Low-rank Matrix Recovery from Errors and Erasures

Yudong Chen, Ali Jalali, Sujay Sanghavi and Constantine Caramanis
Department of Electrical and Computer Engineering
The University of Texas at Austin, Austin, TX 78712 USA
ydchen@utexas.edu, (alij, sanghavi and caramanis)@mail.utexas.edu

***Abstract*—This paper considers the recovery of a low-rank matrix from an observed version that simultaneously contains both *(a) erasures:* most entries are not observed, and *(b) errors:* values at a constant fraction of (unknown) locations are arbitrarily corrupted. We provide a new unified performance guarantee on when a (natural) recently proposed method, based on convex optimization, succeeds in exact recovery. Our result allows for the simultaneous presence of random and deterministic components in both the error and erasure patterns. On the one hand, corollaries obtained by specializing this one single result in different ways recovers (upto poly-log factors) all the existing works in matrix completion, and sparse and low-rank matrix recovery. On the other hand, our results also provide the *first guarantees* for (a) deterministic matrix completion, and (b) recovery when we observe a vanishing fraction of entries of a corrupted matrix.**

I. Introduction

Low-rank matrices play a central role in large-scale data analysis and dimensionality reduction. They arise in a variety of application areas, among them Principal Component Analysis (PCA) [1], [2], [3], Multi-dimensional scaling (MDS), Correlation Clustering [4], [5], Spectral Clustering and related methods [6], [7], ranking and collaborative filtering [8], [9], etc. In all these problems, low-rank structure is used to either approximate a general matrix, or to correct for corrupted or missing data.

This paper considers the recovery of a low-rank matrix in the simultaneous presence of *(a) erasures:* most elements are not observed, and *(b): errors:* among the ones that are observed, a significant fraction at unknown locations are grossly/maliciously corrupted. It is now well recognized that the standard, popular approach to low-rank matrix recovery using SVD as a first step fails spectacularly in this setting [10]. Low-rank matrix completion, which considers only random erasures ([11], [12]) will also fail with even just a few maliciously corrupted entries. In light of this, several recent works have studied an alternate approach based on a (now natural) convex optimization problem. One approach [13], [14] provides deterministic/worst case guarantees for the fully observed setting (i.e. only errors). Another avenue [15], [1] provides probabilistic guarantees for the case when the supports of the error and erasure patterns are chosen uniformly at random. Our work provides (often order-wise) stronger guarantees on the performance of this convex formulation, as compared to all of these papers.

This work is supported by NSF Grants EFRI-0735905, CNS-0831580, 0954059, 1017525 and DTRA grant HDTRA1-08-0029.

We present one main result, and two other theorems. Our main result, Theorem 1, is a *unified performance guarantee* that allows for the simultaneous presence of both errors and erasures, and deterministic and random support patterns for each. In order/scaling terms, this single result recovers as corollaries all the existing results on low-rank matrix completion [11], [12], worst-case error patterns [13], and random error and erasure patterns [15], [1]; we provide detailed comparisons in Section II. More significantly, our result goes *beyond* the existing literature by providing the first guarantees for random support patterns for the case when the fraction of entries observed vanishes as n (the size of the problem) grows – an important regime in many applications, including collaborative filtering. In particular, we show that exact recovery is possible with as few as $\Theta(n \log^4 n)$ entries, even when a constant fraction of these entries are errors.

Theorem 2 is also a unified guarantee, but with the additional assumption that the *signs* of the error matrix are equally likely to be positive or negative. We are now able to show that it is possible to recover the low-rank matrix even when *almost all* entries are corrupted. Again, our results go beyond the existing work [15] on this case, because we allow for a vanishing fraction of observations.

Theorem 3 concentrates on the deterministic/worst-case analysis, providing the first guarantees when there are both errors and erasures. Its specialization to the erasures-only case provides the first deterministic guarantees for low-rank matrix completion (where existing work [11], [12] has concentrated on randomly located observations). Specialization to the errors-only case provides an order improvement over the previous deterministic results in [13], and matches the scaling of [14] but with a much simpler proof.

Besides improving on known guarantees, all our results involve several technical innovations beyond existing proofs. Several of these innovations may be of interest in their own right, for other related high-dimensional problems.

