

Exact Matrix Completion via Convex Optimization

By Emmanuel Candès and Benjamin Recht

Abstract

Suppose that one observes an incomplete subset of entries selected from a low-rank matrix. When is it possible to complete the matrix and recover the entries that have not been seen? We demonstrate that in very general settings, one can perfectly recover all of the missing entries from most sufficiently large subsets by solving a convex programming problem that finds the matrix with the minimum nuclear norm agreeing with the observed entries. The techniques used in this analysis draw upon parallels in the field of compressed sensing, demonstrating that objects other than signals and images can be perfectly reconstructed from very limited information.

1. INTRODUCTION

In many practical problems of interest, one would like to recover a matrix from a sampling of its entries. As a motivating example, consider the task of inferring answers in a partially filled out survey in which questions are asked to a collection of individuals. Then we can form a matrix where the rows index the individuals and the columns index the questions. We collect data to fill out this table, but unfortunately, many questions are left unanswered. Is it possible to make an educated guess about what the missing answers should be? How can one make such a guess? Formally, we may view this problem as follows. We are interested in recovering a data matrix \mathbf{M} with n_1 rows and n_2 columns but have access to only m of its entries, where m is much smaller than the total number of entries, $n_1 n_2$. Can one recover the matrix \mathbf{M} from m of its entries? In general, everyone would agree that this is impossible without some additional information.

In many instances, however, the matrix we wish to recover is known to be structured in the sense that it is low-rank or approximately low-rank. (We recall for completeness that a matrix has rank r if its rows or columns span an r -dimensional space.) Consider the following two scenarios as prototypical examples.

► *The Netflix problem.* In the area of recommender systems, users submit ratings on a subset of entries in a database, and the vendor provides recommendations based on the user's preferences.³¹ Because users only rate a few items, one would like to infer their preference for unrated items. A special instance of this problem is the now famous Netflix problem.²⁴ Users (rows of the data matrix) are given the opportunity to rate movies (columns of the data matrix), but users typically rate only very few movies so that there are very few scattered observed entries of this data matrix. Yet, one would like to complete this matrix so that

the vendor (here Netflix) might recommend titles that any particular user is likely to be willing to order. In this case, the data matrix of all user-ratings may be approximately low-rank, because only a few factors contribute to an individual's tastes or preferences.

► *Triangulation from incomplete data.* Suppose we are given partial information about the distances between objects and would like to reconstruct the low-dimensional geometry describing their locations. For example, we may have a network of low-power, wirelessly networked sensors scattered randomly across a region. Suppose each sensor only has the ability to construct distance estimates based on signal strength readings from its nearest fellow sensors. From these local distance estimates, we can form a partially observed distance matrix. We can then estimate the true distance matrix whose rank will be equal to 2 if the sensors are located in a plane or 3 if they are located in three-dimensional space.^{26, 32} In this case, we only need to observe a few distances per node to have enough information to reconstruct the positions of the objects.

These examples are of course far from exhaustive and there are many other problems which fall in this general category.

Suppose for simplicity that we wish to recover a square $n \times n$ matrix \mathbf{M} of rank r . Although \mathbf{M} contains n^2 numbers, our assumption that its rank is r means that it can be represented exactly by its singular value decomposition (SVD)

$$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_{k=1}^r \sigma_k \mathbf{u}_k \mathbf{v}_k^T \quad (1.1)$$

where \mathbf{V}^T denotes the transpose of \mathbf{V} . $\mathbf{\Sigma}$ is an $r \times r$ diagonal matrix with real, positive elements $\sigma_k > 0$. \mathbf{U} is an $n \times r$ matrix with orthonormal columns $\mathbf{u}_1, \dots, \mathbf{u}_r$. That is, $\mathbf{u}_k^T \mathbf{u}_k = 1$ and $\mathbf{u}_i^T \mathbf{u}_j = 0$ if $i \neq j$. \mathbf{V} is also $n \times r$ with orthonormal columns $\mathbf{v}_1, \dots, \mathbf{v}_r$. The column space of \mathbf{M} is spanned by the columns of \mathbf{U} , and the row space is spanned by the columns of \mathbf{V} .

The number of degrees of freedom associated with a rank r matrix \mathbf{M} is $r(2n - r)$. To see this, note that $\mathbf{\Sigma}$ has r nonzero entries, and \mathbf{U} and \mathbf{V} each have nr total entries. Since \mathbf{U} and \mathbf{V} each satisfy $r(r + 1)/2$ orthogonality constraints, the total number of degrees of freedom is $r + 2nr - r(r + 1) = r(2n - r)$. Thus, when r is much smaller than n , there are significantly

The original version of this paper was published in *Foundations of Computational Mathematics* 9, 6 (2009), 717–772.

fewer degrees of freedom than the size of \mathbf{M} would suggest. The question is then whether \mathbf{M} can be recovered from a suitably chosen sampling of its entries without collecting n^2 measurements.

In this paper, we demonstrate that most low-rank matrices can be indeed recovered from a very sparse sampling of their entries. In Section 2, we summarize the main results of our paper, highlighting the necessary assumptions, algorithmic ingredients, and theoretical foundations of reconstructing matrices from a presented collection of entries. In Section 3, we survey the subsequent developments in this area, including refinements and important extensions of our theory. We close with a discussion of further progress and advances in low-rank and sparse modeling.

2. MATRIX COMPLETION

Which matrices?

In general, one cannot hope to be able to recover a low-rank matrix from a sample of its entries. Consider the rank 1 matrix \mathbf{M} equal to

$$\mathbf{M} = \mathbf{e}_1 \mathbf{x}^T = \begin{bmatrix} x_1 & x_2 & \cdots & x_{n-1} & x_n \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}, \quad (2.1)$$

where here and throughout, \mathbf{e}_i is the i th canonical basis vector in Euclidean space (the vector with all entries equal to 0 but the i th equal to 1). The matrix \mathbf{M} has the entries of \mathbf{x} along its first row and all the other entries are 0. Clearly, this matrix cannot be recovered from a sampling of its entries unless we see *all* of the entries in the first row. As another example, the matrix $\mathbf{e}_1 \mathbf{e}_n^T$ is a matrix with a 1 in the $(1, n)$ entry and 0s everywhere else. If we do not see this upper right corner, then we cannot distinguish the matrix from the all 0s matrix.

Even if it is impossible to recover *all* low-rank matrices from a set of sampled entries, can one recover *most* of them? To investigate this possibility, we introduce a simple model of low-rank matrices.

