

## 1 Introduction to the Matrix Completion Problem

(Notes heavily borrow from Fall 2013 notes by Kristan Temme and Yun William Yu)

This is sometimes called the Netflix problem. A motivation for the matrix completion problem comes from user ratings of some products which are put into a matrix  $M$ . The entries  $M_{ij}$  of the matrix correspond to the  $j$ 'th user's rating of product  $i$ . We assume that there exists an ideal matrix that encodes the ratings of all the products by all the users. However, it is not possible to ask every user his opinion about every product. We are only given some ratings of some users and we want to recover the actual ideal matrix  $M$  from this limited data. So matrix completion is the following problem:

**Problem:** Suppose you are given some matrix  $M \in \mathbb{R}^{n_1 \times n_2}$ . Moreover, you also are given some entries  $(M_{ij})_{ij \in \Omega}$  with  $|\Omega| \ll n_1 n_2$ .

**Goal:** We want to recover the missing elements in  $M$ .

This problem is impossible if we don't make any additional assumptions on the matrix  $M$  since the missing  $M_{ij}$  could in principle be arbitrary. We will discuss a recovery scheme that relies on the following three assumptions.

1.  $M$  is (approximately) low rank.
2. Both the column space and the row space are "incoherent". We say a space is incoherent, when the projection of any vector onto this space has a small  $\ell_2$  norm.
3. If  $M = U\Sigma V^T$  then all the entries of  $UV^T$  are bounded.
4. The subset  $\Omega$  is chosen uniformly at random.

**Note 1.** *There is work on adversarial recovery where the values are not randomly chosen but carefully picked to trick us by an adversary.*

Under these assumptions we show that there exists an algorithm that needs a number of entries in  $M$  bounded by  $|\Omega| \leq (n_1 + n_2) r \text{ poly}(\log(n_1 n_2)) \cdot \mu$ . Here  $\mu$  captures to what extent properties 2 and 3 above hold. One would naturally consider the following recovery method for the matrix  $M$ :

$$\begin{aligned} & \text{minimize} && \text{rank}(X) \\ & \text{subject to:} && X_{ij} = M_{ij} \quad \forall i, j \in \Omega. \end{aligned}$$

Unfortunately this optimization problem is *NP*-hard. We will therefore consider the following alternative optimization problem in trace norm, or *nuclear norm*.

$$\begin{aligned} & \text{minimize} && \|X\|_* \\ & \text{subject to:} && X_{ij} = M_{ij} \quad \forall i, j \in \Omega, \end{aligned}$$

where the nuclear norm of  $X$  defined as the sum of the singular values of  $X$ , i.e.  $\|X\|_* = \sum_i \sigma_i(X)$ . This problem is an SDP (semi- definite program), and can be solved in time polynomial in  $n_1 n_2$ .

## 2 Work on Matrix Completion

Let's now go through the history of prior work on this problem. Recall the setup and model:

- Matrix completion setup:
  - Want to recover  $M \in \mathbb{R}^{n_1 \times n_2}$ , under the assumption that  $\text{rank}(M) = r$ , where  $r$  is small.
  - Only some small subset of the entries  $(M_{ij})_{ij \in \Omega}$  are revealed, where  $\Omega \subset [n_1] \times [n_2]$ ,  $|\Omega| = m \ll n_1, n_2$
- Model:
  - $m$  times we sample  $i, j$  uniformly at random + insert into  $\Omega$  (so  $\Omega$  is a multiset).
  - Note that the same results hold if we choose  $m$  entries without replacement, but it's easier to analyze this way. In fact, if you can show that if recovery works with replacement, then that implies that recovery works without replacement, which makes sense because you'd only be seeing more information about  $M$ .
- We recover  $M$  by Nuclear Norm Minimization (NNM):
  - Solve the program  $\min \|X\|_*$  s.t.  $\forall i, j \in \Omega, X_{ij} = M_{ij}$
- [Recht, Fazel, Parrilo '10] [RFP10] was first to give some rigorous guarantees for NNM.
- [Candés, Recht, '09] [CR09] was the first paper to show provable guarantees for NNM applied to matrix completion.
- There were some quantitative improvements (in the parameters) in two papers: [Candés, Tao '09] [CT10] and [Keshavan, Montanari, Oh '09] [KMO10]
- **Today** we're going to cover an analysis given in [Recht, 2011] [Rec11], which has a couple of advantages.
  - First, it has the laxest of all the conditions.
  - Second, it's also the simplest of all the analyses in the papers.
  - Thus, it's really better in every way there is.

