

Low-Rank Matrix Approximation with Weights or Missing Data is NP-hard

Nicolas Gillis¹ and François Glineur¹

Abstract

Weighted low-rank approximation (WLRA), a dimensionality reduction technique for data analysis, has been successfully used in several applications, such as in collaborative filtering to design recommender systems or in computer vision to recover structure from motion. In this paper, we prove that computing an optimal weighted low-rank approximation is NP-hard, already when a rank-one approximation is sought. In fact, we show that it is hard to compute approximate solutions to the WLRA problem with some prescribed accuracy. Our proofs are based on reductions from the maximum-edge biclique problem, and apply to strictly positive weights as well as to binary weights (the latter corresponding to low-rank matrix approximation with missing data).

Keywords: low-rank matrix approximation, weighted low-rank approximation, missing data, matrix completion with noise, PCA with missing data, computational complexity, maximum-edge biclique problem.

1 Introduction

Approximating a matrix with one of lower rank is a key problem in data analysis and is widely used for linear dimensionality reduction. Numerous variants exist emphasizing different constraints and objective functions, e.g., principal component analysis (PCA) [16], independent component analysis [6], nonnegative matrix factorization [18], and other refinements are often imposed on these models, e.g., sparsity to improve interpretability or increase compression [7].

In some cases, it may be necessary to attach a weight to each entry of the data matrix, expressing its relative importance [8]. This is for example the case in the following situations:

- ◊ The matrix to be approximated is obtained via a sampling procedure and the number of samples and/or the expected variance vary among the entries. For example, it has been shown that using a weighted norm gives better results in 2-D digital filter design [19] and microarray data analysis [20].
- ◊ Some data is missing/unknown, which can be taken into account by assigning zero weights to the missing/unknown entries of the data matrix. This is for example the case in collaborative filtering, notably used to design recommender systems [24] (in particular, the Netflix prize competition has demonstrated the effectiveness of low-rank matrix factorization techniques [17]), or in computer vision to recover structure from motion [26, 15], see also [4]. This problem is often referred to as *PCA with missing data* [26, 13], and can be viewed as a *low-rank matrix completion problem with noise*, i.e., approximate a given noisy data matrix featuring missing entries with a low-rank matrix¹.

¹Université catholique de Louvain, CORE, B-1348 Louvain-la-Neuve, Belgium. E-mail: nicolas.gillis@uclouvain.be and francois.glineur@uclouvain.be. Nicolas Gillis is a research fellow of the Fonds de la Recherche Scientifique (F.R.S.-FNRS). This text presents research results of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The scientific responsibility is assumed by the authors.

¹In our settings, the rank of the approximation is fixed a priori.

- ◇ A greater emphasis must be placed on the accuracy of the approximation on a localized part of the data, a situation encountered for example in image processing [14, Chapter 6].

Finding a low-rank matrix which is closest to the input matrix according to these weights is an optimization problem called *weighted low-rank approximation* (WLRA). Formally, it can be formulated as follows: first, given an m -by- n nonnegative weight matrix $W \in \mathbb{R}_+^{m \times n}$, we define the weighted Frobenius norm of an m -by- n matrix A as $\|A\|_W = (\sum_{i,j} W_{ij} A_{ij}^2)^{\frac{1}{2}}$. Then, given an m -by- n real matrix $M \in \mathbb{R}^{m \times n}$ and a positive integer $r \leq \min(m, n)$, we seek an m -by- n matrix X with rank at most r that approximates M as closely as possible, where the quality of the approximation is measured by the weighted Frobenius norm of the error:

$$p^* = \inf_{X \in \mathbb{R}^{m \times n}} \|M - X\|_W^2 \text{ such that } X \text{ has rank at most } r.$$

Since any m -by- n matrix with rank at most r can be expressed as the product of two matrices of dimensions m -by- r and r -by- n , we will use the following more convenient formulation featuring two unknown matrices $U \in \mathbb{R}^{m \times r}$ and $V \in \mathbb{R}^{r \times n}$ but no explicit rank constraint:

$$p^* = \inf_{U \in \mathbb{R}^{m \times r}, V \in \mathbb{R}^{r \times n}} \|M - UV^T\|_W^2 = \sum_{ij} W_{ij} (M - UV^T)_{ij}^2. \quad (\text{WLRA})$$

Even though (WLRA) is suspected to be NP-hard [15, 27], this has never, to the best of our knowledge, been studied formally. In this paper, we analyze the computational complexity in the rank-one case (i.e., for $r = 1$) and prove the following two results.

Theorem 1. *When $M \in \{0, 1\}^{m \times n}$, and $W \in]0, 1]^{m \times n}$, it is NP-hard to find an approximate solution of rank-one (WLRA) with objective function accuracy less than $2^{-11}(mn)^{-6}$.*

Theorem 2. *When $M \in [0, 1]^{m \times n}$, and $W \in \{0, 1\}^{m \times n}$, it is NP-hard to find an approximate solution of rank-one (WLRA) with objective function accuracy less than $2^{-12}(mn)^{-7}$.*

In other words, it is NP-hard to find an approximate solution to rank-one (WLRA) with positive weights, and to the rank-one matrix approximation problem with missing data. Note that these results can be easily generalized to any fixed rank r , see Remark 3.

The paper is organized as follows. We first review existing results about the complexity of (WLRA) in Section 2. In Section 3.1, we introduce the maximum-edge biclique problem (MBP), which is NP-hard. In Sections 3.2 and 3.3, we prove Theorems 1 and 2 respectively, using polynomial-time reductions from MBP. We conclude with a discussion and some open questions.

1.1 Notation

The set of real m -by- n matrices is denoted $\mathbb{R}^{m \times n}$, or $\mathbb{R}_+^{m \times n}$ when all the entries are required to be nonnegative. For $A \in \mathbb{R}^{m \times n}$, we note $A_{:j}$ the j^{th} column of A , $A_{i:}$ the i^{th} row of A , and A_{ij} or $A(i, j)$ the entry at position (i, j) ; for $b \in \mathbb{R}^{m \times 1} = \mathbb{R}^m$, we note b_i the i^{th} entry of b . The transpose of A is A^T . The Frobenius norm of a matrix A is defined as $\|A\|_F^2 = \sum_{i,j} (A_{ij})^2$, and $\|\cdot\|_2$ is the usual Euclidean norm with $\|b\|_2^2 = \sum_i b_i^2$. For $W \in \mathbb{R}_+^{m \times n}$, the weighted Frobenius ‘norm’ of a matrix A is defined² by $\|A\|_W^2 = \sum_{i,j} W_{ij} (A_{ij})^2$. The m -by- n matrix of all ones is denoted $\mathbf{1}_{m \times n}$, the m -by- n matrix of all zeros $\mathbf{0}_{m \times n}$, and I_n is the identity matrix of dimension n . The smallest integer larger or equal to x is denoted $\lceil x \rceil$.

² $\|\cdot\|_W$ is a matrix norm if and only if $W > 0$, else it is a semi-norm.

2 Previous Results

Weighted low-rank approximation is suspected to be much more difficult than the corresponding unweighted problem (i.e., when W is the matrix of all ones), which is efficiently solved using the singular value decomposition (SVD) [12]. In fact, it has been previously observed that the weighted problem might have several local minima which are not global [27], while this cannot occur in the unweighted case (i.e., when W is the matrix of all ones), see, e.g., [14, p.29, Th.1.14].

