

# Improved Deterministic Conditions for Sparse and Low-Rank Matrix Decomposition

Ali Jalali

Department of Electrical and Computer Engineering  
University of Texas at Austin  
Email: alij@mail.utexas.edu

Sujay Sanghavi

Department of Electrical and Computer Engineering  
University of Texas at Austin  
Email: sanghavi@mail.utexas.edu

## Abstract

In this paper, the problem of splitting a given matrix into sparse and low-rank matrices is investigated. The problem is when and how we can exactly do this decomposition. This problem is ill-posed in general and we need to impose some (sufficient) conditions to be able to decompose a matrix into sparse and low-rank matrices. These conditions can be categorized into two general classes: (a) deterministic conditions and (b) probabilistic conditions. Deterministic conditions guarantee the success of decomposition for a given fixed matrix. In contrast, probabilistic conditions guarantee the success of the decomposition with certain probability for a certain class of random matrices. We improved the best existing result for deterministic conditions in this paper by introducing an alternative projection method for dual matrix construction.

## I. INTRODUCTION

Decomposition of matrices arises in analysis of complex systems. Decomposing a matrix (system) into matrices (sub-systems) with certain properties provides the opportunity to understand a complex system by understanding simpler pieces. More recently, sparse and low-rank decomposition became more interesting to researchers due to their applications in different areas such as but not limited to

**Graphical Models** In statistical learning of graphical model, the problem of learning a graphical model can be viewed as a sparse and low-rank matrix decomposition problem. Sparse matrix corresponds to the Gaussian graphical model of a number of variables with low dependencies and low-rank matrix corresponds to the hidden random variables that potentially have high correlation with all other variables. See [1] for more information.

**Matrix Completion** It has been shown that the low-rank assumption for matrix completion is a valid assumption. This assumption can be improved if we allow a small perturbation to the low-rank matrix. This perturbation can be modeled as a sparse matrix and the whole problem will be a sparse and low-rank decomposition problem.

**Graph Clustering** If we add the identity matrix to the adjacency matrix of a graph that we want to cluster, decomposing the modified adjacency matrix into sparse and low-rank matrices is equivalent to find the edges between clusters (in sparse matrix) and edges inside clusters (in low-rank matrix) [2].

The sparse and low-rank matrix decomposition problem can be considered as a deterministic problem, where we want to answer the question of when and how we can exactly recover a particular sparse and low-rank problem from their sum [3]. On the other hand, it can be considered as a probabilistic problem, where we want to answer the question of with what level of confidence can we make sure that the recovery of a sparse and low-rank matrix from their sum is possible given that they are drawn from certain probability distributions [4].

In both deterministic and probabilistic research papers, the authors propose an algorithm for decomposition, mainly based on convex optimization, and then try to show that their algorithm succeeds under certain conditions. Since the problem is ill-posed, as a sparse matrix can be low-rank and vice versa, we need to impose extra conditions to make the problem tractable. The main result of both research trends is to impose weaker (and consequently more general) sufficient conditions that guarantee – in deterministic or probabilistic sense – the success of the proposed decomposition algorithm.

In this paper, we mention improved conditions for the deterministic case for the algorithm proposed in [3], [5]. This algorithm is based on the minimization of the combination of nuclear norm of the low-rank matrix and  $\ell_1$  norm of the sparse matrix. We improve the deterministic conditions both in low-dimensional regimes, by increasing the constant upper-bound, and high-dimensional regime, by improving the scaling of the rank and sparsity with the size of the matrix.

The rest of the paper is organized as follows: Problem setup and main result are discussed in Section II. The three step proof technique is included in Section III. The paper is concluded in Section IV.

## II. PROBLEM SETUP AND MAIN RESULT

Suppose matrix  $C \in \mathbb{R}^{n_1 \times n_2}$  is the sum of a sparse matrix  $A^* \in \mathbb{R}^{n_1 \times n_2}$  and a low-rank matrix  $B^* \in \mathbb{R}^{n_1 \times n_2}$ . We consider the following problem: given  $C$ , when and how can we exactly recover  $A^*$  and  $B^*$ ? Clearly this will not always be possible, since there exist matrices that are both sparse and low-rank. The general problem, as stated, is fundamentally ill-posed. As discussed in [3], [5], one thus needs to impose additional conditions on  $A^*$  and  $B^*$  to ensure exact recovery. Thus, [3] imposes deterministic conditions that ensure exact recovery with convex optimization, while [6] considers random ensembles and recovery with high probability for large matrices with similar imposed conditions. This paper establishes new deterministic conditions, weaker and thus more general than those in [3], for exact recovery via a natural convex program.

