



**HAL**  
open science

## Exact solutions in low-rank approximation with zeros

Kaie Kubjas, Luca Sodomaco, Elias Tsigaridas

► **To cite this version:**

Kaie Kubjas, Luca Sodomaco, Elias Tsigaridas. Exact solutions in low-rank approximation with zeros. Linear Algebra and its Applications, 2022, 641, pp.67-97. 10.1016/j.laa.2022.01.021 . hal-03670848

**HAL Id: hal-03670848**

**<https://u-paris.hal.science/hal-03670848v1>**

Submitted on 22 Jul 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

# Exact solutions in low-rank approximation with zeros

Kaie Kubjas<sup>1</sup>, Luca Sodomaco<sup>1</sup>, Elias Tsigaridas<sup>b</sup>

<sup>a</sup>*Department of Mathematics and Systems Analysis, Aalto University*

<sup>b</sup>*Inria Paris and IMJ-PRG,  
Sorbonne Université and Paris Université*

---

## Abstract

Low-rank approximation with zeros aims to find a matrix of fixed rank and with a fixed zero pattern that minimizes the Euclidean distance to a given data matrix. We study the critical points of this optimization problem using algebraic tools. In particular, we describe special linear, affine, and determinantal relations satisfied by the critical points. We also investigate the number of critical points and how this number is related to the complexity of nonnegative matrix factorization problem.

*Keywords:* Structured low-rank approximation, zero patterns, Euclidean distance degree, nonnegative matrix factorization

*2000 MSC:* 14M12, 14P05, 13P25, 90C26, 68W30

---

---

*Email addresses:* [kaie.kubjas@aalto.fi](mailto:kaie.kubjas@aalto.fi) (Kaie Kubjas), [luca.sodomaco@aalto.fi](mailto:luca.sodomaco@aalto.fi) (Luca Sodomaco), [elias.tsigaridas@inria.fr](mailto:elias.tsigaridas@inria.fr) (Elias Tsigaridas)

*Preprint submitted to ...*

*January 24, 2022*

## 1. Introduction

The *best rank- $r$  approximation problem* aims to find a real rank- $r$  matrix that minimizes the Euclidean distance to a given real data matrix. The solution of this problem is completely addressed by the Eckart-Young-Mirsky theorem which states that the best rank- $r$  approximation is given by the first  $r$  components of the *singular value decomposition* (SVD) of the data matrix.

We study the *structured best rank- $r$  approximation problem*, namely we consider additional linear constraints on rank- $r$  matrices. We focus on *coordinate subspaces*, i.e., linear spaces that are defined by setting some entries to zero. Let  $S \subset [m] \times [n]$  denote the indices of zero entries. Given  $U = (u_{ij}) \in \mathbb{R}^{m \times n}$ , our optimization problem becomes

$$\begin{aligned} \min_X d_U(X) &:= \sum_{i=1}^m \sum_{j=1}^n (x_{ij} - u_{ij})^2 \\ \text{s.t. } x_{ij} &= 0 \quad \forall (i, j) \in S \text{ and } \text{rank}(X) \leq r. \end{aligned} \tag{1.1}$$

Structured low-rank approximation problem has been studied in [CFP03, Mar08, Mar19]; see also [GG11] for low rank approximations with weights. Exact solutions to this problem have been investigated by Golub, Hoffman and Stewart [GHS87], and by Ottaviani, Spaenlehauer and Sturmfels [OSS14]. In [GHS87], rank- $r$  critical points are studied under the constraint that entries in a set of rows or in a set of columns of a matrix stay fixed. This situation is more general than ours in the aspect that the fixed entries are not required to be zero but more restrictive when it comes to the indices of the entries that are fixed. In [OSS14], rank- $r$  critical points restricted to generic subspaces of matrices are studied. In our paper, the linear spaces set some entries equal to zero and hence are not generic. Because of this, we cannot use many powerful tools from algebraic geometry and intersection theory and we have to come up with algebraic and computational techniques that exploit this special structure. For some properties of determinantal ideals of matrices with 0 entries and their relations to problems in graph theory we refer the reader to [CW19] and references therein. Horobet and Rodriguez study the problem when at least one solution of a certain family of optimization problems satisfies given polynomial conditions, and address the structured low-rank approximation as a particular case [HR20, Example 15].

The global minimum of the optimization problem (1.1) always exists, because we can select any point  $X$  in the feasible region and consider the feasible region intersected with the closed ball  $B_{d_U(X)}(U)$  centered at  $U$  and with radius  $\frac{1}{2}$

$d_U(X)$ . Since the feasible region is a closed semialgebraic set, then the intersection is closed and bounded, and hence compact. The distance function is continuous, thus achieves its minimum on this set. This minimum is a global minimum of (1.1). The optimization problem (1.1) is nonconvex and often local methods are used to solve it. They return a local minimum of the optimization problem. There are heuristics for finding a global minimum, but these heuristics do not guarantee that a local minimum is indeed a global minimum. We refer to [Mar08] for various algorithms and to [SS16] for an algorithm with locally quadratic convergence. Cifuentes recently introduced convex relaxations for structured low-rank approximation that under certain assumptions have provable guarantees [Cif21]. Another interesting direction, closely related but not directly applicable to our problem, is to employ recent optimization techniques for simultaneously sparse and low rank approximation [PS17, SRV12].

To compute a global minimizer of (1.1) algebraically, we need to look at all the complex critical points of the polynomial function  $d_U: \mathbb{C}^{m \times n} \rightarrow \mathbb{C}$  on the intersection  $\mathcal{L}_r^S := \mathcal{X}_r \cap \mathcal{L}^S$ , where

$$\mathcal{X}_r := \{X \in \mathbb{C}^{m \times n} \mid \text{rank}(X) \leq r\}, \quad \mathcal{L}^S := \{X \in \mathbb{C}^{m \times n} \mid x_{ij} = 0 \forall (i, j) \in S\}, \quad (1.2)$$

and then select the real solution that minimizes the Euclidean distance. The problem of finding critical points of  $d_U$  on  $\mathcal{L}_r^S$  can be considered in the more general setting when  $U$  is a complex data matrix. This setting includes the practically meaningful setting when  $U$  has real entries. If  $U \in \mathbb{C}^{m \times n}$  is generic, namely if it belongs to the complement of a Zariski closed set, then the number of critical points is constant and is called the *Euclidean Distance degree* (ED degree) of  $\mathcal{L}_r^S$ . We denote this invariant by  $\text{EDdegree}(\mathcal{L}_r^S)$ . The importance of the ED degree is that it measures the algebraic complexity of writing the optimal solution as a function of  $U$ . More generally, the ED degree of an algebraic variety is introduced in [DHO<sup>+</sup>16]. The main goal of this paper is to study the critical points and the ED degree of the minimization problem (1.1).

When rank is one, then characterizing critical points becomes a combinatorial problem. More precisely, listing all critical points translates to the problem of listing minimal vertex covers of a bipartite graph. The complexity of counting vertex covers in a bipartite graph is known to be #P-complete [PB83]. Our main result about rank-one critical points is Proposition 3.3 which gives the ED degree of  $\mathcal{L}_1^S$  in terms of the minimal covers. For row/column and diagonal zero patterns this results in explicit formulas (Corollaries 3.8 and 3.9).

Our first main result for rank- $r$  critical points is Theorem 4.3 which studies the linear span of rank- $r$  critical points of  $d_U$ . We call it the *critical space* in the

structured setting. This is motivated by the notion of critical space of a tensor in the unstructured setting defined by Draisma, Ottaviani and Tocino [DOT18]. From the algebraic perspective, Theorem 4.3 provides a lower bound on the minimal number of generators of degree one in the zero dimensional ideal of rank- $r$  critical points of  $d_U$ . When  $\mathcal{L}_r^S$  is an irreducible variety, we expect this lower bound to be also an upper bound, as stated in Conjecture 4.4.

In the unstructured setting, the rank-one critical points form a basis of the critical space and the rank- $r$  critical points are linear combinations of the basis vectors with coefficients in  $\{0, 1\}$ . In the structured setting, there are not enough rank-one critical points to give a basis of the critical space. We leave it as an open question, whether there is a natural extension to a basis and whether the coefficients that give rank- $r$  critical points as linear combinations of basis elements have a nice description.

Our second main result is Proposition 4.12 that describes affine linear relations that are satisfied by the rank- $r$  critical points of  $d_U$  in the unstructured setting. In the structured setting, we conjecture the affine linear relations satisfied by the rank- $r$  critical points of  $d_U$ . The last kind of constraints satisfied by the rank- $r$  critical points that we consider are nonlinear determinantal constraints given in Proposition 4.18. The ED degree of  $d_U$  is studied in Section 5. Our experiments indicate that the ED degree is exponential in  $|S|$ .

The optimization problem (1.1) is motivated by the *nonnegative matrix factorization* (NMF) problem. Given a nonnegative matrix  $X \in \mathbb{R}_{\geq 0}^{m \times n}$ , the *nonnegative rank* of  $X$  is the smallest  $r$  such that

$$X = AB, \text{ where } A \in \mathbb{R}_{\geq 0}^{m \times r} \text{ and } B \in \mathbb{R}_{\geq 0}^{r \times n}.$$

NMF aims to find a matrix  $X$  of nonnegative rank at most  $r$  that minimizes the Euclidean distance to a given data matrix  $U \in \mathbb{R}_{\geq 0}^{m \times n}$ , see [Gil20] for further details.

In Section 6, we apply the structured best rank-two approximation problem to NMF. Let  $\mathcal{M}_2$  be the set of matrices of nonnegative rank at most two and consider a matrix  $U \in \mathbb{R}_{\geq 0}^{3 \times 3}$ . In order to compute the best nonnegative rank-2 approximation of  $U$ , we need to compute the critical points of the Euclidean distance function  $d_U$  over  $\mathcal{M}_2 \cap \mathcal{L}^S$  for all zero patterns  $S \subset [3] \times [3]$ . We show that the minimal number of critical points needed to determine the global minimum of  $d_U$  is 756 for a generic  $U$ . For the same case, we show experimentally that the optimal critical point may have a few zeros.

The rest of the paper is organized as follows. In Section 2 we set our no-

tations (Section 2.1), we recall the basics of ED minimization on an algebraic variety (Section 2.2) and we discuss Frobenius distance minimization on a variety of low-rank matrices (Section 2.3). In Section 3 we address the best rank-one approximation problem with assigned zero patterns (Section 3.1) and best rank- $r$  approximation for rectangular and block diagonal matrices (Section 3.2). In Section 4 we investigate special polynomial relations among the critical points of  $d_U$ . In particular, in Sections 4.1 and 4.3 we concentrate on particular linear and affine relations among critical points respectively, and in Section 4.4 on some special nonlinear relations. Observations for generic linear constraints not necessarily coming from assigned zero patterns are given in Section 4.2. In Section 5 we provide conjectural ED degree formulas for special formats and zero patterns  $S$ , obtained from computational experiments. In Section 6 we relate the minimization problem (1.1) to nonnegative matrix factorization. The results of Sections 5 and 6 are supported by computations that use the HomotopyContinuation.jl [BT18] software package as well as the software Macaulay2 [GS] and Maple™ 2016 [Map]. The code can be found at [github.com/kaiekubjas/exact-solutions-in-low-rank-approximation-with-zeros](https://github.com/kaiekubjas/exact-solutions-in-low-rank-approximation-with-zeros).

## 2. Preliminaries

The preliminaries section consists of three subsections on algebra basics and notations (Section 2.1), Euclidean distance minimization (Section 2.2) and unstructured low-rank approximation (Section 2.3).

### 2.1. Algebra basics and notation

We start by setting up the notations used throughout the paper. We always work over the field  $\mathbb{F} = \mathbb{R}$  or  $\mathbb{F} = \mathbb{C}$ . Without loss of generality, we always assume that  $m \leq n$  when considering the vector space  $\mathbb{F}^{m \times n}$ . Given an  $m \times n$  matrix  $X \in \mathbb{F}^{m \times n}$ , we consider the subsets  $I \subset [m]$  and  $J \subset [n]$ . We always assume that the elements of  $I$  and  $J$  are ordered in increasing order. We denote by  $X_{I,J}$  the submatrix obtained by selecting the rows of  $X$  with indices in  $I$  and the columns of  $X$  with indices in  $J$ . Moreover, if  $|I| = |J|$  we denote by  $M_{I,J}(X)$  the minor of  $X$  corresponding to rows in  $I$  and columns in  $J$ . If  $I = J = \emptyset$ , we set  $M_{\emptyset,\emptyset}(X) := 1$ .

We use the notation  $\mathbb{F}[X]$  to denote the ring of polynomials with variables  $x_{ij}$  of  $X$  and with coefficients in  $\mathbb{F}$ . For a given subset  $Y \subset \mathbb{F}[X]$ , we indicate with  $\mathcal{V}(Y)$  the (*algebraic*) *variety* of  $Y$ , that is the zeros of  $Y$  in  $\mathbb{F}^{m \times n}$ . In particular  $\mathcal{V}(Y) = \mathcal{V}(\langle Y \rangle)$ , where  $\langle Y \rangle \subset \mathbb{F}[X]$  is the ideal generated by elements of  $Y$ . The subsets of  $\mathbb{F}^{m \times n}$  of the form  $\mathcal{V}(Y)$  for some  $Y \subset \mathbb{F}[X]$  constitute the closed sets

of the Zariski topology of  $\mathbb{F}^{m \times n}$ . In particular, a closed set  $\mathcal{V}(Y)$  is *irreducible* if it is not the union of two non-empty proper subsets that are closed in the Zariski topology. Finally, for a given subset  $Z \subset \mathbb{F}^{m \times n}$ , we denote by  $\mathcal{I}(Z)$  the *ideal of  $Z$* , namely the ideal generated by all polynomials vanishing at the points of  $Z$ . We refer the reader to [CLO92] for a more general commutative algebra background.