Example 1. *Let*

$$M = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad \text{and} \quad W = \begin{pmatrix} 1 & 100 & 2 \\ 100 & 1 & 2 \\ 1 & 1 & 1 \end{pmatrix}.$$

In the case of a rank-one factorization ($r = 1$) and a nonnegative matrix M , one can impose without loss of generality that the solutions of (WLRA) are nonnegative. In fact, one can easily check that any rank-one solution uv^T of (WLRA) can only be improved by taking its component-wise absolute value $|uv^T| = |u||v|^T$. Moreover, we can impose without loss of generality that $\|u\|_2 = 1$, so that only two degrees of freedom remain. Indeed, for a given

$$u(x, y) = \begin{pmatrix} x \\ y \\ \sqrt{1 - x^2 - y^2} \end{pmatrix}, \quad \text{with} \quad \begin{cases} x \geq 0, y \geq 0 \\ x^2 + y^2 \leq 1 \end{cases},$$

the corresponding optimal $v^(x, y) = \operatorname{argmin}_v \|M - u(x, y)v^T\|_W^2$ can be computed easily³. Figure 1 displays the graph of the objective function $\|M - u(x, y)v^*(x, y)^T\|_W$ with respect to parameters x and y ; we observe four local minima, close to $(\frac{\sqrt{2}}{2}, 0)$, $(0, \frac{\sqrt{2}}{2})$, $(0, 0)$ and $(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$. We will see later in*

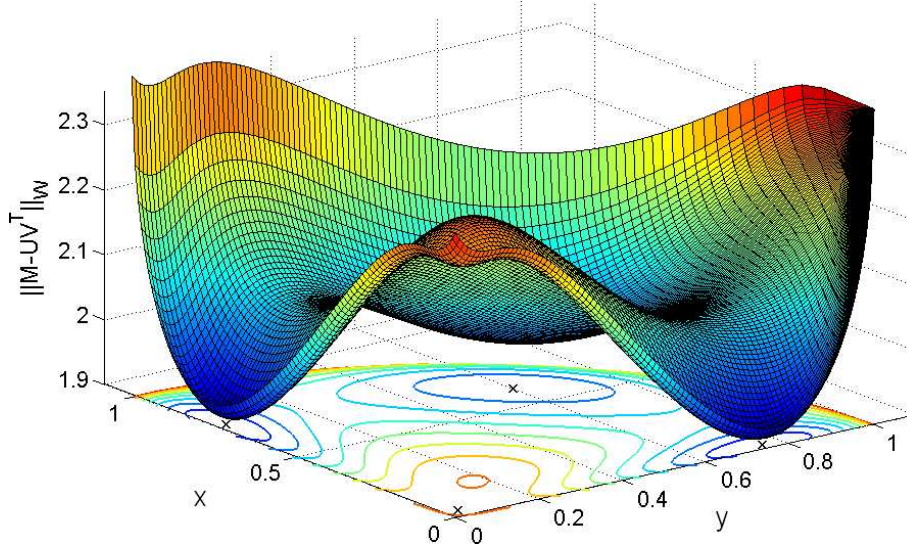


Figure 1: Objective function of (WLRA) with respect to the parameters (x, y) .

Section 3 how this example has been generated.

³This problem can be decoupled into n independent quadratic programs in one variable, and admits the following closed-form solution: $v^*(x, y) = [(M \circ W)^T u] / [W^T (u \circ u)]$, where \circ (resp. $/$) is the component-wise multiplication (resp. division).

However, if the rank of the weight matrix $W \in \mathbb{R}_+^{m \times n}$ is equal to one, i.e., $W = st^T$ for some $s \in \mathbb{R}_+^m$ and $t \in \mathbb{R}_+^n$, (WLRA) can be reduced to an unweighted low-rank approximation. In fact,

$$\begin{aligned} \|M - UV^T\|_W^2 &= \sum_{i,j} W_{ij} (M - UV^T)_{ij}^2 = \sum_{i,j} s_i t_j (M - UV^T)_{ij}^2 \\ &= \sum_{i,j} \left(\sqrt{s_i t_j} M_{ij} - (\sqrt{s_i} U_{i:})(\sqrt{t_j} V_{j:}^T) \right)^2. \end{aligned}$$

Therefore, if we define a matrix M' such that $M'_{ij} = \sqrt{s_i t_j} M_{ij} \forall i, j$, an optimal weighted low-rank approximation (U, V) of M can be recovered from a solution (U', V') to the unweighted problem for matrix M' using $U_{i:} = U'_{i:}/\sqrt{s_i} \forall i$ and $V_{j:} = V'_{j:}/\sqrt{t_j} \forall j$.

When the weight matrix W is binary, WLRA amounts to approximating a matrix with missing data. This problem is closely related to *low-rank matrix completion*, see [2] and the references therein, which can be defined as

$$\min_X \text{rank}(X) \quad \text{such that } X_{ij} = M_{ij} \text{ for } (i, j) \in \Omega, \quad (\text{MC})$$

where $\Omega \subseteq \{1, 2, \dots, m\} \times \{1, 2, \dots, n\}$ is the set of entries for which the values of M are known. (MC) has been shown to be NP-hard [5], and it is clear that an optimal solution X^* of (MC) can be obtained by solving a sequence of (WLRA) problems with the same matrix M , with

$$W_{ij} = \begin{cases} 1 & \text{if } (i, j) \in \Omega \\ 0 & \text{otherwise} \end{cases},$$

and for different values of the target rank ranging from $r = 1$ to $r = \min(m, n)$. The smallest value of r for which the objective function $\|M - UV^T\|_W^2$ of (WLRA) vanishes provides an optimal solution for (MC). This observation implies that it is NP-hard to solve (WLRA) for each possible value of r from 1 to $\min(m, n)$, since it would solve (MC). However, this does not imply that (WLRA) is NP-hard when r is fixed, and in particular when r equals one. In fact, checking whether (MC) admits a rank-one solution can be done easily⁴.

Rank-one (WLRA) can be equivalently reformulated as

$$\inf_A \|M - A\|_W^2 \quad \text{such that} \quad \text{rank}(A) \leq 1,$$

and, when W is binary, is the problem of finding, if possible, the best rank-one approximation of a matrix with missing entries. To the best of our knowledge, the complexity of this problem has never been studied formally; it will be shown to be NP-hard in the next section.

Another closely related result is the NP-hardness of the structure from motion problem (SFM), in the presence of noise and missing data [21]. Several points of a rigid object are tracked with cameras (we are given the projections of the 3-D points on the 2-D camera planes)⁵, and the aim is to recover the structure of the object and the positions of the 3-D points. SFM can be written as a rank-four (WLRA) problem with a binary weight matrix⁶ [15]. However, this result does not imply anything on the complexity of rank-one (WLRA).

An important feature of (WLRA) is exposed by the following example.

⁴The solution $X = uv^T$ can be constructed observing that the vector u must be a multiple of each column of M .

⁵Missing data arise because the points may not always be visible by the cameras, e.g., in the case of a rotation.

⁶With the additional constraint that the last row of V must be all ones, i.e., $V_{r:} = \mathbf{1}_{1 \times n}$.

Example 2. Let

$$M = \begin{pmatrix} 1 & ? \\ 0 & 1 \end{pmatrix},$$

where $?$ indicates that an entry is missing, i.e., that the weight associated with this entry is 0 (1 otherwise). Observe that $\forall (u, v) \in \mathbb{R}^m \times \mathbb{R}^n$,

$$\text{rank}(M) = 2 \text{ and } \text{rank}(uv^T) = 1 \Rightarrow \|M - uv^T\|_W > 0.$$

However, we have

$$\inf_{(u,v) \in \mathbb{R}^m \times \mathbb{R}^n} \|M - uv^T\|_W = 0.$$

In fact, one can check that with

$$u(\epsilon) = \begin{pmatrix} 1 \\ \epsilon \end{pmatrix} \text{ and } v(\epsilon) = \begin{pmatrix} 1 \\ \epsilon^{-1} \end{pmatrix}, \text{ we have } \lim_{\epsilon \rightarrow 0} \|M - u(\epsilon)v(\epsilon)^T\|_W = 0.$$

This indicates that, when W has zero entries, the set of optimal solutions of (WLRA) might be empty. In other words, the (bounded) infimum of the objective function might be unattained. On the other hand, the infimum is always attained for $W > 0$ since $\|\cdot\|_W$ is then a norm.

