Analysis of low rank matrix recovery via Mendelson's small ball method

Maryia Kabanava
RWTH Aachen University
Chair for Mathematics C (Analysis)
Pontdriesch 10
52062 Aachen
kabanava@mathc.rwth-aachen.de

Holger Rauhut
RWTH Aachen University
Chair for Mathematics C (Analysis)
Pontdriesch 10
52062 Aachen
rauhut@mathc.rwth-aachen.de

Ulrich Terstiege
RWTH Aachen University
Chair for Mathematics C (Analysis)
Pontdriesch 10
52062 Aachen
terstiege@mathc.rwth-aachen.de

***Abstract*—We study low rank matrix recovery from undersampled measurements via nuclear norm minimization. We aim to recover an $n_1 \times n_2$ matrix X from m measurements (Frobenius inner products) $\langle X, A_j \rangle$, $j = 1 \dots m$. We consider different scenarios of independent random measurement matrices A_j and derive bounds for the minimal number of measurements sufficient to uniformly recover any rank r matrix X with high probability. Our results are stable under passing to only approximately low rank matrices and under noise on the measurements.**

In the first scenario the entries of the A_j are independent mean zero random variables of variance 1 with bounded fourth moments. Then any X of rank at most r is stably recovered from m measurements with high probability provided that $m \geq Cr \max\{n_1, n_2\}$. The second scenario studies the physically important case of rank one measurements. Here, the matrix X to recover is Hermitian of size $n \times n$ and the measurement matrices A_j are of the form $A_j = a_j a_j^*$ for some random vectors a_j. If the a_j are independent standard Gaussian random vectors, then we obtain uniform stable and robust rank-r recovery with high probability provided that $m \geq crn$. Finally we consider the case that the a_j are independently sampled from an (approximate) 4-design. Then we require $m \geq crn \log n$ for uniform stable and robust rank-r recovery.

In all cases, the results are shown via establishing a stable and robust version of the rank null space property. To this end, we employ *Mendelson's small ball method*.

I. Low rank matrix recovery via nuclear norm minimization

Low rank matrix recovery aims at reconstructing an $n_1 \times n_2$ matrix X of small rank from incomplete linear measurements $\mathcal{A}(X)$, where $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ is a linear map [19], [12]. We may write $\mathcal{A}$ as

$$\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m, \quad Z \mapsto \sum_{j=1}^{m} \operatorname{tr}\left(Z A_j^*\right) e_j,$$

where $A_1, \dots, A_m$ are suitable $n_1 \times n_2$ matrices and $e_1, \dots, e_m$ is the standard basis in $\mathbb{R}^m$. If we allow noise, the entire measurement vector is of the form

$$b = \mathcal{A}(X) + \varepsilon, \tag{1}$$

where $\varepsilon \in \mathbb{R}^m$ denotes additive noise. Since the naive approach of rank minimization is NP-hard, several tractable alternatives have been suggested, the most prominent being nuclear norm minimization [6], [19]. Assuming $\|\varepsilon\|_2 \leq \eta$, this convex optimization strategy consists in computing the minimizer of

$$\min_{Z \in \mathbb{R}^{n_1 \times n_2}} \|Z\|_* \quad \text{subject to } \|\mathcal{A}(Z) - b\|_2 \leq \eta, \tag{2}$$

where $\|\cdot\|_*$ denotes the nuclear norm, see below. Efficient algorithms are available for this task.

In certain physical applications, in particular in quantum experiments, the problem arises to recover an Hermitian matrix of low rank. Denoting by $\mathcal{H}_n$ the space of complex Hermitian $n \times n$ matrices, we obtain the following analogue of the above nuclear norm minimization problem. For $X \in \mathcal{H}_n$ and for noisy data

$$b = \mathcal{A}(X) + \varepsilon, \quad \|\varepsilon\|_2 \leq \eta$$

with measurement matrices $A_j \in \mathcal{H}_n$, solve the nuclear norm minimization problem

$$\min_{Z \in \mathcal{H}_n} \|Z\|_* \quad \text{subject to } \|\mathcal{A}(Z) - b\|_2 \leq \eta. \tag{3}$$

In certain situations such as the phase retrieval problem [4], it is natural to additionally restrict to positive semidefinite matrices X and A_j, but we do not go into detail on this aspect.

