RICE UNIVERSITY

A Study on Conditions for Sparse Solution Recovery in Compressive Sensing

by

Anatoly Eydelzon

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

Yin Zhang, Chairman
Professor of Computational and Applied Mathematics

Richard G. Baraniuk
Victor E. Cameron Professor of Electrical and Computer Engineering

Mark Embree
Associate Professor of Computational and Applied Mathematics

Richard Tapia
University Professor, Maxfield-Oshman Professor in Engineering

HOUSTON, TEXAS

AUGUST 30, 2007
INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI®

UMI Microform 3309865
Copyright 2008 by ProQuest LLC.
All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest LLC
789 E. Eisenhower Parkway
PO Box 1346
Ann Arbor, MI 48106-1346
Abstract

A Study on Conditions for Sparse Solution Recovery in Compressive Sensing

by

Anatoly Eydelzon

It is well-known by now that under suitable conditions $\ell_1$ minimization can recover sparse solutions to under-determined linear systems of equations. More precisely, by solving the convex optimization problem $\min\{\|x\|_1 : Ax = b\}$, where $A$ is an $m \times n$ measurement matrix with $m < n$, one can obtain the sparsest solution $x^*$ to $Ax = b$ provided that the measurement matrix $A$ has certain properties and the sparsity level $k$ of $x^*$ is sufficiently small. This fact has led to active research in the area of compressive sensing and other applications.

The central question for this problem is the following. Given a type of measurements, a signal's length $n$ and sparsity level $k$, what is the minimum measurement size $m$ that ensures recovery? Or equivalently, given a type of measurements, a signal length $n$ and a measurement size $m$, what is the maximum recoverable sparsity level $k$?

The above fundamental question has been answered, with varying degrees of precision, by a number of researchers for a number of different random or semi-random measurement matrices. However, all the existing results still involve unknown constants of some kind and thus are unable to provide precise answers to specific situations. For example, let $A$ be an $m \times n$ partial DCT matrix with $n = 10^7$ and
\( m = 5 \times 10^3 \ (n/m = 20) \). Can we provide a reasonably good estimate on the maximum recoverable sparsity \( k \)?

In this research, we attempt to provide a more precise answer to the central question raised above. By studying new sufficient conditions for exact recovery of sparse solutions, we propose a new technique to estimate recoverable sparsity for different kinds of deterministic, random and semi-random matrices. We will present empirical evidence to show the practical success of our approach, though further research is still needed to formally establish its effectiveness.
Dedicated to the memory of my mother

Irene Eydelzon (1947 – 2007)
Acknowledgements

It is my great pleasure to thank the people who made this thesis possible.

It is difficult to overstate my gratitude to my thesis committee, especially to my supervisor, Dr. Yin Zhang. Throughout my thesis-writing period he provided encouragement, good teaching and lots of good ideas.

I would like to thank my high school teachers: math teacher Isai Lodenzon and physics teacher Abraham Rosenblatt. I am thankful to my Master Thesis advisor Dr. Lev Brutman from University of Haifa.

I wish to thank my wife Victoria and my son David for providing a loving environment for me.

Lastly, I wish to thank my parents, Yakov and Irene Eydelzon. Obviously, this thesis would have not been possible without them.
Contents

Abstract

Acknowledgements

1 Introduction

1.1 What is Compressive Sensing? ........................................... 1
1.2 Problem Definition ...................................................... 2
1.3 LP approach to Compressive Sensing ................................. 3
1.4 Applications .............................................................. 3
   1.4.1 Sparse Data Compression ......................................... 3
   1.4.2 Error Correction .................................................. 4
   1.4.3 An Unusual Application ......................................... 4
1.5 Organization of This Thesis ........................................... 5

2 Review of Existing Results ................................................. 6

2.1 An Equivalent Problem .................................................. 6
2.2 Necessary and Sufficient Conditions
   for Recovery .............................................................. 7
2.3 Deterministic Approach to Recovery .................................. 10
   2.3.1 Sparse Representation in a Union of Bases ................... 10
   2.3.2 Recovery for the Nonnegative Case .......................... 11
2.4 Recovery by Random Spaces ......................................... 13
2.4.1 Why Random Spaces? ........................................ 13
2.4.2 Recovery From Gaussian Measurements .................. 13
2.4.3 Recovery by Other Types of Random Matrices .......... 15
2.4.4 Recovery from Fourier Measurements ................... 15
2.5 Garnaev-Gluskin Theorem ................................... 16

3 New Sufficient Conditions ........................................ 19
3.1 Sufficient Conditions for Recovery .......................... 19
3.2 Computational Aspects of $\Gamma_\infty$ and $\Gamma_2$ ....... 20
  3.2.1 $\Gamma_\infty$ Optimization Problem ......................... 20
  3.2.2 $\Gamma_2$ Optimization Problem ............................ 21
3.3 Properties of $\Gamma_\infty$ and $\Gamma_2$ ....................... 22
3.4 Comparison of $\Gamma_\infty$ and Coherence ..................... 25

4 $\Gamma_2$ and Numerical Experiments ............................ 28
4.1 Necessary Conditions for $\Gamma_2$ ............................. 28
4.2 Preliminary Experiments ..................................... 30
  4.2.1 Experiment with Small Gaussian Matrices ............... 30
  4.2.2 Experiment with Gaussian Matrices ....................... 32
  4.2.3 Comparison of $\Gamma_2$ and $\Gamma_\infty$ .................. 40
  4.2.4 Experiments with Other Types of Random Matrices .... 41
4.3 Practical Technique to Estimate
  Recoverability .................................................. 41
  4.3.1 Some Preliminaries ..................................... 41
  4.3.2 Distribution of $\gamma_2$ on the Necessary Set ........ 42
  4.3.3 The Necessary Set vs. The Whole Space ................. 45
  4.3.4 A Single Constant vs. A Family of Constants .......... 45
  4.3.5 A Sampling Technique .................................. 49
4.4 Estimating Recoverable Sparsity ............................ 49
4.4.1 Estimation Method .............................................. 49
4.4.2 Experiments with DCT Matrices .............................. 50
4.4.3 Experiments with Other Types of Random Matrices ....... 54
4.5 Why Is Our Estimate Tighter for Smaller $m/n$? ............... 56

5 Compressive Sensing and Missing Data Recovery .......... 58
  5.1 Is Missing Data Recoverable? ............................. 58
  5.2 Examples ................................................ 60
    5.2.1 Recovery of a damaged image ...................... 60
    5.2.2 Recovery of a text image ........................... 64

6 Conclusions ....................................................... 73
  6.1 Summary .................................................. 73
  6.2 A Prediction ............................................ 74
  6.3 Future Research ........................................... 75

Bibliography ....................................................... 76
Chapter 1

Introduction

1.1 What is Compressive Sensing?

In modern society the amount of data around us has become enormous. On a daily basis we acquire, store, process and transmit gigabits of data. In some applications, like medical imaging, the amount of data is so big that costs associated with storing and processing it become prohibitive. Fortunately, most of the data types contain a lot of redundancy and can be compressed, easing the burden of storing and transmitting them.

The usual processes of data compression and transmission are encoding and decoding. The encoding stage consists of two main steps: (a) to acquire (or sense) a sufficient amount of data including redundancy, and (b) to compress the acquired data, getting rid of redundancy. The encoding process could be summarized as sensing and compressing. After data transmission comes the decoding process: to recover the original data from the compressed data with as little error as possible. With a traditional sense and compress approach, the encoding process demands many resources such as computing power and time, while the decoding process is relatively inexpensive.

Compressive sensing combines the two steps of encoding, sensing and compressing
into a single step: *compressed sensing*. It takes a smaller amount of measurements with less redundancy. Consequently, no more compression is required. On the other hand, the decoding process becomes more costly in order to recover the original data from the *incomplete* measurements.

In other words compressive sensing is a shift of the workload from pre-transmission of data to post-transmission of data. For some applications such a shift could be of great benefit. For example, a lower power demand of encoding can greatly extend the life of the battery in a digital camera, while a large increase in decoding power on a computer is affordable. Compressive sensing represents a potential breakthrough for certain applications where both the degree of data redundancy and the cost of data acquisition are relatively high. Such beneficiaries could include applications where the source of data has limited computing and power resources.

### 1.2 Problem Definition

Given a data vector $\tilde{x} \in \mathbb{R}^n$, the linear measurements $b_i$ of the data $\tilde{x}$ consist of the inner products of $\tilde{x}$ with a number of measurement vectors $a_i \in \mathbb{R}^n, i = 1, 2, \ldots, m$, that is $b_i = \langle a_i, \tilde{x} \rangle$. In matrix form $b = A\tilde{x}$, where $A$ is an $m \times n$ matrix, called the measurement or encoding matrix, that consists of $a_i$'s as its rows and $m$ is the number of measurements. In practice, taking such linear measurements is often necessary because the data vector $\tilde{x}$ is not directly observable, but can be probed through its interactions with probing vectors. Other times, if $m < n$, the transformation $b = A\tilde{x}$ is done for the purpose of compression.

If the number of measurements is less than the dimension of the data, that is, $m < n$, the linear system $A\tilde{x} = b$ is underdetermined, and therefore has infinitely many solutions, which makes the recovery of $\tilde{x}$ impossible. However if (a) the data vector $\tilde{x}$ is sufficiently sparse and (b) the encoding matrix $A$ contains a sufficient number of measurements and satisfies certain properties, then $\tilde{x}$ can be recovered.
(exactly or to a given accuracy) at a polynomial time complexity.

1.3 LP approach to Compressive Sensing

We consider the following recovery problem of a sparse vector $\tilde{x} \in \mathbb{R}^n$ from its linear measurement $b = A\tilde{x} \in \mathbb{R}^m$, where $A$ is a known $m \times n$ full rank matrix and $m < n$.

The associated optimization problem could be stated as

$$\min_{x \in \mathbb{R}^n} \{\|x\|_0 : Ax = b\}, \tag{1.1}$$

where $\|x\|_0$ is the number of nonzero entries of $x$. This problem is non-convex and therefore can not be solved by conventional optimization methods.

On the other hand we can solve the following problem which can be written as a linear program (LP) via a standard transformation,

$$\min_{x \in \mathbb{R}^n} \{\|x\|_1 : Ax = b\} \tag{1.2}$$

and ask a question:

*Under what conditions on $A$ and $\tilde{x}$ are problems (1.1) and (1.2) equivalent (have the same solution)?*

But we do not stop here, because we are interested in the exact recovery of $\tilde{x}$, that is we want the solution to (1.1) and (1.2) to be equal to $\tilde{x}$.

1.4 Applications

1.4.1 Sparse Data Compression

Suppose we have a vector $\tilde{x} \in \mathbb{R}^n$ with only a small (relative to $n$) number of nonzero entries. It is very inefficient to save or transmit a vector where most entries do not carry any information. In this case it makes sense to work with a vector’s linear measurement $b = A\tilde{x} \in \mathbb{R}^m$ instead of a vector itself, where $A$ is a known $m \times n$
matrix and $m < n$. If $m$ is considerably less than $n$, the advantage of such approach is obvious. This, of course, is true if we can recover $\tilde{x}$ from $b$ exactly or to a given accuracy.

1.4.2 Error Correction

Suppose we want to transmit a vector $\tilde{y} \in \mathbb{R}^p$. Noise is inevitable part of data transmission. To correct errors caused by noise instead of transmitting $\tilde{y}$ we will transmit a redundant linear measurement of $\tilde{y}$ given by $B^T \tilde{y}$ where $B^T$ is a $n \times p$ matrix with $p < n$. After transmission we want to recover $\tilde{y}$ from $c = B^T \tilde{y} + h$, where $h \in \mathbb{R}^n$ is an unknown (sufficiently sparse) vector of errors.

1.4.3 An Unusual Application

While Sparse Data Compression and Error Correction are standard questions of signal and image processing, Compressive Sensing should be seen rather as an innovative method giving new possibilities to technology than a complement to the existing approaches.

Recently, a single pixel camera based on Compressive Sensing approach was developed at the Department of Electrical and Computer Engineering at Rice University (http://www.dsp.ece.rice.edu/cscamera/) by Michael Wakin, Jason Laska, Marco Duarte, Dror Baron, Shriram Sarvotham, Dharmpal Takhar, Kevin Kelly, and Richard Baraniuk [34, 36]. While conventional cameras collect pixel information first, the new camera directly acquires random projections (linear measurements) of a signal. The ability to obtain an image with a single detection element and measuring (probing) the image fewer times than the number of pixels can significantly reduce the computation required for video acquisition/encoding.

Compressive Sensing can be used in Shape Estimation [37], Rapid MR Imaging [29], Missing Data Recovery [41] (see also Chapter 5) and other applications.
1.5 Organization of This Thesis

Besides of the introductory chapter this thesis is organized into five chapters.

Chapter 2 is a review of existing theoretical results in compression and decoding by linear programming.

Chapter 3 gives new theoretical results in the deterministic approach to compressive sensing.

In Chapter 4 we propose a way to estimate recovery properties of different kinds of random and semi-random matrices.

In Chapter 5 we describe an application of compressive sensing for Missing Data Recovery.

Chapter 6 is a summary of our results.
Chapter 2

Review of Existing Results

2.1 An Equivalent Problem

We need to mention that the problem (1.2) is underdetermined. Let us consider the
$l_1$-norm approximation of an overdetermined linear system

$$\min_{y \in \mathbb{R}^p} \| B^T y - c \|_1,$$

(2.1)

where $B$ is a $p \times n$ full rank matrix and $p < n$.

**Theorem 2.1.1.** Let both $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{p \times n}$ be of full rank and $p + m = n$.
Then (1.2) and (2.1) are equivalent if and only if

$$AB^T = 0 \text{ and } b = Ac.$$ 

(2.2)

Moreover, under this equivalence, if $y^*$ solves (2.1), then $c - B^T y^*$ solves (1.2), and
if $x^*$ solves (1.2), then $(BB^T)^{-1} B(c - x^*)$ solves (2.1).

