Entry-wise Matrix Completion from Noisy Entries

Zahra Sabetsarvestani*, Franz Kiraly*, Miguel R. D. Rodrigues*
*University College London, U.K.

Abstract—We address the problem of entry-wise low-rank matrix completion in the noisy observation model. We propose a new noise robust estimator where we characterize the bias and variance of the estimator in a finite sample setting. Utilizing this estimator, we provide a new robust local matrix completion algorithm that outperforms other classic methods in reconstructing large rectangular matrices arising in a wide range of applications such as athletic performance prediction and recommender systems. The simulation results on synthetic and real data show that our algorithm outperforms other state-of-the-art and baseline algorithms in matrix completion in reconstructing rectangular matrices.

I. Introduction

Matrix completion is a task that involves the reconstruction of a matrix from a subset of its entries. This problem is relevant in a wide range of applications, including recommender systems and collaborative filtering [1]; performance prediction [2]; dimensionality reduction [3]; network traffic analysis [4]; multi-task learning [5] to name a few. Usually, the matrix is to be completed in the supervised setting of predicting an unseen entry given its row and column, and a set of known entires; for instance predict a movie rating (entry) for a user (row) - movie (column) pair in a stylized recommender setting. However without any constraint on the number of degrees of freedom, the general matrix completion problem is under-determined. Therefore, it is necessary to impose a low-dimensional structure on the matrix, one particularly popular example being a low rank constraint [6], [7].

Our focus is on the matrix completion under the common low-rank assumption in the presence of noise. In particular, we study the reconstruction of large scale rectangular datasets where one of the matrix dimensions is considerably larger than the other - which is in fact by far the most common case in practical data analytics where the set of feature (columns) is fixed while the number of samples (rows) may grow, as opposed to the common setting in matrix completion literature where it is usually implicitly assumed that both number of rows and columns grow asymptotically per some scaling law - which is plausible and frequent in a recommender setting, but in general unusual. As we outline in more detail in section II, many common guarantees in literature rely crucially on this latter asymptotic setting and fall apart - both theoretically and empirically - in the usual rectangular setting. While some of these guarantees are in fact tight, they are only under the further implicit assumption that the task is to reconstruct all missing entries; hence it is not implausible that an improvement is possible for predicting a single or a small subset of missing entries, while other entries may remain impossible to reconstruct.

A. Contribution

Our approach to address this issue is based on the use of single-entry algebraic estimator where one completes unknown entries individually as a function of other known entries [8]. This class of estimators is the only one coming with guarantees for single entries, hence guarantees useful in a supervised or rectangular setting; however, the exact behaviour in the presence of noise has so far been studied only in the limited case of rank one [9]. Our contribution involves:

- Proposing a (noise-)robustified version of the single-entry estimator where we provide the bias and variance of estimation for each entry in a finite sample setting.
- Proposing a new robust local matrix completion algorithm which is shown to outperform the "classical" matrix completion algorithm in the common supervised setting for a rectangular data matrix.
- Providing a set of simulation results using synthetic and real data which proves the superiority of the proposed algorithm in predicting missing entries in a rectangular data matrix.

B. Notation

We adopt the following notation throughout the paper: Matrices are denoted by upper-case bold characters ($\mathbf{A}$), the vectors are denoted by lower-case bold characters ($\mathbf{u}$) and the (i, j) element of a matrix is denoted by A_{ij}. We denote the set of integers $\{1, 2, ...n\}$ by $[n]$. Sets are denoted by calligraphic letter ($\mathcal{C}$) and the cardinality of a set $\mathcal{C}$ is denoted by $|\mathcal{C}|$. We define $\mathbf{A}$ to be a matrix of dimensions $m \times n$ and rank r. We define $\mathcal{E} := [m] \times [n]$. The sub-matrix of $\mathbf{A}$ specified by row indices $\mathcal{I} \subseteq [m]$ and column indices $\mathcal{J} \subseteq [n]$ is $\mathbf{A}_{\mathcal{I}\mathcal{J}}$. Throughout this paper, with no loss of generality we assume that $m > n$. The expectation, the standard deviation and variance of a random variable are denoted by $\mathbb{E}()$, $\mathrm{sd}()$ and $\mathrm{Var}()$, respectively. The operators $\det()$ and $\mathrm{adj}()$ give the determinant and the adjugate of matrix argument.

