




# The Cone of $5 \times 5$ Completely Positive Matrices

Max Pfeffer<sup>1</sup>  · José Alejandro Samper<sup>2</sup>

Received: 30 September 2021 / Revised: 13 July 2023 / Accepted: 10 November 2023 /  
Published online: 24 January 2024  
© The Author(s) 2024

## Abstract

We study the cone of completely positive (cp) matrices for the first interesting case  $n = 5$ . This is a semialgebraic set for which the polynomial equalities and inequalities that define its boundary can be derived. We characterize the different loci of this boundary and we examine the two open sets with cp-rank 5 or 6. A numerical algorithm is presented that is fast and able to compute the cp-factorization even for matrices in the boundary. With our results, many new example cases can be produced and several insightful numerical experiments are performed that illustrate the difficulty of the cp-factorization problem.

**Keywords** Geometry of convex cones · Convex algebraic geometry · Completely positive matrices · Copositive optimization · Nonnegative matrix factorization

**Mathematics Subject Classification** 14P10 · 15B48 · 65F30

## 1 Introduction

Conic optimization is the problem of minimizing the value of a linear function over the intersection of a cone and a linear space. Many problems in optimization and geometry can be framed in this form and a wide variety of minimization methods have been developed for different types of cones. Cones that are also semialgebraic sets are of particular interest, because their boundary can be described by polynomial inequalities. Special relevant cases include polyhedral cones, the cone of positive semidefinite

---

Editor in Charge: Kenneth Clarkson

---

Max Pfeffer  
m.pfeffer@math.uni-goettingen.de

José Alejandro Samper  
jsamper@mat.uc.cl

<sup>1</sup> Institute of Numerical and Applied Mathematics, University of Göttingen, Göttingen, Germany

<sup>2</sup> Departamento de Matemáticas, Pontificia Universidad Católica de Chile, Santiago, Chile

matrices, the cone of homogeneous nonnegative polynomials in any number of variables, or the cone of homogeneous polynomials that are sums of squares.

There are many natural questions that may be asked about such semialgebraic cones. In order to assess the effectiveness of potential optimization algorithms, we need measures of the "complicatedness" of the cone. For instance, it is often hard to determine if a given vector is in a specific cone (the membership problem), and one may wonder if there are deep reasons for this. One potential approach consists of understanding the system of polynomial inequalities defining the cone. Then the number of monomials and their degree can serve as a measure for this hardness. Particularly interesting is the cone of (homogeneous) sums of squares, which is contained in the cone of nonnegative polynomials, but the algebraic description of their differences is difficult even in the smallest interesting cases, see [5, 6].

In this article, we aim to carry out such an analysis for a more complicated cone that is also of interest in optimization: the collection of completely positive real symmetric  $n \times n$  matrices  $\mathcal{CP}_n$ , i.e., real nonnegative  $n \times n$  matrices  $A$  that can be written as  $A = BB^T$ , where  $B$  is also a nonnegative matrix. This convex cone is dual to the copositive cone  $\mathcal{COP}_n$  of symmetric matrices whose associated quadratic form is nonnegative in the nonnegative orthant. These cones appear quite naturally in various contexts and conic programs on them model a number of concrete applied problems. See for instance [2] and [3] for thorough introductions to the topic. The cones  $\mathcal{CP}_n$  and  $\mathcal{COP}_n$  are notably complicated to work with. For instance, determining whether a given matrix is in  $\mathcal{CP}_n$  is a co-NP-complete problem [20]. There are several algorithms [14, 15] that attempt to factorize these matrices in order to test if they are in  $\mathcal{CP}_n$ , but these algorithms (both exact and approximate) have a well established number of drawbacks: they are often slow and/or fail to detect matrices near the boundary of the cone.

We will look at the cone using tools from algebra and geometry: both  $\mathcal{CP}_n$  and  $\mathcal{COP}_n$  are known to be semialgebraic. Hence we can use algebraic invariants of the varieties defining their boundaries as avatars to measure their complexity. We aim to study low dimensions in detail. For  $n = 1, 2, 3, 4$ , the cone  $\mathcal{CP}_n$  coincides with the cone  $\mathcal{DN}\mathcal{N}_n$  of matrices that are positive semidefinite and have nonnegative entries. Starting at  $n = 5$ , the straightforward containment  $\mathcal{CP}_n \subseteq \mathcal{DN}\mathcal{N}_n$  starts being strict. The boundary of  $\mathcal{DN}\mathcal{N}_n$  is generally easy to describe (matrices of low rank and/or with some zero entries) and it is hence not hard to describe which part of that boundary is also in  $\mathcal{CP}_n$ . The interesting part is then to describe the boundary of  $\mathcal{CP}_n$  contained in the interior of  $\mathcal{DN}\mathcal{N}_n$ . Additionally, there is a partition of  $\mathcal{CP}_n$  according to what is called the completely positive rank (cp-rank), dealing with the smallest size of a nonnegative matrix factorization. The boundaries of the regions are also semialgebraic and in general quite difficult: it is not known how many parts are in this splitting.

We focus therefore on the smallest interesting case, namely  $n = 5$ . We study the boundary  $\partial\mathcal{CP}_5$  from an algebraic point of view, finding explicit equations for part of the boundary and implicit ones for the rest. We use the Hildebrand characterization of the extreme rays of  $\mathcal{COP}_5$  to show that the Zariski-closure of the part of the boundary of  $\mathcal{CP}_5$  in the interior of  $\mathcal{DN}\mathcal{N}_5$  is a degree 3900 hypersurface with 24 irreducible components, 12 of which have degree 320 and 12 degree 5. The explicit polynomials for the degree 320 parts are probably impossible to compute, but the parametrization

comes from a simple toric variety, which yields the correct scenario for the novel implicitization techniques from numerical algebraic geometry [12]. We remark that the rest of the boundary components are in the boundary of  $\mathcal{DN}\mathcal{N}_5$  that are either low rank (and hence have vanishing determinant) or have a zero entry. All in all, the degree of the Zariski closure of these parts is 5 for the determinantal part and 1 for each potential entry equal to zero, which is quite low compared to the degree of the other part.

The algebraic description of the boundary yields a few interesting results for computational experiments: we obtain that the set of matrices with rational entries is dense in the boundary  $\partial\mathcal{CP}_5$  with a method of finding an approximation that is (typically) exactly in the boundary. That is, if we try to find a factorization of a matrix next to the boundary, we can easily test for which part of the boundary it comes from using our description of the boundary. The equations we find also yield a recipe to construct many exact examples in all the components of the boundary as well as computing the tangent space at any given point. This data allows us to find matrices in  $\mathcal{DN}\mathcal{N}_5 \setminus \mathcal{CP}_5$  together with their closest point in  $\partial\mathcal{CP}_5$ . We include a discussion of the cp-rank of the matrices in the interior of  $\mathcal{CP}_5$ . The possible cp-ranks are known to be 5 or 6 and the boundary has a Zariski closure that is the vanishing locus of a polynomial. Although not much is known about this boundary, we highlight some of its properties in order to pursue some computational experiments.

Next we present a novel numerical method for the approximation of the cp-factorization. This method is very fast and it is able to approximate factorizations even of matrices in the boundary of  $\mathcal{CP}_5$ . We carry out a number of experiments to estimate its performance in the small-dimensional setting  $n = 5$ . For instance, in the part of the boundary that does not coincide with the boundary of  $\mathcal{DN}\mathcal{N}_5$ , the factorizations of the matrices are forced to have some zeros. The algorithm easily detects these zero entries in the factorization and generally finds the correct factorization. Using this, and with additional knowledge about the boundary separating matrices of cp-rank 5 and 6, the results of the experiments allow us to formulate a couple of relevant questions and conjectures.

In short, the strength of this paper is that it combines the theoretical progress (the derivation of the algebraic equations for the boundary) with an experimental investigation of some interesting cases. Since the completely positive cone is of great importance in optimization, it is helpful to obtain this practical insight.

## 1.1 Notation

We will consider several convex cones contained in the space of real symmetric  $n \times n$  matrices. The following cones will be relevant:

- $Sym_n$  denotes the space of  $n \times n$  real symmetric matrices.
- $S_n$  is the cone of all positive semidefinite matrices, i.e., matrices  $M$  that can be written as  $M = XX^T$  for some  $n \times k$  matrix  $X$ .
- $\mathcal{N}_n$  is the cone of all symmetric matrices with nonnegative entries.
- $\mathcal{DN}\mathcal{N}_n = S_n \cap \mathcal{N}_n$  denotes the cone of all positive semidefinite matrices with nonnegative entries. This is sometimes called the *doubly nonnegative cone*.

- $\mathcal{COP}_n$  is the *copositive cone* of all matrices  $M$  such that  $v^T M v \geq 0$  for all  $v \in \mathbb{R}_{\geq 0}^n$ .
- $\mathcal{CP}_n$  is the cone of all matrices  $A$  such that there is a  $n \times k$  matrix  $B$  with nonnegative entries and  $A = BB^T$ . This cone is called the *completely positive cone*.

These cones are semialgebraic sets, meaning that they can be described by polynomial inequalities. We are interested in understanding the difference between the cones  $\mathcal{CP}_n \subseteq \mathcal{DN}\mathcal{N}_n$  as good as we can. Maxfield and Minc [19] showed that the two cones are equal if and only if  $n \leq 4$ . Thus we will focus in understanding the case  $n = 5$ , i.e., the smallest value of  $n$  in which the two cones are different. A particularly interesting question is to understand the subset  $\partial\mathcal{CP}_5 \cap \mathcal{DN}\mathcal{N}_5^\circ$ , i.e., the elements in the boundary of  $\mathcal{CP}_5$  that lie in the interior of  $\mathcal{DN}\mathcal{N}_5$ .

Endow the space of symmetric matrices with the inner product  $\langle A, B \rangle = \text{trace}(A^T B)$  and the corresponding Frobenius norm  $\|A\| = \sqrt{\langle A, A \rangle}$ . Then  $\mathcal{CP}$  and  $\mathcal{COP}$  are dual cones in this setting. This will be exploited when we study the boundary of the cones.