DEFINITION 2.1. *Let \mathbf{M} be a rank r matrix with SVD defined by (1.1). Then we say that \mathbf{M} belongs to the random orthogonal model if the family $\{\mathbf{u}_k\}_{1 \leq k \leq r}$ is selected uniformly at random among all families of r orthonormal vectors, and similarly for $\{\mathbf{v}_k\}_{1 \leq k \leq r}$. The two families may or may not be independent of each other. We make no assumptions about the singular values, σ_k .*

If a matrix is sampled from the random orthogonal model, then we would expect most of the entries to be nonzero. This model is convenient in the sense that it is both very concrete and simple, and useful in the sense that it will help us fix the main ideas. In the sequel, however, we will consider far more general models. The question for now is whether or not one can recover such a generic matrix from a sampling of its entries.

Which sampling sets?

Clearly, one cannot hope to reconstruct any low-rank matrix \mathbf{M} —even of rank 1—if the sampling set avoids any column or row of \mathbf{M} . Suppose that \mathbf{M} is of rank 1 and of the form $\mathbf{x}\mathbf{y}^T$, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ so that the (i, j) entry is given by $M_{ij} = x_i y_j$. Then, if we do not have samples from the first row, one could never infer the value of the first component x_1 as no information about x_1 is observed. There is, of course, nothing special about the first row and this argument extends to any row or column. To have any hope of recovering an unknown matrix, one needs to have access to at least one observation per row and one observation per column.

This example demonstrates that there are sampling sets where one would not even be able to recover matrices of rank 1. But what happens for typical sampling sets? Can one recover a low-rank matrix from almost all sampling sets of cardinality m ? Formally, suppose that the set Ω of locations corresponding to the observed entries ($(i, j) \in \Omega$ if M_{ij} is observed) is a set of cardinality m sampled uniformly at random. Then, can one recover a generic low-rank matrix \mathbf{M} , perhaps with very large probability, from the knowledge of the value of its entries in the set Ω ?

Which algorithm?

If the number of measurements is sufficiently large, and if the entries are close to uniformly distributed, one might hope that there is only *one* low-rank matrix with these entries. If this were true, one would want to recover the data matrix by solving the optimization problem

$$\begin{aligned} & \text{minimize} && \text{rank}(\mathbf{X}) \\ & \text{subject to} && X_{ij} = M_{ij} \quad (i, j) \in \Omega \\ & && \mathbf{X} \in \mathbb{R}^{n \times n}, \end{aligned} \quad (2.2)$$

where \mathbf{X} is the decision variable and $\text{rank}(\mathbf{X})$ is equal to the rank of the matrix \mathbf{X} . The program (2.2) is a common sense approach which simply seeks the simplest explanation fitting the observed data. If there were only one low-rank object fitting the data, the solution of (2.2) would recover \mathbf{M} perfectly. This is unfortunately of little practical use, because not only is this optimization problem NP-hard but also all known algorithms which provide exact solutions require time doubly exponential in the dimension n of the matrix in both theory and practice.

If a matrix has rank r , then it has exactly r nonzero singular values so that the rank function in (2.2) is simply the number of nonvanishing singular values. In this paper, we consider an alternative which minimizes the sum of the singular values over the constraint set. This sum is called the *nuclear norm*,

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

where, here and below, $\sigma_k(\mathbf{X})$ denotes the k th largest singular value of \mathbf{X} . The heuristic optimization we study is then given by

$$\begin{aligned} & \text{minimize} && \|\mathbf{X}\|_* \\ & \text{subject to} && X_{ij} = M_{ij} \quad (i, j) \in \Omega. \end{aligned} \quad (2.3)$$

Whereas the rank function is equal to the number of non-vanishing singular values, the nuclear norm equals their sum. The nuclear norm is to the rank functional what the convex ℓ_1 norm is to the ℓ_0 norm in the area of sparse signal recovery. The main point here is that the nuclear norm is a convex function and can be optimized efficiently via semi-definite programming.¹⁴

There are many norms one could define for a given matrix. The *operator norm* is the largest singular value. The *Frobenius norm* is equal to the square root of the sum of the squares of the entries. This norm is akin to the standard Euclidean norm on a real vector space. Why should the nuclear norm provide lower rank solutions than either of these two more commonly studied norms?

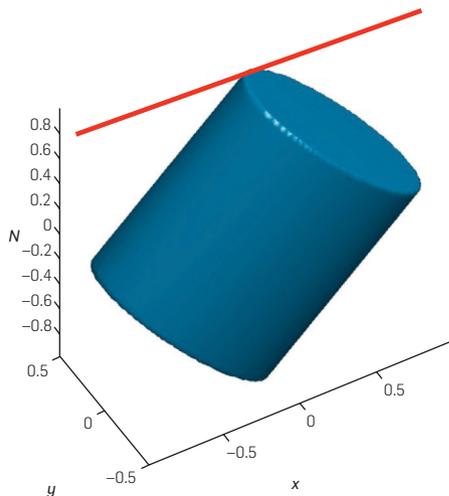
One can gain further intuition by analyzing the geometric structure of the nuclear norm ball. The unit nuclear norm ball is precisely the convex hull of the rank 1 matrices of unit Frobenius norm. The nuclear norm minimization problem (2.3) can be interpreted as inflating the unit ball until it just touches the affine space $X_{ij} = M_{ij}$. Such an intersection will occur at an extreme point of the nuclear norm ball, and these extreme points are sparse convex combinations of rank 1 matrices. That is, the extreme points of the nuclear norm ball have low rank. This phenomenon is depicted graphically in Figure 1. There, we plot the unit ball of the nuclear norm for matrices parametrized as

$$\mathbf{X} = \begin{bmatrix} x & y \\ y & z \end{bmatrix}.$$

The extreme points of this cylindrical object are the rank 1 matrices with unit Frobenius norm. The red line in this figure is a “random,” one-dimensional, affine subspace which, as expected, intersects the nuclear norm ball at a rank 1 matrix.

