LOW RANK MATRIX RECOVERY FROM RANK ONE MEASUREMENTS

RICHARD KUENG, HOLGER RAUHUT, AND ULRICH TERSTIEGE

Abstract. We study the recovery of Hermitian low rank matrices $X \in \mathbb{C}^{n\times n}$ from undersampled measurements via nuclear norm minimization. We consider the particular scenario where the measurements are Frobenius inner products with random rank-one matrices of the form $a_j a_j^*$ for some measurement vectors $a_1,\ldots,a_m$, i.e., the measurements are given by $y_j = \mathrm{tr}(X a_j a_j^*)$. The case where the matrix $X = xx^*$ to be recovered is of rank one reduces to the problem of phaseless estimation (from measurements, $y_j = |\langle x, a_j\rangle|^2$ via the PhaseLift approach, which has been introduced recently. We derive bounds for the number m of measurements that guarantee successful uniform recovery of Hermitian rank r matrices, either for the vectors a_j, $j = 1,\ldots,m$, being chosen independently at random according to a standard Gaussian distribution, or a_j being sampled independently from an (approximate) complex projective t-design with $t = 4$. In the Gaussian case, we require $m \geq Crn$ measurements, while in the case of 4-designs we need $m \geq Crn\log(n)$. Our results are uniform in the sense that one random choice of the measurement vectors a_j guarantees recovery of all rank r-matrices simultaneously with high probability. Moreover, we prove robustness of recovery under perturbation of the measurements by noise. The result for approximate 4-designs generalizes and improves a recent bound on phase retrieval due to Gross, Kueng and Krahmer. In addition, it has applications in quantum state tomography. Our proofs employ the so-called bowling scheme which is based on recent ideas by Mendelson and Koltchinskii.

1. Introduction

1.1. **The phase retrieval problem.** The problem of retrieving a complex signal from measurements that are ignorant towards phases is abundant in many different areas of science, such as X-ray cristallography [40, 57], astronomy [29] diffraction imaging [67, 57] and more [8, 12, 76]. Mathematically formulated, the problem consists of recovering a complex signal (vector) $x \in \mathbb{C}^n$ from measurements of the form

$$|\langle a_j, x\rangle|^2 = b_j \quad \text{for} \quad j = 1,\ldots,m, \tag{1}$$

where $a_1,\ldots,a_m \in \mathbb{C}^n$ are sampling vectors. This ill-posed inverse problem is called *phase retrieval* and has attracted considerable interest over the last few decades. An important feature of this problem is that the signal x enters the measurement process (1) quadratically. This leads to a non-linear inverse problem. Classical approaches to numerically solving it include alternating projection methods [30, 34]. However, these methods usually require extra constraints and careful selection of parameters, and in particular, no rigorous convergence or recovery guarantees seem to be available.

As Balan et al. pointed out in [7], this apparent obstacle of having nonlinear measurements can be overcome by noting that the measurement process – while quadratic in x – is linear in the outer product xx^*:

$$|\langle a_j, x\rangle|^2 = \mathrm{tr}\left(a_j a_j^* x x^*\right).$$

This "lifts" the problem to a matrix space of dimension n^2, where it becomes linear and can be solved explicitly, provided that the number of measurements m is at least n^2 [7]. However, there is additional structure present, namely the matrix $X = xx^*$ is guaranteed to have rank one. This connects the phase retrieval problem to the young but already extensive field of *low-rank matrix recovery*. Indeed, it is just a special case of low-rank matrix recovery, where both the signal $X = xx^*$ and the measurement matrices $A_j = a_j a_j^*$ are constrained to be proportional to rank-one projectors.

Date: October 25, 2014.

It should be noted, however, that such a reduction to a low rank matrix recovery problem is just one possibility to retrieve phases. Other approaches use polarization identities [2] or alternate projections [60]. Yet another recent method is phase retrieval via Wirtinger flow [14].

1.2. **Low rank matrix recovery.** Building on ideas of compressive sensing [18, 27, 33], low rank matrix recovery aims to reconstruct a matrix of low rank from incomplete linear measurements via efficient algorithms [63]. For our purposes we concentrate on Hermitian matrices $X \in \mathbb{C}^{n\times n}$ and consider measurements of the form

$$\operatorname{tr}(XA_j) = b_j \quad j = 1, \ldots, m \tag{2}$$

where the $A_j \in \mathbb{C}^{n\times n}$ are some Hermitian matrices. For notational simplicity, we define the measurement operator

$$\mathcal{A} : \mathcal{H}_n \to \mathbb{R}^m \quad Z \mapsto \sum_{j=1}^{m} \operatorname{tr}(ZA_j)\, e_j,$$

where $e_1, \ldots, e_m$ denotes the standard basis in $\mathbb{R}^m$. This summarizes an entire (possibly noisy) measurement process via

$$b = \mathcal{A}(X) + \epsilon. \tag{3}$$

Here $b = (b_1, \ldots, b_m)^T$ contains all measurement outcomes and $\epsilon \in \mathbb{R}^m$ denotes additive noise. Low rank matrix recovery can be regarded as a non-commutative version of compressive sensing. Indeed, the structural assumption of low rank assures that the matrix is sparse in its eigenbasis. In parallel to the prominent role of ℓ_1-norm minimization in compressive sensing [33], it is by now well-appreciated [1, 17, 16, 63, 35] that in many relevant measurement scenarios, the sought for matrix X can be efficiently recovered via convex programming, although the corresponding rank minimization problem is NP hard in general [28].

In order to formulate this convex program, we introduce the standard ℓ_p-norm on $\mathbb{R}^n$ or $\mathbb{C}^n$ by $\|x\|_{\ell_p} = (\sum_{\ell=1}^n |x_\ell|^p)^{1/p}$ for $1 \le p < \infty$ and the Schatten-p-norm on the space $\mathcal{H}_n$ of Hermitian $n \times n$ matrices as

$$\|Z\|_p = \left(\sum_{\ell=1}^{n} \sigma_\ell(Z)^p \right)^{1/p} = \operatorname{tr}\left(|Z|^p\right)^{1/p}, \quad p \ge 1,$$

where $\sigma_\ell(Z)$, $\ell = 1, \ldots, n$, denote the singular values of Z, tr is the trace and $|Z| = (Z^*Z)^{1/2}$. Important special cases are the nuclear norm $\|Z\|_* = \|Z\|_1$, the Frobenius norm $\|Z\|_F = \|Z\|_2$ and the spectral norm $\|Z\|_\infty = \|Z\|_{2\to 2} = \sigma_{\max}(Z)$ being the largest singular value. More information, concerning Schatten-p norms can be found in Appendix 5.1.

Assuming the upper bound $\|\epsilon\|_{\ell_2} \le \eta$ on the noise for some $\eta \ge 0$, recovery via nuclear norm minimization corresponds to

$$\underset{Z\in\mathcal{H}_n}{\text{minimize}}\ \|Z\|_1 \text{ subject to } \|\mathcal{A}(Z) - b\|_{\ell_2} \le \eta. \tag{4}$$

This is a convex optimization problem which can be solved computationally efficiently with various strategies [33, Chapter 15], [10, 23, 62, 71]. We note that several alternatives to nuclear norm minimization may also be applied including iteratively reweighted least squares [32], iterative hard thresholding [47, 70], greedy approaches [51] and algorithms specialized to certain measurement maps $\mathcal{A}$ [43], but our analysis is geared towards nuclear norm minimization and does not provide guarantees for these other algorithms.

Up to date, a number of measurement instances have been identified for which nuclear norm minimization (4) – and potentially other algorithms – provably recovers the sought for low-rank matrix from considerably fewer than n^2 measurements [17, 16, 20, 35, 32, 52, 63, 74]. All these constructions are based on randomness, the simplest being a random Gaussian measurement map where all entries $\mathcal{A}_{j,k,\ell}$ in the representation $\mathcal{A}(X)_j = \sum_{k,\ell=1}^n \mathcal{A}_{j,k,\ell} X_{k,\ell}$ are independent mean zero variance one Gaussian random variables. It is shown in [16, 63] that

$$m \ge Crn$$

measurements suffice in order to (stably) reconstruct a matrix $X \in \mathbb{C}^{n\times n}$ of rank at most r with probability at least $1 - \exp(-cm)$, where the constants $C, c > 0$ are universal. This result is based on a version of the by-now classical restricted isometry property so that this result is uniform in the sense that a random draw of $\mathcal{A}$ enables reconstruction of *all* rank r matrices simultaneously with high probability. A corresponding nonuniform result, holding only for a fixed rank r matrix X is stated in [20], see also [4, 74], which shows that essentially $m > 6rn$ measurements are sufficient, thus providing also good constants.

While unstructured Gaussian measurements provide optimal guarantees, which are comparably easy to derive, many applications demand for more structure in the measurement process. A particular instance is the matrix completion problem [22, 17, 19, 35, 21], which aims at recovering missing entries of a matrix which is known to be of low rank. Here, the source of randomness is in the selection of the known entries. In contrast to the unstructured measurements, additional incoherence properties of the matrix to be recovered are required and the bounds on the number of measurements are slightly worse [22, 35], namely $m \geq Crn\log^2(n)$. The matrix completion setup generalizes to measurements with respect to an arbitrary operator basis. The incoherence assumption on the matrix to be recovered can be dropped if in turn the operator basis is incoherent, which is the case for the particular example of Pauli measurements arising in quantum tomography [35, 52]. Here, a sufficient and necessary number of measurements scales like $m \geq Crn\log(n)$.

Rank-one measurements, however, in general fail to be sufficiently incoherent for directly applying proof techniques of the same type. For the particular case of phase retrieval (where the matrix of interest is by construction a rank-one projector) this obstacle could be overcome by providing problem specific recovery guarantees that either manifestly rely on (rank one) Gaussian measurements [13, 74] or result in a non-optimal sampling rate [38, 15, 37].

1.3. Weighted complex projective designs.

The concept of real spherical designs was introduced by Delsarte Goethals and Seidel in a seminal paper [26] and has been studied in algebraic combinatorics [68] and coding theory [26, 59]. Recently, complex projective designs – the natural extension of real spherical designs to the complex unit sphere – have been of considerable interest in quantum information theory [79, 65, 41, 36, 53, 11, 48].

Roughly speaking, a complex projective t-design is a finite subset of the complex unit sphere in $\mathbb{C}^n$ with the particular property that the discrete average of any polynomial of degree (t,t) (i.e., a polynomial $p(z,\bar{z})$ of total degree t both in $z = (z_1,\ldots,z_n)$ and in $\bar{z} = (\bar{z}_1,\ldots,\bar{z}_n)$) or less equals its uniform average. Many equivalent definitions capture this essence, but the following one best serves our purpose.

Definition 1 (*exact, weighted t-design*, Definition 3 in [65])**.** For $t \in \mathbb{N}$, a finite set $\{w_1,\ldots,w_N\} \subset \mathbb{C}^n$ of normalized vectors with corresponding weights $\{p_1,\ldots,p_N\}$ such that $p_i \geq 0$ and $\sum_{i=1}^N p_i = 1$ is called a *weighted complex projective t-design* of dimension n and cardinality N if

$$\sum_{i=1}^N p_i\,(w_i w_i^*)^{\otimes t} = \int_{\mathbb{C}P^{n-1}} (ww^*)^{\otimes t}\,\mathrm{d}w, \tag{5}$$

where the integral on the right hand side is taken with respect to the unique unitarily-invariant probability measure on the complex projective space $\mathbb{C}P^{n-1}$ and the integrand is computed using arbitrary preimages of the $w \in \mathbb{C}P^{n-1}$ in the unit sphere in $\mathbb{C}^n$. (Note that if w_1 and w_2 are elements of the unit sphere that have the same image w in $\mathbb{C}P^{n-1}$ then $w_1w_1^* = w_2w_2^*$.) This definition in particular shows that uniform sampling from a t-design mimics the first $2t$ moments of sampling uniformly according to the Haar measure, which is equivalent to sampling standard Gaussian vectors followed by renormalization.

A simple application of Schur's Lemma – see e.g. [65, Lemma 1] – reveals that the integral on the right hand side of (5) amounts to

$$\int_{\mathbb{C}P^{n-1}} (ww^*)^{\otimes t}\,\mathrm{d}w = \binom{n+t-1}{t}^{-1} P_{\mathrm{Sym}^t}, \tag{6}$$

where P_{Sym^t} denotes the projector onto the totally symmetric subspace Sym^t of $(\mathbb{C}^n)^{\otimes t}$ defined in the appendix – see equation (40).

In accordance with [55], we call a t-design *proper*, if all the weights are equal, i.e., $p_i = 1/N$ for all $i = 1, \ldots, N$.

Although exact, proper t-designs exist and can be constructed in any dimension n for any $t \in \mathbb{N}$ [66, 6, 45, 41], these constructions are typically inefficient in the sense that they require vector sets of exponential size. For example, the construction in [41] requires on the order of $\mathcal{O}\left(t\right)^n$ vectors which scales exponentially in the dimension n. Constructions of *exact, proper* designs with significantly smaller number of vectors (scaling only polynomially in n) are notoriously difficult to find.

By introducing weights, it becomes simpler to obtain designs with a number of elements that scales polynomially in the dimension n. Some existence results can be found in [25], where weighted t-designs appear under the notion of cubatures of strength t. It seems that one can construct weighted t-designs by drawing sufficiently many vectors at random and afterwards solving a linear system for the weights. Further note, that generalizations of cubatures to higher dimensional projections were used in [5] in the context of a generalized phase retrieval problem, where the measurements are given as norms of projections onto higher dimensional subspaces.

