1 Introduction

Principal Component Analysis (PCA) is a central problem in statistics which finds a low-rank approximation to a matrix [7, 9, 10]. If each column of a matrix M represents a variable, and each entry within a column represents a measurement of the variable, then PCA finds a smaller set of variables that retain much of the information present in the original variables. PCA provides insight into data by identifying the coordinate directions along which the data vary the most. By approximating a dense matrix with a low-rank matrix, PCA also generates a compressed representation of data.

PCA is defined formally as the optimization problem of finding the best rank-r approximation to an $n \times n$ matrix. The approximation error is measured using the induced ℓ_2 norm. For an $n \times n$ matrix M, and a fixed rank r less than or equal to n, PCA can be expressed as the following optimization problem:

$$\begin{array}{ll} \text{minimize} & \|M - L\|_2 \\ \text{subject to} & \text{rank}(L) \leq r. \end{array}$$

PCA is sensitive to errors added to the matrix M. In particular, PCA is highly sensitive to sparse errors with large magnitudes [2]. Corrupting a single entry of the matrix can markedly change the solution to the PCA optimization problem. This phenomenon can magnify the effects of data corruption, which often occurs in practice through round-off error, malicious tampering, or other sources. Researchers have searched for methods to make PCA less sensitive to errors.

To make PCA less sensitive to errors, the PCA optimization problem can be replaced with a different optimization problem that separates out the errors. Robust PCA is the problem of performing low-rank approximation in a way that is insensitive to sparse errors with large magnitudes. More concretely, robust PCA is the problem of separating a matrix into the sum of a low-rank matrix and a sparse matrix of errors. Given a low-rank matrix L_0 and a sparse matrix of errors S_0, the goal of robust PCA is to recover L_0 and S_0 from their sum $M = L_0 + S_0$. Performing this separation prevents the errors from obscuring the low-rank component.

Like PCA, which has applications in multiple fields, robust PCA also has numerous applications. Applications of robust PCA are surveyed in [2]. Robust PCA can be used to analyze videos by separating the background of a video from moving objects. If each column of a matrix encodes a video frame, then a low-rank approximation represents the background and the sparse "errors" represent moving objects. Another example made famous recently by NetFlix is collaborative filtering, in which entry (i, j) of a matrix represents the preference of user i for category j. A low-rank approximation to the matrix can be used to predict users' preferences for categories that they have not explicitly ranked. Robust PCA can be used to separate out errors that arise in the matrix due to data tampering or people with atypical preferences.

To solve the robust PCA problem, Candès et al. introduce an optimization problem called Principal Component Pursuit (PCP) [2]. This paper investigates the conditions

under which PCP successfully recovers a low-rank matrix L_0 and a sparse matrix S_0 from the sum $M = L_0 + S_0$.

To understand robust PCA and its solution via PCP, it is helpful to first consider a related problem, sparse signal recovery (SSR). SSR is the problem of identifying the sparsest vector in the set of solutions to an underdetermined system of linear equations. SSR can be solved by replacing an intuitively clear but intractable optimization problem with a tractable convex optimization problem. The concepts used to solve SSR are analogous to the concepts used to solve robust PCA. We will discuss SSR before robust PCA to establish the background for robust PCA.

The paper is organized as follows. Section 2 introduces the sparse signal recovery problem. Section 3 discusses robust PCA by analogy with sparse signal recovery. Section 4 investigates the conditions under which Principal Component Pursuit solves the robust PCA problem and provides the results of numerical experiments that test whether the conditions presented in [2] are necessary.

2 Sparse Signal Recovery

2.1 Overview

The sparse signal recovery problem is the problem of identifying the sparsest solution (i.e. the one with fewest nonzero entries) to an underdetermined system of equations $Ax = y$, where x and y are vectors and A is a wide matrix. Because A has fewer rows than columns, the system of equations has multiple solutions. Recovering the sparsest solution is only possible if there is a unique sparsest vector in the solution set to $Ax = y$.

We can also understand the SSR problem by thinking of A as a linear transformation. In an SSR problem instance, the matrix A maps the vector x_0 to the vector y. We observe y, but not x_0. Because A has a nonzero null space, A maps multiple vectors to y. The SSR problem is to recover x_0 from among the set of vectors that A maps to y. If A maps multiple equally sparse vectors to y, then we have no way to distinguish x_0 from those other vectors, so it is impossible to recover x_0. Hence we need a condition on A to ensure that x_0 is the unique sparsest vector in the set of solutions to $Ax = y$.

We call a vector S-sparse if it has S or fewer nonzero entries:

$$\|x\|_0 = \#\{i : x_i \neq 0\} \leq S,$$

where the ℓ_0 quasi-norm $\|x\|_0$ denotes the number of nonzero entries of x. For convenience, we refer to this quasi-norm as a norm. The condition on A is that we require A to be one-to-one on S-sparse vectors. If A is one-to-one on S-sparse vectors, and x_0 is an S-sparse vector such that $Ax_0 = y$, then A does not map any other S-sparse vectors to y. Hence x_0 is the unique sparsest solution to $Ax = y$, so the SSR problem has a unique solution. Later we discuss a condition that can be imposed on A to ensure that A is one-to-one on S-sparse vectors.

2.2 Solving SSR by Formulating an Optimization Problem

SSR can be solved by formulating a tractable optimization problem. To explain the optimization problem, we first discuss a simpler optimization problem using the ℓ_0 norm. The optimization problem that directly corresponds to the SSR problem is

$$\min \|x\|_0 \quad \text{subject to } Ax = y.$$

This optimization problem has no efficient solution, and is NP-hard in general [11]. We adopt the strategy of replacing the ℓ_0 objective function with a related objective function that yields the same solution (under appropriate conditions), but that can be minimized efficiently.

Convex functions can be minimized efficiently using convex optimization software [1], so we replace the ℓ_0 norm with a convex function. The convex norm that we choose comes from the family of norms called ℓ_p norms, defined as follows:

$$\|x\|_p := \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}.$$

Examples of the ℓ_p norm are the ℓ_2 norm,

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2},$$

which is Euclidean length, and the $\ell_1$norm,

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|,$$

which is the sum of the absolute values of the components of x.

To choose an ℓ_p norm for the SSR objective function, we examine the unit balls of the ℓ_p norms. The unit ball for a norm is the set of points of norm 1. The unit ball for the ℓ_0 norm is the set of points along the coordinate axes. We are looking for a convex norm that approximates the ℓ_0 norm. The unit balls for the ℓ_p norms are shown in Figure 1, with p decreasing from the outermost ball to the innermost. As p decreases, the ℓ_p unit balls shrink down to hug the coordinate axes more closely, thus better approximating the ℓ_0 unit ball. As shown in Figure 1, for $p \geq 1$, the ℓ_p norm is convex, but for $p < 1$, the ℓ_p quasi-norm is not convex, so $p = 1$ is the smallest value of p for which the ℓ_p norm is convex.

