

## Quantum State Tomography via Compressed Sensing

David Gross,<sup>1</sup> Yi-Kai Liu,<sup>2</sup> Steven T. Flammia,<sup>3</sup> Stephen Becker,<sup>4</sup> and Jens Eisert<sup>5</sup>

<sup>1</sup>*Institute for Theoretical Physics, Leibniz University Hannover, 30167 Hannover, Germany*

<sup>2</sup>*Institute for Quantum Information, California Institute of Technology, Pasadena, California, USA*

<sup>3</sup>*Perimeter Institute for Theoretical Physics, Waterloo, Ontario, N2L 2Y5 Canada*

<sup>4</sup>*Applied and Computational Mathematics, California Institute of Technology, Pasadena, California, USA*

<sup>5</sup>*Institute of Physics und Astronomy, University of Potsdam, 14476 Potsdam, Germany*

(Received 21 October 2009; published 4 October 2010)

We establish methods for quantum state tomography based on compressed sensing. These methods are specialized for quantum states that are fairly pure, and they offer a significant performance improvement on large quantum systems. In particular, they are able to reconstruct an unknown density matrix of dimension  $d$  and rank  $r$  using  $O(rd\log^2 d)$  measurement settings, compared to standard methods that require  $d^2$  settings. Our methods have several features that make them amenable to experimental implementation: they require only simple Pauli measurements, use fast convex optimization, are stable against noise, and can be applied to states that are only approximately low rank. The acquired data can be used to certify that the state is indeed close to pure, so no *a priori* assumptions are needed.

DOI: 10.1103/PhysRevLett.105.150401

PACS numbers: 03.65.Wj, 03.67.-a

The tasks of reconstructing the quantum states and processes produced by physical systems—known respectively as quantum state and process tomography [1]—are of increasing importance in physics and especially in quantum information science. Tomography has been used to characterize the quantum state of trapped ions [2] and an optical entangling gate [3] among many other implementations. But a fundamental difficulty in performing tomography on many-body systems is the exponential growth in the state space dimension. For example, to get a maximum-likelihood estimate of a quantum state of 8 ions, Ref. [2] required hundreds of thousands of measurements and weeks of postprocessing.

Still, one might hope to overcome this obstacle because the vast majority of quantum states are not of physical interest. Rather, one is often interested in states with special properties: pure states, states with particular symmetries, ground states of local Hamiltonians, etc., and tomography might be more efficient in such special cases [4].

In particular, consider pure or nearly pure quantum states, i.e., states with low entropy. More precisely, consider a quantum state that is essentially supported on an  $r$ -dimensional space, meaning the density matrix is close (in a given norm) to a matrix of rank  $r$ , where  $r$  is small. Such states arise in very common physical settings, e.g., a pure state subject to a local noise process [5].

A standard implementation of tomography [6,7] would use  $d^2$  or more measurement settings, where  $d = 2^n$  for an  $n$ -qubit system. But a simple parameter counting argument suggests that  $O(rd)$  settings could possibly suffice—a significant improvement. However, it is not clear how to achieve this performance in practice, i.e., how to choose these measurements, or how to efficiently reconstruct the density matrix. For instance, the problem of finding a

minimum-rank matrix subject to linear constraints is NP-hard in general [8].

In addition to a reduction in experimental complexity, one might hope that a postprocessing algorithm taking as input only  $O(rd) \ll d^2$  numbers could be tuned to run considerably faster than standard methods. Since the output of the procedure is a low-rank approximation to the density operator and only requires  $O(rd)$  numbers be specified, it becomes conceivable that the run time scales better than  $O(d^2)$ , clearly impossible for naive approaches using dense matrices.

In this Letter, we introduce a method to achieve such drastic reductions in measurement complexity, together with efficient algorithms for postprocessing. The approach further develops ideas that have recently been studied under the label of “compressed sensing.” Compressed sensing [9] provides techniques for recovering a sparse vector from a small number of measurements [10]. Here, sparsity means that this vector contains only a few nonzero entries in a specified basis, and the measurements are linear functions of its entries. When the measurements are chosen at random (in a certain precise sense), then with high probability two surprising things happen: the vector is uniquely determined by a small number of measurements, and it can be recovered by an efficient convex optimization algorithm [9].