## 2 The Boundary of $\mathcal{CP}_5$

We will use the extreme rays in  $\mathcal{COP}_5$  to parametrize the factorizations of elements in  $\partial\mathcal{CP}_5 \cap \mathcal{DN}\mathcal{N}_5^\circ$ . After that we manipulate these parametrizations to obtain information about the algebraic boundary of  $\mathcal{CP}_5$ , i.e., about the Zariski closure of  $\partial\mathcal{CP}_5 \cap \mathcal{DN}\mathcal{N}_5^\circ$  and its irreducible components. In other words, we reduce the problem to a computation of images of certain varieties under algebraic maps. This allows us to compute polynomials defining some of the irreducible components of the algebraic boundary and to compute the degree of the other components. We further discuss the uniqueness of completely positive factorizations in the boundary, rational factorizations of matrices  $\mathcal{CP}_5$  and the cp-rank partition of  $\mathcal{CP}_5$ .

### 2.1 The Boundary of $\mathcal{COP}_5$

Hildebrand classified all extreme rays of  $\mathcal{COP}_5$ . The theorem goes as follows:

**Theorem 2.1** ([17]) *Every extreme ray of  $\mathcal{COP}_5$  is generated by a symmetric matrix  $M$  of one the following four types:*

- (1)  $M = vv^T$ , where  $v \in \mathbb{R}^5$  has positive and negative entries.
- (2)  $M = e_i e_j^T + e_j e_i^T$ , where  $\{e_1, \dots, e_5\}$  is the standard basis of  $\mathbb{R}^5$ .
- (3)  $M = PDHDP^T$ , where  $H$  is the Horn matrix below,  $D$  is a positive diagonal matrix and  $P$  is a permutation matrix.

$$H = \begin{pmatrix} 1 & -1 & 1 & 1 & -1 \\ -1 & 1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 & 1 \end{pmatrix}$$

(4)  $M = PDT(\Theta)DP^T$ , where  $T(\Theta)$  is a matrix defined in terms of five parameters below,  $D$  is a positive diagonal matrix and  $P$  is a permutation matrix. Here

$$T(\Theta) = \begin{pmatrix} 1 & -\cos(\theta_1) & \cos(\theta_1 + \theta_2) & \cos(\theta_4 + \theta_5) & -\cos(\theta_5) \\ -\cos(\theta_1) & 1 & -\cos(\theta_2) & \cos(\theta_2 + \theta_3) & \cos(\theta_5 + \theta_1) \\ \cos(\theta_1 + \theta_2) & -\cos(\theta_2) & 1 & -\cos(\theta_3) & \cos(\theta_3 + \theta_4) \\ \cos(\theta_4 + \theta_5) & \cos(\theta_2 + \theta_3) & -\cos(\theta_3) & 1 & -\cos(\theta_4) \\ -\cos(\theta_5) & \cos(\theta_5 + \theta_1) & \cos(\theta_3 + \theta_4) & -\cos(\theta_4) & 1 \end{pmatrix}, \tag{1}$$

with  $\Theta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)$  a tuple of positive real numbers satisfying  $\sum_{i=1}^5 \theta_i < \pi$ .

This parametrization of the extreme rays in  $\mathcal{COP}_5$  suggests an approach to understand the algebraic boundary of  $\mathcal{CP}_5$ . In fact,  $\partial\mathcal{CP}_5$  is the set of matrices  $A$  such that  $\langle A, X \rangle \geq 0$  for all the extreme rays described, with the additional constraint that equality holds for at least one ray. The part of the boundary shared by  $\mathcal{CP}_5$  and  $\mathcal{DN}\mathcal{N}_5$  is well understood: It consists of low rank matrices and matrices with some zero entries and corresponds to the types (1) and (2) in the theorem above. We will focus mainly on the other part of the boundary, namely, the matrices in the interior of  $\mathcal{DN}\mathcal{N}_5$  and the boundary of  $\mathcal{CP}_5$ . Therefore, any such  $A$  must be invertible, since it would otherwise be on the boundary of  $\mathcal{S}_5$  and thus also on the boundary of  $\mathcal{DN}\mathcal{N}_5$ . Furthermore, it is necessary that the entries of  $A$  are all strictly positive to avoid the boundary of  $\mathcal{N}_5$ . All the matrices in this part of the boundary are either orthogonal to matrices of the type (3), which we call the *Horn orbit*, or (4), which we call the *Hildebrand orbit*. We will first work out the orthogonality to the parts that ignore the permutation matrices. With that in mind, the sets of completely positive matrices in  $\partial\mathcal{CP}_5 \cap \mathcal{DN}\mathcal{N}_5^\circ$  orthogonal to matrices in the Horn or Hildebrand orbit will be called the Horn and Hildebrand *locus* respectively.

We begin by mentioning a relevant theorem.

**Theorem 2.2** [[22] Section 4] *Assume  $A \in \partial\mathcal{CP}_5$  is orthogonal to a matrix in the Horn orbit or a Hildebrand orbit. Then  $A = BB^T$  for a nonnegative square matrix  $B \in \mathbb{R}_{\geq}^{5 \times 5}$ .*

We rely on the above lemma and the following simple remark, exploited heavily in [19] to bound the completely positive rank in  $\mathcal{CP}_5$ . Assume that  $A \in \partial\mathcal{CP}_5$  is orthogonal to a matrix  $M$  as in parts (3) or (4) in Theorem 2.1. Let  $B$  be a nonnegative factorization of  $A$ , i.e.,  $A = BB^T$ . If  $v_1, \dots, v_k$  are the columns of  $B$ , then  $v_i^T M v_i = 0$ , i.e., each column of  $B$  is a zero of the quadratic form associated to  $M$ . Since  $M$  is copositive, this is equivalent to saying that every column is a global minimum of the quadratic form in the positive orthant  $\mathbb{R}_{\geq 0}^5$ .

### 2.2 The Dual of the Orbit of the Horn Matrix

The following theorem is hidden in a proof in [22]:

**Theorem 2.3** [[22] Theorem 4.1] *Let  $H$  be the Horn matrix. A vector  $v \in \mathbb{R}_{\geq 0}^5$  is a solution to the equation  $v^T H v = 0$  if and only if it is in the union of cones*

$$K = \bigcup_{i=1}^5 \text{cone}(e_i + e_{i+1}, e_{i+1} + e_{i+2}),$$

where the indices are taken modulo 5. Consequently, every matrix in the Horn locus of  $\partial \mathcal{CP}_5 \cap \mathcal{DN}\mathcal{N}_5^\circ$  is an invertible matrix  $A$  such that the columns of any nonnegative matrix  $B$  for which  $A = B B^T$  are (linearly independent) elements of  $K$ .

Notice that this restricts significantly the possible factor matrices  $B$ : the columns are a choice of five vectors in the union of such cones. Notice furthermore that if two of the columns are in the same cone, then the product  $B B^T$  has at least one entry equal to zero and is consequently in  $\partial \mathcal{DN}\mathcal{N}_5$ . It follows that in the Horn part of the boundary, the nonnegative factorizations of the matrices must contain exactly one vector in each cone. Notice that if  $A = B B^T$  and  $\tilde{B}$  is obtained by permuting the columns of  $B$ , then  $A = \tilde{B} \tilde{B}^T$ . Hence, the structure of any factorization is as follows:

**Lemma 2.4** *Any matrix  $A$  orthogonal to  $H$  and in  $\partial \mathcal{CP}_5 \cap \mathcal{DN}\mathcal{N}_5^\circ$  has a cp-factorization  $A = B B^T$ , where  $B$  is of the form:*

$$B = \begin{pmatrix} z_1 & 0 & 0 & y_4 & y_5 + z_5 \\ y_1 + z_1 & z_2 & 0 & 0 & y_5 \\ y_1 & y_2 + z_2 & z_3 & 0 & 0 \\ 0 & y_2 & y_3 + z_3 & z_4 & 0 \\ 0 & 0 & y_3 & y_4 + z_4 & z_5 \end{pmatrix}. \tag{2}$$

Here  $y_1, y_2, y_3, y_4, y_5, z_1, z_2, z_3, z_4, z_5$  are positive real numbers.

The set of all matrices in Lemma 2.4 is a 10-dimensional relatively open cone in the space of  $5 \times 5$  matrices. The left action of the diagonal matrices on the factorization increases the dimension to create a hypersurface in the 15-dimensional space where the zero patterns are fixed, but the nonzero entries are free. The variety  $V_{Horn}$  consisting of the Zariski closure of the positive matrices orthogonal to  $D H D$  for some diagonal matrix  $D$  therefore is a variety in a special 15 dimensional affine subspace of the space of  $5 \times 5$  matrices. This produces one of the components of the Zariski closure of  $\partial \mathcal{CP}_5 \cap \mathcal{DN}\mathcal{N}_5^\circ$ . The parameters of the factorization matrix and the diagonal matrix  $D$  can be chosen to be positive, which gives us a nonempty part of the boundary: every choice yields a positive factorization (hence an element of  $\mathcal{CP}_5$ ) and an invertible matrix orthogonal to an extreme ray conjugate to  $H$  (hence in the boundary of  $\mathcal{CP}_5$ ). In order to study this further we define some relevant varieties.

**Definition 2.5** Let  $W \subseteq \mathbb{R}^{5 \times 5}$  be the linear subspace of matrices of the form

$$\begin{pmatrix} y_{11} & 0 & 0 & y_{41} & y_{51} \\ y_{12} & y_{22} & 0 & 0 & y_{52} \\ y_{13} & y_{23} & y_{33} & 0 & 0 \\ 0 & y_{24} & y_{34} & y_{44} & 0 \\ 0 & 0 & y_{35} & y_{45} & y_{55} \end{pmatrix}.$$

Let  $Z_{Horn} \subset W$  be the Zariski closure of all matrices of the form  $DB$ , where  $B$  is a matrix as in (2) and  $D$  is a diagonal matrix with positive entries.

**Theorem 2.6** *The variety  $V_{Horn}$ , i.e. the subvariety of the Zariski closure of  $\partial\mathcal{CP}_5 \cap \mathcal{DN}_5^o$  generated by matrices orthogonal to the matrices of the form  $DHD$ , is a hypersurface contained in the Zariski closure of the image of  $Z_{Horn}$  under the map  $\varphi : W \rightarrow \mathcal{S}_n$  given by  $\varphi(X) = XX^T$ .*

**Proof**  $Z_{Horn}$  contains all matrices factorizing the matrices orthogonal to the matrices of the form  $DHD$ , which in turn generate  $V_{Horn}$ . The map  $\varphi$  encodes these factorizations so  $V_{Horn}$  is contained in the image of  $Z_{Horn}$ .

