New Hardness Results for Low-Rank Matrix Completion

Dror Chawin
The Academic College of Tel Aviv-Yaffo, Tel Aviv, Israel

Ishay Haviv
The Academic College of Tel Aviv-Yaffo, Tel Aviv, Israel

─── **Abstract** ───

The low-rank matrix completion problem asks whether a given real matrix with missing values can be completed so that the resulting matrix has low rank or is close to a low-rank matrix. The completed matrix is often required to satisfy additional structural constraints, such as positive semi-definiteness or a bounded infinity norm. The problem arises in various research fields, including machine learning, statistics, and theoretical computer science, and has broad real-world applications.

This paper presents new NP-hardness results for low-rank matrix completion problems. We show that for every sufficiently large integer d and any real number $\varepsilon \in [2^{-O(d)}, \frac{1}{7}]$, given a partial matrix A with exposed values of magnitude at most 1 that admits a positive semi-definite completion of rank d, it is NP-hard to find a positive semi-definite matrix that agrees with each given value of A up to an additive error of at most ε, even when the rank is allowed to exceed d by a multiplicative factor of $O(\frac{1}{\varepsilon^2 \cdot \log(1/\varepsilon)})$. This strengthens a result of Hardt, Meka, Raghavendra, and Weitz (COLT, 2014), which applies to multiplicative factors smaller than 2 and to ε that decays polynomially in d. We establish similar NP-hardness results for the case where the completed matrix is constrained to have a bounded infinity norm (rather than be positive semi-definite), for which all previous hardness results rely on complexity assumptions related to the Unique Games Conjecture. Our proofs involve a novel notion of nearly orthonormal representations of graphs, the concept of line digraphs, and bounds on the rank of perturbed identity matrices.

1 Introduction

In the matrix completion problem, the input is a partially observed real matrix, where some entries are marked by $\perp$ to indicate missing values. The goal is to fill in these values in such a way that the completed matrix satisfies certain prescribed properties. This algorithmic task is prevalent in various research fields, including machine learning, statistics, and theoretical computer science, and is motivated by a range of real-world applications, such as recommendation systems, medical imaging, computer vision, and signal processing. Typically, the completed matrix is required to have low rank or to be a slight perturbation of a low-rank matrix. In certain cases, the matrix may also need to be positive semi-definite or constrained by a bounded infinity norm.

50th International Symposium on Mathematical Foundations of Computer Science (MFCS 2025).
Editors: Paweł Gawrychowski, Filip Mazowiecki, and Michał Skrzypczak; Article No. 37; pp. 37:1–37:18
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

A central objective in the study of low-rank matrix completion is to identify the conditions under which the problem can be solved efficiently. A successful line of work, initiated by Candès and Recht [7], has applied convex optimization methods to efficiently recover a low-rank matrix from the values of a subset of its entries (see also [8, 36]). The recovery is guaranteed to succeed if (a) the number of given entries is sufficiently large, (b) the completed matrix satisfies a certain incoherence condition (i.e., the row and column spaces of the matrix are not aligned with the vectors of the standard basis), and (c) the subset of exposed entries is drawn uniformly at random. In fact, these conditions further ensure that the solution is unique. However, in most applications, the given entries cannot be chosen at random. It is therefore of interest to determine the computational complexity of recovering a low-rank incoherent matrix, when the observed entries are selected in a worst-case manner.

The hardness of low-rank matrix completion can be traced back to a 1996 paper by Peeters [33], who explored the complexity of determining a graph quantity known as minrank, introduced by Haemers [18] in the context of the Shannon capacity of graphs. Peeters' work implies that for every integer $d \geq 3$, deciding whether a given partial matrix can be completed to a matrix of rank at most d is NP-hard. His proof technique further implies that the same NP-hardness result holds when the completed matrix is required to be positive semi-definite, and this was extended to the case of $d = 2$ in 2013 by Eisenberg-Nagy, Laurent, and Varvitsiotis [14]. This research avenue was then extended by Hardt, Meka, Raghavendra, and Weitz [20], who explored two relaxations of the problem: first, allowing some slackness in the rank of the completed matrix, and second, permitting a bounded additive error on the given entries of the partial matrix. Specifically, they proved that for every integer $d \geq 6$, given a partial matrix A whose observed entries have magnitude at most 1, it is NP-hard to distinguish between the case where A admits a positive semi-definite completion of rank at most d, and the case in which any positive semi-definite completion of A has rank at least $2d$. Moreover, they showed that the problem remains NP-hard, when the matrix in the latter case is allowed to approximate each given value up to an additive error of ε, as long as ε decays polynomially with the desired rank, namely, for $\varepsilon = O(d^{-6})$. More recently, the paper [10] addressed the related problem of determining a graph measure known as the orthogonality dimension. The results of [10] imply that for every sufficiently large integer d, it is NP-hard to decide whether an input partial matrix can be completed to a positive semi-definite matrix of rank at most d, or any positive semi-definite completion has rank at least $2^{(1-o(1)) \cdot d/2}$, with the $o(1)$ term tending to 0 as d tends to infinity. It thus follows that it is NP-hard to approximate to within any constant factor the minimum possible rank of a positive semi-definite matrix that agrees with an input partial matrix. However, this hardness result does not apply to the more tolerant setting that allows an additive perturbation in each entry of the partial matrix.

Another version of the low-rank matrix completion problem considered by Hardt et al. [20] imposes a fixed bound on the infinity norm of the completed matrix (rather than requiring positive semi-definiteness). For this setting, their hardness results were not based on the standard assumption $\mathsf{P} \neq \mathsf{NP}$, but on the hardness of appropriate gap versions of the Coloring and Independent Set problems, the intractability of which is supported by certain variants of the Unique Games Conjecture [12, 13]. Under these assumptions, they showed that for all positive integers $d_2 > d_1 \geq 3$ and real numbers $\varepsilon \in [0, \frac{1}{2})$ and $\theta \geq 1$, there is no efficient algorithm to decide whether an input partial matrix A can be completed to one with an infinity norm at most θ and rank at most d_1, or any completion of A with an infinity norm at most θ must have rank at least d_2, even when an additive error of ε is allowed in each entry. Remarkably, this hardness result persists when a constant fraction of the matrix entries is revealed (as is the case for all the above hardness results) and, in addition, when

the completed matrix is required to be incoherent (in instances admitting a valid completion). In light of the aforementioned algorithmic results for the problem [7, 8, 36], these findings highlight the worst-case choice of the revealed entries as a substantial obstacle to efficient recovery.

1.1 Our Contribution

This paper presents several new hardness results for low-rank matrix completion problems. We begin by considering the case where the completed matrix is required to be positive semi-definite. Specifically, we study the gap problem (d_1, d_2, ε)-PSD-Completion, formally defined below. Here, we let $\mu(B)$ denote the coherence of a matrix B, a measure that is always bounded from below by 1 (see Definition 9).

▶ **Definition 1** (The (d_1, d_2, ε)-PSD-Completion Problem). *For positive integers $d_1 < d_2$ and for a real number $\varepsilon \in [0, 1)$, the (d_1, d_2, ε)-PSD-Completion problem asks, given a partial matrix $A \in ([-1, +1] \cup \{\bot\})^{n \times n}$, to distinguish between the following cases.*
- YES*: There exists a positive semi-definite matrix $B \in \mathbb{R}^{n \times n}$, such that $A_{i,j} = B_{i,j}$ for all $i, j \in [n]$ with $A_{i,j} \neq \bot$, $\mu(B) = 1$, and $\operatorname{rank}(B) \leq d_1$.*
- NO*: Every positive semi-definite matrix $B \in \mathbb{R}^{n \times n}$, such that $|A_{i,j} - B_{i,j}| \leq \varepsilon$ for all $i, j \in [n]$ with $A_{i,j} \neq \bot$, satisfies $\operatorname{rank}(B) \geq d_2$.*

Note that the definition restricts the magnitudes of the values in the input partial matrix to at most 1. Such a restriction is essential when allowing an additive error in the completed matrix, as rank is invariant under scaling.

We first point out that, under plausible complexity assumptions related to the Unique Games Conjecture [12, 13], the (d_1, d_2, ε)-PSD-Completion problem is intractable for all positive integers $d_2 > d_1 \geq 3$ and real numbers $\varepsilon \in [0, \frac{1}{2})$. Our primary contribution lies in establishing hardness results based solely on the more standard assumption $\mathsf{P} \neq \mathsf{NP}$, as stated below.

▶ **Theorem 2** (Simplified). *For every sufficiently large integer d and any real $\varepsilon \in [2^{-O(d)}, \frac{1}{7}]$, the $(d, O(\frac{d}{\varepsilon^2 \cdot \log(1/\varepsilon)}), \varepsilon)$-PSD-Completion problem is NP-hard.*

For admissible values of d and ε, Theorem 2 implies that given a partial matrix A with exposed values of magnitude at most 1, which admits a positive semi-definite completion with coherence 1 and rank d, it is NP-hard to find a positive semi-definite matrix that agrees with each given value of A up to an additive error of at most ε, even when the rank is allowed to exceed d by a multiplicative factor of $O(\frac{1}{\varepsilon^2 \cdot \log(1/\varepsilon)})$. The theorem encompasses various parameter settings of interest. First, for any fixed approximation factor $\alpha > 1$, there exists some constant $\varepsilon > 0$, for which the $(d, \alpha \cdot d, \varepsilon)$-PSD-Completion problem is NP-hard for any sufficiently large integer d. Next, letting ε decrease polynomially with d results in a hardness of approximation factor that is polynomial in d. Finally, setting $\varepsilon = 2^{-\Theta(d)}$ yields a hardness of approximation to within a factor of the form $2^{\Omega(d)}$. In fact, for an ε that decays sufficiently rapidly with d, we obtain the following refined hardness result.

