Permutationally invariant state reconstruction

Moroder, Tobias; Hyllus, Philipp; Tóth, Géza; Schwemmer, Christian; Niggebaum, Alexander; Gaile, Stefanie Bianca; Gühne, Otfried; Weinfurter, Harald

Published in:
New Journal of Physics

Link to article, DOI:
10.1088/1367-2630/14/10/105001

Publication date:
2012

Document Version
Publisher's PDF, also known as Version of record

Permutationally invariant state reconstruction

2012 New J. Phys. 14 105001

(http://iopscience.iop.org/1367-2630/14/10/105001)

New Journal of Physics
The open-access journal for physics

Permutationally invariant state reconstruction

Tobias Moroder[1,2,9], Philipp Hyllus[3], Géza Tóth[3,4,5], Christian Schwemmer[6,7], Alexander Niggebaum[6,7], Stefanie Gaile[8], Otfried Gühne[1,2] and Harald Weinfurter[6,7]

[1] Naturwissenschaftlich-Technische Fakultät, Universität Siegen, Walter-Flex-Straße 3, D-57068 Siegen, Germany
[2] Institut für Quantenoptik und Quanteninformation, Österreichische Akademie der Wissenschaften, Technikerstraße 21A, A-6020 Innsbruck, Austria
[3] Department of Theoretical Physics, University of the Basque Country UPV/EHU, PO Box 644, E-48080 Bilbao, Spain
[4] IKERBASQUE, Basque Foundation for Science, E-48011 Bilbao, Spain
[5] Wigner Research Centre for Physics, Hungarian Academy of Sciences, PO Box 49, H-1525 Budapest, Hungary
[6] Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Straße 1, D-85748 Garching, Germany
[7] Fakultät für Physik, Ludwig-Maximilians-Universität, D-80797 München, Germany
[8] Technical University of Denmark, Department of Mathematics, Matematiktorvet Building 303 B, 2800 Kgs. Lyngby, Denmark
E-mail: moroder@physik.uni-siegen.de

New Journal of Physics **14** (2012) 105001 (25pp)
Received 21 May 2012
Published 1 October 2012
Online at http://www.njp.org/
doi:10.1088/1367-2630/14/10/105001

Abstract. Feasible tomography schemes for large particle numbers must possess, besides an appropriate data acquisition protocol, an efficient way to reconstruct the density operator from the observed finite data set. Since state reconstruction typically requires the solution of a nonlinear large-scale optimization problem, this is a major challenge in the design of scalable tomography schemes. Here we present an efficient state reconstruction scheme

[9] Author to whom any correspondence should be addressed.

for permutationally invariant quantum state tomography. It works for all common state-of-the-art reconstruction principles, including, in particular, maximum likelihood and least squares methods, which are the preferred choices in today's experiments. This high efficiency is achieved by greatly reducing the dimensionality of the problem employing a particular representation of permutationally invariant states known from spin coupling combined with convex optimization, which has clear advantages regarding speed, control and accuracy in comparison to commonly employed numerical routines. First prototype implementations easily allow reconstruction of a state of 20 qubits in a few minutes on a standard computer.

Contents

1. Introduction **2**
2. Background **4**
2.1. Permutationally invariant tomography 4
2.2. Statistical state reconstruction 5
3. Method **6**
3.1. Reduction of the dimensionality 6
3.2. Optimization 9
4. Examples **11**
5. Details **13**
5.1. Reduction 13
5.2. Convex optimization 17
5.3. Additional tools 19
6. Conclusion and outline **22**
Acknowledgments **23**
References **23**

1. Introduction

Full information about the experimental state of a quantum system is naturally highly desirable because it enables one to determine the mean value of each observable and thus also of every other property of the quantum state. Abstractly, such a complete description is given, for example, by the density operator, a positive semidefinite matrix ρ with unit trace. Quantum state tomography [1] refers to the task to determine the density operator for a previously unknown quantum state by means of appropriate measurements. Via the respective outcomes, more and more information about the true state generating the data is collected up to the point where this information uniquely specifies the particular state. Quantum state tomography has successfully been applied in many experiments using different physical systems, including trapped ions [2] or photons [3], as prominent instances.

Unfortunately, tomography comes with a very high price due to the exponential scaling of the number of parameters required to describe composed quantum systems. For an N-qubit system the total number of parameters of the associated density operator is $4^N - 1$ and any standard tomography protocol is naturally designed to determine all these variables. The

most common scheme used in experiments [4] consists of locally measuring in the basis of all different Pauli operators and requires an overall amount of 3^N different measurement settings with 2^N distinct outcomes each. Other schemes, e.g. using an informationally complete measurement [5] locally, would require just one setting but, nevertheless, the statistics for 4^N give different outcomes. Hence the important figure of merit to compare different methods is given by the combination of settings and outcomes.

For such a scaling, the methods rapidly become intractable, already for present state-of-the-art experiments: recording, for example, the data of 14 trapped ions [6], currently the record for quantum registers, would require about 150 days, although 100 measurement outcomes can be collected for a single setting in only about 3 s. In photonic experiments this scaling is even worse because count rates are typically much lower; for example, in recent eight-photon experiments [7, 8], a coincidence of single click events occurs only on the order of minutes; hence, it would require about 7 years to collect an adequate data set. This directly shows that more sophisticated tomography techniques are mandatory.

New tomography protocols equipped with better scaling behaviour exploit the idea that the measurement scheme is explicitly optimized only for particular kinds of states rather than for all possible ones. If the true unknown state lies within this designed target class, then full information about the state can be obtained with much less effort, and if the underlying density operator is not a member, then a certificate signals that tomography is impossible in this given case. Recent results along this direction include tomography schemes designed for states with a low rank [9–11], particularly for important low rank states such as matrix product [12, 13] or multi-scale entanglement renormalization ansatz states [14]. Other schemes include a tomography scheme based on information criteria [15, 16] or—the topic of this paper—states with permutation invariance [17].

However, it needs to be stressed that in any real experiment all these tomography schemes must cope up with another, purely statistical challenge: since only a finite number of measurements can be carried out in any experiment, one cannot access the true probabilities predicted by quantum mechanics $p_k = \mathrm{tr}(\rho_{\mathrm{true}} M_k)$, operator M_k describing the measurements, but merely relative frequencies f_k. Although these deviations might be small, the approximation $f_k \approx p_k$ causes severe problems in the actual state reconstruction process, i.e. the task to determine the density operator from the observed data. If one naïvely uses the frequencies according to Born's rule $f_k = \mathrm{tr}(\hat{\rho}_{\mathrm{lin}} M_k)$ and solves for the unknown operator $\hat{\rho}_{\mathrm{lin}}$, then, apart from possible inconsistencies in the set of linear equations, the reconstructed operator $\hat{\rho}_{\mathrm{lin}} \not\geqslant 0$ is often not a valid density operator anymore because some of its eigenvalues are negative. Hence, in such cases, this reconstruction called linear inversion delivers an unreasonable answer. It should be kept in mind that the inconsistencies can also be due to systematic errors, e.g. if the true measurements are aligned wrongly relative to the respective operator representation, but such effects are typically ignored [18].

Therefore, statistical state reconstruction relies on principles other than linear inversion. In general, these methods require the solution of a nonlinear optimization problem, which is much harder to solve than just a system of linear equations. For large system sizes this becomes, besides the exponential scaling of the number of settings and outcomes, a second major problem, again due to the exponential scaling of the number of parameters of the density operator. In fact for the current tomography record of eight ions in a trap [2], this actual reconstruction took even longer than the experiment itself (one week versus a couple of hours). Hence, feasible quantum state tomography schemes for large systems must, in addition to an efficient

measurement procedure, also possess an efficient state reconstruction algorithm; otherwise they are not scalable.

In this paper, we develop a scalable reconstruction algorithm for the proposed permutationally invariant tomography scheme [17]. It works for common reconstruction principles, including, among others, maximum likelihood and least squares methods. This scheme becomes possible once more by taking advantage of the symmetry of permutationally invariant states, which provides an efficient and operational way to store, characterize and even process those states. This method enables a large dimension reduction in the underlying optimization problem such that it gets into the feasible regime. The final low dimensional optimization is performed via nonlinear convex optimization which offers a great advantage in contrast to commonly used numerical routines, in particular regarding numerical stability and accuracy. Already a first prototype implementation of this algorithm allows state reconstruction for 20 qubits in a few minutes on a standard desktop computer.

The outline of the paper is as follows. Section 2 summarizes the background on permutationally invariant tomography and on statistical state reconstruction. The key method is explained in section 3 and highlighted via examples in section 4, which are generated by our current implementation. Section 5 collects all the technical details: the mentioned toolbox, more notes about convex optimization, additional information about the pretest or certificate and the measurement optimization, both addressed for large qubit numbers. Finally, in section 6 we conclude and provide an outlook on directions for further research.

2. Background

2.1. Permutationally invariant tomography

Permutationally invariant tomography has been introduced as a scalable reconstruction protocol for multi-qubit systems in [17]. It is designed for all states of the system that remain invariant under all possible interchanges of its different particles. Abstractly, such a permutationally invariant state ρ_{PI} of N qubits can be expressed in the form

$$\rho_{\mathrm{PI}} = [\rho]_{\mathrm{PI}} = \frac{1}{N!} \sum_{p \in S_N} V(p)\, \rho\, V(p)^{\dagger}, \tag{1}$$

where $V(p)$ is the unitary operator which permutes the N different subsystems according to the particular permutation p and the summation runs over all possible elements of the permutation group S_N. Many important states, such as the Greenberger–Horne–Zeilinger states or Dicke states, fall within this special class.

As shown in [17], full information on such states can be obtained by using in total $\binom{N+2}{N} = (N^2+3N+2)/2$ different local binary measurement settings, while for each setting only the count rates of $(N+1)$ different outcomes need to be registered. This finally leads to a cubic scaling in contrast to the exponential scaling of standard tomography schemes.

The measurement strategy that attains this number runs as follows: each setting is described by a unit vector $\hat{a} \in \mathbb{R}^3$ which defines associated eigenstates $|i\rangle_a$ of the trace-less operator $\hat{a} \cdot \vec{\sigma}$. Each part locally measures in this basis and registers the outcomes '0' or '1', respectively. The permutationally invariant part can be reconstructed from the collective outcomes, i.e. only the number of '0' or '1' results at the different parties matters but not the individual site information.

