APPROXIMATE TENSOR DECOMPOSITIONS: DISAPPEARANCE OF MANY SEPARATIONS

GEMMA DE LAS CUEVAS, ANDREAS KLINGLER, AND TIM NETZER

ABSTRACT. It is well-known that tensor decompositions show separations, that is, that constraints on local terms (such as positivity) may entail an arbitrarily high cost in their representation. Here we show that many of these separations disappear in the approximate case. Specifically, for every approximation error ε and norm, we define the approximate rank as the minimum rank of an element in the ε-ball with respect to that norm. For positive semidefinite matrices, we show that the separations between rank, purification rank, and separable rank disappear for a large class of Schatten p-norms. For nonnegative tensors, we show that the separations between rank, positive semidefinite rank, and nonnegative rank disappear for all ℓ_p-norms with $p > 1$. For the trace norm ($p = 1$), we obtain upper bounds that depend on the ambient dimension. We also provide a deterministic algorithm to obtain the approximate decomposition attaining our bounds. Our main tool is an approximate version of Carathéodory's Theorem. Our results imply that many separations are not robust under small perturbations of the tensor, with implications in quantum many-body systems and communication complexity.

CONTENTS

INSTITUTE FOR THEORETICAL PHYSICS, TECHNIKERSTR. 21A, A-6020 INNSBRUCK, AUSTRIA

INSTITUTE FOR THEORETICAL PHYSICS, TECHNIKERSTR. 21A, A-6020 INNSBRUCK, AUSTRIA

DEPARTMENT OF MATHEMATICS, TECHNIKERSTR. 13, A-6020 INNSBRUCK, AUSTRIA

1. Introduction

The rank of a matrix is a central notion in mathematics, physics and other sciences. It is defined as the minimum number of rank-1 matrices (i.e. matrices of the form vw^t where v, w are column vectors) needed to decompose a matrix as a sum thereof. This notion has been extended to include constraints. For example, for nonnegative matrices (i.e. matrices with nonnegative entries), the *nonnegative rank* is the minimum number of nonnegative rank-1 matrices, and the *positive semidefinite rank* is the smallest integer r such that the entries of the matrix can be written as $M_{ij} = \text{Tr}(E_i F_j^t)$ where E_i and F_j are positive semidefinite matrices of size $r \times r$. Similarly, for positive semidefinite matrices, generalisations of these two ranks give rise to the separable rank and the purification rank [4]. For tensors (i.e. multilinear maps), there is no unique way of defining 'a' rank or a rank with constraints — instead, there are many meaningful ways of doing so. The various decompositions and their corresponding ranks can be described by a (weighted) simplicial complex whose vertices are associated to the indices of the tensor [3]. These decompositions of matrices or tensors with constraints have applications in many areas, including classical and quantum information theory [7, 13], quantum many body systems [18], electrical engineering, data analysis (see [1] and references therein) and machine learning (see [9] and references therein).

Now, it is known that there are *separations* between many of these ranks, that is, that one can be arbitrarily more costly than the other. For example, there is a separation between the rank and the positive semidefinite rank, and between the latter and the nonnegative rank [7, 10]. (See also [20] for an upper bound of the approximate positive semidefinite rank with respect to the entrywise maximum-norm). This implies that the random communication complexity can be arbitrarily larger than the quantum communication complexity [13]. Similar separations hold for the operator Schmidt rank, the purification rank and the separable rank [5, 4, 3]. This implies that there cannot exist a local transformation from the operator Schmidt decomposition to the local purification form [5, 4, 3]. (See also [14] for a recent upper bound of the approximate purification rank with respect to the trace norm).

In this paper, we show that many of these separations disappear in the approximate case. This means that many of these separations are not robust to small perturbations of the matrix or tensor in question. Our analysis applies to matrices and tensors, possibly with an explicit invariance built-in (see below), and holds for a large class of norms.

More precisely, we work in the framework of (Ω, G)-decompositions, introduced in [3]. Specifically, we consider elements in tensor product spaces — especially, we consider (positive semidefinite matrices in) the tensor product of the space of $d \times d$ complex matrices $\mathcal{M}_d$,

$$\mathcal{M}_d \otimes \cdots \otimes \mathcal{M}_d,$$

and (nonnegative tensors in) the space

$$\mathbb{C}^d \otimes \cdots \otimes \mathbb{C}^d.$$

Obviously, each such element can be expressed as a sum of elementary tensor factors, but the summation indices can be arranged in many different ways. We describe the summation indices as facets of a *weighted simplicial complex* (wsc). In addition, if the element is invariant under a group action G, the decomposition explicitly reflects this invariance, giving rise to a (Ω, G)-decomposition with an associated (Ω, G)-rank.

The *approximate (Ω, G)-rank* is defined as the minimal rank of an element within an ε-ball of the original element. The ball is measured with respect to the Schatten p-norm for elements in $\mathcal{M}_d \otimes \cdots \otimes \mathcal{M}_d$, and the ℓ_p-norm for elements in $\mathbb{C}^d \otimes \cdots \otimes \mathbb{C}^d$. Specifically, for positive semidefinite (psd) matrices, we define the approximate versions of the (Ω, G)-rank, -purification rank, -quantum square root rank, and -separable rank. For nonnegative tensors, we define the approximate versions of the (Ω, G)-rank, -psd rank, -square root rank, and -nonnegative rank.

Our main result is that many separations between exact (Ω, G)-ranks disappear in the approximate case. Specifically, for psd matrices, we prove that the separations between the operator Schmidt rank, the purification rank, and the separable rank disappear in the approximate case for Schatten p-norms with $p \in (1, 4/3] \cup \{2\} \cup [4, \infty)$ (Corollary 25). For nonnegative tensors, we prove that the separation between rank, nonnegative rank, and psd rank disappear in the approximate case for ℓ_p-norms with $p > 1$ (Corollary 29).

Our central tool is an approximate version of Carathéodory's Theorem [12], which allows to upper bound the approximate rank of an element of a convex set in a dimension independent way. We leverage this result to prove dimension independent upper bounds for the approximate (Ω, G)-rank, -purification rank and -separable rank. It follows that none of these approximate ranks can diverge, and thus the separations disappear.

One crucial point is that the approximate Carathéodory Theorem needs to be applied to the convex hull of a bounded set. Separable states are by definition in the convex hull of all product states, and hence the theorem is directly applicable. Yet, more general elements (such as entangled states, or

not normalised positive semidefinite matrices) have to be normalized by a *gauge function* in order to be studied with this theorem. The gauge function essentially says how much the element needs to rescaled in order to be inside the convex hull of normalized elementary tensors. Hence, this disappearance of separations holds up to this rescaling. We also show that for entangled states this gauge function is related to the robustness of entanglement [21].

We also present an algorithm (Algorithm 32) to obtain the approximate decompositions, apply it to several examples, and show how it attains the upper bounds proven in Theorem 5.

This paper is organised as follows. In Section 2 we the present the approximate Carathéodory Theorem for Schatten classes. In Section 3 we present several notions of (exact) (Ω, G)-decompositions and their associated ranks, which we define more rigorously in Appendix A. In Section 4 we define and study the approximate (Ω, G)-ranks. In Section 5 we show the disappearance of separations for several approximate ranks. In Section 6 we present an algorithm to obtain the approximate decomposition. Finally, in Section 7 we present the conclusions and outlook.

2. The approximate Carathéodory Theorem for Schatten classes

In this section we introduce the approximate Carathéodory Theorem for Schatten classes, which is a crucial tool in this paper. Throughout this paper, we denote the set of complex $d \times d$ matrices by $\mathcal{M}_d$. We also denote the unnormalised Schatten p-norm by $\| \cdot \|_p$, i.e. for $A \in \mathcal{M}_d$ we have

$$\|A\|_p := \mathrm{Tr}(|A|^p)^{1/p} = \left(\sum_{i=1}^{d} s_i(A)^p \right)^{1/p},$$

where $|A| := \sqrt{A^*A}$, and A^* denotes the complex conjugate transpose of A. Further, $\{s_i(A)\}_{i=1}^{d}$ denotes the set of singular values of A.

The main result of this section is Theorem 6. We start with a version of the approximate Carathéodory Theorem [12], which holds for uniformly smooth Banach spaces.

Definition 1. *Let X with $\|\cdot\|$ be a Banach space. The* modulus of smoothness $\rho_X : [0, \infty] \to [0, \infty]$ *is given by*

$$\rho_X(t) := \sup\left\{ \frac{1}{2}(\|x + ty\| + \|x - ty\|) - 1 : \|x\| = \|y\| \le 1 \right\} \ \textit{for} \ t \in [0, \infty].$$

A Banach space is called uniformly smooth *if $\rho_X(t) = o(t)$, i.e. $\rho_X(t)/t \to 0$ as $t \to 0$.*

Note that a finite dimensional Banach space X with $\| \cdot \|$ is uniformly smooth if and only if for all $x, y \in X$ with $\|x\| = \|y\| = 1$ the limit

$$\lim_{t \to 0} \frac{\|x + ty\| - \|x\|}{t}$$

exists. In other words, for all $x \neq 0$ all directional derivatives of the norm exist [17]. Intuitively, this implies that the unit ball of uniformly smooth spaces is smooth.

Note that ρ_X is convex, increasing and $\rho_X(0) = 0$. This implies that ρ_X is bijective and we denote the inverse function by ρ_X^{-1}.

Theorem 2 (Approximate Carathéodory [12])**.** *Let S be a bounded set in a uniformly smooth Banach space X equipped with the norm $\|\cdot\|$, and $a \in \mathrm{conv}(S)$. Then there exists a sequence $\{x_i\}_{i=1}^{\infty} \subseteq S$ such that for $a_k = \frac{1}{k}\sum_{i=1}^{k} x_i$ the following inequality holds*

$$\|a - a_k\| \leq \frac{2\exp(2)}{k \cdot \rho_X^{-1}(1/k)} \cdot \mathrm{diam}(S)$$

where $\rho_X(\cdot)$ is the modulus of smoothness.

Note that the upper bound of Theorem 2 is independent of a, and the only dependence on the space X is given through the diameter of S and the inverse of the modulus of smoothness. We will see that for the special case $X = \mathcal{M}_d$ equipped with the Schatten p-norm the upper bound is also dimension independent.

In the following we let $\rho_p(\cdot)$ denote the modulus of smoothness of $\mathcal{M}_d$ with Schatten p-norm. In the following we calculate an upper bound for the expression $1/\rho_p^{-1}(1/k)$ following [12, 16]. A necessary tool for this calculation are *Hanner's inequalities*.

Theorem 3 (Hanner's inequalities for Schatten norms [15])**.** *Let $A, B \in \mathcal{M}_d$. For $4 \leq p < \infty$ the following inequality holds:*

$$\left(\|A\|_p + \|B\|_p\right)^p + \left|\|A\|_p - \|B\|_p\right|^p \geq \|A + B\|_p^p + \|A - B\|_p^p.$$

For $1 \leq p \leq 4/3$ the inequality is reversed and for $p = 2$ equality holds.

Note that Theorem 3 is an extension of the original Hanner's inequalities for ℓ_p-spaces, i.e. spaces $X = \mathbb{C}^d$ equipped with the entrywise p-norm [16]. It is widely believed that Theorem 3 is true for $1 < p < \infty$ as in the ℓ_p case, but it is only proven for the given range of p.

Corollary 4. *The following inequalities hold:*

$$\rho_p(t) \leq \begin{cases} \frac{1}{p} \cdot t^p & \text{if } 1 \leq p \leq 4/3 \\[2em] \frac{p-1}{2} \cdot t^2 & \text{if } p = 2 \text{ or } 4 \leq p < \infty \end{cases}$$

This implies that for $1 < p \leq 4/3$, $p = 2$ and $4 \leq p < \infty$, $\mathcal{M}_d$ equipped with the Schatten p-norm is uniformly smooth. Furthermore, $\mathcal{M}_d$ with the Schatten 1-norm is not uniformly smooth.

Proof. Let $1 \leq p \leq 4/3$. Applying Hanner's inequality for $A = X + tY$ and $B = X - tY$ we obtain

$$\rho_p(t) = \sup\left\{\frac{1}{2}(\|X + tY\|_p + \|X - tY\|_p) - 1 : \|X\|_p = \|Y\|_p = 1\right\}$$

$$\leq (1 + t^p)^{1/p} - 1 \leq \frac{t^p}{p}.$$

where the last inequality can be obtained by a basic calculation.