## 2.2. Euclidean distance minimization on a real algebraic variety

Let  $V^{\mathbb{R}}$  be a finite dimensional real vector space equipped with a positive definite symmetric bilinear form  $\langle \cdot, \cdot \rangle: V^{\mathbb{R}} \times V^{\mathbb{R}} \rightarrow \mathbb{R}$ . Without loss of generality, we can assume that  $\langle \cdot, \cdot \rangle$  is the standard Euclidean inner product, that is  $\langle x, y \rangle = \sum_i x_i y_i$ , for  $x, y \in V^{\mathbb{R}}$ . Consider a real algebraic variety  $\mathcal{X}^{\mathbb{R}} \subset V^{\mathbb{R}}$ . The generalized version of (1.1) is the problem of minimizing the squared Euclidean distance  $d_u(\cdot) := \langle u - \cdot, u - \cdot \rangle$  over  $\mathcal{X}^{\mathbb{R}}$ . For a generic data point  $u$ , the minimum of  $d_u$  is attained at a smooth point  $x \in \mathcal{X}^{\mathbb{R}}$ , such that the vector  $u - x$  is orthogonal to the tangent space  $T_x \mathcal{X}^{\mathbb{R}}$  with respect to  $\langle \cdot, \cdot \rangle$ . Such a point  $x$  is called a *critical point* of the function  $d_u$  on  $\mathcal{X}^{\mathbb{R}}$ . We can formulate the whole problem algebraically, since all relations among critical points are described by polynomials in the coordinates of  $x$  and  $u$ .

A complete study of the ideal of relations among critical points of  $d_u$  requires us to work over an algebraically closed field. For this reason we extend  $V^{\mathbb{R}}$  to a complex vector space  $V = V^{\mathbb{R}} \otimes \mathbb{C}$  and we let  $\mathcal{X}$  be the complex variety associated to  $\mathcal{X}^{\mathbb{R}}$ . Finally, we denote again with  $d_u$  the complex-valued function defined by  $d_u(x) := \langle u - x, u - x \rangle$ . We stress that  $\langle \cdot, \cdot \rangle$  is not the standard inner product in the complex vector space  $V$ . After this extension, one might study all complex critical points of the function  $d_u$  on  $\mathcal{X}$  and then look for the real global minimizer among all the complex solutions. The interesting property is that the number of complex critical points of  $d_u$  is an invariant of  $\mathcal{X}$  for a generic data point  $u \in V$  and is called the *Euclidean Distance degree* (ED degree) of  $\mathcal{X}$ . It is studied in detail in [DHO<sup>+</sup>16].

The bilinear form  $\langle \cdot, \cdot \rangle$  induces a smooth quadric hypersurface  $Q \subset V$  of equation  $\langle x, x \rangle = 0$ . It is called the *isotropic quadric* of  $V$ . In particular, for every  $x \in Q$ , the tangent space  $T_x Q$  is the hyperplane of vectors  $y \in V$  such that  $\langle x, y \rangle = 0$ . An affine subspace  $W \subset V$  is *transversal* to  $Q$  if for every  $x \in W \cap Q$ , the span between  $T_x Q$  and  $W$  generates  $V$ . In such case, the space  $V$  is the direct sum between  $W$  and its orthogonal complement. In turn, this yields a good notion of *orthogonal projection*  $\pi_W: V \rightarrow W$  onto  $W$  with respect to  $Q$ . We refer to [Pie15, Section 4] and [Sod20, Section 1.1] for more details on orthogonal projections with respect to the isotropic quadric  $Q$ .

The following two basic lemmas are needed in the following sections.

**Lemma 2.1.** *In the setting explained before, consider a proper affine subspace  $W \subset V$  transversal to the isotropic quadric  $Q$  and an algebraic variety  $X \subset W$ . Let  $\pi_W: V \rightarrow W$  be the orthogonal projection onto  $W$ . Then, for all  $u \in V$ , the critical points on  $X$  of the squared distance functions  $d_u$  and  $d_{\pi_W(u)}$  coincide.*

*Proof.* Let  $x \in X$  be a critical point of  $d_u$ . In particular  $\langle u - x, y \rangle = 0$  for all  $y \in T_x X$ . Furthermore  $\langle \pi_W(u) - x, y \rangle = \langle u - x, y \rangle - \langle u - \pi_W(u), y \rangle = 0 - 0 = 0$  since  $u - \pi_W(u)$  and  $y$  sit in orthogonal subspaces.  $\square$

**Lemma 2.2.** *Consider the affine varieties  $X_1, \dots, X_p$  in  $V$  and suppose that  $X_i \not\subset X_j$  for all  $i \neq j$ . Then  $\text{EDdegree}(X_1 \cup \dots \cup X_p) = \sum_{i=1}^p \text{EDdegree}(X_i)$ .*

*Proof.* Let  $u \in V$  be a generic data point and  $X = X_1 \cup \dots \cup X_p$ . A smooth point  $x \in X$  is critical for  $d_u$  if the vector  $u - x$  is orthogonal to the tangent space  $T_x X$ . Since  $x$  is a smooth point of  $X$ , there exists a unique index  $i \in [p]$  such that  $x$  is a smooth point of  $X_i$  and  $T_x X = T_x X_i$ . Here we are using the hypothesis  $X_i \not\subset X_j$  for all  $i \neq j$ . Therefore  $x$  is a critical point for  $d_u$  on  $X_i$ . This yields the inequality  $\text{EDdegree}(X) \leq \sum_{i=1}^p \text{EDdegree}(X_i)$ .

To prove the reverse inequality, consider an index  $i \in [p]$  and a critical point  $x \in X_i$  for  $d_u$ . If  $u$  is sufficiently generic, then  $x$  lies outside the singular locus of  $X$  and  $T_x X_i = T_x X$ . Therefore  $x$  is a critical point for  $d_u$  on  $X$ . Repeating the argument for every  $i \in [p]$ , we derive the desired inequality  $\text{EDdegree}(X) \geq \sum_{i=1}^p \text{EDdegree}(X_i)$ .  $\square$

### 2.3. Unstructured low-rank approximation

We consider the notation of Section 2.2. We specialize to the Euclidean space  $(V^{\mathbb{R}}, \langle \cdot, \cdot \rangle) = (\mathbb{R}^{m \times n}, \langle \cdot, \cdot \rangle_F)$ , where  $\langle \cdot, \cdot \rangle_F$  is the *Frobenius inner product* in  $\mathbb{R}^{m \times n}$ . It is defined by  $\langle A, B \rangle_F := \text{trace}(AB^T)$  for any two  $m \times n$  matrices  $A$  and  $B$ . Given  $U \in \mathbb{R}^{m \times n}$ , we minimize the squared Frobenius distance

$$d_U(X) = \langle X - U, X - U \rangle_F = \sum_{i=1}^m \sum_{j=1}^n (x_{ij} - u_{ij})^2,$$

over the real variety  $X^{\mathbb{R}} = X_r^{\mathbb{R}}$  of  $m \times n$  matrices of rank at most  $r$ . Note that  $d_U(X)$  coincides with the squared Euclidean distance between  $X$  and  $U$  in  $\mathbb{R}^{m \times n}$ .

An important tool to study low-rank approximation problem is given by the following decomposition of a real matrix.



**Theorem 2.3** (Singular Value Decomposition). *Any matrix  $U \in \mathbb{R}^{m \times n}$  admits the Singular Value Decomposition (SVD)*

$$U = A \Sigma B^T, \quad (2.1)$$

where  $A \in \mathbb{R}^{m \times m}$  and  $B \in \mathbb{R}^{n \times n}$  are orthogonal matrices and  $\Sigma \in \mathbb{R}^{m \times n}$  is such that  $\Sigma_{ii} = \sigma_i$  for some real numbers  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_m \geq 0$ , otherwise  $\Sigma_{ij} = 0$ . The numbers  $\sigma_i$  are called singular values of  $U$ . Denoting by  $a_i$  and  $b_i$  the columns of  $A$  and  $B$  respectively, for all  $i \in [m]$  the pair  $(a_i, b_i)$  is called a singular vector pair of  $U$ . If the singular values are all distinct, then all singular vector pairs are unique up to a simultaneous change of sign.

**Remark 2.4.** *Theorem 2.3 extends to complex matrices using unitary matrices and their conjugates, but complex conjugation is not an algebraic operation. If a matrix  $U \in \mathbb{C}^{m \times n}$  factors as in (2.1), then  $U$  admits an algebraic SVD. The complex matrices admitting an algebraic SVD are characterized in [CH87, Theorem 2 and Corollary 3], see also [DLOT17, Section 3].*

Similarly as in Section 2.2, we rephrase our low-rank approximation problem over  $\mathbb{C}$ . In particular, we work in the space  $\mathbb{C}^{m \times n}$  and we compute all critical points of the complex-valued function  $d_U(X) = \langle U - X, U - X \rangle_F$  on the variety  $\mathcal{X} = \mathcal{X}_r$  of complex  $m \times n$  matrices of rank at most  $r$ . The following result characterizes such critical points when the data matrix  $U$  is real. A similar statement holds for every complex data matrix  $U$  admitting an algebraic SVD. We report the statement used in [OP15, Theorem 2.9].

**Theorem 2.5** (Eckart-Young-Mirsky). *Consider a matrix  $U \in \mathbb{R}^{m \times n}$  of rank  $k$  and its SVD as in (2.1). Let  $r \in [k]$ . Then all the critical points of  $d_U$  on  $\mathcal{X}_r$  are of the form*

$$A(\Sigma_{i_1} + \dots + \Sigma_{i_r})B^T \quad (2.2)$$

for all subsets  $\{i_1 < \dots < i_r\} \subset [k]$ , where  $\Sigma_j$  is the  $m \times n$  matrix whose only non-zero entry is  $\Sigma_{j,j} = \sigma_j$ . If the non-zero singular values of  $U$  are distinct, then there are  $\binom{k}{r}$  critical points.

Therefore, Theorem 2.5 solves the best rank- $r$  approximation problem and the nice structure of critical points leads to various interesting consequences. In particular, assuming that  $U$  is full rank, their number is independent from the largest dimension  $n$ . Moreover, their linear span does not depend on the rank  $r$ ; it is studied in [OP15, DOT18] in the more general context of tensor spaces, see also Proposition 2.6.

Consider a data matrix  $U \in \mathbb{R}^{m \times n}$  and let  $Z_{U,r}$  be the set of critical points of  $d_U$  on  $\mathcal{X}_r$ . We denote by  $\langle Z_{U,r} \rangle$  the linear span of  $Z_{U,r}$  in  $\mathbb{R}^{m \times n}$ . A consequence of Theorem 2.5 is that  $\langle Z_{U,r} \rangle$  does not depend on the rank  $r$ , namely  $\langle Z_{U,r} \rangle = \langle Z_{U,1} \rangle$  for all  $r \in [m]$ . The following result is a special case of [DOT18, Theorem 1.1] and gives the equations of  $\langle Z_{U,1} \rangle$ .

**Proposition 2.6.** *Given  $U \in \mathbb{R}^{m \times n}$ , the ideal  $\mathcal{I}(Z_{U,1}) = \mathcal{I}(\langle Z_{U,1} \rangle)$  is generated by the linear forms in  $\mathbb{R}[X]_1$*

$$r_U^{(i,j)}(X) := (XU^T - UX^T)_{ij} \quad \forall i, j \in [m], \quad c_U^{(i,j)}(X) := (X^T U - U^T X)_{ij} \quad \forall i, j \in [n]. \quad (2.3)$$

*In particular, if  $U$  is sufficiently generic, then (assuming  $m \leq n$ )*

1. *the  $\binom{m}{2}$  forms  $r_U^{(i,j)}(X)$  are linearly independent,*
2.  *$m(n-1) - \binom{m}{2}$  of the forms  $c_U^{(i,j)}(X)$  are linearly independent,*
3. *none of the forms  $r_U^{(i,j)}(X)$  is linear combination of some of the forms  $c_U^{(i,j)}(X)$ , and viceversa.*

*Therefore  $\dim(\langle Z_{U,1} \rangle) = m$  if  $U$  is sufficiently generic.*

We explain an implication of Proposition 2.6 which is used in the proof of Corollary 4.8.

**Remark 2.7.** *Let  $U \in \mathbb{R}^{m \times n}$  and let  $U_{I,J}$  be a submatrix of  $U$ . Consider the system*

$$r_{U_{I,J}}^{(i,j)} = 0 \quad \forall i, j \in [I], \quad c_{U_{I,J}}^{(i,j)} = 0 \quad \forall i, j \in [J].$$

*If  $|I| \leq |J|$  and  $U$  is sufficiently generic, then*

1. *the first  $\binom{|I|}{2}$  equations are linearly independent,*
2. *there are  $|I|(|J| - 1) - \binom{|I|}{2}$  linearly independent equations in the second set,*
3. *no equation in one set is linear combination of equations in the other set.*

### 3. Rank-one structured approximation and beyond

This section is divided into two subsections: In Section 3.1, we focus on rank-one approximation with zeros, and in Section 3.2, on the simplest cases of rank- $r$  approximation with zeros for rectangular and block-diagonal matrices.

