# A Partial Derandomization of PhaseLift using Spherical Designs

D. Gross[1], F. Krahmer[2], R. Kueng[*1]

[1]Institute for Physics, University of Freiburg, Rheinstraße 10, 79104 Freiburg, Germany
[2]Institute for Numerical and Applied Mathematics, University of Göttingen, Lotzestraße 16-18, 37083 Göttingen, Germany

October 9, 2013

ABSTRACT. The problem of retrieving phase information from amplitude measurements alone has appeared in many scientific disciplines over the last century. *PhaseLift* is a recently introduced algorithm for phase recovery that is computationally efficient, numerically stable, and comes with rigorous performance guarantees. PhaseLift is optimal in the sense that the number of amplitude measurements required for phase reconstruction scales linearly with the dimension of the signal. However, it specifically demands Gaussian random measurement vectors — a limitation that restricts practical utility and obscures the specific properties of measurement ensembles that enable phase retrieval. Here we present a partial derandomization of PhaseLift that only requires sampling from certain polynomial size vector configurations, called *t-designs*. Such configurations have been studied in algebraic combinatorics, coding theory, and quantum information. We prove reconstruction guarantees for a number of measurements that depends on the degree $t$ of the design. If the degree is allowed to to grow logarithmically with the dimension, the bounds become tight up to polylog-factors. Beyond the specific case of PhaseLift, this work highlights the utility of spherical designs for the derandomization of data recovery schemes.

## 1. INTRODUCTION

In this work we are interested in the problem of recovering a complex signal (vector) $x \in \mathbb{C}^d$ from an *intensity* measurement $y_0 = \|x\|_{\ell_2}^2$ and *amplitude* measurements

$$y_i = |\langle a_i, x\rangle|^2 \quad i = 1, \dots, m,$$

where $a_1, \dots, a_m \in \mathbb{C}^d$ are sampling vectors. Problems of this type are abundant in many different areas of science, where capturing phase information is hard or even infeasible, but obtaining amplitudes is comparatively easy. Prominent examples for this case occur in X-ray cristallography, astronomy and diffraction imaging – see for example [1]. This inverse problem is called *phase retrieval* and has attracted considerable interest over the last decades.

It is by no means clear how many such amplitude measurements are necessary to allow for recovery. Thus from the very beginning, there have been a number of works regarding injectivity conditions for this problem in the context of the specific applications [2].

More recently this question has been studied in more abstract terms, asking for the minimal number of amplitude measurements of the form (1) – without imposing structural assumptions on the $a_i$'s – that are required to make the above map injective. In [3], the authors showed that in the real case ($x \in \mathbb{R}^d$), at least $2d - 1$ such measurements are necessary and generically sufficient to guarantee injectivity, while in the complex case a generic sample size of $m \geq 4d - 2$ suffices. Here generic is to be understood in the sense that the sets of measurements of such size which do not allow for recovery form an algebraic variety in the space of all frames. Also, the latter bound is close to optimal:

*Corresponding author: richard.kueng@physik.uni-freiburg.de

as shown in [4], it follows from the results derived in [5] that a sample size of $m \geq (4 + o(1))\, d$ is necessary (cf. [6]). However, finding the precise bound is still in open problem.

Balan et al. [7] consider the scenario of $\mathcal{O}(d^2)$ measurements, which form a complex projective 2-design (cf. Def. 3 below). They derive an explicit reconstruction formula for this setup based on the following observation well known in conic programming. Namely, the quadratic constraints on $x$ are linear in the outer product $xx^*$:

$$y_i = |\langle a_i, x\rangle|^2 = \operatorname{tr}\left((a_i a_i^*)(xx^*)\right). \tag{1}$$

This "lifts" the problem to matrix space of dimension $d^2$, where it becomes linear and can be explicitly solved to find the unique solution.

As we will show in Theorem 2, it is, without making additional assumptions on the 2-design, not possible to use as measurements a random subset of the design which is of size $o(d^2)$. In other words, for the measurement scenario described in [7], the quadratic scaling in $d$ is basically unavoidable.

To contrast these two extreme approaches, ref. [3] works with a number of measurements close to the absolute minimum, but there are no tractable reconstruction schemes provided, the question of numerical stability is not considered, and it is unclear whether non-generic measurements – i.e., vectors with additional structural properties – can be employed. On the other hand, the number of measurements in [7] is much larger, while the measurements are highly structured and there is an explicit reconstruction method. A number of recent works including this paper aim to balance between these two approaches, working with a number of measurements only slightly larger while having at least some of the desired properties mentioned above.

Ref. [8] introduces a reconstruction method called *polarization* that works for $\mathcal{O}(d \log d)$ measurements and can handle structured measurement vectors, including the *masked illumination* setup that appears in diffraction imaging [9], where the measurements are generated by the discrete Fourier transform preceded by a random diagonal matrix. For Gaussian measurements, the polarization approach has also shown to be stable with respect to measurement noise [8]. While simulations seem to suggest stability also for the derandomized masked illumination setup, a proof of stability is – to our knowledge – not available yet.

An alternative approach, which we will also follow in this paper, is the *PhaseLift* algorithm, which is based on the lifted formulation (1). The algorithm was introduced in [10] and reconstruction guarantees have been provided in [11, 12]. The central observation is that the matrix $xx^*$, while unknown, is certainly of rank one. This connects the phase retrievel problem with the young but already extensive field of *low-rank matrix recovery* [13, 14, 15, 16]. Over the past years, this research program has rigorously identified many instances in which low-rank matrices can be efficiently reconstructed from few linear measurements. The existing results on low-rank matrix recovery were not directly applicable to phase retrieval, because the measurement matrices $a_i a_i^*$ failed to be sufficiently *incoherent* in the sense of [14, 15] (the incoherence parameter captures the well-posedness of a low-rank recovery problem). For the case of Gaussian measurement vectors $a_i$, Candès, Strohmer, Voroninski and Li were able to circumvent this problem, providing problem-specific stable recovery guarantees [11, 12] for a number of measurements of optimal order $\mathcal{O}(d)$. For recovery, they use a convex relaxation of the rank minimization problem, which makes the reconstruction algorithm tractable.

It should be noted, however, that because of the significantly increased problem dimensions, PhaseLift is not as efficient as many phase retrieval algorithms developed over the last decades in the physics literature (such as [17]) and the optimization literature (for

example [18]). Recently there have been attempts to provide recovery guarantees for alternating minimization algorithms [19], which are somewhat closer to the algorithms used in practice, but these direction of research is only at its beginnings.

While the above mentioned recovery guarantees for PhaseLift address the issues of tractable reconstruction and stability with respect to noise, these results leave open the question of whether measurement systems with additional structure and less randomness still allow for guaranteed recovery. There are both practical and theoretical motivations for pursuing such generalizations: A practitioner may be constrained in the choice of measurements by the application at hand or reduce the amount of randomness required for implementation purposes. The most prominent example are again masked Fourier measurements, which appear as a natural model in diffraction imaging, but a lot of different scenarios imposing different structure are conceivable. From a theoretical point of view, the use of Gaussian vectors obscures the specific properties that make phase retrieval possible. As discussed in the following subsection, it is a common thread in randomized signal processing that results are first established for Gaussian measurements and later generalized to structured ensembles.

A different direction of research, which will not be pursued in this paper, is to ask how additional structural assumptions on the signal to be recovered, such as sparsity, can be incorporated into the theory. A general analysis based on the Gaussian width of how many measurements are needed to allow for stable recovery of a signal known to lie in a set $T \subset \mathbb{R}^d$ is provided in [20]. Notably the results allow for measurements with arbitrary subgaussian rather than just Gaussian entries. Efficient algorithms for recovery, however, are not provided. For the case of $s$-sparse signals, also tractable recovery algorithms are available: It has been shown in [21] that PhaseLift can recover $x$ with high probability from Gaussian measurements for a number of measurements $m$ proportional to $s^2$ (up to logarithmic factors), which, for small $s$, can be considerably less than the dimension. In [22], it is shown that only a number of subgaussian measurements scaling linearly in the sparsity (up to logarithmic factors) is needed if recovery proceeds using certain greedy algorithms.

1.1. **Designs as a general-purpose tool for de-randomization.** In this paper, we focus on the theoretical aspect: which properties of a measurements are sufficient for PhaseLift to succeed? We prove recovery guarantees for ensembles of measurement vectors drawn at random from a finite set whose first $2t$ moments agree with those of Haar-random vectors (or, essentially, Gaussian vectors). A configuration of finite vectors which gives rise to such an ensemble is known as a *complex projective $t$-design*[2]. Designs were introduced by Delsarte, Goethals and Seidel in a seminal paper [23] and have been studied in algebraic combinatorics [24], coding theory [23, 25], and recently in quantum information theory [26, 27, 28, 29, 30]. Furthermore, complex projective 2-designs were the key ingredient for the reconstruction formula for phase retrieval proposed in [7].

One may see a more general philosophy behind this approach. In the field of sparse and low-rank reconstruction, a number of recovery results had first been established for Gaussian measurements. In subsequent works, it has then been proven that measurements drawn at random from certain fixed orthonormal bases are actually sufficient. Examples include uniform recovery guarantees for compressed sensing ([31, 32] vs. [33, 34]) and

[2] The definition of a $t$-design varies between authors. In particular, what is called a $t$-design here (and in most of the physics literature), would sometime be referred to as a $2t$ or even a $(2t+1)$-design. See Section 2.3 for our precise definition.

low-rank matrix recovery ([13] vs. [16]), respectively. Typically, the de-randomized proofs require much higher technical efforts and deliver slightly weaker results.

As the number of measurements needed for phase retrieval is larger than the signal space dimension, one cannot expect these results to exactly carry over to the phase retrieval setting. Nevertheless, the question remains whether there is a larger, but preferably not too large, set such that measurements drawn from it uniformly at random allow for phase retrieval reconstruction guarantees. In some sense, the sampling scenario we seek can be interpreted as an interpolation between the maximally random setup of Gaussian measurement with an optimal order of measurements and the construction in [7], which is completely deterministic, but suboptimal in terms of the embedding dimension. While in this paper, we will focus on the phase retrieval problem, we remark that such an interpolating approach between measurements drawn from a basis and maximally random measurements may also be of interest in other situations where constructions from bases are known, but lead to somewhat suboptimal embedding dimensions.

The concept of $t$-designs, as defined in Section 2.3, provides such an interpolation. The intuition behind that definition is that with growing $t$, more and more moments of the random vector corresponding to a random selection from the $t$-design agree with the Haar measure on the unit sphere. In that sense, as $t$ scales up further, $t$-designs give better and better approximations to Haar-random vectors.

The utility of this concept as a general-purpose de-randomization tool for Hilbert-space valued random construtions has been appreciated for example in quantum information theory [27, 35]. It has been compared [27] to the notion of *$t$-wise independence*, which plays a role for example in the analysis of discrete randomized algorithms [36], seems to have been long appreciated in coding theory. The smallest $t$-design in $\mathbb{C}^d$ consists of $\mathcal{O}(d^{2t})$ elements. Thus, whenever that lower bound is met, drawing a single element from a design requires $2t \log d$ bits, as opposed to $2d$ bits for a complex Bernoulli vector – an exponential gap.

From a practical point of view, the usefulness of these concepts hinges on the availability of constructions for designs. Explicit constructions for any order $t$ and any dimension $d$ are known [28, 37, 38, 39] – however, they are typically "inefficient" in the sense that they require a vector set of exponential size. For example, the construction in [28] uses $\mathcal{O}(t)^d$ vectors which is exponential in the dimension $d$.

Tighter analytic expressions for *exact* designs are notoriously difficult to find. Designs of degree 2 are widely known [40, 41, 42, 43]. A concrete example is used for the converse bound in Section 6 (as well as for the converse bounds for low-rank matrix recovery from Fourier-type bases in [15]). For degree 3, both real[3] [24] and complex [44] designs are known. For higher $t$, there are numerical methods based on the notion of the *frame potential* [45, 43, 44], non-constructive existence proofs [39], and constructions in sporadic dimensions (c.f. [46] and references thererin).

Importantly, almost-tight randomized constructions for *approximate designs* for arbitrary degrees and dimensions are known [27, 28, 30]. The simplest results [28] show that collections of Haar-random vectors form approximate $t$-designs. This indeed can reduce randomness: One only needs to expend a considerable amount of randomness *once* to generated a design – for subsequent applications it is sufficient to sample small subsets from it[4]. Going further, there have been recent deep results on designs obtained from certain

[3] While stated only for dimensions that are a power of 2, the results can be used for construtions in arbitrary dimensions [44].

[4] The situation is comparable to the use of random graphs as randomness expanders [47].

structured ensembles [30]. We do not describe the details here, as they are geared toward quantum problems and may have to be substantially modified to be applicable to the phase retrivial. The only connection to phase retrieval to date is the estimation of pure quantum states [4, 48].

1.2. **Main results.** In this paper, we show that spherical designs can indeed be used to partially derandomize recovery guarantees for underdetermined estimation problems; we generalize the recovery guarantee in [11] to measurements drawn uniformly at random from complex projective designs, at the cost of a slightly higher number of measurements.