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

$$\begin{aligned} (\hat{A}, \hat{B}) = \arg \min_{A, B} \quad & \gamma \|A\|_1 + \|B\|_* \\ \text{s.t.} \quad & A + B = C, \end{aligned} \quad (1)$$

where,  $\|B\|_* = \sum_i \sigma_i(B)$  is the nuclear norm, defined to be the sum of the singular values of the matrix, and  $\|A\|_1 = \sum_{i,j} |a_{ij}|$  is the elementwise  $\ell_1$  norm. Intuitively, the nuclear norm acts as a convex surrogate for the rank of a matrix, and  $\ell_1$  norm as a convex surrogate for its sparsity. We are interested in characterizing when the optimum recovers the underlying truth, i.e. when  $(\hat{A}, \hat{B}) = (A^*, B^*)$ . Towards this end we need some definitions; these are the same as those made in the literature [3], [6]–[9] and we just include them for completeness.

**Sparse Matrix Setup:** Given a matrix  $A \in \mathbb{R}^{n_1 \times n_2}$ , let the support of  $A$  to be  $\text{Supp}(A) = \{(i, j) : A_{i,j} \neq 0\}$ . Define the sub-space of sparse matrices with respect to the matrix  $A^*$  to be  $\Omega = \{A : \text{Supp}(A) \subseteq \text{Supp}(A^*)\}$ . The orthogonal projection of a matrix  $M \in \mathbb{R}^{n_1 \times n_2}$  to the space  $\Omega$  is the matrix whose  $(i, j)^{th}$  entry is given by

$$(P_\Omega(M))_{i,j} = \begin{cases} M_{i,j} & (i, j) \in \text{Supp}(A^*) \\ 0 & \text{ow} \end{cases}.$$

Let  $\Omega^\perp$  be the orthogonal complement sub-space with projection  $P_{\Omega^\perp}(M) = M - P_\Omega(M)$ . As discussed in [3], for exact recovery the 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. Correspondingly, we say a given matrix  $A^*$  is  $(\eta, d)$ -**spread** for some  $d \in \{1, \dots, \max(n_1, n_2)\}$  and  $\eta \in [\frac{1}{d}, 1]$  iff (i) it has at most  $d$  non-zero entries 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 elementwise maximum magnitude of the elements of the matrix.

**Low-Rank Matrix Setup:** 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  $\Sigma \in \mathbb{R}^{r \times r}$ . Let  $\mathcal{T} = \{B = UX^* + YV^* : X \in \mathbb{R}^{n_2 \times r}, Y \in \mathbb{R}^{n_1 \times r}\}$  be the sub-space spanned by matrices that share the same column space, or the same row space, as  $B^*$ . The orthogonal projection of a matrix  $M \in \mathbb{R}^{n_1 \times n_2}$  to the sub-space  $\mathcal{T}$  is

$$P_{\mathcal{T}}(M) = UU^*M + MVV^* - UU^*MVV^*.$$

Let  $\mathcal{T}^\perp$  be the orthogonal complement sub-space with projection  $P_{\mathcal{T}^\perp}(M) = M - P_{\mathcal{T}}(M)$ . As discussed in [3], for exact recovery, we need the low-rank matrix to have incoherent – both left and right – eigen vectors with respect to standard basis. This guarantees that the low-rank matrix is not too sparse. We say a given matrix  $B^*$  is  $(r, \mu)$ -**incoherent** for some  $r \in \{1, \dots, \min(n_1, n_2)\}$  and  $\mu \in [1, \frac{\max(n_1, n_2)}{r}]$  iff (i)  $\text{rank}(B^*) = r$ , and, (ii)

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

where,  $\mathbf{e}_i$ 's are standard basis vectors with proper length. Here  $\|\cdot\|$  represents the 2-norm of the vector and should not be mistaken by operator norm defined before.

