A Partial Derandomization of PhaseLift Using Spherical Designs

D. Gross · F. Krahmer · R. Kueng

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 tractable, 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 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.
Keywords  Phase retrieval · PhaseLift · Semidefinite relaxations of nonconvex quadratic programs · Non-commutative large deviation estimates · Spherical designs · Quantum information

Mathematics Subject Classification  90C25 · 49N30 · 62H12 · 60F10

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 $y_0 = \|x\|_{l_2}^2$ and amplitude measurements

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

where $a_1, \ldots, a_m \in \mathbb{C}^d$ 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 crystallography, astronomy and diffraction imaging—see for example [57]. 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 [16].

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 [8], 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: as shown in [41], it follows from the results derived in [66] that a sample size of $m \geq (4 + o(1))d$ is necessary (cf. [58]). However, finding the precise bound is still an open problem.

Balan et al. [9] consider the scenario of $O(d^2)$ measurements, which form a complex projective 2-design (cf. Definition 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 $xx^*$:

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

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 this 2-design
which is of size $o(d^2)$. In other words, for the measurement scenario described in [9], the quadratic scaling in $d$ is basically unavoidable.

To contrast these two extreme approaches, Ref. [8] 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 [9] 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.

Reference [3] introduces a reconstruction method called polarization that works for $O(d \log d)$ measurements and can handle structured measurement vectors, including the masked illumination setup that appears in diffraction imaging [10], 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 [3]. 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 [22] and reconstruction guarantees have been provided in [18, 23]. The central observation is that the matrix $xx^*$, while unknown, is certainly of rank one. This connects the phase retrieval problem with the young but already extensive field of low-rank matrix recovery [21, 35, 54, 63]. 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 $a_ia_i^*$ failed to be sufficiently incoherent in the sense of [21, 35] (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 [18, 23] for a number of measurements of optimal order $O(d)$. For recovery, they use a convex relaxation of the rank minimization problem, which makes the reconstruction algorithmically 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 [30]) and the optimization literature (for example [13]). Recently there have been attempts to provide recovery guarantees for alternating minimization algorithms [61], which are somewhat closer to the algorithms used in practice, but this 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 for stable recovery of a signal known to lie in a set \( T \subset \mathbb{R}^d \) is provided in \([28]\). 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 \([53]\) 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 \([27]\), 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.\(^1\) Designs were introduced by Delsarte, Goethals and Seidel in a seminal paper \([25]\) and have been studied in algebraic combinatorics \([70]\), coding theory \([25,60]\), and recently in quantum information theory \([4,15,36,40,68]\). Furthermore, complex projective 2-designs were the key ingredient for the reconstruction formula for phase retrieval proposed in \([9]\).

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 (\([11,19]\) vs. \([20,65]\)) and low-rank matrix recovery (\([63]\) vs. \([54]\)), respectively. Typically, the de-randomized proofs require much higher technical efforts and deliver slightly weaker results. For a recent survey on structured random measurements in signal processing see \([48]\).

\(^1\) 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 sometimes be referred to as a \( 2t \) or even a \((2t+1)\)-design. See Sect. 3.3 for our precise definition.
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 measurements with an optimal order of measurements and the construction in [9], 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 reconstructions from bases are known, but lead to somewhat suboptimal embedding dimensions.

The concept of $t$-designs, as defined in Sect. 3.3, 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 constructions has been appreciated for example in quantum information theory [4, 55]. It has been compared [4] to the notion of $t$-wise independence, which plays a role for example in the analysis of discrete randomized algorithms [56], seems to have been long appreciated in coding theory. The smallest $t$-design in $\mathbb{C}^d$ consists of $O(d^{2t})$ 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 [7, 40, 47, 69]—however, they are typically “inefficient” in the sense that they require a vector set of exponential size. For example, the construction in [40] uses $O(t)^d$ 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 [45, 46, 67, 75]. A concrete example is used for the converse bound in Sect. 7 (as well as for the converse bounds for low-rank matrix recovery from Fourier-type bases in [35]). For degree 3, both real [70] and complex [49] designs are known. For higher $t$, there are numerical methods based on the notion of the frame potential [45, 49, 64], non-constructive existence proofs [69], and constructions in sporadic dimensions (c.f. [5] and references therein).

Importantly, almost-tight randomized constructions for approximate designs for arbitrary degrees and dimensions are known [4, 15, 40]. The simplest results [40] 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 generate a design—for subsequent applications it is sufficient to sample small

---

2 While stated only for dimensions that are a power of 2, the results can be used for constructions in arbitrary dimensions [49].
subsets from it.\(^3\) Going further, there have been recent deep results on designs obtained from certain structured ensembles [15]. 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 phase retrieval. The only connection to phase retrieval to date is the estimation of pure quantum states [41,59].

Finally we point out that the notion of the frame potential above is no coincidence. In [6] a frame-theoretic approach to designs is provided, underlining their close connection.

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 [23] 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 \(\|x\|_2^2\) is known and that \(m\) measurement vectors \(a_1, \ldots, a_m\) 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 \(1 - e^{-\omega}\), PhaseLift (the convex optimization problem (24) below) recovers \(x\) up to a global phase, provided that the sampling rate exceeds

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

(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 \(O(d^{1+2/t})\). This is sub-quadratic for the regime \(t \geq 3\) where our arguments apply. If the degree is allowed to grow logarithmically with the dimension (as \(t = 2 \log d\)), we recover an optimal, linear scaling up to a polylog overhead, \(m = O(d \log^3 d)\).

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 larger than 2. 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 \(y_1, \ldots, y_m\) are sampled independently and uniformly at random from \(D_2\). Then, for any \(\omega \geq 0\), the number of measurements must obey

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

\(^3\) The situation is comparable to the use of random graphs as randomness expanders [43].
\[ m \geq \frac{\omega}{4} d(d + 1), \]

or the event

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

will occur with probability at least \( e^{-\omega} \).

It is worthwhile to put this statement in perspective with other advances in the field. Throughout our work, we have only demanded that the set of all possible measurement vectors forms a \( t \)-design and have not made any further assumptions. Theorem 2 has to be interpreted in this regard: The 2-design property alone does not allow for a sub-quadratic scaling when a “reasonably small” probability of failure is required in the recovery process.

Note that this does not exclude the possibility that certain realizations of 2-designs can perform better, if additional structural properties can be exploited. A good example for such a measurement process is the multi-illumination setup provided in [17]. In [37] the authors of this paper verified that the set of all measurement vectors used in the framework of [17] does constitute a 2-design (Lemma 6). Additional structural properties—most notably a certain correlated Fourier basis structure in the individual measurements—allowed for establishing recovery guarantees already for \( m = \mathcal{O}(d \log^4 d) \) measurements [17] and \( m = \mathcal{O}(d \log^2 d) \) [37], respectively—which both clearly are sub-quadratic sampling rates.

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 already 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 structured 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 Numerical Experiments

In this section we complement our theoretical results with numerical experiments, which we have implemented in Matlab using CVX [33,34]. As may have been expected, these experiments suggest that PhaseLift from designs actually works much
Fig. 1 Phase diagram for PhaseLift from (projected) stabilizer states, which form an exact 3-design in dimensions $2^n$ and a weighted one else. The x-axis indicates the problem’s dimension, while the y-axis denotes the number of independent design measurements performed. The frequency of a successful recovery over 30 independent runs of the experiment appears color-coded from black (zero) to white (one). To guide the eye, we have furthermore included a red line indicating $m = 4d - 4$

better than our main theorem suggests. To be concrete, we use stabilizer states—a highly structured vector set which is very prominent in quantum information theory [31,32]. Stabilizer states exist in any dimension, though their properties are somewhat better-behaved in prime power dimensions. In this case, there exists $O(d \log d)$ stabilizer state vectors. Due to their rich combinatorial structure, these vectors can be constructed efficiently. For dimensions $d = 2^n$ that are a power of two, it is known [49] that the set of stabilizer states forms a 3-design. This statement is false for other prime power dimensions ($d \neq 2^n$ for some $n$), where they only form an exact 2-design. However, weighted 3-designs can be constructed for arbitrary dimensions $d$ by projecting down stabilizer states from the next largest power-of-2-dimension $2^n$ obeying $2^{n-1} < d < 2^n$ [49]. For further clarification of the concept of exact and weighted $t$-designs we defer the reader to [68] and references therein.