II. The Matrix Completion Problem

We first introduce the matrix completion problem, review the main existing algorithms and discuss their theoretical and practical limitations.

Define the *masking* function Ω to be a map from matrix $\mathbf{A}$ to a subset of its entries: $\Omega : \mathbf{A} \mapsto (A_{ij})_{(i,j) \in \mathcal{E}_\Omega}$ where $\mathcal{E}_\Omega \subseteq \mathcal{E}$. The matrix completion problem involves reconstruction of the missing entries from the set of observed ones $\Omega(\mathbf{A})$. In most practical settings, we only have access to a set of noisy observations $\Omega(\mathbf{B})$, where $B_{ij} = A_{ij} + N_{ij}$ and N_{ij} denotes the observation noise.

The three main approaches to address the low-rank matrix completion problem are: 1) convex optimization based approaches that involve a convex relaxation of the rank constraint (e.g [6], [10]); 2) spectral methods (e.g. [7], [11]) and more recently 3) algebraic combinatorial approach [8], [9], [12]. These methods along with reconstruction guarantees are described next.

Nuclear Norm Minimization: Convex optimization based methods use nuclear norm as a proxy to enforce low-rank. Reference [6] studied sufficient conditions for the unique reconstruction of $\mathbf{A}$ given a subset of its entries. The results were later refined by [10], [13]. In the absence of noise, it has been proven in [13] that if the locations in $\mathcal{E}_\Omega$ are sampled uniformly at random with

$$|\mathcal{E}_\Omega| \gtrsim \mu_0 r(m+n)\log^2(m), \quad (1)$$

then $\mathbf{A}$ can be uniquely recovered with high probability. Here μ_0 quantifies the *coherence* measure. In the presence of noise, the results of [14] state that under the same condition, approximate recovery is guaranteed. In the case of square matrices where $m = n$, the measurement bound in (1) leads to considerable reduction in the required number of entries; but we note that the sufficient number of observations for perfect reconstruction with nuclear norm minimization is ineffective for extremely rectangular matrices. In particular, in the case where $n \ll m$ the bound in (1) reduces to the requirement that we observe nearly every entry. For example, in extremely rectangular matrices, e.g. where $n \simeq r \log m$, almost all of the entries are required for unique reconstruction.

Spectral Methods: In a parallel line of work, [7], [15] study a model based on thresholding the singular values of the matrix in both noisy and noiseless observation models under the uniform sampling condition. We will refer to their algorithm as OptSpace. In addition to the incoherency condition, [7] requires bounds on the matrix rank and also on the maximum value of entries. It has been proven in [7] that if $\alpha = m/n \geq 1$ for a bounded rank r, roughly, $|\mathcal{E}_\Omega| \geq C(\alpha) rn \max(\log n, r)$ number of observations guarantees the convergence of OptSpace to the true value of the matrix. For the noiseless case, the provided reconstruction error bound depends linearly on $\alpha^{3/2}$ and when the observed entries are noisy the reconstruction error contains an additional term which also depends linearly on α. In comparison to square matrices, rectangular matrices with large α, require a larger fraction of known entries to reconstruct the unknown ones in order to attain a certain level of accuracy.

Algebraic Combinatorial Approach: More recently, [8] presents a novel algebraic combinatorial view on matrix completion which is based on studying the relation between entries by using tools from algebraic geometry and matroid theory. In contrast to convex based and spectral methods that study the unique reconstruction of the whole matrix, algebraic approach studies the necessary and sufficient condition for the unique reconstruction of each entry in a low-rank setting. The results in [8] apply to any fixed set of observations and provide information about every missing entry. Since the

algebraic approach does not require any restrictive condition on sampling distribution or sampling complexity in terms of matrix dimension, the completion of extremely rectangular matrices is potentially feasible by the algebraic approach.

In particular, for reconstruction, a single entry estimator was proposed which accurately reconstructs the missing entry from a few observed entries in the noiseless setting while performing poorly in the presence of noise. The estimation of single entry in the presence of noise has been addressed in [9] only for rank 1.

III. Algebraic Matrix Completion

We now review some concepts in matroid theory and also Local Matrix Completion method which form the basis of our new single entry estimator.