The corresponding coarse-grained measurements are given by

$$M_k^a = \sum_{p'} V(p')|0\rangle_a\langle 0|^{\otimes k} \otimes |1\rangle_a\langle 1|^{\otimes N-k} V(p')^\dagger, \tag{2}$$

$$= \binom{N}{k} \left[|0\rangle_a\langle 0|^{\otimes k} \otimes |1\rangle_a\langle 1|^{\otimes N-k}\right]_{\mathrm{PI}} \tag{3}$$

with $k = 0, \ldots, N$, and where the summation p' is over all permutations that give distinct terms. In total, one needs the above stated number of different settings $\hat{a}$. These settings can be optimized in order to minimize the total variance which provides an advantage in contrast to random selection.

In addition to this measurement strategy there is also a pretest which estimates the 'closeness' of the true, unknown state with respect to all permutationally invariant states from just a few measurement results [17]. This provides a way to test in advance whether permutationally invariant tomography is a good method for the unknown state.

Restricting oneself to the permutationally invariant part of a density operator has already been discussed in the literature: for example, for spins in a Stern–Gerlach experiment [19] or in terms of the polarization density operator [20–22]. Here, due to the restricted class of possible measurements, only the permutationally invariant part of, in principle distinguishable, particles is accessible [21, 22]. This is a strong conceptual difference compared to permutationally invariant tomography where one intentionally constrains oneself to this invariant part. Nevertheless, it should be emphasized that the employed techniques are similar.

2.2. *Statistical state reconstruction*

Since standard linear inversion of the observed data typically results in unreasonable estimates as explained in the introduction, one employs other principles for actual state reconstruction. In general, one uses a certain fit function $F(\rho)$ that penalizes deviations between the observed frequencies f_k and the true probabilities predicted by quantum mechanics $p_k(\rho) = \mathrm{tr}(\rho M_k)$ if the state of the system is ρ. The reconstructed density operator $\hat{\rho}$ is then given by the (often unique) state that minimizes this fit function,

$$\hat{\rho} = \arg \min_{\rho \geqslant 0} F(\rho); \tag{4}$$

hence the reconstructed state is precisely the one which best fits the observed data. Since the optimization explicitly restricts itself to physical density operators this ensures validity of the final estimate $\hat{\rho}$ in contrast to linear inversion. Depending on the functional form of the fit function, different reconstruction principles are distinguished. A list of the most common choices is provided in table 1.

The presumably best-known and most widely employed method is called the maximum likelihood principle [23]. Given a set of measured frequencies, f_k, the maximum likelihood state, $\hat{\rho}_{\mathrm{ml}}$, is exactly the one with the highest probability of generating these data. Other common fit functions, usually employed in photonic state reconstruction, are different variants of least squares [4, 24] that originate from the likelihood function using Gaussian approximations for the probabilities. There, this is often called maximum likelihood principle, but we distinguish these, indeed different, functions here explicitly. Typically, the weights in the least squares function are set to $w_k = 1/f_k$ because f_k represents an estimate of the variance in a multinomial distribution,

Table 1. Common reconstruction principles and their corresponding fit functions $F(\rho)$ used in the optimization given by equation (4); see text for further details.

Reconstruction principle	Fit function $F(\rho)$
Maximum likelihood [23]	$-\sum_k f_k \log[p_k(\rho)]$
Least squares [24]	$\sum_k w_k[f_k - p_k(\rho)]^2$, $w_k > 0$
Free least squares [4]	$\sum_k 1/p_k(\rho)[f_k - p_k(\rho)]^2$
Hedged maximum likelihood [25]	$-\sum_k f_k \log[p_k(\rho)] - \beta \log[\det(\rho)]$, $\beta > 0$

cf the free least squares principle. However, this leads to a strong bias if the count rates are extremal; for example, if one of the outcomes is never observed, this method naturally leads to difficulties. A method to circumvent this is given by the free least squares function [4] or using improved error analysis for rare events. Let us stress that all these principles have the property that if linear inversion delivers a valid estimate $\hat{\rho}_{\text{lin}} \geqslant 0$, it is also the estimate given by these reconstruction principles[10].

Finally, hedged maximum likelihood [25] represents a method that circumvents low rank state estimates. Via this, one obtains more reasonable error bars using parametric bootstrapping methods [26]; for other error estimates, see the recently introduced confidence regions for quantum states [27, 28]. In principle, many more fit functions are possible, such as generic loss functions [29], but considering these is beyond the scope of this work.

3. Method

From the above, it is apparent that permutationally invariant state reconstruction requires the solution of

$$\hat{\rho}_{\text{PI}} = \arg \min_{\rho_{\text{PI}} \geqslant 0} F(\rho_{\text{PI}}) \tag{5}$$

for the preferred fit function. This large-scale optimization becomes feasible along the following lines.

Firstly, one reduces the dimensionality of the underlying optimization problem because one cannot work with full density operators anymore. This requires an operational way to characterize permutationally invariant states $\rho_{\text{PI}} \geqslant 0$ over which the optimization is performed, and additionally, demands an efficient way to compute probabilities $p_k(\rho_{\text{PI}})$ which appear in the fit function. Secondly, one needs a method for performing the final optimization. We employ convex optimization for this.

3.1. Reduction of the dimensionality

This reduction relies on an efficient toolbox to handle permutationally invariant states, which exploits the particular symmetry. Here we will explain this method and the final structure; for

[10] For the least squares fit functions this follows because $F = 0$ in this case and clearly $F \geqslant 0$ for those functions. In the case of the likelihood it follows from positivity of the classical relative entropy between probability distributions.

more details see section 5.1. These techniques are well proven and established; we employ and adapt them here for the permutationally invariant tomography scheme such that we finally reach state reconstruction of larger qubits.

The methods of this toolbox are obtained via the concept of spin coupling that describes how individual, distinguishable spins couple to a combined system if they become indistinguishable. Since we deal with qubits we only need to focus on spin-1/2 particles. In the simplest case, two spin-1/2 particles can couple to a spin-1 system, called the triplet, if both spins are aligned symmetric, or to a spin-0 state, the singlet, if the spins are aligned anti-symmetric. Abstractly, this can be denoted as $\mathbb{C}^2 \otimes \mathbb{C}^2 = \mathbb{C}^3 \oplus \mathbb{C}^1$. If one considers now three spins, then of course all spins can point in the same direction giving a total spin-3/2 system. There is also a spin-1/2 system possible if two particles form already a spin-0 and the remaining one stays untouched. This can be achieved, however, by more than one possibility, in fact by two inequivalent choices[11], and is expressed by $\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2 = \mathbb{C}^4 \oplus (\mathbb{C}^2 \otimes \mathbb{C}^2)$.

This scheme can be extended to N spin-1/2 particles to obtain the following decomposition of the total Hilbert space,

$$\mathcal{H} = (\mathbb{C}^2)^{\otimes N} = \bigoplus_{j=j_{\min}}^{N/2} \mathcal{H}_j \otimes \mathcal{K}_j, \tag{6}$$

where the summation runs over different total spin numbers $j = j_{\min}, j_{\min}+1, \ldots, N/2$ starting from $j_{\min} \in \{0, 1/2\}$, depending on whether N is even or odd. Here, $\mathcal{H}_j$ are called the spin Hilbert spaces with dimensions $\dim(\mathcal{H}_j) = 2j+1$, while $\mathcal{K}_j$ are referred to as multiplicative spaces that account for the different possibilities to obtain a spin-j state. They are, generally, of a much larger dimension, cf equation (22).

Permutationally invariant states have a simpler form on this Hilbert space decomposition, namely

$$\rho_{\mathrm{PI}} = \bigoplus_{j=j_{\min}}^{N/2} p_j \rho_j \otimes \frac{\mathbb{1}}{\dim(\mathcal{K}_j)}, \tag{7}$$

with density operators ρ_j called spin states and according probabilities p_j. Thus a permutationally invariant density operator only contains non-trivial parts on the spin Hilbert spaces while carrying no information on the multiplicative spaces. Note, further, that there are no coherences between different spin states. This means that any permutationally invariant state can be parsed into a block structure as schematically represented in figure 1. The main diagonal is made up of unnormalized spin states $\tilde{\rho}_j = p_j \rho_j / \dim(\mathcal{K}_j)$, which appear several times, the number being equal to the dimension of the corresponding multiplicative space. This block-decomposition represents a natural way to treat permutationally invariant states and has, for example, already been employed in the aforementioned related works on permutationally invariant tomography [19–22] but also in other contexts [30–33].

This structure shows that if we work with permutationally invariant states, we do not need to consider the full density operator but rather that it is sufficient to deal only with this ensemble

[11] More precisely, all states of the form $|\psi\rangle = V(p)|\psi^-\rangle \otimes |0\rangle$, with p being any possible permutation, are states of total spin $j = 1/2$ and projection $m = 1/2$ to the collective spin operators $J_i = \sum_{n=1}^{3} \sigma_{i;n}/2$, $\sigma_{i;n}$ being the corresponding Pauli operator on qubit n. However, as can be checked, these states only span a two-dimensional subspace.

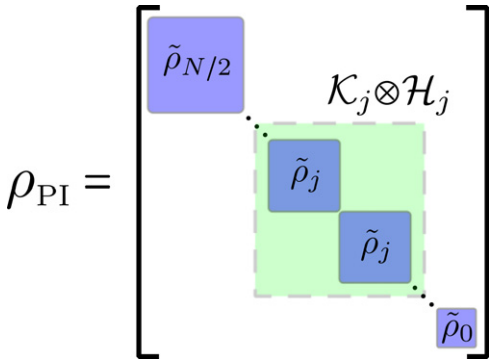

Figure 1. Block decomposition for a generic permutationally invariant state as given by equation (7) with $\tilde{\rho}_j = p_j \rho_j / \dim(\mathcal{K}_j)$. Note that we interchanged the spin- and multiplicative spaces.

of spin states. Therefore we identify from now on

$$\rho_{\mathrm{PI}} \Longleftrightarrow p_j \rho_j, \ j = j_{\min}, j_{\min} + 1, \ldots, N/2. \tag{8}$$

This provides already an efficient way to store and to visualize such states. More importantly, it also provides an operational way to characterize valid states, since any permutationally invariant operator ρ_{PI} represents a true state if and only if all these spin operators ρ_j are density operators and p_j a probability distribution. This is in contrast to the generalized Bloch vector employed in the original proposal of permutationally invariant tomography [17] given by equation (52), which also provides efficient storage and processing of permutationally invariant states, but where Bloch vectors of physical states are not as straightforward to characterize.