We have used these vectors in our numerical simulations, the results of which are depicted in Fig. 1. For each dimension $d$ between 1 and 40 (x-axis) and for each number of measurements $m$ ranging from 1 to 200 (y-axis), we ran a total of 30 independent experiments. Each such experiment consisted in choosing a Haar-random (normalized Gaussian) complex vector $x$ as test signal. Then, we drew $m$ projected stabilizer states uniformly at random and calculated their squared overlap with the test signal. We then ran PhaseLift on this data and declared the recovery a “success” if the Forbenius distance between the reconstructed matrix $\tilde{X}$ and the true projection $X = xx^*$ was smaller than $10^{-3}$. Figure 1 depicts the empirical success probability: Black corresponds to only failures, white to only successes.
We obtain the picture of a relatively sharp phase transition along a line that scales linearly in the problem dimension. In fact, the transition seems to occur in the vicinity of the line \( m = 4d - 4 \) – drawn in red in Fig. 1. This seems to agree with the conjecture that \( 4d - 4 \) measurements are required for injectivity (see e.g. [41]). However, there are a few differences in the problem setup: Firstly, the conjecture only asks whether there is a unique solution, while the numerical simulations study whether the PhaseLift algorithm can find it. Secondly, the conjecture concerns unique solutions for all possible inputs, while numerically, we estimate the success probability. And thirdly, the conjecture states that generic measurements work, while our simulations use a specific random procedure (drawn uniformly from a 3-design) to generate them.

3 Technical Background and Notation

3.1 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

\[
\| z \|_2 = \sqrt{\langle z, z \rangle} \quad \forall z \in V^d.
\]

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

\[
(Z, Y) = \text{tr}(ZY),
\]

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

\[
\| Z \|_1 = \text{tr}(|Z|) \quad \text{(trace norm)},
\]

\[
\| Z \|_2 = \sqrt{\text{tr}(Z^2)} \quad \text{(Frobenius norm)},
\]

\[
\| Z \|_\infty = \sup_{y \in V^d} \frac{\| Z y \|_2}{\| y \|_2} \quad \text{(operator norm)},
\]

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 [12] (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 encounter matrix-valued operators acting on the space $H^d$. We label such objects with capital calligraphic letters and introduce the operator norm

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

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

$$\mathcal{I} : H^d \rightarrow H^d \quad Z \mapsto Z \quad \forall Z \in H^d$$

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

$$\Pi_Y : H^d \rightarrow H^d \quad Z \mapsto Y(Y, Z) = Y \text{tr}(YZ) \quad \forall Z \in H^d.$$

The operator

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

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

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

The notion of positive-semidefiniteness directly translates to matrix valued operators. Concretely, we call $M$ positive-semidefinite ($M \geq 0$) if $(Z, MZ) \geq 0$ for all $Z \in H^d$. Again, this induces a partial ordering. Like in the matrix case, we write $N \geq M$, if $N - M \geq 0$. It is easy to check that all the operators introduced so far are positive semidefinite and in particular we obtain the ordering
\[ 0 \leq \Pi_1 \leq dI, \quad (5) \]

by using (4).

### 3.2 Multilinear Algebra

The properties of \( t \)-designs are most naturally stated in the framework of \((t\text{-fold})\) 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 [51,74].

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 \( f : V_1 \times \cdots \times V_k \rightarrow \mathbb{C} \)

is \textit{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 \textit{tensor product} of \( V_1^*, \ldots, V_k^* \). Consequently, the tensor product \( (V^d)^{\otimes k} = \bigotimes_{i=1}^k V^d \) is the space of all multilinear functions

\[
f : \underbrace{(V^d)^* \times \cdots \times (V^d)^*}_{k \text{ times}} \rightarrow \mathbb{C},
\]

and we call the elementary elements \( z_1 \otimes \cdots \otimes z_k \) the \textit{tensor product} of the vectors \( z_1, \ldots, z_k \in V^d \). Such an element can alternatively be defined more concretely via the \textit{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 \( V^d \rightarrow V^d \) (\( d \times d \)-matrices) corresponds to the tensor product \( M^d := V^d \otimes (V^d)^* \) which is spanned by \( \{ y \otimes z^* : y, z \in V^d \} \)—the set of all rank-1 matrices. For this generating set of \( M^d \), we define the \textit{trace} to be the natural bilinear map

\[ \text{tr} : V^d \otimes (V^d)^* \rightarrow \mathbb{C} \]

\[ (y \otimes z^*) \mapsto z^* y = \langle z, y \rangle \]

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

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

\[
f : \underbrace{((V^d)^* \times V^d) \times \cdots \times ((V^d)^* \times V^d)}_{k \text{ times}} \rightarrow \mathbb{C}
\]
in complete analogy to the above. We call the elements $Z_1 \otimes \cdots \otimes Z_k$ the tensor product of the matrices $Z_1, \ldots, Z_k \in M^d$.

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

$$\text{tr}_i : \left(M^d\right)^\otimes k \to \left(M^d\right)^\otimes (k-1)$$

$$Z_1 \otimes \cdots \otimes Z_k \mapsto \text{tr}(Z_i) \left(Z_1 \otimes \cdots \otimes Z_{i-1} \otimes Z_{i+1} \otimes \cdots \otimes Z_k\right).$$

Note that with the identification $M^d = V^d \otimes (V^d)^*$, $\text{tr}_i$ 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 $1 \leq i < j \leq k$

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

In particular, the full trace then corresponds to

$$\text{tr} := \text{tr}_{1,\ldots,k} : \left(M^d\right)^\otimes k \to \mathbb{C}$$

$$(Z_1 \otimes \cdots \otimes Z_k) \mapsto \text{tr}(Z_1) \ldots \text{tr}(Z_k).$$

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

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

where $S_k$ denotes the group of permutations of $k$ elements. This map projects $(V^d)^\otimes k$ onto the totally symmetric subspace $\text{Sym}^k$ of $(V^d)^\otimes k$ whose dimension [51] is

$$\dim \text{Sym}^k = \binom{d+k-1}{k}.$$

3.3 Complex Projective Designs

The idea of (real) spherical designs originates in coding theory [25] and has been extended to more general spaces in [42,52,62]. We refer the interested reader to Levienstein [52] 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. [42,46,62]—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 [68]) A finite set \( \{w_1, \ldots, w_N\} \subset V^d \) of normalized vectors is called a \( t \)-design of dimension \( d \) if and only if

\[
\frac{1}{N} \sum_{i=1}^{N} (w_i w_i^*)^{\otimes t} = \dim(\text{Sym}^t)^{-1} P_{\text{Sym}^t},
\]

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

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

\[
\frac{1}{N} \sum_{i=1}^{N} (w_i w_i^*)^{\otimes t} = \int_w (ww^*)^{\otimes t},
\]

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. Seymour and T. Zaslavsky proved in [69] 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. [42,46]—for the \( t \)-design property is that the number of vectors \( N \) obeys

\[
N \geq \left( d + \lceil t/2 \rceil - 1 \right) \left( d + \lfloor t/2 \rfloor - 1 \right) = O(d^{2t}).
\]

However, the proof in [69] is non-constructive and known constructions are “inefficient” in the sense that the number of vectors required greatly exceeds (10). Hayashi et al. [40] proposed a construction requiring \( O(t) \) vectors. For real spherical designs other “inefficient” constructions have been proposed \([7,47]\) \((N = t O(d^2))\) which can be used to obtain complex projective designs.

Addressing this apparent lack of efficient constructions, Ambainis and Emerson [4] proposed the notion of approximate desings. These vector sets 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 \( \epsilon = O(d^{-1/3}) \) approximate \( t \)-design consisting of \( O(d^{3t}) \) vectors only.

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

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

\copyright Birkhäuser
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 [23].

3.4 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 [2]. Later the first author of this paper and his coworkers successfully applied these methods to the problem of low rank matrix recovery [35,38]. By now these methods are widely used and we borrow them in their most recent (and convenient) form from Tropp [71,72].

**Theorem 4** (Uniform Operator Bernstein inequality, [35,72]) Consider a finite sequence \(\{M_k\}\) of independent, random self-adjoint operators. Assume that each random variable satisfies \(\mathbb{E}[M_k] = 0\) and \(\|M_k\|_{\infty} \leq R\) (for some finite constant \(R\)) almost surely and define the norm of the total variance \(\sigma^2 := \| \sum_k \mathbb{E}[M_k^2] \|_{\infty}\). Then the following chain of inequalities holds for all \(t \geq 0\).

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

**Theorem 5** (Smallest Eigenvalue Bernstein Inequality, [71]) Let \(S = \sum_k M_k\) be a sum of iid random matrices \(M_k\) which obey \(\mathbb{E}[M_K] = 0\) and \(\lambda_{\min}(M_k) \geq -R\) almost surely for some fixed \(R\). With the variance parameter \(\sigma^2(S) = \| \sum_k \mathbb{E}[M_k^2] \|_{\infty}\) the following chain of inequalities holds for all \(t \geq 0\).

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

3.5 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 [51] for an introduction and also [24,73] (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 \(A : V^d \rightarrow V^d\) can be seen as a two-indexed tensor \(A_{i,j}\). It will thus be represented by a node \(\left[\begin{array}{c} A \\ i \end{array}\right]_{j}\) 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

$$AB = \begin{array}{c} A \\ B \end{array}$$

The trace operation

$$A \mapsto \operatorname{tr} A = \sum_k A^k_k$$

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

$$\operatorname{tr}(A) = \begin{array}{c} A \end{array}.$$  

Tensor products are arranged in parallel:

$$A \otimes B = \begin{array}{c} A \\ B \end{array}.$$ 

Hence, a partial trace takes the following form:

$$\operatorname{tr}_2 (A \otimes B) = \begin{array}{c} A \\ B \end{array}.$$  

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

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

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

$$1 = \begin{array}{c} & \\ & \end{array} \quad \text{(trivial permutation)} \quad \text{and} \quad \sigma_{(1,2)} = \begin{array}{c} & \\ & \end{array},$$

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 $(V^d)^{\otimes 2}$ as an example. Using wiring diagrams we can derive the standard result

$$\sigma^2_{(1,2)} = \begin{array}{c}
\end{array} = 1$$

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

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

(11)

We remark that in general,

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

which is, in our experience, a common misconception.

