
Universal low-rank matrix recovery from Pauli measurements

Yi-Kai Liu

Applied and Computational Mathematics Division
National Institute of Standards and Technology
Gaithersburg, MD, USA
yi-kai.liu@nist.gov

Abstract

We study the problem of reconstructing an unknown matrix M of rank r and dimension d using $O(rd \text{ poly } \log d)$ Pauli measurements. This has applications in quantum state tomography, and is a non-commutative analogue of a well-known problem in compressed sensing: recovering a sparse vector from a few of its Fourier coefficients.

We show that almost all sets of $O(rd \log^6 d)$ Pauli measurements satisfy the rank- r restricted isometry property (RIP). This implies that M can be recovered from a fixed (“universal”) set of Pauli measurements, using nuclear-norm minimization (e.g., the matrix Lasso), with nearly-optimal bounds on the error. A similar result holds for any class of measurements that use an orthonormal operator basis whose elements have small operator norm. Our proof uses Dudley’s inequality for Gaussian processes, together with bounds on covering numbers obtained via entropy duality.

1 Introduction

Low-rank matrix recovery is the following problem: let M be some unknown matrix of dimension d and rank $r \ll d$, and let A_1, A_2, \dots, A_m be a set of measurement matrices; then can one reconstruct M from its inner products $\text{tr}(M^* A_1), \text{tr}(M^* A_2), \dots, \text{tr}(M^* A_m)$? This problem has many applications in machine learning [1, 2], e.g., collaborative filtering (the Netflix problem). Remarkably, it turns out that for many useful choices of measurement matrices, low-rank matrix recovery is possible, and can even be done efficiently. For example, when the A_i are Gaussian random matrices, then it is known that $m = O(rd)$ measurements are sufficient to uniquely determine M , and furthermore, M can be reconstructed by solving a convex program (minimizing the nuclear norm) [3, 4, 5]. Another example is the “matrix completion” problem, where the measurements return a random subset of matrix elements of M ; in this case, $m = O(rd \text{ poly } \log d)$ measurements suffice, provided that M satisfies some “incoherence” conditions [6, 7, 8, 9, 10].

The focus of this paper is on a different class of measurements, known as Pauli measurements. Here, the A_i are randomly chosen elements of the Pauli basis, a particular orthonormal basis of $\mathbb{C}^{d \times d}$. The Pauli basis is a non-commutative analogue of the Fourier basis in \mathbb{C}^d ; thus, low-rank matrix recovery using Pauli measurements can be viewed as a generalization of the idea of compressed sensing of sparse vectors using their Fourier coefficients [11, 12]. In addition, this problem has applications in quantum state tomography, the task of learning an unknown quantum state by performing measurements [13]. This is because most quantum states of physical interest are accurately described by density matrices that have low rank; and Pauli measurements are especially easy to carry out in an experiment (due to the tensor product structure of the Pauli basis).

In this paper we show stronger results on low-rank matrix recovery from Pauli measurements. Previously [13, 8], it was known that, for every rank- r matrix $M \in \mathbb{C}^{d \times d}$, almost all choices of $m = O(rd \text{ poly } \log d)$ random Pauli measurements will lead to successful recovery of M . Here we show a stronger statement: there is a fixed (“universal”) set of $m = O(rd \text{ poly } \log d)$ Pauli measurements, such that for all rank- r matrices $M \in \mathbb{C}^{d \times d}$, we have successful recovery.¹ We do this by showing that the random Pauli sampling operator obeys the “restricted isometry property” (RIP). Intuitively, RIP says that the sampling operator is an approximate isometry, acting on the set of all low-rank matrices. In geometric terms, it says that the sampling operator embeds the manifold of low-rank matrices into $O(rd \text{ poly } \log d)$ dimensions, with low distortion in the 2-norm.

RIP for low-rank matrices is a very strong property, and prior to this work, it was only known to hold for very unstructured types of random measurements, such as Gaussian measurements [3], which are unsuitable for most applications. RIP was known to fail in the matrix completion case, and whether it held for Pauli measurements was an open question. Once we have established RIP for Pauli measurements, we can use known results [3, 4, 5] to show low-rank matrix recovery from a universal set of Pauli measurements. In particular, using [5], we can get nearly-optimal universal bounds on the error of the reconstructed density matrix, when the data are noisy; and we can even get bounds on the recovery of arbitrary (not necessarily low-rank) matrices. These RIP-based bounds are qualitatively stronger than those obtained using “dual certificates” [14] (though the latter technique is applicable in some situations where RIP fails).

In the context of quantum state tomography, this implies that, given a quantum state that consists of a low-rank component M_r plus a residual full-rank component M_c , we can reconstruct M_r up to an error that is not much larger than M_c . In particular, let $\|\cdot\|_*$ denote the nuclear norm, and let $\|\cdot\|_F$ denote the Frobenius norm. Then the error can be bounded in the nuclear norm by $O(\|M_c\|_*)$ (assuming noiseless data), and it can be bounded in the Frobenius norm by $O(\|M_c\|_F \text{ poly } \log d)$ (which holds even with noisy data²). This shows that our reconstruction is nearly as good as the best rank- r approximation to M (which is given by the truncated SVD). In addition, a completely arbitrary quantum state can be reconstructed up to an error of $O(1/\sqrt{r})$ in Frobenius norm. Lastly, the RIP gives some insight into the optimal design of tomography experiments, in particular, the tradeoff between the number of measurement settings (which is essentially m), and the number of repetitions of the experiment at each setting (which determines the statistical noise that enters the data) [15].