▶ **Theorem 3** (Simplified). *For every sufficiently large integer d and any real $\varepsilon \in [0, 2^{-\Omega(d)}]$, the $(d, 2^{(1-o(1)) \cdot d/2}, \varepsilon)$-PSD-Completion problem is NP-hard.*

Theorems 2 and 3 substantially strengthen the previously known NP-hardness results for low-rank matrix completion in the positive semi-definite setting. As demonstrated above, our results offer flexibility in the choice of parameters, enabling us to establish hardness for several scenarios of interest. For comparison, the result of [20] is specific to ε decaying polynomially in the rank and achieves NP-hardness of approximation to within factors smaller

than 2. In contrast, for this regime of ε, Theorem 2 establishes NP-hardness of approximation to within factors that grow polynomially in the rank. Furthermore, all our hardness results hold even when the completed matrices of YES instances are required to have coherence 1, a property not guaranteed by the NP-hardness result of [20]. Finally, while the hardness result of [10] achieves the same gap as Theorem 3, it is restricted to the non-error setting of $\varepsilon = 0$.

We turn to our NP-hardness results for the low-rank matrix completion problem, where the completed matrix is constrained by a bounded infinity norm. Consider the gap $(d_1, d_2, \varepsilon, \theta)$-Completion problem, defined as follows.

▶ **Definition 4** (The $(d_1, d_2, \varepsilon, \theta)$-Completion Problem)**.** *For positive integers $d_1 < d_2$ and real numbers $\varepsilon \geq 0$ and $\theta \geq 1$, the $(d_1, d_2, \varepsilon, \theta)$-Completion problem asks, given a partial matrix $A \in ([-\theta, +\theta] \cup \{\perp\})^{n \times n}$, to distinguish between the following cases.*

- YES*: There exists a matrix $B \in [-\theta, +\theta]^{n \times n}$, such that $A_{i,j} = B_{i,j}$ for all $i, j \in [n]$ with $A_{i,j} \neq \perp$, $\mu(B) = 1$, and $\mathrm{rank}(B) \leq d_1$.*
- NO*: Every matrix $B \in [-\theta, +\theta]^{n \times n}$, such that $|A_{i,j} - B_{i,j}| \leq \varepsilon$ for all $i, j \in [n]$ with $A_{i,j} \neq \perp$, satisfies $\mathrm{rank}(B) \geq d_2$.*

Our hardness results for this problem are stated as follows.

▶ **Theorem 5** (Simplified)**.** *For every sufficiently large integer d and any real numbers $\varepsilon \in [2^{-O(d)}, \frac{1}{7}]$ and $\theta \in [1, 2^{2^{O(d)}}]$, the $(d, O(\frac{d}{\varepsilon^2 \cdot \log(1/\varepsilon)}), \varepsilon, \theta)$-Completion problem is NP-hard.*

▶ **Theorem 6** (Simplified)**.** *For every sufficiently large integer d and any real numbers $\varepsilon \in [0, 2^{-\Omega(d)}]$ and $\theta \in [1, 2^{2^{o(d)}}]$, the $(d, 2^{(1-o(1)) \cdot d/2}, \varepsilon, \theta)$-Completion problem is NP-hard.*

As mentioned earlier, the previously known hardness results for the $(d_1, d_2, \varepsilon, \theta)$-Completion problem, given in [20], rely on complexity assumptions stronger than the standard conjecture $\mathsf{P} \neq \mathsf{NP}$.

Our NP-hardness results are obtained via an efficient reduction from a gap coloring problem, whose hardness was proved by Krokhin, Opršal, Wrochna, and Zivný [28]. Their proof employed the concept of line digraphs, introduced by Harary and Norman [19] in 1960, which lies at the heart of the present paper as well. Specifically, we introduce extensions of the notions of orthogonality dimension and minrank of graphs (see Definitions 15 and 16), and as our main technical contribution, we show that for line digraphs, these quantities are intimately related to the chromatic number. The analysis involves bounds on the rank of perturbed identity matrices that were proved by Alon [2] in 2003 and have found diverse applications (see Theorem 12 and [3]). After establishing hardness results for our extensions of orthogonality dimension and minrank, we derive our hardness results for low-rank matrix completion problems. We believe that these novel graph quantities may be of independent interest, and we encourage their further study by pointing out a close relation to the notion of circular chromatic number, introduced by Vince [37] in 1988 (see Section 3.3).

1.2   Proof Technique

We provide here an overview of the techniques and ideas behind the proofs of our NP-hardness results for the low-rank matrix completion problem. For concreteness, we focus on the PSD-Completion problem, where the completed matrix is required to be positive semi-definite. We then briefly discuss the setting of a bounded infinity norm.

Our hardness proofs are anchored in the classic gap coloring problem. Recall that the chromatic number of a graph G, denoted $\chi(G)$, is the smallest number of colors needed for a vertex coloring of G in which adjacent vertices receive distinct colors. For fixed positive

integers $k_1 < k_2$, the (k_1, k_2)-Coloring problem asks, given an input graph G, to distinguish between the case where $\chi(G) \leq k_1$ and the case where $\chi(G) \geq k_2$. This problem is known to be intractable for all integers $k_2 > k_1 \geq 3$, under complexity assumptions related to the Unique Games Conjecture [12, 13]. However, identifying the values of k_1 and k_2 for which the problem is NP-hard has long been considered notoriously difficult. The current state of the art shows that for every integer $k_1 \geq 3$, the problem is NP-hard for $k_2 = 2k_1$, as proved by Barto, Bulín, Krokhin, and Opršal [5], and for $k_2 = \binom{k_1}{\lfloor k_1/2 \rfloor} = 2^{(1-o(1)) \cdot k_1}$, as proved by Krokhin, Opršal, Wrochna, and Zivný [28]. The latter result, which improves upon the former for all integers $k_1 \geq 6$, serves as the starting point of our hardness proofs.

The NP-hardness proof of Krokhin et al. [28] for the gap coloring problem is based on the concept of line digraphs [19], which allows to efficiently shrink the chromatic number of a given graph in a controlled manner. Specifically, for a graph G, let $\tilde{\delta}G$ denote the underlying graph of the line digraph of G, namely, the graph whose vertices are all the ordered pairs of adjacent vertices in G (whose number is twice the number of edges in G), where two vertices (u_1, u_2) and (v_1, v_2) are adjacent in $\tilde{\delta}G$ if $u_2 = v_1$ or $u_1 = v_2$ (see Definition 20). A result of Poljak and Rödl [34] shows that the chromatic number of the graph $\tilde{\delta}G$ is logarithmic in that of G (see Theorem 21). The hardness result of [28] was derived by repeatedly applying this transformation to instances of the $(k, 2^{O(k^{1/3})})$-Coloring problem, the NP-hardness of which was previously proved by Huang [25]. Since the appearance of [28], line digraphs have been effectively utilized in several hardness proofs, e.g., in [17, 10, 24], and they also form a key ingredient in the present paper.

To give a glimpse of the relationship between the chromatic numbers of a graph G and its associated graph $\tilde{\delta}G$, and as a gentle warm-up to our actual argument, let us briefly explain why $\chi(\tilde{\delta}G) \leq k$ implies that $\chi(G) \leq 2^k$. Indeed, given a proper k-coloring of $\tilde{\delta}G$, we define a coloring of G by assigning to each vertex v the set $c(v)$ of colors used at the vertices of the form $(\cdot, v)$ in $\tilde{\delta}G$ (i.e., vertices whose head is v). Clearly, the number of colors used does not exceed 2^k. To verify that the coloring is proper, observe that if u and v are adjacent vertices in G, the color of the vertex (u, v) in $\tilde{\delta}G$ lies in $c(v)$ but not in $c(u)$, ensuring that $c(u) \neq c(v)$. A slightly better upper bound on the chromatic number of G, along with a matching lower bound, is provided in [34].

A natural extension of graph colorings, originally proposed by Lovász [31] in the introduction of his renowned ϑ-function, is that of orthonormal representations of graphs. A d-dimensional orthonormal representation of a graph is an assignment of a unit vector in $\mathbb{R}^d$ to each vertex, such that the vectors assigned to adjacent vertices are orthogonal.[1] The orthogonality dimension of a graph G, denoted $\overline{\xi}(G)$, is the smallest integer d for which G admits a d-dimensional orthonormal representation (see Definition 15). Note that for every graph G, it holds that

$$\log_3 \chi(G) \leq \overline{\xi}(G) \leq \chi(G). \tag{1}$$

Indeed, for the upper bound on $\overline{\xi}(G)$, notice that a proper k-coloring of G yields a k-dimensional orthonormal representation by assigning the ith vector of the standard basis of $\mathbb{R}^k$ to the vertices of the ith color class. For the lower bound, consider a k-dimensional orthonormal representation of G, and observe that replacing its vectors by their sign vectors in $\{0, \pm 1\}^k$ results in a proper coloring of G with 3^k colors. For a construction of graphs for

---

[1] Strictly speaking, the definition in [31] requires vectors assigned to *non-adjacent* vertices to be orthogonal. This corresponds to an orthonormal representation of the complement graph in our terminology, which we adopt from, e.g., [33, 6] for convenience.

which the left-hand side of (1) is tight up to a multiplicative constant, see, e.g., [23]. Now, by associating with each orthonormal representation of G the Gram matrix of its vectors, it follows that $\overline{\xi}(G)$ is the smallest possible rank of a positive semi-definite matrix with ones on the diagonal and zeros in entries that correspond to pairs of adjacent vertices. This formulation naturally connects the problem of determining the orthogonality dimension of a graph to that of determining the minimum rank of a positive semi-definite completion of a suitably defined partial matrix.