**Remark 1.** The smaller the parameter  $\mu$  is, the more spread out the singular vectors are. Also, since  $UU^*$  and  $VV^*$  are orthogonal projections to the column spaces of  $U$  and  $V$ , respectively,  $B^*$  is  $\mu$ -incoherent low-rank matrix iff

$$\begin{aligned} \|UU^*\|_{\infty,2} &:= \max_i \|UU^* \mathbf{e}_i\| \leq \sqrt{\frac{\mu r}{n_1}} & \|VV^*\|_{\infty,2} &:= \max_i \|VV^* \mathbf{e}_i\| \leq \sqrt{\frac{\mu r}{n_2}} \\ \|UV^*\|_{\infty} &\leq \sqrt{\frac{\mu r}{n_1 n_2}}. \end{aligned} \quad (2)$$

**Existing deterministic conditions:** It has been shown (see Corrolary 3 in [3]) that if  $\eta d \sqrt{\frac{\mu r}{\min(n_1, n_2)}} \leq \frac{1}{6}$  then there exist a range of  $\gamma \in \mathbb{R}^+$  such that  $A^* = \hat{A}$  and  $B^* = \hat{B}$  can be exactly recovered from their sum. As we see, using this existing bound,  $n$  (for square matrices) should scale with  $d^2 r$  in order to guarantee the exact recovery. In this paper, we want to improve this bound in two senses: one is to find a better scaling for parameters  $r$ ,  $n$  and  $d$  and two is to find a better constant (on the left hand-side of the inequality) that guarantees the recovery for a larger set of matrices.

**Main Result:** The improved conditions for deterministic sparse and low-rank matrix decomposition – namely, the main result of this paper – is stated in the following theorem.

**Theorem 1.** Suppose  $\sqrt{\eta d} \sqrt{\frac{\mu r}{\min(n_1, n_2)}} \left(1 + \sqrt{\frac{\min(n_1, n_2)}{\max(n_1, n_2)}} + \sqrt{\frac{\eta d}{\max(n_1, n_2)}}\right) \leq \frac{1}{2}$  and  $\gamma \in \left[\frac{1}{1-2\alpha} \sqrt{\frac{\mu r}{n_1 n_2}}, \frac{1-\alpha}{d} - \sqrt{\frac{\mu r}{n_1 n_2}}\right]$ , where,  $\alpha = \sqrt{\frac{\mu r \eta d}{n_1}} + \sqrt{\frac{\mu r \eta d}{n_2}}$ . Then, the solution to the problem (1) is unique and equal to  $(A^*, B^*)$ .

**Remark 2.** Notice that if  $\eta d \sqrt{\frac{\mu r}{\min(n_1, n_2)}} \leq \frac{1}{6}$  then the condition of Theorem 1 is satisfied. This shows that our result is an improvement of the result in [3] 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 [3].

### III. PROOF

The proof follows along the lines of that in [3] and has three steps: (a) writing down a sufficient optimality condition, stated in terms of a dual certificate, for  $(A^*, B^*)$  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  $(A^*, B^*)$  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 [3].

#### A. Optimality conditions

For the sake of completeness, we restate here a first-order sufficient condition that need to be satisfied for  $(A^*, B^*)$  to be the optimum of (1). The reader is referred to [3] for a proof.

**Lemma 1 (A Sufficient Optimality Condition).** The pair  $(A^*, B^*)$  is the unique optimal solution of (1) if

(a)  $\Omega \cap \mathcal{T} = \{\mathbf{0}\}$ .

(b) There exists a dual matrix  $Q \in \mathbb{R}^{n_1 \times n_2}$  satisfying

$$\begin{aligned} P_{\mathcal{T}}(Q) &= UV^* & \|P_{\mathcal{T}^\perp}(Q)\| &< 1 \\ P_{\Omega}(Q) &= \gamma \text{sgn}(A^*) & \|P_{\Omega^\perp}(Q)\|_{\infty} &< \gamma, \end{aligned} \quad (3)$$

where,  $\text{sgn}(\cdot)$  is the element-wise signum function.

Condition (a) guarantees that the sparse matrices and low-rank matrices can be distinguished without ambiguity. In other words, any given matrix can not be both sparse and low-rank except the zero matrix. The following lemma gives a sufficient guarantee for the condition (a). We construct the dual matrix  $Q$  in the next subsection and prove condition (b) afterwards.

**Lemma 2.** If  $\alpha < 1$  then  $\Omega \cap \mathcal{T} = \{\mathbf{0}\}$ .

*Proof:* It is clear that  $\{\mathbf{0}\} \in \Omega \cap \mathcal{T}$ . In contrary assume that there exists a non-zero matrix  $M \in \Omega \cap \mathcal{T}$ . By idempotency of orthogonal projections, we have  $M = P_\Omega(M) = P_{\mathcal{T}}(P_\Omega(M))$  and hence

$$\begin{aligned} \|P_{\mathcal{T}}(P_\Omega(M))\|_\infty &= \|UU^*P_\Omega(M) + P_\Omega(M)V V^* - UU^*P_\Omega(M)V V^*\|_\infty \\ &\leq \|UU^*P_\Omega(M)\|_\infty + \|(I - UU^*)P_\Omega(M)V V^*\|_\infty \\ &\leq \max_i \|UU^* \mathbf{e}_i\| \|P_\Omega(M)\| + \|I - UU^*\| \|P_\Omega(M)\| \max_i \|V V^* \mathbf{e}_i\| \\ &\leq \alpha \|P_\Omega(M)\|_\infty = \alpha \|P_{\mathcal{T}}(P_\Omega(M))\|_\infty. \end{aligned} \quad (4)$$