Via this identification one can demonstrate once more the origin of the cubic scaling of the permutationally invariant tomography scheme. The largest spin state is of dimension $N+1$ which requires parameters of the order of N^2 for characterization. Since one has of the order of N of these states, this results in a cubic scaling.

Fixing the ensemble of spin states as parameterization, it is now necessary to obtain an efficient procedure to compute the probabilities $p_k^a(\rho_{\mathrm{PI}})$ for the optimized measurement scheme. This is achieved as follows: let us first stress that a similar block decomposition to that given by equation (7) holds for all permutationally invariant operators. Hence also the measurements M_k^a given by equation (2) can be cast into this form. Using the convention

$$M_k^a = \bigoplus_{j=j_{\min}}^{N/2} M_{k,j}^a \otimes \mathbb{1} \tag{9}$$

leads to

$$\mathrm{tr}(\rho_{\mathrm{PI}} M_k^a) = \sum_{j=j_{\min}}^{N/2} p_j \mathrm{tr}(\rho_j M_{k,j}^a). \tag{10}$$

Therefore the problem is shifted to the computation of the spin-j operators $M_{k,j}^a$ for each setting $\hat{a}$. As we show in proposition 5.2 below, using the idea that measurements can be transformed into each other by a local operation $U_a|i\rangle = |i\rangle_a$, this provides the relation

$$M_{k,j}^a = W_j^a M_{k,j}^{e_3} W_j^{a,\dagger}. \tag{11}$$

Here $M^{e_3}_{k,j}$ corresponds to the measurement in the standard basis (that need to be computed once) and W^a_j is a unitary transformation determined by the rotation U_a. This connection is given by

$$U_a = \exp\left(-\mathrm{i}\sum_l t_l \sigma_l/2\right) \Longrightarrow W^a_j = \exp\left(-\mathrm{i}\sum_l t_l S_{l,j}\right), \tag{12}$$

where $S_{l,j}$ stands for the spin operators in dimension $2j+1$. This finally provides an efficient way to compute probabilities.

3.2. Optimization

As a second step one still needs to cope up with the optimization itself. Although there are different numerical routines for statistical state reconstruction like maximum likelihood [34] or least squares [4, 35], we prefer nonlinear convex optimization [36] to obtain the final solution. Quantum state reconstruction problems are known to be convex [35, 37], but convex optimization has hardly been used for this task. However, convex optimization has several advantages: first of all it is a systematic approach that works for any convex fit function, including maximum likelihood and least squares. In contrast to other algorithms such as the fixed-point algorithm proposed in [34], it gives a precise stopping condition via an appropriate error control (see, however, [38]) and still exploits all the favourable, convex, structure in comparison with re-parameterization ideas as in [4]. Moreover, it is guaranteed to find the global optimum and the accuracy obtained is typically much higher than that obtained with other methods.

Quantum state reconstruction as defined via equation (4) can be formulated as a convex optimization problem as follows: all fit functions listed in table 1 are convex on the set of states. Via a linear parameterization of the density operator $\rho(x) = \mathbb{1}/\dim(\mathcal{H}) + \sum x_i B_i$, using an appropriate operator basis B_i such that normalization is fulfilled directly, the required optimization problem becomes

$$\begin{aligned} &\min_x F[\rho(x)] \\ &\text{s.t.}\ \ \rho(x) = \frac{\mathbb{1}}{\dim(\mathcal{H})} + \sum_i x_i B_i \geqslant 0, \end{aligned} \tag{13}$$

with a convex objective function $F(x) = F[\rho(x)]$ and a linear matrix inequality as the constraint, i.e. precisely the structure of a nonlinear convex optimization problem [36]. For permutationally invariant states, one uses $\rho(x) = \oplus_j \bar{\rho}_j(x)$ with $\bar{\rho}_j = p_j \rho_j$ by using an appropriate block-diagonal operator basis B_i; therefore we continue this discussion with the more general form.

The optimization given by equation (13) can be performed, for instance, with the help of a barrier function [36][12]. Rather than considering the constrained problem, one solves the unconstrained convex task given by

$$\min_x F[\rho(x)] - t \log[\det \rho(x)], \tag{14}$$

[12] Let us stress that both least squares options can be parsed into a simpler convex problem, called a semidefinite program, as shown in, for instance, [24, 35], but that this does not work with the true maximum likelihood function to our best knowledge.

where the constraint is now directly included in the objective function. This so-called barrier term acts precisely as its name suggests: if one tries to leave the strictly feasible set, i.e. all parameters x that satisfy $\rho(x) > 0$, one always reaches a point where at least one of the eigenvalues vanish. Since the barrier term is large within this neighbourhood, in fact singular at the crossing, it penalizes points close to the border and thus ensures that one searches for an optimum well inside the region where the constraint is satisfied. The penalty parameter $t > 0$ plays the role of a scaling factor. If it becomes very small the effect of the barrier term becomes negligible within the strictly feasible set and only remains at the border. Therefore a solution of equation (14) with a very small value of t provides an excellent approximation to the real solution. As shown in section 5.2 this statement can be made more precise by

$$F[\rho(x^t_{\mathrm{sol}})] - F[\rho(x_{\mathrm{sol}})] \leqslant t \dim(\mathcal{H}), \tag{15}$$

which follows from convexity and which relates the true solution x_{sol} of the original problem given by equation (13) to the solution x^t_{sol} of the unconstrained problem with penalty parameter t. This condition represents the above-mentioned error control and serves as a stopping condition, i.e. as a quantitative error bound for a given t. Note that for permutationally invariant tomography $\dim(\mathcal{H})$ is not the dimension of the true N-qubit Hilbert space but instead the dimension of $\rho(x) = \oplus_j \bar{\rho}_j(x)$, i.e. $\sum_{j=j_{\min}}(2j+1) = (N+1)(N+2j_{\min}+1)/4$ which increases only quadratically.

Small penalty parameters are approached by an iterative process: for a given starting point x^n_{start} and a certain value of the parameter $t_n > 0$, one solves equation (14). Its solution will be the starting point for $x^{n+1}_{\mathrm{start}} = x^n_{\mathrm{sol}}$ for the next unconstrained optimization with a lower penalty parameter $t_{n+1} < t_n$. As the starting point for the first iteration, we employ $t_0 = 1$ and the point x^0_{start} corresponding to the totally mixed state. This procedure is repeated until one has reached small enough penalty parameters. The penalty parameter is decreased step-wise. Then each unconstrained problem can be solved very efficiently since one starts already quite close to the true solution.

Let us point out that via the above-mentioned barrier method, one additionally obtains solutions to the hedged state reconstruction with $\beta = t$ since the unconstrained problem given by equation (14) is precisely the fit function of the hedged version of table 1.

Finally, for comparison purposes, we would also like to mention the iterative fixed-point algorithm of [39]; for a modification see [40]. It is designed for maximum likelihood estimation and is straightforward to implement, since it only requires matrix multiplication; however, it has deficits regarding control and accuracy. For permutationally invariant tomography the algorithm can be adapted as follows: given a valid iterate ρ^n_{PI} characterized by the ensemble of spin states $\bar{\rho}^n_j = p^n_j \rho^n_j$, one evaluates the probabilities $p^a_k(\rho^n_{\mathrm{PI}})$ using equation (10). Next, one computes the operators

$$R^n_j = \sum_{a,k} \frac{f^a_k}{p^a_k(\rho^n_{\mathrm{PI}})} M^a_{k,j}, \tag{16}$$

which determine the next iterate $\bar{\rho}^{n+1}_j = R^n_j \bar{\rho}^n_j R^{n\dagger}_j / \mathcal{N}$ with $\mathcal{N} = \sum_j \mathrm{tr}(R^n_j \bar{\rho}^n_j R^{n\dagger}_j)$. This iteration is started, for example, from the totally mixed state and repeated until a sufficiently good solution is obtained.

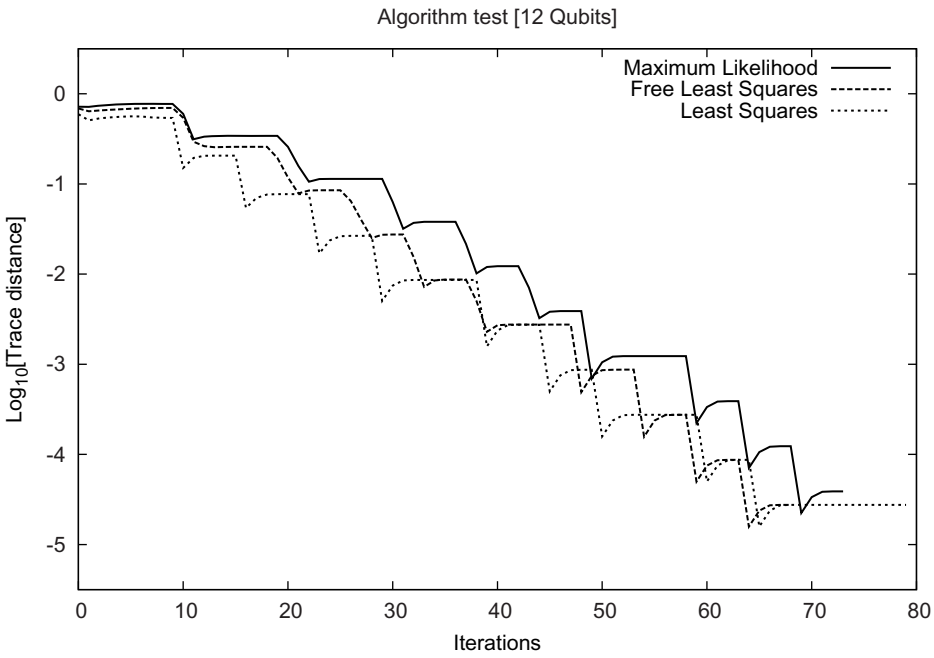

Figure 2. Trace distance between the analytic solution and the estimate after n algorithm steps for the three most common reconstruction principles from table 1.

4. Examples

The two methods from the previous section are employed in a prototype implementation under MATLAB, which already enables state reconstruction of about 20 qubits on a standard desktop computer.