978-1-4577-0594-6/11/$26.00 ©2011 IEEE

II. Main Contributions

A. Setup

The problem: Suppose matrix $C \in \mathbb{R}^{n_1 \times n_2}$ is the sum of an underlying low-rank matrix $B^* \in \mathbb{R}^{n_1 \times n_2}$ and a sparse "errors" matrix $A^* \in \mathbb{R}^{n_1 \times n_2}$. Neither the number, locations or values of A^* are known a-priori; indeed by "sparse" we just mean that it A^* has at least a constant fraction of its entries being 0 – it is allowed to have a significant fraction of its entries being non-0 as well. We consider the following problem: suppose we only observe a subset $\Omega_{\text{obs}} \subset [n_1] \times [n_2]$ of the entries of C; the remaining entries are erased. When and how can we exactly recover B^* ? (and, by simple implication, the entries of A^* that are in Ω_{obs})

The Algorithm: In this paper we are interested in the performance of the following convex program

$$(\hat{A}, \hat{B}) = \arg\min_{A,B} \quad \gamma\|A\|_1 + \|B\|_* \quad \text{s.t.} \quad \mathcal{P}_{\Omega_{\text{obs}}}(A+B) = \mathcal{P}_{\Omega_{\text{obs}}}(C), \tag{1}$$

where the notation is that for any matrix M, $\|M\|_* = \sum_i \sigma_i(M)$ is the nuclear norm, defined to be the sum of the singular values of the matrix, $\|M\|_1 = \sum_{i,j} |a_{ij}|$ is the elementwise ℓ_1 norm, and $\mathcal{P}_{\Omega_{\text{obs}}}(M)$ is the set of elements of M that are in the observed set Ω_{obs}. Intuitively, the nuclear norm acts as a convex surrogate for the rank of a matrix, and ℓ_1 norm as a convex surrogate for its sparsity.

Incoherence: We are interested in characterizing when the optimum of (1) recovers the underlying (observed) truth, i.e., when $(\mathcal{P}_{\Omega_{\text{obs}}}(\hat{A}), \hat{B}) = (\mathcal{P}_{\Omega_{\text{obs}}}(A^*), B^*)$. Clearly, not all low-rank matrices B^* can be recovered exactly; in particular, if B^* is both low-rank *and* sparse, it would be impossible to unambiguously identify it from an added sparse matrix. To prevent such a scenario, we follow the approach taken in all the recent work [13], [1], [11], [12], [16] and define *incoherence* parameters for B^*. Suppose the matrix B^* with rank $r \leq \min(n_1, n_2)$ has the singular value decomposition $U\Sigma V^*$, where $U \in \mathbb{R}^{n_1 \times r}$, $V \in \mathbb{R}^{n_2 \times r}$ and $\Sigma \in \mathbb{R}^{r \times r}$. We say a given matrix B^* is (r, μ)**-incoherent** for some $r \in \{1, \cdots, \min(n_1, n_2)\}$ and $\mu \in \left[1, \frac{\max(n_1,n_2)}{r}\right]$ iff (i) $rank(B^*) = r$, and, (ii)

$$\max_i \|U^* \mathbf{e}_i\| \leq \sqrt{\frac{\mu r}{n_1}} \qquad \max_i \|V^* \mathbf{e}_i\| \leq \sqrt{\frac{\mu r}{n_2}}$$
$$\|UV^*\|_\infty \leq \sqrt{\frac{\mu r}{n_1 n_2}},$$

where, $\mathbf{e}_i$'s are standard basis vectors with proper length. Here $\|\cdot\|$ represents the 2-norm of the vector.

B. Unified Guarantee

Our first main result is a unified guarantee that allows for the simultaneous presence of random and adversarial patterns, for both errors and erasures. As mentioned in the introduction, this recovers all existing results in matrix completion, and sparse and low-rank matrix decomposition, up to constants or $\log$ factors. We now define three bound quantities: p_0, τ and d.

Let $\Omega_{\text{obs, d}}$ be any (i.e. deterministic) set of observed entries, and additionally let $\Omega_{\text{obs, r}}$ be a randomly chosen set such that each element is in $\Omega_{\text{obs, r}}$ with probability *at least* p_0. Thus, the overall set of observed entries is $\Omega_{\text{obs}} = \Omega_{\text{obs, r}} \cap \Omega_{\text{obs, d}}$. Let $\Omega = \Omega_{\text{r}} \cup \Omega_{\text{d}}$ be the support of A^*, again composed of the union of a deterministic component Ω_{d}, and a random component Ω_{r} generated by having each element be in Ω_{r} independently with probability *at most* τ. Finally, consider the union $\Omega^c_{\text{obs, d}} \cup \Omega_{\text{d}}$ of all deterministic errors and erasures, and let d be an upper bound on the maximum number of elements this set has in any row, or in any column.