In the sequel, we assume that the measurement matrices $A_1, \dots, A_m$ are independent samples of a random matrix Φ.

The present paper is a summary of results of [11] (in preparation), the results on rank one measurements are extensions of results of [12].

A. Notation

We denote the Schatten-p-norm of a real or complex matrix by $\|Z\|_p$. Thus

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

where the $\sigma_\ell(Z)$ denote the singular values of Z and tr is the trace. In particular, the nuclear norm is $\|Z\|_* = \|Z\|_1$, the Frobenius norm is $\|Z\|_F = \|Z\|_2$ and the spectral norm $\|Z\|_\infty = \|Z\|_{2\to 2} = \sigma_{\max}(Z)$ is the largest singular value.

978-1-4673-7353-1/15/$31.00 ©2015 IEEE

II. Main results

In this section we present our results for the different scenarios of the form of the random matrix Φ (recalling that the A_j are independent copies of Φ).

A. Measurement matrices with independent entries

Here the random matrix $\Phi = (X_{ij})_{i,j}$ is assumed to have the following properties.

- The X_{ij} are independent and have mean zero
- $\mathbb{E}X_{ij}^2 = 1$ and $\mathbb{E}X_{ij}^4 \leq C_4$ for all i, j and some constant C_4.

Let $n = \max\{n_1, n_2\}$.

Theorem 1. *Let $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ be obained from m independent samples $A_1, \ldots, A_m$ of Φ as above. For $r \leq n_1, n_2$ and $0 < \rho < 1$, suppose that*

$$m \geq c_1 \rho^{-2} nr.$$

Then with probability at least $1 - e^{-c_2 m}$, for any $X \in \mathbb{R}^{n_1 \times n_2}$ any solution $\hat{X}$ of (2) with $b = \mathcal{A}(X) + \varepsilon$, $\|\varepsilon\|_2 \leq \eta$, approximates X with the error

$$\|X - \hat{X}\|_2 \leq \frac{2(1+\rho)^2}{(1-\rho)\sqrt{r}} \sum_{j=r+1}^{n} \sigma_j(X) + \frac{(3+\rho)c_3}{(1-\rho)} \cdot \frac{\eta}{\sqrt{m}}.$$

Here c_1, c_2, c_3 are positive constants that only depend on C_4.

It may be surprising that only fourth moments are required. In the existing literature, see e.g. [19], [3], much stronger assumptions such as Gaussian distributions or subgaussian tails are made. We note, however, that an analogue for recovery of sparse vectors in $\mathbb{R}^N$ (compressive sensing) of the above result has been shown by Lecué and Mendelson, where only $\log(N)$ bounded moments are required, see [14] for details.

B. Rank one measurements

Here we focus on the recovery of Hermitian matrices and we assume that our measurement matrices are of the form $A_j = a_j a_j^*$. We first consider the situation that the a_j are standard Gaussian distributed and then pass to the physically important case of 4-designs. The notions of a design resp. an approximative design here are the same as in [12] and the theorems parallel the results in [12]. More precisely, we extend the result of [12] by showing that the recovery results are also stable with respect to the rank. However, also in [12] Mendelson's small ball method is the crucial ingredient in the proofs. Note that if the matrix X we want to recover is also of the form $X = xx^*$ we are (by the Phase Lift argument [4]) in the situation of phase retrieval.

The driving motivation for our results in the design setup are possible applications to the problem of quantum state tomography, i.e. estimating the density operator of a quantum system. Theorem 7 below allows indeed efficient low rank quantum state tomography for different types of measurements, see [12, section 3] for more information.