For this reason, these two cases will be analyzed separately: in Section 3.2, we study the computational complexity of the problem when $W > 0$, and, in Section 3.3, the case of a binary W (i.e., the problem with missing data).

3 Complexity of rank-one (WLRA)

In this section, we use polynomial-time reductions from the maximum-edge biclique problem to prove Theorems 1 and 2.

3.1 Maximum-Edge Biclique Problem

A *bipartite graph* is a graph whose vertices can be partitioned into two disjoint sets such that there is no edge between two vertices in the same set. The maximum-edge biclique problem (MBP) in a bipartite graph is the problem of finding a complete bipartite subgraph (a *biclique*) with the maximum number of edges.

Let $M \in \{0, 1\}^{m \times n}$ be the biadjacency matrix of a bipartite graph $G_b = (V_1 \cup V_2, E)$ with $V_1 = \{s_1, \dots, s_m\}$, $V_2 = \{t_1, \dots, t_n\}$ and $E \subseteq (V_1 \times V_2)$, i.e.,

$$M_{ij} = 1 \iff (s_i, t_j) \in E.$$

The cardinality of E will be denoted $|E| = \|M\|_F^2 \leq mn$.

For example, Figure 2 displays the graph G_b generated by the matrix M of Example 1.

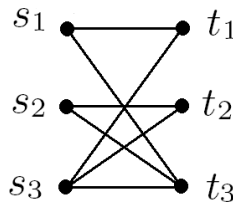


Figure 2: Graph corresponding to the matrix M of Example 1.

With this notation, the maximum-edge biclique problem in a bipartite graph can be formulated as follows [11]

$$\begin{aligned} \min_{u,v} \quad & \|M - uv^T\|_F^2 \\ & u_i v_j \leq M_{ij}, \forall i, j \\ & u \in \{0, 1\}^m, v \in \{0, 1\}^n, \end{aligned} \tag{MBP}$$

where $u_i = 1$ (resp. $v_j = 1$) means that node s_i (resp. t_j) belongs to the solution, $u_i = 0$ (resp. $v_j = 0$) otherwise. The first constraint guarantees feasible solutions of (MBP) to be bicliques of G_b . In fact, it is equivalent to the implication

$$M_{ij} = 0 \quad \Rightarrow \quad u_i = 0 \quad \text{or} \quad v_j = 0,$$

i.e., if there is no edge between vertices s_i and t_j , they cannot simultaneously belong to a solution. The objective function minimizes the number of edges outside the biclique, which is equivalent to maximizing the number of edges inside the biclique. Notice that the minimum of (MBP) is $|E| - |E^*|$, where $|E^*|$ denotes the number of edges in an optimal biclique.

The decision version of the MBP problem:

Given K , does G_b contain a biclique with at least K edges?

has been shown to be NP-complete [23] in the usual Turing machine model [9], which is our framework in this paper. Therefore, computing $|E| - |E^*|$, the optimal value of (MBP), is NP-hard.

3.2 Low-Rank Matrix Approximation with Positive Weights

In order to prove NP-hardness of rank-one (WLRA) with positive weights ($W > 0$), let us consider the following instance:

$$p^* = \min_{u \in \mathbb{R}^m, v \in \mathbb{R}^n} \|M - uv^T\|_W^2, \tag{W-1d}$$

with $M \in \{0, 1\}^{m \times n}$ the biadjacency matrix of a bipartite graph $G_b = (V, E)$ and the weight matrix defined as

$$W_{ij} = \begin{cases} 1 & \text{if } M_{ij} = 1 \\ d & \text{if } M_{ij} = 0 \end{cases}, \quad 1 \leq i \leq m, 1 \leq j \leq n,$$

where $d \geq 1$ is a parameter.

Intuitively, increasing the value of d makes the zero entries of M more important in the objective function, which leads them to be approximated by small values. This observation will be used to show that, for d sufficiently large, the optimal value p^* of (W-1d) will be close to $|E| - |E^*|$, the optimal value of (MBP) (Lemma 3).

A maximal biclique in G_b is a biclique not contained in a larger biclique, and can be seen as a ‘locally’ optimal solutions of (MBP). We will show that, as the value of parameter d increases, the local minima of (W-1d) get closer to binary vectors describing maximal bicliques in G_b .

Example 1 illustrates the situation: the graph G_b corresponding to matrix M (cf. Figure 2) contains four maximal bicliques $\{s_1, s_3, t_1, t_3\}$, $\{s_2, s_3, t_2, t_3\}$, $\{s_3, t_1, t_2, t_3\}$ and $\{s_1, s_2, s_3, t_3\}$, and the weight matrix W that was used is similar to the case $d = 100$ in problem (W-1d). We now observe that (W-1d) has four local optimal solutions as well (cf. Figure 1) close to $(\frac{\sqrt{2}}{2}, 0)$, $(0, \frac{\sqrt{2}}{2})$, $(0, 0)$ and $(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$. There is a one to one correspondence between these solutions and the four maximal bicliques listed above (in this order). For example, for $(x, y) = (\frac{\sqrt{2}}{2}, 0)$ we have $u(x, y) = (\frac{\sqrt{2}}{2}, 0, \frac{\sqrt{2}}{2})^T$, $v^*(x, y)$ is approximately equal to $(\sqrt{2}, 0, \sqrt{2})^T$, and this solution corresponds to the maximal biclique $\{s_1, s_3, t_1, t_3\}$.

Notice that a similar idea was used in [10] to prove NP-hardness of the rank-one nonnegative factorization problem $\min_{u \in \mathbb{R}_+^m, v \in \mathbb{R}_+^n} \|M - uv^T\|_F$, where the zero entries of M were replaced by sufficiently large negative ones.

Remark 1 (Link with classical quadratic penalty method). *It is worth noting that (W-1d) can be viewed as the application of the classical quadratic penalty approach to the biclique problem, see, e.g., [22, §17.1]. In fact, defining $F = \{(i, j) | M_{ij} = 1\}$ and its complement $\bar{F} = \{(i, j) | M_{ij} = 0\}$, the biclique problem can be formulated as*

$$\min_{u, v} \sum_{(i, j) \in F} (1 - u_i v_j)^2 \quad \text{such that } u_i v_j = 0 \text{ for all } (i, j) \in \bar{F}. \quad (3.1)$$

Indeed, in this formulation, it is clear that any optimal solution can be chosen such that vectors u and v are binary, from which the equivalence with problem (MBP) easily follows. Penalizing (quadratically) the equality constraints in the objective, we obtain

$$P_d(u, v) = \sum_{(i, j) \in F} (1 - u_i v_j)^2 + d \sum_{(i, j) \in \bar{F}} (u_i v_j)^2,$$

where $d \geq 0$ is the penalty parameter. We now observe that our choice of W at the beginning of this section gives $P_d(u, v) = \|M - uv^T\|_W^2$, i.e., (W-1d) is exactly equivalent to minimizing $P_d(u, v)$. This implies that, as d grows, minimizers of problem (W-1d) will tend to solutions of the biclique problem (MBP). Our goal is now to prove a more precise statement about the link between these two problems: we provide (in Lemma 3) an explicit value for d that guarantees a small difference between the optimal values of these two problems.

First, we establish that for any (u, v) such that $\|M - uv^T\|_W^2 \leq |E|$, the absolute value of the row or the column of uv^T corresponding to a zero entry of M must be smaller than a constant inversely proportional to $\sqrt[4]{d}$.