In the algebraic framework the question of reconstructability has been addressed separately from the reconstruction process. It has been proven in [8] that whether or not the missing entry at position (i, j) is uniquely completable from observed ones or more generally, how many completions exist, depends only on the positions of (i, j), the rank of true matrix and the observation mask. Algorithm 1 in [8] decides with probability one whether the missing entry at positions (i, j) is reconstructable or not. The set of finitely completable entries in rank r is called the completable closure of the observed set in rank r and is denoted by $\mathrm{cl}_r(\mathcal{E}_\Omega)$. Furthermore, reference [8] provides an algorithm which reconstructs the reconstructable entries from known ones. The algorithm leverages the notion of a circuit which corresponds to the minimally dependent set of entries compatible with rank r. There is at least one circuit going through each reconstructable entry to which we associate a circuit polynomial that vanishes if and only if the circuit is compatible with rank-r (See Theorem 29 in [8]). We denote the circuit polynomial associated to circuit $\mathcal{C}$ by $\theta_\mathcal{C}$. Let $\mathcal{C} \subseteq \mathcal{E}$ be a circuit which is observed in all positions but in one, e.g $(k, \ell) \in \mathcal{C}$ is not observed but $\mathcal{C} \backslash (k, \ell) \in \mathcal{E}_\Omega$ is observed; then $\theta_\mathcal{C}(A_{k\ell}) = 0$ is linear in $A_{k\ell}$ and leads to a unique solution for $A_{k\ell}$. The simplest example of a circuit in a matrix of rank r is an $(r+1) \times (r+1)$ sub-matrix whose associated polynomial is its determinant which vanishes.

In general, finding the reconstructing circuits for the missing entry (k, ℓ) in a partially observed matrix of rank-r is not feasible in practice; but for the special case where the circuit is a $r+1$-minor, algorithm 4 in [8] recovers all completable entries based on finding $(r+1) \times (r+1)$ sub-matrices. We recall a simple result on the form of circuit polynomials for determinants:

Lemma III.1. Let $\mathcal{C}$ be the circuit consisting of all indices forming a $(r+1) \times (r+1)$ sub-matrix S and let (k, ℓ) be the index to reconstruct. Then,

$$\theta_\mathcal{C}(A) = A_{k\ell} \cdot (\det S(1) - \det S(0)) + \det S(0)$$

where $S(x)$ is the $(r+1) \times (r+1)$ matrix S which has the entry $A_{k\ell}$ replaced by the scalar x. In particular,

$$A_{k\ell} = \det S(0)/(\det S(0) - \det S(1))$$

if the denominator is non-zero.

Proof. This follows from the Laplace expansion theorem, and an elementary computation. □

Definition III.2. Let (k, ℓ) be an index to reconstruct, let $\mathcal{I}$ and $\mathcal{J}$ be an ordered tuple of r row resp. column indices, such that $\mathcal{C} = (k, \mathcal{I}) \times (\ell, \mathcal{J})$ is the reconstruction circuit. Let S be the corresponding sub-matrix of A. We define functions:

$$P[(k, \ell), \mathcal{I}, \mathcal{J}](A) := \det S(0) = A_{k\mathcal{J}} \text{adj}(A_{\mathcal{I}\mathcal{J}}) A_{\mathcal{I}\ell} \quad (2)$$

$$Q[(k, \ell), \mathcal{I}, \mathcal{J}](A) := \det S(1) - \det S(0) = \det(A_{\mathcal{I}\mathcal{J}}) \quad (3)$$

We would like to note that P and Q are functions in entries of the true matrix with indices $\mathcal{C} \setminus (k, \ell)$, with parameters in squared brackets and the argument in the round brackets, into which we could also substitute a noisy version of A. In the absence of noise, one reconstructing circuit accurately reconstructs the missing entry. If the observation is noisy, a new estimator which is robust to noise is required to predict the missing entry.

IV. ALGEBRAIC MATRIX COMPLETION FROM NOISY ENTRIES

We now propose a new single-entry estimator which is robust to noise and applicable in the standard (rectangular) supervised learning setting. The idea is to find different candidate solutions for one missing entry and use them to reconstruct the entry under consideration as accurately as possible.