**Theorem 1** (Main Theorem)**.** *Let $x \in \mathbb{C}^d$ be the unknown signal. Suppose that $\|x\|_{\ell_2}^2$ is known and that $m$ measurement vectors $a_1, \dots, a_m$ have been sampled independently and uniformly at random from a $t$-design $D_t \subset \mathbb{C}^d$ ($t \geq 3$). Then, with probability at least $1 - \mathrm{e}^{-\omega}$, PhaseLift (the convex optimization problem (25) below) recovers $x$ up to a global phase, provided that the sampling rate exceeds*

$$m \geq \omega \, C t \, d^{1+2/t} \log^2 d. \tag{2}$$

*Here $\omega \geq 1$ is an arbitrary parameter and $C$ is a universal constant.*

As the discussion of the previous subsection suggests, the bounds on the sampling rate decrease as the order of the design increases. For fixed $t$, and up to poly-log factors, it is proportional to $\mathcal{O}(d^{1+2/t})$. This is sub-quadratic for the regime $t \geq 3$ where our arguments apply. If the degree is allowed to grow logarithmialy with the dimension (as $t = 2 \log d$), we recover an optimal, linear scaling up to a polylog overhead, $m = \mathcal{O}(d \, \log^3 d)$.

The constant $C$ can be bounded by $9394$, but we believe this large size to be an artifact of our proof, as we have made no attempt to optimize it. The interested reader will see that mild assumptions on the size of $d$ already admit a much smaller constant.

In light of the highly structured, analytical and exact designs known for degree 2 and 3, it is of great interest to ask whether a linear scaling can already be achieved for some small, fixed $t$. As shown by the following theorem, however, for $t = 2$ not even a subquadratic scaling is possible if no additional assumptions are made, irrespective of the reconstruction algorithm used.

**Theorem 2** (Converse bound)**.** *Let $d$ be a prime power. Then there exists a 2-design $D_2 \subset \mathbb{C}^d$ and orthogonal, normalized vectors $x, z \in \mathbb{C}^d$ which have the following property.*

*Suppose that $m$ measurement vectors $y_1, \dots, y_m$ are sampled independently and uniformly at random from $D_2$. Then, for any $\omega \geq 0$, the number of measurements must obey*

$$m \geq \frac{\omega}{4} d(d+1),$$

*or the event*

$$|\langle a_i, x \rangle|^2 = |\langle a_i, z \rangle|^2 \quad \forall \, i \in \{1, \dots, m\}$$

*will occur with probability at least $\mathrm{e}^{-\omega}$.*

1.3. **Outlook.** There are a number of problems left open by our analysis. First, recall that our results achieve linear scaling up to logarithmic factors only when samples are drawn from a set of superpolynomial size. Thus it would be very interesting to find out whether there are polynomial size sets such that sampling from them achieves such a scaling, in particular, if $t$-designs for some fixed $t$ can be used. The case of $t = 3$ seems particularly important in that regard, since the converse bound (Theorem 2) shows that a design order of at least 3 is necessary. Also, highly structued 3-designs are known to exist (see above).

Another important follow-up problem concerns approximate $t$-designs. While our main result is phrased for exact $t$-designs, certain scenarios will only exhibit approximate design properties. We expect that our proofs can be generalized to such a setup, but also leave this problem for future work. Lastly, the reconstruction quality for noisy measurements is also an important issue yet to be investigated.

## 2. Technical Background and Notation

2.1. **Vectors, Matrices and matrix valued Operators.** In this work we require three different objects of linear algebra: vectors, matrices and operators acting on matrices.

We will work with vectors in a $d$-dimensional complex Hilbert space $V^d$ equipped with an inner product $\langle \cdot, \cdot \rangle$. We refer to the associated induced norm by

$$\|z\|_{\ell_2} = \sqrt{\langle z, z \rangle} \quad \forall z \in V^d.$$

We will denote such vectors by latin characters. For $z \in V^d$, we define the dual vector $z^* \in (V^d)^*$ via

$$z^* y = \langle z, y \rangle \quad \forall y \in V^d.$$

On the level of matrices we will exclusively consider $d \times d$ dimensional hermitian matrices, which we denote by capital latin characters. Endowed with the Hilbert-Schmitt (or Frobenius) scalar product

$$(Z, Y) = \mathrm{tr}(ZY), \tag{3}$$

the space $H^d$ becomes a Hilbert space. In addition to that, we will require the 3 different Schatten-norms

$$\begin{aligned} \|Z\|_1 &= \mathrm{tr}(|Z|) \quad \text{(trace norm)}, \\ \|Z\|_2 &= \sqrt{\mathrm{tr}(Z^2)} \quad \text{(Frobenius norm)}, \\ \|Z\|_\infty &= \sup_{y \in V^d} \frac{\|Zy\|_{\ell_2}}{\|y\|_{\ell_2}} \quad \text{(operator norm)}, \end{aligned}$$

where the second one is induced by the scalar product (3). These three norms are related via the inequalities

$$\|Z\|_2 \leq \|Z\|_1 \leq \sqrt{d}\|Z\|_2 \quad \text{and} \quad \|Z\|_\infty \leq \|Z\|_2 \leq \sqrt{d}\|Z\|_\infty \quad \forall Z \in H^d.$$

We call a hermitian matrix $Z$ positive-semidefinite ($Z \geq 0$), if $\langle y, Zy \rangle \geq 0$ for all $y \in V^d$. Positive semidefinite matrices form a cone [49] (Chapter II,12), which induces a partial ordering of matrices. Concretely, for $Z, Y \in H^d$ we write $Y \geq Z$ if $Y - Z$ is positive-semidefinite ($Y - Z \geq 0$).

In this work, the identity matrix $\mathbb{1}$ and rank-1 projectors are of particular importance. They are positive semidefinite and any matrix of the latter kind can be decomposed as $Z = zz^*$ for some $z \in V^d$. Up to a global phase, they correspond to vectors $z \in V^d$. The most important cases are the projection onto the unknown signal $x$ and onto the $i$th measurement vector $a_i$ respectively. They will be denoted by

$$X = xx^* \quad \text{and} \quad A_i = a_i a_i^*.$$

Finally, we will frequently encouter *matrix-valued operators* acting on the space $H^d$. We label such objects with capital caligraphic letters and introduce the operator norm

$$\|\mathcal{M}\|_{\mathrm{op}} = \sup_{Z \in H^d} \frac{\|\mathcal{M}Z\|_2}{\|Z\|_2}$$

induced by the Frobenius norm on $H^d$. It turns out that only very few matrix-valued operators will appear below. These are: the identity map

$$\begin{aligned} \mathcal{I} : H^d &\to H^d \\ Z &\mapsto Z \quad \forall Z \in H^d \end{aligned}$$

and (scalar multiples of) projectors onto some matrix $Y \in H^d$. The latter corresponds to

$$\begin{aligned} \Pi_Y : H^d &\to H^d \\ Z &\mapsto Y(Y, Z) = Y \operatorname{tr}(YZ) \quad \forall Z \in H^d. \end{aligned}$$

The operator

$$\Pi_{\mathbb{1}} : Z \mapsto \mathbb{1} \operatorname{tr}(\mathbb{1}Z) = \mathbb{1} \operatorname{tr}(Z) \quad \forall Z \in H^d,$$

is a very important example for this subclass of operators. Note that it is not a normalized projection, but $\frac{1}{d}\Pi_{\mathbb{1}}$ is. Indeed, for $Z \in H^d$ arbitrary

$$\left(d^{-1}\Pi_{\mathbb{1}}\right)^2 Z = d^{-2}\mathbb{1} \operatorname{tr}(\mathbb{1}\Pi_{\mathbb{1}} Z) = d^{-2} \operatorname{tr}(\mathbb{1})\mathbb{1} \operatorname{tr}(Z) = d^{-1}\Pi_{\mathbb{1}} Z. \tag{4}$$

The notion of positive-semidefiniteness directly translates to matrix valued operators. Concretely, we call $\mathcal{M}$ positive-semidefinite ($\mathcal{M} \geq 0$) if $(Z, \mathcal{M}Z) \geq 0$ for all $Z \in H^d$. Again, this induces a partial ordering. Like in the matrix case, we write $\mathcal{N} \geq \mathcal{M}$, if $\mathcal{N} - \mathcal{M} \geq 0$. It is easy to check that all the operators introduced so far are positive semidefinite and in particular we obtain the ordering

$$0 \leq \Pi_{\mathbb{1}} \leq d\mathcal{I}, \tag{5}$$

by using (4).

2.2. **Multilinear Algebra.** The properties of $t$-designs are most naturally stated in the framework of ($t$-fold) tensor product spaces. This motivates recapitulating some basic concepts of multilinear algebra that are going to greatly simplify our analysis later on. The concepts presented here are standard and can be found in any textbook on multilinear algebra. Our presentation has been influenced in particular by [50, 51].

Let $V_1, \ldots, V_k$ be (finite dimensional, complex) vector spaces, and let $V_1^*, \ldots, V_k^*$ be their dual spaces. A function

$$f : V_1 \times \cdots \times V_k \to \mathbb{C}$$

is *multilinear*, if it is linear in each $V_i$, $i = 1, \ldots, k$. 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, the tensor product $\left(V^d\right)^{\otimes k} = \bigotimes_{i=1}^k V^d$ is the space of all multilinear functions

$$f : \underbrace{\left(V^d\right)^* \times \cdots \times \left(V^d\right)^*}_{k \text{ times}} \mapsto \mathbb{C}, \tag{6}$$

and we call the elementary elements $z_1 \otimes \cdots \otimes z_k$ the *tensor product* of the vectors $z_1, \ldots, z_k \in V^d$. Such an element can alternatively be defined more concretely via the *Kronecker product* of the individual vectors. However, such a construction requires an explicit choice of basis in $V^d$ which is not the case in (6).

With this notation, the space of linear maps $V^d \to V^d$ ($d \times d$-matrices) corresponds to the tensor product $M^d := V^d \otimes \left(V^d\right)^*$ which is spanned by $\left\{y \otimes z^* : \ y, z \in V^d\right\}$ –

the set of all rank-1 matrices. For this generating set of $M^d$, we define the *trace* to be the natural bilinear map

$$\begin{aligned} \mathrm{tr} : V^d \otimes \left(V^d\right)^* &\to \mathbb{C} \\ (y \otimes z^*) &\mapsto z^* y = \langle z, y \rangle \end{aligned}$$

for all $y, z \in V^d$. The familiar notion of trace is obtained by extending this definition linearly to $M^d$.

Using $M^d = V^d \otimes \left(V^d\right)^*$ allows us to define the (matrix) tensor product $\left(M^d\right)^{\otimes k}$ to be the space of all multilinear functions

$$f : \underbrace{\left(\left(V^d\right)^* \times V^d\right) \times \cdots \times \left(\left(V^d\right)^* \times V^d\right)}_{k \text{ times}} \to \mathbb{C}$$

in complete analogy to the above. We call the elements $Z_1 \otimes \cdots \otimes Z_k$ the tensor product of the matrices $Z_1, \cdots, Z_k \in M^d$.

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

$$\begin{aligned} \mathrm{tr}_i : \left(M^d\right)^{\otimes k} &\to \left(M^d\right)^{\otimes (k-1)} \\ Z_1 \otimes \cdots \otimes Z_k &\mapsto \mathrm{tr}(Z_i) \left(Z_1 \otimes \cdots \otimes Z_{i-1} \otimes Z_{i+1} \otimes \cdots \otimes Z_k\right). \end{aligned}$$

Note that with the identification $M^d = V^d \otimes \left(V^d\right)^*$, $\mathrm{tr}_i$ corresponds to the natural contraction at position $i$. The partial trace over more than one system can be obtained by concatenating individual traces of this form, e.g. for $1 \leq i < j \leq k$

$$\mathrm{tr}_{i,j} := \mathrm{tr}_i \circ \mathrm{tr}_j : \left(M^d\right)^{\otimes k} \to \left(M^d\right)^{\otimes (k-2)}.$$

In particular, the *full trace* then corresponds to

$$\begin{aligned} \mathrm{tr} := \mathrm{tr}_{1,\ldots,k} : \left(M^d\right)^{\otimes k} &\to \mathbb{C} \\ (Z_1 \otimes \cdots \otimes Z_k) &\mapsto \mathrm{tr}(Z_1) \ldots \mathrm{tr}(Z_k). \end{aligned}$$

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

$$P_{\mathrm{Sym}^k} \left(z_1 \otimes \cdots \otimes z_k\right) := \frac{1}{k!} \sum_{\pi \in S_k} z_{\pi(1)} \otimes \cdots \otimes z_{\pi(k)}, \tag{7}$$

where $S_k$ denotes the group of permutations of $k$ elements. This map projects $\left(V^d\right)^{\otimes k}$ onto the totally symmetric subspace $\mathrm{Sym}^k$ of $\left(V^d\right)^{\otimes k}$ whose dimension [50] is

$$\dim \mathrm{Sym}^k = \binom{d+k-1}{k}. \tag{8}$$

2.3. **Complex projective designs.** The idea of (real) spherical designs originates in coding theory [23] and has been extended to more general spaces in [52, 53, 54]. We refer the interested reader to Levenshtein [54] for a unified treatment of designs in general metric spaces and from now on focus on designs in the complex vector space $V^d$.