1) Gaussian Measurements: The following result extends [12, Theorem 2] to stability of reconstruction under passing from exactly low rank to approximately low rank matrices.

Theorem 2. *Consider the measurement process described in (1) with measurement matrices $A_j = a_j a_j^*$, where $a_1, \ldots, a_m \in \mathbb{C}^n$ are independent standard complex Gaussian random vectors. For $r \leq n$ and $0 < \rho < 1$, suppose that*

$$m \geq C_1 \rho^{-2} nr.$$

Then with probability at least $1 - \mathrm{e}^{-C_2 m}$ it holds that, for any $X \in \mathcal{H}_n$, any solution $\hat{X}$ to the convex optimization problem (3) with noisy measurements $b = \mathcal{A}(X) + \varepsilon$, where $\|\varepsilon\|_{\ell_2} \leq \eta$, obeys

$$\|X - \hat{X}\|_2 \leq \frac{2(1+\rho)^2}{(1-\rho)\sqrt{r}} \sum_{j=r+1}^{n} \sigma_j(X) + \frac{(3+\rho)C_3}{(1-\rho)} \cdot \frac{\eta}{\sqrt{m}}. \tag{4}$$

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

Remark 3. This result generalizes a result by Candès and Li [2] on phase retrieval when $r = 1$. A non-uniform result in that spirit was obtained in [4].

Our proof also applies in the real valued analogue and also in the case of subgaussian vectors a_j at least if we assume that the k-th moments for $k \leq 8$ are as in the Gaussian case. However, it should be noted that a slightly stronger result for the real subgaussian case was obtained in [5] using arguments based on the rank restricted isometry property.

2) 4-Designs: Let us finally consider the situation that the a_j are sampled (independently) from a 4-design, where we obtain significant generalizations of results due to Gross, Krahmer and Kueng [8] on phase retrieval with t-designs. We first recall the definition of an exact weighted t-design, see [20] and [12]. As in [12] we use the following definition by Scott.

Definition 4 (*exact, weighted t-design*, Definition 3 in [20])**.** Let $t \in \mathbb{N}$. A finite set $\{w_1, \ldots, w_N\} \subset \mathbb{C}^n$ of normalized vectors ($\|w_i\|_2 = 1$) with associated weights $\{p_1, \ldots, p_N\}$ such that $p_i \geq 0$ and $\sum_{i=1}^N p_i = 1$ is called a *weighted complex projective t-design* of dimension n and cardinality N if

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

Here $\mathbb{C}P^{n-1}$ denotes the complex projective space of dimension $n-1$ and the integral is computed with respect to the unique unitarily invariant probability measure on $\mathbb{C}P^{n-1}$.

The following result extends [12, Theorem 3] to stability of the recovery, and it strengthens and generalizes the main result of [8] from the rank-one case (phase retrieval) to arbitrary rank.

Theorem 5. *Let $\{p_i, w_i\}_{i=1}^N$ be a weighted 4-design and consider the measurement process described in (1) with measurement matrices $A_j = \sqrt{n(n+1)} a_j a_j^*$, where $a_1, \ldots, a_m \in \mathbb{C}^n$*

are drawn independently from $\{p_i, w_i\}_{i=1}^N$. For $r \leq n$ and $0 < \rho < 1$, suppose that

$$m \geq C_4 \rho^{-2} \log n \; nr.$$

Then with probability at least $1 - \mathrm{e}^{-C_5 m}$ it holds that, for any $X \in \mathcal{H}_n$, any solution $\hat{X}$ to the convex optimization problem (3) with noisy measurements $b = \mathcal{A}(X) + \varepsilon$, where $\|\varepsilon\|_{\ell_2} \leq \eta$, obeys

$$\|X - \hat{X}\|_2 \leq \frac{2(1+\rho)^2}{(1-\rho)\sqrt{r}} \sum_{j=r+1}^{n} \sigma_j(X) + \frac{(3+\rho)C_6}{(1-\rho)} \cdot \frac{\eta}{\sqrt{m}}. \tag{6}$$

Here, C_4, C_5 and C_6 denote absolute positive constants.

