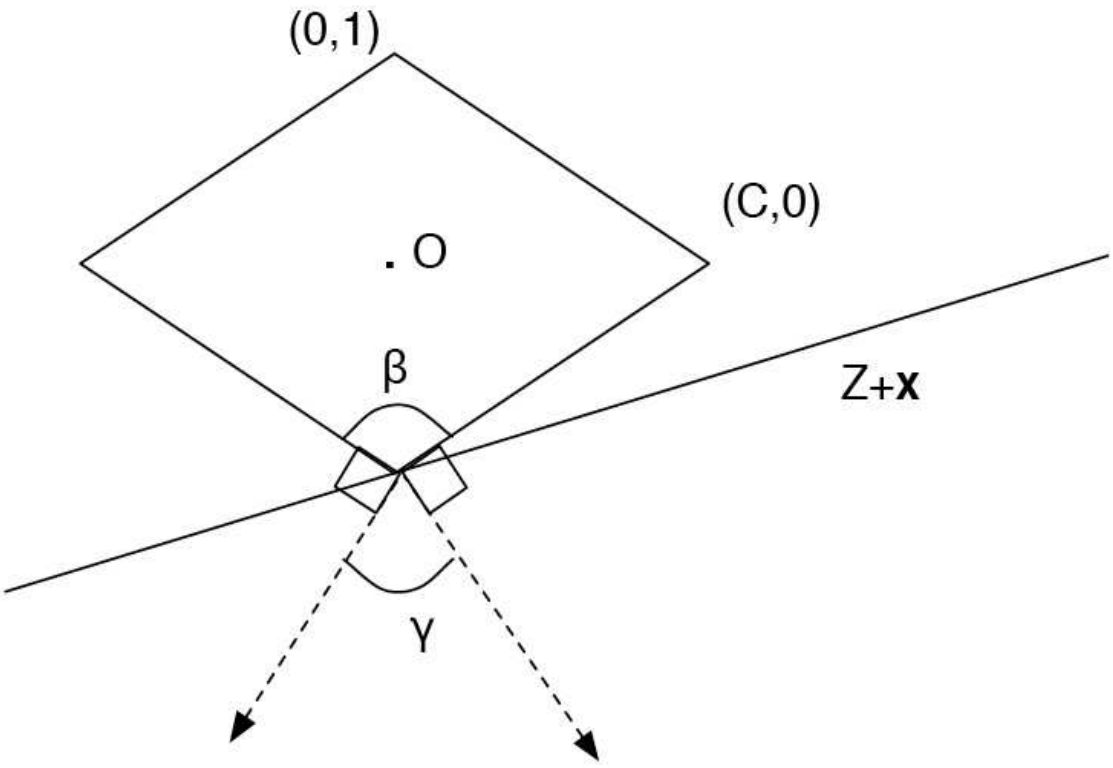


Figure 3.2: The Grassmann angle for a skewed crosspolytope

Now the probability $P_{K,-}$ is the probability that there exists an $\mathbf{x} \in F$, and there exists a $\mathbf{w} \in Z$ ($\mathbf{w} \neq 0$) such that

$$\|\mathbf{x}_K + \mathbf{w}_K\|_1 + \|\frac{\mathbf{w}_{\bar{K}}}{C}\|_1 \leq \|\mathbf{x}_K\|_1 = 1. \tag{3.3.6}$$

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

$$\|\mathbf{x}_K + \mathbf{w}_K\|_1 + \|\frac{\mathbf{w}_{\bar{K}}}{C}\|_1 \leq \|\mathbf{x}_K\|_1 = 1, \tag{3.3.7}$$

is essentially the probability that a uniformly chosen $(n-m)$-dimensional subspace Z shifted by the point $\mathbf{x}$, namely $(Z + \mathbf{x})$, intersects the skewed crosspolytope

$$\text{SP} = \{\mathbf{y} \in R^n \mid \|\mathbf{y}_K\|_1 + \|\frac{\mathbf{y}_{\bar{K}}}{C}\|_1 \leq 1\}, \tag{3.3.8}$$

non-trivially. Namely, at some other point besides $\mathbf{x}$.

From the linear property of the subspace Z, the event that $(Z + \mathbf{x})$ intersects the skewed crosspolytope SP is equivalent to the event that Z intersects nontrivially with the cone SP-Cone($\mathbf{x}$) obtained by observing the skewed polytope SP from the point $\mathbf{x}$. (Namely, SP-Cone($\mathbf{x}$) is conic hull of the point set (SP $- \mathbf{x}$) and of course SP-Cone($\mathbf{x}$) has the origin of the coordinate system as its apex.) However, as noticed in the geometry for convex polytopes [Grü68][Grü03], the SP-Cone($\mathbf{x}$) are identical for any $\mathbf{x}$ lying in the relative interior of the face F. This means that the probability $P_{K,-}$ is equal to $P'_{\mathbf{x}}$, regardless of the fact $\mathbf{x}$ is only a single point in the relative interior of the face F. (The acute reader may have noticed some singularities here because $\mathbf{x} \in F$ may not be in the relative interior of F, but it turns out that the SP-Cone($\mathbf{x}$) in this case is only a subset of the cone we get when $\mathbf{x}$ is in the relative interior of F. So we do not lose anything if we restrict $\mathbf{x}$ to be in the relative interior

of the face 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*** [Grü68] for the face F with respect to the polytope SP under the Grassmann manifold $\mathrm{Gr}_{(n-m)}(n)$:[1] the probability of a uniformly distributed $(n-m)$-dimensional subspace Z from the Grassmannian manifold $\mathrm{Gr}_{(n-m)}(n)$ intersecting non-trivially with the cone SP-Cone($\mathbf{x}$) formed by observing the skewed crosspolytope SP from the relative interior point $\mathbf{x} \in F$.

Building on the works by L. A. Santalö [San52] and P. McMullen [McM75] in high dimensional integral geometry and convex polytopes, the complementary Grassmann angle for the $(k-1)$-dimensional face F can be explicitly expressed as the sum of products of internal angles and external angles [Grü03]:

$$2 \times \sum_{s \geq 0} \sum_{G \in \Im_{m+1+2s}(\mathrm{SP})} \beta(F,G)\gamma(G,\mathrm{SP}), \tag{3.3.9}$$

where s is any nonnegative integer, G is any $(m+1+2s)$-dimensional face of the skewed crosspolytope ($\Im_{m+1+2s}(\mathrm{SP})$ is the set of all such faces), $\beta(\cdot,\cdot)$ stands for the internal angle and $\gamma(\cdot,\cdot)$ stands for the external angle.

The internal angles and external angles are basically defined as follows [Grü03][McM75]:

- An internal angle $\beta(F_1,F_2)$ is the fraction of the hypersphere S covered by the cone obtained by observing the face F_2 from the face F_1.[2] The internal angle $\beta(F_1,F_2)$ is defined to be zero when $F_1 \nsubseteq F_2$ and is defined to be one if $F_1 = F_2$.

- An external angle $\gamma(F_3,F_4)$ is the fraction of the hypersphere S covered by the cone of outward normals to the hyperplanes supporting the face F_4 at the face

[1] Grassman angle and its corresponding complementary Grassmann angle always sum up to 1. There is apparently inconsistency in terms of the definition of which is "Grassmann angle" and which is "complementary Grassmann angle" between [Grü68],[AS92] and [VS92] etc. But we will stick to the earliest definition in [Grü68] for Grassmann angle: the measure of the subspaces that intersect trivially with a cone.

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

F_3. The external angle $\gamma(F_3, F_4)$ is defined to be zero when $F_3 \nsubseteq F_4$ and is defined to be one if $F_3 = F_4$.

Let us take for example the 2-dimensional skewed crosspolytope,

$$\text{SP} = \{(y_1, y_2) \in R^2 |\ \|\mathbf{y}_2\|_1 + \|\frac{\mathbf{y}_1}{C}\|_1 \leq 1\},$$

(namely the diamond) in Figure 3.2, where $n = 2$, $(n - m) = 1$ and $k = 1$. Then the point $\mathbf{x} = (0, -1)$ is a 0-dimensional face (namely a vertex) of the skewed polytope SP. Now from their definitions, the internal angle $\beta(\mathbf{x}, \text{SP}) = \beta$ and the external angle $\gamma(\mathbf{x}, \text{SP}) = \gamma$, $\gamma(\text{SP}, \text{SP}) = 1$. The complementary Grassmann angle for the vertex $\mathbf{x}$ with respect to the polytope SP is the probability that a uniformly sampled 1-dimensional subspace (namely a line, we denote it by Z) shifted by $\mathbf{x}$ intersects non-trivially with $\text{SP} = \{(y_1, y_2) \in R^2 |\ \|\mathbf{y}_2\|_1 + \|\frac{\mathbf{y}_1}{C}\|_1 \leq 1\}$ (or equivalently the probability that Z intersects non-trivially with the cone obtained by observing SP from the point $\mathbf{x}$). It is obvious that this probability is 2β. The readers can also verify the correctness of the formula (3.3.9) very easily for this toy example.

For a general polytope, it might be hard to give explicit formula for the external and internal angles involved. Fortunately in the skewed crosspolytope case, both the internal angles and the external angles can be explicitly computed.

First, let us look at the internal angle $\beta(F, G)$ between the $(k-1)$-dimensional face F and a $(l-1)$-dimensional face G. Notice that the only interesting case is when $F \subseteq G$ since $\beta(F, G) = 0$ if $F \subseteq G$. We will see if $F \subseteq G$, the cone formed by observing G from F is the direct sum of a $(k-1)$-dimensional linear subspace and a convex polyhedral cone formed by $(l-k)$ unit vectors with inner product $\frac{1}{1+C^2k}$ between each other. In this case, the internal angle is given by

$$\beta(F, G) = \frac{V_{l-k-1}(\frac{1}{1+C^2k}, l-k-1)}{V_{l-k-1}(S^{l-k-1})}, \tag{3.3.10}$$

where $V_i(S^i)$ denotes the i-th dimensional surface measure on the unit sphere S^i, while $V_i(\alpha', i)$ denotes the surface measure for regular spherical simplex with $(i+1)$

vertices on the unit sphere S^i and with inner product as α' between these $(i+1)$ vertices. Thus (3.3.10) is equal to $B(\frac{1}{1+C^2k}, l-k)$, where

$$B(\alpha', m') = \theta^{\frac{m'-1}{2}}\sqrt{(m'-1)\alpha'+1}\pi^{-m'/2}{\alpha'}^{-1/2}J(m', \theta), \tag{3.3.11}$$

with $\theta = (1-\alpha')/\alpha'$ and

$$J(m', \theta) = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}(\int_0^{\infty} e^{-\theta v^2+2iv\lambda}\, dv)^{m'} e^{-\lambda^2}\, d\lambda. \tag{3.3.12}$$

Also, we can derive the external angle $\gamma(G, \mathrm{SP})$ between the $(l-1)$-dimensional face G and the skewed crosspolytope SP as

$$\gamma(G, \mathrm{SP}) = \frac{2^{n-l}}{\sqrt{\pi}^{n-l+1}}\int_0^{\infty} e^{-x^2}(\int_0^{\frac{x}{C\sqrt{k+\frac{l-k}{C^2}}}} e^{-y^2}\, dy)^{n-l}\, dx. \tag{3.3.13}$$

The derivations of these expressions involve the computations of the volumes of cones in high dimensional geometry and will be presented in the appendix.

3.4 Evaluating the Bound ζ

In summary, we finally get the probability

$$P \leq \binom{n}{k} \times 2^k \times 2 \times \sum_{s\geq 0}\sum_{G\in\Im_{m+1+2s}(\mathrm{SP})} \beta(F, G)\gamma(G, \mathrm{SP}). \tag{3.4.1}$$

First we define $l = (m+1+2s)+1$ and $\mu = \frac{l}{n}$. In the skewed crosspolytope SP, we notice that there are in total $\binom{n-k}{l-k}2^{l-k}$ faces G of dimension $(l-1)$ such that $F \subset G$ and $\beta(F, G) \neq 0$.In the skewed crosspolytope SP , we notice that there are in total $\binom{n-k}{l-k}2^{l-k}$ faces G of dimension $(l-1)$ such that $F \subset G$ and $\beta(F, G) \neq 0$. Because of the symmetry in the skewed crosspolytope SP, it follows from (3.4.1) that

$$P \leq \binom{n}{k}2^k \times 2\sum_{s\geq 0}\binom{n-k}{l-k}2^{l-k}\beta(F, G)\gamma(G, \mathrm{SP}), \tag{3.4.2}$$

where $l = (m + 1 + 2s) + 1$ and G is any single face of dimension $(l - 1)$ such that $F \subseteq G$.

We plan now to estimate $n^{-1} \log(P)$, decomposing it into a sum of terms involving logarithms of the combinatorial prefactor, the internal angle and the external angle. Define the Shannon entropy:

$$H(p) = p \log(1/p) + (1 - p) \log(1/(1 - p));$$

noting that here the logarithm base is e, rather than the customary base 2. It is easy to see that

$$n^{-1} \log \binom{n}{\lfloor pn \rfloor} \to H(p), p \in [0, 1], n \to \infty \tag{3.4.3}$$

so this provides a convenient summary for combinatorial terms. Defining $\nu = l/n \leq \delta$, we have

$$n^{-1} \log(P) = \nu \log_e(2) + H(\rho\delta) + H(\frac{\nu - \rho\delta}{1 - \rho\delta})(1 - \rho\delta) + R_1 \tag{3.4.4}$$

with remainder $R_1 = R_1(s, k, d, n)$.

Define then the growth exponent,

$$\psi_{com}(\nu; \rho, \delta) \equiv \nu \log_e(2) + H(\rho\delta) + H(\frac{\nu - \rho\delta}{1 - \rho\delta})(1 - \rho\delta), \tag{3.4.5}$$

describing the exponential growth of the combinatorial factors. Applying (3.4.3), we will see that the remainder R_1 in (3.4.4) is o(1) uniformly in the range $k - l > (\delta - \rho)n$, $n > n_0$.

We will define a so-called decay exponent $\psi_{ext}(\nu)$ and show that $\gamma(G, \mathrm{SP})$ decays exponentially at least at the rate $\psi_{ext}(\nu)$; for each $\epsilon > 0$,

$$n^{-1} \log(\gamma(G, \mathrm{SP})) \leq -\psi_{ext}(\nu) + \epsilon,$$

uniformly in $l \geq \delta n$, $n \geq n_0(\delta, \epsilon)$.

Similarly, Section 4.2 below defines a decay exponent $\psi_{int}(\nu; \rho\delta)$. Section 6 be-

low shows that the internal angle $\beta(F, G)$ indeed decays with this exponent; along sequences $k \sim \rho\delta n$, $l \sim \nu n$,

$$n^{-1} \log(\beta(F, G)) = -\psi_{int}(\nu; \rho\delta) + R_2,$$

where the remainder $R_2 < o(1)$ uniformly in $k - l \geq (\delta - \rho)n$.

Hence for any fixed choice of ρ, δ, for $\epsilon > 0$, and for $n \geq n_0(\rho, \delta, \epsilon)$, we have the inequality

$$n^{-1} \log(Ds) \leq \psi_{com}(\nu; \rho, \delta) - \psi_{int}(\nu; \rho, \delta) - \psi_{ext}(\nu; \rho, \delta) + 3\epsilon, \tag{3.4.6}$$

valid uniformly in s.

3.4.1 Defining ρ_N

Define now the *net exponent* $\psi_{net} = \psi_{com}(\nu; \rho, \delta) - \psi_{int}(\nu; \rho, \delta) - \psi_{ext}(\nu; \rho, \delta)$. We can define at last the mysterious ρ_N as the threshold where the net exponent changes sign. We will see that the components of ψ_{net} are all continuous over sets $\rho \in [\rho_0, 1], \delta \in [\delta_0, 1], \nu \in [\delta, 1]$, and so ψ_{net} has the same continuity properties.

Definition 3.4.1. *Let $\delta \in (0, 1]$. The critical proportion $\rho_N(\delta)$ is the supremum of $\rho \in [0, 1]$ obeying*

$$\psi_{net}(\nu; \rho, \delta) < 0, \qquad \nu \in [\delta, 1).$$

Continuity of ψ_{net} shows that if $\rho < \rho_N$ then, for some $\epsilon > 0$,

$$\psi_{net}(\nu; \rho, \delta) < -4\epsilon, \qquad \nu \in [\delta, 1).$$

Combine this with (3.4.6). Then for all $s = 0, 2, \ldots, (n - d)/2$ and all $n > n_0(\delta, \rho, \epsilon)$

$$n^{-1} \log(D_s) \leq -\epsilon.$$

This implies our main result.

3.5 Properties of Exponents

We now define the exponents ψ_{int} and ψ_{ext} and discuss properties of ρ_N.

3.5.1 Exponent for External Angle

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 $G(x) = Prob\{X \leq x\}$. It has density $g(x) = 2/\sqrt{\pi}\exp(-x^2)$. Writing this out,

$$G(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2}\, dy; \tag{3.5.1}$$

so G is just the classical error function *erf*. For $\nu \in (0, 1]$, define x_ν as the solution of

$$\frac{2xG(x)}{g(x)} = \frac{1-\nu}{\nu'}, \tag{3.5.2}$$

where

$$\nu' = (C^2 - 1)\rho\delta + \nu.$$

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 $2xG(x)/g(x)$ is one-one on the positive axis, and x_ν is well-defined, and a smooth, decreasing function of ν. This has limiting behavior $x_\nu \to 0$ as $\nu \to 1$ and $x_\nu \sim \sqrt{\log((1-\nu)/\nu)}$ as $\nu \to 0$. Define now

$$\psi_{ext}(\nu) = -(1-\nu)\log(G(x_\nu)) + \nu x_\nu^2.$$

This function is smooth on the interior of (0, 1), with endpoints $\psi_{ext}(1) = 0$, $\psi_{ext}(0) = 0$. When $C = 1$, a useful fine point is the asymptotic [Don06c]

$$\psi_{ext}(\nu) \sim \nu\log(\frac{1}{\nu}) - \frac{1}{2}\nu\log(\log(\frac{1}{\nu})) + O(\nu), \nu \to 0. \tag{3.5.3}$$

3.5.2 Exponent for Internal Angle

Let Y be the standard half-normal random variable $HN(0,1)$; this has cumulant generating function $\Lambda(s) = \log(E(\exp(sY))$. Very convenient for us is the exact formula

$$\Lambda(s) = \frac{s^2}{2} + \log(2\Phi(s)),$$

where Φ is the usual cumulative distribution function of a standard Normal $N(0,1)$. The cumulant generating function Λ has a rate function (Fenchel-Legendre dual)

$$\Lambda^*(y) = \max_s sy - \Lambda(y)$$

This is smooth and convex on $(0,\infty)$, strictly positive except at $\mu = E(Y) = \sqrt{2/\pi}$. More details are provided in the following sections. For $\gamma' \in (0,1)$ let

$$\xi_{\gamma'}(y) = \frac{1-\gamma'}{\gamma'} y^2/2 + \Lambda^*(y). \tag{3.5.4}$$

The function $\xi_{\gamma'}(y)$ is strictly convex and positive on $(0,\infty)$ and has a unique minimum at a unique $y_{\gamma'}$ in the interval $(0, \sqrt{2/\pi})$. We define

$$\gamma' = \frac{\rho\delta}{\frac{C^2-1}{C^2}\rho\delta + \frac{\nu}{C^2}},$$

then we have the external angle exponent as

$$\Psi_{int}(\nu;\rho,\delta) = \xi_{\gamma'}(y_{\gamma'})(\nu - \rho\delta) + \log(2)(\nu - \rho\delta). \tag{3.5.5}$$

For fixed ρ, δ, Λ_{int} is continuous in $\nu \geq \delta$. Most importantly, in the section below, we get the asymptotic formula

$$\xi_{\gamma'}(y_{\gamma'}) \sim \frac{1}{2}\log(\frac{1-\gamma'}{\gamma'}), \gamma' \to 0. \tag{3.5.6}$$

Because $\gamma' = \frac{\rho\delta}{\frac{C^2-1}{C^2}\rho\delta + \frac{\nu}{C^2}}$, (3.5.6) means for small ρ, $\nu \in [\delta, 1]$ and any given $\eta > 0$

$$\Psi_{int}(\nu, \rho\delta) \geq (\frac{1}{2} \cdot \log(\frac{1-\gamma'}{\gamma'})(1-\eta) + \log(2))(\nu - \rho\delta). \tag{3.5.7}$$

3.5.3 Combining the Exponents

We now consider the combined behavior of Ψ_{com}, Ψ_{int} and Ψ_{net}. We think of these as functions of ν with ρ, δ as parameters. The combinatorial exponent Ψ_{com} is the sum of a linear function in ν, and a scaled, shifted version of the Shannon entropy, which is a symmetric, roughly parabolic shaped function. This is the exponent of a growing function which must be outweighed by the sum $\Psi_{int} + \Psi_{net}$.

3.5.4 Properties of ρ_N

The asymptotic relations (3.5.3) and (3.5.6) allow us to see two key facts about ρ_N, both proved in the appendix. Firstly, the concept is nontrivial:

Lemma 3.5.1. *For any $\delta > 0$ and any $C > 1$, we have*

$$\rho_N > 0, \delta \in (0, 1). \tag{3.5.8}$$

Secondly, one can show that, although $\rho_N \to 0$ as $\delta \to 0$, it goes to zero slowly.

Lemma 3.5.2. *For all $\eta > 0$,*

$$\rho_N(\delta) \geq \log(\frac{1}{\delta})^{-(1+\eta)}, \quad \delta \to 0. \tag{3.5.9}$$

Lemma 3.5.3. *For a fixed $\delta > 0$, define $\rho_N(\delta; C)$ as the $\rho(N; \delta)$ for a certain $C > 1$. Then*

$$\Omega(\frac{1}{C^2}) \leq \rho_N(\delta; C) \leq \frac{1}{C+1}, \quad \text{as} \;\; C \to 0, \tag{3.5.10}$$

where $\Omega(\frac{1}{C^2}) \leq \rho_N(\delta; C)$ *means that there exists a constant* $\iota(\delta)$ *and a* C_0 *such that for all* $C > C_0$*,*

$$\frac{\iota(\delta)}{C^2} \leq \rho_N(\delta; C).$$

3.6 Bounds on the External Angle

We now justify the use of Ψ_{ext}.

Lemma 3.6.1. *Fix* δ,$\epsilon > 0$*,*

$$n^{-1} \log(\gamma(G, SP)) < -\Psi_{net}(l/n) + \epsilon, \tag{3.6.1}$$

uniformly in $l > \delta n$*,* $n \geq n_0(\delta, \epsilon)$*.*

We start from an exact identity. We know that we have the explicit integral formula

$$\gamma(G, \mathrm{SP}) = \frac{2^{n-l}}{\sqrt{\pi}^{n-l+1}} \int_0^\infty e^{-x^2} (\int_0^{\frac{x}{C\sqrt{k+\frac{l-k}{C^2}}}} e^{-y^2}\, dy)^{n-l}\, dx. \tag{3.6.2}$$

After a changing of integral variables, we have

$$\gamma(G, \mathrm{SP}) = \sqrt{\frac{(C^2-1)k + l}{\pi}} \int_0^\infty e^{-((C^2-1)k+l)x^2} (\frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2}\, dy)^{n-l}\, dx. \tag{3.6.3}$$

We notice the term in braces as the error function G from (3.5.1). Let $\nu = l/n$, $\nu' = (C^2-1)\rho\delta + \nu$ then the integral formula can be written as

$$\sqrt{\frac{n\nu'}{\pi}} \int_0^\infty e^{-n\nu' x^2 + n(1-\nu)\log(G(x))}\, dx. \tag{3.6.4}$$

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

$$f_{\rho,\delta,\nu,n}(y) = e^{-n\psi_{\rho,\delta,\nu}(y)} \cdot \sqrt{\frac{n\nu'}{\pi}}, \tag{3.6.5}$$

with

$$\psi_{\rho,\delta,\nu}(y) = \nu' y^2 - (1-\nu)\log(G(y)).$$

We note that $\psi_{\rho,\delta,\nu}$ is smooth and convex and (in the appendix) develop expressions for its second and third derivatives. Applying Laplaces method to $\psi_{\rho,\delta,\nu}$ in the usual way, but taking care about regularity conditions and remainders, gives a result with the uniformity in ν, which is crucial for us.

Lemma 3.6.2. *For $\nu \in (0,1)$, let x_ν denote the minimizer of $\psi_{\rho,\delta,\nu}$. Then*

$$\int_0^\infty f_{\rho,\delta,\nu,n}(x)\,dx \le e^{-n\psi_{\rho,\delta,\nu}(x_\nu)}(1+R_n(\nu)),$$

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

$$\sup_{\nu\in[\delta,1-\eta]} R_n(\nu) = o(1) \quad as \quad n \to \infty.$$

The minimizer x_ν is exactly the same x_ν defined earlier in (3.5.2) and the minimum value in this lemma is the same as the defined exponent Ψ_{ext}:

$$\Psi_{ext}(\nu) = \psi_{\rho,\delta,\nu}(x_\nu). \tag{3.6.6}$$

In fact, we can derive Lemma 3.6.2 from Lemma 3.6.1. We note that as $\nu \to 1$, $x_\nu \to 0$ and $\Psi_{ext}(\nu) \to 0$. For given $\epsilon > 0$ in the statement of Lemma 3.6.1, there is a largest $\nu_\epsilon < 1$ with $\Psi_{ext}(\nu) \to 0$. Note that $\gamma(G, \mathrm{SP}) \le 1$, so that for $l > \nu_\epsilon n$,

$$n^{-1}\log(\gamma(G,\mathrm{SP})) \le 0 < -\Psi_{ext}(\nu) + \epsilon,$$

for $n \ge 1$. Consider now $l \in [\delta n, \nu_\epsilon n]$. Based on (3.6.4),

$$\gamma(G,\mathrm{SP}) = \int_0^\infty f_{\rho,\delta,\nu,n}(y)\,dx.$$

Applying the uniformity in ν given in Lemma 3.6.2, we have as $n \to \infty$,

$$n^{-1} \log(\gamma(G, \mathrm{SP})) = \psi_\nu(x_\nu) + o(1), l \geq \delta n.$$

So from the identity (3.6.6), we get

$$n^{-1} \log(\gamma(G, \mathrm{SP})) \leq -\Psi_{net}(l/n) + o(1). \tag{3.6.7}$$

Then Lemma 3.6.1 follows.

Now it remains to prove the uniform Laplace method Lemma (3.6.2). We will follow the same line of reasoning given in [Don06c]. First, we state explicitly the key lemma from [Don06c].

Lemma 3.6.3. *[Don06c] Let $\psi(x)$ be convex in x and C^2 on an interval I and suppose that it takes its minimum at an interior point $x_0 \in I$, where $\psi'' > 0$ and that in a vicinity $(x_0 - \epsilon, x_0 + \epsilon)$ of x_0:*

$$|\psi''(x) - \psi''(x_0)| \leq D|\psi''(x_0)||x - x_0|. \tag{3.6.8}$$

Let $\bar{\psi}$ be the quadratic approximation $\psi(x_0) + \psi''(x_0)(x - x_0)^2/2$. Then

$$\int_I \exp(-n\psi(x))\, dx \leq \int_{-\infty}^{\infty} \exp -n\bar{\psi}(x)\, dx \cdot (S_{1,n} + S_{2,n}),$$

where

$$S_{1,n} = \exp(n\psi''(x_0)D\epsilon^3/6),$$
$$S_{2,n} = 2/\left(n\epsilon(2\pi|\psi''(0)|)^{\frac{1}{2}}(1 - \frac{1}{2}D\epsilon^2)\right).$$

The constant D in this lemma can be as a scaled third derivative, since if ψ is C^3, we can take

$$D = \sup_{(x_0-\epsilon, x_0+\epsilon)} \psi^{(3)}(x)/\psi''(x).$$

Based on Lemma 3.6.3, we can derive the uniformity in Lemma 3.6.2. In fact, if we pick $\epsilon_n = n^{-\frac{2}{5}}$ and let $n \geq n_0(\psi''(x_0), D)$, where $n_0(\psi''(x_0), D)$ is a number depending only on $\psi''(x_0)$ and D, we can have

$$\int_I e^{-n\psi(x)}\,dx \leq \int_{-\infty}^{\infty} e^{-n\bar{\psi}(x)}\,dx \cdot (1 + o(1)). \tag{3.6.9}$$

Here the term $o(1)$ is uniform over any collection of convex functions with a given $\psi''(x_0)$ and D. Now we consider the collection of convex functions $\psi(\nu)$ $(\nu \in [\delta, 1-\eta])$ in Lemma 3.6.2. Following the derivations in [Don06c], if we can show that there exist a certain $\epsilon > 0$ so that ψ''_{x_0} and D is bounded for the function $\psi_\nu(x)$ uniformly over the range $\nu \in [\delta, 1-\eta]$). Indeed, this is true based on the following Lemma 3.6.4.

Lemma 3.6.4. *The function ψ_ν is C^∞ with second derivative at the minimum,*

$$\psi''_\nu(x_\nu) = 2\nu' + 4x_\nu^2\nu' + \frac{4x_\nu^2\nu'}{1-\nu}, \tag{3.6.10}$$

and third derivative at the minimum,

$$\psi^{(3)}_\nu(x_\nu) = (1-\nu)\left((2 - 4x_v^2)z - 6x_\nu z^2 - 2z^3\right), \tag{3.6.11}$$

where $z = z_\nu = 2\nu' x_\nu/(1-\nu)$. We have

$$0 < 2\delta \leq \inf_{\nu\in[\delta,1]} \psi''_\nu(x_\nu),$$

and

$$\sup_{\nu\in[\delta,1-\eta]} \psi''_\nu(x_\nu) < \infty.$$

Moreover, for small enough $\epsilon > 0$, the ratio

$$D(\epsilon;\delta,\eta) = \sup_{\nu\in(\delta,1-\eta]} \sup_{|x-x_\nu|<\epsilon} \left|\psi^{(3)}_\nu(x)/\psi''_\nu(x)\right|$$

is finite.

Proof. We can get the following first, second, third derivatives of the function $\psi_\nu(x)$:

$$\psi'_\nu(x) = -(1-\nu)g/G + 2\nu' x;$$

$$\psi''_\nu(x) = -(1-\nu)(g'/G - g^2/G^2) + 2\nu';$$

$$\psi^{(3)}_\nu(x) = -(1-\nu)(g''/G - 3g'g/G^2 + 2g^3/G^3).$$

Because $g' = (-2x)g$, $g'' = (-2+4x^2)g$, and

$$g(x_\nu)/G(x_\nu) = \frac{2\nu x_\nu}{1-\nu} = z_\nu$$

at the point x_ν, we can immediately have (3.6.10) and (3.6.11).

Notice that $\psi''_\nu(x_\nu) \geq 2\nu'$, so it is bounded away from zero on any interval $\nu \in [\delta, 1]$, $\delta > 0$. Also, since x_ν is a continuous function bounded away from zero over ν on the interval $[\delta, 1-\eta]$ $(\delta, \eta > 0)$, we have $\psi''_\nu(x_\nu)$ is also bounded above over $[\delta, 1-\eta]$.

Now as for $\psi^{(3)}$, we note that clearly x_ν and z_v are continuous functions on $[\delta, 1)$. And both are bounded on the interval $\nu \in [\delta, 1-\eta]$. As a polynomial in ν, x_ν and Z_ν, $\psi^{(3)}_v$ is also bounded. If we consider the interval $(x_\nu - \epsilon, x_\nu - \epsilon)$, the boundness of the ratio $D(\epsilon; \delta, \eta)$ also holds uniformly over $\nu \in [\delta, 1-\eta]$ by inspection if $\epsilon > 0$ is small enough.

■

3.7 Bounds on the Internal Angle

In this section, we will show how to get the internal angle decay exponent, namely proving the following lemma:

Lemma 3.7.1. *For $\epsilon > 0$ and $n > n_0(\epsilon, \delta, \rho)$,*

$$n^{-1} \log(\beta(F, G)) \leq \Psi_{int}(l/n; k/n) + \epsilon,$$

uniformly in $l \geq \delta n$, $k \geq \rho n$, $(l-k) \geq (\delta - \rho)n$.

Recall that the decaying exponent is

$$n^{-1}\log(\beta(F,G)) = n^{-1}\log(B(\frac{1}{1+C^2k}, l-k)), \tag{3.7.1}$$

where

$$B(\alpha', m') = \theta^{\frac{m'-1}{2}}\sqrt{(m'-1)\alpha'+1}\pi^{-m'/2}{\alpha'}^{-1/2}J(m',\theta), \tag{3.7.2}$$

with $\theta = (1-\alpha')/\alpha'$, and

$$J(m',\theta) = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}(\int_0^{\infty} e^{-\theta v^2+2iv\lambda}\,dv)^{m'}e^{-\lambda^2}\,d\lambda. \tag{3.7.3}$$

To evaluate (3.7.1), we need to evaluate the complex integral in $J(m',\theta')$. A saddle point method based on contour integration was sketched for similar integral expressions in [VS92]. A probabilistic method using large deviation theory for evaluating similar integrals was developed in [Don06c]. Both of these two methods can be applied in our case and of course they will produce the same final results. So we will follow the probabilistic method from [Don06c] in this chapter. The basic idea then is to see the integral in $J(m',\theta')$ as the convolution of $(m+1)$ probability densities being expressed in the Fourier domain. More explicitly, we have the following lemma.

Lemma 3.7.2. *let* $\theta = (1-\alpha')/\alpha'$, *where* $\alpha' = \frac{1}{C^2k+1}$. *Let* T *be a random variable with the* $N(0,\frac{1}{2})$ *distribution, and let* $W_{m'}$ *be a sum of* m' *i.i.d. half normals* $U_i \sim HN(0,\frac{1}{2\theta})$. *Let* T *and* $W_{m'}$ *be stochastically independent, and let* $g_{T+W_{m'}}$ *denote the probability density function of the random variable* $T+W_{m'}$. *Then*

$$B(\alpha', m') = \sqrt{\frac{\alpha'(m'-1)+1}{1-\alpha'}}\cdot 2^{-m'}\cdot\sqrt{\pi}\cdot g_{T+W_{m'}}(0). \tag{3.7.4}$$

Applying this probabilistic interpretation and large deviation techniques, it is

evaluated as in [Don06c] that

$$g_{T+W_{m'}} \leq \frac{2}{\sqrt{\pi}} \cdot \left(\int_0^{\mu_{m'}} v e^{-v^2 - m'\Lambda^*(\frac{\sqrt{2\theta}}{m'}v)} \, dv + e^{-\mu_m^2} \right), \tag{3.7.5}$$

where Λ^* is the rate function for the standard half-normal random variable $HN(0,1)$ and $\mu_{m'}$ is the expectation of $W_{m'}$, namely $\mu_{m'} = EW_{m'}$. In fact, the second term in the sum is argued to be negligible [Don06c]. And after changing variables $y = \frac{\sqrt{2\theta}}{m'}v$, we know that the first term is upperbounded by

$$\frac{2}{\sqrt{\pi}} \cdot \frac{m'^2}{2\theta} \cdot \int_0^{\sqrt{2/\pi}} y e^{-m'(\frac{m'}{2\theta})y^2 - m'\Lambda^*(y)} \, dy. \tag{3.7.6}$$

3.7.1 Laplace's Method for Ψ_{int}

As we know, m' in the exponent of (3.7.6) is defined as $(l-k)$ for our case. Similar to evaluating the external angle decay exponent, we will resort to the Laplace's method in evaluating the internal angle decay exponent. In fact, we can see the function $\xi_{\gamma'}$ of (3.5.4) in the exponent of (3.7.6), with $\gamma' = \frac{\theta}{m'+\theta}$. Since $\theta = \frac{1-\alpha'}{\alpha'} = C^2 k$, we have

$$\gamma' = \frac{\theta}{m'+\theta} = \frac{C^2 k}{(C^2-1)k + l}.$$

Since $k \sim \rho\delta n$, $l \sim \nu n$,

$$\gamma' = \frac{k}{\frac{l}{C^2} + \frac{C^2-1}{C^2}k} = \frac{\rho\delta}{\frac{C^2-1}{C^2}\rho\delta + \frac{\nu}{C^2}}.$$

Define the integral

$$f_{\gamma',m'}(y) = y e^{-m'\xi_{\gamma'}(y)}$$

If we apply similar arguments as in proving Lemma 3.6.2 and take care of the uniformity, we will have the following lemma.

Lemma 3.7.3. *For $\gamma' \in (0,1]$ let $y_{\gamma'} \in (0,1)$ denote the minimizer of $\xi_{\gamma'}$. Then*

$$\int_0^\infty f_{\gamma',m'}(x)\,dx \le e^{-m'\xi_{\gamma'}(y_{\gamma'})} \cdot R_{m'}(\gamma),$$

where, for $\eta > 0$,

$$m'^{-1} \sup_{\gamma\in[\eta,1]} \log(R'_m(\gamma)) = o(1) \quad as \quad m' \to \infty.$$

This means that

$$g_{T+W_{m'}}(0) \le e^{-m'\xi_{\gamma'}(y_{\gamma'})} R_{m'}(\gamma).$$

So applying (3.7.4), we get

$$n^{-1}\log(\beta(F,G)) \le (-\xi_{\gamma'}(y_{\gamma'}) + \log(2))\,(\nu - \rho\delta) + o(1),$$

where the $o(1)$ is uniform over a range of k and l.

3.7.2 Asymptotics of $\xi_{\gamma'}$

As in our previous discussion, we define $\gamma' = \frac{\rho\delta}{\frac{C^2-1}{C^2}\rho\delta + \frac{\nu}{C^2}}$, so γ' can take any value in the range $(0,1]$. As in [Don06c], using the convex duality associated to the cumulant generating function $\Lambda(s)$ and its dual Λ^*, we have

$$y = \Lambda'(s), \qquad s = (\Lambda^*)'(y),$$

defining a one-one relationship $s = s(y)$ and $y = y(s)$ between $s < 0$ and $0 < y < \sqrt{\frac{2}{\pi}}$.

From these relations, following the same line of reasoning in [Don06c], we can get the minimizer $y_{\gamma'}$ of $\xi_{\gamma'}$,

$$\frac{1-\gamma'}{\gamma'} y_{\gamma'} = -s_{\gamma'}, \tag{3.7.7}$$

where $s_{\gamma'} = s(y_{\gamma'})$.

Because the cumulant generating function for a standard half-normal random

variable $HN(0,1)$ is $\Lambda(s) = s^2/2 + \log(2\Phi(s))$ (recall ϕ and Φ are the normal density and cumulative), we have from $y = \Lambda'(s)$ that

$$y(s) = s \cdot (1 - \frac{1}{M(s)}), s < 0, \tag{3.7.8}$$

where the function of $M(s)$ is defined on $s < 0$ with $0 < M(s) < 1$ and $M(s) \to 1$ as $s \to -\infty$ so that

$$\Phi(s) = M(s) \cdot \frac{\phi(s)}{|s|}.$$

Combining (3.7.7) and (3.7.8), we know that

$$M(s_{\gamma'}) = 1 - \gamma'. \tag{3.7.9}$$

Further, we can derive that

$$\xi_{\gamma'}(y_{\gamma'}) = -\frac{1}{2} y_{\gamma'}^2 \frac{1-\gamma'}{\gamma'} - \log(2/\pi)/2 + \log(y_{\gamma'}/\gamma'). \tag{3.7.10}$$

So by the property of the function $M(s)$ and (3.7.9), as $\gamma' \to 0$, $s_{\gamma'} \to -\infty$. We have

$$Ee^{sY} = \frac{2}{2\sqrt{\pi}} \frac{M(s)}{s} \sim \frac{2}{2\sqrt{\pi}} \frac{1}{|s|}, \qquad s \to -\infty.$$

Hence $\Lambda(s) \sim -\log|s|$ and $\Lambda'(s) \sim -\frac{1}{s}$ as $s \to -\infty$. So by $y = \Lambda'(s)$, we have

$$y(s) \sim \frac{1}{|s|} \quad s \to -\infty,$$

and by combining this with (3.7.7), we get further

$$y_{\gamma'} \sim \sqrt{\frac{\gamma'}{1-\gamma'}}.$$

This then leads to the asymptotics of $\xi_{\gamma'}(y_{\gamma'})$ in this section.

3.8 "Weak", "Sectional" and "Strong" Robustness

So far, we have discussed the robustness of ℓ_1 minimization for sparse signals recovery in the "strongest" case, namely the robustness for all the approximately k-sparse signal vectors $\mathbf{x}$. But in applications or analysis, we are also often interested in the signal recovery robustness in weaker senses. As we shall see, the framework given in the previous sections can be naturally extended to the analysis of other notions of robustness for sparse signal recovery, resulting in a coherent analysis scheme. In this section, we will introduce our null-space conditions on the matrix A to guarantee the solution of (3.1.2) approximates the solution of (3.1.1) in the "weak", "sectional" and "strong" senses in a unified way.