Theorem 1. *Set $n = \min\{n_1, n_2\}$. There exist constants c, ρ_r, ρ_s and ρ_d – each independent of n – such that, with probability greater than $1 - cn^{-10}$, the unique optimal solution of* (1) *with $\gamma = \frac{1}{32\sqrt{p_0 n(d+1)}}$ is equal to $(\mathcal{P}_{\Omega_{obs}}(A^*),\ B^*)$ provided that*

$$\begin{aligned} p_0 &\geq \rho_r \max\left\{\frac{\mu r \log^2 n}{n}, \frac{\mu r \tau}{n}\log^3 n\right\} \\ \tau &\leq \rho_s \\ d &\leq \rho_d \frac{n p_0^2}{\mu r \log^2 n} \end{aligned}$$

Remark. *The conclusion of the theorem holds for a range of values of γ. We have chosen one of these valid values. Moreover, in the theorem one can replace p_0 (d) with any lower-bound (upper-bound) of it.*

Remark. *It is possible to remove some of the logarithmic factors in the above bounds. We choose not to dwell on such refinements due to space limit.*

C. Improved Guarantee for Errors with Random Sign

If we further assume the errors in the entries in Ω_{r} have random signs, then one can recover from a overwhelming fraction of corruptions.

Theorem 2. *Under the same setup of Theorem 1, further assume that the signs of entries of A^* in Ω_r are symmetric Bernoulli random variables. Then there exist constants ρ_r and ρ_d such that, with probability at least $1 - cn^{-10}$, the unique optimal solution of* (1) *with $\gamma = \frac{1}{32\sqrt{p_0 n(d+1)}}$ is equal to $(\mathcal{P}_{\Omega_{obs}}(A^*),\ B^*)$ provided that*

$$\begin{aligned} p_0(1-\tau)^2 &\geq \rho_r \max\left\{\frac{\mu r \log^2 n}{n}, \frac{\mu r \tau}{n}\log^3 n\right\} \\ d &\leq \rho_d \frac{n p_0^2 (1-\tau)^2}{\mu r \log^2 n} \end{aligned}$$

Remark. *We would like to point out that in the theorem, p_0 can approach zero faster than $1 - \tau$; this agrees with the intuition that correcting erasures with known locations is easier than correcting errors with unknown locations.*

Remark. *Again, it is possible to remove some of the logarithmic factors in the above bounds.*

Comparison with previous work We compare Theorem 1 and 2 with existing results. [1] provides guarantees for random errors and erasures. Their guarantees require that $\tau = O(1)$ and $p_0 = \Omega(1)$. This is later extended in [15], which shows that τ can be a constant arbitrarily close to 1 provided A^* has random signs and n is sufficiently large. Here Theorem 1 and 2 provide stronger results, as both allow p_0 to be vanishingly small, in particular, $\Theta\left(\frac{\log^3 n}{n}\right)$ (the proof techniques used in [1] and [15] do not seem to be able to cover this case, even when the rank is small). Moreover, when A^* has random signs, Theorem 2 gives explicit scaling between τ and n as $\tau = O\left(1 - \sqrt{\frac{\log^3 n}{n}}\right)$, with γ independent of the usually unknown quantity τ. In contrast, [15] requires $\tau \leq f(n)$ for some unspecified function $f(\cdot)$, and uses a τ-dependent γ.

Deterministic errors are first studied in [13], [17], which stipulate $d = O\left(\sqrt{\frac{n}{\mu r}}\right)$. Theorem 1 and 2 improve this bound to $d = O\left(\frac{n}{\mu r \log^2 n}\right)$. In the next sub-section, we provide a more refined analysis for the deterministic case, which gives $d = O\left(\frac{n}{\mu r}\right)$. As this manuscript was being prepared, we learned of an independent investigation of the deterministic case [14], which gives similar guarantees. Our results also handle the case of partial observations, which is not discussed in previous works [13], [17], [14].