These results can be generalized beyond the class of Pauli measurements. Essentially, one can replace the Pauli basis with any orthonormal basis of $\mathbb{C}^{d \times d}$ that is *incoherent*, i.e., whose elements have small operator norm (of order $O(1/\sqrt{d})$, say); a similar generalization was noted in the earlier results of [8]. Also, our proof shows that the RIP actually holds in a slightly stronger sense: it holds not just for all rank- r matrices, but for all matrices X that satisfy $\|X\|_* \leq \sqrt{r}\|X\|_F$.

To prove this result, we combine a number of techniques that have appeared elsewhere. RIP results were previously known for Gaussian measurements and some of their close relatives [3]. Also, restricted strong convexity (RSC), a similar but somewhat weaker property, was recently shown in the context of the matrix completion problem (with additional “non-spikiness” conditions) [10]. These results follow from covering arguments (i.e., using a concentration inequality to upper-bound the failure probability on each individual low-rank matrix X , and then taking the union bound over all such X). Showing RIP for Pauli measurements seems to be more delicate, however. Pauli measurements have more structure and less randomness, so the concentration of measure phenomena are weaker, and the union bound no longer gives the desired result.

Instead, one must take into account the favorable correlations between the behavior of the sampling operator on different matrices — intuitively, if two low-rank matrices M and M' have overlapping supports, then good behavior on M is positively correlated with good behavior on M' . This can be done by transforming the problem into a Gaussian process, and using Dudley’s entropy bound. This is the same approach used in classical compressed sensing, to show RIP for Fourier measurements [12, 11]. The key difference is that in our case, the Gaussian process is indexed by low-rank matrices, rather than sparse vectors. To bound the correlations in this process, one then needs to bound the covering numbers of the nuclear norm ball (of matrices), rather than the ℓ_1 ball (of vectors). This

¹Note that in the universal result, m is slightly larger, by a factor of $\text{poly } \log d$.

²However, this bound is not universal.

requires a different technique, using entropy duality, which is due to Guédon et al [16]. (See also the related work in [17].)

As a side note, we remark that matrix recovery can sometimes fail because there exist large sets of up to d Pauli matrices that all commute, i.e., they have a simultaneous eigenbasis ϕ_1, \dots, ϕ_d . (These ϕ_i are of interest in quantum information — they are called stabilizer states [18].) If one were to measure such a set of Pauli's, one would gain complete knowledge about the diagonal elements of the unknown matrix M in the ϕ_i basis, but one would learn nothing about the off-diagonal elements. This is reminiscent of the difficulties that arise in matrix completion. However, in our case, these pathological cases turn out to be rare, since it is unlikely that a random subset of Pauli matrices will all commute.

Finally, we note that there is a large body of related work on estimating a low-rank matrix by solving a regularized convex program; see, e.g., [19, 20].

This paper is organized as follows. In section 2, we state our results precisely, and discuss some specific applications to quantum state tomography. In section 3 we prove the RIP for Pauli matrices, and in section 4 we discuss some directions for future work. Some technical details appear in sections A and B in the supplementary material [21].

Notation: For vectors, $\|\cdot\|_2$ denotes the ℓ_2 norm. For matrices, $\|\cdot\|_p$ denotes the Schatten p -norm, $\|X\|_p = (\sum_i \sigma_i(X)^p)^{1/p}$, where $\sigma_i(X)$ are the singular values of X . In particular, $\|\cdot\|_* = \|\cdot\|_1$ is the trace or nuclear norm, $\|\cdot\|_F = \|\cdot\|_2$ is the Frobenius norm, and $\|\cdot\| = \|\cdot\|_\infty$ is the operator norm. Finally, for matrices, A^* is the adjoint of A , and (\cdot, \cdot) is the Hilbert-Schmidt inner product, $(A, B) = \text{tr}(A^*B)$. Calligraphic letters denote superoperators acting on matrices. Also, $|A)(A|$ is the superoperator that maps every matrix $X \in \mathbb{C}^{d \times d}$ to the matrix $A \text{tr}(A^*X)$.

2 Our Results

We will consider the following approach to low-rank matrix recovery. Let $M \in \mathbb{C}^{d \times d}$ be an unknown matrix of rank at most r . Let W_1, \dots, W_{d^2} be an orthonormal basis for $\mathbb{C}^{d \times d}$, with respect to the inner product $(A, B) = \text{tr}(A^*B)$. We choose m basis elements, S_1, \dots, S_m , iid uniformly at random from $\{W_1, \dots, W_{d^2}\}$ (“sampling with replacement”). We then observe the coefficients (S_i, M) . From this data, we want to reconstruct M .