Replacing the ℓ_0 norm with the ℓ_1 norm, the resulting optimization problem is

$$\min \|x\|_1 \quad \text{subject to } Ax = y.$$

An advantage of the ℓ_1 norm is that the minimization problem can be solved using linear programming [1]. Under appropriate conditions, the solution to this optimization problem

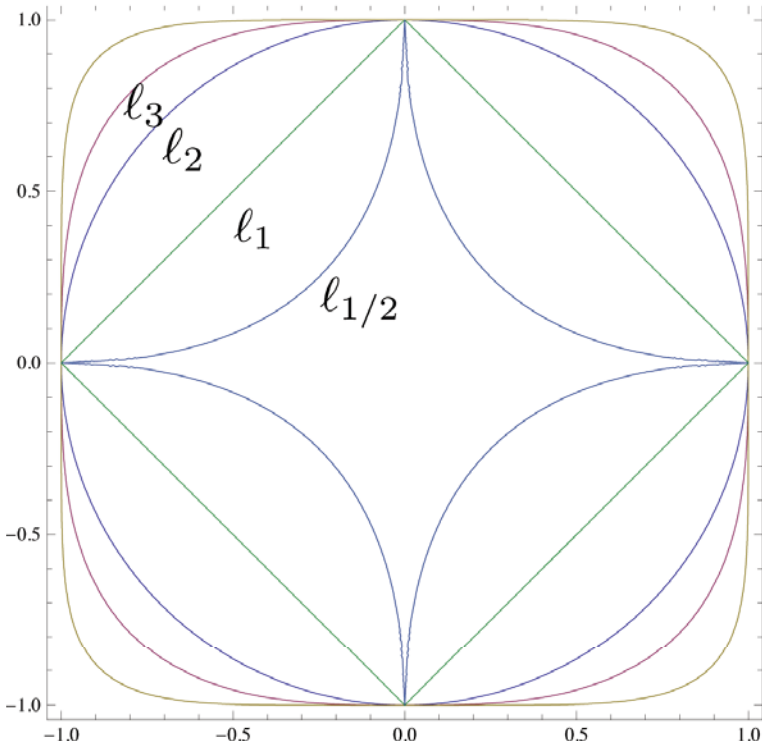

Figure 1: This figure shows ℓ_p balls for several values of p. For $p \geq 1$, the ℓ_p norm is convex, and for $0 \leq p < 1$, the ℓ_p quasi-norm is not convex. The ℓ_0 unit ball is the coordinate axes.

is the same as the solution to the ℓ_0 optimization problem [3, 4, 5, 6]. If A is one-to-one on S-sparse vectors, then the SSR problem is well-defined (that is, it has at most one solution). If A satisfies a stronger condition, called the restricted isometry property, then in addition to being well-defined, ℓ_1 minimization has the same solution as ℓ_0 minimization, so SSR can be solved efficiently [3, 5].

2.3 Conditions Under Which the SSR Problem Is Well-Defined

We have seen that if A is one-to-one on S-sparse vectors, then there is at most one S-sparse solution to $Ax = y$, so the SSR problem is well-defined. Although this condition clarifies when the SSR problem is well-defined, it is not intuitively obvious which matrices are one-to-one on S-sparse vectors. We now discuss a necessary and sufficient condition for a matrix A to be one-to-one on S-sparse vectors. The condition helps us to identify matrices that are one-to-one on S-sparse vectors and gives some geometric intuition regarding what it means for a matrix to be one-to-one on S-sparse vectors.

Proposition 2.1. *The matrix A is one-to-one on S-sparse vectors if and only if every subset of $2S$ columns of A is linearly independent.*

Proof. Assume that every subset of $2S$ columns of A is linearly independent. We show that A is one-to-one on S-sparse vectors. Let x_1 and x_2 be S-sparse vectors, and assume that $Ax_1 = Ax_2$. Then $A(x_1 - x_2) = 0$. Because x_1 and x_2 are S-sparse vectors, $x_1 - x_2$ is a $2S$-sparse vector, so $A(x_1 - x_2)$ is a linear combination of at most $2S$ columns of A. Because every subset of $2S$ columns of A is linearly independent, it must be that $x_1 - x_2 = 0$, so $x_1 = x_2$. Thus A is one-to-one on S-sparse vectors.

Now assume that A is one-to-one on S-sparse vectors. We show that every subset of $2S$ columns of A is linearly independent. Let B be a subset of $2S$ columns of A. Because A is one-to-one on S-sparse vectors, and B is a subset of the columns of A, B must be one-to-one on S-sparse vectors. Let x be a nonzero $2S$-dimensional vector. The vector x can be written as $x = x_1 - x_2$, where x_1 and x_2 are distinct S-sparse vectors. It follows that $Bx = Bx_1 - Bx_2$. If $Bx = 0$, then $Bx_1 = Bx_2$, which contradicts the assumption that B is one-to-one on S-sparse vectors. Therefore for any nonzero $2S$-dimensional vector x, Bx is nonzero, so the columns of B are linearly independent. □

We can also characterize matrices A that are one-to-one on S-sparse vectors in terms of their nullspaces.

Corollary 2.2. *The matrix A is one-to-one on S-sparse vectors if and only if the nullspace of A does not contain a nonzero $2S$-sparse vector.*

Proof. A nonzero $2S$-sparse vector x in the nullspace of A corresponds to a linearly dependent subset of at most $2S$ columns of A. This can be seen by interpreting the components of x as the coefficients of a linear combination of columns of A. By Proposition 2.1, it follows that A is one-to-one on S-sparse vectors if and only if there is no $2S$-sparse vector in the nullspace of A. □

2.4 Conditions Under Which SSR Can Be Solved Efficiently

The condition that A is one-to-one on S-sparse vectors guarantees that $Ax = y$ has at most one S-sparse solution but does not guarantee that an S-sparse solution can be found efficiently. If A satisfies a stronger condition, called the restricted isometry property (RIP), then the sparsest solution also achieves the minimum ℓ_1 norm, so SSR can be solved efficiently by ℓ_1 minimization. We present the restricted isometry property [5], which is defined in terms of the S-restricted isometry property and the isometry constant δ_S. An "isometry" is a mapping that preserves length, so the restricted isometry property is connected with orthonormal matrices, which preserve the length of vectors that they act on.

The S-restricted isometry property requires that all submatrices of A consisting of S columns approximately act as isometries. That is, all submatrices of S columns of A approximately preserve the length of vectors to which they are applied. The S-restricted isometry constant, δ_S, is a property of A that measures the extent to which submatrices of S columns of A deviate from being isometries. If the value of δ_S is sufficiently small, then subsets of S columns of A are approximately orthonormal, so A satisfies the S-restricted isometry property.