In other words, Theorem 2.1.1 provides necessary and sufficient conditions under
which underdetermined and overdetermined systems
\[(O): \min_{y \in \mathbb{R}^p} \|B^T y - c\|_1, \quad (2.3)\]
\[(U): \min_{x \in \mathbb{R}^n} \{\|x\|_1 : Ax = b\} \quad (2.4)\]

are equivalent.

The (2.4) $\Rightarrow$ (2.3) part was proved by Candès and Tao in [7]. Zhang in [39] proved the (2.4) $\Leftarrow$ (2.3) part and stated Theorem 2.1.1 in its current form.

We would like to mention that the overdetermined problem (2.3) is associated with an error correction, while the undetermined problem (2.4) is associated with a compression. Having Theorem 2.1.1 as a tool we can concentrate on solving one problem to obtain solutions for both.

### 2.2 Necessary and Sufficient Conditions for Recovery

**Definition 2.2.1 (Partition).** By a partition $(S, Z)$ we mean a partition of the index set \(\{1, 2, \ldots, n\}\) into two disjoint subsets \(S\) and \(Z\) such that \(S \cup Z = \{1, 2, \ldots, n\}\) and \(S \cap Z = \emptyset\). In particular, for any \(\tilde{x} \in \mathbb{R}^n\), the partition \((S(\tilde{x}), Z(\tilde{x}))\) refers to the support \(S(\tilde{x})\) of \(\tilde{x}\) and its complement – the zero set \(Z(\tilde{x})\), namely

\[S(\tilde{x}) = \{i : \tilde{x}_i \neq 0, 1 \leq i \leq n\}, \quad Z(\tilde{x}) = \{i : \tilde{x}_i = 0, 1 \leq i \leq n\}. \quad (2.5)\]

**Definition 2.2.2 (k-sparcity).** We say that a vector \(\tilde{x}\) is k-sparse if \(|S(\tilde{x})| \leq k\).

**Definition 2.2.3 (k-balancedness, k-thickness and k-neighborness).** A subspace \(V \subseteq \mathbb{R}^n\) is k-balanced (in \(l_1\) norm) if for any partition \((S, Z)\) with \(|S| = k\)

\[\|v_S\|_1 \leq \|v_Z\|_1, \forall v \in V.\]

It is strictly k-balanced if the strict inequality holds for all \(v \neq 0\).
A subspace $V \subseteq \mathbb{R}^n$ is $k$-thick if it intersects with all $(n-k)$-dimensional faces of the unit cube $\{v \in \mathbb{R}^n : \|v\|_\infty \leq 1\}$. It is strictly $k$-thick if all the intersections lie in the relative interiors of the $(n-k)$-dimensional faces.

Let $A := [a_1 \cdots a_n] \in \mathbb{R}^{m \times n}$ be of full rank and $m < n$. The polytope
\[ P(A) := \text{conv}(\{\pm a_j : j = 1, 2, \ldots, n\}) \subset \mathbb{R}^m \]
is called (centrally) $k$-neighborly if every set of $k$ vertices of $P(A)$ not including any antipodal pair is the vertex set of a face of $P(A)$.

Definitions of $k$-balancedness and $k$-thickness were introduced by Zhang in [39]. However, $k$-balancedness was used by Donoho and Huo in [14] and by Elad and Bruckstein in [19], and $k$-thickness was used by Rudelson and Veshynin in [31] without being explicitly defined. Definition of $k$-neighborliness was introduced by Donoho in [11].

**Theorem 2.2.4** (Necessary and Sufficient Conditions for Recovery). Let $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{p \times n}$ be full rank such that $p + m = n$ and $AB^T = 0$. Let $c = B^T \hat{y} + \hat{x}$ and $b = Ac$. Then for any $\hat{x}$ with $|S(\hat{x})| \leq k$, $\hat{y}$ and $\hat{x}$ uniquely solve (1.2) and (2.1), respectively, if and only if one of the following three equivalent conditions holds: (1) range($B^T$) $\subset \mathbb{R}^n$ is strictly $k$-balanced; or (2) range($A^T$) $\subset \mathbb{R}^n$ is strictly $k$-thick; or (3) $P(A) \subset \mathbb{R}^m$ is a $k$-neighborly polytope of $2n$ vertices.

In [39] Zhang stated Theorem 2.2.4 in its current form and gave a simple proof by connecting equivalent recoverability conditions for different spaces. Part (1) of the Theorem was used without being stated explicitly by Donoho and Huo in [14] and by Elad and Bruckstein in [19] and was stated as Lemma by Grinboval and Nielsen in [23]. Part (2) of the Theorem was discussed and proved by Rudelson and Veshynin in [31]. Part (3) of the Theorem was proved by Donoho in [11].

Part (1) of Theorem 2.2.4 is of particular importance for us. As we will see in the next chapter, the number $k$ for which a given matrix $B^T$ has a $k$-balanced range can be estimated.
**Definition 2.2.5** (Restricted Isometry Constant). Let $A$ be an $m \times n$ full rank matrix and $m < n$. For a $k < n$ we define the restricted isometry constant $\delta_k$ to be the smallest positive number such that the inequality

$$C(1 - \delta_k)\|z\|^2_2 \leq \|A_P z\|^2_2 \leq C(1 + \delta_k)\|z\|^2_2$$

(2.6)

holds for some $C > 0$ and for all $z$ and all subsets $P \subset \{1, 2, \ldots, n\}$ of a size $|P| \leq k$, where $A_P$ is a $m \times |P|$ matrix that consists of the columns of $A$ indexed by $P$.

In other words this means that $A$ acts as an almost-isometry on all $O(k)$-sparse vectors.

The restricted Isometry Constant 2.2.5 was introduced by Candès and Tao in [7] and in [8] the authors proved the following theorem:

**Theorem 2.2.6** (Restricted Isometry Condition (RIC)). Let $A$ be an $m \times n$ full rank measurement matrix with $m < n$ whose restricted isometry constant satisfies

$$\delta_{3k} + 3\delta_{4k} < 2.$$  

(2.7)

Let $\tilde{x} \in \mathbb{R}^n$ be such that $|S(\tilde{x})| \leq k$. Then the solution to

$$(U) : \min_{x \in \mathbb{R}^n} \{\|x\|_1 : Ax = A\tilde{x}\}$$

(2.8)

is unique and equal to $\tilde{x}$.

The theorem says that under the restricted isometry condition on the measurement matrix $A$, the non-convex problem

$$(N) : \min_{x \in \mathbb{R}^n} \{\|x\|_0 : Ax = A\tilde{x}\}$$

(2.9)

is equivalent to a linear program for all $k$-sparse vectors $\tilde{x}$.

For a fixed measurement matrix it is impossible to use Theorem 2.2.6 directly due to the exponential number of subsets $P$ involved.
Recently, DeVore in [10] proposed a deterministic way to construct matrices satisfying RIC with $k \leq C \sqrt{m} \log m / \log(n/m)$, "which is the largest range of $k$ that is known for deterministic constructions. However, it falls far short of the range $k = Cm / \log(n/m)$ known for probabilistic constructions", as was mentioned by the author. As a result, construction of measurement matrices is randomized as we will see below.

2.3 Deterministic Approach to Recovery

2.3.1 Sparse Representation in a Union of Bases

Let $A\tilde{x} = b$. Consider again the problem defined in (1.3).

\[
(N) : \min_{x \in \mathbb{R}^n} \{\|x\|_0 : Ax = b\}, \quad (2.10)
\]

\[
(U) : \min_{x \in \mathbb{R}^n} \{\|x\|_1 : Ax = b\} \quad (2.11)
\]

where $A \in \mathbb{R}^{m \times n}$ is of full rank and $m < n$.

Definition 2.3.1 (Dictionary). We say that $A$ is a dictionary if the columns of $A$ are unit vectors.

Definition 2.3.2 (Coherence of a Dictionary). Let $A \in \mathbb{R}^{m \times n}$ be a dictionary. The coherence of a dictionary $M(A)$ is defined by

\[
M(A) = \max_{i \neq j} |\langle a_i, a_j \rangle|, \quad (2.12)
\]

where $a_i, 1 \leq i \leq n$, is the $i$-th column of $A$.

Next theorem is due to Gribnoval and Nielsen [23].

Theorem 2.3.3. Let $k$ be a natural number and let $|S(\tilde{x})| \leq k$. For any dictionary $A$, if $k < \frac{1}{2} \left( 1 + \frac{1}{M(A)} \right)$ and $b = A\tilde{x}$, then $\tilde{x}$ is the unique solution to both (2.10) and (2.11).
One can ask a natural question: How small can the coherence of a dictionary be. The next theorem, proved in [4] and [33], answers this question.

**Theorem 2.3.4.** If $m$ is a power of 2, then there exists a dictionary $A$ such that $M(A) \leq \frac{1}{\sqrt{m}}$.

In [18] Elad proposed a method to optimize a given measurement matrix $A$. As a result of such optimization, a new measurement matrix $\tilde{A}$ with a smaller coherence than $A$ is constructed, leading to a larger number $k$ for guaranteed recovery.

It is necessary to mention that computing of the coherence $M(A)$ of a dictionary $A$ is very simple. Theorem 2.3.4 proves existence of dictionaries for which $\tilde{x}$ is the unique solution to both (2.10) and (2.11) for $k$ as large as $(1 + \sqrt{m})/2$.

In the next chapter we will prove that if $m \leq n/2$, then $k \leq (1 + \sqrt{2m})/2$ is, in fact, the best the Coherence Approach can do. In other words, using the Coherence Approach we can guarantee recovery of vectors with an order of sparsity not exceeding the square root of number of measurements. The low order of guaranteed recovery is a significant disadvantage of the Coherence Approach.

### 2.3.2 Recovery for the Nonnegative Case

In the case $\tilde{x} \geq 0$, equations (2.3) and (2.4) have the following representation:

\[(O+) : \{ e^T(c - B^Ty) : B^Ty \leq c \}, \quad (2.13) \]

\[(U+) : \min \{ e^T x : Ax = b, x \geq 0 \}, \quad (2.14) \]

where $e \in \mathbb{R}^n$ is the vector of all ones. Observe, that if $x \geq 0$, then $\|x\|_1 = e^T x$, and if $c - B^Ty = x \geq 0$, then $\|c - B^Ty\|_1 = e^T(c - B^Ty)$.

When using $l_1$ minimization to recover a sparse, nonnegative solution to a underdetermined system (2.14), the highest sparsity level at which recovery can still be guaranteed equals half of the number of measurements. This fact was discovered by
Donoho and Tanner [15] by invoking classic results from the theory of convex polytopes. Zhang in [40] presented an elementary proof for this result by constructing a basis $A^T$ for an $m$-dimensional subspace of $\mathbb{R}^n$, having a special property of half \( k \)-thickness (defined below) which guarantees recovery.

We would like to mention that the equivalence Theorem 2.1.1 still applies. However, to present necessary and sufficient conditions for recovery for the nonnegative case we need to define half \( k \)-balancedness and half \( k \)-thickness of a subspace, which is somewhat similar to \( k \)-balancedness and \( k \)-thickness of a subspace defined earlier.

**Definition 2.3.5 (Half \( k \)-balancedness and half \( k \)-thickness).** A subspace $V \subseteq \mathbb{R}^n$ is half \( k \)-balanced (in $l_1$ norm) if for any partition $(S, Z)$ with $|S| = k$

$$
\epsilon^T_S v_S \leq \|v_Z\|_1, \forall v \in V.
$$

It is strictly half \( k \)-balanced if the strict inequality holds for all $v \neq 0$.

A subspace $V \subseteq \mathbb{R}^n$ is half \( k \)-thick if it intersects with all the $(n - k)$-dimensional faces of the set \( \{v \in \mathbb{R}^n : \|v\|_\infty \leq 1\} \). It is strictly half \( k \)-thick if all the intersections lie in the relative interiors of the $(n - k)$-dimensional faces.

Definitions of half \( k \)-balancedness and half \( k \)-thickness were introduced by Zhang in [40], where the author extended Theorem 2.2.4 for the nonnegative case.

**Theorem 2.3.6 (Necessary and Sufficient Conditions for Recovery for the Nonnegative Case).** Let $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{p \times n}$ be full rank such that $p + m = n$ and $AB^T = 0$. Let $c = B^T\tilde{y} + \tilde{x}$ and $b = Ac$. Then for any $\hat{x} \geq 0$ with $|S(\hat{x})| \leq k$, $\tilde{y}$ and $\hat{x}$ uniquely solve (2.13) and (2.14), respectively, if and only if one of the following three equivalent conditions holds: (1) range($B^T$) $\subseteq \mathbb{R}^n$ is strictly half \( k \)-balanced; or (2) range($A^T$) $\subseteq \mathbb{R}^n$ is strictly half \( k \)-thick; or (3) $P(A) \subset \mathbb{R}^m$ is a $k$-neighborly polytope of $2n$ vertices.
2.4 Recovery by Random Spaces

2.4.1 Why Random Spaces?

Let $A\tilde{x} = b$. Consider again the problem defined in (1.3).

\[(N) : \min_{\tilde{x} \in \mathbb{R}^n} \{\|x\|_0 : Ax = b\}, \quad (2.15)\]

\[(U) : \min_{\tilde{x} \in \mathbb{R}^n} \{\|x\|_1 : Ax = b\} \quad (2.16)\]

where $A \in \mathbb{R}^{m \times n}$ is of full rank and $m < n$. Let $B \in \mathbb{R}^{p \times n}$ be of full rank such that $p + m = n$ and $AB^T = 0$. Let $\tilde{x} \in \mathbb{R}^n$ be such that $|S(\tilde{x})| \leq k$.

The question to be answered is, what is the smallest number of measurements (or number of rows of $A$) $m = m(k, n)$ which still guarantees exact recovery of $\tilde{x}$?

Summarizing the results of Section 2.2, in order to guarantee exact recovery of $\tilde{x}$ it is enough to show that one of the four sufficient conditions holds: (1) range$(B^T) \subset \mathbb{R}^n$ is strictly $k$-balanced; or (2) range$(A^T) \subset \mathbb{R}^n$ is strictly $k$-thick; or (3) $P(A) \subset \mathbb{R}^m$ is a $k$-neighborly polytope of $2n$ vertices; or (4) $A$ satisfies the restricted isometry condition. Please note that since conditions (1), (2) and (3) are also necessary for exact recovery of $\tilde{x}$, if condition (4) holds, conditions (1), (2) and (3) will be satisfied automatically by mathematical logic.