We recall the definition of an approximative t-design, see [1] and [12]. Approximative t-designs have the advantage of being easier to construct. Certain of these constructions can be realized (at least in principle) in physical experiments with low (quantum) complexity [1], [21].

Definition 6 (*Approximate t-design*[1], [12])**.** We call a weighted set $\{p_i, w_i\}_{i=1}^N$ of normalized vectors an approximate t-design of p-norm accuracy θ_p, if

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

The next result extends [12, Theorem 5].

Theorem 7. *Let $1 \leq r \leq n$ arbitrary and let $\{p_i, w_i\}_{i=1}^N$ be an approximate 4-design satisfying*

$$\left\| \sum_{i=1}^{N} p_i w_i w_i^* - \frac{1}{n} \mathrm{id} \right\|_\infty \quad \leq \quad \frac{1}{n}, \tag{7}$$

that admits either operator norm accuracy $\theta_\infty \leq 1/(16r^2)$, or trace-norm accuracy $\theta_1 \leq 1/4$, respectively. Then, the recovery guarantee from Theorem 5 still holds (with possibly different constants $\tilde{C}_4, \tilde{C}_5$ and $\tilde{C}_6$).

III. Proofs

To prove our results, we will use Mendelson's small ball method [17], [22], [9], [16] to show that $\mathcal{A}$ fulfills with high probability a robust and stable version of the rank null space property [7], see (8) below. (In particular, we do not work with the restricted isometry property.) Stability and robustness of recovery via nuclear norm minimization (2) (resp. (3)) is then obtained via the following theorem.

Theorem 8. *Let $r \in [n]$, and let ρ, τ be constants with $0 < \rho < 1$ and $\tau > 0$. Let $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ (resp. $\mathcal{A} : \mathcal{H}_n \to \mathbb{R}^m$) satisfy the condition*

$$\left(\sum_{j=1}^{r} \sigma_j(M)^2 \right)^{1/2} \leq \frac{\rho}{\sqrt{r}} \sum_{j=r+1}^{n} \sigma_j(M) + \tau \|\mathcal{A}(M)\|_2 \tag{8}$$

for any $M \in \mathbb{R}^{n_1 \times n_2}$ (resp. for any $M \in \mathcal{H}_n$). Then for any $X \in \mathbb{R}^{n_1 \times n_2}$ (resp. $X \in \mathcal{H}_n$) any solution $\hat{X}$ of (2) (resp. of (3)) with $b = \mathcal{A}(X) + \varepsilon$, $\|\varepsilon\|_2 \leq \eta$, approximates X with error

$$\|X - \hat{X}\|_2 \leq \frac{2(1+\rho)^2}{(1-\rho)\sqrt{r}} \sum_{j=r+1}^{n} \sigma_j(X) + \frac{2\tau(3+\rho)}{1-\rho} \eta.$$

Condition (8) is the robust and stable rank null space property. This theorem is well known, see for example [7, exercise 4.20]. It can easily been shown using the corresponding theorem for vector recovery from compressed sensing together with results from [18] which allow to translate the vector case to the matrix case. (Alternatively, one may proceed directly with tools from matrix analysis). In order to verify the rank null space property (8), we introduce the following notation. Let

$$T_{\rho,r,1} := \Big\{ M \in \mathbb{R}^{n_1 \times n_2} : \left(\sum_{i=1}^{r} \sigma_i(M)^2 \right)^{1/2} > \frac{\rho}{\sqrt{r}} \sum_{i=r+1}^{n} \sigma_i(M), \; \|M\|_F = 1 \Big\}.$$