Matrix completion [11–13] is a generalization of compressed sensing from vectors to matrices. Here, one recovers certain “incoherent” low-rank matrices  $X$  from a small number of matrix elements  $X_{i,j}$ . The problem of low-rank quantum state tomography bears a strong resemblance to matrix completion. However, there are important differences. We wish to use measurements that can be more easily implemented in an experiment than obtaining elements  $\rho_{i,j}$  of density matrices. Previous results [11–13]

cannot be applied to this more general situation. We would also like to avoid any unnatural incoherence assumptions crucial in prior work [11].

Our first result is a protocol for tomography that overcomes both of these difficulties: it uses Pauli measurements only, and it works for arbitrary density matrices. We prove that only  $O(rd\log^2 d)$  measurement settings suffice. What is more, our proof introduces some new techniques, which both generalize and vastly simplify the previous work on matrix completion.

In a real experiment, the measurements are noisy, and the true state is only approximately low rank. We show that our method is robust to these sources of error. We also describe ways to certify that a state is nearly pure without any *a priori* assumptions.

Finally, we present fast algorithms for reconstructing the density matrix from the measurement statistics based on semidefinite programming—a feature not present in earlier methods for pure-state tomography [4,6,7]. Reconstructing a low-rank density matrix for 8 qubits takes about 1 min on an ordinary laptop computer.

While our methods do not overcome the exponential growth in measurement complexity (which is provably impossible for any protocol capable of handling generic pure states), they do significantly push the boundary of what can be done in a realistic setting [14].

**Matrix recovery using Pauli measurements.**—We consider the case of  $n$  spin-1/2 systems in an unknown state  $\rho$  [15]. An  $n$ -qubit Pauli matrix is of the form  $w = \bigotimes_{i=1}^n w_i$ , where  $w_i \in \{\mathbb{1}, \sigma^x, \sigma^y, \sigma^z\}$ . There are  $d^2$  such matrices, labeled  $w(a)$ ,  $a \in [1, d^2]$ . The protocol proceeds as follows: choose  $m$  integers  $A_1, \dots, A_m \in [1, d^2]$  at random and measure the expectation values  $\text{tr}\rho w(A_i)$ . One then solves a convex optimization problem: minimize  $\|\sigma\|_{\text{tr}}$  [16] subject to

$$\text{tr } \sigma = 1, \quad \text{tr} w(A_i) \sigma = \text{tr} w(A_i) \rho. \quad (1)$$

**Theorem 1 (low-rank tomography)**—Let  $\rho$  be an arbitrary state of rank  $r$ . If  $m = cdr\log^2 d$  randomly chosen Pauli expectations are known, then  $\rho$  can be uniquely reconstructed by solving the convex optimization problem (1) with probability of failure exponentially small in  $c$ .

The proof is inspired by, but technically very different from, earlier work on matrix completion [11]. Our methods are more general, can be tuned to give tighter bounds, and are much more compact, allowing us to present a fairly complete argument in this Letter. A more detailed presentation will be published elsewhere [17].

**Proof.**—Here we sketch the argument and explain the main ideas; detailed calculations are in the supplementary material [18].

Note that the linear constraints (1) depend only on the projection of  $\rho$  onto the span of the measured observables  $w(A_1), \dots, w(A_m)$ . This is precisely the range of the “sampling operator”  $\mathcal{R}$ :  $\rho \mapsto \frac{d}{m} \sum_{i=1}^m w(A_i) \text{tr} \rho w(A_i)$ . (Note that  $\mathbb{E}[\mathcal{R}(\rho)] = \rho$ .) Indeed, the convex program can be written as  $\min_{\sigma} \|\sigma\|_{\text{tr}}$  s.t.  $\mathcal{R}\sigma = \mathcal{R}\rho$ . Evidently,

the solution is unique if for all deviations  $\Delta := \sigma - \rho$  away from  $\rho$  either  $\mathcal{R}\Delta \neq 0$  or  $\|\rho + \Delta\|_{\text{tr}} > \|\rho\|_{\text{tr}}$ .