The computational hardness of determining the orthogonality dimension of graphs was speculated by Lovász, Saks, and Schrijver [32] in 1989 and has been studied in several recent works, e.g., [6, 16, 10]. We first mention that one may combine the inequalities in (1) with the hardness results of the gap coloring problem from [12, 13] to conclude that deciding whether an input graph G satisfies $\overline{\xi}(G) \leq d_1$ or $\overline{\xi}(G) \geq d_2$ is intractable for all integers $d_2 > d_1 \geq 3$, assuming some variant of the Unique Games Conjecture. This reasoning, however, is insufficient to derive any NP-hardness result for orthogonality dimension, even from the strongest known NP-hardness of gap coloring. Still, it was shown in [10] that for every sufficiently large integer d, it is NP-hard to decide whether an input graph G satisfies $\overline{\xi}(G) \leq d$ or $\overline{\xi}(G) \geq 2^{(1-o(1)) \cdot d/2}$. This result was proved based on the hardness of the $(k, 2^{(1-o(1)) \cdot k})$-Coloring problem from [28] through the reduction that maps a given graph G to the graph $\tilde{\delta}G$. On the one hand, it follows from (1) that the orthogonality dimension of $\tilde{\delta}G$ does not exceed its chromatic number, which is logarithmic in $\chi(G)$. To complete the correctness of the reduction, it was shown in [10] that $\overline{\xi}(\tilde{\delta}G) \leq d$ implies that $\chi(G) \leq d^{O(d^2)}$. This, in turn, leads to the NP-hardness of the d vs. $2^{(1-o(1)) \cdot d/2}$ gap for the orthogonality dimension problem, as well as for the PSD-Completion problem when no error is allowed.

The argument in [10] was based on the idea outlined below. Suppose that $\overline{\xi}(\tilde{\delta}G) \leq d$, and consider a d-dimensional orthonormal representation of $\tilde{\delta}G$, which assigns to each ordered pair e of adjacent vertices in G a vector $x_e \in \mathbb{R}^d$. Associate with each vertex v of G the linear subspace $c(v) \subseteq \mathbb{R}^d$ spanned by the vectors of the form $x_{(\cdot, v)}$. Observe that if u and v are adjacent vertices in G, their subspaces $c(u)$ and $c(v)$ are quite distant from each other, in the sense that there exists a vector – the vector $x_{(u,v)}$ associated with the vertex (u, v) in $\tilde{\delta}G$ – that lies in $c(v)$ but is orthogonal to the entire subspace $c(u)$. Now, every subspace of $\mathbb{R}^d$ can be represented by an orthonormal basis of at most d vectors. To obtain a coloring of G with finitely many colors, the basis vectors are replaced by their representatives from a sufficiently dense net of the unit sphere in $\mathbb{R}^d$, resulting in a proper coloring of G with $d^{O(d^2)}$ colors, as desired.

In order to adapt this approach to the more tolerant setting of PSD-Completion, where an additive error of ε is allowed at each entry of the input partial matrix, we introduce the notion of d-dimensional ε-orthonormal representations of graphs. Here, each vertex of the graph G at hand is assigned a unit vector in $\mathbb{R}^d$, such that the inner product of vectors assigned to adjacent vertices is at most ε in absolute value. Letting $\overline{\xi}_\varepsilon(G)$ denote the smallest possible dimension of such an assignment, one can show that approximating the value of $\overline{\xi}_\varepsilon(G)$ for an input graph G is efficiently reducible to approximating the smallest rank of a positive semi-definite matrix that agrees with a given partial matrix, even when an additive error of $O(\varepsilon)$ is allowed per entry. To prove the NP-hardness of the former, we apply again the reduction that transforms a graph G into the graph $\tilde{\delta}G$. For correctness, we aim to establish an upper bound on $\chi(G)$ in terms of $\overline{\xi}_\varepsilon(\tilde{\delta}G)$. It turns out, though, that the proof idea of [10], described above, does not extend to this setting. To see why, consider a d-dimensional ε-orthonormal representation of $\tilde{\delta}G$, which assigns to each ordered pair e of adjacent vertices in G a vector $x_e \in \mathbb{R}^d$, and as before, associate with each vertex v of G the

linear subspace $c(v) \subseteq \mathbb{R}^d$ spanned by the vectors of the form $x_{(\cdot,v)}$. Now, let u and v be adjacent vertices in G. While the vector $x_{(u,v)}$ still lies in $c(v)$, it is no longer guaranteed to be orthogonal to the subspace $c(u)$. In fact, this vector is only guaranteed to have an inner product of at most ε in absolute value with the vectors of a basis of $c(u)$, which does not even preclude the possibility that it lies in $c(u)$, making it impossible to deduce that $c(u)$ and $c(v)$ are far apart. As an example, consider the two unit vectors e_1 and $\sqrt{1-\varepsilon^2} \cdot e_1 + \varepsilon \cdot e_2$, where e_i denotes the ith vector of the standard basis of $\mathbb{R}^d$. Notice that the vector e_2 has an inner product of at most ε in absolute value with each of the two vectors, yet it lies within the subspace they span.

We overcome the above difficulty through a different coloring of G. Namely, for each vertex v of G, we consider a maximal set $c(v) \subseteq \mathbb{R}^d$ of vectors of the form $x_{(\cdot,v)}$ with pairwise inner products of absolute value at most ε', for some appropriately chosen $\varepsilon' > \varepsilon$. Now, if u and v are adjacent vertices in G, then the vector $x_{(u,v)}$ has an inner product of at most ε in absolute value with each vector of $c(u)$. However, by the maximality of $c(v)$, there must exist a vector in $c(v)$ whose inner product with $x_{(u,v)}$ is larger than ε' in absolute value: either $x_{(u,v)}$ itself or some other vector that prevented it from being added to $c(v)$. This, in a sense, implies that the sets $c(u)$ and $c(v)$ are sufficiently far apart, so that they remain distinct once their vectors are replaced by representatives from a sufficiently dense net. Yet, the number of vectors in the sets $c(v)$ has a significant effect on the number of colors used. In contrast to the $\varepsilon = 0$ setting, here we do not handle bases of subspaces of $\mathbb{R}^d$, and therefore, we cannot ensure that the size of each set $c(v)$ is bounded by d. We do know that the vectors of each set $c(v)$ are nearly orthogonal to each other, with the absolute value of their pairwise inner products at most ε'. To bound their size, we apply bounds of Alon [2] on the rank of perturbed identity matrices (see Theorem 12). When ε' is sufficiently small, namely, $\varepsilon' \leq O(1/\sqrt{d})$, it turns out that each set $c(v)$ includes fewer than $2d$ vectors. For larger values of ε', the bound weakens, leading to the dependence of the hardness gap on the error ε, as stated in Theorem 2.

To extend our approach to the low-rank matrix completion problem with a bounded infinity norm, we introduce an extension of the notion of graph-fitting matrices, proposed by Haemers in [18]. Here, for a graph G, we consider matrices (not necessarily positive semi-definite) that have ones on the diagonal and values of magnitude at most ε in entries corresponding to adjacent vertices (see Definition 16). Our main technical result provides an upper bound on the chromatic number of a graph G, assuming that the graph $\tilde{\delta}G$ admits such a matrix with a bounded rank and a bounded infinity norm (see Theorem 22). The proof generalizes the technique described above, drawing on the fact that every matrix has a rank factorization involving matrices with rows of bounded norm (see Lemma 8). The detailed argument is presented in the subsequent technical sections.

1.3 Outline

The rest of the paper is organized as follows. In Section 2, we collect several notations and tools that will be used throughout the paper. In Section 3, we define and study nearly orthonormal representations of graphs, with particular attention to line digraphs. In Section 4, we apply our insights to establish the hardness of a problem we call Graph Fitness. This, in turn, leads to the hardness of the (d_1, d_2, ε)-PSD-Completion and $(d_1, d_2, \varepsilon, \theta)$-Completion problems, thereby verifying Theorems 2, 3, 5, and 6. Due to space limitations, some proofs and results are deferred to the full version of the paper.

2 Preliminaries

Throughout the paper, we omit floor and ceiling signs when they are not essential, and all logarithms are taken in base 2, unless otherwise specified. Graphs refer to undirected graphs, and digraphs refer to directed graphs, with all graphs and digraphs being simple (i.e., with no loops or parallel edges). For a positive integer n, we denote $[n] = \{1, \ldots, n\}$.

2.1 Linear Algebra

For a positive integer d, let $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ stand for the standard inner product and Euclidean norm on $\mathbb{R}^d$, respectively. As is customary, a vector $x \in \mathbb{R}^d$ with $\|x\| = 1$ is referred to as a unit vector. The following simple claim will be used.

▷ **Claim 7.** For every positive integer d and any three vectors $x, y, z \in \mathbb{R}^d$, it holds that

$$\Big| |\langle x, y \rangle| - |\langle z, y \rangle| \Big| \leq \|x - z\| \cdot \|y\|.$$

For a real matrix $A = (a_{i,j})$, let $\mathrm{rank}(A)$ denote its rank over $\mathbb{R}$, and let $\|A\|_\infty$ denote its infinity norm, defined by $\|A\|_\infty = \max_{i,j} |a_{i,j}|$. It is well known that every matrix $A \in \mathbb{R}^{n \times n}$ of rank d can be expressed as $A = X \cdot Y^t$ for some matrices $X, Y \in \mathbb{R}^{n \times d}$. The following lemma guarantees the existence of such a factorization with matrices X and Y whose rows have bounded norms. Its proof relies on John's classical theorem from Banach space theory and can be found, e.g., in [35, Corollary 2.2].