We define $T_{\rho,r,1}^{\mathcal{H}_n}$ analogous to $T_{\rho,r,1}$ by replacing $\mathbb{R}^{n_1 \times n_2}$ by $\mathcal{H}_n$. It is easy to show (see for example [11]) that (8) is satisfied for any $M \in \mathbb{R}^{n_1 \times n_2}$ (resp. any $M \in \mathcal{H}_n$) if

$$\inf_{M \in T_{\rho,r,1}} \|\mathcal{A}(M)\|_2 \geq \tau^{-1}, \tag{9}$$

resp. in the Hermitian case if

$$\inf_{M \in T_{\rho,r,1}^{\mathcal{H}_n}} \|\mathcal{A}(M)\|_2 \geq \tau^{-1}. \tag{10}$$

A. Applying Mendelson's small ball method

Recall that our measurement matrices $A_1, \ldots, A_m$ are independent samples of a random matrix Φ. We show that with high probability condition (9) resp. (10) is fulfilled so that the recovery statement of Theorem 8 holds. For this we will use the following theorem which is based on ideas due to Mendelson, [17], [9], [22].

Theorem 9. *(Mendelson [16], [17], Koltchinskii, Mendelson [9]; Tropp's version [22]) Fix $E \subset \mathbb{R}^d$ and let $\phi_1, \ldots, \phi_m$ be independent copies of a random vector ϕ in $\mathbb{R}^d$. For $\xi > 0$ let*

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

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

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

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

[1]Recall that a Rademacher vector $\epsilon = (\epsilon_j)_{j=1}^m$ is a vector of independent Rademacher random variables which take the values ± 1 with equal probability.

Applying this theorem and estimating $Q_\xi(E;\phi)$ (using e.g. the Payley-Zygmund inequality) and $W_m(E,\phi)$ is referred to as the *Mendelson's small ball method*, see [22].

We want to apply this as follows. Identify $\mathbb{R}^m$ with $\mathbb{R}^{n_1\times n_2}$ resp. with $\mathcal{H}_n$. Supposing the A_i are independent samples of a random matrix Φ, let $\phi=\Phi$. Finally, let $E=T_{\rho,r,1}$ resp. $E=T^{\mathcal{H}_n}_{\rho,r,1}$. Suppose $\Phi_1,\dots,\Phi_m$ are independent copies of Φ. Then Theorem 9 yields an estimate for $\inf_{M\in T_{\rho,r,1}}\|\mathcal{A}(M)\|_2$ (resp. of $\inf_{M\in T^{\mathcal{H}_n}_{\rho,r,1}}\|\mathcal{A}(M)\|_2$) in terms of

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

and of

$$W_m(E,\Phi)=\mathbb{E}\sup_{Y\in E}\langle H,Y\rangle,\qquad\text{where } H=\frac{1}{\sqrt{m}}\sum_{j=1}^m\varepsilon_j\Phi_j.$$

It remains to estimate $Q_\xi(E;\Phi)$ and $W_m(E,\Phi)$ in the particular cases. In order to estimate $W_m(E,\Phi)$, we use the following lemma.

Lemma 10. *Let E be either $T_{\rho,r,1}$ or $T^{\mathcal{H}_n}_{\rho,r,1}$. Then*

$$W_m(E,\Phi)\leq\sqrt{1+(1+\rho^{-1})^2}\sqrt{r}\cdot\mathbb{E}\|H\|_\infty.$$

Proof. Let D_r be the convex hull of all matrices in $\mathbb{R}^{n_1\times n_2}$ resp. $\mathcal{H}_n$ of Frobenius norm 1 and rank at most r. Then arguing similarly as in [10] (see also [11]), we obtain

$$E\subseteq\sqrt{1+(1+\rho^{-1})^2}D_r. \tag{11}$$

Suppose now that Y has Frobenius norm 1 and rank at most r and let B be any $n_1\times n_2$ (resp. complex $n\times n$) matrix. Then

$$\langle B,Y\rangle\leq\|Y\|_*\|B\|_\infty\leq\sqrt{r}\|B\|_\infty.$$

By convexity, the same estimate holds for any $Y\in D_r$. Combining this with (11), the claim follows. ☐

Hence, in order to estimate $W_m(E,\Phi)$ it is enough to bound $\mathbb{E}\|H\|_\infty$.