Roughly speaking, a complex projective $t$-design is a finite subset of the complex unit sphere in $V^d$ with the property that the discrete average of any polynomial of degree $t$ or less equals its uniform average. Many equivalent definitions – see e.g. [52, 53, 42] – capture this essence. However, there is a more explicit definition of a $t$-design that is much more suitable for our purpose:

**Definition 3** (Definition 2 in [26]). *A finite set $\{w_1, \dots, w_N\} \subset V^d$ of normalized vectors is called a* $t$-design *of dimension $d$ if and only if*

$$\frac{1}{N}\sum_{i=1}^{N}(w_i w_i^*)^{\otimes t} = \dim(\mathrm{Sym}^t)^{-1} P_{\mathrm{Sym}^t}, \tag{9}$$

*where $P_{\mathrm{Sym}^t}$ denotes the projector onto the totally symmetric subspace (7) of $(V^d)^{\otimes t}$ and consequently* $\dim(\mathrm{Sym}^t) = \binom{d+t-1}{t}$.

Note that the defining property (9) is invariant under global phase changes $w_i \mapsto \mathrm{e}^{i\phi} w_i$, thus it matches the symmetry of the phase retrieval problem. The definition above is equivalent to demanding

$$\frac{1}{N}\sum_{i=1}^{N}(w_i w_i^*)^{\otimes t} = \int_w \mathrm{d}w\, (ww^*)^{\otimes t},$$

where the right hand side is integrated with respect to the Haar measure. This form makes the statement that $t$-designs mimic the first $2t$ moments of Haar measure more explicit.

P. Seymor and T. Zaslavsky proved in [39] that $t$-designs on $V^d$ exist for every $t, d \geq 1$, provided that $N$ is large enough ($N \geq N(d,t)$), but they do not give an explicit construction. A necessary criterion – cf. [53, 42] – for the $t$-design property is that the number of vectors $N$ obeys

$$N \geq \binom{d+\lceil t/2\rceil - 1}{\lceil t/2\rceil}\binom{d+\lfloor t/2\lfloor - 1}{\lfloor t/2\rfloor} = \mathcal{O}(d^{2t}). \tag{10}$$

However, the proof in [39] is non-constructive and known constructions are "inneficient" in the sense that the number of vectors required greatly exceeds (10). Hayashi et al. [28] proposed a construction requiring $\mathcal{O}(t)^d$ vectors. For real spherical designs other "inefficient" constructions have been proposed [37, 38] ($N = t^{\mathcal{O}(d^2)}$) which can be used to obtain complex projective designs.

Adressing this apparant lack of efficient constructions, Ambainis and Emerson [27] proposed the notion of *approximate desings*. These vector sets only fulfill property (9) only up to an $\epsilon$-precision, but their great advantage is that they can be constructed efficiently. Concretely, they show that for every $d \geq 2t$, there exists an $\epsilon = \mathcal{O}(d^{-1/3})$ approximate $t$-design consisting of $\mathcal{O}(d^{3t})$ vectors only.

The great value of $t$-designs is due to the following fact: If we sample $m$ vectors $a_i, \dots, a_m$ iid from a $t$-design $D_t = \{w_1, \dots, w_N\}$, the design property guarantees (with $A_i = a_i a_i^*$ and $W_i = w_i w_i^*$)

$$\mathbb{E}\left[\frac{1}{m}\sum_{i=1}^{m} A_i^{\otimes k}\right] = \mathbb{E}\left[A_1^{\otimes k}\right] = \frac{1}{N}\sum_{i=1}^{N} W_i^{\otimes k} = \binom{d+k-1}{k}^{-1} P_{\mathrm{Sym}^k}$$

for all $1 \leq k \leq t$. This knowledge about the first $t$ moments of the sampling procedure is the key ingredient for our partial derandomization of Gaussian PhaseLift [11].

2.4. **Large Deviation Bounds.** This approach makes heavy use of operator-valued large deviation bounds. They have been established first in the field of quantum information by Ahlswede and Winter [55]. Later the first author of this paper and his coworkers successfully applied these methods to the problem of low rank matrix recovery [15, 56]. By now these methods are widely used and we borrow them in their most recent (and convenient) form from Tropp [57, 58].

**Theorem 4** (Uniform Operator Bernstein inequality, [57, 15])**.** *Consider a finite sequence $\{M_k\}$ of independent, random self-adjoint operators. Assume that each random variable satisfies $\mathbb{E}[M_k] = 0$ and $\|M_k\|_\infty \leq \overline{R}$ (for some finite constant $\overline{R}$) almost surely and define the norm of the total variance $\sigma^2 := \|\sum_k \mathbb{E}\left[M_k^2\right]\|_\infty$. Then the following chain of inequalities holds for all $t \geq 0$.*

$$\Pr\left[\|\sum_k M_k\|_\infty \geq t\right] \leq d\ \exp\left(-\frac{t^2/2}{\sigma^2 + \overline{R}t/3}\right) \leq \begin{cases} d\ \exp(-3t^2/8\sigma^2) & t \leq \sigma^2/\overline{R} \\ d\ \exp(-3t/8\overline{R}) & t \geq \sigma^2/\overline{R}. \end{cases}$$

**Theorem 5** (Smallest Eigenvalue Bernstein Inequality, [58])**.** *Let $S = \sum_k M_k$ be a sum of iid random matrices $M_k$ which obey $\mathbb{E}[M_K] = 0$ and $\lambda_{min}(M_k) \geq -\underline{R}$ almost surely for some fixed $\underline{R}$. With the variance parameter $\sigma^2(S) = \|\sum_k \mathbb{E}\left[M_k^2\right]\|_\infty$ the following chain of inequalities holds for all $t \geq 0$.*

$$\Pr\left[\lambda_{\min}(S) \leq -t\right] \leq d\exp\left(-\frac{t^2/2}{\sigma^2 + \underline{R}t/3}\right) \leq \begin{cases} d\ \exp(-3t^2/8\sigma^2) & t \leq \sigma^2/\underline{R} \\ d\ \exp(-3t/8\underline{R}) & t \geq \sigma^2/\underline{R}. \end{cases}$$

2.5. **Wiring Diagrams.** The defining property (9) of $t$-designs is phrased in terms of tensor spaces. To work with these notions practically, we need tools for efficiently computing contractions between high-order tensors. The concept of *wiring diagrams* provides such a method – see [50] for an introduction and also [59, 60] (however, they use a slightly different notation). Here, we give a brief description that should suffice for our calculations.

Roughly, the calculus of wiring diagrams associates with every tensor a box, and with every index of that tensor a line emanating from the box. Two connected lines represent contracted indices. (More precisely, we place contravariant indices of a tensor on top of the associated box and covariant ones at the bottom. However, one should be able to digest our calculations without reference to this detail). A matrix $A : V^d \to V^d$ can be seen as a two-indexed tensor $A^i{}_j$. It will thus be represented by a node  with the upper line corresponding to the index $i$ and the lower one to $j$. Two matrices $A, B$ are multiplied by contracting $B$'s "contravariant" index with $A$'s "covariant" one:

$$(AB)^i{}_j = \sum_k A^i{}_k B^k{}_j$$

Pictographically, we write

$$AB = $$ 

The trace operation

$$A \mapsto \operatorname{tr} A = \sum_k A^k{}_k$$

corresponds to a contraction of the two indices of a matrix:

$$\operatorname{tr}(A) = $$  .

Tensor products are arranged in parallel:

$$A \otimes B = $$  .

Hence, a partial trace takes the following form:

$$\operatorname{tr}_2(A \otimes B) = $$  .

The last ingredient we need are the *transpositions* $\sigma_{(i,j)}$ on $(V^d)^{\otimes t}$ which act by interchanging the $i$th and the $j$th tensor factor. For example

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

with $x, y \in V^d$ arbitrary. Transpositions suffice, because they generate the full group of permutations. For $\left(V^d\right)^{\otimes 2}$ we only have

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

but for higher tensor systems more permutations can occur. Consequently, permutations act by interchanging different input and output lines and the wiring diagram representation allows one to keep track of this pictorially. In fact, only the input and output position of a line matters. We can use diagrams to simplify expressions by disentangling the corresponding lines. Take $\sigma_{(1,2)}$ on $\left(V^d\right)^{\otimes 2}$ as an example. Using wiring diagrams we can derive the standard result

$\sigma_{(1,2)}^2 =$  $=$  $= \underline{1}$

pictorially. We are now ready to prove some important auxiliary results.

**Lemma 6.** *Let $A, B \in H^d$ be arbitrary. Then it holds that*

$$\operatorname{tr}_2\left(P_{\mathrm{Sym}^2} A \otimes B\right) = \frac{1}{2}\left(\operatorname{tr}(B)A + BA\right). \tag{11}$$

We remark that in general,

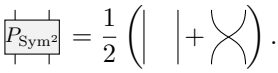
$$P_{\mathrm{Sym}^2}\left(X \otimes Y\right) \neq \frac{1}{2}\left(X \otimes Y + Y \otimes X\right),$$

which is, in our experience, a common misconception.

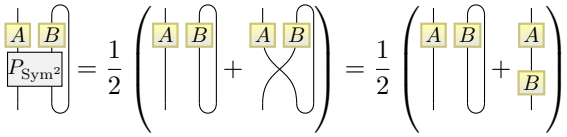
*Proof of Lemma 6.* The basic formula (7) for $P_{\mathrm{Sym}^2}$ is given by

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

and the concepts from above allow us to translate this into the following wiring diagram:



(Note that this operator acts on the full tensor space $\left(V^d\right)^{\otimes 2}$, hence in the wiring diagram it is represented by a two-indexed box.) Applying the graphical calculus yields

$\operatorname{tr}_2\left(P_{\mathrm{Sym}^2} A \otimes B\right) =$ 

$$= \frac{1}{2}\left(\operatorname{tr}(B)A + BA\right),$$

which is the desired result. □

Obviously, it is also possible to obtain (11) by direct calculation. We have included such a calculation in the appendix (Section 8.1) to demonstrate the complexity of direct calculations as compared to graphical ones.

We conclude this section with the following slightly more involved result.

**Lemma 7.** *Let $A, B, C \in H^d$ be arbitrary. Then it holds that*

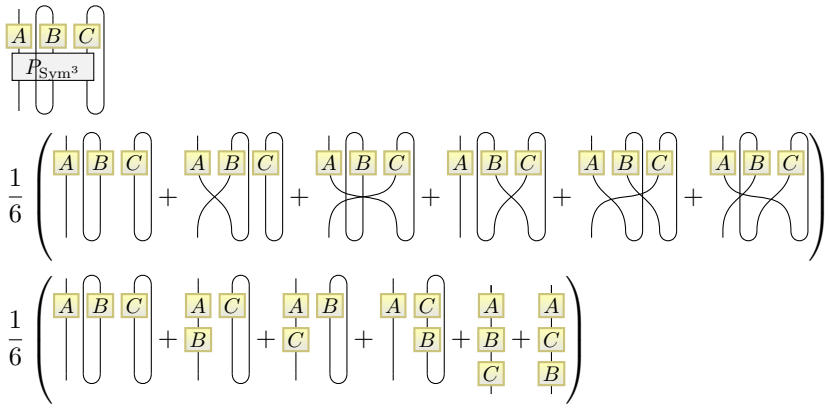
$$\begin{aligned} &\mathrm{tr}_{2,3}\left(P_{\mathrm{Sym}^3} A \otimes B \otimes C\right) \\ &= \frac{1}{6}\left(A\,\mathrm{tr}(B)\mathrm{tr}(C) + BA\,\mathrm{tr}(C) + CA\mathrm{tr}(B) + A\,\mathrm{tr}(BC) + CBA + BCA\right). \end{aligned} \tag{12}$$

The proof can in principle be obtained by evaluating all permutations of 3 tensor systems algebraically and taking the partial trace afterwards. However, a pictorial calculation using wiring diagrams is much faster and more elegant.

*Proof.* For permutations of three elements, formula (7) implies

$$P_{\mathrm{Sym}^3} = \frac{1}{6}\sum_{\pi\in S_3} \sigma_{\pi(1),\pi(2),\pi(3)} = \frac{1}{6}\left(\sigma_{1,2,3} + \sigma_{2,1,3} + \sigma_{3,2,1} + \sigma_{1,3,2} + \sigma_{2,3,1} + \sigma_{3,1,2}\right),$$

where. $\sigma_{2,1,3}(u \otimes v \otimes w) = (v \otimes u \otimes w)$, etc. This in turn allows us to write




$$= \frac{1}{6}\left(A\,\mathrm{tr}(B)\mathrm{tr}(C) + BA\,\mathrm{tr}(C) + CA\mathrm{tr}(B) + A\,\mathrm{tr}(BC) + CBA + BCA\right)$$

and we are done. □

## 3. Problem Setup

3.1. **Modelling the sampling process.** In the sampling process, we start by measuring the intensity of the signal:

$$y_0 = \|x\|_{\ell_2}^2 = \mathrm{tr}(\mathbb{1}X). \tag{13}$$

This allows us to assume w.l.o.g. $\|x\|_{\ell_2} = 1$. Next, we choose $m$ vectors $a_1, \ldots, a_m$ iid at random from a $t$-design $D_t \subset V^d$ and evaluate

$$y_i = \mathrm{tr}(A_i X) = |\langle x, a_i\rangle|^2 \quad \text{for } i = 1, \ldots m, \tag{14}$$

and consequently the vector $y = (y_1, \ldots, y_m)^T \in \mathbb{R}_+^m$ captures all the information we obtain from the sampling process. This process can be represented by a measurement operator

$$\begin{aligned} \mathcal{A} : H^d &\to \mathbb{R}^m, \\ Z &\mapsto \sum_{i=1}^m \mathrm{tr}(A_i Z) e_i, \end{aligned} \tag{15}$$

where $e_1, \dots, e_m$ denotes the standard basis of $\mathbb{R}^m$. Therefore $\mathcal{A}(X) = y$ completely encodes the measurement process. For technical reasons we also consider the measurement operator

$$\begin{aligned} \mathcal{R} : H^d &\to H^d, \\ Z &\mapsto m^{-1} \sum_{i=1}^m (d+1)d \, \Pi_{A_i} Z = m^{-1} \sum_{i=1}^m (d+1)d \, A_i \operatorname{tr}(A_i Z), \end{aligned} \tag{16}$$

which is a renormalized version of $\mathcal{A}^*\mathcal{A} : H^d \to H^d$. Concretely

$$\mathcal{R} = \frac{(d+1)d}{m} \mathcal{A}^*\mathcal{A}.$$

The scaling is going to greatly simplify our analysis, because it guarantees that $\mathcal{R}$ is "nearly isotropic", as the following result shows.