As further motivation, an interesting connection exists between the nuclear norm and popular algorithms in

Figure 1. Unit ball of the nuclear norm for symmetric 2×2 matrices. The red line depicts a random one-dimensional affine space. Such a subspace will generically intersect a sufficiently large nuclear norm ball at a rank one matrix.



data-mining and collaborative filtering. In these fields, researchers commonly aim to find an explicit factorization $\mathbf{X} = \mathbf{L}\mathbf{R}^T$ that agrees with the measured entries. Here \mathbf{L} and \mathbf{R} are $n \times k$ matrices. Since there are many possible such factorizations that might agree with the observations, a common approach searches for matrices \mathbf{L} and \mathbf{R} that have Frobenius norm as small as possible, that is, the solution of the optimization problem

$$\begin{aligned} & \text{minimize} && \frac{1}{2}(\|\mathbf{L}\|_F^2 + \|\mathbf{R}\|_F^2) \\ & \text{subject to} && \mathbf{L}\mathbf{R}^T = \mathbf{X} \\ & && X_{ij} = M_{ij} \quad (i, j) \in \Omega, \end{aligned} \quad (2.4)$$

where we are minimizing with respect to $\mathbf{L} \in \mathbb{R}^{n \times k}$, $\mathbf{R} \in \mathbb{R}^{n \times k}$, and $\mathbf{X} \in \mathbb{R}^{n \times n}$, and $\|\cdot\|_F$ denotes the Frobenius norm. Surprisingly, the optimization problem (2.4) is equivalent to minimization of the nuclear norm subject to the same equality constraints provided k is chosen to be larger than the rank of the optimum of the nuclear norm problem (2.3).³⁰

To get an intuition for this equivalence, take any matrix \mathbf{X} of rank k . Suppose the SVD is $\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^T$. If we set $\mathbf{L} := \mathbf{U}\Sigma^{\frac{1}{2}}$ and $\mathbf{R} := \mathbf{V}\Sigma^{\frac{1}{2}}$, we see that

$$\frac{1}{2}\|\mathbf{L}\|_F^2 + \frac{1}{2}\|\mathbf{R}\|_F^2 = \frac{1}{2}\sum_{i=1}^n \sum_{j=1}^k (U_{ij}^2 \sigma_j + V_{ij}^2 \sigma_j) = \sum_{j=1}^k \sigma_j = \|\mathbf{X}\|_*$$

because $\sum_i U_{ij}^2 = \sum_i V_{ij}^2 = 1$ for all j . Thus, the optimal solution of (2.3) is suboptimal for (2.4). The full equivalence can be seen via an appeal to semidefinite programming and can be found in Recht et al.³⁰

The main advantage of this reformulation (2.4) is to substantially decrease the number of decision variables from n^2 to $2nr$. For large problems, this leads to a significant reduction in computation time, such that very large instances can be solved on a desktop computer. On the other hand, the formulation (2.4) is nonconvex and thus potentially has local minima that are not globally optimal. Nonetheless, this factored approximation (2.4) of the nuclear norm is one of the most successful stand-alone approaches to solving the Netflix Prize problem.^{16,24} Indeed, it was one of the foundational components of the winning team’s prediction engine.

2.1. Main results

As seen in our first example (2.1), it is impossible to recover a matrix which is equal to 0 in nearly all of its entries unless we see all the entries of the matrix. This is particularly likely if the singular vectors of a matrix \mathbf{M} have most of their mass concentrated in a few coordinates. For instance, consider the rank 2 symmetric matrix \mathbf{M} given by

$$\mathbf{M} = \sum_{k=1}^2 \sigma_k \mathbf{u}_k \mathbf{u}_k^T, \quad \begin{aligned} \mathbf{u}_1 &= (\mathbf{e}_1 + \mathbf{e}_2)/\sqrt{2}, \\ \mathbf{u}_2 &= (\mathbf{e}_1 - \mathbf{e}_2)/\sqrt{2}, \end{aligned}$$

where the singular values are arbitrary. Then, this matrix vanishes everywhere except in the top-left 2×2 corner, and one would basically need to see all the entries of \mathbf{M} to be able to recover this matrix exactly. There is an endless list of examples of this sort. Hence, we arrive at the notion that the singular vectors need to be sufficiently spread across

all components—that is, uncorrelated with the standard basis—in order to minimize the number of observations needed to recover a low-rank matrix. This motivates the following definition.

DEFINITION 2.2. Let U be a subspace of \mathbb{R}^n of dimension r and \mathbf{P}_U be the orthogonal projection onto U . Then the coherence of U (vis-à-vis the standard basis (\mathbf{e}_i)) is defined to be

$$\mu(U) \equiv \frac{n}{r} \max_{1 \leq i \leq n} \|\mathbf{P}_U \mathbf{e}_i\|^2.$$

Note that for any subspace, the smallest $\mu(U)$ can be is 1, achieved, for example, if U is spanned by vectors whose entries all have magnitude $1/\sqrt{n}$. The largest possible value for $\mu(U)$ is n/r which would correspond to any subspace that contains a standard basis element. Matrices whose column and row spaces have low coherence are likely not to vanish in too many entries and are our most likely candidates for matrices that are recoverable from a few samples. As we discuss below, subspaces sampled from the random orthogonal model (Definition 2.1) have nearly minimal coherence.

To state our main result, we introduce two assumptions about an $n_1 \times n_2$, rank r matrix \mathbf{M} whose SVD is given by (1.1) and with column and row spaces denoted by U and V , respectively.

- A0** The coherences obey $\max(\mu(U), \mu(V)) \leq \mu_0$ for some $\mu_0 > 0$.
- A1** The $n_1 \times n_2$ matrix $\sum_{1 \leq k \leq r} \mathbf{u}_k \mathbf{v}_k^T$ has a maximum entry bounded by $\mu_1 \sqrt{r/(n_1 n_2)}$ in absolute value for some $\mu_1 > 0$.

These definitions implicitly define two critical parameters, μ_0 and μ_1 . These μ 's may depend on r and n_1, n_2 . Moreover, note that **A1** always holds with $\mu_1 = \mu_0 \sqrt{r}$ since the (i, j) th entry of the matrix $\sum_{1 \leq k \leq r} \mathbf{u}_k \mathbf{v}_k^T$ is given by $\sum_{1 \leq k \leq r} u_{ik} v_{jk}$ and by the Cauchy–Schwarz inequality,

$$\left| \sum_{1 \leq k \leq r} u_{ik} v_{jk} \right| \leq \sqrt{\sum_{1 \leq k \leq r} |u_{ik}|^2} \sqrt{\sum_{1 \leq k \leq r} |v_{jk}|^2} \leq \frac{\mu_0 r}{\sqrt{n_1 n_2}}.$$

Hence, for sufficiently small ranks, μ_1 is comparable to μ_0 . We say that a subspace $U \subset \mathbb{R}^n$ is *incoherent* with the standard basis if $\mu(U)$ is at most logarithmic in n . As we show in the full version of this paper that, for larger ranks, both subspaces selected from the uniform distribution and spaces constructed as the span of singular vectors with bounded entries are not only incoherent with the standard basis but also obey **A1** with high probability for values of μ_1 at most logarithmic in n_1 and/or n_2 .

We are now in a position to state our main result: if a matrix has row and column spaces that are incoherent with the standard basis, then nuclear norm minimization can recover this matrix from a random sampling of a small number of entries.