Let $p = 2$ or $4 \leq p < \infty$. Since $x \mapsto |x|^p$ is convex we have for $a, b \geq 0$

$$(a + b)^p \leq 2^{p-1} \cdot (a^p + b^p).$$

Using this fact and Hanner's inequality for $A := X$ and $B := tY$ we obtain

$$\rho_p(t) = \sup\left\{\frac{1}{2}(\|X + tY\|_p + \|X - tY\|_p) - 1 : \|X\|_p = \|Y\|_p = 1\right\}$$

$$\leq \sup\left\{\left(\frac{\|X + tY\|_p^p + \|X - tY\|_p^p}{2}\right)^{1/p} - 1 : \|X\|_p = \|Y\|_p = 1\right\}$$

$$\leq \left(\frac{(1 + t)^p + |1 - t|^p}{2}\right)^{1/p} - 1 \leq \frac{p - 1}{2} \cdot t^2.$$

To prove that the Schatten 1-class is not uniformly smooth we refer to the ℓ_p-case (see [16] for details). □

Using the fact that the modulus of smoothness is increasing with t, we can formulate the approximate Carathéodory Theorem for several Schatten classes.

Theorem 5 (Approximate Carathéodory for Schatten classes)**.** *Let $\mathcal{S}$ be a bounded set in $\mathcal{M}_d$ equipped with Schatten p-norm, where $1 < p \leq 4/3$, $p = 2$ or $4 \leq p < \infty$, and $A \in \text{conv}(\mathcal{S})$. Then there exists a sequence $\{X_i\}_{i=1}^\infty \subseteq \mathcal{S}$ such that for $A_k = \frac{1}{k}\sum_{i=1}^k X_i$ the following inequalities hold:*

(a) $\|A - A_k\|_p \leq \dfrac{2\exp(2)}{p^{1/p}} \cdot k^{1/p-1} \cdot \text{diam}(\mathcal{S})$ *if $1 < p \leq 4/3$*

(b) $\|A - A_k\|_p \leq \exp(2) \cdot \sqrt{\dfrac{2(p - 1)}{k}} \cdot \text{diam}(\mathcal{S})$ *if $p = 2$ or $4 \leq p < \infty$*

Proof. Using the fact that ρ_p is increasing in its argument, and Theorem 2, we obtain

$$\rho_p^{-1}(1/k) \geq \begin{cases} \sqrt[p]{\dfrac{p}{k}} & \text{if } 1 < p \leq 4/3 \\ \\ \sqrt{\dfrac{2}{k \cdot (p-1)}} & \text{if } p = 2 \text{ or } 4 \leq p < \infty \end{cases}$$

This proves the statement. □

By fixing some approximation error $\varepsilon > 0$ and p, the previous result can be equivalently formulated as an upper bound on the number of summands k necessary to attain an ε-approximation with respect to the Schatten p-norm.

Theorem 6. *Let $\mathcal{S} \subseteq \mathcal{M}_d$ bounded, $A \in conv(\mathcal{S})$ and $\varepsilon > 0$ be given. Then in the ε-ball with respect to the Schatten p-norm around A there is a point B which is a convex combination of at most*

(a) $\left\lceil C_p \cdot \left(\dfrac{diam(\mathcal{S})}{\varepsilon} \right)^{\frac{p}{p-1}} \right\rceil$ *if $1 < p \leq 4/3$*

(b) $\left\lceil D_p \cdot \left(\dfrac{diam(\mathcal{S})}{\varepsilon} \right)^{2} \right\rceil$ *if $p = 2$ or $4 \leq p < \infty$*

points from S, where

$$C_p := \left(\frac{2\exp(2)}{p^{1/p}} \right)^{\frac{p}{p-1}} \quad and \quad D_p := 2(p-1) \cdot \exp(4).$$

Proof. Follows directly from Theorem 5. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

Note that in both cases (a) and (b) of Theorem 6 the upper bound is dimension independent. The upper bound is best for $p = 2$, and diverges for fixed $\varepsilon > 0$ if we approach $p = 1$. The bounds also diverge in the limit $\varepsilon \to 0$ for arbitrary p. Hence, the approximation error needs to be fixed in order to apply this result.

Remark 7. If we assume that $\mathcal{M}_d$ is equipped with the ℓ_p-norm $\| \cdot \|_{\ell_p}$, i.e. for $A \in \mathcal{M}_d$ we have

$$\|A\|_{\ell_p} := \left(\sum_{i,j=1}^{d} |A_{ij}|^p \right)^{1/p},$$

then the upper bounds of Theorem 5 and Theorem 6 hold for $1 < p \leq 2$ and $p \geq 2$, instead of $1 < p \leq 4/3$ and $4 \leq p < \infty$, respectively. This is due to the fact that Hanner's inequalities are proven for $1 < p < \infty$ in ℓ_p-spaces [16]. $\qquad\qquad\triangle$

3. Notions of (Ω, G)-decompositions and ranks

In this section we present the relevant notions of tensor decompositions based on weighted simplicial complexes introduced in [3]. We will give a brief overview of the framework of (Ω, G)-*decompositions* and their corresponding (Ω, G)-*ranks* (Section 3.1), as well as their application to psd matrices (Section 3.2) and nonnegative tensors (Section 3.3). The technical parts of this section are not essential for the main results of this paper, which are presented in Section 4. Only the definitions of exact (Ω, G)-ranks will be necessary for the subsequent definitions of approximate (Ω, G)-ranks.

3.1. **Exact decompositions.** The goal of this section is to explain the notions of weighted simplicial complexes (wsc) Ω and group actions G on Ω to finally define an (Ω, G)-decomposition (Definition 10). In order to motivate them, we will first give an intuitive explanation of these concepts and then illustrate them with Example 8. For a rigorous definition of wsc and group action we refer to Appendix A and [3].

For each index $i \in [n] := \{0, \ldots, n\}$, we fix a $\mathbb{C}$-vector space $\mathcal{V}_i$ (called the *local vector space*). We denote the *global vector space* as the tensor product space

$$\mathcal{V} := \mathcal{V}_0 \otimes \cdots \otimes \mathcal{V}_n.$$

By definition every $v \in \mathcal{V}$ can be expressed as a sum of elementary tensors

$$v^{[0]} \otimes \cdots \otimes v^{[n]}.$$

The different ways of arranging the summation indices will be reflected in the wsc — specifically, the summation indices will be associated to the facets of the wsc. A weighted simplicial complex is intuitively a 'well-formed' multi-hypergraph. A multi-hypergraph is generalization of a graph where connections of possibly more than two vertices are allowed, thus giving rise to a hypergraph. In addition, every facet can be repeated, i.e. there can be multiple copies thereof — this gives rise to a *multi*-hypergraph.

Furthermore, if v is invariant with respect to permutations of the indices $[n]$ of elementary tensors, we are interested in obtaining explicitly invariant decompositions, i.e. decompositions whose elementary tensors themselves are invariant under the permutation of the indices $[n]$. Thus, for a given group action G on the set $[n]$, we consider the induced linear group action on $\mathcal{V}$, i.e.

$$g : v^{[0]} \otimes \cdots \otimes v^{[n]} \mapsto v^{[g0]} \otimes \cdots \otimes v^{[gn]}$$

for $g \in G$. An element $v \in \mathcal{V}$ is called *G-invariant* if it is invariant under the action of G on $\mathcal{V}$. The subspace of invariant elements is denoted $\mathcal{V}_{\mathrm{inv}}$.

Example 8.   (i) Consider a decomposition of the form

$$v = \sum_{\alpha=1}^{r} v_\alpha^{[0]} \otimes \cdots \otimes v_\alpha^{[n]}.$$

The minimal number r of elementary tensors is called the *tensor rank* of v. Setting $\mathcal{I} = \{1, \ldots, r\}$ and $\mathcal{F} = \{\{0, \ldots, n\}\}$ we can equivalently write the decomposition as a sum over all functions $\alpha : \mathcal{F} \to \mathcal{I}$, denoted $\alpha \in \mathcal{I}^{\mathcal{F}}$, i.e.

(1) $$v = \sum_{\alpha \in \mathcal{I}^{\mathcal{F}}} v_{\alpha_{|0}}^{[0]} \otimes \cdots \otimes v_{\alpha_{|n}}^{[n]},$$

where $\alpha_{|i}$ denotes the restriction of α to the set $\mathcal{F}_i = \{F \in \mathcal{F} : i \in F\}$. Intuitively, every element $F \in \mathcal{F}$ corresponds to a summation index α_i, and the elements of F correspond to the positions where α_i appears.

(ii) Let G be a transitive group action on $[n]$, i.e. there is only one orbit, and hence $Gi = [n]$ for all $i \in [n]$. For elements v which are invariant under the group action G, we consider the G-invariant decomposition

$$v = \sum_{\alpha=1}^{r} v_\alpha \otimes \cdots \otimes v_\alpha$$

which is called the *symmetric tensor decomposition*. The smallest number r among all possible decompositions is called the *symmetric tensor rank* (see Example 9 (i) for an explicit example with $n = 2$).

(iii) Consider a decomposition of the form

$$v = \sum_{\alpha_0,\ldots,\alpha_{n-1}=1}^{r} v_{\alpha_0}^{[0]} \otimes v_{\alpha_0,\alpha_1}^{[1]} \otimes \cdots \otimes v_{\alpha_{n-2},\alpha_{n-1}}^{[n-1]} \otimes v_{\alpha_{n-1}}^{[n]}.$$

This decomposition is called the *matrix product operator form* and the minimal r among all possible decompositions the *operator Schmidt rank*. This corresponds to the decomposition of Equation (1) with

$$\mathcal{F} = \{\{0,1\},\{1,2\},\ldots,\{n-1,n\}\}$$

and $\mathcal{I} = \{1,\ldots,r\}$.

(iv) Consider a decomposition of the form

$$v = \sum_{\alpha_0,\ldots,\alpha_{n-1}=1}^{r} v_{\alpha_0,\alpha_1}^{[0]} \otimes v_{\alpha_1,\alpha_2}^{[1]} \otimes \cdots \otimes v_{\alpha_{n-2},\alpha_{n-1}}^{[n-2]} \otimes v_{\alpha_{n-1},\alpha_0}^{[n-1]}.$$

This is similar to (iii), but with periodic boundary conditions. This corresponds to choosing

$$\mathcal{F} = \{\{i,i+1\} : i \in \{0,\ldots,n-2\}\} \cup \{\{n-1,0\}\}$$

in Equation (1).

Additionally considering the symmetry operation given by the cyclic group $G = C_n$ (i.e. the group generated by a mapping $i \mapsto i+1$, where addition is modulo n) we obtain the *translational invariant matrix product operator form* [4]

$$v = \sum_{\alpha_0,\ldots,\alpha_{n-1}=1}^{r} v_{\alpha_0,\alpha_1} \otimes v_{\alpha_1,\alpha_2} \otimes \cdots \otimes v_{\alpha_{n-2},\alpha_{n-1}} \otimes v_{\alpha_{n-1},\alpha_0}.$$

$\triangle$

Example 8 shows a unified framework for various decompositions. In the following, we give examples of simplicial complexes, and relate their facets to the sets $\mathcal{F}$ constructed in Example 8. Since facets can appear multiple times we will denote the multiset of all facets by $\widetilde{\mathcal{F}}$.

Example 9. (i) The simplicial complex (sc) Σ_n which contains all possible connections between vertices is called the *n-simplex*. For $n = 2$ this can be depicted as follows:
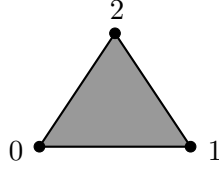
Note that the gray area shows the facet which connects all three vertices. Σ_n has only one (multi-)facet, i.e. $\mathcal{F} = \widetilde{\mathcal{F}} = \{[n]\}$ and hence gives rise to a tensor decomposition shown in (i) or (ii) of Example 8.

(ii) For $n \geq 1$, the *line of length* n (i.e. composed of $n+1$ points) is the sc Λ_n corresponding to the following graph:
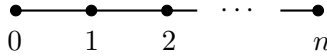
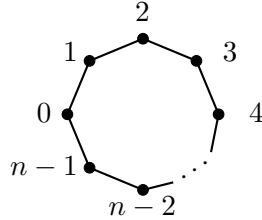The set $\mathcal{F} = \widetilde{\mathcal{F}}$ has n elements and generates a matrix product operator form given in (iii) of Example 8. Intuitively the sc describes the connections between the local vector spaces through shared summation indices.

(iii) For $n \geq 3$, the *circle of length* n is the sc Θ_n corresponding to the following graph:

It has n facets and generates the decomposition given in Example 8 (iv).
$$\triangle$$

A group action of G on the wsc Ω is a natural extension of the group action introduced at the beginning of Section 3.1. It is defined on the power set of $[n]$ and maps subsets

$$[n] \supseteq \{i_1, i_2, \ldots, i_k\} \mapsto \{gi_1, gi_2, \ldots, gi_k\}.$$

Additionally it respects the structure of the wsc, i.e. facets in $\widetilde{\mathcal{F}}$ are mapped to facets in $\widetilde{\mathcal{F}}$. For a rigorous definition of this notion we refer to Appendix A and [3]; for examples we refer to Example 8. Throughout this paper, we assume that every group action G is a valid group action on Ω.