The approach of [Rec11] was inspired by work in quantum tomography [GLF+10]. A more general theorem than the one proven in class today was later proven by Gross [Gross].

It is worth noting that there have been other important works on matrix completion which we will not get to in the course. In particular, one particular paradigm is *Alternating Minimization (AM)*.

The basic idea behind AM is as follows. It is an iterative algorithm. We try to find an approximate rank- $k$  factorization  $M \approx X \cdot Y$ , where  $X$  has  $k$  columns and  $Y$  has  $k$  rows. We start off with initial  $X_0, Y_0$ . Then we do as follows:

1. initialize  $X_0, Y_0$  (somehow)
2. **for**  $\ell = 1, \dots, T$ :
  - (a)  $X_\ell \leftarrow \operatorname{argmin}_X \|R_\Omega(M - XY_{\ell-1})\|_F^2$
  - (b)  $Y_\ell \leftarrow \operatorname{argmin}_Y \|R_\Omega(M - X_\ell Y)\|_F^2$
3. **return**  $X_T, Y_T$

Rigorous analyses of modifications of the above AM template have been carried out in [1, 2]. The work [3] has also shown some performance guarantees when the revealed entries are *adversarial* except for random (though in this case, many more entries have to be revealed).

### 3 Theorem Statement

We're almost ready to formally state the main theorem, but need a couple of definitions first.

**Definition 2.** Let  $M = U\Sigma V^*$  be the singular value decomposition. (Note that  $U \in \mathbb{R}^{n_1 \times r}$  and  $V \in \mathbb{R}^{n_2 \times r}$ .)

**Definition 3.** Define the incoherence of the subspace  $U$  as  $\mu(U) = \frac{n_1}{r} \cdot \max_i \|P_U e_i\|^2$ , where  $P_U$  is projection onto  $U$ . Similarly, the incoherence of  $V$  is  $\mu(V) = \frac{n_2}{r} \cdot \max_i \|P_V e_i\|^2$ , where  $P_V$  is projection onto  $V$ .

**Definition 4.**  $\mu_0 \stackrel{\text{def}}{=} \max\{\mu(U), \mu(V)\}$ .

**Definition 5.**  $\mu_1 \stackrel{\text{def}}{=} \|UV^*\|_\infty \sqrt{n_1 n_2 / r}$ , where  $\|UV\|_\infty$  is the largest magnitude of an entry of  $UV$ .

**Theorem 6.** If  $m \gtrsim \max\{\mu_1^2, \mu_0\} \cdot n_2 r \log^2(n_2)$  then with high probability  $M$  is the unique solution to the semi-definite program  $\min \|X\|_*$  s.t.  $\forall i, j \in \Omega, X_{ij} = M_{ij}$ .

Note that  $1 \leq \mu_0 \leq \frac{n_2}{r}$ . The way  $\mu_0$  can be  $\frac{n_2}{r}$  is if a standard basis vector appears in a column of  $V$ , and the way  $\mu_0$  can get all the way down to 1 is like the best case scenario where all the entries of  $V$  are like  $\frac{1}{\sqrt{n_2}}$  and all the entries of  $U$  are like  $\frac{1}{\sqrt{n_1}}$ , so for example if you took a Fourier matrix and cut off some of its columns. Thus, the condition on  $m$  is a good bound if the matrix has low incoherence.