For this to be possible, the measurement matrices W_i must be “incoherent” with respect to M . Roughly speaking, this means that the inner products (W_i, M) must be small. Formally, we say that the basis W_1, \dots, W_{d^2} is *incoherent* if the W_i all have small operator norm,

$$\|W_i\| \leq K/\sqrt{d}, \quad (1)$$

where K is a constant.³ (This assumption was also used in [8].)

Before proceeding further, let us sketch the connection between this problem and quantum state tomography. Consider a system of n qubits, with Hilbert space dimension $d = 2^n$. We want to learn the state of the system, which is described by a density matrix $\rho \in \mathbb{C}^{d \times d}$; ρ is positive semidefinite, has trace 1, and has rank $r \ll d$ when the state is nearly pure. There is a class of convenient (and experimentally feasible) measurements, which are described by Pauli matrices (also called Pauli observables). These are matrices of the form $P_1 \otimes \dots \otimes P_n$, where \otimes denotes the tensor product (Kronecker product), and each P_i is a 2×2 matrix chosen from the following four possibilities:

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (2)$$

One can estimate expectation values of Pauli observables, which are given by $(\rho, (P_1 \otimes \dots \otimes P_n))$. This is a special case of the above measurement model, where the measurement matrices W_i are the (scaled) Pauli observables $(P_1 \otimes \dots \otimes P_n)/\sqrt{d}$, and they are incoherent with $\|W_i\| \leq K/\sqrt{d}$, $K = 1$.

³Note that $\|W_i\|$ is the maximum inner product between W_i and any rank-1 matrix M (normalized so that $\|M\|_F = 1$).

Now we return to our discussion of the general problem. We choose S_1, \dots, S_m iid uniformly at random from $\{W_1, \dots, W_{d^2}\}$, and we define the *sampling operator* $\mathcal{A} : \mathbb{C}^{d \times d} \rightarrow \mathbb{C}^m$ as

$$(\mathcal{A}(X))_i = \frac{d}{\sqrt{m}} \operatorname{tr}(S_i^* X), \quad i = 1, \dots, m. \quad (3)$$

The normalization is chosen so that $\mathbb{E}\mathcal{A}^* \mathcal{A} = \mathcal{I}$. (Note that $\mathcal{A}^* \mathcal{A} = \sum_{j=1}^m |S_j\rangle\langle S_j| \cdot \frac{d^2}{m}$.)

We assume we are given the data $y = \mathcal{A}(M) + z$, where $z \in \mathbb{C}^m$ is some (unknown) noise contribution. We will construct an estimator \hat{M} by minimizing the nuclear norm, subject to the constraints specified by y . (Note that one can view the nuclear norm as a convex relaxation of the rank function — thus these estimators can be computed efficiently.) One approach is the matrix Dantzig selector:

$$\hat{M} = \arg \min_X \|X\|_* \text{ such that } \|\mathcal{A}^*(y - \mathcal{A}(X))\| \leq \lambda. \quad (4)$$

Alternatively, one can solve a regularized least-squares problem, also called the matrix Lasso:

$$\hat{M} = \arg \min_X \frac{1}{2} \|\mathcal{A}(X) - y\|_2^2 + \mu \|X\|_*. \quad (5)$$

Here, the parameters λ and μ are set according to the strength of the noise component z (we will discuss this later). We will be interested in bounding the error of these estimators. To do this, we will show that the sampling operator \mathcal{A} satisfies the restricted isometry property (RIP).

2.1 RIP for Pauli Measurements

Fix some constant $0 \leq \delta < 1$. Fix d , and some set $U \subset \mathbb{C}^{d \times d}$. We say that \mathcal{A} satisfies the *restricted isometry property* (RIP) over U if, for all $X \in U$, we have

$$(1 - \delta) \|X\|_F \leq \|\mathcal{A}(X)\|_2 \leq (1 + \delta) \|X\|_F. \quad (6)$$

(Here, $\|\mathcal{A}(X)\|_2$ denotes the ℓ_2 norm of a vector, while $\|X\|_F$ denotes the Frobenius norm of a matrix.) When U is the set of all $X \in \mathbb{C}^{d \times d}$ with rank r , this is precisely the notion of RIP studied in [3, 5]. We will show that Pauli measurements satisfy the RIP over a slightly larger set (the set of all $X \in \mathbb{C}^{d \times d}$ such that $\|X\|_* \leq \sqrt{r} \|X\|_F$), provided the number of measurements m is at least $\Omega(rd \text{ poly } \log d)$. This result generalizes to measurements in any basis with small operator norm.

Theorem 2.1 *Fix some constant $0 \leq \delta < 1$. Let $\{W_1, \dots, W_{d^2}\}$ be an orthonormal basis for $\mathbb{C}^{d \times d}$ that is incoherent in the sense of (1). Let $m = CK^2 \cdot rd \log^6 d$, for some constant C that depends only on δ , $C = O(1/\delta^2)$. Let \mathcal{A} be defined as in (3). Then, with high probability (over the choice of S_1, \dots, S_m), \mathcal{A} satisfies the RIP over the set of all $X \in \mathbb{C}^{d \times d}$ such that $\|X\|_* \leq \sqrt{r} \|X\|_F$. Furthermore, the failure probability is exponentially small in $\delta^2 C$.*

We will prove this theorem in section 3. In the remainder of this section, we discuss its applications to low-rank matrix recovery, and quantum state tomography in particular.