**Lemma 8** ($\mathcal{R}$ is nearly isotropic)**.** *The operator $\mathcal{R}$ defined in (16) is* near-isotropic *in the sense that*

$$\mathbb{E}[\mathcal{R}] = \mathcal{I} + \Pi_{\mathbb{1}}. \tag{17}$$

*Furthermore, setting $\mathcal{S} := \mathcal{I} - \frac{1}{d+1}\Pi_{\mathbb{1}}$, we have that*

$$\mathcal{S}\mathbb{E}[\mathcal{R}] = \mathbb{E}[\mathcal{R}]\mathcal{S} = \mathcal{I}. \tag{18}$$

Note that $\mathcal{S}$ is a contraction. Indeed, as $\frac{1}{d}\Pi_{\mathbb{1}}$ is a projection, it holds that $\frac{1}{d}\Pi_{\mathbb{1}} \leq \mathcal{I}$ and consequently

$$0 \leq \frac{1}{d+1}\mathcal{I} = \mathcal{I} - \frac{d}{d+1}\mathcal{I} \leq \mathcal{I} - \frac{1}{d+1}\Pi_{\mathbb{1}} = \mathcal{S} \leq \mathcal{I}.$$

This in turn implies $\operatorname{spec}(\mathcal{S}) \in [1/(d+1), 1]$.

*Proof of Lemma 8.* Let us start with deriving (17). For $Z \in H^d$ arbitrary we have

$$\begin{aligned} \mathbb{E}[\mathcal{R}]Z &= \frac{(d+1)d}{m} \sum_{i=1}^m \mathbb{E}[A_i \operatorname{tr}(A_i Z)] \\ &= (d+1)d \operatorname{tr}_2 \left( \mathbb{E}[A_1^{\otimes 2}] \mathbb{1} \otimes Z \right) \qquad (19) \\ &= 2 \operatorname{tr}_2 \left( P_{\mathrm{Sym}^2} \mathbb{1} \otimes Z \right) \qquad (20) \\ &= Z + \mathbb{1}(\operatorname{tr} Z) = \left( \mathcal{I} + \Pi_{\mathbb{1}} \right) Z. \end{aligned}$$

Here, (19) follows from the fact that the $a_i$'s are chosen iid from a $t$-design, (20) uses the fact that $\dim(\mathrm{Sym}^2) = \binom{d+1}{2}^{-1}$ together with Definition 3, and the final line is an application of Lemma11.

For the second claim, note that $\mathbb{E}[\mathcal{R}]$ and $\mathcal{S}$ commute and we get

$$\begin{aligned} \mathcal{S}\mathbb{E}[\mathcal{R}] &= \mathbb{E}[\mathcal{R}]\mathcal{S} = (\mathcal{I} + \Pi_{\mathbb{1}})(\mathcal{I} - \frac{1}{d+1}\Pi_{\mathbb{1}}) \\ &= \mathcal{I} + \Pi_{\mathbb{1}} - \frac{1}{d+1}\Pi_{\mathbb{1}} - \frac{1}{d+1}\Pi_{\mathbb{1}}^2 \\ &= \mathcal{I} + (1 - \frac{1}{d+1} - \frac{d}{d+1})\Pi_{\mathbb{1}} = \mathcal{I} \end{aligned}$$

as intended. Here we have used (5) ($\Pi_{\mathbb{1}}^2 = d\Pi_{\mathbb{1}}$). □

Let now $x \in V^d$ be the signal we want to recover. As in [11] we consider the space

$$T := \left\{ xz^* + zx^* : \ z \in V^d \right\} \subset H^d \tag{21}$$

(which is the tangent space of the manifold of all hermitian matrices at the point $X = xx^*$). This space is of crucial importance for our analysis. The orthogonal projection onto this space can be given explicitly:

$$\begin{aligned} \mathcal{P}_T : H^d &\to T, \\ Z &\mapsto XZ + ZX - XZX \qquad (22) \\ &= XZ + ZX - (X, Z)X. \qquad (23) \end{aligned}$$

We denote the projection onto its orthogonal complement with respect to the Frobenius inner product by $\mathcal{P}_T^\perp$. Then for any matrix $Z \in H^d$ the decomposition

$$Z = \mathcal{P}_T Z + \mathcal{P}_T^\perp Z =: Z_T + Z_T^\perp$$

is valid. We point out that in particular

$$\mathcal{P}_T \Pi_{\mathbb{1}} \mathcal{P}_T = \Pi_X \tag{24}$$

holds. We will frequently use this fact. For a proof, consider $Z \in H^d$ arbitrary and insert the relevant definitions:

$$\begin{aligned} \mathcal{P}_T \Pi_{\mathbb{1}} \mathcal{P}_T Z &= \mathcal{P}_T \mathbb{1} \operatorname{tr}(\mathbb{1} \mathcal{P}_T Z) = (X\mathbb{1} + \mathbb{1}X - X\mathbb{1}X) \operatorname{tr}(XZ + ZX - XZX) \\ &= X \operatorname{tr}(XZ) = \Pi_X Z. \end{aligned}$$

3.2. **Convex Relaxation.** Following [3, 11, 12] the measurements (13) and (14) can be translated into matrix form by applying the following "lifts":

$$X := xx^*, \quad \text{and} \quad A_i := a_i a_i^*.$$

By doing so the measurements assume the a linear form:

$$\begin{aligned} y_0 &= \|x\|_2^2 = (\mathbb{1}, X) = \operatorname{tr}(X), \\ y_i &= (A_i, X) = \operatorname{Tr}(A_i X) \quad i = 1, \ldots, m. \end{aligned}$$

Hence, the phase retrivial problem becomes a matrix recovery problem. The solution to this is guaranteed to have rank 1 and encodes (up to a global phase) the unknown vector $x$ via $X = xx^*$. Relaxing the rank minimization problem (which would output the correct solution) to a trace norm minimization yields the now-familiar convex optimization problem

$$\begin{aligned} &\text{minarg}_{X'} && \|X'\|_1 \\ &\text{subject to} && (A_i, X') = y_i \quad i = 1, \ldots m, \\ & && X' = (X')^\dagger, \\ & && \operatorname{tr}(X') = 1, \\ & && X' \geq 0. \end{aligned} \tag{25}$$

While this convex program is formally equivalent to the previously studied general-purpose matrix recovery algorithms [13, 14, 15], there are two important differences:

- The measurement matrices $A_i$ are rank-1 projectors: $A_i = a_i a_i^*$.
- The unknown signal is known to be proportional to a rank-1 projector ($X = xx^*$) as well.

While the second fact is clearly of advantage for us, the first one makes the problem considerably harder: In the language of [15], it means that the "incoherence parameter" $\mu = d \max_{i=1,\dots,m} \|A_i\|_\infty = d\|a_i\|_{\ell_2}^2 = d$ is as large as it can get! Higher values of $\mu$ correspond to more ill-posed problems and as a result, a direct application of previous low-rank matrix recovery results fails. It is this problem that Refs. [11, 12] first showed how to circumvent for the case of Gaussian measurements. Below, we will adapt these ideas to the case of measurements drawn from designs, which necessitates following more closely the approach of [15].

3.3. **Well-posedness / Injectivity.** In this section, we follow [11, 15] to establish a certain injectivity property of the measurement operator $\mathcal{A}$. Compared to [11], our injectivity properties are somewhat weaker. Their proof used the independence of the components of the Gaussian measurement operator, which is not available in this setting, where individual vector components might be strongly correlated. We will pay the price for these weaker bounds in Section 5. There, we construct an "approximate dual certificate" that proves that the sought-for signal indeed minimizes the nuclear norm. Owing to the weaker bounds found here, the construction is more complicated than in [11]. In the language of [15], we will have to carry out the full "golfing scheme", as opposed to the "single leg" that proved sufficient in [11].

**Proposition 9.** *With probability of failure smaller than* $d^2 \exp(-\frac{3m}{384d})$ *the inequality*

$$0.25 d^{-2} \|Z\|_2^2 < m^{-1} \|\mathcal{A}(Z)\|_2^2 \tag{26}$$

*is valid for all matrices* $Z \in T$ *simultaneously.*

*Proof.* We aim to show the more general statement

$$\Pr\left[m^{-1}\|\mathcal{A}(Z)\|_2^2 < 0.5(1-\delta)\|Z\|_2^2 \ \ \forall Z \in T\right] \leq d^2 \exp\left(-\frac{3m\delta^2}{96d}\right)$$

for any $\delta \in (0,1)$.

For $Z \in T$ abritrary use near-isotropicity of $\mathcal{R}$ ($\mathbb{E}[\mathcal{R}] = \mathcal{I} + \Pi_{\mathbb{1}}$) and observe

$$\begin{aligned}
& m^{-1}\|\mathcal{A}(Z)\|_2^2 \\
= \ & m^{-1} \sum_{i=1}^m \left(\operatorname{tr}(ZA_i)\right)^2 = \operatorname{tr}(Z m^{-1} \sum_i A_i \operatorname{tr}(A_i Z)) = \frac{1}{(d+1)d} \operatorname{tr}(Z\mathcal{R}Z) \\
= \ & \frac{1}{(d+1)d} \operatorname{tr}(Z(\mathcal{R} - \mathbb{E}[\mathcal{R}])Z) + \frac{1}{(d+1)d} \operatorname{tr}(Z(\mathcal{I} + \Pi_{\mathbb{1}})Z) \\
= \ & \frac{1}{(d+1)d} \operatorname{tr}(Z\mathcal{P}_T(\mathcal{R} - \mathbb{E}[\mathcal{R}])\mathcal{P}_T Z) + \frac{1}{(d+1)d}(\operatorname{tr}(Z^2) + (\operatorname{tr} Z)^2) \\
\geq \ & 0.5 d^{-2} \left(\operatorname{tr}(Z\mathcal{P}_T(\mathcal{R} - \mathbb{E}[\mathcal{R}])\mathcal{P}_T Z) + \operatorname{tr}(Z^2)\right) \\
\geq \ & 0.5 d^{-2} (1 + \lambda_{\min}\left(\mathcal{P}_T(\mathcal{R} - \mathbb{E}[\mathcal{R}])\mathcal{P}_T\right) \|Z\|_2^2, \qquad (27)
\end{aligned}$$

where we have used $\mathcal{P}_T Z = Z$ as well as $\mathcal{M} \geq \lambda_{\min}(\mathcal{M})\mathcal{I}$ for any operator $\mathcal{M}$. Therefore everything boils down to bounding the smallest eigenvalue of $\mathcal{P}_T(\mathcal{R} - \mathbb{E}[\mathcal{R}])\mathcal{P}_T$. To this end we aim to apply Theorem 5 and decompose

$$\mathcal{P}_T(\mathcal{R} - \mathbb{E}[\mathcal{R}])\mathcal{P}_T = \sum_{i=1}^m (\mathcal{M}_i - \mathbb{E}[\mathcal{M}_i]) \quad \text{with} \quad \mathcal{M}_i = \frac{(d+1)d}{m} \mathcal{P}_T \Pi_{A_i} \mathcal{P}_T.$$

Note that these summands have mean zero by construction. Furthermore observe that the auxiliary result (24) implies

$$\begin{aligned} -\frac{2}{m}\mathcal{I} &\leq -\frac{1}{m}\mathcal{I} - \frac{1}{m}\Pi_X \leq -\frac{1}{m}\mathcal{P}_T\mathcal{I}\mathcal{P}_T - \frac{1}{m}\mathcal{P}_T\Pi_{\mathbb{1}}\mathcal{P}_T \\ &= -\mathcal{P}_T\mathbb{E}[\mathcal{M}_i]\mathcal{P}_T \leq \mathcal{P}_T(\mathcal{M}_i - \mathbb{E}[\mathcal{M}_i])\mathcal{P}_T \end{aligned}$$

and the a priori bound

$$\lambda_{\min}(\mathcal{M}_i - \mathbb{E}[\mathcal{M}_i]) \geq -2/m =: -\underline{R}$$

follows. For the variance we use the standard identity

$$0 \leq \mathbb{E}[(\mathcal{M}_i - \mathbb{E}[\mathcal{M}_i])^2] = \mathbb{E}[\mathcal{M}_i^2] - \mathbb{E}[\mathcal{M}_i]^2 \leq \mathbb{E}[\mathcal{M}_i^2]$$

and focus on the last expression. Writing it out explicitly yields

$$\begin{aligned} 0 \leq \mathbb{E}[\mathcal{M}_i^2] &= \frac{(d+1)^2d^2}{m^2}\mathcal{P}_T\mathbb{E}\left[\Pi_{A_i}\mathcal{P}_T\Pi_{A_i}\right]\mathcal{P}_T \\ &= \frac{(d+1)^2d^2}{m^2}\mathcal{P}_T\mathbb{E}\left[\operatorname{tr}(A_i\mathcal{P}_TA_i)\Pi_{A_i}\right]\mathcal{P}_T. \end{aligned}$$