We now formally define the δ_S-restricted isometry constant. If T is a set of column indices, we define A_T to be the submatrix of A consisting of the columns specified by T. The δ_S-restricted isometry constant [5] is the smallest quantity that satisfies the inequality

$$(1 - \delta_S)\|c\|_2^2 \leq \|A_T c\|_2^2 \leq (1 + \delta_S)\|c\|_2^2,$$

for all submatrices A_T of A with S columns and all vectors c in the domain of A_T. Dividing through by $\|c\|^2$, we get the inequality

$$1 - \delta_S \leq \frac{\|A_T c\|_2^2}{\|c\|_2^2} \leq 1 + \delta_S.$$

This form of the inequality reveals that if the S-restricted isometry constant of A is δ_S, then each submatrix consisting of S columns of A rescales the magnitude of each vector in its domain by a factor of at least $\sqrt{1 - \delta_S}$ and at most $\sqrt{1 + \delta_S}$.

We now prove that the $2S$-restricted isometry property implies that A is one-to-one on S-sparse vectors.

Proposition 2.3. *If $\delta_{2S} < 1$, then A is one-to-one on S-sparse vectors.*

Proof. Let A_T be an arbitrary submatrix of A with $2S$ columns. If $\delta_{2S} < 1$, then $1 - \delta_{2S} > 0$, so

$$0 < 1 - \delta_{2S} \leq \frac{\|A_T c\|_2^2}{\|c\|_2^2}$$

for all nonzero vectors c in the domain of A_T. By the above inequality, it follows that the vector $A_T c$ is nonzero for every nonzero vector c in the domain of A_T. Thus the columns of A_T are linearly independent. Because A_T is an arbitrary submatrix of A with $2S$ columns, it follows that every subset of $2S$ columns of A is linearly independent. By Proposition 2.1, A is one-to-one on S-sparse vectors. □

Next we prove two propositions that provide geometric insight into RIP. In Proposition 2.4, we show that if δ_S is small, then the columns of A have norm approximately equal to one. In Proposition 2.6, we show that if δ_S is small, then pairs of distinct columns of A are approximately orthogonal.

Proposition 2.4. *Let A be a matrix with S-restricted isometry constant δ_S for some $S \geq 1$. Then the norm of the ith column of A satisfies*

$$\sqrt{1-\delta_S} \leq \|a_i\|_2 \leq \sqrt{1+\delta_S}.$$

Proof. The standard basis vector e_i is S-sparse, so by the S-restricted isometry property, $\|Ae_i\|_2 \leq \sqrt{1+\delta_S}\|e_i\|_2$. Therefore

$$\|a_i\|_2 = \|Ae_i\|_2 \leq \sqrt{1+\delta_S}\|e\|_2 = \sqrt{1+\delta_S}.$$

Likewise, $\|Ae_i\|_2 \geq \sqrt{1-\delta_S}\|e_i\|_2$, so

$$\|a_i\|_2 = \|Ae_i\|_2 \geq \sqrt{1-\delta_S}\|e\|_2 = \sqrt{1-\delta_S}.$$

□

The following lemma will help us prove Proposition 2.6.

Lemma 2.5. *The quantity $\|Ax\|_2^2$ can be expressed as*

$$\|Ax\|_2^2 = \sum_i x_i^2\|a_i\|_2^2 + \sum_{i\neq j} x_i x_j \langle a_i, a_j\rangle.$$

Proof. Using the relationship between norms and inner products, we can express $\|Ax\|_2^2$ as

$$\|Ax\|_2^2 = \langle Ax, Ax\rangle = \left\langle \sum_{i=1}^n x_i a_i, \sum_{i=1}^n x_i a_i \right\rangle = \sum_{i,j} x_i x_j \langle a_i, a_j\rangle.$$

The last step follows from the bilinearity of inner products. We group the summation into terms wtih $i = j$ and terms with $i \neq j$:

$$\sum_{i,j} x_i x_j \langle a_i, a_j\rangle = \sum_{i=1}^n x_i^2 \langle a_i, a_i\rangle + \sum_{i\neq j} x_i x_j \langle a_i, a_j\rangle.$$

Because $\langle a_i, a_i\rangle = \|a_i\|_2^2$, the right-hand side can be rewritten as

$$\|Ax\|_2^2 = \sum_{i=1}^n x_i^2\|a_i\|_2^2 + \sum_{i\neq j} x_i x_j \langle a_i, a_j\rangle.$$

□

We show that if δ_S is small, and A has real entries, then distinct columns of A are approximately orthogonal. Lemma 2.1 of [4] implies an analogous result for matrices with complex entries.

Proposition 2.6. *Let A be a matrix with real entries and S-restricted isometry constant δ_S. Then any two distinct columns of A satisfy*

$$|\langle a_i, a_j\rangle| \leq 2\delta_s.$$

Proof. Let A be a matrix with real entries, and let δ_S be the S-restricted isometry constant of A. Let x be a vector with $x_i = x_j = \frac{1}{\sqrt{2}}$ and all other components equal to zero. Note that $\|x\|_2 = 1$. By Lemma 2.5,

$$\begin{aligned}\|Ax\|_2^2 &= x_i^2\|a_i\|_2^2 + x_j^2\|a_j\|_2^2 + \sum_{k\neq l} x_k x_l \langle a_k, a_l\rangle \\ &= \left(\frac{1}{\sqrt{2}}\right)^2 \|a_i\|_2^2 + \left(\frac{1}{\sqrt{2}}\right)^2 \|a_j\|_2^2 + \left(\frac{1}{\sqrt{2}}\right)^2 \langle a_i, a_j\rangle + \left(\frac{1}{\sqrt{2}}\right)^2 \langle a_j, a_i\rangle \\ &= \frac{1}{2}\left(\|a_i\|_2^2 + \|a_j\|_2^2\right) + \langle a_i, a_j\rangle.\end{aligned}$$

By the S-restricted isometry property and the fact that $\|x\|_2^2 = 1$, it follows that, $1-\delta_S \leq \|Ax\|_2^2 \leq 1+\delta_S$, so

$$1 - \delta_S \leq \frac{1}{2}\left(\|a_i\|_2^2 + \|a_j\|_2^2\right) + \langle a_i, a_j\rangle \leq 1 + \delta_S.$$

First we prove an upper bound on $\langle a_i, a_j\rangle$. By the above inequality, it follows that

$$\langle a_i, a_j\rangle \leq 1 + \delta_S - \frac{1}{2}\left(\|a_i\|_2^2 + \|a_j\|_2^2\right).$$

By Proposition 2.4, $\|a_i\|_2^2 \geq 1-\delta_S$ and $\|a_j\|_2^2 \geq 1-\delta_S$, so

$$\begin{aligned}\langle a_i, a_j\rangle &\leq 1 + \delta_S - \frac{1}{2}(1-\delta_S + 1 - \delta_S) \\ &= 2\delta_S.\end{aligned}$$

The lower bound on $\langle a_i, a_j\rangle$ follows by an analogous argument:

$$\begin{aligned}\langle a_i, a_j\rangle &\geq 1 - \delta_S - \frac{1}{2}\left(\|a_i\|_2^2 + \|a_j\|_2^2\right) \\ &\geq 1 - \delta_S - \frac{1}{2}(1+\delta_S + 1 + \delta_S) \\ &= -2\delta_S.\end{aligned}$$

We have shown that $-2\delta_S \leq \langle a_i, a_j\rangle \leq 2\delta_S$, so $|\langle a_i, a_j\rangle| \leq 2\delta_S$. □

The key consequence of RIP is that if a matrix satisfies RIP, then ℓ_1 minimization recovers the sparsest vector in the solution set to $Ax = y$. The following theorem states this result.