2.2 Applications

By combining Theorem 2.1 with previous results [3, 4, 5], we immediately obtain bounds on the accuracy of the matrix Dantzig selector (4) and the matrix Lasso (5). In particular, for the first time we can show *universal* recovery of low-rank matrices via Pauli measurements, and near-optimal bounds on the accuracy of the reconstruction when the data is noisy [5]. (Similar results hold for measurements in any incoherent operator basis.) These RIP-based results improve on the earlier results based on dual certificates [13, 8, 14]. See [3, 4, 5] for details.

Here, we will sketch a couple of these results that are of particular interest for quantum state tomography. Here, M is the density matrix describing the state of a quantum mechanical object, and $\mathcal{A}(M)$ is a vector of Pauli expectation values for the state M . (M has some additional properties: it is positive semidefinite, and has trace 1; thus $\mathcal{A}(M)$ is a real vector.) There are two main issues that arise. First, M is not precisely low-rank. In many situations, the *ideal* state has low rank (for instance, a pure state has rank 1); however, for the *actual* state observed in an experiment, the density matrix M is full-rank with decaying eigenvalues. Typically, we will be interested in obtaining a good low-rank approximation to M , ignoring the tail of the spectrum.

Secondly, the measurements of $\mathcal{A}(M)$ are inherently noisy. We do not observe $\mathcal{A}(M)$ directly; rather, we estimate each entry $(\mathcal{A}(M))_i$ by preparing many copies of the state M , measuring the Pauli observable S_i on each copy, and averaging the results. Thus, we observe $y_i = (\mathcal{A}(M))_i + z_i$, where z_i is binomially distributed. When the number of experiments being averaged is large, z_i can be approximated by Gaussian noise. We will be interested in getting an estimate of M that is stable with respect to this noise. (We remark that one can also reduce the statistical noise by performing more repetitions of each experiment. This suggests the possibility of a tradeoff between the accuracy of estimating each parameter, and the number of parameters one chooses to measure overall. This will be discussed elsewhere [15].)

We would like to reconstruct M up to a small error in the nuclear or Frobenius norm. Let \hat{M} be our estimate. Bounding the error in nuclear norm implies that, for any measurement allowed by quantum mechanics, the probability of distinguishing the state \hat{M} from M is small. Bounding the error in Frobenius norm implies that the difference $\hat{M} - M$ is highly “mixed” (and thus does not contribute to the coherent or “quantum” behavior of the system).

We now sketch a few results from [4, 5] that apply to this situation. Write $M = M_r + M_c$, where M_r is a rank- r approximation to M , corresponding to the r largest singular values of M , and M_c is the residual part of M (the “tail” of M). Ideally, our goal is to estimate M up to an error that is not much larger than M_c . First, we can bound the error in nuclear norm (assuming the data has no noise):

Proposition 2.2 (Theorem 5 from [4]) *Let $\mathcal{A} : \mathbb{C}^{d \times d} \rightarrow \mathbb{C}^m$ be the random Pauli sampling operator, with $m = Crd \log^6 d$, for some absolute constant C . Then, with high probability over the choice of \mathcal{A} , the following holds:*

Let M be any matrix in $\mathbb{C}^{d \times d}$, and write $M = M_r + M_c$, as described above. Say we observe $y = \mathcal{A}(M)$, with no noise. Let \hat{M} be the Dantzig selector (4) with $\lambda = 0$. Then

$$\|\hat{M} - M\|_* \leq C'_0 \|M_c\|_*, \quad (7)$$

where C'_0 is an absolute constant.

We can also bound the error in Frobenius norm, allowing for noisy data:

Proposition 2.3 (Lemma 3.2 from [5]) *Assume the same set-up as above, but say we observe $y = \mathcal{A}(M) + z$, where $z \sim N(0, \sigma^2 I)$. Let \hat{M} be the Dantzig selector (4) with $\lambda = 8\sqrt{d}\sigma$, or the Lasso (5) with $\mu = 16\sqrt{d}\sigma$. Then, with high probability over the noise z ,*

$$\|\hat{M} - M\|_F \leq C_0 \sqrt{rd}\sigma + C_1 \|M_c\|_* / \sqrt{r}, \quad (8)$$

where C_0 and C_1 are absolute constants.

This bounds the error of \hat{M} in terms of the noise strength σ and the size of the tail M_c . It is universal: one sampling operator \mathcal{A} works for all matrices M . While this bound may seem unnatural because it mixes different norms, it can be quite useful. When M actually is low-rank (with rank r), then $M_c = 0$, and the bound (8) becomes particularly simple. The dependence on the noise strength σ is known to be nearly minimax-optimal [5]. Furthermore, when some of the singular values of M fall below the “noise level” $\sqrt{d}\sigma$, one can show a tighter bound, with a nearly-optimal bias-variance tradeoff; see Theorem 2.7 in [5] for details.