### 3.1. Rank-1 structured approximation

Rank-one approximation with zeros is built on the observation that non-zero entries of a rank-one matrix form a rectangular submatrix. After introducing the notions of rectangular zero patterns in Definition 3.1 and minimal covers in Definition 3.2, Proposition 3.4 gives the procedure for determining the best rank-one approximation in the structured setting. The main result in this subsection is Proposition 3.3 which gives the ED degree of  $\mathcal{L}_1^S$  in terms of the minimal covers. We recall that the variety  $\mathcal{L}_r^S = \mathcal{X}_r \cap \mathcal{L}^S$  was defined in (1.2) for any  $r \in [m]$ . For row/column and diagonal zero patterns, this results in explicit formulas (Corollaries 3.8 and 3.9). We end the section with Algorithm 1 for finding all minimal covers of a zero pattern  $S$ .

**Definition 3.1.** We say that a zero pattern  $S \subset [m] \times [n]$  is rectangular if the indices that are not in  $S$  form a rectangular matrix. More precisely, a rectangular zero pattern has the form  $S = (S_1 \times [n]) \cup ([m] \times S_2)$ , for some  $S_1 \subset [m]$  and  $S_2 \subset [n]$ . Sometimes we denote this zero pattern also by  $(S_1, S_2)$ .

**Definition 3.2.** Let  $S, T \subset [m] \times [n]$  be two zero patterns.

1. The zero pattern  $T$  is a cover of the zero pattern  $S$ , if  $S \subset T$  and if  $T$  is rectangular.
2. The zero pattern  $T$  is a minimal cover of  $S$ , if it is minimal among all covers of the zero pattern  $S$  with respect to inclusion. We denote by  $MC(S, m, n)$  the set of all minimal covers of  $S \subset [m] \times [n]$ .

For example, if  $S = \{(1, 1), (1, 2), (2, 2)\}$ , then  $MC(S, 3, 3) = \{([2], \emptyset), (\emptyset, [2]), (\{1\}, \{2\})\}$  (see Figure 1).

$$\left( \begin{array}{|c|c|c|} \hline 0 & 0 & 2 \\ \hline 1 & 0 & 1 \\ \hline 2 & 1 & 3 \\ \hline \end{array} \right) \quad \left( \begin{array}{|c|c|} \hline 0 & 0 & 2 \\ \hline 1 & 0 & 1 \\ \hline 2 & 1 & 3 \\ \hline \end{array} \right) \quad \left( \begin{array}{|c|c|c|} \hline 0 & 0 & 2 \\ \hline 1 & 0 & 1 \\ \hline 2 & 1 & 3 \\ \hline \end{array} \right)$$

Figure 1: Minimal covers of a  $3 \times 3$  matrix with zero pattern  $S = \{(1, 1), (1, 2), (2, 2)\}$ .

The main difficulty in studying the best rank-one approximation problem with assigned zero pattern  $S$  lies in identifying all the minimal covers of  $S$ .

**Proposition 3.3.** Let  $S \subset [m] \times [n]$  be a zero pattern. Consider the variety  $\mathcal{X}_1$  of  $m \times n$  matrices of rank at most one and the intersection  $\mathcal{L}_1^S = \mathcal{X}_1 \cap \mathcal{L}^S$ . Then

$$\text{EDdegree}(\mathcal{L}_1^S) = \sum_{(A_r, A_c) \in MC(S, m, n)} \min(m - |A_r|, n - |A_c|). \quad (3.1)$$

*Proof.* There is a bijection between the irreducible components of  $\mathcal{L}_1^S$  and the elements of  $MC(S, m, n)$ . More precisely, for every pair  $(A_r, A_c) \in MC(S, m, n)$ , the corresponding component of  $\mathcal{L}_1^S$  is isomorphic to the variety of  $(m - |A_r|) \times (n - |A_c|)$  matrices of rank at most one, and by Theorem 2.5 this component has ED degree equal to  $\min(m - |A_r|, n - |A_c|)$ . Moreover, given two distinct minimal covers  $T_1$  and  $T_2$  in  $MC(S, m, n)$  and their corresponding irreducible components  $\mathcal{V}_1$  and  $\mathcal{V}_2$ , then neither  $\mathcal{V}_1 \subset \mathcal{V}_2$  nor  $\mathcal{V}_2 \subset \mathcal{V}_1$ . The statement follows by Lemma 2.2.  $\square$

**Corollary 3.4.** *Let  $S \subset [m] \times [n]$ . Given  $U \in \mathbb{R}^{m \times n}$ , its best rank-one approximation with zeros in  $S$  is found by the following procedure:*

1. *Identify all minimal covers of  $S$ .*
2. *Find the best rank-one approximation for each of the minimal covers of  $S$ .*
3. *Choose the best rank-one approximation over all the minimal covers of  $S$ .*

**Corollary 3.5.** *Let  $S = \{(1, 1)\}$ . It has two minimal covers, i.e.,  $MC(S, m, n) = \{(\{1\}, \emptyset), (\emptyset, \{1\})\}$ . Given  $U \in \mathbb{R}^{m \times n}$ , its best rank-one approximation with a zero in  $S$  is found by first identifying the best rank-one approximations for the two minimal covers of  $S$  and then choosing out of the two rank-one matrices the one that minimizes the Euclidean distance to  $U$ .*

**Definition 3.6.** *Let  $S \subset [m] \times [n]$  be a zero pattern. The mask matrix of  $S$  is the  $m \times n$  matrix  $M^S$  with entries in  $\{0, 1\}$  whose  $(i, j)$ -th entry is the characteristic function  $\chi_S(i, j)$  of  $S$ .*

**Definition 3.7.** *Two zero patterns  $S_1$  and  $S_2$  are said to be permutationally equivalent if there exist permutation matrices  $P_1$  and  $P_2$  such that we can write  $M^{S_1} = P_1 M^{S_2} P_2$ .*

**Corollary 3.8.** *Let  $S \subset [m] \times [n]$  be a zero pattern which is permutationally equivalent to the row zero pattern  $\{(1, 1), \dots, (1, |S|)\}$ . Then*

$$\text{EDdegree}(\mathcal{L}_1^S) = \min\{m, n - |S|\} + \min\{m - 1, n\}. \quad (3.2)$$

*Similarly, let  $S \subset [m] \times [n]$  be a zero pattern which is permutationally equivalent to the column zero pattern  $\{(1, 1), \dots, (|S|, 1)\}$ . Then*

$$\text{EDdegree}(\mathcal{L}_1^S) = \min\{m, n - 1\} + \min\{m - |S|, n\}. \quad (3.3)$$

**Corollary 3.9.** *Let  $S \subset [m] \times [n]$  be a zero pattern which is permutationally equivalent to the diagonal zero pattern  $\{(1, 1), \dots, (|S|, |S|)\}$ . Then*

$$\text{EDdegree}(\mathcal{L}_1^S) = \sum_{j=0}^{|S|} \binom{|S|}{j} \min\{m - j, n - |S| + j\}. \quad (3.4)$$

Enumerating minimal covers of a zero pattern translates to the problem of enumerating minimal vertex covers of a bipartite graph. A bipartite graph  $G$  can be associated to a zero pattern of an  $m \times n$ -matrix  $X = (x_{ij})$  in the following way: the bipartite graph  $G$  has  $m$  and  $n$  vertices in the two parts, corresponding to the rows and columns of the matrix. The edges of  $G$  correspond to the zero entries of the matrix, i.e.  $(i, j) \in E(G)$  if and only if  $x_{ij} = 0$ . A (minimal) cover of a zero pattern is then equivalent to a (minimal) vertex cover of the corresponding bipartite graph. Since counting vertex covers in a bipartite graph is #P-complete [PB83], it follows that counting covers of a zero pattern is #P-complete. Recall that #P is the class of counting problems, where the goal is to count the number of solutions of a problem. We know that  $NP \subseteq \#P$  but we are not aware how high in the hierarchy #P is. We refer the interested reader to [AB09, Chapter 9] for further details.

We suggest Algorithm 1, that is based on dynamic programming, to find all minimal covers of a zero pattern  $S$ . To simplify notation, we present the algorithm for the bipartite graph  $G$  corresponding to the zero pattern  $S$ . In particular, we use the following notation. Let  $G = (U, V, E)$ , where  $U = \{u_1, \dots, u_m\}$  and  $V = \{v_1, \dots, v_n\}$  are the two parts of vertices. For  $u \in U$ , we denote by  $\mathcal{N}(u)$  the set of neighbors of  $u$ . Let  $U' \subset U$  and  $V' \subset V$ . We denote by  $G[U', V']$  the induced subgraph of  $G$ , i.e., the graph whose vertex set is  $U' \cup V'$  and whose edge set is the subset of  $E$  that consists of edges whose both endpoints are in  $U' \cup V'$ . We consider the graph  $G[U', V']$  as a bipartite graph with  $U'$  and  $V'$  being the two parts of vertices.

<p><b>Algorithm 1:</b> Minimal covers of a bipartite graph  <math>G = (U, V, E)</math></p> <pre> 1: <b>procedure</b> MINIMALCOVERS(<math>G = (U, V, E)</math>) 2: <b>if</b> <math>G</math> is null graph <b>then</b> 3:   <b>return</b> <math>\{(\emptyset, \emptyset)\}</math> 4: <b>else</b> 5:   <math>MC = \emptyset</math> 6:   <math>MC_1 = \text{MINIMALCOVERS}(G[U \setminus \{u_1\}, V])</math> 7:   <b>for</b> <math>(S_1, S_2) \in MC_1</math> <b>do</b> 8:     append <math>(S_1 \cup \{u_1\}, S_2)</math> to <math>MC</math> 9:   <b>end for</b> 10:  <math>MC_2 = \text{MINIMALCOVERS}(G[U \setminus \{u_1\}, V \setminus \mathcal{N}(u_1)])</math> 11:  <b>for</b> <math>(S_1, S_2) \in MC_2</math> <b>do</b> 12:    append <math>(S_1, S_2 \cup \mathcal{N}(u_1))</math> to <math>MC</math> 13:  <b>end for</b> 14:  <b>return</b> <math>MC</math> 15: <b>end if</b> 16: <b>end procedure</b> </pre>
--

The following example illustrates that it is not enough, even to consider minimal covers with the least number of elements.

**Example 3.10.** Consider the  $3 \times 4$  matrix

$$U = \begin{pmatrix} 1 & -1 & -2 & -2 \\ 1 & 0 & 1 & -2 \\ 2 & 0 & 0 & 2 \end{pmatrix}.$$

We look for the closest rank-one matrix to  $U$  with zero pattern  $S = \{(1, 1), (1, 2)\}$ . We have  $MC(S, 3, 4) = \{(\{1\}, \emptyset), (\emptyset, [2])\}$ . In particular, the first minimal cover consists of four elements, while the second minimal cover consists of six elements. One verifies that the closest critical point to  $U$  is of the second type and is equal to

$$X = \begin{pmatrix} 0 & 0 & -0.627896 & -2.36438 \\ 0 & 0 & -0.430261 & -1.62017 \\ 0 & 0 & 0.496139 & 1.86824 \end{pmatrix}.$$

### 3.2. Rank- $r$ structured approximation for rectangular and block-diagonal matrices

We finish this section with two results for structured best rank- $r$  approximation problem in the simplest cases where non-zero entries form rectangular and block-diagonal submatrices.

**Lemma 3.11** (Rectangular low-rank approximation). *Let  $S \subseteq [m] \times [n]$  be such that  $([m] \times [n]) \setminus S = T' \times T''$  for some  $T' \subseteq [m]$  and  $T'' \subseteq [n]$ . Given  $U \in \mathbb{R}^{m \times n}$ , let  $U' \in \mathbb{R}^{m \times n}$  be the matrix that is obtained from  $U$  by setting the entries in  $S$  equal to zero. The best rank- $r$  approximation of  $U$  with zeros in  $S$  is equal to the best (unstructured) rank- $r$  approximation of  $U'$ . If the non-zero singular values of  $U'$  are distinct, then by Theorem 2.5 there are  $\binom{\min(|T'|, |T''|)}{r}$  critical points and they are given by the Singular Value Decomposition (2.1) of  $U'$ .*

**Lemma 3.12** (Block-diagonal low-rank approximation). *Let  $S \subseteq [m] \times [n]$  be such that  $([m] \times [n]) \setminus S = (T'_1 \times T''_1) \cup (T'_2 \times T''_2) \cup \dots \cup (T'_s \times T''_s)$  for some non-empty pairwise disjoint  $T'_1, T'_2, \dots, T'_s \subseteq [m]$  and non-empty pairwise disjoint  $T''_1, T''_2, \dots, T''_s \subseteq [n]$ . Consider all vectors  $(r_1, \dots, r_s) \in \mathbb{N}_0^s$  such that  $\sum r_i = r$ . For a fixed vector  $(r_1, \dots, r_s)$ , let  $Z(r_1, \dots, r_s)$  be the set of matrices such that for  $1 \leq i \leq s$ , the projection to  $T'_i \times T''_i$  is a critical point of the unstructured rank- $r_i$  approximation problem for  $U|_{T'_i \times T''_i}$ . The union of  $Z(r_1, \dots, r_s)$  over all vectors  $(r_1, \dots, r_s) \in \mathbb{N}_0^s$  satisfying  $\sum r_i = r$  gives the set of all critical points for the the structured best rank- $r$  approximation problem (1.1) for  $U$ .*

**Example 3.13.** *Consider the  $3 \times 3$  matrix*

$$U = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$

*Let  $S = \{(1, 3), (2, 3), (3, 1), (3, 2)\}$ . We look for the closest rank-two matrix to  $U$  with zero pattern  $S$ . The support of this closest rank-two matrix is block-diagonal, i.e., the non-zero entries are  $([3] \times [3]) \setminus S = (\{1, 2\} \times \{1, 2\}) \cup (\{3\} \times \{3\})$ . There are  $s = 2$  blocks and three vectors  $(r_1, r_2) \in \mathbb{N}_0^2$  satisfying  $r_1 + r_2 = 2$ , namely the vectors  $(2, 0), (1, 1)$  and  $(0, 2)$ . The critical point corresponding to the vector  $(2, 0)$  is*

$$C_1 = \begin{pmatrix} 1 & 2 & 0 \\ 4 & 5 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The critical points corresponding to the vector  $(1, 1)$  are

$$C_2 = \begin{pmatrix} 1.3332 & 1.7455 & 0 \\ 3.8857 & 5.0873 & 0 \\ 0 & 0 & 9 \end{pmatrix} \quad \text{and} \quad C_3 = \begin{pmatrix} -0.3332 & 0.2545 & 0 \\ 0.1143 & -0.0873 & 0 \\ 0 & 0 & 9 \end{pmatrix}.$$

There are no critical points corresponding to the vector  $(0, 2)$ , since the size of the last block is one and hence it cannot have rank two. The critical point that minimizes the squared Frobenius distance to  $U$  is  $C_2$ .