Theorem 2.7 (from [3, 4, 5])**.** *If $\delta_{3S} + 3\delta_{4S} < 2$, then for any S-sparse vector x^* that satisfies $Ax^* = y$, x^* is the unique solution to*

$$\arg\min_x \|x\|_1 \quad \text{subject to } Ax = y.$$

This theorem provides the conditions under which the ℓ_1 minimization problem, which can be solved efficiently, has the same solution as the ℓ_0 minimization problem. If a matrix satisfies the conditions of Theorem 2.7, and there is an S-sparse solution to the system of equations $Ax = y$, then SSR can be solved efficiently by ℓ_1 minimization.

2.5 The Structure of SSR

Having looked at the details of SSR, we now summarize the structure of the SSR problem. SSR starts with an underdetermined system $Ax = y$. The true solution, x_0, satisfies an additional constraint: it is S-sparse. For a general underdetermined system, there may be multiple S-sparse vectors in the set of solutions, but we impose an additional condition on A, ensuring that there is at most one S-sparse solution: we require A to be one-to-one on S-sparse vectors. If x_0 is an S-sparse vector such that $Ax_0 = y$, and A is one-to-one on S-sparse vectors, then x_0 is the unique solution to the optimization problem

$$\min \|x\|_0 \quad \text{subject to } Ax = y.$$

In general, this optimization problem cannot be solved efficiently; it is NP hard [11]. If A satisfies RIP—a stronger condition than being one-to-one on S-sparse vectors—then the ℓ_1 minimum is the same as the ℓ_0 minimum, so SSR can be solved efficiently with the following optimization problem:

$$\min \|x\|_1 \quad \text{subject to } Ax = y.$$

The table in Figure 2 summarizes the structure of SSR. When we discuss robust PCA, we will see that it has a parallel structure.

3 Robust Principal Component Analysis

3.1 The Robust PCA Problem

The matrix M is formed as the sum of a low-rank matrix L_0 and a sparse matrix S_0:

$$M = L_0 + S_0.$$

We observe M, but we do not observe L_0 and S_0. The goal of robust PCA is to recover the matrices L_0 and S_0. We can view robust PCA as solving the underdetermined system

1.	Underdetermined system	$Ax = y$
2.	Extra constraint	x_0 sparse
3.	Conditions on problem instance to ensure problem is well-defined (that is, has at most one solution)	A is one-to-one on S-sparse vectors
4.	Actual optimization problem	$\min\|x\|_0$ subject to $Ax = y$
5.	A sufficient condition for an *efficient* solution	RIP
6.	Tractable optimization problem	$\min\|x\|_1$ subject to $Ax = y$

Figure 2: Structure of the sparse signal recovery problem.

of equations $M = L + S$, where M is a given matrix, and L and S are the unknowns. We must impose additional conditions to ensure that L_0 and S_0 are the unique solution. We solve the system of equations subject to the additional constraints that L_0 is low-rank and S_0 is sparse. As stated, the problem is not yet well defined, because there may be multiple ways to decompose M into the sum of a low-rank matrix and a sparse matrix. As we will see later, there are conditions on L_0 and S_0 such that for a range of values of the rank of L_0 and the sparsity of S_0, we can recover L_0 and S_0 using a tractable optimization problem.

We now present an example that demonstrates the need for more precise conditions in defining the robust PCA problem. The example illustrates a trade-off between rank and sparsity when decomposing M: by increasing the rank of the low-rank component, we can decrease the number of nonzero entries in the sparse component. Without additional conditions on the low-rank and sparse components, it is unclear which decomposition is the "correct" one. Consider the two matrices L_1 and S_1, in which L_1 is rank one and S_1 has four nonzero entries:

$$L_1 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad S_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}.$$

Let $M = L_1 + S_1$. If we subtract the 1 from row one, column one of S_1 and add it to the corresponding entry of L_1, we get the matrices L_2 and S_2:

$$L_2 = \begin{bmatrix} 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad S_2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}.$$

The matrix L_2 has rank two, and S_2 has three nonzero entries. Because $L_2 + S_2 = L_1 + S_1$, we

have found distinct decompositions of M that exhibit a trade-off between rank and sparsity.

Accordingly, we need conditions on the low-rank matrix L_0 and the sparse matrix S_0 that distinguish the decomposition $M = L_0 + S_0$ from other decompositions with different rank-sparsity trade-offs. We also need conditions to guarantee that there are no decompositions of M with equivalent rank and sparsity. Motivated by the need to ensure a unique solution, we will focus on the conditions under which a particular optimization problem, Principal Component Pursuit, recovers the matrices L_0 and S_0 from the sum $M = L_0 + S_0$. Under the conditions of Theorem 3.4, Principal Component Pursuit recovers L_0 and S_0 for a range of values of the rank of L_0 and the number of nonzero entries of S_0 [2].

3.2 Solving Robust PCA by Formulating an Optimization Problem

To solve the robust PCA problem, Candès et al. introduced an optimization problem called Principal Component Pursuit (PCP) [2]. Under appropriate conditions on L_0 and S_0, the PCP optimization problem recovers the original decomposition (L_0, S_0).

To motivate Principal Component Pursuit, we introduce a simpler optimization problem. An "ideal" optimization problem for separating a matrix into low-rank and sparse components is

$$\min \operatorname{rank}(L) + \lambda \|S\|_0 \quad \text{subject to } L + S = M,$$

because this optimization problem searches for a low-rank matrix and a sparse matrix that add up to M. The weighting parameter λ determines which term influences the objective function more, achieving a balance between minimizing the rank of L and maximizing the sparsity of S.

This optimization problem cannot in general be solved efficiently [11]. As with the sparse signal recovery problem, we adopt the strategy of replacing an objective function with a different objective function that is convex and can therefore be minimized efficiently. In the SSR problem, we replaced the ℓ_0 norm with the ℓ_1 norm. Likewise, for robust PCA we replace the ℓ_0 norm of S with the ℓ_1 norm. Note that we are not using the induced ℓ_1 norm; we are using the entrywise ℓ_1 norm, which is the sum of the absolute values of the components of the matrix.

In addition to replacing $\|S\|_0$ with a convex function, we replace $rank(L)$ with a convex function. Let the singular value decomposition of L be $L = U\Sigma V^*$, and let σ_i be the ith singular value of L. The rank of a matrix is equal to the number of nonzero singular values. Instead of minimizing the number of nonzero singular values of L, we minimize the sum of the singular values, which is a convex function. In fact, the sum of the singular values is the entrywise ℓ_1 norm of Σ. The sum of the singular values is called the "nuclear norm," denoted

$$\|L\|_* := \sum_{i=1}^{r} \sigma_i$$

where r is the rank of L.