2. Main results

2.1. Low rank matrix recovery from rank one Gaussian projections.

Our first main result gives a uniform and stable guarantee for recovering rank-r matrices with $\mathcal{O}(rn)$ rank one measurements that are proportional to projectors onto standard Gaussian random vectors.

Theorem 2. *Consider the measurement process described in (3) with measurement matrices $A_j = a_j a_j^*$, where $a_1, \ldots, a_m \in \mathbb{C}^n$ are independent standard Gaussian distributed random vectors. Furthermore assume that the number of measurements m obeys*

$$m \geq C_1 nr,$$

for $1 \leq r \leq n$ arbitrary. Then with probability at least $1 - \mathrm{e}^{-C_2 m}$ it holds that for any positive semidefinite matrix $X \in \mathcal{H}_n$ with rank at most r, any solution $X^\#$ to the convex optimization problem (4) with noisy measurements $b = \mathcal{A}(X) + \epsilon$, where $\|\epsilon\|_{\ell_2} \leq \eta$, obeys

$$\|X - X^\#\|_2 \leq \frac{C_3 \eta}{\sqrt{m}}. \tag{7}$$

Here, C_1, C_2 and C_3 denote universal positive constants. (In particular, for $\eta = 0$ one has exact reconstruction.)

For the rank one case $r = 1$, Theorem 2 essentialy reproduces the main result in [13] which uses completely different proof techniques. (More precisely, for X of rank 1 the estimate in loc. cit. is $\|X - X^\#\|_2 \leq \frac{C\|\epsilon\|_1}{m}$ with high probability.) A variant of the above statement was shown in [74] to hold (in the real case) for a fixed matrix X of rank one. (More precisely, in loc. cit. it is assumed that X is positive semidefinite and the optimization is performed wrt. the function f given by (9) below.) In fact, our proof reorganizes and extends the arguments of [74, Section 8] in such a way, that Theorem 8.1 of loc. cit. is shown to hold even uniformly (that is simultaneously for all X) and for arbitrary rank. On the contrary to [13], we will not need ε-nets to show uniformity.

2.2. Recovery with 4-designs.

As we will see, the proof method for Theorem 2 can also be applied to measurements drawn independently from a weighted complex projective 4-design in the sense of Definition 1. In [38] exact complex projective t-designs have been applied to the problem of phase retrieval. The main result (Theorem 1) in [38] is a non-uniform exact recovery guarantee for phase retrieval via the convex optimization problem (4) that requires $m = \mathcal{O}\left(tn^{1+2/t}\log^2 n\right)$ measurement vectors that are drawn uniformly from a proper t-design ($t \geq 3$). The proof technique which we are going to employ here, allows for considerably generalizing and improving this statement. We will draw the measurement vectors $a_1, \ldots, a_m \in \mathbb{C}^n$ independently at random from a weighted 4-design $\{p_i, w_i\}_{i=1}^N$, which means that for each draw of a_j, the design element w_i is selected with probability p_i. In the sequel we assume that $n \geq 2$.

Theorem 3. *Let $\{p_i, w_i\}_{i=1}^N$ be a weighted 4-design and consider the measurement process described in (3) with measurement matrices $A_j = \sqrt{n(n+1)} a_j a_j^*$, where $a_1, \ldots, a_m \in \mathbb{C}^n$ are drawn independently from $\{p_i, w_i\}_{i=1}^N$. Furthermore assume that the number of measurements m obeys*

$$m \geq C_4 nr \log n,$$

for $1 \leq r \leq n$ arbitrary. Then with probability at least $1 - \mathrm{e}^{-C_5 m}$ it holds that for any $X \in \mathcal{H}_n$ with rank at most r, any solution $X^\#$ to the convex optimization problem (4) with noisy measurements $b = \mathcal{A}(X) + \epsilon$, where $\|\epsilon\|_{\ell_2} \leq \eta$, obeys

$$\|X - X^\#\|_2 \leq \frac{C_6 \eta}{\sqrt{m}}. \tag{8}$$

Here, $C_4, C_5, C_6 > 0$ again denote universal positive constants.

The normalization factor $\sqrt{n(n+1)}$ leads to approximately the same normalization of the A_j (wrt. the Frobenius norm) as in expectation in the Gauss case. The theorem is a stable, uniform guarantee for recovering arbitrary Hermitian matrices of rank at most r with high probability using the convex optimization problem (4) and $m = \mathcal{O}(nr \log(n))$ measurements drawn independently (according to the design's weights) from a weighted 4-design. It obviously covers sampling from *proper* 4-designs as a special case.

Also, Theorem 3 is close to optimal in terms of the design order t required. In the context of the phase retrieval problem[1] it was shown in [38, Theorem 2], that choosing measurements uniformly from a proper 2-design does not allow for a sub-quadratic sampling rate m without additional structural assumptions on the measurement ensemble. It is presently open whether Theorem 3 also holds for 3-designs.

Finally, note that the results for Gaussian measurement vectors and 4-designs are remarkably similar. They only differ by a logarithmic factor. This underlines the usefulness of complex projective designs as a general-purpose tool for de-randomization – see e.g. [38, Section 1.1.] for further reading on this topic. Also, Theorem 3 resembles insights in the context of distinguishing quantum states [55, 3], where it was pointed out that (approximate) 4-designs "perform almost as good" as uniform measurements (projectors onto random Gaussian vectors). Note that we will generalize Theorem 3 to approximate 4-designs in Theorem 5 below.

2.3. **Extensions.** In this section we state variants of the main theorems which can be proved in a similar way.

2.3.1. *Real-valued case.* Theorem 2 is also valid in the real case, i.e., assuming that the a_j are real standard Gaussian distributed and $\mathcal{H}_n$ is replaced by the space $\mathcal{S}_n$ of real symmetric $n \times n$-matrices. The proof of the corresponding statement is very similar to the one of Theorem 2 and we sketch the necessary adaptations in Subsection 4.3.

2.3.2. *Recovery of positive semidefinite matrices.* The matrix X to be recovered may be known to be positive semidefinite ($X \succcurlyeq 0$) in advance. In this case, one can enforce the reconstructed matrix to be positive semidefinite by considering the optimization program

$$\underset{Z \succcurlyeq 0}{\text{minimize}} \ \operatorname{tr}(Z) \quad \text{subject to} \quad \|\mathcal{A}(Z) - b\|_{\ell_2} \leq \eta$$

instead of the nuclear norm minimization program (4). Then analog versions of Theorems 2, 3 and 5 hold. In particular, the error bounds (7), (8) remain valid. In the noisy case $\eta > 0$, this does not follow directly from these theorems, since the minimizer of the nuclear norm minimization (4) is not guaranteed to be positive semidefinite in the noisy case. The proof proceeds similarly as the ones for the case $X \in \mathcal{H}_n$. Instead of the nuclear norm one has to consider (as in [74]) the function

$$f : \mathcal{H}_n \to \mathbb{R} \cup \{\infty\}, \quad f(X) = \begin{cases} \operatorname{tr}(X), & \text{if } X \succcurlyeq 0 \\ \infty, & \text{otherwise.} \end{cases} \tag{9}$$

[1] i.e., recovering unknown Hermitian matrices of rank one

3. Applications to quantum state tomography

A particular instance of matrix recovery is the task of reconstructing a finite n-dimensional quantum mechanical system which is fully characterized by its *density operator* ρ – an $n \times n$-dimensional positive semidefinite matrix with trace one. Estimating the density operator of an actual (finite dimensional) quantum system is an important task in quantum physics known as *quantum state tomography.*

One is often interested in performing tomography for quantum systems that have certain structural properties. An important structural property – on which we shall focus here – is *purity.* A quantum system is called *pure*, if its density operator has rank one and *almost pure* if it is well approximated by a matrix of low rank $\mathrm{rank}(\rho) = r \ll n$. Assuming this structural property, quantum state tomography is a low-rank matrix recovery problem [39, 35, 31, 52]. An additional requirement for tomography is the fact that the measurement process has to be "experimentally realizable" and – preferably – "efficiently" so.

Any "experimentally realizable" quantum mechanical measurement corresponds to a *positive operator-valued measure* (POVM). In the special case of (finite) n-dimensional quantum systems, a POVM is a set of positive semidefinite matrices $\{M_j\}_{j=1}^N \subset \mathcal{H}_n$ that sum up to the identity, i.e., $\sum_{j=1}^N M_j = \mathrm{id}$ – see e.g. [61, Chapter 2.2.6] for further information.

For practical reasons, it is highly desirable that a quantum measurement (represented by a POVM) can be implemented with reasonable effort. In accordance with [61], we call a POVM-measurement *efficient* (or *practical*), if it can be carried out by performing a number of $\mathcal{O}(\mathrm{polylog}(n))$ elementary steps[2]. Making this notion precise would go beyond the scope of this work and we refer to [3, 61] for further reading.

Below we will concentrate on random constructions of the vectors a_j. We note, however, that implementing the POVM element $a_j a_j^*$ corresponding to the projection onto a Gaussian random vector is *not* efficient as it requires $\mathcal{O}(\mathrm{poly}(n))$ steps. This renders all low rank matrix recovery guarantees which rely on Gaussian measurements – like in Theorem 2 above – inefficient (and therefore impractical) for low rank quantum state tomography. Utilizing a weakened concept of t-designs discussed next, we partly overcome this obstackle with Theorem 5 below and its possible implementations outlined in Sections 3.2.1, 3.2.2.

3.1. An analogue of Theorem 3 for approximate designs.

While Theorem 3 is a substantial derandomization of Theorem 2 and therefore interesting from a theoretical point of view, its usefulness hinges on the availability of constructions of exact weighted 4-designs. Unfortunately, such constructions are notoriously difficult to find unless one relies on randomness, for which, however, the resulting designs are *not* efficient in the sense described in the previous section. One way to circumvent these difficulties is to relax the defining property (5) of a t-design. This approach was – up to our knowledge – introduced by A. Ambainis and J. Emerson [3] and resulted in the notion of approximate designs which is by now well established in quantum information science.

Definition 4 (*Approximate t-design*)**.** We call a weighted set $\{p_i, w_i\}_{i=1}^N$ of normalized vectors an approximate t-design of p-norm accuracy θ_p, if

$$\left\| \sum_{i=1}^N p_i \left(w_i w_i^*\right)^{\otimes t} - \int_{\mathbb{C}P^{n-1}} \left(w w^*\right)^{\otimes t} \mathrm{d}w \right\|_p \leq \binom{n+t-1}{t}^{-1} \theta_p. \tag{10}$$

While accuracy measured in arbitrary Schatten-p-norms is conceivable, the ones measured in operator norm ($p = \infty$) [42, 3, 54, 11] and nuclear norm ($p = 1$) [58] are the ones most commonly used – at least in quantum information theory. For these two accuracies, the definition in particular assures that every approximate t-design is in particular also a k-design for any $1 \leq k \leq t$ with the same p-norm accuracy θ_p [3, 54]. For the sake of being self-contained we provide a proof of this statement in the appendix – see Lemma 16.

[2] This notion is comparable to the *circuit depth* in classical computer science.

A slightly refined analysis reveals that Theorem 3 also holds for sufficiently accurate approximate 4-designs.

Theorem 5. *Fix $1 \leq r \leq n$ arbitrary and let $\{p_i, w_i\}_{i=1}^N$ be an approximate 4-design satisfying*

$$\left\| \sum_{i=1}^N p_i w_i w_i^* - \frac{1}{n}\mathrm{id} \right\|_\infty \leq \frac{1}{n}, \tag{11}$$

that admits either operator norm accuracy $\theta_\infty \leq 1/(16r^2)$, or trace-norm accuracy $\theta_1 \leq 1/4$, respectively. Then, the recovery guarantee from Theorem 3 is still valid (possibly with slightly worse absolute constants $\tilde{C}_4, \tilde{C}_5$ and $\tilde{C}_6$).

3.2. **Protocols for efficient low rank matrix recovery.** Up to now, efficient recovery of low rank density operators by means of the convex optimization problem (4) has been established for random measurements of (generalized) Pauli observables [39, 35]. For this type of measurements, the statistical issues are well understood [31] and Y.K. Liu managed to prove a uniform recovery guarantee [52] which is comparable to the results presented here. Also, this procedure has been tested in experiments [64].

Theorem 5 is similar in spirit and we show here that it permits efficient low rank quantum state tomography for different types of measurements. Indeed, in the field of quantum information theory, various ways of constructing approximate t-designs are known. Most of these methods are inspired by "realistic" quantum mechanical setups (e.g. the circuit model [61, Chapter 4]) and can therefore be – in principle – implemented efficiently in an actual experiment.

Introducing these constructions in full detail would go beyond the scope of this work and we content ourselves with sketching two possible ways of generating approximate 4-design measurements which meet the requirements of Theorem 5. For further clarification on the concepts used here, we refer directly to the stated references.

From now on we shall assume that the dimension $n = 2^d$ is a power of two (d-qubit density operators).

3.2.1. *The Ambainis-Emerson POVM.* In [3], the authors provide a way of constructing a normalized approximate 4-design of operator-norm accuracy $\theta_\infty = \mathcal{O}\left(1/n^{1/3}\right)$, which in addition is a tight frame. They furthermore present a way to generate the corresponding POVM-measurements efficiently – i.e., involving only $\mathcal{O}\left(\mathrm{polylog}(n)\right)$ elementary steps. It therefore meets the requirements of Theorem 5, provided that the maximal rank r of the unknown density operator obeys

$$r \leq C_7 n^{1/6}, \tag{12}$$

where C_7 is a sufficiently small absolute constant. The additional rank requirement stems from the fact that the resulting design only has limited accuracy.