The trace can be bounded from above by

$$\begin{aligned} \operatorname{tr}(A_i\mathcal{P}_TA_i) &= \operatorname{tr}\left(A_i(XA_i + A_iX - \operatorname{tr}(A_iX)X)\right) \\ &= 2\operatorname{tr}(A_iX) - \operatorname{tr}(A_iX)^2 \leq 2\operatorname{tr}(A_iX), \end{aligned}$$

where we have used the basic definition of $\mathcal{P}_T$ and $0 \leq \operatorname{tr}(A_iX) = |\langle a_i, x\rangle|^2 \leq 1$. Consequently, for $Z \in T$ arbitrary

$$\begin{aligned} & \mathcal{P}_T\mathbb{E}[\mathcal{M}_i^2]\mathcal{P}_TZ \\ \leq\ & \frac{2(d+1)^2d^2}{m^2}\mathcal{P}_T\mathbb{E}\left[A_i\operatorname{tr}(A_iX)\operatorname{tr}(A_iZ)\right] \\ =\ & \frac{2(d+1)^2d^2}{m^2}\mathcal{P}_T\operatorname{tr}_{2,3}\left(\mathbb{E}[A_i^{\otimes 3}]\mathbb{1}\otimes X\otimes Z\right) \\ =\ & \frac{12(d+1)^2d^2}{m^2(d+2)(d+1)d}\mathcal{P}_T\operatorname{tr}_{2,3}\left(P_{\mathrm{Sym}^3}\mathbb{1}\otimes X\otimes Z\right) \\ \leq\ & \frac{2d}{m^2}\mathcal{P}_T\left(\mathbb{1}\operatorname{tr}(Z) + X\operatorname{tr}(Z) + Z + \mathbb{1}\operatorname{tr}(XZ) + ZX + XZ\right) \\ =\ & \frac{2d}{m^2}\left(X\operatorname{tr}(XZ) + X\operatorname{tr}(XZ) + Z + X\operatorname{tr}(XZ) + \mathcal{P}_TZ + X\operatorname{tr}(XZ)\right) \\ =\ & \frac{2d}{m^2}\left(4\Pi_X + 2\mathcal{I}\right)Z \leq \frac{12d}{m^2}\mathcal{I}Z. \end{aligned}$$

Here we have applied $\dim\mathrm{Sym}^3 = \binom{d+2}{3}^{-1}$ and Lemma 7 in lines 3 and 4, respectively. Furthermore we used $Z \in T$ – hence $\mathcal{P}_TZ = Z$ and $\operatorname{tr}(Z) = \operatorname{tr}(XZ)$ – as well as the basic definition (23) of $\mathcal{P}_T$ to simplify the terms occuring in the fourth line. Putting everything together yields

$$\mathbb{E}[(\mathcal{M}_i - \mathbb{E}[\mathcal{M}_i])^2] \leq \mathbb{E}[\mathcal{M}_i^2] \leq \frac{12d}{m^2}\mathcal{I}$$

and we can safely set $\sigma^2 := \frac{12d}{m}$. Now Theorem 5 tells us

$$\Pr\left[\lambda_{\min}\left(\mathcal{P}_T(\mathcal{R} - \mathbb{E}[\mathcal{R}])\mathcal{P}_T\right) \leq -\delta\right] \leq d^2\exp\left(-\frac{3m\delta^2}{8\times 12d}\right)$$

for all $0 \leq \delta \leq 1 \leq 6d = \sigma^2/\underline{R}$. This gives the desired bound on the event

$$\{\lambda_{\min}(\mathcal{P}_T(\mathcal{R} - \mathbb{E}[\mathcal{R}])\mathcal{P}_T) \leq -\delta\}$$

occuring. If this is not the case, (27) implies

$$m^{-1}\|\mathcal{A}(Z)\|_{\ell_2}^2 > 0.5d^{-2}(1-\delta)\|Z\|_2^2$$

for all matrices $Z \in T$ simultaneously. This is the general statement at the beginning of the proof and setting $\delta = 1/2$ yields Proposition 9. □

**Proposition 10.** *Let $\mathcal{A}$ be as above with vectors sampled from a $t$-design ($t \geq 1$). Then the statement*

$$m^{-1}\|\mathcal{A}(Z)\|_{\ell_2}^2 \leq \|Z\|_2^2 \tag{28}$$

*holds with probability one for all matrices $Z \in H^d$ simultaneously.*

*Proof.* Pick $Z \in H^d$ arbitrary and observe

$$\|\mathcal{A}(Z)\|_{\ell_2}^2 = \frac{1}{m}\sum_{i=1}^m \left(\operatorname{tr}(A_i Z)\right)^2 = \operatorname{tr}\left(Z\left(\frac{1}{m}\sum_{i=1}^m \Pi_{A_i}\right)Z\right) \leq \operatorname{tr}(Z\mathcal{I}Z) = \|Z\|_2^2,$$

where we have used $0 \leq \Pi_{A_i} \leq \mathcal{I}$. □

Note that equation (28) can be improved. Indeed, a standard application of the Operator Bernstein inequality (Theorem 4) gives

$$m^{-1}\|\mathcal{A}(Z)\|_{\ell_2}^2 \leq 2d^{-1}\|Z\|_2^2$$

for all matrices $Z \in T$ with probability of failure smaller than $d^2 \exp\left(-Cm/d\right)$ for some $0 < C \leq 1$. However, we actually do not require this tighter bound.

## 4. Proof of the Main Theorem / Convex Geometry

In this section, we will follow [15, 14] to prove that the convex program (25) indeed recovers the sought for signal $x$, provided that a certain geometric object – an *approximate dual certificate* – exists.

**Definition 11** (Approximate dual certificate)**.** *Assume that the sampling process corresponds to (13) and (14). Then we call $Y \in H^d$ an* approximate dual certificate*, provided that $Y \in \operatorname{span}\left(\mathbb{1}, A_1, \ldots, A_m\right)$ and*

$$\|Y_T - X\|_2 \leq \frac{1}{4d} \quad \textit{as well as} \quad \|Y_T^\perp\|_\infty \leq \frac{1}{2}. \tag{29}$$

**Proposition 12.** *Suppose that the measurement gives us access to $\|x\|_{\ell_2}^2$ and $y_i = |\langle a_i, x\rangle|^2$ for $i = 1, \ldots, m$. Then the convex optimization (25) recovers the unknown $x$ (up to a global phase) provided that (26) holds and an approximate dual certificate $Y$ exists.*

*Proof.* Let $\tilde{X} \in H^d$ be an arbitrary feasible point of (25) and decompose it as $\tilde{X} = X + \Delta$. Feasibility then implies $\mathcal{A}(\tilde{X}) = \mathcal{A}(X)$ and $\mathcal{A}(\Delta) = 0$ must in turn hold for any feasible displacement $\Delta$. Now the pinching inequality [61] (Problem II.5.4) implies

$$\|\tilde{X}\|_1 = \|X + \Delta\|_1 \geq \|X\|_1 + \operatorname{tr}(\Delta_T) + \|\Delta_T^\perp\|_1.$$

Consequently $X$ is guaranteed to be the unique minimum of (25), if

$$\operatorname{tr}(\Delta_T) + \|\Delta_T^\perp\|_1 > 0 \tag{30}$$

is true for every feasible $\Delta$. In order to show this we combine feasibility of $\Delta$ with inequalities (26) and (28) to obtain

$$\|\Delta_T\|_2 < 2dm^{-1/2}\|\mathcal{A}(\Delta_T)\|_{\ell_2} = 2dm^{-1/2}\|\mathcal{A}(\Delta_T^\perp)\|_{\ell_2} \leq 2d\|\Delta_T^\perp\|_2. \tag{31}$$

Feasibility of $\Delta$ also implies $(Y, \Delta) = 0$, because by defnition $Y$ is in the range of $\mathcal{A}^*$. Combining this insight with the defining property (29) of $Y$ and (31) yields

$$\begin{aligned} 0 &= (Y, \Delta) = (Y_T - X, \Delta_T) + (X, \Delta_T) + (Y_T^\perp, \Delta_T^\perp) \\ &\leq \|Y_T - X\|_2\|\Delta_T\|_2 + \operatorname{tr}(\Delta_T) + \|Y_T^\perp\|_\infty\|\Delta_T^\perp\|_1 \\ &< \operatorname{tr}(\Delta_T) + \|Y_T - X\|_2 2d\|\Delta_T^\perp\|_2 + \|Y_T^\perp\|_\infty\|\Delta_T^\perp\|_1 r \\ &\leq \operatorname{tr}(\Delta_T) + 1/2\|\Delta_T^\perp\|_2 + 1/2\|\Delta_T^\perp\|_1 \\ &\leq \operatorname{tr}(\Delta_T) + \|\Delta_T^\perp\|_1, \end{aligned}$$

which is just the desired optimality criterion (30). □

## 5. Constructing the Dual Certificate

A straightforward approach to construct an approximate dual certificate would be to set

$$Y = \mathcal{S}\mathcal{R}X = \frac{(d+1)d}{m}\sum_{i=1}^{m}\mathcal{S}A_i \operatorname{tr}(A_iX) \in \operatorname{span}\left(\mathbb{1}, A_1, \ldots, A_m\right). \tag{32}$$

In expectation, $\mathbb{E}[Y] = X$, which is the "perfect dual certificate" in the sense that the norm bounds in (29) vanish. The hope would be to use the Operator Bernstein inequality to show that with high probablity, $Y$ will be sufficiently close to its expectation. It has been shown that a slight refinement of the ansatz (32) indeed achieves this goal Ref. [15, 62]. However, the Bernstein bounds depend on the worst-case operator norm of the summands. In our case, they can be as large as $d^2|\langle a_i, x\rangle|^2$, which can reach $d^2$. This is far larger than in previous low-rank matrix recovery problems. Ref. [11] relied on the fact that large overlaps $|\langle a_i, x\rangle|^2 \gg \mathcal{O}(d^{-1})$ are "rare" for Gaussian $a_i$.

The key observation here is that the $t$-design property provides one with useful information about the first $t$ moments of the random variable $|\langle x, a_i\rangle|^2$. This knowledge allows us to explicitly bound the probability of "dangerously large overlaps" or "coherent measurement vectors" occurring.

**Lemma 13** (Undesired events)**.** *Let $x \in V^d$ be an arbitrary vector of unit length. If $a$ is chosen uniformly at random from a $t$-design ($t \geq 1$) $D_t \subset V^d$, then the following is true for every $\gamma \leq 1$:*

$$\Pr\left[|\langle a, x\rangle|^2 \geq 5td^{-\gamma}\right] \leq 4^{-t}d^{-t(1-\gamma)}. \tag{33}$$

*Proof.* We aim to prove the slightly more general statement

$$\Pr\left[|\langle a, x\rangle|^2 \geq (\delta+1)td^{-\gamma}\right] \leq \delta^{-t}d^{-t(1-\gamma)},$$

which is valid for any $\delta \geq 1$. Setting $\delta = 4$ then yields (33). The $t$-design property provides us with useful information about the first $t$ moments of the non-negative random

variable $\xi = |\langle a, x\rangle|^2$. Indeed, with $A = aa^*$ it holds for every $k \leq t$ that

$$\begin{aligned}
\mathbb{E}\left[\xi^k\right] &= \mathbb{E}\left[\mathrm{tr}(AX)^k\right] \\
&= \mathrm{tr}\left(\mathbb{E}\left[A^{\otimes k}\right] X^{\otimes k}\right) \\
&= \binom{d+k-1}{k}^{-1} \mathrm{tr}\left(P_{\mathrm{Sym}^k} X^{\otimes k}\right) \\
&= \binom{d+k-1}{k}^{-1} \mathrm{tr}\left(X^{\otimes k}\right) \\
&\leq d^{-k} k!,
\end{aligned}$$

because $X^{\otimes k}$ is invariant under $P_{\mathrm{Sym}^k}$. One way of seing this[5] is to note that $\mathrm{range}(X^{\otimes k}) = \mathrm{span}(x^{\otimes k})$ and the latter is already contained in $\mathrm{Sym}^k$. Therefore the $k$-th moment $\tau_k$ of $\xi$ is bounded by

$$\tau_k = \left(\mathbb{E}[\xi^k]\right)^{1/k} \leq (d^{-k}k!)^{1/k} \leq k/d.$$

These inequalities are tight for the mean $\mu = \tau_1$ of $\xi$ and hence

$$\mu = \mathbb{E}[\xi] = d^{-1}.$$

Now we aim to use the well-known $t$-th moment bound

$$\Pr\left[|\xi - \mu| \geq s\tau_t\right] \leq s^{-t},$$

which is a straightforward generalization of Chebyshev's inequality. Applying it, yields the desired result. Indeed,

$$\begin{aligned}
\Pr\left[|\langle a, x\rangle|^2 \geq (\delta+1)td^{-\gamma}\right] &= \Pr\left[\xi - \mu \geq (\delta+1)td^{-\gamma} - d^{-1}\right] \\
&\leq \Pr\left[\xi - \mu \geq \delta td^{-\gamma}\right] \\
&\leq \Pr\left[|\xi - \mu| \geq \delta d^{1-\gamma}\tau_t\right] \\
&\leq \delta^{-t} d^{-t(1-\gamma)},
\end{aligned}$$

and we are done. □

The previous lemma bounds the probability of the undesired events

$$E_i^c = \left\{|\langle a_i, x\rangle|^2 \geq 5td^{-\gamma}\right\}, \tag{34}$$

where $0 \leq \gamma \leq 1$ is a fixed parameter which we refer to as the *truncation rate*. It turns out that a single truncation of this kind does not quite suffice yet for our purpose. We need to introduce a second truncation step.