Replacing the rank of L with the nuclear norm, and the ℓ_0 norm of S with the ℓ_1 norm, the optimization problem becomes

$$\min \|L\|_* + \lambda \|S\|_1 \quad \text{subject to } M = L + S$$

for some fixed constant λ [2]. This optimization problem is Principal Component Pursuit.

3.3 Conditions Under which Robust PCA Is Well-Defined

Candès et al. impose conditions on L_0 and S_0 under which PCP recovers the original decomposition [2]. To understand the conditions, it is helpful to examine the motivations for them presented in [2]. Each condition is motivated by an obstacle that hinders recovery of (L_0, S_0). Looking at the obstacles helps us understand why the conditions work and whether there are matrices that can be recovered even if they do not satisfy the conditions.

3.4 Conditions on the Low-Rank Component L_0

The first obstacle occurs if the low-rank matrix L_0 is sparse. Because we are trying to separate a matrix M into a low-rank component and a sparse component, the separation is not well-defined if the low-rank component is also sparse.

We show that if the low-rank component is also sparse, then there are multiple ways to decompose the matrix M into the sum of a low-rank matrix and a sparse matrix.

Proposition 3.1. *Suppose M is an $n \times n$ matrix with $M = L_0 + S_0$, where L_0 is a rank-r matrix and S_0 has S nonzero components. If there is a k-sparse vector in the column space of L_0, then we can express M as $M = L_1 + S_1$, where L_1 has rank $r - 1$ and S_1 has at most kn additional nonzero components.*

Proof. Let v_1 be a k-sparse vector in the column space of L_0. Extend v_1 to an orthonormal basis $\{v_1, \cdots, v_r\}$ of the column space of L_0. Any vector in the column space of L_0 can be written as a vector in $span(v_1)$ plus a vector in $span(v_2, \cdots, v_r)$. Let l_i be the ith column of L_0, and let s_i be the ith column of S_0. We write the column l_i as a vector in $span(v_1)$ plus a vector in $span(v_2, \cdots, v_r)$. That is, let $l_i = x_i + w_i$, where $x_i \in span(v_1)$ and $w_i \in span(v_2, \cdots v_r)$.

Given the above notation, we define the matrices L_1 and S_1 as follows. The ith column of L_1 is w_i, and the ith column of S_1 is $x_i + s_i$. The ith column of L_1 and the ith column of S_1 sum to form the ith column of M, so that $L_1 + S_1 = M$:

$$w_i + (x_i + s_i) = (w_i + x_i) + s_i = l_i + s_i = m_i.$$

Because each w_i is in the span of $v_2, \cdots, v_r$, it follows that L_1 has rank $r - 1$. Because x_i is in the span of v_1, and v_1 is k-sparse, it follows that x_i is k-sparse. The original S_0 is S-sparse, and S_1 consists of k-sparse vectors added to the columns of S_0, so S_1 has at most $S + kn$ nonzero components. We have thus provided a way to decompose M into a matrix of rank $r - 1$ plus a matrix with at most $S + kn$ nonzero components. □

To rule out sparse matrices L_0, we impose a condition on L_0 that ensures that there is no sparse vector in the column space of L_0. The conditions we impose are called the incoherence conditions [2]. The incoherence conditions are expressed in terms of the singular value decomposition of L_0. Stating the incoherence conditions requires the following notation. The singular value decomposition of L_0 is

$$L_0 = U\Sigma V^*.$$

Let r be the rank of L_0. Let U_r be the matrix consisting of the first r columns of U, and let V_r be the matrix consisting of the first r columns of V.

The incoherence conditions impose three constraints on the left and right singular vectors of L_0:

1. The projections of the standard basis vectors onto the column space of L_0 are small.
2. The projections of the standard basis vectors onto the row space of L_0 are small.
3. The projections of the rows of U_r onto the rows of V_r are small.

We will show that the incoherence conditions rule out matrices L_0 with sparse vectors in their column spaces. Proposition 3.3 states that if the projections of the standard basis vectors onto a unit vector x are small, then x is not sparse. Thus Proposition 3.3 implies that if the first incoherence condition holds, then the columns of L_0 are not sparse. The following lemma helps in the proof of Proposition 3.3.

Lemma 3.2. *Let x be a vector with ℓ_2 norm equal to 1 and at most S nonzero components. Then x has a component whose absolute value is at least $\frac{1}{\sqrt{S}}$.*

Proof. Let x be a vector with at most S nonzero components, and let I be the set of indices corresponding to the nonzero components of x. If the magnitude of each nonzero component of x is strictly smaller than $1/\sqrt{S}$, then

$$\|x\|_2 = \sqrt{\sum_{i\in I} |x_i|^2} < \sqrt{\sum_{i\in I} \left(\frac{1}{\sqrt{S}}\right)^2} \leq \sqrt{S\left(\frac{1}{\sqrt{S}}\right)^2} = 1.$$

The first inequality holds because the nonzero components of x have magnitude strictly less than $1/\sqrt{S}$. The second inequality holds because x has at most S nonzero components, so there are at most S terms in the summation. We have shown that if x has at most S nonzero components, and the magnitudes of the nonzero components are strictly smaller than $1/\sqrt{S}$, then $\|x\|_2$ is strictly smaller than 1. It follows that if $\|x\|_2 = 1$ and x has at most S nonzero components, then x has at least one component whose magnitude is at least $1/\sqrt{S}$. □

We now show that if the projections of the standard basis vectors onto a unit vector x are sufficiently small, then x is not sparse.

Proposition 3.3. *If the magnitude of the projection of each standard basis vector onto x is smaller than $\frac{1}{\sqrt{S}}$ and $\|x\|_2 = 1$, then x has at least $S+1$ nonzero components.*

Proof. Because e_i is a unit vector, $|\langle e_i, x\rangle| = |x_i|$ is the magnitude of the projection of e_i onto x. If $|x_i| < \frac{1}{\sqrt{S}}$ for each component of x, then because $\|x\|_2 = 1$, it follows from Lemma 3.2 that x has at least $S+1$ nonzero components. □

The first incoherence condition guarantees that the projections of the standard basis vectors onto the column space of L_0 are small. In particular, the projections of the standard basis vectors onto the columns of L_0 are small. Therefore, by Proposition 3.3, the columns of L_0 are not sparse, so the incoherence conditions ensure that L_0 is not sparse.

We now present the formal definition of the incoherence conditions. Recall that L_0 is an $m \times n$ matrix of rank r with SVD $L_0 = U\Sigma V^*$. The matrices U_r and V_r are the matrices consisting of the first r columns of U and V, respectively. The incoherence parameter μ is a property of the matrix L_0. It is the smallest value that satisfies all three inequalities (from [2]):

$$\max_i \|U_r^* e_i\|_2^2 \leq \frac{\mu r}{m}, \tag{1}$$

$$\max_i \|V_r^* e_i\|_2^2 \leq \frac{\mu r}{n}, \tag{2}$$

$$\|U_r V_r^*\|_\infty \leq \sqrt{\frac{\mu r}{mn}}. \tag{3}$$

The third inequality uses the entrywise ℓ_∞ norm, written $\|\cdot\|_\infty$, which is the largest absolute value of any entry in the matrix.