One might wonder about the necessity of all the funny terms in the condition on  $m$ . Unfortunately, [Candes, Tao, '09] [CT10] showed  $m \gtrsim \mu_0 n_2 r \log(n_2)$  is needed (that is, there is a family of examples

$M$  that need this). If you want to have any decent chance of recovering  $M$  over the random choice of  $\Omega$  using this SDP, then you need to sample at least that many entries. The condition isn't completely tight because of the square in the log factor and the dependence on  $\mu_1^2$ . However, Cauchy-Schwarz implies  $\mu_1^2 \leq \mu_0^2 r$ .

Just like in compressed sensing, there are also some iterative algorithms to recover  $M$ , but we're not going to analyze them in class. For example, the **SparSA** algorithm given in [Wright, Nowak, Figueiredo '09] [WNF09] (thanks for Ben Recht for pointing this out to me). That algorithm roughly looks as follows when one wants to minimize  $\|AX - M\|_F^2 + \mu\|X\|_*$ :

Pick  $X_0$ , and a stepsize  $t$  and iterate (a)-(d) some number of times:

- (a)  $Z = X_k - t \cdot A^T(AX_k - M)$
- (b)  $[U, \mathbf{diag}(s), V] = \mathbf{svd}(Z)$
- (c)  $r = \max(s - \mu t, 0)$
- (d)  $X_{k+1} = U \mathbf{diag}(r) V^T$

As an aside, trace-norm minimization is actually tolerant to noise, but I'm not going to cover that.

## 4 Analysis

The way that the analysis is going to go is we're going to condition on lots of good events all happening, and if those good events happen, then the minimization works. The way I'm going to structure the proof is I'll first state what all those events are, then I'll show why those events make the minimization work, and finally I'll bound the probability of those events not happening.

### 4.1 Background and more notation

Before I do that, I want to say some things about the trace norm.

**Definition 7.**  $\langle A, B \rangle \stackrel{\text{def}}{=} \text{Tr}(A^*B) = \sum_{i,j} A_{ij}B_{ij}$

**Claim 8.** *The dual of the trace norm is the operator norm:*

$$\|A\|_* = \sup_{\substack{B \text{ s.t.} \\ \|B\| \leq 1}} \langle A, B \rangle$$

This makes sense because the dual of  $\ell_1$  for vectors is  $\ell_\infty$  and this sort of looks like that because the trace norm and operator norm are respectively like the  $\ell_1$  and  $\ell_\infty$  norm of the singular value vector. More rigorously, we can prove it by proving inequality in both directions. One direction is not so hard, but the other requires the following lemma.

**Lemma 9.**

$$\underbrace{\|A\|_*}_{(1)} = \underbrace{\min_{\substack{X, Y \text{ s.t.} \\ A=XY^*}} \|X\|_F \cdot \|Y\|_F}_{(2)} = \underbrace{\min_{\substack{X, Y \text{ s.t.} \\ A=XY^*}} \frac{1}{2} \left( \|X\|_F^2 + \|Y\|_F^2 \right)}_{(3)}$$

*Proof of lemma.*

**(2)  $\leq$  (3):**

AM-GM inequality:  $xy \leq \frac{1}{2}(x^2 + y^2)$ .

**(3)  $\leq$  (1):**

We basically just need to exhibit an  $X$  and  $Y$  which give something that is at most the  $\|A\|_*$ . Set  $X = Y^* = A^{1/2}$ . In general, given  $f : \mathbb{R}^+ \mapsto \mathbb{R}^+$ , then  $f(A) = Uf(\Sigma)V^*$ . i.e. write the SVD of  $A$  and apply  $f$  to each diagonal entry of  $\Sigma$ . You can easily check that  $A^{1/2}A^{1/2} = A$  and that the square of the Frobenius norm of  $A^{1/2}$  is exactly the trace norm.