To see that  $V_{Horn}$  is a hypersurface notice that the parameters of the matrix in Lemma 2.4 can be chosen to be arbitrary positive numbers and the diagonal matrix  $D$  can also be chosen to be arbitrary positive. The diagonal matrices form a 5-dimensional algebraic torus acting by conjugation on the Horn rays. The stabilizer of this action consists of scalar multiples of the identity, so the rays (orbits) are parametrized by a 4-dimensional torus. By taking any chart of the 4-dimensional Torus (for example, choose the last entry to be equal to 1) we get a smooth map from a 14-dimensional neighborhood in  $\mathbb{R}_{\geq 0}^{14}$  to the boundary of  $\mathcal{CP}_5$ . A straightforward computation shows that the rank of this map at a generic point is 14, hence the map is a local homeomorphism to its image around that point. This implies that the (Zariski closure of the) image is 14-dimensional thus  $V_{Horn}$  is a hypersurface. Since  $Z_{horn}$  is not 15-dimensional and contains a copy of the domain of this map it is also a hypersurface. □

**Theorem 2.7** *The variety  $V_{Horn}$  is the vanishing locus of the polynomial*

$$s(X) = \det(H \circ X). \tag{3}$$

Here  $\circ$  denotes the Hadamard product of matrices.

**Proof** One may use the parametrization from Theorem 2.6 to verify that  $V_{Horn}$  is contained in the vanishing locus of  $s(X)$ . This computation is done by a computer algebra system. Furthermore, since  $s(X)$  is irreducible (because determinants are) its vanishing locus is a hypersurface containing  $Z_{Horn}$  and hence equal to its Zariski closure. □

**Remark 2.8** There is a good conceptual reason that leads one to guess the equation of Theorem 2.7. Let  $K$  be the convex cone generated by the matrices of the form  $DHD$

where  $D$  is a diagonal matrix. And let  $K_+$  be the subcone generated by restricting to positive diagonal matrices. Notice that  $K_+$  is a subcone of  $K$  and that the extreme rays of  $K$  are all extreme rays of  $\mathcal{COP}_5$  and of  $K$ . Thus  $K_+ \subseteq \mathcal{COP}_5$ . It follows that the dual  $K_+^*$  contains  $\mathcal{CP}_5$  and  $K^*$ . Furthermore, the three cones  $K, K_+, \mathcal{DN}\mathcal{N}_5$  have many common rays, so one expects them to share a part of their boundary with nonempty Euclidean interior (in the induced topology of the boundary). Among the three cones we can see that  $K$  has a particularly simple boundary. Consider the invertible linear transformation  $f$  of  $Sym_5$  on itself given by  $f(X) = H \circ X$ . Notice that  $f(DHD) = D1D$ , thus  $f$  yields a linear cone  $K$  isomorphism to the convex cone generated by elements of the form  $D1D$ , i.e., the cone whose extreme rays are rank one matrices. This cone is  $\mathcal{S}_5$  which is self dual and its algebraic boundary is known to be given by matrices of low rank, i.e., solutions to the equation  $\det(X) = \det(\mathbf{1} \circ X) = 0$ . The linear change of variables yields that  $K^*$  has a boundary described by (3). Thus the shared part of the boundary of  $K^*, K_+^*$  and  $\mathcal{CP}_5$  satisfies the desired equation of the theorem.

### 2.3 The Dual of the Orbit of Hildebrand Matrices

The following is implicit in the work of Hildebrand [17, Section 3.2.2]:

**Theorem 2.9** *Let  $T(\Theta)$  be the five parameter matrix defined in (1) with  $\Theta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)$  a tuple of positive real numbers satisfying  $\sum_{i=1}^5 \theta_i < \pi$ . A vector  $v$  is a solution to  $v^T T(\Theta)v = 0$  if and only if it is a positive multiple of a column of the matrix*

$$S(\Theta) := \begin{pmatrix} \sin(\theta_5) & 0 & 0 & \sin(\theta_2) & \sin(\theta_3 + \theta_4) \\ \sin(\theta_4 + \theta_5) & \sin(\theta_1) & 0 & 0 & \sin(\theta_3) \\ \sin(\theta_4) & \sin(\theta_5 + \theta_1) & \sin(\theta_2) & 0 & 0 \\ 0 & \sin(\theta_5) & \sin(\theta_1 + \theta_2) & \sin(\theta_3) & 0 \\ 0 & 0 & \sin(\theta_1) & \sin(\theta_3 + \theta_2) & \sin(\theta_4) \end{pmatrix}.$$

As a consequence, we can parametrize the hypersurface dual to all the matrices of the form  $DT(\Theta)D$  by the factor matrices of the form  $D_1 S(\Theta) D_2$  where  $D_1$  and  $D_2$  are diagonal matrices. In fact, the actions of  $D_1$  and  $D_2$  scale the diagonal products proportionally. Thus, with  $W$  as in the previous section, we have the following parametrization:

**Theorem 2.10** *Let  $V_{Hi}$  be the subvariety of the Zariski closure of  $\partial\mathcal{CP}_5 \cap \mathcal{DN}\mathcal{N}_5^o$  generated by the matrices orthogonal to some element in the (torus) orbit of Hildebrand matrices. Then  $V_{Hi}$  is the Zariski closure of the image of the variety  $Z_{Hi}$  associated to the ideal  $\langle y_{11}y_{22}y_{33}y_{44}y_{55} - y_{13}y_{24}y_{35}y_{41}y_{52} \rangle$  in the coordinate ring of  $W$ , under the map  $\varphi$  from Theorem 2.6.*

**Proof** Notice that the variety  $\tilde{Z}_{Hi} \subseteq W$  generated by matrices of the form  $D_1 S(\Theta) D_2$  is a hypersurface: there are fifteen parameters in the factorization: Fixing the last entry of  $D_2$  to be equal to one, we get a fourteen dimensional parameter space that maps to  $\tilde{Z}_{Hi}$ . One can check with a computer algebra system that the derivative of this map



has full rank at one point and hence generically. It follows by the implicit function theorem that the map is a local homeomorphism to its image. Hence the dimension of  $\tilde{Z}_{Hi}$  is at least 14. On the other hand, every element of  $D_1S(\Theta)D_2$  vanishes at the polynomial  $y_{11}y_{22}y_{33}y_{44}y_{55} - y_{13}y_{24}y_{35}y_{41}y_{52}$ . The variety  $Z_{Hi}$  is an irreducible hypersurface (it is affine toric) and contains  $\tilde{Z}_{Hi}$ . The dimension of  $Z_{Hi}$  is fourteen and by irreducibility it contains no proper subvariety of dimension fourteen. It must therefore coincide with  $\tilde{Z}_{Hi}$ .  $\square$

The theorem above yields a parametrization of an algebraic component of the desired boundary. The variety  $V_{Hi}$  is therefore the implicitization of  $Z_{Hi}$  under the map  $\varphi$ . However, the standard techniques using Gröbner bases yield no answer. This is for a good reason, since the numerical implicitization algorithm [12] implemented in the HomotopyContinuation package in Julia [11], yields the following:

**Corollary 2.11** *The degree of  $V_{Hi}$  is 320.*

This computation can be certified using interval arithmetics [10]. The degree being equal to 320 means that the expected number of monomials in the polynomial defining  $V_{Hi}$  is close to  $\binom{334}{14} \approx 10^{24}$ . In particular, the polynomial defining  $V_{Hi}$  is likely impossible to write down. Of course, we could be optimistic and hope that the defining polynomial be very sparse and tractable, but chances of this seem very small.

## 2.4 The Number of the Factorizations

Our experiments below need control on the number of factorizations. Theorem 4.4 in [22] says that up to reordering of the columns, there are at most two factorizations for matrices in the Horn locus and one unique factorization for matrices in the Hildebrand locus. Up to numerical errors, the experiments below capture this phenomenon.

## 2.5 Rational Points on $\partial\mathcal{CP}_5$

We now observe that the matrices in  $\partial\mathcal{CP}_5$  with rational entries are dense in that part of the boundary. This is a consequence of the theorems in the previous two sections and in particular this allows for the exact computation of rational matrices in  $\partial\mathcal{CP}_5 \cap \mathcal{DNN}_5^\circ$ . With this in mind, let  $Sym_5(\mathbb{Q})$  be the set of symmetric matrices with rational entries. We summarize the result as follows:

**Theorem 2.12** *The set of rational matrices in  $\partial\mathcal{CP}_5$  is dense (with the Euclidean topology), that is,  $\partial\mathcal{CP}_5 \subseteq \partial\mathcal{CP}_5 \cap Sym_5(\mathbb{Q})$ .*

**Proof** Every point in  $\partial\mathcal{CP}_5$  vanishes in at least one of the irreducible polynomials generating a component of the Zariski closure of  $\partial\mathcal{CP}_5$ . The points that vanish in at least two such polynomials lie in a proper subvariety of  $\partial\mathcal{CP}_5$  that has measure zero. The complement is therefore a Euclidean dense open subset  $U$  of  $\partial\mathcal{CP}_5$ . For each such polynomial  $p$  the points of  $\partial\mathcal{CP}_5$  that vanish in  $p$  and no other polynomial form a Euclidean open subset  $U_p$  of  $\partial\mathcal{CP}_5$ .  $U$  is the union of each of the  $U_p$ . Thus it suffices

to show density of the rational points in each  $U_p$ . The result is clear in the parts of the boundary that have low rank or the positive semidefinite cone, so it suffices to prove density in the Hildebrandt and Horn parts of the boundary.

Let  $A$  be a matrix in  $\partial\mathcal{CP}_5 \cap \mathcal{DN}\mathcal{N}_5^\circ$  that is (only) in the Horn locus. Since  $A$  is in the interior of  $\mathcal{DN}\mathcal{N}_5^\circ$  all the entries are strictly positive. We can approximate it by a sequence of rational matrices  $A_n$  that is defined as follows: Every entry of  $A_n$  except for the first one is a rational number of distance less than  $1/n$  to the corresponding entry in  $A$  and the first entry is forced by the determinantal Equation 2.7. This sequence converges to  $A$ .