Hence,  $\|M\|_\infty = \mathbf{0}$  or equivalently,  $M = \mathbf{0}$ . This is a contradiction. ■

### B. New Dual Certificate

We now describe our main innovation, a new way to construct the candidate dual certificate  $Q$ . This procedure is different from the ones in [3], [6], [9]. As a first step, consider two matrices  $Q_a$  and  $Q_b$  defined as follows: with  $M^* = \gamma \text{sgn}(A^*)$  and  $N^* = UV^*$ , let

$$\begin{aligned} Q_a &= M^* - P_{\mathcal{T}}(M^*) + P_\Omega(P_{\mathcal{T}}(M^*)) - P_{\mathcal{T}}(P_\Omega(P_{\mathcal{T}}(M^*))) + \dots \\ Q_b &= N^* - P_\Omega(N^*) + P_{\mathcal{T}}(P_\Omega(N^*)) - P_\Omega(P_{\mathcal{T}}(P_\Omega(N^*))) + \dots \end{aligned}$$

Lemma 3 below establishes that  $Q_a$  and  $Q_b$  as described above are well-defined, i.e. it establishes that the infinite summations converge, under the conditions of the theorem. Note that when this is the case, we have that

$$\begin{aligned} P_{\mathcal{T}}(Q_b) &= UV^* & P_{\mathcal{T}}(Q_a) &= \mathbf{0} \\ P_\Omega(Q_a) &= \gamma \text{sgn}(A^*) & P_\Omega(Q_b) &= \mathbf{0}. \end{aligned} \quad (5)$$

Motivated by these properties, if we let  $Q = Q_a + Q_b$ , it is easy to see that the equality conditions in (3) are satisfied. In the next subsection, we will show that the inequality conditions are also satisfied under the assumptions of the theorem 1.

**Lemma 3.** *If  $\alpha < 1$ , then  $Q_a$  and  $Q_b$  converge.*

*Proof:* For any matrix  $W \in \mathbb{R}^{n_1 \times n_2}$ , let  $\mathbf{S}_W = W + P_{\mathcal{T}}(P_\Omega(W)) + P_{\mathcal{T}}(P_\Omega(P_{\mathcal{T}}(P_\Omega(W)))) + \dots$ . It suffices to show that  $\mathbf{S}_W$  converges for all  $W$  since  $Q_a = M^* - P_{\Omega^\perp}(\mathbf{S}_{P_{\mathcal{T}}(M^*)})$  and  $Q_b = \mathbf{S}_{N^* - P_\Omega(N^*)}$ . Notice that  $\|P_{\mathcal{T}}(P_\Omega(W))\|_\infty \leq \alpha \|P_\Omega(W)\|_\infty \leq \alpha \|W\|_\infty$  as shown in (4) and hence  $\mathbf{S}_W$  geometrically converges. ■

### C. Certification

Considering  $Q = Q_a + Q_b$  as a candidate for dual matrix, we need to show the conditions in (3) are satisfied under the conditions of the theorem. As we showed in the previous subsection, the equality conditions are satisfied by construction of  $Q_a$  and  $Q_b$ . To prove the inequality conditions, we first bound the projection of  $Q$  into orthogonal complement spaces in next lemma.

**Lemma 4.** *If  $\alpha < 1$  then*

$$\begin{aligned} \|P_{\Omega^\perp}(Q)\|_\infty &\leq \frac{1}{1-\alpha} \left( \sqrt{\frac{\mu r}{n_1 n_2}} + \alpha \gamma \right) \\ \|P_{\mathcal{T}^\perp}(Q)\| &\leq \frac{\eta d}{1-\alpha} \left( \sqrt{\frac{\mu r}{n_1 n_2}} + \gamma \right). \end{aligned}$$