▶ **Lemma 8.** *Let $d \leq n$ be positive integers. For every matrix $A \in \mathbb{R}^{n \times n}$ of rank d, there exist two matrices $X, Y \in \mathbb{R}^{n \times d}$ satisfying $A = X \cdot Y^t$, such that every row of X and Y has norm at most $d^{1/4} \cdot \|A\|_\infty^{1/2}$.*

A matrix $A \in \mathbb{R}^{n \times n}$ is said to be *positive semi-definite* if $x^t A x \geq 0$ for all vectors $x \in \mathbb{R}^n$. This condition is equivalent to the existence of a matrix $X \in \mathbb{R}^{n \times d}$ such that $A = X \cdot X^t$ where $d = \mathrm{rank}(A)$. The coherence of a symmetric matrix measures the extent to which its row (or column) space aligns with the vectors of the standard basis. It is formally defined as follows.

▶ **Definition 9** (Coherence). *For positive integers $d \leq n$, let U be a d-dimensional subspace of $\mathbb{R}^n$, and let P_U be the orthogonal projection onto U. The* coherence *of U is defined as $\mu(U) = \frac{n}{d} \cdot \max_{i \in [n]} \|P_U e_i\|^2$, where e_i stands for the ith vector of the standard basis of $\mathbb{R}^n$. Note that $\mu(U) \in [1, \frac{n}{d}]$. The* coherence *of a symmetric matrix $A \in \mathbb{R}^{n \times n}$, denoted by $\mu(A)$, is defined as the coherence of its row (or column) space.*

2.2 Nets

For a positive integer d and a real number $\theta > 0$, let $B_d(\theta)$ denote the closed d-dimensional ball of radius θ centered at the origin, i.e., $B_d(\theta) = \{x \in \mathbb{R}^d \mid \|x\| \leq \theta\}$. We define a net for a closed ball as follows.

▶ **Definition 10.** *For a positive integer d and real numbers $\eta, \theta > 0$, an η-net for $B_d(\theta)$ is a set $K \subseteq \mathbb{R}^d$ such that for any $x \in B_d(\theta)$, there exists a point $y \in K$ satisfying $\|x - y\| < \eta$.*

Note that Definition 10 requires every point in the ball to admit a point in the η-net at a distance *strictly* smaller than η. We need the following standard lemma on the existence of bounded-size nets for balls (see, e.g., [15]).

▶ **Lemma 11.** *For every positive integer d and any real numbers $\eta, \theta > 0$, there exists an η-net for $B_d(\theta)$ of size at most $(\frac{2\theta}{\eta} + 1)^d$.*

2.3 The Rank of Perturbed Identity Matrices

It is well known and easy to verify that if a matrix $A = (a_{i,j}) \in \mathbb{R}^{m \times m}$ satisfies $a_{i,i} = 1$ for all $i \in [m]$ and $|a_{i,j}| \leq \frac{1}{m}$ for all distinct $i, j \in [m]$, then A has full rank. The following theorem, proved by Alon [2], provides lower bounds on the rank of a symmetric matrix under a weaker assumption on its off-diagonal entries. For a variety of applications of these bounds, the reader is referred to [3] (see also [4]).

▶ **Theorem 12** ([2]). *There exists an absolute constant $c > 0$ for which the following holds. For positive integers $d \leq m$ and for a real number $\varepsilon \in [0, 1)$, let $A = (a_{i,j}) \in \mathbb{R}^{m \times m}$ be a symmetric matrix of rank d satisfying $a_{i,i} = 1$ for all $i \in [m]$ and $|a_{i,j}| \leq \varepsilon$ for all distinct $i, j \in [m]$. Then*
1. *$d \geq \frac{m}{1 + \varepsilon^2 \cdot (m-1)}$, and*
2. *if $\varepsilon \in [\frac{1}{\sqrt{m}}, \frac{1}{2}]$, then $d \geq c \cdot \frac{\log m}{\varepsilon^2 \cdot \log(1/\varepsilon)}$.*

In light of Theorem 12, we introduce the quantities $m(d, \varepsilon)$, defined as follows.

▶ **Definition 13.** *For a positive integer d and a real number $\varepsilon \in [0, \frac{1}{2}]$, let $m(d, \varepsilon)$ denote the maximum integer m for which there exists a symmetric matrix $A = (a_{i,j}) \in \mathbb{R}^{m \times m}$ of rank (at most) d satisfying $a_{i,i} = 1$ for all $i \in [m]$ and $|a_{i,j}| \leq \varepsilon$ for all distinct $i, j \in [m]$.*

Note that for all positive integers d and real numbers $\varepsilon \leq \varepsilon'$, it holds that $m(d, \varepsilon) \leq m(d, \varepsilon')$.

As a direct corollary of Theorem 12, we obtain the following bounds on $m(d, \varepsilon)$. A proof is provided in the full version of the paper.

▶ **Corollary 14.** *There exists an absolute constant c such that the following holds for all positive integers d.*
1. *If $\varepsilon \in [0, \frac{1}{\sqrt{d}})$, then $m(d, \varepsilon) \leq d \cdot \frac{1 - \varepsilon^2}{1 - d \cdot \varepsilon^2}$. In particular, if $\varepsilon \leq \frac{1}{\sqrt{2d}}$, then $m(d, \varepsilon) < 2d$.*
2. *If $\varepsilon \in [\frac{1}{\sqrt{d}}, \frac{1}{2}]$, then $m(d, \varepsilon) \leq 2^{c \cdot d \cdot \varepsilon^2 \cdot \log(1/\varepsilon)}$.*

3 Nearly Orthonormal Representations of Graphs

A d-dimensional orthonormal representation of a graph is an assignment of a unit vector in $\mathbb{R}^d$ to each vertex, such that adjacent vertices receive orthogonal vectors (see [31]). We introduce the following relaxation of this concept, where adjacent vertices receive vectors that are *nearly* orthogonal.

▶ **Definition 15.** *Let $G = (V, E)$ be a graph. For a positive integer d and a real number $\varepsilon \in [0, 1)$, a d-dimensional ε-orthonormal representation of G is an assignment of a unit vector $x_v \in \mathbb{R}^d$ to each vertex $v \in V$, such that for every pair of adjacent vertices u and v in G, it holds that $|\langle x_u, x_v \rangle| \leq \varepsilon$. For any real number $\varepsilon \in [0, 1)$, the ε-orthogonality dimension of G, denoted $\overline{\xi}_\varepsilon(G)$, is the smallest positive integer d for which G admits a d-dimensional ε-orthonormal representation. We omit ε from the notation and terminology when $\varepsilon = 0$.*

A well-studied extension of orthonormal representations, introduced in [18], is that of graph-fitting matrices (see also [33]). We propose the following relaxation of this notion.

▶ **Definition 16.** *Let $G = (V, E)$ be a graph. For a real number $\varepsilon \in [0, 1)$, a matrix $A = (a_{u,v}) \in \mathbb{R}^{|V| \times |V|}$, whose rows and columns are indexed by V, is said to ε-fit the graph G if $a_{v,v} = 1$ for all $v \in V$ and $|a_{u,v}| \leq \varepsilon$ whenever u and v are adjacent vertices in G. When $\varepsilon = 0$, A is said to fit G.*

For a given graph G and a real number $\varepsilon \in [0, 1)$, we are concerned with the minimum possible rank of a matrix that ε-fits G. When $\varepsilon = 0$, this quantity coincides with the minrank of the complement of G over the reals (see [18, 33]).

▶ **Remark 17.** The notions of ε-orthonormal representations and ε-fitting matrices, given in Definitions 15 and 16, are closely related. To see this, consider a graph $G = (V, E)$, and associate with each d-dimensional ε-orthonormal representation $(x_v)_{v \in V}$ of G the Gram matrix $A = (a_{u,v}) \in \mathbb{R}^{|V| \times |V|}$ of its vectors, defined by $a_{u,v} = \langle x_u, x_v \rangle$ for all $u, v \in V$. Note that such a matrix A is positive semi-definite, has rank at most d, and ε-fits the graph G. Therefore, d-dimensional ε-orthonormal representations of a graph G may be regarded as the special case of matrices of rank at most d that ε-fit G, with the additional property of positive semi-definiteness.

In the rest of this section, we relate the chromatic number of a graph to the rank of matrices that nearly fit it. We first establish such relations for general graphs and then proceed to the case of underlying graphs of line digraphs. Finally, we link the notion of nearly orthonormal representations to the circular chromatic number.

3.1 Chromatic Number

For a positive integer k, a k-coloring of a graph G is a mapping from the vertex set of G to a set of size k. A coloring c of G is called *proper* if $c(u) \neq c(v)$ whenever u and v are adjacent vertices in G. The graph G is called k-colorable if it admits a proper k-coloring, and the smallest integer k for which G is k-colorable is called the *chromatic number* of G and is denoted by $\chi(G)$. Observe that any proper k-coloring of G induces a k-dimensional orthonormal representation of G, in which the vertices of the ith color class are assigned the ith vector of the standard basis of $\mathbb{R}^k$. Consequently, every graph G satisfies $\overline{\xi}(G) \leq \chi(G)$ and thus admits a positive semi-definite matrix of rank at most $\chi(G)$ that fits it (see Remark 17). The following simple lemma, whose argument is borrowed from [20], shows that a slight modification of G ensures the existence of such a matrix with minimal coherence (recall Definition 9). The proof is deferred to the full version of the paper.