Furthermore, we call a group action G on Ω *free* if for all $F \in \widetilde{\mathcal{F}}$ the only element $g \in G$ mapping F to itself is the neutral element.

Finally, we define the notion of (Ω, G)-decomposition and the (Ω, G)-rank. For two sets X, Y, the set Y^X contains, by definition, all functions $f : X \to Y$. For any set $\mathcal{I}$, any $\alpha \in \mathcal{I}^{\widetilde{\mathcal{F}}}$ and any $i \in \mathcal{I}$, we call the restriction to $\widetilde{\mathcal{F}}_i$

$$\alpha_{|_{\widetilde{\mathcal{F}}_i}} \in \mathcal{I}^{\widetilde{\mathcal{F}}_i}$$

the *restriction of α to vertex i*, and write $\alpha_{|i}$ instead.

Definition 10. *For $v \in \mathcal{V}$, an (Ω, G)-decomposition is given by a finite index set $\mathcal{I}$ and families*

$$V^{[i]} := \left(v_\beta^{[i]} \right)_{\beta \in \mathcal{I}^{\widetilde{\mathcal{F}}_i}},$$

where $v_\beta^{[i]} \in \mathcal{V}_i$ for $i \in [n]$ such that:

(a) We have

$$v = \sum_{\alpha \in \mathcal{I}^{\widetilde{\mathcal{F}}}} v_{\alpha_{|0}}^{[0]} \otimes v_{\alpha_{|1}}^{[1]} \otimes \cdots \otimes v_{\alpha_{|n}}^{[n]}. \tag{2}$$

(b) For all $i \in [n]$, $g \in G$ and $\beta \in \mathcal{I}^{\widetilde{\mathcal{F}}_i}$ it holds

$$v_\beta^{[i]} = v_{g\beta}^{[gi]},$$

where $^g\beta \in \mathcal{I}^{\widetilde{\mathcal{F}}_{gi}}$ and $^g\beta(F) := \beta(g^{-1}F)$ for $F \in \widetilde{\mathcal{F}}_{gi}$.

The smallest cardinality of $\mathcal{I}$ among all possible (Ω, G)-decompositions of v is called the (Ω, G)-rank of v, denoted

$$\mathrm{rank}_{(\Omega, G)}(v).$$

If no (Ω, G)-decomposition of v exists, we set $\mathrm{rank}_{(\Omega, G)}(v) := \infty$.

Note that if G is a free group action on a connected wsc Ω, and $v \in \mathcal{V}_{\mathrm{inv}}$, there always exists an (Ω, G)-decomposition of v [3, Thm. 13].

For simplicity, if G is the trivial group, we will call an (Ω, G)-decomposition just an Ω-decomposition, and write rank_Ω for the rank. The same simplification will also be used for all ranks defined in the following two subsections.

3.2. **Exact ranks for psd matrices.** Separability (or its negation, entanglement), and purifications are central notions in quantum information theory. In the next two definitions we will formulate these notions in the framework of (Ω, G)-decompositions.

Throughout this section we will fix a connected wsc Ω together with a free group action G. We will also assume that the local vector space is given by

$$\mathcal{V}_i := \mathcal{M}_{d_i}$$

and hence

$$\mathcal{V} := \mathcal{M}_{d_0} \otimes \cdots \otimes \mathcal{M}_{d_n} \cong \mathcal{M}_{d_0 \cdots d_n},$$

whose hermitian part is $\mathrm{Her}_{d_0} \otimes \cdots \otimes \mathrm{Her}_{d_n} \cong \mathrm{Her}_{d_0 \cdots d_n}$. The d_i can be chosen differently as long as there are no further restrictions given by the group action of G on $\mathcal{V}$; that is, whenever i, j are in the same orbit of the group action of G on $[n]$, then $d_i = d_j$. Further we will define the *cone of (complex) psd matrices* as $\mathcal{M}_{d_0 \cdots d_n}^+$. If $\rho \in \mathcal{M}_{d_0 \cdots d_n}^+$ fulfills $\mathrm{Tr}(\rho) = 1$, we call it a *state*.

Let us now define (Ω, G)-purifications and (Ω, G)-square root decompositions.

Definition 11. *Let $\rho \in \mathcal{M}_{d_0 \cdots d_n}^+$.*

(i) An (Ω, G)-purification is an element

$$\sigma \in \mathcal{M}_{d_0', d_0} \otimes \cdots \otimes \mathcal{M}_{d_n', d_n}$$

with

$$\rho = \sigma^* \sigma \quad \text{and} \quad \text{rank}_{(\Omega, G)}(\sigma) < \infty,$$

where $\mathcal{M}_{d_i, d_i'}$ denotes the space of all complex $d_i' \times d_i$ matrices and * the adjoint. The smallest (Ω, G)-rank among all (Ω, G)-purifications is called (Ω, G)-purification rank of ρ, denoted

$$\text{puri-rank}_{(\Omega, G)}(\rho).$$

(ii) $\sigma \in Her_{d_0} \otimes \cdots \otimes Her_{d_n}$ is called square root *of ρ if $\sigma^2 = \rho$. The smallest (Ω, G)-rank among all square roots of ρ is called (Ω, G)-quantum square root rank of ρ, denoted*

$$\text{q-sqrt-rank}_{(\Omega, G)}(\rho).$$

Remark 12. (i) Note that every square root is a purification and hence

$$\text{puri-rank}_{(\Omega, G)}(\rho) \leq \text{q-sqrt-rank}_{(\Omega, G)}(\rho).$$

(ii) Note that not every matrix σ which fulfills $\sigma^2 = \rho$ is automatically hermitian. From the spectral decomposition, $\rho = UDU^*$ with the diagonal matrix $D = \text{diag}(\lambda_1, \lambda_2, \ldots)$, we see that all hermitian square roots are of the form [11]

$$\sigma = UD^{1/2}U^*, \quad D^{1/2} = \text{diag}\left(\pm\sqrt{\lambda_1}, \pm\sqrt{\lambda_2}, \ldots\right).$$

(iii) If G is a free group action on a connected wsc Ω and $\rho \in \mathcal{V}_{\text{inv}}$ and is psd, there always exists an (Ω, G)-purification of ρ and a square root with finite (Ω, G)-rank [3, Thm. 27]. $\triangle$

The next step is the definition of the separable (Ω, G)-rank. We call the matrix $\rho \in \mathcal{M}_{d_0 \cdots d_n}^+$ *separable* if it admits a decomposition

$$\rho = \sum_j \rho_j^{[0]} \otimes \cdots \otimes \rho_j^{[n]}$$

where $\rho_j^{[i]} \in \mathcal{M}_{d_i}^+$. If additionally $\text{Tr}(\rho) = 1$, we call ρ a *separable state*. From now on, we will denote the set of separable states

$$\text{SEP}_{d_0, d_1, \ldots, d_n} := \{\rho \in \mathcal{M}_{d_0 \cdots d_n}^+ : \rho \text{ separable state}\}.$$

If $d_i = d$ for all $i \in [n]$, we will write for simplicity $\text{SEP}_{n,d} := \text{SEP}_{d_0, \ldots, d_n}$.

Definition 13. *A separable (Ω, G)-decomposition of $\rho \in \mathcal{M}_{d_0 \cdots d_n}$ is given by an (Ω, G)-decomposition*

$$\rho = \sum_{\alpha \in \mathcal{I}^{\tilde{\mathcal{F}}}} \rho_{\alpha_{|0}}^{[0]} \otimes \rho_{\alpha_{|1}}^{[1]} \otimes \cdots \otimes \rho_{\alpha_{|n}}^{[n]}$$

in which $\rho_\beta^{[i]} \in \mathcal{M}_{d_i}^+$ *for* $\beta \in \mathcal{I}^{\widetilde{\mathcal{F}}_i}$ *and* $i \in [n]$. *The smallest cardinality of an index set* $\mathcal{I}$ *among all possible separable* (Ω, G)-*decompositions of* ρ *is called the* separable (Ω, G)-rank of ρ, *denoted*

$$\text{sep-rank}_{(\Omega, G)}(\rho).$$

If there exists no separable (Ω, G)-*decomposition of* v, *we set sep-rank*$_{(\Omega, G)}(v)$ *to* ∞.

Note that if G is a free group action on a connected wsc Ω and $\rho \in \mathcal{V}_{\text{inv}}$ is separable, there always exists a separable (Ω, G)-decomposition of ρ [3, Thm. 21].

3.3. **Exact ranks for nonnegative tensors.** In the following we will consider the set of nonnegative tensors, and will define different notions of (Ω, G)-ranks based on [3, Sec. 5].

For simplicity we consider the local space $\mathcal{V}_i = \mathbb{C}^d$ and define the global space $\mathcal{V} := \mathcal{K}_{n,d}$, where

$$\mathcal{K}_{n,d} := \bigotimes_{i=0}^{n} \mathbb{C}^d.$$

If n and d are clear from the context, we write $\mathcal{K}$ instead of $\mathcal{K}_{n,d}$. Any element $M \in \mathcal{K}$ can be uniquely written as

$$M = \sum_{i_0,\ldots,i_n} m_{i_0,\ldots,i_n} e_{i_0} \otimes \cdots \otimes e_{i_n}$$

where e_j denotes the j-th standard basis vector in the corresponding vector space $\mathbb{C}^d$. M is said to be *nonnegative* if $m_{i_0,\ldots,i_n} \geq 0$ for all $i_0, \ldots, i_n$.

Recall that every tensor $M \in \mathcal{K}_{n,d}$ can be associated with a diagonal matrix $\sigma \in \mathcal{M}_d \otimes \cdots \otimes \mathcal{M}_d \cong \mathcal{M}_{d^{n+1}}$ by setting

$$(3) \qquad \sigma = \sum_{i_0\ldots,i_n} m_{i_0,\ldots,i_n} E_{i_0 i_0} \otimes \cdots \otimes E_{i_n i_n},$$

where E_{jk} is the matrix which has value 1 on position (j, k) and 0 elsewhere. Obviously σ is psd if and only if M is nonnegative.

We now give a brief description of several (Ω, G)-decompositions of nonnegative tensors and their corresponding ranks. For a more detailed discussion we refer to [3].

Definition 14. *(i) A nonnegative* (Ω, G)-*decomposition of* $M \in \mathcal{K}_{n,d}$ *is an* (Ω, G)-*decomposition as in Equation* (2) *where all* $v_{\alpha_{|i}}^{[i]} \in \mathbb{C}^d$ *have nonnegative entries. The corresponding rank is called the* nonnegative (Ω, G)-rank of M, *denoted*

$$\text{nn-rank}_{(\Omega, G)}(M).$$

(ii) A positive semidefinite (Ω, G)-*decomposition of* $M \in \mathcal{K}_{n,d}$ *consists of psd matrices*

$$E_j^{[i]} \in \mathcal{M}_{k_i}^+$$

where $k_i = \left|\mathcal{I}^{\widetilde{\mathcal{F}}_i}\right|$ for $i \in [n]$ and $j \in \{1, \dots, d\}$, such that

$$\left(E_j^{[i]}\right)_{\beta,\beta'} = \left(E_j^{[gi]}\right)_{{}^g\beta, {}^g\beta'}$$

for all i, g, j, β, β', where ${}^g\beta(F) := \beta(g^{-1}F)$ for $F \in \widetilde{\mathcal{F}}_{gi}$, and

$$m_{i_0,\dots,i_n} = \sum_{\alpha,\alpha' \in \mathcal{I}^{\widetilde{\mathcal{F}}}} \left(E_{i_0}^{[0]}\right)_{\alpha_{|0},\alpha'_{|0}} \cdots \left(E_{i_n}^{[n]}\right)_{\alpha_{|n},\alpha'_{|n}}$$

for all $i_0, \dots, i_n$. The smallest cardinality of an index set $\mathcal{I}$ among all possible positive semidefinite (Ω, G)-decompositions is called the positive semidefinite (Ω, G)-rank of M, denoted

$$\text{psd-rank}_{(\Omega,G)}(M).$$

(iii) $N \in \mathcal{K}_{n,d}$ with $n_{i_0,\dots,i_n} \in \mathbb{R}$ is called a square root of M, if $M = N \circ N$, where $\circ$ denotes the Hadamard product (i.e. entrywise multiplication, $m_{i_0,\dots,i_n} = n_{i_0,\dots,i_n}^2$). The smallest (Ω, G)-rank among all square roots of M is called (Ω, G)-square root rank of M, denoted

$$\text{sqrt-rank}_{(\Omega,G)}(M).$$

Remark 15. For the psd matrix σ and the nonnegative tensor M of Equation (3) the following relations hold (see [3, Thm. 43] for details):

(a) $\text{rank}_{(\Omega,G)}(M) = \text{rank}_{(\Omega,G)}(\sigma)$
(b) $\text{nn-rank}_{(\Omega,G)}(M) = \text{sep-rank}_{(\Omega,G)}(\sigma)$
(c) $\text{psd-rank}_{(\Omega,G)}(M) = \text{puri-rank}_{(\Omega,G)}(\sigma)$
(d) $\text{sqrt-rank}_{(\Omega,G)}(M) = \text{q-sqrt-rank}_{(\Omega,G)}(\sigma)$.