#### 4. Special relations among critical points

In this section we provide (some of) the generators of the ideal of critical points on  $\mathcal{L}_r^S$  of  $d_U$ . In particular, in Sections 4.1 and 4.3 we concentrate on particular linear and affine relations among critical points respectively, and in Section 4.4 on some special nonlinear relations. Observations for generic linear constraints not necessarily coming from assigned zero patterns are given in Section 4.2.

We stress that in our statements we always consider a real  $m \times n$  matrix  $U$ . However, as we explained in Section 2.3, our optimization problem is defined in the ambient space of  $m \times n$  complex matrices, since in the structured setting the critical points of  $d_U$  are not necessarily real.

##### 4.1. Linear relations among critical points

Let  $Z_{U,r}^S$  be the set of critical points of  $d_U$  on  $\mathcal{L}_r^S$ . When  $S = \emptyset$  we simply write  $Z_{U,r}$  as in Section 2.3. The study of the linear span  $\langle Z_{U,r}^S \rangle$  is more involved when  $S \neq \emptyset$ . The main result in this section is Theorem 4.3 which states that certain linear equations from the unstructured setting in Proposition 2.6 are satisfied by the rank- $r$  critical points of  $d_U$  in the structured setting. In the rest of the section, we investigate when do these linear equations define  $\langle Z_{U,r}^S \rangle$  (Conjecture 4.4, Examples 4.5-4.7) and give a lower bound on the codimension of  $\langle Z_{U,r}^S \rangle$  (Corollary 4.8).

**Definition 4.1.** Let  $S \subset [m] \times [n]$  be a zero pattern. Recall that  $M^S$  denotes the mask matrix of  $S$  as introduced in Definition 3.6. We introduce the following equivalence relations  $\sim_R^S$  and  $\sim_C^S$  in the sets  $[m]$  and  $[n]$ , respectively:

- $i \sim_R^S j$  if and only if the  $i$ -th and the  $j$ -th rows of  $M^S$  coincide,
- $i \sim_C^S j$  if and only if the  $i$ -th and the  $j$ -th columns of  $M^S$  coincide.



**Definition 4.2.** We define critical space of  $U \in \mathbb{R}^{m \times n}$  to be the linear space  $H_U^S$  defined by relations

$$\langle r_U^{(i,j)} \mid i \sim_R^S j \rangle + \langle c_U^{(i,j)} \mid i \sim_C^S j \rangle + \langle x_{ij} \mid (i,j) \in S \rangle. \quad (4.1)$$

Observe that  $H_U^S$  does not depend on the rank  $r$ . Moreover, the previous definition is inspired by [DOT18, Definition 2.8] in the context of unstructured low-rank tensor approximation. In particular, for  $S = \emptyset$  we denote the critical space simply by  $H_U$ .

Proposition 2.6 tells us that  $\langle Z_{U,r} \rangle = H_U$  for all  $r \in [m-1]$ . The next result shows that at least one of the two inclusions in the last equality is still true when  $S \neq \emptyset$ .

**Theorem 4.3.** Let  $S \subset [m] \times [n]$ ,  $U \in \mathbb{R}^{m \times n}$  and  $r \in [m-1]$ . Then  $\langle Z_{U,r}^S \rangle \subset H_U^S$ .

*Proof.* Let  $L_1, \dots, L_s$  be the constraints that set  $s$  entries of the matrix to be equal to zero. We denote by  $\text{Jac}_{\mathcal{X}_r}(X)$  and  $\text{Jac}_{\mathcal{L}^S}(X)$  the Jacobian matrices of  $\mathcal{X}_r$  and  $\mathcal{L}^S$  evaluated at  $X$ , respectively. The rank- $r$  critical points  $X \in \mathcal{L}_r^S$  of  $d_U$  satisfy the equality constraints

$$\begin{cases} M_{I,J}(X) = 0 & \forall |I| = |J| = r+1 \\ L_k(X) = 0 & \forall k \in [s] \end{cases} \quad \left[ \lambda \mid \mu \mid 1 \right] \begin{bmatrix} \text{Jac}_{\mathcal{X}_r}(X) \\ \text{Jac}_{\mathcal{L}^S}(X) \\ X - U \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}, \quad (4.2)$$

where  $\lambda = (\lambda_{I,J})_{I,J}$  and  $\mu = (\mu_1, \dots, \mu_s)$  are vectors of Lagrange multipliers.

We denote by  $v$  the vector of polynomials that is obtained when multiplying the vector and the augmented Jacobian matrix in (4.2). Its entries are naturally indexed by  $(1, 1), \dots, (m, n)$ . Let

$$X^{i \leftrightarrow j} := \left[ 0 \quad \dots \quad 0 \quad -x_{j1} \quad \dots \quad -x_{jn} \quad 0 \quad \dots \quad 0 \quad x_{i1} \quad \dots \quad x_{in} \quad 0 \quad \dots \quad 0 \right]^T$$

be the vector with the entry  $-x_{jk}$  at the position  $(i, k)$  and the entry  $x_{ik}$  at the position  $(j, k)$  for all  $k \in [n]$ .

We show that  $v \cdot X^{i \leftrightarrow j}$  is equal to a linear constraint in (4.1) plus some  $(r+1)$ -minors  $M_{I,J}(X)$  and linear constraints  $L_k(X)$  multiplied with Lagrange multipliers. To do this, we study the products of the rows of the augmented Jacobian with the vector  $X^{i \leftrightarrow j}$ .

First, observe that the last row of the augmented Jacobian multiplied with  $X^{i \leftrightarrow j}$  is precisely the linear form  $r_U^{(i,j)}$ . Secondly, we show that the rows of the augmented Jacobian corresponding to minors multiplied with  $X^{i \leftrightarrow j}$  are either zero or

a sum of  $(r+1)$ -minors. Let  $A = \{a_1, \dots, a_{r+1}\} \subset [m]$  and  $B = \{b_1, \dots, b_{r+1}\} \subset [n]$ . We consider the product

$$\left[ \frac{\partial M_{A,B}(X)}{\partial x_{i_1}} \quad \dots \quad \frac{\partial M_{A,B}(X)}{\partial x_{i_r}} \right] \cdot X^{(j)}. \quad (4.3)$$

If  $i \notin A$ , then the product (4.3) is equal to zero. Otherwise  $i \in A$  and the product (4.3) can be seen as the Laplace expansion of the matrix with rows in  $(A \setminus \{i\}) \cup \{j\}$  considered as a multiset and columns in  $B$ . Hence if  $j \notin A$ , then the product (4.3) is equal to the minor corresponding to rows in  $(A \setminus \{i\}) \cup \{j\}$  and columns in  $B$ . Finally, if  $j \in A$ , then the product (4.3) is zero again, because the row indexed by  $j$  appears twice.

Finally, we consider the rows of the augmented Jacobian corresponding to constraints that  $x_{kl} = 0$ . The Jacobian of  $x_{kl}$  consists of the entry  $(k, l)$  being equal to one and all other entries being equal to zero. If  $i \neq k$  and  $j \neq k$ , then the Jacobian of this constraint multiplied by  $X^{i \leftrightarrow j}$  is clearly zero. Otherwise the Jacobian of  $x_{kl}$  multiplied by  $X^{i \leftrightarrow j}$  is either  $x_{il}$  or  $x_{jl}$ . However both  $x_{il} = x_{jl} = 0$  by the assumption  $i \sim_R^S j$ .

This proves that  $r_U^{(i,j)} \in \mathcal{I}(Z_{U,r}^S)$  for all  $i \sim_R^S j$ . Similarly, one verifies that  $c_U^{(i,j)} \in \mathcal{I}(Z_{U,r}^S)$  for all  $i \sim_C^S j$  by applying the same argument with the vector

$$X_{i \leftrightarrow j} := \left[ -x_{1j} \quad 0 \quad \dots \quad 0 \quad x_{1i} \quad 0 \quad \dots \quad 0 \quad -x_{nj} \quad 0 \quad \dots \quad 0 \quad x_{ni} \quad 0 \quad \dots \quad 0 \right]^T. \quad \square$$

**Conjecture 4.4.** *Let  $S \subset [m] \times [n]$ ,  $U \in \mathbb{R}^{m \times n}$  and  $r \in [m-1]$ . Then  $\langle Z_{U,r}^S \rangle = H_U^S$  if and only if  $\mathcal{L}_r^S$  is irreducible.*

**Example 4.5** ( $m = n = 3, r = 2, S = \{(1, 1)\}$ ). *By experimental computation, we observe that  $\text{EDdegree}(\mathcal{L}_2^S) = 8$  (see Table 2). If linearly independent, the eight critical points should span an eight-dimensional linear space  $\langle Z_{U,2}^S \rangle \subset \mathbb{C}^{3 \times 3}$ . We verified symbolically using Gröbner bases and elimination that  $\mathcal{I}(\langle Z_{U,2}^S \rangle) = \langle r_U^{(2,3)}, c_U^{(2,3)}, x_{11} \rangle$ , thus confirming Conjecture 4.4.*

We will show in Examples 4.6 and 4.7 that if  $\mathcal{L}_r^S$  is not irreducible, then  $\langle Z_{U,r}^S \rangle$  can be strictly contained in  $H_U^S$ . Example 4.6 is for structured rank-one approximation. We know from Proposition 3.3 that  $\mathcal{L}_1^S$  is never irreducible if  $S \neq \emptyset$ , and hence Conjecture 4.4 suggests that in the structured setting  $\langle Z_{U,1}^S \rangle$  is always strictly contained in  $H_U^S$ . Example 4.7 is for rank-two approximation.

**Example 4.6** ( $m = n = 3, r = 1, S = \{(1, 1)\}$ ). *In this case, the variety  $\mathcal{L}_1^S$  has two irreducible components corresponding to the minimal coverings  $(\{1\}, \emptyset)$  and*

$(\emptyset, \{1\})$ . Moreover  $\text{EDdegree}(\mathcal{L}_1^S) = 4$  by Corollary 3.9. The four critical points on  $\mathcal{L}_1^S$  are obtained in this way:

- (i) by computing the SVD of the  $3 \times 3$  matrix having zero first row and coinciding with  $U$  elsewhere (two critical points  $C_1, C_2$ ),
- (ii) by computing the SVD of the  $3 \times 3$  matrix having zero first column and coinciding with  $U$  elsewhere (two critical points  $C_3, C_4$ ).

One verifies immediately that the critical space  $H_U^S$  is six-dimensional. Therefore  $\langle Z_{U,1}^S \rangle$  is strictly contained in  $H_U^S$  and motivates our hypothesis in Conjecture 4.4.

A higher rank example is showed below.

**Example 4.7** ( $m = n = 3, r = 2, S = \{(1, 1), (1, 2)\}$ ). The determinant of a  $3 \times 3$  matrix  $X = (x_{ij})$  with zero pattern  $S$  is  $\det(X) = x_{13}(x_{21}x_{32} - x_{31}x_{22})$ . Then the variety  $\mathcal{L}_2^S$  has two components  $\mathcal{V}_1 = \mathcal{V}(x_{11}, x_{12}, x_{13})$ ,  $\mathcal{V}_2 = \mathcal{V}(x_{11}, x_{12}, x_{21}x_{32} - x_{31}x_{22})$  and by Lemma 2.2

$$\text{EDdegree}(\mathcal{L}_2^S) = \text{EDdegree}(\mathcal{V}_1) + \text{EDdegree}(\mathcal{V}_2) = 1 + 2 = 3.$$

The critical point on  $\mathcal{V}_1$  is the projection of  $U$  onto  $\mathcal{V}_1$ . The two critical points on  $\mathcal{V}_2$  come by projecting the third column of  $U$  and computing the SVD of the non-zero  $2 \times 2$  block. In particular, the linear span  $\langle Z_{U,2}^S \rangle$  is three-dimensional, whereas the critical space  $H_U^S$  has dimension five.

**Corollary 4.8.** Let  $S \subset [m] \times [n]$ . If  $U$  is sufficiently generic, then the codimension of  $H_U^S \subset \mathbb{C}^{m \times n}$  is

$$\text{codim}(H_U^S) = \sum_{C \in [m]/\sim_R^S} \binom{|C|}{2} + \sum_{D \in [n]/\sim_C^S} \gamma_D + |S|,$$

where

$$\gamma_D = \begin{cases} \binom{|D|}{2} & \text{if } |D| \leq m \\ m(|D| - 1) - \binom{m}{2} & \text{if } |D| \geq m. \end{cases}$$

*Proof of Corollary 4.8.* We have to show how many of all the linear polynomials appearing in (4.1) are linearly independent.