▶ **Lemma 18.** *For positive integers k and n, let G be a k-colorable graph on n vertices, and let H denote the disjoint union of k copies of G. Then there exists a positive semi-definite matrix $B \in \{0, 1\}^{kn \times kn}$ that fits the graph H, such that $\mathrm{rank}(B) = k$ and $\mu(B) = 1$.*

The following theorem relates the chromatic number of a graph to the rank of a matrix that nearly fits it. A similar reasoning appears in [20]. The proof is given in the full version of the paper.

▶ **Theorem 19.** *Let $G = (V, E)$ be a graph. For a positive integer d and real numbers $\varepsilon \in [0, 1)$ and $\theta \geq 1$, suppose that there exist two matrices $X, Y \in \mathbb{R}^{|V| \times d}$, where each row of X and Y has norm at most θ, such that the matrix $X \cdot Y^t$ ε-fits the graph G. Then, it holds that $\chi(G) \leq (\frac{4\theta^2}{1-\varepsilon} + 1)^d$. Furthermore, if $X = Y$, then $\chi(G) \leq (\frac{2\sqrt{2}}{\sqrt{1-\varepsilon}} + 1)^d$.*

3.2 Chromatic Number of Line Digraphs

The concept of line digraphs, introduced in [19], is defined as follows.

▶ **Definition 20** (Line Digraph). *For a digraph $G = (V, E)$, the* line digraph of G, *denoted δG, is the digraph on the vertex set E, where there is a directed edge from a vertex (u_1, u_2) to a vertex (v_1, v_2) whenever $u_2 = v_1$. For an (undirected) graph G, its line digraph δG is*

defined as the line digraph of the digraph obtained from G by replacing each edge with two oppositely directed edges. Let $\tilde{\delta}G$ denote the underlying graph of δG, i.e., the graph obtained from δG by ignoring the directions of the edges.

The following result, proved by Poljak and Rödl [34] (see also [21]), shows that the chromatic number of a graph G determines the chromatic number of $\tilde{\delta}G$. The statement involves the function $b : \mathbb{N} \to \mathbb{N}$, defined by $b(n) = \binom{n}{\lfloor n/2 \rfloor}$.

▶ **Theorem 21** ([34]). *For every graph G, $\chi(\tilde{\delta}G) = \min\{n \in \mathbb{N} \mid \chi(G) \leq b(n)\}$.*

The following theorem ties the chromatic number of a graph to the rank of a symmetric matrix that nearly fits the underlying graph of its line digraph. It plays a crucial role in our NP-hardness results. The statement involves the quantities $m(d, \varepsilon)$ given in Definition 13.

▶ **Theorem 22.** *Let $G = (V, E)$ be a graph, and let $\tilde{\delta}G = (V', E')$ be the underlying graph of its line digraph. For a positive integer d and real numbers $\varepsilon \in [0, \frac{1}{2})$ and $\theta \geq 1$, suppose that there exist two matrices $X, Y \in \mathbb{R}^{|V'| \times d}$, where each row of X and Y has norm at most θ, such that $X \cdot Y^t$ is a symmetric matrix that ε-fits the graph $\tilde{\delta}G$. Then, for any $\eta \in (0, \frac{1-2\varepsilon}{4\theta}]$, it holds that*

$$\chi(G) \leq \left(\frac{2\theta}{\eta} + 1 \right)^{d \cdot m(d, 2\eta\theta + \varepsilon)}.$$

Proof. Consider a graph G, an integer d, real numbers $\varepsilon, \theta, \eta$, and matrices X, Y as in the statement of the theorem. By Lemma 11, there exists an η-net K for the closed d-dimensional ball $B_d(\theta)$ of radius θ, such that $|K| \leq (\frac{2\theta}{\eta} + 1)^d$. Let $f : B_d(\theta) \to K$ be a function that maps each point $x \in B_d(\theta)$ to a point in K that is closest to x. Since K is an η-net for $B_d(\theta)$, it holds that $\|f(x) - x\| < \eta$ for every $x \in B_d(\theta)$.

Recall that every vertex of $\tilde{\delta}G$ is a pair $e = (u, v) \in V'$ of vertices $u, v \in V$ that are adjacent in G. For each such vertex e, let x_e and y_e denote the rows associated with e in the given matrices X and Y, respectively. By assumption, $\|x_e\| \leq \theta$ and $\|y_e\| \leq \theta$ for every $e \in V'$. Since the matrix $X \cdot Y^t$ ε-fits $\tilde{\delta}G$, it follows that $\langle x_e, y_e \rangle = 1$ for every $e \in V'$, and that $|\langle x_e, y_{e'} \rangle| \leq \varepsilon$ whenever e and e' are adjacent in $\tilde{\delta}G$.

We define a coloring of G as follows. Set $\varepsilon' = 2\eta\theta + \varepsilon \leq \frac{1}{2}$. For every vertex $v \in V$, consider the set E_v of vertices of $\tilde{\delta}G$ whose head is v, that is,

$$E_v = \{e \in V' \mid e = (u, v) \text{ for some } u \in V\}.$$

Let E_v' be a maximal subset of E_v (with respect to containment), such that for all distinct $e, e' \in E_v'$, it holds that $|\langle x_e, y_{e'} \rangle| \leq \varepsilon'$. Equivalently, we require the sub-matrix of $X \cdot Y^t$, restricted to the rows and columns corresponding to the vertices of E_v', to have off-diagonal values at most ε' in absolute value. Notice that $X \cdot Y^t$ is a symmetric matrix of rank at most d, and thus so is each of its principal sub-matrices. Letting $m = m(d, \varepsilon')$, it follows that $|E_v'| \leq m$. Now, we assign to each vertex $v \in V$ the color $c(v)$, defined as the set of all vectors $f(x_e)$ with $e \in E_v'$. The number of colors used by the coloring c does not exceed the number of m-tuples of vectors from K, which is $|K|^m \leq (\frac{2\theta}{\eta} + 1)^{d \cdot m}$. It remains to show that the coloring c is proper.

Let $u, v \in V$ be adjacent vertices in G, and consider the vector $y_{(u,v)}$ associated with the vertex (u, v) of $\tilde{\delta}G$ in the matrix Y. Since every vertex of E_u is adjacent in $\tilde{\delta}G$ to the vertex (u, v), it follows that every $e \in E_u$ satisfies $|\langle x_e, y_{(u,v)} \rangle| \leq \varepsilon$. Using Claim 7, this yields that every $e \in E_u' \subseteq E_u$ satisfies

$$|\langle f(x_e), y_{(u,v)} \rangle| \leq |\langle x_e, y_{(u,v)} \rangle| + \|f(x_e) - x_e\| \cdot \|y_{(u,v)}\| < \varepsilon + \eta\theta.$$

We next argue that there exists a vertex $e \in E'_v$ such that $|\langle x_e, y_{(u,v)} \rangle| > \varepsilon'$. Indeed, the vertex (u, v) lies in E_v. If (u, v) lies in E'_v, then we have $|\langle x_{(u,v)}, y_{(u,v)} \rangle| = 1 > \varepsilon'$, and otherwise, the maximality of E'_v combined with the symmetry of $X \cdot Y^t$ implies the existence of the desired vertex e. Using Claim 7 again, it follows that this vertex e satisfies

$$|\langle f(x_e), y_{(u,v)} \rangle| \geq |\langle x_e, y_{(u,v)} \rangle| - \|f(x_e) - x_e\| \cdot \|y_{(u,v)}\| > \varepsilon' - \eta\theta = \varepsilon + \eta\theta.$$

We conclude that some vector $f(x_e)$ in the set $c(v)$ is different from all the vectors in the set $c(u)$, hence $c(u) \neq c(v)$. Therefore, the coloring c of G is proper, and we are done.   ◀

3.3   Circular Chromatic Number

The circular chromatic number of graphs, introduced by Vince [37], has several equivalent definitions, one of which is presented below. For a comprehensive introduction to the topic, the reader is referred to the survey [38].

▶ **Definition 23.** *The* circular chromatic number *of a graph $G = (V, E)$, denoted $\chi_c(G)$, is the infimum of all real numbers $r \geq 1$ that admit a mapping $f : V \to [0, 1)$, such that for every pair of adjacent vertices u and v in G, it holds that $\frac{1}{r} \leq |f(u) - f(v)| \leq 1 - \frac{1}{r}$.*

It is known that the infimum in the definition of $\chi_c(G)$ is always attained (at a rational number), hence the infimum can be replaced by a minimum. It is also known that any graph G satisfies $\chi(G) = \lceil \chi_c(G) \rceil$. The following observation relates the circular chromatic number to 2-dimensional nearly orthonormal representations. Note that every graph G with at least one edge satisfies $\chi_c(G) \geq 2$.

