Research Article **Open Access**

Martin Ehler*, Manuel Gräf, and Franz J. Király

Phase retrieval using random cubatures and fusion frames of positive semidefinite matrices

DOI 10.1515/wwfaa-2015-0005
Received May 29, 2015; accepted October 6, 2015

Abstract: As a generalization of the standard phase retrieval problem, we seek to reconstruct symmetric rank-1 matrices from inner products with subclasses of positive semidefinite matrices. For such subclasses, we introduce random cubatures for spaces of multivariate polynomials based on moment conditions. The inner products with samples from sufficiently strong random cubatures allow the reconstruction of symmetric rank-1 matrices with a decent probability by solving the feasibility problem of a semidefinite program.

1 Introduction

Many signal processing problems in engineering such as X-ray crystallography and coherent diffraction imaging require the reconstruction of symmetric rank-1 matricies from inner products with rank-1 projectors, often called phase retrieval, cf. [4, 5, 12, 33, 34] and references therein. Signal recovery from inner products with higher rank positive semidefinite matrices is a suitable model when diffraction patterns are weighted averages of k wavefields, which occurs with incoherent diffraction [21].

Classical reconstruction algorithms for the rank-1 phase retrieval problem are based on iterated projection schemes [22, 24] but there is a lack of stringent mathematical recovery guarantees. Signal reconstruction with high probability is guaranteed in [12, 14] by solving the feasibility problem of a semidefinite program when sufficiently many rank-1 projectors are chosen in a uniformly distributed fashion, see also [38]. Similarly, higher rank phase retrieval was solved in [2] by using uniformly distributed rank-k orthogonal projectors.

In order to match the measurement process in optical physics more appropriately, the requirement of uniform distribution must be relaxed. For $k = 1$, such an important relaxation was recently obtained in [26], where random sampling rank-1 projectors from so-called spherical designs of strength $t \geq 3$ has been addressed. Increasing t yields higher recovery probability and allows for fewer measurements. Asymptotic existence results of strong spherical designs were obtained in [9, 10]. Deriving actual constructions, however, is a delicate issue. Recently, such issues were overcome for $k = 1$ in [30] involving so-called cubatures (weighted designs) with strength $t \geq 4$, whose existence is well-understood and many linear algebra based constructions are known [16].

Here, we shall generalize [26] to the range of positive semidefinite measurement matrices, and we do not require designs but only so-called cubatures. In contrast to [30], we only need strength $t \geq 3$. We address the real setting and point out some specialities that are due to the higher rank, partially based on earlier observations in [1, 2]. To summarize, we generalize the results in [26] from rank-1 projectors to positive semidefinite

***Corresponding Author: Martin Ehler:** University of Vienna, Department of Mathematics, Oskar-Morgenstern-Platz 1 A-1090 Vienna, E-mail: martin.ehler@univie.ac.at
Manuel Gräf: University of Vienna, Department of Mathematics, Oskar-Morgenstern-Platz 1 A-1090 Vienna, E-mail: manuel.graef@univie.ac.at
Franz J. Király: Department of Statistical Science, University College London, E-mail: f.kiraly@ucl.ac.uk

matrices and at the same time provide significant improvements for the rank-1 case because we only require cubatures of strength $t \geq 3$ and not designs. For cubatures of strength $t \geq 4$, the rank-1 case is already covered in [30].

The overall structure of our proofs related to the reconstruction of symmetric rank-1 matricies from inner products with positive semidefinite matricies is guided by the approach in [26]. Our generalizations are based on the computation of trace moments of matrix distributions induced by the Haar measure on the orthogonal group. The use of zonal polynomials as discussed in [27, 28] enables us to compute all trace moments, and we explicitly provide the first 3 of them. The remaining parts of the signal reconstruction proofs essentially follow the approach in [26] with adjusted parameters and constants combined with our results on the first three trace moments, cubatures [16], and random tight t-fusion frames [1, 18].

Outline

We introduce the general phase retrieval problem in Section 2, where we also state the result from [2] about uniformly distributed rank-k-projectors. The findings in [26] for sampling spherical designs (hence the setting of rank-1 projectors) are stated in Section 3. Deterministic conditions for signal recovery with positive semidefinite measurement matrices through solving the feasibility problem of a semidefinite program are verified in Section 4 and are based on near isometry properties and a so-called approximate dual certificate.

Our main result on phase retrieval is stated in Section 5 and is based on random cubatures. The remaining part of the present paper is dedicated for providing its proof. We outline the structure of the proof in Section 6 that is based on trace moments. The trace moments are computed in Section 7, special moments in Section 7.1 and the general case is treated in Section 7.2. We compute the first 3 trace moments explicitly in Section 7.3, which are an important ingredient of the proof of our main result on phase retrieval. Most of the technical details of the complete proof are contained in the appendix. Conclusions are given in Section 8.

2 Phase retrieval and uniform sampling

Let $\mathscr{H}_d$ denote the space of symmetric matrices in $\mathbb{R}^{d\times d}$ endowed with the Hilbert-Schmidt inner product $\langle X, Y\rangle := \operatorname{trace}(XY^*)$, for $X, Y \in \mathscr{H}_d$. In our phase retrieval problem, we seek to recover some unknown signal $x \in \mathbb{R}^d$ from the knowledge of n matrices $\{P_j\}_{j=1}^n \subset \mathscr{H}_d$ and the associated measurements

$$\{\langle xx^*, P_j\rangle\}_{j=1}^n.$$

Clearly, x can at best be recovered up to a global phase factor and so we simply aim to recover the rank-1 matrix xx^*. Uniqueness of xx^* was discussed in [3–5, 8] for rank-1 orthogonal projectors $\{P_j\}_{j=1}^n$ and in [2, 11] for more general choices of $\{P_j\}_{j=1}^n$.

Besides injectivity, we also need an efficient algorithm to eventually reconstruct the signal. We consider the set of measurement matrices

$$\mathcal{G}_{\lambda,d} := \{OD_\lambda O^* : O \in \mathcal{O}_d\},$$

where $D_\lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_d)$ and $\lambda = (\lambda_1, \dots, \lambda_d)^*$ is a fixed vector with

$$1 \geq \lambda_1 \geq \ldots \geq \lambda_k > \lambda_{k+1} = \ldots, \lambda_d = 0.$$

Throughout the entire manuscript we make this assumption and without loss of generality we additionally assume $\lambda_1 = 1$, which can easily be achieved by rescaling.

To derive asymptotic recovery results, we shall later increase the dimension d while we keep k and $\lambda_1, \dots, \lambda_k$ fixed. Note that $\mathcal{G}_{\lambda,d}$ is simply the set of all rank-k positive semidefinite matrices with nonzero eigenvalues $\lambda_1, \dots, \lambda_k$.

The Haar measure dO on the orthogonal group $\mathcal{O}(d)$ acts transitively on $\mathcal{G}_{\lambda,d}$ by definition and induces a probability measure $\sigma_{\lambda,d}$ on $\mathcal{G}_{\lambda,d}$ that is invariant under the orthogonal group. When $\lambda_1 = \cdots = \lambda_k = 1$, then

$\mathcal{G}_{\lambda,d}$ can be identified with the set of all k-dimensional linear subspaces in $\mathbb{R}^d$, known as the (real) Grassmann space.

Choose $\{P_j\}_{j=1}^n \subset \mathcal{G}_{\lambda,d}$, and let us consider

$$\text{find } X \in \mathscr{H}_d, \quad \text{subject to} \quad \langle X, P_j\rangle = \langle xx^\star, P_j\rangle,\ j=1,\dots,n, \quad X \succeq 0, \tag{1}$$

where $X \succeq 0$ means that X is positive semidefinite. This is the feasibility problem of a semidefinite program and efficient algorithms based on interior point methods are available. For $\lambda = (1,0,\dots,0)^\star$, there is a constant $c > 0$, such that the choice of $n \geq cd$ uniformly distributed subspaces yields that, with high probability, $xx^\star$ is the only feasible point of (1), cf. [13, 14, 17]. This result was generalized to rank-k orthogonal projectors in [2]:

Theorem 2.1 ([2]). *Let $\lambda = (1,\dots,1,0,\dots,0)^\star$, where 1 is repeated k times. Then there are constants $c_1, c_2 > 0$ such that, if*

$$n \geq c_1 d$$

and $\{P_j\}_{j=1}^n \subset \mathcal{G}_{\lambda,d}$ are chosen independently identically distributed according to the normalized Haar measure $\sigma_{\lambda,d}$ on $\mathcal{G}_{\lambda,d}$, then, for all $x \in \mathbb{R}^d$, the matrix $xx^\star$ is the unique solution to (1) *with probability at least $1 - e^{-c_2 n}$.*

For rank-1 projectors, i.e., $\lambda = e_1 = (1,0,\dots,0)^\star$, the sampling from the uniform measure $\sigma_{\lambda,d}$ has been relaxed in [26] , which is the topic of the subsequent section.

3 Signal reconstruction for rank-1 projectors

This section deals with $\lambda = e_1$ only and before we cite some reconstruction results, we need to introduce further concepts and notation. For $x \in \mathbb{S}^{d-1} := \{z \in \mathbb{R}^d : \|z\| = 1\}$, we denote $P_x := xx^\star$. A collection $\{P_{x_j}\}_{j=1}^n \subset \mathcal{G}_{e_1,d}$ is called a *projective t-design* if

$$\frac{1}{n}\sum_{j=1}^n \langle P_{x_j}, P_x\rangle^t = \int_{\mathcal{G}_{e_1,d}} \langle P, P_x\rangle^t d\sigma_{e_1,d}(P), \quad \text{for all } x \in \mathbb{S}^{d-1}. \tag{2}$$

The latter is equivalent to

$$\frac{1}{n}\sum_{j=1}^n |\langle x_j, x\rangle|^{2t} = \int_{\mathbb{S}^{d-1}} |\langle y, x\rangle|^{2t} dy, \quad \text{for all } x \in \mathbb{S}^{d-1},$$

where dy denotes, as usual, the canonical measure on the sphere that we additionally assume to be normalized.