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

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

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

$$P_{\text{Sym}^2} = \frac{1}{2} \begin{array}{c}
\end{array}.$$

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

$$\text{tr}_2 \left( P_{\text{Sym}^2} A \otimes B \right) = \frac{1}{2} \left( \begin{array}{c}
\end{array} + \begin{array}{c}
\end{array} \right) = \frac{1}{2} \left( \begin{array}{c}
\end{array} + \begin{array}{c}
\end{array} \right) = \frac{1}{2} (\text{tr}(B)A + BA),$$

which is the desired result. \qed
Obviously, it is also possible to obtain (11) by direct calculation. We have included such a calculation in the appendix (Sect. 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

$$
\text{tr}_{2,3} \left( P_{\text{Sym}^3} A \otimes B \otimes C \right) = \frac{1}{6} \left( A \text{ tr}(B)\text{ tr}(C) + BA \text{ tr}(C) + C A \text{ tr}(B) \\
+ A \text{ tr}(B C) + C B A + B C A \right).
$$

(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

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

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

$$
P_{\text{Sym}^3} = \frac{1}{6} \left( \begin{array}{c}
A \ B \ C \\
A \ B \ C \\
A \ B \ C \\
A \ B \ C \\
\end{array} \right)
$$

and we are done.

\[ \square \]

## 4 Problem Setup

### 4.1 Modelling the Sampling Process

In the sampling process, we start by measuring the intensity of the signal:

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

(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

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

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

\[
A : H^d \to \mathcal{R}^m, \quad Z \mapsto \sum_{i=1}^{m} \text{tr}(A_i Z) e_i,
\]

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

\[
\mathcal{R} : H^d \to H^d, \quad Z \mapsto m^{-1} \sum_{i=1}^{m} (d + 1) d \, \Pi_A Z = m^{-1} \sum_{i=1}^{m} (d + 1) d \, A_i \, \text{tr}(A_i Z),
\]

which is a renormalized version of \( A^* A : H^d \to H^d \). Concretely

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

The scaling is going to greatly simplify our analysis, because it guarantees that \( \mathcal{R} \) is “near-isotropic”, as the following result shows.

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

\[
\mathbb{E}[\mathcal{R}] = \mathcal{I} + \Pi_\perp \quad \text{or} \quad \mathbb{E}[\mathcal{R}] Z = Z + \text{tr}(Z) \perp \quad \forall Z \in H^d
\]

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

\[
\mathbb{E}[\mathcal{R}] Z = \frac{(d + 1) d}{m} \sum_{i=1}^{m} \mathbb{E}[A_i \, \text{tr}(A_i Z)]
\]

\[
= (d + 1) d \, \text{tr}_2 \left( \mathbb{E}[A_i^\otimes 2] \perp \otimes Z \right)
\]

\[
= 2 \text{tr}_2 \left( P_{\text{Sym}^2} \perp \otimes Z \right)
\]

\[
= Z + \perp (\text{tr} Z) = (\mathcal{I} + \Pi_\perp) Z.
\]
Here, (18) follows from the fact that the $a_i$’s are chosen iid from a $t$-design, (19) uses the fact that $\dim(\text{Sym}^2) = \binom{d+1}{2}^{-1}$ together with Definition 3, and the final line is an application of Lemma 6. 

Let now $x \in V^d$ be the signal we want to recover. As in [23] we consider the space

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

(20)

(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:

$$P_T : H^d \rightarrow T, \quad Z \mapsto XZ + ZX - XZX$$

(21)

$$= XZ + ZX - (X, Z)X.$$  

(22)

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

$$Z = P_T Z + P_T^\perp Z =: Z_T + Z_T^\perp$$

is valid. We point out that in particular

$$P_T \Pi_{\perp} P_T = \Pi_X$$

(23)

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

$$P_T \Pi_{\perp} P_T Z = P_T \Pi (1P_T Z) = (X \Pi + \Pi X - X \Pi X) \text{tr} (XZ + ZX - XZX)$$

$$= X \text{tr} (XZ) = \Pi_X Z.$$

4.2 Convex Relaxation

Following [8, 18, 23] 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 a linear form:

$$y_0 = \|x\|_2^2 = (1, X) = \text{tr}(X),$$

$$y_i = (A_i, X) = \text{Tr}(A_i X) \quad i = 1, \ldots, m.$$
Hence, the phase retrieval 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

$$\min_{X'} \|X'\|_1 \quad \text{subject to} \quad (A_i, X') = y_i \quad i = 1, \ldots, m,$$

$$X' = (X')^+, \quad \text{tr}(X') = 1,$$

$$X' \succeq 0.$$  \hfill (24)

While this convex program is formally equivalent to the previously studied general-purpose matrix recovery algorithms [21,35,63], 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 [35], it means that the “incoherence parameter” $\mu = d \max_{i=1, \ldots, m} \|A_i\|_\infty = d \|d_i\|_2 = d$ 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. [18,23] 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 [35].

### 4.3 Well-Posedness/Injectivity

In this section, we follow [23,35] to establish a certain injectivity property of the measurement operator $\mathcal{A}$. Compared to [23], 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 Sect. 6. 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 [23]. In the language of [35], we will have to carry out the full “golfing scheme”, as opposed to the “single leg” that proved sufficient in [23].

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

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

is valid for all matrices $Z \in T$ simultaneously.
Proof We aim to show the more general statement

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

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

For $Z \in T$ arbitrary use near-isotropy of $\mathcal{R}$ ($\mathbb{E}[\mathcal{R}] = \mathcal{I} + \Pi_1$) and observe

$$m^{-1}\|A(Z)\|_2^2 = m^{-1} \sum_{i=1}^{m} (\text{tr}(ZA_i))^2 = \text{tr}(Zm^{-1} \sum_i A_i \text{tr}(A_i Z)) = \frac{1}{(d+1)d} \text{tr}(ZRZ)$$

$$= \frac{1}{(d+1)d} \text{tr}(Z(\mathcal{R} - \mathbb{E}[\mathcal{R}])Z) + \frac{1}{(d+1)d} \text{tr}(Z(\mathcal{I} + \Pi_1)Z)$$

$$= \frac{1}{(d+1)d} \text{tr}(ZP_T(\mathcal{R} - \mathbb{E}[\mathcal{R}])P_T Z) + \frac{1}{(d+1)d} (\text{tr}(Z^2) + (\text{tr} Z)^2)$$

$$\geq 0.5d^{-2} \left( \text{tr}(ZP_T(\mathcal{R} - \mathbb{E}[\mathcal{R}])P_T Z) + \text{tr}(Z^2) \right)$$

$$\geq 0.5d^{-2}(1 + \lambda_{\text{min}}(P_T(\mathcal{R} - \mathbb{E}[\mathcal{R}])P_T)\|Z\|_2^2. \quad \text{(26)}$$

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

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

Note that these summands have mean zero by construction. Furthermore observe that the auxiliary result (23) implies

$$-\frac{2}{m} \mathcal{I} \leq -\frac{1}{m} \mathcal{I} - \frac{1}{m} \Pi_X \leq -\frac{1}{m} P_T \mathcal{I} P_T - \frac{1}{m} P_T \Pi_1 P_T$$

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

and the a priori bound

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

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

$$0 \leq \mathbb{E}[\mathcal{M}_i^2] = \frac{(d+1)^2 d^2}{m^2} P_T \mathbb{E} \left[ \Pi_{A_i} P_T \Pi_{A_i} \right] P_T$$

$$= \frac{(d+1)^2 d^2}{m^2} P_T \mathbb{E} \left[ \text{tr}(A_i P_T A_i) \Pi_{A_i} \right] P_T.$$
The trace can be bounded from above by
\[
\text{tr}(A_i \mathcal{P}_T A_i) = \text{tr} \left( A_i (X A_i + A_i X - \text{tr}(A_i) X) \right) \\
= 2 \text{tr}(A_i X) - \text{tr}(A_i X)^2 \leq 2 \text{tr}(A_i X),
\]
where we have used the basic definition of \( \mathcal{P}_T \) and \( 0 \leq \text{tr}(A_i X) = | \langle a_i, x \rangle |^2 \leq 1. \)