On the other hand, when M is full-rank, then the error of \hat{M} depends on the behavior of the tail M_c . We will consider a couple of cases. First, suppose we do not assume anything about M , besides the fact that it is a density matrix for a quantum state. Then $\|M\|_* = 1$, hence $\|M_c\|_* \leq 1 - \frac{r}{d}$, and we can use (8) to get $\|\hat{M} - M\|_F \leq C_0 \sqrt{rd}\sigma + \frac{C_1}{\sqrt{r}}$. Thus, even for *arbitrary* (not necessarily low-rank) quantum states, the estimator \hat{M} gives nontrivial results. The $O(1/\sqrt{r})$ term can be interpreted as the penalty for only measuring an incomplete subset of the Pauli observables.

Finally, consider the case where M is full-rank, but we do know that the tail M_c is small. If we know that M_c is small in nuclear norm, then we can use equation (8). However, if we know that M_c is small in Frobenius norm, one can give a different bound, using ideas from [5], as follows.

Proposition 2.4 *Let M be any matrix in $\mathbb{C}^{d \times d}$, with singular values $\sigma_1(M) \geq \dots \geq \sigma_d(M)$.*

Choose a random Pauli sampling operator $\mathcal{A} : \mathbb{C}^{d \times d} \rightarrow \mathbb{C}^m$, with $m = Crd \log^6 d$, for some absolute constant C . Say we observe $y = \mathcal{A}(M) + z$, where $z \sim N(0, \sigma^2 I)$. Let \hat{M} be the Dantzig selector (4) with $\lambda = 16\sqrt{d}\sigma$, or the Lasso (5) with $\mu = 32\sqrt{d}\sigma$. Then, with high probability over the choice of \mathcal{A} and the noise z ,

$$\|\hat{M} - M\|_F^2 \leq C_0 \sum_{i=1}^r \min(\sigma_i^2(M), d\sigma^2) + C_2(\log^6 d) \sum_{i=r+1}^d \sigma_i^2(M), \quad (9)$$

where C_0 and C_2 are absolute constants.

This bound can be interpreted as follows. The first term expresses the bias-variance tradeoff for estimating M_r , while the second term depends on the Frobenius norm of M_c . (Note that the $\log^6 d$ factor may not be tight.) In particular, this implies: $\|\hat{M} - M\|_F \leq \sqrt{C_0} \sqrt{rd}\sigma + \sqrt{C_2}(\log^3 d) \|M_c\|_F$. This can be compared with equation (8) (involving $\|M_c\|_*$). This bound will be better when $\|M_c\|_F \ll \|M_c\|_*$, i.e., when the tail M_c has slowly-decaying eigenvalues (in physical terms, it is highly mixed).

Proposition 2.4 is an adaptation of Theorem 2.8 in [5]. We sketch the proof in section B in [21]. Note that this bound is not universal: it shows that for all matrices M , a random choice of the sampling operator \mathcal{A} is likely to work.

3 Proof of the RIP for Pauli Measurements

We now prove Theorem 2.1. The general approach involving Dudley's entropy bound is similar to [12], while the technical part of the proof (bounding certain covering numbers) uses ideas from [16]. We summarize the argument here; the details are given in section A in [21].

3.1 Overview

Let $U_2 = \{X \in \mathbb{C}^{d \times d} \mid \|X\|_F \leq 1, \|X\|_* \leq \sqrt{r}\|X\|_F\}$. Let \mathcal{B} be the set of all self-adjoint linear operators from $\mathbb{C}^{d \times d}$ to $\mathbb{C}^{d \times d}$, and define the following norm on \mathcal{B} :

$$\|\mathcal{M}\|_{(r)} = \sup_{X \in U_2} |(X, \mathcal{M}X)|. \quad (10)$$

(Suppose $r \geq 2$, which is sufficient for our purposes. It is straightforward to show that $\|\cdot\|_{(r)}$ is a norm, and that \mathcal{B} is a Banach space with respect to this norm.) Then let us define

$$\varepsilon_r(\mathcal{A}) = \|\mathcal{A}^* \mathcal{A} - \mathcal{I}\|_{(r)}. \quad (11)$$

By an elementary argument, in order to prove RIP, it suffices to show that $\varepsilon_r(\mathcal{A}) < 2\delta - \delta^2$. We will proceed as follows: we will first bound $\mathbb{E}\varepsilon_r(\mathcal{A})$, then show that $\varepsilon_r(\mathcal{A})$ is concentrated around its mean.

Using a standard symmetrization argument, we have that $\mathbb{E}\varepsilon_r(\mathcal{A}) \leq 2\mathbb{E}\left\|\sum_{j=1}^m \varepsilon_j |S_j\rangle\langle S_j| \frac{d^2}{m}\right\|_{(r)}$, where the ε_j are Rademacher (iid ± 1) random variables. Here the round ket notation $|S_j\rangle$ means we view the matrix S_j as an element of the vector space \mathbb{C}^{d^2} with Hilbert-Schmidt inner product; the round bra $\langle S_j|$ denotes the adjoint element in the (dual) vector space.