A. OLS estimator

We suggest the ordinary least square (OLS) estimator as the single entry estimator from noisy observations. In particular, we adopt the following generative observation model to describe the (rectangular) supervised learning setting: We will assume that all observed entries are entries of the noise-corrupted matrix $\mathbf{B} = \mathbf{A} + \mathbf{N}$, where $\mathbf{N} \in \mathbb{R}^{m \times n}$ is a real valued, centred (= expectation zero) random matrix. We will further assume that expectations exist. Our aim is to estimate an unknown entry $A_{k\ell}$ as accurately as possible, given n samples of $(P, Q)[(k, \ell), \mathcal{I}, \mathcal{J}](\mathbf{B})$ denoted by $(p_1, q_1), \dots (p_n, q_n)$ where the ratio $\theta_i = \frac{p_i}{q_i}$, $i \in \{1, \dots, n\}$ represents a noisy estimation of $A_{k\ell}$. We denote the OLS estimation of A_{kl} from n samples of $(P, Q)[(k, \ell), \mathcal{I}, \mathcal{J}](\mathbf{B})$ by $\hat{\theta}_n^{OLS}$ where

$$\hat{\theta}_n^{OLS}(\mathbf{B}) = \min_\theta \sum_{i=1}^n (p_i - \theta q_i)^2. \quad (4)$$

Here θ is a dummy variable associated with (4) which represents the true value for A_{kl}. The values (p_i, q_i) are the noisy observations of (p_i^*, q_i^*) with $q_i = q_i^* + e_{q_i}$, $p_i = p_i^* + e_{p_i}$ where $\theta = \frac{p_i^*}{q_i^*}$. The estimator in (4) estimates θ by minimizing the sum of squared errors $e_i = e_{p_i} - \theta e_{q_i}$. This leads to the following propositions.

Proposition IV.1. Consider the above data generating process where we have access to samples (p_i, q_i), $i \in 1, \dots n$ of $(P, Q)[(k, \ell), \mathcal{I}, \mathcal{J}](\mathbf{B})$, then,

$$\hat{\theta}_n^{OLS}(\mathbf{B}) = \frac{\sum_{i=1}^n q_i p_i}{\sum_{i=1}^n q_i^2}. \quad (5)$$

Proof. This immediately follows from taking the derivative of (4) with respect to θ and setting the derivative to zero. □

Proposition IV.2. Consider the OLS estimator in Proposition IV.1. Then, the followings hold:

i Let $q_i^*, \dots, q_n^*$ be the samples from $Q[(k, \ell), \mathcal{I}, \mathcal{J}](\mathbf{A})$, then

$$\text{Var}(\hat{\theta}_n^{OLS}(\mathbf{B}|\mathbf{A})) = \frac{\sigma_p^2}{n \bar{q}^{*2}}, \quad (6)$$

where $\bar{q}^{*2} = \frac{\sum_{i=1}^n q_i^{*2}}{n}$ and $\sigma_p^2 = \text{Var}(p_i)$.

ii The bias of the OLS estimator is given by:

$$\mathbb{E}\{\hat{\theta}_n^{OLS}(\mathbf{B})\} - \theta = \mathbb{E}\left(\frac{\sum_{i=1}^N q_i e_i}{\sum_{i=1}^N q_i^2} \right) \quad (7)$$

Proof.
- The proof of i follows immediately from reformulating (4) and (5) in terms of q_i^* and p_i and taking the variance of $\hat{\theta}_n(\mathbf{B}|\mathbf{A})$ with respect to the variable $p_i | q_i^*$.
- The proof of ii follows by reformulating the estimator (5) leading to:

$$\hat{\theta}_n^{OLS}(\mathbf{B}) = \frac{\sum_{i=1}^N q_i(\theta q_i + e_i)}{\sum_{i=1}^N q_i^2} = \theta + \frac{\sum_{i=1}^N q_i e_i}{\sum_{i=1}^N q_i^2}.$$

Taking the expectation concludes the claim. □

As shown in Proposition IV.2, the proposed single entry estimator has finite variance which decreases with $\mathcal{O}(n^{-1})$. Note that all the previous works recover the matrix by minimizing the reconstruction error over the whole matrix and provide a bound on the overall MSE [7], [14]. By contrast, our approach reconstructs the missing entry by minimizing the reconstruction error for the entry under consideration and provides the reconstruction accuracy for the individual entries. This is particularly important in applications such as recommender systems, where the reconstruction of only a subset of unknown entries is of interest (instead of reconstructing the whole matrix).