Consequently, for \( Z \in T \) arbitrary
\[
\mathcal{P}_T \mathbb{E}[M_i^2] \mathcal{P}_T Z \leq \frac{2(d + 1)^2 d^2}{m^2} \mathcal{P}_T \mathbb{E}[A_i \text{tr}(A_i X) \text{tr}(A_i Z)] \\
= \frac{2(d + 1)^2 d^2}{m^2} \mathcal{P}_T \text{tr}_{2,3} \left( \mathbb{E}[A_i^3] \mathbb{1} \otimes X \otimes Z \right) \\
= \frac{12(d + 1)^2 d^2}{m^2(d + 2)(d + 1)d} \mathcal{P}_T \text{tr}_{2,3} \left( \mathbb{P}_{\text{Sym}^3} \mathbb{1} \otimes X \otimes Z \right) \\
\leq \frac{2d}{m^2} \mathcal{P}_T \left( \mathbb{1} \text{tr}(Z) + X \text{tr}(Z) + Z + \mathbb{1} \text{tr}(XZ) + ZX + XZ \right) \\
= \frac{2d}{m^2} \left( X \text{tr}(X Z) + X \text{tr}(X Z) + Z + X \text{tr}(X Z) + \mathcal{P}_T Z + X \text{tr}(X Z) \right) \\
= \frac{2d}{m^2} \left( 4 \mathbb{I} X + 2 \mathbb{I} \right) Z \leq \frac{12d}{m^2} \mathbb{I} Z.
\]

Here we have applied \( \dim \text{Sym}^3 = \left( \frac{d + 2}{3} \right)^{-1} \) and Lemma 7 in lines 3 and 4, respectively. Furthermore we used \( Z \in T \)—hence \( \mathcal{P}_T Z = Z \) and \( \text{tr}(Z) = \text{tr}(XZ) \)—as well as the basic definition (22) of \( \mathcal{P}_T \) to simplify the terms occurring in the fourth line. Putting everything together yields
\[
\mathbb{E}[ (M_i - \mathbb{E}[M_i])^2 ] \leq \mathbb{E}[M_i^2] \leq \frac{12d}{m^2} \mathbb{I}
\]
and we can safely set \( \sigma^2 := \frac{12d}{m^2} \). Now Theorem 5 tells us
\[
\Pr[ \lambda_{\min}(\mathcal{P}_T (\mathbb{R} - \mathbb{E}[\mathbb{R}]) \mathcal{P}_T) \leq -\delta ] \leq d^2 \exp \left( -\frac{3m \delta^2}{8 \times 12d} \right)
\]
for all \( 0 \leq \delta \leq 1 \leq 6d = \sigma^2 / \mathbb{R} \). This gives the desired bound on the event
\[
\{ \lambda_{\min}(\mathcal{P}_T (\mathbb{R} - \mathbb{E}[\mathbb{R}]) \mathcal{P}_T) \leq -\delta \}
\]
occuring. If this is not the case, (26) implies
\[
m^{-1} \| A(Z) \|_{\ell_2}^2 > 0.5d^{-2}(1 - \delta) \| Z \|_{\ell_2}^2
\]
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. \( \square \)
Proposition 10 Let $A$ be as above with vectors sampled from a $t$-design ($t \geq 1$). Then the statement
\[ m^{-1} \|A(Z)\|_{\ell_2}^2 \leq \|Z\|_2^2 \] (27)
holds with probability one for all matrices $Z \in H^d$ simultaneously.

Proof Pick $Z \in H^d$ arbitrary and observe
\[ \|A(Z)\|_{\ell_2}^2 = \frac{1}{m} \sum_{i=1}^{m} (\text{tr}(A_i Z))^2 = \text{tr} \left( Z \left( \frac{1}{m} \sum_{i=1}^{m} \Pi_{A_i} \right) Z \right) \leq \text{tr}(ZZ) = \|Z\|_2^2, \]
where we have used $0 \leq \Pi_{A_i} \leq I$.

Note that equation (27) can be improved. Indeed, a standard application of the Operator Bernstein inequality (Theorem 4) gives
\[ m^{-1} \|A(Z)\|_{\ell_2}^2 \leq 2d^{-1} \|Z\|_2^2 \]
for all matrices $Z \in T$ with probability of failure smaller than $d^2 \exp(-Cm/d)$ for some $0 < C \leq 1$. However, we actually do not require this tighter bound.

5 Proof of the Main Theorem/Convex Geometry

In this section, we will follow [21,35] to prove that the convex program (24) 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 $Y \in \text{span}(\mathbb{1}, A_1, \ldots, A_m)$ and
\[ \|Y_T - X\|_2 \leq \frac{1}{4d} \quad \text{as well as} \quad \|Y_T^\perp\|_\infty \leq \frac{1}{2}. \] (28)

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 (24) recovers the unknown $x$ (up to a global phase) provided that (25) holds and an approximate dual certificate $Y$ exists.

Proof Let $\tilde{X} \in H^d$ be an arbitrary feasible point of (24) and decompose it as $\tilde{X} = X + \Delta$. Feasibility then implies $A(\tilde{X}) = A(X)$ and $A(\Delta) = 0$ must in turn hold for any feasible displacement $\Delta$. Now the pinching inequality [14] (Problem II.5.4) implies
\[ \|\tilde{X}\|_1 = \|X + \Delta\|_1 \geq \|X\|_1 + \text{tr}(\Delta_T) + \|\Delta_T^\perp\|_1. \]
Consequently $X$ is guaranteed to be the unique minimum of (24), if
\[ \text{tr}(\Delta_T) + \| \Delta_T^1 \|_1 > 0 \] (29)
is true for every feasible $\Delta$. In order to show this we combine feasibility of $\Delta$ with inequalities (25) and (27) to obtain
\[ \| \Delta_T \|_2 < 2dm^{-1/2} \| A(\Delta_T) \|_{\ell_2} = 2dm^{-1/2} \| A(\Delta_T^1) \|_{\ell_2} \leq 2d \| \Delta_T^1 \|_2. \] (30)
Feasibility of $\Delta$ also implies $(Y, \Delta) = 0$, because by definition $Y$ is in the range of $A^*$. Combining this insight with the defining property (28) of $Y$ and (30) yields
\[
0 = (Y, \Delta) = (Y_T - X, \Delta_T) + (X, \Delta_T) + (Y^1_T, \Delta_T^1) \\
\leq \| Y_T - X \|_2 \| \Delta_T \|_2 + \text{tr}(\Delta_T) + \| Y^1_T \|_{\ell_\infty} \| \Delta_T^1 \|_1 \\
< \text{tr}(\Delta_T) + \| Y_T - X \|_2 2d \| \Delta_T^1 \|_2 + \| Y^1_T \|_{\ell_\infty} \| \Delta_T^1 \|_1 \\
\leq \text{tr}(\Delta_T) + 1/2 \| \Delta_T^1 \|_2 + 1/2 \| \Delta_T^1 \|_1 \\
\leq \text{tr}(\Delta_T) + \| \Delta_T^1 \|_1,
\]
which is just the desired optimality criterion (29).

\section{Constructing the Dual Certificate}

A straightforward approach to construct an approximate dual certificate would be to set
\[
Y = R X - \text{tr}(X) \mathbb{1} = \frac{(d + 1)d}{m} \sum_{i=1}^{m} A_i \text{tr}(A_i X) - \text{tr}(X) \mathbb{1} \in \text{span} (\mathbb{1}, A_1, \ldots, A_m).
\] (31)
In expectation, $\mathbb{E}[Y] = X$, which is the “perfect dual certificate” in the sense that the norm bounds in (28) vanish. The hope would be to use the Operator Bernstein inequality to show that with high probability, $Y$ will be sufficiently close to its expectation. It has been shown that a slight refinement of the ansatz (31) indeed achieves this goal Ref. [35,50]. 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. Reference [23] relied on the fact that large overlaps $|\langle a_i, x \rangle|^2 \gg O(d^{-1})$ 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.