The reconstruction results in [26] were only derived for complex signals and measurements, but can also be checked in the real case by analogous arguments. For consistency with the presentation here, we shall therefore recall this result in the real setting. So, for $x \in \mathbb{R}^d$ and $\{P_j\}_{j=1}^n \subset \mathcal{G}_{e_1,d}$, we consider the optimization problem

$$\arg\min_{X \in \mathscr{H}_d,\ X \succeq 0} \|X\|_\star, \quad \text{s.t. } \operatorname{trace}(X) = \|x\|^2,\ \big(\langle P_j, X\rangle\big)_{j=1}^n = \big(\langle P_j, xx^\star\rangle\big)_{j=1}^n, \tag{3}$$

where $\|X\|_\star$ denotes the nuclear norm of X, i.e., the sum of the absolute values of its singular values, then [26] yields:

Theorem 3.1 ([26]). *Let $x \in \mathbb{R}^d$ be an unknown signal and suppose that $\|x\|^2$ is known. If $\{P_j\}_{j=1}^n \subset \mathcal{G}_{e_1,d}$ is independently sampled in a uniform fashion from some projective design of strength $t \geq 3$, then with probability at least $1 - e^{-\omega}$, the rank-one matrix $xx^\star$ is the unique solution to* (3) *provided that*

$$n \geq c_1 \omega t d^{1+2/t} \log^2(d),$$

where $\omega \geq 1$ is an arbitrary parameter and c_1 is a universal constant.

Note that the above Theorem 3.1 is restricted to uniform sampling of a projective t-design. The latter is a rather inconvenient restriction as shown in the following example:

Example 3.2. The classical phase retrieval problem stemming from optical physics involves Fourier measurements, meaning that the rank-1 projectors $P_j = x_j x_j^*$ are generated by Fourier vectors

$$x_j = \frac{1}{\sqrt{d}}(-e^{2\pi i l_1 j/m}, \dots, e^{-2\pi i l_d j/m})^\star \in \mathbb{C}^d, \tag{4}$$

where $\{l_i\}_{i=1}^d \subset \mathbb{Z}$. Often, magnitude measurements in time are also available, expressed as additional measurements $\{P_{e_k}\}_{k=1}^n$, where $\{e_k\}_{k=1}^d$ is the canonical orthogonal basis of $\mathbb{C}^d$. It turns out that the combination of special Fourier vectors with time measurements yield a formula similar to (2), hence, yields almost a projective design. In fact, these ideas are inspired by [29, Proposition 4] and [35, Section 2.1.2], see also [20]: let q be a prime and let $d = q^r + 1$ for some $r \in \mathbb{N}$. For $m = d^2 - d + 1$, there exist integers $0 \le l_1 < \cdots < l_d < m$ such that all numbers $1, \dots, m-1$ occur as residues mod m of the $d(d-1)$ differences $(l_k - l_\ell)$, for $k \ne \ell$, cf. [29]. Then the following formula holds, for all $x \in \mathbb{C}^d$ with $\|x\| = 1$,

$$\frac{d}{d^3+1}\sum_{j=1}^{d^2-d+1}\langle P_{x_j}, P_x\rangle^2 + \frac{1}{d(d+1)}\sum_{k=1}^{d}\langle P_{e_k}, P_x\rangle^2 = \int_{\mathcal{G}_{e_1,d}^{\mathbb{C}}} \langle P, P_x\rangle^2 d\sigma_{e_1,d}^{\mathbb{C}}(P),$$

where $\mathcal{G}_{e_1,d}^{\mathbb{C}}$ denotes the complex projective space and $\sigma_{e_1,d}^{\mathbb{C}}$ its normalized canonical measure induced by the Haar measure on the unitary group. Thus, the combined Fourier and time measurements provide some sort of weighted projective design.

Remark 3.3. Although our presentation is focused on the real case, we want to point out that all results can be derived in the complex setting as well, so that Example 3.2 can still guide us. It shows that the structural requirement of a design in Theorem 3.1 is still too restrictive. We shall generalize Theorem 3.1 in several aspects. First, it is restricted to $\lambda = e_1$, and we shall address the general case λ. Moreover, we can handle weighted designs, which is a significant structural generalization, so that our results also yield significant improvements for $\lambda = e_1$.

4 Deterministic conditions for signal reconstruction with general λ

This section is dedicated to consider phase retrieval when λ is arbitrary. To model the knowledge of $\|x\|^2$, we make the convention that $P_0 = I_d$ and, hence, $\langle xx^*, P_0\rangle = \operatorname{trace}(xx^*) = \|x\|^2$ holds, and we consider the problem

$$\text{find } X \in \mathcal{H}_d, \quad \text{such that} \quad \big(\langle X, P_j\rangle\big)_{j=0}^n = \big(\|P_j x\|^2\big)_{j=0}^n, \quad X \succeq 0, \tag{5}$$

where $\{P_j\}_{j=1}^n \subset \mathcal{G}_{\lambda,d}$. Note that (5) is the feasibility problem of a semidefinite programm. In comparison to (3), the actual minimization is void. In fact, $X \succeq 0$ yields $\|X\|_\star = \operatorname{trace}(X)$, so that the minimization in (3) was superfluous too.

To establish deterministic conditions that ensure solvability of (5), we use the notion of dual certificates that require some preparation. For a fixed $x \in \mathbb{R}^d$, we consider the subspace

$$T_x := \{xz^* + zx^* : z \in \mathbb{R}^d\} \subset \mathcal{H}_d,$$

which is the tangent space of the rank-one symmetric matrices at the point xx^*. For some $Y \in \mathcal{H}_d$, let Y_{T_x} denote the orthogonal projection of Y onto T_x and $Y_{T_x^\perp}$ the orthogonal projection onto the orthogonal complement of T_x. Moreover, let $\|\cdot\|_F$ denote the Frobenius norm and $\|\cdot\|_{Op}$ the spectral norm.

Definition 4.1. For $\{P_j\}_{j=1}^n \subset \mathcal{G}_{\lambda,d}$, we call $Y \in \mathcal{H}_d$ a (γ, δ)-*dual certificate* with respect to $x \in \mathbb{R}^d$ if $Y \in \operatorname{span}\{I_d, P_1, \dots, P_n\}$ and

$$\|Y_{T_x} - xx^*\|_F \le \gamma \quad \text{and} \quad \|Y_{T_x^\perp}\|_{Op} \le \delta. \tag{6}$$

For notational convenience, we introduce the mapping

$$\mathcal{A}_n : \mathscr{H}_d \to \mathbb{R}^n, \quad X \mapsto \big(\langle X, P_j\rangle\big)_{j=1}^n, \tag{7}$$

for $\{P_j\}_{j=1}^n \subset \mathcal{G}_{\lambda,d}$.

Now, we can formulate deterministic recovery guarantees:

Theorem 4.2. *Suppose that there are $\alpha, \beta > 0$ and $\{P_j\}_{j=1}^n \subset \mathcal{G}_{\lambda,d}$ satisfying*

$$\alpha\|X\|_F^2 \le \frac{1}{n}\|\mathcal{A}_n(X)\|^2 \le \beta\|X\|_F^2, \tag{8}$$

where $\mathcal{A}_n$ is given by (7) *and the lower inequality holds for all matrices $0 \neq X \in T_x$, and the upper one for all $X \in \mathscr{H}_d$. If a (γ, δ)-dual certificate Y with respect to x exists and*

$$\sqrt{\frac{\beta}{\alpha}} < \frac{1-\delta}{\gamma},$$

then xx^* is the unique solution to (5).

Proof. We know that xx^* solves (5). Suppose that X is another solution and put $\Delta := X - xx^*$. As in [26], we apply the pinching inequality, cf. [7, 26], to obtain

$$\operatorname{trace}(X) = \operatorname{trace}(xx^* + \Delta) \ge \operatorname{trace}(xx^*) + \operatorname{trace}(\Delta_{T_x}) + \|\Delta_{T_x^\perp}\|_\star.$$

Since $\operatorname{trace}(X) = \operatorname{trace}(xx^*) = \|x\|^2$, we obtain

$$0 \ge \operatorname{trace}(\Delta_{T_x}) + \|\Delta_{T_x^\perp}\|_\star. \tag{9}$$

If $\Delta_{T_x} = 0$, then we derive $\Delta_{T_x^\perp} = 0$, so that $\Delta = 0$ and hence $X = xx^*$. If $\Delta_{T_x} \neq 0$, then (8) implies

$$\|\Delta_{T_x}\|_F \le \sqrt{\frac{1}{\alpha n}}\|\mathcal{A}_n(\Delta_{T_x})\| = \sqrt{\frac{1}{\alpha n}}\|\mathcal{A}_n(\Delta_{T_x^\perp})\| \le \sqrt{\frac{\beta}{\alpha}}\|\Delta_{T_x^\perp}\|_F. \tag{10}$$

Next, we observe that $\langle xx^*, \Delta_{T_x}\rangle = \operatorname{trace}(\Delta_{T_x})$ and obtain

$$\begin{aligned}
0 = \langle Y, \Delta\rangle &= \langle Y_{T_x} - xx^*, \Delta_{T_x}\rangle + \langle xx^*, \Delta_{T_x}\rangle + \langle Y_{T_x^\perp}, \Delta_{T_x^\perp}\rangle \\
&\le \|Y_{T_x} - xx^*\|_F\|\Delta_{T_x}\|_F + \operatorname{trace}(\Delta_{T_x}) + \|Y_{T_x^\perp}\|_{op}\|\Delta_{T_x^\perp}\|_\star \\
&\le \operatorname{trace}(\Delta_{T_x}) + \|Y_{T_x} - xx^*\|_F\sqrt{\frac{\beta}{\alpha}}\|\Delta_{T_x^\perp}\|_F + \delta\|\Delta_{T_x^\perp}\|_\star \\
&\le \operatorname{trace}(\Delta_{T_x}) + \gamma\sqrt{\frac{\beta}{\alpha}}\|\Delta_{T_x^\perp}\|_F + \delta\|\Delta_{T_x^\perp}\|_\star \\
&\le \operatorname{trace}(\Delta_{T_x}) + \Big(\gamma\sqrt{\frac{\beta}{\alpha}} + \delta\Big)\|\Delta_{T_x^\perp}\|_\star \\
&\le \operatorname{trace}(\Delta_{T_x}) + \|\Delta_{T_x^\perp}\|_\star.
\end{aligned}$$

Since $\Delta_{T_x} \neq 0$, the inequalities (10) yield $\Delta_{T_x^\perp} \neq 0$, so that the inequality of the last line is strict, which is a contradiction to (9). Therefore, we have $\Delta = 0$ and hence $X = xx^*$, so that xx^* is the unique solution to (5). □