Lemma 1. *Let (i, j) be such that $M_{ij} = 0$, then $\forall (u, v)$ such that $\|M - uv^T\|_W^2 \leq |E|$,*

$$\min \left(\max_{1 \leq k \leq n} |u_i v_k|, \max_{1 \leq p \leq m} |u_p v_j| \right) \leq \sqrt[4]{\frac{4|E|^2}{d}}.$$

Proof. Without loss of generality u and v can be scaled such that $\|u\|_2 = \|v\|_2$ without changing the product uv^T , i.e., we replace u by $u' = \sqrt{\frac{\|v\|_2}{\|u\|_2}} u$ and v by $v' = \sqrt{\frac{\|u\|_2}{\|v\|_2}} v$ so that $\|u'\|_2 = \|v'\|_2 = \sqrt{\|u\|_2 \|v\|_2}$ and $u'v'^T = uv^T$. First, observe that since $\|\cdot\|_W$ is a norm,

$$\|uv^T\|_W - \sqrt{|E|} = \|uv^T\|_W - \|M\|_W \leq \|M - uv^T\|_W \leq \sqrt{|E|}.$$

Since all entries of W are larger than 1 ($d \geq 1$), we have

$$\|u\|_2 \|v\|_2 = \|uv^T\|_F \leq \|uv^T\|_W \leq \sqrt{4|E|},$$

and then $\|u\|_2 = \|v\|_2 \leq \sqrt[4]{4|E|}$.

Moreover $d(0 - u_i v_j)^2 \leq \|M - uv^T\|_W^2 \leq |E|$, so that $|u_i v_j| \leq \sqrt{\frac{|E|}{d}}$ which implies that either $|u_i| \leq \sqrt[4]{\frac{|E|}{d}}$ or $|v_j| \leq \sqrt[4]{\frac{|E|}{d}}$. Combining the above inequalities with the fact that $(\max_{1 \leq k \leq n} |v_k|)$ and $(\max_{1 \leq p \leq m} |u_p|)$ are bounded above by $\|u\|_2 = \|v\|_2 \leq \sqrt[4]{4|E|}$ completes the proof. \square

We now prove the following general lemma which, combined with Lemma 1 above, will allow us to derive a lower bound on the objective function of (W-1d) (it will also be used for the proof of the problem with missing data in Section 3.3).

Lemma 2. Let $M \in \{0, 1\}^{m \times n}$ be the biadjacency matrix of a bipartite graph $G_b = (V, E)$, $W \in \mathbb{R}_+^{m \times n}$ a weight matrix such that $W_{ij} = 1$ for each pair (i, j) satisfying $M_{ij} = 1$, and (u, v) be such that

$$\min \left(\max_{1 \leq k \leq n} |u_i v_k|, \max_{1 \leq p \leq m} |u_p v_j| \right) \leq c, \quad (3.2)$$

for each pair (i, j) satisfying $M_{ij} = 0$, where $0 < c \leq 1$. Let also $p = |E| - |E^*|$ be the optimal objective function value of (MBP). Then, if $p > 0$, we have

$$\|M - uv^T\|_W > p(1 - 2c).$$

Proof. Define the biclique corresponding to the following set $\Omega_c(u, v) \subseteq \{1, 2, \dots, m\} \times \{1, 2, \dots, n\}$

$$\Omega_c(u, v) = \{i \mid \exists j \text{ s.t. } |u_i v_j| > c\} \times \{j \mid \exists i \text{ s.t. } |u_i v_j| > c\}.$$

This biclique is part of the original graph, i.e., every edge in $\Omega_c(u, v)$ belongs to G_b . Indeed, if $M_{ij} = 0$, the pair (i, j) cannot belong to $\Omega_c(u, v)$ since, by Equation (3.2), the absolute value of either the i^{th} row or the j^{th} column of uv^T is smaller than c . By construction, we also have that the entries M_{ij} corresponding to pairs (i, j) not in the biclique $\Omega_c(u, v)$ are approximated by values smaller than c . The error corresponding to a unit entry of M_{ij} not in the biclique $\Omega_c(u, v)$ is then at least $(1 - c)^2$ (because the corresponding weight W_{ij} is equal to one). Since there are at least $p = |E| - |E^*|$ such entries (because there are $|E|$ unit entries in M and at most $|E^*|$ pairs in biclique $\Omega_c(u, v)$), we have

$$\|M - uv^T\|_W^2 \geq (1 - c)^2 p > p(1 - 2c) = p - 2pc.$$

□

We can now provide lower and upper bounds on the optimal value p^* of (W-1d), and show that it is not too different from the optimal value $|E| - |E^*|$ of (MBP).

Lemma 3. Let $0 < \epsilon \leq 1$. For any value of parameter d such that $d \geq \frac{2^6 |E|^6}{\epsilon^4}$, the optimal value p^* of (W-1d) satisfies

$$|E| - |E^*| - \epsilon < p^* \leq |E| - |E^*|.$$

Proof. Let (u, v) be an optimal solution of (W-1d) (since $W > 0$, there always exists at least one optimal solution, cf. Section 2), and let us note $p = |E| - |E^*| \geq 0$. If $p = 0$, then $p^* = 0$ and the result is trivial (it is the case when the rank of M is one, i.e., G_b contains only one biclique). Otherwise, since any optimal solution of (MBP) plugged in (W-1d) achieves an objective function equal to p , we must have

$$p^* = \|M - uv^T\|_W^2 \leq p = |E| - |E^*|,$$

which gives the upper bound.

Since d is greater than $4|E|^2$ for any $0 < \epsilon \leq 1$, the constant $\alpha = \sqrt[4]{\frac{4|E|^2}{d}}$ appearing in Lemma 1 is smaller than one. This means that Lemma 2 is applicable, so that we have

$$\|M - uv^T\|_W^2 > p - 2\alpha p \geq p - 2\alpha |E| \geq p - \epsilon,$$

which gives the lower bound (the last inequality follows from the fact that $2\alpha |E| \leq \epsilon$ is equivalent to the condition $d \geq \frac{2^6 |E|^6}{\epsilon^4}$). □

This result implies that for $\epsilon = 1$, i.e., for $d \geq (2|E|)^6$, we have $|E| - |E^*| - 1 < p^* \leq |E| - |E^*|$, and therefore computing p^* exactly would allow to recover $|E^*|$ (since $|E^*| = |E| - \lceil p^* \rceil$), which is NP-hard. Since the reduction from (MBP) to (W-1d) is polynomial (it uses the same matrix M and a weight matrix W whose description has polynomial length), we conclude that solving (W-1d) exactly is NP-hard. The next result shows that even solving (W-1d) approximately is NP-hard.

Corollary 1. For any $d \geq (2mn)^6$, $M \in \{0,1\}^{m \times n}$ and $W \in \{1,d\}^{m \times n}$, it is NP-hard to find an approximate solution of rank-one (WLRA) with objective function accuracy less than $1 - \frac{(2mn)^{3/2}}{d^{1/4}}$.

Proof. Let $d \geq (2mn)^6$, $0 < \epsilon = \frac{(2mn)^{3/2}}{d^{1/4}} < 1$, and (\bar{u}, \bar{v}) be an approximate solution of (W-1d) with objective function accuracy $(1 - \epsilon)$, i.e., $p^* \leq \bar{p} = \|M - \bar{u}\bar{v}^T\|_W^2 \leq p^* + 1 - \epsilon$. Since $d = \frac{(2mn)^6}{\epsilon^4} \geq \frac{(2|E|)^6}{\epsilon^4}$, Lemma 3 applies and we have

$$|E| - |E^*| - \epsilon < p^* \leq \bar{p} \leq p^* + 1 - \epsilon \leq |E| - |E^*| + 1 - \epsilon.$$

We finally observe that knowing \bar{p} allows to recover $|E^*|$, which is NP-hard. In fact, adding ϵ to the above inequalities gives $|E| - |E^*| < \bar{p} + \epsilon \leq |E| - |E^*| + 1$, and therefore

$$|E^*| = |E| - \lceil \bar{p} + \epsilon \rceil + 1.$$

□

We are now in position to prove Theorem 1, which deals with the hardness of rank-one (WLRA) with bounded weights.