Theorem 3.8.1. *Let the $m \times n$ measurement matrix be A, $\mathbf{x}$ be a n-element vector and $\mathbf{y} = A\mathbf{x}$. Let K be a subset of $\{1, 2, \ldots, n\}$ such that $|K| = k$ and further, let $\bar{K} = \{1, 2, \ldots, n\} \setminus K$. Let $\mathbf{w}$ denote an $n \times 1$ vector.*

- *(Weak Robustness) Suppose the part of $\mathbf{x}$ supported on K, namely $\mathbf{x}_K$ is fixed. With $C > 1$, the solution $\hat{\mathbf{x}}$ produced by (3.1.2) will satisfy $\forall\ \mathbf{x}_{\bar{K}}$,*

$$\|\mathbf{x}_K\|_1 - \|\hat{\mathbf{x}}_K\|_1 \leq \frac{2}{C-1}\|\mathbf{x}_{\bar{K}}\|_1,$$

and

$$\|(\mathbf{x} - \hat{\mathbf{x}})_{\bar{K}}\|_1 \leq \frac{2C}{C-1}\|\mathbf{x}_{\bar{K}}\|_1,$$

if and only if $\forall\ \mathbf{w} \in \boldsymbol{R}^n$ such that $A\mathbf{w} = 0$, we have

$$\|\mathbf{x}_K + \mathbf{w}_K\|_1 + \|\frac{\mathbf{w}_{\bar{K}}}{C}\|_1 \geq \|\mathbf{x}_K\|_1. \tag{3.8.1}$$

- *(Sectional Robustness) With $C > 1$, for a specific set $K \subseteq \{1, 2, \ldots, n\}$, the solution $\hat{\mathbf{x}}$ produced by (3.1.2) will satisfy $\forall\ \mathbf{x} \in \boldsymbol{R}^n$*

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq \frac{2(C+1)}{C-1}\|\mathbf{x}_{\bar{K}}\|_1,$$

if and only if $\forall\ \mathbf{x}' \in \boldsymbol{R}^n$, $\forall\ \mathbf{w} \in \boldsymbol{R}^n$ *such that* $A\mathbf{w} = 0$,

$$\|\mathbf{x}'_K + \mathbf{w}_K\|_1 + \|\frac{\mathbf{w}_{\bar{K}}}{C}\|_1 \geq \|\mathbf{x}'_K\|_1. \tag{3.8.2}$$

- *(Strong Robustness) With* $C > 1$, $\forall K \subseteq \{1, 2, \ldots, n\}$, $\forall\ \mathbf{x} \in \boldsymbol{R}^n$, *the solution* $\hat{\mathbf{x}}$ *produced by (3.1.2) satisfies*

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq \frac{2(C+1)}{C-1}\|\mathbf{x}_{\bar{K}}\|_1,$$

if and only if $\forall\ K \subseteq \{1, 2, \ldots, n\}, \forall\ \mathbf{x}' \in \boldsymbol{R}^n$, $\forall\ \mathbf{w} \in \boldsymbol{R}^n$ *such that* $A\mathbf{w} = 0$,

$$\|\mathbf{x}'_K + \mathbf{w}_K\|_1 + \|\frac{\mathbf{w}_{\bar{K}}}{C}\|_1 \geq \|\mathbf{x}'_K\|_1. \tag{3.8.3}$$

Proof. Sufficiency: First we consider the "weak" norm robustness. Let $\mathbf{w} = \hat{\mathbf{x}} - \mathbf{x}$ and we must have $A\mathbf{w} = A(\hat{\mathbf{x}} - \mathbf{x}) = 0$. From the triangular inequality for ℓ_1 norm and the fact that $\|\mathbf{x}\|_1 \geq \|\mathbf{x} + \mathbf{w}\|_1$, we have

$$\begin{aligned} &\|\mathbf{x}_K\|_1 - \|\mathbf{x}_K + \mathbf{w}_K\|_1 \\ \geq\ &\|\mathbf{w}_{\bar{K}} + \mathbf{x}_{\bar{K}}\|_1 - \|\mathbf{x}_{\bar{K}}\|_1 \\ \geq\ &\|\mathbf{w}_{\bar{K}}\|_1 - 2\|\mathbf{x}_{\bar{K}}\|_1. \end{aligned}$$

But the condition (3.8.1) guarantees that

$$\|\mathbf{w}_{\bar{K}}\|_1 \geq C(\|\mathbf{x}_K\|_1 - \|\mathbf{x}_K + \mathbf{w}_K\|_1),$$

so we have

$$\|\mathbf{w}_{\bar{K}}\|_1 \leq \frac{2C}{C-1}\|\mathbf{x}_{\bar{K}}\|_1,$$

and

$$\|\mathbf{x}_K\|_1 - \|\hat{\mathbf{x}}_K\|_1 \leq \frac{2}{C-1}\|\mathbf{x}_{\bar{K}}\|_1.$$

For the "sectional" robustness, again, we let

$$\mathbf{w} = \hat{\mathbf{x}} - \mathbf{x}.$$

Then there must exist an $\mathbf{x}' \in \mathbf{R}^n$ such that

$$\|\mathbf{x}'_K + \mathbf{w}_K\|_1 = \|\mathbf{x}'_K\|_1 - \|\mathbf{w}_K\|_1.$$

Following the condition (3.8.2), we have

$$\|\mathbf{w}_K\|_1 \leq \|\frac{\mathbf{w}_{\bar{K}}}{C}\|_1.$$

Since

$$\|\mathbf{x}\|_1 \geq \|\mathbf{x} + \mathbf{w}\|_1,$$

following the proof of Theorem 1, we have

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq \frac{2(C+1)}{C-1}\|\mathbf{x}_{\bar{K}}\|_1.$$

The sufficiency of the condition (3.8.3) for strong robustness also follows.

Necessity: Since in the proof of the sufficiency, equalities can be achieved in the triangular equalities, the conditions (3.8.1), (3.8.2) and (3.8.3) are also necessary conditions for the respective robustness to hold for every $\mathbf{x}$ (otherwise, for certain $\mathbf{x}$'s, there will be $\mathbf{x}' = \mathbf{x} + \mathbf{w}$ with $\|\mathbf{x}'\|_1 < \|\mathbf{x}\|_1$ while violating the respective robustness definitions. Also, such $\mathbf{x}'$ can be the solution to (3.1.2)). ∎

The conditions for "weak", "sectional" and "strong" robustness are very similar except that the "weak" robustness condition is for $\mathbf{x}$ with a specific $\mathbf{x}_k$ on a specific subset K, the "sectional" robustness condition is for $\mathbf{x}$ with all possible $\mathbf{x}_K$'s on a specific subset K, and the "strong" robustness conditions are for $\mathbf{x}$ with all possible $\mathbf{x}_K$'s on all possible subsets. Basically, the "weak norm robustness" condition (3.8.1) guarantees that the ℓ_1 norm of $\hat{\mathbf{x}}_K$ is not too far away from the ℓ_1 norm of $\mathbf{x}_K$ and

the error vector $\mathbf{w}_{\bar{K}}$ is small in ℓ_1 norm when $\|\mathbf{x}_{\bar{K}}\|_1$ is small. Notice that if we define

$$\kappa = \max_{A\mathbf{w}=0, \mathbf{w}\neq 0} \frac{\|\mathbf{w}_K\|_1}{\|\mathbf{w}_{\bar{K}}\|_1},$$

then

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq \frac{2C(1+\kappa)}{C-1}\|\mathbf{x}_{\bar{K}}\|_1.$$

That means, if κ is not ∞ for a measurement matrix A, $\|\mathbf{x} - \hat{\mathbf{x}}\|_1$ is also small when $\|\mathbf{x}_{\bar{K}}\|_1$ is small. Indeed, it is not hard to see that, for a given matrix A

$$\kappa < \infty$$

as long as the rank of the measurement matrix A is no smaller than $|K| = k$, which is generally satisfied for $k \leq m$.

Despite the fact that the "weak" robustness condition is only for one specific signal $\mathbf{x}$, the "sectional" robustness condition instead guarantees that given *any* approximately k-sparse signal mainly supported on the subset K, the ℓ_1-minimization gives a solution $\hat{\mathbf{x}}$ close to the original signal by satisfying (3.1.3). When we measure an approximately k-sparse signal $\mathbf{x}$ (the support of the k largest-magnitude components is fixed though unknown to the decoder) using a randomly generated measurement matrix A, the "sectional" robustness conditions characterize the probability that the ℓ_1 minimization solution satisfies (3.1.3) for *any* signals for the set K. If that probability goes to 1 as $n \to \infty$ for any subset K, we know that there exist measurement matrices A's that guarantee (3.1.3) on "almost all" support sets (namely, (3.1.3) is "almost always" satisfied). The "strong" robustness conditions instead guarantee the recovery robustness for approximately sparse signals mainly supported on *any* subset K. The "strong" robustness conditions are useful in guaranteeing robustness *universally* for *all* approximately k-sparse signals *under a single measurement matrix* A.

Interestingly, after we take $C = 1$ and let (3.8.1), (3.8.2) and (3.8.3) take strict inequality for all $\mathbf{w} \neq 0$ in the null space of A, the conditions (3.8.1), (3.8.2) and

(3.8.3) are also sufficient and necessary conditions for unique exact recovery of ideally k-sparse signals in "weak", "sectional" and "strong" senses [Don06c], namely the unique exact recovery of a specific ideally k-sparse signal, the unique exact recoveries of all ideally k-sparse signal on a specific support set K and the unique exact recoveries of all ideally k-sparse signal on all possible support sets K. In fact, if $\|\mathbf{x}_{\bar{K}}\|_1 = 0$, from similar triangular inequality derivations in Theorem 1, we have $\hat{\mathbf{x}} = \mathbf{x}$ under all the three conditions.

For a given value $\alpha = \frac{m}{n}$ and any value $C \geq 1$, we will determine the value of feasible $\zeta = \frac{k}{n}$ for which there exist a sequences of A's such that these three conditions are satisfied when $n \to \infty$ and $\frac{m}{n} = \alpha$. It turns out that for a specific A, it is very hard to check whether these three conditions hold. Instead, we consider randomly choosing A from a certain distribution, and analyze for what ζ, these three conditions are satisfied with overwhelming probability. The standard results on compressed sensing assume that the matrix A has i.i.d. $\mathcal{N}(0, 1)$ entries. In this case, the null-space of A is rotationally invariantly distributed and sampling from this rotationally invariant distribution is equivalent to uniformly sampling a random $(n - m)$-dimensional subspace from the Grassmann manifold $\mathrm{Gr}_{(n-m)}(n)$, the set of $(n - m)$-dimensional subspaces in the n-dimensional Euclidean space R^n [Boo86]. For such a distribution on A and ideally sparse signals, the sharp bounds of [Don06c], for example, apply. In this chapter, as manifested by the statements of the three conditions (3.8.1), (3.8.2) and (3.8.3) and the previous discussions in Section 3.3, we can naturally extend the Grassmann angle approach to analyze the bounds for the probabilities that (3.8.1), (3.8.2) and (3.8.3) fail. Here we will denote these probabilities as P_1, P_2 and P_3 respectively. Note that there are $\binom{n}{k}$ possible support sets K and there are 2^k possible sign patterns for signal $\mathbf{x}_K$. From previous discussions, we know that the event that the condition (3.8.1) fails is the same for all $\mathbf{x}_K$'s of a specific support set and a specific sign pattern. Then following the same line of reasoning as in Section 3.3, we

have

$$P_1 = P_{K,-}, \tag{3.8.4}$$

$$P_2 \leq 2^k \times P_1, \tag{3.8.5}$$

$$P_3 \leq \binom{n}{k} \times 2^k \times P_1, \tag{3.8.6}$$

where $P_{K,-}$ is the probability as in (3.3.4).

We have the following lemma about the P_1, P_2 and P_3.

Lemma 3.8.2. *For any $C > 1$, we define $\zeta_W(\delta)$, $\zeta_{Sec}(\delta)$, and $\zeta_S(\delta)$ to be the largest fraction $\zeta = \frac{k}{n}$ such that the conditions (3.8.1), (3.8.2), and (3.8.3) are satisfied with overwhelming probability as $n \to 0$ if we sample the $(n-m)$-dimensional null-space uniformly, where $\frac{m}{n} = \delta$. Then*

$$\begin{aligned} \zeta_W(\delta) &> 0, \\ \zeta_{Sec}(\delta) &> 0, \\ \zeta_S(\delta) &> 0, \end{aligned}$$

for any $C > 1$ and $\delta > 0$. Also,

$$\lim_{\delta \to 1} \zeta_W(\delta) = 1,$$

for any $C > 1$.

The proof of this lemma is given in the appendix. The numerical results of ζ making sure that P_1, P_2, P_3 converge to zero overwhelmingly are presented in Section 3.10.

3.9 Analysis of ℓ_1 Minimization under Noisy Measurements

In the previous sections, we have analyzed the ℓ_1 minimization algorithm for decoding the approximately sparse signals. In this section, we will discuss the effect of noisy measurements on the ℓ_1 minimization for approximately sparse signals, using the null-space characterization.

Theorem 3.9.1. *Assume that an $m \times n$ measurement matrix A with rank m is given and denote its minimum non-zero singular value as σ_{min}. Further, assume that $\mathbf{y} = A\mathbf{x} + \mathbf{b}$, with its ℓ_1-norm $\|\mathbf{b}\| \leq \epsilon$, and that $\mathbf{w}$ is an $n \times 1$ vector. Let K be any subset of $\{1, 2, \ldots, n\}$ such that its cardinality $|K| = k$ and let K_i denote the i-th element of K. Further, let $\bar{K} = \{1, 2, \ldots, n\} \setminus K$. Then the solution $\hat{\mathbf{x}}$ produced by (3.1.2) will satisfy*

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq \frac{2(C+1)}{C-1}\|\mathbf{x}_{\bar{K}}\|_1 + \frac{(3C+1)\sqrt{n}\epsilon}{(C-1)\sigma_{min}},$$

with $C > 1$, if $\forall\ \mathbf{w} \in \boldsymbol{R}^n$ such that

$$A\mathbf{w} = 0\ ,$$

and $\forall\ K$ such that $|K| = k$, we have

$$C\sum_{i=1}^{k}|\mathbf{w}_{K_i}| \leq \sum_{i=1}^{n-k}|\mathbf{w}_{\bar{K}_i}|. \tag{3.9.1}$$

Proof. Since

$$\mathbf{y} = A\mathbf{x} + \mathbf{b},$$

we can write

$$\mathbf{y} = A\mathbf{x}^*,$$

where

$$\|\mathbf{x}^* - \mathbf{x}\| \leq \frac{\epsilon}{\sigma_{\min}}.$$

By the Cauchy-Schwarz inequality, we have

$$\|\mathbf{x}^* - \mathbf{x}\|_1 \leq \frac{\sqrt{n}\epsilon}{\sigma_{\min}}.$$

Suppose the matrix A has the claimed null-space property. Now the solution $\hat{\mathbf{x}}$ of (3.1.2) satisfies $\|\hat{\mathbf{x}}\|_1 \leq \|\mathbf{x}^*\|_1$. Since $A\hat{\mathbf{x}} = \mathbf{y}$, it easily follows that $\mathbf{w} = \hat{\mathbf{x}} - \mathbf{x}^*$ is in the null space of A. Therefore we can further write $\|\mathbf{x}^*\|_1 \geq \|\mathbf{x}^* + \mathbf{w}\|_1$. Using the triangular inequality for the ℓ_1 norm we obtain

$$\begin{aligned}
\|\mathbf{x}^*_K\|_1 + \|\mathbf{x}^*_{\bar{K}}\|_1 &= \|\mathbf{x}^*\|_1 \\
&\geq \|\hat{\mathbf{x}}\|_1 = \|\mathbf{x}^* + \mathbf{w}\|_1 \\
&\geq \|\mathbf{x}^*_K\|_1 - \|\mathbf{w}_K\|_1 + \|\mathbf{w}_{\bar{K}}\|_1 - \|\mathbf{x}^*_{\bar{K}}\|_1 \\
&\geq \|\mathbf{x}^*_K\|_1 + \frac{C-1}{C}\|\mathbf{w}_K\|_1 - \|\mathbf{x}^*_{\bar{K}}\|_1 \\
&\geq \|\mathbf{x}^*_K\|_1 - \|\mathbf{x}^*_{\bar{K}}\|_1 + \frac{C-1}{C+1}\|\mathbf{w}\|_1,
\end{aligned}$$

where the last two inequalities are from the claimed null-space property. Relating the first equality and the last inequality above, we have $2\|\mathbf{x}^*_{\bar{K}}\|_1 \geq \frac{(C-1)}{C+1}\|\mathbf{w}\|_1$.

Since

$$\|\mathbf{x}^*_{\bar{K}}\|_1 \leq \|\mathbf{x}_{\bar{K}}\|_1 + \|\mathbf{x}^* - \mathbf{x}\|_1,$$

we get

$$\begin{aligned}
\|\mathbf{w}\|_1 &\leq \frac{2(C+1)}{C-1}\|\mathbf{x}^*_{\bar{K}}\|_1 \\
&\leq \frac{2(C+1)}{C-1}\|\mathbf{x}_{\bar{K}}\|_1 + \frac{2(C+1)}{C-1}\|\mathbf{x}^* - \mathbf{x}\|_1.
\end{aligned}$$

From the triangular inequality,

$$\begin{aligned} \|\mathbf{x}-\hat{\mathbf{x}}\|_1 &\leq \|\mathbf{x}-\mathbf{x}^*\|_1 + \|\mathbf{w}\|_1 && (3.9.2)\\ &\leq \frac{2(C+1)}{C-1}\|\mathbf{x}_{\bar{K}}\|_1 + \frac{3C+1}{C-1}\|\mathbf{x}^*-\mathbf{x}\|_1, && (3.9.3)\\ &\leq \frac{2(C+1)}{C-1}\|\mathbf{x}_{\bar{K}}\|_1 + \frac{(3C+1)\sqrt{n}\epsilon}{(C-1)\sigma_{\min}}. && (3.9.4) \end{aligned}$$

■

If the elements in the measurement matrix A are i.i.d. as the unit real Gaussian random variable $N(0,1)$, following upon the work of Marchenko and Pastur [MP67], Geman [Gem80] and Silverstein [Sil85] proved that for $m/n = \delta$, as $n \to \infty$,

$$\frac{1}{\sqrt{n}}\sigma_{min} \to 1 - \sqrt{\delta}$$

almost surely.

Then almost surely as $n \to \infty$, $\frac{(3C+1)\sqrt{n}\epsilon}{(C-1)\sigma_{\min}} \to \frac{(3C+1)\epsilon}{(C-1)(1-\sqrt{\delta})}$. So in this case, we have $\|\mathbf{x}^*-\mathbf{x}\|_1$ is upperbounded by some constant times ϵ.

3.10 Numerical Computations on the Bounds of ζ

In this section, we will numerically evaluate the performance bounds on $\zeta = \frac{k}{n}$ such that the conditions (3.2.1), (3.8.1), (3.8.2) and (3.8.3) are satisfied with overwhelming probability as $n \to \infty$.

First, we know that the condition (3.2.1) fails with probability

$$P \leq \binom{n}{k} \times 2^k \times 2 \times \sum_{s\geq 0} \sum_{G \in \Im_{m+1+2s}(\mathrm{SP})} \beta(F,G)\gamma(G,\mathrm{SP}), \qquad (3.10.1)$$

Recall that we assume $\frac{m}{n} = \delta$, $l = (m+1+2s)+1$ and $\nu = \frac{l}{n}$. In order to make P overwhelmingly converge to zero as $n \to \infty$, following the discussions in Section 3.4, one sufficient condition is to make sure that the exponent for the combinatorial

factors,

$$\psi_{com} = \lim_{n\to\infty} \frac{\log\left(\binom{n}{k} 2^k 2 \binom{n-k}{l-k} 2^{l-k}\right)}{n}, \tag{3.10.2}$$

and the negative exponent for the angle factors,

$$\psi_{angle} = -\lim_{n\to\infty} \frac{\log\left(\beta(F,G)\gamma(G,\mathrm{SP})\right)}{n}, \tag{3.10.3}$$

satisfy $\psi_{com} - \psi_{angle} < 0$ uniformly over $\nu \in [\delta, 1)$.

Following [Don06c] we take $m = 0.5555n$. By analyzing the decaying exponents of the external angles and internal angles through the Laplace methods as in Sections 3.6 and 3.7, we can compute the numerical results as shown in Figures 3.1, 4.2 and 3.4. In Figure 3.1, we show the largest sparsity level $\zeta = \frac{k}{n}$ (as a function of C) which makes the probability of the condition (3.3.2) failing approach zero asymptotically as $n \to \infty$. As we can see, when $C = 1$, we get the same bound $\zeta = 0.095 \times 0.5555 = 0.0528$ as obtained for the ideally sparse signals case in [Don06c]. As expected, as C grows, the ℓ_1 minimization requires a smaller sparsity level ζ to achieves higher signal recovery accuracy.

In Figure 4.2, we show the exponents ψ_{com}, ψ_{int}, ψ_{ext} under the parameters $C = 2$, $\delta = 0.5555$ and $\zeta = 0.0265$. For the same set of parameters, in Figure 3.4, we compare the exponents ψ_{com} and ψ_{angle}: the solid curve denotes ψ_{angle} and the dashed curve denotes ψ_{com}. It shows that, under $\zeta = 0.0265$, $\psi_{com} - \psi_{angle} < 0$ uniformly over $\delta \le \mu \le 1$. Indeed, $\zeta = 0.0265$ is the bound shown in Figure 3.1 for $C = 2$. In Figure 3.5, for the parameter $\delta = 0.5555$, we give the bounds ζ as a function of C for satisfying the signal recovery robustness conditions (3.8.1), (3.8.2) and (3.8.3) respectively in the "weak", "sectional" and "strong" senses. In Figure 3.6, fixing $C = 2$, we plot how large $\rho = \zeta/\delta$ can be for different δ's for satisfying the signal recovery robustness conditions (3.8.1), (3.8.2) and (3.8.3) respectively in "weak", "sectional" and "strong" senses.

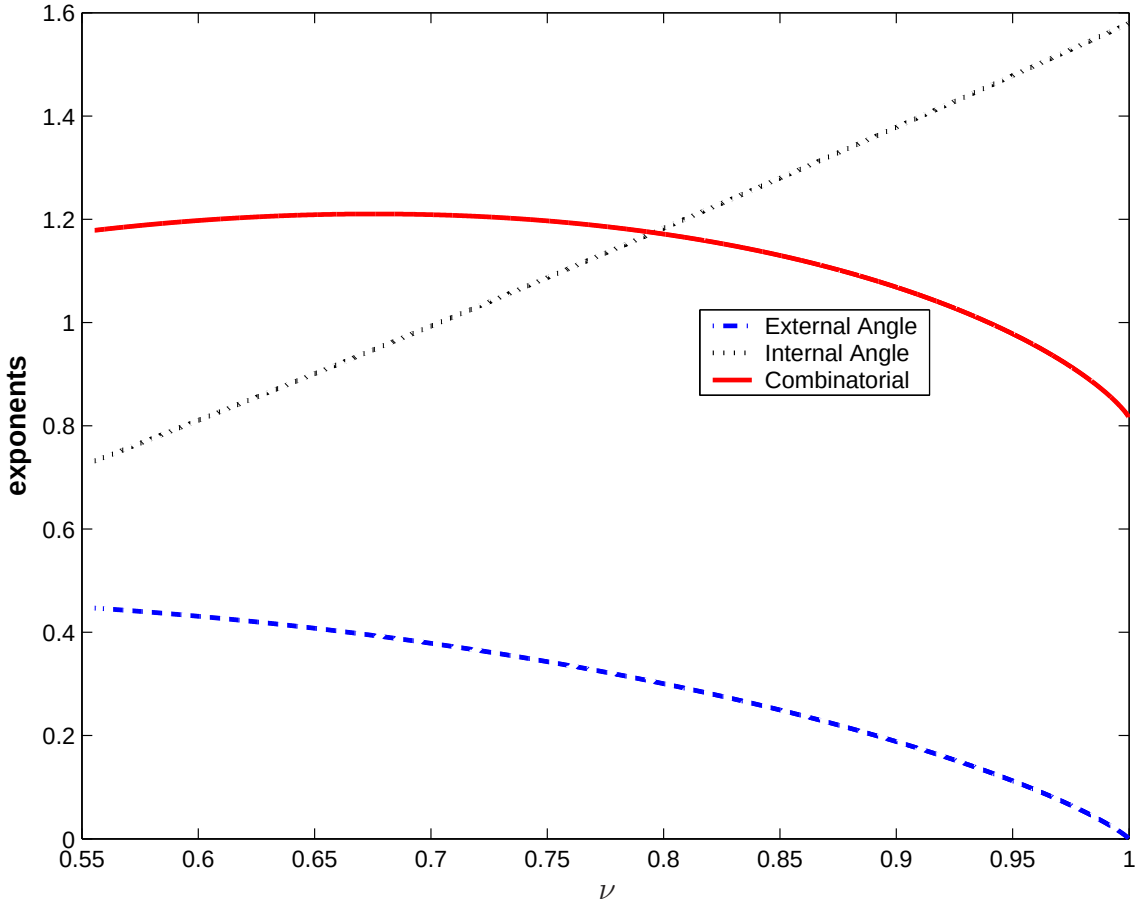


Figure 3.3: The combinatorial, internal and external angle exponents

3.11 Conclusion

It is well known that ℓ_1 optimization can be used to recover ideally sparse and approximately sparse signals in compressed sensing, if the underlying signal is sparse enough. While in the ideally sparse case the recent results of [Don06c] have given us sharp bounds on how sparse the signal can be, sharp bounds for the recovery of approximately sparse signals were not available.

In this chapter we developed and analyzed a null-space characterization of the necessary and sufficient conditions for the success of ℓ_1-norm optimization in compressed sensing of the approximately sparse signals. Using high-dimensional geometry, we give a unified *null-space Grassmann angle*-based analytical framework for compressive sensing. This new framework gives sharp quantitative tradeoffs between the

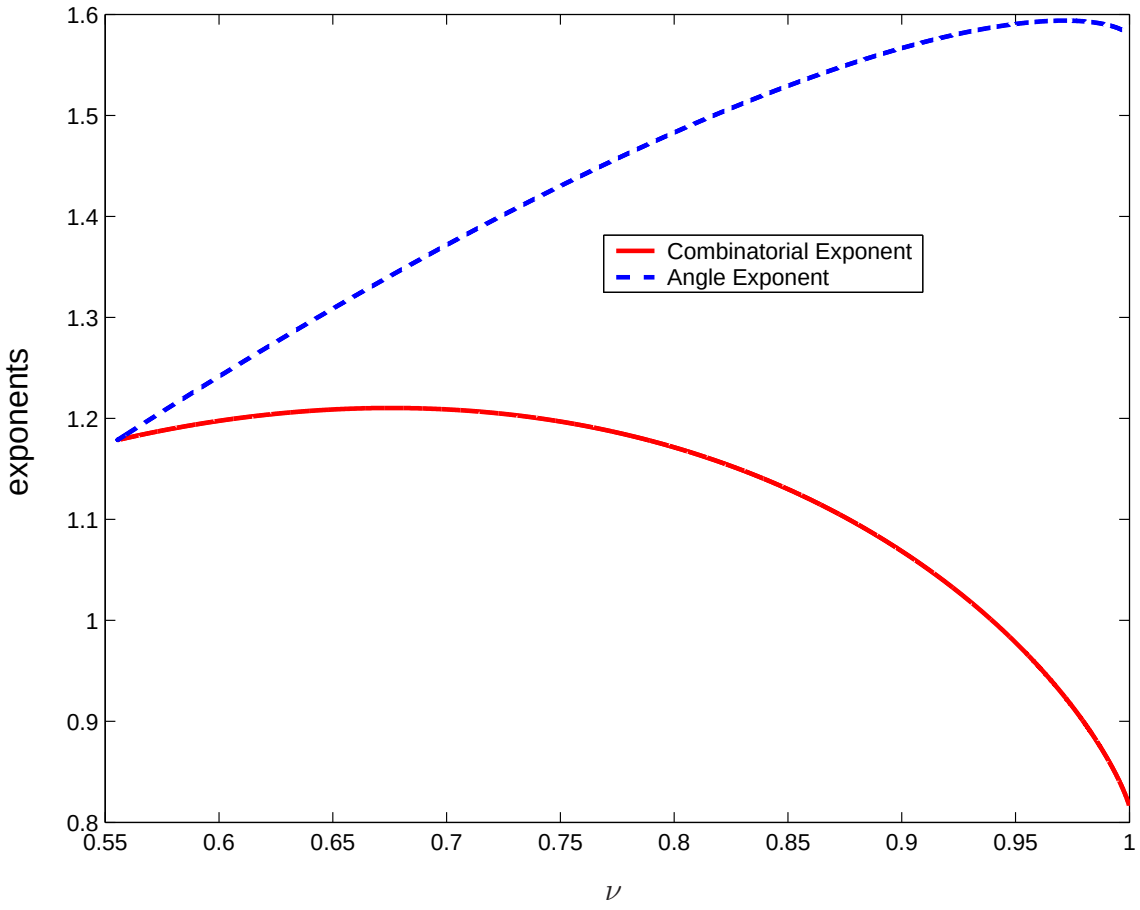


Figure 3.4: The combinatorial exponents and the angle exponents

signal sparsity and the recovery accuracy of the ℓ_1 optimization for approximately sparse signals. It can therefore be of practical use in applications where the underlying signal is not ideally sparse and where we are interested in the quality of the recovered signal. As expected, the neighborly polytopes result of [Don06c] for ideally sparse signals follows as a special case of our results. This work investigates the fundamental "balancedness" property of linear subspaces, and may be of independent mathematical interest. It is also possible to generalize these results to the analysis of compressive sensing noisy measurements in future work.

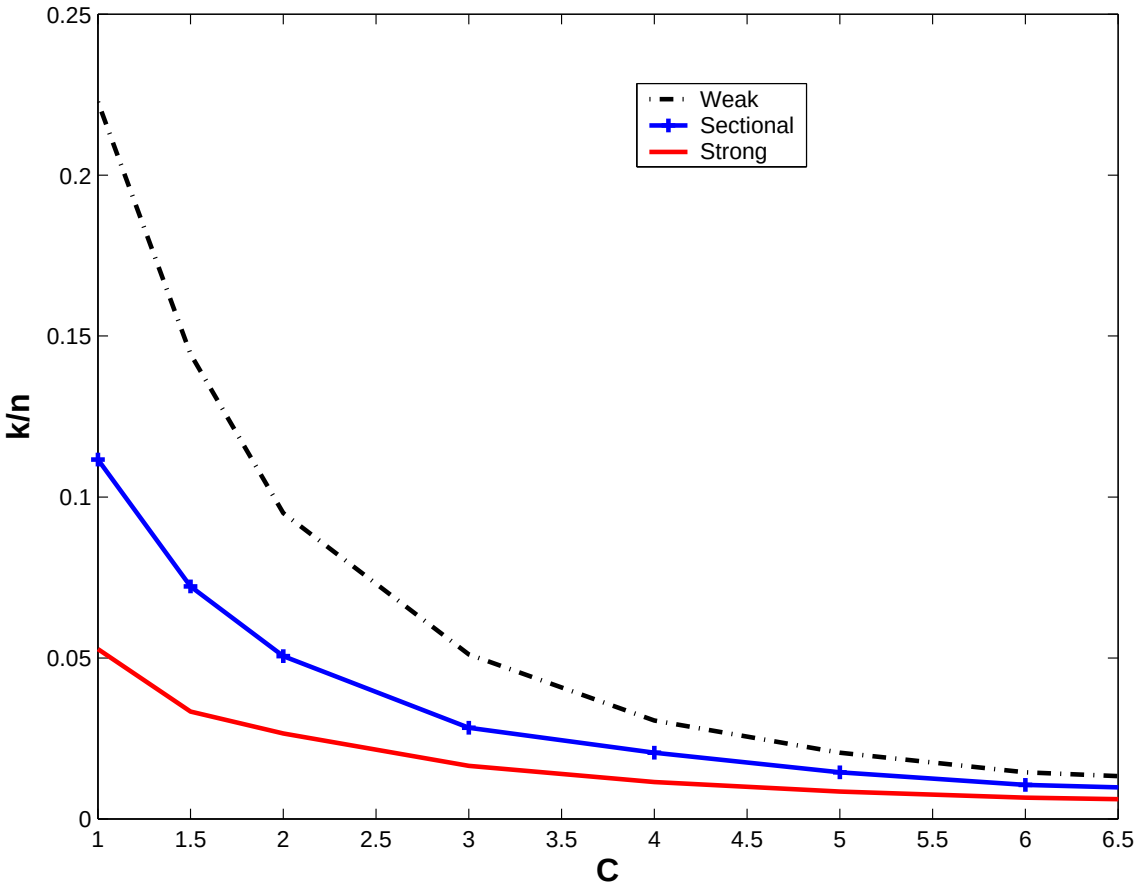


Figure 3.5: The weak, sectional and strong robustness bounds

3.12 Appendix

3.12.1 Derivation of the Internal Angles

There are two situations in the derivations of the internal angles for the skewed crosspolytope, which are respectively given in the following two lemmas.

Lemma 3.12.1. *Suppose that F is a $(k-1)$-dimensional face of the skewed crosspolytope*

$$SP = \{\mathbf{y} \in R^n \mid \|\mathbf{y}_K\|_1 + \|\frac{\mathbf{y}_{\bar{K}}}{C}\|_1 \leq 1\},$$

supported on the subset K with $|K| = k$. Then the internal angle $\beta(F, G)$ between the $(k-1)$-dimensional face F and a $(l-1)$-dimensional face G ($F \subseteq G$, $G \neq SP$) is

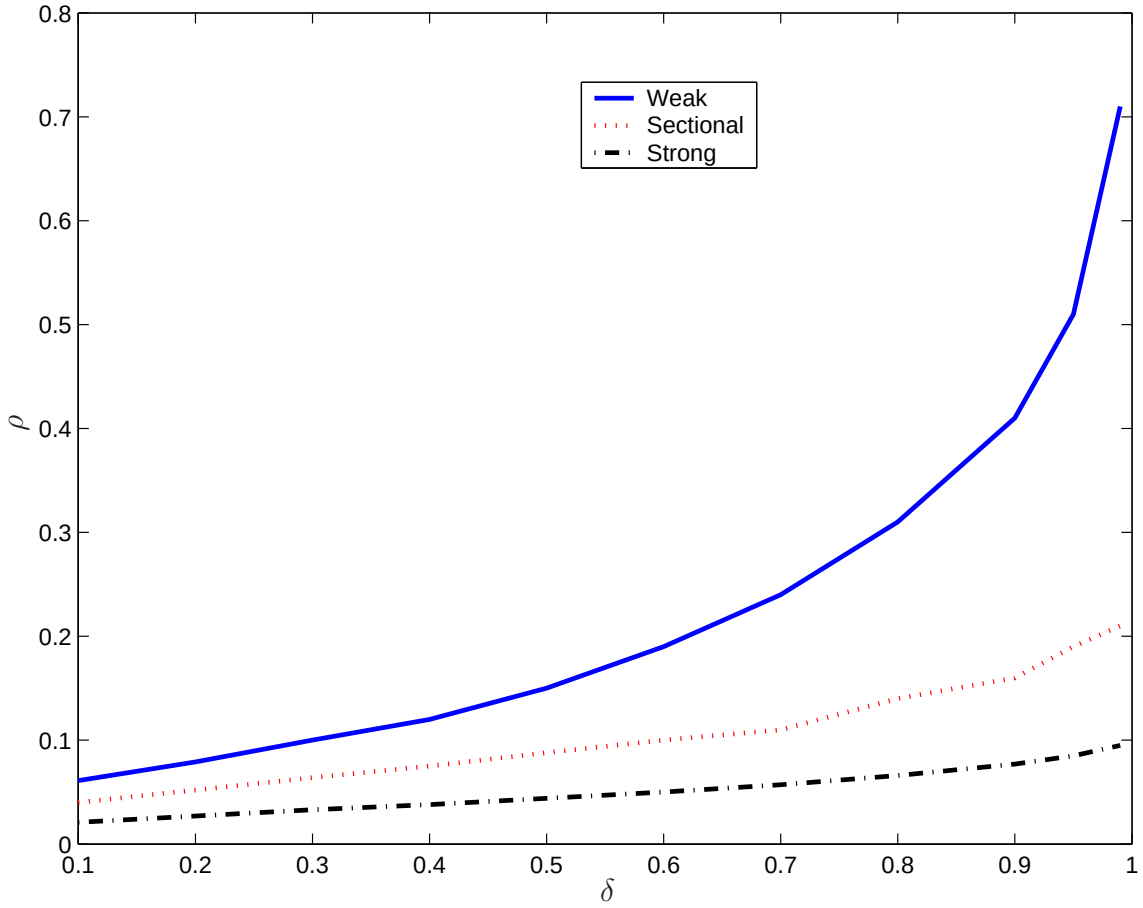


Figure 3.6: The weak, sectional and strong robustness bounds

given by

$$\beta(F,G) = \frac{V_{l-k-1}(\frac{1}{1+C^2k}, l-k-1)}{V_{l-k-1}(S^{l-k-1})}, \tag{3.12.1}$$

where $V_i(S^i)$ *denotes the* i*-th dimensional surface measure on the unit sphere* S^i*, while* $V_i(\alpha', i)$ *denotes the surface measure for regular spherical simplex with* $(i+1)$ *vertices on the unit sphere* S^i *and with inner product as* α' *between these* $(i+1)$ *vertices. (3.3.10) is equal to* $B(\frac{1}{1+C^2k}, l-k)$*, where*

$$B(\alpha', m') = \theta^{\frac{m'-1}{2}}\sqrt{(m'-1)\alpha'+1}\pi^{-m'/2}\alpha'^{-1/2}J(m', \theta), \tag{3.12.2}$$

with $\theta = (1-\alpha')/\alpha'$ *and*

$$J(m',\theta) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} (\int_{0}^{\infty} e^{-\theta v^2 + 2iv\lambda}\, dv)^{m'} e^{-\lambda^2}\, d\lambda. \tag{3.12.3}$$

Proof. Without loss of generality, assume that F is a $(k-1)$-dimensional face with k vertices as $e_p, 1 \le p \le k$, where e_p is the n-dimensional standard unit vector with the p-th element as '1'; and also assume that the $(l-1)$-dimensional face G be the convex hull of the l vertices: $e_p, 1 \le p \le k$ and $Ce_p, (k+1) \le p \le l$. Then the cone $\mathrm{Con}_{F,G}$ formed by observing the $(l-1)$-dimensional face G of the skewed crosspolytope SP from an interior point x^F of the face F is the positive cone of the vectors:

$$Ce_j - e_i, \quad \text{for all} \;\; j \in J \backslash K, \; i \in K, \tag{3.12.4}$$

and also the vectors

$$e_{i_1} - e_{i_2}, \quad \text{for all} \;\; i_1 \in K, \; i_2 \in K, \tag{3.12.5}$$

where $J = \{1, 2, ..., l\}$ is the support set for the face G.

So the cone $\mathrm{Con}_{F,G}$ is the direct sum of the linear hull $L_F = \mathrm{lin}\{F - x^F\}$ formed by the vectors in (3.12.5) and the cone $\mathrm{Con}_{F^\perp,G} = \mathrm{Con}_{F,G} \bigcap L_F^\perp$, where $L_F^\perp$ is the orthogonal complement to the linear subspace L_F. Then $\mathrm{Con}_{F^\perp,G}$ has the same spherical volume as $\mathrm{Con}_{F,G}$.

Now let us analyze the structure of $\mathrm{Con}_{F^\perp,G}$. We notice that the vector

$$e_0 = \sum_{p=1}^{k} e_p$$

is in the linear space $L_F^\perp$ and is also the only such a vector (up to linear scaling) supported on K. Thus a vector $\mathbf{x}$ in the positive cone $\mathrm{Con}_{F^\perp,G}$ must take the form

$$-\sum_{i=1}^{k} b_i \times e_i + \sum_{i=k+1}^{l} b_i \times e_i, \tag{3.12.6}$$

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

$$C\sum_{i=1}^{k} b_i = \sum_{i=k+1}^{l} b_i,$$
$$b_1 = b_2 = \cdots = b_k.$$

That is to say, the $(l-k)$-dimensional $\mathrm{Con}_{F^\perp,G}$ is the positive cone of $(l-k)$ vectors $a^1, a^2, ..., a^{l-k}$, where

$$a^i = C \times e_{k+i} - \sum_{p=1}^{k} e_p/k, \quad 1 \leq i \leq (l-k).$$

The normalized inner products between any two of these $(l-k)$ vectors is

$$\frac{< a^i, a^j >}{\|a^i\|\|a^j\|} = \frac{k \times \frac{1}{k^2}}{C^2 + k \times \frac{1}{k^2}} = \frac{1}{1+kC^2}.$$

(In fact, a^i's are also the vectors obtained by observing the vertices $e_{k+1}, \ldots, e_l$ from $Ec = \sum_{p=1}^{k} e_p/k$, the epicenter of the face F.)