▶ **Proposition 24.** *For every graph G with at least one edge and for any real number $\varepsilon \in [0, 1)$, it holds that $\overline{\xi}_\varepsilon(G) \leq 2$ if and only if $\varepsilon \geq \cos(\frac{\pi}{\chi_c(G)})$.*

Proof. A 2-dimensional ε-orthonormal representation of a graph $G = (V, E)$ assigns a unit vector $x_v \in \mathbb{R}^2$ to each vertex $v \in V$, such that every pair of adjacent vertices u and v in G satisfies $|\langle x_u, x_v \rangle| \leq \varepsilon$. We may assume that each vector x_v lies in the upper half of the unit circle, by multiplying some of the vectors by -1 if needed. Therefore, each vector x_v can be expressed by a real number $\alpha_v \in [0, 1)$, such that the angle between the vectors $(1, 0)$ and x_v is $\alpha_v \cdot \pi$. In this language, the condition $|\langle x_u, x_v \rangle| \leq \varepsilon$ translates to $|\cos(\pi \cdot (\alpha_u - \alpha_v))| \leq \varepsilon$, or equivalently, $\frac{\arccos(\varepsilon)}{\pi} \leq |\alpha_u - \alpha_v| \leq 1 - \frac{\arccos(\varepsilon)}{\pi}$. By the definition of circular chromatic number, such a mapping $v \mapsto \alpha_v$ exists if and only if it holds that $\frac{\arccos(\varepsilon)}{\pi} \leq \frac{1}{\chi_c(G)}$, that is, $\varepsilon \geq \cos(\frac{\pi}{\chi_c(G)})$. The proof is complete.   ◀

As a concluding remark, we raise the question of determining the ε-orthogonality dimension of Kneser graphs. The Kneser graph $K(n, k)$, defined for integers n and k with $n \geq 2k$, has vertices corresponding to all k-subsets of $[n]$ and edges between disjoint sets. Settling a conjecture of Kneser [27], Lovász [30] proved that $\chi(K(n, k)) = n - 2k + 2$ as an application of the Borsuk–Ulam theorem from algebraic topology. This result has been strengthened in various ways over time. For example, it was shown by Chen [11] that the chromatic number of $K(n, k)$ coincides with its circular chromatic number, resolving a conjecture of Johnson, Holroyd, and Stahl [26] (see also [9, 29]). By Proposition 24, this result characterizes the values of $\varepsilon \in [0, 1)$ for which $\overline{\xi}_\varepsilon(K(n, k)) \leq 2$ holds. A more recent result [22, 1] asserts that the chromatic number of $K(n, k)$ coincides with its standard orthogonality dimension (where $\varepsilon = 0$). It would thus be intriguing to determine the quantities $\overline{\xi}_\varepsilon(K(n, k))$ for general parameter choices.

4 Hardness Results

This section presents our hardness results for low-rank matrix completion and for related problems. The starting point of our hardness proofs is the gap coloring problem, defined as follows.

▶ **Definition 25** (The (k_1, k_2)-Coloring Problem). *For positive integers $k_1 < k_2$, the (k_1, k_2)-Coloring problem asks to decide whether an input graph G satisfies $\chi(G) \leq k_1$ or $\chi(G) \geq k_2$.*

We rely on the following hardness result, proved by Krokhin et al. [28]. Recall that the function $b : \mathbb{N} \to \mathbb{N}$ is defined by $b(n) = \binom{n}{\lfloor n/2 \rfloor}$.

▶ **Theorem 26** ([28]). *For every integer $k \geq 4$, the $(k, b(k))$-Coloring problem is* NP*-hard.*

In what follows, we reduce the gap coloring problem to an intermediate problem, termed Graph Fitness, and establish its hardness via Theorem 26. While the definition of the Graph Fitness problem may appear somewhat artificial, its hardness enables us to derive all our hardness results in a unified framework. Due to space limitations, some of these implications, including those concerning the Orthogonality Dimension problem, appear only in the full version of the paper. The following figure summarizes the reductions used throughout the paper.
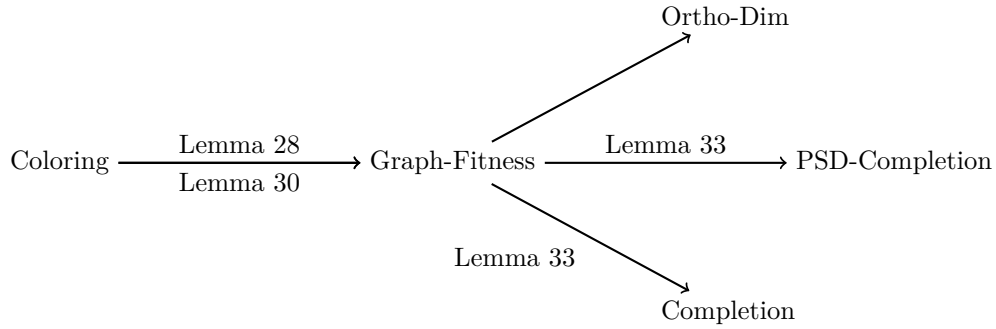
Figure 1 Reductions map.

4.1 Graph Fitness

The Graph Fitness problem is defined as follows.

▶ **Definition 27** (The $(d_1, d_2, \varepsilon, \theta)$-Graph-Fitness Problem). *For positive integers $d_1 < d_2$ and real numbers $\varepsilon \in [0, 1)$ and $\theta \geq 1$, the $(d_1, d_2, \varepsilon, \theta)$-Graph-Fitness problem asks, given a graph G on n vertices, to distinguish between the following cases.*
- YES*: There exists a positive semi-definite matrix $B \in \mathbb{R}^{n \times n}$ that fits the graph G, such that $\mu(B) = 1$ and $\mathrm{rank}(B) \leq d_1$.*
- NO*: For any two matrices $X, Y \in \mathbb{R}^{n \times d}$ whose rows have norm at most θ and for which $X \cdot Y^t$ is a symmetric matrix that ε-fits the graph G, it holds that $d \geq d_2$.*

Note that the condition on YES instances in the above definition implies the existence of a matrix $X \in \mathbb{R}^{n \times d_1}$ whose rows have norm 1, such that $X \cdot X^t$ is a symmetric matrix that fits the graph G. Therefore, the YES and NO instances of the problem do not overlap.

We state two efficient reductions from the (k_1, k_2)-Coloring problem to the $(d_1, d_2, \varepsilon, \theta)$-Graph-Fitness problem for suitable choices of parameters. The proofs can be found in the full version of the paper. The first reduction builds on Theorem 19.

▶ **Lemma 28.** *Let $d_1 < d_2$ be positive integers, and let $\varepsilon \in [0,1)$ and $\theta \geq 1$ be real numbers. Then there exists a polynomial-time reduction from $(d_1, (\frac{4\theta^2}{1-\varepsilon} + 1)^{d_2})$-Coloring to $(d_1, d_2, \varepsilon, \theta)$-Graph-Fitness.*

▶ Remark 29. It was proved in [12, 13] that certain variants of the Unique Games Conjecture imply the hardness of the (k_1, k_2)-Coloring problem for all integers $k_2 > k_1 \geq 3$. By Lemma 28, it follows that the same complexity assumptions imply the hardness of the $(d_1, d_2, \varepsilon, \theta)$-Graph-Fitness problem for all integers $d_2 > d_1 \geq 3$ and real numbers $\varepsilon \in [0, 1)$ and $\theta \geq 1$.

The next reduction between the problems relies on Theorem 22 and is crucial for deriving our NP-hardness results from Theorem 26. Note that the reduction from Lemma 28 is insufficient for this purpose. The statement involves the quantities $m(d, \varepsilon)$ from Definition 13 and the function $b : \mathbb{N} \to \mathbb{N}$ from Theorem 26.

▶ **Lemma 30.** *Let $k_1 < k_2$ and $d_1 < d_2$ be positive integers, and let $\varepsilon \in [0, \frac{1}{2})$, $\theta \geq 1$, and η be real numbers, such that*

$$\eta \in \left(0, \frac{1-2\varepsilon}{4\theta}\right], \quad k_1 \leq b(d_1), \quad and \quad k_2 \geq \left(\frac{2\theta}{\eta} + 1\right)^{d_2 \cdot m(d_2, 2\eta\theta + \varepsilon)}.$$

Then there exists a polynomial-time reduction from (k_1, k_2)-Coloring to $(d_1, d_2, \varepsilon, \theta)$-Graph-Fitness.

We next state an NP-hardness result for the Graph Fitness problem. The (somewhat tedious) proof, which is omitted here, integrates the hardness of the gap coloring problem from Theorem 26, the reduction to Graph Fitness provided by Lemma 30, and the bounds on the quantities $m(d, \varepsilon)$ from Corollary 14.

▶ **Theorem 31.** *There exists an absolute constant $c > 0$ for which the following holds. Let d and g be positive integers with d sufficiently large and $d < g$, and let ε and $\theta \geq 1$ be real numbers. Suppose that either*
1. *$\varepsilon \in [0, \frac{1}{3\sqrt{g}}]$ and $g \leq c \cdot \frac{2^{d/2}}{d^{1/4} \cdot \max(\log \theta, d)^{1/2}}$, or*
2. *$\varepsilon \in [\frac{1}{3\sqrt{g}}, \frac{1}{6}]$, $\theta \leq 2^{2^{c \cdot d}}$, and $g \leq c \cdot \frac{d}{\varepsilon^2 \cdot \log(1/\varepsilon)}$.*
Then the $(d, g, \varepsilon, \theta)$-Graph-Fitness problem is NP-hard.

Theorem 31 establishes the NP-hardness of the Graph Fitness problem for general parameter settings. To illustrate its applicability, we present three implications below. The first provides an exponential gap in the rank for an exponentially small error, the second offers a polynomial gap in the rank along with a polynomial decay of the error, and the third shows a constant multiplicative gap in the rank for a constant error.