For a fixed measurement matrix $A$ there is not a known way to check if $A$ satisfies one of the sufficient conditions (1)-(4). One of the ideas of recovery by random spaces is to show that a randomly generated measurement matrix $A$ will satisfy one of the sufficient conditions with high probability.

2.4.2 Recovery From Gaussian Measurements

In [7] Candès and Tao introduced the Restricted Isometry Constant 2.2.5 and proved that with high probability iid Gaussian matrices with mean 0 and variance 1 satisfy the Restricted Isometry Condition 2.2.6. They also showed that if measurement
matrix \( A \in \mathbb{R}^{m \times n} \) is an iid Gaussian matrix, then \( \tilde{x} \) can be exactly recovered with high probability if \( |S(\tilde{x})| \leq k \), where \( k < \rho m / \log \frac{n}{m} \) and \( \rho \) is some positive constant.

While the constant \( \rho \) remained unspecified at the time of publication, recovery by iid Gaussian matrices, described by Candès and Tao in the paper *Decoding by Linear Programming* [7], was a breakthrough in the subject of Compressive Sensing, because the order of (sparse) recovery by iid Gaussian matrices is \( m / \log \frac{n}{m} \), compared to only \( \sqrt{m} \) for the Coherence Approach.

The next theorem due to Rudelson and Vershynin [32] was the first result, where a reasonable value of an unknown constant was estimated. We should mention that the theorem was proved directly, without use of the Restricted Isometry Condition 2.2.6.

**Theorem 2.4.1** (Recovery from Gaussian measurements). An \( m \times n \) Gaussian matrix \( A \) with \( m > m(k, n) \), \( m(k, n) \leq c_1 k [c_2 + \log(n/k)](1 + o(1)) \) where \( c_1 = 6 + 4\sqrt{2} \) and \( c_2 = 1.5 \) with probability

\[
1 - 3.5 \exp(-(\sqrt{k} - \sqrt{m(k, n)})^2 / 18)
\]  

satisfies the following: Let \( \tilde{x} \in \mathbb{R}^n \) be such that \( |S(\tilde{x})| \leq k \). Then \( \tilde{x} \) can be exactly recovered from the measurements \( b = A\tilde{x} \) as a unique solution to LP (2.16).

In [16] Donoho and Tanner by answering the question *How Neighborly can a Polytope Be?* (sufficient condition (3) of Theorem 2.2.4), obtained the following estimate for iid Gaussian measurement matrices:

*If \( k \) and \( n \) are large then we only need \( m \) measurements, where

\[
m \geq 2k \log(n/m)(1 + o(1)).
\]  

The estimate (2.18) is asymptotic and as the authors mentioned in [16], page 44, "for \( n \) and \( m/n \) simultaneously small, our bounds become weak or useless." Another question is how large should \( n \) be in order to omit the \( o(1) \) term in (2.18)? With some uncertainty still left, the asymptotic formula (2.18) so far is the best result in Compressive Sensing.
2.4.3 Recovery by Other Types of Random Matrices

In [1] Baraniuk, Davenport, DeVore and Wakin proved that the Restricted Isometry Condition 2.2.6 holds for any distribution satisfying the so-called concentration of measure inequality (see for example [27]); namely they proved the following theorem:

Theorem 2.4.2. Suppose that $m$, $n$ and $0 < \delta < 1$ are given. If the probability distribution generating the $m \times n$ measurement matrices satisfies the concentration inequality, then there exist constants $c_1, c_2 > 0$ depending only on $\delta$ such that the Restricted Isometry Condition holds with the prescribed $\delta$ and any $k \leq c_1 m / \log(n/k)$ with probability $\geq 1 - e^{-c_2 m}$.

This theorem allows us to use other classes of random matrices with the same (order) recovery properties as iid Gaussian matrices (see [1] for details). For example, let $A$ be a measurement matrix whose entries $A_{i,j}$ are independent realizations of discrete distribution: $-1$ with probability $1/6$; 0 with probability $2/3$; 1 with probability $1/6$). Such matrix $A$ requires fewer additions and multiplications than iid Gaussian matrix.

2.4.4 Recovery from Fourier Measurements

Consider recovery of a vector $\hat{x} \in \mathbb{C}^n$ such that $|S(\hat{x})| \leq k$ from its discrete Fourier transform evaluated at $m = m(k,n)$ points. These points will be chosen at random and uniformly from the set $\{1, 2, \ldots, n\}$.

The Discrete Fourier Transform $\hat{x} \rightarrow \Psi \hat{x}$ is defined by the DFT matrix $\Psi$ where

$$\Psi_{\omega,t} = \frac{1}{\sqrt{n}} \exp(-i2\pi\omega t/n), \quad 0 \leq \omega, t \leq n - 1.$$  \hfill (2.19)

Next two theorems are due to Rudelson and Vershynin [32].

Theorem 2.4.3 (Recovery From Fourier Measurements). A random set $\Omega \subset \{1, 2, \ldots, n\}$ with $|\Omega| = m(k,n)$ satisfies the following with high probability. Let $\hat{x} \in \mathbb{C}^n$ be such that
$|S(\tilde{x})| \leq k$. Then $\tilde{x}$ can be exactly recovered from the values of its Fourier Transform on $\Omega$ as a solution to the LP

$$
\min_{x \in \mathbb{C}^n} \{ ||x||_1 : \Psi x = \Psi \tilde{x} \}. \quad (2.20)
$$

The next theorem gives the order of $m$ for which Theorem 2.4.3 holds.

**Theorem 2.4.4 (Sample size).** Theorem 2.4.3 holds with

$$
m(k, n) = O(k \log(n) \log^2(k) \log(k \log n)). \quad (2.21)
$$

Despite an unknown constant this theorem has a significant practical importance, since vector multiplication is fast for DFT matrices.

**Remark 2.4.5.** As authors mentioned in [32], both Theorems 2.4.3 and 2.4.4 hold for transforms more general than the Discrete Fourier Transform. The DFT matrix $\Psi$ can be replaced by any orthogonal matrix with entries of magnitude $O(1/\sqrt{n})$.

Theorem 2.4.4 improves result of Candès and Tao [9], namely $m(k, n) = O(k \log^6(n))$. The conjectured optimal estimate for Fourier measurements is $m(k, n) = O(k \log(n))$, which holds for nonuniversal measurements, that is for one sparse signal $\tilde{x}$ and for a random set $\Omega$. The latest result was established by Candès, Romberg and Tao in [6].

### 2.5 Garnaev-Gluskin Theorem

Let $V$ be a $p$-dimensional subspace of $\mathbb{R}^n$ and let $VU_1$ be the intersection of $V$ with 1-norm unit ball $U_1 = \{ u \in \mathbb{R}^n : ||u||_1 = 1 \}$. The next theorem, due to Garnaev and Gluskin [22] says that for some $p$-dimensional subspaces of $\mathbb{R}^n$ the intersection $VU_1$ can be very close to a $p$-dimensional Euclidean ball. As we will show in the next chapter, if this is the case, a subspace $V$ will be $k$-balanced for some $k > 0$ and, therefore, a basis for its orthogonal complement would be a good choice for a measurement matrix.
Theorem 2.5.1 (Garnaev-Gluskin). For any natural numbers \( m \) and \( n \) with \( 0 < m < n \), there exists a set of \( p = n - m \)-dimensional subspaces of \( \mathbb{R}^n \) that has the following property. For any \( p \)-dimensional subspace \( V \) in this set,

\[
C \sqrt{\frac{m}{1 + \log \frac{n}{m}}} \leq \frac{\|v\|_1}{\|v\|_2}, \quad \forall v \in V \setminus \{0\},
\]

where \( C \) is an absolute constants independent of \( m \) and \( n \).

Theorem 2.5.1 is an existence theorem, which does not provide a way to construct a subspace having the above mentioned property. Moreover, although the constant \( C \) is absolute, its value remains unknown. On the other hand, the authors based their proof of the theorem on iid Gaussian matrices, so we can say that a \( p = n - m \)-dimensional subspace of \( \mathbb{R}^n \) spanned by the nullspace of an \( m \) by \( n \) iid Gaussian matrix will satisfy the above mentioned property with high probability.

It is impossible to illustrate the theorem in a high dimensional space, however it can be done on the plane (see Figure 2.1). Consider three one-dimensional subspaces of \( \mathbb{R}^2 \) defined by \( V_1 = \text{span}\{(1,1)^T\} \), \( V_2 = \text{span}\{(1,2)^T\} \) and \( V_3 = \text{span}\{(1,0)^T\} \).

Let \( v_1 \in V_1 \), \( v_2 \in V_2 \) and \( v_3 \in V_3 \). Then \( \frac{\|v_1\|_1}{\|v_1\|_2} = \frac{\sqrt{1^2 + 1^2}}{\sqrt{1^2 + 1^2}} = \sqrt{2} \), \( \frac{\|v_2\|_1}{\|v_2\|_2} = \frac{\sqrt{1^2 + 2^2}}{\sqrt{1^2 + 2^2}} = \frac{3}{\sqrt{5}} \) and \( \frac{\|v_3\|_1}{\|v_3\|_2} = \frac{\sqrt{1^2 + 0^2}}{\sqrt{1^2 + 0^2}} = 1 \).

In \( V_1 \), the 1-norm “outperform” 2-norm by \( \sqrt{2} \). In fact, this is the best ratio for a one-dimensional subspace of \( \mathbb{R}^2 \). In \( V_2 \), the 1-norm “outperform” 2-norm by \( \frac{3}{\sqrt{5}} \). On the other hand, in \( V_3 \) both 1 and 2-norms are equal and it is the worst possible case. We will say that \( V_1 \) and \( V_2 \) are good subspaces, while \( V_3 \) is a bad subspace.

In general, a subspace is good if it does not contain very sparse (or close to sparse) vectors, because for such vectors 1 and 2-norms are close.
Figure 2.1: The bubble-shaped curve around the Euclidean ball is the $\frac{\|u\|}{\|u\|_2}$ ratio.
Chapter 3

New Sufficient Conditions

3.1 Sufficient Conditions for Recovery

Definition 3.1.1 ($\gamma_\infty$ and $\gamma_2$). For a $B \in \mathbb{R}^{p \times n}$ of full rank with $p < n$ and $y \in \mathbb{R}^p, y \neq 0$ we define

$$\gamma_\infty(B, y) = \frac{\|B^T y\|_1}{\|B^T y\|_\infty},$$

$$\gamma_2(B, y) = \frac{\|B^T y\|_1}{\|B^T y\|_2}.$$  \hfill (3.1)

Next two lemmas present sufficient conditions for recovery. They follow directly from the sufficient (and necessary) condition (1) of Theorem 2.2.4.

Let $B \in \mathbb{R}^{p \times n}$ be a full rank matrix. Let $k$ be a natural number. In view of Theorem 2.2.4, to prove a sufficient condition, it is enough to show that the range of $B^T$ is strictly $k$-balanced, that is, for every partition $(S, Z)$ with $|S| \leq k$,

$$\|v_S\|_1 < \|v_Z\|_1, \forall v \in \text{range}(B^T), \ v \neq 0$$  \hfill (3.3)

or equivalently

$$\|v_S\|_1 < \frac{1}{2}\|v\|_1, \forall v \in \text{range}(B^T), \ v \neq 0.$$  \hfill (3.4)
Lemma 3.1.2 (Sufficient Condition for Recovery). Recovery is guaranteed whenever 
\( k < \frac{1}{4}(\Gamma_2(B))^2 \), where

\[
\Gamma_2(B) = \min_y \gamma_2(B, y). \tag{3.5}
\]

Proof. We have:

\[
\|v_S\|_1 \leq \sqrt{k}\|v_S\|_2 \leq \sqrt{k}\|v\|_2 < \frac{1}{2} \Gamma_2(B)\|v\|_2 \leq \frac{1}{2} \|v\|_1 \|v\|_2 \leq \frac{1}{2} \|v\|_1. \tag{3.6}
\]

\(\square\)

Lemma 3.1.3 (Weak Sufficient Condition for Recovery). Recovery is guaranteed whenever \( k < \frac{1}{2} \Gamma_\infty(B) \), where

\[
\Gamma_\infty(B) = \min_y \gamma_\infty(B, y). \tag{3.7}
\]

Proof. We have:

\[
\|v_S\|_1 \leq k\|v_S\|_\infty \leq k\|v\|_\infty \leq \frac{1}{2} \Gamma_\infty(B)\|v\|_\infty \leq \frac{1}{2} \|v\|_1 \|v\|_\infty \leq \frac{1}{2} \|v\|_1. \tag{3.8}
\]

\(\square\)

3.2 Computational Aspects of \( \Gamma_\infty \) and \( \Gamma_2 \)

3.2.1 \( \Gamma_\infty \) Optimization Problem

Recall that

\[
\Gamma_\infty(B) = \min_y \frac{\|B^T y\|_1}{\|B^T y\|_\infty} = \min_{\|B^T y\|_\infty = 1} \|B^T y\|_1. \tag{3.9}
\]

The optimization set \( \{y \in \mathbb{R}^p : \|B^T y\|_\infty = 1\} \) is non-convex, however, it is a union of \( 2n \) convex sets

\[
\{y \in \mathbb{R}^p : \|B^T y\|_\infty = 1\} = \bigcup_{i=1}^{n} \pm F_i, \tag{3.10}
\]
where

\[ F_i = \{ y \in \mathbb{R}^p : [B^T y]_i = 1; \|[B^T y]_j\| \leq 1, j \neq i \}. \tag{3.11} \]

Taking into consideration that the objective function in (3.9) does not depend on the sign of \( y \), we conclude that

\[ \Gamma_\infty(B) = \min_{1 \leq i \leq n} \min_{y \in F_i} \| B^T y \|_1. \tag{3.12} \]

For every \( 1 \leq i \leq n \), \( \min_{y \in F_i} \| B^T y \|_1 \) could be rewritten as a linear program via a standard transformation. Therefore, in order to compute \( \Gamma_\infty \) we have to solve \( n \) linear programs. While it requires considerable computational efforts for a large \( n \), the problem is solvable in polynomial time.