We have so far reduced the computation of the internal angle to evaluating (3.12.1), the relative spherical volume of the cone $\mathrm{Con}_{F^\perp,G}$ with respect to the sphere surface S^{l-k-1}. This was computed as given in this lemma [VS92, KBH99] for the positive cones of vectors with equal inner products by using a transformation of variables and 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.

Instead, in this chapter, we will give a proof of (3.12.2) which can directly lead to the probabilistic large deviation method of evaluating the internal angle exponent in [Don06c].

First, we notice that $\mathrm{Con}_{F^\perp,G}$ is a $(l-k)$-dimensional cone. Also, all the vectors

$(x_1, \ldots, x_n)$ in the cone $\mathrm{Con}_{F^\perp,G}$ take the form in (3.12.6). From [Had79],

$$\int_{\mathrm{Con}_{F^\perp,G}} e^{-\|x\|^2}\, dx = \beta(F,G) V_{l-k-1}(S^{l-k-1}) \times \int_0^\infty e^{-r^2} r^{l-k-1}\, dx = \beta(F,G) \cdot \pi^{(l-k)/2}, \tag{3.12.7}$$

where $V_{l-k-1}(S^{l-k-1})$ is the spherical volume of the $(l-k-1)$-dimensional sphere S^{l-k-1}. Now define $U \subseteq R^{l-k+1}$ as the set of all nonnegative vectors satisfying:

$$x_p \geq 0,\ 1 \leq p \leq l-k+1,\ \sum_{p=2}^{l-k+1} x_p = Ckx_1$$

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

$$f(x_1,\ \ldots,\ x_{l-k+1}) \ = \ -\sum_{p=1}^{k} x_1 e_p + \sum_{p=k+1}^{l} x_{p-k} \times e_p.$$

Then

$$\begin{aligned} &\int_{\mathrm{Con}_{F^\perp,G}} e^{-\|x'\|^2}\, dx' \\ &= \sqrt{\frac{l+(C^2-1)k}{C^2 k}} \int_{\sum_{p=2}^{l-k+1} x_p = Ckx_1, x_p \geq 0,\ 2 \leq p \leq l-k+1} \\ &\qquad e^{-\|f(x)\|^2}\, dx_2 \cdots dx_{l-k+1} \\ &= \sqrt{\frac{l+(C^2-1)k}{C^2 k}} \int_{\sum_{p=2}^{l-k+1} x_p = Ckx_1, x_p \geq 0,\ 2 \leq p \leq l-k+1} \\ &\quad e^{-kx_1^2 - x_2^2 - \cdots - x_{l-k+1}^2}\, dx_2 \cdots dx_{l-k+1}, \end{aligned} \tag{3.12.8}$$

where $\sqrt{\frac{l+(C^2-1)k}{C^2 k}}$ is due to the change of integral variables.

Now we define a random variable,

$$Z = X_2 + X_3 + \cdots + X_{l-k+1} - CkX_1,$$

where $X_1, X_2, \cdots, X_n$ are independent random variables, with $X_p \sim HN(0, \frac{1}{2})$, $2 \leq$

$p \leq (n-k+1)$, as half-normal distributed random variables and $X_1 \sim N(0, \frac{1}{2k})$ as a normal distributed random variable. Then by inspection, (3.12.8) is equal to

$$\frac{\sqrt{\pi}^{l-k+1}}{2^{l-k}} \times \sqrt{l + (C^2-1)k} p_Z(0),$$

where $p_Z(\cdot)$ is the probability density function for the random variable $Z = X_2 + X_3 + \cdots + X_{l-k+1} - CKX_1$ and $p_Z(0)$ is the probability density function $p_Z(\cdot)$ evaluated at the point $Z = 0$.

Use the notation

$$G_X(\lambda) = \int_{-\infty}^{\infty} e^{i\lambda x} p_X(x)\, dx$$

as the characteristic function for any random variable X, where $p_X(x)$ is the probability density function of X. Then from the independence of $X_1, X_2, \ldots, X_{l-k+1}$, the characteristic function for Z is equal to

$$G_Z(\lambda) = G_{X_2}(\lambda)^{l-k} \times G_{X_1}(\lambda).$$

Expressing the probability density function $p_Z(x)$ in the Fourier domain, we have

$$p_Z(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G_Z(\lambda).$$

Combining this with (3.12.7) gives us the desired result. ∎

Now we have only considered the internal angle $\beta(F, G)$ when G is not the whole skewed crosspolytope. The following lemma discusses the special case when $G = \mathrm{SP}$.

Lemma 3.12.2. *Suppose F, K and SP are defined in the same way as in the statement of Lemma 3.12.1. Then the internal angle $\beta(F, SP)$ between the $(k-1)$-dimensional face F and the n-dimensional skewed crosspolytope SP is given by*

$$\int_{-\infty}^{0}\int_{-\infty}^{\infty} e^{-\frac{C^2k\lambda^2}{4}}\left(\int_0^\infty e^{-\mu^2+i\mu\lambda}\,d\mu\right)^{n-k} \frac{2^{n-k-1}}{\sqrt{\pi}^{n-k+2}} e^{-i\lambda z}\,d\lambda\,dz.$$

Proof. We use the same set of notations as in the proof of Lemma 3.12.1. Without loss of generality, assume $K=\{1,\ldots,k\}$, F is the $(k-1)$-dimensional face supported on K and $G=\mathrm{SP}$. So the cone $\mathrm{Con}_{F,G}$ is the direct sum of the linear hull $L_F=\mathrm{lin}\{F-x^F\}$ formed by the vectors in (3.12.5) and the cone $\mathrm{Con}_{F^\perp,G}=\mathrm{Con}_{F,G}\bigcap L_F^\perp$, where $L_F^\perp$ is the orthogonal complement to the linear subspace L_F. Then $\mathrm{Con}_{F^\perp,G}$ has the same spherical volume as $\mathrm{Con}_{F,G}$.

Following similar analysis in Lemma 3.12.1, the cone $\mathrm{Con}_{F^\perp,G}$ is the positive cone of $2(n-k)$ vectors $a_\pm^1, a_\pm^2, ..., a_\pm^{n-k}$, where

$$a_\pm^i = \pm C\times e_{k+i} - \sum_{p=1}^{k} e_p/k,\quad 1\le i\le (n-k).$$

This also means that $\mathrm{Con}_{F^\perp,G}$ is a $(n-k+1)$-dimensional cone. Also, all the vectors $(x_1,\ldots,x_n)$ in the cone $\mathrm{Con}_{F^\perp,G}$ take the form

$$x_1=x_2=\cdots=x_k\le 0,$$
$$\sum_{p=k+1}^{n}|x_p|\le Ck|x_1|.$$

From [Had79],

$$\int_{\mathrm{Con}_{F^\perp,G}} e^{-\|x\|^2}\,dx = \beta(F,G)V_{n-k}(S^{n-k}) \times\int_0^\infty e^{-r^2}r^{n-k}\,dx = \beta(F,G)\cdot\pi^{(n-k+1)/2}, \tag{3.12.9}$$

where $V_{n-k}(S^{n-k})$ is the spherical volume of the $(n-k)$-dimensional sphere S^{n-k}. Now define $U \subseteq R^{n-k+1}$ as the set of all the vectors taking the form:

$$\{x_1 \geq 0, \sum_{p=2}^{n-k+1} |x_p| \leq Ckx_1\},$$

and define $f(x_1, \ \ldots, \ x_{n-k+1}) : U \to \mathrm{Con}_{F^\perp,G}$ to be the linear and bijective map,

$$f(x_1, \ \ldots, \ x_{n-k+1}) \ = \ -\sum_{p=1}^{k} x_1 e^p + \sum_{p=k+1}^{n} x_{p-k} \times e^p.$$

Then

$$\begin{aligned} &\int_{\mathrm{Con}_{F^\perp,G}} e^{-\|x'\|^2}\, dx' = \sqrt{k} \int_U e^{-\|f(x)\|^2}\, dx \\ &= \sqrt{k} \int_0^\infty \int_{\sum_{p=2}^{n-k+1} |x_p| \leq Ckx_1} \\ &\quad e^{-kx_1^2 - x_2^2 - \cdots - x_{n-k+1}^2}\, dx_2 \cdots dx_{n-k+1}\, dx_1, \end{aligned} \tag{3.12.10}$$

where $\sqrt{k}$ is due to the change of integral variables.

By inspection, (3.12.10) is equal to

$$\sqrt{\pi}^{n-k+1} P(X_2 + X_3 + \cdots + X_{n-k+1} - CkX_1 \leq 0),$$

where $X_1, X_2, \ldots, X_n$ are independent random variables, with $X_p \sim HN(0, \frac{1}{2})$, $2 \leq p \leq (n-k+1)$, as half-normal distributed random variables and $X_1 \sim N(0, \frac{1}{2k})$ as a normal distributed random variable.

Expressing the probability density function of $Z = X_2 + X_3 + \cdots + X_{n-k+1} - CKX_1$ in the Fourier domain, we can simplify (3.12.10) to

$$\int_{-\infty}^{0} \int_{-\infty}^{\infty} e^{-\frac{C^2 k \lambda^2}{4}} \left(\int_0^\infty e^{-\mu^2 + i\mu\lambda}\, d\mu \right)^{n-k} \frac{2^{n-k-1}}{\sqrt{\pi}} e^{-i\lambda z}\, d\lambda\, dz$$

Combining this with (3.12.9) gives us the desired result. ∎

3.12.2 Derivation of the External Angles

Lemma 3.12.3. *Suppose that F is a $(k-1)$-dimensional face of the skewed crosspolytope*

$$SP = \{\mathbf{y} \in R^n \mid \|\mathbf{y}_K\|_1 + \|\frac{\mathbf{y}_{\bar{K}}}{C}\|_1 \leq 1\}$$

supported on a subset K with $|K| = k$. Then the external angle $\gamma(G, SP)$ between a $(l-1)$-dimensional face G ($F \subseteq G$) and the skewed crosspolytope SP is given by

$$\gamma(G, SP) = \frac{2^{n-l}}{\sqrt{\pi}^{n-l+1}} \int_0^\infty e^{-x^2} \left(\int_0^{\frac{x}{C\sqrt{k+\frac{l-k}{C^2}}}} e^{-y^2}\, dy\right)^{n-l} dx. \tag{3.12.11}$$

Proof. Without loss of generality, assume $K = \{n-k+1, \cdots, n\}$. Consider the $(l-1)$-dimensional face

$$G = \text{conv}\{C \times e^{n-l+1}, ..., C \times e^{n-k}, e^{n-k+1}, ..., e^n\}$$

of the skewed crosspolytope SP. The 2^{n-l} outward normal vectors of the supporting hyperplanes of the facets containing G are given by

$$\{\sum_{p=1}^{n-l} j_p e^p/C + \sum_{p=n-l+1}^{n-k} e^p/C + \sum_{p=n-k+1}^{n} e^p, j_p \in \{-1, 1\}\}.$$

Then the outward normal cone $c(G, \text{SP})$ at the face G is the positive hull of these normal vectors. Thus

$$\begin{aligned} &\int_{c(G,\text{SP})} e^{-\|x\|^2}\, dx = \gamma(G, SP) V_{n-l}(S^{n-l}) \\ &\times \int_0^\infty e^{-r^2} r^{n-l}\, dx = \gamma(G, \text{SP}).\pi^{(n-l+1)/2}, \end{aligned} \tag{3.12.12}$$

where $V_{n-l}(S^{n-l})$ is the spherical volume of the $(n-l)$-dimensional sphere S^{n-l}. Now

define U to be the set

$$\{x \in R^{n-l+1} \mid x_{n-l+1} \geq 0, |x_p| \leq \frac{x_{n-l+1}}{C}, 1 \leq p \leq (n-l)\}$$

and define $f(x_1,\ \cdots,\ x_{n-l+1}) : U \to c(G, \mathrm{SP})$ to be the linear and bijective map

$$\begin{aligned} f(x_1,\ \cdots,\ x_{n-l+1}) &= \sum_{p=1}^{n-l} x_p e^p + \sum_{p=n-l+1}^{n-k} \frac{x_{n-l+1}}{C} e^p \\ &\quad + \sum_{p=n-k+1}^{n} x_{n-l+1} \times e^p. \end{aligned}$$

Then

$$\begin{aligned} &\int_{c(G,\mathrm{SP})} e^{-\|x'\|^2}\, dx' \\ &= \sqrt{k + \frac{l-k}{C^2}} \int_U e^{-\|f(x)\|^2}\, dx \\ &= \sqrt{k + \frac{l-k}{C^2}} \int_0^\infty \int_{-\frac{x_{n-l+1}}{C}}^{\frac{x_{n-l+1}}{C}} \cdots \int_{-\frac{x_{n-l+1}}{C}}^{\frac{x_{n-l+1}}{C}} \\ &e^{-x_1^2 - \cdots - x_{n-l}^2 - (k + \frac{l-k}{C^2}) x_{n-l+1}^2}\, dx_1 \cdots dx_{n-l+1} \\ &= \sqrt{k + \frac{l-k}{C^2}} \int_0^\infty e^{-(k+\frac{l-k}{C^2})x^2} \\ &\qquad \times \left(\int_{-\frac{x}{C}}^{\frac{x}{C}} e^{-y^2}\, dy \right)^{n-l} dx \\ &= 2^{n-l} \int_0^\infty e^{-x^2} \left(\int_0^{\frac{x}{C\sqrt{k+\frac{l-k}{C^2}}}} e^{-y^2}\, dy \right)^{n-l} dx, \end{aligned}$$

where $\sqrt{k + \frac{l-k}{C^2}}$ is due to the change of integral variables. Combining it with (3.12.12) leads to the desired result. ■

3.12.3 Proof of Lemma 3.5.1

Consider any fixed $\delta > 0$. First, we consider the internal angle exponent Ψ_{int}, where we define $\gamma' = \frac{\rho\delta}{\frac{C^2-1}{C^2}\rho\delta + \frac{\nu}{C^2}}$. Then for this fixed δ,

$$\frac{1-\gamma'}{\gamma'} \geq \frac{\frac{C^2-1}{C^2}\rho\delta + \frac{\delta}{C^2}}{\rho\delta} - 1,$$

uniformly over $\nu \in [\delta, 1]$.

Now if we take ρ small enough, $\frac{\frac{C^2-1}{C^2}\rho\delta + \frac{\delta}{C^2}}{\rho\delta}$ can be arbitrarily large. By the asymptotic expression (3.5.7), this leads to large enough internal decay exponent Ψ_{int}. At the same time, the external angle exponent Ψ_{ext} is lowerbounded by zero and the combinatorial exponent is upperbounded by some finite number. Then if ρ is small enough, we will get the net exponent Ψ_{net} to be negative uniformly over the range $\nu \in [\delta, 1]$.

3.12.4 Proof of Lemma 3.5.2

We will show that for fixed $C > 1$, with $\rho(\delta) = \log(1/\delta)^{-(1+\eta)}$ and some $\delta_0 > 0$,

$$\Psi_{net}(\nu; \rho(\delta), \delta) < -\delta, \qquad \delta < \delta_0, \qquad \nu \in [\delta, 1).$$

To this end, we need to get the asymptotic of $\Psi_{int}(\nu)$, $\Psi_{ext}(\nu)$ and $\Psi_{ext}(\nu)$ as $\delta \to 0$ and $\rho(\delta) = \log(1/\delta)^{-(1+\eta)}$.

With

$$H(\nu) + H(\rho\delta/\nu)\nu = H(\rho\delta) + H(\frac{\nu - \rho\delta}{1-\rho\delta})(1-\rho\delta),$$

from its definition, $\Psi_{net}(\nu; \rho, \delta)$ is equal to

$$H(\nu) - \Psi_{ext}(\nu) - \xi_{\gamma'}(y_{\gamma'})(\nu - \rho\delta) + \rho\delta\log(2) + H(\rho\delta/\nu)\nu.$$

From the derivation (or the expression) of the external angle $\gamma(G, \mathrm{SP})$ in this chapter, $\gamma(G, \mathrm{SP})$ is a decreasing function in C. So we can upperbound $\gamma(G, \mathrm{SP})$

uniformly in $\nu \in [\delta, 1]$, for any $C \geq 1$, by

$$\frac{2^{n-l}}{\sqrt{\pi}^{n-l+1}} \int_0^\infty e^{-x^2} (\int_0^{\frac{x}{\sqrt{l}}} e^{-y^2}\, dy)^{n-l}\, dx,$$

namely the expression for the external angle when $C = 1$.

Now define $\Omega(\nu) = H(\nu) - \Psi_{ext}^{C=1}(\nu)$, where $\Psi_{ext}^{C=1}(\nu)$ is the external exponent when $C = 1$. Then from the asymptotic formula (3.5.3), we have

$$H(\nu) - \Psi_{ext}(\nu) \leq \Omega(\nu) \sim \frac{1}{2} \log(\log(\frac{1}{\nu}))\nu,$$

as $\nu \to 0$.

So $\Psi_{net}(\nu; \rho, \delta)$ is no bigger than

$$\Omega(v) - \xi_{\gamma'}(y_{\gamma'})(\nu - \rho\delta) + \rho\delta \log(2) + H(\rho\delta/\nu)\nu.$$

From [Don06c], for a certain δ_1, if $\delta < \delta_1$,

$$H(\rho\delta/\nu) \leq H(\rho)\delta + 2\rho(\nu - \delta),$$

so we have $\Psi_{net}(v)$

$$\Psi_{net}(v) \leq K(\nu; \rho, \delta) + [\rho\delta \log(2) + H(\rho)\delta] + 2\rho(\nu - \delta),$$

where $K(\nu; \rho, \delta) \doteq \Omega(v) - \xi_{\gamma'}(y_{\gamma'})(\nu - \rho\delta)$.

As we will show later in Lemma 3.12.4, $K(v; \rho, \delta)$ is a concave function in $\nu \in [\delta, 1]$ if $\delta < \delta_2$. Also, we will show that for $\delta < \delta_3$,

$$K'(\delta; \rho, \delta) \leq -\eta/4 \log(\log(\frac{1}{\delta})), \tag{3.12.13}$$

$$K(\delta; \rho, \delta) \leq -\delta \times \eta/4 \log(\log(\frac{1}{\delta})), \tag{3.12.14}$$

where $K'(\nu;\rho,\delta) \doteq \frac{\partial K(\nu;\rho,\delta)}{\partial \nu}$. So there exists a $\delta_4 > 0$ so that for any $0 < \delta < \delta_4$,

$$K'(\delta;\rho,\delta) < -2\rho.$$

Also there exists a $\delta_5 > 0$ so that for any $0 < \delta < \delta_5$,

$$K(\delta;\rho,\delta) + \rho\delta\log(2) + H(\rho)\delta < -\delta.$$

Then by the concavity of $K(\nu;\rho,\delta)$, if $\delta < \min(\delta_1,\delta_2,\delta_3,\delta_4,\delta_5)$,

$$\Psi_{net}(\nu) \leq -\delta,$$

uniformly over the interval $\nu \in [\delta, 1]$.

Now we need to prove (3.12.13) and (3.12.14). As computed in [Don06c],

$$\Omega'(v) \sim \frac{1}{2}\log(\log(\frac{1}{\delta})), \qquad \nu \to 0.$$

By (3.5.7), we know that as $\delta \to 0$, and with $\nu = \delta$,

$$\xi_{\gamma'}(y_{\gamma'}) \sim \frac{1}{2}\log(\frac{1}{C^2\rho}) \sim \frac{1}{2}\log(\log(\frac{1}{\delta}))(1+\eta).$$

Hence for $\delta < \delta_6$,

$$\xi_{\gamma'}(y_{\gamma'})(\nu - \rho\delta) \geq (1 + \frac{\eta}{2}\Omega(\delta)). \tag{3.12.15}$$

Following this, there exists a $\delta_7 > 0$ so that for $\delta < \delta_7$,

$$K(\delta;\rho,\delta) \leq \frac{\eta}{4}\log\log(\frac{1}{\delta}).$$

Also, from the asymptotic of $\Omega'(\nu)$ and the asymptotic of the derivative of $\xi_{\gamma'}(y_{\gamma'})(\nu - \rho\delta)$ in the following Lemma 3.12.4, we can further have

$$K'(\delta;\rho,\delta) \leq -\eta/4\log(\log(\frac{1}{\delta})).$$

Lemma 3.12.4. *If ρ and δ are small enough, $K(\nu;\rho,\delta)$ is concave as a function of ν.*

Proof. We define

$$\Upsilon(\nu;\rho,\delta) = \xi_{\gamma'}(y_{\gamma'})(\nu - \rho\delta).$$

Since $K(\nu;\rho,\delta) = \Omega(\nu) - \Upsilon(\nu;\rho,\delta)$ and $\Omega(\nu)$ is a concave function in ν, we only need to show that $\Upsilon(\nu;\rho,\delta)$ is a convex function in ν.

Recall that $\gamma' = \frac{\rho\delta}{\frac{C^2-1}{C^2}\rho\delta+\frac{\nu}{C^2}}$ and we first look at $\frac{\partial\gamma'}{\partial\nu}$:

$$\frac{\partial\gamma'}{\partial\nu} = -\frac{\rho\delta}{\frac{C^2-1}{C^2}\rho\delta + \frac{\nu}{C^2}} \times \frac{1}{(C^2-1)\rho\delta + \nu}.$$

So

$$\begin{aligned}
\frac{\partial\Upsilon(\nu;\rho,\delta)}{\partial\nu} &= \xi_{\gamma'}(y_{\gamma'}) + \frac{\partial\xi_{\gamma'}(y_{\gamma'})}{\partial\gamma'} \cdot \frac{\partial\gamma'}{\partial\nu} \cdot (\nu - \rho\delta) \\
&= \xi_{\gamma'}(y_{\gamma'}) - \frac{\partial\xi_{\gamma'}(y_{\gamma'})}{\partial\gamma'} \cdot \gamma' \cdot \frac{1-\frac{\rho\delta}{\nu}}{(C^2-1)\frac{\rho\delta}{\nu}+1}.
\end{aligned}$$

If we define $\Xi = \frac{1-\frac{\rho\delta}{\nu}}{(C^2-1)\frac{\rho\delta}{\nu}+1}$ and $\Pi = \frac{1}{(C^2-1)\rho\delta+\nu}$, we can have

$$\begin{aligned}
&\frac{\partial^2\Upsilon(\nu;\rho,\delta)}{\partial\nu^2} \\
= \;& \frac{\partial\xi_{\gamma'}(y_{\gamma'})}{\partial\gamma'}\frac{\partial\gamma'}{\partial\nu} - \frac{\partial^2\xi_{\gamma'}(y_{\gamma'})}{\partial\gamma'^2} \cdot \frac{\partial\gamma'}{\partial\nu} \cdot \gamma' \cdot \Xi \\
& - \frac{\partial\xi_{\gamma'}(y_{\gamma'})}{\partial\gamma'} \cdot \frac{\partial\gamma'}{\partial\nu} \cdot \Xi - \frac{\partial\xi_{\gamma'}(y_{\gamma'})}{\partial\gamma'} \cdot \frac{\partial\Xi}{\partial\nu} \cdot \gamma' \\
= \;& \frac{\partial\xi_{\gamma'}(y_{\gamma'})}{\partial\gamma'} \cdot (-\gamma') \cdot (1-\Xi) \cdot \Pi \\
& - \frac{\partial\xi_{\gamma'}(y_{\gamma'})}{\partial\gamma'} \cdot \gamma' \cdot \frac{\partial\Xi}{\partial\nu} + \frac{\partial^2\xi_{\gamma'}(y_{\gamma'})}{\partial\gamma'^2} \cdot \gamma'^2 \cdot \Pi \\
= \;& -\frac{\partial\xi_{\gamma'}(y_{\gamma'})}{\partial\gamma'} \cdot \gamma'^2 \cdot \Pi - \frac{\partial\xi_{\gamma'}(y_{\gamma'})}{\partial\gamma'} \cdot \gamma'^2 \cdot \Pi \\
& + \frac{\partial^2\xi_{\gamma'}(y_{\gamma'})}{\partial\gamma'^2} \cdot \gamma'^2 \cdot \Xi \cdot \Pi.
\end{aligned}$$

It has been shown in [Don06c] that as $\gamma \to 0$,

$$\frac{\partial \xi_\gamma y_\gamma}{\partial \gamma} \sim -\frac{\gamma^{-1}}{2},$$
$$\frac{\partial^2 \xi_\gamma y_\gamma}{\partial \gamma^2} \sim \frac{\gamma^{-4}}{4}.$$

So from the definition of γ', there exists a small enough ρ_0 such that for any $\rho < \rho_0$,

$$\frac{\partial^2 \Upsilon(\nu;\rho,\delta)}{\partial \nu^2} > 0, \nu \in [\delta, 1], \tag{3.12.16}$$

which then implies the concavity of $K(\nu;\rho,\delta)$. ■

3.12.5 Proof of Lemma 3.5.3

Proof. Suppose instead that $\rho_N(\delta;C) > \frac{1}{C+1}$. Then for every vector $\mathbf{w}$ from the null-space of the measurement matrix A, any $\rho_{N;C}$ fraction of the n components in $\mathbf{w}$ take no more than $\frac{1}{C+1}$ fraction of $\|\mathbf{w}\|_1$. But this can not be true if we consider the $\rho_N(\delta;C)$ fraction of $\mathbf{w}$ with largest magnitudes.

Now we only need to prove the lower bound for $\rho_N(\delta;C)$. we argue that

$$\rho_N(\delta;C) \geq \frac{\rho_N(\delta;C=1)}{C^2}.$$

We know from Lemma 3.5.1 that $\rho_N(\delta;C) > 0$ for any $C \geq 1$. Denote $\psi_{net}(C)$, $\psi_{com}(\nu;\rho,\delta,C)$, $\psi_{int}(\nu;\rho,\delta,C)$ and $\psi_{ext}(\nu;\rho,\delta,C)$ as the respective exponents for a certain C. Because $\rho_N(\delta;C=1) > 0$, the net exponent $\psi_{net}(C=1)$ is negative uniformly over $\nu \in [\delta,1]$. By examining the formula 3.3.13 for the external angle $\gamma(G,\mathrm{SP})$, where G is a $(l-1)$-dimensional face of the skewed crosspolytope SP, we have $\gamma(G,\mathrm{SP})$ is a decreasing function in both k and C for a fixed l. So $\gamma(G,\mathrm{SP})$ is upperbounded by

$$\frac{2^{n-l}}{\sqrt{\pi}^{n-l+1}} \int_0^\infty e^{-x^2} \left(\int_0^{\frac{x}{\sqrt{l}}} e^{-y^2}\, dy\right)^{n-l} dx, \tag{3.12.17}$$

namely the expression for the external angle when $C = 1$. Then for any $C > 1$ and any k, $\psi_{ext}(\nu; \rho, \delta, C)$ is lowerbounded by $\psi_{ext}(\nu; \rho, \delta, C = 1)$.

Now let us check $\psi_{int}(\nu; \rho, \delta, C)$ by using the formula 3.5.5. With

$$\gamma' = \frac{\rho\delta}{\frac{C^2-1}{C^2}\rho\delta + \frac{\nu}{C^2}},$$

we have

$$\frac{1-\gamma'}{\gamma'} = -\frac{1}{C^2} + \frac{\nu}{C^2\rho\delta}. \tag{3.12.18}$$

Then for fixed δ and ν, if we take $\rho = \frac{\rho_N(\delta;C=1)}{C^2}$, then $\frac{1-\gamma'}{\gamma'}$ is an increasing function in C. So from its definition $\xi_{\gamma'}(y_{\gamma'})$ is an increasing function in C. This further implies that $\psi_{int}(\nu; \rho, \delta)$ is an increasing function in C if we take $\rho = \frac{\rho_N(\delta;C=1)}{C^2}$.

Also, for fixed ν and δ, it is not hard to show that $\psi_{com}(\nu; \rho, \delta, C)$ is a decreasing function in C if $\rho = \frac{\rho_N(\delta;C=1)}{C^2}$.

Thus for any $C > 1$, if $\rho = \frac{\rho_N(\delta;C=1)}{C^2}$, the net exponent $\psi_{net}(C)$ is also negative uniformly over $\nu \in [\delta, 1]$. Lemma 3.5.3 then follows. ∎

3.12.6 Proof of Lemma 3.8.2

Proof. First, we notice that for any $C > 1$,

$$\zeta_W(\delta) \geq \rho_N\delta,$$
$$\zeta_{Sec}(\delta) \geq \rho_N\delta,$$
$$\zeta_S(\delta) = \rho_N\delta,$$

so by Lemma 3.5.1,

$$\zeta_W(\delta) > 0,$$
$$\zeta_{Sec}(\delta) > 0,$$
$$\zeta_S(\delta) > 0.$$

Now we will prove

$$\lim_{\delta \to 1} \zeta_W(\delta) = \delta.$$

As discussed in previous sections, we know that the decay exponent for the probability that the condition (3.8.1) is violated is equal to

$$H(\frac{\nu - \rho\delta}{1 - \rho\delta})(1 - \rho\delta) - \xi_{\gamma'}(y_{\gamma'})(\nu - \rho\delta) - \Psi_{ext}(x_\nu).$$

But from the derivations of the exponents, we know that

$$0 = \lim_{\delta \to 1} \sup_{\nu \in [\delta, 1)} \Psi_{ext}(x_\nu),$$

$$0 = \lim_{\delta \to 1} \sup_{\nu \in [\delta, 1)} H(\frac{\nu - \rho\delta}{1 - \rho\delta}),$$

$$\xi_{\gamma'}(y_{\gamma'})(\nu - \rho\delta) \geq \xi_{\gamma'(\delta)}(y_{\gamma'(\delta)})(1 - \rho)\delta > 0, \ \nu \geq \delta,$$

where

$$\gamma'(\delta) = \frac{\rho\delta}{\frac{C^2-1}{C^2}\rho\delta + \frac{\delta}{C^2}} = \frac{\rho}{\frac{C^2-1}{C^2}\rho + \frac{1}{C^2}}.$$

Noticing that $\xi_{\gamma'(\delta)}(y_{\gamma'(\delta)})(1 - \rho) > 0$ is only determined by ρ, for any $0 < \rho < 1$, there exists a big enough $\delta < 1$, such that

$$H(\frac{\nu - \rho\delta}{1 - \rho\delta})(1 - \rho\delta) - \xi_{\gamma'}(y_{\gamma'})(\nu - \rho\delta) - \Psi_{ext}(x_\nu) > 0,$$

uniformly over $[\delta, 1]$. Then it follows that

$$\lim_{\delta \to 1} \zeta_W(\delta) = 1.$$

■

Chapter 4

The Weighted ℓ_1 Minimization Algorithm

4.1 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 [RIC]. The whole premise is that sparse signals (signals with many zero or negligible elements in 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 [CT05].

The conventional approach to compressed sensing assumes no prior information on the unknown signal other than the fact that it is sufficiently sparse in a particular basis. In many applications, however, additional prior information is available. In fact, in many cases the signal recovery problem (which compressed sensing attempts to address) is a detection or estimation problem in some statistical setting. Some recent work along these lines can be found in [MDB] (which considers compressed detection and estimation) and [JXC08] (on Bayesian compressed sensing). In other cases, 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.

In this chapter we will consider a particular model for the sparse signal that assigns a probability of being zero or nonzero to each entry of the unknown vector. The standard compressed sensing model is therefore a special case where these probabilities are all equal (for example, for a k-sparse vector the probabilities will all be $\frac{k}{n}$, where n is the number of entries of the unknown vector). 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 where the signal is often *block sparse*, i.e., the signal is more likely to be nonzero in certain blocks rather than in others [SPH].

While it is possible (albeit cumbersome) to study this model in full generality, in this chapter we will focus on the case where the entries of the unknown signal fall into a fixed number T of categories: in the ith set K_i (with cardinality n_i) the probability of being nonzero is P_i (Clearly, in this case the sparsity[1] will, with high probability, be around $\sum_{i=1}^{T} n_i P_i$.) This model is rich enough to capture many of the salient features regarding prior information. The signal generated based on this model could be the vector representation of a natural image in some linear transform domain (e.g., DFT, DCT, DWT ...) or the spatial representation of some biomedical image, e.g., a brain fMRI image. Although the latter is not essentially sparse, the difference of the brain image at any moment during an experiment and an initial baseline 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 in the form of which regions of brain are more likely to be involved in the decision making process. This can be captured in the above *nonuniform sparse* model. DPCM encoders are example of systems that are based on encoding only the difference of consecutive samples which results in more efficient coding rates [Cut]. In a similar fashion, this model is applicable to other problems like network monitoring (see [CPR] for an application

[1] Quantitatively speaking, by sparsity we mean the number of nonzero elements of a vector $\mathbf{x}$.

of compressed sensing and nonlinear estimation in compressed network monitoring), DNA microarrays [MBSR, ES, VPMH], astronomy, satellite imaging and a lot more.

In this chapter we do the analysis for the case where there are only two categories of entries ($T = 2$) and show that even for that case the performance is going to be boosted significantly by making use of the additional information. While it is in principle possible to analyze this model with more than two categories of entries ($T > 2$), the analysis becomes increasingly tedious and we leave it as feature work. An interesting question would be to characterize the gain in recovery percentage as a function of the classes T to which we can classify the signal entries.

The contributions of this chapter are the following. We propose a weighted ℓ_1 minimization approach for sparse recovery where the ℓ_1 norms of each set are given different weights w_i ($i = 1, 2$). Clearly, one would want to give a larger weight to those entries whose probability of being nonzero is less (thus further forcing them to be zero).[2] The second contribution is to compute explicitly the relationship between the p_i, the w_i, the $\frac{n_i}{n}$, $i = 1,\ 2$ and the number of measurements so that the unknown signal can be recovered with overwhelming probability as $n \to \infty$ (the so-called weak threshold) 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 [Don06c, DT05a] (e.g., Grassman angles) to obtain sharp thresholds for compressed sensing. However, rather than use the *neighborliness* condition used in [Don06c, DT05a], we find it more convenient to use the null space characterization of Xu and Hassibi [XH08, SXH08a]. The resulting Grassmanian manifold approach is a general framework for incorporating additional factors into compressed sensing: in [XH08] it was used to incorporate measurement noise; here it is used to incorporate prior information and weighted ℓ_1 optimization. Our analytic results allow us to compute the optimal weights for any p_1, p_2, n_1, n_2. We also provide simulation results to show the advantages of the weighted method over standard ℓ_1 minimization.

[2] A somewhat related method that uses weighted ℓ_1 optimization is Candes et al [CWB08]. The main difference is that there is no prior information and at each step the ℓ_1 optimization is re-weighted using the estimates of the signal obtained in the last minimization step.

This chapter is organized as follows. In the next section we describe the model and state the principle assumptions of nonuniform sparsity. 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 4.3, we go briefly over our critical theorems and try to present a big picture of the main results. Sections 4.4 and 4.4.4 are dedicated to the derivation of these results in concrete. Finally, in section 4.5 some simulation results are presented and compared to the analytical bounds of the previous sections.

4.2 Problem Description

The signal is represented by a $n \times 1$ vector $\mathbf{x} = (x_1, x_2, ..., x_n)^T$ of real valued numbers, and is *non-uniformly sparse* with sparsity factors $P_1, P_2, ..., P_T$ over the (index) sets $K_1, K_2, ..., K_T$, $K_i \cap K_j = \emptyset$ $i \neq j$ and $\bigcup_{i=1}^{T} K_i = \{1, 2, ..., n\}$. By this, we mean that for each index $1 \leq i \leq n$, if $i \in K_j$ then x_i is a nonzero element (with an arbitrary distribution say $\mathcal{N}(0, 1)$) with probability P_j, and zero with probability $1 - P_j$ independent of all other elements of $\mathbf{x}$. The signal is thus non-homogenously (non-uniformly) sparse over the sets $K_1, ..., K_T$. In Figure 4.1, the support set of a sample signal generated based on the described nonuniform sparse model is schematically depicted. The number of classes is consider to be $T = 2$ in that case with the two classes having the same size $\frac{n}{2}$. The sparsity factor in the first class (K_1) is $P_1 = 0.3$, and in 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 nonnegative entries. However, as expected, it is not independent of the properties of the measurement matrix. We assume that the measurement matrix $\mathcal{A}$ is a $m \times n$ ($\frac{m}{n} = \delta < 1$) matrix with i.i.d standard Gaussian distributed $\mathcal{N}(0, 1)$ entries. The observation vector is denoted by $\mathbf{y}$ and obeys the following:

$$\mathbf{y} = A\mathbf{x}. \tag{4.2.1}$$

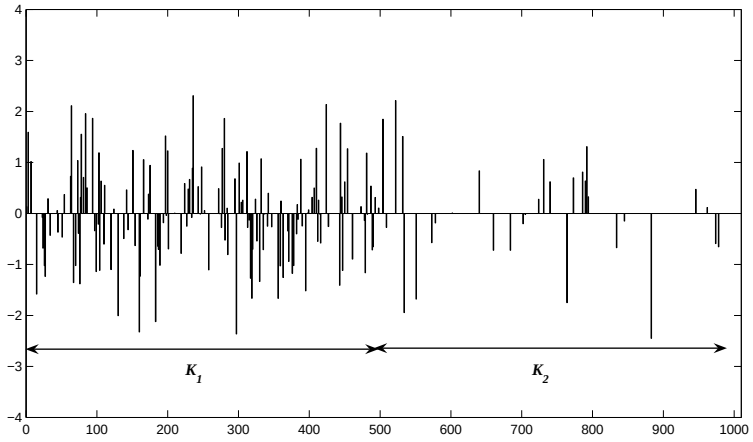


Figure 4.1: Illustration of a non-uniformly sparse signal

As mentioned in Section 4.1, ℓ_1-minimization can recover a vector $\mathbf{x}$ with $k = \mu n$ non-zero entries, provided μ is less than a known function of δ. ℓ_1 minimization has the following form:

$$\min_{\mathbf{Ax=y}} \|\mathbf{x}\|_1. \tag{4.2.2}$$

Please see [Don06c] for the exact relationship between μ and δ in the case of Gaussian measurements. (4.2.2) is a linear programming and can be solved polynomially fast ($O(n^3)$). However, it fails to encapsulate additional prior information of the signal nature, might there be any such information. One might simply think of modifying (4.2.2) to a weighted ℓ_1 minimization as follows:

$$\min_{\mathbf{Ax=y}} \|\mathbf{x}\|_{\mathbf{w}1} = \min_{Ax=y} \sum_{i=1}^{n} w_i |x_i|. \tag{4.2.3}$$

The index $\mathbf{w}$ is an indication of the $n \times 1$ positive weight vector. Now the question is what is the optimal set of weights, and can one improve the recovery threshold using the weighted ℓ_1 minimization of (4.2.3) with those weights rather than (4.2.2). We have to be more clear with the objective at this point and what we mean by extending the recovery threshold. First of all note that the vectors generated based on the model

described above can have any arbitrary number of nonzero entries. However, their support size is typically (with probability arbitrary close to one) around $n_1P_1 + n_2P_2$. Therefore, there is no such notion of strong threshold as in the case of [Don06c]. We are asking the question of for what P_1 and P_2 signals generated based on this model can be recovered with overwhelming probability as $n \to \infty$. Moreover we are wondering if by adjusting w_i's according to P_1 and P_2 can one extend the typical sparsity to dimension ratio ($\frac{n_1P_1+n_2P_2}{n}$) for which reconstruction is successful with high probability.

4.3 Summary of Main Results

We address two main questions in this chapter. First we want to know how much the weighted ℓ_1 minimization approach help improve the performance of the recovery (decrease the misdetection probability). Second, using the answer to the first question, we are interested to know what the optimal choice of the set of weights (w_is) is. Given that the signal is generated based on the model of section 4.2, the natural question is for which regimes of the problem parameters is the recovery with weighted ℓ_1 minimization almost surely successful. In other words given that the ratio, the vector $\mathbf{w}$ and the probabilities $P1$ and P_2 are fixed, what is the minimum number of measurements to dimension ratio (i.e. minimum ratio $\delta = \frac{m}{n}$) that guarantees the weighted ℓ_1 minimization of 4.2.3 successfully retrieves the signal almost surely as $n \to \infty$. Based on that characterization, the *optimal* set of weights would be those that result in smaller recovery threshold for δ.

To this end, in the first place, we try to understand how the misdetection (failure recovery) event is related to the properties of the measurement matrix. For the non-weighted case, this has been considered in [SXH08a] and is known by the null space property. we generalize this result to the case of weighted ℓ_1 minimization, and mention a necessary and sufficient condition for (4.2.3) to recover the original signal of interest. The theorem is as follows.

Theorem 4.3.1. *Let* $\mathbf{x}_0$ *be a* $n \times 1$ *vector supported on the set* $K \subseteq \{1, 2, ..., n\}$. *Then* $\mathbf{x}_0$ *is the unique solution to the linear programming* $\min_{\mathcal{A}\mathbf{x}=\mathbf{y}} \sum_{i=1}^{n} w_i|x_i|$ *with* $\mathbf{y} = \mathbf{x}_0$, *if and only if for every* Z *in the null space of* $\mathcal{A}$ *the following holds* $\sum_{i\in K} w_i|Z_i| \leq \sum_{i\in \overline{K}} w_i|Z_i|$.