THEOREM 2.3. Let \mathbf{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 \mathbf{M} with locations sampled uniformly at random. Then there exist constants C, c such that if

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

for some $\beta > 2$, then the minimizer to the problem (2.3) is unique and equal to \mathbf{M} with probability at least $1 - cn^{-\beta}$. For $r \leq \mu_0^{-1} n^{1/5}$ this estimate can be improved to

$$m \geq C \mu_0 n^{6/5} r (\beta \log n)$$

with the same probability of success.

Theorem 2.3, proven in the full version of this paper, asserts that if the coherence is low, few samples are required to recover \mathbf{M} . For example, if μ_0 is a small constant and the rank is not too large, then the recovery is exact with large probability provided that

$$m \geq C n^{6/5} r \log n. \quad (2.5)$$

We give two illustrative examples of matrices with incoherent column and row spaces. This list is by no means exhaustive.

1. The first example is the random orthogonal model (see Definition 2.1). For values of the rank r greater than $\log n$, $\mu(U)$ and $\mu(V)$ are $O(1)$, $\mu_1 = O(\log n)$ both with very large probability. Hence, the recovery is exact on most sampling sets provided that $m \leq C n^{5/4} r \log n$. When $r \leq n^{1/5}$, we can strengthen this bound to $m \leq C n^{6/5} r \log n$.
2. The second example is more general and simply requires that the components of the singular vectors of \mathbf{M} are small. Assume that the \mathbf{u}_j and \mathbf{v}_j 's obey

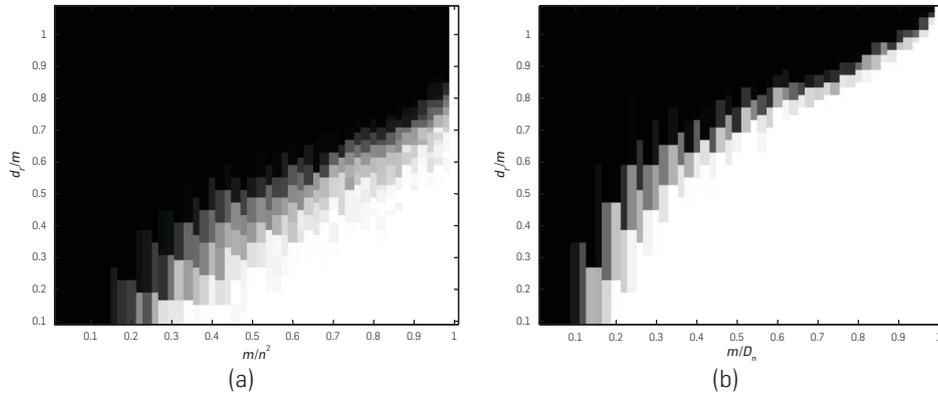
$$\max_{ij} |\langle \mathbf{e}_i, \mathbf{u}_j \rangle|^2 \leq \mu_b / n, \quad \max_{ij} |\langle \mathbf{e}_i, \mathbf{v}_j \rangle|^2 \leq \mu_b / n, \quad (2.6)$$

for some value of $\mu_b = O(1)$. Then, the maximum coherence is at most μ_b since $\mu(U) \leq \mu_b$ and $\mu(V) \leq \mu_b$. Further, we show in the full version of this paper that **A1** holds most of the time with $\mu_1 = O(\sqrt{\log n})$. Thus, for matrices with singular vectors obeying (2.6), the recovery is exact provided that m obeys (2.5) for values of the rank not exceeding $\mu_b^{-1} n^{1/5}$.

2.2. Numerical validation

To demonstrate the practical applicability of the nuclear norm heuristic for recovering low-rank matrices from their entries, we conducted a series of numerical experiments for a variety of the matrix sizes n , ranks r , and numbers of entries m . For each (n, m, r) triple, we repeated the following procedure 50 times. We generated \mathbf{M} , an $n \times n$ matrix of rank r , by sampling two $n \times r$ factors \mathbf{M}_L and \mathbf{M}_R with i.i.d. Gaussian entries and setting $\mathbf{M} = \mathbf{M}_L \mathbf{M}_R^T$. We sampled a subset Ω of m entries uniformly at random. Then, the nuclear norm minimization problem was solved using the semi-definite programming solver, SeDuMi.³³ We declared \mathbf{M} to be recovered if the solution returned by the solver, \mathbf{X}_{opt} , satisfied $\|\mathbf{X}_{\text{opt}} - \mathbf{M}\|_F / \|\mathbf{M}\|_F < 10^{-3}$. Figure 2 shows the results of these experiments for $n = 50$. The x -axis corresponds to the fraction of the entries of the matrix that are revealed to the SDP solver. The y -axis corresponds to the ratio between the

Figure 2. Recovery of full matrices from their entries. For each (n, m, r) triple, we repeated the following procedure 50 times. A matrix \mathbf{M} of rank r and a subset of m entries were selected at random. Then, we solved the nuclear norm minimization for \mathbf{X} subject to $X_{ij} = M_{ij}$ on the selected entries. We declared \mathbf{M} to be recovered if $\|\mathbf{X}_{\text{opt}} - \mathbf{M}\|_F / \|\mathbf{M}\|_F < 10^{-3}$. The results are shown for (a) general 50×50 matrices (b) 50×50 positive definite matrices. The color of each cell reflects the empirical recovery rate (scaled between 0 and 1). White denotes perfect recovery in all experiments, and black denotes failure for all experiments.



dimension of the rank r matrices, $d_r = r(2n - r)$, and the number of measurements m .

Note that the axes range from 0 to 1 as a value > 1 on the x -axis corresponds to an overdetermined linear system where the semidefinite program always succeeds, and a value > 1 on the y -axis corresponds to when there are an infinite number of rank r matrices with the provided entries. The color of each cell in the figures reflects the empirical recovery rate of the 50 runs (scaled between 0 and 1). Interestingly, the experiments reveal very similar plots for different n , suggesting that our theoretical upper bounds on recovery may be rather conservative.