Theorem 1. Let us use Corollary 1 with $W \in \{1,d\}^{m \times n}$, and define $W' = \frac{1}{d}W \in \{\frac{1}{d}, 1\}^{m \times n}$. Clearly, replacing W by W' in (W-1d) simply amounts to multiplying the objective function by $\frac{1}{d}$, with $\|M - uv^T\|_{W'}^2 = \frac{1}{d}\|M - uv^T\|_W^2$. Taking $d^{1/4} = 2(2mn)^{3/2}$ in Corollary 1, we obtain that for $M \in \{0,1\}^{m \times n}$ and $W \in [0,1]^{m \times n}$, it is NP-hard to find an approximate solution of rank-one (WLRA) with objective function accuracy less than $\frac{1}{d} \left(1 - \frac{(2mn)^{3/2}}{d^{1/4}}\right) = \frac{1}{2d} = 2^{-11}(mn)^{-6}$. □

Remark 2. The above bounds on d have been crudely estimated, and can be improved. Our main goal here was to show existence of a polynomial-time reduction from (MBP) to rank-one (WLRA).

Remark 3. Using the same construction as in [11, Theorem 3], this rank-one NP-hardness result can be generalized to any factorization rank, i.e., approximate (WLRA) for any fixed rank r is NP-hard. The idea is the following: given a bipartite graph G_b with biadjacency matrix $M \in \{0,1\}^{m \times n}$, we construct a larger bipartite graph G'_b which is made of r disconnected copies of G_b , whose biadjacency matrix is therefore given by

$$M' = \begin{pmatrix} M & \mathbf{0}_{m \times n} & \dots & \mathbf{0}_{m \times n} \\ \mathbf{0}_{m \times n} & M & & \mathbf{0}_{m \times n} \\ \vdots & & \ddots & \vdots \\ \mathbf{0}_{m \times n} & \dots & & M \end{pmatrix} \in \{0,1\}^{rm \times rn}.$$

Clearly, no biclique in this graph can be larger than a maximum biclique in G_b , and there are (at least) r disjoint bicliques with such maximum size in G'_b . Letting $(U, V) \in \mathbb{R}^{rm \times r} \times \mathbb{R}^{r \times rn}$ be an optimal solution of the rank- r (WLRA) problem with M' above and weights $W' = M' + d(\mathbf{1}_{rm \times rn} - M')$ defined as before, it can be shown that, for d sufficiently large, each rank-one factor $U_{:,k}V_{:,k}^T$ must correspond to a maximum biclique of G_b .

3.3 Low-Rank Matrix Approximation with Missing Data

The above NP-hardness proof does not cover the case when W is binary, corresponding to missing data in the matrix to be approximated (or to low-rank matrix completion with noise). This corresponds to the following problem

$$\inf_{U \in \mathbb{R}^{m \times r}, V \in \mathbb{R}^{n \times r}} \|M - UV^T\|_W^2 = \sum_{ij} W_{ij} (M - UV^T)_{ij}^2, \quad W \in \{0,1\}^{m \times n}.$$

In the same spirit as before, we consider the following rank-one version of the problem

$$p^* = \inf_{u \in \mathbb{R}^m, v \in \mathbb{R}^n} \|M - uv^T\|_W^2, \quad (\text{MD-1d})$$

with input data matrices M and W defined as follows

$$M = \left(\begin{array}{c|c} M_b & \mathbf{0}_{s \times Z} \\ \hline \mathbf{0}_{Z \times t} & dI_Z \end{array} \right) \text{ and } W = \left(\begin{array}{c|c} \mathbf{1}_{s \times t} & B_1 \\ \hline B_2 & I_Z \end{array} \right),$$

where $M_b \in \{0,1\}^{s \times t}$ is the biadjacency matrix of the bipartite graph $G_b = (V, E)$, $d > 1$ is a parameter, $Z = st - |E|$ is the number of zero entries in M_b , and $m = s + Z$ and $n = t + Z$ are the dimensions of M and W .

Binary matrices $B_1 \in \{0,1\}^{s \times Z}$ and $B_2 \in \{0,1\}^{Z \times t}$ are constructed as follows: assume the Z zero entries of M_b can be enumerated as

$$\{M_b(i_1, j_1), M_b(i_2, j_2), \dots, M_b(i_Z, j_Z)\},$$

and let k_{ij} be the (unique) index k ($1 \leq k \leq Z$) such that $(i_k, j_k) = (i, j)$ (therefore k_{ij} is only defined for pairs (i, j) such that $M_b(i, j) = 0$, and establishes a bijection between these pairs and the set $\{1, 2, \dots, Z\}$). We now define matrices B_1 and B_2 as follows: for every index $1 \leq k_{ij} \leq Z$, we have

$$B_1(i, k_{ij}) = 1, B_1(i', k_{ij}) = 0 \ \forall i' \neq i \text{ and } B_2(k_{ij}, j) = 1, B_2(k_{ij}, j') = 0 \ \forall j' \neq j.$$

Equivalently, each column of B_1 (resp. row of B_2) corresponds to a different zero entry $M_b(i, j)$, and contains only zeros except for a one at position i within the column (resp. at position j within the row). Hence the matrix B_1 (resp. B_2) contains only zero entries except Z entries equal to one, one in each column (resp. row).

In the case of Example 1, we get

$$M = \left(\begin{array}{ccc|c} 1 & 0 & 1 & \\ 0 & 1 & 1 & \mathbf{0}_{3 \times 2} \\ 1 & 1 & 1 & \\ \hline \mathbf{0}_{2 \times 3} & & & dI_2 \end{array} \right) \text{ and } W = \left(\begin{array}{ccc|cc} & & & 1 & 0 \\ & & & 0 & 1 \\ & & & 0 & 0 \\ \hline 0 & 1 & 0 & & \\ 1 & 0 & 0 & I_2 & \end{array} \right),$$

i.e., the matrix to be approximated can be represented as

$$\left(\begin{array}{ccc|cc} 1 & 0 & 1 & 0 & ? \\ 0 & 1 & 1 & ? & 0 \\ 1 & 1 & 1 & ? & ? \\ \hline ? & 0 & ? & d & ? \\ 0 & ? & ? & ? & d \end{array} \right). \quad (3.3)$$

For any feasible solution (u, v) of (MD-1d), we also note

$$u = \begin{pmatrix} u^{(b)} \\ u^{(d)} \end{pmatrix} \in \mathbb{R}^m, \ u^{(b)} \in \mathbb{R}^s \text{ and } u^{(d)} \in \mathbb{R}^Z,$$

$$v = \begin{pmatrix} v^{(b)} \\ v^{(d)} \end{pmatrix} \in \mathbb{R}^n, \ v^{(b)} \in \mathbb{R}^t \text{ and } v^{(d)} \in \mathbb{R}^Z.$$

We will show that this formulation ensures that, as d increases, the zero entries of matrix M_b (the biadjacency matrix of G_b which appears as the upper left block of matrix M) have to be approximated with smaller values. Hence, as for (W-1d), we will be able to prove that the optimal value p^* of

(MD-1d) will have to get close to the minimal value $|E| - |E^*|$ of (MBP), implying NP-hardness of its computation.

Intuitively, when d is large, the lower right matrix dI_Z of M will have to be approximated by a matrix with large diagonal entries, since they are weighted by unit entries in matrix W . Hence $u_{k_{ij}}^{(d)}v_{k_{ij}}^{(d)}$ has to be large for all $1 \leq k_{ij} \leq Z$. We then have at least either $u_{k_{ij}}^{(d)}$ or $v_{k_{ij}}^{(d)}$ large for all k_{ij} (recall that each k_{ij} corresponds to a zero entry in M at position (i, j) , cf. definition of B_1 and B_2 above). By construction, we also have two entries $M(s + k_{ij}, j) = 0$ and $M(i, t + k_{ij}) = 0$ with unit weights corresponding to the nonzero entries $B_1(i, k_{ij})$ and $B_2(k_{ij}, j)$, which then also have to be approximated by small values. If $u_{k_{ij}}^{(d)}$ (resp. $v_{k_{ij}}^{(d)}$) is large, then $v_j^{(b)}$ (resp. $u_i^{(b)}$) will have to be small since $u_{k_{ij}}^{(d)}v_j^{(b)} \approx 0$ (resp. $u_i^{(b)}v_{k_{ij}}^{(d)} \approx 0$). Finally, either $u_i^{(b)}$ or $v_j^{(b)}$ has to be small, implying that $M_b(i, j)$ is approximated by a small value, because $(u^{(b)}, v^{(b)})$ can be bounded independently of the value of d .