We will ascertain this by using a basic idea from convex optimization: constructing a *strict subgradient*  $Y$  for the norm. A matrix  $Y$  is a strict subgradient if  $\|\rho + \Delta\|_{\text{tr}} > \|\rho\|_{\text{tr}} + \text{tr} Y \Delta$  for all  $\Delta \neq 0$ . The main contribution below is a method for constructing such a  $Y$  which is also in the range of  $\mathcal{R}$ . For then  $\mathcal{R}\Delta = 0$  implies that  $\Delta$  is orthogonal to the range of  $\mathcal{R}$ , thus  $\text{tr} Y \Delta = 0$  and the subgradient condition reads  $\|\rho + \Delta\|_{\text{tr}} > \|\rho\|_{\text{tr}}$ . This implies uniqueness. (In fact, it suffices to approximate the condition.)

Let  $E$  be the projection onto the range of  $\rho$ , let  $T$  be the space spanned by those operators whose row or column space is contained in range  $\rho$ . Let  $\mathcal{P}_T$  be the projection onto  $T$ ,  $\mathcal{P}_T^\perp$  onto the orthogonal complement. Decompose  $\Delta = \Delta_T + \Delta_T^\perp$ , the parts of  $\Delta$  that lie in the subspaces  $T$  and  $T^\perp$ . We distinguish two cases: (i)  $\|\Delta_T\|_2 > d^2 \|\Delta_T^\perp\|_2$ , and (ii)  $\|\Delta_T\|_2 \leq d^2 \|\Delta_T^\perp\|_2$  [16].

Case (i) is easier. In this case,  $\Delta$  is well approximated by  $\Delta_T$  and essentially we only have to show that the restriction  $\mathcal{A} := \mathcal{P}_T \mathcal{R} \mathcal{P}_T$  of  $\mathcal{R}$  to  $T$  is invertible. Using a noncommutative large-deviation bound (see Refs. [18,19]),

$$\Pr[\|\mathcal{A} - \mathbb{1}_T\| > t] < 4dre^{-t^2 \kappa/8} \quad (2)$$

where  $\kappa = m/(dr)$  [16]. Hence the probability that  $\|\mathcal{A} - \mathbb{1}_T\| > \frac{1}{2}$  is smaller than  $4dre^{-\kappa/32} =: p_1$ . If that is not the case, one easily sees that  $\|\mathcal{R}\Delta\|_2 > 0$ , concluding the proof for this case.

Case (ii) is more involved. A matrix  $Y \in \text{span}(w(A_1), \dots, w(A_m))$  is an *almost subgradient* [20] if

$$\|\mathcal{P}_T Y - E\|_2 \leq 1/(2d^2), \quad \|\mathcal{P}_T^\perp Y\| < 1/2. \quad (3)$$

First, suppose such a  $Y$  exists. Then a simple calculation (see Ref. [18]) using the condition (ii) shows that  $\mathcal{R}\Delta = 0$  indeed implies  $\|\rho + \Delta\|_{\text{tr}} > \|\rho\|_{\text{tr}}$  as hinted at above. This proves uniqueness in case (ii). The difficult part consists in showing that an almost subgradient exists.