### 3.2.2 \( \Gamma_2 \) Optimization Problem

Let \( B \in \mathbb{R}^{p \times n} \) be of full rank with \( p < n \). Without loss of generality suppose that rows of \( B \) are orthonormal. Consider the optimization problem associated with computing of \( \Gamma_2(B) \):

\[ \Gamma_2(B) = \min_{y \in \mathbb{R}^p} \gamma_2(B, y) = \min_{y \in \mathbb{R}^p} \frac{\|B^T y\|_1}{\|B^T y\|_2} = \min_{y \in \mathbb{R}^p} \frac{\|B^T y\|_1}{\|y\|_2}, \tag{3.13} \]

or equivalently

\[ \frac{1}{\Gamma_2(B)} = \max_{y \in \mathbb{R}^p} \frac{\|y\|_2}{\|B^T y\|_1} = \max_{y \in \mathbb{R}^p} \{\|y\|_2 : \|B^T y\|_1 \leq 1\}. \tag{3.14} \]

By introducing auxiliary variable \( t \in \mathbb{R}^n_+ \) we can write

\[ \frac{1}{\Gamma_2(B)} = \max_{y \in \mathbb{R}^p, t \in \mathbb{R}^n_+} \{\|y\|_2 : e^T t = 1, -t \leq B^T y \leq t\}, \tag{3.15} \]

where \( e \) is the vector of ones. Obviously,

\[ \arg \max_{y \in \mathbb{R}^p, t \in \mathbb{R}^n_+} \{\|y\|_2 : e^T t = 1, -t \leq B^T y \leq t\} = \tag{3.16} \]

\[ \arg \max_{y \in \mathbb{R}^p, t \in \mathbb{R}^n_+} \{\|y\|_2^2 : e^T t = 1, -t \leq B^T y \leq t\} \tag{3.17} \]
Let an optimal solution to (3.17) be \((y^*, t^*)\). We consider a simpler problem:
\[
\max_{y \in \mathbb{R}^p} \{\|y\|_2^2 : -t^* \leq B^T y \leq t^*\}. \tag{3.18}
\]
When \(p = n\), the feasible set of (3.18) is an arbitrary parallelogram in \(\mathbb{R}^n\). When \(p < n\), the feasible set of (3.18) is an arbitrary centrally symmetric polytope in \(\mathbb{R}^p\), that has more faces than a parallelogram in \(\mathbb{R}^p\). In both cases the problem (3.18) is NP-hard according to [2, 3].

It follows that the \(\Gamma_2\) optimization problem (3.13) is NP-hard.

### 3.3 Properties of \(\Gamma_\infty\) and \(\Gamma_2\)

Next lemma shows that \(\Gamma_\infty\) and \(\Gamma_2\) can be used as indicators of recovery properties of a matrix \(B\), otherwise one can ask the question: *Maybe \(\Gamma_\infty(B)\) and \(\Gamma_2(B)\) are always numbers slightly greater than 1."

**Lemma 3.3.1.** For any natural numbers \(m\) and \(n\) with \(0 < m < n\) there exists \(B \in \mathbb{R}^{(n-m) \times n}\) such that both \(\Gamma_\infty(B)\) and \(\Gamma_2(B)\) have order \(\sqrt{\frac{m}{1 + \log(n/m)}}\).

**Proof.** It follows from Garnaev-Gluskin Theorem (2.5.1) that there exists \(B \in \mathbb{R}^{(n-m) \times n}\) such that
\[
\Gamma_\infty(B) = \min_{y \in \mathbb{R}^{n-m}} \frac{\|B^T y\|_1}{\|B^T y\|_\infty} \geq \Gamma_2(B) = \min_{y \in \mathbb{R}^{n-m}} \frac{\|B^T y\|_1}{\|B^T y\|_2} \geq C_2 \sqrt{\frac{m}{1 + \log \frac{n}{m}}} \tag{3.19}
\]
for some constant \(C_2\) independent of \(m\) and \(n\). \(\square\)

To estimate an upper bound for \(\Gamma_\infty\) we need to prove two lemmas. Let \(V\) be a \(p\)-dimensional subspace of \(\mathbb{R}^n\) and \(B^T \in \mathbb{R}^{n \times p}\) an orthonormal basis for \(V\). Let \(b_i, 1 \leq i \leq n\) be the \(i\)-th column of \(B\).

**Lemma 3.3.2.** The following holds:
\[
\max_{v \in V} \frac{\|v\|_\infty}{\|v\|_2} = \max_{y \in \mathbb{R}^p} \frac{\|B^T y\|_\infty}{\|B^T y\|_2} = \max_{1 \leq i \leq n} \|b_i\|_2. \tag{3.20}
\]
Proof. We can write $b_i^T y = \|b_i\|_2 y_2 \cos \alpha_i$. Then,

$$
\max_{y \in \mathbb{R}^p} \frac{\|B^T y\|_\infty}{\|B^T y\|_2} = \max_{y \in \mathbb{R}^p} \frac{\|B^T y\|_\infty}{\|y\|_2} = \max_{y \in \mathbb{R}^p} \frac{\max_i \{\|b_i\|_2 y_2 \cos \alpha_i\}}{\|y\|_2} = \max_i \{\max \{\max b_i \|_2 \cos \alpha_i\}\} = \max_i \|b_i\|_2
$$

(3.21)

□

Lemma 3.3.3. The following inequality holds for any $p$-dimensional subspace of $\mathbb{R}^n$:

$$
\min_{v \in V} \frac{\|v\|_2}{\|v\|_\infty} \leq \sqrt{\frac{n}{p}}.
$$

(3.22)

In particular, for $p \geq \frac{n}{2}$, the inequality

$$
\min_{v \in V} \frac{\|v\|_2}{\|v\|_\infty} \leq \sqrt{2}
$$

(3.23)

holds for all $p$-dimensional subspaces of $\mathbb{R}^n$.

Proof.

$$
\min_{v \in B^T y} \frac{\|v\|_2}{\|v\|_\infty} = \left( \max_{v \in B^T y} \frac{\|v\|_\infty}{\|v\|_2} \right)^{-1} \leq \sqrt{\frac{n}{p}},
$$

(3.24)

because

$$
\max_{v \in B^T y} \frac{\|v\|_\infty}{\|v\|_2} \geq \sqrt{\frac{p}{n}}.
$$

(3.25)

To prove the above inequality let us assume that $\max_{v \in B^T y} \frac{\|v\|_\infty}{\|v\|_2} < \sqrt{p/n}$, which means that $\max_i \|b_i\|_2 < \sqrt{p/n}$ by Lemma 3.3.2. It follows that $\|b_i\|_2 < \sqrt{p/n}$ for every $1 \leq i \leq n$. Since matrix $B^T$ is orthonormal it follows that $p = \sum_{i=1}^{n} \|b_i\|_2^2 < (p/n)n = p$, which leads to a contradiction. □

It is necessary to mention that the inequality (3.22) is tight. For $n$ equal to a power of 2 there exists a matrix for which the equality holds. For example, see the recursive matrix defined by Feuer and Nemirovski in [20].

Lemma 3.3.4 (Upper Bound for $\Gamma_\infty$). The following holds for any $B \in \mathbb{R}^{p \times n}$:

$$
\Gamma_\infty(B) = \min_{y \in \mathbb{R}^p} \frac{\|B^T y\|_1}{\|B^T y\|_\infty} \leq 1 + \sqrt{\frac{n-1}{p}} \sqrt{n-p}.
$$

(3.26)

In particular, for $p \geq \frac{n}{2}$ the following inequality holds:

$$
\Gamma_\infty(B) \leq 1 + \sqrt{n}.
$$

(3.27)
Proof. Let \( V = \text{range}(B^T) \). It follows from Lemma 3.3.3 that there exists \( v^* \in V \) such that \( \|v^*\|_2 \leq \sqrt{n/p} \). Without loss of generality we can assume that \( \|v^*\|_\infty = v_1^* = 1 \). We can write \( v^* = (1; z^*) \), where \( z^* \) is a vector of dimension \( n - 1 \).

\[
\|v^*\|_2^2 = 1 + \|z^*\|_2^2 \leq \frac{n}{p} \quad (3.28)
\]

\[
\|z^*\|_2^2 \leq \frac{n}{p} - 1 = \frac{n - p}{p} \quad (3.29)
\]

\[
\|z^*\|_2 \leq \sqrt{\frac{n - p}{p}}. \quad (3.30)
\]

Therefore,

\[
\|z^*\|_1 \leq \sqrt{n - 1}\|z^*\|_2 = \sqrt{n - 1}\sqrt{\frac{n - p}{p}}, \quad (3.31)
\]

and so

\[
\Gamma_\infty(B) \leq \frac{\|v^*\|_1}{\|v^*\|_\infty} \leq 1 + \frac{\sqrt{n - 1}\sqrt{n - p}}{1} = 1 + \sqrt{\frac{n - 1}{p}}\sqrt{n - p}, \quad (3.32)
\]

which completes the proof. \( \square \)

Lemma 3.3.5 (Upper Bound for \( \Gamma_2 \)). The following holds for any \( B \in \mathbb{R}^{p \times n} \):

\[
\Gamma_2(B) = \min_{y \in \mathbb{R}^p} \frac{\|B^T y\|_1}{\|B^T y\|_2} \leq \sqrt{n - p + 1}. \quad (3.33)
\]

In particular, for \( p \geq \frac{n}{2} \) the following inequality holds:

\[
\Gamma_2(B) \leq 1 + \sqrt{\frac{n}{2}} + 1. \quad (3.34)
\]

Proof. There exists a vector \( y^* \) such that \( B^T y^* \) has \( p - 1 \) zero entries. Therefore,

\[
\Gamma_2(B) = \min_{y} \frac{\|B^T y\|_1}{\|B^T y\|_2} \leq \frac{\|B^T y^*\|_1}{\|B^T y^*\|_2} \leq \sqrt{n - p + 1}. \quad (3.35)
\]

\( \square \)

For compression the number of measurements \( m = n - p \) does not exceed \( \frac{n}{2} \), or equivalently \( p \geq \frac{n}{2} \). It follows from Lemmas 3.3.4 and 3.3.5 that if \( p \geq \frac{n}{2} \), then

\[
\Gamma_\infty(B) \leq 1 + \sqrt{n} \quad (3.36)
\]

\[
\Gamma_2(B) \leq 1 + \sqrt{\frac{n}{2}} + 1, \quad (3.37)
\]
which means that if \( p \geq \frac{3}{2} \) both \( \Gamma_\infty(B) \) and \( \Gamma_2(B) \) always have upper bounds of the same order.

Recall that sufficient conditions for recovery are \( k < \frac{1}{2} \Gamma_\infty(B) \) or \( k < \frac{1}{4} (\Gamma_2(B))^2 \), that is, the sufficient condition associated with \( \Gamma_2 \) is stronger than the sufficient condition associated with \( \Gamma_\infty \). This is the reason why Lemma 3.1.3 is called \textit{Weak Sufficient Condition for Recovery}.

### 3.4 Comparison of \( \Gamma_\infty \) and Coherence

In this section we will show that the estimated sparsity \( k \) for guaranteed recovery computed using \( \Gamma_\infty \) is always greater or equal to the estimated sparsity \( k \) computed using coherence. We will also show that in some cases both methods give an estimate of the same order.

Let \( A \in \mathbb{R}^{m \times n} \) be a dictionary and \( m < n \). Let \( B \in \mathbb{R}^{p \times n} \), such that \( p + m = n \) and \( AB^T = 0 \). Let \( M(A) \) be the coherence of the dictionary \( A \) as defined in Subsection 2.3.1. Recall that the columns of \( A \) are unit vectors.

#### Lemma 3.4.1

Let \( k_1 \) and \( k_2 \) be the sparsities for guaranteed recovery estimated by \( \Gamma_\infty(B) \) and \( M(A) \) respectively. Then \( k_1 \geq k_2 \).

#### Proof

According to Theorem 2.3.3 and Lemma 3.1.3, it is enough to show that

\[
1 + \frac{1}{M(A)} \leq \Gamma_\infty(B). \tag{3.38}
\]

We will follow the proof of Theorem 1 by Gribnoval and Nielsen [23].

Let \( v \in \text{range}(B^T) \), then \( Av = 0 \), or, in vector form \( \sum_{i=1}^n v_i a_i = 0 \), where \( a_i \), \( 1 \leq i \leq n \) is the \( i\)-th column of \( A \). Then, \( v_1 a_1 = -\sum_{i=2}^n v_i a_i \). Taking the inner product of both sides with \( a_1 \), we get \( v_1 = -\sum_{i=2}^n v_i \langle a_i, a_1 \rangle \). It follows that

\[
|v_1| = \left| -\sum_{i=2}^n v_i \langle a_i, a_1 \rangle \right| \leq M(A) \sum_{i=2}^n |v_i| = M(A)(\|v\|_1 - |v_1|), \tag{3.39}
\]
or

$$|v_i|(1 + M(A)) \leq \|v\|_1 M(A).$$  \hfill (3.40)

The same way for $2 \leq i \leq n$, we get

$$|v_i|(1 + M(A)) \leq \|v\|_1 M(A).$$  \hfill (3.41)

Since this is true for every index $1 \leq i \leq n$, it follows that for every vector $v \in \text{range}(B^T)$ the following inequality holds:

$$1 + \frac{1}{M(A)} \leq \frac{\|v\|_1}{\|v\|_{\infty}}.$$  \hfill (3.42)

Now if we take minimum over all $v \in \text{range}(B^T)$ we will get:

$$1 + \frac{1}{M(B)} \leq \min_v \frac{\|v\|_1}{\|v\|_{\infty}} = \Gamma_{\infty}(B).$$  \hfill (3.43)

which completes the proof. \hfill \Box

Suppose that we have a good dictionary, that is, a dictionary with a coherence satisfying $M(A) \leq \frac{1}{\sqrt{m}}$. For such a dictionary $A$, using Lemma 3.3.4 we conclude that

$$\frac{1}{2}(1 + \sqrt{m}) \leq \frac{1}{2} \left(1 + \frac{1}{M(A)}\right) \leq \frac{1}{2} \Gamma_{\infty}(B) \leq \frac{1}{2} \left(1 + \sqrt{\frac{n-1}{p} \sqrt{m}}\right).$$  \hfill (3.44)

According to (3.44), for a good dictionary $A$, the $\Gamma_{\infty}$ approach is better than the coherence approach by no more than a factor of

$$\sqrt{\frac{n-1}{p}} = \sqrt{\frac{n-1}{n-m}} \leq \sqrt{\frac{1}{1 - \frac{m}{n}}},$$  \hfill (3.45)

where $m$ is the number of measurements.