▶ **Theorem 32.** *There exists an absolute constant $c > 0$ for which the following holds.*
1. *There exists an absolute constant $c' > 0$, such that for every sufficiently large positive integer d, the $(d, c \cdot \frac{2^{d/2}}{d^{3/4}}, 2^{-c' \cdot d}, 2^d)$-Graph-Fitness problem is NP-hard.*
2. *For every $\beta > 1$, there exists some $c' > 0$, such that for every sufficiently large positive integer d, the $(d, d^\beta, c' \cdot \frac{1}{(d^{\beta-1} \cdot \log d)^{1/2}}, 2^{2^{c \cdot d}})$-Graph-Fitness problem is NP-hard.*
3. *For every $\alpha > 1$, there exists some $\varepsilon \in (0, 1)$, such that for every sufficiently large positive integer d, the $(d, \alpha \cdot d, \varepsilon, 2^{2^{c \cdot d}})$-Graph-Fitness problem is NP-hard.*

4.2 Low-Rank Matrix Completion

We finally turn to proving our hardness results for the low-rank matrix completion problems described in Definitions 1 and 4, thereby strengthening the NP-hardness results of [20]. This will be accomplished through the reductions outlined in the following lemma.

▶ **Lemma 33.** *For all positive integers $d_1 < d_2$ and real numbers $\varepsilon \in [0,1)$, the following holds.*

1. *There exists a polynomial-time reduction from $(d_1, d_2, \varepsilon, 1)$-Graph-Fitness to*

$$\left(d_1, d_2, \frac{\varepsilon}{1+\varepsilon}\right)\text{-}PSD\text{-}Completion.$$

2. *For every real number $\theta \geq (1+\varepsilon)^{1/2} \cdot d_2^{1/4}$, if $d_1 < \lfloor \frac{d_2}{2} \rfloor$ then there exists a polynomial-time reduction from $(d_1, d_2, \varepsilon, \theta)$-Graph-Fitness to*

$$\left(d_1, \left\lfloor \frac{d_2}{2} \right\rfloor, \frac{\varepsilon}{1+\varepsilon}, \frac{\theta^2}{(1+\varepsilon) \cdot d_2^{1/2}}\right)\text{-}Completion.$$

Proof. Fix integers d_1, d_2 and real numbers ε, θ as in the statement of the lemma, and set

$$\varepsilon' = \frac{\varepsilon}{1+\varepsilon} \quad \text{and} \quad \theta' = \frac{\theta^2}{(1+\varepsilon) \cdot d_2^{1/2}}.$$

Note that $\varepsilon' \in [0,1)$ and $\theta' \geq 1$. Both parts of the lemma are established through the same reduction, which is described next. For a given graph $G = (V, E)$ on n vertices, the reduction produces and returns the partial matrix $A \in \{0, 1, \perp\}^{n \times n}$, whose rows and columns are indexed by V, defined by $A_{u,v} = 1$ if $u = v$, $A_{u,v} = 0$ if $\{u, v\} \in E$, and $A_{u,v} = \perp$ otherwise. This reduction can clearly be implemented in polynomial time (in fact, in logarithmic space).

We now prove the correctness of the reduction. We begin with the forward direction, which applies to both parts of the lemma. Suppose that G is a YES instance of $(d_1, d_2, \varepsilon, \theta)$-Graph-Fitness for some $\theta \geq 1$. Then, there exists a positive semi-definite matrix $B \in \mathbb{R}^{n \times n}$ that fits the graph G, such that $\mu(B) = 1$ and $\text{rank}(B) \leq d_1$. Since B fits G, the diagonal entries of B are all ones, and the entries of B associated with adjacent vertices are zeros. This implies that $A_{u,v} = B_{u,v}$ whenever $A_{u,v} \neq \perp$. It follows that A is a YES instance of both (d_1, d_2, ε')-PSD-Completion and $(d_1, \lfloor \frac{d_2}{2} \rfloor, \varepsilon', \theta')$-Completion, as needed.

For the reverse direction, we prove the contrapositive. For Item 1 of the lemma, we show that if A is not a NO instance of (d_1, d_2, ε')-PSD-Completion, then G is not a NO instance of $(d_1, d_2, \varepsilon, 1)$-Graph-Fitness. Suppose that for some integer $d < d_2$, there exists a positive semi-definite matrix $B \in \mathbb{R}^{n \times n}$ of rank d, such that $|A_{u,v} - B_{u,v}| \leq \varepsilon'$ whenever $A_{u,v} \neq \perp$. Since B is positive semi-definite and of rank d, one may write $B = X \cdot X^t$ for a matrix $X \in \mathbb{R}^{n \times d}$. For each vertex $v \in V$, let x_v denote the row of X associated with v. Notice that for every vertex $v \in V$, it holds that $\|x_v\|^2 = \langle x_v, x_v \rangle = B_{v,v} \in [1 - \varepsilon', 1 + \varepsilon']$, and that for every pair of adjacent vertices u and v in G, it holds that $\langle x_u, x_v \rangle = B_{u,v} \in [-\varepsilon', +\varepsilon']$. For each vertex $v \in V$, let $x'_v = \frac{x_v}{\|x_v\|}$. It follows that for every $v \in V$, x'_v is a unit vector, and that for every pair of adjacent vertices u and v in G, it holds that

$$\langle x'_u, x'_v \rangle = \frac{\langle x_u, x_v \rangle}{\|x_u\| \cdot \|x_v\|} \in \left[-\frac{\varepsilon'}{1 - \varepsilon'}, +\frac{\varepsilon'}{1 - \varepsilon'} \right] = [-\varepsilon, +\varepsilon].$$

Letting $X' \in \mathbb{R}^{n \times d}$ denote the matrix in which the row associated with a vertex v is x'_v, we obtain that every row of X' has norm 1 and that $X' \cdot (X')^t$ is a symmetric matrix that ε-fits the graph G. By $d < d_2$, it follows that G is not a NO instance of $(d_1, d_2, \varepsilon, 1)$-Graph-Fitness, as desired.

For Item 2 of the lemma, we show that if A is not a NO instance of $(d_1, \lfloor \frac{d_2}{2} \rfloor, \varepsilon', \theta')$-Completion, then G is not a NO instance of $(d_1, d_2, \varepsilon, \theta)$-Graph-Fitness. Suppose that for some integer $d < \lfloor \frac{d_2}{2} \rfloor$, there exists a matrix $B \in [-\theta', +\theta']^{n \times n}$ of rank d, such that $|A_{u,v} - B_{u,v}| \leq \varepsilon'$ whenever $A_{u,v} \neq \perp$. Put $C = \frac{1}{2} \cdot (B + B^t)$, and notice that C is a symmetric matrix that lies in $[-\theta', +\theta']^{n \times n}$ and satisfies $\mathrm{rank}(C) \leq 2d < d_2$. By the symmetry of A, we observe that for all $u, v \in V$ with $A_{u,v} \neq \perp$, it holds that

$$
\begin{aligned}
|A_{u,v} - C_{u,v}| &= \left| \tfrac{1}{2} \cdot (A_{u,v} - B_{u,v}) + \tfrac{1}{2} \cdot (A_{v,u} - B_{v,u}) \right| \\
&\leq \tfrac{1}{2} \cdot |A_{u,v} - B_{u,v}| + \tfrac{1}{2} \cdot |A_{v,u} - B_{v,u}| \leq \varepsilon'.
\end{aligned}
\tag{2}
$$

By Lemma 8, there exist two matrices $X, Y \in \mathbb{R}^{n \times (2d)}$ satisfying $C = X \cdot Y^t$, such that every row of X and Y has norm at most

$$
(2d)^{1/4} \cdot \theta'^{1/2} < d_2^{1/4} \cdot \theta'^{1/2} = \frac{\theta}{(1+\varepsilon)^{1/2}} = (1-\varepsilon')^{1/2} \cdot \theta.
\tag{3}
$$

For each vertex $v \in V$, let x_v and y_v denote the rows associated with v in X and Y, respectively. By (2), for every vertex $v \in V$, it holds that $\langle x_v, y_v \rangle = C_{v,v} \in [1 - \varepsilon', 1 + \varepsilon']$, and for every pair of adjacent vertices u and v in G, it holds that $\langle x_u, y_v \rangle = C_{u,v} \in [-\varepsilon', +\varepsilon']$. For each vertex $v \in V$, let $x'_v = \frac{x_v}{\langle x_v, y_v \rangle^{1/2}}$ and $y'_v = \frac{y_v}{\langle x_v, y_v \rangle^{1/2}}$, and observe using (3) that $\|x'_v\| \leq \theta$ and $\|y'_v\| \leq \theta$. For every $v \in V$, we have $\langle x'_v, y'_v \rangle = 1$, and for every pair of adjacent vertices u and v in G, it holds that

$$
\langle x'_u, y'_v \rangle = \frac{\langle x_u, y_v \rangle}{\langle x_u, y_u \rangle^{1/2} \cdot \langle x_v, y_v \rangle^{1/2}} \in \left[-\frac{\varepsilon'}{1-\varepsilon'}, +\frac{\varepsilon'}{1-\varepsilon'} \right] = [-\varepsilon, +\varepsilon].
$$

Let $X', Y' \in \mathbb{R}^{n \times (2d)}$ denote the matrices in which the rows associated with a vertex v are x'_v and y'_v, respectively. The above discussion implies that every row of X' and Y' has norm at most θ and that $X' \cdot (Y')^t$ is a symmetric matrix that ε-fits the graph G. By $2d < d_2$, it follows that G is not a NO instance of $(d_1, d_2, \varepsilon, \theta)$-Graph-Fitness, completing the proof. ◀

By combining Theorem 31 with the first item of Lemma 33, we obtain the following result. A similar consequence, omitted here, can be derived from the second item of the lemma.