This accuracy can be improved if we construct an approximate design in a much larger space – say $\mathbb{C}^{n^6}$ – and project it down onto an arbitrary n-dimensional subspace. The reason for such an approach is that the projected design's accuracy corresponds to $\theta_\infty = \mathcal{O}\left(\left(n^6\right)^{-1/3}\right) = \mathcal{O}(1/n^2)$. This allows for replacing (12) by the much weaker rank constraint

$$r \leq C_8 n, \tag{13}$$

(where C_8 is again a sufficiently small absolute constant) in order to assure that the design's operator-norm accuracy obeys $\theta_\infty \leq 1/(16r^2)$.

Also, the projected design vectors still form a tight frame, but are sub-normalized, i.e. $\|\tilde{w}_i\|_{\ell_2}^2 = \|P w_i\|_{\ell_2}^2 \leq \|w_i\|_2^2 = 1$. Here, $P : \mathbb{C}^{n^6} \to \mathbb{C}^n$ denotes the projection. However, since they are an approximate design's projection onto a smaller space, they maintain all properties of an approximate 4-design – most notably Lemma 16 – except normalization. In the proof of Theorem 5, normalization is only used once, namely in (28) and sub-normalization is sufficient to guarantee this estimate. Consequently, Theorem 5 is applicable and guarantees universal quantum state tomography via the convex optimization problem (4), provided that (13) holds and $m = C_4 rn \log n$ randomly chosen measurements $\mathrm{tr}\left(\tilde{w}_i \tilde{w}_i^* \rho\right)$ are known.

3.2.2. *Approximate unitary designs.* Another way to generate approximate t-designs is to consider arbitrary orbits of unitary t-designs. Unitary t-designs $\{p_i, U_i\}_{i=1}^N$ are a natural generalization of the spherical design concept to unitary matrices [24, 36]. They have the particular property that every weighted orbit $\{p_i, U_i x\}$ with $\|x\|_{\ell_2} = 1$ of an approximate unitary design forms an approximate complex projective t-design of the same accuracy.

It was shown in [11] that unitary t-designs of arbitrary operator-norm accuracy θ_∞ can be constructed efficiently by using local random circuits. This approach allows for generating an approximate unitary 4-design of operator-norm accuracy $\theta_\infty \leq 1/(16n^2)$ by means of local random circuits of length $C_9 \log(n)^2$, where C_9 is a sufficiently large absolute constant. Consequently, every orbit of the union of all such local random circuits of length $C_9 \log(n)^2$ forms a normalized approximate 4-design which meets the requirements of Theorem 5. One way of implementing such a measurement consists in choosing a local quantum circuit U_i at random, applying its adjoint circuit U_i^* to the density operator ρ and then measuring the two-outcome POVM $\{xx^*, \mathrm{id} - xx^*\}$, where $x \in \mathbb{C}^n$ is arbitrary (but fixed and normalized) to obtain

$$y_i = \mathrm{tr}\left(xx^* U_i^* \rho U_i\right) = \mathrm{tr}\left(U_i xx^* U_i^* \rho\right) = \mathrm{tr}\left(w_i w_i^* \rho\right).$$

According to Theorem 5, $m = \tilde{C}_4 nr \log n$ random measurements of this kind are sufficient to reconstruct any density operator ρ of rank at most r with very high probability via the convex optimization problem (4).

Remark 6. One should note that the approximate unitary designs of [11] are not of a finite nature, because the set of all local random unitaries is continuous. Nevertheless, assuming that such local random unitaries are available as "basic building blocks", local random circuits are efficiently implementable in terms of circuit length. Replacing the atomic expectation values $\sum_{i=1}^N p_i \left(w_i w_i\right)^{\otimes t}$ by their continuous counterparts does not change the argument and Theorem 5 remains valid.

It is worthwhile to point out that the two possible applications of Theorem 5 to the problem of low rank quantum state tomography, as presented here, are not yet optimal. The implementation using the Ambainis-Emerson POVM – presented in 3.2.1 – suffers from the drawback that it demands either a very strong criterion on the density operator's rank – condition (12) – or generating the design in a much larger space and projecting it down. The latter construction is highly unlikely to be optimal and it is furthermore a priori not clear where the corresponding POVM-measurements can be implemented efficiently.

The second approach, on the other hand, suffers from the drawback that carrying out each of the $Crn \log n$ random measurements requires terminating with a very coarse two-outcome POVM measurement. It is very likely that a more fine grained-output statistics could be obtained with comparable effort. The recovery protocol stated here, however, does not allow for advantageously taking into account such refined information about the unknown state.

However, we still feel that mentioning these protocols is worthwhile, as they substantially narrow down the gap between what can be proved (Theorem 5 and the protocols presented in subsection 3.2) and what can be implemented efficiently in an actual quantum state tomography experiment. Next, we provide ideas for further narrowing this gap and finding more protocols that allow for efficient low rank quantum state tomography.

3.3. **Outlook.** The construction of approximate t-designs in Section 3.2.1 via projections from higher-dimensional designs would be much stronger if an efficient protocol for the corresponding POVM measurements could be provided. We leave this for future work. Alternatively, the authors of [3] mention results by Kuperberg [46] who managed to construct exact t-designs containing only $\mathcal{O}\left(n^{2t}\right)$ vectors. They furthermore conjecture that their method of efficiently implementing the corresponding POVM measurement also works for Kuperberg's exact construction. Trying to find such an implementation and combining it with Theorem 3 also does constitute an intriguing follow up-project.

Diagonal-unitary designs are yet another generalization of the spherical design concept to a more restrictive family of unitaries [58]. The notion of a diagonal-unitary design depends on choosing a reference basis and is therefore weaker than the unitary design notation from above. Nevertheless, in [58, Proposition 1] it was shown that the orbit[3] of a particular vector $f_1 \in \mathbb{C}^n$ under a diagonal-unitary t-designs still forms approximate complex projective t-designs with trace-norm accuracy

$$\theta_1 = \binom{n+t-1}{t} \left(\frac{t(t-1)}{n} + \mathcal{O}\left(\frac{1}{n^2} \right) \right). \tag{14}$$

A quick calculation reveals that this orbit forms a normalized tight frame. Unfortunately, the trace-norm accuracy (14) is too weak for a direct application of Theorem 5. However, in [58, Theorem 1] it is shown that the union of all 3-qubit phase-random circuits forms an exact diagonal-unitary 4-design. Similar to local random circuits, such 3-qubit phase-random circuits can in principle be implemented efficiently [58, Proposition 3] in an actual quantum mechanical setup. Furthermore, comparing (14) with the accuracy relation $\theta_\infty \leq \theta_1 \leq n^t \theta_\infty$ – see Lemma 16 in the appendix – suggests that particular orbits of diagonal-unitary designs might possess a much tighter operator-norm accuracy, if the spectrum of their (t-fold tensored) average were sufficiently flat. Such a result, combined with Theorem 5, would lead to a tomography procedure that is similar to the one of Section 3.2.2, but uses random 3-qubit phase gates instead of local random circuits.

4. Proofs

Our proof technique consists in the application of a uniform version of Tropp's bowling scheme, see [74]. The crucial ingredient is a new method due to Mendelson [56] and Koltchiskii, Mendelson [44] (see also [50]) to obtain lower bounds for quantities of the form $\inf_{u \in E} \sum_{j=1}^m |\langle \phi_j, x \rangle|^2$ where the ϕ_j are independent random vectors in $\mathbb{R}^d$ and E is a subset of $\mathbb{R}^d$. We start by recalling from [74] the notions and results underlying this technique.

Suppose we measure $x_0 \in \mathbb{R}^d$ via measurements $y = \Phi x_0 + \epsilon \in \mathbb{R}^m$, where Φ is an $m \times d$ measurement matrix and $\epsilon \in \mathbb{R}^m$ vector of unknown errors. Let $\eta \geq 0$ and assume $\|\epsilon\|_{\ell_2} \leq \eta$. For $f : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ proper convex we aim at recovering x_0 by solving the convex program

$$\text{minimize } f(x) \quad \text{subject to} \quad \|\Phi x - y\|_{\ell_2} \leq \eta. \tag{15}$$

Here, *proper convex* means that f is convex and attains at least one finite value.
Let $K \subseteq \mathbb{R}^d$ be a cone. Then we define the minimum singular value of Φ with respect to K as

$$\lambda_{\min}(\Phi; K) = \inf\{\|\Phi u\|_{\ell_2} : u \in K \cap \mathbb{S}^{d-1}\},$$

where $\mathbb{S}^{d-1}$ is the unit sphere in $\mathbb{R}^d$. For $x \in \mathbb{R}^d$, we consider the (convex) *descent cone*

$$\mathcal{D}(f, x) = \bigcup_{\tau > 0} \{y \in \mathbb{R}^d : \; f(x + \tau y) \leq f(x)\}.$$

With these notions, the success of the convex program (15) can be estimated as follows.

Proposition 7. *([74], see also [20]) Let $x_0 \in \mathbb{R}^d$, $\Phi \in \mathbb{R}^{m \times d}$ and $y = \Phi x_0 + \epsilon$ with $\|\epsilon\|_{\ell_2} \leq \eta$. Let $f : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ be proper convex and let $x^\sharp$ be a solution of the corresponding convex program (15). Then*

$$\|x^\sharp - x_0\|_{\ell_2} \leq \frac{2\eta}{\lambda_{\min}(\Phi; \mathcal{D}(f, x_0))}.$$

The crucial point for us is that in the situation that Φ is a random matrix with i.i.d. rows, the following theorem can be applied to estimate $\lambda_{\min}(\Phi; \mathcal{D}(f, x_0))$ (see also [44, 74, 56]).

[3] For a diagonal-unitary design with respect to the standard basis $e_1, \ldots, e_n$, their result requires the first Fourier vector $f_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i$ as a fiducial. This vector is isomorphic to the $|+\rangle^{\otimes d} = \left(\frac{1}{\sqrt{2}} (e_1 + e_2) \right)^{\otimes d}$ state which is well-known in quantum information theory.

Theorem 8. *(Koltchinskii, Mendelson; Tropp's version [74]) Fix $E \subset \mathbb{R}^d$ and let $\phi_1, \ldots, \phi_m$ be independent copies of a random vector ϕ in $\mathbb{R}^d$. For $\xi > 0$ let*

$$Q_\xi(E;\phi) = \inf_{u \in E} \mathbb{P}\{|\langle \phi, u\rangle| \geq \xi\}$$

$$\text{and} \qquad W_m(E,\phi) = \mathbb{E} \sup_{u \in E} \langle h, u\rangle, \qquad \text{where } h = \frac{1}{\sqrt{m}} \sum_{j=1}^m \varepsilon_j \phi_j$$

with (ε_j) being a Rademacher sequence[4]*. Then for any $\xi > 0$ and any $t \geq 0$ with probability at least $1 - e^{-2t^2}$*

$$\inf_{u \in E} \left(\sum_{i=1}^m |\langle \phi_i, u\rangle|^2 \right)^{1/2} \geq \xi \sqrt{m} Q_{2\xi}(E;\phi) - 2W_m(E,\phi) - \xi t.$$

Remark 9. We note that the above theorem is stated in [74] to hold with probability $1 - e^{-t^2/2}$. Inspecting the proof, however, reveals that the probability estimate can actually be improved to $1 - e^{-2t^2}$.

We will apply the notions in these results in the context of Theorems 2 and 3 as follows:

- identify $\mathcal{H}_n$ with $\mathbb{R}^d = \mathbb{R}^{n^2}$
- Φ is the matrix of $\mathcal{A}$ in the standard basis, i.e., $\Phi(X)_i = \mathrm{tr}(a_i a_i^* X)$
- $f : \mathcal{H}_n \to \mathbb{R} \cup \{\infty\}$ is the nuclear norm, i.e., $f(X) = \|X\|_1$.

In particular,

$$\mathcal{D}(f,X) = \bigcup_{\tau > 0} \{Y \in \mathcal{H}_n : \ f(X + \tau Y) \leq f(X)\}.$$

In Topp's original *bowling scheme*, [74, Sections 7 and 8], a positive semidefinite matrix X of rank 1 is fixed and Theorem 8 is then applied to $E_X = \mathcal{D}(f,X) \cap \mathbb{S}^{d-1}$, where $\mathbb{S}^{d-1} = \{Z \in \mathcal{H}_n : \ \|Z\|_2 = 1\}$. He then uses the Payley-Zygmund inequality to obtain a lower bound for $Q_{2\xi}$ (after choosing some appropriate ξ) and finally applies arguments like conic duality to bound W_m from above.

Our approach differs from the original bowling scheme in one aspect: instead of fixing one rank r-matrix and focusing on E_X, we are going to consider the union $E_r = \{X \in \mathcal{H}_n : \ \mathrm{rank}(X) \leq r, \ X \neq 0\}$ of all low rank matrices. The rest of the proof essentially parallels the bowling scheme from [74]. However, we are going to require an auxiliary statement – Lemma 10 below – in order to obtain a comparable upper bound on W_m. This slightly refined analysis is going to result in a uniform recovery result whose probability of success equals the one for non-uniform recovery of a single fixed X. Note that with such an approach, we do not need to use ε-nets in order to establish uniformity.