For compression, the number of measurements $m = n - p$ does not exceed $\frac{n}{2}$ and usually $m$ is a small fraction of $n$. Note, that $\sqrt{\frac{1}{1 - \frac{m}{n}}}$ approaches 1 as $\frac{m}{n}$ goes to 0. This means that in practice, for a good dictionary and small ratios $\frac{m}{n}$, the $\Gamma_{\infty}$ approach is not better than the coherence approach. In the best case, when $\frac{m}{n} = \frac{1}{2}$, the $\Gamma_{\infty}$ approach is better than the coherence approach by no more than $\sqrt{2}$. 
In the best case when \( m = \frac{n}{2} \), the \( \Gamma_\infty \) approach (and therefore the coherence approach) can recover no more than

\[
\frac{1}{2} \sqrt{\frac{n-1}{p}} \sqrt{n-p} = \frac{1}{2} \sqrt{\frac{n-1}{n-m}} \sqrt{m} \leq \frac{1}{2} \sqrt{n} = \frac{\sqrt{2}}{2} \sqrt{\frac{n}{2}} \approx 0.707 \sqrt{\frac{n}{2}} \tag{3.46}
\]

nonzero entries.

We have to mention that computation of \( M(A) \) is straightforward, while computation of \( \Gamma_\infty(B) \) requires the solution of \( n \) linear programs of dimension \( n \).

For \( m \leq \frac{n}{2} \) both methods will estimate sparsity of an order not exceeding \( \sqrt{m} \), "which is of little practical use since we are interested in procedures that might recover a signal when a constant fraction of the output is unreliable", as was mentioned by Candès and Tao in [7].

The coherence approach to recovery depends on the matrix representation of a subspace, while the \( \Gamma_\infty \) approach is matrix invariant. We believe that \( \Gamma_\infty \) approach to recovery still may be useful for applications where the measurement matrix is fixed and guaranteed recovery is required.
Chapter 4

Γ₂ and Numerical Experiments

Below we prove a theorem which will allow us to develop a technique to estimate Γ₂. Although solving the Γ₂ optimization problem exactly is generally NP-hard according to Subsection 3.2.2, we will show in this Chapter that for random and semi-random matrices a random-sampling algorithm can provide us with approximate solutions that are empirically adequate in practice.

4.1 Necessary Conditions for Γ₂

Let \( B \in \mathbb{R}^{p \times n} \) be of full rank and \( p < n \). Without loss of generality suppose that the rows of \( B \) are orthonormal.

**Theorem 4.1.1** (Necessary conditions for Γ₂). Consider the optimization problem associated with computing Γ₂:

\[
\Gamma_2(B) = \min_{y \in \mathbb{R}^p} \gamma_2(B, y) = \min_{y \in \mathbb{R}^p} \frac{\|B^Ty\|_1}{\|B^Ty\|_2} = \min_{y \in \mathbb{R}^p} \frac{\|B^Ty\|_1}{\|y\|_2}. \tag{4.1}
\]

Let \( y^* \) be a minimizer and let \( u^* = B^Ty^* \). Then \( u^* \) has at least \( p - 1 \) zero entries.

**Proof.** Suppose that this is not true, that is, \( y^* \) is a minimizer but \( u^* = B^Ty^* \) has at most \( p - 2 \) zero entries. Let \( z \in \mathbb{R}^p \) with \( \|z\|_2 = 1 \) such that \( z \) is orthogonal to \( y^* \) and to all \( b_i \)'s corresponding to \( Z(u^*) \), where \( b_i, 1 \leq i \leq n \), is the \( i \)-th column of \( B \).
Since $0 \neq z \in \mathbb{R}^p$ can be orthogonal to at most $p - 1$ linearly independent vectors in $\mathbb{R}^p$, such a vector $z$ exists. Let $t$ be sufficiently small such that $|b_i^T y^*| > |tb_i^T z|$ for all $i \in S(u^*)$. Consider $\tilde{y} = y^* + tz$ and $\tilde{u} = B^T \tilde{y}$. Observe, that $S(\tilde{u}) = S(u^*)$.

Then,

$$
\gamma_2(B, \tilde{y}) = \frac{\|B^T \tilde{y}\|_1}{\|\tilde{y}\|_2} = \frac{\|B^T (y^* + tz)\|_1}{\|y^* + tz\|_2} = \frac{\sum_{i \in S(u^*)} |b_i^T (y^* + tz)|}{\|y^* + tz\|_2} \leq \frac{\sum_{i \in S(u^*)} |b_i^T y_i| + t \delta_i b_i^T z}{\|y^*\|_2} = \frac{\|B^T y^*\|_1}{\|y^*\|_2} + t \frac{\sum_{i \in S(u^*)} \delta_i b_i^T z}{\|y^*\|_2},
$$

(4.2)

where $\delta_i, i \in S(u^*)$ is either 1 or $-1$. We always can choose $t$ either positive or negative such that $t \sum_{i \in S(u^*)} \delta_i b_i^T z \leq 0$. Then,

$$
\gamma_2(B, \tilde{y}) < \gamma_2(B, y^*),
$$

(4.3)

which is a contradiction to the fact that $y^*$ is a minimizer of $\gamma_2(B, y)$.

\[\square\]

**Definition 4.1.2 (Necessary Set).** Let $B \in \mathbb{R}^{p \times n}$ be of full rank and $p < n$. We define the necessary set of $B$ to be

$$
NS(B) = \{y \in \mathbb{R}^p, \|y\|_2 = 1 : |Z(B^T y)| \geq p - 1\}.
$$

(4.4)

Suppose that any $p$ columns of $B$ are linearly independent. Then

$$
|NS(B)| = \binom{n}{p-1} = \frac{n!}{(p-1)!((n-p+1))!}.
$$

(4.5)

$|NS(B)|$ is exponential in $p$, which is not surprising, since we mentioned in Subsection 3.2.2 that the optimization problem associated with $\Gamma_2$ can not be solved (exactly) in polynomial time.

However, we can ask the question:

_Maybe for some classes of matrices we can find a good estimate for $\Gamma_2$ within a reasonable time?_
4.2 Preliminary Experiments

In this section we will study behavior of $\gamma_2(B, y)$ on the necessary set for iid Gaussian matrices.

4.2.1 Experiment with Small Gaussian Matrices

We start our experiment with small matrices for which sets of necessary points can be computed in a reasonable time.

![Distribution of $\gamma_2(n = 20, m = 10)$](image)

Figure 4.1: Distribution of the values of $\gamma_2(B, y)$ at necessary points of an iid Gaussian matrix $B$ of the size 10 by 20. The number of measurements $m = n - p = 10$. The total number of necessary points is 167,960.
Figure 4.2: Distribution of the values of $\gamma_2(B, y)$ at necessary points of an iid Gaussian matrix $B$ of the size 16 by 20. The number of measurements $m = n - p = 4$. The total number of necessary points is 15,504.

By looking on Figures 4.1 and 4.2 we can see that for the case $n = 20$ and $m = 10$ values of $\gamma_2(B, NS(B))$ are between 2.25 and 3.25 and for the case $n = 20$ and $m = 4$ values of $\gamma_2(B, NS(B))$ are between 1.35 and 2.25. While both examples are too small to make a meaningful conclusion, we see that in both examples values of $\gamma_2(B, y)$ are close for all necessary points.
4.2.2 Experiment with Gaussian Matrices

In this subsection we will show by computational experiments that for iid Gaussian matrices the sparsity $k$ for guaranteed recovery can be estimated within a reasonable time.

Let $A \in \mathbb{R}^{m \times n}$. In order to say how good the measurement matrix $A$ is, we have to estimate $\Gamma_2(B)$, where $B$ is a $p$ by $n$ matrix with $p = n - m$ and $AB^T = 0$.

Let $A \in \mathbb{R}^{m \times n}$ be an iid Gaussian matrix. We want to know how good Gaussian matrices of the size $m$-by-$n$ are. According to [7], the nullspace of $A$ is spanned by an iid Gaussian matrix. This means that instead of generating a Gaussian matrices of the size $m$-by-$n$ and computing their nullspaces, we can directly generate Gaussian matrices of size $p$-by-$n$ with $p = n - m$ and estimate $\Gamma_2$.

Our experiment is as follows. We generate a $p$-by-$n$ Gaussian matrix $B$ and compute $\gamma_2(B; y)$ at some reasonable number of points from the necessary set $NS(B)$.

As we can see in Figures 4.3, 4.4 and 4.5, values of $\gamma_2(B; y)$ are close for all sampled necessary points. Not only this, but if we keep generating Gaussian matrices of the same size, the distribution will remain unchanged.
Figure 4.3: Distribution of the values of $\gamma_2(B, y)$ at necessary points of an iid Gaussian matrix $B$ of the size 500 by 1000. The sample size is 500,000.
Figure 4.4: Distribution of the values of $\gamma_2(B, y)$ at necessary points of an iid Gaussian matrix $B$ of the size 750 by 1000. The sample size is 75,000.
Figure 4.5: Distribution of the values of $\gamma_2(B, y)$ at necessary points of an iid Gaussian matrix $B$ of the size 900 by 1000. The sample size is 90,000.
While we cannot compute $\Gamma_2$ exactly, we can estimate the sparsity $k$ for guaranteed recovery by computing values of $\gamma_2$ at some reasonable number of necessary points and estimate $\Gamma_2$ by a minimum of $\gamma_2$ on the sampled necessary set.

Results of some of our experiments are presented in Tables 4.1, 4.2 and 4.3.

<table>
<thead>
<tr>
<th></th>
<th>min</th>
<th>max</th>
<th>mean</th>
<th>variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16.8965</td>
<td>18.8591</td>
<td>17.8616</td>
<td>0.043230</td>
</tr>
<tr>
<td>2</td>
<td>16.8477</td>
<td>18.7961</td>
<td>17.8663</td>
<td>0.045426</td>
</tr>
<tr>
<td>3</td>
<td>16.8722</td>
<td>18.8203</td>
<td>17.8687</td>
<td>0.044872</td>
</tr>
<tr>
<td>4</td>
<td>16.6897</td>
<td>18.8218</td>
<td>17.8688</td>
<td>0.046795</td>
</tr>
<tr>
<td>5</td>
<td>16.8389</td>
<td>18.8242</td>
<td>17.8734</td>
<td>0.044806</td>
</tr>
<tr>
<td>6</td>
<td>16.8418</td>
<td>18.7480</td>
<td>17.8627</td>
<td>0.047383</td>
</tr>
<tr>
<td>7</td>
<td>16.6851</td>
<td>18.7248</td>
<td>17.8660</td>
<td>0.044615</td>
</tr>
<tr>
<td>8</td>
<td>16.8749</td>
<td>18.7608</td>
<td>17.8697</td>
<td>0.045249</td>
</tr>
<tr>
<td>9</td>
<td>16.8805</td>
<td>18.8472</td>
<td>17.8698</td>
<td>0.043457</td>
</tr>
<tr>
<td>10</td>
<td>16.8080</td>
<td>18.7765</td>
<td>17.8718</td>
<td>0.045808</td>
</tr>
</tbody>
</table>

Table 4.1: For $n = 1000$, $m = 500$ we generated 10 Gaussian matrices of the size 500 by 1000. For each of the matrices we computed $\gamma_2$ at 500,000 necessary points.

Estimated sparsities $k$ for guaranteed recovery for Gaussian matrices of different sizes are presented in Table 4.4. For every pair $(n, m)$ we generated 10 Gaussian matrices of the size $n$ by $n - m$. For each of 10 matrices we found the minimum $\gamma_2$ on a reasonable number of necessary points. Numbers $k$ in Table 4.4 are rounded down averages of the minima of $\frac{(\Gamma_2)^2}{4}$ based on 10 trials.
<table>
<thead>
<tr>
<th></th>
<th>min</th>
<th>max</th>
<th>mean</th>
<th>variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11.7648</td>
<td>13.4596</td>
<td>12.6468</td>
<td>0.049385</td>
</tr>
<tr>
<td>2</td>
<td>11.7381</td>
<td>13.4870</td>
<td>12.6526</td>
<td>0.039431</td>
</tr>
<tr>
<td>3</td>
<td>11.6469</td>
<td>13.4270</td>
<td>12.6627</td>
<td>0.041874</td>
</tr>
<tr>
<td>4</td>
<td>11.7682</td>
<td>13.4606</td>
<td>12.6604</td>
<td>0.045707</td>
</tr>
<tr>
<td>5</td>
<td>11.7230</td>
<td>13.4612</td>
<td>12.6483</td>
<td>0.042421</td>
</tr>
<tr>
<td>6</td>
<td>11.7517</td>
<td>13.5120</td>
<td>12.6552</td>
<td>0.049560</td>
</tr>
<tr>
<td>7</td>
<td>11.7161</td>
<td>13.5738</td>
<td>12.6661</td>
<td>0.044856</td>
</tr>
<tr>
<td>8</td>
<td>11.6701</td>
<td>13.4685</td>
<td>12.6286</td>
<td>0.046612</td>
</tr>
<tr>
<td>9</td>
<td>11.6062</td>
<td>13.4604</td>
<td>12.6789</td>
<td>0.044172</td>
</tr>
<tr>
<td>10</td>
<td>11.7283</td>
<td>13.5958</td>
<td>12.6533</td>
<td>0.047523</td>
</tr>
</tbody>
</table>

Table 4.2: For $n = 1000$, $m = 250$ we generated 10 Gaussian matrices of the size 750 by 1000. For each of the matrices we computed $\gamma_2$ at 75,000 necessary points.