$\triangle$

4. Approximate (Ω, G)-decompositions and ranks

In this section we will define the notions of approximate (Ω, G)-decompositions for both psd matrices and nonnegative tensors, and will apply the results from Section 2 to obtain upper bounds for ranks of approximate (Ω, G)-decompositions.

The section is organized as follows. In Section 4.1 and Section 4.2 we define the approximate analogues of the (Ω, G)-ranks for psd matrices and nonnegative tensors, respectively. In Section 4.3 we introduce gauge functions, a relevant tool to obtain the upper bounds. Subsequently, we show upper bounds for general matrices (Section 4.4), psd matrices (Section 4.5) and separable states (Section 4.6).

As before we consider

$$\mathcal{M}_{d_0} \otimes \mathcal{M}_{d_1} \otimes \cdots \otimes \mathcal{M}_{d_n} \cong \mathcal{M}_{d_0 \cdots d_n}.$$

We further fix a connected wsc Ω and a free group action G on Ω.

4.1. **Approximate decompositions of psd matrices.** We now introduce the different notions of approximate (Ω, G)-ranks in the space $\mathcal{M}_{d_0} \otimes \cdots \otimes \mathcal{M}_{d_n}$. Generally speaking, given a matrix ρ, the approximate rank is the minimal rank of all matrices contained in the ε-ball of ρ with respect to the Schatten p-norm $\|\cdot\|_p$. Note that this is different to the notion of *border rank*, which is the minimial $k \in \mathbb{N}$ among all possible sequences $(\rho_n)_{n \in \mathbb{N}}$ consisting of rank-k matrices ρ_n converging to ρ [2].

Definition 16. *Let $p \in [1, \infty)$ and $\varepsilon > 0$. Further let $M \in \mathcal{M}_{d_0} \otimes \cdots \otimes \mathcal{M}_{d_n}$ and $\rho \in \mathcal{M}_{d_0 \cdots d_n}^+$. We define*

$$rank_{(\Omega,G)}^{\varepsilon,p}(M) := \min\{rank_{(\Omega,G)}(N) : \|M - N\|_p \leq \varepsilon, N \in \mathcal{M}_{d_0 \cdots d_n}\},$$

and similarly

$$puri\text{-}rank_{(\Omega,G)}^{\varepsilon,p}(\rho) := \min\{puri\text{-}rank_{(\Omega,G)}(\sigma) : \|\rho - \sigma\|_p \leq \varepsilon, \sigma \in \mathcal{M}_{d_0 \cdots d_n}\},$$

$$q\text{-}sqrt\text{-}rank_{(\Omega,G)}^{\varepsilon,p}(\rho) := \min\{q\text{-}sqrt\text{-}rank_{(\Omega,G)}(\sigma) : \|\rho - \sigma\|_p \leq \varepsilon, \sigma \in \mathcal{M}_{d_0 \cdots d_n}\}$$

and

$$sep\text{-}rank_{(\Omega,G)}^{\varepsilon,p}(\rho) := \min\{sep\text{-}rank_{(\Omega,G)}(\sigma) : \|\rho - \sigma\|_p \leq \varepsilon, \sigma \in \mathcal{M}_{d_0 \cdots d_n}\}.$$

Recall that a non-psd matrix ρ does not admit a purification, and thus its purification rank is ∞. In contrast, $puri\text{-}rank_{(\Omega,G)}^{\varepsilon,p}(\rho)$ might be finite — it is finite if and only if there is a psd matrix in the ε-ball of ρ with respect to the Schatten p-norm. Similar statements hold for the quantum square root rank and the separable rank.

Let us now revisit Example 8 and Example 9, and explain the notions of approximate ranks in the cases therein.

Example 17. Let $M \in \mathcal{M}_{d_0 \cdots d_n}$, $\varepsilon > 0$ and $p \in [1, \infty)$.

(i) An approximate Σ_n-decomposition is given by a matrix $N \in \mathcal{M}_{d_0 \cdots d_n}$ attaining a decomposition

$$N = \sum_{\alpha=1}^r N_\alpha^{[0]} \otimes \cdots \otimes N_\alpha^{[n]}$$

with $\|M - N\|_p \leq \varepsilon$. The approximate Σ_n-rank, $rank_{\Sigma_n}^{\varepsilon,p}(\rho)$, is called *approximate tensor rank* of M and is the smallest integer r among all matrices N in the ε-ball of M.

(ii) Consider the line $\Omega = \Lambda_n$ of length n. The *approximate operator Schmidt rank*, $rank_{\Lambda_n}^{\varepsilon,p}(M)$, is the minimal integer r among all $N \in \mathcal{M}_{d_0 \cdots d_n}$ with $\|M - N\|_p \leq \varepsilon$ and all decompositions of matrix product operator form.

(iii) For $n \geq 3$ consider the circle $\Omega = \Theta_n$ of length n together with the cyclic group $G = C_n$ whose elements are translations of the points on the line. Further let $M \in \mathcal{M}_{d^n}$ be C_n-invariant. In this example C_n-invariance corresponds to translational invariance of M. The *approximate translational invariant operator Schmidt rank* of M, $rank_{(\Theta_n, C_n)}^{\varepsilon,p}(M)$, is the minimal integer

r among all $N \in \mathcal{M}_{d^n}$ with $\|M - N\|_p \leq \varepsilon$ and all decompositions of N of translational invariant matrix product operator form. $\triangle$

4.2. Approximate decompositions of nonnegative tensors. In the following we define the notions of approximate (Ω, G)-ranks similar to Section 4.1 using the exact (Ω, G)-decompositions defined in Section 3.3.

Motivated by the correspondence between psd matrices and nonnegative tensors given in Equation (3), we will use for $p \geq 1$ the ℓ_p-norm, which is defined for $M \in \mathcal{K}_{n,d}$ as

$$\|M\|_{\ell_p} = \left(\sum_{i_0,\dots,i_n} |m_{i_0,\dots,i_n}|^p \right)^{1/p}.$$

Since in the following definition of the approximate (Ω, G)-decompositions of nonnegative tensors the relevant norm is the ℓ_p-norm, we will indicate this fact by using ℓ_p instead of p.

Definition 18. *Let $p \in [1, \infty)$, $\varepsilon > 0$ and $M \in \mathcal{K}$. We define*

(i) $rank_{(\Omega,G)}^{\varepsilon,\ell_p}(M) := \min\{rank_{(\Omega,G)}(N) : \|M - N\|_{\ell_p} \leq \varepsilon, N \in \mathcal{K}\}$

(ii) $nn\text{-}rank_{(\Omega,G)}^{\varepsilon,\ell_p}(M) := \min\{nn\text{-}rank_{(\Omega,G)}(N) : \|M - N\|_{\ell_p} \leq \varepsilon, N \in \mathcal{K}\}$

(iii) $psd\text{-}rank_{(\Omega,G)}^{\varepsilon,\ell_p}(M) := \min\{psd\text{-}rank_{(\Omega,G)}(N) : \|M - N\|_{\ell_p} \leq \varepsilon, N \in \mathcal{K}\}$

(iv) $sqrt\text{-}rank_{(\Omega,G)}^{\varepsilon,\ell_p}(M) := \min\{sqrt\text{-}rank_{(\Omega,G)}(N) : \|M - N\|_{\ell_p} \leq \varepsilon, N \in \mathcal{K}\}$

Recall that the nonnegative decomposition, psd decomposition and square root decomposition only exist if M is nonnegative, and thus the corresponding ranks are ∞ if M is not nonnegative. In contrast, the approximate ranks might be finite, even if M is not nonnegative. More precisely, they are finite if and only if there exists a nonnegative tensor in the ε-ball of M with respect to the ℓ_p-norm. In Example 31 we will illustrate the behavior of these ranks.

4.3. More norms for matrices. For the (non-scaled) Schatten norms, where $1 \leq p \leq q \leq \infty$, we have the following inequalities [24]:

(4) $\|M\|_q \leq \|M\|_p \leq \text{rank}(M)^{\frac{1}{p} - \frac{1}{q}} \|M\|_q.$

We denote the space of all complex hermitian $d \times d$ matrices by $\text{Her}_d \subseteq \mathcal{M}_d$. For any $1 \leq p \leq \infty$ we define

$$P_p := \{\pm \rho^{[0]} \otimes \cdots \otimes \rho^{[n]} | \text{ all } \rho^{[i]} \in \mathcal{M}_{d_i}^+ \text{ and } \| \otimes_{i=0}^n \rho^{[i]} \|_p \leq 1\}$$

and consider

$$B_p := \text{conv}(P_p) \subseteq \text{Her}_{d_0 \cdots d_n}.$$

Note that for $p \leq q$ we have $B_p \subseteq B_q$ and already B_1 contains all separable states. Each B_p is compact, convex, centrally symmetric and contains the

origin in its interior. We can thus understand it as the unit ball of a norm: For a set S in a real vector space V, the *gauge function* μ_S is defined by

$$(5) \qquad \mu_S(v) := \inf\left\{\lambda > 0 \mid \frac{1}{\lambda}v \in S\right\}$$

for $v \in V$. If S is compact, convex, centrally symmetric and has nonempty interior, the gauge function μ_S is in fact a norm (see for example Theorem 15.2 in [19]), and S is clearly its unit ball.

We denote the gauge function of B_p by μ_p. We now relate these gauge functions to a multipartite version of the robustness of entanglement [21].

Definition 19. *Let $\rho \in \mathcal{M}_{d_0} \otimes \cdots \otimes \mathcal{M}_{d_n}$ be a state. The* robustness of entanglement *of ρ is defined as*

$$R(\rho) := \inf\{\lambda \geq 1 \mid \rho = (1-\lambda)\rho_1 + \lambda\rho_2, \rho_i \text{ separable states}\}.$$

Note that this definition differs from the original robustness of entanglement by the addition of a constant 1.

Proposition 20. *For $\rho \in \mathrm{Her}_{d_0} \otimes \cdots \otimes \mathrm{Her}_{d_n} \cong \mathrm{Her}_{d_0 \cdots d_n}$ and $1 \leq p \leq q \leq \infty$ we have:*

(i) $\|\rho\|_p \leq \mu_p(\rho)$.
(ii) $\mu_q(\rho) \leq \mu_p(\rho) \leq (d_0 \cdots d_n)^{1/p-1/q}\mu_q(\rho)$.
(iii) If ρ is a separable state, then $\mu_p(\rho) \leq 1$.
(iv) If ρ is a state, then $R(\rho) \leq \mu_1(\rho) \leq 2R(\rho)$.
(v) Define $\mu_{\sqrt{},p}(v) = \min\{\mu_p(\sqrt{v})\}$. If $v^2 = v$ then $\mu_{\sqrt{},p}(v) \leq \mu_p(v)$.

Proof. Since the unit ball of the p-norm is convex, it contains B_p. Statement (i) follows directly from this. (ii) is a direct consequence of the corresponding inequalities for the p-norms. (iii) is clear from $\rho \in B_1$. For the first inequality in (iv) we express

$$\frac{1}{\mu_1(\rho)}\rho = \sum_i \lambda_i \xi_i$$

as a convex combination of elements $\xi_i \in P_1$, with all $\lambda_i > 0, \|\xi_i\| > 0$. From the minimality of $\mu_1(\rho)$ it then follows that $\|\xi_i\|_1 = 1$ holds for all i. Sorting the positive and negative terms we obtain

$$\frac{1}{\mu_1(\rho)}\rho = r\sigma_1 - (1-r)\sigma_2$$

with separable states σ_i, where $0 \leq r \leq 1$ is the sum over those λ_i with ξ_i psd. Taking the trace on both sides shows $2\mu_1(\rho)r = 1 + \mu_1(\rho)$. Thus $\lambda := \mu_1(\rho)r$ yields $\rho = (1-\lambda)\sigma_2 + \lambda\sigma_1$ and thus $R(\rho) \leq \mu_1(\rho)$.

For the second inequality we express $\rho = (1-\lambda)\rho_1 + \lambda\rho_2$ where ρ_i are separable states and $\lambda \geq 1$. Since $\mu_1(\rho_i) \leq 1$ the second inequality follows from the triangle inequality for μ_1.