For $r \leq n$ let

$$K_r = \bigcup_X \mathcal{D}(f,X),$$

where the union runs over all $X \in \mathcal{H}_n \setminus \{0\}$ of rank at most r. We further define

$$E_r = K_r \cap \mathbb{S}^{d-1} = \bigcup_X E_X,$$

where $E_X = \mathcal{D}(f,X) \cap \mathbb{S}^{d-1}$. We recall that for a convex cone $K \subseteq \mathbb{R}^d$, its *polar cone* is defined to be the closed convex cone

$$K^\circ = \{v \in \mathbb{R}^d : \langle v, x\rangle \leq 0 \text{ for all } x \in K\}.$$

A crucial ingredient for Theorems 2 and 3 is the following lemma.

[4] A Rademacher vector $\epsilon = (\epsilon_j)_{j=1}^m$ is a vector of independent Rademacher random variables, taking the values ±1 with equal probability.

Lemma 10. *Let $A \in \mathcal{H}_n$ be a Hermitian $n \times n$-matrix. Then*

$$\sup_{Y \in E_r} \operatorname{tr}(A \cdot Y) \leq 2\sqrt{r}\|A\|_\infty.$$

By duality and the matrix Hölder inequality this statement is equivalent to

$$\|Y\|_1 \leq 2\sqrt{r} \quad \text{for all } Y \in E_r. \tag{16}$$

The following proof is inspired by [74, Section 8], where similar arguments are used.

Proof. It is enough to show that, for any $X \in \mathcal{H}_n \setminus \{0\}$ of rank at most r, we have

$$\sup_{Y \in E_X} \operatorname{tr}(A \cdot Y) \leq 2\sqrt{r}\|A\|_\infty.$$

We may assume that X has precisely rank $r \geq 1$. By weak duality for cones, see [74, Proposition 4.2] or [33, eq. (B.40)], we have $\sup_{Y \in E_X} \operatorname{tr}(A \cdot Y) \leq \operatorname{dist}_F(A, \mathcal{D}(f, X)^\circ)$, where as usual $\operatorname{dist}_F(A, \mathcal{D}(f, X)^\circ) = \inf_{B \in \mathcal{D}(f,X)^\circ} \|A - B\|_2$. By [74, Fact 4.3], we know that the polar cone $\mathcal{D}(f, X)^\circ$ is the closure of $\bigcup_{\tau \geq 0} \tau \cdot \partial f(X)$. For $S \in \partial f(X)$ and $\tau \geq 0$, it follows that

$$\sup_{Y \in E_X} \operatorname{tr}(A \cdot Y) \leq \|A - \tau \cdot S\|_2.$$

Write $X = \sum_{i=1}^r \lambda_i x_i x_i^*$, where the x_i are orthonormal and the λ_i are non-zero. Extend $x_1, \ldots, x_r$ to an orthonormal basis $x_1, \ldots, x_n$ of $\mathbb{C}^n$ and write A in the form

$$A = \sum \tilde{a}_{i,j} x_i x_j^*.$$

(Hence the $\tilde{a}_{i,j}$ form the matrix obtained from A by a basis change to $x_1, \ldots, x_n$.) Define the four blocks $A_1 = \sum_{i,j \leq r} \tilde{a}_{i,j} x_i x_j^*$, $A_2 = \sum_{i \leq r, j > r} \tilde{a}_{i,j} x_i x_j^*$, $A_3 = \sum_{i > r, j \leq r} \tilde{a}_{i,j} x_i x_j^* = A_2^*$ and $A_4 = \sum_{i,j > r} \tilde{a}_{i,j} x_i x_j^*$. It is well known that $\partial\|X\|_1$ consists of all matrices of the form

$$S = \sum_{i=1}^r \operatorname{sgn}(\lambda_i) x_i x_i^* + S_2,$$

where $S_2 \in \mathcal{H}_n$ has the property that $S_2 x_i = 0$ for all $i \in \{1, \ldots, r\}$ and $\|S_2\|_\infty \leq 1$. (See for example [78], where the real analogue is shown.) Consider now

$$S = \sum_{i=1}^r \operatorname{sgn}(\lambda_i) x_i x_i^* + \tau^{-1} A_4 \in \partial\|X\|_1, \quad \text{where } \tau = \|A_4\|_\infty.$$

(If $\tau = 0$, let $S = \sum_{i=1}^r \operatorname{sgn}(\lambda_i) x_i x_i^*$.) To simplify the notation, write $S_1 = \sum_{i=1}^r \operatorname{sgn}(\lambda_i) x_i x_i^*$. Then

$$\begin{aligned}
\|A - \tau S\|_2 &= \|A - A_4 - \tau S_1\|_2 = \left(\operatorname{tr}(A_1 - \tau S_1)^2 + 2\operatorname{tr}(A_2^* A_2)\right)^{1/2} \\
&= \left(\|A_1 - \tau S_1)\|_2^2 + 2\|A_2^*\|_2^2\right)^{1/2} \leq \left(2\|A_1\|_2^2 + 2\|\tau \cdot S_1\|_2^2 + 2\|A_2^*\|_2^2\right)^{1/2} \\
&= \left(2\|A \cdot x_1\|_2^2 + \ldots + 2\|A \cdot x_r\|_2^2 + 2\|\tau \cdot S_1\|_2^2\right)^{1/2} \\
&\leq \left(2r\|A\|_\infty^2 + 2r\tau^2\right)^{1/2} \leq 2\sqrt{r}\|A\|_\infty,
\end{aligned}$$

since $\tau = \|A_4\|_\infty \leq \|A\|_\infty = \lambda$. □

4.1. **Proof of Theorem 2.** In order to prove both statements of Theorem 2, it is enough by Proposition 7 to show that for $m \geq cnr$ with probability at least $1 - e^{-\gamma m}$

$$\inf_{Y \in E_r} \left(\sum_{j=1}^m \operatorname{tr}(a_j a_j^* Y)^2\right)^{1/2} \geq c_1\sqrt{m}$$

for suitable positive constants c, c_1, γ. For $\xi > 0$ let

$$Q_\xi = \inf_{Z \in E_r} \mathbb{P}(|\operatorname{tr}(a_j a_j^* Z)| \geq \xi). \tag{17}$$

Further let

$$H = \frac{1}{\sqrt{m}} \sum_{j=1}^{m} \varepsilon_j a_j a_j^*, \tag{18}$$

where the ε_j form a Rademacher sequence independent of everything else, and introduce

$$W_m = \mathbb{E} \sup_{Y \in E_r} \operatorname{tr}(H \cdot Y).$$

By Theorem 8, for any $\xi > 0$ and any $t \geq 0$ with probability at least $1 - e^{-2t^2}$,

$$\inf_{Y \in E_r} \left(\sum_{j=1}^{m} (\operatorname{tr}(a_j a_j^* Y))^2 \right)^{1/2} \geq \xi \sqrt{m} Q_{2\xi} - 2W_m - \xi t.$$

Following Tropp's bowling scheme, we first estimate $Q_{2\xi}$ for a suitable ξ. As in [74], we conclude from the Payley-Zygmund inequality (see e.g. [33, Lemma 7.16]) that

$$\mathbb{P}\{|\langle aa^*, U\rangle|^2 \geq \frac{1}{2}(\mathbb{E}|\langle aa^*, U\rangle|^2)\} \geq \frac{1}{4} \cdot \frac{(\mathbb{E}|\langle aa^*, U\rangle|^2)^2}{\mathbb{E}|\langle aa^*, U\rangle|^4}. \tag{19}$$

(Here a follows the standard Gaussian distribution on $\mathbb{C}^n$.) Assume now $\|U\|_2 = 1$ and write $U = \sum_i \lambda_i u_i u_i^*$, where $\sum_i \lambda_i^2 = 1$ and the u_i are orthonormal. Then $\langle aa^*, U\rangle = \operatorname{tr}(aa^*U) = \sum_j \lambda_j \operatorname{tr}(aa^* u_j u_j^*) = \sum_j \lambda_j |u_j^* a|^2$ and hence,

$$|\langle aa^*, U\rangle|^2 = \sum_{i,j} \lambda_i \lambda_j |u_i^* a|^2 |u_j^* a|^2.$$

The $u_j^* a$ form independent standard (complex) Gaussian random variables. To compute the moments of a standard complex Gaussian random variable Z, write $Z = X + iY$ where X, Y are independent and $\mathcal{N}(0, \frac{1}{2})$ distributed. The $2k$-th moment of X resp. Y is $\frac{(2k)!}{2^{2k} k!}$, which allows us to compute higher moment of Z, for example, $\mathbb{E}|Z|^2 = \mathbb{E}X^2 + \mathbb{E}Y^2 = 1$ and $\mathbb{E}|Z|^4 = \mathbb{E}X^4 + 2\mathbb{E}X^2\mathbb{E}Y^2 + \mathbb{E}Y^4 = 2$. Similarly, we obtain $\mathbb{E}|Z|^6 = 6$ and $\mathbb{E}|Z|^8 = 24$ (and more generally $\mathbb{E}|Z|^{2k} = k!$). Thus, we conclude that

$$\mathbb{E}|\langle aa^*, U\rangle|^2 = \sum_{i \neq j} \lambda_i \lambda_j + 2\sum_i \lambda_i^2 = \sum_{i,j} \lambda_i \lambda_j + \sum_i \lambda_i^2 = (\sum_i \lambda_i)^2 + 1 \geq 1 \tag{20}$$

and

$$(\mathbb{E}|\langle aa^*, U\rangle|^2)^2 = (\sum_i \lambda_i)^4 + 2(\sum_i \lambda_i)^2 + 1.$$

Expanding $\mathbb{E}|\langle aa^*, U\rangle|^4$ in a similar way, we obtain

$$\begin{aligned}\mathbb{E}|\langle aa^*, U\rangle|^4 &= \sum_{i,j,k,\ell} \lambda_i \lambda_j \lambda_k \lambda_\ell + \sum_{i,k,\ell} \lambda_i^2 \lambda_k \lambda_\ell + 2\sum_{i,k} \lambda_i^2 \lambda_k^2 + 4\sum_{i,k} \lambda_i^3 \lambda_k + 16\sum_i \lambda_i^4 \\ &= (\sum_i \lambda_i)^4 + (\sum_i \lambda_i)^2 + 2 + 4(\sum_i \lambda_i)(\sum_i \lambda_i^3) + 16\sum_i \lambda_i^4,\end{aligned}$$

where we used that $\sum_i \lambda_i^2 = 1$. Again because of $\sum_i \lambda_i^2 = 1$ we have $|\lambda_i| \leq 1$ for all i and hence $|\sum_i \lambda_i^3| \leq \sum_i \lambda_i^2 = 1$ and similarly $\sum_i \lambda_i^4 \leq \sum_i \lambda_i^2 = 1$. Also observe that $|\sum_i \lambda_i| \leq 1 + (\sum_i \lambda_i)^2$. Combining these inequalities with the above expressions for $\mathbb{E}|\langle aa^*, U\rangle|^4$ and $(\mathbb{E}|\langle aa^*, U\rangle|^2)^2$, we obtain the inequality

$$\mathbb{E}|\langle aa^*, U\rangle|^4 \leq 24(\mathbb{E}|\langle aa^*, U\rangle|^2)^2.$$

Combining this with (19) and (20), we obtain

$$Q_{1/\sqrt{2}} \geq \frac{1}{96}.$$

Thus we choose $\xi = \frac{1}{2\sqrt{2}}$.

In order to estimate W_m, we use Lemma 10 to obtain

$$W_m = \mathbb{E} \sup_{Y \in E_r} \operatorname{tr}(H \cdot Y) \leq 2\sqrt{r} \cdot \mathbb{E}\|H\|_\infty. \tag{21}$$

By the arguments in [75, Section 5.4.1] we have $\mathbb{E}\|H\|_\infty \leq c_2\sqrt{n}$ if $m \geq c_3 n$ for suitable constants c_2, c_3, see also [74, Section 8]. Choosing $t = c_4\sqrt{m}$ and $m \geq cnr$ for suitable constants c, c_4, the proof of Theorem 2 is completed.

Remark 11. In [13], a uniform result for phase retrieval in the Gaussian case is proved using an inexact dual certificate. One can write down a generalization of this dual certificate for the rank r-case, but following the arguments of loc. cit., the resulting number of required measurements then seems to depend significantly worse than linearly on r. It might be possible to rather adapt the arguments in [38, 37] based on a different construction of a dual certificate in order to derive linear scaling of m in r, but the resulting proof would be more complicated than ours (and likely lead to more logarithmic factors).

4.2. **Proof of Theorem 3.** Let us now turn to proving the analogous result for complex projective 4-designs. It is convenient to rescale the (normalized) 4-design vectors as

$$\tilde{w}_i := \sqrt[4]{(n+1)n}\, w_i \quad \forall i = 1, \ldots, N. \tag{22}$$

This mimics the expected length of random Gaussian vectors (which corresponds to $\mathbb{E}\|a_j\|_2^2 = n$) and we will call the system $\{\tilde{w}_i\}$ a *super-normalized* 4-design. We can apply the same technique as in the proof of Theorem 2, provided that we can derive a suitable lower bound for $Q_{2\xi}$ for some $0 < \xi < 1/2$ and an upper bound for $\mathbb{E}\|H\|_\infty$. The following two technical propositions serve this purpose.

Proposition 12. *Assume that a is drawn at random from a super-normalized weighted 4-design. Then*

$$Q_\xi = \inf_{Z \in E_r} \mathbb{P}\left(|\mathrm{tr}\,(aa^* Z)| \geq \xi\right) \geq \frac{(1-\xi^2)^2}{24} \tag{23}$$

for all $\xi \in [0,1]$.