<table>
<thead>
<tr>
<th></th>
<th>min</th>
<th>max</th>
<th>mean</th>
<th>variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.0720</td>
<td>8.7697</td>
<td>8.0556</td>
<td>0.043585</td>
</tr>
<tr>
<td>2</td>
<td>7.0442</td>
<td>8.8399</td>
<td>8.0550</td>
<td>0.044138</td>
</tr>
<tr>
<td>3</td>
<td>7.0367</td>
<td>8.8916</td>
<td>8.0451</td>
<td>0.039490</td>
</tr>
<tr>
<td>4</td>
<td>7.0113</td>
<td>8.8343</td>
<td>8.0364</td>
<td>0.043960</td>
</tr>
<tr>
<td>5</td>
<td>6.8897</td>
<td>8.8386</td>
<td>8.0453</td>
<td>0.045022</td>
</tr>
<tr>
<td>6</td>
<td>7.0534</td>
<td>8.8438</td>
<td>8.0117</td>
<td>0.049278</td>
</tr>
<tr>
<td>7</td>
<td>6.9924</td>
<td>8.7701</td>
<td>8.0300</td>
<td>0.042540</td>
</tr>
<tr>
<td>8</td>
<td>7.0968</td>
<td>8.7710</td>
<td>8.0199</td>
<td>0.043147</td>
</tr>
<tr>
<td>9</td>
<td>7.1756</td>
<td>8.9313</td>
<td>8.0451</td>
<td>0.039640</td>
</tr>
<tr>
<td>10</td>
<td>7.0211</td>
<td>8.8011</td>
<td>8.0339</td>
<td>0.047394</td>
</tr>
</tbody>
</table>

Table 4.3: For $n = 1000$, $m = 100$ we generated 10 Gaussian matrices of the size 900 by 1000. For each of the matrices we computed $\gamma_2$ at 90,000 necessary points.
<table>
<thead>
<tr>
<th>$n/m$ = 2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 50$</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>150</td>
<td>9</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>200</td>
<td>12</td>
<td>7</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>250</td>
<td>15</td>
<td>10</td>
<td>7</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>300</td>
<td>19</td>
<td>12</td>
<td>9</td>
<td>7</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>350</td>
<td>23</td>
<td>14</td>
<td>10</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>400</td>
<td>26</td>
<td>16</td>
<td>12</td>
<td>9</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>450</td>
<td>30</td>
<td>19</td>
<td>14</td>
<td>11</td>
<td>9</td>
<td>7</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>500</td>
<td>33</td>
<td>21</td>
<td>16</td>
<td>12</td>
<td>10</td>
<td>8</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>550</td>
<td>37</td>
<td>23</td>
<td>17</td>
<td>13</td>
<td>11</td>
<td>9</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>600</td>
<td>41</td>
<td>26</td>
<td>19</td>
<td>15</td>
<td>12</td>
<td>10</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>650</td>
<td>44</td>
<td>29</td>
<td>21</td>
<td>16</td>
<td>13</td>
<td>11</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>700</td>
<td>48</td>
<td>31</td>
<td>23</td>
<td>18</td>
<td>15</td>
<td>12</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>750</td>
<td>51</td>
<td>33</td>
<td>25</td>
<td>19</td>
<td>16</td>
<td>13</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>800</td>
<td>55</td>
<td>36</td>
<td>26</td>
<td>21</td>
<td>17</td>
<td>14</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>850</td>
<td>59</td>
<td>38</td>
<td>28</td>
<td>22</td>
<td>18</td>
<td>15</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>900</td>
<td>63</td>
<td>40</td>
<td>30</td>
<td>24</td>
<td>19</td>
<td>16</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>950</td>
<td>66</td>
<td>43</td>
<td>32</td>
<td>25</td>
<td>20</td>
<td>17</td>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>1000</td>
<td>70</td>
<td>45</td>
<td>34</td>
<td>26</td>
<td>21</td>
<td>18</td>
<td>16</td>
<td>13</td>
</tr>
</tbody>
</table>

Table 4.4: Estimated sparsity $k$ for guaranteed recovery for Gaussian matrices of different sizes.
Figure 4.6: Estimated recoverability for Gaussian matrices. This picture is based on the results presented in Table 4.4.

Table 4.4 and Figure 4.6 allow us approximately answer two important questions: 1) Given dimension $n$ and number of measurements $m$, what is the largest sparsity $k$ for which recovery is still possible? and 2) Given dimension $n$ and sparsity $k$, what is the smallest number of measurements $m$ which will guarantee recovery?
4.2.3 Comparison of $\Gamma_2$ and $\Gamma_\infty$

Let $B \in \mathbb{R}^{p \times n}$ with $p < n$. According to Lemma 3.3.4, the upper bound of $\Gamma_\infty(B)$ is given by

$$\Gamma_\infty(B) \leq 1 + \sqrt{\frac{m}{1 - m/n}},$$

where $m = n - p$.

As Figure 4.7 shows, recoverability for Gaussian matrices based on $\Gamma_2$ is larger (in order) than recoverability based on $\Gamma_\infty$. It is not a surprise, because according to Lemmas 3.1.2 and 3.1.3 recovery is guaranteed if $k < \frac{1}{4}(\Gamma_2(B))^2$ or $k < \frac{1}{2}\Gamma_\infty(B)$.

![Graph](image.png)

**Figure 4.7:** Recoverability for Gaussian matrices based on $\Gamma_2$ and $\Gamma_\infty$. Line for $\Gamma_2$-estimated sparsity (blue) was obtained from Figure 4.4. Upper bound for $\Gamma_\infty$-estimated sparsity (red) is given by inequality (4.6).
4.2.4 Experiments with Other Types of Random Matrices

We conducted the same experiment described in Subsection 4.2.2 for other types of random matrices, namely Bernoulli \(\{0, 1\}\) matrices and matrices which entries are \(\{-1, 0, 1\}\) with equal probability \(1/3\). We also experimented with Partial Discrete Fourier Transform (PDCT) matrices. To our surprise, the numerical results for all types of matrices appear to be identical.

Based on our experiments we found a practical technique to estimate recoverability of different types of random matrices.

4.3 Practical Technique to Estimate Recoverability

4.3.1 Some Preliminaries

Let \(A \in \mathbb{R}^{m \times n}\) and \(B \in \mathbb{R}^{p \times n}\) be of full rank with \(p + m = n\) and \(AB^T = 0\). For practical purposes we are interested in recovery of vectors in \(\mathbb{R}^n\), where sparsity \(k\) is a fraction of \(n\).

Most results in this area deal with some unknown or uncertain absolute constants. The Garnaev-Gluskin inequality [22],

\[
\gamma_2(B, y) = \frac{\|B^Ty\|_1}{\|B^Ty\|_2} \geq C\sqrt{\frac{m/n}{(1 + \log(n/m))n}},
\]

contains a single unknown absolute constant \(C\). It is known to hold with high probability for Gaussian matrices.

The Donoho-Tanner formula [16] for Gaussian matrices,

\[
k < \frac{1}{2(1 + o(1))} \frac{m/n}{\log(n/m)n},
\]

also has a single constant of the form \(\frac{1}{2(1 + o(1))}\), which contains a lesser degree of uncertainty, but still is not definitive.
The formula of Baraniuk, Davenport, DeVore and Wakin [1], which is valid for random matrices satisfying the concentration of measure inequality ([1], formula (4.3)), is

\[ k \leq C_1 \frac{m/n}{\log \left( \frac{n^2}{k} \right)} n, \]  

(4.9)

where \( C_1 \) is an unknown constant.

Rudelson and Vershynin's result [32] for partial-orthonormal matrices such as DCT matrices,

\[ m = O(k \log(n) \log^2(k) \log(k \log n)). \]  

(4.10)

We propose practical techniques to compute empirical but specific bounds for all those matrices.

4.3.2 Distribution of \( \gamma_2 \) on the Necessary Set

Figures 4.8 and 4.9 show that empirical distributions of \( \gamma_2 \) on the sampled necessary sets for iid Gaussian and PDCT matrices appear to be identical.

Based on these experiments we conjecture that \( \gamma_2(B, y) \) have an identical distribution on the necessary sets for any “reasonable” random matrices (iid Gaussian, Bernoulli, PDCT, etc).
Figure 4.8: Empirical distributions of $\gamma_2$ on their necessary sets, respectively, for two 500 by 1000 matrices, one Gaussian and another PDCT. The sample size is 500,000. The two distributions appear to be identical.
Figure 4.9: Empirical distributions of $\gamma_2$ on their necessary sets, respectively, for two 100 by 1000 matrices, one Gaussian and another PDCT. The sample size is 900,000. The two distributions appear to be identical.
4.3.3 The Necessary Set vs. The Whole Space

As Figure 4.10 shows, values of $\gamma_2$ on the necessary set are strongly concentrated around mean.

![Distributions of $\gamma_2$ (m = 100, n = 1000)](image)

Figure 4.10: Empirical distributions of $\gamma_2$ on the necessary set and the whole space, respectively, for a 100 by 1000 Gaussian matrix. The sample size is 900,000.

4.3.4 A Single Constant vs. A Family of Constants

As we mentioned in Subsection 4.3.1, all formulas in Compressive Sensing contain a single absolute constant, regardless of the $n/m$ ratio. For example, Garnaev-Gluskin's formula (2.22) together with Lemma 3.1.2 suggest that for Gaussian matrices recovery is guaranteed if the sparsity $k$ satisfies

$$k \leq \frac{1}{4} C_2 \frac{m/n}{(1 + \log(n/m))} n = \left( C \frac{m/n}{(1 + \log(n/m))} \right) n. \quad (4.11)$$
However, numerical experiments indicate, that for different $n/m$ ratios, constants are different (see Figure 4.11). We can, of course, choose the smallest estimated constant which will be valid for all ratios, but using different constants for different $n/m$ ratios allows us to estimate recoverability more precisely.

Figure 4.11: Computed $C$ constants for $n/m = 2, 3, \ldots, 10$ and for $n = 50, 100, \ldots, 1000$. This picture shows numerical evidence for the existence of a family of limits for different $n/m$ ratios. To obtain this picture we simply divided estimated values of $\Gamma^2_2$ for iid Gaussian matrices of different sizes by corresponding factors $m/(1 + \log(n/m))$.

We will treat the term $C \frac{m/n}{(1+\log(n/m))}$ as constant $S_{\frac{n}{m}}$ for the ratio $n/m$ and will say that recovery is guaranteed if $k$ is a small fraction of dimension $n$, that is $k \leq S_{\frac{n}{m}} n$.

Estimated values of constants $S_{\frac{n}{m}}$ are presented in Figure 4.12
Figure 4.12: Computed $\hat{S}_r$ constants for $r \equiv n/m = 2, 3, \cdots, 10$ and for $n = 50, 100, \cdots, 1000$. The "o" marks are for iid Gaussian matrices and "x" for PDCT matrices, while the line is the average of the two.
A “zoomed-in” figure is in Figure 4.13, which shows that as \( n \) increases, although the computed constants are still increasing at \( n = 4000 \), the slopes appears to be decreasing to zero. Hence in all likelihood there appears to exist a separate upper limit for each ratio. Most surprisingly, those limits appear to be the same regardless what type of random or semi-random matrices is involved.

![Graph showing constants for 3 types of matrices and 2 ratios](image)

Figure 4.13: Computed \( \tilde{S} \) constants for \( m/n = 1/10 \) (top group) and \( 8/100 \) (bottom group), \( n = 1000, 1200, \ldots, 4000 \), and for iid Gaussian, Bernoulli and PDCT matrices. Each constant is taken as the minimum of those for 10 different random matrices of the same class. Corresponding to a given \( n \) value, the sample size is \( 900n \) for each matrix; hence the total sample size for each \( n \) value is \( 9000n \) (36 millions for \( n = 4000 \)).
4.3.5 A Sampling Technique

Let matrix $A \in \mathbb{R}^{m \times n}$ have a partition

$$A \rightarrow [A_1 \ A_2]$$

where $A_2 \in \mathbb{R}^{m \times m}$ is nonsingular. With necessary permutation, we can write without loss of generality

$$A = [A_1 \ A_2]. \quad (4.12)$$

Then the matrix

$$B = [-I \ A_1^T A_2^{-T}] \in \mathbb{R}^{(n-m) \times n}$$

satisfies

$$AB^T = 0.$$  

Hence $B^T$ spans the null space of $A$. Moreover, all the $n-m$ columns of $B^T$ correspond to necessary points for minimizing $\|B^Ty\|_1/\|B^Ty\|_2$.

For each partition of $A$, this procedure takes $n - m$ samples of necessary points at a cost of computing $A_2^{-1}A_1$ where $A_2$ is $m$-by-$m$. Relatively speaking, for a fixed $n$ the smaller the $m$ value is, the faster this procedure is. We can repeat this calculation for different partitions as many times as is necessary and computationally feasible.

4.4 Estimating Recoverable Sparsity

4.4.1 Estimation Method

For $A \in \mathbb{R}^{m_0 \times n_0}$ let $B \in \mathbb{R}^{(n_0-m_0) \times n_0}$ be such that $AB^T = 0$ where $n_0$ and $m_0$ are fixed. We know that a $k$-sparse solution is recoverable if

$$k < \frac{(\gamma_2(B))^2}{4} := S(B) n_0$$

where

$$S(B) := \frac{(\gamma_2(B))^2}{4n_0}. \quad (4.13)$$
Our computational results suggest that recovery can be obtained if

\[ k \lesssim S(B) n \]  

(4.14)

for \( A\)-matrices of the same random class with the same ratio \( m/n = m_0/n_0 \) and \( n \geq n_0 \). In practically important cases where the \( (m/n)\)-ratio is small the bound appears to be quite tight when \( n_0 \) is sufficiently large (say, \( n_0 \geq 1000 \)) and \( n \) is not too much larger than \( n_0 \) (say, \( n \in [n_0, 2n_0] \)).