(v) Immediate from the definition. $\qquad\qquad\square$

4.4. **Upper bounds for approximate ranks.** We now prove an upper bound for the approximate (Ω, G)-rank which only depends on the gauge function value of the matrix, and the approximation error ε. Recall that Ω is a connected wsc and G a free group action on Ω.

Theorem 21. *Let $1 < p \leq 4/3$, $p = 2$ or $4 \leq p < \infty$ and $\varepsilon > 0$. Assume $M \in \mathrm{Her}_{d_0 \cdots d_n}$ is G-invariant. Then*

(a) $rank_{(\Omega,G)}^{\varepsilon,p}(M) \leq \left\lceil C_p \cdot \left(\dfrac{2\mu_p(M)}{\varepsilon} \right)^{\frac{p}{p-1}} \right\rceil \cdot |G|$ *if $1 < p \leq 4/3$*

(b) $rank_{(\Omega,G)}^{\varepsilon,p}(M) \leq \left\lceil D_p \cdot \left(\dfrac{2\mu_p(M)}{\varepsilon} \right)^{2} \right\rceil \cdot |G|$ *if $p = 2$ or $4 \leq p < \infty$*

where C_p, D_p are constants defined in Theorem 6.

The proof is a straightforward application of Theorem 6 with set P_p to construct an Ω-decomposition. To obtain a G-invariant decomposition, we apply the construction used in [3, Thm. 13].

Proof. First let $p = 2$ or $4 \leq p < \infty$. We prove the statement for the case that $\mu_p(M) = 1$; the general case follows by replacing M with $M/\mu_p(M)$. First we consider the case of trivial group action. By assumption $M \in \mathrm{conv}(P_p)$, and $\mathrm{diam}(P_p) = 2$. By Theorem 6 we find an M' such that $\|M - M'\| \leq \varepsilon$ and M' is a convex combination of at most $\lceil 4D_p/\varepsilon^2 \rceil$ elements of P_p. This implies that

$$\mathrm{rank}_\Omega(M') \leq \mathrm{rank}_{\Sigma_n}(M') \leq \lceil 4D_p/\varepsilon^2 \rceil,$$

where the first inequality follows from [3, Prop. 36]. We now involve a group action and apply the construction of [3, Thm. 13] to M'. This gives rise to an (Ω, G)-decomposition of an element $M'' = \frac{1}{|G|} \sum_g gM'$ with

$$\mathrm{rank}_{(\Omega,G)}(M'') \leq \lceil 4D_p/\varepsilon^2 \rceil |G|.$$

It remains to prove that M'' is contained in the ε-neighborhood of M:

$$
\begin{aligned}
\|M'' - M\|_p &= \left\| \frac{1}{|G|} \sum_g (gM' - gM) \right\|_p \\
&\leq \frac{1}{|G|} \sum_g \|gM' - gM\|_p \leq \varepsilon
\end{aligned}
$$

where we used the invariance $M = \frac{1}{|G|} \sum_g gM$, that the norm is unitarily invariant $\|gM' - gM\|_p = \|M' - M\|_p$, and that M' is in the ε-ball around M. The case $1 < p \leq 4/3$ is analogous. $\square$

Remark 22. Note that Theorem 21 and all following theorems state upper bounds for convex combinations of the type

$$M_k = \frac{1}{k} \sum_{i=1}^{k} X_i,$$

where $X_i \in P_p$. The sequence $\{X_i\}_{i=1}^k$ might contain a repetition of elements which would decrease the rank of the decomposition even more, but our estimates do not exploit this fact. $\triangle$

4.5. **Upper bounds for approximate ranks of psd matrices.** We now provide upper bounds for the purification rank and the quantum square root rank of psd matrices. More specifically, we provide an upper bound for the quantum square root rank, which is itself an upper bound of the purification rank (see Remark 12 (i)). Recall that Ω is a connected wsc and G a free group action on Ω.

Corollary 23. *Let* $1 < p \leq 4/3$, $p = 2$ *or* $4 \leq p < \infty$ *and* $\varepsilon > 0$. *Let* $\rho \in \mathcal{M}^+_{d_0 \cdots d_n}$ *be* G-invariant. *Then*

(a) $\text{puri-rank}^{\varepsilon,p}_{(\Omega,G)}(\rho) \leq \left\lceil C_p \cdot \left(\frac{2}{\sqrt{1+\varepsilon/\mu^2_{\sqrt{},p}(\rho)} - 1} \right)^{\frac{p}{p-1}} \right\rceil \cdot |G|$ *if* $1 < p \leq 4/3$

(b) $\text{puri-rank}^{\varepsilon,p}_{(\Omega,G)}(\rho) \leq \left\lceil D_p \cdot \left(\frac{2}{\sqrt{1+\varepsilon/\mu^2_{\sqrt{},p}(\rho)} - 1} \right)^2 \right\rceil \cdot |G|$ *if* $\begin{matrix} p = 2 \ or \\ 4 \leq p < \infty \end{matrix}$

and the same upper bounds hold for $\text{q-sqrt-rank}^{\varepsilon,p}_{(\Omega,G)}$.

Proof. We apply Theorem 21 to the square root of ρ (called M) which realises $\mu_{\sqrt{},p}(\rho)$. We have that $M/\mu_{\sqrt{},p}(\rho) \in \text{conv}(P_p)$ and thus $\|M\|_p \leq \mu_{\sqrt{},p}(\rho)$ (defined in Proposition 20 (v)). We now choose an element M' which is $\mu_{\sqrt{},p}(\rho) \cdot \delta$ - close to M with $\delta = \sqrt{1 + \varepsilon/\mu^2_{\sqrt{},p}(\rho)} - 1$. We define $\rho' = M'^2$ and compute

$$\begin{aligned} \|\rho - \rho'\|_p &= \|M^2 - M'^2\|_p \\ &\leq \|M - M'\|_p \cdot (2\|M\|_p + \|M - M'\|_p) \\ &\leq \mu_{\sqrt{},p}(\rho)^2 \cdot \delta \cdot (2 + \delta) \leq \varepsilon. \end{aligned}$$

Using that

$$\text{puri-rank}_{(\Omega,G)}(\rho') \leq \text{q-sqrt-rank}_{(\Omega,G)}(\rho') \leq \text{rank}_{(\Omega,G)}(M')$$

and the fact that $\text{rank}_{(\Omega,G)}(M')$ is upper bounded by Theorem 21 we obtain the result. $\square$

Note that this result in fact upper bounds the quantum square root rank, which may be arbitrarily larger than the purification rank [4].

4.6. **Upper bounds for approximate ranks of separable states.** An upper bound for separable states can be obtained without the use of a gauge function. This is possible because we can directly conclude that a separable state is in the convex hull of

$$\left\{ +\rho^{[0]} \otimes \cdots \otimes \rho^{[n]} : \rho^{[i]} \in \mathcal{M}^+_{d_i} \text{ and } \| \otimes_{i=0}^n \rho^{[i]}\|_p \leq 1 \right\} \subseteq P_p.$$

Using that sep-rank$_{(\Omega,G)}$ upper bounds rank$_{(\Omega,G)}$ and puri-rank$_{(\Omega,G)}$, we can conclude that this upper bound holds for the other two ranks in the approximate case too. Recall that Ω is a connected wsc and G a free group action on Ω.

Proposition 24. *Let $1 < p \leq 4/3$, $p = 2$ or $4 \leq p < \infty$ and $\varepsilon > 0$. Further let $\rho \in \text{SEP}_{d_0,\ldots,d_n}$ be a G-invariant separable state. Then*

(a) *sep-rank$_{(\Omega,G)}^{\varepsilon,p}(\rho) \leq \lceil C_p \cdot (2/\varepsilon)^{\frac{p}{p-1}} \rceil \cdot |G|$ if $1 < p \leq 4/3$*

(b) *sep-rank$_{(\Omega,G)}^{\varepsilon,p}(\rho) \leq \lceil D_p \cdot (2/\varepsilon)^2 \rceil \cdot |G|$ if $p = 2$ or $4 \leq p < \infty$*

The same upper bound holds for rank$_{(\Omega,G)}^{\varepsilon,p}(\rho)$ and puri-rank$_{(\Omega,G)}^{\varepsilon,p}(\rho)$, too.

Proof. It is proven exactly as Theorem 21 by using that separable states are a convex combination of product states, which are a subset of P_p. The rest follows from [3, Thm. 43] and [3, Cor. 44].   □

Note again that the separable rank may be arbitrarily larger than the rank and the purification rank [4].

5. Disappearance of separations in the approximate case

We now turn to study relations between the different notions of ranks, especially their (lack of) separations in the approximate case.

Let $\mathcal{X}$ be an arbitrary set and $f, g \colon \mathcal{X} \to \mathbb{N}$ two functions. We say that there is a *separation* between f and g, and write $f \ll g$, if there exists a sequence in $\mathcal{X}$ along which f is bounded but g is not. This implies that the values of g cannot be upper bounded by a function that only depends on f. In the exact case, there are several examples of separations between different ranks — for example on the set

$$\mathcal{X} := \bigcup_{d \in \mathbb{N}} \mathcal{M}_{d^{n+1}}^+$$

one has rank$_{(\Omega,G)} \ll$ puri-rank$_{(\Omega,G)}$, for suitable choices of (Ω, G) [7, 4, 3]. In the following we will show that many separations of ranks of psd matrices disappear (Section 5.1), and the same happens for nonnegative tensors (Section 5.2). This is an immediate consequence of the fact the approximate ranks admit an upper bound which is independent of the dimension of the ambient space.

As in previous section, we fix a connected wsc Ω and a free group action G acting on Ω.

5.1. **Positive semidefinite matrices.** In the following we show that the separation between several ranks of psd matrices vanishes in the approximate case. The strategy is simple: we will use Theorem 21, Proposition 24 and Corollary 23 to upper bound the ranks independently of the matrix dimension. Since all ranks are bounded functions, it follows that many separations vanish.

For simplicity we assume that $d_i = d$ and hence consider the space

$$\mathcal{M}_d \otimes \cdots \otimes \mathcal{M}_d \cong \mathcal{M}_{d^{n+1}}.$$

The result can be extended in a straightforward manner to the case that the parts of the tensor product have different dimensions.

Corollary 25. *Let G be a free group action on Ω. Let $\varepsilon > 0$, $K \in \mathbb{N}$ and $1 < p \leq 4/3$, $p = 2$ or $4 \leq p < \infty$. We define the set*

$$\mathcal{X}_K := \{\rho \in \mathcal{M}_{d^{n+1}}^+ : d \in \mathbb{N} \text{ and } \mu_{\sqrt{},p}(\rho) \leq K\}.$$

Then the following holds on $\mathcal{X}_K$:

$$
\begin{array}{llll}
(i) & rank_{(\Omega,G)}^{\varepsilon,p} & \not\ll & puri\text{-}rank_{(\Omega,G)}^{\varepsilon,p} \\[2mm]
(ii) & puri\text{-}rank_{(\Omega,G)}^{\varepsilon,p} & \not\ll & q\text{-}sqrt\text{-}rank_{(\Omega,G)}^{\varepsilon,p}
\end{array}
$$

Further, we define the set

$$\mathcal{X}_{sep} := \bigcup_{d \in \mathbb{N}} \mathrm{SEP}_{n,d}.$$

The following holds on $\mathcal{X}_{sep}$:

$$
\begin{array}{llll}
(iii) & rank_{(\Omega,G)}^{\varepsilon,p} & \not\ll & sep\text{-}rank_{(\Omega,G)}^{\varepsilon,p} \\[2mm]
(iv) & puri\text{-}rank_{(\Omega,G)}^{\varepsilon,p} & \not\ll & sep\text{-}rank_{(\Omega,G)}^{\varepsilon,p}
\end{array}
$$

Note that (i) and (ii) imply that $\mathrm{rank}_{(\Omega,G)}^{\varepsilon,p} \not\ll \text{q-sqrt-rank}_{(\Omega,G)}^{\varepsilon,p}$ on $\mathcal{X}_K$.

Proof. (i)-(ii) For $\rho \in \mathcal{M}_{d^{n+1}}^+$ we have $\mathrm{rank}_{(\Omega,G)}(\rho) \leq \text{puri-rank}_{(\Omega,G)}(\rho)^2$ [3, Prop. 29]. By the basic properties of q-sqrt-rank$_{(\Omega,G)}$ (Remark 12 (i)) we obtain

$$\sqrt{\mathrm{rank}_{(\Omega,G)}^{\varepsilon,p}(\rho)} \leq \text{puri-rank}_{(\Omega,G)}^{\varepsilon,p}(\rho) \leq \text{q-sqrt-rank}_{(\Omega,G)}^{\varepsilon,p}(\rho).$$

Since Corollary 23 upper bounds q-sqrt-rank$_{(\Omega,G)}^{\varepsilon,p}$ by a constant which is independent of the dimension of $\rho \in \mathcal{X}_K$, all ranks mentioned above are upper bounded on $\mathcal{X}_K$.