The set of matrices that are only in the Hildebrand locus is a Euclidean open subset of the variety  $V_{Hi}$ . Thus, it suffices to show that we can approximate any  $A \in \partial\mathcal{CP}_5 \cap \mathcal{DN}\mathcal{N}_5^\circ$  that is only in the Hildebrand locus by rational matrices in  $V_{Hi}$ . We assume that  $A = BB^T$  is the cp-factorization of  $A$ , and we approximate  $B$  by a sequence of rational matrices  $B_n$  where again all entries except the first one are rational numbers of distance less than  $1/n$  to the corresponding entry in  $B$  and the first entry is uniquely determined by the equation defining  $Z_{Hi}$  as in Theorem 2.10. Then this sequence converges to  $B$  and  $B_n B_n^T$  is a sequence of rational matrices converging to  $A$ .  $\square$

## 2.6 The Action of the Permutation Matrices

Notice that in parts (3) and (4) of Theorem 2.1, there are some permutation matrices  $P$  that we have ignored so far. To understand all the components of the boundary of  $\mathcal{CP}_5$  contained in  $\mathcal{DN}\mathcal{N}_5^\circ$ , we have to consider the effect of these matrices, which permute rows and endow algebraic components of the boundary with an action of the symmetric group  $\mathfrak{S}_5$ . The group acts on the algebraic components of the boundary by conjugation, and on the space of factorizations by multiplication on the left. This makes the parametrizations into equivariant maps. Thus what we need to count is the number of hypersurfaces obtained by acting on  $Z_{Horn}$  and  $Z_{Hi}$  with  $\mathfrak{S}_5$  by multiplying with permutation matrices on the left. The orbit-stabilizer theorem says we need to find how many such matrices fix either variety.

To make this more precise recall that if  $A \in \mathcal{CP}_5 \cap \mathcal{DN}\mathcal{N}_5^\circ$  and  $A = BB^T$ , then any matrix  $Q \in \mathcal{O}(5)$  yields a factorization  $A = (BQ)(BQ)^T$ . In general, the matrices of  $\mathcal{O}(5)$  have negative entries, so preserving positivity is nontrivial and known to fail in  $\partial\mathcal{CP}_5 \cap \mathcal{DN}\mathcal{N}_5^\circ$  unless  $Q$  is a permutation matrix. In this case  $Q$  permutes columns of  $B$  and preserves positivity.

Let two generic points in  $\partial\mathcal{CP}_5 \cap \mathcal{DN}\mathcal{N}_5^\circ$  be either in the Zariski closures of the Horn or the Hildebrand locus. Then they are in the same algebraic component if the factorizations share the same zero pattern up to this column permutation. This is because each matrix has at most two nonnegative factorizations and they are known to satisfy this property [22].

The left action of the symmetric group permutes rows and the condition is satisfied precisely if the matrices shift rows cyclically or reflect the rows and then shift them cyclically. It follows that the stabilizer of the algebraic components of  $\partial\mathcal{CP}_5 \cap \mathcal{DN}\mathcal{N}_5^\circ$  under the  $\mathfrak{S}_5$ -action is a dihedral group and has order 10. As a consequence, the orbit

of each algebraic component consists of 12 varieties. Putting this together we obtain the following:

**Theorem 2.13** *The Zariski closure of the hypersurface  $\partial\mathcal{CP}_5 \cap \mathcal{DN}\mathcal{N}_5^\circ$  is the vanishing locus of a 15 variable polynomial of degree  $3900 = (320 + 5)12$ .*

### 2.7 Completely Positive Rank and the Interior of $\mathcal{CP}_5$

Another rather interesting aspect of the cone  $\mathcal{CP}_n$  comes with regard to the so called cp-rank. This will actually help us understand the cone  $\mathcal{CP}_5$  in more detail. Throughout this section, the topology on  $\mathcal{CP}_5$  and all of its subsets is the Euclidean topology.

**Definition 2.14** The cp-rank  $cp(A)$  of a matrix  $A \in \mathcal{CP}_n$  is the minimal size  $k$  such that there is a nonnegative  $n \times k$  matrix  $B$  with  $A = BB^T$ . The  $cp^+$ -rank  $cp^+(A)$  of  $A$  is the smallest such  $k$  for which  $B$  can be taken to have strictly positive entries.

As a corollary of Theorem 2.2, we know that all matrices in  $\partial\mathcal{CP}_5 \cap \mathcal{DN}\mathcal{N}_5^\circ$  can have cp-rank at most 5. We would like to understand the cp-ranks and  $cp^+$ -ranks in the interior of  $\mathcal{CP}_5$ . It is known [7, Thm. 5.1] that they agree generically on an open subset of the interior of  $\mathcal{CP}_5$ . For that, we consider a few extra subsets of  $\mathcal{CP}_n$ .

**Definition 2.15** Let  $\mathcal{CP}\mathcal{R}_n(k)$  be the set of all matrices  $A \in \mathcal{CP}_n$  such that  $cp(A) \leq k$ .

**Theorem 2.16** *For every  $n, k$  the set  $\mathcal{CP}\mathcal{R}_n(k)$  is a closed (not necessarily convex) cone.*

**Proof** Let  $\mathbb{R}_{\geq}^{n \times k}$  be the set of all  $n \times k$  matrices with nonnegative entries and let  $\varphi : \mathbb{R}_{\geq}^{n \times k} \rightarrow \mathcal{CP}_n$  be given by  $\varphi(X) = XX^T$ . This is a continuous map whose image is precisely  $\mathcal{CP}\mathcal{R}_n(k)$ . Now let  $A$  be a matrix in the closure of  $\mathcal{CP}\mathcal{R}_n(k)$  and choose a sequence of matrices  $\{A_i\}_{i=1}^\infty$  that converges to  $A$ . For each  $i$ , pick a matrix  $B_i$  in the preimage of  $A_i$  under  $\varphi$ . Notice that  $\{B_i\}_{i=1}^\infty$  is a sequence of matrices that are entrywise bounded: indeed,  $(B_i)_{j,\ell} \leq \sqrt{\sum_m (B_i)_{j,m}^2} = \sqrt{(A_i)_{j,j}}$  and the diagonal entries of the matrices  $A_i$  are bounded since they converge to  $A$ . Thus  $\{B_i\}_{i=1}^\infty$  has a convergent subsequence in the Euclidean topology and this subsequence converges to a matrix  $B$  such that  $\varphi(B) = A$ . □

Our goal is to understand  $\partial\mathcal{CP}\mathcal{R}_5(5) \cap \mathcal{CP}_5^\circ$ . Since the matrices in this set are in the interior of  $\mathcal{CP}_5$ , they are all invertible and by Theorem 2.16 their cp-rank is exactly 5.

**Corollary 2.17** *The intersection  $\partial\mathcal{CP}\mathcal{R}_5(5) \cap \mathcal{CP}_5^\circ$  consists entirely of matrices whose cp-rank is equal to 5 and whose  $cp^+$ -rank is equal to 6.*

**Proof** The interior of  $\mathcal{CP}_5$  consists of invertible matrices, meaning that every matrix in the interior has cp-rank at least 5. Combining [7, Theorem 4.1] together with the fact that maximal cp-rank in  $\mathcal{CP}_5$  is equal to 6 [22] tells us that

$$\max\{cp^+(X) \mid X \in \mathcal{CP}_5^\circ\} = \max\{cp(X) \mid X \in \mathcal{CP}_5^\circ\} = 6.$$

By [7, Corollary 2.7] and Theorem 2.16, all the points of  $\overline{\mathcal{CPR}_5(5)} \setminus (\mathcal{CPR}_5(5)^\circ \cup \partial\mathcal{CP}_5)$  are matrices whose cp-rank is 5 and whose cp<sup>+</sup>-rank is 6. According to Theorem 2.16, every point in  $\overline{\mathcal{CPR}_5(5)} \setminus (\mathcal{CPR}_5(5)^\circ \cup \mathcal{CP}_5)$  is an accumulation point and hence an element of  $\partial\mathcal{CPR}_5(5) \cap \mathcal{CP}_5^\circ$ .  $\square$

We are interested in the factorization of elements in  $\partial\mathcal{CPR}_5(5) \cap \mathcal{CP}_5^\circ$ . Corollary 2.17 implies that the interior of the cone is contained in the closure of two disjoint open sets, whose boundary is described in terms of matrices whose factorizations are quite special. In particular, all of them must have a large number of entries that are zero.

While figuring out the exact pattern is complicated, there are still a few things that could be said. By manipulating the zeros as in [13], we can assume that every column of a  $5 \times 5$  factor matrix contains at least one zero. Furthermore, in a factorization with the smallest possible number of zeros, there can be no pair of columns such that the zero entries of one column are also zero entries of the other column, i.e., the zero pattern of one column cannot be contained in the zero pattern of another.

Since the dimension of the variety  $\partial\mathcal{CPR}_5(5) \cap \mathcal{CP}_5^\circ$  is 14, the number of zeros in the factorization is between 5 and 11. The possible patterns that fulfill all these constraints can be classified combinatorially and are many. But even if the zero patterns of such matrices are known, we still need to work out a number of polynomial equations and an unknown number of inequalities that parametrize them. This is difficult also due to the lack of convexity of the relevant cones making it impossible to use duality techniques as for the boundary of  $\mathcal{CP}_5$ . Furthermore, even though the factorizations should in principle be unique (in the spirit of Theorem 2.3), there could be other factorizations that are not nonnegative but very close. This is an issue even in the boundary  $\partial\mathcal{CP}_5$ , as will be discussed in Sec. 3.2, and it makes the generation of possible examples very difficult.

Nevertheless, we believe that we can produce exact examples of such matrices, as will be shown in Sec. 5.5. The numerical experiments suggest that the number of zeros in the square factorization would be 10 or 11. We remark that having factorizations with 11 zeros seems implausible and worth further consideration. On the one hand, since  $\partial\mathcal{CPR}_5(5) \cap \mathcal{CP}_5^\circ$  is a hypersurface, its dimension is 14 and the matrices with 11 zeros would already provide a parametrization of a component of the Zariski closure of  $\partial\mathcal{CPR}_5(5) \cap \mathcal{CP}_5^\circ$ . On the other hand, this boundary remains elusive: Simply inserting random values into a given zero pattern will not produce a desired factor matrices, because of the unknown polynomial inequalities. We discuss this in detail in Sec. 5.5.

### 3 Computing the cp-Factorization

Deciding whether a matrix allows a cp-factorization, much less than computing its cp-rank, is a co-NP-complete problem [20, 23]. A number of problems that are themselves NP-hard, such as the clique problem or the standard quadratic problem [16], can be reformulated to conic programs over  $\mathcal{CP}_n$ . Therefore, it is desirable to compute the cp-factorization, at least to a certain accuracy.