B. Robust Local Matrix Completion

We now suggest to use the OLS estimator - which admits a closed form expression for the bias and variance of the final estimation of the missing entry from $(p_i, q_i), i \in \{1, ..., n\}$- in conjunction with local matrix completion algorithms in order to build a Robust Local Matrix Completion (RLMC) algorithm. See Algorithm 1. In step 3, N possible reconstructing minors for the missing entry (k, l) are obtained. In step 4, (p_i, q_i) corresponding to each circuit is obtained from equations (2) and (3). Finally, in step 5, the missing entry is imputed by $\frac{\sum_{i=1}^n q_i p_i}{\sum_{i=1}^n q_i^2}$. The algorithm for finding reconstructing minors is detailed in algorithm 4 in [8] whose worst case complexity is $\mathcal{O}(|\mathbf{I}|^r |\mathbf{J}|^r)$ where $\mathbf{I}$ and $\mathbf{J}$ are the dimensions of the sub-matrix of partially observed matrix; but this can be done much more efficiently as some rows and columns can be safely removed from the search space, e.g. the rows and columns with less than r observed entries. For extremely rectangular matrices finding multiple minors can be

Algorithm 1 Robust Local Matrix Completion
Input: The set of observed entries $\Omega(\mathbf{A})$, rank r
Output: Estimate for entries in $\mathrm{cl}_r(\mathcal{E}_\Omega)\backslash\mathcal{E}_\Omega$

1: **Repeat**
2: Find an unobserved entry $(k,l) \in \mathrm{cl}_r(\mathcal{E}_\Omega)\backslash\mathcal{E}_\Omega$
3: Find the set $\mathcal{C} = \{\mathcal{C}_1, ..., \mathcal{C}_N\}$ containing (k,l)
4: Compute (p_i, q_i), $i \in \{1, ..., N\}$ corresponding the set $\mathcal{C}$ form (2) and (3)
5: $\hat{A}_{kl} \leftarrow \frac{\sum_{i=1}^{n} q_i p_i}{\sum_{i=1}^{n} q_i^2}$
6: $\mathcal{E}_\Omega \leftarrow \mathcal{E}_\Omega \cup (k,l)$
7: **Until** $\mathcal{E}_\Omega = \mathrm{cl}_r(\mathcal{E}_\Omega)$

---

implemented very efficiently by simply sampling rows and computing the degree of each column in the corresponding bi-partite graph.

V. EXPERIMENTS

We finally evaluate the performance of various algorithms - including the proposed one - on a number of matrix completion problems. We run a set of experiments on real and synthetic data of various dimensions.

Synthetic Data: To compare the performance of algorithms, we run the experiment on both the square and rectangular matrices. We generate rectangular random matrices of rank 2 and 3 and of dimensions 2000×20 by a product of two random matrices $\mathbf{A} = \mathbf{C}\mathbf{F}^T$, where the entries of $\mathbf{C}$ and $\mathbf{F}$ are sampled from a uniform distribution on the interval between 0 and 1. We also generate rank-2 and rank-3 random square matrices of dimension 200×200. We then observe 33% and 40% of the entries in rank-2 and rank-3 matrices uniformly at random, respectively, where the observed entries are contaminated by a centered multiplicative noise as $B_{ij} = A_{ij}(1 + \sigma N_{ij})$ where $\sigma > 0$ and N_{ij} is sampled from a uniform distribution on the interval between -1 and 1. In this set of experiments, σ is varied from 0.05 to 0.5 in the step of 0.05.

Real Data: We present the performance of the matrix completion algorithms on Movielens [1] and athletic performance [2] datasets. Movielens 100K contains 100k ratings (1-5) from 1682 on 943 movies. Movielens 1m is a larger matrix of dimensions (approximately) 6000×4000 where rows represent the users, columns represent the movies and 1m entries are known. Athletic Performance Dataset contains records of 16466 individuals of male athletes, comprising a total of 52350 individual performances over 10 different distances. We note that the Movielens dataset is rather a square matrix while the athletes dataset is an extremely rectangular matrix.