Previous work in low-rank matrix completion deals with the case when there are no errors or deterministic erasure (i.e., $d, \tau = 0$). For this problem, Theorem 1 recovers the best existing bound $p_0 = O\left(\frac{\mu r \log^2 n}{n}\right)$ in [12], [16], [18]. Our theorem also provides the first guarantee for deterministic matrix completion under potentially adversarial erasures.

One prominent feature of our guarantees is that we allows adversarial and random erasures/errors to exist *simultaneously*. To the best of our knowledge, this is the first such result in low-rank matrix recovery/robust PCA.

D. Improved Deterministic Guarantee

Our second main result deals with the case where the errors and erasures are arbitrary but fixed. As discussed in [13], for the exact recovery, the error matrix A^* needs to be not only sparse but also "spread out", i.e. to not have any row or column with too many non-zero entries. The same holds for unobserved entries. Correspondingly, we require the following: (i) there are at most d errors and erasures on each row/column, and, (ii) $\|A^*\| \leq \eta d \|A^*\|_\infty$; where $\|A^*\| = \sigma_{\max}(A^*)$ is the operator norm of the matrix and is defined to be the largest singular value of the matrix and $\|A^*\|_\infty = \max_{i,j} |a^*_{ij}|$ is the element-wise maximum magnitude of the elements of the matrix.

Theorem 3. *For $\gamma \in \left[\frac{1}{1-2\alpha}\sqrt{\frac{\mu r}{n_1 n_2}}, \frac{1-\alpha}{\eta d} - \sqrt{\frac{\mu r}{n_1 n_2}}\right]$, suppose*

$$\sqrt{\frac{\mu r d}{\min(n_1,n_2)}}\left(1+\sqrt{\frac{\min(n_1,n_2)}{\max(n_1,n_2)}}+\eta\sqrt{\frac{d}{\max(n_1,n_2)}}\right)\leq\frac{1}{2},$$

where, $\alpha = \sqrt{\frac{\mu r d}{n_1}} + \sqrt{\frac{\mu r d}{n_2}}$. Then, the solution to the problem (1) is unique and equal to $(\mathcal{P}_{\Omega_{obs}}(A^*), B^*)$.

Remark. *Notice that we have $\sqrt{d}$ in the bound while [13] has d in their bound. This improvement is achieved by a different construction of dual variable presented in this paper.*

Remark. *If $\eta d\sqrt{\frac{\mu r}{\min(n_1,n_2)}} \leq \frac{1}{6}$ (the condition provided for exact recovery in [13]) is satisfied then the condition of Theorem 3 is satisfied as well. This shows that our result is an improvement to the result in [13] in the sense that this result guarantees the recovery of a larger set of matrices A^* and B^*. Moreover, this bound implies that n (for square matrices) should scale with dr which is another improvement compare to the $d^2 r$ scaling in [13].*

Remark. *This theorem provides the same scaling result for d in terms of r and n as the result of [14]. However, our assumptions are closer to existing ones in matrix completion and sparse and low-rank decomposition papers [11], [12], [1], [13].*

III. Proof of Theorem 1 and 2

In a high level, the proof is along the same lines of those in the low-rank matrix recovery literature [1], [11], [16] and is consist of two steps: (a) write down the first-order sufficient condition for $(\mathcal{P}_{\Omega_{\text{obs}}}(A^*), B^*)$ to be the unique solution to (1), which involves the existence of a dual certificate, (b) construct the dual certificate and show that it has the desired property.

To state the optimality condition in step (a), we need to introduce some notations. Define the support of A^* as $\Omega = \{(i,j) : A^*_{i,j} \neq 0\}$. The orthogonal projection of a matrix $M \in \mathbb{R}^{n_1 \times n_2}$ to Ω is the matrix whose $(i,j)^{th}$ entry is given by

$$\left(\mathcal{P}_\Omega(M)\right)_{i,j} = \begin{cases} M_{i,j} & (i,j) \in \Omega \\ 0 & \text{otherwise.} \end{cases}$$