For a second experiment, we generated random *positive semidefinite* matrices and tried to recover them from their entries using the nuclear norm heuristic. As above, we repeated the same procedure 50 times for each (n, m, r) triple. We generated \mathbf{M} , an $n \times n$ positive semidefinite matrix of rank r , by sampling an $n \times r$ factor \mathbf{M}_F with i.i.d. Gaussian entries and setting $\mathbf{M} = \mathbf{M}_F \mathbf{M}_F^T$. We sampled a subset Ω of m entries uniformly at random. Then, we solved the nuclear norm minimization problem with an additional constraint that the decision variable be positive definite. Figure 2(b) shows the results of these experiments for $n = 50$. The x -axis again corresponds to the fraction of the entries of the matrix that are revealed to the solver, but, in this case, the number of measurements is divided by $D_n = n(n + 1)/2$, the number of unique entries in a positive-semidefinite matrix, and the dimension of the rank r matrices is $d_r = nr - r(r - 1)/2$. The color of each cell is chosen in the same fashion as in the experiment with full matrices. Interestingly, the recovery region is much larger for positive semidefinite matrices, and future work is needed to investigate if the theoretical scaling is also more favorable in this scenario of low-rank matrix completion.

These phase transition diagrams reveal a considerably smaller region of parameter space than the Gaussian models studied in Recht et al.³⁰ In the experiments in Recht et al.,³⁰ \mathbf{M} was generated in the same fashion as above, but, in the

place of sampling entries, we generated m random Gaussian projections of the data (see the discussion in Section 2.4). In these experiments, the recovery regime is far larger than that in the case of sampling entries, but this is not particularly surprising as each Gaussian observation measures a contribution from every entry in the matrix \mathbf{M} .

2.3. More general bases

Our main result (Theorem 2.3) extends to a variety of other low-rank matrix completion problems beyond the sampling of entries. Indeed, suppose we have two orthonormal bases $\mathbf{f}_1, \dots, \mathbf{f}_n$ and $\mathbf{g}_1, \dots, \mathbf{g}_n$ of \mathbb{R}^n , and that we are interested in solving the rank minimization problem

$$\begin{aligned} & \text{minimize} && \text{rank}(\mathbf{X}) \\ & \text{subject to} && \mathbf{f}_i^T \mathbf{X} \mathbf{g}_j = \mathbf{f}_i^T \mathbf{M} \mathbf{g}_j, \quad (i, j) \in \Omega. \end{aligned} \quad (2.7)$$

The machine learning community's interest in specialized algorithms for *multiclass* and *multitask* learning provides a motivating example (see, e.g., Amit et al.¹ and Argyriou et al.²). In multiclass learning, the goal is to build multiple classifiers with the same training data to distinguish between more than two categories. For example, in face recognition, one might want to classify whether an image patch corresponds to an eye, nose, or mouth. In multitask learning, we have a large set of data and a variety of different classification tasks, but, for each task, only partial subsets of the data are relevant. For instance, in activity recognition, we may have acquired sets of observations of multiple subjects and want to determine if each observed person is walking or running. However, a different classifier is desired for each individual, and it is not clear how having access to the full collection of observations can improve classification performance. Multitask learning aims to take advantage of access to the full database to improve performance on individual tasks. A description of how to apply our results to the multiclass setting can be

found in the full version of this paper.

To see that our theorem provides conditions under which (2.7) can be solved via nuclear norm minimization, note that there exist unitary transformations \mathbf{F} and \mathbf{G} such that $\mathbf{e}_j = \mathbf{F}\mathbf{f}_j$ and $\mathbf{e}_j = \mathbf{G}\mathbf{g}_j$ for each $j = 1, \dots, n$. Hence,

$$\mathbf{f}_i^T \mathbf{X} \mathbf{g}_j = \mathbf{e}_i^T (\mathbf{F} \mathbf{X} \mathbf{G}^T) \mathbf{e}_j.$$

Then, if the conditions of Theorem 2.3 hold for the matrix $\mathbf{F} \mathbf{X} \mathbf{G}^T$, it is immediate that nuclear norm minimization finds the unique optimal solution of (2.7) when we are provided a large enough random collection of the inner products $\mathbf{f}_i^T \mathbf{M} \mathbf{g}_j$. In other words, all that is needed is that the column and row spaces of \mathbf{M} be, respectively, incoherent with the bases (\mathbf{f}_i) and (\mathbf{g}_j) .

2.4. Connections, alternatives, and prior art

Nuclear norm minimization is a recent heuristic introduced by Fazel¹⁴ and is an extension of the trace heuristic often used in control theory; see, for example, Beck and D’Andrea.³ Indeed, when the matrix variable is symmetric and positive semidefinite, the nuclear norm of \mathbf{X} is the sum of the (nonnegative) eigenvalues and thus equal to the trace of \mathbf{X} . Hence, for positive semidefinite unknowns, \mathbf{X} , (2.3) becomes the semidefinite program

$$\begin{array}{ll} \text{minimize} & \text{trace}(\mathbf{X}) \\ \text{subject to} & X_{ij} = M_{ij} \quad (i, j) \in \Omega \\ & \mathbf{X} \succeq 0. \end{array}$$

Even for the general matrix \mathbf{M} , which may not be positive definite or even symmetric, the nuclear norm heuristic can be formulated in terms of semidefinite programming. The program (2.3) is equivalent to

$$\begin{array}{ll} \text{minimize} & \text{trace}(\mathbf{W}_1) + \text{trace}(\mathbf{W}_2) \\ \text{subject to} & X_{ij} = M_{ij} \quad (i, j) \in \Omega \\ & \begin{bmatrix} \mathbf{W}_1 & \mathbf{X} \\ \mathbf{X}^T & \mathbf{W}_2 \end{bmatrix} \succeq 0, \end{array}$$

with optimization variables \mathbf{X} , \mathbf{W}_1 , and \mathbf{W}_2 (see, e.g., Fazel¹⁴). There are many efficient algorithms and high-quality software packages available for solving these types of problems.

Our work is inspired by results in the emerging field of *compressive sampling* or *compressed sensing*, a new paradigm for acquiring information about objects of interest from what appears to be a highly incomplete set of measurements.^{8, 13} In practice, this means that high-resolution images can be captured with fewer sensors or that signal acquisition can be accelerated by orders of magnitude in biomedical applications, simply by taking far fewer specially coded samples. Mathematically speaking, we wish to reconstruct a signal $\mathbf{x} \in \mathbb{R}^n$ from a small number of measurements $\mathbf{y} = \Phi \mathbf{x}$, $\mathbf{y} \in \mathbb{R}^m$ with m much smaller than n ; that is, we have far fewer equations than unknowns. In general,

one cannot hope to reconstruct \mathbf{x} but assume now that the object we wish to recover is known to be structured in the sense that it is sparse (or approximately sparse). This means that the unknown object depends upon a smaller number of unknown parameters. Then, it has been shown that ℓ_1 minimization—minimizing the sum of the absolute values of \mathbf{x} , subject to the linear constraints $\mathbf{y} = \Phi \mathbf{x}$ —allows recovery of sparse signals from remarkably few measurements.⁸ If Φ is chosen randomly from a suitable distribution, then with very high probability, all sparse signals with about k nonzero entries can be recovered from on the order of $k \log n$ measurements. For instance, if \mathbf{x} is k -sparse in the Fourier domain, that is, \mathbf{x} is a superposition of k sinusoids, then it can be perfectly recovered with high probability—by ℓ_1 minimization—from the knowledge of about $k \log n$ of its entries sampled uniformly at random.

