# Final Project: Reading Report

Brandyn Tucknott

Due 10 December 2025

This is a reading report on Exact Matrix Completion Via Convex Optimization by Emmanuel J. Candes and Benjamin Recht in 2008.

## Background

As a motivator, suppose we issue a survey to a population and collect responses, forming a data matrix where the rows represent individuals and columns denote questions. Our goal is to gather as much information as possible, but we also know that not all individuals will answer all questions. This naturally leads us to ask: can we make a reasonable guess as to what the unfilled values in our data matrix should be? A more familiar version of this problem is *The Netflix Problem*, where each user has only rated a subset of all available movies, and based on those ratings a recommender system would like to guess new movies the user might enjoy.

## Problem Formulation Rationale

It turns out we can classify such problems as matrix completion problems. Suppose we have an $n \times n$ matrix $M$ with rank $r$. (Here a square matrix is used for simplicity, but neither the argument nor intuition depends on $M$ being square) It is clear $M$ can be represented with $n^2$ numbers, so our goal is to recover $M$ using no more than $n^2$ measurements. Even with low rank $M$, it is clear this is not generally true. For example, consider $M$ a 1-sparse matrix (matrix with only 1 element non-zero). Then any way you sample the matrix (especially for large $n$), it is unlikely that you would stumble upon the non-zero entry and correctly recover the original matrix. Fortunately, while it is impossible to recover all low-rank matrices, we can still recover a significant amount.

Consider the singular value decomposition (SVD) of $M$ as

$$M = \sum_{k=1}^{n} \sigma_k u_k v_k^*$$

where $u_k, v_k$ are the left and right singular vectors, and $\sigma_k$ are the singular values. Let $\mathcal{U}_r = \{u_k : 1 \leq k \leq r\}$ and $\mathcal{V}_r = \{v_k : 1 \leq k \leq r\}$ and define a *generic* low-rank matrix as selecting uniformly at random a family $\mathcal{U}_r, \mathcal{V}_r$ from all families of $r$ orthonormal vectors. We make no assumptions about the depedence or independence of $\mathcal{U}_r, \mathcal{V}_r$ nor about the singular values $\sigma_k$. This is referred to as a *random orthogonal model*.

Now to address sampling, let $\Omega = \{(i, j) : 0 \leq i, j \leq n\}$ where $M_{ij}$ is observed. Let $m = |\Omega|$ denote the number of samples and assume our sampling has a random uniform distribution. Intuitively, it would be helpful if there were only one such low-rank matrix satisfying the samples, in which case we could solve for the optimization problem

$$\underset{X}{\text{minimize}} \ \text{rank}(X)$$
$$\text{subject to } X_{ij} = M_{ij}, \text{ for all } (i, j) \in \Omega.$$

This is unfortunately NP-hard, and all known algorithms have a doubly-exponential time complexity for an exact solution[1]. Instead, we consider the *nuclear norm*, defined to be

$$||X||_* = \sum_{k=1}^{r} \sigma_k(X)$$

where $M$ has rank $r$ and $\sigma_1(X), \ldots, \sigma_r(X)$ are the singular values of $M$ in decreasing order. Now instead of optimizing for rank, we choose to optimize over the nuclear norm, giving us

$$\underset{X}{\text{minimize}} \ ||X||_*$$
$$\text{subject to } X_{ij} = M_{ij}, (i,j) \in \Omega.$$

Intuitively the rank function counts the number of non-vanishing singular values, and the nuclear norm measures the amplitude of our singular values. We refer to this optimization problem as OPT 1. It turns out this small change is still valid with enough samples (sufficiently large $m = |\Omega|$), but to see why we need to introduce some theorems.