The current algorithm is tested along the following lines: for a randomly generated permutationally invariant state $\rho_{\mathrm{PI}}^{\mathrm{true}}$ we compute the true probabilities $p_{k,\mathrm{true}}^{a}$ for the chosen measurement settings. Rather than sampling we set the observed frequencies equal to this distribution, i.e. $f_k^a = p_{k,\mathrm{true}}^a$. In this way linear inversion would return the original state; hence also every other reconstruction principle from table 1 has this state as the solution. We now start the algorithm and compare the trace distance between the analytic solution $\rho_{\mathrm{PI}}^{\mathrm{true}}$ and the state after n iterations ρ_{PI}^{n}. This distance $\frac{1}{2}\mathrm{tr}|\rho_{\mathrm{PI}}^{\mathrm{true}} - \rho_{\mathrm{PI}}^{n}|$ quantifies the probability with which the two states, the true analytic solution and the iterate after n steps in the algorithm, could be distinguished [41].

A typical representative of this process is depicted in figure 2 for 12 qubits using optimized settings. The randomly generated state $\rho_{\mathrm{PI}}^{\mathrm{true}}$ was chosen to lie at the boundary of the state space since such rank-reduced solutions better resemble the case of state reconstruction of real data. More precisely, each spin state of the true density operator is given by a pure state $\rho_j^{\mathrm{true}} = |\psi_j\rangle\langle\psi_j|$ chosen according to the Haar measure, while the p_j are selected by the symmetric Dirichlet distribution with concentration measure $\alpha = 1/2$ [42]. As is apparent the algorithm behaves similarly for all three reconstruction principles and rapidly obtains a good solution after about 70 iterations. The steps in this plot are points where the penalty parameter is reduced by a factor of 10 starting from $t = 1$ and decreased down to $t = 10^{-10}$. The slight rise

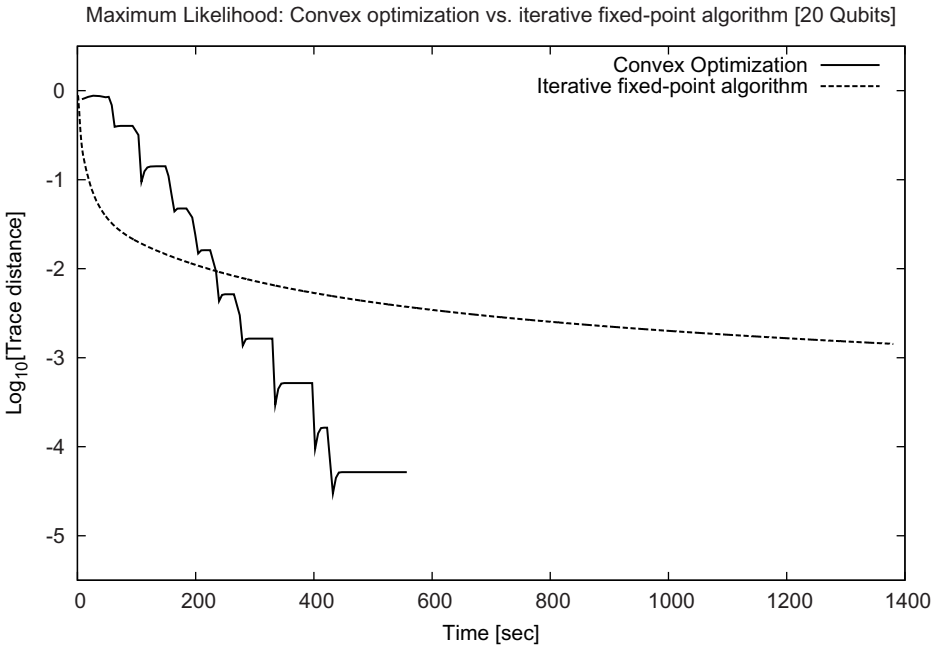

Figure 3. Comparison of the maximum likelihood principle between convex optimization and the iterative fixed-point algorithm for the described testing procedure with respect to accuracy and algorithm time.

after these points comes from the fact that we plot the trace distance and not the actual function (fit function plus penalty term) that is minimized.

Figure 3 shows a similar comparison for the maximum likelihood reconstruction of 20 qubits but now plotted versus algorithm time[13]. For comparison we include the performance of the iterative fixed-point algorithm, which requires much more iterations in general (3000 in this case versus about 90 for convex optimization). Let us emphasize that a similar behaviour between these two algorithms appears also for smaller qubit numbers. As one can see, convex optimization delivers a faster and in particular more accurate solution. In contrast, the iterative fixed-point algorithm shows a bad convergence rate although it initially starts off better. This was one of the main reasons for us to switch to convex optimization.

The current algorithm times of this test are listed in table 2 which are averaged over 50 randomly generated states. Thus already this prototype implementation enables state reconstruction of larger qubit numbers in moderate times. The small time difference between reconstruction principles is because least squares as a quadratic fit function provides some advantages in the implementation. More details of this difference are given in section 5.

Table 2 also contains the algorithm times for reconstructions using simulated frequencies $f_k^a = n_k^a / N_\mathrm{r}$. For each setting they are deduced from the count rates n_k^a sampled from a multinomial distribution using the true event distribution $p_{k,\mathrm{true}}^a$ and $N_\mathrm{r} = 1000$ repetitions. The true probabilities correspond to the same states as already employed in the algorithm test. From the table, one observes that state reconstruction for data with count statistics requires only slightly more time than the algorithm test with the correct probabilities. We attribute this to the fact that a few more iterations are typically required in order to achieve the desired accuracy.

[13] All simulations were performed on an Intel Core i5-650, 3.2 GHz, 8 GB RAM using MATLAB 7.12.

Table 2. Current performance of the convex optimization algorithm on the described test procedure and on frequencies from simulated experiments; free least squares provides similar results to the maximum likelihood principle.

	$N = 8$	$N = 12$	$N = 16$	$N = 20$
Maximum likelihood				
Algorithm test	8.5 s	47 s	2.7 min	9.2 min
Simulated experiment	9.2 s	48 s	2.9 min	9.3 min
Least squares				
Algorithm test	8.4 s	39 s	2.5 min	6 min
Simulated experiment	9.2 s	43 s	2.7 min	6.7 min

Finally, let us perform the reconstruction of a simulated experiment of $N = 14$ qubits. Suppose that one intends to create a Dicke state $|D_{k,N}\rangle$ as given by equation (26) for some k and N, but that the preparation suffers from some imperfections such that at best one can prepare states of the form

$$\rho_{\text{dicke mix}} = \sum_{l=0}^{N} \binom{N}{l} p^l (1-p)^{N-l} |D_{l,N}\rangle\langle D_{l,N}|, \tag{17}$$

where $p = 0.6$ characterizes some asymmetry. As the true state prepared in the experiment we now model some further imperfection in the form of an additional small misalignment $U^{\otimes N}$, with $U = \exp(-\mathrm{i}\theta\sigma_y/2)$, $\theta = 0.2$, and some permutationally invariant noise σ_{PI} (chosen via the aforementioned method but using Hilbert–Schmidt instead of the Haar measure), i.e.

$$\rho_{\text{true}} = 0.6\, U^{\otimes N} \rho_{\text{dicke mix}} U^{\dagger\otimes N} + 0.4\, \sigma_{\text{PI}}. \tag{18}$$

The frequencies are obtained via sampling from the state given by equation (18) using intentionally only $N_{\text{r}} = 200$ repetitions per setting (to see some differences). Finally, we reconstruct the state according to the maximum likelihood principle. Figure 4 shows the tomography bar plots of one of these examples for the largest spin state $p_j\rho_j$, $j = N/2 = 7$ for both states. Although this state might be artificial, this example should highlight once more that this state reconstruction algorithm works also for realistic data and for qubit sizes where clearly any non-tailored state reconstruction scheme would fail. Moreover, it demonstrates that the spin ensemble $p_j\rho_j$ represents a very convenient graphical representation of such states (compared to a $2^{14} \times 2^{14}$ matrix in this case).

5. Details

5.1. Reduction

Let us first give more details regarding the reduction step. This starts by recalling a group theoretic result summarized in the next section, which is then used to show how the stated simplifications with respect to states and measurements are obtained.

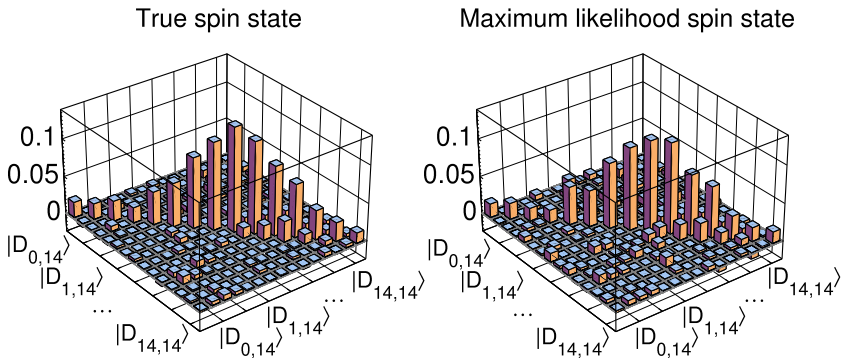

Figure 4. The real part of the true and reconstructed (according to maximum likelihood) largest spin state ensemble $p_j\rho_j$, $j = N/2 = 7$ using the optimal measurement setting. The basis is given by the Dicke basis $|D_{k,14}\rangle$, cf equation (26).