From this viewpoint, the results in this paper greatly extend the theory of compressed sensing by showing that other types of interesting objects or structures, beyond sparse signals and images, can be recovered from a limited set of measurements. Moreover, the techniques for proving our main results build upon ideas from the compressed sensing literature together with powerful probabilistic tools for bounding norms of operators between Banach spaces.

Also, our notion of coherence generalizes the concept of the same name in compressive sensing. Notably, the authors Candès and Romberg⁷ introduce the notion of the coherence of a unitary transformation \mathbf{U} ; the coherence of \mathbf{U} is simply proportional to $\max_{j,k} |U_{j,k}|^2$. This quantity plays a crucial role in determining the minimal sampling rate necessary to recover a k -sparse signal by ℓ_1 minimization.

In Recht et al.,³⁰ the authors studied the nuclear norm heuristic applied to a related problem where partial information about a matrix \mathbf{M} is available from m equations of the form

$$\langle \mathbf{A}^{(k)}, \mathbf{M} \rangle = \sum_{ij} A_{ij}^{(k)} M_{ij} = b_k, \quad k = 1, \dots, m. \quad (2.8)$$

Here, for each k , $\{A_{ij}^{(k)}\}_{ij}$ is an i.i.d. sequence of Gaussian or Bernoulli random variables and the sequences $\{\mathbf{A}^{(k)}\}$ are also independent of each other (the sequences $\{\mathbf{A}^{(k)}\}$ and $\{b_k\}$ are available to the analyst). Building on the concept of *restricted isometry* in the context of sparse signal recovery, Recht et al.³⁰ establish the first sufficient conditions for which the nuclear norm heuristic returns the minimum rank element in the constraint set. The authors prove that the heuristic succeeds with large probability whenever the number m of available measurements is greater than a constant times $2nr \log n$ for $n \times n$ matrices of rank r . These results do not generalize to the matrix completion problem of interest to us in this paper. The measurements in (2.8) give some information about *all* the entries of \mathbf{M} , whereas in our problem information about most of the entries is simply not available. As a consequence, our methods are quite different and require more involved probabilistic analysis.

Our work also has close connections with the study of stochastic algorithms for low-rank matrix *approximation*. In this body of work, one is interested in sampling some entries of a matrix in order to construct an approximate factorization of this matrix. Typically, it is assumed that one may sample any subset of entries but would like to minimize the computational complexity involved in constructing an approximation. Pioneering work in this area appears in Frieze et al.¹⁵ and Liberty et al.,²⁵ and an extensive survey of these methods can be found in Halko et al.¹⁹ While this body of work also uses similar foundational theory of random matrices, our modeling assumptions are fundamentally different. Here, we are primarily concerned with the scenario where we have very limited control over which entries of the matrix we can observe. In the examples described in the introduction, one does not have access to all of the entries of the matrix due to systemic constraints. Surprisingly, our results demonstrate that low-rank matrices can be recovered *exactly* from almost all sufficiently large subsets of entries. However, when we have the ability to sample entries at will, the algorithmic recovery schemes become considerably more efficient. We see our results as complementary extremes of the sort of access one may have to the entries of a matrix.

Indeed, when the sampling can be chosen in specially designed patterns, the exact recovery problem becomes dramatically simpler. For example, suppose that \mathbf{M} is generic and that we precisely observe every entry in the first r rows and columns of the matrix.

Write \mathbf{M} in block form as

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{bmatrix},$$

with \mathbf{M}_{11} an $r \times r$ matrix. In the special case that \mathbf{M}_{11} is invertible and \mathbf{M} has rank r , it is easy to verify that $\mathbf{M}_{22} = \mathbf{M}_{21} \mathbf{M}_{11}^{-1} \mathbf{M}_{12}$. One can prove this identity by forming the SVD of \mathbf{M} . That is, if \mathbf{M} is generic, the upper $r \times r$ block is invertible, and we observe *every* entry in the first r rows and columns, we can recover \mathbf{M} . This result immediately generalizes to the case where one observes r rows and r columns and the $r \times r$ matrix at the intersection of the observed rows and columns is invertible. Algorithms in the stochastic low-rank matrix approximation literature are essentially no more complicated than this simple algorithm. They use randomness to add numerical robustness and to guarantee that the sampled entries span the row/column space of the matrix to be acquired.

3. RECENT ADVANCES IN LOW-RANK MODELING

Our original article announced the possibility of various refinements and extensions, and invited researchers to develop the new field of matrix completion. We are pleased to see that the area of low-rank modeling and matrix completion has been quite active, and the field is growing at a very fast pace. In fact, there are so many new and exciting results recently developed that it is unfortunately impossible to review them all here. Below, we

survey selected progress that has occurred since our original submission.

3.1. Improvements and other approaches

The results discussed in Section 2.1 show that under suitable conditions, one can reconstruct an $n \times n$ matrix of rank r from a small number, m , of its sampled entries provided that m is on the order of $n^{1.2} r \log n$, at least for moderate values of the rank. One would like to know whether better results hold, in the sense that exact matrix recovery would be guaranteed with a reduced number of measurements. In particular, recall that an $n \times n$ matrix of rank r depends on $(2n - r)r$ degrees of freedom; is it possible to recover most low-rank matrices from on the order of nr randomly selected entries? Can the sample size be merely proportional to the true complexity of the low-rank object we wish to recover?

In this direction, we would like to emphasize that there is nothing in the approach of our original paper that stands in the way of stronger results. Our proof architecture requires bounding an infinite matrix series in the operator norm. We develop a bound on the spectral norm of each of the first four terms of this series and a general argument to bound the remainder of the series in the full version of this paper. Presumably, one could bound higher order terms by the same techniques. Getting an appropriate bound on the fifth term would lower the exponent of n from $6/5$ to $7/6$. The appropriate bound on the sixth term would further lower the exponent to $8/7$, and so on. To obtain an optimal result, one would need to bound $O(\log n)$ terms.

Following this main idea, the authors Candès and Tao⁹ reduced the upper bound on the number of required measurements to $O(nr \log^6(n))$ using a combinatorial argument to bound precisely this particular series. Their results depend on some additional assumptions, including a “strong incoherence condition” that is more restrictive than the one defined in Section 2.1. However, they also show that *no algorithm* could succeed with high probability if less than $\Theta(nr \log(n))$ entries were observed.