This theorem will be stated and proven in section 4.4. As will be explained in section 4.4.1, Theorem 4.3.1 along with the known facts on the null space of random Gaussian matrices, help us interpret the probability of recovery error in terms of a high dimensional geometrical event called the complementary Grassman angle; namely that a uniformly chosen $(n - m)$-dimensional subspace Ψ shifted by the point x, $(\Psi + \mathbf{x})$, intersects the skewed weighted crosspolytope $SP_{\mathbf{w}} = \{\mathbf{y} \in \mathbb{R}^n \mid \sum_{i=1}^{n} w_i|y|_i \leq 1\}$ nontrivially at some other point besides $\mathbf{x}$. The fact that we can take for granted without proving is that due to identical distribution of all entries of $\mathbf{x}$ in each of the sets K_1 and K_2, the entries of the optimal weight vector only takes two values W_1 and W_2 depending on their index. In other words

$$\forall i \in \{1, 2, \ldots, n\} \quad w_i = \begin{cases} W_1 \ if \ i \in K_1, \\ W_2 \ if \ i \in K_2. \end{cases} \tag{4.3.1}$$

Leveraging on the existing techniques for computation of complementary Grassman manifold [San52, McM75] and theorems of typicality we will be able to state and prove the following theorems along the same lines, which essentially provides the answer to our first key question.

Theorem 4.3.2. *Recall that* $n_1 = |K_1|$ *and* $n_2 = |K_2|$ *are the set sizes defined earlier. Also, let* E *be the event that a random vector* $\mathbf{x}_0$ *generated based on the sparsity model of section 4.2 is recovered by the linear programming of (4.2.3) with* $\mathbf{y} = \mathbf{x}_0$. *For every*

$\epsilon > 0$ there exists a positive constant c_ϵ so that:

$$\mathbb{P}(E^c) \leq O(e^{-c_\epsilon n}) + \sum_{k_1=n_1(P_1-\epsilon)}^{n_1(P_1+\epsilon)} \sum_{k_2=n_2(P_2-\epsilon)}^{n_2(P_2+\epsilon)} \left(\sum_{\substack{0 \leq t_1 \leq n_1-k_1 \\ 0 \leq t_2 \leq n_2-k_2 \\ t_1+t_2 > m-k_1-k_2+1}} 2^{t_1+t_2+1} \binom{n_1-k_1}{t_1} \binom{n_2-k_2}{t_2} \beta(k_1,k_2|t_1,t_2)\gamma(t_1+k_1,t_2+k_2) \right) \tag{4.3.2}$$

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 skewed crosspolytope $SP_{\mathbf{w}} = \{y \in \mathbb{R}^n | \sum_{i=1}^n w_i|y_i| \leq 1\}$ with k_1 vertices supported on K_1 and k_2 vertices supported on K_2 and a $(k_1+k_2+t_1+t_2-1)$-dimensional face $\mathcal{G}$ that includes $\mathcal{F}$ and has t_1+k_1 faces supported on K_1 and the remaining vertices supported on K_2. $\gamma(d_1,d_2)$ 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 skewed crosspolytope $SP_{\mathbf{w}}$.

We are actually interested in the regimes that lead the above upper bound to decay to zero as $n \to \infty$, which entails the cumulative exponent in (4.4.12) be negative. We are able to calculate upper bounds on the exponents of the terms in (4.4.12) by using large deviations of sums of normal and half-normal variables. More precisely, for small enough ϵ_1 and ϵ_2, if we assume that the sum of the terms corresponding to a particular t_1 and t_2 in (4.4.12) is denoted by $F(t_1,t_2)$ then we are able to find and compute an exponent function $\psi_{tot}(t_1,t_2) = \psi_{com}(t_1,t_2) - \psi_{int}(t_1,t_2) - \psi_{ext}(t_1,t_2)$ so that $\frac{1}{n}\log F(t_1,t_2) \sim \psi(t_1,t_2)$ as $n \to \infty$. Note that $\psi_{com}(.,.)$, $\psi_{int}(.,.)$ and $\psi_{ext}(.,.)$ are the contributions to the exponent by combinatorial, internal angle and external angle terms respectively. Next, we state a key theorem that enables us to provide the answer to the second main question. Note that we will denote by δ the ratio $\frac{m}{n}$ and by γ_1 and γ_2 the ratios $\frac{n_1}{n}$ and $\frac{n_2}{n}$ respectively.

Theorem 4.3.3. *If γ_1, γ_2, P_1, P_2, W_1 and W_2 are fixed, there exists a critical threshold $\delta_c = \delta_c(P_1, P_2, \frac{W_2}{W_1})$ such that if $\delta = \frac{m}{n} \geq \delta_c$, then the R.H.S. of (4.3.2) (the upper bound on the probability of failure) decays exponentially to zero as $n \to \infty$. Furthermore, δ_c is given by*

$$\delta_c = \min\{\delta \mid \psi_{com}(t_1', t_2') - \psi_{int}(t_1', t_2') - \psi_{ext}(t_1', t_2') < 0 \ \forall \ 0 \leq t_1' \leq \gamma_1(1-P_1), \\ 0 \leq t_2' \leq \gamma_2(1-P_2), t_1' + t_2' > \delta - \gamma_1 P_1 - \gamma_2 P_2\},$$

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

1. *(Combinatorial exponent)*

$$\psi_{com}(t_1', t_2') = \log 2 \left(\sum_{i=1}^{2} \left(\gamma_i(1-P_i) H(\frac{t_i'}{\gamma_i(1-P_i)}) + t_i' \right) \right) \tag{4.3.3}$$

   where $H(.)$ is the Shannon entropy function defined by $H(x) = -x \log_2 x - (1-x)\log_2(1-x)$.

2. *(External angle exponent) Let $g(x) = \frac{2}{\sqrt{\pi}} e^{-\frac{x^2}{2}}$, $G(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2} dy$. Also define $C = (t_1' + \gamma_1 P_1) + W^2(t_2' + \gamma_2 P_2)$, $D_1 = \gamma_1(1-P_1) - t_1'$ and $D_2 = \gamma_2(1-P_2) - t_2'$. Let x_0 be the unique solution to x of the following:*

$$2C - \frac{g(x)D_1}{xG(x)} - \frac{Wg(Wx)D_2}{xG(Wx)} = 0.$$

   Then

$$\psi_{ext}(t_1', t_2') = Cx_0^2 - D_1 \log G(x_0) - D_2 \log G(Wx_0). \tag{4.3.4}$$

3. *(Internal angle exponent) Let $b = \frac{t_1' + W^2 t_2'}{t_1' + t_2'}$ and $\varphi(.)$ and $\Phi(.)$ be the standard Gaussian pdf and cdf functions respectively. Also let $\Omega' = \gamma_1 P_1 + W^2 \gamma_2 P_2$ and $Q(s) = \frac{t_1' \varphi(s)}{(t_1'+t_2')\Phi(s)} + \frac{W t_2' \varphi(Ws)}{(t_1'+t_2')\Phi(Ws)}$. Define the function $\hat{M}(s) = -\frac{s}{Q(s)}$ and solve for s in $\hat{M}(s) = \frac{t_1'+t_2'}{(t_1'+t_2')b+\Omega'}$. Let the unique solution be s^* and set $y = s^*(b - \frac{1}{\hat{M}(s^*)})$. Compute the rate function $\Lambda^*(y) = sy - \frac{t_1'}{t_1'+t_2'}\Lambda_1(s) - \frac{t_2'}{t_1'+t_2'}\Lambda_1(Ws)$ 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

$$\psi_{int}(t'_1, t'_2) = (\Lambda^*(y) + \frac{t'_1 + t'_2}{2\Omega'} y^2 + \log 2)(t'_1 + t'_2). \tag{4.3.5}$$

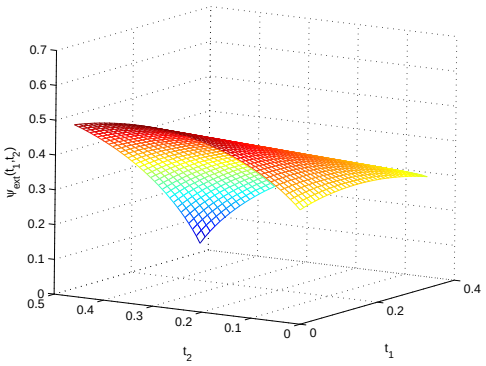
Theorem 4.3.3 is a very powerful result, since it allows us to analytically find the optimal set of weights for which the fewest possible measurements are needed to recover the signals almost surely. In fact, all we have to do is to find for fixed values of P_1 and P_2, the ratio $\frac{W_2}{W_1}$ for which the critical threshold $\delta_c(P_1, P_2, \frac{W_2}{W_1})$ from Theorem 4.3.3 is the smallest. We discuss this by some examples in Section 4.5. An example illustrations of the combinatorial, internal angle and external angle exponent as functions of t'_1 and t'_2 is given in Figure 4.2. There, it has been assumed that $\gamma_1 = \gamma_2 = 0.5$, $P_1 = 0.05$, $P_2 = 0.3$ and $\frac{W_2}{W_1} = 1.5$. Note that δ is not directly involved in the value of ψ_{com}, ψ_{int} and ψ_{ext} for fixed t'_1 and t'_2. However, it plays an important role as for the constraints that it imposes on the region of valid t'_1 and t'_2 (see Theorem 4.3.3).

As mentioned earlier, using Theorem 4.3.3, it is possible to find analytically the optimal ratio $\frac{W_2}{W_1}$. For $P_1 = 0.3$ and $P_2 = 0.05$, we have numerically computed $\delta_c(P_1, P_2, \frac{W_2}{W_1})$ as a function of $\frac{W_2}{W_1}$ and depicted the resulting curve in Figure 4.3. This suggests that $\frac{W_2}{W_1} = 2.5$ is almost the optimal ratio we can choose. We compare this to the simulation results.

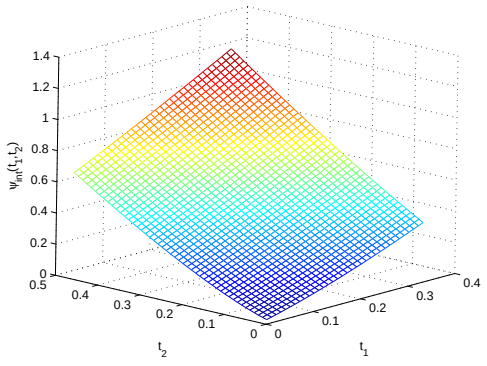
4.4 Derivation of the Main Results

We first mention the typicality and bound the error probability assuming that the nontypical portion decays exponentially. We then state the null space condition and relate the failure event to a corresponding event on the skewed weighted polytope $SP_{\mathbf{w}}$ and consequently the Grassman manifold and the summation formula in which internal and external angles show up.

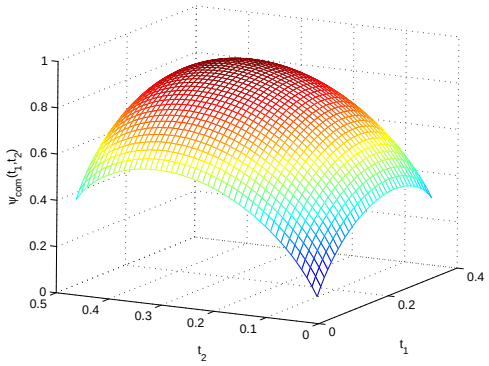
Let $\mathbf{x}$ be a random sparse signal generated based on the non-uniformly sparse model of section 4.2 and be supported on the set K. For $\epsilon > 0$ we call K *ϵ-typical* if $||K \cap K_1| - n_1 P_1| \leq \epsilon n$ and $||K \cap K_2| - n_2 P_2| \leq \epsilon n$. Interchangeably, we may



(a) External angle exponent.



(b) Internal angle exponent.



(c) Combinatorial factor exponent.

Figure 4.2: A plot of asymptotes of external angle, internal angle and combinatorial factor exponents for $\gamma_1 = \gamma_2 = 0.5$, $P_1 = 0.05$, $P_2 = 0.3$ and $\frac{W_2}{W_1} = 1.5$

also call $\mathbf{x}$ ϵ-typical. Let E be the event that $\mathbf{x}$ is recovered by (4.2.3). Then by conditioning we have

$$\begin{aligned}\mathbb{P}(E^c) &= \mathbb{P}(E^c|K \textit{ is } \epsilon\textit{-typical})\mathbb{P}(K \textit{ is } \epsilon\textit{-typical}) \\ &\quad + \mathbb{P}(E^c|K \textit{ not } \epsilon\textit{-typical})\mathbb{P}(K \textit{ not } \epsilon\textit{-typical}).\end{aligned}$$

According to the law of large numbers, for any fixed $\epsilon > 0$, $\mathbb{P}(K \textit{ not } \epsilon\textit{-typical})$ will decay exponentially as n grows. So, in order to bound the probability of failed recovery, we may assume that K is ϵ-typical for any small enough ϵ. In other words for any $\epsilon > 0$,

$$\mathbb{P}(E^c) = \mathbb{P}(E^c \wedge K \textit{ is } \epsilon\textit{-typical}) + O(e^{-c_\epsilon n}). \tag{4.4.1}$$

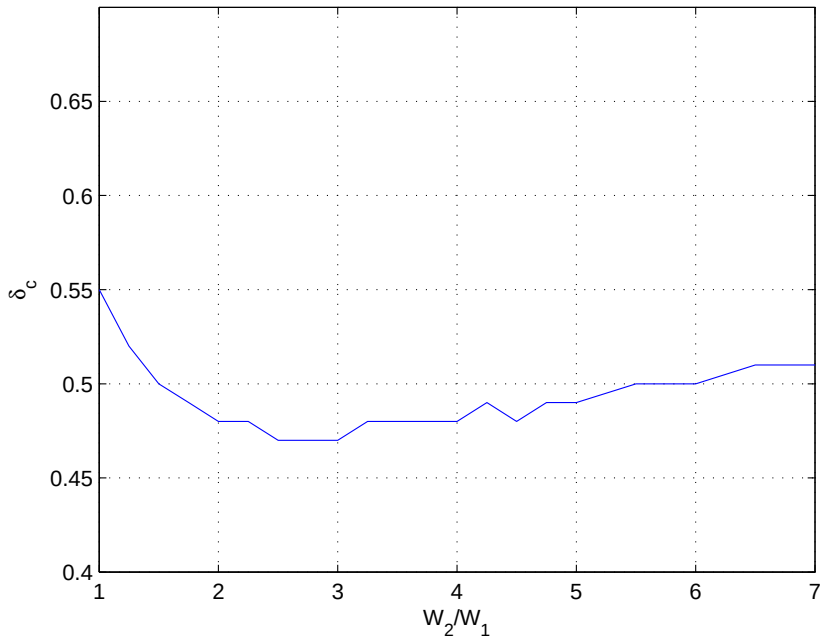


Figure 4.3: δ_c as a function of $\frac{W_2}{W_1}$ for $P_1 = 0.3$ and $P_2 = 0.05$

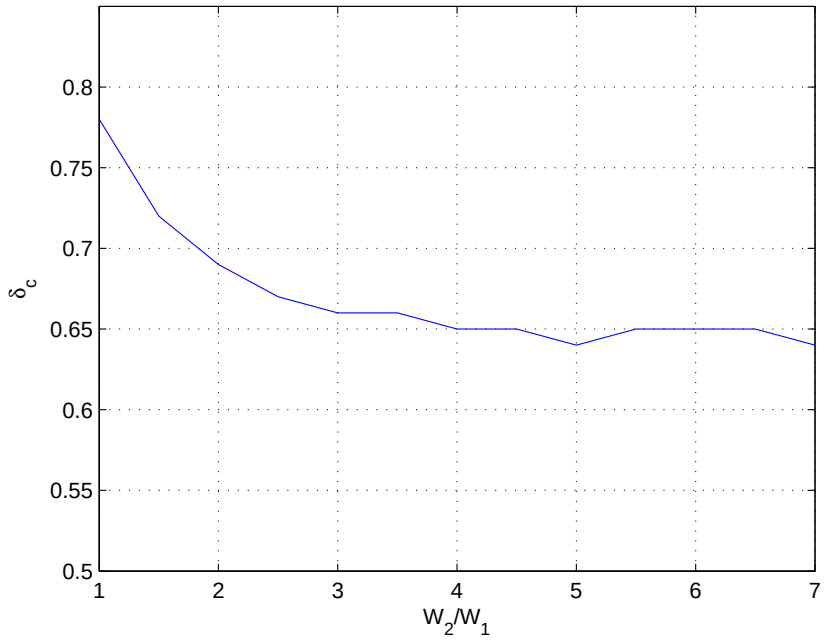


Figure 4.4: δ_c as a function of $\frac{W_2}{W_1}$ for $P_1 = 0.65$ and $P_2 = 0.1$

In order to bound the conditional error probability $\mathbb{P}(E^c|K \textit{ is } \epsilon\textit{-typical})$ we adopt the idea of [SXH08a] to interpret the failure recovery event (E^c) in terms of an event on the null space of the measurement matrix $\mathcal{A}$.

Theorem 4.4.1. *Let $\mathbf{x}_0$ be a $n \times 1$ vector supported on the set $K \subseteq \{1, 2, ..., n\}$. Then $\mathbf{x}_0$ is the unique solution to the linear programming $\min_{\mathcal{A}\mathbf{x}=\mathbf{y}} \sum_{i=1}^{n} w_i|x_i|$ with $\mathbf{y} = \mathbf{x}_0$, if and only if for every Z in the null space of $\mathcal{A}$ the following holds $\sum_{i\in K} w_i|Z_i| \leq \sum_{i\in\overline{K}} w_i|Z_i|$.*

Proof. This is almost identical to the proof of Theorem 1 of [SXH08a] or Theorem 1 of [XH08] in which ℓ_1 norm is replaced by the weighted ℓ_1 norm (which is still a valid norm). ∎

From this point on, we follow closely the steps towards calculating the upper bound on the failure probability from [XH08], but with appropriate modification. Particularly, we are dealing with a *weighted* skewed crosspolytope $SP_{\mathbf{w}}$ instead of the regular skewed crosspolytope $SP_{\mathbf{w}}$, in which all the weights (w_is) are equal to one. The key in our derivation is the following lemma.

Lemma 4.4.1. *For a certain subset $K \subseteq \{1, 2, ..., n\}$ with $|K| = k$, the event that the null-space $\mathcal{N}(A)$ satisfies*

$$\sum_{i\in K} w_i|Z_i| \leq \sum_{i\in\overline{K}} w_i|Z_i|, \forall Z \in \mathcal{N}(A). \tag{4.4.2}$$

is equivalent to the event that $\forall \mathbf{x}$ supported on the k-set K (or supported on a subset of K):

$$\sum_{i\in K} w_i|x_i + Z_i| + \sum_{i\in\overline{K}} w_i|Z_i| \geq \sum_{i\in K} w_i|x_i|, \forall Z \in \mathcal{N}(A), \tag{4.4.3}$$

Proof. First, let us assume that $\sum_{i\in K} w_i|Z_i| \leq \sum_{i\in\overline{K}} w_i|Z_i|, \forall Z \in \mathcal{N}(A)$. 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 L.H.S.) we obtain

$$\begin{aligned}\sum_{i\in K} w_i|x_i + Z_i| + \sum_{i\in\overline{K}} w_i|Z_i| &\geq \sum_{i\in K} w_i|x_i| - \sum_{i\in K} w_i|Z_i| + \sum_{i\in\overline{K}} w_i|Z_i| \\ &\geq \sum_{i\in K} w_i|x_i|.\end{aligned}$$

thus proving the forward part of this lemma. Now let us assume instead that $\exists Z \in \mathcal{N}(A)$, 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 $\mathbf{x}_K = -Z_K$ (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 inverse part of this lemma. ■

4.4.1 Upper Bound on the Failure Probability

Knowing Lemma 4.4.1, we are now in a position to derive the probability that condition (4.4.2) holds for the sparsity $|K| = k$, if we uniformly sample a random $(n-m)$-dimensional subspace Ψ from the Grassmann manifold $\mathrm{Gr}_{(n-m)}(n)$. From the previous discussions, we can equivalently consider the complementary probability $P = P(E^c \wedge K\ is\ \epsilon\text{-}typical)$, namely the probability that a ϵ-typical subset $K \subset \{1, 2, ..., n\}$ with $|K| = k$, and a vector $\mathbf{x} \in \mathbb{R}^n$ (with a random sign pattern) supported on the set K (or a subset of K) fail the condition (4.4.3). Due to the vector linear proportionality in the linear subspace Ψ, we can restrict our attention to those vectors $\mathbf{x}$ from the weighted crosspolytope

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

that are only supported on the set K (or a subset of K).

Since K is assumed to be ϵ-typical set and x has one particular sign pattern (i.e.

is drawn randomly), unlike [XH08] we do not need to upper bound the probability P by a union bound over all the possible support sets $K \subset \{1, 2, ..., n\}$ and all the sign patterns of the k-sparse vector $\mathbf{x}$. Instead, we can write

$$P \leq \sum_{K \ \epsilon\text{-typical}} P_{K,-}, \tag{4.4.4}$$

where $P_{K,-}$ is the probability that for a specific *support set* K (which is epsilon typical here), there exist a k-sparse vector $\mathbf{x}$ of a specific *sign pattern* which fails the condition (4.4.3). By symmetry, without loss of generality, we assume the signs of the elements of $\mathbf{x}$ to be non-positive.

So 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 (or a subset of K) and can be restricted to the weighted crosspolytope $\{\mathbf{x} \in \mathbb{R}^n \mid \sum_{i=1}^n w_i|x_i| = 1\}$, $\mathbf{x}$ is also on a $(k-1)$-dimensional face, denoted by $\mathcal{F}$, of the weighted skewed crosspolytope $SP_{\mathbf{w}}$:

$$SP_{\mathbf{w}} = \{\mathbf{y} \in \mathbb{R}^n \mid \sum_{i=1}^{n} w_i|y_i| \leq 1\}. \tag{4.4.5}$$

The subscript $\mathbf{w}$ in $SP_{\mathbf{w}}$ is an indication of the weight vector $\mathbf{w} = (w_1, w_2, \ldots, w_n)^T$.

Now the probability $P_{K,-}$ is the probability that there exists an $\mathbf{x} \in \mathcal{F}$, and there exists a $Z \in \Psi$ $(Z \neq 0)$ such that

$$\sum_{i \in K} w_i|x_i + Z_i|_1 + \sum_{i \in \bar{K}} w_i|Z_i|_1 \leq \sum_{i \in K} w_i|x_i| = 1. \tag{4.4.6}$$

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 $\exists Z \in \Psi$ $(Z \neq 0)$ such that

$$\sum_{i \in K} w_i|x_i + Z_i|_1 + \sum_{i \in \bar{K}} w_i|Z_i|_1 \leq \sum_{i \in K} w_i|x_i| = 1, \tag{4.4.7}$$

is essentially the probability that a uniformly chosen $(n-m)$-dimensional subspace Ψ

shifted by the point $\mathbf{x}$, namely $(\Psi + \mathbf{x})$, intersects the weighted skewed crosspolytope

$$SP_{\mathbf{w}} = \{\mathbf{y} \in \mathbb{R}^n \mid \sum_{i=1}^{n} w_i |y_i| \leq 1\} \tag{4.4.8}$$

non-trivially, namely, at some other point besides $\mathbf{x}$.

From the linear property of the subspace Ψ, the event that $(\Psi + \mathbf{x})$ intersects the skewed crosspolytope $SP_{\mathbf{w}}$ is equivalent to the event that Ψ intersects nontrivially with the cone $SPCone_{\mathbf{w}}(\mathbf{x})$ obtained by observing the weighted skewed polytope $SP_{\mathbf{w}}$ from the point $\mathbf{x}$. (Namely, $SPCone_{\mathbf{w}}(\mathbf{x})$ is conic hull of the point set $(SP_{\mathbf{w}} - \mathbf{x})$ and of course $SPCone_{\mathbf{w}}(\mathbf{x})$ has the origin of the coordinate system as its apex.) However, as noticed in the geometry for convex polytopes [Grü68, Grü03], the $SPCone_{\mathbf{w}}(\mathbf{x})$ is 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 $\mathbf{x}$ is only a single point in the relative interior of the face $\mathcal{F}$. (The acute reader may have noticed 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 $SPCone_{\mathbf{w}}(\mathbf{x})$ 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*** [Grü68] for the face $\mathcal{F}$ with respect to the polytope SP under the Grassmann manifold $\mathrm{Gr}_{(n-m)}(n)$:[3] the probability of a uniformly distributed $(n-m)$-dimensional subspace Ψ from the Grassmannian manifold $\mathrm{Gr}_{(n-m)}(n)$ intersecting non-trivially with the cone SP-Cone($\mathbf{x}$) formed by observing the skewed crosspolytope SP from the relative interior point $\mathbf{x} \in \mathcal{F}$.

Building on the works by L. A. Santalö [San52] and P. McMullen [McM75] in high

[3] A Grassman angle and its corresponding complementary Grassmann angle always sum up to 1. There is apparently inconsistency in terms of the definition of which is "Grassmann angle" and which is "complementary Grassmann angle" between [Grü68],[AS92] and [VS92] etc. But we will stick to the earliest definition in [Grü68] for Grassmann angle: the measure of the subspaces that intersect trivially with a cone.

dimensional integral 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 [Grü03]:

$$2 \times \sum_{s \geq 0} \sum_{G \in \Im_{m+1+2s}(\mathrm{SP})} \beta(\mathcal{F}, \mathcal{G}) \gamma(\mathcal{G}, SP_{\mathbf{w}}), \tag{4.4.9}$$

where s is any nonnegative integer, $\mathcal{G}$ is any $(m+1+2s)$-dimensional face of the skewed crosspolytope ($\Im_{m+1+2s}(\mathrm{SP})$ is the set of all such faces), $\beta(\cdot,\cdot)$ stands for the internal angle and $\gamma(\cdot,\cdot)$ stands for the external angle.

The internal angles and external angles are basically defined as follows [Grü03, McM75]:

- 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 $\mathcal{F}_1$.[4] 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 $\gamma(\mathcal{F}_3, \mathcal{F}_4)$ 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 $\gamma(\mathcal{F}_3, \mathcal{F}_4)$ 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 crosspolytope $SP_{\mathbf{w}}$. First of all, $SP_{\mathbf{w}}$ is nothing but the convex hull of the following set of $2n$ vertices in $\mathbb{R}^n$

$$SP_{\mathbf{w}} = conv\{\pm\frac{\varepsilon_i}{w_i} \mid 1 \leq i \leq n\} \tag{4.4.10}$$

where ε_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 $SP_{\mathbf{w}}$ is just the convex hull of k of the linearly

[4]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.

independent vertices of $SP_{\mathbf{w}}$. We then say that $\mathcal{F}$ is supported on the index set K of the k indices corresponding to these vertex indices. More precisely, if $\mathcal{F} = conv\{j_1\frac{\varepsilon_{i_1}}{w_{i_1}}, j_2\frac{\varepsilon_{i_2}}{w_{i_2}}, \ldots, j_n\frac{\varepsilon_{i_k}}{w_{i_k}}\}$ with $j_i \in \{-1, +1\}$ $\forall 1 \leq i \leq k$, then $\mathcal{F}$ is supported on the set $K = \{i_1, i_2, \ldots, i_k\}$. The particular choice of w_is as in (4.3.1) makes $SP_{\mathbf{w}}$ partially symmetric. Two faces $\mathcal{F}$ and $\mathcal{F}'$ of $SP_{\mathbf{w}}$ that are respectively supported on K and K', are geometrically identical.[5] if $|K \cap K_1| = |K' \cap K_1|$ and $|K \cap K_2| = |K' \cap K_2|$.[6] In other words the only thing that distinguishes the faces is the proportion of their support sets that is located in K_1 or K_2. Therefore for two faces $\mathcal{F}$ and $\mathcal{G}$ with $\mathcal{F}$ supported on K and $\mathcal{G}$ supported on L ($K \subseteq L$), $\beta(\mathcal{F}, \mathcal{G})$ is only a function of the parameters $k_1 = |K \cap K_1|$, $k_2 = |K \cap K_2|$, $k_1 + t_1 = |L \cap K_1|$ and $k_2 + t_1 = |K \cap K_2|$. So, instead of $\beta(\mathcal{F}, \mathcal{G})$ we write $\beta(k_1, k_2|t_1, t_2)$ and similarly instead of $\gamma(\mathcal{G}, SP_{\mathbf{w}})$ we just write $\gamma(t_1 + k_1, t_2 + k_2)$. Using this notation and recalling the formula (4.4.9), we can write

$$\begin{aligned} P_{K,-} &= 2\sum_{s\geq 0}\sum_{G\in\Im_{m+1+2s}(\mathrm{SP})} \beta(\mathcal{F},\mathcal{G})\gamma(\mathcal{G}, SP_{\mathbf{w}}) \\ &= \sum_{\substack{0\leq t_1\leq n_1-k_1 \\ 0\leq t_2\leq n_2-k_2 \\ t_1+t_2>m-k_1-k_2+1}} 2^{t_1+t_2+1}\binom{n_1-k_1}{t_1}\binom{n_2-k_2}{t_2}\beta(k_1,k_2|t_1,t_2)\gamma(t_1+k_1,t_2+k_2) \end{aligned} \tag{4.4.11}$$

where in (4.4.11) we have used the fact that the number of faces $\mathcal{G}$ of $SP_{\mathbf{w}}$ of dimension $l - 1 = t_1 + t_2$ that encompass $\mathcal{F}$ and have $k_1 + t_1$ vertices supported on K_1 and the rest $k_2 + t_2$ vertices supported on K_2 is $2^{t_1+t_2}\binom{n_1-k_1}{t_1}\binom{n_2-k_2}{t_2}$. Now we can apply the union bound of (4.4.4) to get the following result.

Theorem 4.4.2. *Let E be the event that a random vector $\mathbf{x}_0$ generated based on the sparsity model of section 4.2 is recovered by the linear programming of (4.2.3) with*

[5]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., $\mathcal{F}' = \theta\mathcal{F}$.

[6]Remember that K_1 and K_2 are the same sets as defined in the model description of section 4.2.

$\mathbf{y} = \mathbf{x}_0$. *For every* $\epsilon > 0$ *there exists a positive constant* c_ϵ *so that:*

$$\mathbb{P}(E^c) \leq O(e^{-c_\epsilon n}) + \sum_{k_1=n_1(P_1-\epsilon)}^{n_1(P_1+\epsilon)} \sum_{k_2=n_2(P_2-\epsilon)}^{n_2(P_2+\epsilon)} \left(\sum_{\substack{0 \leq t_1 \leq n_1-k_1 \\ 0 \leq t_2 \leq n_2-k_2 \\ t_1+t_2 > m-k_1-k_2+1}} 2^{t_1+t_2+1} \binom{n_1-k_1}{t_1} \binom{n_2-k_2}{t_2} \beta(k_1,k_2|t_1,t_2)\gamma(t_1+k_1,t_2+k_2) \right) \tag{4.4.12}$$

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* $SP_{\mathbf{w}}$ *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 includes* $\mathcal{F}$ *and has* t_1+k_1 *faces supported on* K_1 *and the remaining vertices supported on* K_2*.* $\gamma(d_1,d_2)$ *is the external angle* $\gamma(d_1,d_2)$ *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 skewed crosspolytope* $SP_{\mathbf{w}}$*.*

Proof. Apply (4.4.11) to the R.H.S. of (4.4.4) and then replace in (4.4.1) to get the desired result. ■

In the following subsections we will try to evaluate the internal and external angles for a typical face $\mathcal{F}$, and a face $\mathcal{G}$ containing $\mathcal{F}$, and try to give closed-form upper bounds for them. We combine the terms together and compute the exponents using Laplace method in section 4.4.4 and derive thresholds for nonnegativity of the cumulative exponent.

4.4.2 Computation of Internal Angle

In summary, the main result of this section is the following theorem.

Theorem 4.4.3. *Let* Z *be a random variable defined as*

$$Z = (k_1W_1^2 + k_2W_2^2)X_1 - W_1^2 \sum_{i=1}^{t_1} X_1' - W_2^2 \sum_{i=1}^{t_2} X_1'',$$

where $X_1 \sim N(0, \frac{1}{2(k_1W_1^2+k_2W_2^2)})$ *is a normal distributed random variable,* $X_i' \sim HN(0, \frac{1}{2W_1^2})$, $1 \leq i \leq t_1$, *and* $X_i'' \sim HN(0, \frac{1}{2W_2^2})$ $1 \leq i \leq t_2$ *are independent (from each other and from* X_1*) half normal distributed random variables. Let* $p_Z(.)$ *denote the probability distribution function of* Z *and* $C_0 = \frac{\sqrt{\pi}}{2^{l-k}}\sqrt{(k_1+t_1)W_1^2+(k_2+t_2)W_2^2}$. *Then,*

$$\beta(k_1, k_2|t_1, t_2) = C_0 p_Z(0). \tag{4.4.13}$$

We will try in this whole section to prove this theorem. Suppose that $\mathcal{F}$ is a ϵ-*typical* $(k-1)$-dimensional face of the skewed crosspolytope

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

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

First we can prove the following lemma:

Lemma 4.4.2. *Let* $Con_{\mathcal{F}^\perp,\mathcal{G}}$ *be the positive cone of all the vectors* $\mathbf{x} \in \mathbb{R}^n$ *that take the form:*

$$-\sum_{i=1}^{k} b_i \times e_i + \sum_{i=k+1}^{l} b_i \times e_i, \tag{4.4.14}$$

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

$$\int_{Con_{\mathcal{F}^\perp,\mathcal{G}}} e^{-\|\mathbf{x}\|^2}\, d\mathbf{x} = \beta(\mathcal{F},\mathcal{G}) V_{l-k-1}(S^{l-k-1}) \int_0^\infty e^{-r^2} r^{l-k-1}\, dx = \beta(\mathcal{F},\mathcal{G}) \cdot \pi^{(l-k)/2}, \tag{4.4.15}$$

where $V_{l-k-1}(S^{l-k-1})$ *is the spherical volume of the* $(l-k-1)$*-dimensional sphere*

S^{l-k-1}.

Proof. Given in Appendix 4.6.1. ■

From (4.4.15) we can find the expression for the internal angle. Define $U \subseteq \mathbb{R}^{l-k+1}$ as the set of all nonnegative vectors $(x_1, x_2, \ldots, x_{l-k+1})$ satisfying:

$$x_p \geq 0,\ 1 \leq p \leq l-k+1\ (\textstyle\sum_{p=1}^{k} w_p^2)x_1 = \sum_{p=k+1}^{l} w_p^2 x_{p-k+1}$$

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

$$f(x_1, \ldots, x_{l-k+1}) = -\sum_{p=1}^{k} x_1 w_p \varepsilon_p + \sum_{p=k+1}^{l} x_{p-k+1} w_p \varepsilon_p$$

Then

$$\begin{aligned}
&\int_{\mathrm{Con}_{\mathcal{F}^\perp,\mathcal{G}}} e^{-\|\mathbf{x}'\|^2}\, d\mathbf{x}' = \int_U e^{-\|f(\mathbf{x})\|^2}\, df(\mathbf{x}) \\
&= |J(A)| \int_\Gamma e^{-\|f(x)\|^2}\, dx_2 \cdots dx_{l-k+1} \\
&= |J(A)| \int_\Gamma e^{-(\sum_{p=1}^{k} w_p^2)x_1^2 - \sum_{p=k+1}^{l} w_p^2 x_{p-k+1}^2}\, dx_2 \cdots dx_{l-k+1}. \qquad (4.4.16)
\end{aligned}$$

Γ is the region described by

$$(\sum_{p=1}^{k} w_p^2)x_1 = \sum_{p=k+1}^{l} w_p^2 x_{p-k+1}, x_p \geq 0\ 2 \leq p \leq l-k+1, \qquad (4.4.17)$$

where $|J(A)|$ 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 A given by

$$A_{i,j} = \begin{cases} -\frac{1}{\Omega} w_i w_{k+j}^2 & 1 \leq i \leq k, 1 \leq j \leq l-k, \\ w_i & k+1 \leq i \leq l, j = i-k, \\ 0 & \text{otherwise.} \end{cases} \qquad (4.4.18)$$

where $\Omega = \sum_{p=1}^{k} w_p^2$. Now $|J(A)| = \sqrt{\det(A^T A)}$. By finding the eigenvalues of $A^T A$ we obtain

$$|J(A)| = W_1^{t_1} W_2^{t_2} \sqrt{\frac{\Omega + t_1 W_1^2 + t_2 W_2^2}{\Omega}} \tag{4.4.19}$$

Now we define a random variable

$$Z = (\sum_{p=1}^{k} w_p^2) X_1 - \sum_{p=k+1}^{l} w_p^2 X_{p-k+1},$$

where $X_1, X_2, \ldots, X_{l-k+1}$ are independent random variables, with $X_p \sim HN(0, \frac{1}{2w_{p+k-1}^2})$, $2 \leq p \leq (l-k+1)$, as half-normal distributed random variables and $X_1 \sim N(0, \frac{1}{2\sum_{p=1}^{k} w_p^2})$ as a normal distributed random variable. Then by inspection, (4.4.16) is equal to $Cp_Z(0)$, 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

$$C = \frac{\sqrt{\pi}^{l-k+1}}{2^{l-k}} \prod_{q=k+1}^{l} \frac{1}{w_q} \sqrt{\sum_{p=1}^{k} w_p^2}\ |J(A)| = \frac{\sqrt{\pi}^{l-k+1}}{2^{l-k}} \sqrt{(k_1+t_1)W_1^2 + (k_2+t_2)W_2^2}. \tag{4.4.20}$$

Combining these results, the statement of Theorem 4.4.3 follows to be true.

4.4.3 Computation of External Angle

Theorem 4.4.4. *The external angle $\gamma(\mathcal{G}, SP_{\mathbf{w}}) = \gamma(d_1, d_2)$ between the face $\mathcal{G}$ and $SP_{\mathbf{w}}$, where G is supported on the set L with $|L \cap K_1| = d_1$ and $|L \cap K_2| = d_2$ is given by*

$$\gamma(d_1, d_2) = \pi^{-\frac{n-l+1}{2}} 2^{n-l} \int_0^{\infty} e^{-x^2} \left(\int_0^{\frac{W_1 x}{\xi(d_1,d_2)}} e^{-y^2}\, dy \right)^{r_1} \left(\int_0^{\frac{W_2 x}{\xi(d_1,d_2)}} e^{-y^2}\, dy \right)^{r_2} dx, \tag{4.4.21}$$

where $\xi(d_1, d_2) = \sqrt{\sum_{i \in L} w_i^2} = \sqrt{d_1 W_1^2 + d_2 W_2^2}$ $r_1 = n_1 - d_1$ $r_2 = n_2 - d_2$.

Proof. Without loss of generality, assume $L = \{n-l+1, n-l+2, \ldots, n\}$ and consider

the $(l-1)$-dimensional face,

$$\mathcal{G} = \text{conv}\{\frac{\varepsilon_{n-l+1}}{w_{n-l+1}}, \ldots, \frac{\varepsilon_{n-k}}{w_{n-k}}, \frac{\varepsilon_{n-k+1}}{w_{n-k+1}}, \ldots, \frac{\varepsilon_n}{w_n}\}$$

of the skewed crosspolytope SP. The 2^{n-l} outward normal vectors of the supporting hyperplanes of the facets containing $\mathcal{G}$ are given by

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

Then the outward normal cone $c(\mathcal{G}, \text{SP})$ at the face $\mathcal{G}$ is the positive hull of these normal vectors. Thus

$$\begin{aligned}\int_{c(\mathcal{G},\text{SP})} e^{-\|x\|^2}\, dx &= \gamma(\mathcal{G}, SP) V_{n-l}(S^{n-l}) \int_0^\infty e^{-r^2} r^{n-l}\, dx \\ &= \gamma(\mathcal{G}, \text{SP}).\pi^{(n-l+1)/2}, \end{aligned} \tag{4.4.22}$$

where $V_{n-l}(S^{n-l})$ is the spherical volume of the $(n-l)$-dimensional 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 $f(x_1,\ \ldots,\ x_{n-l+1}) : U \to c(\mathcal{G}, \text{SP})$ to be the linear and bijective map

$$f(x_1,\ \ldots,\ x_{n-l+1}) \ = \ \sum_{i=1}^{n-l} x_i \varepsilon_i + \sum_{i=n-l+1}^{n} w_i x_{n-l+1} \varepsilon_i.$$

Then

$$\begin{aligned}&\int_{c(\mathcal{G},\text{SP})} e^{-\|x'\|^2}\, dx' = |J(A)| \int_U e^{-\|f(x)\|^2}\, dx \\ &= |J(A)| \int_0^\infty \int_{-w_1 x_{n-l+1}}^{w_1 x_{n-l+1}} \cdots \int_{-w_{n-l} x_{n-l+1}}^{w_{n-l} x_{n-l+1}} e^{-x_1^2 - \cdots - x_{n-l}^2 - (\sum_{i=n-l+1}^n w_i^2) x_{n-l+1}^2}\, dx_1 \cdots dx_{n-l+1} \\ &= |J(A)| \int_0^\infty e^{-(\sum_{i=n-l+1}^n w_i^2) x^2} \left(\int_{-W_1 x}^{W_1 x} e^{-y^2}\, dy\right)^{n_1 - d_1} \left(\int_{-W_2 x}^{W_2 x} e^{-y^2}\, dy\right)^{n_2 - d_2} dx. \end{aligned} \tag{4.4.23}$$

A is the $n \times (n-l+1)$ change of variable matrix given by

$$A = \begin{bmatrix} \mathbf{I_{n-l}} & 0 \\ 0 & \mathbf{w}_L \end{bmatrix},$$

where $\mathbf{w}_L = (w_{n-l+1}, w_{n-l+2}, \ldots, w_n)^T$. Therefore

$$J(A) = \sqrt{\det(A^T A)} = \sqrt{d_1 W_1^2 + d_2 W_2^2}.$$

Using this equation and a change of variable for x (replace ξx with x) in (4.4.23), along with (4.4.22), completes the proof.