Now we use the following lemma, which we will prove later. This bounds the expected magnitude in (r) -norm of a Rademacher sum of a fixed collection of operators V_1, \dots, V_m that have small operator norm.

Lemma 3.1 *Let $m \leq d^2$. Fix some $V_1, \dots, V_m \in \mathbb{C}^{d \times d}$ that have uniformly bounded operator norm, $\|V_i\| \leq K$ (for all i). Let $\varepsilon_1, \dots, \varepsilon_m$ be iid uniform ± 1 random variables. Then*

$$\mathbb{E}_\varepsilon \left\| \sum_{i=1}^m \varepsilon_i |V_i\rangle\langle V_i| \right\|_{(r)} \leq C_5 \cdot \left\| \sum_{i=1}^m |V_i\rangle\langle V_i| \right\|_{(r)}^{1/2}, \quad (12)$$

where $C_5 = \sqrt{r} \cdot C_4 K \log^{5/2} d \log^{1/2} m$ and C_4 is some universal constant.

After some algebra, one gets that $\mathbb{E}\varepsilon_r(\mathcal{A}) \leq 2(\mathbb{E}\varepsilon_r(\mathcal{A}) + 1)^{1/2} \cdot C_5 \cdot \sqrt{\frac{d}{m}}$, where $C_5 = \sqrt{r} \cdot C_4 K \log^3 d$. By finding the roots of this quadratic equation, we get the following bound on $\mathbb{E}\varepsilon_r(\mathcal{A})$. Let $\lambda \geq 1$. Assume that $m \geq \lambda d(2C_5)^2 = \lambda \cdot 4C_4^2 \cdot dr \cdot K^2 \log^6 d$. Then we have the desired result:

$$\mathbb{E}\varepsilon_r(\mathcal{A}) \leq \frac{1}{\lambda} + \frac{1}{\sqrt{\lambda}}. \quad (13)$$

It remains to show that $\varepsilon_r(\mathcal{A})$ is concentrated around its expectation. For this we use a concentration inequality from [22] for sums of independent symmetric random variables that take values in some Banach space. See section A in [21] for details.

3.2 Proof of Lemma 3.1 (bounding a Rademacher sum in (r) -norm)

Let $L_0 = \mathbb{E}_\varepsilon \|\sum_{i=1}^m \varepsilon_i |V_i\rangle \langle V_i| \|_{(r)}$; this is the quantity we want to bound. Using a standard comparison principle, we can replace the ± 1 random variables ε_i with iid $N(0, 1)$ Gaussian random variables g_i ; then we get

$$L_0 \leq \mathbb{E}_g \sup_{X \in U_2} \sqrt{\frac{\pi}{2}} |G(X)|, \quad G(X) = \sum_{i=1}^m g_i |(V_i, X)|^2. \quad (14)$$

The random variables $G(X)$ (indexed by $X \in U_2$) form a Gaussian process, and L_0 is upper-bounded by the expected supremum of this process. Using the fact that $G(0) = 0$ and $G(\cdot)$ is symmetric, and Dudley's inequality (Theorem 11.17 in [22]), we have

$$L_0 \leq \sqrt{2\pi} \mathbb{E}_g \sup_{X \in U_2} G(X) \leq 24\sqrt{2\pi} \int_0^\infty \log^{1/2} N(U_2, d_G, \varepsilon) d\varepsilon, \quad (15)$$

where $N(U_2, d_G, \varepsilon)$ is a covering number (the number of balls in $\mathbb{C}^{d \times d}$ of radius ε in the metric d_G that are needed to cover the set U_2), and the metric d_G is given by

$$d_G(X, Y) = \left(\mathbb{E}[(G(X) - G(Y))^2] \right)^{1/2}. \quad (16)$$

Define a new norm (actually a semi-norm) $\|\cdot\|_X$ on $\mathbb{C}^{d \times d}$, as follows:

$$\|M\|_X = \max_{i=1, \dots, m} |(V_i, M)|. \quad (17)$$

We use this to upper-bound the metric d_G . An elementary calculation shows that $d_G(X, Y) \leq 2R\|X - Y\|_X$, where $R = \|\sum_{i=1}^m |V_i\rangle \langle V_i| \|_{(r)}^{1/2}$. This lets us upper-bound the covering numbers in d_G with covering numbers in $\|\cdot\|_X$:

$$N(U_2, d_G, \varepsilon) \leq N(U_2, \|\cdot\|_X, \frac{\varepsilon}{2R}) = N(\frac{1}{\sqrt{r}}U_2, \|\cdot\|_X, \frac{\varepsilon}{2R\sqrt{r}}). \quad (18)$$

We will now bound these covering numbers. First, we introduce some notation: let $\|\cdot\|_p$ denote the Schatten p -norm on $\mathbb{C}^{d \times d}$, and let B_p be the unit ball in this norm. Also, let B_X be the unit ball in the $\|\cdot\|_X$ norm.