Algorithms: We compare the performance of RLMC, other state-of-the-art and baseline algorithms in matrix completion. For nuclear norm minimization, we use the implementation from softImpute [3] R package. SoftImpute requires λ and an initial guess of the rank as inputs. For spectral methods we

---

implement the OptSpace algorithm in [7] which requires the rank as input. To tune the free-parameters of the algorithms a grid-tuning strategy which performs a nested re-sampling algorithm is used. We set the search space for the rank parameter to $1 - 20$ and for the regularizer parameter, λ to $0 - 10$. For comparison, we implement the so called *Double-Mean* and *Two Factor Model* algorithms as baselines. In Double-Mean the missing entry A_{ij} is imputed by a linear combination of the mean of observed entries in row i, mean of observed entries in column j and the mean of all observed entries. Two Factor Model assumes that the matrix $\mathbf{A}$ can be represented by a linear regression model, $A_{ij} = w_i + z_j$ where w_i and z_i are elements of the vectors $\mathbf{w} \in \mathbb{R}^{m \times 1}$ and $\mathbf{z} \in \mathbb{R}^{n \times 1}$ that are the factors of the model. These are estimated by minimizing the residual sum of squares over all observed entries.

Validation Setup: We employ a cross-validation strategy to obtain the performance of algorithms. For parameter tunning we use the grid-tuning strategy which performs a nested resampling algorithm. We use 5-folds cross-validation for outer and 3-folds cross-validation for inner re-sampling loop. At each level the folds are sampled uniformly at random from the entries represented in data frame format.

A. Results and Discussion

As performance measure, we compute root mean squared error (RMSE) and mean absolute error (MAE). The results in Table I show that, interestingly, none of the advanced algorithms in matrix completion can perform better than baselines on Movielens dataset. While the results are consistent across both error measures, MAE and RMSE, the best performance on predicting the movie rating is provided by the two factor model algorithm followed by OptSpace algorithm. RLMC with rank 1 and softImpute perform equally on Movielens-100k while RLMC with rank 1 performs slightly better on Movielens-1m. As expected OptSpace and SoftImpute perform poorly in predicting athletics performance. As shown in table I, RLMC algorithm with $r = 2$, significantly outperforms the other stat-of-the-art and baseline algorithms in terms of prediction accuracy across both measures RMSE and MAE. In this set of experiments we find the reconstructing minors by randomly sampling rows and columns.

Figure 1 demonstrate the benchmarking results on the synthetic data. We repeat each experiment on 20 random matrices and report the average result. Although the fraction of observed/missing entries in both square and rectangular matrices are the same, there is a significant difference between the performance of the algorithms. We note that the performance of the algorithms remain consistent across both MAE and RMSE measures. While OptSpace has the poorest performance in reconstructing both square and rectangular matrices, its performance remains stable by the change in noise level. SoftImpute performs better in square matrices, whereas RLMC works better in tall (rectangular) matrices. This implies that RLMC is the superior algorithm for real scenario data analytics such as recommender systems/athletic performance prediction

Algorithm	MovieLens-100k		MovieLens-1m		Athletic Performance	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
DoubleMean	0.983±0.005	0.760±0.004	0.936±0.001	0.736±0.001	0.913± 0.004	0.7070±0.0043
TwofacMod	0.971±0.004	0.744±0.005	0.913±0.001	0.7238±0.0009	0.5045± 0.0042	0.383±0.003
SoftImpute	1.001± 0.006	0.775±0.004	1.007±0.001	0.770±0.001	0.752± 0.013	0.229±0.005
OptSpace	0.978± 0.005	0.747 ± 0.004	0.917±0.001	0.7224 ± 0.0009	1.292± 0.008	1.4606 ± 0.0079
RLMC -$r = 1$	1.004±0.005	0.745±0.004	0.999 ±0.001	0.735±0.001	0.3818 ± 0.0009	0.2044 ± 0.0004
RLMC -$r = 2$	1.028±0.005	0.807±0.004	1.014±0.001	0.803±.001	0.1798 ± 0.0003	0.0212 ± 0.0003
RLMC -$r = 3$	1.033±0.004	0.819±0.005	1.025±0.001	0.807±0.001	0.2409 ± 0.0009	0.04159± 0.0004