Satisfying the incoherence conditions means having a small value of μ. For the first incoherence condition (the first inequality), a small value of μ means that the projections of the standard basis vectors onto the column space of L_0 are small. For the second incoherence condition, a small value of μ means that the projections of the standard basis vectors onto the row space of L_0 are small. For the third incoherence condition, a small value of μ means that the projections of the rows of U_r onto the rows of V_r are small.

3.5 Conditions on the Sparse Component S_0

The second obstacle that undermines our ability to recover the original matrices (L_0, S_0) is if S_0 is low-rank. If the sparse component is also low-rank, then there is ambiguity regarding whether columns of the sparse component should actually be included within the low-rank component.

Suppose that L_0 has rank r and S_0 has rank k. Then the matrix $L_1 = L_0 + S_0$ has rank at most $r + k$. Therefore the decomposition $M = L_1 + S_1$, where S_1 is the zero matrix, is another decomposition of M into a low-rank component and a sparse component. More generally, if a subset of columns of S_0 spans a k-dimensional subspace, then those columns can be subtracted from S_0 and added to the corresponding columns of L_0 to produce a new

decomposition (L_1, S_1). The matrix L_1 has rank at most $r + k$, and S_1 has fewer nonzero components than S_0 (because a subset of columns of S_0 has been replaced with zeros).

This argument demonstrates that if S_0 is low-rank, then we can produce a different decomposition of M into the sum of a low-rank matrix and a sparse matrix. Because we do not know the rank of L_0 in advance, and we do not know the number of nonzero components of S_0 in advance, we have no way to distinguish the original decomposition (L_0, S_0) from the other decomposition (L_1, S_1).

We now discuss an assumption that rules out matrices S_0 that are low-rank. In [2], the authors adopt the assumption that the locations of the nonzero entries in S_0 have been chosen randomly. Let S be the number of nonzero entries in the matrix S_0. We assume that each set of S entries in the matrix has an equal probability of being chosen as the subset of nonzero entries. Because the locations of the nonzero entries are randomly distributed, the probability that S_0 is low-rank is small.

In practice, the locations of the errors are not necessarily chosen randomly, but this assumption makes the analysis easier. Candès et al. show that if this assumption is satisfied (together with the incoherence conditions), then the PCP optimization problem recovers the original decomposition [2].

3.6 The Main Theorem

The main theorem of [2] states that if the above conditions on L_0 and S_0 are satisfied, then PCP recovers the original decomposition with high probability. (Recall that the only source of randomness is the assumption that the locations of the nonzero entries of S_0 are randomly chosen.) The theorem is as follows:

Theorem 3.4 (from [2])**.** *Let L_0 and S_0 be $n \times n$ matrices that satisfy conditions (1) – (3). Let the locations of the nonzero entries of S_0 be uniformly distributed. Then PCP succeeds with high probability if the following two inequalities hold,*

$$rank(L_0) \leq \rho_r \mu^{-1} \frac{n}{log^2(n)}, \tag{4}$$

$$\|S_0\|_0 \leq \rho_s n^2, \tag{5}$$

where ρ_r and ρ_s are constants.

The first inequality states that the rank of L_0 is smaller than some quantity that depends on the incoherence parameter. Hence the first inequality ensures that the rank is low and the incoherence conditions are satisfied. The second inequality states that the number of nonzero components of S_0 is smaller than some fraction of the number of entries.

These conditions are sufficient for PCP to successfully recover the original decomposition, but they may not be necessary. There may be matrices (L_0, S_0) that are successfully recovered by PCP even though they do not satisfy the above conditions. We investigate the conditions in more detail in the next section.

3.7 The Structure of Robust PCA

We summarize the structure of the robust PCA problem and highlight its parallels with sparse signal recovery.

The goal of robust PCA is to separate a matrix M into a low-rank matrix L_0 and a sparse matrix of errors S_0. This goal corresponds to solving an underdetermined system of equations $M = L + S$, with the additional constraints that L is "low-rank" and S is "sparse." Because there can be multiple decompositions of a matrix into low-rank and sparse components, these constraints must be made more precise for the robust PCA problem to be well-defined. Instead of precisely defining the robust PCA problem, we focus on the conditions under which Principal Component Pursuit recovers matrices L_0 and S_0 from their sum $M = L_0 + S_0$.

To separate the matrix M into low-rank and sparse components, we could formulate the following optimization problem

$$\min \operatorname{rank}(L) + \lambda \|S\|_0 \quad \text{subject to } M = L + S.$$

This objective function is "ideal" because it is conceptually simple—it is a direct formalization of our intuition that L_0 is low-rank and S_0 is sparse. However, it has no efficient solution, so Candès et al. replace it with the tractable convex optimization problem Principal Component Pursuit (PCP) [2]:

$$\min \|L\|_* + \lambda \|S\|_1 \quad \text{subject to } M = L + S.$$

Under conditions on L_0 and S_0 imposed by Theorem 3.4, which ensure that L_0 is not sparse and S_0 is not low-rank, PCP recovers L_0 and S_0 from their sum.

4 Principal Component Pursuit and the Incoherence Conditions

We now investigate the conditions under which PCP successfully solves the robust PCA problem. The incoherence conditions discussed in the previous section are sufficient to guarantee that the PCP algorithm recovers the original low-rank matrix. The motivation behind the incoherence conditions is to ensure that the low-rank matrix L_0 is not sparse. While the motivation is intuitively clear, it is difficult to verify that the incoherence conditions are satisfied without observing L_0. Furthermore, the incoherence conditions may not be necessary. We present experiments that demonstrate that recovery is possible even if the conditions of Theorem 3.4 are violated. Ideally, we would like to identify conditions on (L_0, S_0) that are

1. necessary and sufficient,
2. easily checked,
3. satisfied by a broad class of matrices.

	Robust PCA	SSR
1. Underdetermined system	$L + S = M$	$Ax = y$
2. Extra constraints	L_0 low-rank, S_0 sparse	x_0 sparse
3. Conditions on problem instance to ensure problem is well defined (that is, has at most one solution)	Existence of a unique decomposition of M into $M = L+S$ such that $rank(L) \leq r$ and $\|S\|_0 \leq s$, with r and s fixed constants	A is one-to-one on S-sparse vectors
4. "Ideal" optimization problem	$\min rank(L) + \lambda\|S\|_0$ subject to $L + S = M$	$\min\|x\|_0$ subject to $Ax = y$
5. Tractable optimization problem	$\min\|L\|_* + \lambda\|S\|_1$ subject to $L + S = M$	$\min\|x\|_1$ subject to $Ax = y$
6. Conditions under which the tractable optimization problem recovers the correct solution	Incoherence Conditions (Theorem 3.4)	Restricted Isometry Property

Figure 3: This table presents the structure of robust PCA and SSR, illustrating the parallels between the two problems.

In Section 4.1, we consider whether randomly generated matrices satisfy the incoherence conditions. In Section 4.2, we examine the relationship between the incoherence conditions and the success of PCP.