Let $\Gamma = \Omega_{\text{obs}}/\Omega$ be the set of observed *and* clean entries; then Γ^c is the set of corrupted *or* unobserved entries. $\mathcal{P}_\Gamma$ and $\mathcal{P}_{\Gamma^c}$ are defined accordingly. Set $E^* := \mathcal{P}_{\Omega_{\text{obs}}}(\text{sgn}(A^*))$, where $\text{sgn}(\cdot)$ is the element-wise signum function. We also define a sub-space $\mathcal{T}$ of the span of all matrices that share either the same column space or the same row space as B^*:

$$\mathcal{T} = \left\{UX^\top + YV^\top : X \in \mathbb{R}^{n_2 \times r}, Y \in \mathbb{R}^{n_1 \times r}\right\}.$$

For any matrix $M \in \mathbb{R}^{n_1 \times n_2}$, we can define its *orthogonal projection* to the space $\mathcal{T}$ as follows:

$$\mathcal{P}_\mathcal{T}(M) = UU^\top M + MVV^\top - UU^\top MVV^\top.$$

We also define the projections onto $\mathcal{T}^\perp$, the complement orthogonal space of $\mathcal{T}$, as follows:

$$\mathcal{P}_{\mathcal{T}^\perp}(M) \;=\; M - \mathcal{P}_{\mathcal{T}}(M).$$

In the sequel, by *with high probability* we mean with probability at least $1 - c\min\{n_1, n_2\}^{-10}$. The optimality condition is given in the following lemma.

Lemma 1. *Assume that* $\|\mathcal{P}_{\mathcal{T}}\mathcal{P}_{\Gamma^\perp}M\|_F \leq n\,\|\mathcal{P}_{\mathcal{T}^\perp}\mathcal{P}_{\Gamma^\perp}M\|_F$ *for any matrix. Suppose* $\gamma < 1$. *Then* $(\mathcal{P}_{\Omega_{obs}}(A^*),\ B^*)$ *is the unique solution if there is a dual certificate* Q *obeying*

$$Q = UV^\top + W + \mathcal{P}_{\mathcal{T}}D = \gamma(E^* + F) \in \Omega_{obs},$$

with $W \in T^\perp$, $\|W\| < 1/2$, $F \in \Gamma$ *and,* $\|F\|_\infty < 1/2$, *and* $\|\mathcal{P}_{\mathcal{T}}D\|_F \leq \frac{\gamma}{\sqrt{\min\{n_1,n_2\}}}$.

The second step, constructing the desired Q, is the "art" of the proof: a new dual certificate and a new proof of it are needed to handle dense erasures/errors and the co-existence of deterministic and random components. We will do so using a variation of the so-called Golfing Scheme [16], [18]; details can be found in the full version of our paper [19].

IV. Proof of Theorem 3

The proof follows along the lines of that in [13] and has three steps: *(a)* writing down a sufficient optimality condition, stated in terms of a dual certificate, for $(\mathcal{P}_{\Omega_{\text{obs}}}(A^*),\ B^*)$ to be the optimum of the convex program (1), *(b)* constructing a particular candidate dual certificate, and, *(c)* showing that under the imposed conditions this candidate does indeed certify that $(\mathcal{P}_{\Omega_{\text{obs}}}(A^*),\ B^*)$ is the optimum. Part *(b)* is the "art" in this method; different ways to devise dual certificates can yield different sufficient conditions for exact recovery. Indeed this is the main difference between this paper and [13]. The details of the proof can be found in the full version of our paper [19].

For the sake of completeness, we restate here a first-order sufficient condition for $(\mathcal{P}_{\Omega_{\text{obs}}}(A^*),\ B^*)$ to be the optimum of (1). The proof is almost identical to that in [13]; only minor change is needed to handle erasures.

Lemma 2 (A Sufficient Optimality Condition [13]**).** *The pair* $(\mathcal{P}_{\Omega_{obs}}(A^*),\ B^*)$ *is the unique optimal solution of* (1) *if*

(a) $\Gamma^c \cap \mathcal{T} = \{\mathbf{0}\}$.

(b) *There exists a dual matrix* $Q \in \mathbb{R}^{n_1 \times n_2}$ *satisfying* $\mathcal{P}_{\Omega^c_{obs}}(Q) = 0$ *and*

$$\begin{aligned} &\mathcal{P}_{\mathcal{T}}(Q) = UV^* && \|\mathcal{P}_{\mathcal{T}^\perp}(Q)\| < 1 \\ &\mathcal{P}_{\Gamma^c}(Q) = \gamma\mathcal{P}_{\Omega_{obs}}(\operatorname{sgn}(A^*)) && \|\mathcal{P}_{\Gamma}(Q)\|_\infty < \gamma. \end{aligned} \tag{2}$$

V. Experiments

In this section, we illustrate our results via simulation. In particular, we investigate how the algorithm performs as the size of the low-rank matrix gets larger. In other words, we try to see how the requirements for the success of our algorithm changes as the size of the matrix grows. These simulation results show that the conditions get relaxed more and more as n increases, which matches our theoretical results.