The proof of this statement is similar to the proof of Theorem 4 in [3] and – likewise – equation (15) in [55]. However, since we are interested in a bound on the probability of an event happening, rather than bounding an expectation value, we use the Payley-Zygmund inequality instead of Berger's one [9] (which states $\mathbb{E}\left[|S|\right] \geq \mathbb{E}\left[S^2\right]^{3/2}\mathbb{E}\left[S^4\right]^{-1/2}$).

Proof. The desired statement follows, if we can show that

$$\mathbb{P}\left(|\mathrm{tr}\,(aa^* Z)| \geq \xi\right) \geq \frac{(1-\xi^2)^2}{24} \tag{24}$$

holds for any matrix $Z \in \mathcal{H}_n$ obeying $\|Z\|_2 = 1$. For such Z we define the random variable $S := |\mathrm{tr}\,(aa^* Z)|$. Since a is chosen at random from a (super-normalized) complex projective 4-design, we can use the design's defining property (5) together with (6) to evaluate the second and fourth moment of S. Indeed,

$$\begin{aligned}
\mathbb{E}S^2 &= \mathbb{E}\mathrm{tr}\,(aa^*Z)^2 = \mathrm{tr}\left(\mathbb{E}\,(aa^*)^{\otimes 2} Z^{\otimes 2}\right) = \mathrm{tr}\left(\sum_{i=1}^N p_i\,(\tilde{w}_i\tilde{w}_i^*)^{\otimes 2} Z^{\otimes 2}\right) \\
&= (n+1)n\;\mathrm{tr}\left(\sum_{i=1}^N p_i\,(w_i w_i^*)^{\otimes 2} Z^{\otimes 2}\right) = (n+1)n\binom{n+1}{2}^{-1}\mathrm{tr}\left(P_{\mathrm{Sym}^2} Z^{\otimes 2}\right) \\
&= 2\mathrm{tr}\left(P_{\mathrm{Sym}^2} Z^{\otimes 2}\right)
\end{aligned}$$

and likewise

$$\mathbb{E}S^4 = \mathbb{E}\mathrm{tr}\,(aa^*Z)^4 = \mathrm{tr}\left(\sum_{i=1}^N p_i\,(\tilde{w}_i\tilde{w}_i^*)^{\otimes 4} Z^{\otimes 4}\right) = \frac{4!(n+1)n}{(n+3)(n+2)}\mathrm{tr}\left(P_{\mathrm{Sym}^4} Z^{\otimes 4}\right).$$

The remaining right hand sides are standard expressions in multilinear algebra and can for instance be calculated using wiring calculus. Indeed, Lemma 17 in the appendix implies that

$$\mathbb{E}S^2 = 2\mathrm{tr}\left(P_{\mathrm{Sym}^2} Z^{\otimes 2}\right) = \mathrm{tr}(Z)^2 + \mathrm{tr}(Z^2) = \mathrm{tr}(Z)^2 + 1, \tag{25}$$

because $\mathrm{tr}(Z^2) = \|Z\|_F^2 = 1$ by assumption, hence,

$$(\mathbb{E}S^2)^2 \geq \max\{1, \mathrm{tr}(Z)^4\}.$$

Similarly, Lemma 17 assures

$$\begin{aligned}
\mathbb{E}S^4 &= \frac{4!(n+1)n}{(n+3)(n+2)} \mathrm{tr}\left(P_{\mathrm{Sym}^4} Z^{\otimes 4}\right) \\
&= \frac{(n+1)n}{(n+3)(n+2)} \left(6\mathrm{tr}(Z^4) + 8\mathrm{tr}(Z)\mathrm{tr}(Z^3) + 6\mathrm{tr}(Z)^2\mathrm{tr}(Z^2) + 3\mathrm{tr}(Z^2)^2 + \mathrm{tr}(Z)^4\right) \\
&\leq \left(6\mathrm{tr}(Z^4) + 8\mathrm{tr}(Z)\mathrm{tr}(Z^3) + 6\mathrm{tr}(Z)^2 + \mathrm{tr}(Z)^4 + 3\right),
\end{aligned}$$

where the simplifications in the last line are due to $\mathrm{tr}(Z^2) = \|Z\|_F^2 = 1$ and $\frac{(n+1)n}{(n+3)(n+2)} \leq 1$. Using the hierarchy of Schatten-p-norms – in particular $\mathrm{tr}(Z^4) = \|Z\|_4^4 \leq \|Z\|_2^4 = 1$ and $\mathrm{tr}(Z^3) \leq \|Z\|_3^3 \leq \|Z\|_2^3 = 1$ – yields

$$\begin{aligned}
\mathbb{E}S^4 &\leq 6\mathrm{tr}(Z^4) + 8\mathrm{tr}(Z)\mathrm{tr}(Z^3) + 6\mathrm{tr}(Z)^2 + \mathrm{tr}(Z)^4 + 3 \\
&\leq \left(6\|Z\|_4^4 + 8\|Z\|_3^3 + 10\right) \max\left\{1, \mathrm{tr}(Z)^4\right\} \leq 24 \max\left\{1, \mathrm{tr}(Z)^4\right\}.
\end{aligned}$$

Having precise knowledge of the second and fourth moments and the trivial fact that $\mathrm{tr}(Z)^2 \geq 0$ allows us to use the Payley-Zygmund inequality (for the random variable S^2) to bound

$$\begin{aligned}
\mathbb{P}\left(|\mathrm{tr}\left(aa^* Z\right)| \geq \xi\right) = \mathbb{P}\left(S^2 \geq \xi^2\right) &\geq \mathbb{P}\left(S^2 \geq \xi^2\left(1 + \mathrm{tr}(Z)^2\right)\right) \\
&= \mathbb{P}\left(S^2 \geq \xi^2 \mathbb{E}S^2\right) \geq \left(1-\xi^2\right)^2 \frac{(\mathbb{E}S^2)^2}{\mathbb{E}S^4} \\
&\geq (1-\xi^2)^2 \frac{\max\{1, \mathrm{tr}(Z)^4\}}{24 \max\{1, \mathrm{tr}(Z)^4\}} = \frac{(1-\xi^2)^2}{24}.
\end{aligned}$$

This completes the proof. □

Proposition 13. *Let H be the matrix defined in (18), where the a_j's are chosen independently at random from a super-normalized weighted 1-design. Then it holds that*

$$\mathbb{E}\|H\|_\infty \leq c_4 \sqrt{n \log(2n)} \quad \text{with } c_4 = 3.1049, \tag{26}$$

provided that $m \geq 2n \log n$.

Proof. Since the ϵ_j's in the definition of H form a Rademacher sequence, the non-commutative Khintchine inequality [75, p. 19], see also [33, Exercise 8.6(d)], is applicable and yields

$$\begin{aligned}
\mathbb{E}\|H\|_\infty = \mathbb{E}_a \mathbb{E}_\epsilon \frac{1}{\sqrt{m}} \left\| \sum_{j=1}^m \epsilon_j a_j a_j^* \right\|_\infty &\leq \sqrt{\frac{2\log(2n)}{m}} \mathbb{E}_a \left\| \left(\sum_{j=1}^m \left(a_j a_j^*\right)^2 \right)^{1/2} \right\|_\infty \\
= \sqrt{\frac{2\log(2n)}{m}} \mathbb{E}_a \left\| \sqrt{(n+1)n} \sum_{j=1}^m a_j a_j^* \right\|_\infty^{1/2} &\leq \sqrt{\frac{2\sqrt{2}n\log(2n)}{m}} \left(\mathbb{E}_a \left\| \sum_{j=1}^m a_j a_j^* \right\|_\infty \right)^{1/2}.
\end{aligned} \tag{27}$$

Here we have used super-normalization of our design vectors $(a_j a_j^*)^2 = \|a_j\|_2^2 a_j a_j^* = \sqrt{(n+1)n} a_j a_j^*$ according to (22), the fact that $\|Z^{1/2}\|_\infty = \|Z\|_\infty^{1/2}$ holds for $Z \in \mathcal{H}_d$ arbitrary and Jensen's inequality in the last estimate. It remains to bound $\mathbb{E}\|\sum_j a_j a_j^*\|_\infty$. To this end, we will use the matrix Chernoff inequality of Theorem 15 for $X_j = a_j a_j^*$ and calculate

$$\begin{aligned} \|X_j\|_\infty &= \|a_j a_j^*\|_\infty = \|a_j\|_2^2 \leq \max_{1\leq i\leq N} \|\tilde{w}_i\|_2^2 = \sqrt{(n+1)n} \leq \sqrt{2}n =: R, \qquad (28) \\ \|\sum_{j=1}^m \mathbb{E}X_j\|_\infty &= \|\sum_{j=1}^m \sum_{i=1}^N p_i \tilde{w}_i \tilde{w}_i^* \|_\infty = m\sqrt{n(n+1)} \left\| \sum_{i=1}^N p_i w_i w_i^* \right\|_\infty \\ &= m\sqrt{(n+1)n} \left\| \frac{1}{n} \mathrm{id} \right\|_\infty = \frac{m\sqrt{(n+1)n}}{n} \leq \sqrt{2}m, \qquad (29) \end{aligned}$$

where we once more have taken into account super-normalization and used the 1-design property. Theorem 15 together with the assumption $m \geq 2n\log n$ implies that, for any $\tau > 0$,

$$\begin{aligned} \mathbb{E}\|\sum_{j=1}^m a_j a_j^* \|_\infty &\leq \frac{e^\tau - 1}{\tau}\sqrt{2}m + \tau^{-1}\sqrt{2}n\log(n) \leq \frac{e^\tau - 1}{\tau}\sqrt{2}m + \tau^{-1}\sqrt{2}m/2 \\ &= \left(\frac{e^\tau - 1}{\tau}\sqrt{2} + \frac{1}{\sqrt{2}\tau} \right) m. \end{aligned}$$

The choice $\tau = 1.27$ approximately minimizes the above expression and yields

$$\mathbb{E}\|\sum_{j=1}^m a_j a_j^* \|_\infty \leq c_5 m \quad \text{with } c_5 = 3.4084.$$

Combining this estimate with (27) yields the desired statement with $c_4 = 2^{3/4}\sqrt{c_5} = 3.1049$. □

Now we are ready to prove the second main theorem of this work.

Proof of Theorem 3. The proof of Theorem 2 shows that we only need suitable bounds for $Q_{2\xi}$ and for $\mathbb{E}\|H\|_\infty$ (both notions are defined analogously to the Gaussian case). Fix $0 < \xi < 1/2$ arbitrary. For any such ξ, a lower bound for $Q_{2\xi}$ is provided by Proposition 12 and an upper bound for $\mathbb{E}\|H\|_\infty$ in this case can be obtained from Proposition 13. Setting $m = C_4 nr\log n$, choosing the constants C_4, C_5 and C_6 appropriately (depending on the particular choice of ξ) and applying Theorem 8 then yields the desired result in complete analogy to the Gaussian case (proof of Theorem 2). □

Remark 14. The difference in the sampling rate m by a factor proportional to $\log n$ in Theorems 2 and 3 stems from the fact that Proposition 13 is by a factor of $\sqrt{\log(n)}$ weaker than its Gaussian analogue [75, Section 5.4.1], where $\mathbb{E}\|H\|_\infty \leq c_2\sqrt{n}$.

4.3. **Proof of Theorem 2 for real Gaussian vectors.** As already mentioned in paragraph 2.3.1 the proof of this statement is almost identical to the proof of Theorem 2. The only difference is the estimate of $Q_{2\xi}$. Using the moments of the real instead of the complex standard Gaussian distribution, the reasoning in the proof of Theorem 2 yields the estimates $\mathbb{E}|\langle aa^*, U\rangle|^2 \geq 2$, (compare also with [74]). Using real moments, one further obtains $\mathbb{E}|\langle aa^*, U\rangle|^4 \leq 27(\mathbb{E}|\langle aa^*, U\rangle|^2)^2$ (alternatively one can use Gaussian hypercontractivity as done in [74], which gives the factor 81 instead of 27.) This yields $Q_1 \geq \frac{1}{108}$, and the rest of the proof is the same as before.

4.4. **Proof for recovery of positive semidefinite matrices.** The only part in the proof of the recovery result for positive semidefinite matrices stated in Section 2.3.2 that slightly differs from the one for arbitrary Hermitian matrices, is the proof of a corresponding version of Lemma 10. The subdifferential of the function f introduced in (9) slightly differs from the subdifferential of the nuclear norm. For $X = \sum_{i=1}^r \lambda_i x_i x_i^*$, where all λ_i are nonzero, $\partial f(X)$ consists of all matrices of the form

$$S = \sum_{i=1}^r x_i x_i^* + S_2,$$

where $S_2 \in \mathcal{H}_n$ has the property that $S_2 x_i = 0$ for all $i \in \{1, \ldots, r\}$ and all eigenvalues of S_2 do not exceed 1. Hence we choose (in the notation of the proof of Lemma 10)

$$S = \sum_{i=1}^{r} x_i x_i^* + \tau^{-1} A_4 \in \partial f(X).$$

Then the remainder of the proof of Lemma 10 is the same.

4.5. **Proof of Theorem 5.** The proof of this generalized statement proceeds along the same lines as the one of Theorem 3. However, Propositions 12 and 13 – as well as their respective proofs – have to be slightly altered due to the weaker requirements imposed by Theorem 5.