5 Cubatures for phase retrieval with general λ

We aim to verify that certain random samples in $\mathcal{G}_{\lambda,d}$ satisfy the conditions of Theorem 4.2 with a decent probability, so that signal recovery is guaranteed. To characterize the type of random distributions involved, we need to define some sort of weighted design on $\mathcal{G}_{\lambda,d}$, for which we shall first introduce trace moments:

Definition 5.1. The *t-th trace moments* (or trace moments of degree t) of some random matrix $\mathcal{P} \in \mathcal{G}_{\lambda,d}$ are

$$\mu_{\mathcal{P}}^t(X) := \mathbb{E}\big(\langle \mathcal{P}, X\rangle^t\big), \quad X \in \mathscr{H}_d.$$

The trace moments of $\mathcal{P}$ distributed according to $\sigma_{\lambda,d}$ are denoted by

$$\mu_{\lambda,d}^t(X) := \int_{\mathcal{G}_{\lambda,d}} \langle P, X\rangle^t d\sigma_{\lambda,d}(P).$$

Similarly, for $\beta \in \mathbb{N}^s$, we define cross-moments by

$$\mu_{\mathcal{P}}^\beta(X_1, \dots, X_s) = \mathbb{E}\big(\langle X_1, \mathcal{P}\rangle^{\beta_1} \cdots \langle X_s, \mathcal{P}\rangle^{\beta_s}\big), \quad X_1, \dots, X_s \in \mathscr{H}_d$$

and make use of the expression $\mu_{\lambda,d}^\beta(X_1, \dots, X_s)$, respectively. If β consists of ones only, then we simply write $\mu_{\mathcal{P}}(X_1, \dots, X_s)$ and $\mu_{\lambda,d}(X_1, \dots, X_s)$.

Next, we can introduce cubatures for $\mathcal{G}_{\lambda,d}$:

Definition 5.2. A $\mathcal{G}_{\lambda,d}$-valued random variable $\mathcal{P}$ is called a *random cubature* of strength t (in $\mathcal{G}_{\lambda,d}$) if its t-th trace moments coincide with those of $\sigma_{\lambda,d}$, i.e.,

$$\mu_{\mathcal{P}}^t(X) = \mu_{\lambda,d}^t(X), \quad \text{for all } X \in \mathscr{H}_d. \tag{11}$$

If $\mathcal{P}$ satisfies (11) at least for all $X = xx^\star$, $x \in \mathbb{R}^d$, then it is called a *random tight t-fusion frame*.

Remark 5.3. In the literature, the term tight t-fusion frame usually refers to the case when the entries in λ are ones and zeros, so that the measurement matrices are orthogonal projectors. Here, we use this term in a slightly more general sense.

If $\lambda = e_1$ holds, then any random tight t-fusion frame is already a random cubature of strength t. Still for $\lambda = e_1$, let us consider a random cubature $\mathcal{P} \in \mathcal{G}_{e_1,d}$ with finite support, say $\{P_j\}_{j=1}^n$, and corresponding weight distribution $\{\omega_j\}_{j=1}^n$. Then strength t implies that

$$\sum_{j=1}^n \omega_j \langle P_j, P_x\rangle^t = \mathbb{E}(\langle \mathcal{P}, P_x\rangle^t) = \mu_{\mathcal{P}}^t(P_x) = \mu_{e_1,d}^t(P_x) = \int_{\mathcal{G}_{e_1,d}} \langle P_V, P_x\rangle^t d\sigma_{e_1,d}(V)$$

holds for all P_x, which becomes formula (2) when the weights are constant. Thus, (random) cubatures are a more flexible concept than designs.

The trace moments as functions on $\mathcal{G}_{\lambda,d}$ generate polynomial function spaces, and we define

$$\mathrm{Pol}_t(\mathcal{G}_{\lambda,d}) := \mathrm{span}\{\langle \cdot, X_1\rangle \cdots \langle \cdot, X_t\rangle\big|_{\mathcal{G}_{\lambda,d}} : X_1, \dots, X_t \in \mathscr{H}_d\}. \tag{12}$$

We also define the subspace

$$\mathrm{Pol}_t^1(\mathcal{G}_{\lambda,d}) := \mathrm{span}\{\langle \cdot, P_x\rangle^t\big|_{\mathcal{G}_{\lambda,d}} : x \in \mathbb{S}^{d-1}\}. \tag{13}$$

Existence of cubatures is quite well-understood, and the following results are based on findings in [16]. In fact, the second part of the following proposition is completely contained in [16]. The first part is an analogous proof, cf. [19]:

Proposition 5.4. *There exists a random tight t-fusion frame $\mathcal{P} \in \mathcal{G}_{\lambda,d}$ distributed according to some probability measure ν such that*

$$\#\,\mathrm{supp}(\nu) \le \dim(\mathrm{Pol}_t^1(\mathcal{G}_{\lambda,d})) + 1.$$

There exists a random cubature $\mathcal{P} \in \mathcal{G}_{\lambda,d}$ of strength t distributed according to some probability measure ν such that

$$\#\,\mathrm{supp}(\nu) \le \dim(\mathrm{Pol}_t(\mathcal{G}_{\lambda,d})) - 1.$$

It is noteworthy that the dimension of $\mathrm{Pol}_t(\mathcal{G}_{\lambda,d})$ can be bounded by the number of monomials of degree t in $\frac{1}{2}d(d+1)$ variables, i.e.,

$$\dim(\mathrm{Pol}_t^1(\mathcal{G}_{\lambda,d})) \le \dim(\mathrm{Pol}_t(\mathcal{G}_{\lambda,d})) \le \binom{\frac{1}{2}d(d+1)+t-1}{t}.$$

There is also a tighter bound for $\dim(\mathrm{Pol}_t^1(\mathcal{G}_{\lambda,d}))$, i.e.,

$$\dim(\mathrm{Pol}_t^1(\mathcal{G}_{\lambda,d})) \le \binom{d+2t-1}{2t}, \tag{14}$$

which is a consequence of the following result showing that the dimension can be bounded by the dimension of the homogeneous polynomials of degree $2t$ in d variables.

Lemma 5.5. *For $t \in \mathbb{N}$, we obtain*

$$\dim(\mathrm{Pol}_t^1(\mathcal{G}_{\lambda,d})) = \dim(\mathrm{span}\{\|P\cdot\|^{2t}\big|_{\mathbb{S}^{d-1}} : P \in \mathcal{G}_{\sqrt{\lambda},d}\}),$$

where $\sqrt{\lambda} = (\sqrt{\lambda_1},\dots,\sqrt{\lambda_d})^\star$.

Proof. Let $\{x_i\}_{i=1}^n \subset S^{d-1}$ be such that $\langle x_i x_i^\star, \cdot\rangle^t\big|_{\mathcal{G}_{\lambda,d}}$, $i = 1,\dots,n$ are linearly independent. By classical arguments, there are $\{P_j\}_{j=1}^n \subset \mathcal{G}_{\lambda,d}$, such that the matrix $(\langle x_i x_i^\star, P_j\rangle^t)_{i,j}$ is invertible. Therefore, the functions $\|P_j^{1/2}\cdot\|^{2t}\big|_{S^{d-1}}$, $j = 1,\dots,n$, are linearly independent since $\|P_j^{1/2}x\|^{2t} = \langle xx^\star, P_j\rangle^t$, for all $x \in S^{d-1}$. The same arguments apply vice versa, which concludes the proof. □

It should also be noted that existence of cubatures on the sphere when their support is fixed, i.e., designing the mass distribution according to some fixed locations, have been investigated in [31], and we refer to [23] for more general manifolds. However, general existence results for designs with specific bounds similar to Proposition 5.4 are not known.

After having established existence of cubatures, we can now state our main result on phase retrieval, which generalizes Theorem 3.1.

Theorem 5.6. *Suppose that $\|x\|^2$ is known and that $\{\mathcal{P}_j\}_{j=1}^n \subset \mathcal{G}_{\lambda,d}$ are independently sampled from a random cubature of strength 3, which is also a random tight t-fusion frame for some $t \ge 3$. Then with probability at least $1-e^{-\omega}$, the rank-one matrix $xx^\star$ is the unique solution to (5) provided that*

$$n \ge c_1 \omega t d^{1+2/t}\log^2(d), \tag{15}$$

where $\omega \ge 1$ is an arbitrary parameter and c_1 is a constant, which does not depend on d.

Few comments are in order. In contrast to Theorem 3.1, we allow random cubatures that are not uniformly distributed on their support. Furthermore, we can separate the cubature condition of strength 3 from the tight frame requirements for $t > 3$, which are indeed different concepts when k is bigger than 1. Note that the number of measurements n scales linearly in the ambient dimension d up to logarithmic factors if we choose $t = \log(d)$ because then (15) yields $n \ge c_1\omega d\log^3(d)$.

6 General structure of the proof of Theorem 5.6

The proof of Theorem 5.6 is guided by the structure provided in [26] and based on the following two results about near isometry properties and the existence of dual certificates as required by Theorem 4.2.

Theorem 6.1. *If $\{P_j\}_{j=1}^n \subset \mathcal{G}_{\lambda,d}$ are independent and identical copies of a random matrix $\mathcal{P} \in \mathcal{G}_{\lambda,d}$ that is a random cubature of strength 3, then, for any sufficiently large constant C_0, there is a constant $c > 0$ such that*

$$\frac{1}{C_0 d^2}\|X\|_F^2 \le \frac{1}{n}\|\mathcal{A}_n(X)\|^2 \tag{16}$$

holds for all matrices $X \in T_x$ simultaneously with probability of failure at most $d^2 e^{-c\frac{n}{d}}$.

Note that the constant C_0 will be used in the remaining part of the present paper. We still need an approximate dual certificate though.

Theorem 6.2. *Suppose that $0 \neq x \in \mathbb{R}^d$, that $\omega \ge 1$, and that $\mathcal{P} \in \mathcal{G}_{\lambda,d}$ is a random cubature of strength 3 and a random tight t-fusion frame for some $t \ge 3$. Then, for any sufficiently large constant c_0, there is a constant $c > 0$ such that if the number of measurements satisfies*

$$n \ge c\omega t d^{1+2/t}\log^2(d), \tag{17}$$

then with probability of failure at most $\frac{1}{2}e^{-\omega}$, there exists a $(\frac{1}{c_0 d}, \frac{1}{c_0})$-dual certificate with respect to x.