■

4.4.4 Derivation of the Critical δ_c Threshold

We begin by summarizing all the results that we have proven so far and what we are trying to show next. We proved in Section 4.4.1 that the probability of the failure event is bounded by the formula,

$$\mathbb{P}(E^c) \leq O(e^{-c_\epsilon n}) + \sum_{k_1=n_1(P_1-\epsilon)}^{n_1(P_1+\epsilon)} \sum_{k_2=n_2(P_2-\epsilon)}^{n_2(P_2+\epsilon)} \left(\sum_{\substack{0 \leq t_1 \leq n_1 - k_1 \\ 0 \leq t_2 \leq n_2 - k_2 \\ t_1 + t_2 > m - k_1 - k_2 + 1}} 2^{t_1+t_2+1} \binom{n_1 - k_1}{t_1} \binom{n_2 - k_2}{t_2} \beta(k_1, k_2 | t_1, t_2) \gamma(t_1 + k_1, t_2 + k_2) \right). \tag{4.4.24}$$

Furthermore, in Sections 4.4.2 and 4.4.3, we gave respectively expressions for the internal angle $\beta(t_1, t_2 | k_1, k_2)$ and external angle $\gamma(t_1 + k_1, t_2, k_2)$ terms that appear in the upper bound for the probability of misdetection in (4.4.24). Now our objective is to show that the R.H.S of (4.4.24) will exponentially grow 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 (4.4.24), 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 $\epsilon \to 0$,

$$\frac{1}{n}\log\left(2^{t_1+t_2+1}\binom{n_1-k_1}{t_1}\binom{n_2-k_2}{t_2}\right) \to$$
$$\log 2\left(\gamma_1(1-P_1)H(\frac{t_1'}{\gamma_1(1-P_1)}) + \gamma_2(1-P_2)H(\frac{t_2'}{\gamma_2(1-P_2)}) + t_1' + t_2'\right) \tag{4.4.25}$$

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

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

1. Let $g(x) = \frac{2}{\sqrt{\pi}}e^{-\frac{x^2}{2}}$, $G(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-y^2}dy$. Also define $C = (t_1' + \gamma_1 P_1) + W^2(t_2' + \gamma_2 P_2)$, $D_1 = \gamma_1(1-P_1) - t_1'$ and $D_2 = \gamma_2(1-P_2) - t_2'$. Let x_0 be the unique solution to x of the following:

$$2C - \frac{g(x)D_1}{xG(x)} - \frac{Wg(Wx)D_2}{xG(Wx)} = 0.$$

Define

$$\psi_{ext}(t_1', t_2') = Cx_0^2 - D_1\log G(x_0) - D_2\log G(Wx_0). \tag{4.4.26}$$

2. Let $b = \frac{t_1' + W^2 t_2'}{t_1' + t_2'}$ and $\varphi(.)$ and $\Phi(.)$ be the standard Gaussian pdf and cdf functions respectively. Also let $\Omega' = \gamma_1 P_1 + W^2\gamma_2 P_2$ and $Q(s) = \frac{t_1'\varphi(s)}{(t_1'+t_2')\Phi(s)} + \frac{Wt_2'\varphi(Ws)}{(t_1'+t_2')\Phi(Ws)}$. Define the function $\hat{M}(s) = -\frac{s}{Q(s)}$ and solve for s in $\hat{M}(s) = \frac{t_1'+t_2'}{(t_1'+t_2')b+\Omega'}$. Let the unique solution be s^* and set $y = s^*(b - \frac{1}{\hat{M}(s^*)})$. Compute the rate function $\Lambda^*(y) = sy - \frac{t_1'}{t_1'+t_2'}\Lambda_1(s) - \frac{t_2'}{t_1'+t_2'}\Lambda_1(Ws)$ 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:

$$\psi_{int}(t_1', t_2') = (\Lambda^*(y) + \frac{t_1' + t_2'}{2\Omega'}y^2 + \log 2)(t_1' + t_2'). \tag{4.4.27}$$

and we prove the following lemmas

Lemma 4.4.3. *Fix δ, $\epsilon > 0$. There exists a finite number $n_0(\delta, \epsilon)$ such that*

$$\frac{1}{n}\log(\gamma(t_1 + k_1, t_2 + k_2)) < -\psi_{ext}(t_1', t_2') + \epsilon, \tag{4.4.28}$$

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

Similarly, we have

Lemma 4.4.4. *Fix δ, $\epsilon > 0$. There exists a finite number $n_1(\delta, \epsilon)$ such that*

$$\frac{1}{n}\log(\beta(t_1, t_2|k_1, k_2)) < -\psi_{int}(t_1', t_2') + \epsilon, \tag{4.4.29}$$

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

These lemmas are proven in the Appendix 4.6.2 and 4.6.3 using Laplace method and large deviation method. Combining Lemmas 4.4.3 and 4.4.4, Eq. (4.4.25) and the bound in (4.4.24) we readily get the critical bound for δ_c as in the Theorem 4.3.3. We simply repeat the statement here once more.

Theorem 4.4.5. *If γ_1, γ_2, P_1, P_2, W_1 and W_2 are fixed, there exists a critical threshold $\delta_c = \delta_c(P_1, P_2, \frac{W_2}{W_1})$ such that if $\delta = \frac{m}{n} \geq \delta_c$, then the R.H.S of (4.3.2) (the upper bound on the probability of failure) decays exponentially to zero as $n \to \infty$. Furthermore, δ_c is given by*

$$\delta_c = \min\{\delta \mid \psi_{com}(t_1', t_2') - \psi_{int}(t_1', t_2') - \psi_{ext}(t_1', t_2') < 0 \ \forall\ 0 \leq t_1' \leq \gamma_1(1 - P_1), \\ 0 \leq t_2' \leq \gamma_2(1 - P_2), t_1' + t_2' > \delta - \gamma_1 P_1 - \gamma_2 P_2\}$$

where ψ_{com}, ψ_{ext}, ψ_{int} are computed as in (4.4.25), (4.4.27) and (4.4.26) respectively.

4.5 Simulation Results

We demonstrate by some examples that appropriate weights can boost the recovery percentage. 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 4.5a shows one such comparison for $P_2 = 0.05$ and different values of w_2. Note that the optimal value of w_2 varies as P_1 changes. Figure 4.5b 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 4.5a for each single value of P_1. Figure 4.6 shows the result of simulations in another setting where $P_2 = 0.1$ and $m = 0.75n$ (similar to the setting of the previous section). It is clear from the figure that the recovery success threshold for P_1 has been shifted higher when using weighted ℓ_1 minimization rather than standard ℓ_1 minimization. Note that this result very well matches the theoretical results of Figures 4.3 and 4.4.

4.6 Appendix. Proof of Important Lemmas

4.6.1 Proof of Lemma 4.4.2

Without loss of generality, assume that $K = 1, 2, ..., k$ and F is a $(k-1)$-dimensional face with k vertices as $\frac{\varepsilon_p}{w_p}, 1 \leq p \leq k$, where ε_p is the n-dimensional standard unit vector with the p-th element as '1'; and also assume that the $(l-1)$-dimensional face G be the convex hull of the l vertices: $\frac{\varepsilon_p}{w_p}, 1 \leq p \leq l$. Then the cone $\mathrm{Con}_{F,G}$ formed by observing the $(l-1)$-dimensional face G of the skewed crosspolytope SP from an interior point x^F of the face F is the positive cone of the vectors:

$$\frac{\varepsilon_j}{w_j} - \frac{\varepsilon_i}{w_i}, \quad \text{for all} \quad j \in J \backslash K, \ i \in K, \tag{4.6.1}$$

and also the vectors

$$\frac{\varepsilon_{i_1}}{w_{i_1}} - \frac{\varepsilon_{i_2}}{w_{i_2}}, \quad \text{for all} \quad i_1 \in K,\ i_2 \in K, \tag{4.6.2}$$

where $L = \{1, 2, ..., l\}$ is the support set for the face G.

So the cone $\mathrm{Con}_{F,G}$ is the direct sum of the linear hull $L_F = \mathrm{lin}\{F - x^F\}$ formed by the vectors in (4.6.2) and the cone $\mathrm{Con}_{F^\perp,G} = \mathrm{Con}_{F,G} \bigcap L_F^\perp$, where $L_F^\perp$ is the orthogonal complement to the linear subspace L_F. Then $\mathrm{Con}_{F^\perp,G}$ has the same (relative) spherical volume as $\mathrm{Con}_{F,G}$, and by definition the internal angle $\beta(F, G)$ is the relative spherical volume of the cone $\mathrm{Con}_{F,G}$. Now let us analyze the structure of $\mathrm{Con}_{F^\perp,G}$. We notice that the vector

$$\varepsilon_0 = \sum_{p=1}^{k} w_p \varepsilon_p$$

is in the linear space $L_F^\perp$ and is also the only such a vector (up to linear scaling) supported on K. Thus a vector $\mathbf{x}$ in the positive cone $\mathrm{Con}_{F^\perp,G}$ must take the form

$$-\sum_{i=1}^{k} b_i \times e_i + \sum_{i=k+1}^{l} b_i \times e_i, \tag{4.6.3}$$

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

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

Now that we have identified $\mathrm{Con}_{F^\perp,G}$ we try to calculate its relative spherical volume with respect to the sphere surface S^{l-k-1} to derive $\beta(F, G)$. First, we notice that $\mathrm{Con}_{F^\perp,G}$ is a $(l-k)$-dimensional cone. Also, all the vectors $(x_1, \ldots, x_n)$ in the

cone $\text{Con}_{F^\perp,G}$ take the form in (4.6.3). From [Had79],

$$\int_{\text{Con}_{F^\perp,G}} e^{-\|\mathbf{x}\|^2}\, d\mathbf{x} = \beta(F,G)V_{l-k-1}(S^{l-k-1}) \times \int_0^\infty e^{-r^2} r^{l-k-1}\, dx = \beta(F,G)\cdot \pi^{(l-k)/2}, \tag{4.6.4}$$

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.

4.6.2 Proof of Lemma 4.4.3

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 $G(x) = \mathbb{P}\{X \le x\}$. Since X has density function $g(x) = \frac{2}{\sqrt{\pi}}\exp(-x^2)$, we know that

$$G(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-y^2}\, dy; \tag{4.6.5}$$

and so G is just the classical error function *erf(·)*. We now justify the external angle exponent computations in Theorem 4.3.3 and Lemma 4.4.3 using Laplace methods [DT05a]. Using the same set of notations as in Theorem 4.3.3, let $t_1 = t_1' n$, $t_2 = t_2' n$. Also define $C = (t_1' + \gamma_1 P_1) + W^2(t_2' + \gamma_2 P_2)$, $D_1 = \gamma_1(1-P_1) - t_1'$ and $D_2 = \gamma_2(1-P_2) - t_2'$. Let x_0 be the unique solution to x of the following:

$$2C - \frac{g(x)D_1}{xG(x)} - \frac{Wg(Wx)D_2}{xG(Wx)} = 0 \tag{4.6.6}$$

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 $\frac{g(x)D_1}{xG(x)} + \frac{Wg(Wx)D_2}{xG(Wx)}$ is one-one on the positive axis, and x_0 is a well-defined function of t_1' and t_2'. Hence, we denote it

as $x_0(t_1', t_2')$.

Then

$$\psi_{ext}(t_1', t_2') = Cx_0^2 - D_1 \log G(x_0) - D_2 \log G(Wx_0). \tag{4.6.7}$$

.

To prove Lemma 4.4.3, we start from the explicit integral formula

$$\gamma(d_1, d_2) = \pi^{-\frac{n-l+1}{2}} 2^{n-l} \int_0^\infty e^{-x^2} \left(\int_0^{\frac{W_1 x}{\xi(d_1,d_2)}} e^{-y^2}\, dy \right)^{r_1} \left(\int_0^{\frac{W_2 x}{\xi(d_1,d_2)}} e^{-y^2}\, dy \right)^{r_2} dx, \tag{4.6.8}$$

After a changing of integral variables (noticing that $W_1 = 1$, $W_2 = W$, $\frac{n_1 - d_1}{n} = D_1$, and $\frac{n_2 - d_2}{n} = D_2$), we have

$$\gamma(t_1 + k_1, t_2 + k_2) = \sqrt{Cn/\pi} \int_0^\infty e^{-n(Cx^2 - D_1 \log(G(x)) - D_2 \log(G(Wx))}\, dx. \tag{4.6.9}$$

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

$$f_{t_1', t_2', n} = e^{-n\psi_{t_1', t_2'}(y)} \cdot \sqrt{Cn/\pi}, \tag{4.6.10}$$

with

$$\psi_{t_1', t_2'}(y) = Cy^2 - D_1 \log(G(y)) - D_2 \log(G(Wy)).$$

We note that the function $\psi_{t_1', t_2'}$ is smooth and convex. Applying Laplaces method to $\psi_{t_1', t_2'}$, but taking care about regularity conditions and remainders as in [DT05a], gives a result with the uniformity in (t_1', t_2').

Lemma 4.6.1. *For t_1', t_2', let $x_0(t_1', t_2')$ denotes 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 $\delta, \eta > 0$,

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

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 $x_o(t_1', t_2')$ is exactly the same $x_0(t_1', t_2')$ defined earlier in (4.6.6) and the corresponding minimum value is the same as the defined exponent Ψ_{ext}:

$$\psi_{ext}(t_1', t_2') = \psi_{t_1', t_2'}(x_{t_1'}, x_{t_2'}). \tag{4.6.11}$$

In fact, we can derive Lemma 4.4.3 from Lemma 4.6.1. 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 4.4.3, there is a largest $\eta_\epsilon < 1$ such that as long as $t_1' + t_2' + \rho_1 + \rho_2 > \eta_\epsilon$, $\psi_{ext}(t_1', t_2') < \epsilon$. Note that $\gamma(\mathcal{G}, \mathrm{SP}) \le 1$, so that for $t_1' + t_2' + \rho_1 + \rho_2 > \eta_\epsilon$,

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

for $n \ge 1$. Applying the uniformity in t_1', t_2' given in Lemma 4.6.1, we have as $n \to \infty$, uniformly over the feasible region for t_1', t_2',

$$n^{-1} \log(\gamma(t_1 + k_1, t_2 + k_2)) \le -\psi_{ext}(t_1', t_2') + o(1). \tag{4.6.12}$$

Then Lemma 4.4.3 follows.

4.6.3 Proof of Lemma 4.4.4

Recall Theorem 4.4.3. By applying the large deviation techniques as in [DT05a], we have

$$p_Z(0) \le \frac{2}{\sqrt{\pi}} \times \frac{1}{\sqrt{\Omega}} \cdot \left(\int_0^{\mu_{m'}} v e^{-v^2 - m' \Lambda^*(\frac{\sqrt{2\Omega}}{m'} v)} \, dv + e^{-\mu_{m'}^2} \right), \tag{4.6.13}$$

where Ω is the same as defined in Section 4.4.2, $W_1 = 1$, $W_2 = W$, $m' = t_1 + t_2$, $\mu_{m'} = (t_1 + t_2 W)\sqrt{\frac{1}{\pi\Omega}}$ is the expectation of $\frac{1}{\sqrt{\Omega'}}(W_1^2 \sum_{i=1}^{t_1} X_i' - W_2^2 \sum_{i=1}^{t_2} X_i'')$, ($X_i'$ and X_i'' are defined as in Theorem 4.4.3), 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

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

$$\Lambda_2(s) = \Lambda_1(Ws).$$

In fact, the second term in the sum can be argued to be negligible [DT05a]. And after changing variables $y = \frac{\sqrt{2\Omega}}{m'}v$, we know that the first term of (4.6.13) is upperbounded by

$$\frac{2}{\sqrt{\pi}} \cdot \frac{1}{\sqrt{\Omega}} \cdot \frac{m'^2}{2\Omega} \cdot \int_0^{\frac{t_1+t_2W}{t_1+t_2}\sqrt{2/\pi}} y e^{-m'(\frac{m'}{2\Omega})y^2 - m'\Lambda^*(y)}\, dy. \tag{4.6.14}$$

As we know, m' in the exponent of (4.6.14) 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 4.6.1 and take care of the uniformity, we have the following lemma.

Lemma 4.6.2. *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

$$m'^{-1} \sup_{\frac{(t1+t_2)}{n} \in [\delta-\rho_1-\rho_2,1]} \log(R_{m'}(t_1,t_2)) = o(1) \quad as \quad m' \to \infty.$$

Now in order to find a lower bound on the decay exponent for $p_Z(0)$,(ultimately the decay exponent $\psi_{int}(t_1', t_2')$), we need to focus on finding the minimizer $y_{t_1,t_2}*$ for $(\frac{m'}{2\Omega})y^2 + \Lambda^*(y)$. In 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

$$s = -\frac{m'}{\Omega}y. \tag{4.6.15}$$

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

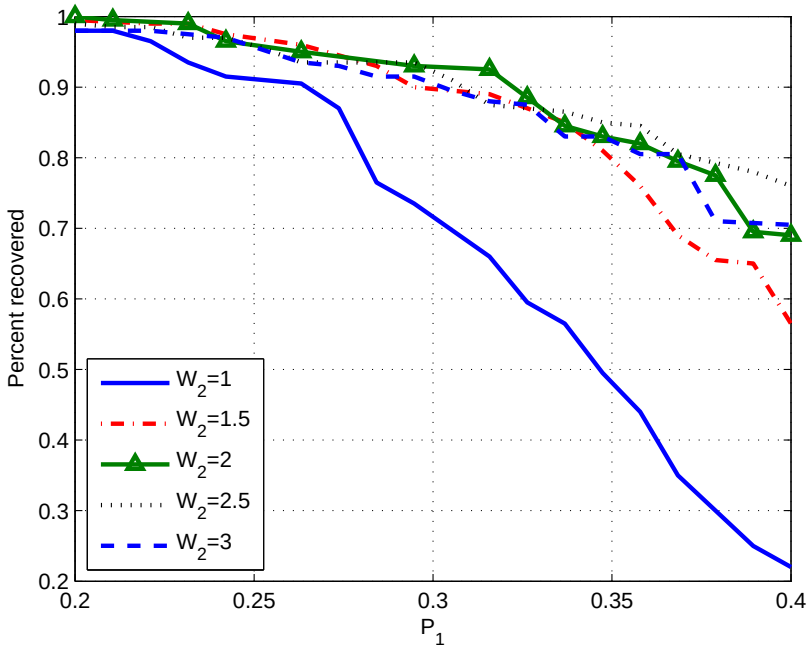
$$y = \frac{t_1}{t_1 + t_2}\Lambda_1'(s) + \frac{t_2}{t_1 + t_2}\Lambda_2'(s), \tag{4.6.16}$$

namely, (by writing out (4.6.16)),

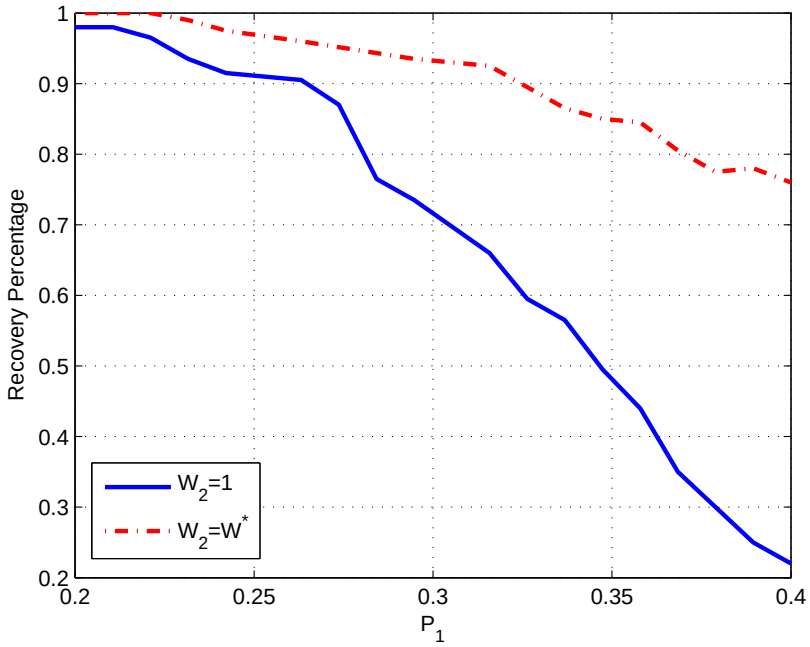
$$y = \frac{t_1 + W^2 t_2}{t_1 + t_2}s + Q(s), \tag{4.6.17}$$

where $Q(s)$ is defined as in Theorem 4.4.3.

By combining (4.6.15) and (4.6.16), we can solve for the s and y, thus resulting in the decay exponent for $\psi_i nt(t_1', t_2')$ as calculated in Theorem 4.4.3.



(a)



(b)

Figure 4.5: Successful recovery percentage for weighted ℓ_1 minimization with different weights in a nonuniform sparse setting. $P_2 = 0.05$ and $m = 0.5n$

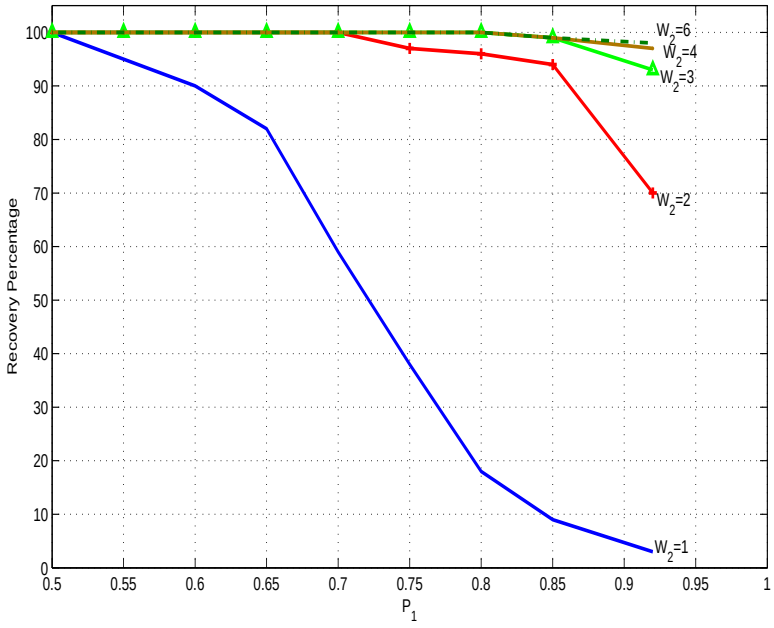


Figure 4.6: Successful recovery percentage for different weights. $P_2 = 0.1$ and $m = 0.75n$

Chapter 5

An Analysis for Iterative Reweighted ℓ_1 Minimization Algorithm

It is now well understood that ℓ_1 minimization algorithm is able to recover sparse signals from incomplete measurements [CT05, Don06c, DT06b] and sharp recoverable sparsity thresholds have also been obtained for the ℓ_1 minimization algorithm. However, even though iterative reweighted ℓ_1 minimization algorithms or related algorithms have been empirically observed to boost the recoverable sparsity thresholds for certain types of signals, no rigorous theoretical results have been established to prove this fact. In this chapter, we try to provide a theoretical foundation for analyzing the iterative reweighted ℓ_1 algorithms. In particular, we show that for a nontrivial class of signals, the iterative reweighted ℓ_1 minimization can indeed deliver recoverable sparsity thresholds larger than that given in [Don06c, DT06b]. Our results are based on a high-dimensional geometrical analysis (Grassmann angle analysis) of the null-space characterization for ℓ_1 minimization and weighted ℓ_1 minimization algorithms.

5.1 Introduction

In this chapter we are interested in compressed sensing problems. Namely, we would like to find $\mathbf{x}$ such that

$$A\mathbf{x} = \mathbf{y}, \tag{5.1.1}$$

where A is an $m \times n$ ($m < n$) measurement matrix, $\mathbf{y}$ is a $m \times 1$ measurement vector and $\mathbf{x}$ is an $n \times 1$ unknown vector with only k ($k < m$) nonzero components. We will further assume that the number of the measurements is $m = \delta n$ and the number of the nonzero components of $\mathbf{x}$ is $k = \zeta n$, where $0 < \zeta < 1$ and $0 < \delta < 1$ are constants independent of n (clearly, $\delta > \zeta$).

A particular way of solving (5.1.1) which has recently generated a large amount of research is called ℓ_1-optimization (basis pursuit) [CT05]. It proposes solving the following problem

$$\begin{aligned} \min \quad & \|\mathbf{x}\|_1 \\ \text{subject to} \quad & A\mathbf{x} = \mathbf{y}. \end{aligned} \tag{5.1.2}$$

Quite remarkably in [CT05] the authors were able to show that if the number of the measurements is $m = \delta n$ and if the matrix A satisfies a special property called the restricted isometry property (RIP), then any unknown vector $\mathbf{x}$ with no more than $k = \zeta n$ (where ζ is an absolute constant which is a function of δ, but independent of n, and explicitly bounded in [CT05]) non-zero elements can be recovered by solving (5.1.2). Instead of characterizing the $m \times n$ matrix A through the RIP condition, in [Don06c, DT06b] the authors assume that A constitutes a k-neighborly polytope. It turns out (as shown in [Don06c]) that this characterization of the matrix A is in fact a necessary and sufficient condition for (5.1.2) to produce the solution of (5.1.1). Furthermore, using the results of [VS92][AS92][KBH99], it can be shown that if the matrix A has i.i.d. zero-mean Gaussian entries with overwhelming probability it also constitutes a k-neighborly polytope. The precise relation between m and k in order for this to happen is characterized in [Don06c] as well.

In this chapter we will be interested in providing the theoretical guarantees for the emerging iterative reweighted ℓ_1 algorithms [CWB08]. These algorithms iteratively updated weights for each element of $\mathbf{x}$ in the objective function of ℓ_1 minimization, based on the decoding results from previous iterations. Experiments showed that the iterative reweighted ℓ_1 algorithms can greatly enhance the recoverable sparsity

threshold for certain types of signals, for example, sparse signals with Gaussian entries. However, no rigorous theoretical results have been provided for establishing this phenomenon. To quote from [CWB08], "any result quantifying the improvement of the reweighted algorithm for special classes of sparse or nearly sparse signals would be significant." In this chapter, we try to provide a theoretical foundation for analyzing the iterative reweighted ℓ_1 algorithms. In particular, we show that for a nontrivial class of signals, (It is worth noting that empirically, the iterative reweighted ℓ_1 algorithms do not always improve the recoverable sparsity thresholds, for example, they often fail to improve the recoverable sparsity thresholds when the non-zero elements of the signals are "flat" [CWB08]), a modified iterative reweighted ℓ_1 minimization algorithm can indeed deliver recoverable sparsity thresholds larger than those given in [Don06c, DT06b] for unweighted ℓ_1 minimization algorithms. Our results are based on a high-dimensional geometrical analysis (Grassmann angle analysis) of the null-space characterization for ℓ_1 minimization and weighted ℓ_1 minimization algorithms. The main idea is to show that the preceding ℓ_1 minimization iterations can provide certain information about the support set of the signals and this support set information can be properly taken advantage of to perfectly recover the signals even though the sparsity of the signal $\mathbf{x}$ itself is large.

This chapter is structured as follows. In Section 5.2, we present the iterative reweighted ℓ_1 algorithm for analysis. The signal model for $\mathbf{x}$ will be given in Section 5.3. In Section 5.4 and Section 5.5, we will show how the iterative reweighted ℓ_1 minimization algorithm can indeed improve recoverable sparsity thresholds. Numerical results will be given in Section 5.6.

5.2 The Modified Iterative Reweighted ℓ_1 Minimization Algorithm

Let w_i^t, $i = 1, ..., n$, denote the weights for the i-th element $\mathbf{x}_i$ of $\mathbf{x}$ in the t-th iteration of the iterative reweighted ℓ_1 minimization algorithm and let $\mathbf{W}^t$ be the diagonal

matrix with $w_1^t, w_2^t, ..., w_n^t$ on the diagonal. In the paper [CWB08], the following iterative reweighted ℓ_1 minimization algorithm is presented:

Algorithm 6 [CWB08]

1. Set the iteration count t to zero and $w_i^t = 1$, $i = 1, ..., n$.
2. Solve the weighted ℓ_1 minimization problem
$$\mathbf{x}^t = \arg\min \|\mathbf{W}^t\mathbf{x}\|_1 \text{ subject to } \mathbf{y} = A\mathbf{x}. \tag{5.2.1}$$
3. Update the weights: for each $i = 1, ..., n$,
$$w_i^{t+1} = \frac{1}{|\mathbf{x}_i^t| + \epsilon'}, \tag{5.2.2}$$
where ϵ' is a tunable positive number.
4. Terminate on convergence or when t attains a specified maximum number of iterations $t_{\max}$. Otherwise, increment t and go to step 2.

For the sake of tractable analysis, we will give another iterative reweighted ℓ_1 minimization algorithm , but it still captures the essence of the reweighted ℓ_1 algorithm presented in [CWB08]. In our modified algorithm, we only do two ℓ_1 minimization programming, namely we stop at the time index $t = 1$.

This modified algorithm is certainly different from the algorithm from [CWB08], but the important thing is that both algorithms assign bigger weights to those elements of $\mathbf{x}$ which are more likely to be 0.

5.3 Signal Model for x

In this chapter, we consider the following model for the n-dimensional sparse signal $\mathbf{x}$. First of all, we assume that there exists a set $K \subset \{1, 2, ..., n\}$ with cardinality $|K| = (1 - \epsilon)\rho_F(\delta)\delta n$ such that each of the elements of $\mathbf{x}$ over the set K is large in amplitude. W.L.O.G., those elements are assumed to be all larger than $a_1 > 0$. For a given signal $\mathbf{x}$, one might take such set K to be the set corresponding to the

Algorithm 7 The Modified Iterative Reweighted ℓ_1 Minimization Algorithm

1. Set the iteration count t to zero and $w_i^t = 1$, $i = 1, ..., n$.

2. Solve the weighted ℓ_1 minimization problem

$$\mathbf{x}^t = \arg\min \|\mathbf{W}^t \mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{y} = A\mathbf{x}. \tag{5.2.3}$$

3. Update the weights: find the index set $K' \subset \{1, 2, ..., n\}$ which corresponds to the largest $(1-\epsilon)\rho_F(\delta)\delta n$ elements of $\mathbf{x}^0$ in amplitudes, where $0 < \epsilon < 1$ is a specified parameter and $\rho_F(\delta)$ is the weak threshold for perfect recovery defined in [Don06c] using ℓ_1 minimization (thus $\zeta = \rho_F(\delta)\delta$ is the weak sparsity threshold). Then assign the weight $W_1 = 1$ to those w_i^{t+1} corresponding to the set K' and assign the weight $W_2 = W$, $W > 1$, to those w_i^{t+1} corresponding to the complementary set $\bar{K}' = \{1, 2, ..., n\} \setminus K'$.

4. Terminate on convergence or when $t = 1$. Otherwise, increment t and go to step 2.

$(1-\epsilon)\rho_F(\delta)\delta n$ largest elements of $\mathbf{x}$ in amplitude.

Secondly, (let $\bar{K} = \{1, 2, ..., n\} \setminus K$), we assume that the ℓ_1 norm of $\mathbf{x}$ over the set $\bar{K}$, denoted by $\|\mathbf{x}_{\bar{K}}\|_1$, is upperbounded by Δ, though Δ is allowed to take a non-diminishing portion of the total ℓ_1 norm $\|\mathbf{x}\|_1$ as $n \to \infty$. We further denote the support set of $\mathbf{x}$ as K_{total} and its complement as $\bar{K}_{\text{total}}$. The sparsity of the signal $\mathbf{x}$, namely the total number of nonzero elements in the signal $\mathbf{x}$ is then $|K_{\text{total}}| = k_{\text{total}} = \xi n$, where ξ can be above the weak sparsity threshold $\zeta = \rho_F(\delta)\delta$ achievable using the ℓ_1 algorithm.

In the following sections, we will show that if certain conditions on a_1, Δ and the measurement matrix A are satisfied, we will be able to recover perfectly the signal $\mathbf{x}$ using Algorithm 7 even though its sparsity level is above the sparsity threshold for ℓ_1 minimization. Intuitively, this is because the weighted ℓ_1 minimization puts larger weights on the signal elements which are more likely to be zero, and puts smaller weights on the signal support set, thus promoting sparsity at the right positions. In order to achieve this, we need some prior information about the support set of $\mathbf{x}$, which can be obtained from the decoding results in previous iterations. We will first

argue that the equal-weighted ℓ_1 minimization of Algorithm 7 can sometimes provide very good information about the support set of signal $\mathbf{x}$.

5.4 Estimating the Support Set from the ℓ_1 Minimization

Since the set K' corresponds to the largest elements in the decoding results of ℓ_1 minimization, one might guess that most of the elements in K' are also in the support set K_{total}. The goal of this section is to get an upper bound on the cardinality of the set $\bar{K}_{\text{total}} \cap K'$, namely the number of zero elements of $\mathbf{x}$ over the set K' . To this end, we will first give the notion of "weak" robustness for the ℓ_1 minimization.

Let K be fixed and $\mathbf{x}_K$, the value of $\mathbf{x}$ on this set, be also fixed. Then the solution produced by (5.1.2), $\hat{\mathbf{x}}$, will be called *weakly robust* if, for some $C > 1$ and all possible $\mathbf{x}_{\bar{K}}$, it holds that

$$\|(\mathbf{x} - \hat{\mathbf{x}})_{\bar{K}}\|_1 \leq \frac{2C}{C-1}\|\mathbf{x}_{\bar{K}}\|_1,$$

and

$$\|\mathbf{x}_K\|_1 - \|\hat{\mathbf{x}}_K\|_1 \leq \frac{2}{C-1}\|\mathbf{x}_{\bar{K}}\|_1.$$

The above "weak" notion of robustness allows us to bound the error $\|\mathbf{x} - \hat{\mathbf{x}}\|_1$ in the following way. If the matrix A_K, obtained by retaining only those columns of A that are indexed by K, has full column rank, then the quantity

$$\kappa = \max_{A\mathbf{w}=0, \mathbf{w}\neq 0} \frac{\|\mathbf{w}_K\|_1}{\|\mathbf{w}_{\bar{K}}\|_1},$$

must be finite ($\kappa < \infty$). In particular, since $\mathbf{x} - \hat{\mathbf{x}}$ is in the null space of A ($\mathbf{y} = A\mathbf{x} =$

$A\hat{\mathbf{x}}$), we have

$$\begin{aligned} \|\mathbf{x}-\hat{\mathbf{x}}\|_1 &= \|(\mathbf{x}-\hat{\mathbf{x}})_K\|_1 + \|(\mathbf{x}-\hat{\mathbf{x}})_{\bar{K}}\|_1 \\ &\leq (1+\kappa)\|(\mathbf{x}-\hat{\mathbf{x}})_{\bar{K}}\|_1 \\ &\leq \frac{2C(1+\kappa)}{C-1}\|\mathbf{x}_{\bar{K}}\|_1, \end{aligned}$$

thus bounding the recovery error. We can now give necessary and sufficient conditions on the measurement matrix A to satisfy the notion of weak robustness for ℓ_1 minimization.

Theorem 5.4.1. *For a given $C > 1$, support set K, and $\mathbf{x}_K$, the solution $\hat{\mathbf{x}}$ produced by (5.1.2) will be* weakly robust *if, and only if, $\forall \mathbf{w} \in \boldsymbol{R}^n$ such that $A\mathbf{w} = 0$, we have*

$$\|\mathbf{x}_K + \mathbf{w}_K\|_1 + \|\frac{\mathbf{w}_{\bar{K}}}{C}\|_1 \geq \|\mathbf{x}_K\|_1. \tag{5.4.1}$$

Proof. Sufficiency: Let $\mathbf{w} = \hat{\mathbf{x}} - \mathbf{x}$, for which $A\mathbf{w} = A(\hat{\mathbf{x}} - \mathbf{x}) = 0$. Since $\hat{\mathbf{x}}$ is the minimum ℓ_1 norm solution, we have $\|\mathbf{x}\|_1 \geq \|\hat{\mathbf{x}}\|_1 = \|\mathbf{x}+\mathbf{w}\|_1$, and therefore $\|\mathbf{x}_K\|_1 + \|\mathbf{x}_{\bar{K}}\|_1 \geq \|\hat{\mathbf{x}}_K\|_1 + \|\hat{\mathbf{x}}_{\bar{K}}\|_1$. Thus,

$$\begin{aligned} \|\mathbf{x}_K\|_1 - \|\mathbf{x}_K + \mathbf{w}_K\|_1 &\geq \|\mathbf{w}_{\bar{K}} + \mathbf{x}_{\bar{K}}\|_1 - \|\mathbf{x}_{\bar{K}}\|_1 \\ &\geq \|\mathbf{w}_{\bar{K}}\|_1 - 2\|\mathbf{x}_{\bar{K}}\|_1. \end{aligned}$$

But the condition (5.4.1) guarantees that

$$\|\mathbf{w}_{\bar{K}}\|_1 \geq C(\|\mathbf{x}_K\|_1 - \|\mathbf{x}_K + \mathbf{w}_K\|_1),$$

so we have

$$\|\mathbf{w}_{\bar{K}}\|_1 \leq \frac{2C}{C-1}\|\mathbf{x}_{\bar{K}}\|_1,$$

and

$$\|\mathbf{x}_K\|_1 - \|\hat{\mathbf{x}}_K\|_1 \leq \frac{2}{C-1}\|\mathbf{x}_{\bar{K}}\|_1,$$

as desired.

Necessity: Since in the above proof of the sufficiency, equalities can be achieved in the triangular inequalities, the condition (5.4.1) is also a necessary condition for the weak robustness to hold for every $\mathbf{x}$. (Otherwise, for certain $\mathbf{x}$'s, there will be $\mathbf{x}' = \mathbf{x} + \mathbf{w}$ with $\|\mathbf{x}'\|_1 < \|\mathbf{x}\|_1$ while violating the respective robustness definitions. Also, such $\mathbf{x}'$ can be the solution to (5.1.2)). ■

We should remark (without proof for the interest of space) that for any $\delta > 0$, $0 < \epsilon < 1$, let $|K| = (1-\epsilon)\rho_F(\delta)\delta n$, and suppose each element of the measurement matrix A is sampled from i.i.d. Gaussian distribution, then there exists a constant $C > 1$ (as a function of δ and ϵ), such that the condition (5.4.1) is satisfied with overwhelming probability as the problem dimension $n \to \infty$. At the same time, the parameter κ defined above is upperbounded by a finite constant (independent of the problem dimension n) with overwhelming probability as $n \to \infty$. These claims can be shown by using the Grasamann angle approach for the balancedness property of random linear subspaces in [XH08].

In Algorithm 7, after equal-weighted ℓ_1 minimization, we pick the set K' corresponding to the $(1-\epsilon)\rho_F(\delta)\delta$ largest elements in amplitudes from the decoding result $\hat{\mathbf{x}}$ (namely $\mathbf{x}^0$ in the algorithm description) and assign the weights $W_1 = 1$ to the corresponding elements in the next iteration of reweighted ℓ_1 minimization. Now we can show that an overwhelming portion of the set K' are also in the support set K_{total} of $\mathbf{x}$ if the measurement matrix A satisfies the specified weak robustness property.