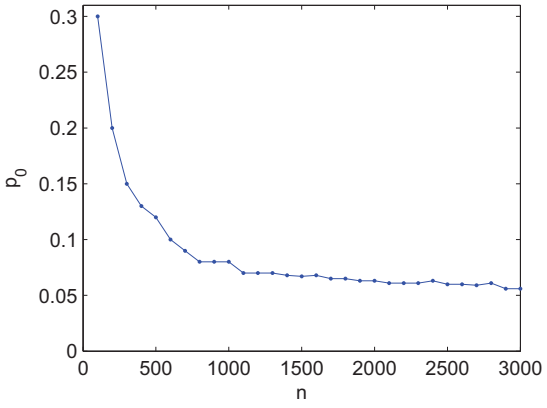

Fig. 1. For a rank two matrix of size n, with probability of corruptions $\tau = 0.1$ and no adversarial noise ($d = 0$), we plot the minimum probability of observation p_0 required for successful recovery of the low-rank matrix as n gets larger.

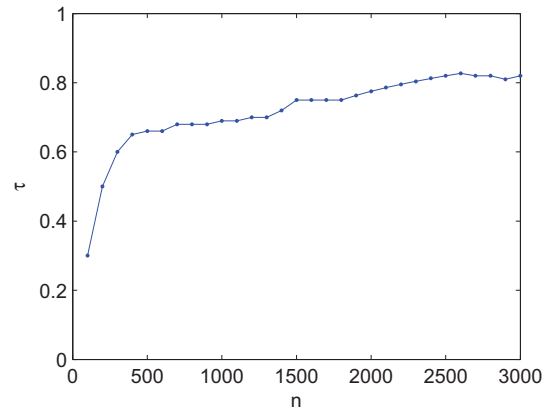

Fig. 2. For a rank two matrix of size n, with probability of observation $p_0 = 0.9$ and no adversarial noise ($d = 0$), we plot the maximum probability of corruptions τ tolerable for successful recovery of the low-rank matrix as n gets larger.

We run three experiments as follows:

(1) **Minimum Required Observing Probability:** We generate a rank two matrix ($r = 2$) of size n by multiplying a random $n \times 2$ matrix and a random $2 \times n$ matrix, and then corrupt the entries randomly with probability $\tau = 0.1$, without any adversarial noise ($d = 0$). The entries of the corrupted matrix are observed independently with probability p_0. We then solve (1) using the method in [20]. Success is declared if we recover the low-rank matrix with a relative error less than 10^{-6}, measured in Frobenius norm. The experiment is repeated 10 times and we count the frequency of success. For any fixed number n, if we start from $p_0 = 1$ and decrease p_0, at some point, the probability of success jumps from one to zero, i.e., we observe a phase transition. In Fig. 1, we plot the p_0 at which the

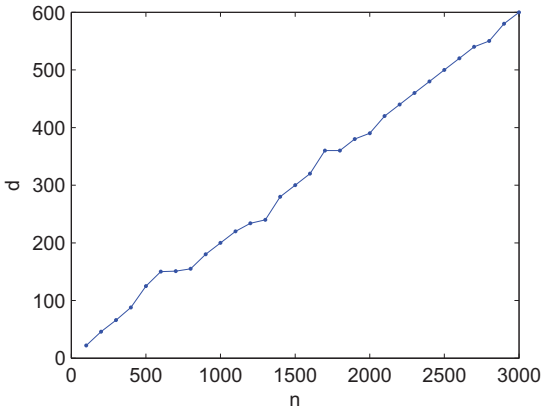

Fig. 3. For a rank two matrix of size n, with probability of observation $p_0 = 0.5$ and probability of corruptions $\tau = 0.1$, and with adversarial/deterministic noise in the form of a $d \times d$ block of 1's lying on the diagonal of the matrix, we plot the maximum size of the adversarial noise d tolerable for successful recovery of the low-rank matrix as n gets larger.

phase transition happens versus the size of the matrix. This experiment shows that the phase transition p_0 goes to zero as n increases, as predicted by our theorem.