B. Proof of Theorem 1

By the discussion in the last paragraph, we only need to find suitable bounds for $Q_{2\xi}(E;\Phi)$ from below and for $\mathbb{E}\|H\|_\infty$ from above.

Lemma 11. *The quantity $Q_{\frac{1}{\sqrt{2}}}(E;\Phi)$ (where $E=T_{\rho,r,1}$) can be estimated as*

$$Q_{\frac{1}{\sqrt{2}}}(E;\Phi)\geq\inf_{\{Y,\|Y\|_2=1\}}\mathbb{P}(|\langle\Phi,Y\rangle|\geq\frac{1}{\sqrt{2}})\geq\frac{1}{4C_5},$$

where $C_5=\max\{3,C_4\}$.

Proof. ([11]) Suppose we are given Y with $\|Y\|_2=1$. By the Payley-Zygmund inequality (see e.g. [7, Lemma 7.16], comp. also [22]),

$$\mathbb{P}\{|\langle\Phi,Y\rangle|^2\geq\frac{1}{2}(\mathbb{E}|\langle\Phi,Y\rangle|^2)\}\geq\frac{1}{4}\cdot\frac{(\mathbb{E}|\langle\Phi,Y\rangle|^2)^2}{\mathbb{E}|\langle\Phi,Y\rangle|^4}. \tag{12}$$

Now

$$\begin{aligned}\mathbb{E}|\langle\Phi,Y\rangle|^2&=\sum_{i,j,k,l}\mathbb{E}(X_{ij}X_{kl})\cdot Y_{ij}Y_{kl}\\&=\sum_{i,j}\mathbb{E}X_{ij}^2\cdot Y_{ij}^2=\sum_{i,j}Y_{ij}^2=1.\end{aligned}$$

Similarly one checks that

$$\mathbb{E}|\langle\Phi,Y\rangle|^4\leq C_5$$

Combining this with $\mathbb{E}|\langle\Phi,Y\rangle|^2=1$ and the estimate (12), the claim follows. ☐

Lemma 12. *Let $\Phi_1,\dots,\Phi_m$ be independent copies of a random matrix Φ as above. Let $\varepsilon_1,\dots,\varepsilon_m$ be Rademacher variables independent of everything else and let $H=\frac{1}{\sqrt{m}}\sum_{k=1}^m\varepsilon_k\Phi_k$. Then*

$$\mathbb{E}\|H\|_\infty\leq C_1\sqrt{n}.$$

Here C_1 is a constant that only depends on C_4.

Proof. ([11]) Let $S=\sum_{k=1}^m\Phi_k$. Since the entries of the Φ_k all have mean zero we may desymmetrize the sum H (see [15, Lemma 6.3]) to obtain

$$\mathbb{E}\|H\|_\infty\leq\frac{2}{\sqrt{m}}\mathbb{E}\|S\|_\infty.$$

Thus it is enough to show that $\mathbb{E}\|S\|_\infty\leq c_3\sqrt{mn}$ for a suitable constant c_3. Since S is a matrix with independent mean zero entries, a result of Latała (see [13]) tells us that, for some universal constant C_2,

$$\mathbb{E}\|S\|_\infty\leq C_2\left(\max_i\sqrt{\sum_j\mathbb{E}S_{ij}^2}+\max_j\sqrt{\sum_i\mathbb{E}S_{ij}^2}+\sqrt[4]{\sum_{i,j}\mathbb{E}S_{ij}^4}\right).$$