(iii)-(iv) Let $\rho \in \mathcal{X}_{\mathrm{sep}}$. Using again [3, Prop. 29] we have that

$$
\begin{aligned}
\mathrm{rank}_{(\Omega,G)}^{\varepsilon,p}(\rho) &\leq \text{sep-rank}_{(\Omega,G)}^{\varepsilon,p}(\rho), \\
\text{puri-rank}_{(\Omega,G)}^{\varepsilon,p}(\rho) &\leq \text{sep-rank}_{(\Omega,G)}^{\varepsilon,p}(\rho).
\end{aligned}
$$

By Proposition 24 sep-rank$_{(\Omega,G)}$ is upper bounded by a constant which is again independent of the dimension of ρ. $\qquad\square$

Note that the previous result is not making any statement about the Schatten 1-norm, since the approximate Carathéodory Theorem is not applicable for this norm. Using the results from Theorem 21, Proposition 24 and Equation (4), we now give an upper bound of the approximate rank and

approximate separable rank in Schatten 1-norm, which is however dimension dependent.

Corollary 26. *Let $M \in \mathrm{Her}_{d^{n+1}}$ and $\rho \in \mathrm{SEP}_{n,d}$ both be G-invariant. Then:*

(i) $\mathrm{rank}^{\varepsilon,1}_{(\Omega,G)}(M) \leq d^{n+1} \cdot \mathrm{rank}^{\varepsilon,2}_{(\Omega,G)}(M) \leq \left\lceil D_2 \cdot \left(\frac{2\mu_2(M)}{\varepsilon} \right)^2 \right\rceil \cdot |G| \cdot d^{n+1}$

(ii) $\mathrm{sep\text{-}rank}^{\varepsilon,1}_{(\Omega,G)}(\rho) \leq d^{n+1} \cdot \mathrm{sep\text{-}rank}^{\varepsilon,2}_{(\Omega,G)}(\rho) \leq \lceil D_2 \cdot (2/\varepsilon)^2 \rceil \cdot |G| \cdot d^{n+1}$

Proof. Let $A, A_k \in \mathcal{M}_{d^{n+1}}, \mathcal{S} \subseteq \mathcal{M}_{d^{n+1}}$ be chosen as in Theorem 5. Using Equation (4) and optimizing over all valid p-values gives the bound

$$\|A - A_k\|_1 \leq \sqrt{d^{n+1}} \cdot \|A - A_k\|_2,$$

or equivalently

$$\|A - A_k\|_1 \leq \varepsilon \quad \text{if} \quad k \geq d^{n+1} \cdot \left\lceil D_2 \cdot \left(\frac{\mathrm{diam}(\mathcal{S})}{\varepsilon} \right)^2 \right\rceil.$$

Applying this statement to the proofs of Theorem 21 and Proposition 24 shows the statement. $\qquad\qquad\square$

Example 27. We now apply Corollary 26 to the running examples of Example 8 and Example 17.

(i) Consider the line $\Omega = \Lambda_n$ of size n and G the trivial group. For the approximate operator Schmidt rank in Schatten 1-norm of $\rho \in \mathcal{M}^+_{d^{n+1}}$ it holds that

$$\mathrm{rank}^{\varepsilon,1}_{\Lambda_n}(\rho) \leq \left\lceil D_2 \cdot \left(\frac{2\mu_2(\rho)}{\varepsilon} \right)^2 \right\rceil \cdot d^{n+1}.$$

If $\rho \in \mathrm{SEP}_{n,d}$, then

$$\mathrm{sep\text{-}rank}^{\varepsilon,1}_{\Lambda_n}(\rho) \leq \lceil D_2 \cdot (2/\varepsilon)^2 \rceil \cdot d^{n+1}.$$

Note that, in the exact case, $\mathrm{rank}_{\Lambda_n}$ is also bounded by a constant times d^{n+1} [4, Prop. 49]. In contrast, the upper bound of $\mathrm{sep\text{-}rank}_{\Lambda_n}$ in the exact case is linear in $d^{2(n+1)}$. Hence, for fixed ε and sufficiently large dimension d this yields a better upper bound in the approximate case.

(ii) Consider the circle of n elements, $\Omega = \Theta_n$, together with the cyclic group C_n. Further let $\rho \in \mathcal{M}^+_{d^n}$ be C_n-invariant. Since $|C_n| = n$ we obtain

$$\mathrm{rank}^{\varepsilon,1}_{(\Theta_n,C_n)}(\rho) \leq \left\lceil D_2 \cdot \left(\frac{2\mu_2(\rho)}{\varepsilon} \right)^2 \right\rceil \cdot n \cdot d^n$$

and

$$\mathrm{sep\text{-}rank}^{\varepsilon,1}_{(\Theta_n,C_n)}(\rho) \leq \lceil D_2 \cdot (2/\varepsilon)^2 \rceil \cdot n \cdot d^n.$$

$$\triangle$$

5.2. **Nonnegative tensors.** In the following we use the correspondence between nonnegative tensors and diagonal psd matrices given in Equation (3) to show that the separation between nn-rank$_{(\Omega,G)}$, rank$_{(\Omega,G)}$ and psd-rank$_{(\Omega,G)}$ of nonnegative tensors vanishes in the approximate case (Corollary 29).

As in Section 3.3, we consider the space $\mathcal{K}_{n,d} := \bigotimes_{i=0}^{n} \mathbb{C}^d$ equipped with the ℓ_p-norm. In contrast to psd matrices, which are equipped with the Schatten p-norm, the disappearance of separations for nonnegative tensors equipped with the ℓ_p-norm can be proven for all $1 < p < \infty$. Recall that Ω is a connected wsc and G a free group action.

We start with a preparatory lemma, which is an immediate consequence of [3, Prop. 29] and the relations given in Remark 15.

Lemma 28. *Let $M \in \mathcal{K}_{n,d}$ and $p \in [1, \infty)$. Then the following holds:*

(i) $\quad rank_{(\Omega,G)}^{\varepsilon,\ell_p}(M) \quad\quad \leq \quad nn\text{-}rank_{(\Omega,G)}^{\varepsilon,\ell_p}(M)$

(ii) $\quad psd\text{-}rank_{(\Omega,G)}^{\varepsilon,\ell_p}(M) \quad \leq \quad nn\text{-}rank_{(\Omega,G)}^{\varepsilon,\ell_p}(M)$

(iii) $\quad rank_{(\Omega,G)}^{\varepsilon,\ell_p}(M) \quad\quad \leq \quad psd\text{-}rank_{(\Omega,G)}^{\varepsilon,\ell_p}(M)^2$

(iv) $\quad psd\text{-}rank_{(\Omega,G)}^{\varepsilon,\ell_p}(M) \quad \leq \quad sqrt\text{-}rank_{(\Omega,G)}^{\varepsilon,\ell_p}(M)$

Note that it is not clear if a similar result to Remark 15 holds for approximate ranks. This is due to the fact that the ε-ball of diagonal matrices also contains non-diagonal matrices with a possibly smaller rank than the diagonal ones.

Corollary 29. *Let G be a free group action on Ω. Let $\varepsilon > 0$, $K \in \mathbb{N}$ and $1 < p < \infty$. We define the set*

$$\mathcal{Y}_K := \{M \in \mathcal{K}_{n,d} : d \in \mathbb{N}, M \text{ nonnegative and } \|M\|_{\ell_p} \leq K\}.$$

Then the following holds on $\mathcal{Y}_K$:

(i) $\quad rank_{(\Omega,G)}^{\varepsilon,\ell_p} \quad\quad \not\ll \quad psd\text{-}rank_{(\Omega,G)}^{\varepsilon,\ell_p}$

(ii) $\quad psd\text{-}rank_{(\Omega,G)}^{\varepsilon,\ell_p} \quad \not\ll \quad nn\text{-}rank_{(\Omega,G)}^{\varepsilon,\ell_p}$

Further, we define the set

$$\mathcal{Y}_{\sqrt{},K} := \{M \in \mathcal{K}_{n,d} : d \in \mathbb{N}, M \text{ nonnegative and } \mu_{\sqrt{},p}(\sigma) \leq K\}$$

where σ denotes the corresponding diagonal matrix of M. Then the following holds on $\mathcal{Y}_{\sqrt{},K}$:

(iii) $\quad psd\text{-}rank_{(\Omega,G)}^{\varepsilon,\ell_p} \quad \not\ll \quad sqrt\text{-}rank_{(\Omega,G)}^{\varepsilon,\ell_p}$

(iv) $\quad rank_{(\Omega,G)}^{\varepsilon,\ell_p} \quad\quad \not\ll \quad sqrt\text{-}rank_{(\Omega,G)}^{\varepsilon,\ell_p}$

Note that (i) and (ii) imply that rank$_{(\Omega,G)}^{\varepsilon,\ell_p} \not\ll$ nn-rank$_{(\Omega,G)}^{\varepsilon,\ell_p}$ on $\mathcal{Y}_K$.

Proof. (i)-(ii) Let $M \in \mathcal{Y}_K$. By Proposition 28, we have that $\mathrm{rank}_{(\Omega,G)}^{\varepsilon,\ell_p}(M)$ and $\mathrm{psd\text{-}rank}_{(\Omega,G)}^{\varepsilon,\ell_p}(M)$ are upper bounded by $\mathrm{nn\text{-}rank}_{(\Omega,G)}^{\varepsilon,\ell_p}(M)$. To obtain an upper bound of $\mathrm{nn\text{-}rank}_{(\Omega,G)}^{\varepsilon,\ell_p}(M)$ we restrict to $p = 2$, as the other cases are analogous. We set $M' = M/\|M\|_{\ell_2}$ and $\varepsilon' = \varepsilon/K$. Then the corresponding diagonal matrix σ' is separable with $\|\sigma'\|_2 \leq 1$. Note that any ε'-approximation of M' immediately leads to an ε-approximation of M. Using

$$P_2' := \left\{ \rho_0 \otimes \cdots \otimes \rho_n : \rho_i \in \mathcal{M}_d^+ \text{ diagonal and } \| \bigotimes_{i=0}^{n} \rho_i \|_2 \leq 1 \right\} \subseteq P_2$$

instead of P_2 in the proof of Theorem 21, we obtain that

$$\mathrm{nn\text{-}rank}_{(\Omega,G)}^{\varepsilon,\ell_2}(M) \leq \lceil D_2 \cdot (2K/\varepsilon)^2 \rceil \cdot |G|$$

which is independent of the ambient dimension.

Since the Schatten p-norm on the space of diagonal matrices is equivalent to the ℓ_p-norm, we can apply Theorem 6 and Theorem 21 for $1 < p < \infty$ (see Remark 7 for details).

(iii)-(iv) Let $M \in \mathcal{Y}_{\sqrt{,K}}$. Again, by Proposition 28, $\mathrm{psd\text{-}rank}_{(\Omega,G)}^{\varepsilon,p}(M)$ is upper bounded by $\mathrm{sqrt\text{-}rank}_{(\Omega,G)}^{\varepsilon,p}(M)$. Applying Corollary 23 to the corresponding diagonal matrix σ gives an upper bound for $\mathrm{q\text{-}sqrt\text{-}rank}_{(\Omega,G)}^{\varepsilon,p}(\sigma)$. This is again independent of the ambient dimension. $\square$

Remark 30. Similarly to psd matrices (Corollary 26) we can upper bound the approximate ranks for nonnegative tensors in the ℓ_1-norm. The correspondence between diagonal psd matrices and nonnegative matrices yields for a nonnegative and G-invariant $M \in \mathcal{K}_{n,d}$ the inequality

$$\mathrm{rank}_{(\Omega,G)}^{\varepsilon,\ell_1}(M) \leq d^{n+1} \cdot \mathrm{rank}_{(\Omega,G)}^{\varepsilon,\ell_2}(M).$$

This is the same d-dependence as the one of the exact Carathéodory Theorem [23]. Similar results appear for all other ranks of nonnegative tensors. $\triangle$

Example 31. We now give some concrete examples of decompositions whose separations between ranks disappear in the approximate case. From now on we consider the space $\mathcal{K}_{1,d} = \mathbb{C}^d \otimes \mathbb{C}^d \cong \mathcal{M}_d$ and Λ_1-decompositions, i.e. decompositions of the form

$$M = \sum_{\alpha=1}^{r} v_\alpha^{[0]} \otimes v_\alpha^{[1]},$$

where $v_\alpha^{[i]} \in \mathbb{C}^d$.