The first two sets of generators in (4.1) are precisely of the type explained in Remark 2.7. On one hand, since  $m \leq n$ , then  $|C| \leq n$  for every equivalence class  $C \in [m]/\sim_R^S$ . This means that all relations  $r_U^{(i,j)}$ , where  $i, j \in C$ , are linearly

independent. By Remark 2.7 and since equivalence classes on rows are disjoint, this gives in total the first  $\sum_{C \in [m]/\sim_R^S} \binom{|C|}{2}$  independent conditions. On the other hand, if  $D \in [n]/\sim_C^S$  and  $|D| \leq m$ , then again by Remark 2.7 all relations  $c_U^{(i,j)}$ , where  $i, j \in D$ , are linearly independent, thus giving  $\binom{|D|}{2}$  independent conditions. Otherwise if  $|D| \geq m$ , then  $m(|D| - 1) - \binom{m}{2}$  among the last equations are linearly independent. Since equivalence classes on columns are disjoint, this gives in total the second  $\sum_{D \in [n]/\sim_C^S} \gamma_D$  independent conditions. Moreover, again by Remark 2.7, each condition on rows is not a linear combination of equations involving columns, and vice versa.

Finally, consider the last  $|S|$  conditions coming from the zero pattern. Trivially each relation  $x_{ij}$  is independent from the other variables  $x_{rs}$  with  $(r, s) \in S$ . Moreover, all the conditions  $\{x_{ij} \mid (i, j) \in S\}$  are independent from the first two sets of equations because the first two contain no variables with indices in  $S$ .  $\square$

If Conjecture 4.4 is true, then the statement in Corollary 4.8 holds for  $\langle Z_{U,r}^S \rangle$  when  $\mathcal{L}_r^S$  is irreducible.

**Remark 4.9.** Consider again the situation of Example 4.5 with  $m = n = 3$ ,  $r = 2$ ,  $S = \{(1, 1)\}$ . We recall that in this case  $\mathcal{I}(\langle Z_{U,2}^S \rangle) = \langle r_U^{(2,3)}, c_U^{(2,3)}, x_{11} \rangle$ , thus confirming Conjecture 4.4. In particular  $\dim(\langle Z_{U,2}^S \rangle) = \dim(H_U^S) = 6$ . One might try to extend the basis  $\{C_1, C_2, C_3, C_4\}$  of  $\langle Z_{U,1}^S \rangle$  given in Example 4.6 to form a basis of  $\langle Z_{U,2}^S \rangle$ , in the most “natural” way. In this example, we consider the additional two rank-one matrices  $C_5, C_6$  obtained by computing the SVD of the  $3 \times 3$  matrix having zero first row and column and coinciding with  $U$  elsewhere. One might check that the six rank-one matrices  $C_1, \dots, C_6$  are linearly independent and form a basis of  $H_U^S$ .

The matrices  $C_5$  and  $C_6$  are “good” in the sense that they are computed directly from the data matrix  $U$  via projections and SVDs. Any critical point  $X \in \mathcal{L}_2^S$  may be written as  $X = \alpha_1 C_1 + \dots + \alpha_6 C_6$  for some complex coefficients  $\alpha_1, \dots, \alpha_6$ . In Table 1 we display the coefficients  $\alpha_i$  of the eight critical points on  $\mathcal{L}_2^S$  with respect to the data matrix

$$U = \begin{pmatrix} 78.57 & 93.47 & 51.33 \\ -58.54 & -7.64 & 34.34 \\ 53.53 & -89.96 & -87.14 \end{pmatrix}.$$

More generally, knowing the ideal of critical points  $X \in \mathcal{L}_r^S$  for  $d_U$  and a basis  $\{C_1, \dots, C_k\}$  of  $\langle Z_{U,r}^S \rangle$  whose elements depend only on  $U$ , allows to com-

	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	$\alpha_6$
$X_1$	0.96	-0.11	0.01	-0.27	0.04	1.06
$X_2$	0.02	1.05	0.00	1.13	-0.03	-1.06
$X_3$	0.04	-0.12	0.99	-0.29	-0.01	0.77
$X_4$	1.00	0.99	1.00	0.97	-1.00	-0.93
$X_5$	$0.77-0.87i$	$-1.39-0.41i$	$0.73-0.30i$	$-2.67-0.72i$	$-0.28+1.31i$	$3.75+3.41i$
$X_6$	$0.77+0.87i$	$-1.39+0.41i$	$0.73+0.30i$	$-2.67+0.72i$	$-0.28-1.31i$	$3.75-3.41i$
$X_7$	$2.45-0.71i$	$-0.86-3.45i$	$1.65-0.37i$	$2.81+5.93i$	$-3.59+1.46i$	$3.73-4.66i$
$X_8$	$2.45+0.71i$	$-0.86+3.45i$	$1.65+0.37i$	$2.81-5.93i$	$-3.59-1.46i$	$3.73+4.66i$

Table 1: The critical points  $X_i \in \mathcal{L}_2^S \subset \mathbb{C}^{3 \times 3}$  of  $d_U$  in the basis  $\{C_1, \dots, C_6\}$ .

pute the ideal  $\mathcal{J}_\alpha \subset \mathbb{C}[\alpha_1, \dots, \alpha_k]$  of relations among the coefficients  $\alpha_i$  of a representation of  $X$ .

This idea needs further investigation and is motivated by the unstructured case. Indeed, given  $U \in \mathbb{R}^{m \times n}$ , the critical points  $C_1, \dots, C_m$  on  $\mathcal{X}_1$  of  $d_U$  computed from the SVD of  $U$  form a basis of  $\langle Z_{U,r}^S \rangle = H_U^S$  for any  $r \in [m]$ . By Theorem 2.5, the critical points on  $\mathcal{X}_r$  are written uniquely as  $X = \alpha_1 C_1 + \dots + \alpha_m C_m$  for some coefficients  $\alpha_j \in \{0, 1\}$ . In this case, the ideal  $\mathcal{J}_\alpha$  is zero-dimensional in  $\mathbb{C}[\alpha_1, \dots, \alpha_m]$  and its degree is equal to  $\text{EDdegree}(\mathcal{X}_r) = \binom{m}{r}$ .

#### 4.2. Linear relations among critical points for generic subspaces

The statement of Theorem 4.3 can be adapted to arbitrary linear subspaces  $\mathcal{L}$  of  $\mathbb{C}^{m \times n}$ .

**Proposition 4.10.** *Let  $U \in \mathbb{R}^{m \times n}$ . Let  $\mathcal{L} \subset \mathbb{C}^{m \times n}$  be the linear subspace defined by the linear forms  $L_1, \dots, L_s$  and let  $Z_{U,r}^\mathcal{L}$  be the set of critical points of  $d_U$  on the variety  $\mathcal{L}_r$ . For all  $i < j$ ,  $i, j \in [m]$  and  $k \in [s]$ , if  $\nabla L_k \cdot X^{i \leftrightarrow j} \in \langle L_1, \dots, L_s \rangle$ , then  $r_U^{(i,j)} \in \mathcal{I}(Z_{U,r}^\mathcal{L})$ . Similarly, for all  $i < j$ ,  $i, j \in [n]$  and  $k \in [s]$ , if  $\nabla L_k \cdot X_{i \leftrightarrow j} \in \langle L_1, \dots, L_s \rangle$ , then  $c_U^{(i,j)} \in \mathcal{I}(Z_{U,r}^\mathcal{L})$ .*

Symbolic computations in Macaulay2 motivate the following question for generic subspaces.

**Question 4.11.** *Let us replace  $\mathcal{L}^S$  with a non-zero proper subspace  $\mathcal{L} \subset \mathbb{C}^{m \times n}$ . Is it true that  $\mathcal{L} = \langle Z_{U,r}^\mathcal{L} \rangle$  for every  $r \in [m-1]$  if  $\mathcal{L}$  is generic?*

We have checked symbolically with a Macaulay2 code that  $\mathcal{L} = \langle Z_{U,r}^\mathcal{L} \rangle$  when  $\mathcal{L}$  is a generic subspace of codimension  $\text{codim}(\mathcal{L}) \in \{1, 2\}$  and for small formats. This suggests that the answer to Question 4.11 is positive. As showed in Theorem 4.3, the same is not true for  $\mathcal{L} = \mathcal{L}^S$  for some zero pattern  $S$ .

### 4.3. The affine span of critical points

In the previous sections, we regarded structured critical points on  $\mathcal{L}_r^S$  as vectors in  $\mathbb{C}^{m \times n}$  and studied all linear relations among them. In this section, we look for their affine span. This is important to understand how the critical points of  $d_U$  are distributed inside the critical space  $H_U^S$ . The affine span in the unstructured case is characterized in Proposition 4.12 and Corollary 4.13. Conjecture 4.15 suggests affine relations satisfied by the rank- $r$  critical points in the structured setting.

**Proposition 4.12.** *Assume  $m = n$  and  $S = \emptyset$ . Given  $U \in \mathbb{R}^{m \times m}$  of rank  $k$ , for every  $r \in [k]$  the  $\binom{k}{r}$  critical points of  $d_U$  on  $\mathcal{X}_r$  span an affine hyperplane  $W_{U,r}^S \subset H_U^S$  of equation*

$$W_{U,r}^S: \langle X, C(U) \rangle_F - r \det(U) = 0,$$

where  $C(U)$  is the cofactor matrix of  $U$  whose  $(i, j)$ -th entry is  $(-1)^{i+j} M_{[m] \setminus \{i\}, [m] \setminus \{j\}}(U)$ . In particular, if  $\det(U) \neq 0$ , then  $\{W_{U,r}^S\}_{r \in [m]}$  is a finite family of parallel hyperplanes contained in  $H_U^S$ . Otherwise if  $\det(U) = 0$ , then  $W_{U,r}^S$  is a linear hyperplane whose equation does not depend on the rank  $r$ .

*Proof.* Let  $U \in \mathbb{R}^{m \times m}$  of rank  $k$ , written in SVD form as  $U = A\Sigma B^T$ . One verifies immediately that  $C(U) = AC(\Sigma)B^T$  and that the  $j$ -th diagonal entry of  $C(\Sigma)$  is equal to  $\prod_{l \neq j} \sigma_l$  for all  $j \in [m]$ . Given  $r \in [k]$ , by Theorem 2.5 a critical point  $X \in \mathcal{X}_r$  of  $d_U$  is of the form  $X = A\Sigma_I B^T$  for some  $I = \{i_1 < \dots < i_r\} \subset [m]$ , where  $\Sigma_I = \text{diag}(0, \dots, \sigma_{i_1}, \dots, \sigma_{i_r}, \dots, 0)$ . Without loss of generality, assume  $I = [r] \subset [m]$ . We denote by  $a_1, \dots, a_m$  and  $b_1, \dots, b_m$  the orthonormal columns of  $A$  and  $B$ , respectively. Then

$$\begin{aligned} \langle X, C(U) \rangle_F &= \left\langle A\Sigma_{[r]}B^T, AC(\Sigma)B^T \right\rangle_F \\ &= \left\langle \sum_{i=1}^r \sigma_i a_i b_i^T, \sum_{j=1}^m \left( \prod_{l \neq j} \sigma_l \right) a_j b_j^T \right\rangle_F \\ &= \sum_{i=1}^r \sum_{j=1}^m \sigma_i \left( \prod_{l \neq j} \sigma_l \right) \langle a_i b_i^T, a_j b_j^T \rangle_F \\ &= \sum_{i=1}^r \sigma_i \left( \prod_{l \neq i} \sigma_l \right) = r \det(\Sigma) = r \det(U), \end{aligned}$$

where we have applied the identity ( $\delta_{ij}$  is the Kronecker delta)

$$\langle a_i b_i^T, a_j b_j^T \rangle_F = \text{trace}(a_i b_i^T b_j a_j^T) = \delta_{ij} \text{trace}(a_i a_i^T) = \delta_{ij} a_i^T a_i = \delta_{ij}. \quad \square$$

The following corollary generalizes Proposition 4.12 to non-squared case.

**Corollary 4.13.** *Assume  $m \leq n$  and  $S = \emptyset$ . Given  $U \in \mathbb{R}^{m \times n}$  of rank  $k$ , for every  $r \in [k]$  the  $\binom{k}{r}$  critical points of  $d_U$  on  $\mathcal{X}_r$  span an affine hyperplane  $W_{U,r}^S \subset H_U^S$  of equation*

$$W_{U,r}^S: \langle X_{[m],I}, C(U_{[m],I}) \rangle_F - r \det(U_{[m],I}) = 0,$$

where  $I = \{i_1 < \dots < i_m\} \subset [n]$  and, modulo the ideal  $\mathcal{I}(H_U^S)$ , the above equation does not depend on the particular choice of  $I$ .

In the next example we observe which of the affine relations of Corollary 4.13 still hold true in the structured case.

**Example 4.14.** *Let  $U \in \mathbb{R}^{3 \times 4}$ . In this example we investigate the best rank-two approximation of  $U$  with zero pattern  $S = \{(1, 1)\}$ . We verified symbolically that  $\text{EDdegree}(\mathcal{L}_r^S) = 8$ . The eight critical points of  $d_U$  span a seven-dimensional linear space  $\langle Z_{U,2}^S \rangle = H_U^S$  whose ideal is*

$$\mathcal{I}(H_U^S) = \langle r_U^{(2,3)}, c_U^{(2,3)}, c_U^{(2,4)}, c_U^{(3,4)}, x_{11} \rangle.$$

Moreover, the critical points satisfy the affine relation in  $H_U^S$

$$\langle X_{[3],I}, C(U_{[3],I}) \rangle_F - 2 \det(U_{[3],I}) = 0 \text{ for } I = \{2, 3, 4\}.$$

The previous example motivates the following conjecture in structured setting.