Note that the constant c_0 is used in the remaining part of the present paper. Now, we have all ingredients for the proof of our main result on phase retrieval:

Proof of Theorem 5.6. Guided by the structure provided in [26], we aim to apply Theorem 4.2, and the upper bound in the near isometry property can easily be verified. Indeed, for any collection $\{P_j\}_{j=1}^n \subset \mathcal{G}_{\lambda,d}$, the Cauchy-Schwartz inequality and the assumptions on λ yield

$$\frac{1}{n}\|\mathcal{A}_n(X)\|^2 = \frac{1}{n}\sum_{j=1}^n \langle X, P_j\rangle^2 \le \frac{1}{n}\sum_{j=1}^n \|X\|_F^2\|P_j\|_F^2 \le k\|X\|_F^2, \quad \text{for all } X \in \mathcal{H}_d.$$

Hence, we can choose $\beta := k$. According to Theorem 6.1, we can select $\alpha = \frac{1}{C_0 d^2}$. By choosing $\gamma := \frac{1}{c_0 d}$ and $\delta := \frac{1}{c_0}$ with $c_0 > \sqrt{kC_0} + 1$, Theorem 6.2 yields a (γ, δ)-dual certificate for the required number of measurements, and we have the estimate

$$\sqrt{\frac{\beta}{\alpha}} = \sqrt{C_0}d\sqrt{k} < d(c_0 - 1) = \frac{1-\delta}{\gamma}.$$

Thus, the assumptions in Theorem 4.2 are satisfied. The corresponding probabilities work out nicely by applying $d^2 e^{-c\frac{n}{d}} = \frac{1}{2}e^{\log(2)+2\log(d)-c\frac{n}{d}}$ and $\omega \ge 1$, which concludes the proof. □

In order to complete the proof of Theorem 5.6, we must still verify Theorems 6.1 and 6.2. Their proofs require the actual computation of t-th trace moments of $\mathcal{P} \sim \sigma_{\lambda,d}$ for $t = 1, 2, 3$ that we shall discuss in the subsequent sections. In fact, we shall present a closed formula for the t-th trace moments for all t based on zonal polynomials.

7 Computing trace moments

7.1 Some special trace moments

For special choices of λ and X, the trace moments of $\mathcal{P} \sim \sigma_{\lambda,d}$ are already known. If $\lambda = (1, \dots, 1, 0, \dots, 0)^\star$, where 1 is repeated k times, an explicit expression for the moments of rank-1 matrices $X = xx^\star$ can be derived,

$$\mu_{\lambda,d}^t(xx^\star) = \frac{(k/2)_t}{(d/2)_t}\cdot\|x\|^{2t}, \quad \text{for all } x \in \mathbb{R}^d, \tag{18}$$

where $(a)_t := a(a+t)\cdots(a+t-1)$ denotes the Pochhammer symbol, cf. [1]. Recall that those are the moments needed for the characterization of random tight t-fusion frames.

Moreover, we further restrict λ to derive explicit formulas for more general moments. If $\mathcal{P} \sim \sigma_{e_1,d}$ and $\{x_i\}_{i=1}^d$ is an orthonormal basis for $\mathbb{R}^d$, then one can verify that the vector $\left(\langle \mathcal{P}, x_1 x_1^\star \rangle, \ldots, \langle \mathcal{P}, x_d x_d^\star \rangle\right)^\star$ is Dirichlet distributed with parameter vector $(1/2, \ldots, 1/2)$, [6]. The generalized moments of such Dirichlet distributed random vectors are known [39], and, indeed, if $\beta \in \mathbb{N}^d$, we obtain

$$\mu^\beta_{e_1,d}(x_1 x_1^\star, \ldots, x_d x_d^\star) = \frac{\prod_{i=1}^d (1/2)_{\beta_i}}{(d/2)_{|\beta|}}, \tag{19}$$

where $|\beta| = \sum_{i=1}^d \beta_i$. Since $\sigma_{e_1,d}$ is invariant under the orthogonal group, the terms in (19) do not depend on the special choice of the orthonormal basis. Hence, the spectral decomposition of $X \in \mathscr{H}_d$ yields a closed formula,

$$\mu^t_{e_1,d}(X) = \sum_{\substack{\beta \in \mathbb{N}^d \\ |\beta|=t}} \binom{t}{\beta} \alpha^\beta \frac{\prod_{i=1}^d (1/2)_{\beta_i}}{(d/2)_t}, \quad \text{for } X \in \mathscr{H}_d, \tag{20}$$

where $\alpha = (\alpha_1, \ldots, \alpha_d)$ are the eigenvalues of X.

7.2 Trace moments for general λ, t, and X

Computing trace moments when λ is more general requires the theory of zonal polynomials as developed in [27, 28], see also the textbooks [15, 32]. Zonal polynomials are homogeneous polynomials in $\mathscr{H}_d$, which are invariant under conjugation with respect to the orthogonal group. According to representation theory, those polynomials C_π are indexed by all partitions π of the natural numbers. Here, a partition of t are integer vectors $\pi = (\pi_1, \ldots, \pi_t)$ with $\pi_1 \geq \ldots \geq \pi_t \geq 0$ and $\sum_{i=1}^t \pi_i = t$. The number of parts of π is the number of nonzero entries. The set of partitions of t with no more than d parts is denoted by $\mathscr{P}_{t,d}$.

To compute cross-moments of a random matrix $\mathcal{P} \in \mathcal{G}_{\lambda,d}$, we shall make use of the following combinatorial fact:

Lemma 7.1. *For any integer $t \geq 1$ and $x_1 \ldots, x_t \in \mathbb{R}$, we have*

$$t! x_1 \cdots x_t = \sum_{J \subset \{1,\ldots,t\}} (-1)^{t+\#J} \Big(\sum_{j \in J} x_j\Big)^t.$$

Proof. Consider the homogeneous and symmetric polynomials

$$S^\ell_t(x_1, \ldots, x_t) := \sum_{\substack{J \subset \{1,\ldots,t\} \\ \#J=\ell}} \Big(\sum_{j \in J} x_j\Big)^t$$

of degree t. The coefficient of the monomial x^β, for $\beta \in \mathbb{N}^t$, $|\beta| = t$, in $\big(\sum_{j \in J} x_j\big)^t$ is

$$\begin{cases} \binom{t}{\beta}, & \operatorname{supp}(\beta) \subset J, \\ 0, & \text{otherwise}, \end{cases}$$

where $\binom{t}{\beta} = \frac{t!}{\beta_1! \cdots \beta_t!}$. Together with

$$\#\{J \subset \{1, \ldots, t\} : \#J = \ell,\ \operatorname{supp}(\beta) \subset J\} = \binom{t - \#\operatorname{supp}(\beta)}{\ell - \#\operatorname{supp}(\beta)},$$

we can conclude

$$S^\ell_t(x_1, \ldots, x_t) = \sum_{\substack{\beta \in \mathbb{N}^t \\ |\beta|=t}} \binom{t}{\beta} \binom{t - \#\operatorname{supp}(\beta)}{\ell - \#\operatorname{supp}(\beta)} x^\beta.$$

This yields

$$\begin{aligned}\sum_{J\subset\{1,\dots,t\}}(-1)^{t+\#J}\Big(\sum_{j\in J}x_j\Big)^t &= \sum_{\substack{\beta\in\mathbb{N}^t\\|\beta|=t}}\binom{t}{\beta}x^\beta\sum_{\ell=\#\operatorname{supp}(\beta)}^{t}(-1)^{t+\ell}\binom{t-\#\operatorname{supp}(\beta)}{t-\ell}\\ &= \sum_{\substack{\beta\in\mathbb{N}^t\\|\beta|=t}}\binom{t}{\beta}x^\beta(-1)^{t+\#\operatorname{supp}(\beta)}\sum_{\ell=0}^{t-\#\operatorname{supp}(\beta)}(-1)^{\ell}\binom{t-\#\operatorname{supp}(\beta)}{\ell}\\ &= \sum_{\substack{\beta\in\mathbb{N}^t\\|\beta|=t}}\binom{t}{\beta}x^\beta(-1)^{t+\#\operatorname{supp}(\beta)}(1-1)^{t-\#\operatorname{supp}(\beta)},\end{aligned}$$

with $0^0 := 1$, which concludes the proof. □

The latter lemma enables us to actually compute trace moments:

Theorem 7.2. *The cross-moments of any random matrix $\mathcal{P}\in\mathcal{G}_{\lambda,d}$ satisfy*

$$\mu_{\mathcal{P}}(X_1,\dots,X_t)=\frac{1}{t!}\sum_{J\subset\{1,\dots,t\}}(-1)^{t+\#J}\mu_{\mathcal{P}}^t\Big(\sum_{j\in J}X_j\Big),\quad X_1,\dots,X_t\in\mathscr{H}_d. \tag{21}$$

In particular, if $\mathcal{P}\sim\sigma_{\lambda,d}$, then (21) *can be computed by*

$$\mu_{\lambda,d}^t(X)=\sum_{\pi\in\mathscr{P}_{t,d}}\frac{C_\pi(X)C_\pi(D_\lambda)}{C_\pi(I_d)},\quad \textit{for all } X\in\mathscr{H}_d, \tag{22}$$

where $D_\lambda=\operatorname{diag}(\lambda_1,\dots,\lambda_d)^\star$.

Proof. The formula (21) is a direct consequence of Lemma 7.1.