4.5.1. *Generalized version of Proposition 12.* Under the assumptions of Theorem 5, a weaker version of (23), namely

$$Q_\xi = \inf_{Z \in E_r} \mathbb{P}\left(|\operatorname{tr}(aa^* Z)| \geq \xi \right) \geq \frac{(1-2\xi^2)^2}{192} \tag{30}$$

for all $0 \leq \xi \leq 1/\sqrt{2}$ is still valid. This statement can be shown analogously to Proposition 12. However, one has to establish bounds on the second and fourth moments in a slightly more involved way, depending also on the type of design accuracy. Let us start with generalizing the second moment estimate of $S := |\operatorname{tr}(aa^* Z)|$ for an approximate 4-design with operator norm accuracy $\theta_\infty \leq 1/(16r^2)$:

$$\begin{aligned}
\mathbb{E}S^2 &= (n+1)n \left(\sum_{i=1}^{N} p_i \left(w_i w_i^*\right)^{\otimes 2}, Z^{\otimes 2} \right) \\
&= 2\left(P_{\mathrm{Sym}^2}, Z^{\otimes 2}\right) + (n+1)n \left(\sum_{i=1}^{N} p_i \left(w_i w_i^*\right)^{\otimes 2} - \binom{n+1}{2}^{-1} P_{\mathrm{Sym}^2}, Z^{\otimes 2} \right) \\
&\geq 2\left|\left(P_{\mathrm{Sym}^2}, Z^{\otimes 2}\right)\right| - (n+1)n \left\| \sum_{i=1}^{N} p_i \left(w_i w_i^*\right)^{\otimes 2} - \binom{n+1}{2}^{-1} P_{\mathrm{Sym}^2} \right\|_\infty \left\| Z^{\otimes 2} \right\|_1 \qquad (31) \\
&\geq 2\left|\left(P_{\mathrm{Sym}^2}, Z^{\otimes 2}\right)\right| - 2\theta_\infty \|Z\|_1^2 \geq 2\left|\left(P_{\mathrm{Sym}^2}, Z^{\otimes 2}\right)\right| - \frac{8r}{16r^2}, \\
&> 2\left|\left(P_{\mathrm{Sym}^2}, Z^{\otimes 2}\right)\right| - 1/2, \qquad (32)
\end{aligned}$$

where we have used the fact that $\left(P_{\mathrm{Sym}^2}, Z^{\otimes 2}\right) = \left|\left(P_{\mathrm{Sym}^2}, Z^{\otimes 2}\right)\right|$ (see Lemma 17), the matrix Hölder inequality and the fact that $\|Z\|_1 \leq 2\sqrt{r}$ – see (16). The estimates for designs with nuclear norm accuracy $\theta_1 \leq 1/4$ is very similar. Replacing the matrix Hölder inequality in (31) by

$$\left(\sum_{i=1}^{N} p_i \left(w_i w_i^*\right)^{\otimes 2} - \binom{n+1}{2}^{-1} P_{\mathrm{Sym}^2}, Z^{\otimes 2} \right) \geq - \left\| \sum_{i=1}^{N} p_i \left(w_i w_i^*\right)^{\otimes 2} - \binom{n+1}{2}^{-1} P_{\mathrm{Sym}^2} \right\|_1 \left\| Z^{\otimes 2} \right\|_\infty$$

yields the same lower bound (32) due to $\|Z^{\otimes 2}\|_\infty = \|Z\|_\infty^2 \leq \|Z\|_2^2 = 1$ (where the last equality follows from $Z \in E_r$). Applying Lemma 17 then yields

$$\mathbb{E}S^2 \geq \operatorname{tr}(Z)^2 + 1/2 \quad \text{and} \quad (\mathbb{E}S^2)^2 \geq \frac{1}{4} \max\{1, \operatorname{tr}(Z)^4\}.$$

which is the (slightly weaker) analogue of (25). Likewise we derive a fourth moment bound:

$$\begin{aligned}\mathbb{E}S^4 &= \left(\mathbb{E}\left[(aa^*)^{\otimes 4}\right], Z^{\otimes 4}\right) = (n+1)^2 n^2 \left(\sum_{i=1}^{N} p_i \left(w_i w_i^*\right)^{\otimes 4}, Z^{\otimes 4}\right) \\ &\leq (n+1)^2 n^2 \binom{n+3}{4}^{-1} \left| \left(P_{\mathrm{Sym}^4}, Z^{\otimes 4}\right) \right| \\ &\quad + (n+1)^2 n^2 \left\| \sum_{i=1}^{N} p_i \left(w_i w_i^*\right)^{\otimes 4} - \binom{n+3}{4}^{-1} P_{\mathrm{Sym}^4} \right\|_\infty \left\| Z^{\otimes 4} \right\|_1 \\ &\leq \frac{4!(n+1)n}{(n+3)(n+2)} \left(\left| \left(P_{\mathrm{Sym}^4}, Z^{\otimes 4}\right) + \theta_\infty \|Z\|_1^4 \right) \leq \left|4! \left(P_{\mathrm{Sym}^4}, Z^{\otimes 4}\right) \right| + 4! \frac{16r^2}{16r^2}.\end{aligned}$$

As above, using the nuclear norm accuracy $\theta_1 \leq 1/4$ instead of the operator norm accuracy yields the bound $\mathbb{E}\left[S^4\right] \leq \left|4! \left(P_{\mathrm{Sym}^4}, Z^{\otimes 4}\right)\right| + 4!/4 < \left|4! \left(P_{\mathrm{Sym}^4}, Z^{\otimes 4}\right)\right| + 4!$. Lemma 17 yields then in both cases

$$\begin{aligned}\mathbb{E}\left[S^4\right] &\leq \left|4! \mathrm{tr}\left(P_{\mathrm{Sym}^4} Z^{\otimes 4}\right)\right| + 24 \leq 6\mathrm{tr}(Z^4) + 8|\mathrm{tr}(Z)\mathrm{tr}(Z^3)| + 6\mathrm{tr}(Z)^2 + \mathrm{tr}(Z)^4 + 27 \\ &\leq 48 \max\{1, \mathrm{tr}(Z)^4\},\end{aligned}$$

compare the proof of Proposition 12. Having these bounds at hand, allows for applying the Payley Zygmund inequality to obtain

$$\begin{aligned}\mathbb{P}\left(\left|\mathrm{tr}\left(aa^* Z\right)\right| \geq \xi\right) &= \mathbb{P}\left(S^2 \geq \xi^2\right) \geq \mathbb{P}\left(S^2 \geq 2\xi^2 \left(1/2 + \mathrm{tr}(Z)^2\right)\right) \geq \mathbb{P}\left(S^2 \geq 2\xi^2 \mathbb{E}\left[S^2\right]\right) \\ &\geq (1-2\xi^2)^2 \frac{(\mathbb{E}S^2)^2}{\mathbb{E}S^4} \geq (1-2\xi^2)^2 \frac{\max\{1, \mathrm{tr}(Z)^4\}/4}{48 \max\{1, \mathrm{tr}(Z)^4\}} = \frac{(1-2\xi^2)^2}{192}.\end{aligned}$$

The proof is completed.

4.5.2. *Generalized version of Proposition 13.* The assumptions in Theorem 5 assure that (26) is still valid, possibly with a larger absolute constant c_4. Again, the proof of this generalized statement is very similar to the proof of Proposition 13. Indeed, only the bound (29) for the matrix Chernoff inequality needs to be slightly altered. The assumption (11) implies that

$$\| \sum_{j=1}^{m} \mathbb{E}\left[X_j\right] \|_\infty \leq m\sqrt{(n+1)n} \left(\| \frac{1}{n} \mathrm{id} \|_\infty + \| \sum_{i=1}^{N} p_i w_i w_i^* - \frac{1}{n} \mathrm{id} \|_\infty \right) \leq 2\sqrt{2} m.$$

Consequently, applying the matrix Chernoff inequality yields (26) with a slightly larger absolute constant c_4.

5. Appendix

5.1. **Schatten p-norms.** Recall from Section 1.2 that for $1 \leq p < \infty$, the Schatten-p-norm on $\mathcal{H}_n$ is defined as

$$\|Z\|_p = \mathrm{tr}\left(|Z|^p\right)^{1/p} = \left(\sum_{i=1}^{n} |\lambda_i|^p\right)^{1/p},$$

where $\lambda_1, \ldots, \lambda_n$ denote the n eigenvalues of $Z \in \mathcal{H}_n$. For $p = \infty$ one defines similarly

$$\|Z\|_\infty = \max\{|\lambda_1|, \ldots, |\lambda_n|\},$$

i.e., $\|Z\|_\infty$ is the spectral norm of Z. The Frobenius norm $\|\cdot\|_F = \|\cdot\|_2$ is induced by the the Hilbert-Schmitt (or Frobenius) scalar product

$$(X, Y) = \mathrm{tr}\left(XY\right),$$

which makes $\mathcal{H}_n$ a Hilbert space. The Schatten-p norms are non-increasing in p, i.e. for any $0 < p \leq p' \leq \infty$

$$\|Z\|_p \geq \|Z\|_{p'} \tag{33}$$

holds for all $Z \in \mathcal{H}_n$. The following relations provide converse inequalities for particular instances of Schatten p-norms that are used frequently in our work:

$$\|Z\|_1 \leq \sqrt{\operatorname{rank}(Z)}\|Z\|_2 \quad \text{and} \quad \|Z\|_2 \leq \sqrt{\operatorname{rank}(Z)}\|Z\|_\infty \quad \text{for all } Z \in \mathcal{H}_n. \tag{34}$$

In addition, we often use a particular instance of the matrix Hölder inequality, namely

$$|\,(X,Y)\,| \leq \|X\|_1 \|Y\|_\infty \quad \text{for all } X,Y \in \mathcal{H}_n. \tag{35}$$

5.2. **Matrix Chernoff inequality.** The matrix version of the classical Chernoff inequality for the expection of a sum of independent random matrices shown in [73, Theorem 5.1.1] (see also [72]) reads as follows.

Theorem 15. *Let $X_1,\ldots,X_m$ be a sequence of independent random positive definite matrices in $\mathcal{H}_n$ satisfying*

$$\|X_\ell\|_\infty \leq L \quad \textit{almost surely for all } \ell = 1,\ldots,m.$$

Then, for any $\tau > 0$, their sum obeys

$$\mathbb{E}\|\sum_{\ell=1}^m X_\ell\|_\infty \leq \frac{e^\tau - 1}{\tau}\|\sum_{\ell=1}^m \mathbb{E}X_\ell\|_\infty + \tau^{-1} L \log n.$$

5.3. **Multilinear algebra.** We briefly repeat some standard concepts in multilinear algebra which are convenient for our proof of Proposition 12. They can be found in any textbook on multilinear algebra – e.g. [49] – but we nonetheless include them here for the sake of being self-contained.

Let $V_1,\ldots,V_k$ be (finite dimensional, complex) vector spaces and let $V_1^*,\ldots,V_k^*$ denote their duals. A function $f : V_1 \times \cdots \times V_k \to \mathbb{C}$ is *multilinear*, if it is linear in each space V_i. We denote the space of such functions by $V_1^* \otimes \cdots \otimes V_k^*$ and call it the *tensor product* of $V_1^*,\ldots,V_k^*$. Consequently, for one fixed n-dimensional vector space V, the tensor product $(V)^{\otimes k} = \bigotimes_{i=1}^k V$ is the space of all multilinear functions

$$f : \underbrace{(V)^* \times \cdots \times (V)^*}_{k \text{ times}} \mapsto \mathbb{C}, \tag{36}$$

and we call the elementary elements $z_1 \otimes \cdots \otimes z_k$ the *tensor product* of the vectors $z_1,\ldots,z_k \in V$.

With this notation, the space of linear maps $V \to V$ ($n \times n$-matrices) corresponds to the tensor product $\mathcal{M}_n := V \otimes V^*$ which is spanned by $\{x \otimes y^* : \; x,y \in V\}$ – the set of all rank-1 matrices. Using this tensor product description of $\mathcal{M}_n$ allows for defining the (matrix) tensor product $\mathcal{M}_n^{\otimes k}$ in complete analogy to above. We refer to its elements $Z_1 \otimes \cdots \otimes Z_k$ as the tensor product of the matrices $Z_1,\ldots,Z_k \in \mathcal{M}_n$.