A number of algorithms for the computation of the completely positive factorization of a matrix  $A$  have been proposed. These algorithms must assume that the cp-rank  $r$  is known and preset. If this choice turns out to be wrong, the algorithm will either not terminate or it will present a solution that does not fulfill all criteria, i.e., it is either not symmetric, not nonnegative, or it does not yield the desired matrix.

There are several problems that a numerical algorithm can aim to solve, to some accuracy or certainty:

- Factorization problem: Given  $A \in \mathcal{CP}_n$ , compute  $A = BB^\top$  for given rank  $r$ .
- Membership problem: Given  $A \in \mathcal{DN}\mathcal{N}_n$ , decide if  $A \in \mathcal{CP}_n$ .
- Approximation problem: Given  $A \in \mathcal{DN}\mathcal{N}_n \setminus \mathcal{CP}_n$ , find best approximation in  $\mathcal{CP}_n$ .
- Boundary problem: Given  $A \in \mathcal{CP}_n$ , decide if  $A \in \partial\mathcal{CP}_n$ . In particular, find the unique factorization  $A = BB^\top$ .

We can distinguish two general classes of algorithms. The first approach is to begin with any symmetric factorization  $A = BB^\top$  and then to iteratively alter  $B \in \mathbb{R}^{n \times r}$  such that it becomes nonnegative. This can for example be done by picking an initial orthogonal matrix  $Q \in \mathcal{O}(r)$  and projecting it onto the polyhedral cone

$$\mathcal{P} = \{R \in \mathbb{R}^{r \times r} : BR \geq 0\}.$$

After this, the result will in turn be projected back onto  $\mathcal{O}(r)$ , whereupon the procedure repeats until  $BQ$  remains nonnegative and therefore constitutes a solution. This method has been proposed in [15], it is very fast and it returns an exact solution since  $BQ(BQ)^\top = BB^\top = A$ , at least up to numerical accuracy for the representation of  $Q$ . However, it has the severe drawback that the set of orthogonal matrices  $Q$  with  $BQ \geq 0$  needs to be sufficiently large, or else the algorithm often fails to converge. In the boundary  $\mathcal{CP}_n$ , the factor matrices have many zeros and therefore this set is of high codimension in  $\mathcal{O}(r)$ . Thus, if  $A$  is (very) close to the boundary  $\partial\mathcal{CP}_n$ , the algorithm will fail and it cannot be used to distinguish these matrices from the ones that do not allow a completely positive factorization of rank  $r$ .

The alternative approach consists of algorithms that approximate the exact factorization while maintaining symmetry and nonnegativity, see for example [14]. In this paper, we propose a version of these methods that to our knowledge has not been applied to the problem at hand, although it uses only standard tools of numerical approximation. The goal is to minimize the function

$$f(B) = \|A - BB^\top\|^2.$$

In order to guarantee nonnegativity of  $B$ , we can write its entries as squares, resulting in  $B = C \circ C$ . Together with a factor that simplifies the gradient, we aim to minimize the smooth function

$$g(C) = \frac{1}{8} \|A - (C \circ C)(C \circ C)^\top\|^2.$$

---

**Algorithm 1:** Trust region scheme for the computation of the cp-factorization.

---

**Input:**  $A \in \mathcal{DN}\mathcal{N}_n$ , initial point  $C_0 \in \mathbb{R}_{\geq}^{n \times r}$ , tolerance  $\varepsilon > 0$ , maximal number of iterations  $\text{maxit}$

$k = 0$ ;

**while**  $\|\nabla g(C_k)\| > \varepsilon$  **and**  $k < \text{maxit}$  **do**

Perform trust region step  $C_{k+1} := \text{TrustRegion}(C_k, g, \nabla g(C_k))$ ;

$k \leftarrow k + 1$ ;

**end**

$B = C_k \circ C_k$ ;

**Output:** Approximate factor matrix  $B \in \mathbb{R}_{\geq}^{n \times r}$

---

This is done using standard tools from numerical optimization. Since we deal with matrices, the MATLAB-toolbox `manopt` allows for an easy implementation and good performance [9]. We applied the provided trust region scheme, because it proves to be faster than equally applicable methods like gradient descent or nonlinear conjugate gradients.<sup>1</sup> The only other ingredient that we need is the Euclidean gradient of our function  $g$ , which can be readily given as

$$\nabla g(C) = \left( (C \circ C)(C \circ C)^T(C \circ C) \right) \circ C - \left( A(C \circ C) \right) \circ C.$$

See Algorithm 1 for the implementation in pseudo-code.

### 3.1 Enforcing Zeros

We know that the concept of zero patterns in the factorization plays an important role in the characterization of the boundary  $\partial\mathcal{CP}_5$  or  $\mathcal{CPR}_5(5) \cap \mathcal{CP}_5^{\circ}$ . For the purpose of numerical experiments, it can therefore be beneficial to enforce a specific zero pattern in the solution, for example in order to find a matrix in these sets. Hence, we note that if the initial point of the optimization, say  $C_0$ , has a given zero pattern, then so does the gradient  $\nabla g(C_0)$ . Since we use a finite difference approximation of the Hessian, any next iterate  $C_1$  will still have the given zero pattern, and so on. This means that enforcing a zero pattern in the algorithm can be done by simply starting out with this pattern. However, we also need to take into consideration that we effectively look for a solution on the intersection of a linear space with a hypersurface (e.g., a part of the boundary), which can lead to undesired effects.

### 3.2 Condition of the Reconstruction Problem

The above approach essentially means that we are trying to find a global minimum of a polynomial of order 8, resulting in many local minima and possibly in an ill-

---

<sup>1</sup> The trust region method consists of solving a quadratic approximation of the cost function  $g$  on a small *trusted* region around the current iterate. Its size is adapted throughout the procedure according to prior performance. In `manopt`, the Hessian is approximated using finite differences of the gradient. We do not go into more detail here and simply set the next iterate as  $C_{k+1} := \text{TrustRegion}(C_k, g, \nabla g(C_k))$ . See [21] for an accessible introduction into these methods.

conditioned problem. However, we can simply restart the method if it does not produce an actual solution (meaning  $g(C) = 0$ ). This works as long as we know the cp-rank of our matrix. Otherwise, the algorithm will produce an approximation of our matrix and we can use several tries to find the best one. This seems to work well in practice.

The issue of conditioning is more problematic. Naïvely, if  $g(C) = \varepsilon$  for a small error  $\varepsilon > 0$ , one would expect the error  $\|B - (C \circ C)\|$  to be of order  $O(\varepsilon^{\frac{1}{8}})$ , where  $B$  is an exact factorization of  $A$ . It seems however that the error of the factor matrices can sometimes be even larger than that, suggesting that the conditioning of the problem is worse. We will investigate this issue numerically in Sec. 5.4. Ultimately, the problem of ill-conditioning is inherent in the problem structure and not in the algorithm itself. Any algorithm that is subject to numerical noise will suffer from it.

### 4 Constructing Examples

We aim to test the above algorithm’s performance for the factorization problem, the membership problem, the approximation problem, and the boundary problem. Using our knowledge about the boundary in the case  $n = 5$ , we can construct more or less generic matrices in many different parts of the cone  $\mathcal{DN}\mathcal{N}_5$ . First of all, let it be stated that picking a matrix  $M \in \mathcal{DN}\mathcal{N}_5$ , say of fixed Frobenius norm  $\|M\| = 1$ , uniformly at random is not entirely trivial. If we just pick 15 nonnegative entries of  $M$  on the upper triangle and normalize, we will most likely not have a positive semidefinite matrix. This could be remedied with the so-called rejection algorithm, where such a randomly chosen matrix  $M \in \mathcal{N}_5$  is rejected precisely when it has some negative eigenvalues. In any case, a generic matrix in  $\mathcal{DN}\mathcal{N}_5$  will either have cp-rank 5 or 6, or it will not allow for a completely positive factorization. In the following, we will discuss several other strategies of how to obtain interesting matrices.

#### 4.1 Matrices in $\partial\mathcal{CP}_5$

The parts of the boundary  $\partial\mathcal{CP}_5$  that are also in the boundary of  $\mathcal{DN}\mathcal{N}_5$  are very easy to reproduce, either by picking a matrix of rank 4 or less, or by keeping one or many entries equal to zero. We have identified the remaining parts as the Horn and Hildebrand locus respectively. Since the algebraic and semialgebraic equations that describe these sets are rather complicated, we have to use an indirect approach and choose the factor matrices.

In order to obtain points in the Horn locus, one can simply choose 15 variables  $x_1, \dots, x_5, y_1, \dots, y_5, z_1, \dots, z_5$  (e.g. uniformly at random) and produce a factor matrix

$$B_{Horn} = \begin{pmatrix} x_1 & 0 & 0 & 0 & 0 \\ 0 & x_2 & 0 & 0 & 0 \\ 0 & 0 & x_3 & 0 & 0 \\ 0 & 0 & 0 & x_4 & 0 \\ 0 & 0 & 0 & 0 & x_5 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & y_4 & y_5 + 1 \\ y_1 + 1 & 1 & 0 & 0 & y_5 \\ y_1 & y_2 + 1 & 1 & 0 & 0 \\ 0 & y_2 & y_3 + 1 & 1 & 0 \\ 0 & 0 & y_3 & y_4 + 1 & 1 \end{pmatrix} \begin{pmatrix} z_1 & 0 & 0 & 0 & 0 \\ 0 & z_2 & 0 & 0 & 0 \\ 0 & 0 & z_3 & 0 & 0 \\ 0 & 0 & 0 & z_4 & 0 \\ 0 & 0 & 0 & 0 & z_5 \end{pmatrix}. \tag{4}$$

If all variables are chosen to be positive, this will yield an element in the boundary  $A_{Horn} = B_{Horn} B_{Horn}^T \in V_{Horn} \cap \partial \mathcal{CP}_5$ .

In the case of the Hildebrand locus, this is done similarly by choosing  $x_1, \dots, x_5, \theta_1, \dots, \theta_5, z_1, \dots, z_5$  and replacing the middle matrix in (4) by  $S(\Theta)$  in (2.9), keeping in mind that  $\sum_i \theta_i < \pi$  must hold. The resulting matrix  $A_{Hi} = B_{Hi} B_{Hi}^T$  will however be only approximately in the boundary  $V_{Hi}$ , since the trigonometric functions can only be approximated in general. In order to obtain matrices in  $V_{Hi}$  with rational entries, we can round all entries of  $B_{Hi}$  to a given number of digits except for one, which will then be determined by the fact that  $y_{11}y_{22}y_{33}y_{44}y_{55} - y_{13}y_{24}y_{35}y_{41}y_{52} = 0$  must hold.