Denote the entries of Φ_k by $X_{k;ij}$. Then $S_{ij}=\sum_k X_{k;ij}$. Thus $\mathbb{E}S_{ij}^2=\mathbb{E}(\sum_k X_{k;ij})^2=\sum_k\mathbb{E}X_{k;ij}^2=m$, where we used the independence of the $X_{k;ij}$. Hence $\sqrt{\sum_j\mathbb{E}S_{ij}^2}\leq\sqrt{nm}$ for any i and $\sqrt{\sum_i\mathbb{E}S_{ij}^2}\leq\sqrt{nm}$ for any j. It remains to estimate $\sqrt[4]{\sum_{i,j}\mathbb{E}S_{ij}^4}$. We calculate $\mathbb{E}S_{ij}^4=\mathbb{E}(\sum_k X_{k;ij})^4$. Again since the $X_{k;ij}$ are independent and have mean zero we obtain

$$\mathbb{E}S_{ij}^4=\sum_k\mathbb{E}X_{k;ij}^4+3\sum_{k_1\neq k_2}\mathbb{E}X_{k_1;ij}^2\mathbb{E}X_{k_2;ij}^2.$$

Since $\mathbb{E}X_{k;ij}^2=1$ for all i,j,k, we obtain $\mathbb{E}S_{ij}^4\leq C_5m^2$, where $C_5=\max\{3,C_4\}$. Hence,

$$\sqrt[4]{\sum_{i,j}\mathbb{E}S_{ij}^4}\leq\sqrt[4]{C_5m^2n^2}=\sqrt[4]{C_5}\sqrt{mn}$$

and consequently

$$\mathbb{E}\|S\|_\infty\leq c_3\sqrt{mn}$$

for a suitable constant c_3 that depends only on C_4. ☐

Now we can finish the proof of Theorem 1. Let $H = \frac{1}{\sqrt{m}}\sum_{j=1}^m \varepsilon_j \Phi_j$ and let $\xi = \frac{1}{2\sqrt{2}}$ and $E = T_{\rho,r,1}$. Then it follows from Theorem 9 that for any $t \geq 0$ with probability at least $1 - e^{-2t^2}$

$$\inf_{Y \in T} \left(\sum_{i=1}^m |\langle \Phi_i, Y \rangle|^2 \right)^{1/2} \geq \frac{\sqrt{m}}{2\sqrt{2}} Q_{\frac{1}{\sqrt{2}}}(T;\Phi) \tag{13}$$

$$- 2W_m(T,\Phi) - \frac{1}{2\sqrt{2}}t. \tag{14}$$

By Lemma 11,

$$Q_{\frac{1}{\sqrt{2}}}(T;\Phi) \geq \frac{1}{4C_5}. \tag{15}$$

Combining now Lemma 10 and Lemma 12, we obtain

$$W_m(T,\Phi) \leq C_1 \sqrt{1 + (1+\rho^{-1})^2}\sqrt{r}\sqrt{n}. \tag{16}$$

Combining (13), (15) and (16) we see that choosing $m \geq c_1'(1+(1+\rho^{-1})^2)nr \asymp c_1 \rho^{-2} nr$ and $t = c_4\sqrt{m}$ for suitable constants c_1, c_4, we obtain with probability at least $1 - e^{-c_2 m}$

$$\inf_{Y \in T} \left(\sum_{i=1}^m |\langle \Phi_i, Y \rangle|^2 \right)^{1/2} \geq c_3^{-1}\sqrt{m}$$

for suitable constants c_2, c_3. Now the claim follows from Theorem 8.

IV. Proofs of Theorems 2, 5 and 7

These Theorems are proved along the lines of the proof of Theorem 1 if we can bound $Q_\xi(E, aa^*)$ (where $E = T_{\rho,r,1}^{\mathcal{H}_n}$) suitably from below and $\mathbb{E}\|H\|_\infty$ suitably from above. This was in all cases already done in [12], so we cite the results.

A. Bounds for Theorem 2

Here $Q_\xi(E, aa^*)$ and $\mathbb{E}\|H\|_\infty$ can be bounded as follows.