5.1.1. Background. Consider the following two unitary representations defined on the N qubit Hilbert space: the permutation or symmetric group $V(p)$ which is defined by their action onto a standard tensor product basis by $V(p)|i_1, \ldots, i_N\rangle = |i_{p^{-1}(1)}, \ldots, i_{p^{-1}(N)}\rangle$ according to the given permutation p, and the tensor product representation $W(U) = U^{\otimes N}$ of the special unitary group. A result known as the Schur–Weyl duality [43, 44] states that the action of these two groups is dual, which means that the total Hilbert space can be divided into blocks on which the two representations commute. More precisely, one has

$$(\mathbb{C}^2)^{\otimes N} = \bigoplus_{j=j_{\min}}^{N/2} \mathcal{H}_j \otimes \mathcal{K}_j, \tag{19}$$

$$V(p) = \bigoplus_{j=j_{\min}}^{N/2} \mathbb{1} \otimes V_j(p), \tag{20}$$

$$W(U) = \bigoplus_{j=j_{\min}}^{N/2} W_j(U) \otimes \mathbb{1}. \tag{21}$$

Here V_j and W_j are the respective irreducible representations, and $j_{\min} \in \{0, 1/2\}$ depending on whether N is even or odd. The dimensions of the appearing Hilbert spaces are $\dim(\mathcal{H}_j) = 2j + 1$ and

$$\dim(\mathcal{K}_j) = \binom{N}{N/2 - j} - \binom{N}{N/2 - j - 1} \tag{22}$$

for all $j < N/2$ and $\dim(\mathcal{K}_{N/2}) = 1$. Let us note that equation (20) already ensures the block-diagonal structure of permutationally invariant operators, while equation (21) becomes important for the measurement computation.

A basis of the Hilbert space $\mathcal{H}_j \otimes \mathcal{K}_j$ is formed by the spin states $|j, m, \alpha\rangle = |j, m\rangle \otimes |\alpha_j\rangle$ with $m = -j, -j+1, \ldots, j$ and $\alpha_j = 1, \ldots, \dim(\mathcal{K}_j)$. These are obtained by starting with the states having the largest spin number $m = j$, which are given by

$$|j, j, 1\rangle = |0\rangle^{\otimes 2j} \otimes |\psi^-\rangle^{\otimes N/2-j}, \tag{23}$$

$$|j, j, \alpha\rangle = \sum_p c_{j,p} V(p) |j, j, 1\rangle, \tag{24}$$

for all $\alpha \geqslant 2$. The coefficients $c_{j,p}$ must ensure that the states $|j, j, \alpha\rangle$ are orthogonal; otherwise their choice is completely free since the detailed structure of different α's is not important. The full basis is obtained by subsequently applying the ladder operator $J_- = \sum_{n=1}^{N} \sigma_{-;n}$ to decrease the spin number m. Here $\sigma_{-;n}$ refers to the operator with $\sigma_- = (\sigma_x - i\sigma_y)/2$ on the nth qubit and identity on the rest. Thus in total the basis becomes

$$|j, m, \alpha\rangle = \mathcal{N} J_-^{j-m} |j, j, \alpha\rangle, \tag{25}$$

with appropriate normalizations $\mathcal{N}$. Note that the subspace corresponding to the highest spin number $j = N/2$ is also called the symmetric subspace, which contains many important states such as the Greenberger–Horne–Zeilinger or Dicke states, which using the spin states read as

$$|\mathrm{GHZ}\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle^{\otimes N} + |1\rangle^{\otimes N}\right) = \frac{1}{\sqrt{2}}(|N/2, N/2, 1\rangle + |N/2, -N/2, 1\rangle), \tag{26}$$

$$|D_{k,N}\rangle = \mathcal{N}\left[|1\rangle^{\otimes k} \otimes |0\rangle^{\otimes N-k}\right]_{\mathrm{PI}} = |N/2, N/2 - k, 1\rangle. \tag{27}$$

5.1.2. Permutationally invariant states and measurement operators. Let us now employ this result in order to derive a generic form for permutationally invariant states; we give the proof for completeness.

Proposition 5.1 (Permutationally invariant states). *Any permutationally invariant state of N qubits ρ_{PI} defined via equation (1) can be written as*

$$\rho_{\mathrm{PI}} = \bigoplus_{j=j_{\min}}^{N/2} p_j \rho_j \otimes \frac{\mathbb{1}}{\dim(\mathcal{K}_j)}; \tag{28}$$

hence it is fully characterized by the ensemble $p_j \rho_j$. Moreover, ρ_{PI} is a density operator if and only if all ρ_j are density operators and p_j a probability distribution.

Proof. The proposition follows using the representation $V(p)$ given by equation (20) in the definition of the states, cf equation (1), and then applying Schur's lemma [43, 44]. This lemma states that any linear operator A from $\mathcal{K}_j$ to $\mathcal{K}_i$ which commutes with all elements p of the group $V_i(p)A = AV_j(p)$ must either be zero if i and j label different irreducible representations or $A = c\mathbb{1}$ if they are unitarily equivalent. Since $A_{\mathrm{PI}} = 1/N! \sum_p V_i(p) A V_j(p)^\dagger$ fulfils this relation, one obtains

$$\frac{1}{N!} \sum_p V_i(p) A V_j(p)^\dagger = \delta_{ij} \operatorname{tr}(A) \frac{\mathbb{1}}{\dim(\mathcal{K}_j)}. \tag{29}$$

The normalization can be checked taking the trace on both sides. Adding appropriate identities provides

$$\frac{1}{N!} \sum_p \mathbb{1} \otimes V_i(p) B \mathbb{1} \otimes V_j(p)^\dagger = \delta_{ij} \operatorname{tr}_{\mathcal{K}_j}(B) \otimes \frac{\mathbb{1}}{\dim(\mathcal{K}_j)}, \tag{30}$$

where B should now be a linear operator from $\mathcal{H}_j \otimes \mathcal{K}_j$ to $\mathcal{H}_i \otimes \mathcal{K}_i$.

Finally, let P_j denote the projector onto $\mathcal{H}_j \otimes \mathcal{K}_j$, and using equation (30) yields

$$\rho_{\mathrm{PI}} = \frac{1}{N!}\sum_p V(p)\rho V(p)^\dagger = \sum_{i,i',j,j'} \frac{1}{N!}\sum_p P_i V(p) P_{i'}\rho P_j V(p)^\dagger P_{j'} \tag{31}$$

$$= \sum_{i,j}\left\{\frac{1}{N!}\sum_p P_i[\mathbb{1}\otimes V_i(p)]P_i\rho P_j[\mathbb{1}\otimes V_j(p)]^\dagger P_j\right\} \tag{32}$$

$$= \sum_{i,j} P_i\left[\delta_{ij}\,\mathrm{tr}_{\mathcal{K}_j}(P_i\rho P_j)\otimes\frac{\mathbb{1}}{\dim(\mathcal{K}_j)}\right]P_j \tag{33}$$

$$= \bigoplus_j \mathrm{tr}_{\mathcal{K}_j}(P_j\rho P_j)\otimes\frac{\mathbb{1}}{\dim(\mathcal{K}_j)}, \tag{34}$$

which provides the general structure.

The state characterization part follows because positivity of a block matrix is equivalent to positivity of each block. □

Next let us concentrate on the measurement part. Although the block decomposition follows already from the previous proposition, it is here more important to obtain an efficient computation of each measurement block for the selected setting.

Proposition 5.2 (Measurement operators). *The POVM elements M_k^a as defined in equation (2) for any local setting $\hat{a}\in\mathbb{R}^3$ can be decomposed as $M_k^a = \bigoplus_j M_{k,j}^a \otimes \mathbb{1}$ with*

$$M_{k,j}^a = W_j(U_a)M_{k,j}^{e_3}W_j(U_a)^\dagger. \tag{35}$$

The unitary is given by $W_j(U_a) = \exp(-\mathrm{i}\sum_l t_l S_{l,j})$ using the spin operators $S_{l,j}$ in dimension $2j+1$, while the coefficients t_l are determined by $U_a = \exp(-\mathrm{i}\sum_l t_l\sigma_l/2)$ which satisfies $\hat{a}\cdot\vec{\sigma} = U_a\sigma_z U_a^\dagger$. For the measurement in the standard basis $\hat{a}=\hat{e}_3$ one obtains

$$M_{k,j}^{e_3} = |j, N/2-k\rangle\langle j, N/2-k| \tag{36}$$

if $-j\leqslant N/2-k\leqslant j$ and zero otherwise.

Proof. The basic idea is to consider the measurement in an arbitrary local basis $\hat{a}$ by a rotation followed by the collective measurement in the standard basis. The block decomposition is obtained as follows:

$$M_k^a = U_a^{\otimes N}M_k^{e_3}U_a^{\dagger\otimes N} = W(U_a)\left[\bigoplus_j M_{k,j}^{e_3}\otimes\mathbb{1}\right]W(U_a)^\dagger \tag{37}$$

$$= \bigoplus_j W_j(U_a)M_{k,j}^{e_3}W_j(U_a)^\dagger\otimes\mathbb{1}. \tag{38}$$

The first step holds because U_a satisfies $|i\rangle_a\langle i| = U_a|i\rangle\langle i|U_a^\dagger$, while the block decomposition of the standard basis measurement $M_k^{e_3}$ is employed afterwards. In the last part one uses the tensor product representation given by equation (21).

Since one knows that W_j is irreducible it can be uniquely written in terms of its Lie algebra representation $\mathrm{d}W_j$ as $W_j(U_a) = W_j(\mathrm{e}^{-\mathrm{i}X}) = \mathrm{e}^{-\mathrm{i}\mathrm{d}W_j(X)}$, which is given by the spin operators in this case, i.e. $\mathrm{d}W_j(\sigma_l/2) = S_{l,j}$ [45].

Thus we are left to compute the measurement blocks $M^{e_3}_{k,j}$ for the standard basis. Note it is sufficient to evaluate $M^{e_3}_{k,j} \otimes |1_j\rangle\langle 1_j|$ such that one can employ the spin basis states $|j, m, 1\rangle$ as introduced in section 5.1.1. At first, note that $M^{e_3}_k$ exactly contains k projections onto $|0\rangle$, while each basis state $|j, m, 1\rangle$ possesses $(N/2+m)$ zeros. Therefore one obtains $M^{e_3}_{k,j}|j, m, 1\rangle \propto \delta_{k,N/2+m}|j, m, 1\rangle$. Since each POVM has to resolve the identity this is only possible if each $M^{e_3}_{k,j}$ is the stated rank-1 projector on the basis states. □

Finally one still needs to express $U_a = \exp(-\mathrm{i}\sum_l t_l\sigma_l/2)$ for the chosen setting $\hat{a} \in \mathbb{R}^3$. Since this can be related to a familiar rotation [45] these coefficients can be expressed as $t_l = (\theta\hat{n})_l$ via a rotation about an angle θ around the axis $\hat{n}$. Since this rotation should turn $\hat{e}_3$ into $\hat{a}$, its parameters are given by

$$\hat{n} = \frac{\hat{e}_3 \times \hat{a}}{\|\hat{e}_3 \times \hat{a}\|_2}, \tag{39}$$

$$\theta = \arccos(\hat{e}_3 \cdot \hat{a}), \tag{40}$$

and $\hat{n} = \hat{e}_1$ (or any other orthogonal vector) if $\hat{a} = \pm\hat{e}_3$.