We now proceed as in Section 3.2. Let us first give an upper bound for the optimal value p^* of (MD-1d).

Lemma 4. *For $d > 1$, the optimal value p^* of (MD-1d) is bounded above by $|E| - |E^*|$, i.e.,*

$$p^* = \inf_{u \in \mathbb{R}^m, v \in \mathbb{R}^n} \|M - uv^T\|_W^2 \leq |E| - |E^*|. \quad (3.4)$$

Proof. Let us build the following feasible solution (u, v) of (MD-1d): $(u^{(b)}, v^{(b)})$ is a (binary) optimal solution of (MBP) and $(u^{(d)}, v^{(d)})$ is defined as⁷

$$u_{k_{ij}}^{(d)} = \begin{cases} d^K & \text{if } v_j^{(b)} = 0, \\ d^{1-K} & \text{if } v_j^{(b)} = 1, \end{cases} \quad \text{and} \quad v_{k_{ij}}^{(d)} = \begin{cases} d^{1-K} & \text{if } v_j^{(b)} = 0, \\ d^K & \text{if } v_j^{(b)} = 1, \end{cases} \quad (3.5)$$

where K is a real parameter and k_{ij} is the index of the column of B_1 and the row of B_2 corresponding to the zero entry (i, j) of M_b (i.e., $(i, j) = (i_{k_{ij}}, j_{k_{ij}})$).

We have that

$$(uv^T) \circ W = \begin{pmatrix} u^{(b)}v^{(b)T} & D_1 \\ D_2 & dI_Z \end{pmatrix},$$

where \circ is the component-wise (or Hadamard) product between two matrices, and matrices D_1 and D_2 satisfy

$$\begin{cases} D_i(l, p) &= 0 & \text{if } B_i(l, p) = 0, \\ D_i(l, p) &\in \{0, d^{1-K}\} & \text{if } B_i(l, p) = 1, \end{cases} \quad i = 1, 2.$$

In fact, let us analyze the four blocks of $(uv^T) \circ W$:

1. Upper-left: the upper-left block of W and uv^T are respectively the all-one matrix and $u^{(b)}v^{(b)T}$.
2. Lower-right: since the lower-right block of W is the identity matrix, we only need to consider the diagonal entries of the lower-right block of uv^T , which are given by $u_{k_{ij}}^{(d)}v_{k_{ij}}^{(d)} = d$ for $k_{ij} = 1, 2, \dots, Z$, cf. Equation (3.5).
3. Upper-right and lower-left: by definition of B_1 and B_2 , the only entries of D_1 and D_2 which may be different from zero are given by

$$D_1(i, k_{ij}) = u_i^{(b)}v_{k_{ij}}^{(d)} \quad \text{and} \quad D_2(k_{ij}, j) = u_{k_{ij}}^{(d)}v_j^{(b)},$$

⁷Notice that this construction is not symmetric, and the variant using $u^{(b)}$ instead of $v^{(b)}$ to define $u^{(d)}$ and $v^{(d)}$ is also possible.

for all (i, j) such that $M_b(i, j) = 0$. By construction, we have either $v_j^{(b)} = 0$ or $v_j^{(b)} = 1$. If $v_j^{(b)} = 0$, then $v_{k_{ij}}^{(d)} = d^{1-K}$ by Equation (3.5) and we have $D_1(i, k_{ij}) = u_i^{(b)} v_{k_{ij}}^{(d)} \in \{0, d^{1-K}\}$ and $D_2(k_{ij}, j) = 0$. If $v_j^{(b)} = 1$, we have $u_i^{(b)} = 0$ (since $M_b(i, j) = 0$) and $u_{k_{ij}}^{(d)} = d^{1-K}$ by Equation (3.5) whence $D_1(i, k_{ij}) = 0$ and $D_2(k_{ij}, j) = d^{1-K}$.

Finally, D_1 and D_2 have at most Z non-zero entries (recall Z is the number of zero entries in M_b), which are all equal to d^{1-K} ; therefore,

$$p^* \leq \|M - uv^T\|_W^2 \leq |E| - |E^*| + 2Zd^{2(1-K)}, \quad \forall K. \quad (3.6)$$

Since $d > 1$, taking the limit $K \rightarrow +\infty$ gives the result. \square

Example 3. Let us illustrate the construction of Lemma 4 on the matrix from Example 1, which contains two maximum bicliques with 4 edges, including the one corresponding to $u^{(b)} = (0, 1, 1)^T$ and $v^{(b)} = (0, 1, 1)^T$. Taking $u = (0, 1, 1, d^{1-K}, d^K)^T$ and $v = (0, 1, 1, d^K, d^{1-K})^T$, we obtain

$$\left(\begin{array}{ccc|cc} 1 & 0 & 1 & 0 & ? \\ 0 & 1 & 1 & ? & 0 \\ 1 & 1 & 1 & ? & ? \\ \hline ? & 0 & ? & d & ? \\ 0 & ? & ? & ? & d \end{array} \right) \approx uv^T = \left(\begin{array}{ccc|cc} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & d^K & d^{1-K} \\ 0 & 1 & 1 & d^K & d^{1-K} \\ \hline 0 & d^{1-K} & d^{1-K} & d & d^{2(1-K)} \\ 0 & d^K & d^K & d^{2K} & d \end{array} \right),$$

with $\|M - uv^T\|_W^2 = (|E| - |E^*|) + 2d^{2(1-K)} = 3 + 2d^{2(1-K)}$, which is less than the bound $3 + 4d^{2(1-K)}$ guaranteed by Equation (3.6).

We now prove a property similar to Lemma 1 for any solution with objective value smaller than $|E|$.

Lemma 5. Let $d > \sqrt{|E|}$ and (i, j) be such that $M_b(i, j) = 0$, then the following holds for any pair (u, v) such that $\|M - uv^T\|_W^2 \leq |E|$:

$$\min \left(\max_{1 \leq k \leq n} |u_i v_k|, \max_{1 \leq p \leq m} |u_p v_j| \right) \leq \frac{\sqrt{2}|E|^{\frac{3}{4}}}{(d - \sqrt{|E|})^{\frac{1}{2}}}. \quad (3.7)$$

Proof. Without loss of generality we set $\|u^{(b)}\|_2 = \|v^{(b)}\|_2$ by scaling u and v without changing uv^T . Observing that

$$\begin{aligned} \|u^{(b)}\|_2 \|v^{(b)}\|_2 - \sqrt{|E|} &= \|u^{(b)} v^{(b)T}\|_F - \|M_b\|_F \leq \|M_b - u^{(b)} v^{(b)T}\|_F \\ &\leq \|M - uv^T\|_W \leq \sqrt{|E|}, \end{aligned}$$

we have $\|u^{(b)}\|_2 \|v^{(b)}\|_2 \leq 2\sqrt{|E|}$, and $\|u^{(b)}\|_2 = \|v^{(b)}\|_2 \leq \sqrt{2}|E|^{\frac{1}{4}}$.