**(1)  $\leq$  (2):**

Let  $X, Y$  be some matrices such that  $A = XY^*$ . Then

$$\begin{aligned}
 \|A\|_* &= \|XY^*\|_* \\
 &\leq \sup_{\substack{\{a_i\} \text{ orthonormal basis} \\ \{b_i\} \text{ orthonormal basis}}} \sum_i \langle XY^*a_i, b_i \rangle && \text{This can be seen to be true by letting} \\
 &= \sup_{\dots} \sum_i \langle Y^*a_i, X^*b_i \rangle && \begin{array}{l} a_i=v_i \text{ and } b_i=u_i \\ \text{(from the SVD), when we get equality.} \end{array} \\
 &\leq \sup_{\dots} \sum_i \|Y^*a_i\| \cdot \|X^*b_i\| \\
 &\leq \sup_{\dots} \left( \sum_i \|Y^*a_i\|^2 \right)^{1/2} \left( \sum_i \|X^*b_i\|^2 \right)^{1/2} && \text{(by Cauchy-Schwarz)} \quad (1) \\
 &= \|X\|_F \cdot \|Y\|_F && \begin{array}{l} \text{because } \{a_i\}, \{b_i\} \text{ are orthonormal bases} \\ \text{and the Frobenius norm is rotationally invariant} \end{array}
 \end{aligned}$$

□

*Proof of claim.*

**Part 1:**

$$\|A\|_* \leq \sup_{\|B\|=1} \langle A, B \rangle.$$

We show this by writing  $A = U\Sigma V^*$ . Then take  $B = \sum_i u_i v_i^*$ . That will give you something on the right that is at least the trace norm.

**Part 2:**

$$\|A\|_* \geq \langle A, B \rangle \quad \forall B \text{ s.t. } \|B\| = 1.$$

We show this using the lemma.

- Write  $A = XY^*$  s.t.  $\|A\|_* = \|X\|_F \cdot \|Y\|_F$  (lemma guarantees that there exists such an  $X$  and  $Y$ ).
- Write  $B = \sum_i \sigma_i a_i b_i, \forall i, \sigma_i \leq 1$ .

Then using a similar argument to (1),

$$\begin{aligned}
\langle A, B \rangle &= \left\langle XY^*, \sum_i \sigma_i a_i b_i \right\rangle \\
&= \sum_i \sigma_i \langle Y^* a_i, X^* b_i \rangle \\
&\leq \sum_i |\langle Y^* a_i, X^* b_i \rangle| \\
&\leq \|X\|_F \|Y\|_F = \|A\|_*
\end{aligned}$$

which concludes the proof of the claim.  $\square$

Recall that the set of matrices that are  $n_1 \times n_2$  is itself a vector space. I'm going to decompose that vector space into  $T$  and the orthogonal complement of  $T$  by defining the following projection operators.

- $P_{T^\perp}(Z) \stackrel{\text{def}}{=} (I - P_U)Z(I - P_V)$
- $P_T(Z) \stackrel{\text{def}}{=} Z - P_{T^\perp}(Z)$

So basically, the matrices that are in the vector space  $T^\perp$  are the matrices that can be written as the sum of rank 1 matrices  $a_i b_i^*$  where the  $a_i$ 's are orthogonal to all the  $u$ 's and the  $b_i$ 's are orthogonal to all the  $v$ 's. Also define  $R_\Omega(Z)$  as only keeping entries in  $\Omega$ , multiplied by multiplicity in  $\Omega$ . If you think of the operator  $R_\Omega : \mathbb{R}^{n_1 n_2} \mapsto \mathbb{R}^{n_1 n_2}$  as a matrix, it is a diagonal matrix with the multiplicity of entries in  $\Omega$  on the diagonal.