\textbf{Lemma 13} (Undesired events) \textit{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 $\gamma \leq 1$:}
\[
\Pr \left[ |⟨a, x⟩|^2 \geq 5td^{-\gamma} \right] \leq 4^{-t} d^{-(1-\gamma)}. \tag{32}
\]

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

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

which is valid for any \( \delta \geq 1 \). Setting \( \delta = 4 \) then yields (32). The \( t \)-design property provides us with useful information about the first \( t \) moments of the non-negative random variable \( ξ = |⟨a, x⟩|^2 \). Indeed, with \( A = aa^* \) it holds for every \( k \leq t \) that

\[
E[ξ^k] = E[tr(AX)^k] = tr(E[A^k]X^k) = \binom{d+k-1}{k}^{-1} tr(P_{Sym^k}X^k) = \binom{d+k-1}{k}^{-1} tr(X^k) \leq d^{-k} k!,
\]

because \( X^k \) is invariant under \( P_{Sym^k} \). One way of seeing this is to note that \( \text{range}(X^k) = \text{span}(x^k) \) and the latter is already contained in \( \text{Sym}^k \). Therefore the \( k \)-th moment \( τ_k \) of \( ξ \) is bounded by

\[
τ_k = \left( E[ξ^k] \right)^{1/k} \leq (d^{-k} k!)^{1/k} \leq k/d.
\]

These inequalities are tight for the mean \( μ = τ_1 \) of \( ξ \) and hence

\[
μ = E[ξ] = d^{-1}.
\]

Now we aim to use the well-known \( t \)-th moment bound

\[
\Pr \left[ |ξ - μ| \geq sτ_t \right] \leq s^{-t},
\]

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

---

4 Alternatively one could also rearrange tensor systems: \( X^k = (xx^*)^k \simeq x^k (x^*)^k \) and use \( P_{Sym^k} x^k \simeq x^k \).
\[
\Pr \left[ |\langle a, x \rangle|^2 \geq (\delta + 1)td^{-\gamma} \right] = \Pr \left[ |\xi - \mu|^2 \geq (\delta + 1)td^{-\gamma} - d^{-1} \right] \\
\leq \Pr \left[ |\xi - \mu| \geq \delta td^{-\gamma} \right] \\
\leq \Pr \left[ |\xi - \mu| \geq \delta d^{1-\gamma} \tau_t \right] \\
\leq \delta^{-t}d^{-(1-\gamma)},
\]
and we are done. \(\Box\)

The previous lemma bounds the probability of the undesired events

\[ E_i^c = \left\{ |\langle a_i, x \rangle|^2 \geq 5td^{-\gamma} \right\}, \tag{33} \]

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

\[ Z = \zeta (xz^* + zx^*) , \]

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

\[ R_Z := R_{x, y} = (d + 1)d m \sum_{i=1}^{m} 1_{E_i} 1_{G_i} \Pi_{A_i}, \tag{34} \]

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 \(R\).

**Proposition 15** Fix \(Z \in T\) arbitrary and let \(R_Z\) be as in (34). Then

\[ \|E[R_Z - R]\|_{\text{op}} \leq 4^{-t}d^{2-t(1-\gamma)} \tag{35} \]

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

\[ R_{\text{aux}} := (d + 1)d m \sum_{i=1}^{m} 1_{E_i} \Pi_{A_i} \]

\(\text{Birkhäuser}\)
and observe
\[
\| \mathbb{E} [\mathcal{R} Z - \mathcal{R}] \|_{\text{op}} \leq \| \mathbb{E} [\mathcal{R} - \mathcal{R}_{\text{aux}}] \|_{\text{op}} + \| \mathbb{E} [\mathcal{R} Z - \mathcal{R}_{\text{aux}}] \|_{\text{op}}. \tag{36}
\]
Now use Lemma 13 to bound the first term:
\[
\| \mathbb{E} [\mathcal{R} - \mathcal{R}_{\text{aux}}] \|_{\text{op}} = \left\| \frac{(d + 1)d}{m} \sum_{i=1}^{m} \mathbb{E} \left[ (1 - 1_{E_i}) \Pi A_i \right] \right\|_{\text{op}} \\
\leq \frac{(d + 1)d}{m} \sum_{i=1}^{m} \mathbb{E} \left[ 1_{E_i} \| \Pi A_i \|_{\text{op}} \right] \\
\leq \frac{2d^2}{m} \sum_{i=1}^{m} \mathbb{E} \left[ 1_{G_i} \right] = \frac{2d^2}{m} \sum_{i=1}^{m} \text{Pr} \left[ E_i \right] \\
\leq 2d^2 \times 4^{-t} d^{-t(1 - \gamma)} = 2^{1-2t} d^{2-t(1-\gamma)}.
\]
Similarly,
\[
\| \mathbb{E} [\mathcal{R}_{\text{aux}} - \mathcal{R} Z] \|_{\text{op}} \leq \frac{(d + 1)d}{m} \left\| \sum_{i=1}^{m} \mathbb{E} \left[ 1_{G_i} \Pi A_i \right] \right\|_{\text{op}} \\
\leq \frac{2d^2}{m} \sum_{i=1}^{m} \text{Pr} \left[ G_i \right] \leq 2^{1-2t} d^{2-t(1-\gamma)}
\]
and inserting these bounds into (36) yields the desired statement. \(\Box\)

We now establish a technical result which will allow us to find a suitable approximate
dual certificate using the “golfing scheme” construction \([35, 50]\).

**Proposition 16** Fix \(Z \in T\) arbitrary, let \(\mathcal{R} Z\) be as in (34). 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
\[
\| P_T^{\perp} (\mathcal{R} Z - \text{tr}(Z) \mathbb{1}) \|_{\infty} \leq b \| Z \|_2 \quad \text{and} \tag{37}
\]
\[
\| P_T (\mathcal{R} Z - Z - \text{tr}(Z) \mathbb{1}) \|_2 \leq c \| Z \|_2. \tag{38}
\]

**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
\[
Z = \zeta (z x^* + x z^*)
\]
with some fixed \( z \in V^d \) obeying \( \|z\|_{\ell_2} = 1 \) and \( 0 < \xi \leq 1 \). Near-Isotropicty (Lemma 8) of \( \mathcal{R} \) guarantees \( \mathcal{P}_T^\perp \mathbb{E}[\mathcal{R}] Z = \text{tr}(Z) \mathcal{P}_T^\perp \mathbb{1} \) as well as \( \mathcal{P}_T \mathbb{E}[\mathcal{R}] Z = Z + \text{tr}(Z) \mathcal{P}_T \mathbb{1} \). Let us now focus on (37) and use Proposition 15 in order to write

\[
\|\mathcal{P}_T^\perp (\mathcal{R} Z - \text{tr}(Z) \mathbb{1})\|_\infty = \|\mathcal{P}_T^\perp (\mathcal{R} Z - \mathbb{E}[\mathcal{R}] Z)\|_\infty
\]

\[
\leq \|\mathcal{P}_T^\perp (\mathcal{R} Z - \mathbb{E}[\mathcal{R}] Z)\|_\infty + \|\mathcal{P}_T^\perp \mathbb{E}[\mathcal{R} Z - \mathcal{R}] Z\|_\infty
\]

\[
\leq \|\mathcal{P}_T^\perp\|_{\text{op}} \|\mathcal{R} Z - \mathbb{E}[\mathcal{R} Z\right\|_\infty + 4^{1-t}d^{2-2t(1-\gamma)} \|\mathcal{P}_T^\perp\|_{\text{op}} \|Z\|_2
\]

\[
\leq (\mathcal{R} Z - \mathbb{E}[\mathcal{R} Z])\|_\infty + b/4.
\]

Here we have used \( \|\mathcal{P}_T^\perp\|_{\text{op}} \leq 1 \) as well as

\[
\|\mathbb{E}[\mathcal{R} Z - \mathcal{R}]\|_{\text{op}} \leq 4^{1-t}d^{2-2t(1-\gamma)} \leq 4^{1-t} \leq 1/16 \leq b/4,
\]

which follows from \( \gamma \leq 1 - 2/t, t \geq 3 \) and \( b \geq 1/4 \). To obtain (38) we use a similar reasoning:

\[
\|\mathcal{P}_T (\mathcal{R} Z - Z - \text{tr}(Z) \mathbb{1})\|_2 = \|\mathcal{P}_T (\mathcal{R} Z - \mathbb{E}[\mathcal{R}] Z)\|_2
\]

\[
\leq \sqrt{2}\|\mathcal{P}_T (\mathcal{R} Z - \mathbb{E}[\mathcal{R}] Z)\|_\infty + \|\mathcal{P}_T \mathbb{E}[\mathcal{R} Z - \mathcal{R}] Z\|_2
\]

\[
\leq \sqrt{2}\|\mathcal{P}_T\|_{\text{op}} \|\mathcal{R} Z - \mathbb{E}[\mathcal{R} Z\right\|_\infty + b/4\|\mathcal{P}_T\|_{\text{op}} \|Z\|_2
\]

\[
\leq \sqrt{2}\| (\mathcal{R} Z - \mathbb{E}[\mathcal{R} Z])\|_\infty + b/4,
\]

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 (39) in the fourth. This motivates to define the event

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

which guarantees both (37) and (38) due to the assumption on \( c \) and \( \|Z\|_2 = 1 \). So everything boils down to bounding the probability of \( E^c \). We decompose

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

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

\[
\|M_i\|_{\infty} = \frac{(d + 1)d}{m} E_i 1_{G_i} \|A_i\|_{\infty} |\text{tr}(A_i Z)| \leq \frac{2d^2}{m} E_i 1_{G_i} 2|\langle x, a_i \rangle||\langle z, a_i \rangle|
\]

\[
\leq \frac{4d^2}{m} 5t d^{-\gamma} = \frac{20}{m} t d^{2-\gamma} =: \overline{R},
\]

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

$$\mathbb{E} \left[ M_i^2 \right] = \frac{(d+1)^2 d^2}{m^2} \mathbb{E} \left[ 1_{E_i} 1_{G_i} \right] \text{tr}(A_i Z)^2 A_i^2 \leq \frac{(d+1)^2 d^2}{m^2} \text{tr} \left( A_i Z \right)^2 A_i^2$$

$$= \frac{(d+1)^2 d^2}{m^2} \text{tr}_{2,3} \left( \mathbb{E} [A_i \otimes] \otimes Z \otimes Z \right) = \frac{6(d+1)d}{m^2(d+2)} \text{tr}_{2,3} \left( P_{\text{Sym}^3} \otimes Z \otimes Z \right)$$

$$= \frac{(d+1)d}{m^2(d+2)} \left( \mathbb{1} \text{tr}(Z)^2 + Z \text{tr}(Z) + Z + \mathbb{1} \text{tr}(Z^2) + 2Z^2 \right)$$

$$\leq \frac{8d}{m^2} \|Z\|_2^2 = \frac{8d}{m^2}.$$ 

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

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

Observing that

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

Theorem 4 yields

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

as desired. \( \square \)

With this ingredient we can now construct a suitable approximate dual certificate \( Y \), closely following [50].