**Definition 14.** *Fix $Z \in T$ arbitrary and decompose it as*

$$Z = \zeta\left(xz^* + zx^*\right),$$

*for some unique $\zeta > 0$ and $z \in V^d$ with $\|z\|_{\ell_2} = 1$. For this $z$ we introduce the event*

$$G_i^c := \left\{|\langle z, a_i\rangle|^2 \geq 5td^{-\gamma}\right\}$$

*and define the two-fold truncated operator*

$$\mathcal{R}_Z := \mathcal{R}_z = \frac{(d+1)d}{m} \sum_{i=1}^{m} 1_{E_i} 1_{G_i} \Pi_{A_i}, \tag{35}$$

[5]Alternatively one could also rearange tensor systems: $X^{\otimes k} = (xx^*)^{\otimes k} \simeq x^{\otimes k}(x^*)^{\otimes k}$ and use $P_{\mathrm{Sym}^k} x^{\otimes k} = x^{\otimes k}$.

*where $1_{E_i}$ and $1_{G_i}$ denote the indicator functions associated with the events $E_i$ and $G_i$, respectively.*

The following result shows that due to Lemma 13 this truncated operator is in expectation close to the original $\mathcal{R}$.

**Proposition 15.** *Fix $Z \in T$ arbitrary and let $\mathcal{R}_Z$ be as in (35). Then*

$$\|\mathbb{E}[\mathcal{R}_Z - \mathcal{R}]\|_{\mathrm{op}} \leq 4^{1-t} d^{2-t(1-\gamma)} \tag{36}$$

*Proof.* We start by introducing the auxiliar (singly truncated) operator

$$\mathcal{R}_{\mathrm{aux}} := \frac{(d+1)d}{m} \sum_{i=1}^{m} 1_{E_i} \Pi_{A_i}$$

and observe

$$\|\mathbb{E}\left[\mathcal{R}_Z - \mathcal{R}\right]\|_{\mathrm{op}} \leq \|\mathbb{E}\left[\mathcal{R} - \mathcal{R}_{\mathrm{aux}}\right]\|_{\mathrm{op}} + \|\mathbb{E}\left[\mathcal{R}_Z - \mathcal{R}_{\mathrm{aux}}\right]\|_{\mathrm{op}}. \tag{37}$$

Now use Lemma 13 to bound the first term:

$$\begin{aligned}
\|\mathbb{E}[\mathcal{R} - \mathcal{R}_{\mathrm{aux}}]\|_{\mathrm{op}} &= \left\| \frac{(d+1)d}{m} \sum_{i=1}^{m} \mathbb{E}\left[(1 - 1_{E_i})\Pi_{A_i}\right] \right\|_{\mathrm{op}} \\
&\leq \frac{(d+1)d}{m} \sum_{i=1}^{m} \mathbb{E}\left[1_{E_i^c} \|\Pi_{A_i}\|_{\mathrm{op}}\right] \\
&\leq \frac{2d^2}{m} \sum_{i=1}^{m} \mathbb{E}\left[1_{E_i^c}\right] = \frac{2d^2}{m} \sum_{i=1}^{m} \Pr[E_i^c] \\
&\leq 2d^2 \times 4^{-t} d^{-t(1-\gamma)} = 2^{1-2t} d^{2-t(1-\gamma)}.
\end{aligned}$$

Similarily,

$$\begin{aligned}
\|\mathbb{E}[\mathcal{R}_{\mathrm{aux}} - \mathcal{R}_Z\|_{\mathrm{op}} &= \frac{(d+1)d}{m} \left\| \sum_{i=1}^{m} \mathbb{E}\left[1_{G_i^c} \Pi_{A_i}\right] \right\|_{\mathrm{op}} \leq \frac{2d^2}{m} \sum_{i=1}^{m} \mathbb{E}[1_{G_i^c}] \\
&\leq \frac{2d^2}{m} \sum_{i=1}^{m} \Pr[G_i^c] \leq 2^{1-2t} d^{2-t(1-\gamma)}
\end{aligned}$$

and inserting these bounds into (37) yields the desired statement. □

We now establish a technical result which will allow us to find a suitable approximate dual certificate using the "golfing scheme" construction [15, 62].

**Proposition 16.** *Fix $Z \in T$ arbitrary, let $\mathcal{R}_Z$ be as in (35). Assume that that the design order $t$ is at least 3 and the truncation rate $\gamma$ satisfies*

$$\gamma \leq 1 - 2/t.$$

*Then for $1/4 \leq b \leq 1$ and $c \geq \sqrt{2}b$ with probability at least $1 - d\exp(-\frac{9mb}{640td^{2-\gamma}})$ one has*

$$\|\mathcal{P}_T^{\perp} \mathcal{S} \mathcal{R}_Z Z\|_{\infty} \leq b\|Z\|_2 \quad \textit{and} \tag{38}$$

$$\|\mathcal{P}_T(\mathcal{S}\mathcal{R}_Z - \mathcal{I})Z\|_2 \leq c\|Z\|_2. \tag{39}$$

(Recall the definition $\mathcal{S} := \mathcal{I} - \frac{1}{d+1}\Pi_{\mathbb{1}}$ from Lemma 8).

*Proof.* The statement is invariant under rescaling of $Z$. Therefore it suffices to treat the case $\|Z\|_2 = 1$. In this case we can decompose

$$Z = \zeta(zx^* + zx^*)$$

with some fixed $z \in V^d$ obeying $\|z\|_{\ell_2} = 1$ and $0 < \zeta \leq 1$. Isotropicity (Lemma 8) of $\mathcal{SR}$ guarantees $\mathcal{P}_T^\perp \mathcal{S}\mathbb{E}[\mathcal{R}]Z = 0$ as well as $\mathcal{P}_T \mathcal{S}\mathbb{E}[\mathcal{R}]Z = Z$. Let us now focus on (38) and use Proposition 15 in order to write

$$\begin{aligned}
& \|\mathcal{P}_T^\perp \mathcal{S}\mathcal{R}_Z Z\|_\infty \\
= \; & \|\mathcal{P}_T^\perp \mathcal{S}\left(\mathcal{R}_Z - \mathbb{E}[\mathcal{R}]\right) Z\|_\infty \\
\leq \; & \|\mathcal{P}_T^\perp \mathcal{S}\left(\mathcal{R}_Z - \mathbb{E}[\mathcal{R}_Z]\right) Z\|_\infty + \|\mathcal{P}_T^\perp \mathcal{S}\mathbb{E}[\mathcal{R}_Z - \mathcal{R}]Z\|_\infty \\
\leq \; & \|\mathcal{P}_T^\perp\|_{\mathrm{op}} \|\mathcal{S}\|_{\mathrm{op}} \|(\mathcal{R}_Z - \mathbb{E}[\mathcal{R}_Z])Z\|_\infty + 4^{1-t} d^{2-t(1-\gamma)} \|\mathcal{P}_T^\perp\|_{\mathrm{op}} \|\mathcal{S}\|_{\mathrm{op}} \|Z\|_2 \\
\leq \; & \|(\mathcal{R}_Z - \mathbb{E}[\mathcal{R}_Z])Z\|_\infty + b/4.
\end{aligned}$$

Here we have used $\|\mathcal{S}\|_{\mathrm{op}} \leq 1$, $\|\mathcal{P}_T^\perp\|_{\mathrm{op}} \leq 1$ as well as

$$\|\mathbb{E}\left[\mathcal{R}_Z - \mathcal{R}\right]\|_{\mathrm{op}} \leq 4^{1-t} d^{2-t(1-\gamma)} \leq 4^{1-t} \leq 1/16 \leq b/4, \tag{40}$$

which follows from $\gamma \leq 1 - 2/t$, $t \geq 3$ and $b \geq 1/4$. To obtain (39) we use a similar reasoning:

$$\begin{aligned}
& \|\mathcal{P}_T \mathcal{S}\left(\mathcal{R}_Z - \mathcal{I}\right) Z\|_2 \\
= \; & \|\mathcal{P}_T \mathcal{S}\left(\mathcal{R}_Z - \mathbb{E}[\mathcal{R}]\right) Z\|_2 \\
\leq \; & \sqrt{2}\|\mathcal{P}_T \mathcal{S}\left(\mathcal{R}_Z - \mathbb{E}[\mathcal{R}_Z]\right) Z\|_\infty + \|\mathcal{P}_T \mathcal{S}\mathbb{E}[\mathcal{R}_Z - \mathcal{R}]Z\|_2 \\
\leq \; & \sqrt{2}\|\mathcal{P}_T\|_{\mathrm{op}} \|\mathcal{S}\|_{\mathrm{op}} \|(\mathcal{R}_Z - \mathbb{E}[\mathcal{R}_Z])Z\|_\infty + b/4 \|\mathcal{P}_T\|_{\mathrm{op}} \|\mathcal{S}\|_{\mathrm{op}} \|Z\|_2 \\
\leq \; & \sqrt{2}\|\left(\mathcal{R}_Z - \mathbb{E}[\mathcal{R}_Z]\right) Z\|_\infty + b/4,
\end{aligned}$$

where we have used the fact that $\mathcal{P}_T$ projects onto a subspace of at most rank-2 matrices in the third line and (40) in the fourth. This motivates to define the event

$$E := \{\|\left(\mathcal{R}_Z - \mathbb{E}[\mathcal{R}_Z]\right) Z\|_\infty \leq 3b/4\}$$

which guarantees both (38) and (39) due to the assumption on $c$ and $\|Z\|_2 = 1$. So everything boils down to bounding the probability of $E^c$. We decompose

$$\left(\mathcal{R}_Z - \mathbb{E}[\mathcal{R}_Z]\right) Z = \sum_{i=1}^m \left(M_i - \mathbb{E}[M_i]\right) \quad \text{with} \quad M_i = \frac{(d+1)d}{m} 1_{E_i} 1_{G_i} A_i \operatorname{tr}(A_i Z).$$

We will estimate this sum using the Operator Bernstein inequality (Theorem 4). Thus we need an a priori bound for the summands

$$\begin{aligned}
\|M_i\|_\infty &= \frac{(d+1)d}{m} 1_{E_i} 1_{G_i} \|A_i\|_\infty |\operatorname{tr}(A_i Z)| \leq \frac{2d^2}{m} 1_{E_i} 1_{G_i} 2|\langle x, a_i\rangle||\langle z, a_i\rangle| \\
&\leq \frac{4d^2}{m} 5td^{-\gamma} = \frac{20}{m} td^{2-\gamma} =: \overline{R},
\end{aligned}$$

as well as a bound for the variance. First observe that

$$\mathbb{E}[(M_i - \mathbb{E}[M_i])^2] = \mathbb{E}\left[M_i^2\right] - \mathbb{E}[M_i]^2 \leq \mathbb{E}\left[M_i^2\right].$$

and therefore

$$\begin{aligned}
& \mathbb{E}\left[M_i^2\right] \\
= \;& \frac{(d+1)^2 d^2}{m^2}\mathbb{E}\left[1_{E_i}1_{G_i}\operatorname{tr}(A_i Z)^2 A_i^2\right] \leq \frac{(d+1)^2 d^2}{m^2}\mathbb{E}\left[\operatorname{tr}(A_i Z)^2 A_i^2\right] \\
= \;& \frac{(d+1)^2 d^2}{m^2}\operatorname{tr}_{2,3}\left(\mathbb{E}[A_i^{\otimes 3}]\mathbb{1}\otimes Z\otimes Z\right) = \frac{6(d+1)d}{m^2(d+2)}\operatorname{tr}_{2,3}\left(P_{\mathrm{Sym}^3}\mathbb{1}\otimes Z\otimes Z\right) \\
\leq \;& \frac{d}{m^2}\left(\mathbb{1}\operatorname{tr}(Z)^2 + Z\operatorname{tr}(Z) + Z + \mathbb{1}\operatorname{tr}(Z^2) + 2Z^2\right) \\
\leq \;& \frac{8d}{m^2}\|Z\|_2^2\mathbb{1} = \frac{8d}{m^2}\mathbb{1}.
\end{aligned}$$

Here we have used $\operatorname{tr}(Z) \leq \sqrt{2}\|Z\|_2$, $Z^2 \leq \|Z\|_2^2\mathbb{1}$ and $\|Z\|_2 = 1$. From this we can conclude

$$\left\| \sum_i \mathbb{E}[(M_i - \mathbb{E}[M_i])^2 \right\|_\infty \leq m \max_{i=1,\ldots,m} \|\mathbb{E}[M_i^2]\|_\infty \leq \frac{8d}{m} =: \sigma^2.$$

Observing that

$$\frac{\sigma^2}{\overline{R}} \leq \frac{8}{20t}d^{\gamma-1} \leq \frac{2}{15} \leq \frac{3}{4}b,$$

Theorem 4 yields

$$\Pr\left[E^c\right] = \Pr\left[\|\left(\mathcal{R}_Z - \mathbb{E}[\mathcal{R}_Z]\right)Z\|_\infty > 3b/4\right] \leq d\exp\left(-\frac{3\times 3mb}{8\times 4\times 20td^{2-\gamma}}\right),$$

as desired. □

With this ingredient we can now construct a suitable approximate dual certificate $Y$, closely following [62].