## 4.2 Good events

We will condition on all these events happening in the analysis. It will turn out that with high probability—probability  $1 - \frac{1}{\text{poly}(n_2)}$ , and you can make the  $\frac{1}{\text{poly}(n_2)}$  factor decay as much as you want by increasing the constant in front of  $m$ —all these events will occur:

1.  $\left\| \frac{n_1 n_2}{m} P_T R_\Omega P_T - P_T \right\| \lesssim \sqrt{\frac{\mu_0 r (n_1 + n_2) \log(n_2)}{m}} \ll \frac{1}{2}$

This is simple to understand from the perspective of leverage score sampling for approximate matrix multiplication (AMM) with spectral norm error (as in pset 4, problem 2). Specifically, recall that AMM we have matrices  $A, B$  with the same number  $n$  of rows and want for some  $\Pi$  with  $m$  rows that  $\|(\Pi A)^T (\Pi A) - A^T B\| \leq \varepsilon \|A\| \cdot \|B\|$ . Now, note here that  $P_T = P_T P_T$ , since  $P_T$  is a projection matrix. Thus the above is just an AMM condition for  $A = B = P_T$ , and  $\Pi = R_\Omega^{1/2}$ . Now, typically for row sampling we had  $\Pi$  be a diagonal matrix with  $\Pi_{i,i} = \eta_i / \sqrt{p_i}$ , where  $\eta_i$  is an indicator random variable for the event that we sampled row  $i$ , and  $p_i = \mathbb{E} \eta_i$ . In class we discussed that we should set  $p_i$  to be roughly proportional to the *leverage score* of row  $i$ . The total number of samples is thus on the order of the sum of leverage scores. More specifically, according to pset 4 problem 2, the total number of rows samples will be on the order of  $O(q \log(q/\delta)/\varepsilon^2)$  to succeed with probability  $1 - \delta$ , where  $q$  is the sum of the leverage scores (or equivalently, the maximum rank of  $A, B$ ). In our

case, the rank of  $P_T$  is the sum of leverage scores of  $P_T$ , which is  $\sum_{a,b} \|P_T e_a e_b^*\|_F^2$ , which is  $\sum_{a,b} (\|P_U e_a\|_2^2 + \|P_V e_b\|_2^2 - \|P_U e_a\|_2^2 \cdot \|P_V e_b\|_2^2)$ . Here  $P_U$  is orthogonal projection onto  $U$ , where  $M = U\Sigma V^*$ . One can verify that this sum is  $n_1 r + n_2 r - r^2 = r(n_1 + n_2 - r) = O((n_1 + n_2)r)$ . Thus  $q \log(q/\delta)$  is  $O(r(n_1 + n_2) \log(n_2/\delta)) = O(r n_2 \log(n_2/\delta))$  (note what  $m$  is above!). Now, unfortunately  $R_\Omega$  samples *uniformly* and not according to leverage scores! This is where  $\mu_0$  comes in. We need to make sure we oversample enough so that each row's expected number of occurrences in our sampling is *at least* its leverage score (show via a modification of your pset 4 pset 2 solution that this suffices). To make this oversampling good enough, we need to oversample by a factor related to the maximum leverage score, hence the  $\mu_0$  in  $m$ .

2.  $\|R_\Omega\| \lesssim \log(n_2)$

This one is actually really easy (also the shortest): it's just balls and bins. We've already said  $R_\Omega$  is a diagonal matrix, so the operator norm is just the largest diagonal entry. Imagine we have  $m$  balls, and we're throwing them independently at random into  $n_1 n_2$  bins, namely the diagonal entries, and this is just how loaded is the maximum bin. In particular,  $m < n_1 n_2$ , or else we wouldn't be doing matrix completion since we'd have the whole matrix. In general, when you throw  $t$  balls into  $t$  bins, the maximum load by the Chernoff bound is at most  $\log t$ . In fact, it's at most  $\log t / \log \log t$ , but who cares, since that would save us an extra  $\log \log$  somewhere. Actually, I'm not even sure it would save us that since there are other  $\log$ 's that come into play.

3.  $\exists Y$  in  $range(R_\Omega)$  s.t.

(5a)  $\|P_T(Y) - UV^*\|_F \leq \sqrt{\frac{r}{2n_2}}$

(5b)  $\|P_{T^\perp}(Y)\| < \frac{1}{2}$

We will not justify this one in class; see the paper for the argument for the existence of such a  $Y$ .