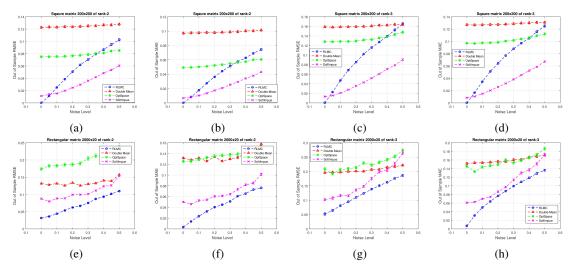TABLE I: Table of performances for benchmarking on Real datasets.

Fig. 1: Noise level versus out of sample prediction error. Figures 1a-1d and Figures 1e-1h show out of sample RMSE and MAE for random square and rectangular matrices, respectively. The results are depicted in comparison to the baseline algorithm, DoubleMean (Red curve).

where the set of items/standard distances is fixed while the number of users/runners may grow. Furthermore, our results are consistent with our expectations (as inferred by the theory of nuclear norm minimization) in section II.

VI. CONCLUSION

This paper proposes a new robust entry-wise matrix completion algorithm. Both analytic and simulation results with synthetic and real data suggest that the proposed algorithm outperforms classical ones in the reconstruction of noisy rectangular matrices arising in wide range of practical applications.

REFERENCES

[1] J. D. M. Rennie and N. Srebro, "Fast maximum margin matrix factorization for collaborative prediction," in *Proceedings of the 22Nd International Conference on Machine Learning*, ser. ICML '05, 2005, pp. 713–719.

[2] D. A. J. Blythe and F. J. Kiràly, "Prediction and quantification of individual athletic performance of runners," *PLoS ONE*, vol. 11, 06 2016.

[3] K. Q. Weinberger and L. K. Saul, "Unsupervised learning of image manifolds by semidefinite programming," *Int. J. Comput. Vision*, vol. 70, no. 1, pp. 77–90, Oct. 2006.

[4] N. Ruchansky, M. Crovella, and E. Terzi, "Matrix completion with queries," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '15. New York, NY, USA: ACM, 2015, pp. 1025–1034. [Online]. Available: http://doi.acm.org/10.1145/2783258.2783259

[5] A. Argyriou, C. A. Micchelli, M. Pontil, and Y. Ying, "A spectral regularization framework for multi-task structure learning," in *Proceedings of the 20th International Conference on Neural Information Processing Systems*, ser. NIPS'07, 2007, pp. 25–32.

[6] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics*, vol. 9, pp. 717–772, 2009.

[7] R. H. Keshavan, A. Montanari, and S. Oh, "Matrix completion from a few entries," *IEEE Trans. Inf. Theor.*, vol. 56, no. 6, pp. 2980–2998, Jun. 2010.

[8] F. Király, L. Theran, and R. Tomioka, "The algebraic combinatorial approach for low-rank matrix completion," *Journal of Machine Learning Research*, vol. 16, pp. 1391–1436, 2015.

[9] F. Kiraly and L. Theran, "Error-minimizing estimates and universal entry-wise error bounds for low-rank matrix completion," pp. 2364–2372, 2013.

[10] E. J. Candès and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Trans. Inf. Theor.*, vol. 56, no. 5, pp. 2053–2080, May 2010.

[11] R. Meka, P. Jainand, and I. S. Dhillon, "Guaranteed rank minimization via singular value projection," in *Advances in Neural Information Processing Systems 23*. Curran Associates, Inc., 2010, pp. 937–945.

[12] F. Király and R. Tomioka, "A combinatorial algebraic approach for the identifiability of low-rank matrix completion," in *Advances in Neural Information Processing Systems 23*. Published in ICML Proceedings., 2012.

[13] B. Recht, "A simpler approach to matrix completion," *J. Mach. Learn. Res.*, vol. 12, pp. 3413–3430, Dec. 2011.

[14] E. J. Candes and Y. Plan, "Matrix completion with noise," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 925–936, June 2010.

[15] R. H. Keshavan, A. Montanari, and S. Oh, "Matrix completion from noisy entries," *J. Mach. Learn. Res.*, vol. 11, pp. 2057–2078, Aug. 2010.