Observe that $\frac{1}{\sqrt{r}}U_2 \subseteq B_1 \subseteq K \cdot B_X$. (The second inclusion follows because $\|M\|_X \leq \max_{i=1, \dots, m} \|V_i\| \|M\|_* \leq K \|M\|_*$.) This gives a simple bound on the covering numbers:

$$N(\frac{1}{\sqrt{r}}U_2, \|\cdot\|_X, \varepsilon) \leq N(B_1, \|\cdot\|_X, \varepsilon) \leq N(K \cdot B_X, \|\cdot\|_X, \varepsilon). \quad (19)$$

This is 1 when $\varepsilon \geq K$. So, in Dudley's inequality, we can restrict the integral to the interval $[0, K]$.

When ε is small, we will use the following simple bound (equation (5.7) in [23]):

$$N(K \cdot B_X, \|\cdot\|_X, \varepsilon) \leq (1 + \frac{2K}{\varepsilon})^{2d^2}. \quad (20)$$

When ε is large, we will use a more sophisticated bound based on Maurey's empirical method and entropy duality, which is due to [16] (see also [17]):

$$N(B_1, \|\cdot\|_X, \varepsilon) \leq \exp(\frac{C_1^2 K^2}{\varepsilon^2} \log^3 d \log m), \quad \text{for some constant } C_1. \quad (21)$$

We defer the proof of (21) to the next section.

Using (20) and (21), we can bound the integral in Dudley's inequality. We get

$$L_0 \leq C_4 R \sqrt{r} K \log^{5/2} d \log^{1/2} m, \quad (22)$$

where C_4 is some universal constant. This proves the lemma.

3.3 Proof of Equation (21) (covering numbers of the nuclear-norm ball)

Our result will follow easily from a bound on covering numbers introduced in [16] (where it appears as Lemma 1):

Lemma 3.2 *Let E be a Banach space, having modulus of convexity of power type 2 with constant $\lambda(E)$. Let E^* be the dual space, and let $T_2(E^*)$ denote its type 2 constant. Let B_E denote the unit ball in E .*

Let $V_1, \dots, V_m \in E^$, such that $\|V_j\|_{E^*} \leq K$ (for all j). Define the norm on E ,*

$$\|M\|_X = \max_{j=1, \dots, m} |(V_j, M)|. \quad (23)$$

Then, for any $\varepsilon > 0$,

$$\varepsilon \log^{1/2} N(B_E, \|\cdot\|_X, \varepsilon) \leq C_2 \lambda(E)^2 T_2(E^*) K \log^{1/2} m, \quad (24)$$

where C_2 is some universal constant.

The proof uses entropy duality to reduce the problem to bounding the “dual” covering number. The basic idea is as follows. Let ℓ_p^m denote the complex vector space \mathbb{C}^m with the ℓ_p norm. Consider the map $S : \ell_1^m \rightarrow E^*$ that takes the j 'th coordinate vector to V_j . Let $N(S)$ denote the number of balls in E^* needed to cover the image (under the map S) of the unit ball in ℓ_1^m . We can bound $N(S)$ using Maurey's empirical method. Also define the dual map $S^* : E \rightarrow \ell_\infty^m$, and the associated dual covering number $N(S^*)$. Then $N(B_E, \|\cdot\|_X, \varepsilon)$ is related to $N(S^*)$. Finally, $N(S)$ and $N(S^*)$ are related via entropy duality inequalities. See [16] for details.

We will apply this lemma as follows, using the same approach as [17]. Let S_p denote the Banach space consisting of all matrices in $\mathbb{C}^{d \times d}$ with the Schatten p -norm. Intuitively, we want to set $E = S_1$ and $E^* = S_\infty$, but this won't work because $\lambda(S_1)$ is infinite. Instead, we let $E = S_p$, $p = (\log d)/(\log d - 1)$, and $E^* = S_q$, $q = \log d$. Note that $\|M\|_p \leq \|M\|_*$, hence $B_1 \subseteq B_p$ and

$$\varepsilon \log^{1/2} N(B_1, \|\cdot\|_X, \varepsilon) \leq \varepsilon \log^{1/2} N(B_p, \|\cdot\|_X, \varepsilon). \quad (25)$$

Also, we have $\lambda(E) \leq 1/\sqrt{p-1} = \sqrt{\log d - 1}$ and $T_2(E^*) \leq \lambda(E) \leq \sqrt{\log d - 1}$ (see the Appendix in [17]). Note that $\|M\|_q \leq e\|M\|$, thus we have $\|V_j\|_q \leq eK$ (for all j). Then, using the lemma, we have

$$\varepsilon \log^{1/2} N(B_p, \|\cdot\|_X, \varepsilon) \leq C_2 \log^{3/2} d (eK) \log^{1/2} m, \quad (26)$$

which proves the claim.

4 Outlook

We have showed that random Pauli measurements obey the restricted isometry property (RIP), which implies strong error bounds for low-rank matrix recovery. The key technical tool was a bound on covering numbers of the nuclear norm ball, due to Guédon et al [16].