Since we are not able to compute \( S(B) \) exactly, we will sample necessary points corresponding to a set of \( A \) matrices, say \( \mathcal{A} \), of the same size in a given class (for example, the DCT class) to get an approximate value for \( S(B) \),

\[ \tilde{S}_{r_0} \approx \min\{S(B) : \ BA^T = 0, \ A \in \mathcal{A}\} \]  

(4.15)

where we use the subscript to emphasize the association of \( \tilde{S} \) and the ratio \( r_0 = n_0/m_0 \) (also recall that we can sample for \( S(B) \) directly from \( A \)).

Our computational results indicate that once we have obtained \( \tilde{S}_{r_0} \) by sampling matrices \( A \in \mathbb{R}^{m_0 \times n_0} \), then the \( k \)-sparse solution can be recovered as long as

\[ k \lesssim \tilde{S}_{r_0} n \]  

(4.16)

for \( A \in \mathbb{R}^{m \times n} \) with \( m/n = r_0 \) and \( n \geq n_0 \). We quickly add that when applying this technique, all matrices involved should be from the same random class.

4.4.2 Experiments with DCT Matrices

4.4.2.1 Basic Set-ups

For our computations we use a MATLAB code written by Y. Zhang. The code utilizes a fixed-point iteration algorithm developed by E. T. Hale, W. Yin and Y. Zhang [24] for solving the \( l_1\)-regularized minimization problem

\[ \min_{x \in \mathbb{R}^n} \{\|x\|_1 + \frac{\mu}{2}\|Ax - b\|_2^2\}. \]  

(4.17)
For a given sparsity $k$ we generate a random vector $r \in \mathbb{R}^k$ containing $\{-1, 1\}$ with equal probability $1/2$. To obtain a $k$-sparse vector $\tilde{x} \in \mathbb{R}^n$ we generate a random permutation $\Pi_n$ of the index set $\{1, 2, \ldots, n\}$ and place vector $r$ into entries of the zero vector of length $n$ indexed by the first $k$ entries of $\Pi_n$. We use only $-1$ and $1$ as non-zero entries of $\tilde{x}$ in order to avoid numerical ambiguity and numerical difficulty. For a given matrix $A$ we set $b = A\tilde{x}$ and solve (4.17) to obtain a sparse solution $x^*$.

We consider recovery successful if the index set corresponding to the $2k$ largest, in absolute value, components of the computed solution $x^*$ contains the index set corresponding to the $k$ nonzeros of the exact solution $\tilde{x}$. This is done to avoid numerical errors of the iterative solver we used to compute $x^*$ and in practice it is enough to know the zero set of $\tilde{x}$ in order to compute it.

The computed break point for a given matrix is the largest $k$ at which recovery was 100% successful in 100 trials for 100 random $k$-sparse exact solutions.

### 4.4.2.2 A Simple Experiment

Our first experiment was designed to show the usefulness of our estimation technique. We consider partial DCT (PDCT) matrices of size 100 by 1000. We sampled 1000 different PDCT matrices with randomly chosen rows, each with 1000 different partitions. For such random PDCT matrices, we obtained

$$\hat{S}_{10} = 0.01115.$$  

Using such 100 by 1000 PDCT matrices for recovery, since $1000\hat{S}_{10} \approx 11$ we expect that recovery would fail for $k$ around 11. We ran our recovery code 100 times on $k = 12, 11$ and 10. Indeed, the code failed before the number of trials reached 100 for the first two $k$ values, while not failing for $k = 10$.

Next we set $A$ to the PDCT matrix of size 1000 with the first 100 rows of the DCT matrix. This $A$ is not a random matrix. Indeed, our sampling yielded

$$\hat{S} = 0.0044.$$
Hence, using such a non-random PDCT matrix recovery could fail even for $k = 1$. By running our recovery code with this PDCT matrix $A$ and $b = Ax$ where $x$ has a single nonzero at a random position, we indeed encountered failures from time to time.

This simple experiment indicates that the quantity $\tilde{S}_r$ is a useful indicator to predict the success or failure of recovery using random or semi-random matrices.

### 4.4.2.3 Recoverability of PDCT Matrices

The Donoho-Tanner formula (4.8) is established for iid Gaussian matrices. Does this formula also more or less predict the behavior partial-DCT (PDCT) matrices? Here we demonstrate that this is not the case. On the other hand, we show empirically that our estimation technique appears to provide a more accurate tool for predicting recoverability of PDCT matrices.

In Figure 4.14, we plot the line $k = \tilde{S}_2 n$, corresponding to $n/m = 2$, for $n = 1000$ to 2000 with an increment 100, and two lines correspond to the Donoho-Tanner formula for $o(1) = 1$ and 0, respectively. The value $\tilde{S}_2 = .007$ was obtained by sampling $500 \times 1000$ random PDCT matrices. We also plot the simulated break points in small circles. We do the same in Figure 4.15 for $n/m = 10$.

The break points were simulated as follows. For each $n$, we first generated an $m$ by $n$ random PDCT matrix $A$. For $k$ varying in a pre-determined range (which required some initial experimenting), we generate a series of random $k$-sparse vectors $x \in \mathbb{R}^n$ and form $b = Ax$. In each trial, We used our solver to try to recover $x$ from $A$ and $b$. If recovery failed before 100 trials, we decrease $k$ by 1 and try again, until we stopped at a $k$ value for which recovery was successful after 100 trials. Then we treat such a $k$ value as a simulated break point for the corresponding dimension $n$.

From Figure 4.14, we see that the Donoho-Tanner formula does not correctly predict the break points for random PDCT matrices, perhaps as should be expected. However, our empirical formula correctly predicted recoverability for $k < 0.07n$, even
Figure 4.14: Predicted vs. Computed Recoverability for PDCT matrices with ratio \( n/m = 2 \). The solid line is \( k = \tilde{S}_2 n \), and the other two lines correspond to Donoho-Tanner formula for \( o(1) = 1 \) and 0, respectively. The small circles represent simulated break points.

though it is a bit too conservative in this case (\( n/m = 2 \)). On the other hand, for the case of \( n/m = 10 \) as depicted in Figure 4.15, our empirical formula \( k < 0.0011n \) appears to be extremely tight in predicting the recoverability behavior. The line for our formula could perhaps be lowered a bit by a more extensive sampling (that would likely produce a smaller \( \tilde{S}_{10} \) value).

In Figure 4.15, the Donoho-Tanner line for \( o(1) = 1 \) also agrees well with computed break points for the ratio \( n/m = 10 \). However, it does not appear that asymptotically as \( o(1) \to 0 \) the prediction should still be good, at least this is not so in this range of \( n \) values.
Figure 4.15: Predicted vs. Computed Recoverability for PDCT matrices with ratio $n/m = 10$. The solid line is $k = \tilde{S}_{10} n$, and the other two lines correspond to Donoho-Tanner formula for $o(1) = 1$ and 0, respectively. The small circles represent simulated break points.

### 4.4.3 Experiments with Other Types of Random Matrices

For $n = 4000$ and $n/m = 10$ predicted recoverable sparsity $k = 0.0137n$ (see Figure 4.13). How good this predicted recoverability is? We simulated break points for the range of $n$ between 4000 and 5000 for iid Gaussian matrices, Bernoulli $\{-1,1\}$ matrices, Partial Discrete Cosine Transform (PDCT) matrices and Partial Inverse Discrete Cosine Transform (PIDCT) matrices.

As we can see on Figure 4.16, predicted recoverable sparsity coincides with simulated break points. Moreover, subplots for all types of matrices look identical. This experiment demonstrates that our approach can be used for prediction the success or failure of recovery for random or semi-random matrices.
Donoho-Tanner formula (4.8) contains an uncertain term $o(1)$. Our predicted recoverable sparsity and break points lines lie between lines for Donoho-Tanner formula with $o(1) = 0$ and $o(1) = 1$. It appears that for the range of $n$ between 4000 and 5000 and $n/m = 10$, the $o(1)$ term in Donoho-Tanner formula is a number between 0 and 1.

Figure 4.16: Predicted vs. Computed Recoverability for different types of matrices with ratio $n/m = 10$. Predicted recoverability $k = 0.0137n$ was obtained for $n = 4000$ from Figure 4.13. The solid line is $k = \hat{S}_{10} n$, and the other two lines correspond to Donoho-Tanner formula for $o(1) = 1$ and 0, respectively. The small circles represent simulated break points.
4.5 Why Is Our Estimate Tighter for Smaller $m/n$?

Let $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{(n-m) \times n}$ such that both are of full row rank and $AB^T = 0$. We consider $\min_x \{ \|x\|_1 : Ax = b\}$.

A sufficient condition for guaranteed recovery is

$$2\sqrt{k} < \min_{|S|=k} \min_y \frac{\|B^Ty\|_1}{\|B^S_y\|_2}.$$  \hspace{1cm} (4.18)

A weaker condition is

$$2\sqrt{k} < \min_y \frac{\|B^Ty\|_1}{\|B^T_y\|_2} \left( \min_{|S|=k} \min_y \frac{\|B^Ty\|_2}{\|B^S_y\|_2} \right).$$  \hspace{1cm} (4.19)

Equivalently,

$$2\sqrt{k} < \Gamma_2(B) \omega(B, k)$$  \hspace{1cm} (4.20)

where

$$\omega(B, k) := \min_{|S|=k} \min_y \frac{\|B^Ty\|_2}{\|B^S_y\|_2} \geq 1.$$  \hspace{1cm} (4.21)

If we replace the term $\omega(B, k)$ by 1, we arrive at the weakest sufficient condition involving $\Gamma_2$, namely,

$$k < \frac{(\Gamma_2(B))^2}{4}.$$  \hspace{1cm} (4.22)

If we use an orthonormal basis $B^T$, i.e., $BB^T = I$, for the null space of $A$, then

$$\frac{1}{\omega(B, k)} = \max_{|S|=k} \max_y \frac{\|B^S_y\|_2}{\|x\|_2} = \max_{|S|=k} \|B_S\|_2.$$  \hspace{1cm} (4.23)

From well-known results of concentration of measure, one can prove [30] that for an orthonormal matrix $B \in \mathbb{R}^{(n-m) \times n}$ whose rows span a Gaussian random subspace of $\mathbb{R}^n$, there holds

$$\max_{|S|=k} \|B_S\|_2 \leq \min \left( 1, \sqrt{\frac{n-m}{cn}} \right), \forall k \leq n-m,$$  \hspace{1cm} (4.24)

with prevailing probability as $n-m$ grows, where $c \in (0, 1)$ is some absolute constant. Therefore, with prevailing probability

$$\omega(B, k) \geq \max \left( 1, \sqrt{\frac{c}{1-m/n}} \right), \forall k \leq n-m.$$  \hspace{1cm} (4.25)
Substituting (4.25) into (4.20) and squaring, we obtain a condition that ensures recovery will succeed, with prevailing probability, if
\[ k < \frac{(\Gamma_2(B))^2}{4} \max \left( 1, \frac{c}{1 - m/n} \right). \] (4.26)

Observe that
\[ \max \left( 1, \frac{c}{1 - m/n} \right) = \begin{cases} 
1, & m/n \leq 1 - c, \\
\max(1, 2c), & m/n = 1/2.
\end{cases} \] (4.27)

The above formula implies that when the ratio $m/n$ is sufficiently small, condition (4.22) becomes tight in comparison to (4.20). On the other hand, when $m/n = 1/2$ and $c > 1/2$, the condition is not tight. In this case condition (4.22) under-estimates recoverable sparsity by at least a factor of $2c$ (and potentially at least a factor of 2 in the worst case), considering additional relaxations done in deriving (4.22), beside dropping the $\omega(B, k)$ term.

The above derivation is for iid Gaussian matrices, but empirically the result seems applicable to other random matrices as well.

Finally, we should mention that whenever the estimate $k < \frac{(\Gamma_2)^2}{4}$ is tight, then we should have $k < \frac{m+1}{4}$ based on the upper bound for $\Gamma_2$ in Lemma 3.3.5, that is the number of measurements needs to be at least 4 times of sparsity.
Chapter 5

Compressive Sensing and Missing Data Recovery

5.1 Is Missing Data Recoverable?

Let a data vector be such that a portion of data is available and the rest is missing. Can we recover the missing data from the available one? The question may sound at first like a mission impossible. However, it was shown by Zhang in [41] that under certain favorable conditions recovery of missing data is indeed achievable.

Let \( \hat{v} \in \mathbb{R}^d \) be a vector of data. Suppose, that there exists an \( d \) by \( n \) matrix \( F \), such that \( \hat{v} \) is \textit{sparsely represented} by the matrix \( F \), that there exists a vector \( \hat{x} \in \mathbb{R}^n \) that satisfies the linear system

\[
F\hat{x} = \hat{v}
\]  

(5.1)

and such that \( k = S(\hat{x}) \ll d \). Suppose that \( \hat{x} \) has the fewest nonzero components possible out of all solutions to (5.1). We will call such a sparsest solution \( \hat{x} \) the representation of \( \hat{v} \) under \( F \).
Let
\[ \tilde{v} = \begin{pmatrix} \tilde{v}_I \\ \tilde{v}_J \end{pmatrix}, \]  
(5.2)

where \( \tilde{v}_I \) and \( \tilde{v}_J \) are subvectors of \( \tilde{v} \) corresponding to available and missing entries respectively and let
\[ F = \begin{pmatrix} F_I \\ F_J \end{pmatrix}, \]  
(5.3)

where \( F_I \in \mathbb{R}^{m \times n} \) and \( F_J \in \mathbb{R}^{d-m \times n} \), such that \( F_I \) consists of the rows of \( F \) corresponding to the indices of available data and \( F_J \) consists of the rows of \( F \) corresponding to the indices of missing data.