Theorem 5.4.2. *Supposed that we are given a signal vector $\mathbf{x} \in R^n$ satisfying the signal model defined in Section 5.3. Given $\delta > 0$, and a measurement matrix A which satisfies the weak robustness condition in (5.4.1) with its corresponding $C > 1$ and $\kappa < \infty$, then the set K' generated by the equal-weighted ℓ_1 minimization in Algorithm 2 contains at most $\frac{2C}{(C-1)\frac{a_1}{2}}\|\mathbf{x}_{\bar{K}}\|_1 + \frac{2C\kappa}{(C-1)\frac{a_1}{2}}\|\mathbf{x}_{\bar{K}}\|_1$ indices which are outside the support set of signal $\mathbf{x}$.*

Proof. Since the measurement matrix A satisfies the weak robustness condition for the set K and the signal $\mathbf{x}$,

$$\|(\mathbf{x}-\hat{\mathbf{x}})_{\bar{K}}\|_1 \leq \frac{2C}{C-1}\|\mathbf{x}_{\bar{K}}\|_1.$$

By the definition of the $\kappa < \infty$, namely,

$$\kappa = \max_{A\mathbf{w}=0,\mathbf{w}\neq 0} \frac{\|\mathbf{w}_K\|_1}{\|\mathbf{w}_{\bar{K}}\|_1},$$

we have

$$\|(\mathbf{x}-\hat{\mathbf{x}})_K\|_1 \quad \leq \quad \kappa\|(\mathbf{x}-\hat{\mathbf{x}})_{\bar{K}}\|_1.$$

Then there are at most $\frac{2C}{(C-1)\frac{a_1}{2}}\|\mathbf{x}_{\bar{K}}\|_1$ indices that are outside the support set of $\mathbf{x}$ but have amplitudes larger than $\frac{a_1}{2}$ in the corresponding positions of the decoding result $\hat{\mathbf{x}}$ from the equal-weighted ℓ_1 minimization algorithm. This bound follows easily from the facts that all such indices are in the set $\bar{K}$ and that $\|(\mathbf{x}-\hat{\mathbf{x}})_{\bar{K}}\|_1 \leq \frac{2C}{C-1}\|\mathbf{x}_{\bar{K}}\|_1$.

Similarly, there are at most $\frac{2C\kappa}{(C-1)\frac{a_1}{2}}\|\mathbf{x}_{\bar{K}}\|_1$ indices which are originally in the set K but now have corresponding amplitudes smaller than $\frac{a_1}{2}$ in the decoded result $\hat{\mathbf{x}}$ of the equal-weighted ℓ_1 algorithm.

Since the set K' corresponds to the largest $(1-\epsilon)\rho_F(\delta)\delta n$ elements of the signal $\hat{\mathbf{x}}$, by combining the previous two results, it is not hard to see that the number of indices which are outside the support set of $\mathbf{x}$ but are in the set K' is no bigger than$\frac{2C}{(C-1)\frac{a_1}{2}}\|\mathbf{x}_{\bar{K}}\|_1 + \frac{2C\kappa}{(C-1)\frac{a_1}{2}}\|\mathbf{x}_{\bar{K}}\|_1$.

■

As we can see, Theorem 5.4.2 provides useful information about the support set of the signal $\mathbf{x}$, which can be used in the analysis for the weighted ℓ_1 minimization using the null-space Grassmann Angle analysis approach for weighted ℓ_1 minimization algorithm [KXAH09a].

5.5 The Grassmann Angle Approach for the Reweighted ℓ_1 Minimization

In the previous work [KXAH09a], the authors have shown that by exploiting certain prior information about the original signal, it is possible to extend the threshold of sparsity factor for successful recovery beyond the original bounds of [Don06c, DT06b]. The authors proposed a nonuniform sparsity model in which the entries of the vector $\mathbf{x}$ can be considered as T different classes, where in the ith class, each entry is (independently from others) nonzero with probability P_i, and zero with probability $1-P_i$. The signals generated based on this model will have around $n_1P_1+\cdots+n_TP_T$ nonzero entries with high probability, where n_i is the size of the ith class. Examples of such signals arise in many applications as medical or natural imaging, satellite imaging, DNA micro-arrays, network monitoring and so on. They prove that provided such structural prior information is available about the signal, a proper *weighted* ℓ_1-minimization strictly outperforms the regular ℓ_1-minimization in recovering signals with some fixed average sparsity from underdetermined linear i.i.d. Gaussian measurements.

The detailed analysis in [KXAH09a] is only done for $T = 2$, and is based on the high dimensional geometrical interpretations of the constrained weighted ℓ_1-minimization problem:

$$\min_{A\mathbf{x}=\mathbf{y}} \sum_{i=1}^{n} w_i|\mathbf{x}_i|. \tag{5.5.1}$$

Let the two classes of entries be denoted by K_1 and K_2. Also, due to the partial symmetry, for any suboptimal set of weights $\{w_1,\ldots,\mathbf{w}_n\}$ we have the following,

$$\forall i \in \{1,2,\ldots,n\} \quad w_i = \begin{cases} W_1 \ if \ i \in K_1 \\ W_2 \ if \ i \in K_2 \end{cases}$$

The following theorem is implicitly proven in [KXAH09a] and more explicitly stated and proven in [KXAH09b]

Theorem 5.5.1. *Let* $\gamma_1 = \frac{n_1}{n}$ *and* $\gamma_2 = \frac{n_2}{n}$. *If* γ_1, γ_2, P_1, P_2, W_1 *and* W_2 *are fixed, there exists a critical threshold* $\delta_c = \delta_c(\gamma_1, \gamma_1, P_1, P_2, \frac{W_2}{W_1})$, *totally computable, such that if* $\delta = \frac{m}{n} \geq \delta_c$, *then a vector* $\mathbf{x}$ *generated randomly based on the described nonuniformly sparse model can be recovered from the weighted* ℓ_1*-minimization of 5.5.1 with probability* $1 - o(e^{-cn})$ *for some positive constant* c.

In [KXAH09a] and [KXAH09b], a way for computing δ_c is presented which, in the uniform sparse case (e.g., $\gamma_2 = 0$) and equal weights, is consistent with the weak threshold of Donoho and Tanner [Don06c] for almost sure recovery of sparse signals with ℓ_1-minimization.

In summary, given a certain δ, the two different weights W_1 and W_2 for weighted ℓ_1 minimization, the size of the two weighted blocks, and also the number (or proportion) of nonzero elements inside each weighted block, the framework from [KXAH09a] can determine whether a uniform random measurement matrix will be able to perfectly recover the original signals with overwhelming probability. Using this framework we can now begin to analyze the performance of the modified reweighted algorithm of section 5.2. Although we are not directly given some prior information, as in the nonuniform sparse model for instance, about the signal structure, one might hope to infer such information after the first step of the modified reweighted algorithm. To this end, note that the immediate step in the algorithm after the regular ℓ_1-minimization is to choose the largest $(1-\epsilon)\rho_F(\delta)\delta n$ entries in absolute value. This is equivalent to splitting the index set of the vector $\mathbf{x}$ to two classes K' and K'', where K' corresponds to the larger entries. We now try to find a correspondence between this setup and the setup of [KXAH09a] where sparsity factors on the sets K' and $\bar{K}'$ are known. We claim the following upper bound on the number of nonzero entries of $\mathbf{x}$ with index on K'

Theorem 5.5.2. *There are at least* $(1-\epsilon)\rho_F(\delta)\delta n - \frac{4C(\kappa+1)\Delta}{(C-1)a_1}$ *nonzero entries in* $\mathbf{x}$ *with index on the set* K'.

Proof. Directly from Theorem 5.4.2 and the fact that $\|\mathbf{x}_{\bar{K}}\|_1 \leq \Delta$. ∎

The above result simply gives us a lower bound on the sparsity factor (ratio of nonzero elements) in the vector $\mathbf{x}_{K'}$,

$$P_1 \geq 1 - \frac{4C(\kappa+1)}{(C-1)a_1\rho_F(\delta)\delta}\frac{\Delta}{n}. \tag{5.5.2}$$

Since we also know the original sparsity of the signal, $\|\mathbf{x}\|_0 \leq k_{\text{total}}$, we have the following upper bound on the sparsity factor of the second block of the signal $\mathbf{x}_{\bar{K}'}$,

$$P_2 \leq \frac{k_{\text{total}} - (1-\epsilon)\rho_F(\delta)\delta n + \frac{4C(\kappa+1)\Delta}{(C-1)a_1}}{n - (1-\epsilon)\rho_F(\delta)\delta n}. \tag{5.5.3}$$

Note that if a_1 is large and $1 \gg \frac{\Delta}{a_1 n}$ (Note however, we can let Δ take a non-diminishing portion of $\|\mathbf{x}\|_1$, even though that portion can be very small), then P_1 is very close to 1. This means that the original signal is much denser in the block K' than in the second block $\bar{K}'$. Therefore, as in the last step of the modified re-weighted algorithm, we may assign a weight $W_1 = 1$ to all entries of $\mathbf{x}$ in K' and weight $W_2 = W$, $W > 1$ to the entries of $\mathbf{x}$ in $\bar{K}'$ and perform the weighted ℓ_1-minimization. The theoretical results of [KXAH09a], namely Theorem 5.5.1 guarantee that as long as $\delta > \delta_c(\gamma_1, \gamma_2, P_1, P_2, \frac{W_2}{W_1})$ then the signal will be recovered with overwhelming probability for large n.[1] The numerical examples in the next section do show that the reweighted ℓ_1 algorithm can increase the recoverable sparsity threshold, i.e., $P_1\gamma_1 + P_2\gamma_2$.

5.6 Numerical Computations on the Bounds

Using numerical evaluations similar to those in [KXAH09a], we demonstrate a strict improvement in the sparsity threshold from the weak bound of [Don06c], for which our algorithm is guaranteed to succeed. Let $\delta = 0.555$ and $\frac{W_2}{W_1} = 3$ be fixed, which

[1] We should remark that this only holds if the Gaussian random matrix is sampled independently from the signal to be decoded in the weighted ℓ_1 minimization. In the iterative reweighted ℓ_1 minimization, we do not have this independence. However, this can be accounted for by using a union bound over the possible configurations of the set K'. Using similar arguments as in Theorem 5.4.2, we can show that the exponent for this union bound can be made arbitrarily small if $1 \gg \frac{\Delta}{a_1 n}$, which can be outweighed by the Grassmann Angle exponent.

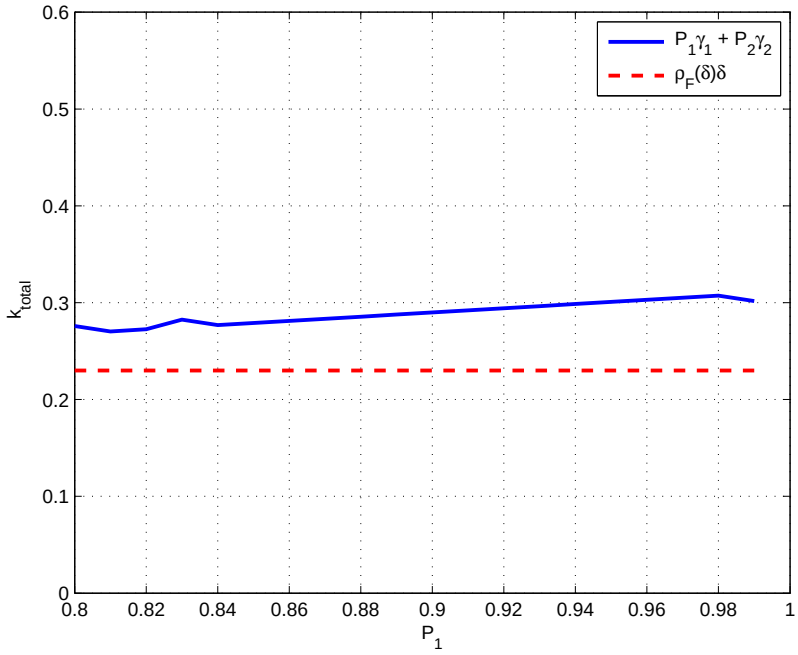


Figure 5.1: Recoverable sparsity factor for $\delta = 0.555$, when the modified reweighted ℓ_1-minimization algorithm is used.

means that $\zeta = \rho_F(\delta)\delta$ is also given. We set $\epsilon = 0.01$. The sizes of the two classes K' and $\overline{K'}$ would then be $\gamma_1 n = (1-\epsilon)\zeta n$ and $\gamma_2 n = (1-\gamma_1)n$ respectively. The sparsity ratios P_1 and P_2 of course depend on other parameters of the original signal, as is given in equations (5.5.2) and (5.5.3). For values of P_1 close to 1, we search over all pairs of P_1 and P_2 such that the critical threshold $\delta_c(\gamma_1, \gamma_2, P_1, P_2, \frac{W_2}{W_1})$ is strictly less than δ. This essentially means that a non-uniform signal with sparsity factors P_1 and P_2 over the sets K' and $\overline{K'}$ is highly probable to be recovered successfully via the weighted ℓ_1-minimization with weights W_1 and W_2. For any such P_1 and P_2, the signal parameters (Δ, a_1) can be adjusted accordingly. Eventually, we will be able to recover signals with average sparsity factor $P_1\gamma_1 + P_2\gamma_2$ using this method. We simply plot this ratio as a function of P_1 in Figure 5.1. The straight line is the weak bound of [Don06c] for $\delta = 0.555$ which is basically $\rho_F(\delta)\delta$.

Chapter 6

Null Space Conditions and Thresholds for Rank Minimization

Evolving from compressive sensing problems, where we are interested in recovering sparse vector signals from compressed linear measurements, in this chapter, we will turn our attention to recovering matrices of low ranks from compressed linear measurements. Minimizing the rank of a matrix subject to constraints is a challenging problem that arises in many applications in machine learning, control theory, and discrete geometry. This class of optimization problems, known as rank minimization, is NP-HARD, and for most practical problems there are no efficient algorithms that yield exact solutions. A popular heuristic replaces the rank function with the nuclear norm—equal to the sum of the singular values—of the decision variable and has been shown to provide the optimal low rank solution in a variety of scenarios. In this chapter, we assess the practical performance of this heuristic for finding the minimum rank matrix subject to linear constraints. Our starting point is the characterization of a necessary and sufficient condition that determines when this heuristic finds the minimum rank solution. We then obtain conditions, as a function of the matrix dimensions and rank and the number of constraints, such that our conditions for success are satisfied for almost all linear constraint sets as the matrix dimensions tend to infinity. Finally, we provide empirical evidence that these probabilistic bounds provide accurate predictions of the heuristic's performance in non-asymptotic scenarios.

6.1 Introduction

The *rank minimization* problem consists of finding the minimum rank matrix in a convex constraint set. Though this problem is NP-Hard even when the constraints are linear, a recent paper by Recht et al. [RFP] showed that most instances of the linearly constrained rank minimization problem could be solved in polynomial time as long as there were sufficiently many linearly independent constraints. Specifically, they showed that minimizing the *nuclear norm* (also known as the Ky Fan 1-norm or the trace norm) of the decision variable subject to the same affine constraints produces the lowest rank solution if the affine space is selected at random. The nuclear norm of a matrix—equal to the sum of the singular values—can be optimized in polynomial time. This initial paper initiated a groundswell of research, and, subsequently, Candès and Recht showed that the nuclear norm heuristic could be used to recover low-rank matrices from a sparse collection of entries [CR09], Ames and Vavasis have used similar techniques to provide average case analysis of NP-HARD combinatorial optimization problems [AV09], and Vandenberghe and Zhang have proposed novel algorithms for identifying linear systems [LV08]. Moreover, fast algorithms for solving large-scale instances of this heuristic have been developed by many groups [CCS08, LB09, MGC08, MJCD08, RFP]. These developments provide new strategies for tackling the rank minimization problems that arise in Machine Learning [YAU07, AMP08, RS05], Control Theory [BD98, EGG93, FHB01], and dimensionality reduction [LLR95, WS06, YELM07].

Numerical experiments in [RFP] suggested that the nuclear norm heuristic significantly out-performed the theoretical bounds provided by their probabilistic analysis. They showed numerically that random instances of the nuclear norm heuristic exhibited a *phase transition* in the parameter space, where, for sufficiently small values of the rank the heuristic always succeeded. Surprisingly, in the complement of this region, the heuristic never succeeded. The transition between the two regions appeared sharp and the location of the phase transition appeared to be nearly independent of the problem size. A similar phase transition was also observed by Candès and Recht

when the linear constraints merely constrained the values of a subset of the entries of the matrix [CR09].

In this chapter we provide an approach to explicitly calculate the location of this phase transition and provide bounds for the success of the nuclear norm heuristic that accurately reflect empirical performance. We present a *necessary* and sufficient condition for the solution of the nuclear norm heuristic to coincide with the minimum rank solution in an affine space. This condition is akin to the one in compressed sensing [SXH08a], first reported in [RXH08b]. The condition characterizes a particular property of the null-space of the linear map which defines the affine space. We then show that when the null space is sampled from the uniform distribution on subspaces, the null-space characterization holds with overwhelming probability provided the dimensions of the equality constraints are of appropriate size. We provide explicit formulas relating the dimension of the null space to the largest rank matrix that can be found using the nuclear norm heuristic. We also compare our results against the empirical findings of [RFP] and demonstrate that they provide a good approximation of the phase transition boundary especially when the number of constraints is large.

6.1.1 Main Results

Let X be an $n_1 \times n_2$ matrix decision variable. Without loss of generality, we will assume throughout that $n_1 \leq n_2$. Let $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathcal{R}^m$ be a linear map, and let $b \in \mathcal{R}^m$. The main optimization problem under study is

$$\begin{array}{ll} \text{minimize} & \text{rank}(X) \\ \text{subject to} & \mathcal{A}(X) = b\,. \end{array} \tag{6.1.1}$$

This problem is known to be NP-HARD and is also hard to approximate [MJCD08]. As mentioned above, a popular heuristic for this problem replaces the rank function with the sum of the singular values of the decision variable. Let $\sigma_i(X)$ denote the i-th largest singular value of X (equal to the square-root of the i-th largest eigenvalue of XX^*). Recall that the rank of X is equal to the number of nonzero singular values.

In the case when the singular values are all equal to one, the sum of the singular values is equal to the rank. When the singular values are less than or equal to one, the sum of the singular values is a convex function that is strictly less than the rank. This sum of the singular values is a unitarily invariant matrix norm, called the *nuclear norm*, and is denoted

$$\|X\|_* := \sum_{i=1}^{r} \sigma_i(X)\,.$$

This norm is alternatively known by several other names including the Schatten 1-norm, the Ky Fan norm, and the trace class norm.

As described in the introduction, our main concern is when the optimal solution of (6.1.1) coincides with the optimal solution of

$$\begin{array}{ll} \text{minimize} & \|X\|_* \\ \text{subject to} & \mathcal{A}(X) = b\,. \end{array} \tag{6.1.2}$$

This optimization is convex, and can be efficiently solved via a variety of methods including semidefinite programming. See [RFP] for a survey and [CCS08, LV08, MGC08] for customized algorithms.

We characterize an affine rank minimization problem (6.1.1) by three dimensionless parameters that take values in $(0, 1]$: the *aspect ratio* γ, the *constraint ratio* μ, the *rank ratio* β. Without loss of generality, we will assume throughout that we are dealing with matrices with fewer rows than columns. The aspect ratio is such that the number of rows is equal to $n_1 = \gamma n_2$. The constraint ratio is the ratio of the number of constraints to the number of parameters needed to fully specify an $n_1 \times n_2$ matrix. That is, the number of measurements is equal to $\mu\gamma n_2^2$. Generically, in the case that $\mu \geq 1$, the linear system describing the constraints is overdetermined and hence the minimum rank solution can be found by least-squares. The rank ratio is the ratio of the number of rows to the rank of the matrix so that the rank is equal to $\beta n_1 = \beta\gamma n_2$. The *model size* is the number of parameters required to define a low rank matrix. An $n_1 \times n_2$ matrix of rank r is defined by $r(n_1 + n_2 - r)$ parameters

(this quantity can be computed by calculating the number of parameters needed to specify the singular value decomposition). In terms of the parameters β and γ, the model size is equal to $\beta(1+\gamma-\beta\gamma)n_2^2$. We will focus our attention to determining for which triples (β,γ,μ) the problem (6.1.2) has the same optimal solution as the rank minimization problem (6.1.1).

Whenever $\mu < 1$, the null space of $\mathcal{A}$, that is the set of Y such that $\mathcal{A}(Y) = 0$, is not empty. Note that X is the unique optimal solution for (6.1.2) if and only if for every Y in the null-space of $\mathcal{A}$

$$\|X+Y\|_* > \|X\|_* . \tag{6.1.3}$$

The following theorem generalizes this null-space criterion to a critical property that guarantees when the nuclear norm heuristic finds the minimum rank solution of $\mathcal{A}(X) = b$ as long as the minimum rank solution is sufficiently small. Our first result is the following.

Theorem 6.1.1. *Let X_0 be the optimal solution of (6.1.1) and assume that X_0 has rank $r < n_1/2$. Then*

1. *If for every Y in the null space of $\mathcal{A}$ and for every decomposition*

$$Y = Y_1 + Y_2,$$

where Y_1 has rank r and Y_2 has rank greater than r, it holds that

$$\|Y_1\|_* < \|Y_2\|_*,$$

then X_0 is the unique minimizer of (6.1.2).

2. *Conversely, if the condition of part 1 does not hold, then there exists a vector $b \in \mathcal{R}^m$ such that the minimum rank solution of $\mathcal{A}(X) = b$ has rank at most r and is not equal to the minimum nuclear norm solution.*

This result is of interest for multiple reasons. First, it gives a necessary and sufficient condition on the mapping $\mathcal{A}$ such that *all* sufficiently low rank X_0 are recoverable from (6.1.2). Second, as shown in [RXH08b], a variety of the rank minimization problems, including those with inequality and semidefinite cone constraints, can be reformulated in the form of (6.1.1). Finally, we now present a family of random equality constraints under which the nuclear norm heuristic succeeds with overwhelming probability. We prove both of the following two theorems by showing that $\mathcal{A}$ obeys the null-space criteria of Equation (6.1.3) and Theorem 6.1.1 respectively with overwhelming probability.

Note that for a linear map $\mathcal{A} : \mathcal{R}^{n_1 \times n_2} \to \mathcal{R}^m$, we can always find an $m \times n_1 n_2$ matrix $\mathbf{A}$ such that

$$\mathcal{A}(X) = \mathbf{A} \operatorname{vec} X \,. \tag{6.1.4}$$

In the case where $\mathbf{A}$ has entries sampled independently from a zero-mean, unit-variance Gaussian distribution, then the null space characterization of Theorem 6.1.1 holds with overwhelming probability provided m is large enough. We define the random ensemble of $d_1 \times d_2$ matrices $\mathfrak{G}(d_1, d_2)$ to be the Gaussian ensemble, with each entry sampled i.i.d. from a Gaussian distribution with zero-mean and variance one. We also denote $\mathfrak{G}(d, d)$ by $\mathfrak{G}(d)$.

In order to state our results, we need to define a function $\varphi : [0, 1] \to \mathcal{R}$ that specifies the asymptotic mean of the nuclear norm of a matrix sampled from $\mathfrak{G}(d_1, d_2)$ $(d_1 \leq d_2)$.

$$\varphi(\gamma) := \frac{1}{2\pi} \int_{(1-\sqrt{\gamma})^2}^{(1+\sqrt{\gamma})^2} \sqrt{\frac{(z - s_1)(s_2 - z)}{z}} dz \tag{6.1.5}$$

The origins of this formula will be described in Section 6.3.4. We can now state our main threshold theorems. The first result characterizes when a particular low-rank matrix can be recovered from a random linear system via nuclear norm minimization.

Theorem 6.1.2 (Weak Bound)**.** *Set $n_1 \leq n_2$, $\gamma = n_1/n_2$, and let X_0 be an $n_1 \times n_2$ matrix with of rank $r = \beta n_1$. Let $\mathcal{A} : \mathcal{R}^{n_1 \times n_2} \to \mathcal{R}^{\mu n_1 n_2}$ denote the random linear*

transformation

$$\mathcal{A}(X) = \mathbf{A}\,\mathrm{vec}(X)\,,$$

where $\mathbf{A}$ *is sampled from* $\mathfrak{G}(\mu n_1 n_2, n_1 n_2)$. *Then whenever*

$$\mu \geq 1 - \left(\varphi\left(\frac{\gamma - \beta\gamma}{1-\beta\gamma}\right)\frac{(1-\beta)^{3/2}}{\gamma} - \frac{8}{3\pi}\gamma^{1/2}\beta^{3/2}\right)^2, \tag{6.1.6}$$

there exists a numerical constant $c_w(\mu,\beta,\gamma) > 0$ *such that with probability exceeding* $1 - e^{-c_w(\mu,\beta,\gamma)n_2^2 + o(n_2^2)}$,

$$X_0 = \arg\min\{\|Z\|_* \ : \ \mathcal{A}(Z) = \mathcal{A}(X_0)\}\,.$$

In particular, if β,γ, *and* μ *satisfy (6.1.6), then nuclear norm minimization will recover* X_0 *from a random set of* $\mu\gamma n_2^2$ *constraints drawn from the Gaussian ensemble almost surely as* $n_2 \to \infty$.

Formula (6.1.6) provides a lower-bound on the empirical phase transition observed in [RFP]. Note that this theorem only depends on the null-space of $\mathcal{A}$ being selected from the uniform distribution of subspaces. From this perspective, the theorem states that the nuclear norm heuristic succeeds for almost all instances of the affine rank minimization problem with parameters (β, γ, μ) satisfying (6.1.6). A particular case of interest is the case of square matrices ($\gamma = 1$). In this case, the Weak Bound (6.1.6) takes the elegant closed form:

$$\mu \geq 1 - \frac{64}{9\pi^2}\left((1-\beta)^{3/2} - \beta^{3/2}\right)^2\,. \tag{6.1.7}$$

The second theorem characterizes when the nuclear norm heuristic succeeds at recovering *all* low rank matrices.

Theorem 6.1.3 (Strong Bound)**.** *Let* $\mathcal{A}$ *be defined as in Theorem 6.1.2. Define the*

two functions

$$f(\gamma,\beta,\epsilon) = \frac{\varphi\left(\frac{\gamma-\beta\gamma}{1-\beta\gamma}\right)\gamma^{-1}(1-\beta)^{3/2} - \frac{8}{3\pi}\gamma^{1/2}\beta^{3/2} - 4\epsilon\varphi(\gamma)}{1+4\epsilon} \tag{6.1.8}$$

$$g(\gamma,\beta,\epsilon) = \sqrt{2\beta\gamma(1+\gamma-\beta\gamma)\log\left(\frac{3\pi}{2\epsilon}\right)}\,. \tag{6.1.9}$$

Then there exists a numerical constant $c_s(\mu,\beta) > 0$ *such that with probability exceeding* $1 - e^{-c_s(\mu,\beta)n^2+o(n^2)}$, *for* all $\gamma n \times n$ *matrices* X_0 *of rank* $r \leq \beta\gamma n$,

$$X_0 = \arg\min\{\|Z\|_* \ : \ \mathcal{A}(Z) = \mathcal{A}(X_0)\},$$

whenever

$$\mu \geq 1 - \sup_{\substack{\epsilon>0 \\ f(\beta,\epsilon)-g(\beta,\epsilon)>0}} \left(f(\beta,\epsilon) - g(\beta,\epsilon)\right)^2\,. \tag{6.1.10}$$

In particular, if β, γ, *and* μ *satisfy (6.1.10), then nuclear norm minimization will recover all rank* r *matrices from a random set of* $\gamma\mu n^2$ *constraints drawn from the Gaussian ensemble almost surely as* $n \to \infty$.

Figure 6.1 plots the bound from Theorems 6.1.2 and 6.1.3 with $\gamma = 1$. We call (6.1.6) the *Weak Bound* because it is a condition that depends on the optimal solution of (6.1.1). On the other hand, we call (6.1.10) the *Strong Bound* as it guarantees the nuclear norm heuristic succeeds, *no matter what the optimal solution*, as long as the minimum of the rank minimization problem is sufficiently small. The Weak Bound is the only bound that can be tested experimentally, and, in Section 6.4, we will show that it corresponds well to experimental data. Moreover, the Weak Bound provides guaranteed recovery over a far larger region of the (β,μ) parameter space. Nonetheless, the mere existence of a Strong Bound is surprising in and of itself and results in a much better bound than what was available from previous results (c.f., [RFP]).

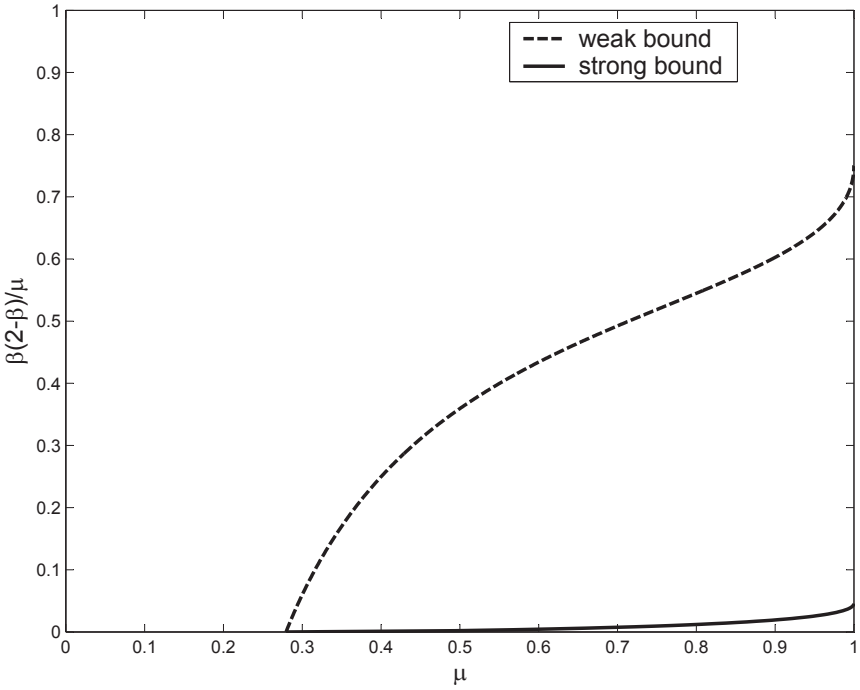


Figure 6.1: The Weak Bound (6.1.6) versus the Strong Bound (6.1.10)

6.1.2 Related Work

Optimization problems involving constraints on the rank of matrices are pervasive in engineering applications. For example, in Machine Learning, these problems arise in the context of inference with partial information [RS05] and Multi-task learning [AMP08]. In control theory, problems in controller design [EGG93, MP97], minimal realization theory [FHB01], and model reduction [BD98] can be formulated as rank minimization problems. Rank minimization also plays a key role in the study of embeddings of discrete metric spaces in Euclidean space [LLR95] and of learning structure in data and manifold learning [WS06].

In certain instances with special structure, the rank minimization problem can be solved via the singular value decomposition or can be reduced to the solution of a linear system [MP97, PK00]. In general, however, minimizing the rank of a matrix subject to convex constraints is NP-HARD. Even the problem of finding the lowest rank matrix in an affine space is NP-HARD. The best exact algorithms for this problem involve quantifier elimination and such solution methods require at least

exponential time in the dimensions of the matrix variables.

Nuclear norm minimization is a recent heuristic for rank minimization introduced by Fazel in [Faz02]. When the matrix variable is symmetric and positive semidefinite, this heuristic is equivalent to the "trace heuristic" from control theory (see, e.g., [BD98, MP97]). Both the trace heuristic and the nuclear norm generalization have been observed to produce very low-rank solutions in practice, but, until very recently, conditions where the heuristic succeeded were only available in cases that could also be solved by elementary linear algebra [PK00]. As mentioned above, the first non-trivial sufficient conditions that guaranteed the success of the nuclear norm heuristic were provided in [RFP].

The initial results in [RFP] build on seminal developments in "compressed sensing" that determined conditions for when minimizing the ℓ_1 norm of a vector over an affine space returns the sparsest vector in that space (see, e.g., [CT05, CRT06, BDDW08]). There is a strong parallelism between the sparse approximation and rank minimization settings. The rank of a diagonal matrix is equal to the number of non-zeros on the diagonal. Similarly, the sum of the singular values of a diagonal matrix is equal to the ℓ_1 norm of the diagonal. Exploiting the parallels, the authors in [RFP] were able to extend much of the analysis developed for the ℓ_1 heuristic to provide guarantees for the nuclear norm heuristic.

Building on this work, Candès and Recht showed that most matrix low rank matrices can be recovered from a sampling of on the order of $(n^{1.2}r)$ of the matrices entries [CR09] using nuclear norm minimization. In another recently provided extension, Meka et al. [MJCD08] have provided an analysis of the multiplicative weights algorithm for providing very low-rank approximate solutions of systems of inequalities. Ames and Vavasis have demonstrated that the nuclear norm heuristic can solve many instances of the NP-Hard combinatorial optimization problems maximum clique and maximum biclique [AV09].

Focusing on the special case where one seeks the lowest rank matrix in an affine subspace, Recht et al generalized the notion of "restricted isometry" from [CT05]

to the space of low rank matrices. They provided deterministic conditions on the linear map defining the affine subspace which guarantees the minimum nuclear norm solution is the minimum rank solution. Moreover, they provided several ensembles of affine constraints where this sufficient condition holds with overwhelming probability. They proved that the heuristic succeeds with large probability whenever the number m of available measurements is greater than a constant times $2nr \log n$ for $n \times n$ matrices. Since a matrix of rank r cannot be specified with less than $r(2n - r)$ real numbers, this is, up to asymptotic scaling, a nearly optimal result. However, the bounds developed in this chapter did not reflect the empirical performance of the nuclear norm heuristic. In particular, it gave vacuous results for practically sized problems where the rank was large. The results in the present work provide bounds that much more closely approximate the practical recovery region of the heuristic.

The present work builds on a different collection of developments in compressed sensing [DT05a, DT05b, SXH08a]. In these papers, the authors study properties of the null space of the linear operator that gives rise to the affine constraints. In [Don06c, DT05a], the authors think of the constraint set as a k-neighborly polytope. It turns out that this characterization of the matrix A is in fact a necessary and sufficient condition for the ℓ_1 minimization to produce the sparest solution. Furthermore, using the results of [VS92], it can be shown that if the matrix A has i.i.d. zero-mean Gaussian entries with overwhelming probability it also constitutes a k-neighborly poly-tope. The precise relation between m and k in order for this to happen is characterized in [Don06c] as well. It should also be noted that for a given value m i.e. for a given value of the constant α, the sparsity bound is significantly better in [Don06c, DT05a] than in [CT05]. Furthermore, the values of sparsity thresholds obtained for different values of α in [Don06c] approach the ones obtained by simulation as $n \longrightarrow \infty$. Our null-space criteria generalizes the concept of the same name in Compressed Sensing.

Unfortunately, the polyhedral analysis of Donoho and Tanner does not extend to the space of matrices as the unit ball in the nuclear norm is not a polyhedral set.

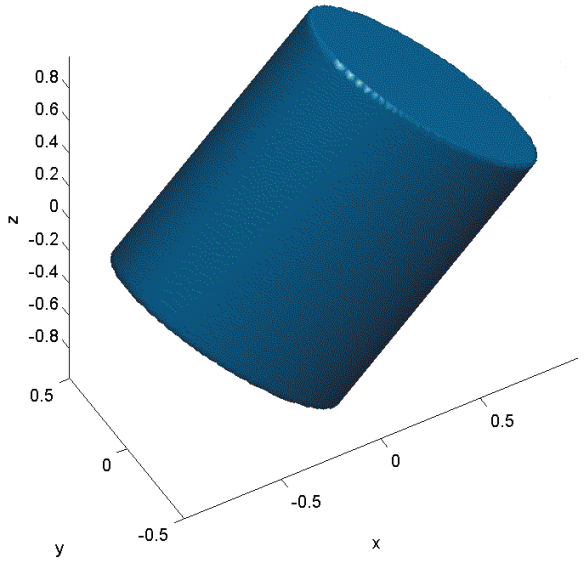


Figure 6.2: The unit ball of the nuclear norm. The figure depicts the set of all matrices of the form of equation (6.1.11) with nuclear norm less than one.

Figure 6.2 plots a simple three dimensional example, depicting the unit ball of the nuclear norm for matrices parameterized as

$$\left\{ X \ : \ X = \begin{bmatrix} x & y \\ y & z \end{bmatrix}, \ \|X\|_* \leq 1 \right\} . \tag{6.1.11}$$

In order to extend null-space analysis to the rank minimization problem, we need to follow a different path. In [SXH08a], the authors provide a probabilistic argument specifying a large region where the minimum ℓ_1 solution is the sparsest solution. This works by directly estimating the probability of success via a simple Chernoff-style argument. Our work follows this latter approach, but requires the introduction of specialized machinery to deal with the asymptotic behavior of the singular values of random matrices. We provide a sufficient statistic that guarantees the heuristic succeeds, and then use comparison lemmas for Gaussian processes to bound the expected value of this heuristic (see, for example, [LT91]). We then show that this random variable is sharply concentrated around its expectation.

6.1.3 Notation and Preliminaries

For a rectangular matrix $X \in \mathcal{R}^{n_1 \times n_2}$, X^* denotes the transpose of X. vec(X) denotes the vector in $\mathcal{R}^{n_1 n_2}$ with the columns of X stacked on top of one and other.

For vectors $v \in \mathcal{R}^d$, the only norm we will ever consider is the Euclidean norm

$$\|v\|_{\ell_2} = \left(\sum_{i=1}^{d} v_i^2 \right)^{1/2} .$$

On the other hand, we will consider a variety of matrix norms. For matrices X and Y of the same dimensions, we define the inner product in $\mathcal{R}^{n_1 \times n_2}$ as $\langle X, Y \rangle := \text{trace}(X^* Y) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} X_{ij} Y_{ij}$. The norm associated with this inner product is called the Frobenius (or Hilbert-Schmidt) norm $|| \cdot ||_F$. The Frobenius norm is also equal to the Euclidean, or ℓ_2, norm of the vector of singular values, i.e.,

$$\|X\|_F := \left(\sum_{i=1}^{r} \sigma_i^2 \right)^{\frac{1}{2}} = \sqrt{\langle X, X \rangle} = \left(\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} X_{ij}^2 \right)^{\frac{1}{2}}$$

The operator norm (or induced 2-norm) of a matrix is equal to its largest singular value (i.e., the ℓ_∞ norm of the singular values):

$$\|X\| := \sigma_1(X).$$

The nuclear norm of a matrix is equal to the sum of its singular values, i.e.,

$$\|X\|_* := \sum_{i=1}^{r} \sigma_i(X) \ .$$

These three norms are related by the following inequalities which hold for any matrix X of rank at most r:

$$||X|| \leq ||X||_F \leq ||X||_* \leq \sqrt{r}||X||_F \leq r||X||. \tag{6.1.12}$$

To any norm, we may associate a *dual norm* via the following variational definition

$$\|X\|_d = \sup_{\|Y\|_p=1} \langle Y, X \rangle .$$

One can readily check that the dual norm of the Frobenius norm is the Frobenius norm. Less trivially, one can show that the dual norm of the operator norm is the nuclear norm (See, for example, [RFP]). We will leverage the duality between the operator and nuclear norm several times in our analysis.

6.2 Necessary and Sufficient Conditions

We first prove our necessary and sufficient condition for success of the nuclear norm heuristic. We will need the following two technical lemmas. The first is an easily verified fact.

Lemma 6.2.1. *Suppose X and Y are $n_1 \times n_2$ matrices such that $X^*Y = 0$ and $XY^* = 0$. Then $\|X + Y\|_* = \|X\|_* + \|Y\|_*$.*

Indeed, if $X^*Y = 0$ and $XY^* = 0$, we can find a coordinate system in which

$$X = \left\| \begin{bmatrix} A & 0 \\ 0 & 0 \end{bmatrix} \right\|_* \quad \text{and} \quad Y = \left\| \begin{bmatrix} 0 & 0 \\ 0 & B \end{bmatrix} \right\|_*,$$

from which the lemma trivially follows. The next lemma allows us to exploit Lemma 6.2.1 in our proof.

Lemma 6.2.2. *Let X be an $n_1 \times n_2$ matrix with rank $r < \frac{n_1}{2}$ and Y be an arbitrary $n_1 \times n_2$ matrix. Let P_X^c and P_X^r be the matrices that project onto the column and row spaces of X respectively. Then if $P_X^c Y P_X^r$ has full rank, Y can be decomposed as*

$$Y = Y_1 + Y_2,$$

where Y_1 has rank r, and

$$\|X + Y_2\|_* = \|X\|_* + \|Y_2\|_*.$$

Proof. Without loss of generality, we can write X as

$$X = \begin{bmatrix} X_{11} & 0 \\ 0 & 0 \end{bmatrix},$$

where X_{11} is $r \times r$ and full rank. Accordingly, Y becomes

$$Y = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix},$$

where Y_{11} is full rank since $P_X^r Y P_X^c$ is. The decomposition is now clearly

$$Y = \underbrace{\begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{21}Y_{11}^{-1}Y_{12} \end{bmatrix}}_{Y_1} + \underbrace{\begin{bmatrix} 0 & 0 \\ 0 & Y_{22} - Y_{21}Y_{11}^{-1}Y_{12} \end{bmatrix}}_{Y_2}.$$

That Y_1 has rank r follows from the fact that the rank of a block matrix is equal to the rank of a diagonal block plus the rank of its Schur complement (see, e.g., [HJ91, S2.2]). That $\|X_1 + Y_2\|_* = \|X_1\|_* + \|Y_2\|*$ follows from Lemma 6.2.1. ∎

We can now provide a proof of Theorem 6.1.1.