**Proposition 17** Let \( x \in V^d \) be an arbitrary normalized vector \( \|x\|_2 = 1 \), \( X = xx^* \) and let \( \omega \geq 1 \) be arbitrary. If the design order \( t \) \( (t \geq 3) \) and the truncation rate \( \gamma \) are chosen such that

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

holds and the total number of measurements fulfills

$$m \geq C \omega t d^{2-\gamma} \log^2(d), \quad (40)$$

then with probability larger than \( 1 - 0.5 e^{-\omega} \), there exists an approximate dual certificate \( Y \) as in Definition 11. Here, \( C \) is a universal constant (which can in principle be recovered explicitly from the proof).
Algorithm 1: Summary of the randomized “golfing scheme” [35] used in the proof of Proposition 17 to show the existence of an approximate dual certificate.

**Input:**
- $X \in H^d$  
  # signal to be recovered  
- $l \in \mathbb{N}$  
  # maximum number of iterations  
- $\{m_i\}_{i=1}^{l} \subset \mathbb{N}$  
  # number of measurement vectors used in $i$th iteration  
- $r$  
  # require $r$ “successful” iterations  
  # (i.e. iterations where we enter the inner if-block)

**Initialize:**
- $Y = [\ ]$  
  # a list of matrices in $H^d$, initially empty  
- $Q = [X]$  
  # a list of matrices in $T$, initialized to hold $X$ as its only element  
- $i = 1$  
  # number of current iteration  
- $\xi = [0, \ldots, 0]$  
  # array of $l$ zeros; $\xi_i$ will be set to 1 if $i$th iteration succeeds

**Body:**

while $i \leq l$ and $\sum_{j=1}^{i} \xi_j \leq r$
do

- set $Q$ to be the last element of $Q$ and $Y$ to be the last element of $Y$,  
- sample $m_i$ vectors uniformly from the $t$-design; construct $R_Q$ according to Definition 14.

  if (37), (38) hold for $R_Q$ and $Q \in T$ with parameters $b = 1/4$, $c = 1/2$ then

    $\xi_i = 1$

    $Y \leftarrow R_Q Q - \text{tr}(Q) \mathbb{1} + Y$,  
    append $Y$ to $Y$

    $Q \leftarrow X - P_T Y$,  
    append $Q$ to $Q$

  end

  $i \leftarrow i + 1$

end

if $\sum_{i=1}^{l} \xi_i = r$ then

  report success and output $Y$, $Q$, $\xi$

else

  report failure

end

**Proof** The randomized construction of $Y$ is summarized in Algorithm 1. If this algorithm succeeds, it outputs three lists:

$$Y = [Y_1, \ldots, Y_r], \quad Q = [X, Q_1, \ldots, Q_r], \quad \text{and} \quad \xi = \{\xi_1, \ldots, \xi_l\}.$$ 

The recursive construction yields the following expressions (c.f. [50, Lemma 14]):

$$Y := Y_r = R_{Q_{r-1}} Q_{r-1} - \text{tr}(Q_{r-1}) \mathbb{1} + Y_{r-1}$$

$$= \sum_{i=1}^{r} \left( R_{Q_{i-1}} Q_{i-1} - \text{tr}(Q_{i-1}) \mathbb{1} \right) \quad \text{and}$$

$$Q_i = X - P_T Y_i = P_T \left( Q_{i-1} + \text{tr}(Q_{i-1}) \mathbb{1} - R_{Q_{i-1}} Q_{i-1} \right)$$

$$= P_T \left( I + \prod_{j=1}^{i} - R_{Q_{j-1}} \right) Q_{i-1} = \cdots = \prod_{j=1}^{i} P_T \left( I + \prod_{j=1}^{i} - R_{Q_{j-1}} \right) X.$$ 

---

5 The use of pseudo-code allows for a compact presentation of this randomized procedure. However, the reader should keep in mind that the construction is purely part of a proof and should not be confused with the recovery algorithm (which is given in Eq. (24)).
We now set
\[ r = \lceil \log_2 d \rceil + 2. \quad (41) \]
Then, in case of success, the validity of properties (37) and (38) for \( c = 1/2 \) and \( b = 1/4 \) in each step (\( Q_i \to Q_{i+1} \) and \( Y_i \to Y_{i+1} \), respectively) guarantee
\[
\| Y_T - X \|_2 = \| Q_r \|_2 \leq \| X \|_2 \leq \frac{2^{-\lceil \log_2 d \rceil - 2} \| X \|_2}{4d},
\]
\[
\| Y_T^\perp \|_\infty \leq \sum_{i=1}^{r} \| \mathcal{P}_T^\perp (\mathcal{R}_{Q_i-1} Q_{i-1} - \text{tr}(Q_{i-1}) \mathbb{I}) \|_\infty \\
\leq \sum_{i=1}^{r} \frac{1}{4} \| Q_{i-1} \|_2 \leq \frac{1}{4} \sum_{i=1}^{r} 2^{1-i} \| Q_0 \|_\infty \\
\leq \| X \|_\infty \frac{1}{4} \sum_{i=0}^{\infty} 2^{-i} = \frac{1}{2},
\]
Thus, \( Y_r \) constitutes an approximate dual certificate in the sense of Definition 11.

What remains to be done is to choose the parameters \( l \) and \( \{ m_i \}_{i=1}^{l} \) such that the probability of the algorithm failing is smaller than \( 0.5e^{-\omega} \). Algorithm 1 fails precisely if
\[
\sum_{i=1}^{l} \xi_i < r. \quad (42)
\]
Recall that the \( \xi_i \)'s are Bernoulli random variables which indicate whether the \( i \)th iteration of the algorithm has been successful (\( \xi_i = 1 \)), or failed (\( \xi_i = 0 \)). Our aim is to bound the probability of the event in (42) by a similar expression involving independent\(^6\) Bernoulli variables \( \xi_i' \). To this end, write
\[
\Pr \left[ \sum_{i=1}^{l} \xi_i < r \right] = \mathbb{E} \left[ \Pr \left[ \xi_l < r - \sum_{i=1}^{l-1} \xi_i \mid \xi_{l-1}, \ldots, \xi_1 \right] \right]. \quad (43)
\]
Conditioned on an arbitrary instance of \( \xi_{l-1}, \ldots, \xi_1 \), the variable \( \xi_l \) follows a Bernoulli distribution with some parameter \( p(\xi_{l-1}, \ldots, \xi_1) \). Note that if \( \xi \sim \text{B}(p) \) is a Bernoulli variable with parameter \( p \), then for every fixed \( t \in \mathbb{R} \), the probability \( \Pr_{\xi \sim \text{B}(p)}[\xi < t] \) is non-increasing as a function of \( p \). Consequently, the estimate
\[
\Pr \left[ \sum_{i=1}^{l} \xi_i < r \right] \leq \Pr \left[ \xi_l' + \sum_{i=1}^{l-1} \xi_i < r \right] \quad (44)
\]
\(^6\) It was pointed out to us by A. Hansen that in some previous papers \([35,50]\) which involve a similar construction to the one presented here, it was tacitly assumed that the \( \xi_i \) are independent. This will of course not be true in general. Fortunately, a more careful argument shows that all conclusions remain valid \([1]\). Our treatment here is similar to the one presented in \([1]\).
is valid if $\xi'_i$ is an independent $p'$-Bernoulli distributed with
\[ p' \leq \min_{\xi_{l-1}, \ldots, \xi_1} p(\xi_{l-1}, \ldots, \xi_1). \]