Equation (22) follows from properties of zonal polynomials, cf. [27], namely

$$\begin{aligned}\operatorname{trace}(X)^t &= \sum_{\pi\in\mathscr{P}_t}C_\pi(X),\\ \int_{\mathcal{O}(d)}C_\pi(XOYO^\star)dO &= \frac{C_\pi(X)C_\pi(Y)}{C_\pi(I_d)},\end{aligned}$$

for all $X, Y\in\mathscr{H}_d$, and that $\sigma_{\lambda,d}$ is induced by the Haar measure on the orthogonal group implying

$$\int_{\mathcal{G}_{\lambda,d}}C_\pi(XP)d\sigma_{\lambda,d}(P)=\int_{\mathcal{O}(d)}C_\pi(XOD_\lambda O^\star)dO.$$ □

7.3 Explicit trace moments for $t = 1, 2, 3$

To make use of Theorem 7.2 enabling us to compute trace moments of $\sigma_{\lambda,d}$ for $t = 1, 2, 3$, we still need explicit forms of the zonal polynomials. Indeed, they were computed in [27]:

$$
\begin{aligned}
C_{(1)}(X) &= \operatorname{trace}(X)\\
C_{(2)}(X) &= \frac{1}{3}\big(\operatorname{trace}^2(X) + 2\operatorname{trace}(X^2)\big)\\
C_{(1,1)}(X) &= \frac{2}{3}\big(\operatorname{trace}^2(X) - \operatorname{trace}(X^2)\big)\\
C_{(3)}(X) &= \frac{1}{15}\big(\operatorname{trace}^3(X) + 6\operatorname{trace}(X)\operatorname{trace}(X^2) + 8\operatorname{trace}(X^3)\big)\\
C_{(2,1)}(X) &= \frac{3}{5}\big(\operatorname{trace}^3(X) + \operatorname{trace}(X)\operatorname{trace}(X^2) - 2\operatorname{trace}(X^3)\big)\\
C_{(1,1,1)}(X) &= \frac{1}{3}\big(\operatorname{trace}^3(X) - 3\operatorname{trace}(X)\operatorname{trace}(X^2) + 2\operatorname{trace}(X^3)\big)
\end{aligned}
$$

We can now apply Theorem 7.2, which yields the trace moments for $t = 1, 2, 3$:

Theorem 7.3. *For all $d \geq 3$ and $X_1, X_2, X_3 \in \mathscr{H}_d$, we have*

$$
\begin{aligned}
\mu_{\lambda,d}(X_1) &= \frac{1}{q_{1,d}}\alpha_{(1)}\operatorname{trace}(X_1),\\
\mu_{\lambda,d}(X_1, X_2) &= \frac{1}{q_{2,d}}\big(\alpha_{(1,1)}\operatorname{trace}(X_1)\operatorname{trace}(X_2) + \alpha_{(2)}\operatorname{trace}(X_1X_2)\big),\\
\mu_{\lambda,d}(X_1, X_2, X_3) &= \frac{1}{q_{3,d}}\big(\alpha_{(1,1,1)}\operatorname{trace}(X_1)\operatorname{trace}(X_2)\operatorname{trace}(X_3) +\\
&\quad \frac{\alpha_{(2,1)}}{3}(\operatorname{trace}(X_1)\operatorname{trace}(X_2X_3) + \operatorname{trace}(X_2)\operatorname{trace}(X_1X_3) + \operatorname{trace}(X_3)\operatorname{trace}(X_1X_2))\\
&\quad \alpha_{(3)}\operatorname{trace}(X_1X_2X_3)\big),
\end{aligned}
$$

where we set $s_i := \operatorname{trace}(D_\lambda^i)$ and

$$
\begin{aligned}
q_{1,d} &= d,\\
\alpha_{(1)} &= s_1,
\end{aligned}
$$

$$
\begin{aligned}
q_{2,d} &= (d-1)d(d+2),\\
\alpha_{(1,1)} &= (d+1)s_1^2 - 2s_2\\
\alpha_{(2)} &= -2s_1^2 + 2ds_2,
\end{aligned}
$$

$$
\begin{aligned}
q_{3,d} &= (d-2)(d-1)d(d+2)(d+4),\\
\alpha_{(1,1,1)} &= (d^2+3d-2)s_1^3 - 6(d+2)s_1s_2 + 16s_3,\\
\alpha_{(2,1)} &= -6(d+2)s_1^3 + 6(d^2+2d+4)s_1s_2 - 24ds_3,\\
\alpha_{(3)} &= 16s_1^3 - 24ds_1s_2 + 8d^2s_3.
\end{aligned}
$$

If we keep the last matrix argument undetermined, then we derive the following result, which is simply a weak formulation of Theorem 7.3:

Corollary 7.4. *Let a random matrix $\mathcal{P} \in \mathcal{G}_{\lambda,d}$ be given. If $\mathcal{P}$ is a random cubature of strength 2, then, for $d \geq 2$ and $X \in \mathscr{H}_d$,*

$$
a_1\mathbb{E}\langle \mathcal{P}, X\rangle\mathcal{P} = X + a_2\operatorname{trace}(X)I_d, \tag{23}
$$

where $a_1 = \frac{d(d+2)(d-1)}{-2s_1^2+2ds_2}$ and $a_2 = \frac{(d+1)s_1^2-2s_2}{-2s_1^2+2ds_2}$. Moreover, if $\mathcal{P}$ is a random cubature of strength 3, then, for $d \geq 3$ and $X_1, X_2 \in \mathscr{H}_d$,

$$
\begin{aligned}
\mathbb{E}\langle \mathcal{P}, X_1\rangle\langle \mathcal{P}, X_2\rangle\mathcal{P} = \frac{1}{q_{3,d}}\big(&\alpha_{(1,1,1)}\operatorname{trace}(X_1)\operatorname{trace}(X_2)I_d + \frac{\alpha_{(2,1)}}{3}(\operatorname{trace}(X_1)X_2 + \operatorname{trace}(X_2)X_1 + \operatorname{trace}(X_1X_2)I_d)\\
&\alpha_{(3)}\operatorname{trace}(X_1X_2)I_d\big).
\end{aligned}
$$

Note that (23) has been derived for $\lambda = (1, \dots, 1, 0, \dots, 0)^\star$ in [2]. It is also worth mentioning that a_1 is on the order of d^2 when d tends to infinity and a_2 behaves like a constant that may depend on k. The coefficients $\frac{\alpha_{(2,1)}}{q_{3,d}}$, $\frac{\alpha_{(3)}}{q_{3,d}}$, and $\frac{\alpha_{(1,1,1)}}{q_{3,d}}$ behave like $1/d^3$ when d tends to infinity.

We establish one more consequence:

Corollary 7.5. *Suppose that $d \geq 2$. If a random matrix $\mathcal{P} \in \mathcal{G}_{\lambda,d}$ is a random cubature of strength 2, then*

$$\mathcal{S}a_1\mathbb{E}(\langle \mathcal{P}, X\rangle \mathcal{P}) = a_1\mathbb{E}(\langle \mathcal{P}, \mathcal{S}X\rangle \mathcal{P}) = X, \quad \text{for all } X \in \mathscr{H}_d,$$

where $\mathcal{S} : \mathscr{H}_d \to \mathscr{H}_d$, $X \mapsto X - \frac{a_2}{1+a_2 d}\operatorname{trace}(X) I_d$.

Note that $\mathcal{S}$ in Corollary 7.5 is a contraction, so that

$$\mathcal{I} \succeq \mathcal{S} \succeq 0, \qquad \|\mathcal{S}\|_{Op} \leq 1, \tag{24}$$

where $\mathcal{I}$ denotes the identity map on $\mathscr{H}_d$.

Remark 7.6. By using the theory of zonal polynomials, we have explicitly computed the t-th trace moments for $t = 1, 2, 3$ to be able to verify the Theorems 6.1 and 6.2. Indeed, the trace moments are an essential ingredient in their proofs. We established the Corollaries 7.4 and 7.5 in our more general setting. Next, we can essentially follow the lines in [26] with adjusted parameters and minor modifications to verify the Theorems 6.1 and 6.2, and their complete proofs are placed in Appendix A and B.

8 Conclusions

Our results generalize findings in [26] from 1-dimensional subspace measurements to the general setting of rank-k positive semidefinite matrices. Moreover, we deal with cubatures in place of the required t-designs in [26].

Our proofs were guided by the approach in [26]. In our general setting, we had to compute the trace moments on $\mathcal{G}_{\lambda,d}$ for $t = 1, 2, 3$ by applying zonal polynomials as discussed in [27]. Based on such findings, we then followed the structure in [26] with adjusted parameters and constants in the appendix to verify the phase retrieval results. We only explicitly addressed the real setting, but the theory of complex zonal polynomials also works in complex space with adjusted coefficients, but the asymptotics in d remain the same, so that our approach can cover the complex phase retrieval setting as well.

A Near isometries: proof of Theorem 6.1

To prove Theorem 6.1, we shall make use of the following deviation bound that was also used in [26]:

Theorem A.1 ([36]). *Let $S = \sum_{j=1}^n M_j$ be a sum of independently identically distributed $d \times d$ random matrices with zero mean and smallest eigenvalue $\lambda_{min} \geq -R$ almost surely. For $\sigma^2 = \|\sum_{j=1}^n \mathbb{E}M_j^2\|_{Op}$, the smallest eigenvalue Λ_{min} of S satisfies, for all $q \geq 0$,*

$$\operatorname{Prob}\left(\Lambda_{min} \leq -q\right) \leq d\exp\left(-\frac{q^2/2}{\sigma^2 + Rq/3}\right) \leq d\begin{cases}\exp(-3q^2/8\sigma^2), & q \leq \sigma^2/R,\\ \exp(-3q/8R), & q \geq \sigma^2/R.\end{cases}$$

Proof of Theorem 6.1. We make use of the mapping

$$\mathcal{R} : \mathscr{H}_d \to \mathscr{H}_d, \quad X \mapsto \frac{a_1}{n}\sum_{j=1}^n \langle X, \mathcal{P}_j\rangle \mathcal{P}_j, \tag{25}$$

where a_1 is as in Corollary 7.4 and whose expectation was derived there too. Without loss of generality, we can assume $x \neq 0$. As in the proof of Proposition 9 in [26], we derive

$$\frac{1}{a_1}\left(1+\Lambda_{min}\right)\|X\|_F^2 \leq \frac{1}{n}\|\mathcal{A}_n(x)\|^2$$

where Λ_{min} is the minimal eigenvalue of $\mathrm{P}_{T_x}(\mathcal{R}-\mathbb{E}\mathcal{R})\mathrm{P}_{T_x}$. Here, P_{T_x} is the orthogonal projector onto T_x, explicitly given by

$$\mathrm{P}_{T_x} : \mathscr{H}_d \to \mathscr{H}_d, \quad X \mapsto P_{x\mathbb{R}}X + XP_{x\mathbb{R}} - \langle X, P_{x\mathbb{R}}\rangle P_{x\mathbb{R}}, \tag{26}$$

where $P_{x\mathbb{R}} = \frac{1}{\|x\|^2}xx^*$ is the orthogonal projector onto $x\mathbb{R}$. Thus, we must find a lower bound on Λ_{min}.