To this end, we design a recursive process (the “golfing scheme” [17]) which converges to a subgradient exponentially fast. Assume we draw  $l$  batches of  $\kappa_0 rd$  Pauli observables independently at random ( $\kappa_0$  will be chosen later). Define recursively  $X_0 = E$ ,

$$Y_i = \sum_{j=1}^i \mathcal{R}_j X_{j-1}, \quad X_i = E - \mathcal{P}_T Y_i, \quad (4)$$

$Y = Y_l$ . Let  $\mathcal{R}_i$  be the sampling operator associated with the  $i$ th batch, and  $\mathcal{A}_i$  its restriction to  $T$ . Assume that in each run  $\|\mathcal{A}_i - \mathbb{1}_T\|_2 < 1/2$ . Denote the probability of this event not occurring by  $p_2$ . Then

$$\begin{aligned} \|X_i\|_2 &= \|X_{i-1} - \mathcal{P}_T \mathcal{R}_i X_{i-1}\|_2 = \|(\mathbb{1}_T - \mathcal{A}_i) X_{i-1}\|_2 \\ &\leq 1/2 \|X_{i-1}\|_2, \end{aligned}$$

so that  $\|X_i\|_2 \leq 2^{-i} \|X_0\| = 2^{-i} \sqrt{r}$ . Hence,  $Y = Y_l$  fulfills the first part of (3), as soon as  $l \geq \log_2(2d^2 \sqrt{r})$ . We turn to the second part. Again using large-deviation techniques [18] we find  $\|\mathcal{P}_T^\perp \mathcal{R}_i X_{i-1}\| \leq 1/(4\sqrt{r}) \|X_{i-1}\|_2$  with some

(high) probability  $(1 - p_3)$ . Therefore:

$$\|\mathcal{P}_T^\perp Y_l\| \leq \sum_{j=1}^l \|\mathcal{P}_T^\perp \mathcal{R}_j X_{j-1}\| \leq \frac{1}{4} \sum_{j=0}^{\infty} 2^{-j} < \frac{1}{2}, \quad (5)$$

which is the second part of (3).

Lastly, we have to bound the total probability of failure  $p_f \leq p_1 + p_2 + p_3$ . Set  $\kappa_0 = 64\mu[1 + \ln(8dl)]$ , which means that  $m = dr(\ln d)^2 O(1)$  coefficients will be sampled in total. A simple calculation gives  $p_f \leq e^{-\mu}$ .  $\square$

In the remaining space, we address the important aspects of resilience against noise, certified tomography, and numerical performance. Owing to space limitations, the presentation will focus on conceptual issues, with the details in [21].

**Robustness to noise.**—Realistic situations will differ from the previous case in two regards. First, the true state  $\rho_t$  may not be low rank, but only well approximated by a state  $\rho$  of rank  $r$ :  $\|\rho_t - \rho\|_2 \leq \varepsilon_1$ . Second, due to systematic and statistical noise, the available estimates for the Pauli expectations are not exactly  $\text{tr} \rho_t w(a)$ , but of the form  $\text{tr} \omega w(a)$  for some matrix  $\omega$ . Assume  $\|\mathcal{R}\omega - \mathcal{R}\rho_t\|_2 \leq \varepsilon_2$  (in practical situations,  $\varepsilon_2$  may be estimated from the error bars associated with the individual Pauli expectation values). In order to get an estimate for  $\rho_t$ , choose some  $\lambda \geq 1$  and  $\varepsilon \geq \lambda(\sqrt{d^2/m})\varepsilon_1 + \varepsilon_2$ , and solve the convex program

$$\min \|\sigma\|_{\text{tr}}, \quad \text{subject to } \|\mathcal{R}\sigma - \mathcal{R}\omega\|_2 \leq \varepsilon. \quad (6)$$

**Observation 1 (robustness to noise)**—Let  $\rho_t$  be an approximately low-rank state as described above. Suppose  $m = cdr \log^2 d$  randomly chosen Pauli expectations are known up to an error of  $\varepsilon$  as in (6), and let  $\sigma^*$  be the solution of (6). Then the difference  $\|\sigma^* - \rho_t\|_{\text{tr}}$  is smaller than  $O(\varepsilon \sqrt{rd})$ . This holds with probability of failure at most  $1/\lambda^2$  plus the probability of failure in Theorem 1.