On this tensor space, we define the *partial trace* (over the i-th tensor system) to be the natural contraction

$$\begin{aligned} \operatorname{tr}_i : \; \mathcal{M}_n^{\otimes k} &\to \mathcal{M}_n^{\otimes(k-1)} \\ Z_1 \otimes \cdots \otimes Z_k &\mapsto \operatorname{tr}(Z_i) Z_1 \otimes \cdots \otimes Z_{i-1} \otimes Z_{i+1} \otimes \cdots \otimes Z_k. \end{aligned}$$

The partial trace over multiple systems can then be obtained by concatenating individual traces of this form, e.g.

$$\operatorname{tr}_{i,j} = \operatorname{tr}_i \circ \operatorname{tr}_j : \mathcal{M}_n^{\otimes k} \to \mathcal{M}_n^{\otimes(k-2)} \tag{37}$$

for $1 \leq i < j \leq k$ arbitrary and so forth. A particular property of arbitrary partial traces is that they preserve positive semidefiniteness – see e.g. [61, Section 8.3.1] or any lecture notes on quantum information theory. If a matrix $Z \in \mathcal{M}_n^{\otimes k}$ is positive semidefinite, then $\operatorname{tr}_i(Z) \in \mathcal{M}^{\otimes(k-1)}$ is again positive semidefinite for any $1 \leq i \leq k$. This behavior naturally extends to multiple partial traces in the sense of (37). The *full trace* corresponds to

$$\begin{aligned} \operatorname{tr} := \operatorname{tr}_{1,\ldots,k} : \; \mathcal{M}_n^{\otimes k} &\to \mathbb{C} \\ Z_1 \otimes \cdots \otimes Z_k &\mapsto \operatorname{tr}(Z_1)\cdots\operatorname{tr}(Z_k). \end{aligned}$$

This implies that the nuclear norm is multiplicative with respect to the tensor structure, i.e.,

$$\|Z_1 \otimes \cdots Z_k\|_1 = \operatorname{tr}\left(|Z_1| \otimes \cdots \otimes |Z_k|\right) = \operatorname{tr}\left(|Z_1|\right)\cdots\operatorname{tr}\left(|Z_k|\right) = \|Z_1\|_1 \cdots \|Z_k\|_1 \tag{38}$$

for $Z_1, \ldots, Z_k \in \mathcal{M}$ arbitrary. A singular value decomposition – see e.g. [77, Lecture 2] – reveals that the same is true for the operator norm, i.e.

$$\left\| Z_1 \otimes \cdots \otimes Z_k \right\|_\infty = \| Z_1 \|_\infty \cdots \| Z_k \|_\infty . \tag{39}$$

Let us now return to the k-fold tensor space $V^{\otimes k}$ of n-dimensional complex vectors. We define the (symmetrizer) map $P_{\mathrm{Sym}^k} : (V)^{\otimes k} \to (V)^{\otimes k}$ via their action on elementary elements:

$$P_{\mathrm{Sym}^k} \left(z_1 \otimes \cdots \otimes z_k \right) := \frac{1}{k!} \sum_{\pi \in S_k} z_{\pi(1)} \otimes \cdots \otimes z_{\pi(k)}, \tag{40}$$

where S_k denotes the group of permutations of k elements. This map projects $(V)^{\otimes k}$ onto the totally symmetric subspace Sym^k of $(V)^{\otimes k}$ whose dimension [49, Exercise 2.6.3.5] is

$$\dim \mathrm{Sym}^k = \binom{n+k-1}{k}. \tag{41}$$

Using these basic concepts of multilinear algebra and (6), we can show that every approximate t-design is also an approximate design of lower order.

Lemma 16. *Every approximate t-design of accuracy measured either in operator- or trace-norm is also an approximate k-design of the same accuracy for any $1 \leq k \leq t$. Furthermore the accuracies θ_∞ and θ_1 are related via*

$$\theta_\infty \leq \theta_1 \leq n^t \theta_\infty . \tag{42}$$

This statement is implicitly proved in [3], where the authors use an equivalent definition of approximate t-designs as averaging sets of complex polynomials of degree at most (t,t). With this alternative definition, Lemma 16 follows naturally from the fact that every polynomial of degree at most (k,k) with $1 \leq k \leq t$ is a particular instance of a degree-(t,t)-polynomial. Here we provide an alternative proof that uses concepts from multilinear algebra and accesses Definition 4 directly. Such a proof idea is mentioned in [54, Section 2.2.3] and we include the full argument here for the sake of being self-contained.

Proof of Lemma 16. Let us start with proving the statement for the accuracy measured in operator norm. In this case, Definition 4 is equivalent to demanding

$$(1-\theta_\infty) \int_{\mathbb{C}P^{n-1}} \left(w w^* \right)^{\otimes t} \mathrm{d}w \leq \sum_{i=1}^N p_i \left(w_i w_i^* \right)^{\otimes t} \leq (1+\theta_\infty) \int_{\mathbb{C}P^{n-1}} \left(w w^* \right)^{\otimes t} \mathrm{d}w . \tag{43}$$

The desired statement follows if we can show that (43) implies a corresponding inequality for smaller tensor powers k. Fix $1 \leq k \leq t$ and note that the inequality chain (43) is preserved under taking arbitrary partial traces, because partial traces respect the positive semidefinite ordering. This in particular implies that

$$\begin{aligned} (1-\theta_\infty) \int_{\mathbb{C}P^{n-1}} \mathrm{tr}_{1,\ldots,(t-k)} \left(\left(w w^* \right)^{\otimes t} \right) \mathrm{d}w \quad & \leq \quad \sum_{i=1}^N p_i \mathrm{tr}_{1,\ldots,(t-k)} \left(\left(w_i w_i^* \right)^{\otimes t} \right) \\ & \leq \quad (1+\theta_\infty) \int_{\mathbb{C}P^{n-1}} \mathrm{tr}_{1,\ldots,(t-k)} \left(\left(w w^* \right)^{\otimes t} \right) \mathrm{d}w \end{aligned}$$

remains valid. Due to normalization $\| w_i \|_{\ell_2} = 1$ and and since we calculate the integrals using preimages of the $w \in \mathbb{C}P^{n-1}$ in the unit sphere, these expressions can be readily calculated. Indeed,

$$\mathrm{tr}_{1,\ldots,(t-k)} \left(\left(w_i w_i^* \right)^{\otimes t} \right) = \left(w_i w_i^* \right)^{\otimes k} \left| \langle w_i, w_i \rangle \right|^{2(t-k)} = \left(w_i w_i^* \right)^{\otimes k}$$

and

$$\int_{\mathbb{C}P^{n-1}} \mathrm{tr}_{1,\ldots,(t-k)} \left(\left(w w^* \right)^{\otimes t} \right) \mathrm{d}w = \int_{\mathbb{C}P^{n-1}} \left(w w^* \right)^{\otimes k} \left| \langle w, w \rangle \right|^{2(t-k)} \mathrm{d}w = \int_{\mathbb{C}P^{n-1}} \left(w w^* \right)^{\otimes k} \mathrm{d}w .$$

The desired statement follows.

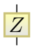
The analogous statement for accuracy measured in trace-norm directly follows from the fact that the nuclear norm is monotonic with respect to partial traces, i.e., $\|\mathrm{tr}_i(Z)\|_1 \leq \|Z\|_1$ for any $Z \in \mathcal{M}_n^{\otimes t}$ and $1 \leq i \leq t$ [77, Lecture 2]. Combining this with the calculations above reveals that

$$\begin{aligned}
&\left\| \sum_{i=1}^N p_i \left(w_i w_i^* \right)^{\otimes k} - \int_{\mathbb{C}P^{n-1}} \left(w w^* \right)^{\otimes k} \mathrm{d}w \right\|_1 \\
&= \left\| \mathrm{tr}_{1,\ldots,t-k} \left(\sum_{i=1}^N p_i \left(w_i w_i^* \right)^{\otimes t} - \int_{\mathbb{C}P^{n-1}} \left(w w^* \right)^{\otimes t} \mathrm{d}w \right) \right\|_1 \\
&\leq \left\| \sum_{i=1}^N p_i \left(w_i w_i^* \right)^{\otimes t} - \int_{\mathbb{C}P^{n-1}} \left(w w^* \right)^{\otimes t} \mathrm{d}w \right\|_1 \leq \theta_1 .
\end{aligned}$$

Finally, inequality (42) directly follows from comparing trace and operator norm on $\mathcal{M}_n^{\otimes t}$ which is isomorphic to the space of all $n^t \times n^t$-dimensional matrices.

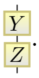
□

5.4. **Wiring calculus in multilinear algebra.** The defining properties (5), (10) of exact and approximate complex projective t-designs are phrased in terms of tensor spaces. For calculations in multilinear algebra – particularly if they involve (partial) traces– *wiring diagrams* [49, Chapter 2.11] are very useful, as they provide a way of computing contractions of tensors pictorially. Here we give a brief introduction that should suffice for our calculations and defer the interested reader to [38] and references therein for further reading.

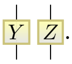
In wiring calculus, every tensor is associated with a box, and every index corresponds to a line emanating from this box. Two connected lines correspond to connected indices. The formalism becomes much clearer when applying it to matrix calculus. A matrix $Z : \mathbb{C}^n \to \mathbb{C}^n$ can be viewed as two-index-tensors $Z^i{}_j$ and is thus represented by a node  with upper line corresponding to the index i and the lower one to j. Two matrices Y, Z are multiplied by contracting Z's upper index with Y's lower one:

$$(YZ)^i{}_j = \sum_{k=1}^n Y^i{}_k Z^k{}_j .$$

In wiring calculus matrix multiplication is therefore represented by

$YZ =$ .

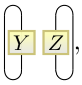
Tensor products of matrices are arranged in parallel, i.e.,

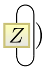
$Y \otimes Z =$ .

Taking traces of tensor products, e.g.,

$$Y \otimes Z \mapsto \mathrm{tr}(Y \otimes Z) = \sum_{i,j=1}^n Y^i{}_i Z^j{}_j$$

just corresponds to contracting parallel matrix indices and therefore

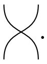
$\mathrm{tr}(Y \otimes Z) =$ ,

which straightforwardly extends to larger (and smaller, namely $\mathrm{tr}(Z) =$ ) tensor systems.

Finally, we are going to require *transpositions* on $(\mathbb{C}^n)^{\otimes t}$ which act by interchanging the i-th and j-th tensor factor. For example

$$\sigma_{(1,2)} \left(x \otimes y \otimes \cdots \right) = y \otimes x \otimes \cdots ,$$

with $x, y \in \mathbb{C}^n$ arbitrary. Note that these transpositions generate the full group of permutations. For $(\mathbb{C}^n)^{\otimes 2}$ there are only two transpositions, namely

$\underline{1} =$ (trivial permutation) and $\sigma_{(1,2)} =$ .

But for higher tensor systems more permutations can occur. In wiring calculus, permutations therefore act by interchanging different input and output lines.

We are now ready to prove the statements required in Proposition 12.

Lemma 17. *For an abritrary Hermitian matrix $Z \in \mathcal{H}_n$ and a positive integer m, it holds*

$$m!\left(P_{\mathrm{Sym}^m} Z^{\otimes m}\right) = \sum_{\substack{(j_1,\ldots,j_m)\in\mathbb{N}_0^m \\ \sum_{k=1}^m k j_k = m}} \frac{m!}{\prod_{k=1}^m j_k!\, k^{j_k}} \prod_{k=1}^m \mathrm{tr}(Z^k)^{j_k}.$$

In particular, for $m = 2$ we obtain

$$2\mathrm{tr}\left(P_{\mathrm{Sym}^2} Z^{\otimes 2}\right) = \mathrm{tr}(Z)^2 + \mathrm{tr}(Z^2),$$

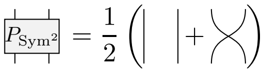
and for $m = 4$ we obtain

$$4!\ \mathrm{tr}\left(P_{\mathrm{Sym}^4} Z^{\otimes 4}\right) = \mathrm{tr}(Z)^4 + 8\mathrm{tr}(Z)\mathrm{tr}(Z^3) + 3\mathrm{tr}(Z^2)^2 + 6\mathrm{tr}(Z)^2\mathrm{tr}(Z^2) + 6\mathrm{tr}(Z^4).$$

Proof. We start with the case $m = 2$ and then extend the argument to the general case. The basic formula for P_{Sym^2} is given by

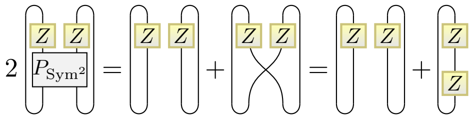
$$P_{\mathrm{Sym}^2} = \frac{1}{2}\sum_{\pi\in S_2} \pi = \frac{1}{2}\left(\underline{1} + \sigma_{(1,2)}\right),$$

and its pictorial counterpart is therefore



.

Applying the graphical calculus introduced above then yields

$2\mathrm{tr}\left(P_{\mathrm{Sym}^2} Z^{\otimes 2}\right) \;=\;$ 

$= \mathrm{tr}(Z)^2 + \mathrm{tr}(Z^2),$

which is the desired statement for $m = 2$.

Expanding $m!\,(P_{\mathrm{Sym}^m} Z^{\otimes m})$ analogously in the general case, we obtain for each $\pi \in S_m$ one summand which corresponds to a wiring diagram in which m copies of the node $\boxed{Z}$ are involved. More precisely, the wiring diagram corresponding to π is obtained by connecting for each $i \in \{1, \ldots, m\}$ the output line of the i-th copy of $\boxed{Z}$ to the input line of the $\pi(i)$-th copy of $\boxed{Z}$. If we write π as a product of k cyclic permutations, $\pi = c_1 \cdots c_k$, then the wiring diagram of π consists of k closed loops, one for each of the cyclic permutations $c_1, \ldots, c_k$. Write $c_i = (i_1, \ldots, i_{r_i})$. Then the loop corresponding to c_i connects r_i copies of $\boxed{Z}$. Hence the contribution of π to the whole sum is $\mathrm{tr}(Z^{r_1}) \cdots \mathrm{tr}(Z^{r_k})$. Thus for a given partition $m = r_1 + \ldots + r_k$ of m, any element of S_m which is the product of k cyclic (and disjoint) permutations of lengths $r_1, \ldots, r_k$ respectively gives the same contribution $\mathrm{tr}(Z^{r_1}) \cdots \mathrm{tr}(Z^{r_k})$.

Note that we can rewrite any partition of m in the form $m = j_1 \cdot 1 + \ldots + j_m \cdot m$, where j_i counts how often the summand i appears in that partition. It remains to count for each partition $m = j_1 \cdot 1 + \ldots + j_m \cdot m$ of m how many elements of S_m there are which are a product of precisely j_1 cyclic permutations of length 1, of precisely j_2 cyclic permutations of length 2 and so on (all the cyclic permutations being disjoint). It is easy to see (and well known, see for example [69, Proposition 1.3.2]) that there are precisely $\frac{m!}{\prod_{k=1}^m j_k!\, k^{j_k}}$ such permutations in S_m. Each of them contributes a summand $\mathrm{tr}(Z^1)^{j_1} \ldots \mathrm{tr}(Z^m)^{j_m}$ to $m!\,(P_{\mathrm{Sym}^m} Z^{\otimes m})$. This gives the claimed formula. □

Acknowledgements. RK is glad to acknowledge inspiring discussions with D. Gross and helpful comments from J. Aaberg.