Proposition 16 provides a uniform lower bound on the success probability $p(\xi_{l-1}, \ldots, \xi_1)$. Indeed, there is a universal constant $C_1$ such that invoking Proposition 16 with
\[ m' := C_1 t d^{2-\gamma} \log d \]
and $Z = Q$ gives a probability of success of at least $9/10$ for any $Q$ (in particular, independently of the $\xi_{l-1}, \ldots, \xi_1$). Thus, choosing $p' = 9/10$ and $m_i = m'$ for all $i$, we can then iterate the estimate (44) to arrive at
\[
\Pr \left[ \sum_{i=1}^{l} \xi_i < r \right] \leq \Pr \left[ \xi'_l + \sum_{i=1}^{l-1} \xi_i < r \right] \leq \cdots \leq \Pr \left[ \sum_{i=1}^{l} \xi'_i < r \right].
\] (45)
where the $\xi'_i$ are independent Bernoulli variables with parameter $9/10$. A standard Chernoff bound (e.g. [39, Sect. Concentration: Theorem 2.1]) gives
\[
\Pr \left[ \sum_{i=1}^{l} \xi'_i \leq l(9/10 - t) \right] \leq e^{-2lt^2}.
\]
Choosing $t = 9/10 - r/l$ we obtain
\[
\Pr \left[ \sum_{i=1}^{l} \xi'_i < r \right] \leq \Pr \left[ \sum_{i=1}^{l} \xi'_i \leq r \right] = \Pr \left[ \sum_{i=1}^{l} \xi'_i \leq l(9/10 - t) \right] \leq \exp \left( -2l \left( \frac{9}{10} - \frac{r}{l} \right)^2 \right). \] (46)

Setting the number of iterations generously to
\[ l = 10\omega r = 10\omega (\lceil \log_2 d \rceil + 2) \]
implies
\[ 2l \left( \frac{9}{10} - \frac{r}{l} \right)^2 \geq 20\omega r \left( \frac{8}{10} \right)^2 \geq 12\omega r \geq \omega + \log 2, \]
where we have used $\omega \geq 1 \geq \log 2$ in the last inequality. Together with (42), (45) and (46) this gives the desired bound
\[
\Pr \left[ \text{algorithm fails} \right] = \Pr \left[ \sum_{i=1}^{l} \xi_i < r \right] \leq \Pr \left[ \sum_{i=1}^{l} \xi'_i < r \right] \leq e^{-\omega - \log(2)} = \frac{1}{2} e^{-\omega},
\]
on our construction of $Y$ failing. The total number of measurement vectors sampled is

$$\sum_{i=1}^{l} m_i = lm' \leq C \omega td^{2-\gamma} \log^2 d,$$

for some constant $C$.

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

**Proof of the Main Theorem** In Sect. 5 (Proposition 12) we have shown that the algorithm (24) recovers the sought-for signal $x$, provided that (25) 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 $0.5e^{-\omega}$, provided that the total sampling rate $m$ obeys

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

for a sufficiently large absolute constant $C$. Provided that this constant is large enough, Proposition 9 implies that the probability of (25) failing is also bounded by $0.5e^{-\omega}$. Theorem 1 now follows from the union bound over these two probabilities of failure. $\square$

7 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 sub-quadratic 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) [45,46,67,75]. 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) [44]. 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. [26] for a detailed survey—maximal sets of MUBs are known to form $2$-designs [45,75].

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 \( a_1, \ldots, a_m \) 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

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

will occur with probability at least \( e^{-\omega} \).

Consequently a scaling of \( O(d^2) \) in general cannot be avoided when demanding only the property of being a 2-design and simultaneously 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 Chapt. 6.2 in [29]):

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 \( 1 - e^{-\omega} \) or larger?

To answer this question, we have to find the smallest integer \( m \) such that

\[
1 - (1 - p)^m \geq 1 - e^{-\omega}, \quad \text{or equivalently} \quad -m \log(1 - p) \geq \omega. \tag{49}
\]

The standard inequality

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

for any \( p \in [0, 1/2] \) implies that (48) is a necessary criterion for (49) and we are done. \( \square \)

8 Conclusion

In this paper we have derived a partly derandomized version of Gaussian PhaseLift \([18, 23]\). 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 \( t \):

\[
m = O\left(t d^{1+2/t} \log^2 d \right). \tag{50}
\]
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.

Acknowledgments  DG and RK are grateful to the organizers and participants of the Workshop on Phaseless Reconstruction, held as part of the 2013 February Fourier Talks at the University of Maryland, where they were introduced to the details of the problem. This extends, in particular, to Thomas Strohmer. The work of DG and RK is supported by the Excellence Initiative of the German Federal and State Governments (Grant ZUK 43), by scholarship funds from the State Graduate Funding Program of Baden-Württemberg, and by the US Army Research Office under contracts W911NF-14-1-0098 and W911NF-14-1-0133 (Quantum Characterization, Verification, and Validation), and the DFG. FK acknowledges support from the German Federal Ministry of Education and Research (BMBF) through the cooperative research project ZeMat.

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.

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

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

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

$$\text{tr}_2(A) = \sum_{i=1}^d \left( 1 \otimes b_i^* \right) A \left( 1 \otimes b_i \right).$$

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

$$\text{tr}_2 \left( P_{\text{Sym}^2} A \otimes B \right) = \frac{1}{2} \text{tr}_2 \left( A \otimes B \right) + \frac{1}{2} \text{tr}_2 \left( \sigma_{(1,2)} A \otimes B \right).$$
The latter term can be evaluated explicitly:

\[
\text{tr}_2 \left( \sigma_{(1,2)} A \otimes B \right) = \sum_{k=1}^{d} (\mathbb{1} \otimes b_k^*) \sum_{i,j=1}^{d} b_i b_j^* \otimes b_j b_i^* A \otimes B (\mathbb{1} \otimes b_k) \\
= \sum_{i,j,k=1}^{d} b_i b_j^* A b_k^* b_j b_i^* B b_k = \sum_{i,j=1}^{d} \langle b_i, B b_j \rangle b_i b_j^* A \\
= \left( \sum_{i=1}^{d} b_i b_i^* \right) B \left( \sum_{j=1}^{d} b_j b_j^* \right) A = \mathbb{1} B \mathbb{1} A = BA,
\]