**Theorem 1.1.** *Let $M$ be an $n_1 \times n_2$ matrix of rank $r$ sampled from the random orthogonal model, and put $n = \max(n_1, n_2)$. Suppose we observe $m$ entries of $M$ with locations sampled uniformly at random. Then there are numerical constants $C, c$ such that if*

$$m \geq Cn^{5/4}r \log n,$$

*the minimizer to the problem OPT 1 is unique and equal to $M$ with probability at least $1 - cn^{-3}$; that is to say, the semi-definite program OPT 1 recovers all the entries of $M$ with no error. In addition, if $r \leq n^{1/5}$, then recovery is exact with probability at least $1 - cn^{-3}$ provided that*

$$m \geq Cn^{6/5}r \log n.$$

There are a few key insights from theorem 1.1:

- There is a unique low-rank matrix consistent with our observations.

- Such a matrix is recoverable via convex optimization (nuclear norm minimization is formally equivalent to rank minimization).

- Only a small number of observations is required, much less than the naive $n_1 \cdot n_2$.

Let $U$ be a subspace of $\mathbb{R}^n$ with dimension $r$ and $P_U$ be the orthogonal projection onto $U$. Then the *coherence* of $U$ with respect to the standard basis $e_i$ is defined to be

$$\mu(U) \equiv \frac{n}{r} \max_{1 \leq i \leq n} ||P_U e_i||^2.$$

In order to obtain our main result (exact low-rank matrix recovery), we make two assumptions[1]. For some $\mu_0, \mu_1 > 0$:
**A0:** The $\max(\mu(U), \mu(V)) \leq \mu_0$.
**A1:** The $n_1 \times n_2$ matrix $\sum_{1 \leq k \leq r} u_k v_k^*$ has a maximum entry in absolute value is bounded by $\mu_1/\sqrt{r/(n_1 n_2)}$. It is easy to verify that **A1** always holds with $\mu_1 = \mu_0\sqrt{r}$, so for small ranks $\mu_1 \approx \mu_0$. For larger ranks, we will later see that the subspaces $U, V$ are incoherent with the standard basis and that $\mu_1$ is at most $\log \max(n_1, n_2)$[1].

**Theorem 1.2.** *Let $M$ be an $n_1 \times n_2$ matrix of rank $r$ obeying **A0** and **A1**, and put $n = \max(n_1, n_2)$. Suppose we observe $m$ entries of $M$ with locations sampled uniformly at random. Then there exist constants $C, c$ such that*

$$m \geq C \max(\mu_1^2, \mu_0^{1/2}, \mu_1, \mu_0 n^{1/4})nr\beta \log n$$

*for some $\beta > 2$. Then the minimizer to OPT 1 is unique and equal to M with probability at least $1 - cn^{-\beta}$. For $r \le \mu_0^{-1} n^{1/5}$, this estimate can be improved to*

$$m \ge C\mu_0 n^{6/5} r\beta \log n$$

*with the same probability of success.*

Our takeaway from theorem 1.2 is that with low coherence, fewer observations are required to recover $M$[1].

# Algorithm Overview

The algorithm itself is quite simple, and is simply OPT 1 (nuclear norm minimization).

$$\underset{X}{\text{minimize}} \ ||X||_*$$
$$\text{subject to } X_{ij} = M_{ij}, (i,j) \in \Omega.$$

If we know our original matrix $M$ is PSD, then we can equivalently solve the following optimization problem:

$$\underset{X}{\text{minimize}} \ \text{trace}(X)$$
$$\text{subject to } X_{ij} = M_{ij}, (i,j) \in \Omega$$
$$X \succeq 0.$$

# Implementations

The algorithm was originally solved using SDPT3, a MATLAB package to solve semidefinite problems (SDPs). We will use CVXPY to achieve similar results. Denote the size of our matrix $M$ with $n$, the rank as $r$, and the number of observations as $m$. Each $(n, m, r)$ triple is repeated 50 times

## Existing Examples - Gaussian

## Existing Examples - PSD

## New Examples

# References

[1]: E. J. Candès and B. Recht. Exact Matrix Completion via Convex Optimization. arXiv preprint arXiv:0805.4471, 2008