### 4.3 Recovery conditioned on good events

Now that we've stated all these things, let's show that they imply trace norm minimization actually works. We want to make sure

$$\underset{R_\Omega(X)=R_\Omega(M)}{\operatorname{argmin}} \quad X \text{ s.t.} \quad \|X\|_*$$

is unique and equal to  $M$ .

Let  $Z \in Ker(R_\Omega)$ , ( $Z \neq 0$ ); we want to show  $\|M + Z\|_* > \|M\|_*$ .

First we want to argue that  $\|P_T(Z)\|_F$  cannot be big.

**Lemma 10.**  $\|P_T(Z)\|_F < \sqrt{\frac{n_2}{2r}} \cdot \|P_{T^\perp}(Z)\|_F$

*Proof.*

$$0 = \|R_\Omega(Z)\|_F \geq \|R_\Omega(P_T(Z))\|_F - \|R_\Omega(P_{T^\perp}(Z))\|_F$$

Also

$$\begin{aligned}
\|R_\Omega(P_T(Z))\|_F^2 &= \langle R_\Omega P_T Z, R_\Omega P_T Z \rangle \\
&\geq \langle P_T Z, R_\Omega P_T Z \rangle \\
&= \langle Z, P_T R_\Omega P_T Z \rangle \\
&= \langle P_T Z, P_T R_\Omega P_T P_T Z \rangle \\
&= \left\langle P_T Z, \frac{m}{n_1 n_2} P_T P_T Z \right\rangle + \left\langle P_T Z, \left( P_T R_\Omega P_T - \frac{m}{n_1 n_2} \right) P_T Z \right\rangle \\
&\geq \frac{m}{n_1 n_2} \|P_T Z\|_F^2 - \left\| P_T R_\Omega P_T - \frac{m}{n_1 n_2} \right\| \cdot \|P_T Z\|_F^2 \\
&\geq \frac{m}{n_1 n_2} \cdot \|P_T Z\|_F^2
\end{aligned}$$

Also have

$$\begin{aligned}
\|R_\Omega(P_{T^\perp}(Z))\|_F^2 &\leq \|R_\Omega\|^2 \cdot \|P_{T^\perp}(Z)\|_F^2 \\
&\lesssim \log^2(n_2) \cdot \|P_{T^\perp}(Z)\|_F^2
\end{aligned}$$

**Summarize:** combining all the inequalities together, and then making use of our choice of  $m$ ,

$$\begin{aligned}
\|P_T(Z)\|_F &< \sqrt{\frac{n_1 n_2 \log^2(n_2)}{m}} \cdot \|P_{T^\perp}(Z)\|_F \\
&< \sqrt{\frac{n_2}{2r}} \cdot \|P_{T^\perp}(Z)\|_F
\end{aligned}$$

□

Pick  $U_\perp, V_\perp$  s.t.  $\langle U_\perp V_\perp^*, P_{T^\perp}(Z) \rangle = \|P_T(Z)\|_*$  and s.t.  $[U, U_\perp], [V, V_\perp]$  orthogonal matrices. We know from claim 8 that the trace norm is exactly the sup over all  $B$  matrices of the inner product. But the  $B$  matrix that achieves the sup has all singular values equal to 1, so  $B = U_\perp V_\perp^*$ , because  $P_{T^\perp}(Z)$  is in the orthogonal space so  $B$  should also be in the orthogonal space.