The proof combines ideas from Ref. [13] with our argument above [22]. The main difference from the noise-free case is that, instead of using  $\text{tr} Y \Delta = 0$ , we must now work with  $|\text{tr} Y \Delta| \leq 2\|Y\|_2 \delta$ . With this estimate, Observation 1 follows from the noise-free proof, together with some elementary calculations [18]. We remark that the above bound is likely to be quite loose; based on related work [23] and more natural noise models, we conjecture that the robustness is substantially stronger than what is shown here.

**Certified tomography of almost pure states.**—The preceding results require an *a priori* promise: that the true state  $\rho_t$  is  $\delta_1$  close to a rank- $r$  state. However, when performing tomography of an unknown state, neither  $r$  nor  $\delta_1$  are known beforehand. There are a few solutions to this quandary. First,  $r$  and  $\delta_1$  may be estimated from other physical parameters of the system, such as the strength of the local noise [5].

Another approach is to estimate  $r$  and  $\delta_1$  from the same data that are used to reconstruct the state. When  $r = 1$ , this approach is particularly effective to facilitate entirely assumption-free tomography. This is because  $\delta_1$  is related to the purity  $\text{Tr} \rho^2$ , which has a simple closed-form

expression in terms of Pauli expectation values. See Ref. [18] for details. We get:

**Observation 2 (certified tomography).**—Assume that the unknown physical state is close to being pure. Then one can find a certificate for that assumption, and reconstruct the state with explicit guarantees on the reconstruction error, from  $O(cd \log^2 d)$  Pauli expectation values. The probability of failure is exponentially small in  $c$ .

Finally, when the state is approximately low rank but not nearly pure ( $r > 1$ ), one may perform tomography using different numbers of random Pauli expectation values  $m$ . When  $m$  is larger than necessary (corresponding to an overestimate of  $r$ ), we are guaranteed to find the correct density matrix. When  $m$  is too small, we find empirically that the algorithms for reconstructing the density matrix fail to converge.

**A hybrid approach to matrix recovery.**—Here we describe a variant of our tomography method that makes the classical postprocessing step (i.e., solving the convex program (1) to reconstruct the density matrix) faster. This method also uses random Pauli measurements, but they are chosen in a structured way. Any Pauli matrix is of the form  $w(u, v) = \bigotimes_{k=1}^n i^{u_k v_k} (\sigma^x)^{u_k} (\sigma^z)^{v_k}$  for  $u, v \in \{0, 1\}^n$ . We choose a random subset  $S \subset \{0, 1\}^n$  of size  $O(r \text{poly} \log(d))$ , and then for all  $u \in S$  and  $v \in \{0, 1\}^n$ , measure the Pauli matrix  $w(u, v)$ . We call this the “hybrid method” because it is equivalent to a certain structured matrix completion problem. This fact implies that certain key computations in solving the convex program (1) can be implemented in time  $O(d)$  rather than  $O(d^2)$  [24]. However, the hybrid method is not covered by the strong theoretical guarantees shown earlier, though it does give accurate results in practice (when combined with twirling by unitary  $k$ -designs [25] to ensure incoherence). For a more complete discussion, see Ref. [18].

**Numerical results.**—We numerically simulated both the random Pauli and hybrid approaches discussed above. For both approaches, we used singular value thresholding (SVT) [24]. Instead of directly solving Eq. (6), SVT minimizes  $\tau \|\sigma\|_{\text{tr}} + \|\sigma\|_2^2/2$  subject to  $|\text{tr}(\sigma - \omega)w(A_i)| \leq \delta$ , which is a good proxy to Eq. (6) when  $\tau$  dominates the second term; the programs are equivalent in the limit  $\tau \rightarrow \infty$  ([provided Eq. (6) has a unique solution] [24]. Estimating the second term for typical states suggests choosing  $2\tau r \gg 1$ ; we use  $\tau = 5$ . To simulate tomography, we chose a random state from the Haar measure on a  $d \times r$  dimensional system and traced out the  $r$ -dimensional ancilla, then applied depolarizing noise of strength  $\gamma$ . We sampled expectation values associated with randomly chosen operators as above, and added additional statistical noise which was i.i.d. Gaussian with variance  $\sigma^2$  and mean zero. We used SVT and quantified the quality of the reconstruction by the fidelity and the trace distance for various values of  $m$ , each averaged over 5 simulations. This dependence is shown in Fig. 1. The reconstruction is remarkably high fidelity, despite severe undersampling and