Proof. We begin by proving the converse. Assume the condition of part 1 is violated, i.e., there exists some Y, such that $\mathcal{A}(Y) = 0$, $Y = Y_1 + Y_2$, $\text{rank}(Y_2) > \text{rank}(Y_1) = r$, yet $\|Y_1\|_* > \|Y_2\|_*$. Now take $X_0 = Y_1$ and $b = \mathcal{A}(X_0)$. Clearly, $\mathcal{A}(-Y_2) = b$ (since Y is in the null space) and so we have found a matrix of higher rank, but lower nuclear norm.

For the other direction, assume the condition of part 1 holds. Now use Lemma 6.2.2 with $X = X_0$ and $Y = X_* - X_0$. That is, let P_X^c and P_X^r be the matrices that project

onto the column and row spaces of X_0 respectively and assume that $P^c_{X_0}(X_* - X_0)P^r_{X_0}$ has full rank. Write $X_* - X_0 = Y_1 + Y_2$ where Y_1 has rank r and $\|X_0 + Y_2\|_* = \|X_0\|_* + \|Y_2\|_*$. Assume further that Y_2 has rank larger than r (recall $r < n/2$). We will consider the case where $P^c_{X_0}(X_* - X_0)P^r_{X_0}$ does not have full rank and/or Y_2 has rank less than or equal to r in the appendix. We now have:

$$\begin{aligned}
\|X_*\|_* &= \|X_0 + X_* - X_0\|_* \\
&= \|X_0 + Y_1 + Y_2\|_* \\
&\geq \|X_0 + Y_2\|_* - \|Y_1\|_* \\
&= \|X_0\|_* + \|Y_2\|_* - \|Y_1\|_* \quad \text{by Lemma 6.2.2.}
\end{aligned}$$

But $\mathcal{A}(Y_1 + Y_2) = 0$, so $\|Y_2\|_* - \|Y_1\|_*$ non-negative and therefore $\|X_*\|_* \geq \|X_0\|_*$. Since X_* is the minimum nuclear norm solution, implies that $X_0 = X_*$. ■

For the interested reader, the argument for the case where $P^r_{X_0}(X_* - X_0)P^c_{X_0}$ does not have full rank or Y_2 has rank less than or equal to r can be found in the appendix.

6.3 Proofs of the Probabilistic Bounds

We now turn to the proofs of the probabilistic bounds (6.1.6) and (6.1.10). We first provide a sufficient condition which implies the necessary and sufficient null-space conditions. Then, noting that the null space of $\mathcal{A}$ is spanned by Gaussian vectors, we use bounds from probability on Banach Spaces to show that the sufficient conditions are met. This will require the introduction of two useful auxiliary functions whose actions on Gaussian processes are explored in Section 6.3.4.

6.3.1 Sufficient Condition for Null Space Characterizations

The following theorem gives us a new condition that implies our necessary and sufficient condition.

Theorem 6.3.1. *Let $\mathcal{A}$ be a linear map of $n_1 \times n_2$ matrices into $\mathcal{R}^m$. Suppose that for every Y in the null-space of $\mathcal{A}$ and any projection operators P and Q onto r-dimensional subspaces of $\mathcal{R}^{n_1}$ and $\mathcal{R}^{n_2}$ respectively that*

$$\|(I-P)Y(I-Q)\|_* \geq \|PYQ\|_* . \tag{6.3.1}$$

Then for every matrix Z with row and column spaces equal to the range of Q and P respectively,

$$\|Z+Y\|_* \geq \|Z\|_*,$$

for all Y in the null-space of $\mathcal{A}$. In particular, if (6.3.1) holds for every pair of projection operators P and Q, then for every Y in the null space of $\mathcal{A}$ and for every decomposition $Y = Y_1 + Y_2$ where Y_1 has rank r and Y_2 has rank greater than r, it holds that

$$\|Y_1\|_* \leq \|Y_2\|_* .$$

We will need the following lemma.

Lemma 6.3.2. *For any block partitioned matrix,*

$$X = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

we have $\|X\|_* \geq \|A\|_* + \|D\|_*$.

Proof. This lemma follows from the dual description of the nuclear norm:

$$\|X\|_* = \sup \left\{ \left\langle \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix}, \begin{bmatrix} A & B \\ C & D \end{bmatrix} \right\rangle \;\middle|\; \left\| \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \right\| = 1 \right\} . \tag{6.3.2}$$

and similarly,

$$\|A\|_* + \|D\|_* = \sup \left\{ \left\langle \begin{bmatrix} Z_{11} & 0 \\ 0 & Z_{22} \end{bmatrix}, \begin{bmatrix} A & B \\ C & D \end{bmatrix} \right\rangle \;\middle|\; \left\| \begin{bmatrix} Z_{11} & 0 \\ 0 & Z_{22} \end{bmatrix} \right\| = 1 \right\}. \tag{6.3.3}$$

Since (6.3.2) is a supremum over a larger set that (6.3.3), the claim follows. ∎

Theorem 6.3.1 now trivially follows.

Proof of Theorem 6.3.1. Without loss of generality, we may choose coordinates such that P and Q both project onto the space spanned by first r standard basis vectors. Then we may partition Y as

$$Y = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix}$$

and write, using Lemma 6.3.2,

$$\begin{aligned} \|Y - Z\|_* - \|Z\|_* &= \left\| \begin{bmatrix} Y_{11} - Z & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} \right\|_* - \|Z\|_* \\ &\geq \|Y_{11} - Z\|_* + \|Y_{22}\|_* - \|Z\|_* \\ &\geq \|Y_{22}\|_* - \|Y_{11}\|_* \end{aligned}$$

which is non-negative by assumption. Note that if the theorem holds for all projection operators P and Q whose range has dimension r, then $\|Z + Y\|_* \geq \|Z\|_*$ for all matrices Z of rank r and hence the second part of the theorem follows. ∎

6.3.2 Proof of the Weak Bound

Now we can turn to the proof of Theorem 6.1.2. The key observation in proving this lemma is the following characterization of the null-space of $\mathcal{A}$ provided by Stojnic et al. [SXH08a]

Lemma 6.3.3. *Let $\mathcal{A}$ be sampled from $\mathfrak{G}(\mu n_1 n_2,\ n_1 n_2)$. Then the null space of $\mathcal{A}$ is identically distributed to the span of $n_1 n_2(1-\mu)$ matrices G_i where each G_i is sampled i.i.d. from $\mathfrak{G}(n_1, n_2)$. In other words, we may assume that $w \in \ker(\mathcal{A})$ can be written as $\sum_{i=1}^{n_1 n_2(1-\mu)} v_i G_i$ for some $v \in \mathcal{R}^{n_1 n_2(1-\mu)}$.*

This is nothing more than a statement that the null-space of $\mathcal{A}$ is a random subspace. However, when we parameterize elements in this subspace as linear combinations of Gaussian vectors, we can leverage Comparison Theorems for Gaussian processes to yield our bounds.

Let $M = n_1 n_2(1-\mu)$ and let $G_1, \ldots, G_M$ be i.i.d. samples from $\mathfrak{G}(n_1, n_2)$. Let X_0 be a matrix of rank βn_1. Let P_{X_0} and Q_{X_0} denote the projections onto the column and row spaces of X_0 respectively. By Theorem 6.3.1 and Lemma 6.3.3, we need to show that for all $v \in \mathcal{R}^M$,

$$\left\| (I - P_{X_0}) \left(\sum_{i=1}^{M} v_i G_i \right) (I - Q_{X_0}) \right\|_* \geq \left\| P_{X_0} \left(\sum_{i=1}^{M} v_i G_i \right) Q_{X_0} \right\|_* . \tag{6.3.4}$$

That is, $\sum_{i=1}^{M} v_i G_i$ is an arbitrary element of the null space of $\mathcal{A}$, and this equation restates the sufficient condition provided by Theorem 6.3.1. Now it is clear by homogeneity that we can restrict our attention to those $v \in \mathcal{R}^M$ with Euclidean norm 1. The following lemma characterizes when the expected value of this difference is nonnegative.

Lemma 6.3.4. *Let $n_1 = \gamma n_2$ for some $\gamma \in (0, 1]$ and $r = \beta n_1$ for some $\beta \in (0, 1]$. Suppose P and Q are projection operators onto r-dimensional subspaces of $\mathcal{R}^{n_1}$ and $\mathcal{R}^{n_2}$ respectively. For $i = 1, \ldots, M$ let G_i be sampled from $\mathfrak{G}(n_1, n_2)$. Then*

$$\begin{aligned} &\mathcal{E} \left[\inf_{\|v\|_{\ell_2}=1} \left\| (I - P) \left(\sum_{i=1}^{M} v_i G_i \right) (I - Q) \right\|_* - \left\| P \left(\sum_{i=1}^{M} v_i G_i \right) Q \right\|_* \right] \\ &\quad \geq \left(\left(\varphi \left(\frac{\gamma - \beta\gamma}{1 - \beta\gamma} \right) + o(1) \right) (1-\beta)^{3/2} - (\varphi(1) + o(1)) \gamma^{3/2} \beta^{3/2} \right) n_2^{3/2} - \sqrt{M n_1}, \end{aligned} \tag{6.3.5}$$

where φ is defined as in (6.1.5).

We will prove this lemma and a similar inequality required for the proof of the Strong Bound in Section 6.3.4 below. But we now show how using this Lemma and a concentration of measure argument, we can prove Theorem 6.1.2.

First note, that if we plug in $M = (1-\mu)n_1 n_2$, divide the right hand side by $n_2^{3/2}$, and ignore the $o(1)$ terms, the right hand side of (6.3.5) is non-negative if (6.1.6) holds. To bound the probability that (6.3.4) is non-negative, we employ a powerful concentration inequality for the Gaussian distribution bounding deviations of smoothly varying functions from their expected value.

To quantify what we mean by smoothly varying, recall that a function f is *Lipshitz* with respect to the Euclidean norm if there exists a constant L such that $|f(x) - f(y)| \leq L\|x-y\|_{\ell_2}$ for all x and y. The smallest such constant L is called the *Lipshitz constant* of the map f. If f is Lipshitz, it cannot vary too rapidly. In particular, note that if f is differentiable and Lipshitz, then L is a bound on the norm of the gradient of f. The following theorem states that the deviations of a Lipshitz function applied to a Gaussian random variable have Gaussian tails.

Theorem 6.3.5. *Let $x \in \mathcal{R}^D$ be a normally distributed random vector with zero-mean variance equal to the identity. Let $f : \mathcal{R}^D \to \mathcal{R}$ be a function with Lipshitz constant L. Then*

$$\mathbb{P}[|f(x) - \mathcal{E}[f(x)]| \geq t] \leq 2\exp\left(-\frac{t^2}{2L^2}\right).$$

See [LT91] for a proof of this theorem with slightly weaker constants and a list of several references to more complicated proofs that give rise to this concentration inequality. The following lemma bounds the Lipshitz constant of interest

Lemma 6.3.6. *For $i = 1, \ldots, M$, let $X_i \in \mathcal{R}^{D_1 \times D_2}$ and $Y_i \in \mathcal{R}^{D_3 \times D_4}$ with $D_1 \leq D_2$ and $D_3 \leq D_4$. Define the function*

$$F_I(X_1, \ldots, X_M, Y_1, \ldots, Y_M) = \inf_{\|v\|_{\ell_2}=1} \left\| \sum_{i=1}^{M} v_i X_i \right\|_* - \left\| \sum_{i=1}^{M} v_i Y_i \right\|_*.$$

Then the Lipshitz constant of F_I is at most $\sqrt{D_1 + D_3}$.

The proof of this lemma is straightforward and can be found in the appendix. Using Theorem 6.3.5 and Lemmas 6.3.4 and 6.3.6, we can now bound

$$\mathbb{P}\left[\inf_{\|v\|_{\ell_2}=1}\left\|(I-P_{X_0})\left(\sum_{i=1}^{M}v_iG_i\right)(I-Q_{X_0})\right\|_*-\left\|P_{X_0}\left(\sum_{i=1}^{M}v_iG_i\right)Q_{X_0}\right\|_*\leq tn_2^{3/2}\right]$$
$$\leq \exp\left(-\tfrac{1}{2}\left\{\varphi\left(\frac{\gamma-\beta\gamma}{1-\beta\gamma}\right)\frac{(1-\beta)^{3/2}}{\gamma}-\frac{8}{3\pi}\gamma^{1/2}\beta^{3/2}-\sqrt{1-\mu}-\frac{t}{\gamma}\right\}n_2^2+o(n_2^2)\right). \tag{6.3.6}$$

Setting $t=0$ completes the proof of Theorem 6.1.2. We will use this concentration inequality with a non-zero t to prove the Strong Bound.

6.3.3 Proof of the Strong Bound

The proof of the Strong Bound is similar to that of the Weak Bound except we prove that (6.3.4) holds for *all* operators P and Q that project onto r-dimensional subspaces. Our proof will require an ϵ-net for the projection operators. By an ϵ-net, we mean a finite set Ω consisting of pairs of r-dimensional projection operators such that for any P and Q that project onto r-dimensional subspaces, there exists $(P',Q')\in\Omega$ with $\|P-P'\|+\|Q-Q'\|\leq\epsilon$. We will show that if a slightly stronger bound than (6.3.4) holds on the ϵ-net, then (6.3.4) holds for all choices of row and column spaces.

Let us first examine how (6.3.4) changes when we perturb P and Q. Let P, Q, P' and Q' all be projection operators onto r-dimensional subspaces of $\mathcal{R}^{n_1}$ and $\mathcal{R}^{n_2}$

respectively. Let W be some $n_1 \times n_2$ matrix and observe that

$$\begin{aligned}
&\|(I-P)W(I-Q)\|_* - \|PWQ\|_* - (\|(I-P')W(I-Q')\|_* - \|P'WQ'\|_*) \\
\leq &\|(I-P)W(I-Q) - (I-P')W(I-Q')\|_* + \|PWQ - P'WQ'\|_* \\
\leq &\|(I-P)W(I-Q) - (I-P')W(I-Q)\|_* \\
&+ \|(I-P')W(I-Q) - (I-P')W(I-Q')\|_* \\
&+ \|PWQ - P'WQ\|_* + \|P'WQ - P'WQ'\|_* \\
\leq &\|P-P'\|\|W\|_*\|I-Q\| + \|I-P'\|\|W\|_*\|Q-Q'\| \\
&+ \|P-P'\|\|W\|_*\|Q\| + \|P'\|\|W\|_*\|Q-Q'\| \\
\leq &2(\|P-P'\| + \|Q-Q'\|)\|W\|_* .
\end{aligned}$$

Here, the first and second lines follow from the triangle inequality, the third line follows because $\|AB\|_* \leq \|A\|\|B\|_*$, and the fourth line follows because P, P', Q, and Q' are all projection operators. Rearranging this inequality gives

$$\begin{aligned}
\|(I-P)W(I-Q)\|_* - \|PWQ\|_* \geq \|(I-P')W(I-Q')\|_* - \|P'WQ'\|_* \\
- 2(\|P-P'\| + \|Q-Q'\|)\|W\|_* .
\end{aligned} \tag{6.3.7}$$

Let us now suppose that with overwhelming probability

$$\|(I-P')W(I-Q')\|_* - \|P'WQ'\|_* - 4\epsilon\|W\|_* \geq 0 \tag{6.3.8}$$

for all (P', Q') in our ϵ-net Ω. Then by (6.3.7), this means that $\|(I-P)W(I-Q)\|_* - \|PWQ\|_* \geq 0$ for any arbitrary pair of projection operators onto r-dimensional subspaces. Thus, if we can show that (6.3.8) holds on an ϵ-net, we will have proved the Strong Bound.

To proceed, we need to know the size of an ϵ-net. The following bound on such a net is due to Szarek.

Theorem 6.3.7 (Szarek [Sza98])**.** *Consider the space of all projection operators on $\mathcal{R}^n$ projecting onto r-dimensional subspaces endowed with the metric*

$$d(P, P') = \|P - P'\|.$$

Then there exists an ϵ-net in this metric space with cardinality at most $\left(\frac{3\pi}{2\epsilon}\right)^{r(n-r/2-1/2)}$.

With this covering number in hand, we now calculate the probability that for a given P and Q in the ϵ-net,

$$\inf_{\|v\|_{\ell_2}=1} \left\| (I-P)\left(\sum_{i=1}^M v_i G_i\right)(I-Q)\right\|_* - \left\|P\left(\sum_{i=1}^M v_i G_i\right)Q\right\|_* \geq 4\epsilon \sup_{\|v\|_{\ell_2}=1} \left\|\sum_{i=1}^M v_i G_i\right\|_* . \tag{6.3.9}$$

As we will show in Section 6.3.4, we can upper bound the right hand side of this inequality using a similar bound as in Lemma 6.3.4.

Lemma 6.3.8. *For $i = 1, \ldots, M$ let G_i be sampled from $\mathfrak{G}(\gamma n, n)$ with $\gamma \in (0, 1]$. Then*

$$\mathcal{E}\left[\sup_{\|v\|_{\ell_2}=1} \left\|\sum_{i=1}^M v_i G_i\right\|_*\right] \leq (\varphi(\gamma) + o(1))\, n^{3/2} + \sqrt{\gamma M n}\,. \tag{6.3.10}$$

Moreover, we prove the following in the appendix.

Lemma 6.3.9. *For $i = 1, \ldots, M$, let $X_i \in \mathcal{R}^{D_1 \times D_2}$ with $D_1 \leq D_2$ and define the function*

$$F_S(X_1, \ldots, X_M) = \sup_{\|v\|_{\ell_2}=1} \left\|\sum_{i=1}^M v_i X_i\right\|_* .$$

Then the Lipshitz constant of F_S is at most $\sqrt{D_1}$.

Using Lemmas 6.3.8 and 6.3.9 combined with Theorem 6.3.5, we have that

$$\mathbb{P}\left[4\epsilon \sup_{\|v\|_{\ell_2}=1} \left\|\sum_{i=1}^M v_i G_i\right\|_* \geq t n_2^{3/2}\right] \leq \exp\left(-\tfrac{1}{2}\left(\frac{\varphi(\gamma)}{\gamma} - \sqrt{1-\mu} - \frac{t}{4\epsilon\gamma}\right)^2 n_2^2 + o(n_2^2)\right). \tag{6.3.11}$$

Let t_0 be such that the exponents of (6.3.6) and (6.3.11) equal to each other. Then

we find after some algebra and the union bound

$$\mathbb{P}\Bigg[\inf_{\|v\|_{\ell_2}=1}\left\|(I-P)\left(\sum_{i=1}^{M}v_iG_i\right)(I-Q)\right\|_*-\left\|P\left(\sum_{i=1}^{M}v_iG_i\right)Q\right\|_*$$

$$\geq 4\epsilon\sup_{\|v\|_{\ell_2}=1}\left\|\sum_{i=1}^{M}v_iG_i\right\|_*\Bigg]$$

$$\geq\mathbb{P}\Bigg[\inf_{\|v\|_{\ell_2}=1}\left\|(I-P)\left(\sum_{i=1}^{M}v_iG_i\right)(I-Q)\right\|_*-\left\|P\left(\sum_{i=1}^{M}v_iG_i\right)Q\right\|_*$$

$$> t_0n_2^{3/2} > 4\epsilon\sup_{\|v\|_{\ell_2}=1}\left\|\sum_{i=1}^{M}v_iG_i\right\|_*\Bigg]$$

$$\geq 1-\mathbb{P}\left[\inf_{\|v\|_{\ell_2}=1}\left\|(I-P)\left(\sum_{i=1}^{M}v_iG_i\right)(I-Q)\right\|_*-\left\|P\left(\sum_{i=1}^{M}v_iG_i\right)Q\right\|_*<t_0n_2^{3/2}\right]$$

$$-\mathbb{P}\left[4\epsilon\sup_{\|v\|_{\ell_2}=1}\left\|\sum_{i=1}^{M}v_iG_i\right\|_*>t_0n_2^{3/2}\right]$$

$$\geq 1-2\times$$

$$\exp\left(\tfrac{1}{2}\left(\frac{\varphi\left(\frac{\gamma-\beta\gamma}{1-\beta\gamma}\right)\gamma^{-1}(1-\beta)^{3/2}-\frac{8}{3\pi}\gamma^{1/2}\beta^{3/2}-4\epsilon\varphi(\gamma)}{1+4\epsilon}-\sqrt{1-\mu}\right)^2n_2^2+o(n_2^2)\right).$$

Now, let Ω be an ϵ-net for the set of pairs of projection operators (P,Q) such that P (resp. Q) projects $\mathcal{R}^{n_1}$ (resp. $\mathcal{R}^{n_2}$) onto an r-dimensional subspace. Again by the

union bound, we have that

$$\mathbb{P}\left[\forall P, Q \inf_{\|v\|_{\ell_2}=1} \left\|(I-P)\left(\sum_{i=1}^{M} v_i G_i\right)(I-Q)\right\|_* - \left\|P\left(\sum_{i=1}^{M} v_i G_i\right)Q\right\|_* \right.$$
$$\left. \geq 4\epsilon \sup_{\|v\|_{\ell_2}=1} \left\|\sum_{i=1}^{M} v_i G_i Q\right\|_* \right]$$
$$\leq 1 - 2\exp\left(-\left\{\tfrac{1}{2}\left(f(\beta,\gamma,\epsilon) - \sqrt{1-\mu}\right)^2 - \tfrac{1}{2}g(\beta,\gamma,\epsilon)^2\right\} n_2^2 + o(n_2^2)\right)$$

where

$$f(\gamma,\beta,\epsilon) = \frac{\varphi\left(\frac{\gamma-\beta\gamma}{1-\beta\gamma}\right)\gamma^{-1}(1-\beta)^{3/2} - \frac{8}{3\pi}\gamma^{1/2}\beta^{3/2} - 4\epsilon\varphi(\gamma)}{1+4\epsilon} \tag{6.3.12}$$

$$g(\gamma,\beta,\epsilon) = \sqrt{2\beta\gamma(1+\gamma-\beta\gamma)\log\left(\frac{3\pi}{2\epsilon}\right)}. \tag{6.3.13}$$

Finding the parameters μ, β, γ, and ϵ that make the terms multiplying n_2^2 negative completes the proof of the Strong Bound.

6.3.4 Comparison Theorems for Gaussian Processes and the Proofs of Lemmas 6.3.4 and 6.3.8

Both of the two following Comparison Theorems provide sufficient conditions for when the expected supremum or infimum of one Gaussian process is greater to that of another. Elementary proofs of both of these Theorems and several other Comparison Theorems can be found in S3.3 of [LT91].

Theorem 6.3.10 (Slepian's Lemma [Sle62])**.** *Let X and Y by Gaussian random vectors in $\mathcal{R}^N$ such that*

$$\begin{cases} \mathcal{E}[X_i X_j] \leq \mathcal{E}[Y_i Y_j] & \text{for all } i \neq j \\ \mathcal{E}[X_i^2] = \mathcal{E}[Y_i^2] & \text{for all } i \end{cases}$$

Then

$$\mathcal{E}[\max_i Y_i] \leq \mathcal{E}[\max_i X_i]\,.$$

Theorem 6.3.11 (Gordan [Gor85, Gor88])**.** *Let $X = (X_{ij})$ and $Y = (Y_{ij})$ be Gaussian random matrices in $\mathcal{R}^{N_1 \times N_2}$ such that*

$$\begin{cases} \mathcal{E}[X_{ij}X_{ik}] \leq \mathcal{E}[Y_{ij}Y_{ik}] & \text{for all } i,j,k \\ \mathcal{E}[X_{ij}X_{lk}] \geq \mathcal{E}[Y_{ij}Y_{lk}] & \text{for all } i \neq l \text{ and } j,k \\ \mathcal{E}[X_{ij}^2] = \mathcal{E}[X_{ij}^2] & \text{for all } j,k \end{cases}$$

Then

$$\mathcal{E}[\min_i \max_j Y_{ij}] \leq \mathcal{E}[\min_i \max_j X_{ij}]\,.$$

The following two lemmas follow from applications of these Comparison Theorems. We prove them in more generality than necessary for the current work because both lemmas are interesting in their own right. Let $\|\cdot\|_p$ be any norm on $D_1 \times D_2$ matrices and let $\|\cdot\|_d$ be its associated dual norm (See Section 6.1.3). Again without loss of generality, we assume $D_1 \leq D_2$. Let us define the quantity $\sigma(\|\cdot\|_p)$ to be the maximum attainable Frobenius norm of an element in the unit ball of the dual norm. That is

$$\sigma(\|\cdot\|_p) = \sup_{\|Z\|_d = 1} \|Z\|_F\,, \tag{6.3.14}$$

and note that by this definition, we have for $G \in \mathfrak{G}(D_1, D_2)$

$$\sigma(\|\cdot\|_p) = \sup_{\|Z\|_d = 1} \mathcal{E}_G\left[\langle G, Z\rangle^2\right]^{1/2}$$

motivating the notation.

This first lemma is now a straightforward consequence of Slepian's lemma

Lemma 6.3.12. *Let $\Delta > 0$ and let g be a Gaussian random vector in $\mathcal{R}^M$. Let*

$G, G_1, \ldots, G_M$ be sampled i.i.d. from $\mathfrak{G}(D_1, D_2)$. Then

$$\mathcal{E}\left[\sup_{\|v\|_{\ell_2}=1}\sup_{\|Y\|_d=1}\Delta\langle g, v\rangle + \left\langle\sum_{i=1}^{M} v_i G_i, Y\right\rangle\right] \leq \mathcal{E}[\|G\|_p] + \sqrt{M(\Delta^2 + \sigma(\|\cdot\|_p)^2)}\,.$$

Proof. We follow the strategy used to prove Theorem 3.20 in [LT91]. Let $G, G_1, \ldots, G_M$ be sampled i.i.d. from $\mathfrak{G}(D_1, D_2)$ and $g \in \mathcal{R}^M$ be a Gaussian random vector and let γ be a zero-mean, unit-variance Gaussian random variable. For $v \in \mathcal{R}^M$ and $Y \in \mathcal{R}^{D_1 \times D_2}$ define

$$\begin{aligned} Q_L(v, Y) &= \Delta\langle g, v\rangle + \left\langle\sum_{i=1}^{M} v_i G_i, Y\right\rangle + \sigma(\|\cdot\|_p)\gamma \\ Q_R(v, Y) &= \langle G, Y\rangle + \sqrt{\Delta^2 + \sigma(\|\cdot\|_p)^2}\langle g, v\rangle\,. \end{aligned}$$

Now observe that for any M-dimensional unit vectors v, $\hat{v}$ and any $D_1 \times D_2$ matrices Y, $\hat{Y}$ with dual norm 1

$$\begin{aligned} &\mathcal{E}[Q_L(v, Y)Q_L(\hat{v}, \hat{Y})] - \mathcal{E}[Q_R(v, Y)Q_R(\hat{v}, \hat{Y})] \\ =&\Delta^2\langle v, \hat{v}\rangle + \langle v, \hat{v}\rangle\langle Y, \hat{Y}\rangle + \sigma(\|\cdot\|_p)^2 - \langle Y, \hat{Y}\rangle - (\Delta^2 + \sigma(\|\cdot\|_p)^2)\langle v, \hat{v}\rangle \\ =&(\sigma(\|\cdot\|_p)^2 - \langle Y, \hat{Y}\rangle)(1 - \langle v, \hat{v}\rangle)\,. \end{aligned}$$

The first quantity is always non-negative because $\langle Y, \hat{Y}\rangle \leq \max(\|Y\|_F^2, \|\hat{Y}\|_F^2) \leq \sigma(\|\cdot\|_p)^2$ by definition. The difference in expectation is thus equal to zero if $v = \hat{v}$ and is greater than or equal to zero if $v \neq \hat{v}$. Hence, by Slepian's lemma and a compactness argument (see Proposition 6.6.1 in the Appendix),

$$\mathcal{E}\left[\sup_{\|v\|_{\ell_2}=1}\sup_{\|Y\|=1} Q_L(v, Y)\right] \leq \mathcal{E}\left[\sup_{\|v\|_{\ell_2}=1}\sup_{\|Y\|=1} Q_R(v, Y)\right]$$

which proves the lemma. ∎

The following lemma can be proved in a similar fashion

Lemma 6.3.13. *Let $\|\cdot\|_p$ be a norm on $\mathcal{R}^{D_1 \times D_1}$ with dual norm $\|\cdot\|_d$ and let $\|\cdot\|_b$ be*

a norm on $\mathcal{R}^{D_2\times D_2}$. Let g be a Gaussian random vector in $\mathcal{R}^M$. Let $G_0, G_1, \ldots, G_M$ be sampled i.i.d. from $\mathfrak{G}(D_1)$ and $G'_1, \ldots, G'_M$ be sampled i.i.d. from $\mathfrak{G}(D_2)$. Then

$$\begin{aligned}\mathcal{E}\left[\inf_{\|v\|_{\ell_2}=1}\inf_{\|Y\|_b=1}\sup_{\|Z\|_d=1}\left\langle\sum_{i=1}^M v_iG_i, Z\right\rangle+\left\langle\sum_{i=1}^M v_iG'_i, Y\right\rangle\right]&\\ \geq\mathcal{E}\left[\|G_0\|_p\right]-\mathcal{E}\left[\sup_{\|v\|_{\ell_2}=1}\sup_{\|Y\|_b=1}\sigma(\|\cdot\|_p)\langle g, v\rangle+\left\langle\sum_{i=1}^M v_iG'_i, Y\right\rangle\right]&.\end{aligned}$$

Proof. Define the functionals

$$\begin{aligned}P_L(v,Y,Z)&=\left\langle\sum_{i=1}^M v_iG_i, Z\right\rangle+\left\langle\sum_{i=1}^M v_iG'_i, Y\right\rangle+\gamma\sigma(\|\cdot\|_p)\\ P_R(v,Y,Z)&=\langle G_0, Z\rangle+\sigma(\|\cdot\|_p)\langle g, v\rangle+\left\langle\sum_{i=1}^M v_iG'_i, Y\right\rangle.\end{aligned}$$

Let v and $\hat{v}$ be unit vectors in $\mathcal{R}^M$, Y and $\hat{Y}$ be $D_2\times D_2$ matrices with $\|Y\|_b=\|\hat{Y}\|_b=1$, and Z and $\hat{Z}$ be $D_1\times D_1$ matrices with $\|Z\|_d=\|\hat{Z}\|_d=1$. Then we have

$$\begin{aligned}&\mathcal{E}[P_L(v,Y,Z)P_L(\hat{v},\hat{Y},\hat{Z})]-\mathcal{E}[P_R(v,Y,Z)P_L(\hat{v},\hat{Y},\hat{Z})]\\ =&\langle v,\hat{v}\rangle\langle Z,\hat{Z}\rangle+\langle v,\hat{v}\rangle\langle Y,\hat{Y}\rangle+\sigma(\|\cdot\|_p)^2-\langle Z,\hat{Z}\rangle-\sigma(\|\cdot\|_p)^2\langle v,\hat{v}\rangle-\langle v,\hat{v}\rangle\langle Y,\hat{Y}\rangle\\ =&(\sigma(\|\cdot\|_p)^2-\langle Z,\hat{Z}\rangle)(1-\langle v,\hat{v}\rangle).\end{aligned}$$

Just as was the case in the proof of Lemma 6.3.12, the first quantity is always non-negative. Hence, the difference in expectations is greater than or equal to zero and equal to zero when $v=\hat{v}$ and $Y=\hat{Y}$. Hence, by Gordan's Lemma and a compactness argument,

$$\mathcal{E}\left[\inf_{\|v\|_{\ell_2}=1}\inf_{\|Y\|_b=1}\sup_{\|Z\|_d=1}Q_L(v,Y,Z)\right]\geq\mathcal{E}\left[\inf_{\|v\|_{\ell_2}=1}\inf_{\|Y\|_b=1}\sup_{\|Z\|_d=1}Q_R(v,Y,Z)\right]$$

completing the proof. ∎

Together with Lemmas 6.3.12 and 6.3.13, we can prove the Lemma 6.3.4.

of Lemma 6.3.4. For $i = 1, \ldots, M$, let $G_i \in \mathfrak{G}((1-\beta)\gamma n_2, (1-\beta\gamma)n_2)$ and $G_i' \in \mathfrak{G}(\gamma\beta n_2, \gamma\beta n_2)$. Then

$$\begin{aligned}
&\mathcal{E}\left[\inf_{\|v\|_{\ell_2}=1}\left\|\sum_{i=1}^M v_i G_i\right\|_* - \left\|\sum_{i=1}^M v_i G_i'\right\|_*\right] \\
&= \mathcal{E}\left[\inf_{\|v\|_{\ell_2}=1}\inf_{\|Y\|=1}\sup_{\|Z\|=1}\left\langle\sum_{i=1}^M v_i G_i, Z\right\rangle + \left\langle\sum_{i=1}^M v_i G_i', Y\right\rangle\right] \\
&\geq \mathcal{E}\left[\|G_0\|_*\right] - \mathcal{E}\left[\sup_{\|v\|_{\ell_2}=1}\sup_{\|Y\|=1}\sigma(\|\cdot\|_*)\langle g, v\rangle + \left\langle\sum_{i=1}^M v_i G_i', Y\right\rangle\right] \\
&\geq \mathcal{E}\left[\|G_0\|_*\right] - \mathcal{E}\left[\|G_0'\|_*\right] - \sqrt{M}\sqrt{\sigma(\|\cdot\|_*)^2 + \sigma(\|\cdot\|_*)^2}
\end{aligned}$$

where the first inequality follows from Lemma 6.3.13, and the second inequality follows from Lemma 6.3.12.

Now we only need to plug in the asymptotic expected value of the nuclear norm and the quantity $\sigma(\|\cdot\|_*)$. Let G be sampled from $\mathfrak{G}(D_1, D_2)$. Then

$$\mathcal{E}\|G\|_* = D_1 \mathbb{E}\sigma_i = \varphi\left(\frac{D_1}{D_2}\right) D_2^{3/2} + q(D_2) \tag{6.3.15}$$

where $\varphi(\cdot)$ is found by integrating the Mar**v**cenko-Pastur distribution (see, e.g., [MP67, Bai99]):

$$\begin{aligned}
\varphi(\gamma) &= \frac{1}{2\pi}\int_{s_1}^{s_2}\sqrt{\frac{(z-s_1)(s_2-z)}{z}}dz \\
s_1 &= (1-\sqrt{\gamma})^2 \\
s_2 &= (1+\sqrt{\gamma})^2 .
\end{aligned}$$

and $q(D_2)/D_2^{3/2} = o(1)$. Note that $\varphi(1)$ can be computed in closed form:

$$\varphi(1) = \frac{1}{2\pi}\int_0^4 \sqrt{4-t}\; dt = \frac{8}{3\pi} \approx 0.85 .$$

For $\sigma(\|\cdot\|_*)$, a straightforward calculation reveals

$$\sigma(\|\cdot\|_*) = \sup_{\|H\|\leq 1} \|G\|_F = \sqrt{D_1}\,.$$

Plugging these values in with the appropriate dimensions completes the proof. ∎

Proof of Lemma 6.3.8. This lemma immediately follows from applying Lemma 6.3.12 with $\Delta = 0$ and from the calculations at the end of the proof above. It is also an immediate consequence of Lemma 3.21 from [LT91]. ∎

6.4 Numerical Experiments

We now show that these asymptotic estimates hold even for moderately sized matrices. For simplicity of presentation, we restrict our attention in this section to square matrices with $n = n_1 = n_2$ (i.e., $\gamma = 1$). We conducted a series of experiments for a variety of the matrix sizes n, ranks r, and numbers of measurements m. As in the previous section, we let $\beta = \frac{r}{n}$ and $\mu = \frac{m}{n^2}$. For a fixed n, we constructed random recovery scenarios for low-rank $n \times n$ matrices. For each n, we varied μ between 0 and 1 where the matrix is completely determined. For a fixed n and μ, we generated all possible ranks such that $\beta(2-\beta) \leq \mu$. This cutoff was chosen because beyond that point there would be an infinite set of matrices of rank r satisfying the m equations.

For each (n, μ, β) triple, we repeated the following procedure 10 times. A matrix of rank r was generated by choosing two random $n \times r$ factors Y_L and Y_R with i.i.d. random entries and setting $Y_0 = Y_L Y_R^*$. A matrix $\mathbf{A}$ was sampled from the Gaussian ensemble with m rows and n^2 columns. Then the nuclear norm minimization

$$\begin{array}{ll} \text{minimize} & \|X\|_* \\ \text{subject to} & \mathbf{A}\,\mathrm{vec}\,X = \mathbf{A}\,\mathrm{vec}\,Y_0 \end{array}$$

was solved using the freely available software SeDuMi [Stu99] using the semidefinite programming formulation described in [RFP]. On a 2.0 GHz Laptop, each semidef-

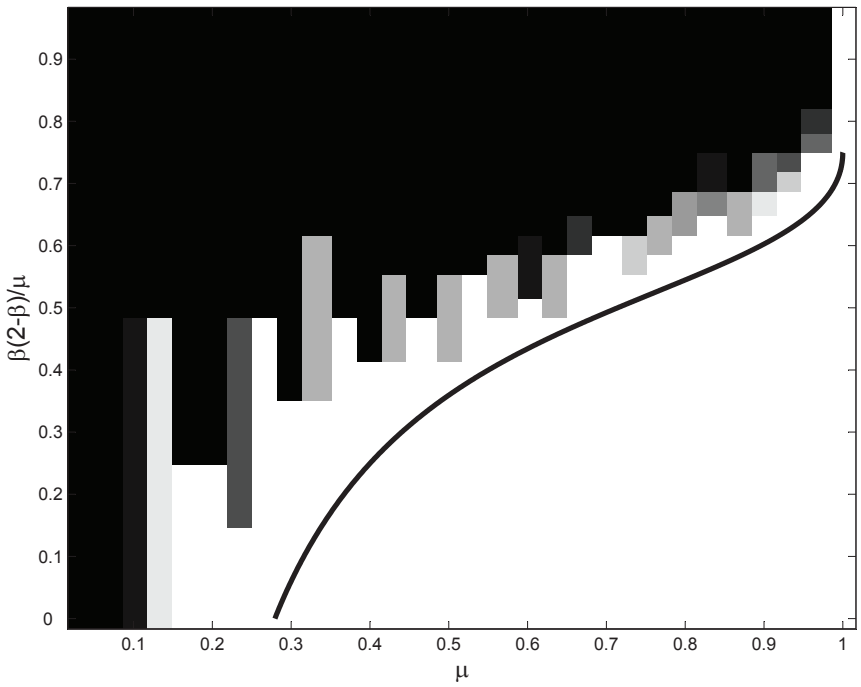
inite program could be solved in less than two minutes for 40×40 dimensional X. We declared Y_0 to be recovered if

$$\|X - Y_0\|_F / \|Y_0\|_F < 10^{-3}\,.$$

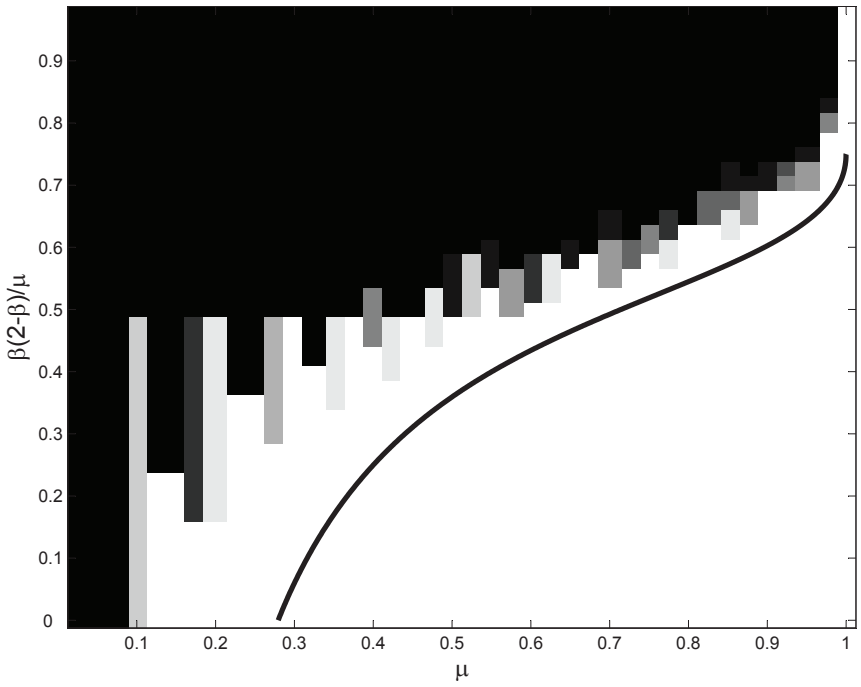
Figure 6.3 displays the results of these experiments for $n = 30$ and 40. The color of the cell in the figures reflects the empirical recovery rate of the 10 runs (scaled between 0 and 1). White denotes perfect recovery in all experiments, and black denotes failure for all experiments. It is remarkable to note that not only are the plots very similar for $n = 30$ and $n = 40$, but that the Weak Bound falls completely within the white region and is an excellent approximation of the boundary between success and failure for large β.