4.1 Incoherence of Randomly Generated Matrices

To investigate whether randomly generated matrices satisfy the incoherence conditions, use the following procedure. Suppose X is an $n \times r$ matrix with each entry generated independently from a normal distribution of mean 0 and variance $\frac{1}{\sqrt{n}}$. Suppose Y is also an $n \times r$ matrix with each entry generated independently from a normal distribution with mean 0 and variance $\frac{1}{\sqrt{n}}$. Then XY^T is an $n \times n$ matrix of rank r. This observation results in a method to generate rank-r matrices. To obtain an $n \times n$ matrix of rank r, we generate matrices X and Y as described above, then take the product XY^T.

We test whether matrices generated by the above procedure satisfy the incoherence conditions. All experiments were performed in Matlab. For every dimension between 10 and 100 and every rank between 1 and the dimension, we generated a matrix according to the above procedure. For each matrix, we determined the incoherence parameter—the smallest value of μ that satisfies the incoherence inequalities (1) – (3). The plot in Figure 4 shows the incoherence parameter as a function of rank and dimension. The only meaningful values in the plot are the values below the diagonal, because the rank cannot exceed the dimension of the matrix.

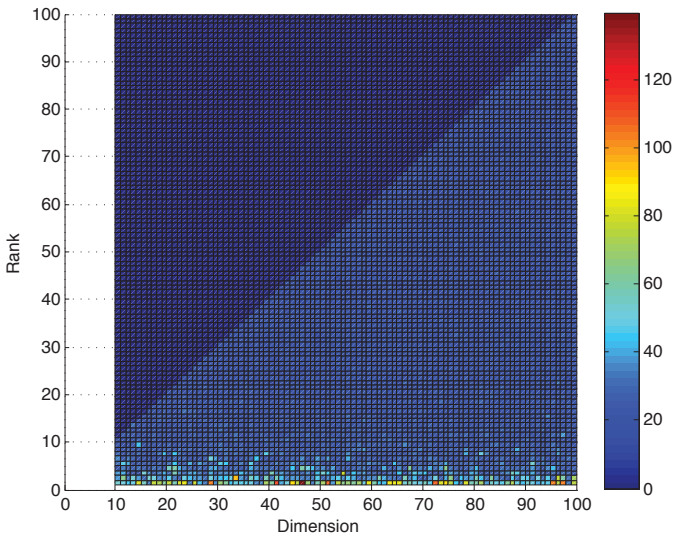

Figure 4: Incoherence parameter as a function of rank and dimension. The incoherence parameter of a matrix is the smallest value of μ that satisfies inequalities (1) – (3). For each dimension between 10 and 100 and each rank between 1 and the dimension, we generated a matrix and calculated the incoherence parameter of the matrix. Each matrix was generated as a product XY^T, where X and Y are $n \times r$ matrices with entries generated from a normal distribution of mean 0 and variance $\frac{1}{\sqrt{n}}$.

As shown in Figure 4, there is a region of high incoherence parameter corresponding to a horizontal band at the bottom of the plot. For matrices with rank between 1 and 5, the incoherence parameter tends to be markedly greater than for matrices with higher rank. The plot suggests that the incoherence parameter is higher for low-rank matrices but decays rapidly as the rank increases. This is expected, because in each of the three inequalities, the incoherence parameter is inversely proportional to the rank:

$$\mu \geq \frac{m}{r} \max_i \|U_r^* e_i\|_2^2,$$
$$\mu \geq \frac{n}{r} \max_i \|V_r^* e_i\|_2^2,$$
$$\mu \geq \frac{mn}{r} \|U_r V_r^*\|_\infty^2.$$

The decay in the incoherence parameter as the rank increases suggests that as the rank increases, the factors $\max_i \|U_r^* e_i\|^2$, $\max_i \|V_r^* e_i\|^2$, and $\|U_r V_r^*\|_\infty^2$ do not increase as rapidly as the rank.

A question that arises from the plot is which of the three inequalities requires μ to have the highest value, thus serving as the binding constraint. We performed an experiment to test which constraint is binding for matrices with normally distributed entries. Following is

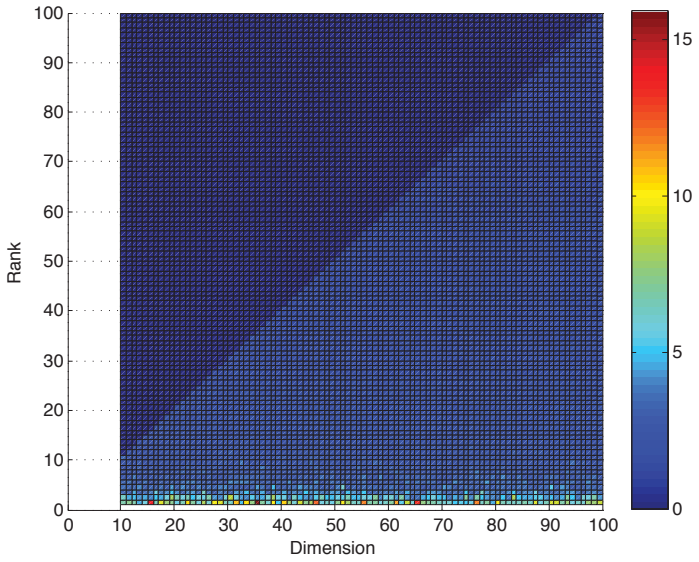

Figure 5: "Left incoherence parameter." Minimum value of μ that satisfies the first incoherence inequality, as a function of rank and dimension. We generated one matrix for each value of dimension and rank. The matrices were generated using the procedure described in Figure 4.

a description of the experiment, with the results displayed in Figure 5. For each dimension between 10 and 100, and each rank between 1 and the dimension, we generated a matrix using the same procedure as for Figure 4. For each matrix, we calculated the smallest value of μ that satisfies the first incoherence inequality. We refer to that value as the "left incoherence parameter," because the first incoherence inequality involves the left singular vectors of the matrix. The plot in Figure 5 shows the left incoherence parameter as a function of rank and dimension for the matrices that we generated. This plot has the same form as the plot in Figure 4. That is, there is a horizontal band at the bottom of the plot corresponding to a region of high left incoherence parameter, and the left incoherence parameter decays as the rank increases. But the magnitudes of the left incoherence parameters in Figure 5 are smaller than the magnitudes of the incoherence parameters in Figure 4. This suggests that for "generic" matrices, the first incoherence inequality is not binding. A similar plot for the second incoherence condition suggests that the second incoherence condition is not binding either. Hence the third inequality is binding. The third incoherence inequality requires that the projections of the rows of U_r onto the rows of V_r are small.