With such partitioning equation (5.1) can be rewritten as
\[ F \tilde{x} = \tilde{v} \Rightarrow F_I \tilde{x} = \tilde{v}_I, \quad F_J \tilde{x} = \tilde{v}_J. \]  
(5.4)

Since \( \tilde{v}_J \) is missing (or unreliable) we have no use for the equation \( F_J \tilde{x} = \tilde{v}_J \). Therefore, we hope to be able to find the solution \( \tilde{x} \) from the equation
\[ F_I \tilde{x} = \tilde{v}_I. \]  
(5.5)

Since the linear system (5.5) is under-determined, we will try to find the sparsest \( \tilde{x} \), satisfying (5.5), namely
\[ \min_{F_I \tilde{x} = \tilde{v}_I} \| x \|_0. \]  
(5.6)
5.2 Examples

5.2.1 Recovery of a damaged image

We will use the Discrete Cosine Transform (DCT) matrix for recovery purposes. As we can see from Figure 5.1, the DCT matrix is a good choice for a sparse representation of the image. Even by setting 90% of the DCT coefficients of the smallest magnitudes to 0, we still have an image of a reasonable quality.

![Original image (left) and compressed image (right).](image)

Figure 5.1: Original image (left) and compressed image (right). The compressed image was obtained from the original image by setting 90% of the DCT coefficients of the smallest magnitudes to 0.
Figure 5.2: Original image (top left) and recovered image (top right). Damaged image was obtained from the original image by removing randomly 25% of the pixels. On the bottom images missing pixels were replaced by black pixels (left) and white pixels (right).
Figure 5.3: Original image (top left) and recovered image (top right). Damaged image was obtained from the original image by removing randomly 50% of the pixels. On the bottom images missing pixels were replaced by black pixels (left) and white pixels (right).
Figure 5.4: Original image (top left) and recovered image (top right). Damaged image was obtained from the original image by removing randomly 75% of the pixels. On the bottom images missing pixels were replaced by black pixels (left) and white pixels (right).
As we can see from Figure 5.2 if we remove 25% of the pixels randomly, we can recover the original image without considerable loss of quality. Even if we remove 50% of the pixels (Figure 5.3), recovered image is still of a reasonable quality.

If we remove 75% of the pixels randomly, the human eye can not recognize what is depicted on the damaged image. While the recovered image is not of the best quality, it is good enough to say that the image is a photo of a man in uniform (Figure 5.4).

5.2.2 Recovery of a text image

Optical character recognition (OCR), is translation of images of handwritten or typewritten text into machine-editable text.

OCR technology is widely used today. Archives worldwide are trying to transfer documents into electronic formats, which will allow quick searches through documents, the main purpose archives are intended for.

The OCR process usually consists of 3 main steps. First, an image of a document is obtained by scan. Second, an OCR (pattern recognition) software is used to convert an image of a text into electronic format. The third, and the most expensive and time consuming step in the OCR process, is text editing by human, because even the best OCR software still makes mistakes. Performance of an OCR software highly depends on the quality of the original paper document, mostly on the contrast between text and background.

Most old documents were printed on paper that contains acid. Such paper has tanned (yellowed) over the years and developed a tiny brown dots on its surface. Book collectors refer to brown dots on a paper as foxing (from ferrum oxid). Foxing appears randomly on a paper and affects both paper and ink.

Foxed documents are easily readable by humans. However, an OCR software makes numerous mistakes reading an image of such document. We propose a way to preprocess an image of a foxed document, such that an OCR software will make less errors during the recognition process.
In our experiment we used ABBYY FineReader OCR software, one of the best OCR softwares on the Market. In May 2006 ABBYY USA was awarded the Fujitsu Quarterly Innovative Leadership Award. PC World Magazine commented: 

*FineReader 8.0 Pro is the best OCR software we have seen.*

The original image of the text (Figure 5.5) is an 8-bit image, where each pixel has integer value between 0 (black) and 255 (white). The contrast of the image is good for FineReader. The software recognized 171 out of 174 words from this image.

The damaged image (Figure 5.6) was obtained from the original image by replacing numerical values of approximately 30% of the pixels. Damaged pixels are grey, that is they are neither black (text) nor white (background). In our experiment we treated grey pixels as *missing* data. To be precise, all pixel with numerical values between 50 and 205 were considered missing.

To make our experiment more appealing we let FineReader to read the damaged image (Figure 5.6), the damaged image with grey pixels replaced by black pixels (Figure 5.7) and the damaged image with grey pixels replaced by white pixels (Figure 5.8).

FineReader failed to recognize a single word from both the damaged image (Figure 5.6) and the damaged image with grey pixels replaced by black pixels (Figure 5.7). It is not a surprise, because both images are hardly readable by human eyes. On the other hand, FineReader recognized 80 words from the damaged image with grey pixels replaced by white pixels (Figure 5.8).

When we let FineReader to read the recovered image (Figure 5.10), the software recognized 145 words.
Researchers are still working out how best to do this. A study at four hospitals, published last year by the University of California, showed a remarkable rate of success in treating sudden cardiac arrest with an approach that involved, among other things, a "cardioplegic" blood infusion to keep the heart in a state of suspended animation. Patients were put on a heart-lung bypass machine to maintain circulation to the brain until the heart could be safely restarted. The study involved just 34 patients, but 80 percent of them were discharged from the hospital alive. In one study of traditional methods, the figure was about 15 percent.

Becker also endorses hypothermia—lowering body temperature from 37 to 33 degrees Celsius—which appears to slow the chemical reactions touched off by reperfusion. He has developed an injectable slurry of salt and ice to cool the blood quickly that he hopes to make part of the standard emergency-response kit. "In an emergency department, you work like mad for half an hour on someone whose

Figure 5.5: Original image. Words count is 174. FineReader recognized 171 words from this image.
Researchers are still working out how best to do this. A study at four hospitals, published last year by the University of California, showed a remarkable rate of success in treating sudden cardiac arrest with an approach that involved, among other things, a "cardioplegic" blood infusion to keep the heart in a state of suspended animation. Patients were put on a heart-lung bypass machine to maintain circulation to the brain until the heart could be safely restarted. The study involved just 34 patients, but 86 percent of them were discharged from the hospital alive. In one study of traditional methods, the figure was about 15 percent.

Becker also endorses hypothermia—lowering body temperature from 37 to 33 degrees Celsius—which appears to slow the chemical reactions touched off by reperfusion. He has developed an injectable slurry of ice and ice to cool the heart quickly that he hopes to make part of the standard emergency-response kit. "In an emergency department, you work like mad for an hour in some instances, but if you could cool the patient down from 37 to 33 with ice and bring it down 10 degrees in 20 minutes..."
Figure 5.7: Damaged text. Grey pixels replaced by black pixels. FineReader failed to recognize a single word from this image.
Researchers are still working out how best to do this. A study at four hospitals, published last year by the University of California, showed a remarkable rate of success in treating sudden cardiac arrest with an approach that involved, among other things, a "cardioplegic" blood infusion to keep the heart in a state of suspended animation. Patients were put on a heart-lung bypass machine to maintain circulation to the brain until the heart could be safely restarted. The study involved just 34 patients, but 80 percent of them were discharged from the hospital alive. In one study of traditional methods, the figure was about 15 percent.

Becker also endorses hypothermia—lowering body temperature from 37 to 33 degrees Celsius—which appears to slow the chemical reactions touched off by reperfusion. He has developed an injectable slurry of hail and ice to cool the blood quickly that he hopes to make part of the standard emergency-response kit. "In an emergency department, you work like mad for half an hour on someone whose..."

Figure 5.8: Damaged text. Grey pixels replaced by white pixels. FineReader recognized 80 out of 174 words from this image.
A study at four hospitals, published last year in the Journal of Cardiac Surgery, showed a remarkable rate of success in treating sudden cardiac arrest with an "electrical" shock that invovled an urgent other thing, a "cardioplegic" blood infusion to keep the heart in a state of suspended animation. Pacemaker-attached defibrillating devices were put into patients on heart-lung bypass machines to maintain circulation to the brain until the heart could be surely restarted. The study involved just 34 patients, but 95 percent of those were discharged from the hospital alive. In one study of the traditional method, the figure was about .15 percent.

Backer gave a lead as hypothermia—lowering of body temperature from 32 to 33 in the home. Celsius, which appears to slow the enzymes that pass through the body. A new injectable drug--that invovled an extreme load quickly that, he hopes to make part of the emergency response kit. In an emergency department, you work like mad for all hair or for scars whose

Figure 5.9: Screenshot of FineReader output of the damaged image with grey pixels replaced by white pixels (Figure 5.8). The software highlights characters which are recognized, but it is not sure of.
Researchers are still working out how best to do this. A study at four hospitals, published last year by the University of California, showed a remarkable rate of success in treating sudden cardiac arrest with an approach that involved, among other things, a "cardioplegic" blood infusion to keep the heart in a state of suspended animation. Patients were put on a heart-lung bypass machine to maintain circulation to the brain until the heart could be safely restarted. The study involved just 34 patients, but 80 percent of them were discharged from the hospital alive. In one study of traditional methods, the figure was about 15 percent.

Becker also endorses hypothermia—lowering body temperature from 37 to 33 degrees Celsius—which appears to slow the chemical reactions touched off by reperfusion. He has developed an injectable slurry of salt and ice to cool the blood quickly that he hopes to make part of the standard emergency-response kit. "In an emergency department, you work like mad for half an hour on someone whose...

Figure 5.10: Recovered text. FineReader recognized 145 out of 174 words from this image.
Researchers at the University of California showed a remarkable rate of success in treating sudden cardiac arrest. A study at four hospitals published last year by the University of California showed that involving, among other things, a "cardioplegic" blood infusion to keep the heart in a state of suspended animation, patients were put on an extracorporeal bypass machine to maintain circulation to the brain until they could be safely restarted. The study involved just 34 patients, but 30 percent of them were discharged from the hospital alive. In one study of traditional methods, the figure was about 15 percent.

Backer also endorses hypothermia—lowering body temperature from 37 to 33 degrees Celsius—which appears to slow the chemical reactions touched off by reperfusion. He has developed an injectable slurry to cool the blood quickly that he hopes to make part of the standard emergency response kit. "In an emergency department, you work like mad for half an hour on someone whose

Figure 5.11: Screenshot of FineReader output of the recovered image (Figure 5.10). The software highlights characters which are recognized, but it is not sure of.
Chapter 6

Conclusions

6.1 Summary

In this dissertation, we studied two new sufficient conditions for recovery of sparse solution to under-determined linear system of equations. The $\Gamma_\infty$ condition is polynomial-time computable and can always provide slightly stronger bounds than the existing coherence factor can. However, we find that it cannot deliver a higher-order bound than what the coherence factor provides, while being more expensive to compute (even though being polynomial-time computable). Hence, we conclude that in general the $\Gamma_\infty$ condition is not useful. Nevertheless, to the best of our knowledge so far it remains the strongest deterministic condition computable in polynomial-time. It may still find use in situations where one needs to determine the absolutely guaranteed recoverability of a given matrix.

On the other hand, the $\Gamma_2$ condition can provide bounds of the best order available. Unfortunately, we show that its computation is NP-hard. Is there any hope for using this condition? A key result we derived is a necessary condition for $\Gamma_2$ that allows us to examine the finite set of points that satisfy the necessary condition (necessary set). Even though the necessary set still contains an exponential number of points, our computations show that it exhibits a strong concentration of measure for random
and semi-random matrices. This phenomenon allows us to sample the set quickly and obtain good estimates for $\Gamma_2$. We showed numerically that this approach is useful in predicting recoverability of different types of random and semi-random matrices.

We proposed a practical approach to estimate recoverable sparsity for any given type of random or semi-random matrices. A novel feature of our approach is to estimate a family of constants rather than a conventional single constant. We presented empirical evidence of success of our approach in precise prediction of recoverable sparsity, especially for the cases when the $m/n$-ratio, the number of measurements over the length of the signal, is small. Using our approach we estimated recoverable sparsity for Partial Discrete Fourier Transform matrices, which are of great practical importance in Compressive Sensing due to fast matrix-vector multiplication. We showed empirical evidence that different kinds of random and semi-random matrices have similar recoverable properties when the $m/n$-ratio becomes small.

Finally, we demonstrated further evidence that Compressive Sensing technique can be successfully used to recover missing data.

6.2 A Prediction

In our abstract, we raise the following question. "Let $A$ be an $m \times n$ partial DCT matrix with $n = 10^7$ and $m = 5 \times 10^5$ ($n/m = 20$). Can we provide a reasonably good estimate on the maximum recoverable sparsity $k$?"

We have sampled the constants $S_{20}$ for Gaussian, Bernoulli and PDCT matrices at $n = 6000$. The values of these three sampled constants are close enough so that we may consider them random perturbations of each other. The mean of the three is about 0.0066329. Our computations lead us to believe that $k = 0.00663n$ is a good estimate of recoverability for Gaussian, Bernoulli and PDCT matrices with $n/m = 20$ and $n \geq 6000$. Such an estimate may become a bit conservative as $n$ goes far beyond 6000. We feel safe to predict that with very high probability the recoverable sparsity
is around $k = 66300$ for $n = 10^7$ and $m = 5 \times 10^5$.

Although as time-consuming as it may be, it is not impossible today to check, say by doing 100 trials on a powerful enough computer, whether or not the above prediction is on target or not.

## 6.3 Future Research

Our computations have raised a number of interesting theoretical questions that demand answers.

From our computational results, we conjectured that $\gamma_2(B, y)$ for Gaussian and PDCT matrices (most likely others as well) have an identical distribution on their necessary sets. A proof or disproof for this conjecture would be interesting and useful. It would also be interesting to derive the theoretical distribution of $\gamma_2(B, y)$ on the necessary sets.

Figure 4.11 gives numerical evidence for the existence of a family of limits for different $n/m$-ratios. It would be desirable to have a proof for this fact. Being able to theoretically obtain these limits would be doubly desirable.

Figures 4.12 and 4.13 suggest that the constants for different ratios approach their limits from below. We would like to have a proof of this fact, which should be very useful for establishing concrete lower bounds for recoverability of large matrices.

A key to understanding these and other open questions seems to be the concentration of measure phenomenon [27] on the necessary sets. We believe that a thorough study of this phenomenon will be an important step towards resolving these questions.
Bibliography


[38] Y. Zhang. Solution-recovery in $l_1$-norm for non-square linear systems: deterministic conditions and open questions. Technical report TR05-06, Department of Computational and Applied Mathematics, Rice University, Houston, TX, 2005.