5.2. *Convex optimization*

In this part, we collect some more details of the described convex optimization algorithm; for complete coverage see the book [36].

Each unconstrained optimization given by equation (14) is solved via a damped Newton algorithm. The minimization of $f(x) = F[\rho(x)] - t\log[\det\rho(x)]$ is obtained by an iterative process. In order to determine a search direction at a given iterate x^n, one minimizes the quadratic approximation

$$f(x^n + \Delta x) \approx f(x^n) + \nabla f(x^n)^T\Delta x + \tfrac{1}{2}\Delta x^T\nabla^2 f(x^n)\Delta x. \tag{41}$$

This reduces to solving a linear set of equations called the Newton equation

$$\nabla^2 f(x^n)\Delta x_{\mathrm{nt}} = -\nabla f(x^n), \tag{42}$$

which determines the search direction Δx_{nt}. The step length s for the next iterate $x^{n+1} = x^n + s\Delta x_{\mathrm{nt}}$ is chosen by a backtracking line search. Here one picks the largest $s = \max_{k\in\mathbb{N}}\beta^k$ with $\beta \in (0, 1)$ such that the iterate stays feasible $\rho(x^{n+1}) > 0$ and that the function value decreases sufficiently, i.e. $f(x^{n+1}) \leqslant f(x^n) + \alpha s\nabla f(x^n)^T\Delta x_{\mathrm{nt}}$ with $\alpha \in (0, 0.5)$. The process is stopped if one has reached an appropriate solution, which can be identified by $\|\nabla f(x^n)\|_2 \leqslant \epsilon$. If the initial point x_{start} is already sufficiently close to the true solution then the whole algorithm converges quadratically, i.e. the precision gets doubled at each step.

At this point, let us give the gradient and Hessian of the appearing functions. For the barrier term $\psi(x) = -\log[\det\rho(x)]$ restricted to the positive domain $\rho(x) = \mathbb{1}/\dim(\mathcal{H}) + \sum_i x_i B_i > 0$, one obtains [36]

$$\frac{\partial\psi(x)}{\partial x_i} = -\mathrm{tr}[\rho(x)^{-1}B_i], \tag{43}$$

$$\frac{\partial^2\psi(x)}{\partial x_j\partial x_i} = \mathrm{tr}[\rho(x)^{-1}B_j\rho(x)^{-1}B_i]. \tag{44}$$

Equation (44) shows that the Hessian of the penalty term $\nabla^2\psi(x) > 0$ is positive definite, such that $\psi(x)$ is indeed convex. The derivatives of the preferred fit function can be computed directly. For instance, using the likelihood function $F_{\mathrm{ml}}(x) = -\sum_k f_k \log p_k(x)$ with $p_k(x) = \mathrm{tr}[\rho(x)M_k]$, they read

$$\frac{\partial F_{\mathrm{ml}}(x)}{\partial x_i} = -\sum_k \frac{f_k}{p_k(x)}\mathrm{tr}(B_iM_k), \tag{45}$$

$$\frac{\partial^2 F_{\mathrm{ml}}(x)}{\partial x_j\partial x_i} = \sum_k \frac{f_k}{p_k^2(x)}\mathrm{tr}(B_jM_k)\mathrm{tr}(B_iM_k). \tag{46}$$

The bottleneck of such an algorithm is the actual computation of the second derivatives. Although the expansion coefficients of each measurement $\mathrm{tr}(B_jM_k)$ can be computed in advance, it is still necessary to compute equation (46) anew at each point x due to the dependence of $p_k(x)$. With respect to that, the least squares fit function bears a great advantage since its Hessian is constant, i.e. $\partial_j\partial_i F_{\mathrm{ls}}(x) = 2\sum_k w_k\mathrm{tr}(B_jM_k)\mathrm{tr}(B_iM_k)$, such that one saves time on this part.

Finally, let us comment on the optimality conditions, known as the Karush–Kuhn–Tucker conditions [36]. A given $x^\star$ is the global solution of the convex problem given by equation (13) if and only if[14] there exists an additional Lagrange multiplier $Z^\star$ (as given in the equations beneath) such that the pair $(x^\star, \Lambda^\star)$ satisfies

$$\frac{\partial}{\partial x_i}F(x^\star) - \mathrm{tr}[\Lambda^\star B_i] = 0, \quad \forall i, \tag{47}$$

$$\Lambda^\star \geqslant 0, \quad \rho(x^\star) \geqslant 0, \tag{48}$$

$$\mathrm{tr}[\Lambda^\star\rho(x^\star)] = 0. \tag{49}$$

Given the solution x^t_{sol} of the corresponding unconstrained problem with penalty parameter t, it follows from $\nabla f(x^t_{\mathrm{sol}}) = 0$ using equation (43) that the gradient conditions are satisfied with $\Lambda_t = t\rho(x^t_{\mathrm{sol}})^{-1}$ being the Lagrange multiplier. This pair $(x^t_{\mathrm{sol}}, \Lambda_t)$ also satisfies equation (48); only the duality gap condition $\mathrm{tr}[\Lambda_t\rho(x^t_{\mathrm{sol}})] = t\dim(\mathcal{H}) > 0$ does not hold exactly. However, this quantity appears in the following inequality:

$$F(x^t_{\mathrm{sol}}) - \mathrm{tr}[\Lambda_t\rho(x^t_{\mathrm{sol}})] = \min_{x:\rho(x)\geqslant 0} F(x) - \mathrm{tr}[\Lambda_t\rho(x)] \tag{50}$$

$$\leqslant \min_{x:\rho(x)\geqslant 0} F(x) = F(x_{\mathrm{sol}}). \tag{51}$$

Here one used that x^t_{sol} is the solution of $F(x) - \mathrm{tr}[\Lambda_t\rho(x)]$ because the gradient vanishes (and the solution is not at the border), and $\mathrm{tr}[\Lambda_t\rho(x)] \geqslant 0$ for the inequality. This is the stated accuracy given by equation (15) which relates the function value of x^t_{sol} to the true solution x_{sol}.

[14] Sufficiency holds under the Slater regularity condition that demands a strictly feasible point $\rho(x) > 0$, which naturally holds for state reconstruction problems.

5.3. Additional tools

5.3.1. Optimization of measurement settings. Measurement settings, each described by a unit vector $\hat{a}_i \in \mathbb{R}^3$ as explained in section 2.1, are chosen to optimize a figure of merit characterizing how well a given permutationally invariant target state ρ_{tar} can be reconstructed. As such a quality measure we use the sum of errors for the tomographically complete operator set of all tensor products of Pauli operators. Note that a permutationally invariant state ρ_{PI} is already uniquely determined by its generalized Bloch vector [17] defined as

$$b_{klmn} = \text{tr}\left(\left[\sigma_x^{\otimes k} \otimes \sigma_y^{\otimes l} \otimes \sigma_z^{\otimes m} \otimes \mathbb{1}^{\otimes n}\right]_{\text{PI}} \rho_{\text{PI}}\right) \tag{52}$$

with natural numbers satisfying $k+l+m+n = N$. Consequently, the total error of all Bloch vector elements is given by

$$\mathcal{E}^2_{\text{total}}(\hat{a}_i, \rho_{\text{tar}}) = \sum_{k,l,m,n} \binom{N}{k, l, m, n} \mathcal{E}^2_{b_{klmn}}(\hat{a}_i, \rho_{\text{tar}}), \tag{53}$$

where the multinomial coefficient weights the error of each generalized Bloch vector by its number of appearance in a generic Pauli product decomposition.

The error of each Bloch vector element must now be related to the carried out measurements. For that, note that each Bloch vector element can be expressed as

$$b_{klmn} = \sum_i c_i^{klmn} \text{tr}([(\hat{a}_i \cdot \vec{\sigma})^{\otimes N-n} \otimes \mathbb{1}^{\otimes n}]_{\text{PI}} \rho_{\text{tar}}) \tag{54}$$

using appropriate coefficients c_i^{klmn} and the expectation values of $[(\hat{a}_i \cdot \vec{\sigma})^{\otimes N-n} \otimes \mathbb{1}^{\otimes n}]_{\text{PI}}$ which can be computed from the coarse-grained measurement outcomes $M_k^{a_i}$ of setting $\hat{a}_i$ as given by equation (2) using linear combinations. Assuming independent errors, one obtains

$$\mathcal{E}^2_{b_{klmn}}(\hat{a}_i, \rho_{\text{tar}}) = \sum_i c_i^{klmn} \mathcal{E}^2_{\rho_{\text{tar}}}\left([(\hat{a}_i \cdot \vec{\sigma})^{\otimes N-n} \otimes \mathbb{1}^{\otimes n}]_{\text{PI}}\right). \tag{55}$$

The detailed form of the error expression $\mathcal{E}^2_{\rho_{\text{tar}}}([(\hat{a}_i \cdot \vec{\sigma})^{\otimes N-n} \otimes \mathbb{1}^{\otimes n}]_{\text{PI}})$ may depend on the actual physical realization, but we assume the following form:

$$\mathcal{E}^2_{\rho_{\text{tar}}}\left([(\hat{a}_i \cdot \vec{\sigma})^{\otimes N-n} \otimes \mathbb{1}^{\otimes n}]_{\text{PI}}\right) = K \Delta_{\rho_{\text{tar}}}\left([(\hat{a}_i \cdot \vec{\sigma})^{\otimes N-n} \otimes \mathbb{1}^{\otimes n}]_{\text{PI}}\right), \tag{56}$$

where $\Delta_\rho[A] = \text{tr}(\rho A^2) - [\text{tr}(\rho A)]^2$ is the standard variance and K is a state-independent factor. This form fits well, for example, to the common error model in photonic experiments where count rates are assumed to follow a Poissonian distribution. For more details of this derivation, see [17].