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

References

1. Adcock, B., Hansen, A.C.: Generalized Sampling and Infinite-Dimensional Compressed Sensing, Technical Report NA2011/02, DAMTP. University of Cambridge, Cambridge (2011)
2. Ahlswede, R., Winter, A.: Strong converse for identification via quantum channels. IEEE Trans. Inform. Theory 48, 569–579 (2002)
3. Alexeev, B., Bandeira, A.S., Fickus, M., Mixon, D.G.: Phase retrieval with polarization. SIAM J. Imaging Sci. 7, 35–66 (2014)
4. Ambainis, A., Emerson, J.: Quantum t-designs: t-wise independence in the quantum world. In: Proceedings of the 22nd Annual IEEE Conference on Computational Complexity, pp. 129–140 (2007)
5. Bachoc, C., Venkov, B.: Modular forms, lattices and spherical designs. In: Euclidean Lattices, Spherical Designs and Modular Forms, pp. 87–111. On the works of Boris Venkov. L’Enseignement Mathématique, Genève (2001)
6. Bachoc, C., Ehler, M.: Tight p-fusion frames. Appl. Comput. Harmon. Anal. 35, 1–15 (2013)
7. Bajnok, B.: Construction of spherical \(t\)-designs. Geom. Dedic. 43, 167–179 (1992)
8. Balan, R., Casazza, P., Edidin, D.: On signal reconstruction without phase. Appl. Comput. Harmon. Anal. 20, 345–356 (2006)
9. Balan, R., Bodmann, B.G., Casazza, P.G., Edidin, D.: Painless reconstruction from magnitudes of frame coefficients. J. Fourier Anal. Appl. 15, 488–501 (2009)
10. Bandeira, A.S., Chen, Y., Mixon, D.G.: Phase retrieval from power spectra of masked signals. Inf. Inference (2014). doi:10.1093/imaiai/iau002
11. Baraniuk, R.G., Davenport, M., DeVore, R.A., Wakin, M.: A simple proof of the restricted isometry property for random matrices. Constr. approx. 28, 253–263 (2008)
12. Barvinok, A.: A Course in Convexity. American Mathematical Society (AMS), Providence (2002)
13. Bauschke, H.H., Combettes, P.L., Luke, D.R.: Hybrid projection–reflection method for phase retrieval. JOSA A 20, 1025–1034 (2003)
14. Bhatia, R.: Matrix Analysis. Springer, New York (1996)
15. Brandao, F.G., Harrow, A.W., Horodecki, M.: Local random quantum circuits are approximate polynomial-designs. arXiv:1208.0692 (2012)
16. Bruck, Y.M., Sodin, L.: On the ambiguity of the image reconstruction problem. Opt. Commun. 30, 304–308 (1979)
17. Candès, E., Li, X., Soltanolkotabi, M.: Phase retrieval from masked fourier transforms. Appl. Comput. Harmon. Anal. arXiv:1310.3240 (to appear)
18. Candès, E., Li, X.: Solving quadratic equations via PhaseLift when there are about as many equations as unknowns. Found. Comput. Math., 1–10 (2013)
19. Candès, E., Tao, T.: Decoding by linear programming. IEEE Trans. Inform. Theory 51, 4203–4215 (2005)
20. Candès, E.J., Romberg, J., Tao, T.: Stable signal recovery from incomplete and inaccurate measurements. Commun. Pure Appl. Math. 59, 1207–1223 (2006)
21. Candès, E.J., Tao, T.: The power of convex relaxation: near-optimal matrix completion. IEEE Trans. Inform. Theory 56, 2053–2080 (2010)
22. Candès, E.J., Eldar, Y.C., Strohmer, T., Voroninski, V.: Phase retrieval via matrix completion. SIAM J. Imaging Sci. 6, 199–225 (2013)
23. Candès, E.J., Strohmer, T., Voroninski, V.: Phaselift: exact and stable signal recovery from magnitude measurements via convex programming. Commun. Pure Appl. Math. 66, 1241–1274 (2013)
24. Cvitanović, P.: Group Theory. Birdtracks, Lie’s, and Exceptional Groups. Princeton University Press, Princeton (2008)
25. Delsarte, P., Goethals, J., Seidel, J.: Spherical codes and designs. Geom. Dedic. 6, 363–388 (1977)
26. Durt, T., Englert, B.-G., Bengtsson, I., Życzkowski, K.: On mutually unbiased bases. Int. J. Quantum Inf. 8, 535–640 (2010)
27. Ehler, M., Fornasier, M., Sigl, J.: Quasi-linear compressed sensing. Multiscale Model. Simul. 12, 725–754 (2014)
28. Eldar, Y.C., Mendelson, S.: Phase retrieval: stability and recovery guarantees. Appl. Comput. Harmon. Anal. 36, 473–494 (2014)
29. Feller, W.: An Introduction to Probability Theory and its Applications. I. Wiley, New York (1968)
30. Fienup, J.R.: Phase retrieval algorithms: a comparison. Appl. Opt. 21, 2758–2769 (1982)
31. Gottesman, D.: Stabilizer codes and quantum error correction. Ph.D. dissertation, California Institute of Technology (1997)
32. Gottesman, D.: The Heisenberg representation of quantum computers. In: Proceedings of the XXII International Colloquium on Group Theoretical Methods in Physics, pp. 32–43. International Press, Cambridge (1999)
33. Grant, M., Boyd, S.: Graph implementations for nonsmooth convex programs. In: Blondel, V., Boyd, S., Kimura, H. (eds.) Recent Advances in Learning and Control, Ser. Lecture Notes in Control and Information Sciences, Springer-Verlag Limited, London, pp. 95–110. http://stanford.edu/~boyd/graph_dcp.html (2008). Accessed 13 Oct 2014
34. Grant, M., Boyd, S.: CVX: Matlab software for disciplined convex programming, version 2.1. http://cvxr.com/cvx (2014). Accessed 13 Oct 2014
35. Gross, D.: Recovering low-rank matrices from few coefficients in any basis. IEEE Trans. Inform. Theory 57, 1548–1566 (2011)
36. Gross, D., Audenaert, K., Eisert, J.: Evenly distributed unitaries: on the structure of unitary designs. J. Math. Phys. 48, pp. 052 104, 22 (2007)
37. Gross, D., Krahmer, F., Kueng, R.: Improved recovery guarantees for phase retrieval from coded diffraction patterns. arXiv:1402.6286 (2014)
38. Gross, D., Liu, Y.-K., Flammini, S.T., Becker, S., Eisert, J.: State tomography via compressed sensing. Phys. Rev. Lett. 105, 150401 (2010)
39. Habib, M., McDiarmid, C., Ramírez Alfonsín, J., Reed, B. (eds.): Probabilistic Methods for Algorithmic Discrete Mathematics. Springer, Berlin (1998)
40. Hayashi, A., Hashimoto, T., Horibe, M.: Reexamination of optimal quantum state estimation of pure states. Phys. Rev. A 72, 032301 (2005)
41. Heinosaari, T., Mazzarella, L., Wolf, M.M.: Quantum tomography under prior information. Commun. Math. Phys. 318, 355–374 (2013)
42. Hoggar, S.: t-Designs in projective spaces. Eur. J. Comb. 3, 233–254 (1982)
43. Hoory, S., Linial, N., Wigderson, A.: Expander graphs and their applications. Bull. Am. Math. Soc. 43, 439–561 (2006)
44. Klappenecker, A., Rötteler, M.: Constructions of mutually unbiased bases. In: Finite Fields and Applications. 7th International Conference, \( \mathbb{F}_q^7 \), pp. 137–144. Springer, Berlin (2004)
45. Klappenecker, A., Rötteler, M.: Mutually unbiased bases are complex projective 2-designs. In: 2005 IEEE International Symposium on Information Theory (ISIT), vols. 1, 2, pp. 1740–1744 (2005)
46. König, H.: Cubature formulas on spheres. In: Advances in Multivariate Approximation. Proceedings of the 3rd International Conference on Multivariate Approximation Theory, pp. 201–211. Wiley-VCH, Berlin (1999)
47. Korevaar, J., Meyers, J.: Chebyshev-type quadrature on multidimensional domains. J. Approx. Theory 79, 144–164 (1994)
48. Krahmer, F., Rauhut, H.: Structured random measurements in signal processing. arXiv:1401.1106 (2014)
49. Kueng, R., Gross, D.: Stabilizer states are complex projective 3-designs in qubit dimensions. (in preparation)
50. Kueng, R., Gross, D.: RIPless compressed sensing from anisotropic measurements. Lin. Alg. Appl. 441, 110–123 (2014)
51. Landsberg, J.M.: Tensors: Geometry and Applications. American Mathematical Society (AMS), Providence (2012)
52. Levenshtein, V.I.: Universal bounds for codes and designs. In: Handbook of Coding Theory. Part 1: Algebraic Coding, vol. 1, pp. 499–648. Elsevier, Amsterdam (1998)
53. Li, X., Voroninski, V.: Sparse signal recovery from quadratic measurements via convex programming. SIAM J. Math. Anal. 45, 3019–3033 (2013)
54. Liu, Y.-K.: Universal low-rank matrix recovery from pauli measurements. Adv. Neural Inf. Process. Syst. 24, 1638–1646 (2011)
55. Low, R.A.: Large deviation bounds for $\ell$-designs. Proc. R. Soc. Lond. Ser. A 465, 3289–3308 (2009)
56. Luby, M., Wigderson, A.: Pairwise independence and derandomization. Print version of Foundations and Trends in Theoretical Computer Science, Vol. 1, No. 4 (2005), print version of foundations and trends in theoretical computer science vol. 1, no. 4 (2005) ed. Boston, MA (2006)
57. Millane, R.: Phase retrieval in crystallography and optics. JOSA A 7, 394–411 (1990)
58. Mixon, D.: Short, Fat Matrices, Blog, 2013. http://dustingmixon.wordpress.com/. Accessed 13 Oct 2014
59. Mondragon, D., Voroninski, V.: Determination of all pure quantum states from a minimal number of observables. arXiv:1306.1214 (2013)
60. Nebe, G., Rains, E., Sloane, N.: The invariants of the Clifford groups. Des. Codes Cryptogr. 24, 99–121 (2001)
61. Netrapalli, P., Jain, P., Sanghavi, S.: Phase retrieval using alternating minimization. In: Advances in Neural Information Processing Systems, pp. 2796–2804 (2013)
62. Neumaier, A.: Combinatorial Configurations in Terms of Distances. Department of Mathematics Memorandum, Eindhoven (1981)
63. Recht, B., Fazel, M., Parrilo, P.A.: Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. SIAM Rev. 52, 471–501 (2010)
64. Renes, J.M., Blume-Kohout, R., Scott, A., Caves, C.M.: Symmetric informationally complete quantum measurements. J. Math. Phys. 45, 2171–2180 (2004)
65. Raudelson, M., Vershynin, R.: On sparse reconstruction from Fourier and Gaussian measurements. Commun. Pure Appl. Math. 61, 1025–1045 (2008)
66. Sanderson, B.: Immersions and embeddings of projective spaces. Proc. Lond. Math. Soc. 3(14), 137–153 (1964)
67. Schwinger, J.: Unitary operator bases. Proc. Natl. Acad. Sci. USA 46, 570–579 (1960)
68. Scott, A.: Tight informationally complete quantum measurements. J. Phys. A 39, 13507–13530 (2006)
69. Seymour, P., Zaslavsky, T.: Averaging sets: a generalization of mean values and spherical designs. Adv. Math. 52, 213–240 (1984)
70. Sidelnikov, V.: Spherical 7-designs in 2$^n$-dimensional Euclidean space. J. Algebr. Comb. 10, 279–288 (1999)
71. Tropp, J. A.: User-friendly tools for random matrices: an introduction, notes. http://users.cms.caltech.edu/~jtropp/notes/Tro12-User-Friendly-Tools-NIPS.pdf (2012). Accessed 13 Oct 2014
72. Tropp, J.A.: User-friendly tail bounds for sums of random matrices. Found. Comput. Math. 12, 389–434 (2012)
73. Turaev, V.: Quantum Invariants of Knots and 3-Manifolds. Walter de Gruyter, Berlin (1994)
74. Watrous, J.: Theory of quantum information, lecture notes. https://cs.uwaterloo.ca/~watrous/LectureNotes.html (2011). Accessed 13 Oct 2014
75. Zauner, G.: Quantendesigns: Grundzüge einer nichtkommutativen Designtheorie. Ph.D. dissertation, University of Vienna (1999)