We now split

$$\mathrm{P}_{T_x}(\mathcal{R}-\mathbb{E}\mathcal{R})\mathrm{P}_{T_x} = \sum_{j=1}^{n}\mathcal{M}_j - \mathbb{E}\mathcal{M}_j, \quad \text{where } \mathcal{M}_j = \frac{a_1}{n}\langle \mathrm{P}_{T_x}\cdot, \mathcal{P}_j\rangle(\mathcal{P}_j)_{T_x}.$$

It is fairly easy to see that $\langle X_{T_x}, I_d\rangle I_{T_x} = \langle X, P_{x\mathbb{R}}\rangle P_{x\mathbb{R}}$, which implies

$$-\frac{2}{n}\mathcal{I} \preceq -\frac{1}{n}\mathcal{I} - \frac{1}{n}\langle\cdot, P_{x\mathbb{R}}\rangle P_{x\mathbb{R}} \preceq \mathrm{P}_{T_x}(\mathcal{M}_j - \mathbb{E}\mathcal{M}_j)\mathrm{P}_{T_x},$$

so that

$$-\frac{2}{n} \leq \lambda_{min}, \tag{27}$$

where λ_{min} is the minimal eigenvalue of $\mathcal{M}_j - \mathbb{E}\mathcal{M}_j$. We have

$$0 \preceq \mathbb{E}((\mathcal{M}_j - \mathbb{E}\mathcal{M}_j)^2) \preceq \mathbb{E}(\mathcal{M}_j^2) = \frac{a_1^2}{n^2}\mathbb{E}\langle\cdot, (\mathcal{P}_j)_{T_x}\rangle\langle(\mathcal{P}_j)_{T_x}, (\mathcal{P}_j)_{T_x}\rangle(\mathcal{P}_j)_{T_x}$$

and according to $\langle(\mathcal{P}_j)_{T_x}, (\mathcal{P}_j)_{T_x}\rangle = \langle\mathcal{P}_j, (\mathcal{P}_j)_{T_x}\rangle$, we can use (26) to derive

$$\langle(\mathcal{P}_j)_{T_x}, (\mathcal{P}_j)_{T_x}\rangle = \mathrm{trace}(\mathcal{P}_j(P_{x\mathbb{R}}\mathcal{P}_j + \mathcal{P}_jP_{x\mathbb{R}} - \langle\mathcal{P}_j, P_{x\mathbb{R}}\rangle P_{x\mathbb{R}}) \leq 2\langle\mathcal{P}_j, P_{x\mathbb{R}}\rangle,$$

so that

$$\mathbb{E}((\mathcal{M}_j - \mathbb{E}\mathcal{M}_j)^2) \preceq \frac{2a_1^2}{n^2}\mathrm{P}_{T_x}\mathbb{E}\langle\mathrm{P}_{T_x}, \mathcal{P}_j\rangle\langle\mathcal{P}_j, P_{x\mathbb{R}}\rangle\mathcal{P}_j.$$

Since we have a cubature of strength 3, we derive the estimates

$$\begin{aligned}
\mathbb{E}((\mathcal{M}_j - \mathbb{E}\mathcal{M}_j)^2) &\preceq \frac{2a_1^2}{n^2}\mathrm{P}_{T_x}\mathbb{E}\langle\mathrm{P}_{T_x}, \mathcal{P}_j\rangle\langle\mathcal{P}_j, P_{x\mathbb{R}}\rangle\mathcal{P}_j \\
&= \frac{2a_1^2}{n^2}\mathrm{P}_{T_x}\big[\alpha_1(\mathrm{P}_{T_x} + \mathrm{trace}(\mathrm{P}_{T_x}\cdot)P_{x\mathbb{R}} + \mathrm{trace}((\mathrm{P}_{T_x}\cdot)P_{x\mathbb{R}})I) \\
&\quad + \alpha_2((\mathrm{P}_{T_x}\cdot)P_{x\mathbb{R}} + P_{x\mathbb{R}}\mathrm{P}_{T_x}) + \alpha_3\,\mathrm{trace}(\mathrm{P}_{T_x}\cdot)I\big] \\
&= \frac{2a_1^2}{n^2}\big[\alpha_1(\mathrm{P}_{T_x} + \mathrm{trace}((\mathrm{P}_{T_x}\cdot)P_{x\mathbb{R}})P_{x\mathbb{R}} + \mathrm{trace}((\mathrm{P}_{T_x}\cdot)P_{x\mathbb{R}})P_{x\mathbb{R}}) \\
&\quad + \alpha_2((\mathrm{P}_{T_x}\cdot)P_{x\mathbb{R}} + P_{x\mathbb{R}}\mathrm{P}_{T_x}) + \alpha_3\,\mathrm{trace}((\mathrm{P}_{T_x}\cdot)P_{x\mathbb{R}})P_{x\mathbb{R}}\big] \\
&= \frac{2a_1^2}{n^2}\big[\alpha_1(\mathrm{P}_{T_x} + \mathrm{trace}((\mathrm{P}_{T_x}\cdot)P_{x\mathbb{R}})P_{x\mathbb{R}} + \mathrm{trace}((\mathrm{P}_{T_x}\cdot)P_{x\mathbb{R}})P_{x\mathbb{R}}) \\
&\quad + \alpha_2(\mathrm{P}_{T_x} + P_{x\mathbb{R}}\,\mathrm{trace}(P_{x\mathbb{R}}\mathrm{P}_{T_x}\cdot)) + \alpha_3\,\mathrm{trace}((\mathrm{P}_{T_x}\cdot)P_{x\mathbb{R}})P_{x\mathbb{R}}\big],
\end{aligned}$$

where we have used (26) twice and $\mathrm{trace}(\mathrm{P}_{T_x}\cdot P_{x\mathbb{R}}) = \mathrm{trace}(\mathrm{P}_{T_x}\cdot)$. Next, we apply $\alpha_i \leq \frac{c}{d^3}$ for sufficiently large d and obtain

$$\begin{aligned}
\mathbb{E}((\mathcal{M}_j - \mathbb{E}\mathcal{M}_j)^2) &\preceq \frac{2ca_1^2}{n^2d^3}\big[\mathrm{P}_{T_x} + \mathrm{trace}((\mathrm{P}_{T_x}\cdot)P_{x\mathbb{R}})P_{x\mathbb{R}} + \mathrm{trace}((\mathrm{P}_{T_x}\cdot)P_{x\mathbb{R}})P_{x\mathbb{R}} \\
&\quad + \mathrm{P}_{T_x} + P_{x\mathbb{R}}\,\mathrm{trace}(P_{x\mathbb{R}}\mathrm{P}_{T_x}\cdot) + \mathrm{trace}((\mathrm{P}_{T_x}\cdot)P_{x\mathbb{R}})P_{x\mathbb{R}}\big] \\
&= \frac{2ca_1^2}{n^2d^3}\big[4\,\mathrm{trace}((\mathrm{P}_{T_x}\cdot)P_{x\mathbb{R}})P_{x\mathbb{R}} + 2\mathrm{P}_{T_x}\big] \\
&\preceq \frac{16ca_1^2}{n^2d^3}\mathcal{I}.
\end{aligned}$$

The rough estimate $a_1 \le c_2 d^2$ implies

$$\mathbb{E}((\mathcal{M}_j - \mathbb{E}\mathcal{M}_j)^2) \preceq \frac{16cc_2^2 d}{n^2}\mathcal{I}.$$

Let $\sigma^2 := \frac{16cc_2^2 d}{n}$, so that Theorem A.1 yields with $R = 2/n$, see (27),

$$\begin{aligned}\text{Prob}\left(\Lambda_{min} \le -\epsilon\right) &\le d^2 \exp(-\frac{n\epsilon^2/2}{16cc_2^2 d + 2\epsilon/3}) \\ &\le d^2 \exp(-\frac{n\epsilon^2}{32cc_2^2 d + 4\epsilon/3}) \le d^2 \exp(-\frac{c_3 n\epsilon^2}{d}),\end{aligned}$$

for all $0 \le \epsilon \le 1 \le 32cc_2^2 d = \sigma^2/R$.

So far, we have verified that

$$\frac{1}{a_1}\left(1-\epsilon\right)\|X\|_F^2 \le \frac{1}{n}\|\mathcal{A}_n(X)\|^2$$

holds with probability of failure at most $d^2 \exp(-\frac{c_3 n\epsilon^2}{d})$. If we choose ϵ fixed such that $\epsilon \le 1 - \frac{a_1}{c_2 d^2}$, then we can conclude the proof. □

B Dual certifcate: proof of Theorem 6.2

We first derive a bound for $\mu_{\lambda,d}^t(xx^*)$:

Proposition B.1. *If $x \in S^{d-1}$, then we have*

$$\mu_{\lambda,d}^t(xx^*) \le \left(\frac{kt}{d}\right)^t.$$

Proof. This bound has been derived in [2] for λ having k ones and $d-k$ zeros. The general conditions on λ, i.e., only k entries are nonzero and $\lambda_i \le 1$, imply the statement. □

We shall now bound $\langle \mathcal{P}, xx^* \rangle$:

Proposition B.2. *Suppose that $x \in S^{d-1}$. If $\mathcal{P} \in \mathcal{G}_{\lambda,d}$ is a random tight t-fusion frame with $t \ge 1$, then we have, for all $0 < r \le 1 \le s$,*

$$\langle \mathcal{P}, xx^* \rangle \le (s+1)tkd^{-r}$$

with probability of failure at most $s^{-t}d^{-t(1-r)}$.