Now we have a long chain of inequalities to show that the trace of any  $M + Z$  is greater than the trace of  $M$ :

$$\begin{aligned}
\|M + Z\|_* &\geq \langle UV^* + U_\perp V_\perp^*, M + Z \rangle && \text{by claim 8} \\
&= \|M\|_* + \langle UV^* + U_\perp V_\perp^*, Z \rangle && \text{since } M \perp U_\perp V_\perp^* \\
&= \|M\|_* + \langle UV^* + U_\perp V_\perp^* - Y, Z \rangle && \text{since } Z \in \ker(R_\Omega) \\
&&& \text{and } Y \in \text{range}(R_\Omega) \\
&= \|M\|_* + \langle UV^* - P_T(Y), P_T(Z) \rangle + \langle U_\perp V_\perp^* - P_{T^\perp}(Y), P_{T^\perp}(Z) \rangle && \text{decomposition into } T \text{ \& } T^\perp \\
&\geq \|M\|_* - \|UV^* - P_T(Y)\|_F \cdot \|P_T(Z)\|_F && \langle x, y \rangle \leq \|x\|_2 \|y\|_2 \\
&\quad + \|P_{T^\perp}(Z)\|_* && \text{by our choice of } UV^* \\
&\quad - \|P_{T^\perp}(Y)\| \cdot \|P_{T^\perp}(Z)\|_* && \text{norm inequality} \\
&\geq \|M\|_* - \sqrt{\frac{r}{2n_2}} \cdot \|P_T(Z)\|_F + \frac{1}{2} \cdot \|P_{T^\perp}(Z)\|_* \\
&> \|M\|_* - \frac{1}{2} \cdot \|P_T^\perp(Z)\|_F + \frac{1}{2} \cdot \|P_{T^\perp}(Z)\|_* && \text{by Lemma 10} \\
&\geq \|M\|_* && \text{since } \|\cdot\|_* \geq \|\cdot\|_F
\end{aligned}$$

Hence, when all of the good conditions hold, minimizing the trace norm recovers  $M$ .

## 5 Concluding remarks

Why would you think of trace minimization as solving matrix completion? Analogously, why would you use  $\ell_1$  minimization for compressed sensing? In some way, these two questions are very similar in that rank is like the support size of the singular value vector, and trace norm is the  $\ell_1$  norm of the singular value vector, so the two are very analogous.  $\ell_1$  minimization seems like a natural choice, since it is the closest convex function to support size from all the  $\ell_p$  norms (and being convex allows us to solve the program in polynomial time).

## References

- [CR09] Emmanuel J Candès and Benjamin Recht, *Exact matrix completion via convex optimization*, Foundations of Computational mathematics **9** (2009), no. 6, 717–772.
- [CT10] Emmanuel J Candès and Terence Tao, *The power of convex relaxation: Near-optimal matrix completion*, Information Theory, IEEE Transactions on **56** (2010), no. 5, 2053–2080.
- [Gross] David Gross, *Recovering low-rank matrices from few coefficients in any basis*, Information Theory, IEEE Transactions on (2011), no. 57, :1548-1566.
- [GLF+10] David Gross, Yi-Kai Liu, Steven T. Flammia, Stephen Becker, and Jens Eisert. *Quantum state tomography via compressed sensing*, Physical Review Letters (2010), 105(15):150401.
- [1] Moritz Hardt. *Understanding Alternating Minimization for Matrix Completion*. FOCS, pages 651–660, 2014.

- [2] Moritz Hardt, Mary Wootters. *Fast matrix completion without the condition number*. COLT, pages 638–678, 2014.
- [KMO10] Raghunandan H Keshavan, Andrea Montanari, and Sewoong Oh, *Matrix completion from noisy entries*, The Journal of Machine Learning Research **99** (2010), 2057–2078.
- [Rec11] Benjamin Recht, *A simpler approach to matrix completion*, The Journal of Machine Learning Research **12** (2011), 3413–3430.
- [RFP10] Benjamin Recht, Maryam Fazel, and Pablo A Parrilo, *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*, SIAM review **52** (2010), no. 3, 471–501.
- [3] Tselil Schramm, Benjamin Weitz. *Low-Rank Matrix Completion with Adversarial Missing Entries*. CoRR abs/1506.03137, 2015.
- [WNF09] Stephen J Wright, Robert D Nowak, and Mário AT Figueiredo, *Sparse reconstruction by separable approximation*, Signal Processing, IEEE Transactions on **57** (2009), no. 7, 2479–2493.