Recall that for M nonnegative, $\mathrm{nn\text{-}rank}_{\Lambda_1}(M)$ is the smallest integer r such that there exists a decomposition of M with $v_\alpha^{[i]} \in \mathbb{R}_+^d$ for all $\alpha = 1, \ldots, r$ (i.e. vectors with nonnegative entries). Furthermore, $\mathrm{psd\text{-}rank}_{\Lambda_1}(M)$ is the

smallest integer r such that there exist $A_j^{[i]} \in \mathcal{M}_r^+$ for $j = 1, \ldots, d$ and $i = 0, 1$ which generate a decomposition

$$M_{kl} = \sum_{\alpha_0, \alpha_1 = 1}^{r} \left(A_k^{[0]}\right)_{\alpha_0, \alpha_1} \cdot \left(A_l^{[1]}\right)_{\alpha_0, \alpha_1} = \mathrm{tr}\left(\left(A_k^{[0]}\right) \cdot \left(A_l^{[1]}\right)^t\right)$$

Assume $1 < p \leq 4/3$, $p = 2$ or $4 \leq p < \infty$ and $\varepsilon > 0$ fixed. The separations are based on examples mentioned in [7].

(i) Consider the normalized Euclidean distance matrix

$$U_d = M_d / \|M_d\|_{\ell_p} \in \mathcal{M}_d,$$

where M_d is defined as

$$M_d := ((i - j)^2)_{i,j=1}^d.$$

It was shown that $\mathrm{sqrt\text{-}rank}_{\Lambda_1}(M_d) = \mathrm{psd\text{-}rank}_{\Lambda_1}(M_d) = 2$ [7] and

$$\mathrm{nn\text{-}rank}_{\Lambda_1}(M_d) \geq \log_2(d).$$

Obviously the same statement is true for U_d. Hence, we obtain

$$\mathrm{sqrt\text{-}rank}_{\Lambda_1} \ll \mathrm{nn\text{-}rank}_{\Lambda_1} \quad \text{and} \quad \mathrm{psd\text{-}rank}_{\Lambda_1} \ll \mathrm{nn\text{-}rank}_{\Lambda_1}.$$

Since $\|U_d\|_{\ell_p} = 1$, Corollary 29 shows that the separations for this example vanish, since $\mathrm{nn\text{-}rank}_{\Lambda_1}^{\varepsilon, \ell_p}(U_d)$ can be upper bounded independently of d.

(ii) Let S_d be the slack matrix of a d-gon in the plane. We define the normalized slack matrix as $V_d := S_d / \|S_d\|_{\ell_p}$. It was shown that $\mathrm{rank}_{\Lambda_1}(S_d) = 3$ and $\mathrm{psd\text{-}rank}_{\Lambda_1}(S_d)$ diverges if d goes to infinity [10]. Obviously, the same holds for V_d. Hence $\mathrm{rank}_{\Lambda_1} \ll \mathrm{psd\text{-}rank}_{\Lambda_1}$.

Since $\|V_d\|_{\ell_p} = 1$, Corollary 29 shows that the separation for this example vanishes, i.e. $\mathrm{psd\text{-}rank}_{\Lambda_1}^{\varepsilon, \ell_p}(V_d)$ can be upper bounded independently of d. $\quad \triangle$

Note that in these examples the normalization is important, because both $\|S_d\|_{\ell_p}$ and $\|M_d\|_{\ell_p}$ are unbounded as a function of d. Further note that this discussion can be extended to Λ_n-decompositions with $n \geq 1$. One application fulfilling the normalization condition is the interpretation of a tensor $M \in \mathcal{K}_{n,d}$ as a probability mass function $P(X_0, \ldots, X_n)$ over $n + 1$ discrete random variables, taking values in $\{1, \ldots, d\}$ and setting

$$m_{i_0, \ldots, i_n} := P(X_0 = i_0, \ldots, X_n = i_n).$$

In this case M is nonnegative and bounded with $\|M\|_{\ell_p} \leq \|M\|_{\ell_1} = 1$. The nonnegative Λ_n-decomposition corresponds to a *hidden Markov model* [9].

6. Algorithm to find the approximate decomposition

The sequence approximating a matrix in the convex hull in Theorem 5 can be computed by a deterministic algorithm presented in [12] for uniformly smooth Banach spaces. In the following we give an explicit description of this algorithm for Schatten p-classes.

The Schatten p-norm is everywhere differentiable except for 0. Let $X, Y \in \mathcal{M}_d \setminus \{0\}$ be two arbitrary matrices. Further, let

$$X = U \cdot \Sigma \cdot V^*$$

be a singular value decomposition, in particular, $U, V \in \mathcal{M}_{d,r}$ are isometries where $r \leq d$ and

$$\Sigma = \mathrm{diag}(\lambda_1, \ldots, \lambda_r) \in \mathcal{M}_r$$

is a diagonal matrix with positive entries. We also denote the ith column of U and V by u_i and v_i, respectively. Then, the directional derivative of $\|\cdot\|_p$ at X in direction Y can be evaluated (see Ref. [22] for details) by

$$(6) \qquad \mathcal{D}_Y \|\cdot\|_p \Big|_X = \frac{1}{\|X\|_p^{p-1}} \cdot \sum_{i=1}^r \lambda_i^{p-1} \cdot u_i^* \cdot Y \cdot v_i.$$

Since $X \cdot v_i = \lambda_i \cdot u_i$, it further holds that

$$\mathcal{D}_X \|\cdot\|_p \Big|_X = \|X\|_p.$$

This is a necessary property of a functional for the algorithm of [12] to be applicable and leads to the following algorithm.

Algorithm 32. The sequence $\{X_i\}_{i=1}^\infty \subseteq \mathcal{S}$ which generates the approximation $A_k = \frac{1}{k} \sum_{i=1}^k X_i$ of $A \in \mathrm{conv}(\mathcal{S})$ satisfying Theorem 5 can be constructed in the following way:

(i) X_1 is an arbitrary point in $\mathcal{S}$.
(ii) For the constructed sequence $\{X_1, \ldots, X_k\} \subseteq \mathcal{S}$, $k \geq 1$ we choose $X_{k+1} \in \mathcal{S}$ such that for $Y = X_{k+1} - A$ the following holds

$$(7) \qquad \mathcal{D}_Y \|\cdot\|_p \Big|_{A_k - A} \leq 0.$$

$\triangle$

Note that as long as $A \in \mathrm{conv}(\mathcal{S})$, there always exists an $X_{k+1} \in \mathcal{S}$ such that the inequality is true. Moreover, Algorithm 32 makes no further constraints on the choice of X_{k+1}, and hence the upper bound given in Theorem 5 is satisfied for any sequence constructed with this algorithm.

If we want to apply this algorithm to the ℓ_p-norm instead, we only have to replace Equation (6) with

$$(8) \qquad \mathcal{D}_Y \|\cdot\|_{\ell_p} \Big|_X = \frac{1}{\|X\|_{\ell_p}^{p-1}} \cdot \sum_{i,j=1}^d y_{ij} \cdot x_{ij} \cdot |x_{ij}|^{p-2}$$

where x_{ij} and y_{ij} the entries of X and Y respectively at position (i,j).

We now apply the above algorithm to a nonnegative Λ_1-decomposition on the space $\mathcal{K}_{1,d}$. We set

$$\mathcal{S} = \{e_i \otimes e_j : i, j \in \{1, \ldots, d\}\}$$

where e_i the standard basis-vectors in $\mathbb{C}^d$. Assume that $A \in \mathcal{K}_{1,d}$ is non-negative and $\|A\|_{\ell_1} = 1$. Obviously we have that $A \in \mathrm{conv}(\mathcal{S})$ and the corresponding convex combination is a valid nonnegative Λ_1-decomposition of A.

Note that in step (ii) of Algorithm 32 there is in general not a unique choice which satisfies Equation (7). Hence, we present a standard and a greedy method of choice.

Method 1. Define an order on $\mathcal{S}$ and choose the smallest element which satisfies Equation (7).

Method 2. Choose the element in $\mathcal{S}$ which attains the smallest value on the left hand side of Equation (7).
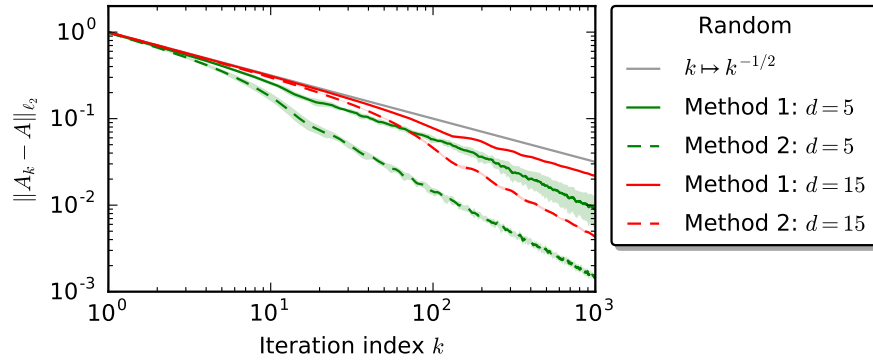

FIGURE 1. Application of Algorithm 32 to random matrices A at two different dimensions $d = 5$ (green) and $d = 15$ (red). The entries are independently uniformly distributed on $[0, 1]$ with the constraint that $\|A\|_{\ell_1} = 1$. The x-axis shows the iteration index and the y-axis shows the error measured in the ℓ_2-norm. Gray shows the function $k \mapsto k^{-1/2}$ as an orientation for the theoretical convergence rate (up to a constant). Method 1 is plotted as a continuous line, Method 2 as dashed line. The sampling size for the random matrices is 20 and the plots show the mean value and the standard deviation.

In the following numerical examples we use for Method 1 the lexicographic ordering
$$(i, j) \preceq (i', j') :\Longleftrightarrow i < i' \text{ or } (i = i' \text{ and } j \leq j').$$
Figure 1 shows the application of the two methods to random matrices with uniformly independently distributed entries normalized to 1. Both methods show a $k^{-1/2}$ convergence for small k and a transition to a faster convergence rate depending on the method. As expected, the Method 2 (the
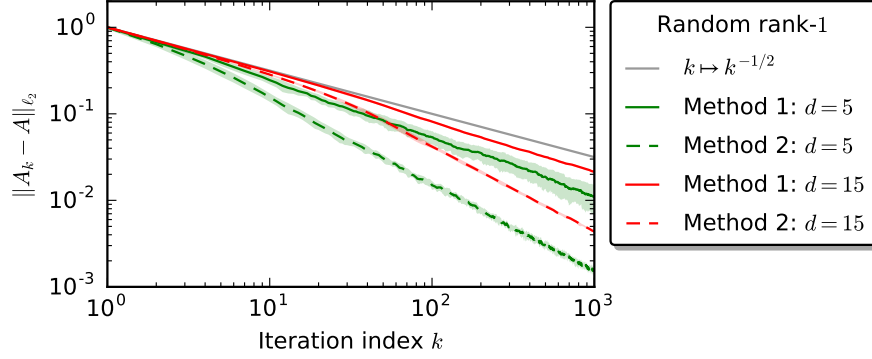
FIGURE 2. For the description of the graph we refer to the caption of Figure 1. $A = a \cdot b^T$ is a nonnegative random rank-1 matrix with $a, b \in [0, 1]^d$ uniformly distributed and A normalized to 1, i.e. $\|A\|_{\ell_1} = 1$. The sampling size is 20 and the plots show the mean value and the standard deviation.
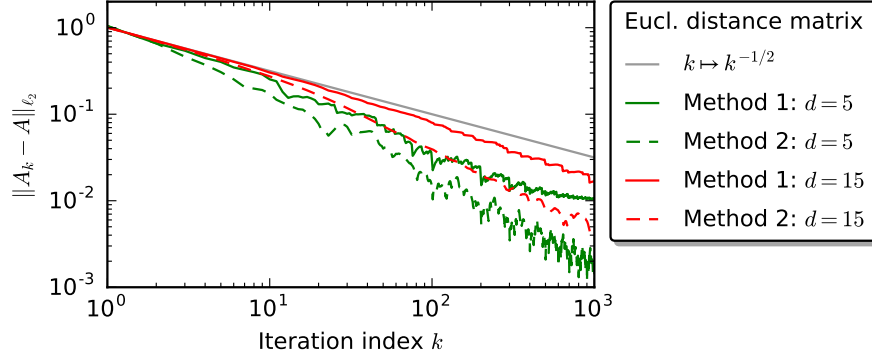

FIGURE 3. For the description of the different lines we refer to the caption of Figure 1. A is the normalized Euclidean distance matrix $M_d/\|M_d\|_{\ell_1}$.

greedy method) converges faster than Method 1. Concerning the choice of the ordering in Method 1 there would be no difference to other orderings in that case since the entries are uniformly independently distributed. The numerical experiments also indicate that the iteration index k where the transition of faster convergence appears grows with increasing dimension d of the matrices.