(2) **Maximum Tolerable Corruption Probability:** Similarly as before, we generate an instance with $r = 2$, $p_0 = 0.9$ and $d = 0$ for some n. For a fixed n, if we start from $\tau = 0$ and increase τ, at some point, the probability of success jumps from one to zero. Fig. 2 illustrates how the phase transition τ changes as the size of the matrix increases. This experiment shows that higher probability of corruptions can be tolerated as the size of the matrix increases as predicted by the theorem.

(3) **Maximum Tolerable Adversarial/Deterministic Noise:** Similarly as before, we set $r = 2$, $p_0 = 0.5$ and $\tau = 0.1$. We add the adversarial noise in the form of a $d \times d$ block of 1's lying on the diagonal of the original matrix. Notice that this is a hard case since all the adversarial corruptions are burst as oppose to be spread over the matrix. We find the maximum possible d such that the probability of success to goes from 1 to 0 (phase transition). In Fig. 3, we plot this phase transition d versus the size of the matrix and as the deterministic theorem predicts, it grows linearly in n.

ACKNOWLEDGEMENT

The authors thank Dr. Huan Xu for insightful discussions and comments.

REFERENCES

[1] E. J. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Available at http://www-stat.stanford.edu/ candes/papers/RobustPCA.pdf*, 2009.

[2] H. Xu, C. Caramanis, and S. Sanghavi, "Robust PCA via outlier pursuit," in *Advances in Neural Information Processing Systems*, 2010.

[3] ——, "Robust PCA via outlier pursuit," *Arxiv preprint arXiv:1010.4237*, 2010.

[4] A. Jalali, Y. Chen, S. Sanghavi, and H. Xu, "Clustering partially observed graphs via convex optimization," *Arxiv preprint arXiv:1104.4803*, Apr. 2011.

[5] ——, "Clustering partially observed graphs via convex optimization," in *The International Conference on Machine Learning (ICML), to appear*, 2011.

[6] A. Ng, M. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Advances in Neural Information Processing Systems (NIPS)*, 2002.

[7] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

[8] Y. Chen, H. Xu, C. Caramanis, and S. Sanghavi, "Robust matrix completion and corrupted columns," in *The International Conference on Machine Learning (ICML), to appear*, 2011.

[9] ——, "Robust matrix completion with corrupted columns," *Arxiv preprint arXiv:1102.2254*, Feb. 2011.

[10] P. Huber, *Robust Statistics*. Wiley, New York, 1981.

[11] E. J. Candes and B. Recht, "Exact matrix completion via convex optimzation," *Foundation of Computational Math*, vol. 9, pp. 717–772, 2009.

[12] E. J. Candes and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Transaction on Information Theory*, 2009.

[13] V. Chandrasekaran, S. Sanghavi, P. Parrilo, and A. S. Willsky, "Rank-sparsity incoherence for matrix decomposition," *SIAM Journal on Optimization, to appear*, 2010.

[14] D. Hsu, S. Kakade, and T. Zhang, "Robust matrix decomposition with outliers," *Available at arXiv:1011.1518*, 2010.

[15] A. Ganesh, J. Wright, X. Li, E. Candes, and Y. Ma, "Dense error correction for low-rank matrices via principal component pursuit," in *IEEE International Symposium on Information Theory (ISIT)*, 2010.

[16] D. Gross, "Recovering low-rank matrices from few coefficients in any basis," *Available at arXiv:0910.1879v4*, 2009.

[17] V. Chandrasekaran, S. Sanghavi, P. Parrilo, and A. S. Willsky, "Sparse and low-rank matrix decompositions," in *15th IFAC Sypmposium on System Identification (SYSID)*, 2009.

[18] B. Recht, "A Simpler Approach to Matrix Completion," *Arxiv preprint arXiv:0910.0651*, 2009.

[19] Y. Chen, A. Jalali, S. Sanghavi, and C. Caramanis, "Low-rank matrix recovery from errors and erasures," *Arxiv preprint arXiv:1104.0354*, 2011.

[20] Z. Lin, M. Chen, L. Wu, and Y. Ma, "The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices," *UIUC Technical Report UILU-ENG-09-2215*, 2009.