An unexpected and clever method for approximating this infinite matrix series was invented in Gross et al.¹⁸ This new approach used powerful large deviation bounds from quantum information theory combined with an iterative construction that circumvented much of the combinatorics necessary for the proof in Candès and Tao.⁹ This approach is dramatically simpler than previous approaches, and, using this technique, it was shown that $O(nr \log^2(n))$ entries were sufficient for exact matrix completion in Gross¹⁷ and Recht.²⁹ In Recht,²⁹ the leading constant was even upper bounded by 64.

From a very different perspective, the authors in Keshavan et al.²¹ provided a non-convex algorithm for low-rank matrix recovery. Here, the authors analyze a gradient descent scheme over the Grassmannian manifold of subspaces. Using some of the techniques developed in the full version of this paper, the authors show that this nonconvex problem is actually convex in a neighborhood of the true low-rank matrix provided the number of observed entries exceeds

$O(n \log(n))$ and the rank is less than $\log(n)$. This provides the asymptotically tightest bound on the number of entries required for recovery, but the authors need to assume that the singular values of the unknown low-rank matrix are all of order unity and that the rank is less than $\log(n)$ for their results to hold.

3.2. Toward a more general theory

More general measurement models. In our original work, we anticipated in Section 1.3 that our results would extend to the case where one observes a small number of arbitrary linear functionals of a hidden matrix \mathbf{M} . Set $N = n^2$ and let $\mathbf{A}_1, \dots, \mathbf{A}_N$ be an orthonormal basis for the linear space of $n \times n$ matrices with the usual inner product $\langle \mathbf{X}, \mathbf{Y} \rangle = \text{trace}(\mathbf{X}^T \mathbf{Y})$. Then, we predicted that our results should also apply to the rank minimization problem

$$\text{minimize rank}(\mathbf{X}) \quad \text{subject to } \langle \mathbf{A}_k, \mathbf{X} \rangle = \langle \mathbf{A}_k, \mathbf{M} \rangle \quad k \in \Omega, \quad (3.1)$$

where $\Omega \subset \{1, \dots, N\}$ is selected uniformly at random. In fact, (3.1) is (2.2) when the orthobasis is the canonical basis $(\mathbf{e}_i \mathbf{e}_j^T)_{1 \leq i, j \leq n}$. We conjectured that those low-rank matrices that have small inner product with all the basis elements \mathbf{A}_k may be recoverable by nuclear norm minimization.

This conjecture was proven to be true by Gross,¹⁷ where a general definition of coherence was provided, and it was shown that the same number of measurements sufficed for reconstruction under this modified definition. Additionally, in Gross et al.,¹⁸ it was shown that the *Pauli basis*, studied in quantum information theory, was incoherent with *any* basis. This fact follows because all of the matrices in the Pauli basis are mutually orthogonal and unitary. Hence, the Pauli basis is a deterministic collection of matrices such that a random subset of these matrices can be used to reconstruct any low-rank matrix. This result has been applied to propose new methods in quantum-state tomography where one aims to determine the state of some quantum mechanical system with as few measurements or experiments as possible.

Matrix completion with noise. All of the results described above concern the problem of *exact* matrix completion, where we have perfect information about the entries of the matrix to be reconstructed. Of course, in almost all real-world problems, we can only gain access to noisy samples of the entries of the matrices we would like to recover. Fortunately, many authors have investigated the stability of matrix completion when the observations are noisy. The first such result⁶ uses a stability argument based on convexity to guarantee accurate recovery from noisy data. Several subsequent works studied this problem under different matrix models. The work in Keshavan et al.²² gives near-optimal bounds provided the unknown matrix obeys additional assumptions which say that the singular values are all about the same size. In Negahban and Wainwright,²⁸ error bounds are derived provided the

matrix is not spiky; that is to say, assuming that all the entries have about the same magnitude. We additionally invite interested readers to peruse Koltchinskii et al.,²³ which very recently introduced powerful results with yet a slightly different matrix model.

Algorithmic innovations. While it was known that the nuclear norm problem could be efficiently solved by semidefinite programming, the results of Recht et al.³⁰ and the full version of this paper have inspired the development of many special purpose algorithms to rapidly minimize the nuclear norm. For example, in Cai et al.⁴ and Ma et al.,²⁷ the authors show that many of the fast first order methods developed for compressed sensing problems can be adapted to solve large-scale matrix completion problems. These algorithms are projected gradient algorithms which operate by alternately correcting the predictions on the observed entries and soft-thresholding the singular values of the iterate. Using accelerated gradient schemes, the authors Ji and Ye²⁰ and Toh and Yun³⁴ have improved these algorithms to provide very fast implementations of nuclear norm minimization. These codes enable the solution of problems with hundreds of thousands of rows and columns in a few hours on a standard workstation. Moreover, the numerical experiments in these works confirm that nuclear minimization can successfully recover very large low-rank matrices with on the order of 3 to 5 times the number of latent degrees of freedom. Additional experiments demonstrate that low-rank matrices can be robustly recovered under significant additive noise using nuclear-norm minimization.

3.3. From sparsity to rank and beyond

In concert with Recht et al.,³⁰ our work on matrix completion crystallized some of the foundational ideas of compressed sensing. We were able to extend the notion of sparsity to the much more general concept of matrix rank, and situate the main ideas of compressed sensing in a dramatically broader context.

One exciting new development since the appearance of our original paper shows that the notions of sparsity and rank are in some sense orthogonal. If a matrix can be written as a sum of a low-rank matrix and a sparse matrix, then these two matrices can be identified and deconvolved from their sum. Deterministic conditions required for such an algorithm to work were provided in Chandrasekaran et al.¹² A randomized analysis in Candes et al.⁵ provided sharper recovery guarantees and furnished a new method for Robust Principal Components Analysis, demonstrating that principal components could be constructed even in the presence of a large number of outliers. Moreover, the results of Chandrasekaran et al.¹² were extended to provide convex algorithms for identifying Gaussian graphical models in Chandrasekaran et al.¹⁰ Prior art in this area had resorted to nonconvex heuristics based on Expectation–Maximization with no provable guarantees. It is quite surprising that, under very modest assumptions, a convex algorithm can solve a hidden variable estimation problem in multivariate statistics.