For large qubit numbers, N, this optimization is carried out iteratively. Starting from randomly chosen measurement directions or from vectors which are uniformly distributed according to some frame potential [46], one searches for updates according to

$$\hat{a}_i' = \frac{p\hat{a}_i^n + (1-p)\hat{r}_i}{\|p\hat{a}_i^n + (1-p)\hat{r}_i\|}. \tag{57}$$

Here $\hat{a}_i^n$ is the current iterate, $p < 1$ a probability close to 1 and $\hat{r}_i$ are randomly chosen unit vectors. If this new set of directions $\hat{a}_i'$ leads to a smaller total error $\mathcal{E}^2_{\text{total}}(\hat{a}_i', \rho_{\text{tar}})$ than the previous set, then these new measurement settings form the next iterate $\hat{a}_i^{n+1} = \hat{a}_i'$; otherwise this procedure is carried out once more. This process is repeated until the total error does not decrease anymore. This way of optimizing the measurements requires a method for computing

the total error $\mathcal{E}^2_{\text{total}}(\hat{a}_i, \rho_{\text{tar}})$ for a given set of measurements $\hat{a}_i$. Using the generalized Bloch vector or the spin ensemble this computation can be carried out again efficiently for larger qubit numbers N.

Although this algorithm is not proven to attain the true global optimum it is still a straightforward technique to obtain good settings. In the end this is often sufficient, recalling that the true unknown state can deviate from the target state and that one considers 'just' an overall single figure of merit. For our simulations we always use the optimized settings for the totally mixed state; a reasonable guess if no extra information is available [47].

Regarding this point, we finally like to add that if one does not employ the minimal number of measurement settings, but rather an over-complete set, e.g. four times as many settings but four times fewer measurements per setting, then the procedure is quite insensitive to the chosen measurement directions. Hence, in many practical situations the search for optimal directions might not even be necessary and randomly chosen measurement directions suffice equally well.

5.3.2. Statistical pretest. Via the pretest one estimates the fidelity between the true ρ_{true} and the best permutationally invariant state $F_{\text{PI}}(\rho_{\text{true}}) = \max_{\rho_{\text{PI}} \geqslant 0} \text{tr}(\sqrt{\sqrt{\rho_{\text{true}}}\rho_{\text{PI}}\sqrt{\rho_{\text{true}}}})$ using only measurement results from a few settings $\hat{a} \in T$, e.g. employing only $T = \{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$. Depending on this quantity, one decides on whether permutationally invariant tomography is worth continuing. As explained in detail in [17], this fidelity can be bounded by

$$F_{\text{PI}}(\rho_{\text{true}}) \geqslant [\text{tr}(\rho_{\text{true}} Z)]^2 \tag{58}$$

with an operator $Z = \sum z^a_k M^a_k$ being built up by the carried out measurements M^a_k given by equation (2) and satisfying $Z \leqslant P_{\text{sym}}$, where P_{sym} denotes the projector onto the symmetric subspace.

The expansion coefficients z^a_k should be optimized to attain the best lower bound. For a given target state ρ_{tar} this problem can be cast into a semidefinite program [36, 48] that can be solved efficiently using standard numerical routines. However, for larger qubit numbers one must again employ the block structure of the measurement operators as given by equation (9) to handle the operator inequality. Note that the projector on the symmetric subspace has a Block structure $P_{\text{sym},j} = \delta_{j,N/2}\mathbb{1}$. Then the final problem reads as

$$\begin{aligned} &\max_z \sum_{a\in T,k} z^a_k \text{tr}(\rho_{\text{tar}} M^a_k) \\ &\text{s.t.} \quad \sum_{a\in T,k} z^a_k M^a_{k,j=N/2} \leqslant \mathbb{1}, \quad \sum_{a\in T,k} z^a_k M^a_{k,j} \leqslant 0, \quad \forall j < N/2. \end{aligned} \tag{59}$$

If one experimentally implements this pretest, one must account for additional statistical fluctuations. For the chosen Z, one can employ the sample mean $\bar{Z} = \sum z^a_k f^a_k$ using the observed frequencies $f^a_k = n^a_k/N_{\text{R}}$ in N_{R} repetitions of setting $\hat{a}$, as an estimate of the true expectation value $\text{tr}(\rho_{\text{true}} Z)$. This sample mean $\bar{Z}$ will fluctuate around the true mean but large deviations will become less likely, such that for an appropriately chosen ϵ the quantity $\text{sign}(\bar{Z} - \epsilon)(\bar{Z} - \epsilon)^2$ is a lower bound to the true fidelity at the desired confidence level. The proof essentially uses the techniques employed in [49, 50].

Proposition 5.3 (Statistical pretest). *For any $Z = \sum z^a_k M^a_k \leqslant P_{\text{sym}}$ let $\bar{Z} = \sum z^a_k n^a_k/N_{\text{R}}$ denote the sample mean using N_{R} repetitions for setting $\hat{a} \in T$. If the data are generated by*

the state ρ_{true}*, then*

$$\text{Prob}\left[F_{\text{PI}}(\rho_{\text{true}}) \geqslant \text{sign}(\bar{Z}-\epsilon)(\bar{Z}-\epsilon)^2\right] \geqslant 1-\exp(-2N_{\text{R}}\epsilon^2/C_z^2) \tag{60}$$

with $C_z^2 = \sum_a \left(z^s_{\max} - z^s_{\min}\right)^2$ *where* $z^a_{\max/\min}$ *are the respective optima for setting* $\hat{a}$ *over all outcomes* k*.*

Proof. The given statement follows along

$$\text{Prob}\left[F_{\text{PI}}(\rho_{\text{true}}) \geqslant \text{sign}(\bar{Z}-\epsilon)(\bar{Z}-\epsilon)^2\right] \geqslant \text{Prob}\left[\text{tr}(\rho_{\text{true}}Z) \geqslant \bar{Z}-\epsilon\right], \tag{61}$$

$$\geqslant 1-\text{Prob}\left[\text{tr}(\rho_{\text{true}}Z) \leqslant \bar{Z}-\epsilon\right] \geqslant 1-\exp(-2N_{\text{R}}\epsilon^2/C_z^2). \tag{62}$$

Here the first inequality holds because the set of outcomes satisfying $\{n_k^a : \text{tr}(\rho_{\text{true}}Z) \geqslant \bar{Z}-\epsilon\}$ is a subset of $\{n_k^a : F_{\text{PI}}(\rho_{\text{true}}) \geqslant \text{sign}(\bar{Z}-\epsilon)(\bar{Z}-\epsilon)^2\}$ using equation (58). In the last inequality we use the Hoeffdings tail inequality [51] to bound Prob[$\text{tr}(\rho_{\text{true}}Z) \leqslant \bar{Z}-\epsilon$]. □

Note that this pretest can also be applied after the whole tomography scheme in which case the projector $P_{\text{sym}} = \sum z_k^a M_k^a$ becomes accessible. Moreover, let us point out that a strong statistical significance, or a low ϵ respectively, might not be achieved with the best expectation value as given by equation (59) [52]; hence, optimizing Z for a rather mixed state is often better.

Finally, let us remark that the pretest can be improved by additional projectors, see the supplementary material of [17]. This leads to the bound $F_{\text{PI}}(\rho_{\text{true}}) \geqslant \sum_j p_j^2$ with p_j being the weight of the corresponding spin-j state of the permutationally invariant part of ρ_{true} as given in equation (7). From this expression one sees that this test only works well for states having a rather large weight on one of these spin states. Others, like the totally mixed state, while clearly being permutationally invariant, are not identified as states close to the permutationally invariant subspace. This is in stark contrast to compressed sensing where the certificate succeeds for the whole class of low-rank target states [10].

5.3.3. Entanglement and the MaxLik–MaxEnt principle. Following the last comment from the previous subsection, we want to argue that even in the case of an inefficient certificate, permutationally invariant state reconstruction as given by equation (5) is meaningful. Firstly we would like to emphasize that the permutationally invariant part of any density operator corresponds to a good representative to investigate the entanglement properties of the true, unknown state. This is because the transformation given by equation (1) can be achieved by permutations or multiple swap operations together with classical mixing, whereby entanglement cannot be created [53]. Thus if the permutationally invariant part of the density operator is entangled this holds true also for the real state. In addition, any symmetric entanglement measure, i.e. all commonly known ones, can only decrease under this operation, thereby even quantification is faithful [54].

Secondly, permutationally invariant state reconstruction also represents the solution of the maximum-likelihood maximum-entropy principle as introduced in [55], which goes as follows: if the carried out measurements are not tomographically complete, then there is, in general, not a single state $\hat{\rho}_{\text{ml}}$ that maximizes the likelihood but rather a complete set of them. In order to single out a 'good' representative, the authors of [55] propose to choose the state that has the largest entropy, which, according to the Jaynes principle [47], is the special state for which one has the fewest information.

Proposition 5.4 (Permutationally invariant MaxLik–MaxEnt principle). *Using the described permutationally invariant tomography scheme, the reconstructed permutationally invariant state given by equation (5) (with the likelihood function) is also the solution of the maximum-likelihood maximum-entropy principle.*

Proof. Since the measurements given by equation (2) are invariant and tomographically complete for permutationally invariant states, all density operators with the same spin ensemble as $\hat{\rho}_{\mathrm{PI}}$ have the same maximum likelihood. According to Ref. [47], the state with maximal entropy and consistent with a given set of expectation values for operators M_k^a has the form $\rho \propto \exp(\sum_{a,k} \lambda_k^a M_k^a)$. The Lagrange multipliers $\lambda_k^a \in \mathbb{R}$ must be chosen such that the given expectation values match. However, because all M_k^a are permutationally invariant we can employ the block decomposition given by equation (9) and finally obtain $\exp(\sum_{a,k} \lambda_k^a M_k^a) = \exp(\oplus_j \sum_{a,k} \lambda_k^a M_{k,j}^a \otimes \mathbb{1}) = \oplus_j \exp(\sum_k^a \lambda_k^a M_{k,j}^a) \otimes \mathbb{1}$. Hence we obtain the same structure as $\hat{\rho}_{\mathrm{PI}}$, which therefore is also the state with maximum entropy. □

6. Conclusion and outline

In this paper, we have provided all the necessary ingredients to carry out permutationally invariant tomography [17] in experiments with large qubit numbers. This includes, besides scheme-specific tasks such as the statistical pretest and the optimization of the measurement settings, in particular the state reconstruction part. Accounting for statistical fluctuations due to a finite number of data, this reconstruction demands the solution of a nonlinear large-scale optimization problem. We achieve this, firstly, by using a convenient toolbox to store, characterize and process permutationally invariant states, which largely reduces the dimension of the underlying problem and, secondly, by using convex optimization, which is superior compared to commonly used numerical routines in many respects. This makes permutationally invariant tomography a complete tomography method requiring only moderate measurement and data analysis effort.