6.5 Discussion and Future Work

Future work should investigate if the probabilistic analysis that provides the bounds in Theorems 6.1.2 and 6.1.3 can be further tightened at all. There are two particular regions where the bounds can be improved. First, when $\beta = 0$, μ should also equal zero. However, in our Weak Bound, $\beta = 0$ tells us that μ must be greater than or equal to 0.2795. In order to provide estimates of the behavior for small values of μ, we will need to find a different lower bound than (6.3.5). When μ is small, M in (6.3.5) is very large causing the bound on the expected value to be negative. This suggests that a different parametrization of the null space of $\mathcal{A}$ could be the key to a better bound for small values of β. For large values of β, the bound is a rather good approximation of empirical results, and it might not be possible to further tighten this bound. However, it is still worth looking to see if some of the techniques in [DT05a, DT05b] on neighborly polytopes can be generalized to yield tighter approximations of the recovery region. It would also be of interest to construct a *necessary* condition, parallel to the sufficient condition of Section 6.3.1, and apply a similar probabilistic analysis to yield an upper bound for the phase transition.



(a)



(b)

Figure 6.3: Random rank recovery experiments for (a) $n = 30$ and (b) $n = 40$. The color of each cell reflects the empirical recovery rate. White denotes perfect recovery in all experiments, and black denotes failure for all experiments. In both frames, we plot the Weak Bound (6.1.6), showing that the predicted recovery regions are contained within the empirical regions, and the boundary between success and failure is well approximated for large values of β.

The comparison theorem techniques in this chapter add a novel set of tools to the behavior of the nuclear norm heuristic, and they may be very useful in the study of other rank minimization scenarios. For example, the structured problems that arise in control theory can be formulated in the form of (6.1.1) with a very structured $\mathcal{A}$ operator (see, e.g., [RXH08b]). It would be of interest to see if these structured problems can also be analyzed within the null-space framework. Using the particular structure of the null-space of $\mathcal{A}$ in these specialized problems may provide sharper bounds for these cases. For example, a problem of great interest is the Matrix Completion Problem where we would like to reconstruct a low-rank matrix from a small subset of its entries. In this scenario, the operator $\mathcal{A}$ reveals a few of the entries of the unknown low-rank matrix, and the null-space of $\mathcal{A}$ is simply the set of matrices that are zero in the specified set. The Gaussian comparison theorems studied above cannot be directly applied to this problem, but it is possible that generalizations exist that could be applied to the Matrix Completion problem and could possibly tighten the bounds provided in [CR09].

6.6 Appendix

6.6.1 Rank-Deficient Case of Theorem 6.1.1

As promised above, here is the completion of the proof of Theorem 6.1.1

Proof. In an appropriate basis, we may write

$$X_0 = \begin{bmatrix} X_{11} & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad X_* - X_0 = Y = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix}.$$

If Y_{11} and $Y_{22} - Y_{21}Y_{11}^{-1}Y_{12}$ have full rank, then all our previous arguments apply. Thus, assume that at least one of them is not full rank. Nonetheless, it is always

possible to find an *arbitrarily small* $\epsilon > 0$ such that

$$Y_{11} + \epsilon I \quad \text{and} \quad \begin{bmatrix} Y_{11} + \epsilon I & Y_{12} \\ Y_{21} & Y_{22} + \epsilon I \end{bmatrix}$$

are full rank. This, of course, is equivalent to having $Y_{22} + \epsilon I - Y_{21}(Y_{11} + \epsilon I)^{-1} Y_{12}$

full rank. We can write

$$
\begin{aligned}
\|X_*\|_* &= \|X_0 + X_* - X_0\|_* \\
&= \left\| \begin{bmatrix} X_{11} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} \right\|_* \\
&\geq \left\| \begin{bmatrix} X_{11} - \epsilon I & 0 \\ 0 & Y_{22} - Y_{21}(Y_{11} + \epsilon I)^{-1} Y_{12} \end{bmatrix} \right\|_* \\
&\quad - \left\| \begin{bmatrix} Y_{11} + \epsilon I & Y_{12} \\ Y_{21} & Y_{21}(Y_{11} + \epsilon I)^{-1} Y_{12} \end{bmatrix} \right\|_* \\
&= \|X_{11} - \epsilon I\|_* + \left\| \begin{bmatrix} 0 & 0 \\ 0 & Y_{22} - Y_{21}(Y_{11} + \epsilon I)^{-1} Y_{12} \end{bmatrix} \right\|_* \\
&\quad - \left\| \begin{bmatrix} Y_{11} + \epsilon I & Y_{12} \\ Y_{21} & Y_{21}(Y_{11} + \epsilon I)^{-1} Y_{12} \end{bmatrix} \right\|_* \\
&\geq \|X_0\|_* - r\epsilon + \left\| \begin{bmatrix} \epsilon I - \epsilon I & 0 \\ 0 & Y_{22} - Y_{21}(Y_{11} + \epsilon I)^{-1} Y_{12} \end{bmatrix} \right\|_* \\
&\quad - \left\| \begin{bmatrix} Y_{11} + \epsilon I & Y_{12} \\ Y_{21} & Y_{21}(Y_{11} + \epsilon I)^{-1} Y_{12} \end{bmatrix} \right\|_* \\
&\geq \|X_0\|_* - 2r\epsilon \quad + \left\| \begin{bmatrix} -\epsilon I & 0 \\ 0 & Y_{22} - Y_{21}(Y_{11} + \epsilon I)^{-1} Y_{12} \end{bmatrix} \right\|_* \\
&\quad - \left\| \begin{bmatrix} Y_{11} + \epsilon I & Y_{12} \\ Y_{21} & Y_{21}(Y_{11} + \epsilon I)^{-1} Y_{12} \end{bmatrix} \right\|_* \\
&\geq \|X_0\|_* - 2r\epsilon,
\end{aligned}
$$

where the last inequality follows from the condition of part 1 and noting that

$$
X_0 - X_* = \begin{bmatrix} -\epsilon I & 0 \\ 0 & Y_{22} - Y_{21}(Y_{11} + \epsilon I)^{-1} Y_{12} \end{bmatrix} + \begin{bmatrix} Y_{11} + \epsilon I & Y_{12} \\ Y_{21} & Y_{21}(Y_{11} + \epsilon I)^{-1} Y_{12} \end{bmatrix},
$$

lies in the null space of $\mathcal{A}(\cdot)$ and the first matrix above has rank more than r. But, since ϵ can be arbitrarily small, this implies that $X_0 = X_*$. ■

6.6.2 Lipshitz Constants of F_I and F_S

We begin with the proof of Lemma 6.3.9 and then use this to estimate the Lipshitz constant in Lemma 6.3.6.

Proof of Lemma 6.3.9. Note that the function F_S is convex as we can write as a supremum of a collection of convex functions

$$F_S(X_1, \ldots, X_M) = \sup_{\|v\|_{\ell_2}=1} \sup_{\|Z\|<1} \langle \sum_{i=1}^{M} v_i X_i, Z \rangle \,. \tag{6.6.1}$$

The Lipshitz constant L is bounded above by the maximal norm of a subgradient of this convex function. That is, if we denote $\bar{X} := (X_1, \ldots, X_M)$, then we have

$$L \leq \sup_{\bar{X}} \sup_{\bar{Z} \in \partial F_S(\bar{X})} \left(\sum_{i=1}^{M} \|Z_i\|_F^2 \right)^{1/2} .$$

Now, by (6.6.1), a subgradient of F_S at $\bar{X}$ is given of the form $(v_1 Z, v_2 Z, \ldots, v_M Z)$ where v has norm 1 and Z has operator norm 1. For any such subgradient

$$\sum_{i=1}^{M} \|v_i Z\|_F^2 = \|Z\|_F^2 \leq D_1$$

bounding the Lipshitz constant as desired. ■

Proof of Lemma 6.3.6. For $i = 1, \ldots, M$, let $X_i, \hat{X}_i \in \mathcal{R}^{D_1 \times D_2}$, and $Y_i, \hat{Y}_i \in \mathcal{R}^{D_3 \times D_4}$. Let

$$w^* = \arg \min_{\|w\|_{\ell_2}=1} \| \sum_{i=1}^{M} w_i \hat{X}_i \|_* - \| \sum_{i=1}^{M} w_i \hat{Y}_i \|_* \,.$$

Then we have that

$$
\begin{aligned}
&F_I(X_1,\ldots,X_M,Y_1,\ldots,Y_M) - F_I(\hat{X}_1,\ldots,\hat{X}_M,\hat{Y}_1,\ldots,\hat{Y}_M)\\
=&\left(\inf_{\|v\|_{\ell_2}=1}\|\sum_{i=1}^M v_iX_i\|_* - \|\sum_{i=1}^M v_iY_i\|_*\right) - \left(\inf_{\|w\|_{\ell_2}=1}\|\sum_{i=1}^M w_i\hat{X}_i\|_* - \|\sum_{i=1}^M w_i\hat{Y}_i\|_*\right)\\
\leq&\|\sum_{i=1}^M w_i^*X_i\|_* - \|\sum_{i=1}^M w_i^*Y_i\|_* - \|\sum_{i=1}^M w_i^*\hat{X}_i\|_* + \|\sum_{i=1}^M w_i^*\hat{Y}_i\|_*\\
\leq&\|\sum_{i=1}^M w_i^*(X_i-\hat{X}_i)\|_* + \|\sum_{i=1}^M w_i^*(Y_i-\hat{Y}_i)\|_*\\
\leq&\sup_{\|w\|_{\ell_2}=1}\|\sum_{i=1}^M w_i(X_i-\hat{X}_i)\|_* + \|\sum_{i=1}^M w_i(Y_i-\hat{Y}_i)\|_*\\
=&\sup_{\|w\|_{\ell_2}=1}\|\sum_{i=1}^M w_i\tilde{X}_i\|_* + \|\sum_{i=1}^M w_i\tilde{Y}_i\|_*
\end{aligned}
$$

where $\tilde{X}_i = X_i - \hat{X}_i$ and $\tilde{Y}_i = Y_i - \hat{Y}_i$. This last expression is a convex function of $\tilde{X}_i$ and $\tilde{Y}_i$,

$$\sup_{\|w\|_{\ell_2}=1}\|\sum_{i=1}^M w_i\tilde{X}_i\|_* + \|\sum_{i=1}^M w_i\tilde{Y}_i\|_* = \sup_{\|w\|_{\ell_2}=1}\sup_{\|Z_X\|<1}\sup_{\|Z_Y\|<1}\langle\sum_{i=1}^M w_i\tilde{X}_i, Z_X\rangle + \langle\sum_{i=1}^M w_i\tilde{Y}_iZ_Y\rangle$$

with Z_X $D_1 \times D_2$ and Z_Y $D_3 \times D_4$. Using an identical argument as the one presented in the proof of Lemma 6.3.9, we have that a subgradient of this expression is of the form

$$(w_1Z_X, w_2Z_X, \ldots, w_MZ_X, w_1Z_Y, w_2Z_Y, \ldots, w_MZ_Y)$$

where w has norm 1 and Z_X and Z_Y have operator norms 1, and thus

$$\sum_{i=1}^M \|w_iZ_X\|_F^2 + \|w_iZ_Y\|_F^2 = \|Z_X\|_F^2 + \|Z_Y\|_F^2 \leq D_1 + D_3$$

completing the proof. ■

6.6.3 Compactness Argument for Comparison Theorems

Proposition 6.6.1. *Let Ω be a compact metric space with distance function ρ. Suppose that f and g are real-valued function on Ω such that f is continuous and for any finite subset $X \subset \Omega$*

$$\max_{x\in X} f(x) \leq \max_{x\in X} g(x)\,.$$

Then

$$\sup_{x\in \Omega} f(x) \leq \sup_{x\in \Omega} g(x)\,.$$

Proof. Let $\epsilon > 0$. Since f is continuous and Ω is compact, f is uniformly continuous on Ω. That is, there exists a $\delta > 0$ such that for all $x, y \in \Omega$, $\rho(x, y) < \delta$ implies $|f(x) - f(y)| < \epsilon$. Let X_δ be a δ-net for Ω. Then, for any $x \in \Omega$, there is a y in the δ-net with $\rho(x, y) < \delta$ and hence

$$f(x) \leq f(y) + \epsilon \leq \sup_{z\in X_\delta} f(z) + \epsilon \leq \sup_{z\in X_\delta} g(z) + \epsilon \leq \sup_{z\in \Omega} g(z) + \epsilon\,.$$

Since this holds for all $x \in \Omega$ and $\epsilon > 0$, this completes the proof. ∎

Chapter 7

Conclusions and Future Work

Compressed sensing is a very active area of research which holds great promise for acquiring and processing massive data efficiently and accurately. Given signals or objects with sparse contents relative to its dimension, compressed sensing seeks to reconstruct the signals from as few non-adaptive linear measurements as possible. With its deep roots in fundamental theories and its great promise for various applications, the area of compressive sensing has seen an explosion of research activities recently and has attracted researchers from applied mathematics, optimization theory, information theory and mathematical statics. By now the area of compressive sensing has continued to thrive and has already had impacts on a diverse area of fields, including signal processing, medical imaging, function approximation, radar, MRI and high-speed analog-to-digital conversion. In this concluding chapter, we summarize our observations and point out directions for future work.

7.1 Summary and Directions for Future Work

7.1.1 Expander Graphs for Compressive Sensing

In Chapter 2, we proposed an expander graph based compressive sensing scheme, which, for the first time, comes, all at the same time, with explicit constructions of sensing matrices, linear complexity decoding algorithms and optimally scaling sparsity recovery capability. In compressive sensing, random measurement matrices are gener-

ally used and ℓ_1 minimization algorithms, such as Basis Pursuit algorithms, often use linear programming or other optimization methods to recover the sparse signal vectors. However, explicitly constructible measurement matrices providing performance guarantees were elusive and the ℓ_1 minimization methods have not yet had strong polynomial-time algorithms and are very demanding in computational complexity for some applications. Later on, it has been shown that the expander-graph based measurement matrices can further lead to fast and robust sparse signal recoveries with ℓ_1-minimization or iterative algorithms [BGI$^+$08, BI08, IR08, BIR08].

However, unlike the case of dense sensing matrices satisfying the restricted isometry conditions [CT05] or matrices with null spaces as *almost Euclidean* subspaces [GLR08], it is not known whether the expander-graph based schemes will be able to provide approximation guarantees with respect to ℓ_2 norm. It is then very interesting to validate or disprove the approximate guarantees in ℓ_2 norm for expander-graph based compressive sensing. Meanwhile, it has been observed empirically that ℓ_1 minimization for binary and sparse matrices has similar sparsity recovery thresholds as dense measurement matrices, but there are not results theoretically proving this phenomenon while the exact sparsity thresholds have been established for dense random Gaussian measurement matrix [Don06c].

7.1.2 Grassmann Angle Analytical Framework for Subspaces Balancedness

ℓ_1 minimization algorithm, namely basis pursuit algorithms, generally give better sparsity recovery performances than known greedy decoding algorithms in compressive sensing. In Chapter 3, starting from a necessary and sufficient null-space balancedness condition for achieving a certain signal recovery accuracy, using high-dimensional geometry, we gave a unified *null-space Grassmann angle*-based analytical framework for ℓ_1 minimization in compressive sensing. Sharp bounds on the balancedness property, namely the property that any small portion of every vector from a certain subspace only takes a small portion of the ℓ_1 norm, have been derived in

this thesis, thus giving sharp quantitative tradeoffs between the signal sparsity and the recovery accuracy of the ℓ_1 optimization for approximately sparse signals. Our result is tight for the weak balancedness discussed in this thesis and while for the strong and sectional balancedness discussed in this thesis, our result is tight up to one use of union bounding.

It is of great theoretical interest to develop methods to get even tighter bounds for the balancedness property of linear subspaces, which is of both mathematical interest and also of practical interest in compressive sensing. In Chapter 3, we are mainly concerned with the achievable (lower) bounds on the sparsity level such that the balancedness of linear subspaces holds. In fact, tight upper bounds on the sparsity level over which the linear subspaces balancedness property can not possibly hold might be interesting as well. One may also wonder how to extend the Grassmann angle methods or to develop new methods in studying the *almost Euclidean* subspaces [GLR08].

7.1.3 Weighted ℓ_1 Minimization Algorithm

In Chapter 4, we looked at the problem of compressive sensing with prior information. 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 Chapter 4, we considered a particular model for the sparse signal that assigns a probability of being zero or nonzero to each entry of the unknown vector. The standard compressed sensing model is therefore a special case where these probabilities are all equal. Following the introduction of the *null-space Grassmann angle*-based analytical framework in this thesis, we were able to characterize the optimal recoverable sparsity thresholds using weighted ℓ_1 minimization algorithms with the prior information.

We should notice that in Chapter 4, we used the widely used Gaussian measurement matrix ensemble, where the Grassmann angle-based framework applies, and it would be of interest to study weigthed ℓ_1 minimization algorithm with other mea-

surement matrix ensembles as well as other forms of prior information.

7.1.4 An Analysis for Iterative Reweighted ℓ_1 Minimization Algorithm

In Chapter 5, we presented a way of analyzing the Iterative Reweighted ℓ_1 Minimization Algorithm and tried to provide a theoretical foundation for analyzing such a type of algorithms. In particular, we showed that for a nontrivial class of compressible signals, the iterative reweighted ℓ_1 minimization can indeed deliver larger recoverable sparsity thresholdsthan the ℓ_1 minimization does. Again, our results were based on the *null-space Grassmann angle*-based analytical framework.

Although some insights have been given about how the iterative reweighted ℓ_1 minimization increases the recoverable sparsity thresholds for a certain class of signals under the Gaussian measurement matrix ensemble, we would like to see eventually the analysis for more general measurement matrices and more general compressible signals such as Gaussian distributed signals.

7.1.5 Null Space Conditions and Thresholds for Rank Minimization

In Chapter 6, we worked on the problem of recovering matrices of low ranks from compressed linear measurements through a heuristic of nuclear norm minimization. We assessed the practical performance of this heuristic for finding the minimum rank matrix subject to linear constraints. By characterizing a necessary and sufficient condition for the nuclear norm minimization to succeed, we provided the probabilistic performance bounds on the ranks as a function of the matrix dimensions and the number of constraints, for which the nuclear norm minimization heuristic succeeds with overwhelming probability. The performance bounds we derived are tight in some regimes, especially the number of measurements and provided accurate predictions of the heuristic's performance in non-asymptotic scenarios.

Future work along this line should investigate if the probabilistic analysis that provides the bounds in Chapter 6 can be further tightened, for Gaussian measurement ensembles or for measurement ensembles from other distributions. A problem of great interest is the Matrix Completion Problem where we would like to reconstruct a low-rank matrix from a small subset of its entries. It is possible that generalizations of our work exist that could be applied to the Matrix Completion problem.

7.2 Discussion

There are many interesting open problems from compressive sensing which are closely related to the topics studied in this thesis. One prominent one is the seeking of sharp analysis for non-Gaussian measurement matrix. It has been observed in empirical experiments that almost all the considered ensembles of measurement matrices seem to have the same sparsity recovery threshold, and it is thus natural to wonder whether a universal law can be established for this phenomenon. Till now, only the sparsity recovery threshold for Gaussian measurement ensemble matrix has been precisely established. We have shown that for a certain class of signals, the iterative reweighted ℓ_1-minimization algorithm can indeed increase the sparsity threshold compared with the standard ℓ_1-minimization algorithm, but it remains a big open problem whether there exists a polynomial-time optimization algorithm which can *provably* deliver larger sparsity recovery thresholds than the standard ℓ_1-minimization algorithm for arbitrary signals.

Bibliography

[AMP08] Andreas Argyriou, Charles A. Micchelli, and Massimiliano Pontil. Convex multi-task feature learning. *Machine Learning*, 2008. Published online first at `http://www.springerlink.com/`.

[AS92] Fernando Affentranger and Rolf Schneider. Random projections of regular simplices. *Discrete Comput. Geom.*, 7(3):219–226, 1992.

[AT07] M. Akcakaya and V. Tarokh. A frame construction and a universal distortion bound for sparse representations. *IEEE Int. Symposium on Information Theory (ISIT), Nice, France*, 2007.

[AV09] Brendan P. W. Ames and Stephen A. Vavasis. Nuclear norm minimization for the planted clique and biclique problems. 2009. Submitted to Mathematical Programming. Preprint available at `http://arxiv.org/abs/0901.3348v1`.

[Bai99] Z. D. Bai. Methodologies in spectral analysis of large dimensional random matrices. *Statistica Sinica*, 9(3):611–661, 1999.

[BD] T. Blumensath and M. E. Davies. Iterative hard thresholding for compressed sensing. to appear in Applied and Computational Harmonic Analysis.

[BD98] Carolyn Beck and Raffaello D'Andrea. Computational study and comparisons of LFT reducibility methods. In *Proceedings of the American Control Conference*, 1998.

[BDDW08] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 2008. Preprint available at `http://dsp.rice.edu/cs/jlcs-v03.pdf`.

[BGI+08] R. Berinde, A. Gilbert, P. Indyk, H. Karloff, and M. Strauss. Combining geometry and combinatorics: a unified approach to sparse signal recovery. *Proceedings of the Forty-sixth Annual Allerton Conference on Communication, Control, and Computing*, 2008.

[BI08] R. Berinde and P. Indyk. Sparse recovery using sparse random matrices. *Technical Report, MIT*, 2008.

[BIR08] R. Berinde, P. Indyk, and M. Ruzic. Practical near-optimal sparse recovery in the ℓ_1 norm. *Proceedings of the Forty-Sixth Annual Allerton Conference on Communication, Control, and Computing*, 2008.

[BM01] D. Burshtein and G. Miller. Expander graph arguments for message passing algorithms. *IEEE Transactions on Information Theory*, 47:782–790, 2001.

[Boo86] W. M. Boothby. *An Introduction to Differential Manifolds and Riemannian Geometry*. Springer-Verlag, 2nd Edition, 1986. San Diego, CA: Academic.

[Bou89] J. Bourgain. Bounded orthogonal systems and the $\lambda(p)$-set problem. *Acta Math.*, 162:227–245, 1989.

[boy] `http://www.stanford.edu/~boyd/cvx/`.

[BP73] L. Bassalygo and M. Pinsker. Complexity of an optimum nonblocking switching network without reconnections. *Problem in Information Transmission, vol 9 no 1, pp. 289-313*, 1973.

[BSW99] E. Ben-Sasson and A. Wigderson. Short proofs are narrow—resolution made simple. *Proceedings of the Annual ACM Symposium on Theory of Computing*, pages 517–526, 1999.

[BT] J. Bourgain and L. Tzafrir. Complements of subspaces of $\ln p$, $p \geq 1$, which are uniquely determined. in *Geometrical Aspects of Functional Analysis*, (1985/86), pages 39-52, number 1267 in Lecture Notes in Math., Berlin, 1987.

[BT87] J. Bourgain and L. Tzafrir. Invertibility of "large" submatrices with applications to the geometry of banach spaces and harmonic analysis. *Israel J. Math.*, 57(2):137–224, 1987.

[BT91] J. Bourgain and L. Tzafrir. On a problem of Kadison and Singer. *J. Reine Angew. Math.*, 420(2):1–43, 1991.

[Can06] E. J. Candès. Compressive sampling. In *International Congress of Mathematicians. Vol. III*, pages 1433–1452. Eur. Math. Soc., Zürich, 2006.

[CCS08] Jian-Feng Cai, Emmanuel J. Candes, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. Technical report, 2008. Preprint available at `http://arxiv.org/abs/0810.3286`.

[CDD08] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k-term approximation. *J. Amer. Math. Soc. 22 (2009), 211-231*, 2008.

[CDS98] Scott S. Chen, David L. Donoho, and Michael A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal of Scientific Computing*, 20(1):33–61, 1998.

[Che95] S.S. Chen. *Basis Pursuit*. PhD thesis, Stanford University, 1995.

[CM06] G. Cormode and S. Muthukrishnan. Combinatorial algorithms for compressed sensing. *In Proc. 40th Ann. Conf. Information Sciences and Systems, Princeton*, 2006.

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

[CR] E. Candès and J. Romberg. Practical signal recovery from random projections. Preprint, Jan. 2005.

[CR09] Emmanuel Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 2009. To Appear.

[CRT06] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.

[CRVW02] M. Capalbo, O. Reingold, S. Vadhan, and A. Wigderson. Randomness Conductors and Constant degree Expansions beyond the degree 2 Barrier. *STOC*, 2002.

[CT05] Emmanuel J. Candès and Terence Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, 2005.

[CT06] E. Candès and T. Tao. Near optimal signal recovery from random projections: Universal encoding strategies. *IEEE Transactions on Information Theory, vol. 52, pp.5406-5425*, 2006.

[CT08] Emmanuel J. Candès and Terence Tao. Reflections on compressed sensing. *IEEE Information Theory Society Newsletter*, 58(4):14–17, 2008.

[Cut] C. Chapin Cutler. Differential quantization of communication signals. U.S. patent 2605361, filed June 29, 1950, issued July 29, 1952.

[CWB08] E. J. Candès, Michel P. Wakin, and Stephen P. Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *Journal of Fourier Analysis and Application*, 14(5):877–905, 2008.

[DeV07] R. DeVore. Derministic constructions of compressed sensing matrices. *J. of Complexity, 23, pp. 918-925*, 2007.

[DL92] D. L. Donoho and B. F. Logan. Signal recovery and the large sieve. *SIAM J.Appl. Math.*, 52(2):577–591, 1992.

[Don06a] D. Donoho. For most large underdetermined systems of linear equations, the minimal ℓ_1-norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics*, 59(6):797–829, 2006.

[Don06b] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

[Don06c] David Donoho. High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension. *Discrete and Computational Geometry*, 35(4):617–652, 2006.

[DS89] D. L. Donoho and P. B. Stark. Uncertainty principles and signal recovery. *SIAM J.Appl. Math.*, 49(3):906–931, 1989.

[DT05a] David L. Donoho and Jared Tanner. Neighborliness of randomly projected simplices in high dimensions. *Proc. Natl. Acad. Sci. USA*, 102(27):9452–9457, 2005.

[DT05b] David L. Donoho and Jared Tanner. Sparse nonnegative solution of underdetermined linear equations by linear programming. *Proc. Natl. Acad. Sci. USA*, 102(27):9446–9451, 2005.

[DT06a] D. L. Donoho and Y. Tsaig. Fast solution of l_1-norm minimization problems when the solution may be sparse. 2006. Preprint.

[DT06b] David Donoho and Jared Tanner. Thresholds for the recovery of sparse solutions via l1 minimization. In *Proceedings of the Conference on Information Sciences and Systems*, 2006.

[DTDS07] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck. Sparse solution of underdetermined linear equations by stagewise Orthogonal Matching Pursuit (StOMP). 2007. Submitted for publication.

[DZ98] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*, volume 38 of *Application of Mathematics*. Springer, Second Edition, 1998.

[EGG93] Laurent El Ghaoui and P. Gahinet. Rank minimization under LMI constraints: A framework for output feedback problems. In *Proceedings of the European Control Conference*, 1993.

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

[Faz02] Maryam Fazel. *Matrix Rank Minimization with Applications*. PhD thesis, Stanford University, 2002.

[FHB01] Maryam Fazel, Haitham Hindi, and Stephen Boyd. A rank minimization heuristic with application to minimum order system approximation. In *Proceedings of the American Control Conference*, 2001.

[FN03] A. Feuer and A. Nemirovski. On sparse representation in pairs of bases. *IEEE Transactions on Information Theory*, 49(6):1579–1581, 2003.

[FR07] M. Fornasier and H. Rauhut. Iterative thresholding algorithms. 2007. Preprint.

[Fuc04] J-J Fuchs. On sparse representations in arbitrary redundant bases. *IEEE Trans. Info. Theory,50(6):1341-1344*, 2004.

[Gem80] Stuart Geman. A limit theorem for the norm of random matrices. *Annals of Probability*, 8(2):252–261, 1980.

[GG] A. Garnaev and E. Gluskin. The widths of a euclidean ball. *Dokl. A. N. USSR, vol. 277, pp. 1048-1052, 1984. In Russian.*

[GGI+02] A.C. Gilbert, S. Guba, P. Indyk, S. Muthukrishnan, and M.J. Strauss. Near-optimal sparse fourier representations via sampling. *34th ACM Symposium on Theory of Computing*, 2002.

[GLR08] V. Guruswami, J. Lee, and A. Razborov. Almost euclidean subspaces of ℓ_1 via expander codes. *Soda*, 2008.

[Gor85] Y. Gordan. Some inequalities for Gaussian processes and applications. *Israel Journal of Math*, 50:265–289, 1985.

[Gor88] Y. Gordan. Gaussian processes and almost spherical sections of convex bodies. *Annals of Probability*, 16:180–188, 1988.

[Grü68] Branko Grünbaum. Grassmann angles of convex polytopes. *Acta Math.*, 121:293–302, 1968.

[Grü03] Branko Grünbaum. *Convex polytopes*, volume 221 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, Second Edition, 2003. Prepared and with a preface by Volker Kaibel, Victor Klee and Günter M. Ziegler.

[GSTV05] A. Gilbert, M. Strauss, J. Tropp, and R. Vershynin. Sublinear, small-space approximation of compressible signals and uniform algorithmic embeddings. 2005. Preprint.

[GUV07] V. Guruswami, C. Umans, and S. Vadhan. Unbalanced expanders and randomness extractors from Parvaresh-Vardy codes. *IEEE Conference on Computational Complexity (CCC)*, 2007.

[Had79] H. Hadwiger. Gitterpunktanzahl im simplex und willssche vermutung. *Math. Ann.*, 239:271–288, 1979.

[HJ91] Roger A. Horn and Charles R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, New York, 1991.

[HLW06] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bulletin (New Series) of the American Mathematical Society, Volume 43*, 2006.

[HN06] J. Haupt and R. Nowak. Signal reconstruction from noisy random projections. *IEEE Trans. on Information Theory*, 52(9):4036–4048, 2006.

[HS07] M. Herman and T. Strohmer. High-resolution radar via compressed sensing. *IEEE Transactions on Signal Processing*, 57(6), 2007.

[Ind08] P. Indyk. Explicit constructions for compressed sensing of sparse signals. *SODA*, 2008.

[IR08] P. Indyk and M. Ruzic. Near-optimal sparse recovery in the $\ell 1$ norm. *FOCS*, 2008.

[JXC08] Shihao Ji, Ya Xue, and Lawrence Carin. Bayesian compressive sensing. *IEEE Transactions on Signal Processing*, 56(6):2346–2356, 2008.

[JXHC] S. Jafarpour, W. Xu, B. Hassibi, and R. Calderbank. Efficient and robust compressed sensing using optimized expander graphs. *IEEE Transactions on Information Theory*. To Appear, 2009.

[Kas77] B. Kashin. The widths of certain finite dimensional sets and classes of smooth functions. *Izvestia, vol. 41, pp. 334-351*, 1977.

[KBH99] Jr. Károly Böröczky and Martin Henk. Random projections of regular polytopes. *Arch. Math. (Basel)*, 73(6):465–473, 1999.

[KT07] Boris S. Kashin and Vladimir N. Temlyakov. A remark on compressed sensing. *Mathematical Notes*, 82(5):748–755, November 2007.

[KXAH09a] M. Amin Khajehnejad, Weiyu Xu, Amir Salman Avestimehr, and Babak Hassibi. Weighted ℓ_1 minimization for sparse recovery with prior informationn. In *Proceedings of the International Symposium on Information Theory*, 2009.

[KXAH09b] M. Amin Khajehnejad, Weiyu Xu, Amir Salman Avestimehr, and Babak Hassibi. Weighted ℓ_1 minimization for sparse recovery with prior informationn. 2009. Journal version, to be submitted.

[LB09] Kiryung Lee and Yoram Bresler. Efficient and guaranteed rank minimization by atomic decomposition. 2009. Submitted to ISIT2009. Preprint available at `http://arxiv.org/abs/0901.1898v1`.

[LDP07] M. Lustig, D. Donoho, , and J. M. Pauly. Sparse Mri: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.

[LLR95] Nathan Linial, Eran London, and Yuri Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15:215–245, 1995.

[LN06] N. Linial and I. Novik. How neighborly can a centrally symmetric polytope be? *Discrete and Computational Geometry*, 36(6):273–281, 2006.

[Log65] B.F. Logan. *Properties of High-pass Signals*. PhD thesis, Columbia University, 1965.

[LT91] Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces*. Springer-Verlag, Berlin, 1991.

[LV08] Zhang Liu and Lieven Vandenberghe. Interior-point method for nuclear norm approximation with application to system identification. 2008. Submitted.

[MBSR] 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.

[McM75] Peter McMullen. Non-linear angle-sum relations for polyhedral cones and polytopes. *Math. Proc. Cambridge Philos. Soc.*, 78(2):247–261, 1975.

[MDB] Michael Wakin Mark Davenport and Richard Baraniuk. Detection and estimation with compressive measurements. Rice ECE Department Technical Report TREE 0610, November 2006.

[MGC08] Shiqian Ma, Donald Goldfarb, and Lifeng Chen. Fixed point and Bregman iterative methods for matrix rank minimization. Technical report, 2008. Preprint available at `http://www.optimization-online.org/DB_HTML/2008/11/2151.html`.

[MJCD08] Raghu Meka, Prateek Jain, Constantine Caramanis, and Inderjit S. Dhillon. Rank minimization via online learning. In *Proceedings of the International Conference on Machine Learning*, 2008.

[MP67] V. A. Mar**v**cenko and L. A. Pastur. Distributions of eigenvalues for some sets of random matrices. *Math. USSR-Sbornik*, 1:457–483, 1967.

[MP97] M. Mesbahi and G. P. Papavassilopoulos. On the rank minimization problem over a positive semidefinite linear matrix inequality. *IEEE Transactions on Automatic Control*, 42(2):239–243, 1997.

[MZ93] S. Mallat and Z. Zhang. Matching Pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.*, 41(12):3397–3415, 1993.

[NT08] D. Needell and J. A. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Appl. Comp. Harmonic Analysis*, 2008.

[NV09] D. Needell and R. Vershynin. Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit. *Foundations of Computational Mathematics,Springer*, 2009.

[PH08] F. Parvaresh and B. Hassibi. Explicit measurements with almost optimal thresholds for compressed sensing. *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2008.

[PK00] P. A. Parrilo and S. Khatri. On cone-invariant linear matrix inequalities. *IEEE Trans. Automat. Control*, 45(8):1558–1563, 2000.

[PU89] D. Peleg and E. Upfal. Constructing disjoint paths on expander graphs. *Combinatorica*, 9(3):289–313, 1989.

[PV05] F. Parvaresh and A. Vardy. Correcting errors beyond the Guruswami-Sudan radius in polynomial time. *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science, pages 285-294*, 2005.

[RFP] Benjamin Recht, Maryam Fazel, and Pablo Parrilo. Guaranteed minimum rank solutions of matrix equations via nuclear norm minimization. Submitted. Preprint Available at `http://www.ist.caltech.edu/~brecht/publications.html`.

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

[RS05] Jason D. M. Rennie and Nathan Srebro. Fast maximum margin matrix factorization for collaborative prediction. In *Proceedings of the International Conference of Machine Learning*, 2005.

[RV05] Mark Rudelson and Roman Vershynin. Geometric approach to error correcting codes and reconstruction of signals. *International Mathematical Research Notices*, 64:4019–4041, 2005.

[RXH08a] Benjamin Recht, Weiyu Xu, and Babak Hassibi. Necessary and sufficient conditions for success of the nuclear norm heuristic for rank minimiza-

tion. Technical report, California Institute of Technology, 2008. Preprint available at `http://arxiv.org/abs/0809.1260`.

[RXH08b] Benjamin Recht, Weiyu Xu, and Babak Hassibi. Necessary and sufficient conditions for success of the nuclear norm heuristic for rank minimization. In *Prooceedings of the 47th IEEE Conference on Decision and Control*, 2008.

[San52] L. A. Santaló. Geometría integral enespacios de curvatura constante. *Rep. Argetina Publ. Com. Nac. Energí Atómica, Ser. Mat 1, No. 1*, 1952.

[SBB06a] S. Sarvotham, D. Baron, and R. Baraniuk. Compressed sensing reconstruction via belief propagation. *Rice ECE Department Technical Report TREE 0601*, 2006.

[SBB06b] S. Sarvotham, Dror Baron, and Richard Baraniuk. Sudocodes - fast measurement and reconstruction of sparse signals. In *Proc. IEEE Int. Symposium on Information Theory(ISIT), Seattle, WA*, 2006.

[Sil85] J. W. Silverstein. The smallest eigenvalue of a large dimensional Wishart matrix. *The Annals of Probability*, 13:1364–1368, 1985.

[Sle62] D. Slepian. The one-sided barrier problem for Gaussian noise. *Bell System Technical Journal*, 41:463–501, 1962.

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

[SS86] F. Santosa and W. W. Symes. Linear inversion of band-limited reflection seismograms. *SIAM J. Sci. Stat. Comput.*, 7(4):1307–1330, 1986.

[SS96] M. Sipser and D. Spielman. Expander codes. *IEEE transaction on Information Theory, 42(6):1710-1722*, 1996.

[Stu99] Jos F. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 11-12:625–653, 1999.

[SXH08a] Mihailo Stojnic, Weiyu Xu, and Babak Hassibi. Compressed sensing - probabilistic analysis of a null-space characterization. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2008.

[SXH08b] Mihailo Stojnic, Weiyu Xu, and Babak Hassibi. Compressed sensing of approximately sparse signals. In *IEEE International Symposium on Information Theory*, 2008.

[Sza98] Stanisław J. Szarek. Metric entropy of homogeneous spaces. In *Quantum probability (Gdańsk, 1997)*, volume 43 of *Banach Center Publ.*, pages 395–410. Polish Acad. Sci., Warsaw, 1998. Preprint available at `arXiv:math/9701213v1`.

[Tal96] M. Talagrand. Majorizing measures: The generic chaining. *Ann. Probab.*, 24(3):1049–1103, 1996.

[Tao07] Terence Tao. Open question: deterministic uup matrices. 2007. Weblog at `http://terrytao.worldpress.com`.

[TBM79] H. L. Taylor, S. C. Banks, and J. F. McCoy. Deconvolution with the ℓ_1 norm. *Geophysics*, 44(1):39–52, 1979.

[TG] J. Tropp and A. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. on Information Theory, 53(12):4655-4666, December 2007.*

[Tib96] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, B*, 58(1):267–288, 1996.

[TWD+06] J. Tropp, M. Wakin, M. Duarte, D. Baron, and R. Baraniuk. Random filters for compressive sampling and reconstruction. *in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2006.

[UW87] E. Upfal and A. Wigderson. How to share memory in a distributed system. *Journal of the Association of Computing Machines*, 34:116–127, 1987.

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

[VS92] A. M. Vershik and P. V. Sporyshev. Asymptotic behavior of the number of faces of random polyhedra and the neighborliness problem. *Selecta Mathematica Sovietica*, 11(2):181–201, 1992.

[Wai06] M. Wainwright. Sharp thresholds for noisy and high-dimensional recovery of sparsity using $\ell 1$-constrained quadratic programming. *Technical report, UC Berkeley, Department of Statistics*, 2006.

[WLD+06] M. Wakin, J. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. Kelly, and R. Baraniuk. An architecture for compressive imaging. In *Proc. International Conference on Image Processing – ICIP 2006*, Oct 2006.

[WNB06] M. Wassermann, A. Neisser, and C. Bruck. Eine serodiagnostische reaktion bei syphilis. *Deutsche Medicinische Wochenschrift*, 32:745–746, 1906.

[WS06] Kilian Q. Weinberger and Lawrence K. Saul. Unsupervised learning of image manifolds by semidefinite programming. *International Journal of Computer Vision*, 70(1):77–90, 2006.

[XH07a] Weiyu Xu and Babak Hassibi. Efficient compressive sensing with deterministic guarantees using expander graphs. *Proceedings of IEEE Information Theory Workshop, Lake Tahoe*, 2007.

[XH07b] Weiyu Xu and Babak Hassibi. Further results on performance analysis for compressive sensing using expander graphs. *Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers, 621-625, November*, 2007.

[XH08] Weiyu Xu and Babak Hassibi. Compressed sensing over the Grassmann manifold: A unified analytical framework. *Proceedings of the Forty-Sixth Annual Allerton Conference on Communication, Control, and Computing*, 2008.

[YAU07] N. Srebro Y. Amit, M. Fink and S. Ullman. Uncovering shared structures in multiclass classification. In *Proceedings of the International Conference of Machine Learning*, 2007.

[YELM07] M. Yuan, A. Ekici, Z. Lu, and R. Monteiro. Dimension reduction and coeffient estimation in multivariate linear regression. *Journal of the Royal Statistical Society: Series B*, 69:329–346, 2007.

[Zha06] Y. Zhang. When is missing data recoverable? 2006. available online at http://www.dsp.ece.rice.edu/cs/.