### 4.2 Tangent Spaces of the Boundary and Their Orthogonal Lines

Examples of matrices in  $\mathcal{DN}\mathcal{N}_5$  that do not allow for a completely positive factorization of any rank, i.e., of matrices in  $\mathcal{DN}\mathcal{N}_5 \setminus \mathcal{CP}_5$ , are rare. Some are given in [3]. In theory, all extreme rays of  $\mathcal{DN}\mathcal{N}_5$  are known and those that are *not* extreme rays of  $\mathcal{CP}_5$  should in principle produce more examples, but obtaining rational matrices this way could be difficult due to the involvement of trigonometric functions.

With the knowledge of the boundary  $\partial \mathcal{CP}_5 \cap \mathcal{DN}\mathcal{N}_5^o$ , we are able to present a more systematic procedure to produce such examples for the case of  $5 \times 5$  matrices. For this, we calculate a normal direction to this part of the boundary. Then, any point in this direction will be outside  $\mathcal{CP}_5$ . If we take a small enough step, we will often find cases that are still in  $\mathcal{DN}\mathcal{N}_5$ . We describe this procedure in more detail for the Hildebrand locus:

Let  $B_{Hi}$  be the factor matrix as in Sec. 4.1. Theorem 2.10 states that then  $y_{52} = \frac{y_{11}y_{22}y_{33}y_{44}y_{55}}{y_{13}y_{24}y_{35}y_{41}}$  and therefore, we have a smooth parametrization  $\varphi : \mathbb{R}^{14} \rightarrow V_{Hi} \subset \mathbb{R}^{15}$ . Taking the orthogonal complement with respect to the Bombieri-norm in  $\mathbb{R}^{15}$  of the Jacobian  $\nabla \varphi(y)$  reveals the normal direction in  $\mathbb{R}^{15}$ .

The same procedure works for the Horn locus, either by using the much more complicated polynomial equation  $\det(H \circ X) = 0$ , or by taking the derivative of the 15-dimensional parametrization (4) and using the fact that the Jacobian will be rank-deficient.

### 4.3 Matrices in $\mathcal{CP}_5 \setminus \mathcal{CPR}_5(5)$

It is also interesting to obtain matrices of cp-rank 6, i.e., matrices in  $\mathcal{CP}_5 \setminus \mathcal{CPR}_5(5)$ . In [22], a Kronecker-structured matrix of cp-rank 6 is given. However, it has rank 4 and it is therefore an element of  $\partial \mathcal{DN}\mathcal{N}_5$ . In order to obtain matrices in the interior, we can use some small perturbations of these matrices by matrices in the interior of  $\mathcal{CP}_5$ . Alternatively, a brute force sampling of  $\mathcal{DN}\mathcal{N}_5$ , as performed in the next section, sometimes (albeit rarely) yields a matrix of cp-rank 6.



**Table 1** Success rate of the factorization for different distances to the Hildebrand locus out of 100 tries (average time in brackets)

$\lambda$	$10^{-5}$	$10^{-4}$	$10^{-3}$	$10^{-2}$	$10^{-1}$
Dür, Groetzner [15]	8 (5.7s)	26 (6.1s)	66 (5.5s)	98 (2.0s)	100 (0.2s)
Ding et al. [14]	16 (4.4s)	63 (3.4s)	98 (1.6s)	100 (0.2s)	100 (0.0s)
Pfeffer, Samper	96 (38.0s)	100 (4.6s)	100 (1.4s)	100 (0.6s)	100 (0.4s)

## 5 Numerical Experiments

We now perform numerical experiments for the aforementioned problems in the case  $n = 5$ , as well as some other experiments that serve to highlight interesting aspects of the cp-cone. We will use our approximation algorithm and we will see that even for the smallest nontrivial case, the cp-factorization problem is very complicated.

### 5.1 Factorization Problem

As a first experiment, we test the performance of our algorithm against the two aforementioned algorithms in [15] and [14] respectively. It is known that the factorization becomes harder the closer we get to the boundary  $\partial\mathcal{CP}_5$ . Since we are able to construct matrices in this boundary, we can test the algorithms for different distances to this boundary. For this, we construct a random matrix in the Hildebrand locus  $A \in V_{Hi}$  and a random matrix  $\tilde{A} \in \mathcal{CP}_5^{\circ}$  (by generating a random factor matrix  $B$  with uniformly distributed entries between 0 and 1 and setting  $\tilde{A} = BB^T$ ). We then apply each algorithm to the matrix  $\lambda A + (1 - \lambda)\tilde{A}$ . This is repeated 100 times for different values of  $\lambda$  and we report the success rate as well as the average CPU time. In each setting, we run each algorithm 10 times (with different initializations), allowing for 10,000,000 iteration steps (our algorithm never reaches that limit) and declaring the factorization successful if the error in the Frobenius norm is smaller than  $10^{-6}$  for one of these runs. We used an Intel i7-10510U CPU with 1.80GHz for each core. See Table 1 for the results. Our algorithm can factorize the matrix for any value of  $\lambda$  (in almost all cases) while the other two fail for small distances to the boundary. We note that the algorithm of Ding et al. would often have eventually found the factorization but it was cut off by the iteration limit. When successful, the three algorithms are comparable in terms of CPU times (note that we report average CPU times only for successful attempts).

Next, we test the three algorithms for different sizes  $n$  of randomly generated matrices  $A \in \mathcal{CP}_n$  (again setting  $A = BB^T$  for uniformly random  $B$ ). We again run each algorithm 10 times with 10,000,000 iterations and declare the factorization successful if the Frobenius error is smaller than  $10^{-6}$ . We used the same CPU and report the success rate out of 100 experiments as well as the average CPU time of the successful attempts. See Table 2 for the results. Our algorithm finds a factorization for each matrix size and it is also slightly faster than the other two algorithms. These are comparable in CPU time but they fail to converge in the given number of iterations for large matrices.

**Table 2** Success rate of the factorization for different matrix sizes  $n$  out of 100 tries (average time in brackets)

$n$	5	10	20	30	50
Dür, Groetzner [15]	100 (0.3s)	100 (6.5s)	6 (27.9s)	0	0
Ding et al. [14]	95 (0.3s)	89 (2.3s)	65 (11.1s)	36 (28.3s)	0
Pfeffer, Samper	100 (0.4s)	100 (1.2s)	100 (6.6s)	100 (169.3s)	100 (2006.2s)

**Table 3** Occurrences of matrices out of 50000 random draws in  $\mathcal{DN}\mathcal{N}_5$  with the rejection algorithm

$A \in \mathcal{DN}\mathcal{N}_5$	$A \in \mathcal{DN}\mathcal{N}_5 \setminus \mathcal{CP}_5$	$A \in \mathcal{CPR}_5(5)$	$A \in \mathcal{CPR}_5(6)$
50000	8	49991	1

In conclusion, our algorithms outperforms the existing algorithms in terms of success rate both when closer to the boundary and for larger matrices. For this reason, we will only use our own algorithm for the remaining experiments.

### 5.2 Membership Problem

Given a generic matrix  $A \in \mathcal{DN}\mathcal{N}_5$ , can we decide if it allows a cp-factorization for a given cp-rank  $r$ ? As discussed above, a generic matrix in  $\mathcal{DN}\mathcal{N}_5$  can be picked using the rejection algorithm. In Table 3, we report on the occurrences of matrices in different full-dimensional parts of  $\mathcal{DN}\mathcal{N}_5$  out of 50000 of such picks. For each pick, we generated the 15 unique entries of the symmetric matrix  $A$  uniformly at random in  $(0, 1)$ . Then we calculated the eigenvalues of this matrix and rejected the matrix if any one of the eigenvalues was negative. Note that we did not normalize the matrix, since linear scaling does not have an effect on this experiment. The resulting matrix  $A$  will be an element of  $\mathcal{DN}\mathcal{N}_5$ . We then ran our algorithm 10 times with cp-rank 5 and we consider  $A \in \mathcal{CPR}_5(5)$  if the resulting factorization  $BB^T$  of any of these runs has an error

$$\|A - BB^T\| < 10^{-8}.$$

If out of these 10 runs none was successful, we increased the cp-rank to 6 and repeated the experiment. If again no run produced an error smaller than  $10^{-8}$ , we consider the matrix  $A \in \mathcal{DN}\mathcal{N}_5 \setminus \mathcal{CP}_5$ .

We can see that the vast majority of generic matrices in  $\mathcal{DN}\mathcal{N}_5$  has cp-rank 5. A small number of matrices does not allow for a cp-factorization and only one matrix seems to have cp-rank 6. Note that our results are inherently approximate, and the membership cannot be established with absolute certainty.

### 5.3 Approximation Problem

Using the method discussed in Sec. 4.2, we can produce examples in  $\mathcal{DN}\mathcal{N} \setminus \mathcal{CP}_5$ . Since the cone is also convex, we know the best approximations in  $\mathcal{CP}_5$  for these matrices and we can use our algorithm to try to find them.

**Table 4** Success rate of the approximation out of 100 tries for the computation of a factorization (a) and for reconstructing the original factor matrix (b) for different distances from the boundary

$10^{-5}$		$10^{-4}$		$10^{-3}$		$10^{-2}$		$10^{-1}$	
a)	b)	a)	b)	a)	b)	a)	b)	a)	b)
87	66	87	71	74	66	63	57	66	62

Table 4 shows the success rate for this approximation problem. For 100 random matrices in  $V_{Hi}$ , we computed the normal directions and perturbed the matrices in this direction with distances  $10^{-5}$  up to  $10^{-1}$ . For each distance, we again performed our algorithm 10 times and reported the approximation successful if the reconstructed matrix has an error of  $10^{-6}$ . As before, we considered the reconstruction of the factor matrix successful if the factor matrices exhibited an error of less than  $10^{-3}$ . If at the beginning we successfully reconstructed the *perturbed* matrix, we took this as a sign that we moved into the interior of the cone and thus changed directions for the following steps.

One can see that the algorithm is able to find the closest matrix in  $\mathcal{CP}_5$  in many cases. Peculiarly, it also finds the correct factorization more often than in the reconstruction problem. This is consistent with all our experiments and we suspect that it happens because the value of the cost function does not tend to zero and the algorithm therefore does not terminate prematurely. However, the approximation problem seems to be more difficult and it runs into many local minima. In many cases, we did not find the original matrix over the 10 attempts. This problem seems to get worse with a larger distance from the boundary.