**Conjecture 4.15.** *Assume that  $\mathcal{L}_r^S$  is irreducible. Given a subset  $I \subset [n]$  with  $|I| = m$ , the complex critical points of  $d_U$  on  $\mathcal{L}_r^S$  satisfy the additional affine relation*

$$\langle X_{[m],I}, C(U_{[m],I}) \rangle_F - r \det(U_{[m],I}) = 0 \quad (4.4)$$

if and only if  $S \cap ([m] \times I) = \emptyset$ .

Conjecture 4.15 is not valid if  $\mathcal{L}_r^S$  is not irreducible, as showed in the next example.

**Example 4.16** ( $m = n = 3, r = 2, S = \{(1, 1), (1, 2)\}$ ). Following up Example 4.7, we observe that the ideal of affine relations among the critical points is

$$\left\langle r_U^{(2,3)}, c_U^{(1,2)}, x_{11}, x_{12}, x_{23} - u_{23}, x_{33} - u_{33}, \langle \widetilde{X}, C(\widetilde{U}) \rangle_F - 2 \det(\widetilde{U}) \right\rangle, \quad (4.5)$$

where the matrices  $\widetilde{U}$  and  $\widetilde{X}$  coincide with  $U$  and  $X$  outside  $S$ , respectively, and are zero otherwise. The seven affine relations are independent and thus define an affine plane in  $\mathbb{C}^{3 \times 3}$ .

The first four relations are linear and define the critical space  $H_U^S$ . The last affine relation in (4.5) is equivalent to  $\langle X, C(U) \rangle_F - r \det(U) = 0$  modulo  $\mathcal{I}(\mathcal{L}^S) = \langle x_{11}, x_{12} \rangle$ , since by Lemma 2.1 the matrix  $U$  shares the same critical points of its projection  $\pi_{\mathcal{L}^S}(U)$  onto  $\mathcal{L}^S$ . On one hand, this affine relation coincides with (4.4) for  $I = J = [3]$ . On the other hand, in this case  $S \cap ([3] \times [3]) \neq \emptyset$ .

#### 4.4. Special nonlinear relations among structured critical points

In Section 4.1, we observed that the special linear constraints  $\mathcal{L}^S$  coming from zero patterns  $S \subset [m] \times [n]$  preserve some of the linear relations among unstructured critical points. In this subsection, we deal with special nonlinear relations among unstructured or structured critical points. The following is a consequence of [DHO<sup>+</sup>16, Theorem 5.2] and Theorem 2.5.

**Proposition 4.17.** *Let  $U \in \mathbb{R}^{m \times n}$ . Every critical point  $X \in \mathcal{X}_r$  of  $d_U$  is such that  $U - X \in \mathcal{X}_r^\vee = \mathcal{X}_{m-r}$  and  $U - X$  is a critical point of  $d_U$  as well. In particular, for every subset  $A \subset [m]$  and  $B \subset [n]$  with  $|A| = |B| \geq m - r + 1$ , we have that*

$$M_{A,B}(U - X) = 0. \quad (4.6)$$

The main result of this subsection is the next proposition which states that some of the relations in (4.6) hold even in the structured case.

**Proposition 4.18.** *Let  $S \subset [m] \times [n]$ ,  $U \in \mathbb{R}^{m \times n}$  and consider a critical point  $X = (x_{ij}) \in \mathcal{L}_r^S$  of  $d_U$ . Let  $A \times B \subset [m] \times [n]$  with  $|A| = |B| \geq m - r + 1$  and such that  $S \cap (A \times B) = \emptyset$ . Then  $M_{A,B}(U - X) = 0$ .*

*Proof.* In the following, we denote by  $L_1, \dots, L_{|S|}$  the constraints that set the entries in  $S \subset [m] \times [n]$  of the structured matrices to be equal to zero. We recall from Theorem 4.3 that the critical points  $X = (x_{ij}) \in \mathcal{L}_r^S$  of  $d_U$  are the solutions



of the system

$$\begin{cases} M_{I,J}(X) = 0 & \forall |I| = |J| = r + 1 \\ L_k(X) = 0 & \forall k \in [S] \\ u_{ij} - x_{ij} = \sum_{I,J} \frac{\partial M_{I,J}}{\partial x_{ij}} \lambda_{I,J} + \sum_{k=1}^t \frac{\partial L_k}{\partial x_{ij}} \mu_k & \forall (i, j) \in [m] \times [n]. \end{cases} \quad (4.7)$$

Let  $A \times B \subset [m] \times [n]$  with  $|A| = |B| = s$  and assume  $S \cap (A \times B) = \emptyset$ . Equivalently we have  $\frac{\partial L_k}{\partial x_{ij}} = 0$  for all  $k \in [t]$  and for all  $(i, j) \in A \times B$ . Define the matrix

$$\partial(f) := \left( \frac{\partial f}{\partial x_{ij}} \right) \in \mathbb{C}^{m \times n} \quad \forall f = f(X) \in \mathbb{C}[X].$$

Using the third set of equations in (4.7), we get the identity

$$M_{A,B}(U - X) = M_{A,B} \left( \sum_{I,J} \partial(M_{I,J}(X)) \lambda_{I,J} \right) =: F_{A,B}. \quad (4.8)$$

Under the assumption  $S \cap (A \times B) = \emptyset$ , the polynomial  $F_{A,B}$  does not depend on the linear constraints  $L_k(X) = 0$ , and thus it is independent of  $S$ . Hence the equality (4.8) for  $S = \emptyset$  involves the same  $F_{A,B}$  as for any other  $S$  satisfying  $S \cap (A \times B) = \emptyset$ . By Proposition 4.17,  $M_{A,B}(U - X) = 0$  in the unstructured case, and hence  $M_{A,B}(U - X) = 0$  in the structured case.  $\square$

**Remark 4.19.** *The polynomial  $F_{A,B}$  is homogeneous in the  $\ell_{I,J}$ 's. The coefficients of the monomials of  $F_{A,B}$  in the  $\ell_{I,J}$ 's are polynomials in  $\mathbb{C}[X]$ . In particular, they belong to the ideal  $\mathcal{I}(\mathcal{X}_r)$  for all  $A \times B \subset [m] \times [n]$  with  $s \geq m - r + 1$ .*

**Remark 4.20.** *The condition  $S \cap (A \times B) = \emptyset$  in Proposition 4.18 is sufficient but not necessary to prove that  $M_{A,B}(U - X) = 0$ . For example, let  $m = n = 3$ ,  $r = 2$ ,  $s = 1$  and  $S = \{(1, 1)\}$ . If  $A = B = [3]$ , we obtain that (here  $\lambda_{[3],[3]} = \lambda$ ,  $\mu_1 = \mu$  and  $M_{[3],[3]}(X) = \det(X)$ )*

$$F_{[3],[3]} = \det(X)^2 \lambda^3 + x_{11} \det(X) \lambda^2 \mu,$$

that is,  $F_{[3],[3]} \in \mathcal{I}(\mathcal{X}_2)$  and consequently  $\det(U - X) = 0$ .

The statement of Proposition 4.18 can be generalized to arbitrary linear sections  $\mathcal{L}_r$  of  $\mathcal{X}_r$ . The condition which replaces  $S \cap (A \times B) = \emptyset$  is simply that  $\frac{\partial L}{\partial x_{ij}} = 0$  for all  $L \in \mathcal{I}(\mathcal{L})$ , namely no linear constraint depends by variables  $x_{ij}$  with indices in  $A \times B$ . Again this condition is far from being necessary. To show this, below we restrict to one constraint  $L = \sum_{i,j} v_{ij} x_{ij}$  and to the variety of corank

one square matrices  $\mathcal{X}_{m-1} \subset \mathbb{C}^{m \times m}$ . We prove that the rank of  $U - X$  is completely characterized by the rank of the coefficient matrix  $V = (v_{ij})$ .

**Proposition 4.21.** *Consider a linear form  $L = \sum_{i,j=1}^m v_{ij}x_{ij}$  for some matrix  $V = (v_{ij}) \in \mathbb{C}^{m \times m}$ . Let  $U \in \mathbb{R}^{m \times m}$  and let  $X \in \mathcal{L}_{m-1} = \mathcal{X}_{m-1} \cap \mathcal{V}(L)$  be a critical point of  $d_U$ . Then for all  $2 \leq k \leq m$*

$$\text{rk}(U - X) \leq k - 1 \text{ if and only if } \text{rank}(V) \leq k - 2.$$

*In particular  $\text{rank}(U - X) \geq 1$  if  $V \neq 0$  and  $\text{rank}(U - X) = 1$  if and only if  $V = 0$ , namely in the unstructured case.*

*Proof.* Recall the notation introduced at the beginning of Section 4.

We have  $\mathcal{I}(\mathcal{L}_{m-1}) = \langle \det(X), L \rangle$ . Suppose that  $\text{rk}(U - X) \leq k - 1$ , namely  $M_{A,B}(U - X) = 0$  for all  $A, B \subset [m]$  with  $|A| = |B| = k$ . Using the system (4.7) we get that

$$M_{A,B}(U - X) = M_{A,B}(\lambda \partial(\det(X)) + \mu \partial(L)) = M_{A,B}(\lambda C(X) + \mu V), \quad (4.9)$$

where  $C(X) = (C_{ij}(X))$  is the cofactor matrix of  $X$ . Assume  $A = B = [k]$ . Our goal is to expand the polynomial at the right-hand side of (4.9), which we call  $G_k$  for brevity. First, we consider the following expansion of  $M_{[k],[k]}(P + Q)$  for all  $P, Q \in \mathbb{C}^{m \times m}$ :

$$M_{[k],[k]}(P + Q) = \sum_{\substack{I, J \subset [k] \\ |I|=|J|}} (-1)^{I+J} M_{I,J}(P) M_{[k] \setminus I, [k] \setminus J}(Q),$$

where  $(-1)^{I+J} = (-1)^{\sum_{i \in I} i + \sum_{j \in J} j}$ . This identity follows from the Laplace expansion of the determinant in multiple columns, that can be found for example in [CSS13, Lemma A.1(f)]. We apply it in the case  $P = \lambda C(X)$  and  $Q = \mu V$ :

$$G_k = \sum_{\substack{I, J \subset [k] \\ |I|=|J|}} (-1)^{I+J} M_{I,J}(C(X)) M_{[k] \setminus I, [k] \setminus J}(V) \lambda^{|I|} \mu^{k-|I|}.$$

Then we apply the identity

$$M_{I,J}(C(X)) = \begin{cases} 1 & \text{if } I = J = \emptyset \\ (-1)^{I+J} \det(X)^{|I|-1} M_{[k] \setminus I, [k] \setminus J}(X) & \text{if } |I| = |J| \geq 1 \end{cases}$$

which follows from the relation  $C(X) = \det(X) X^{-T}$  and the *Jacobi complement-*

tary minor Theorem, see [Lal96] and [CSS13, Lemma A.1(e)]. Hence we get

$$G_k = M_{[k],[k]}(V)\mu^k + \sum_{\substack{I,J \subset [k] \\ |I|=|J| \geq 1}} \det(X)^{|I|-1} M_{[k] \setminus I, [k] \setminus J}(X) M_{[k] \setminus I, [k] \setminus J}(V) \lambda^{|I|} \mu^{k-|I|}.$$

Finally, using the condition  $\det(X) = 0$ , the previous identity simplifies to

$$G_k = M_{[k],[k]}(C)\mu^k + \sum_{i,j \in [k]} M_{[k] \setminus \{i\}, [k] \setminus \{j\}}(X) M_{[k] \setminus \{i\}, [k] \setminus \{j\}}(V) \lambda \mu^{k-1}.$$

Similarly, for arbitrary  $A, B \subset [m]$  with  $|A| = |B| = k$ , we get that

$$M_{A,B}(U - X) = M_{A,B}(C)\mu^k + \sum_{i \in A, j \in B} M_{A \setminus \{i\}, B \setminus \{j\}}(X) M_{A \setminus \{i\}, B \setminus \{j\}}(V) \lambda \mu^{k-1}.$$

The consequence is that a critical point  $X$  of  $d_U$  is such that  $\text{rank}(U - X) \leq k - 1$  if and only if  $M_{A \setminus \{i\}, B \setminus \{j\}}(V) = 0$  for all  $i \in A, j \in B$  and  $A, B \subset [m]$  with  $|A| = |B| = k$ , or equivalently  $\text{rank}(V) \leq k - 2$ .  $\square$

The main observation coming from Proposition 4.21 is that a generic linear constraint  $L$  destroys the structure of critical points coming from Theorem 2.5, in particular the relations  $M_{A,B}(U - X)$  for suitable  $A$  and  $B$ . However, if  $L$  is special, these conditions might still hold, even in the case when  $L$  involves entries  $x_{ij}$  with  $(i, j) \in A \times B$ .

## 5. Computations of Euclidean Distance degrees

In this section we present various experiments that study the ED degree of  $\mathcal{L}_r^S$ , when  $r \geq 2$  and the zero pattern  $S$  involves only elements in the diagonal.

First, we restrict to square matrices and consider the zero pattern  $S = \{(1, 1)\}$ . Since the number of (complex) critical points of  $d_U$  on  $\mathcal{L}_{n-1}^S$  is constant for a generic (complex) data matrix  $U$ , it is reasonable to apply a monodromy technique for computing these critical points numerically. For this, we use the HomotopyContinuation.jl [BT18] software package. The number of solutions obtained (that is, the ED degree of  $\mathcal{L}_{n-1}^S$  with respect to the Frobenius inner product) is reported in Table 2. We checked symbolically in Maple™ 2016 [Map] that the number of numerical solutions obtained coincides with the degree of the ideal of rank- $r$  critical points  $\mathcal{I}(Z_{U,n-1})$ , that is the ED degree of  $\mathcal{L}_{n-1}^S$ .