**Proposition 17.** *Let $x \in V^d$ be an arbitrary normalized vector ($\|x\|_{\ell_2} = 1$), $X = xx^*$ and let $\omega \geq 1$ be arbitrary. If the design order $t$ ($t \geq 3$) and the truncation rate $\gamma$ is chosen such that*

$$\gamma \leq 1 - 2/t$$

*holds and the total number of measurements fulfills*

$$m \geq 9394\omega t d^{2-\gamma}\log^2(d), \tag{41}$$

*then with probability larger than $1 - 0.5\mathrm{e}^{-\omega}$, there exists an approximate dual certificate $Y$ as in Def. 11.*

*Proof.* Our construction of $Y$ follows a recursive scheme of $l$ iterations. The $i$-th iteration depends on 3 parameters $m_i \in \mathbb{N}$ and $b_i, c_i \in (0,1)$ which will be chosen later on. To initialize set

$$Y_0 = 0$$

(the $Y_i$'s, $i \geq 1$, will be defined iteratively below). Define

$$Q_i = X - \mathcal{P}_T Y_i \in T.$$

The $i$-th step proceeds according to the following protocol:

We sample $m_i$ vectors iid from the $t$-desing $D_t$. Let $\tilde{R}_{Q_{i-1}}$ be the measurement operator of length $m_i$ introduced in Definiton 14 (so the summands are conditioned on $E_i$ and $G_i$

for $Q_{i-1} \in T$). Then we check whether for $b = b_i$ and $c = c_i$ equations (38) and (39) are satisfied. If so, set $\mathcal{R}^{(i)}_{Q_{i-1}} = \tilde{\mathcal{R}}_{Q_{i-1}}$ as well as

$$Y_i = \mathcal{S}\mathcal{R}^{(i)}_{Q_{i-1}} \mathcal{P}_T(X - Y_{i-1}) + Y_{i-1}$$

and proceed to step $i+1$. If either of the bounds (38, 39) does not hold, repeat the $i$-th step with a fresh batch of $m_i$ measurements drawn iid from $D_t$. Denote the probability of having to repeat the $i$-th step by $p_{\mathrm{err}}(i)$ and the eventual number of repetitions by $r_i \geq 1$.

The following identities are easily verified (c.f. [62][Lemma 14]):

$$\begin{aligned} Y &:= Y_l = \mathcal{S}\mathcal{R}^{(l)}_{Q_{l-1}} \mathcal{P}_T(X - Y_{l-1}) + Y_{l-1} = \sum_{i=1}^{l} \mathcal{S}\mathcal{R}^{(i)}_{Q_{i-1}} Q_{i-1} \quad \text{and} \\ Q_i &= X - \mathcal{P}_T Y_i = \prod_{j=1}^{i} \mathcal{P}_T(\mathcal{I} - \mathcal{S}\mathcal{R}^{(j)}_{Q_{j-1}}) X. \end{aligned}$$

The validity of properties (38) and (39) in each step guarantee

$$\begin{aligned} \|Y_T - X\|_2 &= \|Q_l\|_2 \leq \prod_{i=1}^{l} c_i \|Q_0\|_2 = \prod_{i=1}^{l} c_i \|X\|_2 = \prod_{i=1}^{l} c_i, \\ \|Y_T^{\perp}\|_\infty &= \left\| \mathcal{P}_T^{\perp} \sum_{i=1}^{l} \mathcal{S}\mathcal{R}^{(i)}_{Q_{i-1}} Q_{i-1} \right\|_\infty \leq \sum_{i=1}^{l} \left\| \mathcal{P}_T^{\perp} \mathcal{S}\mathcal{R}^{(i)}_{Q_{i-1}} Q_{i-1} \right\|_\infty \\ &\leq \sum_{i=1}^{l} b_i \|Q_{i-1}\|_2 = b_1 + \sum_{i=2}^{l} b_i \prod_{j=1}^{i-1} c_j. \end{aligned}$$

Now choose parameters

$$l = \lceil \log_2 d \rceil + 2, \quad b_i = \frac{1}{4}, \quad c_i = \frac{1}{2},$$

which obey the conditions ($1/4 \leq b_i \leq 1$ and $c_i \geq \sqrt{2} b_i$) required for Proposition 16. These constants assure

$$\begin{aligned} \|Y_T - X\|_2 &= \|Q_l\|_2 = 2^{-l} \leq \frac{1}{4d}, \\ \|Y_T^{\perp}\|_\infty &\leq b_1 + \sum_{i=2}^{l} b_i \prod_{j=1}^{i-1} c_j \leq \frac{1}{4} + \frac{1}{4} \sum_{i=1}^{\infty} 2^{-i} = \frac{1}{2}, \end{aligned}$$

which are precisely the requirements on $Y$.

Next, we estimate the probability $p_{\mathrm{err}}$ that the total number of measurements

$$\sum_{i=1}^{l} m_i r_i$$

exceeds the bound (41). To that end, it is fruitful to think of a random walk which advances from position $i$ to $i+1$ if a newly sampled batch fulfills equations (38), (39), and remains at position $i$ if that is not the case, i.e. with probability $p_{\mathrm{err}}(i)$. In that sense, $p_{\mathrm{err}}$ is the probability that the random walker fails to reach position $l$ before exceeding the allowed number of trials.

To obtain concrete numbers, choose $m_i = 854td^{2-\gamma}\log d$. Then Propostion 16 gives

$$p_{\mathrm{err}}(i) \leq d\exp\left(-\frac{9m_i}{2560td^{2-\gamma}}\right) \leq \mathrm{e}^{-3} \leq 1/20.$$

Dividing the advertised total number of measurements (41) by $m_i$ shows that one can sample

$$l' = 2l + 3\omega + 6\log 2 \tag{42}$$

batches. The total failure probability $p_{\mathrm{err}}$ is thus the probability that fewer than $l$ successes occur in $l'$ trials with individual failure probabilty smaller than $1/20$. This can be estimated using a standard concentration bound for binomial random variables, e.g.

$$\Pr\left[|\mathrm{Bin}(n,p) - np| \geq \tau\right] \leq 2\exp\left(-\frac{\tau^2}{3np}\right)$$

from [63, Section Concentration]. In this particular situation $n = l'$, $p = 19/20$ and $\tau = (l' - l)$ is adequate. The choice of $l'$ – equation (42)– then assures

$$p_{\mathrm{err}} \leq \Pr\left[\mathrm{Bin}(l', 19/20) - 19/20l'| \geq l' - l\right] \leq 2\exp\left(-\frac{20(l'-l)^2}{3l'19}\right) \leq 0.5\mathrm{e}^{-\omega},$$

Note that our choice of $l'$ is not tight, but suffices for our purpose. Consequently $p_{\mathrm{err}} \leq 0.5\mathrm{e}^{-\omega}$ which is the desired bound on the probability of our construction failing. □

Finally we are ready to put all pieces together and show or main result – Theorem 1.

*Proof of the Main Theorem.* In section 4 (Proposition 12) we have shown that the algorithm (25) recovers the sought for signal $x$, provided that (26) holds and a suitable approximate dual certificate $Y$ exists. Proposition 17 – with a maximal truncation rate of $\gamma = (1 - 2/t)$ – implies that the probability that no such $Y$ can be constructed is smaller than $0.5\mathrm{e}^{-\omega}$, provided that the sampling rate $m$ obeys

$$m \geq 9394\omega td^{1+2/t}\log^2 d, \tag{43}$$

Furthermore, Proposition 9 implies that the probability of (26) failing is also bounded by $0.5\mathrm{e}^{-\omega}$. Theorem 1 now follows from the union bound over these two probabilities of failure. □

## 6. Converse Bound

In this paper, we require designs of order at least three. Here we prove that this criterion is fundamental in the sense that sampling from 2-designs in general cannot guarantee a subquadtratic sampling rate. In order to do so, we will use a particular sort of 2-design, called a *maximal set of mutually unbiased bases* (MUBs) [40, 41, 42, 43]. Two orthonormal bases $\{u_i\}_{i=1}^d$ and $\{v_i\}_{i=1}^d$ are called *mutually unbiased* if their overlap is uniformly minimal. Concretely, this means that

$$|\langle u_i, v_j\rangle|^2 = \frac{1}{d} \quad \forall i,j = 1,\ldots,d$$

must hold for all $i, j = 1, \ldots, d$. Note that this is just a generalization of the incoherence property between standard and Fourier basis. In prime power dimensions, a maximal set of $(d + 1)$ such MUBs is known to exist (and can be constructed) [64]. Such a set is maximal in the sense that it is not possible to find more than $(d + 1)$ MUBs in any Hilbert space. Among other interesting properties – cf. [65] for a detailed survey – maximal sets of MUBs are known to form 2-designs [41, 43].

The defining properties of a maximal set of MUBs allow us to derive the converse bound – Theorem 2.

**Theorem 18** (Converse bound)**.** *Let $d \geq 2$ be a prime power and let $D_2 \subset \mathbb{C}^d$ be a maximal set of MUBs. Then there exist orthogonal, normalized vectors $x, z \in \mathbb{C}^d$ which have the following property.*

*Suppose that $m$ measurement vectors $y_1, \ldots, y_m$ are sampled independently and uniformly at random from $D_2$. Then, for any $\omega \geq 0$, the number of measurements must obey*

$$m \geq \frac{\omega}{4} d(d+1), \tag{44}$$

*or the event*

$$|\langle a_i, x\rangle|^2 = |\langle a_i, z\rangle|^2 \quad \forall\, i \in \{1, \ldots, m\}$$

*will occur with probability at least $\mathrm{e}^{-\omega}$.*

Consequently a scaling of $\mathcal{O}(d^2)$ in general cannot be avoided when using only 2-designs and requiring a "reasonably small" probability of failure in the recovery process.

*Proof of Theorem 18.* Suppose that $\{u_i\}_{i=1}^d$ is one orthonormal basis contained in the maximal set of MUBs $D_2$ and set $x := u_1$ as well as $z := u_2$. Note that by definition these vectors are orthogonal and normalized. Due to the particular structure of MUBs, $x$ and $z$ can only be distinguished if either $u_1$ or $u_2$ is contained in $\{a_1, \ldots, a_m\}$. Since each $a_i$ is chosen iid at random from $D_2$ containing $(d+1)d$ elements, the probability of obtaining either $u_1$ or $u_2$ is $p = \frac{2}{(d+1)d}$. As a result, the problem reduces to the following standard stopping time problem (cf. for example Example (2) in Chapter 6.2 in [66]):

Suppose that the probability of success in a Bernoulli experiment is $p$. How many trials $m$ are required in order for the probability of at least one success to be $1 - \mathrm{e}^{\omega}$ or larger?

To answer this question, we have to find the smallest integer $m$ such that

$$1 - (1-p)^m \geq 1 - \mathrm{e}^{-\omega}, \quad \text{or equivalently} \quad -m \log(1-p) \geq \omega. \tag{45}$$

The standard inequality

$$p \leq -\log(1-p) \leq \frac{p}{1-p} \leq 2p$$

for any $p \in [0, 1/2]$ implies that (44) is a necessary criterion for (45) and we are done. □

## 7. Conclusion

In this paper we have derived a partly derandomized version of Gaussian PhaseLift [11, 12]. Instead of Gaussian random measurements, our method guarantees recovery for sampling iid from certain finite vector configurations, dubbed $t$-designs. The required sampling rate depends on the design order $t$:

$$m = \mathcal{O}\left(t d^{1+2/t} \log^2 d\right). \tag{46}$$

For small $t$ this rate is worse than the Gaussian analogue – but still non-trivial. However, as soon as $t$ exceeds $2 \log d$, we obtain linear scaling up to a polylogarithmic overhead.

In any case, we feel that the main purpose of this paper is not to present yet another efficient solution heuristics, but to show that the phase retrieval problem can be derandomized using $t$-designs. These finite vector sets lie in the vast intermediate region between random Fourier vectors and Gaussian random vectors (the Fourier basis is a 1-design, whereas normalized Gaussian random vectors correspond to an $\infty$-design). Therefore the design order $t$ allows us to gradually transcend between these two extremal cases.

**Acknowledgements** The work of DG and RK is supported by the Excellence Initiative of the German Federal and State Governments (Grant ZUK 43).