Proof. For $s \ge 1$, we estimate

$$\begin{aligned}\text{Prob}\left(\langle \mathcal{P}, xx^* \rangle \ge (s+1)tkd^{-r}\right) &\le \text{Prob}\left(\langle \mathcal{P}, xx^* \rangle - \mu_{\mathcal{P}}(xx^*) \ge (s+1)tkd^{-r} - \frac{k}{d}\right) \\ &\le \text{Prob}\left(\langle \mathcal{P}, xx^* \rangle - \mu_{\mathcal{P}}(xx^*) \ge stkd^{-r}\right),\end{aligned}$$

where we have used Theorem 7.3 and $\text{trace}(D_\lambda) \le k$. Due to Proposition B.1, $\tau_t := \left(\mu_{\mathcal{P}}^t(xx^*)\right)^{1/t} \le \frac{kt}{d}$ holds, so that we obtain

$$\text{Prob}\left(\langle \mathcal{P}, xx^* \rangle \ge (s+1)tkd^{-r}\right) \le \text{Prob}\left(\left|\langle \mathcal{P}, xx^* \rangle - \mu_{\mathcal{P}}(xx^*)\right| \ge sd^{1-r}\tau_t\right).$$

We can conclude the proof by applying a generalized Chebyshev inequality that was used in the proof of Lemma 13 in [26], i.e.,

$$\text{Prob}\left(|\langle \mathcal{P}, xx^* \rangle - \mu_{\mathcal{P}}(xx^*)| \ge u\tau_t\right) \le u^{-t}$$

and by choosing $u = sd^{1-r}$. □

To introduce a sampled truncation of the operator $\mathcal{R}$ defined in (25), we denote the event

$$E_j := \{\langle \mathcal{P}_j, xx^* \rangle \le (s+1)tkd^{-r}\}$$

where $\{\mathcal{P}_j\}_{j=1}^n \subset \mathcal{G}_{\lambda,d}$ are i.i.d. according to a random tight t-fusion frame. The number $0 < r \le 1$ is referred to as the *truncation rate*. We also decompose a fixed $0 \neq Z \in T_x$ by $Z = \lambda(xz^* + zx^*)$, where $\lambda > 0$ and $z \in S^{d-1}$. For this z, we introduce the event

$$G_j := \{\langle \mathcal{P}_j, zz^* \rangle \le (s+1)tkd^{-r}\}$$

and define

$$\mathcal{R} : \mathscr{H} \to \mathscr{H}, \quad X \mapsto \frac{a_1}{n} \sum_{j=1}^n \langle X, \mathcal{P}_j \rangle \mathcal{P}_j,$$

where a_1 is as in Corollary 7.4, which is the analogue of (25) with its truncated counterpart

$$\mathcal{R}_Z : \mathscr{H} \to \mathscr{H}, \quad X \mapsto \frac{a_1}{n} \sum_{j=1}^n 1_{E_j} 1_{G_j} \langle X, \mathcal{P}_j \rangle \mathcal{P}_j. \tag{28}$$

It turns out that $\mathcal{R}$ and $\mathcal{R}_Z$ are close to each other in expectation:

Proposition B.3. *For $x \in \mathbb{R}^d$, fix $Z \in T_x$ and let $\mathcal{R}_Z$ be as in* (28)*, where $\{\mathcal{P}_j\}_{j=1}^n \subset \mathcal{G}_{\lambda,d}$ are i.i.d. according to a random tight t-fusion frame with $t \ge 2$. Then, for any sufficiently large constant c_0, we have*

$$\left\| \mathbb{E}(\mathcal{R}_Z - \mathcal{R}) \right\|_{Op} \le c_0 s^{-t} d^{2-t(1-r)}.$$

Proof. We first define the auxiliar operator

$$\mathcal{R}_{aux} : \mathscr{H} \to \mathscr{H}, \quad X \mapsto \frac{a_1}{n} \sum_{j=1}^n 1_{E_j} \langle X, \mathcal{P}_j \rangle \mathcal{P}_j.$$

The triangular inequality yields

$$\left\| \mathbb{E}(\mathcal{R}_Z - \mathcal{R}) \right\|_{Op} \le \left\| \mathbb{E}(\mathcal{R}_Z - \mathcal{R}_{aux}) \right\|_{Op} + \left\| \mathbb{E}(\mathcal{R}_{aux} - \mathcal{R}) \right\|_{Op}.$$

Since $\|\langle X, \mathcal{P}_j \rangle \mathcal{P}_j\| \le k\|X\|$, we obtain with Proposition B.2

$$\begin{aligned} \left\| \mathbb{E}(\mathcal{R}_{aux} - \mathcal{R}) \right\|_{Op} &\le \frac{a_1 k}{n} \sum_{j=1}^n \mathrm{Prob}(E_j^c) \\ &\le a_1 k s^{-t} d^{-t(1-r)} \\ &\le \frac{c_0}{2} d^2 s^{-t} d^{-t(1-r)}, \end{aligned}$$

since a_1 behaves like d^2. The analogue estimates for $\left\| \mathbb{E}(\mathcal{R}_{aux} - \mathcal{R}) \right\|_{Op}$ using $c_0/2$ conclude the proof. □

Let $\mathrm{P}_{T_x} : \mathscr{H}_d \to T_x$ be the orthogonal projector onto T_x, i.e., $\mathrm{P}_{T_x}(Y) = Y_{T_x}$, and $\mathcal{P}_{T_x^\perp}$ the orthogonal projector onto the orthogonal complement of $T_x^\perp$.

Proposition B.4. *For $x \in \mathbb{R}^d$, fix $Z \in T_x$ and let $\mathcal{R}_Z$ be as in* (28)*, where $\{V_j\}_{j=1}^n \subset \mathcal{G}_{k,d}$ are i.i.d. according to a cubature of strength $t \ge 3$, and the truncation rate is supposed to satisfy $r \le 1 - 2/t$. Then there is a constant $c_1 > 0$ such that, for $1/c_0 \le A \le 1$ and $\sqrt{2}A \le B$,*

$$\left\| \mathcal{P}_{T_x^\perp} \mathcal{S} \mathcal{R}_Z Z \right\|_{Op} \le A\|Z\|_F, \tag{29}$$

$$\left\| \mathrm{P}_{T_x} (\mathcal{S}\mathcal{R}_Z - \mathcal{I}) Z \right\|_F \le B\|Z\|_F, \tag{30}$$

hold with probability of failure at most $d \exp(-c_1 \frac{nA}{td^{2-r}})$.

For the proof of the above proposition, we need the following concentration bound from [25, 37]

Theorem B.5 ([25, 37]). *Consider a finite sequence $\{M_j\}_{j=1}^n$ of independent, random self-adjoint operators on $\mathbb{C}^d$. Assume that $\mathbb{E}M_j = 0$ and $\|M_j\|_{Op} \le R$ almost surely and let $\sigma^2 = \|\sum_{j=1}^n \mathbb{E}M_j^2\|_{Op}$. Then we have, for all $q \ge 0$,*

$$\text{Prob}\left(\|\sum_{j=1}^n M_j\|_{Op} \ge q\right) \le d\exp(-\frac{q^2/2}{\sigma^2 + Rq/3}) \le \begin{cases} d\exp(-3q^2/8\sigma^2), & q \le \sigma^2/R, \\ d\exp(-3q/8R), & q \ge \sigma^2/R. \end{cases}$$

Proof of Proposition B.4. Without loss of generality, we can assume that $Z = q(zx^\star + xz^\star)$ with $z \in S^{d-1}$ and $0 < q \le 1$.

As in [26], we have with (24) $\|\mathcal{S}\|_{Op} \le 1$, and we can estimate with Proposition B.3

$$\begin{aligned} \left\|\mathcal{P}_{T_x^\perp}\mathcal{S}\mathcal{R}_Z Z\right\|_{Op} &\le \left\|(\mathcal{R}_Z - \mathbb{E}\mathcal{R}_Z)Z\right\|_{Op} + c_0 s^{-t} d^{2-t(1-r)} \\ &\le \left\|(\mathcal{R}_Z - \mathbb{E}\mathcal{R}_Z)Z\right\|_{Op} + c_0 s^{-3} \\ &\le \left\|(\mathcal{R}_Z - \mathbb{E}\mathcal{R}_Z)Z\right\|_{Op} + 1/c_0^2 \\ &\le \left\|(\mathcal{R}_Z - \mathbb{E}\mathcal{R}_Z)Z\right\|_{Op} + A/c_0, \end{aligned}$$

where we have chosen $s = c_0$. As in [26], we obtain in a similar fashion

$$\left\|\mathrm{P}_{T_x}\mathcal{S}(\mathcal{R}_Z - \mathcal{I})Z\right\|_F \le \sqrt{2}\left\|(\mathcal{R}_Z - \mathbb{E}\mathcal{R}_Z)Z\right\|_{Op} + A/c_0,$$

We define the event

$$E := \{\left\|(\mathcal{R}_Z - \mathbb{E}\mathcal{R}_Z)Z\right\|_{Op} \le A - A/c_0\},$$

so that A and B are chosen such that it boils down to bound the probability of E^c. As in [26], we define

$$(\mathcal{R}_Z - \mathbb{E}\mathcal{R}_Z)Z = \sum_{j=1}^n (M_j - \mathbb{E}M_j), \quad \text{where} \quad M_j = \frac{a_1}{n} 1_{E_j} 1_{G_j} \langle Z, \mathcal{P}_j\rangle \mathcal{P}_j$$

and estimate analogously

$$\begin{aligned} \|M_j\|_{Op} &\le \frac{a_1}{n} 1_{E_j} 1_{G_j} |\langle Z, \mathcal{P}_j\rangle| \\ &\le \frac{a_1}{n} 1_{E_j} 1_{G_j} 2|x^\star \mathcal{P}_j z| \\ &\le \frac{a_1}{n} 1_{E_j} 1_{G_j} 2\sqrt{\langle \mathcal{P}_j, xx^\star\rangle \langle \mathcal{P}_j, zz^\star\rangle|} \\ &\le \frac{a_1}{n} 2(s+1)tkd^{-r}, \end{aligned}$$

where we have used the definitions of E_j and G_j. We fix s and knowing that a_1 grows like d^2, we can further derive

$$\|M_j\|_{Op} \le \frac{c_3}{n} t d^{2-r} =: \tilde{R}.$$

Next, we estimate with Corollary 7.4, $\text{trace}(Z) \le \sqrt{2}\|Z\|_F$, $Z \preceq I_d$, $Z^2 \preceq \|Z\|_F I_d$, and $\|Z\|_F \le 1$,

$$\begin{aligned} \mathbb{E}(M_j - \mathbb{E}M_j)^2 &\preceq \mathbb{E}M_j^2 \\ &\preceq \frac{a_1^2}{n^2} \mathbb{E}\langle Z, \mathcal{P}_j\rangle^2 \mathcal{P}_j \\ &\preceq \frac{a_1^2}{n^2}\left[\alpha_1(2\,\text{trace}(Z)Z + \text{trace}(Z^2)I_d) + 2\alpha_2 Z^2 + \alpha_3\,\text{trace}(Z)^2 I_d\right] \\ &\preceq \frac{a_1^2}{n^2}\left[\alpha_1(2\sqrt{2}\|Z\|_F Z + \text{trace}(Z^2)I_d) + 2\alpha_2\|Z\|_F I_d + \alpha_3 2\|Z\|_F^2 I_d\right] \\ &\preceq \frac{a_1^2}{n^2}\left[\alpha_1(2\sqrt{2}I_d + I_d) + 2\alpha_2 I_d + \alpha_3 2 I_d\right] \preceq \frac{cd}{n^2} I_d, \end{aligned}$$

where $c > 0$ is some constant independent of d and we used that a_1 and a_i can be estimated by a constant times d^2 and $1/d^3$, respectively. We can deduce that

$$\Big\| \sum_{j=1}^{n} \mathbb{E}(M_j - \mathbb{E}M_j^2)^2 \Big\|_{Op} \le n \max_{j=1,\dots,n} \|\mathbb{E}M_j^2\|_{Op} \le \frac{cd}{n} =: \sigma^2.$$