## 5.4 Experiments on the Boundary

As shown, in contrast to [15] and [14], our numerical algorithm allows us to factorize matrices in  $\partial\mathcal{CP}_5$ . The factorizations of matrices in the Hildebrand locus are unique and should in principle be reconstructable. Matrices in the Horn locus have exactly two factorizations, meaning a successful reconstruction of the factor matrices should occur in about 50% of the attempts. However, as we have discussed in Sec. 3.2, the factorization problem is ill-conditioned. We will illustrate this with the following experiment.

Table 5 shows an experiment for the reconstructability of different parts of the cone. Since the boundary is of lower dimension, a random pick in  $\mathcal{DN}_5$  will never produce a matrix on the boundary. Therefore, we resort to the procedures presented in Sec. 4 in order to obtain more or less random elements of  $\partial\mathcal{CP}_5$ .

With these methods, we randomly generated 100 matrices in the interior  $\mathcal{CP}_5^\circ$ , the Horn locus, the Hildebrand locus, the rank-deficient locus, and those containing (at least) one zero. For each of the matrices we performed our algorithm 10 times and considered it successful if the reconstructed matrix differs by less than  $10^{-6}$  from the original matrix, in terms of Frobenius error. In Table 5, we can see that our algorithm successfully finds such a factorization in all cases.

**Table 5** Success rate of the algorithm out of 100 tries for the computation of a factorization (a) and for reconstructing the original factor matrix (b) for different parts of  $\mathcal{CP}_5$

$A \in \mathcal{CP}_5^\circ$		$A \in V_{Horn}$		$A \in V_{Hi}$		$\text{rank}(A) < 5$	$\exists(i, j) : A_{i,j} = 0$
a)	b)	a)	b)	a)	b)		
100	0	100	31	100	49	100	100

Furthermore, for matrices in the interior, in the Horn locus, and in the Hildebrand locus, we can also compare the reconstructed factor matrices. This is not possible for rank-deficient matrices, because we applied our algorithm with initial rank 5 and the actual factor matrix has size  $5 \times 4$ . Similarly, we do not know the original factor matrix for matrices with a zero entry, and therefore we cannot compare our results to it. We consider the factorization to be successful if the factor matrices differ by less than an error of  $10^{-3}$  from the original factors. It is unsurprising that we never find the original factorization of a matrix in the interior of the cone since there are infinitely many factorizations. For the Horn locus, we expect a successful reconstruction of the factor matrices in about 50% of the cases. However, as we can see, this number is lower due to ill-conditioning of the problem. Similarly, reconstructing the factor matrices of a matrix in the Hildebrand locus was successful only in about half of the cases, even though this factorization is unique.

### 5.5 Numerical Observations on the Boundary $\partial\mathcal{CPR}_5(5) \cap \mathcal{CP}_5^\circ$

In the previous sections, we have given an exhaustive description of the boundary  $\partial\mathcal{CP}_5$  of the convex cone  $\mathcal{CP}_5$ . Since the maximal cp-rank of a  $5 \times 5$  matrix is 6, the only remaining interesting part of the cone is the intersection  $\partial\mathcal{CPR}_5(5) \cap \mathcal{CP}_5^\circ$ , which we briefly discussed in Sec. 2.7. A detailed description of this “interior boundary” is out of reach even for the case  $n = 5$ , because it is not derived from the extreme rays of the dual problem, nor is the set  $\mathcal{CPR}_5(5)$  convex. With all current tools at our disposal, it seems like a general description of this set comes down to the combinatorial evaluation of all possible nonreducible zero patterns as for example done in [17].

Nevertheless, it is possible to make some numerical observations. Given a matrix  $A$  of cp-rank 6 (see Sec. 4.3) and running our algorithm with rank 5, it returns an approximate factorization of cp-rank 5. If subsequent runs with different random initial inputs return the same factorization, one can reasonably conclude that this approximate factorization is in fact unique and thus an element of  $\partial\mathcal{CPR}_5(5) \cap \mathcal{CP}_5^\circ$ . In this fashion, we derived the following interesting example:

We begin with the matrix

$$A_6 = \begin{pmatrix} 0.4722 \dots & 0.1493 \dots & 0.0225 \dots & 0.1083 \dots & 0.0296 \dots \\ 0.1493 \dots & 0.3519 \dots & 0.1442 \dots & 0.0111 \dots & 0.1316 \dots \\ 0.0225 \dots & 0.1442 \dots & 0.4121 \dots & 0.2120 \dots & 0.0157 \dots \\ 0.1083 \dots & 0.0111 \dots & 0.2120 \dots & 0.2113 \dots & 0.0719 \dots \\ 0.0296 \dots & 0.1316 \dots & 0.0157 \dots & 0.0719 \dots & 0.4366 \dots \end{pmatrix}.$$

This matrix has cp-rank 6. Several runs of our algorithm with rank 5 reveal the best cp-rank 5 approximation  $A_5$ , which has the factorization

$$B_5 = \begin{pmatrix} 0 & 0.6216\dots & 0.2872\dots & 0.0578\dots & 0 \\ 0 & 0 & 0.4455\dots & 0.3692\dots & 0.1309\dots \\ 0.5097\dots & 0 & 0 & 0.3904\dots & 0 \\ 0.4158\dots & 0.1742\dots & 0 & 0 & 0.0902\dots \\ 0.0310\dots & 0 & 0.1036\dots & 0 & 0.6519\dots \end{pmatrix}.$$

The matrix  $A_5$  is an element of  $\partial\mathcal{CP}\mathcal{R}_5(5) \cap \mathcal{CP}_5^\circ$  and its factor matrix  $B_5$  has 11 zeros. This was consistent throughout all our experiments with  $\partial\mathcal{CP}\mathcal{R}_5(5) \cap \mathcal{CP}_5^\circ$ . But since this is a hypersurface of dimension 14, there can be no additional polynomial equation that defines it, meaning that the other entries of  $B_5$  can be altered freely and we will always get a matrix in  $\partial\mathcal{CP}\mathcal{R}_5(5) \cap \mathcal{CP}_5^\circ$  (up to semialgebraic equations, i.e., inside some possibly small intervals).

In order to verify this observation, we can round the entries of  $B_5$  to 2 decimals:

$$\tilde{B}_5 = \begin{pmatrix} 0 & 0.62 & 0.29 & 0.06 & 0 \\ 0 & 0 & 0.45 & 0.37 & 0.13 \\ 0.51 & 0 & 0 & 0.39 & 0 \\ 0.42 & 0.17 & 0 & 0 & 0.09 \\ 0.03 & 0 & 0.10 & 0 & 0.65 \end{pmatrix}.$$

Applying our algorithm to  $\tilde{A}_5 = \tilde{B}_5\tilde{B}_5^T$  recovers the factorization  $\tilde{B}_5$ . This strongly suggests that also  $A_5 \in \partial\mathcal{CP}\mathcal{R}_5(5) \cap \mathcal{CP}_5^\circ$  and that this set is in fact fully described by the zero patterns of the factor matrices.

**Remark 5.1** Since the number of nonnegative factorizations of elements in  $\partial\mathcal{CP}_5 \cap \mathcal{DN}\mathcal{N}_5^\circ$  and  $\partial\mathcal{CP}\mathcal{R}_5(5) \cap \mathcal{DN}\mathcal{N}_5^\circ$  is finite, all these factorizations are locally rigid in the sense of [18]. One may wonder if they are also infinitesimally rigid. One of the main results in Sec. 6 of the article is that any infinitesimally rigid factorization must have at least 11 entries equal to zero. This implies that matrices in  $\partial\mathcal{CP}_5 \cap \mathcal{DN}\mathcal{N}_5^\circ$  do not admit infinitesimally rigid nonnegative factorizations. The situation in  $\mathcal{CP}\mathcal{R}_5(5)$  is different: Some components of the boundary may actually consist of matrices that have a factorization with 11 zeros. As explained above, we have some candidates for such components. However, the examples we found fail to be infinitesimally rigid.

### 5.6 Finding Exact Factorizations

We can also use our algorithm to find exact factorizations of difficult matrices. The article [8] generates a number of matrices of different sizes with high cp-rank. Our algorithm was able to find the (approximate) cp-factorization  $\tilde{B}$  of the  $7 \times 7$  matrix

$$A = \begin{pmatrix} 163 & 108 & 27 & 4 & 4 & 27 & 108 \\ 108 & 163 & 108 & 27 & 4 & 4 & 27 \\ 27 & 108 & 163 & 108 & 27 & 4 & 4 \\ 4 & 27 & 108 & 163 & 108 & 27 & 4 \\ 4 & 4 & 27 & 108 & 163 & 108 & 27 \\ 27 & 4 & 4 & 27 & 108 & 163 & 108 \\ 108 & 27 & 4 & 4 & 27 & 108 & 163 \end{pmatrix}$$

that has cp-rank 14, up to an accuracy  $\|A - \tilde{B}\tilde{B}^T\| \approx 1.179332674168671 \cdot 10^{-8}$ . The peculiar zero pattern together with the fact that many of the entries seemed to be repeated led us to deduce that the exact factorization is

$$B = \begin{pmatrix} \sqrt{27} & \frac{51}{\sqrt{27}} & \sqrt{27} & 0 & 0 & 0 & 0 & \sqrt{\frac{2}{3}} & 0 & 0 & \sqrt{6} & \sqrt{6} & 0 & 0 \\ 0 & \sqrt{27} & \frac{51}{\sqrt{27}} & \sqrt{27} & 0 & 0 & 0 & 0 & \sqrt{\frac{2}{3}} & 0 & 0 & \sqrt{6} & \sqrt{6} & 0 \\ 0 & 0 & \sqrt{27} & \frac{51}{\sqrt{27}} & \sqrt{27} & 0 & 0 & 0 & 0 & \sqrt{\frac{2}{3}} & 0 & 0 & \sqrt{6} & \sqrt{6} \\ 0 & 0 & 0 & \sqrt{27} & \frac{51}{\sqrt{27}} & \sqrt{27} & 0 & \sqrt{6} & 0 & 0 & \sqrt{\frac{2}{3}} & 0 & 0 & \sqrt{6} \\ 0 & 0 & 0 & 0 & \sqrt{27} & \frac{51}{\sqrt{27}} & \sqrt{27} & \sqrt{6} & \sqrt{6} & 0 & 0 & \sqrt{\frac{2}{3}} & 0 & 0 \\ \sqrt{27} & 0 & 0 & 0 & 0 & \sqrt{27} & \frac{51}{\sqrt{27}} & 0 & \sqrt{6} & \sqrt{6} & 0 & 0 & \sqrt{\frac{2}{3}} & 0 \\ \frac{51}{\sqrt{27}} & \sqrt{27} & 0 & 0 & 0 & 0 & \sqrt{27} & 0 & 0 & \sqrt{6} & \sqrt{6} & 0 & 0 & \sqrt{\frac{2}{3}} \end{pmatrix}.$$

We remark that our algorithm did not produce similar results for the other matrices in [8], even over many tries, as it seems to run into local minima.