▶ **Corollary 34.** *There exists an absolute constant $c > 0$ for which the following holds. Let d and g be positive integers with d sufficiently large and $d < g$, and let ε be a real number. Suppose that either*
1. *$\varepsilon \in [0, \frac{1}{3\sqrt{g}}]$ and $g \leq c \cdot \frac{2^{d/2}}{d^{3/4}}$, or*
2. *$\varepsilon \in [\frac{1}{3\sqrt{g}}, \frac{1}{6}]$ and $g \leq c \cdot \frac{d}{\varepsilon^2 \cdot \log(1/\varepsilon)}$.*
Then the $(d, g, \frac{\varepsilon}{1+\varepsilon})$-PSD-Completion problem is NP-hard.

───── **References** ─────

1   Meysam Alishahi and Frédéric Meunier. Topological bounds for graph representations over any field. *SIAM J. Discret. Math.*, 35(1):91–104, 2021. `doi:10.1137/19M1295921`.

2   Noga Alon. Problems and results in extremal combinatorics, I. *Discrete Math.*, 273(1–3):3–15, 2003.

3   Noga Alon. Perturbed identity matrices have high rank: Proof and applications. *Comb. Probab. Comput.*, 18(1–2):3–15, 2009. `doi:10.1017/S0963548307008917`.

4   Noga Alon, Yoshiharu Kohayakawa, Christian Mauduit, Carlos Gustavo Moreira, and Vojtěch Rödl. Measures of pseudorandomness for finite sequences: minimal values. *Comb. Probab. Comput.*, 15(1–2):1–29, 2006. `doi:10.1017/S0963548305007170`.

5 Libor Barto, Jakub Bulín, Andrei A. Krokhin, and Jakub Opršal. Algebraic approach to promise constraint satisfaction. *J. ACM*, 68(4):28:1–66, 2021. Preliminary versions in STOC'19 and LICS'19. `doi:10.1145/3457606`.

6 Jop Briët, Harry Buhrman, Debbie Leung, Teresa Piovesan, and Florian Speelman. Round elimination in exact communication complexity. In *Proc. of the 10th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC'15)*, pages 206–225, 2015. `doi:10.4230/LIPICS.TQC.2015.206`.

7 Emmanuel J. Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Found. Comput. Math.*, 9(6):717–772, 2009. `doi:10.1007/S10208-009-9045-5`.

8 Emmanuel J. Candès and Terence Tao. The power of convex relaxation: Near-optimal matrix completion. *IEEE Trans. Inform. Theory*, 56(5):2053–2080, 2010. `doi:10.1109/TIT.2010.2044061`.

9 Gerard Jennhwa Chang, Daphne Der-Fen Liu, and Xuding Zhu. A short proof for Chen's alternative Kneser coloring lemma. *J. Comb. Theory A*, 120(1):159–163, 2013. `doi:10.1016/J.JCTA.2012.07.009`.

10 Dror Chawin and Ishay Haviv. Improved NP-hardness of approximation for orthogonality dimension and minrank. *SIAM J. Discret. Math.*, 37(4):2670–2688, 2023. Preliminary version in STACS'23. `doi:10.1137/23M155760X`.

11 Peng-An Chen. A new coloring theorem of Kneser graphs. *J. Combin. Theory Ser. A*, 118(3):1062–1071, 2011. `doi:10.1016/J.JCTA.2010.08.008`.

12 Irit Dinur, Elchanan Mossel, and Oded Regev. Conditional hardness for approximate coloring. *SIAM J. Comput.*, 39(3):843–873, 2009. Preliminary version in STOC'06. `doi:10.1137/07068062X`.

13 Irit Dinur and Igor Shinkar. On the conditional hardness of coloring a 4-colorable graph with super-constant number of colors. In *Proc. of the 13th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX'10)*, pages 138–151, 2010. `doi:10.1007/978-3-642-15369-3_11`.

14 Marianna Eisenberg-Nagy, Monique Laurent, and Antonios Varvitsiotis. *Complexity of the Positive Semidefinite Matrix Completion Problem with a Rank Constraint*, volume 69 of *Fields Institute Communications*, pages 105–120. Springer, Heidelberg, 2013.

15 Tadeusz Figiel, Joram Lindenstrauss, and Vitali D. Milman. The dimension of almost spherical sections of convex bodies. *Acta Math.*, 139(1–2):53–94, 1977.

16 Alexander Golovnev and Ishay Haviv. The (generalized) orthogonality dimension of (generalized) Kneser graphs: Bounds and applications. *Theory Comput.*, 18(22):1–22, 2022. Preliminary version in CCC'21. `doi:10.4086/TOC.2022.V018A022`.

17 Venkatesan Guruswami and Sai Sandeep. d-to-1 hardness of coloring 3-colorable graphs with $O(1)$ colors. In *Proc. of the 47th International Colloquium on Automata, Languages, and Programming, (ICALP'20)*, pages 62:1–12, 2020. `doi:10.4230/LIPICS.ICALP.2020.62`.

18 Willem H. Haemers. An upper bound for the Shannon capacity of a graph. In László Lovász and Vera T. Sós, editors, *Algebraic Methods in Graph Theory*, volume 25/I of *Colloquia Mathematica Societatis János Bolyai*, pages 267–272. Bolyai Society and North-Holland, 1981.

19 Frank Harary and Robert Z. Norman. Some properties of line digraphs. *Rend. Circ. Mat. Palermo*, 9(2):161–168, 1960.

20 Moritz Hardt, Raghu Meka, Prasad Raghavendra, and Benjamin Weitz. Computational limits for matrix completion. In *Proc. of the 27th Conference on Learning Theory (COLT'14)*, pages 1017–1032, 2014.

21 Charles C. Harner and Roger C. Entringer. Arc colorings of digraphs. *J. Comb. Theory, Ser. B*, 13(3):219–225, 1972.

22 Ishay Haviv. Topological bounds on the dimension of orthogonal representations of graphs. *Eur. J. Comb.*, 81:84–97, 2019. `doi:10.1016/J.EJC.2019.04.006`.

23 Ishay Haviv. Approximating the orthogonality dimension of graphs and hypergraphs. *Chic. J. Theor. Comput. Sci.*, 2022(2), 2022. Preliminary version in MFCS'19.

24 Yahli Hecht, Dor Minzer, and Muli Safra. NP-hardness of almost coloring almost 3-colorable graphs. In *Proc. of the 27th International Conference on Randomization and Computation (RANDOM'23)*, pages 51:1–12, 2023. `doi:10.4230/LIPICS.APPROX/RANDOM.2023.51`.

25 Sangxia Huang. Improved hardness of approximating chromatic number. In *Proc. of the 16th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX'13)*, pages 233–243, 2013. `doi:10.1007/978-3-642-40328-6_17`.

26 A. Johnson, Fred C. Holroyd, and Saul Stahl. Multichromatic numbers, star chromatic numbers and Kneser graphs. *J. Graph Theory*, 26(3):137–145, 1997. `doi:10.1002/(SICI) 1097-0118(199711)26:3\%3C137::AID-JGT4\%3E3.0.CO;2-S`.

27 Martin Kneser. Aufgabe 360. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 58(2):27, 1955.

28 Andrei A. Krokhin, Jakub Opršal, Marcin Wrochna, and Stanislav Zivný. Topology and adjunction in promise constraint satisfaction. *SIAM J. Comput.*, 52(1):38–79, 2023. Preliminary versions in FOCS'19 and SODA'20. `doi:10.1137/20M1378223`.

29 Daphne Der-Fen Liu and Xuding Zhu. A combinatorial proof for the circular chromatic number of Kneser graphs. *J. Comb. Optim.*, 32:765–774, 2016. `doi:10.1007/S10878-015-9897-3`.

30 László Lovász. Kneser's conjecture, chromatic number, and homotopy. *J. Comb. Theory, Ser. A*, 25(3):319–324, 1978. `doi:10.1016/0097-3165(78)90022-5`.

31 László Lovász. On the Shannon capacity of a graph. *IEEE Trans. Inform. Theory*, 25(1):1–7, 1979. `doi:10.1109/TIT.1979.1055985`.

32 László Lovász, Michael Saks, and Alexander Schrijver. Orthogonal representations and connectivity of graphs. *Linear Algebra Appl.*, 114–115:439–454, 1989.

33 René Peeters. Orthogonal representations over finite fields and the chromatic number of graphs. *Combinatorica*, 16(3):191–206, 1996.

34 Svatopluk Poljak and Vojtech Rödl. On the arc-chromatic number of a digraph. *J. Comb. Theory, Ser. B*, 31(2):190–198, 1981. `doi:10.1016/S0095-8956(81)80024-X`.

35 Cyrus Rashtchian. Bounded matrix rigidity and John's theorem. *Electron. Colloquium Comput. Complex.*, TR16-093, 2016. URL: `https://eccc.weizmann.ac.il/report/2016/093`.

36 Benjamin Recht. A simpler approach to matrix completion. *J. Mach. Learn. Res.*, 12:3413–3430, 2011. `doi:10.5555/1953048.2185803`.

37 Andrew Vince. Star chromatic number. *J. Graph Theory*, 12(4):551–559, 1988. `doi:10.1002/JGT.3190120411`.

38 Xuding Zhu. Circular chromatic number: a survey. *Discrete Math.*, 229(1–3):371–410, 2001. `doi:10.1016/S0012-365X(00)00217-X`.