Our experimental results support the following conjecture.

$n$	3	4	5	6	7	8	9	10
$\text{EDdegree}(\mathcal{L}_{n-1}^S)$	8	13	18	23	28	33	38	43

Table 2: ED degrees for  $n \times n$  matrices of rank  $\leq n - 1$  and  $S = \{(1, 1)\}$ .

**Conjecture 5.1.** Consider the variety  $\mathcal{L}_{n-1}^S \subset \mathbb{C}^{n \times n}$ , where  $S = \{(1, 1)\}$ . Then

$$\text{EDdegree}(\mathcal{L}_{n-1}^S) = 5(n - 1) - 2.$$

Next, we fix the diagonal zero pattern  $S = \{(1, 1), \dots, (s, s)\}$  for  $s \in [4]$ , and we consider the variety  $\mathcal{L}_2^S \subset \mathbb{C}^{m \times n}$ . We present in Tables 3,4,5,6 the values of  $\text{EDdegree}(\mathcal{L}_2^S)$  computed depending on the format  $m \times n$ . Our experiments support the following conjectural formulas.

**Conjecture 5.2.** Consider the variety  $\mathcal{L}_2^S \subset \mathbb{C}^{m \times n}$  with  $S = \{(1, 1), \dots, (s, s)\}$  for  $s \in [4]$ . Let  $l = \min(m, n)$ . Then

$$\text{EDdegree}(\mathcal{L}_2^S) = \begin{cases} 3(l-1)^2 - 2(l-1) & \text{if } s = 1 \\ 18(l-2)^2 + 6(l-2) + 1 & \text{if } s = 2 \text{ and } m \neq n \\ 18(l-2)^2 + 10(l-2) + 1 & \text{if } s = 2 \text{ and } m = n \\ 108(l-3)^2 + 144(l-3) + 30 & \text{if } s = 3 \text{ and } m = n \\ 648(l-4)^2 + 1600(l-4) + 488 & \text{if } s = 4 \text{ and } m = n. \end{cases}$$

$s \backslash n$	3	4	5	6	7	8	9	10	11	12
3	8	8	8	8	8	8	8	8	8	8
4	8	21	21	21	21	21	21	21	21	21
5	8	21	40	40	40	40	40	40	40	40
6	8	21	40	65	65	65	65	65	65	65
7	8	21	40	65	96	96	96	96	96	96
8	8	21	40	65	96	133	133	133	133	133
9	8	21	40	65	96	133	176	176	176	176
10	8	21	40	65	96	133	176	225	225	225
11	8	21	40	65	96	133	176	225	280	280
12	8	21	40	65	96	133	176	225	280	341

Table 3: Values of  $\text{EDdegree}(\mathcal{L}_2^S)$  for  $S = \{(1, 1)\}$ .

$s \backslash n$	3	4	5	6	7	8	9	10	11	12
3	25	29	29	29	29	29	29	29	29	29
4	29	85	93	93	93	93	93	93	93	93
5	29	93	181	193	193	193	193	193	193	193
6	29	93	193	313	329	329	329	329	329	329
7	29	93	193	329	481	501	501	501	501	501

Table 4: Values of  $\text{EDdegree}(\mathcal{L}_2^S)$  for  $S = \{(1, 1), (2, 2)\}$ .

$s \backslash n$	3	4	5	6	7
3	30	62	66	66	66
4	62	282	358	366	366
5	66	358	750	870	882
6	66	366	870	1434	1598
7	66	366	882	1598	2334

Table 5: Values of  $\text{EDdegree}(\mathcal{L}_2^S)$  for  $S = \{(1, 1), (2, 2), (3, 3)\}$ .

$s \backslash n$	4	5	6	7	8
4	488	968	1072	1080	1080
5	968	2736	?	?	?

Table 6: Values of  $\text{EDdegree}(\mathcal{L}_2^S)$  for  $S = \{(1, 1), (2, 2), (3, 3), (4, 4)\}$ .

**Remark 5.3.** The values of  $\text{EDdegree}(\mathcal{L}_2^S)$  in Table 3 are known as octagonal numbers: writing  $0, 1, 2, \dots$  in a hexagonal spiral around 0, then these are numbers on the line starting from 0 and going in the direction of 1 [Slo, A000567]. Also the diagonal entries of Table 4 form another interesting integer sequence, see [Slo, A081272].

Our experiments suggest a formula for  $\text{EDdegree}(\mathcal{L}_2^S)$  in the square case.

**Conjecture 5.4.** Consider the variety  $\mathcal{L}_2^S \subset \mathbb{C}^{m \times m}$ , where  $S$  is the zero pattern  $\{(1, 1), \dots, (s, s)\}$  for some  $s \geq 2$ . Then for some constant  $c$

$$\text{EDdegree}(\mathcal{L}_2^S) = 3^s 2^{s-1} (n-s)^2 + s^{s-1} (s+1)^{\lceil s/2 \rceil} (n-s) + c.$$

**Remark 5.5.** In Tables 3,4,5,6 we observe the following symmetry property for the variety  $\mathcal{L}_r^S \subset \mathbb{C}^{m \times n}$ , where  $S = \{(1, 1), \dots, (s, s)\}$  for some  $s \geq 2$ :

$$\text{EDdegree}(\mathcal{L}_r^S)(m, n, r, s) = \text{EDdegree}(\mathcal{L}_r^S)(n, m, r, s).$$

This identity always holds, because the two structured best rank- $r$  approximation problems are the same after relabeling variables.

We conclude by performing the same experiments showed at the beginning of the section, but restricting our study to the subspace  $\text{Sym}_n(\mathbb{C}) \subset \mathbb{C}^{n \times n}$  of  $n \times n$  symmetric matrices. We denote again by  $\mathcal{L}_2^S$  the variety of symmetric matrices of rank at most 2 with (symmetric) zero pattern  $S \subset [n] \times [n]$ . The values of  $\text{EDdegree}(\mathcal{L}_2^S)$  with respect to the diagonal zero pattern  $S = \{(1, 1), \dots, (s, s)\}$  are reported in Table 7.

$s \backslash n$	2	3	4	5	6	7	8	9	10
1	1	4	7	10	13	16	19	22	25
2	1	7	16	25	34	43	52	61	70
3		4	31	58	85	112	139	166	193
4			28	109	190	271	352	433	514

Table 7: Values of  $\text{EDdegree}(\mathcal{L}_2^S) \subset \text{Sym}_n(\mathbb{C})$  for  $S = \{(1, 1), \dots, (s, s)\}$ .

**Conjecture 5.6.** Consider the variety  $\mathcal{L}_2^S \subset \text{Sym}_n(\mathbb{C})$  with  $S = \{(1, 1), \dots, (s, s)\}$  for all  $s \in [4]$ . Then

$$\text{EDdegree}(\mathcal{L}_2^S) = \begin{cases} 3(n-1) - 2 & \text{if } s = 1 \\ 9(n-2) - 2 & \text{if } s = 2 \\ 27(n-3) + 4 & \text{if } s = 3 \\ 81(n-4) + 28 & \text{if } s = 4. \end{cases}$$

## 6. Nonnegative low-rank matrix approximation

In this section, we apply rank-two approximation with zeros to the problem of nonnegative rank-two approximation. Our goal is to find the best nonnegative rank-two approximation with a guarantee that we have found the correct solution. There are two options for the critical points of the Euclidean distance function over  $\mathcal{M}_2$ :

1. A critical point of the Euclidean distance function over  $\mathcal{M}_2$  is a critical point of the Euclidean distance function over the set  $\mathcal{X}_2$  of matrices of rank at most two.
2. A critical point of the Euclidean distance function over  $\mathcal{M}_2$  lies on the boundary of  $\mathcal{M}_2$ , i.e. the critical point contains one or more zero entries.

In Example 6.2, we consider  $3 \times 3$  matrices and show that computing the Euclidean distance to 756 points guarantees finding the best nonnegative rank-two approximation. This example together with the ED degree computations in Section 5 suggests that the exact nonnegative rank-two approximation problem is highly nontrivial, although in general problems on nonnegative decompositions tend to be easier for decompositions of size at most 2 or 3.

Recall, that for a  $m \times n$  matrix  $M$  there is an algorithm for nonnegative factorization with complexity  $O((mn)^{O(r^2 2^r)})$ , where  $r$  is the nonnegative rank [AGKM16], see also [Moi16]. We can also compute a rank- $r$  approximate nonnegative factorization, under the Frobenius norm  $\|M\|_F$ , in  $2^{\text{poly}(r \lg(1/\epsilon))}$  with relative error  $O(\epsilon^{1/2} r^{1/4})$  [AGKM16]. We notice that both algorithms run in polynomial time when the rank is fixed. Nevertheless, their implementation is far from straightforward and the exact constants hidden in the big-O notation could be rather big.

We end this section with simulations demonstrating that most of the time, the optimal solution is given by a critical point with a few zeros. The reader might wonder whether it is interesting to consider the  $3 \times 3$  case, when in practical

applications much larger matrices are considered. We believe that thoroughly understanding small cases is important for understanding the structure of the problem, and might provide insights for developing better numerical algorithms.

In the following example, we demonstrate that in the case of  $3 \times 3$ -matrices, the critical points of the rank-two approximation with zeros can be often described explicitly.

**Example 6.1.** Let  $U = (u_{ij}) \in \mathbb{R}^{3 \times 3}$ . We consider the best rank-two approximation problem with zeros in  $S = \{(1, 1), (1, 2), (2, 2)\}$ . Then there are three critical points, each of which has in addition to the entries in  $S$  one further entry equal to zero and all other entries are equal to the corresponding entries of the matrix  $U$ . Specifically, the critical points are

$$\begin{pmatrix} 0 & 0 & 0 \\ u_{21} & 0 & u_{23} \\ u_{31} & u_{32} & u_{33} \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & u_{13} \\ 0 & 0 & u_{23} \\ u_{31} & u_{32} & u_{33} \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & u_{13} \\ u_{21} & 0 & u_{23} \\ u_{31} & 0 & u_{33} \end{pmatrix}.$$

The additional entries that are set to zero can be read from the determinant of the matrix

$$\begin{pmatrix} 0 & 0 & x_{13} \\ x_{21} & 0 & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix},$$

that is  $x_{13}x_{21}x_{32}$ . Its factors correspond precisely to the additional entries that are set to zero to obtain the three critical points.

In other words, we consider the Euclidean distance minimization problem to the variety defined by  $x_{13}x_{21}x_{32}$ . Since this variety is reducible, we can consider the Euclidean distance minimization problem to each of its three irreducible components. Each of the irreducible components has ED degree 1.

This example generalizes to  $n \times n$  matrices whose non-zero entries form a lower triangular submatrix. The determinant of such a matrix is the product of the diagonal elements, and hence the critical points of the best rank  $(n - 1)$ -approximation problem are obtained by adding a zero to the diagonal.

A more interesting example is given when  $S = \{(1, 1), (1, 2)\}$ . In this case, the determinant is  $x_{13}(-x_{22}x_{31} + x_{21}x_{32})$ . One of the critical points has the entry  $x_{13}$  equal to zero and other entries equal to the corresponding entries of  $U$ . The two other critical points agree with  $U$  in the third column and the  $2 \times 2$  submatrix defined by the rows 2, 3 and columns 1, 2 is equal to one of the critical points of the rank-one approximation for the corresponding  $2 \times 2$  submatrix of  $U$ .

This example generalizes to a zero pattern that contains all but one entry



in a row or in a column. Then the determinant factors as a variable times a  $(n - 1) \times (n - 1)$  subdeterminant. One critical point is obtained by adding the missing zero and the rest of the critical points are obtained by rank  $(n - 2)$ -approximations for the  $(n - 1) \times (n - 1)$  submatrix whose determinant is a factor in the above product.

In the following example, we discuss how to find the best nonnegative rank-two estimate of a  $3 \times 3$ -nonnegative matrix with guarantee.

**Example 6.2.** Consider the group whose elements are simultaneous permutations of rows and columns of a  $3 \times 3$  matrix, and permutations of rows with columns. This group acts on the set zero patterns of a  $3 \times 3$  matrix. There are 26 orbits of this group action, 13 of which are listed in Table 8. The columns of Table 8 list an orbit representative, the orbit size, the ED degree and the description of critical points if available.

The 13 orbit representatives listed in Table 8 have the property that there is no zero pattern  $S$  with less zeros such that the zero pattern of a critical point for  $S$  is contained in the orbit representative. For example, no zero pattern that contains a row or a column is listed in Table 8, because a critical point on the line three of the table has a row of zeros. If a critical point agrees with  $U$  at all its non-zero entries, then adding more zeros causes the Euclidean distance to the data matrix to increase, so such critical points can be discarded.

Moreover, we can also discard the seven orbits of zero patterns marked with star in Table 8, because their critical points either appear earlier in the table or the zero patterns of critical points contain the zero pattern of a critical point that appears earlier in the table.

In summary, there are five different kinds of critical points to be considered:

1. Sum of any 2 components of the SVD of  $U$ . In total:  $1 \cdot 3 = 3$ .
2. Critical points of diagonal zero patterns. In total:  $9 \cdot 8 + 18 \cdot 25 + 6 \cdot 30 = 702$ .
3. Critical points that set one row or column of  $U$  to zero. In total:  $6 \cdot 1 = 6$ .
4. Critical points where a  $2 \times 2$ -submatrix is given by a rank-one critical point of the corresponding submatrix of  $U$ . These critical points also have two zeros and one row or column equal to the corresponding row or column of  $U$ . In total:  $18 \cdot 2 = 36$ .
5. Critical points where zeros form a  $2 \times 2$ -submatrix. In total:  $9 \cdot 1 = 9$ .