Figure 2 shows the application of the algorithm to random rank-1 matrices. The results are qualitatively similar to the case of random matrices which have almost surely full rank. This is due to the fact that random rank-1 matrices

are almost surely a linear combination of all d^2 elements of S. Hence the algorithm cannot distinguish between random matrices and random rank-1 matrices.

Since the algorithm cannot distinguish matrices with different ranks, Figure 3, which shows the application to the Euclidean distance matrix M_d normalized to 1, also shows a similar convergence rate in comparison to Figure 1 and Figure 2. Note that the fluctuations of the convergence are a natural consequence of the fact that for every iteration step k the prefactor of the linear combinations $1/k$ are fixed. Since the graphs in Figure 1 and Figure 2 show an average convergence rate the fluctuations do not appear therein.

The Python code `ApproximationAlgorithm.py` of this numerical simulation is available together with this submission.

7. Conclusions and Outlook

In this paper, we have defined and studied various notions of approximate (Ω, G)-ranks. The exact versions thereof were introduced in [3] — the idea is to consider elements in a tensor product of matrix spaces $\mathcal{M}_d$ or of column vectors $\mathbb{C}^d$, and to express them as a sum of elementary tensor factors. The arrangement of indices in the sum is determined by the facets of a weighted simplicial complex Ω, and a group action G specifies the permutations of tensor product indices which leave the element invariant. The approximate (Ω, G)-rank is defined as the minimal rank of an element within an ε-ball of the original element. The ball is measured with respect to the Schatten p-norm and the ℓ_p-norm for elements in $\mathcal{M}_d \otimes \cdots \otimes \mathcal{M}_d$ and $\mathbb{C}^d \otimes \cdots \otimes \mathbb{C}^d$, respectively. Specifically, for psd matrices, we have defined the approximate versions of the (Ω, G)-rank, -purification rank, -quantum square root rank, and -separable rank (Definition 16). For nonnegative tensors, we have defined the approximate versions of the (Ω, G)-rank, -psd rank, -square root rank, and -nonnegative rank (Definition 18).

Our main technical result is that an element in a convex hull can be approximated by a dimension independent number of generating elements of the convex hull, for several Schatten p-norms (Theorem 5) and ℓ_p-norms (Remark 7). We have leveraged this result to prove dimension independent upper bounds for the approximate (Ω, G)-rank (Theorem 21), -purification rank (Corollary 23), and -separable rank (Proposition 24) (up to a gauge function defined in Equation (5) and the cardinality of the group action G). It follows that many separations between exact (Ω, G)-ranks disappear in the approximate case, both for psd matrices and separable states (Corollary 25), and for nonnegative tensors (Corollary 29). Specifically, for psd matrices, the separation between rank, purification rank, and quantum square root rank disappear in the approximate case. Similarly, for separable states, the separation between rank, purification rank and separable rank disappear in the approximate case. For nonnegative tensors, the separation between rank,

nonnegative rank, and psd rank disappear in the approximate case. Finally, we have presented a procedure (Algorithm 32) to obtain such approximations, attaining the bounds of Theorem 5, and have illustrated its performance with a few examples.

These results can have an impact in the fields where these tensor decompositions are used. The disappearance of separations between the approximate (Ω, G)-rank and the approximate (Ω, G)-purification rank implies that the separation between operator Schmidt rank and purification rank (see [5]) disappears for states with a bounded gauge function. For this application it would be important to look for effective methods to compute the upper bounds of our approximate ranks, in particular $\mu_{\sqrt{}}$ for the purification rank. Another application is in the field of communication complexities, where the joint probability distribution of two parties is described by a nonnegative matrix. Its nn-rank describes the random communication complexity, and the psd-rank the quantum communication complexity [8]. The results of this paper imply that both the approximate random communication complexity and the approximate quantum communication complexity can be upper bounded by a constant independent of the size of the probability distribution (i.e. the number of outcomes of the random variables generated by the parties).

This work leaves several open questions. The most important one is whether one can exploit the tensor product structure to obtain better upper bounds. So far our bounds just use the fact that elements are in the convex hull of some set, but do not exploit, e.g., their spectral properties (see Remark 22) or their tensor product structure. Taking either of these into account should allow to derive better bounds, which could be relevant to prove a disappearance of separations in the $p = 1$ case, for example.

It would also be interesting to do a similar study with local distance measures, which would exploit the Ω structure of the element. In this case, we again expect tighter bounds, since local distance measures are looser than the global distance measures considered in this paper.

Another interesting question is to study the approximate ranks of this paper in the intersection of all $\varepsilon > 0$. This is related to the notion of border rank [2]. Note that, in contrast to matrices, for tensors of higher order the best low-rank approximation might not exist [6].

References

[1] J. Brachat, P. Comon, B. Mourrain, and E. Tsigaridas. *Symmetric tensor decomposition. Linear Algebra Appl.*, 433:1851, 2010.

[2] W. Bruzda, S. Friedland, and K. Życzkowski. *Tensor rank and entanglement of pure quantum states. arXiv: 1912.06854*, 2019.

[3] G. De las Cuevas, M. Hoogsteder Riera, and T. Netzer. Tensor decompositions on simplicial complexes with invariance. *arXiv:1909.01737*, 2019.

[4] G. De las Cuevas and T. Netzer. Mixed states in one spatial dimension: decompositions and correspondence with nonnegative matrices. *J. Math. Phys.*, 61:041901, 2020.

[5] G. De las Cuevas, N. Schuch, D. Perez-Garcia, and J. I. Cirac. Purifications of multipartite states: limitations and constructive methods. *New J. Phys.*, 15:123021, 2013.

[6] V. de Silva and L.-H. Lim. Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM J. Matrix Anal. Appl.*, 30:1084, 2008.

[7] H. Fawzi, J. Gouveia, P. A. Parrilo, R. Z. Robinson, and R. R. Thomas. Positive semidefinite rank. *Math. Program.*, 15:133, 2015.

[8] S. Fiorini, S. Massar, S. Pokutta, H. R. Tiwary, R. de Wolf. Linear vs. Semidefinite Extended Formulations: Exponential Separation and Strong Lower Bounds. *STOC '12 Proc. of the 44th Symposium on Theory of Computing*, 95, 2012.

[9] I. Glasser, R. Sweke, N. Pancotti, J. Eisert, and J. I. Cirac. Expressive power of tensor-network factorizations for probabilistic modeling, with applications from hidden Markov models to quantum machine learning. *Proc. NeurIPS Conference*, 32, 2019.

[10] J. Gouveia, P. A. Parrilo, and R. R. Thomas. Lifts of convex sets and cone factorizations. *Math. Oper. Res.*, 38:248, 2013.

[11] N. J. Higham. *Functions of Matrices. SIAM, Philadelphia*, 2008.

[12] G. M. Ivanov. Approximate Caratheodory's theorem in uniformly smooth Banach spaces. *Discrete Comput. Geom.*, 2019.

[13] R. Jain, Y. Shi, Z. Wei, and S. Zhang. Efficient protocols for generating bipartite classical distributions and quantum states. *IEEE Trans. Inf. Theory*, 59:5171, 2013.

[14] J.G. Jarkovsky, A. Molnar, N. Schuch, J. I. Cirac. Efficient description of many-body systems with Matrix Product Density Operators. *arXiv:2003.12418*, 2020.

[15] E. H. Lieb, K. Ball, E. A. Carlen Sharp uniform convexity and smoothness inequalities for trace norms. *Inventiones mathematicae*, 115:463, 1994.

[16] J. Lindenstrauss. On the modulus of smoothness and divergent series in banach spaces. *Michigan Math. J.*, 10:241–252, 08 1963.

[17] R. E. Megginson. *An Introduction to Banach Space Theory.* Springer, 1998.

[18] R. Orus. Tensor networks for complex systems. *Nat. Rev. Phys.*, 2019.

[19] R. T. Rockafellar. *Convex analysis.* Princeton Mathematical Series, No. 28, xviii+451, 1970.

[20] C. Stark and A. Harrow. Compressibility of Positive Semidefinite Factorizations and Quantum Models. *IEEE Trans. Inf. Theory* 62, 28672880 (2016).

[21] G. Vidal and R. Tarrach. Robustness of entanglement. *Phys. Rev. A*, 59:141, 1999.

[22] G. A. Watson. *Characterization of the subdifferential of some matrix norms. Linear Algebra Appl.*, 170:3345, 1992.

[23] R. Webster. *Convexity. Oxford Science Publications*, 1994.

[24] M. M. Wolf. *Quantum Channels and Operations. Unpublished lecture notes*, 2012.

Appendix A. Weighted simplicial complexes and group actions

In this appendix we give a rigorous definition of weighted simplicial complexes (wsc) Ω and group actions G [3]. Recall that we write $[n]$ for the set $\{0, \ldots, n\}$, and $\mathcal{P}_n$ for the power set $\mathcal{P}([n])$ (i.e. the set of all subsets of $[n]$, which has 2^{n+1} elements).

Definition 33. *(i) A* weighted simplicial complex (wsc) *on $[n]$ is a function*

$$\Omega : \mathcal{P}_n \to \mathbb{N}$$

such that $S_1 \subseteq S_2$ implies that $\Omega(S_1)$ divides $\Omega(S_2)$.

(ii) A set $S \in \mathcal{P}_n$ is called a simplex *of Ω, if $\Omega(S) \neq 0$. In the following, we will assume that each singleton $\{i\} \in \mathcal{P}_n$ is a simplex, and call the elements*

$i \in [n]$ vertices *of the wsc. A maximal simplex (with respect to inclusion) is called a* facet *of Ω. The set of all facets is denoted*

$$\mathcal{F} := \{F \in \mathcal{P}_n : F \text{ facet of } \Omega\},$$

and the set of all facets which contain vertex i is denoted

$$\mathcal{F}_i := \{F \in \mathcal{F} : i \in F\}.$$

The restriction of Ω to $\mathcal{F}$ gives rise to a multiset, called $\widetilde{\mathcal{F}}$, which contains $F \in \mathcal{F}$ precisely $\Omega(F)$-times. Analogously, we define the multiset $\widetilde{\mathcal{F}}_i$ for $i \in [n]$. There exists a canonical collapse map

$$c : \widetilde{\mathcal{F}} \to \mathcal{F} \quad and \quad c : \widetilde{\mathcal{F}}_i \to \mathcal{F}_i$$

mapping all copies of a facet to the underlying facet.

  (iii) Two vertices i, j are neighbors *if*

$$\mathcal{F}_i \cap \mathcal{F}_j \neq \emptyset$$

Two vertices i, j are connected *if there exists a sequence of vertices $i = i_0, i_1, \ldots, i_k = j$ such that i_m and i_{m+1} are neighbors for all $m \in [k-1]$.*

  The next step is the definition of group actions on weighted simplicial complexes. First we will introduce the notions of G-linearity and G-invariance.

Definition 34. *Let G be a group action on the sets X,Y. A function $f : X \to Y$ is called G-*linear *if*

$$f(gx) = gf(x)$$

*holds for all $x \in X$ and $g \in G$. If G acts trivially on Y, we call f G-*invariant.

  We say that the group action of G on X is *free* if $\mathrm{Stab}(x) = \{e\}$ for all $x \in X$, where

$$\mathrm{Stab}(x) := \{g \in G : gx = x\}.$$

To define a group action on Ω we first consider a group action of G on the set $[n]$. Without loss of generality, G might be assumed to be a subgroup of $S_{[n]}$, the permutation group on the set $[n]$. Every group action on $[n]$ canonically induces a group action on $\mathcal{P}_n$. Further, if Ω is G-invariant, G also induces a group action on $\mathcal{F}$, since for $F \in \mathcal{F}$, $g \in G$ and $F \subsetneq S$ it holds that $\Omega(gS) = \Omega(S) = 0$.

Definition 35. *A group action G on the wsc Ω consists of the following:*

  (i) An action G on $[n]$ such that Ω is G-invariant with respect to the induced action on $\mathcal{P}_n$. This induces an action of G on $\mathcal{F}$.
  (ii) An action of G on $\widetilde{\mathcal{F}}$, such that the collapse map

$$c : \widetilde{\mathcal{F}} \to \mathcal{F}$$

  is G-linear (we also say the action of G on $\widetilde{\mathcal{F}}$ refines the action of G on $\mathcal{F}$).

Further, an action of G on the wsc Ω is called free *if the action of G on $\widetilde{\mathcal{F}}$ is free.*

Note that in order to obtain a group action on a wsc, one does not only have to specify how a group action acts on $\mathcal{F}$, but also how it permutes the copies of facets in the multiset $\widetilde{\mathcal{F}}$. It is shown in [3, Prop. 7] that a group action on $\mathcal{F}$ can always be refined to a free group action on $\widetilde{\mathcal{F}}$.