*Proof:* Using the definition of  $\mathbf{S}_W$  for any matrix  $W \in \mathbb{R}^{n_1 \times n_2}$ , we get  $\|\mathbf{S}_W\|_\infty \leq \frac{1}{1-\alpha} \|W\|_\infty$ , because of the geometrical convergence. Thus, we have

$$\begin{aligned} \|P_{\Omega^\perp}(Q)\|_\infty &= \|P_{\Omega^\perp}(\mathbf{S}_{N^* - P_{\mathcal{T}}(M^*)})\|_\infty \leq \|\mathbf{S}_{N^* - P_{\mathcal{T}}(M^*)}\|_\infty \\ &\leq \frac{1}{1-\alpha} \|N^* - P_{\mathcal{T}}(M^*)\|_\infty \leq \frac{1}{1-\alpha} (\|N^*\|_\infty + \|P_{\mathcal{T}}(M^*)\|_\infty) \\ &\leq \frac{1}{1-\alpha} (\|N^*\|_\infty + \alpha \|M^*\|_\infty) \leq \frac{1}{1-\alpha} \left( \sqrt{\frac{\mu r}{n_1 n_2}} + \alpha \gamma \right). \end{aligned}$$

For the second part, since  $\|\mathbf{I} - UU^*\| \leq 1$  and  $\|\mathbf{I} - VV^*\| \leq 1$ , we have

$$\begin{aligned}
\|P_{\mathcal{T}^\perp}(Q)\| &= \|P_{\mathcal{T}^\perp}(M^* - P_\Omega(\mathbf{S}_{N^* - P_{\mathcal{T}}(M^*)))\| \\
&= \|(\mathbf{I} - UU^*)(M^* - P_\Omega(\mathbf{S}_{N^* - P_{\mathcal{T}}(M^*))) (\mathbf{I} - VV^*)\| \\
&\leq \|M^* - P_\Omega(\mathbf{S}_{N^* - P_{\mathcal{T}}(M^*)))\| \leq \eta d \|M^* - P_\Omega(\mathbf{S}_{N^* - P_{\mathcal{T}}(M^*)))\|_\infty \\
&\leq \eta d (\|M^*\|_\infty + \|\mathbf{S}_{N^* - P_{\mathcal{T}}(M^*)}\|_\infty) \leq \eta d \left( \gamma + \frac{1}{1 - \alpha} \left( \sqrt{\frac{\mu r}{n_1 n_2}} + \alpha \gamma \right) \right) \\
&\leq \frac{\eta d}{1 - \alpha} \left( \sqrt{\frac{\mu r}{n_1 n_2}} + \gamma \right).
\end{aligned}$$

This concludes the proof of the lemma. ■

Finally to satisfy (3), we require

$$\begin{aligned}
\|P_{\mathcal{T}^\perp}(Q)\| &\leq \frac{\eta d}{1 - \alpha} \left( \sqrt{\frac{\mu r}{n_1 n_2}} + \gamma \right) < 1 \\
\|P_{\Omega^\perp}(Q)\|_\infty &\leq \frac{1}{1 - \alpha} \left( \sqrt{\frac{\mu r}{n_1 n_2}} + \alpha \gamma \right) < \gamma.
\end{aligned}$$

Combining these two inequalities, we get  $\frac{1}{1 - 2\alpha} \sqrt{\frac{\mu r}{n_1 n_2}} < \gamma < \frac{1 - \alpha}{d} - \sqrt{\frac{\mu r}{n_1 n_2}}$  as stated in the assumptions of the theorem.

#### IV. CONCLUSION

In this paper, using the alternating projection method, we found an improved deterministic condition for the sparse and low-rank matrix decomposition problem. This result outperforms the previous results both by providing better scaling and better constant limits.

#### REFERENCES

- [1] S. L. Lauritzen, *Graphical Models*. Oxford University Press, 1996.
- [2] "Graph clustering via convex optimization."
- [3] V. Chandrasekaran, S. Sanghavi, P. Parrilo, and A. S. Willsky, "Rank-sparsity incoherence for matrix decomposition," *Submitted*, 2009.
- [4] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections and universal encoding strategies," vol. 52, no. 12, pp. 5406–5425, 2006.
- [5] V. Chandrasekaran, S. Sanghavi, P. Parrilo, and A. S. Willsky, "Sparse and low-rank matrix decompositions," in *15th IFAC Symposium on System Identification (SYSID)*, 2009.
- [6] 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.
- [7] E. J. Candes and B. Recht, "Exact matrix completion via convex optimization," *Foundation of Computational Math*, vol. 9, pp. 717–772, 2009.
- [8] E. J. Candes and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Transaction on Information Theory*, 2009.
- [9] D. Gross, "Recovering low-rank matrices from few coefficients in any basis," Available at [arXiv:0910.1879v4](http://arxiv.org/abs/0910.1879v4), 2009.