### 6 Open Questions and Outlook

Before we discuss the conclusions that we derive from our results, we formulate some of the open problems that we encountered on our way.

The first problem relates to the locus  $\partial\mathcal{CPR}_5(5) \cap \mathcal{CP}_5^{\circ}$ , i.e., those matrices that have cp-rank 5 but cp<sup>+</sup>-rank 6. Because our evidence is the result of a numerical approximation algorithm that is also known to be fallible (as we will state next), we refrain from calling this a conjecture:

**Problem 6.1** *What are the polynomial inequalities that describe  $\partial\mathcal{CPR}_5(5) \cap \mathcal{CP}_5^{\circ}$ ? Are they linear in the entries of the factor matrices, as we suspect according to Sec. 5.5?*

Corollary 5.8 of [7] implies that the number of nonnegative square factorizations of any matrix in  $\mathcal{CPR}_5(5) \cap \mathcal{CP}_5^{\circ}$  is finite. However, since neither  $\mathcal{CPR}_5(5)$  nor  $\mathcal{CPR}_5(6)$  are convex, answering this question requires a substantially different technique.

The next open problem concerns the numerical ill-posedness of the reconstruction problem. We know by our experiments that the factorization of matrices near the boundary is very chaotic: small changes in the matrix  $A$  can result in very large changes in the entries of the factor matrix.

**Problem 6.2** For any  $\varepsilon > 0$ , do there always exist matrices  $A, \tilde{A} \in \partial\mathcal{CP}_5$  with  $\|A - \tilde{A}\| < \varepsilon$ , but, say  $\|BB^T - \tilde{B}\tilde{B}^T\| > 10^{-3}$  for the factor matrices? In other words, can the numerical ill-posedness get arbitrarily bad or are we safe if we allow for a minimal numerical accuracy?

Notice that if  $\mathcal{M} := \{X \in \mathbb{R}^{5 \times 5} : A = XX^T\}$  denotes the set of all  $5 \times 5$  possibly negative factor matrices of  $A$ , the orthogonal group acts on  $\mathcal{M}$  by right multiplication. Any matrix in  $\partial\mathcal{CP}_5 \cap \mathcal{DN}\mathcal{N}_5^\circ$  has exactly 5 nonnegative factorizations that correspond to one orbit of the restricted action by the group of  $5 \times 5$  permutation matrices. The question above pertains to the notion of how close these orbits get to other parts of the boundary, i.e., can the orbits of some matrices of fixed size get arbitrarily close to the boundary at a point far from an actual factorization?

Finally, we address a problem that has been open for some time (see [2, Section 4.1]) concerning rational factorizations. There are at least two notions for rational factorizations, which are somewhat equivalent, but subtle. Say that a matrix  $A \in \mathcal{CP}_n$  with rational entries admits a rational factorization if there is an integer number  $k$ , rational numbers  $q_1, \dots, q_k$  and rational column vectors  $v_1, \dots, v_k$  such that  $A = \sum_{j=1}^k q_j v_j v_j^T$ . Notice that when writing this expression as a factorization the entries may not be rational, as one has to use the square root of  $q_j$  to distribute it among the two vectors. However this turns out to be equivalent (see [4, Section 4]) to the existence of a rational nonnegative  $n \times m$  matrix  $C$ , such that  $A = CC^T$ . We remark that the minimal possible values for  $k$  and  $m$ , in case they exist, may be different and larger than the cp-rank.

**Problem 6.3** Let  $A \in \partial\mathcal{CP}_5 \cap \mathcal{DN}\mathcal{N}_5^\circ$  have rational entries. Are the square factorizations of  $A$  rational in either sense?

Given the parametrization of the boundary and the uniqueness of  $5 \times 5$  factorizations and the parametrization of the algebraic components of the boundary we could hope to answer this question entirely in the  $5 \times 5$  case. However, this simplification of the problem leads to finding rational points on a variety, which is a priori a difficult question.

We now conclude our paper with a discussion of the implications and the future outlook. The algebraic description of the boundary  $\partial\mathcal{CP}_5$  gives us an explanation for the complicatedness of the boundary: the high degree makes it very hard to access. Nonetheless, the algebra allows to systematically construct exact matrices in the boundary which seems to be a useful fact. It furthermore suggests that a description with the same level of precision for  $n \geq 6$  is likely hopeless.

It is shown in [1] that the nontrivial extreme rays of the cone of  $6 \times 6$  copositive matrices come in 42 types. In other words, instead of having to deal with two loci, an analogous analysis of that cone would involve 42 varieties, most of which are expected to yield varieties of much higher degree than  $V_{Hi}$ . In short, as dimension grows, the complexity of the boundary increases in two directions simultaneously: the number of algebraic components will diverge and each of the resulting varieties will become harder.

Furthermore, we have so far not found a description of the part  $\partial\mathcal{CPR}_5(5) \cap \mathcal{CP}_5^\circ$  and we suspect that a full derivation would come down to figuring out all nonreducible

zero patterns, which would be feasible in principle. But if the number of zeros is smaller than 11, one also needs to find the associated algebraic equations. This is a difficult algebraic problem. The lack of convexity of the involved cones makes the description even harder.

**Acknowledgements** We would like to thank Bernd Sturmfels for suggesting this project and for many interesting conversations. We also thank Sascha Timme and Paul Breiding for help with the computation of the degree of  $V_{Hi}$  using HomotopyContinuation. This project started while both authors were employed at the Max Planck Institute for Mathematics in the Sciences. M.P. was partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Projektnummer 448293816. J.A.S was partially supported by ANID-FONDECYT Iniciación grant #11221076.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Afonin, A., Hildebrand, R., Dickinson, P.J.C.: The extreme rays of the  $6 \times 6$  copositive cone. *J. Glob. Optim.* **79**(1), 153–190 (2021)
2. Berman, Abraham, Dür, Mirjam, Shaked-Monderer, Naomi: Open problems in the theory of completely positive and copositive matrices. *Electron. J. Linear Algebra* **29**, 46–58 (2015)
3. Berman, A., Shaked-Monderer, N.: *Completely Positive Matrices*. World Scientific, Singapore (2003)
4. Berman, Abraham, Shaked-Monderer, Naomi: Completely positive matrices: real, rational, and integral. *Acta Math. Vietnam.* **43**(4), 629–639 (2018)
5. Blekherman, Grigoriy: Nonnegative polynomials and sums of squares. *J. Am. Math. Soc.* **25**(3), 617–635 (2012)
6. Blekherman, G., Hauenstein, J., Ottem, J.C., Ranestad, K., Sturmfels, B.: Algebraic boundaries of Hilbert's SOS cones. *Compos. Math.* **148**(6), 1717–1735 (2012)
7. Bomze, I.M., Dickinson, P.J.C., Still, G.: The structure of completely positive matrices according to their CP-rank and CP-plus-rank. *Linear Algebra Appl.* **482**, 191–206 (2015)
8. Bomze, I.M., Schachinger, W., Ullrich, R.: From seven to eleven: completely positive matrices with high cp-rank. *Linear Algebra Appl.* **459**(Complete), 208–221 (2014)
9. Boumal, N., Mishra, B., Absil, P.-A., Sepulchre, R.: Manopt, a Matlab toolbox for optimization on manifolds. *J. Mach. Learn. Res.* **15**(42), 1455–1459 (2014)
10. Breiding, Paul, Rose, K., Timme, S.: Certifying zeros of polynomial systems using interval arithmetic. *ACM Trans. Math. Softw.* **49**(1), 1–14 (2021)
11. Breiding, P., Timme, S.: HomotopyContinuation.jl: A Package for Homotopy Continuation in Julia. In: *International Congress on Mathematical Software*, pp. 458–465. Springer, New York (2018)
12. Chen, Justin, Kileel, Joe: Numerical implicitization. *J. Softw. Algebra Geom.* **9**(1), 55–63 (2019)
13. Dickinson, P.J.C.: An improved characterisation of the interior of the completely positive cone. *Electron. J. Linear Algebra* **20**, 723–729 (2010)
14. Ding, C., He, X., Simon, H.D.: On the equivalence of nonnegative matrix factorization and spectral clustering, pp. 606–610
15. Groetzner, Patrick, Dür, Mirjam: A factorization method for completely positive matrices. *Linear Algebra Appl.* **591**, 1–24 (2020)
16. Groetzner, P.H.: *A Method for Completely Positive and Nonnegative Matrix Factorization*. doctoralthesis, Universität Trier (2018)



17. Hildebrand, Roland: The extreme rays of the  $5 \times 5$  copositive cone. *Linear Algebra Appl.* **437**(7), 1538–1547 (2012)
18. Krone, Robert, Kubjas, Kaie: Uniqueness of nonnegative matrix factorizations by rigidity theory. *SIAM J. Matrix Anal. Appl.* **42**(1), 31 (2021)
19. Maxfield, J.E., Minc, H.: On the matrix equation  $X'X = A$ . *Proc. Edinb. Math. Soc.* **13**, 125–129 (1962/1963)
20. Murty, Katta G., Kabadi, Santosh N.: Some NP-complete problems in quadratic and nonlinear programming. *Math. Program.* **39**(2), 117–129 (1987)
21. Nocedal, Jorge, Wright, Stephen J.: *Numerical Optimization*, 2nd edn. Springer, New York (2006)
22. Shaked-Monderer, Naomi, Bomze, Immanuel M., Jarre, Florian, Schachinger, Werner: On the cp-rank and minimal cp factorizations of a completely positive matrix. *SIAM J. Matrix Anal. Appl.* **34**(2), 355–368 (2013)
23. Sponsel, Julia, Bundfuss, Stefan, Dür, Mirjam: An improved algorithm to test copositivity. *J. Glob. Optim.* **52**(3), 537–551 (2012)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.