In total, the number of critical points is  $3 + 702 + 6 + 36 + 9 = 756$ . Thus, given a nonnegative  $3 \times 3$ -matrix  $U$ , if we construct the 756 critical points described above and choose among the nonnegative critical points the one that is closest

to  $U$ , then it is guaranteed to be the best nonnegative rank-two approximation of the matrix  $U$ .

Example 6.2 suggests that finding the best nonnegative rank-two approximation of a generic matrix with guarantee might be hopeless, because we expect the number of critical points to increase at least exponentially in the matrix size by the conjectures in Section 5. In practice, the best rank-two approximation often has a few zeros.

Using Macaulay2 we sampled uniformly randomly  $10^5$  matrices from the set of  $3 \times 3$  matrices with real nonnegative entries and the sum of entries being equal to 1000 and for each data sample we computed numerically a best nonnegative rank-two approximation. In 88561 cases, the best approximation has no zeros; in 10550 cases, the best approximation has one zero; in 889 cases, the best approximation has two zeros in different rows and columns. Based on this experiment, we observed two interesting phenomena:

- (a) We never encountered a best nonnegative rank-two approximation with three zeros or with two aligned zeros.
- (b) If the best nonnegative rank-two approximation has zero pattern  $S$ , then the best rank-two approximation given by SVD has negative entries in  $S$ .

These facts lead to the following open questions.

- Question 6.3.**
1. Are the experimental observations (a) and (b) true for any nonnegative matrix  $U \in \mathbb{R}^{3 \times 3}$ ?
  2. Given a nonnegative matrix  $U \in \mathbb{R}_{\geq 0}^{m \times n}$  whose best nonnegative rank-2 approximation has zero pattern  $S$ , does the best rank-2 approximation given by SVD have negative entries in  $S$ ?
  3. For which zero patterns  $S \subset [m] \times [n]$  and target ranks  $r \in [m - 1]$  there exists a full rank nonnegative matrix  $U \in \mathbb{R}_{\geq 0}^{m \times n}$  whose best nonnegative rank- $r$  approximation has zero pattern  $S$ ?

## Acknowledgements

We thank Giorgio Ottaviani, Grégoire Sergeant-Perthuis, Pierre-Jean Spaenlehauer, and Bernd Sturmfels for helpful discussions and suggestions. We thank two anonymous reviewers for insightful comments which improved the original manuscript. Kaie Kubjas and Luca Sodomaco are partially supported by the Academy of Finland Grant No. 323416. Elias Tsigaridas is partially supported by ANR JCJC GALOP (ANR-17-CE40-0009), the PGM0 grant ALMA, and the PHC GRAPE.

	$S$	#orbit	EDdegree( $\mathcal{L}_r^S$ )	critical points
1	$\begin{pmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix}$	1	3	sum of any 2 components of SVD
2	$\begin{pmatrix} 0 & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix}$	9	8	no interpretation
3	$\begin{pmatrix} 0 & 0 & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix}$	18	3	$\begin{pmatrix} 0 & 0 & 0 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix}$ or $\text{rank}(X_{\{2,3\},\{1,2\}}) = 1$
4	$\begin{pmatrix} 0 & \cdot & \cdot \\ \cdot & 0 & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix}$	18	25	no interpretation
5*	$\begin{pmatrix} 0 & 0 & \cdot \\ 0 & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix}$	36	3	$\begin{pmatrix} 0 & 0 & \cdot \\ 0 & 0 & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix}$ or $\begin{pmatrix} 0 & 0 & 0 \\ 0 & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix}$ or $\begin{pmatrix} 0 & 0 & \cdot \\ 0 & \cdot & \cdot \\ 0 & \cdot & \cdot \end{pmatrix}$
6*	$\begin{pmatrix} 0 & 0 & \cdot \\ \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot \end{pmatrix}$	36	3	$\begin{pmatrix} 0 & 0 & 0 \\ \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot \end{pmatrix}$ or $\text{rank}(X_{\{2,3\},\{1,2\}}) = 1$
7	$\begin{pmatrix} 0 & \cdot & \cdot \\ \cdot & 0 & \cdot \\ \cdot & \cdot & 0 \end{pmatrix}$	6	30	no interpretation
8	$\begin{pmatrix} 0 & 0 & \cdot \\ 0 & 0 & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix}$	9	1	projection onto $\mathcal{L}^S$
9*	$\begin{pmatrix} 0 & 0 & \cdot \\ 0 & \cdot & 0 \\ \cdot & \cdot & \cdot \end{pmatrix}$	36	3	$\begin{pmatrix} 0 & 0 & 0 \\ 0 & \cdot & 0 \\ \cdot & \cdot & \cdot \end{pmatrix}$ or $\begin{pmatrix} 0 & 0 & \cdot \\ 0 & 0 & 0 \\ \cdot & \cdot & \cdot \end{pmatrix}$ or $\begin{pmatrix} 0 & 0 & \cdot \\ 0 & \cdot & 0 \\ 0 & \cdot & \cdot \end{pmatrix}$
10*	$\begin{pmatrix} 0 & 0 & \cdot \\ 0 & \cdot & \cdot \\ \cdot & \cdot & 0 \end{pmatrix}$	36	3	$\begin{pmatrix} 0 & 0 & 0 \\ 0 & \cdot & \cdot \\ \cdot & \cdot & 0 \end{pmatrix}$ or $\begin{pmatrix} 0 & 0 & \cdot \\ 0 & 0 & \cdot \\ \cdot & \cdot & 0 \end{pmatrix}$ or $\begin{pmatrix} 0 & 0 & \cdot \\ 0 & \cdot & \cdot \\ 0 & \cdot & 0 \end{pmatrix}$
11*	$\begin{pmatrix} 0 & 0 & \cdot \\ \cdot & \cdot & 0 \\ \cdot & \cdot & 0 \end{pmatrix}$	9	3	$\begin{pmatrix} 0 & 0 & 0 \\ \cdot & \cdot & 0 \\ \cdot & \cdot & 0 \end{pmatrix}$ or $\text{rank}(X_{\{2,3\},\{1,2\}}) = 1$
12*	$\begin{pmatrix} 0 & 0 & \cdot \\ 0 & \cdot & 0 \\ \cdot & 0 & \cdot \end{pmatrix}$	36	3	$\begin{pmatrix} 0 & 0 & 0 \\ 0 & \cdot & 0 \\ \cdot & 0 & \cdot \end{pmatrix}$ or $\begin{pmatrix} 0 & 0 & \cdot \\ 0 & 0 & 0 \\ \cdot & 0 & \cdot \end{pmatrix}$ or $\begin{pmatrix} 0 & 0 & \cdot \\ 0 & \cdot & 0 \\ 0 & 0 & \cdot \end{pmatrix}$
13*	$\begin{pmatrix} 0 & 0 & \cdot \\ 0 & \cdot & 0 \\ \cdot & 0 & 0 \end{pmatrix}$	6	3	$\begin{pmatrix} 0 & 0 & 0 \\ 0 & \cdot & 0 \\ \cdot & 0 & 0 \end{pmatrix}$ or $\begin{pmatrix} 0 & 0 & \cdot \\ 0 & 0 & 0 \\ \cdot & 0 & 0 \end{pmatrix}$ or $\begin{pmatrix} 0 & 0 & \cdot \\ 0 & \cdot & 0 \\ 0 & 0 & 0 \end{pmatrix}$

34

Table 8: The 13 orbit representatives of zero patterns of  $3 \times 3$  matrices that have the property that there is no zero pattern  $S$  with less zeros such that the zero pattern of a critical point for  $S$  is contained in the orbit representative.

## References

- [AB09] Sanjeev Arora and Boaz Barak. *Computational complexity: a modern approach*. Cambridge University Press, 2009.
- [AGKM16] Sanjeev Arora, Rong Ge, Ravi Kannan, and Ankur Moitra. Computing a nonnegative matrix factorization—provably. *SIAM J. Comput.*, 45(4):1582–1611, 2016.
- [BT18] Paul Breiding and Sascha Timme. HomotopyContinuation.jl: A Package for Homotopy Continuation in Julia. In *International Congress on Mathematical Software*, pages 458–465. Springer, 2018.
- [CFP03] Moody T. Chu, Robert E. Funderlic, and Robert J. Plemmons. Structured low rank approximation. *Linear Algebra Appl.*, 366:157–172, 2003.
- [CH87] Dipa Choudhury and Roger A. Horn. An analog of the singular value decomposition for complex orthogonal equivalence. *Linear Multilinear Algebra*, 21(2):149–162, 1987.
- [Cif21] Diego Cifuentes. A convex relaxation to compute the nearest structured rank deficient matrix. *SIAM J. Matrix Anal. Appl.*, 42(2):708–729, 2021.
- [CLO92] David Cox, John Little, and Donal O’Shea. *Ideals, varieties, and algorithms*. Undergraduate Texts in Mathematics. Springer-Verlag, New York, 1992. An introduction to computational algebraic geometry and commutative algebra.
- [CSS13] Sergio Caracciolo, Alan D. Sokal, and Andrea Sportiello. Algebraic/combinatorial proofs of Cayley-type identities for derivatives of determinants and Pfaffians. *Adv. in Appl. Math.*, 50(4):474–594, 2013.
- [CW19] Aldo Conca and Volkmar Welker. Lovász–Saks–Schrijver ideals and coordinate sections of determinantal varieties. *Algebra & Number Theory*, 13(2):455–484, 2019.
- [DHO<sup>+</sup>16] Jan Draisma, Emil Horobet, Giorgio Ottaviani, Bernd Sturmfels, and Rekha R. Thomas. The Euclidean distance degree of an algebraic variety. *Found. Comput. Math.*, 16(1):99–149, 2016.
- [DLOT17] Dmitriy Drusvyatskiy, Hon-Leung Lee, Giorgio Ottaviani, and Rekha R. Thomas. The Euclidean distance degree of orthogonally invariant matrix varieties. *Israel J. Math.*, 221(1):291–316, 2017.
- [DOT18] Jan Draisma, Giorgio Ottaviani, and Alicia Tocino. Best rank- $k$  approximations for tensors: generalizing Eckart-Young. *Res. Math. Sci.*, 5(2):27, 2018.
- [GG11] Nicolas Gillis and François Glineur. Low-rank matrix approximation with weights or missing data is np-hard. *SIAM Journal on Matrix Analysis and Applications*, 32(4):1149–1165, 2011.
- [GHS87] G. H. Golub, Alan Hoffman, and G. W. Stewart. A generalization of the Eckart-Young-Mirsky matrix approximation theorem. *Linear Algebra Appl.*, 88/89:317–327, 1987.
- [Gil20] Nicolas Gillis. *Nonnegative Matrix Factorization*. SIAM, 2020.
- [GS] Daniel R. Grayson and Michael E. Stillman. Macaulay2, a software system for research in algebraic geometry. Available at [www.math.uiuc.edu/Macaulay2/](http://www.math.uiuc.edu/Macaulay2/).
- [HR20] Emil Horobet and Jose Israel Rodriguez. Data loci in algebraic optimization. [arXiv:2007.04688](https://arxiv.org/abs/2007.04688), 2020.
- [Lal96] Pierre Lalonde. A non-commutative version of Jacobi’s equality on the cofactors of a matrix. *Discrete Math.*, 158(1-3):161–172, 1996.
- [Map] a division of Waterloo Maple Inc. Maplesoft. Maple.

- [Mar08] Ivan Markovsky. Structured low-rank approximation and its applications. *Automatica*, 44(4):891–909, 2008.
- [Mar19] Ivan Markovsky. *Low-rank approximation*. Communications and Control Engineering Series. Springer, Cham, 2019.
- [Moi16] Ankur Moitra. An almost optimal algorithm for computing nonnegative rank. *SIAM J. Comput.*, 45(1):156–173, 2016.
- [OP15] Giorgio Ottaviani and Raffaella Paoletti. A geometric perspective on the singular value decomposition. *Rend. Istit. Mat. Univ. Trieste*, 47:107–125, 2015.
- [OSS14] Giorgio Ottaviani, Pierre-Jean Spaenlehauer, and Bernd Sturmfels. Exact solutions in structured low-rank approximation. *SIAM J. Matrix Anal. Appl.*, 35(4):1521–1542, 2014.
- [PB83] J. Scott Provan and Michael O. Ball. The complexity of counting cuts and of computing the probability that a graph is connected. *SIAM J. Comput.*, 12(4):777–788, 1983.
- [Pie15] Ragni Piene. Polar varieties revisited. In *Computer algebra and polynomials*, volume 8942 of *Lecture Notes in Comput. Sci.*, pages 139–150. Springer, Cham, 2015.
- [PS17] Ankit Parekh and Ivan W Selesnick. Improved sparse low-rank matrix estimation. *Signal Processing*, 139:62–69, 2017.
- [Slo] Neil J. A. Sloane. The on-line encyclopedia of integer sequences. Available at [oeis.org](http://oeis.org).
- [Sod20] Luca Sodomaco. *The Distance Function from the Variety of partially symmetric rank-one Tensors*. PhD thesis, Università degli Studi di Firenze, 2020. Available at <https://sites.google.com/view/luca-sodomaco/home/publications>.
- [SRV12] Pierre-André Savalle, Emile Richard, and Nicolas Vayatis. Estimation of simultaneously sparse and low rank matrices. In *ICML*, 2012.
- [SS16] Éric Schost and Pierre-Jean Spaenlehauer. A quadratically convergent algorithm for structured low-rank approximation. *Found. Comput. Math.*, 16(2):457–492, 2016.