An interesting question is whether this method can be applied to other problems, such as matrix completion, or constructing embeddings of low-dimensional manifolds into linear spaces with slightly higher dimension. For matrix completion, one can compare with the work of Negahban and Wainwright [10], where the sampling operator satisfies restricted strong convexity (RSC) over a certain set of “non-spiky” low-rank matrices. For manifold embeddings, one could try to generalize the results of [24], which use the sparse-vector RIP to construct Johnson-Lindenstrauss metric embeddings.

There are also many questions pertaining to low-rank quantum state tomography. For example, how does the matrix Lasso compare to the traditional approach using maximum likelihood estimation? Also, there are several variations on the basic tomography problem, and alternative notions of sparsity (e.g., elementwise sparsity in a known basis) [25], which have not been fully explored.

Acknowledgements: Thanks to David Gross, Yaniv Plan, Emmanuel Candès, Stephen Jordan, and the anonymous reviewers, for helpful suggestions. Parts of this work were done at the University of California, Berkeley, and supported by NIST grant number 60NANB10D262. This paper is a contribution of the National Institute of Standards and Technology, and is not subject to U.S. copyright.

References

- [1] M. Fazel. *Matrix Rank Minimization with Applications*. PhD thesis, Stanford, 2002.
- [2] N. Srebro. *Learning with Matrix Factorizations*. PhD thesis, MIT, 2004.
- [3] B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum rank solutions to linear matrix equations via nuclear norm minimization. *SIAM Review*, 52(3):471–501, 2010.
- [4] M. Fazel, E. Candes, B. Recht, and P. Parrilo. Compressed sensing and robust recovery of low rank matrices. In *42nd Asilomar Conference on Signals, Systems and Computers*, pages 1043–1047, 2008.
- [5] E. J. Candes and Y. Plan. Tight oracle bounds for low-rank matrix recovery from a minimal number of random measurements. 2009.
- [6] E. J. Candes and B. Recht. Exact matrix completion via convex optimization. *Found. of Comput. Math.*, 9:717–772.
- [7] E. J. Candes and T. Tao. The power of convex relaxation: Near-optimal matrix completion. *IEEE Trans. Inform. Theory*, 56(5):2053–2080, 2009.
- [8] D. Gross. Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans. Inform. Theory*, to appear. arXiv:0910.1879, 2010.
- [9] B. Recht. A simpler approach to matrix completion. *J. Machine Learning Research* (to appear), 2010.
- [10] S. Negahban and M. J. Wainwright. Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. arXiv:1009.2118, 2010.
- [11] E. J. Candes and T. Tao. Near-optimal signal recovery from random projections: universal encoding strategies. *IEEE Trans. Inform. Theory*, 52:5406–5425, 2004.
- [12] M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Commun. Pure and Applied Math.*, 61:1025–1045, 2008.
- [13] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert. Quantum state tomography via compressed sensing. *Phys. Rev. Lett.*, 105(15):150401, Oct 2010. arXiv:0909.3304.
- [14] E. J. Candes and Y. Plan. Matrix completion with noise. *Proc. IEEE*, 98(6):925 – 936, 2010.
- [15] B. Brown, S. Flammia, D. Gross, and Y.-K. Liu. in preparation, 2011.
- [16] O. Guédon, S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann. Majorizing measures and proportional subsets of bounded orthonormal systems. *Rev. Mat. Iberoamericana*, 24(3):1075–1095, 2008.
- [17] G. Aubrun. On almost randomizing channels with a short Kraus decomposition. *Commun. Math. Phys.*, 288:1103–1116, 2009.
- [18] M. A. Nielsen and I. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2001.
- [19] A. Rohde and A. Tsybakov. Estimation of high-dimensional low-rank matrices. arXiv:0912.5338, 2009.
- [20] V. Koltchinskii, K. Lounici, and A. B. Tsybakov. Nuclear norm penalization and optimal rates for noisy low rank matrix completion. arXiv:1011.6256, 2010.
- [21] Y.-K. Liu. Universal low-rank matrix recovery from Pauli measurements. arXiv:1103.2816, 2011.
- [22] M. Ledoux and M. Talagrand. *Probability in Banach spaces*. Springer, 1991.
- [23] G. Pisier. *The volume of convex bodies and Banach space geometry*. Cambridge, 1999.
- [24] F. Kraher and R. Ward. New and improved Johnson-Lindenstrauss embeddings via the restricted isometry property. *SIAM J. Math. Anal.*, 43(3):1269–1281, 2011.
- [25] A. Shabani, R. L. Kosut, M. Mohseni, H. Rabitz, M. A. Broome, M. P. Almeida, A. Fedrizzi, and A. G. White. Efficient measurement of quantum dynamics via compressive sensing. *Phys. Rev. Lett.*, 106(10):100401, 2011.
- [26] P. Wojtaszczyk. Stability and instance optimality for gaussian measurements in compressed sensing. *Found. Comput. Math.*, 10(1):1–13, 2009.