To investigate the incoherence conditions in more detail, we vary the rank and dimension separately. For every rank between 1 and 200, we generated a 200×200 matrix using the procedure described in Figure 4. Figure 6 shows the incoherence parameter as a function of the rank. After a sharp drop-off, the plot fluctuates around a level value. A future question

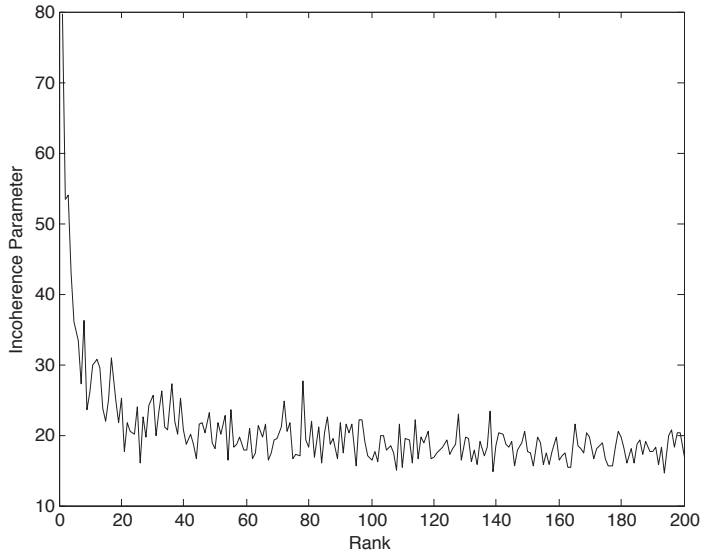

Figure 6: Incoherence parameter as a function of rank, with the dimension fixed at 200×200. The matrices were generated as products XY^T, where X and Y are $200 \times r$ matrices with entries generated from a normal distribution with mean 0 and variance $\frac{1}{\sqrt{n}}$.

to investigate is why the incoherence parameter does not seem to depend on the rank after the initial drop-off.

We also vary the dimension while fixing the rank. For every dimension between 10 and 200, we generated a rank 10 matrix using the above procedure. Figure 7 shows the incoherence parameter as a function of the dimension. The incoherence parameter seems to increase as a function of dimension, as expected from the incoherence inequalities.

Understanding the incoherence of randomly generated matrices gives insight into whether "generic" matrices satisfy the incoherence conditions.

4.2 Incoherence Conditions and the Success of PCP

We investigate how the performance of PCP depends on the incoherence parameter of the low-rank matrix. For each dimension between 10 and 100 and each rank between 1 and the dimension, we generate a matrix using the procedure described in Figure 4. For each matrix we calculate the amount by which the matrix violates the first assumption of the main PCP theorem, Theorem 3.4. The first assumption states that

$$rank(L_0) \leq \rho_r \mu^{-1} \frac{n}{log^2(n)}.$$

The amount by which the inequality is violated is equal to the left-hand side minus the right-hand side. We set the constant ρ_r equal to 1. Figure 8 shows the PCP assumption

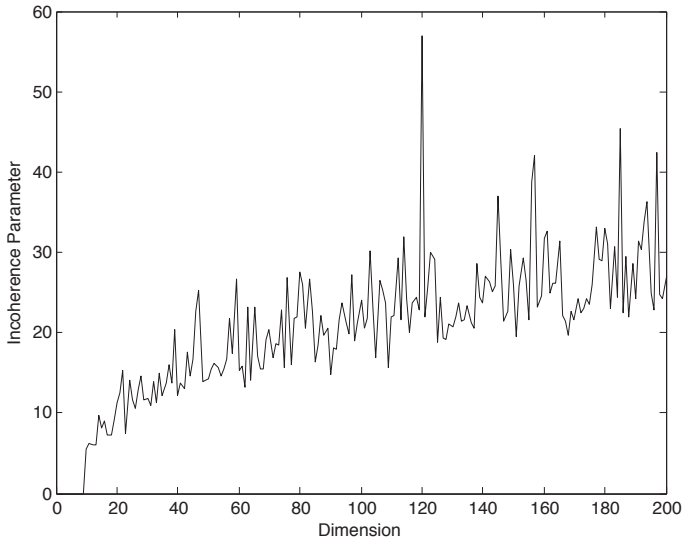

Figure 7: Incoherence parameter as a function of dimension, with the rank fixed at 10. The matrices were generated as products XY^T, where X and Y are $n \times 10$ matrices with entries generated from a normal distribution with mean 0 and variance $\frac{1}{\sqrt{n}}$.

violations as a function of rank and dimension. All of the matrices exhibit violations; none satisfy the first assumption of the main theorem. As the rank increases, the amount by which the inequality is violated increases. The violations do not seem to depend on the dimension.

We performed tests to reveal whether successful recoveries occur despite violations of the sufficient conditions of Theorem 3.4. We generated 30×30 matrices L_0 and S_0, and we tested whether the PCP objective function correctly recovers L_0 and S_0 from their sum $M = L_0 + S_0$. The low-rank matrices L_0 were generated using the procedure described in Figure 4. For the sparse matrices S_0, the locations of the nonzero entires were chosen uniformly at random, the signs of the entries were chosen randomly (positive with probability 1/2), and the magnitudes of the nonzero entries were equal to 10. To solve the PCP optimization problems we used `CVX`, a software package for convex optimization [8]. Figure 9 shows the proportion of successful recoveries as a function of $\text{rank}(L_0)/30$ and of the proportion of nonzero entries in S_0. For each value of the rank of L_0 and the proportion of nonzero entries in S_0, we conducted ten trials of PCP. A trial is considered successful if the relative error of the recovered low-rank matrix is sufficiently small, measured with the Frobenius norm. That is, a trial is successful if the following condition holds

$$\frac{\|L_0 - L_{rec}\|_F}{\|L_0\|_F} < 0.001,$$

where L_{rec} is the low-rank matrix recovered by PCP. Figure 9 shows that despite the violations of the PCP conditions, there are successful recoveries. As the rank of L_0 increases

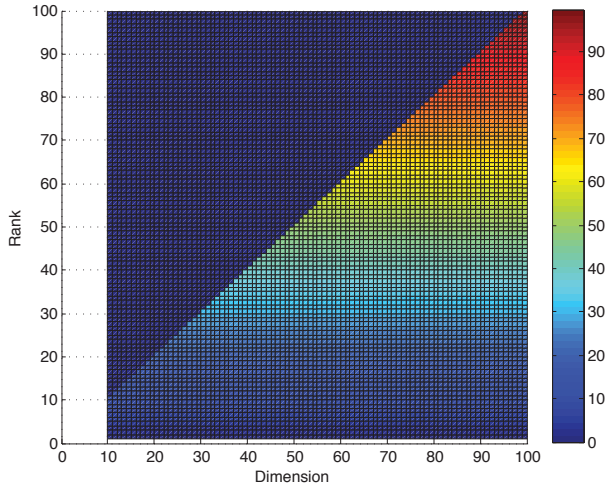

Figure 8: PCP assumption violations as a function of rank and dimension. For each dimension between 10 and 100 and each rank between 1 and the dimension, we generated a low-rank matrix L_0. We measured the violation of inequality (4) in Theorem 3.4. That is, the plot displays the value of $rank(L_0)\mu - \rho_r \frac{n}{\log^2 n}$.

and the proportion of nonzero entries in S_0 increases, the fraction of successful recoveries decreases.

5 Conclusion

[redacted]