The work of RK is supported by scholarship funds from the State Graduate Funding Program of Baden-Württemberg, the Excellence Initiative of the German Federal and State Governments (Grant ZUK 43), the ARO under contracts, W911NF-14-1-0098 and W911NF-14-1-0133 (Quantum Characterization, Verification, and Validation), the Freiburg Research Innovation Fund, and the DFG. HR and UT acknowledge funding by the European Research Council through the Starting Grant StG 258926 (SPALORA). RK and HR would like to thank the *Mathematisches Forschungsinstitut Oberwolfach* and the organizers of the Oberwolfach workshop *Mathematical Physics meets Sparse Recovery* (April 2014, Workshop ID: 1416a), where discussions on the topic of this article have started.

References

[1] A. Ahmed, B. Recht, and J. Romberg. Blind deconvolution using convex programming. *preprint*, 2012.

[2] B. Alexeev, A. S. Bandeira, M. Fickus, and D. G. Mixon. Phase retrieval with polarization. *SIAM J. Imaging Sci.*, 7:35–66, 2014.

[3] A. Ambainis and J. Emerson. Quantum t-designs: t-wise independence in the quantum world. In *22nd Annual IEEE Conference on Computational Complexity, Proceedings*, pages 129–140, 2007.

[4] D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp. Living on the edge: Phase transitions in convex programs with random data. *Inform. Inference*, 3(3):224–294, 2014.

[5] C. Bachoc and M. Ehler. Signal reconstruction from the magnitude of subspace components. *arXiv:1209.5986*, 2012.

[6] B. Bajnok. Construction of spherical t-designs. *Geom. Dedicata*, 43:167–179, 1992.

[7] R. Balan, B. G. Bodmann, P. G. Casazza, and D. Edidin. Painless reconstruction from magnitudes of frame coefficients. *J. Fourier Anal. Appl.*, 15:488–501, 2009.

[8] R. Balan, P. Casazza, and D. Edidin. On signal reconstruction without phase. *Appl. Comput. Harmon. Anal.*, 20(3):345–356, 2006.

[9] B. Berger. The fourth moment method. *SIAM J. on Comp.*, 26(4):1188–1207, 1997.

[10] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge Univ. Press, 2004.

[11] F. G. Brandao, A. W. Harrow, and M. Horodecki. Local random quantum circuits are approximate polynomial-designs. *preprint arXiv:1208.0692*, 2012.

[12] O. Bunk, A. Diaz, F. Pfeiffer, C. David, B. Schmitt, D. K. Satapathy, and J. F. van der Veen. Diffractive imaging for periodic samples: retrieving one-dimensional concentration profiles across microfluidic channels. *Acta Crystallographica Section A: Foundations of Crystallography*, 63(4):306–314, 2007.

[13] E. Candès and X. Li. Solving quadratic efquations via PhaseLift when there are about as many equations as unknowns. *Found. Comput. Math.*, pages 1–10, 2013.

[14] E. Candes, X. Li, and M. Soltanolkotabi. Phase Retrieval via Wirtinger Flow: Theory and Algorithms. *ArXiv e-prints*, jul 2014.

[15] E. Candes, X. Li, and M. Soltanolkotabi. Phase retrieval from coded diffraction patterns. *Appl. Comput. Harmonic Anal.*, to appear.

[16] E. J. Candès and Y. Plan. Tight oracle bounds for low-rank matrix recovery from a minimal number of random measurements. *IEEE Trans. Inform. Theory*, 57(4):2342–2359, 2011.

[17] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Found. Comput. Math.*, 9(6):717–772, 2009.

[18] E. J. Candès and T. Tao. Near optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory*, 52(12):5406–5425, 2006.

[19] E. J. Candès and T. Tao. The power of matrix completion: near-optimal convex relaxation. *IEEE Trans. Information Theory*, 56(5):2053–2080, 2010.

[20] V. Chandrasekaran, B. Recht, P. Parrilo, and A. Willsky. The convex geometry of linear inverse problems. *Found. Comput. Math.*, 12(6):805–849, 2012.

[21] Y. Chen. Incoherence-optimal matrix completion. *preprint arXiv:1310.0154*, 2013.

[22] Y. Chen, S. Bhojanapalli, S. Sanghavi, and R. Ward. Completing any low-rank matrix, provably. *ArXiv:1306.2979*, 2013.

[23] P. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In H. Bauschke, R. Burachik, P. Combettes, V. Elser, D. Luke, and H. Wolkowicz, editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer, 2011.

[24] C. Dankert, R. Cleve, J. Emerson, and E. Livine. Exact and approximate unitary 2-designs: constructions and applications. *arXiv preprint quant-ph/0606161*, 2006.

[25] P. De La Harpe and C. Pache. Cubature formulas, geometrical designs, reproducing kernels, and markov operators. In *Infinite groups: geometric, combinatorial and dynamical aspects*, pages 219–267. Springer, 2005.

[26] P. Delsarte, J. Goethals, and J. Seidel. Spherical codes and designs. *Geom. Dedicata*, 6:363–388, 1977.

[27] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

[28] M. Fazel. *Matrix rank minimization with applications.* PhD thesis, 2002.

[29] C. Fienup and J. Dainty. Phase retrieval and image reconstruction for astronomy. In H. Stark, editor, *Image Recovery: Theory and Application*, pages 231–275. Academic Press, San Diego, 1987.

[30] J. Fienup. Phase retrieval algorithms: A comparison. *Appl. Opt.*, 21(15):2758–2769, 1982.

[31] S. T. Flammia, D. Gross, Y.-K. Liu, and J. Eisert. Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators. *New J. Phys.*, 14:095022, 2012.

[32] M. Fornasier, H. Rauhut, and R. Ward. Low-rank matrix recovery via iteratively reweighted least squares minimization. *SIAM J. Optim.*, 21(4):1614–1640, 2011.

[33] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing.* Applied and Numerical Harmonic Analysis. Birkhäuser, 2013.

[34] R. Gerchberg and W. Saxton. Phase retrieval by iterated projection. *Optik*, 35, 1972.

[35] D. Gross. Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans. Inform. Theory*, 57:1548–1566, 2011.

[36] D. Gross, K. Audenaert, and J. Eisert. Evenly distributed unitaries: on the structure of unitary designs. *J. Math. Phys.*, 48:052104, 22, 2007.

[37] D. Gross, F. Krahmer, and R. Kueng. Improved recovery guarantees for phase retrieval from coded diffraction patterns. *preprint arXiv:1402.6286*, 2014.

[38] D. Gross, F. Krahmer, and R. Kueng. A partial derandomization of PhaseLift using spherical designs. *J. Fourier Anal. Appl.*, to appear.

[39] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert. Quantum state tomography via compressed sensing. *Phys. Rev. Lett.*, 105:150401, Oct 2010.

[40] R. Harrison. Phase problem in crystallography. *JOSA A*, 10(5):1046–1055, 1993.

[41] A. Hayashi, T. Hashimoto, and M. Horibe. Reexamination of optimal quantum state estimation of pure states. *Phys. Rev. A*, 72, SEP 2005.

[42] P. Hayden, D. Leung, P. W. Shor, and A. Winter. Randomizing quantum states: Constructions and applications. *Commun. Math. Phys.*, 250(2):371–391, 2004.

[43] R. KeshavanH., A. Montanari, and S. Oh. Matrix completion from a few entries. *IEEE Trans. Inform. Theory*, 56(6):2980 – 2998, 2010.

[44] V. Koltchinskii and S. Mendelson. Bounding the smallest singular value of a random matrix without concentration. *ArXiv:1312.3580*, dec 2013.

[45] J. Korevaar and J. Meyers. Chebyshev-type quadrature on multidimensional domains. *J. Approx. Theory*, 79:144–164, 1994.

[46] G. Kuperberg. Numerical cubature using error-correcting codes. *SIAM J. Numer. Anal.*, 44(3):897–907, 2006.

[47] A. Kyrillidis and V. Cevher. Matrix recipes for hard thresholding methods. *J. Math. Imaging Vis.*, 48:235–265, 2014.

[48] C. Lancien and A. Winter. Distinguishing multi-partite states by local measurements. *Commun. in Math.l Phys.*, 323(2):555–573, 2013.

[49] J. M. Landsberg. *Tensors: geometry and applications.* Providence, RI: American Mathematical Society (AMS), 2012.

[50] G. Lecué and S. Mendelson. Sparse recovery under weak moment assumptions. *ArXiv:1401.2188*, 2014.

[51] K. Lee and Y. Bresler. ADMiRA: Atomic decomposition for minimum rank approximation. *IEEE Trans. Image Process.*, 56(9):4402 – 4416, 2010.

[52] Y.-K. Liu. Universal low-rank matrix recovery from Pauli measurements. *Adv. Neural Inf. Process. Syst.*, pages 1638–1646, 2011.

[53] R. A. Low. Large deviation bounds for k-designs. *Proc. R. Soc. Lond., Ser. A, Math. Phys. Eng. Sci.*, 465:3289–3308, 2009.

[54] R. A. Low. *Pseudo-randomness and learning in quantum computation.* PhD thesis, University of Bristol, arXiv:1006.5227, 2010.

[55] W. Matthews, S. Wehner, and A. Winter. Distinguishability of quantum states under restricted families of measurements with an application to quantum data hiding. *Commun. Math. Phys.*, 291(3):813–843, 2009.

[56] S. Mendelson. Learning without Concentration. *ArXiv:1401.0304*, jan 2014.

[57] R. Millane. Phase retrieval in crystallography and optics. *JOSA A*, 7(3):394–411, 1990.

[58] Y. Nakata, M. Koashi, and M. Murao. Generating a state t-design by diagonal quantum circuits. *New J. Phys.*, 16(5):053043, 2014.

[59] G. Nebe, E. Rains, and N. Sloane. The invariants of the Clifford groups. *Des. Codes Cryptography*, 24:99–121, 2001.

[60] P. Netrapalli, P. Jain, and S. Sanghavi. Phase retrieval using alternating minimization. In *Advances in Neural Information Processing Systems*, pages 2796–2804, 2013.

[61] M. A. Nielsen and I. L. Chuang. *Quantum computation and quantum information.* Cambridge university press, 2010.

[62] N. Parikh and S. Boyd. Proximal algorithms. *Foundations and Trends in Optimization*, 1(3):123–231, 2013.

[63] B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev.*, 52:471–501, 2010.

[64] C. Schwemmer, G. Tóth, A. Niggebaum, T. Moroder, D. Gross, O. Gühne, and H. Weinfurter. Experimental comparison of efficient tomography schemes for a six-qubit state. *Phys. Rev. Lett.*, 113(4):040503, 2014.

[65] A. Scott. Tight informationally complete quantum measurements. *J. Phys. A-Math. Gen.*, 39:13507–13530, 2006.

[66] P. Seymour and T. Zaslavsky. Averaging sets: A generalization of mean values and spherical designs. *Adv. Math.*, 52:213–240, 1984.

[67] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev. Phase Retrieval with Application to Optical Imaging. *Preprint*, feb 2014. arXiv:1402.7350.

[68] V. Sidelnikov. Spherical 7-designs in 2^n-dimensional Euclidean space. *J. Algebr. Comb.*, 10:279–288, 1999.

[69] R. Stanley. *Enumerative Combinatorics, Volume I.* Cambridge University Press, 1997.

[70] J. Tanner and K. Wei. Normalized iterative hard thresholding for matrix completion. *SIAM J. Sci. Comput.*, 59(11):7491–7508, 2013.

[71] K. Toh and S. Yun. An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems. *Pac. J. Optim.*, 6:615–640, 2010.

[72] J. A. Tropp. User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.*, 12(4):389–434, 2012.

[73] J. A. Tropp. User friendly tools for random matrices. An introduction. *Preprint*, 2012.

[74] J. A. Tropp. Convex recovery of a structured signal from independent random linear measurements. *ArXiv:1405.1102*, 2014.

[75] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. In Y. Eldar and G. Kutyniok, editors, *Compressed Sensing: Theory and Applications*, pages 210–268. Cambridge Univ Press, 2012.

[76] A. Walther. The question of phase retrieval in optics. *Journal of Modern Optics*, 10(1):41–49, 1963.

[77] J. Watrous. Theory of quantum information. lecture notes, 2011.

[78] G. A. Watson. Characterization of the subdifferential of some matrix norms. *Linear Algebra Appl.*, 170:33–45, 1992.

[79] G. Zauner. *Quantendesigns: Grundzüge einer nichtkommutativen Designtheorie.* PhD thesis, University of Vienna, 1999.

Institute for Physics, University of Freiburg, Rheinstrasse 10, 79104 Freiburg, Germany
E-mail address: richard.kueng@physik.uni-freiburg.de

Lehrstuhl C für Mathematik (Analysis), RWTH Aachen University, Pontdriesch 10, 52062 Aachen, Germany
E-mail address: rauhut@mathc.rwth-aachen.de

Lehrstuhl C für Mathematik (Analysis), RWTH Aachen University, Pontdriesch 10, 52062 Aachen, Germany
E-mail address: terstiege@mathc.rwth-aachen.de