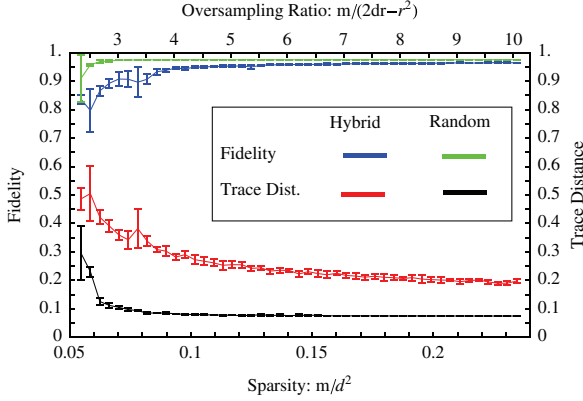


FIG. 1 (color online). Average fidelity and trace distance vs (scaled) number of measurement settings  $m$  for random states of  $n = 8$  qubits, so  $d = 2^n$ . As discussed in the text, the sampled states had rank  $r = 3$ , depolarizing noise of 5% and Gaussian statistical noise with  $\sigma = 0.1/d$ . Both the random Pauli and hybrid approaches are shown.

corruption by both depolarizing and statistical noise [26]. Using the hybrid method with 8 qubits on a rank 3 state plus  $\gamma = 5\%$  depolarizing, and statistical noise strength  $\sigma d = 0.1$ , we typically achieve 95% fidelity reconstructions in under 10 seconds on a modest laptop with 2 GB of RAM and a 2.2 GHz dual-core processor using MATLAB—even though 90% of the matrix elements remain unsampled. Increasing the number of samples only improves our accuracy and speed, so long as sparsity is maintained.

Using truly randomly chosen Pauli observables (instead of the hybrid method) slightly increases the processing time due to the dense matrix multiplications involved: in our setup about 1 min. However, this method achieves even better performance with respect to errors, as seen in Fig. 1.

The simulations above show that our method works for generic low-rank states. Lastly, we demonstrate the functioning of the approach in the experimental context of the state  $\rho$  found in the 8 ion experiment of Ref. [2]. To exemplify the above results, we simulated physical measurements by sampling from the probability distribution computed using the Born rule applied to the reconstructed state  $\rho$ . This state is approximately low rank, with 99% of the weight concentrated on the first 11 eigenvectors. The standard deviation per observable was  $3/d$ . Fewer than 30% of all Pauli matrices were chosen randomly. From this information, a rank = 3 approximation  $\sigma$  with fidelity of 90.5% with respect to  $\rho$  was found in about 3 min on the aforementioned laptop.

We thank E. Candès and Y. Plan for useful discussions. Research at PI is supported by the Government of Canada through Industry Canada and by the Province of Ontario through the Ministry of Research & Innovation. Y.L. is supported by the NSF, J.E. by the EU (QAP, QESSENCE, MINOS, COMPAS) and the EURYI, D.G. by the EU (CORNER). We thank the anonymous referees for many helpful suggestions.