Now, we can choose a sufficiently large constant $c_2 \ge 1$ such that the definition $R := c_2\tilde{R}$ yields

$$\frac{\sigma^2}{R} \le \frac{cd^{r-1}}{c_2 c_3 t} \le \tilde{q}A,$$

with some $\tilde{q} < 1$. As in [26], an application of Theorem B.5 with $q = \tilde{q}A$ concludes the proof. □

We have now completed the preparations for the proof of Theorem 6.2:

Proof of Theorem 6.2. We construct the dual certificate in a recursive manner and begin with $Y_0 = 0$. Suppose that Y_i is constructed, then we put $Q_i := xx^* - (Y_i)_{T_x} \in T_x$. We choose n_i subspaces independently and identically distributed according to the cubature V. Let $\mathcal{R}_{Q_{i-1}}$ be the operator defined in (28). We define

$$A := 1/c_0, \qquad B := \sqrt{2}A$$

and check whether (29) and (30) are satisfied. If so, let $\mathcal{R}^{(i)}_{Q_{i-1}} := \mathcal{R}_{Q_{i-1}}$,

$$Y_i := \mathcal{S}\mathcal{R}^{(i)}_{Q_{i-1}}(xx^* - Y_{i-1})_{T_x} + Y_{i-1},$$

and we proceed to step $i+1$. If one of the bounds (29) and (30) does not hold, then we repeat the i-th step with a new batch of n_i subspaces. We denote the probability of having to repeat the i-th step by p_i and the eventual number of repetitions by r_i. For $l := \lceil \log_{1/B}(d) \rceil + 2$, we define $Y := Y_l$. Analogously to [26], we derive

$$\|Y_{T_x} - xx^*\|_F \le \frac{2}{d}A^2 = \frac{2}{c_0^2 d},$$
$$\|Y_{T_x^\perp}\|_{Op} \le \frac{A}{1-\sqrt{2}A} = \frac{c_0}{c_0(c_0-\sqrt{2})} \le \frac{1}{c_0-\sqrt{2}}.$$

In order to estimate the probability that the total number of measurements $\sum_{i=1}^{l} n_i r_i$ exceeds the bound in (17), we first apply Proposition B.4 to obtain

$$p_i \le d\exp(-c_1 \frac{n_i A}{t d^{2-r}}) \le d\exp(-c_1 \frac{n_i}{c_0 t d^{2-r}}).$$

To get the exact point of contact with the proof in [26], we choose

$$n_i = 3\frac{c_0}{c_1} t d^{2-r} \log(d),$$

which yields

$$p_i \le e^{-3} \le 1/20.$$

This is the same bound as in [26]. The remaining part of the proof is based on a concentration bound for binomial random variables and directly follows the lines in [26], so that we omit the details.

□

Acknowledgement: ME and MG are funded by the Vienna Science and Technology Fund (WWTF) through project VRG12-009. FK is partially supported by Mathematisches Forschungsinstitut Oberwolfach (MFO). This research was partially carried out at MFO, supported by FK's Oberwolfach Leibniz Fellowship.

Author contributions

ME and MG conceived the mathematical approach and wrote the manuscript. ME, MG, and FK jointly verified Lemma 7.1.

References

[1] C. Bachoc and M. Ehler, *Tight p-fusion frames*, Appl. Comput. Harmon. Anal. **35** (2013), no. 1, 1–15.

[2] C. Bachoc and M. Ehler, *Signal reconstruction from the magnitude of subspace components*, IEEE Trans. Inform. Theory **61** (2015), no. 7, 1–13.

[3] R. Balan, *Stability of phase retrievable frames*, arXiv:1308.5465v1 (2013).

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

[5] A. S. Bandeira, J. Cahill, D. G. Mixon, and A. A. Nelson, *Saving phase: Injectivity and stability for phase retrieval*, Appl. Comput. Harmon. Anal. **37** (2014), no. 1, 106–125.

[6] F. Barthe, F. Gamboa, L.-V. Lozada-Chang, and A. Rouault, *Generalized Dirichlet distributions on the ball and moments*, Alea **7** (2010), 319–340.

[7] B. Bhatia, *Matrix analysis*, Springer, New York, 1996.

[8] B. G. Bodmann and N. Hammen, *Stable phase retrieval with low-redundancy frames*, arXiv:1302.5487v1 (2013).

[9] A. Bondarenko, D. Radchenko, and M. Viazovska, *Optimal asymptotic bounds for spherical designs*, arXiv:1009.4407v3 (2011).

[10] A. V. Bondarenko, D. V. Radchenko, and M. S. Viazovska, *On optimal asymptotic bounds for spherical designs*, arXiv:1009.4407v1 (2010).

[11] J. Cahill, P. G. Casazza, J. Peterson, and L. Woodland, *Phase retrieval by projections*, arXiv:1305.6226v3 (2013).

[12] E. J. Candès, Y. Eldar, T. Strohmer, and V. Voroninski, *Phase retrieval via matrix completion*, arXiv:1109.0573v2 (2011).

[13] E. J. Candès and X. Li., *Solving quadratic equations via PhaseLift when there are about as many equations as unknowns*, Foundations of Computational Mathematics **14** (2014), 1017–1026.

[14] E. J. Candès, T. Strohmer, and V. Voroninski, *PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming*, Communications on Pure and Applied Mathematics, DOI:10.1002/cpa.21432 **66** (2013), no. 8, 1241–1274.

[15] Y. Chikuse, *Statistics on special manifolds*, Lecture Notes in Statistics, Springer, New York, 2003.

[16] P. de la Harpe and C. Pache, *Cubature formulas, geometrical designs, reproducing kernels, and Markov operators*, Infinite groups: geometric, combinatorial and dynamical aspects (Basel), vol. 248, Birkhäuser, 2005, pp. 219–267.

[17] L. Demanet and P. Hand, *Stable optimizationless recovery from phaseless linear measurements*, J. Fourier Anal. Appl. **20** (2014), 199–221.

[18] M. Ehler, *Random tight frames*, J. Fourier Anal. Appl. **18** (2012), no. 1, 1–20.

[19] M. Ehler and M. Gräf, *Cubatures and designs in unions of Grassmann spaces*, arXiv (2014).

[20] M. Ehler and S. Kunis, *Phase retrieval using time and Fourier magnitude measurements*, 10th International Conference on Sampling Theory and Applications, 2013.

[21] V. Elser and R. P. Millane, *Reconstruction of an object from its symmetry-averaged diffraction pattern*, Acta Crystallographica Section A **64** (2008), no. 2, 273–279.

[22] J. R. Fienup, *Phase retrieval algorithms: a comparison*, Applied Optics **21** (1982), no. 15, 2758–2769.

[23] F. Filbir and H. N. Mhaskar, *A quadrature formula for diffusion polynomials corresponding to a generalized heat kernel*, J. Fourier Anal. Appl. **16** (2010), no. 5, 629–657.

[24] R. W. Gerchberg and W. O. Saxton, *A practical algorithm for the determination of the phase from image and diffraction plane pictures*, Optik **35** (1972), no. 2, 237–246.

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

[26] D. Gross, F. Krahmer, and R. Kueng, *A partial derandomization of PhaseLift using spherical designs*, J. Fourier Anal. Appl. **21** (2015), no. 2, 229–266.

[27] A. T. James, *Distributions of matrix variates and latent roots derived from normal samples*, Annals of Mathematical Statistics **35** (1964), no. 2, 475–501.

[28] A. T. James and A. G. Constantine, *Generalized Jacobi polynomials as spherical functions of the Grassmann manifold*, Proc. London Math. Soc. **29** (1974), no. 3, 174–192.

[29] H. König, *Cubature formulas on spheres*, Adv. Multivar. Approx. Math. Res. **107** (1999), 201–211.

[30] R. Kueng, H. Rauhut, and U. Terstiege, *Low rank matrix recovery from rank one measurements*, arXiv:1410.6913 (2014).

[31] H. N. Mhaskar, F. J. Narcowich, and J. D. Ward, *Spherical Marcinkiewicz-Zygmund inequalities and positive quadrature*, Math. Comp. **70** (2002), 1113–1130.

[32] R. J. Muirhead, *Aspects of multivariate statistical theory*, John Wiley & Sons, New York, 1982.

[33] F. Philipp, *Phase retrieval from 4n-4 measurements: A proof for injectivity*, Proc. Appl. Math. Mech. **14** (2014), no. 833-834.

[34] E. Riegler and G. Tauböck, *Almost lossless analog compression without phase information*, in Proc. IEEE Int. Symp. Inf. Th. (Hong Kong, China), 2015, pp. 1–5.

[35] T. Strohmer and R. W. Heath, *Grassmannian frames with applications to coding and communication*, Appl. Comput. Harmon. Anal. **14** (2003), no. 3, 257–275.

[36] J.A. Tropp, *User-friendly tools for random matrices: An introduction.*, NIPS.

[37] J.A. Tropp,, *User-friendly tail bounds for sums of random matrices*, Journal Foundations of Computational Mathematics **12** (2012), no. 4, 389–434.
[38] I. Waldspurger, A. d'Aspremont, and S. Mallat, *Phase recovery, maxcut and complex semidefinite programming*, arXiv:1206.0102v2 (2012).
[39] T. Wong, *Generalized Dirichlet distribution in Bayesian analysis*, Appl. Math. Comput. **97** (1998), no. 2-3, 165–181.