## REFERENCES

[1] R. Millane, "Phase retrieval in crystallography and optics," *JOSA A*, vol. 7, no. 3, pp. 394–411, 1990.
[2] Y. M. Bruck and L. Sodin, "On the ambiguity of the image reconstruction problem," *Optics Communications*, vol. 30, no. 3, pp. 304–308, 1979.
[3] R. Balan, P. Casazza, and D. Edidin, "On signal reconstruction without phase." *Appl. Comput. Harmon. Anal.*, vol. 20, no. 3, pp. 345–356, 2006.
[4] T. Heinosaari, L. Mazzarella, and M. M. Wolf, "Quantum tomography under prior information." *Commun. Math. Phys.*, vol. 318, no. 2, pp. 355–374, 2013.
[5] B. Sanderson, "Immersions and embeddings of projective spaces." *Proc. Lond. Math. Soc. (3)*, vol. 14, pp. 137–153, 1964.
[6] D. Mixon, "Short, fat matrices," blog, 2013. [Online]. Available: http://dustingmixon.wordpress.com/
[7] R. Balan, B. G. Bodmann, P. G. Casazza, and D. Edidin, "Painless reconstruction from magnitudes of frame coefficients." *J. Fourier Anal. Appl.*, vol. 15, no. 4, pp. 488–501, 2009.
[8] B. Alexeev, A. S. Bandeira, M. Fickus, and D. G. Mixon, "Phase retrieval with polarization," *preprint arXiv:1210.7752*, 2012.
[9] A. S. Bandeira, Y. Chen, and D. G. Mixon, "Phase retrieval from power spectra of masked signals," *preprint arXiv:1303.4458*, 2013.
[10] E. J. Candes, Y. C. Eldar, T. Strohmer, and V. Voroninski, "Phase retrieval via matrix completion," *SIAM Journal on Imaging Sciences*, vol. 6, no. 1, pp. 199–225, 2013.
[11] E. J. Candès, T. Strohmer, and V. Voroninski, "Phaselift: exact and stable signal recovery from magnitude measurements via convex programming." *Commun. Pure Appl. Math.*, vol. 66, no. 8, pp. 1241–1274, 2013.
[12] E. Candès and X. Li, "Solving quadratic equations via PhaseLift when there are about as many equations as unknowns," *Foundations of Computational Mathematics*, pp. 1–10, 2013.
[13] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization." *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, 2010.
[14] E. J. Candès and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *Information Theory, IEEE Transactions on*, vol. 56, no. 5, pp. 2053–2080, 2010.
[15] D. Gross, "Recovering low-rank matrices from few coefficients in any basis," *Information Theory, IEEE Transactions on*, vol. 57, no. 3, pp. 1548–1566, 2011.
[16] Y.-K. Liu, "Universal low-rank matrix recovery from pauli measurements," *Advances in Neural Information Processing Systems (NIPS)*, pp. 1638–1646, 2011.
[17] J. R. Fienup, "Phase retrieval algorithms: A comparison," *Applied Optics*, vol. 21, no. 15, pp. 2758–2769, 1982.
[18] H. H. Bauschke, P. L. Combettes, and D. R. Luke, "Hybrid projection–reflection method for phase retrieval," *JOSA A*, vol. 20, no. 6, pp. 1025–1034, 2003.
[19] P. Netrapalli, P. Jain, and S. Sanghavi, "Phase retrieval using alternating minimization," *preprint arXiv:1306.0160*, 2013.
[20] Y. C. Eldar and S. Mendelson, "Phase retrieval: Stability and recovery guarantees," *arXiv preprint arXiv:1211.0872*, 2012.
[21] X. Li and V. Voroninski, "Sparse signal recovery from quadratic measurements via convex programming," *preprint arXiv:1209.4785*, 2012.
[22] M. Ehler, M. Fornasier, and J. Sigl, "Quasi-linear compressed sensing," preprint.
[23] P. Delsarte, J. Goethals, and J. Seidel, "Spherical codes and designs." *Geom. Dedicata*, vol. 6, pp. 363–388, 1977.
[24] V. Sidelnikov, "Spherical 7-designs in $2^n$-dimensional Euclidean space." *J. Algebr. Comb.*, vol. 10, no. 3, pp. 279–288, 1999.
[25] G. Nebe, E. Rains, and N. Sloane, "The invariants of the Clifford groups." *Des. Codes Cryptography*, vol. 24, no. 1, pp. 99–121, 2001.
[26] A. Scott, "Tight informationally complete quantum measurements." *J. Phys. A, Math. Gen.*, vol. 39, no. 43, pp. 13 507–13 530, 2006.
[27] A. Ambainis and J. Emerson, "Quantum t-designs: t-wise independence in the quantum world," in *22nd Annual IEEE Conference on Computational Complexity, Proceedings*, 2007, pp. 129–140.
[28] A. Hayashi, T. Hashimoto, and M. Horibe, "Reexamination of optimal quantum state estimation of pure states," *Physical review A*, vol. 72, no. 3, SEP 2005.

[29] D. Gross, K. Audenaert, and J. Eisert, "Evenly distributed unitaries: on the structure of unitary designs." *J. Math. Phys.*, vol. 48, no. 5, pp. 052 104, 22, 2007.

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

[31] E. Candès and T. Tao, "Decoding by linear programming," *Information Theory, IEEE Transactions on*, vol. 51, pp. 4203–4215, 2005.

[32] R. G. Baraniuk, M. Davenport, R. A. DeVore, and M. Wakin, "A simple proof of the Restricted Isometry Property for random matrices," *Constr. Approx.*, vol. 28, no. 3, pp. 253–263, 2008.

[33] E. J. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[34] M. Rudelson and R. Vershynin, "On sparse reconstruction from Fourier and Gaussian measurements," *Comm. Pure Appl. Math.*, vol. 61, pp. 1025–1045, 2008.

[35] R. A. Low, "Large deviation bounds for $k$-designs." *Proc. R. Soc. Lond., Ser. A, Math. Phys. Eng. Sci.*, vol. 465, no. 2111, pp. 3289–3308, 2009.

[36] M. Luby and A. Wigderson, *Pairwise independence and derandomization*. Boston, MA: Now, 2006.

[37] B. Bajnok, "Construction of spherical $t$-designs." *Geom. Dedicata*, vol. 43, no. 2, pp. 167–179, 1992.

[38] J. Korevaar and J. Meyers, "Chebyshev-type quadrature on multidimensional domains." *J. Approx. Theory*, vol. 79, no. 1, pp. 144–164, 1994.

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

[40] J. Schwinger, "Unitary operator bases." *Proc. Natl. Acad. Sci. USA*, vol. 46, pp. 570–579, 1960.

[41] G. Zauner, "Quantendesigns: Grundzüge einer nichtkommutativen Designtheorie," Ph.D. dissertation, University of Vienna, 1999.

[42] H. König, "Cubature formulas on spheres." in *Advances in multivariate approximation. Proceedings of the 3rd international conference on multivariate approximation theory*. Berlin: Wiley-VCH, 1999, pp. 201–211.

[43] A. Klappenecker and M. Rotteler, "Mutually unbiased bases are complex projective 2-designs," in *2005 IEEE International Symposium on Information Theory (ISIT), Vols 1 and 2*, 2005, pp. 1740–1744.

[44] R. Kueng and D. Gross, "Stabilizer states are complex projective 3-designs in qubit dimensions," in preparation, 2013.

[45] J. M. Renes, R. Blume-Kohout, A. Scott, and C. M. Caves, "Symmetric informationally complete quantum measurements." *J. Math. Phys.*, vol. 45, no. 6, pp. 2171–2180, 2004.

[46] C. Bachoc and B. Venkov, "Modular forms, lattices and spherical designs." in *Euclidean lattices, spherical designs and modular forms. On the works of Boris Venkov*. Genève: L'Enseignement Mathématique, 2001, pp. 87–111.

[47] S. Hoory, N. Linial, and A. Widgerson, "Expander graphs and their applications." *Bull. Am. Math. Soc., New Ser.*, vol. 43, no. 4, pp. 439–561, 2006.

[48] D. Mondragon and V. Voroninski, "Determination of all pure quantum states from a minimal number of observables," *preprint arXiv:1306.1214*, 2013.

[49] A. Barvinok, *A course in convexity.* Providence, RI: American Mathematical Society (AMS), 2002.

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

[51] J. Watrous, "Theory of quantum information," lecture notes, 2011. [Online]. Available: https://cs.uwaterloo.ca/~watrous/LectureNotes.html

[52] A. Neumaier, "Combinatorial configurations in terms of distances," *Dept. of Mathematics Memorandum*, pp. 81–09, 1981.

[53] S. Hoggar, "t-designs in projective spaces." *European Journal of Combinatorics*, vol. 3, pp. 233–254, 1982.

[54] V. I. Levenshtein, "Universal bounds for codes and designs." in *Handbook of coding theory. Vol. 1. Part 1: Algebraic coding*. Amsterdam: Elsevier, 1998, pp. 499–648.

[55] R. Ahlswede and A. Winter, "Strong converse for identification via quantum channels." *Information Theory, IEEE Transactions on*, vol. 48, no. 3, pp. 569–579, 2002.

[56] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, "Quantum state tomography via compressed sensing," *Physical review letters*, vol. 105, no. 15, p. 150401, 2010.

[57] J. A. Tropp, "User-friendly tail bounds for sums of random matrices." *Found. Comput. Math.*, vol. 12, no. 4, pp. 389–434, 2012.

[58] J. A. Tropp, "User-friendly tools for random matrices: An introduction," Notes, 2012. [Online]. Available: http://users.cms.caltech.edu/~jtropp/notes/Tro12-User-Friendly-Tools-NIPS.pdf

[59] V. Turaev, *Quantum invariants of knots and 3-manifolds.* Berlin: Walter de Gruyter, 1994.

[60] P. Cvitanović, *Group theory. Birdtracks, Lie's, and exceptional groups.* Princeton, NJ: Princeton University Press, 2008.
[61] R. Bhatia, *Matrix analysis.* New York, NY: Springer, 1996.
[62] R. Kueng and D. Gross, "RIPless compressed sensing from anisotropic measurements," 2013. [Online]. Available: http://dx.doi.org/10.1016/j.laa.2013.04.018
[63] M. Habib, C. McDiarmid, J. Ramírez Alfonsín, and B. Reed, Eds., *Probabilistic methods for algorithmic discrete mathematics.* Berlin: Springer, 1998.
[64] A. Klappenecker and M. Rötteler, "Constructions of mutually unbiased bases." in *Finite fields and applications. 7th international conference,* $\mathbb{F}_{q^7}$. Berlin: Springer, 2004, pp. 137–144.
[65] T. Durt, B.-G. Englert, I. Bengtsson, and K. Życzkowski, "On mutually unbiased bases." *Int. J. Quantum Inf.*, vol. 8, no. 4, pp. 535–640, 2010.
[66] W. Feller, "An introduction to probability theory and its applications. I." New York-London-Sydney: John Wiley and Sons, 1968.

## 8. Appendix

Here we briefly state an elementary proof of Lemma 6. In the main text we proved this result using wiring diagrams. The purpose of this is to underline the relative simplicity of wiring diagram calculations. Indeed, the elementary proof below is considerably more cumbersome than its pictorial counterpart.

8.1. **Elementary proof of Lemma 6.** Let us choose an arbitrary orthonormal basis $b_1, \ldots, b_d$ of $V^d$. In the induced basis $\{b_i \otimes b_j\}_{i,j=1}^d$ of $V^d \otimes V^d$ the transpositions then correspond to

$$\underline{1} = \mathbb{1} \otimes \mathbb{1} = \sum_{i=1}^d b_i b_i^* \otimes \sum_{j=1}^d b_j b_j^* \quad \text{and} \quad \sigma_{(1,2)} = \sum_{i,j=1}^d b_i b_j^* \otimes b_j b_i^*.$$

This choice of basis furthermore allows us to write down $\mathrm{tr}_2(A)$ for $A \in M^d \otimes M^d$ explicity:

$$\mathrm{tr}_2(A) = \sum_{i=1}^d (\mathbb{1} \otimes b_i^*) \, A \, (\mathbb{1} \otimes b_i) \, .$$

Consequently we get for $A, B \in H^d$ arbitrary

$$\mathrm{tr}_2 \left( P_{\mathrm{Sym}^2} A \otimes B \right) = \frac{1}{2} \mathrm{tr}_2 \left( A \otimes B \right) + \frac{1}{2} \mathrm{tr}_2 \left( \sigma_{(1,2)} A \otimes B \right) .$$

The latter term can be evaluated explicitly:

$$\begin{aligned} \mathrm{tr}_2 \left( \sigma_{(1,2)} A \otimes B \right) &= \sum_{k=1}^d (\mathbb{1} \otimes b_k^*) \sum_{i,j=1}^d b_i b_j^* \otimes b_j b_i^* A \otimes B \, (\mathbb{1} \otimes b_k) \\ &= \sum_{i,j,k=1}^d b_i b_j^* A b_k^* b_j b_i^* B b_k = \sum_{i,j=1}^d \langle b_i, B b_j \rangle b_i b_j^* A \\ &= \left( \sum_{i=1}^d b_i b_i^* \right) B \left( \sum_{j=1}^d b_j b_j^* \right) A = \mathbb{1} B \mathbb{1} A = BA, \end{aligned}$$

and the desired result follows. Here we have used the basis representation of the identity, namely $\mathbb{1} = \sum_{i=1}^d b_i b_i^*$.