Assume $M_b(i, j)$ is zero for some pair (i, j) and let $k = k_{ij}$ denote the index of the corresponding column of B_1 and row of B_2 (i.e., such that $B_1(i, k) = B_2(k, j) = 1$). By construction, $u_k^{(d)} v_k^{(d)}$ has to approximate d in the objective function. This implies $(u_k^{(d)} v_k^{(d)} - d)^2 \leq |E|$ and then

$$u_k^{(d)} v_k^{(d)} \geq d - \sqrt{|E|} > 0.$$

Suppose $|u_k^{(d)}|$ is greater than $|v_k^{(d)}|$ (the case where $|v_k^{(d)}|$ is greater than $|u_k^{(d)}|$ is similar), which implies $|u_k^{(d)}| \geq (d - |E|^{\frac{1}{2}})^{\frac{1}{2}}$. Moreover, since $B_2(k, j)$ is a unit weight, we have that $u_k^{(d)} v_j$ has to approximate zero in the objective function, implying

$$(u_k^{(d)} v_j - 0)^2 \leq |E| \quad \Rightarrow \quad |u_k^{(d)} v_j| \leq \sqrt{|E|}.$$

Hence

$$|v_j| \leq \frac{\sqrt{|E|}}{|u_k^{(d)}|} \leq \frac{|E|^{\frac{1}{2}}}{(d - \sqrt{|E|})^{\frac{1}{2}}}, \quad (3.8)$$

and since $(\max_{1 \leq p \leq m} |u_p|)$ is bounded by $\|u^{(b)}\|_2 \leq \sqrt{2}|E|^{\frac{1}{4}}$, the proof is complete. \square

Using Lemma 2, we can now derive a lower bound for the value of p^* .

Lemma 6. *Let $0 < \epsilon \leq 1$. For any value of parameter d strictly greater than $\frac{8|E|^{\frac{7}{2}}}{\epsilon^2} + |E|^{\frac{1}{2}}$, the infimum p^* of (MD-1d) satisfies*

$$|E| - |E^*| - \epsilon < p^*.$$

Proof. Let us note $p = |E| - |E^*|$. If $p = 0$, the result is trivial since $p^* = 0$. Otherwise, suppose $p^* \leq p - \epsilon$ and let $\beta = \frac{\sqrt{2}|E|^{\frac{3}{4}}}{(d - \sqrt{|E|})^{\frac{1}{2}}}$. First observe that $d > \frac{8|E|^{\frac{7}{2}}}{\epsilon^2} + |E|^{\frac{1}{2}}$ is equivalent to $2|E|\beta < \epsilon$.

Then, by continuity of (MD-1d), for any δ such that $\delta < \epsilon$, there exists a pair (u, v) such that

$$\|M_b - u^{(b)}v^{(b)T}\|_W^2 \leq \|M - uv^T\|_W^2 \leq p - \delta \leq |E|.$$

In particular, let us take $\delta = 2|E|\beta < \epsilon$. Observe that $\beta \leq 1$ as soon as $d \geq 2|E|^{\frac{3}{2}} + |E|^{\frac{1}{2}}$ (which is guaranteed because $0 < \epsilon \leq 1$). By Lemma 5 and Lemma 2 (applied on matrix M_b and the solution $(u^{(b)}, v^{(b)})$), we then have

$$p - 2\beta p < \|M_b - u^{(b)}v^{(b)T}\|_W^2 \leq \|M - uv^T\|_W^2 \leq p - \delta.$$

Dividing the above inequalities by $p > 0$, we obtain

$$1 - 2\beta < 1 - \frac{\delta}{p} < 1 - \frac{\delta}{|E|} \Rightarrow \delta < 2|E|\beta,$$

a contradiction. \square

Corollary 2. *For any $d > 8(mn)^{7/2} + \sqrt{mn}$, $M \in \{0, 1, d\}^{m \times n}$, and $W \in \{0, 1\}^{m \times n}$, it is NP-hard to find an approximate solution of rank-one (WLRA) with objective function accuracy $1 - \frac{2\sqrt{2}(mn)^{7/4}}{(d - \sqrt{mn})^{1/2}}$.*

Proof. Let $d > 8(mn)^{7/2} + \sqrt{mn}$, $0 < \epsilon = \frac{2\sqrt{2}(mn)^{7/4}}{(d - \sqrt{mn})^{1/2}} < 1$, and (\bar{u}, \bar{v}) be an approximate solution of (W-1d) with absolute error $(1 - \epsilon)$, i.e., $p^* \leq \bar{p} = \|M - \bar{u}\bar{v}^T\|_W^2 \leq p^* + 1 - \epsilon$. Lemma 6 applies because $d = \frac{8(mn)^{7/2}}{\epsilon^2} + \sqrt{mn} \geq \frac{8(st)^{7/2}}{\epsilon^2} + \sqrt{st} \geq \frac{8|E|^{7/2}}{\epsilon^2} + |E|^{1/2}$. Using Lemmas 4 and 6, the rest of the proof is identical as the one of Theorem 1. Since the reduction from (MBP) to (MD-1d) is polynomial (description of matrices W and M has polynomial length, since the increase in matrix dimensions from M_b to M is polynomial), we conclude that finding such an approximate solution for (MD-1d) is NP-hard. \square

We can now easily derive Theorem 2, which deals with the hardness of rank-one (WLRA) with a bounded matrix M .

Theorem 2. Replacing M by $M' = \frac{1}{d}M$ in (MD-1d) gives an equivalent problem with objective function multiplied by $\frac{1}{d^2}$, since $\frac{1}{d^2}\|M - uv^T\|_W^2 = \|M' - \frac{uv^T}{d}\|_W^2$. Taking $d = 2^5(mn)^{7/2} + \sqrt{mn}$ in Corollary 2, we find that it is NP-hard to compute an approximate solution of rank-one (WLRA) for $M \in [0, 1]^{m \times n}$ and $W \in \{0, 1\}^{m \times n}$, and with objective function accuracy less than $\frac{1}{d^2} \left(1 - \frac{2\sqrt{2}(mn)^{7/4}}{(d - \sqrt{mn})^{1/2}}\right) = \frac{1}{2d^2} \geq 2^{-12}(mn)^{-7}$. \square

4 Concluding Remarks

In this paper, we have studied the complexity of the weighted low-rank approximation problem (WLRA), and proved that computing an approximate solution with some prescribed accuracy is NP-hard, already in the rank-one case, both for positive and binary weights (the latter also corresponding to low-rank matrix completion with noise, or PCA with missing data).

The following more general problem is sometimes also referred to as WLRA:

$$\inf_{U \in \mathbb{R}^{m \times r}, V \in \mathbb{R}^{r \times n}} \|M - UV\|_{(P)}^2, \quad (4.1)$$

where $\|A\|_{(P)}^2 = \text{vec}(A)^T P \text{vec}(A)$, with $\text{vec}(A)$ a vectorization of matrix A and P an mn -by- mn positive semidefinite matrix, see [25] and the references therein. Since our WLRA formulation corresponds to the special case of a diagonal (nonnegative) P , our hardness results also apply to Problem (4.1).

It is also worth pointing out that, when the data matrix M is nonnegative, any optimal solution to rank-one (WLRA) can be assumed to be nonnegative (see discussion for Example 1). Therefore, all the complexity results of this paper apply to the weighted nonnegative matrix factorization problem (weighted NMF), which is the following low-rank matrix approximation problem with nonnegativity constraints on the factors

$$\min_{U \in \mathbb{R}^{m \times r}, V \in \mathbb{R}^{n \times r}} \|M - UV^T\|_W^2 \quad \text{such that} \quad U \geq 0, V \geq 0.$$

Hence, it is NP-hard to find an approximate solution to *rank-one weighted NMF* (used, e.g., in image processing [14, Chapter 6]) and to *rank-one NMF with missing data* (used, e.g., for collaborative filtering [3]). This is in contrast with unweighted rank-one NMF, which is polynomially solvable (e.g., taking the absolute value of the first rank-one factor generated by the singular value decomposition). Note that (unweighted) NMF has been shown to be NP-hard when r is not fixed [29] (i.e., when r is part of the input).