Lemma 13 (see [12])**.** *Assume that the a_i are indepndent and Gaussian. Then*

$$Q_{\frac{1}{\sqrt{2}}}(E;\phi) := \inf_{u \in E} \mathbb{P}\{|\langle aa^*, u\rangle| \geq \frac{1}{\sqrt{2}}\} \geq \frac{1}{96}.$$

Similarly to the situation in Theorem 1, we also have $\mathbb{E}\|H\|_\infty \leq c\sqrt{n}$ if $m \geq \tilde{c}n$ for suitable constants $c, \tilde{c}$, see also [22, Section 8] and [12].

B. Bounds for Theorem 5 and 7

Recall from [12] that a super-normalized weighted 4-design is obtained from a weighted 4-design by multiplication with $\sqrt[4]{n(n+1)}$.

Proposition 14. *[12, Proposition 12] Assume that a is drawn uniformly from a super-normalized weighted 4-design. Then*

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

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

Proposition 15. *[12, Proposition 13] Let $H = \frac{1}{\sqrt{m}}\sum_{k=1}^m \varepsilon_k a_k a_k^*$, where the a_j's are chosen independently at random from a super-normalized weighted 1-design. Then it holds that*

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

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

The bounds for Theorem 7 are similar, see [12, Section 4.5].

Acknowledgment

The authors acknowledge funding from the ERC Starting Grant StG258926.

References

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

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

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

[4] E. J. Candès, T. Strohmer and V. Voroninski. *PhaseLift. Exact and stable signal recovery from magnitude measurements via convex programming* Comm. Pure Appl. Math., 66(8): 1241–1274, 2013.

[5] Y. Chen, Y. Chi, and A. Goldsmith. *Exact and stable covariance estimation from quadratic sampling via convex programming* IEEE Trans. on Information Theory, to appear.

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

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

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

[9] V. Koltchinskii and S. Mendelson. *Bounding the smallest singular value of a random matrix without concentration.* International Mathematics Research Notices, to appear.

[10] M. Kabanava, and H. Rauhut. *Analysis ℓ_1-recovery with frames and Gaussian measurements.* Acta Appl. Math., to appear. doi:10.1007/s10440-014-9984-y.

[11] M. Kabanava, R. Kueng, H. Rauhut, and U. Terstiege. *Low rank matrix recovery.* In preparation.

[12] R. Kueng, H. Rauhut, and U. Terstiege. Low rank matrix recovery from rank one measurements. *ArXiv:1410.6913*.

[13] R. Latała. *Some Estimates of Norms of Random Matrices* Proceedings of the American Mathematical Society Vol. 133, No. 5 (May, 2005), pp. 1273-1282.

[14] G. Lecué and S. Mendelson. *Sparse recovery under weak moment assumptions.* Journal of the European Mathematical Society, to appear.

[15] M. Ledoux and M. Talagrand. *Probability in Banach spaces:* Isoperimetry and Processes. Springer, Berlin, 1991.

[16] S. Mendelson. *A remark on the diameter of random sections of convex bodies.* Geometric Aspects of Functional Analysis (GAFA Seminar Notes, B. Klartag and E. Milman Eds.), Lecture notes in Mathematics 2116, 395-404, 2014.

[17] S. Mendelson. *Learning without Concentration.* Journal of the ACM, to appear.

[18] S. Oymak, K. Mohan, M. Fazel, and B. Hassibi. *A simplified approach to recovery conditions for low rank matrices Proc. Intl. Sympo. Information Theory (ISIT), Aug 2011.*

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

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

[21] Y. Nakata, P. Turner, M. Mio. *Phase-random states: Ensembles of states with fixed amplitudes and uniformly distributed phases in a fixed basis* Phys. Rev. A. 86:012301, 2012.

[22] J. A. Tropp. *Convex recovery of a structured signal from independent random linear measurements.* Sampling Theory, a Renaissance, to appear.