There are many questions one may pursue in this direction: firstly, let us stress that the current prototype implementation is still not optimal. As explained, the bottleneck is the computation of the second derivatives; hence, we strongly believe that Hessian free optimization, like quasi-Newton or conjugate gradients [56], or the use of other barrier functions more tailored to linear matrix inequalities [57] are likely to push the reconstruction limit further. Secondly, it is natural to try to exploit other symmetries in the development of 'symmetry' tomography protocols; that is, tomography should work for all states that remain invariant under the action of a specific group. Clearly, any symmetry decreases the number of state-dependent parameters, but the challenge is to devise efficient local measurement strategies. Interesting classes here are graph-diagonal [58] or, more general, locally maximally entangleable states [59], translation or shift-invariant states [60], or $U^{\otimes N}$ invariant states [61, 62]. Thirdly, it is worth investigating to what extent particularly designed state tomography protocols are useful for further tasks, such as process tomography for quantum gates. For instance, permutationally invariant tomography might be unable to resolve the Toffoli gate [63] directly, but since the operation on all N target qubits is symmetric, a permutationally invariant resolution of this subspace (and the additional control qubit) might be sufficient. Finally, let us point out that permutationally invariant tomography can be further restricted to the symmetric

subspace, which often contains the desired states. This is reasonable since we have seen that the pretest is only good if the unknown state has a large weight in one of the spin states. However, since the symmetric subspace grows only linearly with the number of particles, this tomography scheme can analyse many more qubits efficiently.

Acknowledgments

We thank M Kleinmann, B Kraus, T Monz, L Novo, P Schindler and J Wehr for stimulating discussions on the topic and technicalities. This work was supported by the FWF (START prize Y376-N16), the EU (Project Q-ESSENCE and Marie Curie CIG 293993/ENFOQI), the BMBF (Chist-Era Project QUASAR), the ERC (Starting Grant GEDENTQOPT), the QCCC of the Elite Network of Bavaria, the Spanish MICINN (project no. FIS2009-12773-C02-02), the Basque Government (project no. IT4720-10) and the National Research Fund of Hungary OTKA (contract no. K83858).

References

[1] Paris M G A and Řeháček J (ed) 2004 *Quantum State Estimation* (Berlin: Springer)
[2] Häffner H *et al* 2005 *Nature* **438** 643
[3] Kiesel N, Schmid C, Tóth G, Solano E and Weinfurter H 2007 *Phys. Rev. Lett.* **98** 063604
[4] James D F V, Kwiat P G, Munro W J and White A G 2001 *Phys. Rev.* A **64** 052312
[5] Renes J M, Blume-Kohout R, Scott A J and Caves C M 2004 *J. Math. Phys.* **45** 2171
[6] Monz T, Schindler P, Barreiro J T, Chwalla M, Nigg D, Coish W A, Harlander M, Haensel W, Hennrich M and Blatt R 2011 *Phys. Rev. Lett.* **106** 130506
[7] Yao X C, Wang T X, Xu P, Lu H, Pan G S, Bao X H, Peng C Z, Lu C Y, Chen Y A and Pan J W 2012 *Nature Photon.* **6** 225
[8] Huang Y F, Liu B H, Peng L, Li Y H, Li L, Li C F and Guo G C 2011 *Nature Commun.* **2** 546
[9] Gross D, Liu Y K, Flammia S T, Becker S and Eisert J 2010 *Phys. Rev. Lett.* **105** 150401
[10] Gross D 2011 *IEEE Trans. Inf. Theory* **57** 1548
[11] Flammia S T, Gross D, Liu Y K and Eisert J 2012 Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators arXiv:1205.2300
[12] Cramer M, Plenio M B, Flammia S T, Gross D, Bartlett S D, Somma R, Landon-Cardinal O, Liu Y K and Poulin D 2010 *Nature Commun.* **1** 9
[13] Baumgratz T, Gross D, Cramer M and Plenio M B 2012 Scalable reconstruction of density matrices arXiv:1207.0358
[14] Landon-Cardinal O and Poulin D 2012 Practical learning method for multi-scale entangled states arXiv:1204.0792
[15] Yin J O S and van Enk S J 2011 *Phys. Rev.* A **83** 062110
[16] Guta M, Kypraios T and Dryden I 2012 Rank-based model selection for multiple ions quantum tomography arXiv:1206.4032
[17] Tóth G, Wieczorek W, Gross D, Krischek R, Schwemmer C and Weinfurter H 2010 *Phys. Rev. Lett.* **105** 250403
[18] Moroder T, Kleinmann M, Schindler P, Monz T, Gühne O and Blatt R 2012 Detection of systematic errors in quantum experiments arXiv:1204.3644
[19] D'Ariano G M, Maccone L and Paini M 2003 *J. Opt. B: Quantum Semiclass. Opt.* **5** 77
[20] Karassiov V P 2005 *J. Russ. Laser Res.* **26** 484
[21] Adamson R B A, Shalm L K, Mitchell M W and Steinberg A M 2007 *Phys. Rev. Lett.* **98** 043601

[22] Adamson R B A, Turner P S, Mitchell M W and Steinberg A M 2008 *Phys. Rev.* A **78** 033832
[23] Hradil Z 1997 *Phys. Rev.* A **55** R1561
[24] Langford N K 2007 Encoding, manipulating and measuring quantum information in optics *PhD Thesis* School of Physical Sciences, University of Queensland
[25] Blume-Kohout R 2010 *Phys. Rev. Lett.* **105** 200504
[26] Efron B and Tibshirani R J 1994 *An Introduction to the Bootstrap* (London: Chapman and Hall)
[27] Christandl M and Renner R 2011 Reliable quantum state tomography arXiv:1108.5329
[28] Blume-Kohout R 2012 Robust error bars for quantum tomography arXiv:1202.5270
[29] Mood A F 1974 *Introduction to the Theory of Statistics* (New York: McGraw-Hill)
[30] Cirac J I, Ekert A K and Macchiavello C 1999 *Phys. Rev. Lett.* **82** 4344
[31] Demkowicz-Dobrzański R 2005 *Phys. Rev.* A **71** 062321
[32] Chase B A and Geremia J M 2008 *Phys. Rev.* A **78** 052101
[33] Baragiola B Q, Chase B A and Geremia J M 2010 *Phys. Rev.* A **81** 032104
[34] Hradil Z, Řeháček J, Fiurášek J and Ježek M 2004 *Lect. Notes Phys.* **649** 59
[35] Reimpell M 2008 Quantum information and convex optimization *PhD Thesis* Technische Universität Braunschweig
[36] Boyd S and Vandenberghe S 2004 *Convex Optimization* (Cambridge: Cambridge University Press)
[37] Kosut R L, Walmsley I A and Rabitz H 2004 Optimal experiment design for quantum state and process tomography and Hamiltonian parameter estimation arXiv:quant-ph/0411093
[38] Glancy S, Knill E and Girard M 2012 Gradient-based stopping rules for maximum-likelihood quantum-state tomography arXiv:1205.4043
[39] Řeháček J, Hradil Z and Ježek M 2001 *Phys. Rev.* A **63** 040303
[40] Řeháček J, Hradil Z, Knill E and Lvovsky A I 2007 *Phys. Rev.* A **75** 042108
[41] Nielsen M A and Chuang I L 2000 *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press)
[42] Życzkowski K and Sommers H J 2001 *J. Phys. A: Math. Gen.* **34** 7111
[43] Simon B 1996 *Representations of Finite and Compact Groups* (*Graduate Studies in Mathematics* vol 10) (Providence, RI: American Mathematical Society)
[44] Christandl M 2006 The structure of bipartite quantum states—insights from group theory and cryptography arXiv:0604183
[45] Hall B C 2003 *Lie Groups, Lie Algebras and Representations: An Elementary Introduction* (New York: Springer)
[46] Gross D, Audenaert K and Eisert J 2007 *J. Math. Phys.* **48** 052104
[47] Jaynes E 1957 *Phys. Rev.* **108** 171
[48] Tóth G, Wieczorek W, Krischek R, Kiesel N, Michelberger P and Weinfurter H 2009 *New J. Phys.* **11** 083002
[49] Flammia S T and Liu Y K 2011 *Phys. Rev. Lett.* **106** 230501
[50] da Silva M P, Landon-Cardinal O and Poulin D 2011 *Phys. Rev. Lett.* **107** 210404
[51] Hoeffding W 1963 *J. Am. Stat. Assoc.* **58** 301
[52] Jungnitsch B, Niekamp S, Kleinmann M, Gühne O, Lu H, Gao W B, Chen Y A, Chen Z B and Pan J W 2010 *Phys. Rev. Lett.* **104** 210401
[53] Busch P 2003 *Int. J. Theor. Phys.* **42** 937
[54] Horodecki K, Horodecki M and Horodecki P 2010 *Quantum Inform. Comput.* **10** 0901
[55] Teo Y S, Zhu H B-G, Englert B G, Řeháček J and Hradil Z 2011 Knowledge and ignorance in incomplete quantum state tomography arXiv:1102.2662
[56] Geiger C and Kanzow C 1999 *Numerische Verfahren zur Lösung unrestringierter Optimierungsaufgaben* (Berlin: Springer)
[57] Kočvara M and Stingl M 2003 *Optim. Methods Softw.* **18** 317
[58] Hein M, Dür W, Eisert J, Raussendorf R, Van den Nest M and Briegel H J 2005 *Proc. Int. School of Physics 'Enrico Fermi' on 'Quantum Computers, Algorithms and Chaos', Varenna, Italy*

[59] Kruszynska C and Kraus B 2009 *Phys. Rev.* A **79** 052304
[60] O'Connor K M and Wootters W K 2001 *Phys. Rev.* A **63** 052302
[61] Eggeling T and Werner R F 2001 *Phys. Rev.* A **63** 042111
[62] Cabello A 2003 *J. Mod. Opt.* **50** 10049
[63] Monz T, Kim K, Hänsel W, Riebe M, Villar A S, Schindler P, Chwalla M, Hennrich M and Blatt R 2009 *Phys. Rev. Lett.* **102** 040501