- [1] In *Quantum State Estimation*, edited by M. Paris and J. Reháček, Lect. Notes Phys. Vol. 649 (Springer, Heidelberg, 2004).
- [2] H. Häffner *et al.*, *Nature (London)* **438**, 643 (2005).
- [3] J. L. O'Brien *et al.*, *Nature (London)* **426**, 264 (2003).
- [4] M. S. Kaznady and D. F. V. James, *Phys. Rev. A* **79**, 022109 (2009).
- [5] Consider a pure state of  $n$  qubits subject to local noise that occurs with probability  $p$  on each site. Then the density matrix is well approximated by a matrix of rank  $r = 2^{nH(p)} = d^{H(p)}$ , where  $H(p)$  is the binary entropy of  $p$ , and  $d = 2^n$  is the Hilbert space dimension. When  $p$  is small, we have  $r \ll d$ .
- [6] G. M. D'Ariano, L. Maccone, and M. Painsi, *J. Opt. B* **5**, 77 (2003); V. V. Dodonov and V. I. Man'ko, *Phys. Lett. A* **229**, 335 (1997).
- [7] J.-P. Amiet and S. Weigert, *J. Phys. A* **32**, 2777 (1999).
- [8] B. K. Natarajan, *SIAM J. Comput.* **24**, 227 (1995).
- [9] D. Donoho, *IEEE Trans. Inf. Theory* **52**, 1289 (2006); E. Candès and T. Tao, *IEEE Trans. Inf. Theory* **52**, 5406 (2006).
- [10] R. L. Kosut, *arXiv:0812.4323*.
- [11] E. J. Candès and B. Recht, *Found. Comput. Math.* **9**, 717 (2009).
- [12] E. J. Candès and T. Tao, *arXiv:0903.1476* [IEEE Trans. Inform. Th. (to be published)].
- [13] E. J. Candès and Y. Plan, *Proc. IEEE* **98**, 925 (2010).
- [14] Our techniques also apply to process tomography: to characterize an unknown quantum process  $\mathcal{E}$ , prepare the Jamiołkowski state  $\rho_{\mathcal{E}}$ , and perform state tomography on  $\rho_{\mathcal{E}}$ . Our methods work when  $\mathcal{E}$  can approximately be written as a sum of only a few Kraus operators, because this implies that  $\rho_{\mathcal{E}}$  has small rank.
- [15] The techniques easily generalize to spin- $j$  particles [17].
- [16] We use the usual matrix norms  $\|A\|_{\text{tr}} = \sum_i \sigma_i$ ,  $\|A\|_2^2 = \text{tr} A^\dagger A = \sum_i \sigma_i^2$ ,  $\|A\| = \max_i \sigma_i$ , with  $\sigma_i$  the singular values of  $A$ . The last definition extends to superoperators: if  $\mathcal{A}$  is a superoperator, then  $\|\mathcal{A}\|$  is its largest singular value, or, equivalently  $\|\mathcal{A}\| = \sup_{\sigma, \|\sigma\|_2=1} \|\mathcal{A}\sigma\|_2$  (a.k.a. “ $2 \rightarrow 2$ ”-norm).
- [17] D. Gross, *arxiv:0910.1879* [IEEE Trans. Inform. Th. (to be published)].
- [18] See supplementary material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.105.150401>.
- [19] R. Ahlswede and A. Winter, *IEEE Trans. Inf. Theory* **48**, 569 (2002).
- [20] If the term  $1/(2d^2)$  were zero,  $Y$  would be a strict subgradient.
- [21] S. Becker, S. T. Flammia, D. Gross, Y.-K. Liu, and J. Eisert (to be published).
- [22] Going beyond [13], we bound deviations in 1-norm, as opposed to 2-norm. This carries an operational meaning in terms of statistical distinguishability.
- [23] M. Fazel, E. Candès, B. Recht, and P. Parrilo, *Proc. Asilomar Conf. CA, Nov 2008*.
- [24] J.-F. Cai, E. J. Candès, and Z. Shen, *arXiv:0810.3286*.
- [25] A. W. Harrow and R. A. Low, *Proceedings of 13th Intl. Workshop on Randomization and Computation (RANDOM 2009)*, Berkeley, CA, USA, Aug 2009, Lecture Notes in Comput. Sci. Vol. 5687 (Springer, Berlin, 2009), p. 548; D. Gross, K. Audenaert, and J. Eisert, *J. Math. Phys.* (N.Y.) **48**, 052104 (2007).
- [26] The estimate returned by SVT typically has a subnormalized trace, which we handle by renormalizing or by debiasing [21].