This subsequent research has shown that there is much more work to be done in this area. Work in compressed sensing, matrix completion, and their generalizations have shown that convex optimization can be used to solve a myriad of hard identification problems at nearly optimal rates. But the picture is likely much broader than what we currently understand. There are likely notions of simplicity beyond rank and sparsity that can also be leveraged in high-dimensional data analysis to open new frontiers in low-rate sampling. New work in Chandrasekaran et al.¹¹ develops a unified program for recovering simple signals and objects from incomplete information, illustrating a general approach for translating expert domain knowledge into convex optimization algorithms. This work not only generalizes prior art on compressed sensing and matrix completion but also provides several new models where low-rate sampling can recover specially structured models. Such new developments suggest that we have only begun to scratch the surface of the types of models and objects that may be recovered from highly incomplete information. 

References

- Amit, Y., Fink, M., Srebro, N., Ullman, S. Uncovering shared structures in multiclass classification. In *Proceedings of the Twenty-fourth International Conference on Machine Learning* (2007).
- Argyriou, A., Evgeniou, T., Pontil, M. Multi-task feature learning. In *Neural Information Processing Systems*, 2007.
- Beck, C., D'Andrea, R. Computational study and comparisons of LFT reducibility methods. In *Proceedings of the American Control Conference* (1998).
- Cai, J.F., Candès, E.J., Shen, Z. A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.* 20, 4 (2008), 1956–1982.
- Candès, E.J., Li, X., Ma, Y., Wright, J. Robust principal component analysis? Submitted for publication. Preprint available at <http://www-stat.stanford.edu/~candes/papers/RobustPCA.pdf>, 2009.
- Candès, E.J., Plan, Y. Matrix completion with noise. *Proc. IEEE* 98, 6 (2009), 925–936.
- Candès, E.J., Romberg, J. Sparsity and incoherence in compressive sampling. *Inverse Probl.* 23, 3 (2007), 969–985.
- Candès, E.J., Romberg, J., Tao, T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theor.* 52, 2 (2006), 489–509.
- Candès, E.J., Tao, T. The power of convex relaxation: Near-optimal matrix completion. *IEEE Trans. Inform. Theor.* 56, 5 (2009), 2053–2080.
- Chandrasekaran, V., Parrilo, P.A., Willsky, A.S. Latent variable graphical model selection via convex optimization. Submitted for publication. Preprint available at http://sng.mit.edu/~venkatc/cpw_lgm_preprint10.pdf, 2010.
- Chandrasekaran, V., Recht, B., Parrilo, P.A., Willsky, A. The convex geometry of linear inverse problems. Submitted for publication. Preprint available at arxiv.org/1012.0621, 2010.
- Chandrasekaran, V., Sanghavi, S., Parrilo, P.A., Willsky, A.S. Rank-sparsity incoherence for matrix decomposition. *SIAM J. Optim.* 21, 2 (2011), 572–596.
- Donoho, D.L. Compressed sensing. *IEEE Trans. Inform. Theor.* 52, 4 (2006), 1289–1306.
- Fazel, M. *Matrix Rank Minimization with Applications*. PhD thesis, Stanford University, 2002.
- Frieze, A., Kannan, R., Vempala, S. Fast Monte-Carlo algorithms for finding low-rank approximations. *J.ACM* 51, 6 (2004).
- Funk, S. Netflix update: Try this at home, December 2006. <http://sifter.org/~simon/journal/20061211.html>.
- Gross, D. Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans. Inform. Theor.* 57 (2011), 1548–1566.
- Gross, D., Liu, Y.K., Flammia, S.T., Becker, S., Eisert, J. Quantum state tomography via compressed sensing. *Phys. Rev. Lett.* 105, 15 (2010), 150401.
- Halko, N., Martinsson, P.G., Tropp, J.A. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Rev.* 53, 2 (2011), 217–288.
- Ji, S., Ye, J. An accelerated gradient method for trace norm minimization. In *Proceedings of the ICML* (2009).
- Keshavan, R.H., Montanari, A., Oh, S. Matrix completion from a few entries. *IEEE Trans. Inform. Theor.* 56, 6 (2009), 2980–2998.
- Keshavan, R.H., Montanari, A., Oh, S. Matrix completion from noisy entries. *J. Mach. Learn. Res.* 11 (2009), 2057–2078.
- Koltchinskii, V., Tsybakov, A.B., Lounici, K. Nuclear norm penalization and optimal rates for noisy low rank matrix completion. Submitted for publication. Preprint available at arxiv.org/abs/1011.6256, 2010.
- Koren, Y., Bell, R., Volinsky, C. Matrix factorization techniques for recommender systems. *Computer* 42, 8 (2009), 30–37.
- Liberty, E., Woolfe, F., Martinsson, P.G., Rokhlin, V., Tygert, M. Randomized algorithms for the low-rank approximation of matrices. *Proc. Nat. Acad. Sci.* 104, 51 (2007), 20167–20172.
- Linial, N., London, E., Rabinovich, Y. The geometry of graphs and some of its algorithmic applications. *Combinatorica* 15 (1995), 215–245.
- Ma, S., Goldfarb, D., Chen, L. Fixed point and Bregman iterative methods for matrix rank minimization. *Math. Program.* 128, 1 (2011), 321–353.
- Negahban, S., Wainwright, M.J. Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. Submitted for publication. Preprint available at arxiv.org/abs/1009.2118, 2010.
- Recht, B. A simpler approach to matrix completion. *J. Mach. Learn. Res.* (2009). To Appear 2012. Preprint available at <http://arxiv.org/abs/0910.0651>.
- Recht, B., Fazel, M., Parrilo, P. Guaranteed minimum rank solutions of matrix equations via nuclear norm minimization. *SIAM Rev.* 52, 3 (2010), 471–501.
- Rennie, J.D.M., Srebro, N. Fast maximum margin matrix factorization for collaborative prediction. In *Proceedings of the International Conference of Machine Learning* (2005).
- So, A.M.C., Ye, Y. Theory of semidefinite programming for sensor network localization. *Math. Program., Ser. B* 109 (2007), 367–384.
- Sturm, J.F. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optim. Meth. Software* 11–12 (1999), 625–653.
- Toh, K.-C., Yun, S. An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems. *Pac. J. Math.* 6 (2010), 615–640.

Emmanuel Candès (candes@stanford.edu), Departments of Mathematics and Statistics, Stanford University, Stanford, CA.

Candès was partially supported by a National Science Foundation grant CCF-515362, by the 2006 Waterman Award (NSF) and by an ONR grant.

Benjamin Recht (brecht@cs.wisc.edu) Department of Computer Sciences, University of Wisconsin-Madison, WI.

This work was performed while Recht was a post-doc at the Center for the Mathematics of Information at the California Institute of Technology. He was partially funded by ONR grant N00014-08-1-0749.