Nevertheless, many questions remain open, including the following:

- ◊ Our approximation results are rather weak. In fact, they require the objective function accuracy to *increase* with the dimensions of the input matrix, in proportion with $(mn)^{-6}$, which is somewhat counter-intuitive. The reason is twofold: first, independently of the size of the matrix, we needed the objective function value of approximate solutions of problems (W-1d) and (MD-1d) to be no larger than the objective function of the optimal biclique solution plus one (in order to obtain $|E^*|$ by rounding). Second, parameter d in problems (W-1d) and (MD-1d) depends on the dimensions of matrix M . Therefore, when matrices W or M are rescaled between 0 and 1, the objective function accuracy is affected by parameter d , and hence decreases with the dimensions of matrix M . Strengthening of these bounds is a topic for further research.
- ◊ Moreover, as pointed out to us, these results say nothing about the hardness of approximation within a constant multiplicative factor. It would then be interesting to combine our reductions with inapproximability results for the biclique problem (which have yet to be investigated thoroughly, see, e.g., [28]), or construct reductions from other problems.
- ◊ When W is the matrix of all ones, WLRA can be solved in polynomial-time. We have shown that, when the ratio between the largest and the smallest entry in W is large enough, the problem is NP-hard (Theorem 1). It would be interesting to investigate the gap between these two facts, i.e., what is the minimum ratio between the entries of W that leads to an NP-hard WLRA problem?

- ◊ When $\text{rank}(W) = 1$, WLRA can be solved in polynomial-time (cf. Section 2) while it is NP-hard for a general matrix W (with rank up to $\min(m, n)$). What is the complexity of (WLRA) if the rank of the weight matrix W is fixed and greater than one, e.g., if $\text{rank}(W) = 2$?
- ◊ When data is missing, the rank-one matrix approximation problem is NP-hard in general. Nevertheless, it has been observed [1] that when the given entries are sufficiently numerous, well-distributed in the matrix, and affected by a relatively low level of noise, the original uncorrupted low-rank matrix can be recovered accurately, with a technique based on convex optimization (minimization of the nuclear norm of the approximation, which can be cast as a semidefinite program). It would then be particularly interesting to analyze the complexity of the problem given additional assumptions on the data matrix, for example on the noise distribution, and deal in particular with situations related to applications.

Acknowledgments

We thank Chia-Tche Chang for his helpful comments. We are grateful to the insightful comments of the three anonymous reviewers which helped to improve the paper substantially.

References

- [1] E.J. CANDÈS AND Y. PLAN, *Matrix Completion with Noise*, in Proceedings of the IEEE, 2009.
- [2] E.J. CANDÈS AND B. RECHT, *Exact Matrix Completion via Convex Optimization*, Foundations of Computational Mathematics, 9 (2009), pp. 717–772.
- [3] F. CHEN, G. WANG AND C. ZHANG, *Collaborative filtering using orthogonal nonnegative matrix tri-factorization*, Information Processing & Management, 45(3) (2009), pp. 368–379.
- [4] P. CHEN, *Optimization Algorithms on Subspaces: Revisiting Missing Data Problem in Low-Rank Matrix*, International Journal of Computer Vision, 80(1) (2008), pp. 125–142.
- [5] A.L. CHISTOV AND D.YU. GRIGORIEV, *Complexity of quantifier elimination in the theory of algebraically closed fields*, Proceedings of the 11th Symposium on Mathematical Foundations of Computer Science, Lecture Notes in Computer Science, Springer, 176 (1984), pp. 17–31.
- [6] P. COMON, *Independent component analysis, A new concept?*, Signal Processing, 36 (1994), pp. 287–314.
- [7] A. D’ASPREMONT, L. EL GHAOU, M.I. JORDAN, AND G.R.G. LANCKRIET, *A Direct Formulation for Sparse PCA Using Semidefinite Programming*, SIAM Rev., 49(3) (2007), pp. 434–448.
- [8] K.R. GABRIEL AND S. ZAMIR, *Lower Rank Approximation of Matrices by Least Squares With Any Choice of Weights*, Technometrics, 21(4) (1979), pp. 489–498.
- [9] M.R. GAREY AND D.S. JOHNSON, *Computers and Intractability: A guide to the theory of NP-completeness*, Freeman, San Francisco, 1979.
- [10] N. GILLIS AND F. GLINEUR, *Nonnegative Factorization and The Maximum Edge Biclique Problem*. CORE Discussion paper 2008/64, 2008.
- [11] ———, *Using underapproximations for sparse nonnegative matrix factorization*, Pattern Recognition, 43(4) (2010), pp. 1676–1687.
- [12] G.H. GOLUB AND C.F. VAN LOAN, *Matrix Computation, 3rd Edition*, The Johns Hopkins University Press Baltimore, 1996.
- [13] B. GRUNG AND R. MANNE, *Missing values in principal component analysis*, Chemom. and Intell. Lab. Syst., 42 (1998), pp. 125–139.
- [14] N.-D. HO, *Nonnegative Matrix Factorization - Algorithms and Applications*, PhD thesis, Université catholique de Louvain, 2008.

- [15] D. JACOBS, *Linear fitting with missing data for structure-from-motion*, Vision and Image Understanding, 82 (2001), pp. 57–81.
- [16] I.T. JOLLIFFE, *Principal Component Analysis*, Springer-Verlag, 1986.
- [17] Y. KOREN, R. BELL, AND C. VOLINSKY, *Matrix Factorization Techniques for Recommender Systems*, IEEE Computer, 42(8) (2009), pp. 30–37.
- [18] D.D. LEE AND H.S. SEUNG, *Learning the Parts of Objects by Nonnegative Matrix Factorization*, Nature, 401 (1999), pp. 788–791.
- [19] W.-S. LU, S.-C. PEI, AND P.-H. WANG, *Weighted low-rank approximation of general complex matrices and its application in the design of 2-D digital filters*, IEEE Trans. Circuits Syst. I, 44 (1997), pp. 650–655.
- [20] I. MARKOVSKY AND M. NIRANJAN, *Approximate low-rank factorization with structured factors*, Computational Statistics & Data Analysis, 54 (2010), pp. 3411–3420.
- [21] D. NISTER, F. KAHL, AND H. STEWENIUS, *Structure from Motion with Missing Data is NP-Hard*, in IEEE 11th Int. Conf. on Computer Vision, 2007.
- [22] J. NOCEDAL AND S.J. WRIGHT, *Numerical Optimization, Second Edition*, Springer, New York, 2006.
- [23] R. PEETERS, *The maximum edge biclique problem is NP-complete*, Discrete Applied Mathematics, 131(3) (2003), pp. 651–654.
- [24] B.M. SARWAR, G. KARYPIS, J.A. KONSTAN, AND J. RIEDL, *Item-Based Collaborative Filtering Recommendation Algorithms*, in 10th International WorldWideWeb Conference, 2001.
- [25] M. SCHUERMANS, *Weighted Low Rank Approximation: Algorithms and Applications*, PhD thesis, Katholieke Universiteit Leuven, 2006.
- [26] H. SHUM, K. IKEUCHI, AND R. REDDY, *Principal component analysis with missing data and its application to polyhedral object modeling*, IEEE Trans. Pattern Anal. Mach. Intelligence, 17(9) (1995), pp. 854–867.
- [27] N. SREBRO AND T. JAAKKOLA, *Weighted Low-Rank Approximations*, in 20th ICML Conference Proceedings, 2004.
- [28] J. TAN, *Inapproximability of maximum weighted edge biclique and its applications*, in Proceedings of the 5th Int. Conf. on Theory and Appl. of Models of Comput., 2008.
- [29] S.A. VAVASIS, *On the complexity of nonnegative matrix factorization*, SIAM J. on Optimization, 20 (2009), pp. 1364–1377.