Self-consistent tomography of the state-measurement Gram matrix

Cyril Stark
Theoretische Physik, ETH Zurich, CH-8093 Zurich, Switzerland
and Center for Theoretical Physics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139-4307, USA

(Received 2 December 2013; published 12 May 2014)

State and measurement tomography make assumptions about the experimental states or measurements. These assumptions are often not justified because state preparation and measurement errors are unavoidable in practice. Here we describe how the Gram matrix associated with the states and measurement operators can be estimated via semidefinite programming if the states and the measurements are so-called globally completable. This is, for instance, the case if the unknown measurements are known to be projective and nondegenerate. The computed Gram matrix determines the states, and the measurement operators uniquely up to simultaneous rotations in the space of Hermitian matrices. We prove the reliability of the proposed method in the limit of a large number of independent measurement repetitions.

DOI: 10.1103/PhysRevA.89.052109 PACS number(s): 03.65.Wj, 03.65.Ta, 03.67.−a

I. INTRODUCTION

Analyzing a quantum experiment amounts to fitting a quantum description (i.e., states, measurements, and processes) to measurement data. State tomography accomplishes this task when we ignore the dynamics between the state preparations and the measurements, and when the experimental measurements are known accurately. Hence, from the perspective of state tomography, deviations between the postulated measurement operators and the experimental measurement operators are *systematic* errors (as opposed to *statistical* errors which are a consequence of finite measurement repetitions). These errors are unavoidable in practice [1,2]; they are a serious bottleneck for today's experiments [3–5], and they undermine the recent efforts [6–8] to equip state tomography with meaningful error bars. Systematic errors are addressed in self-consistent tomography [1,2,9–12]: Assuming as little as possible about the experiment, self-consistent tomography aims at fitting quantum models to measured data.

Ideally we would like an algorithm that takes the measured data as input and outputs compatible states and measurements. This task can by no means be compared with state tomography because the underlying computational problem is nonconvex and therefore, it will require a lot of effort to fully understand this problem. In fact, the simultaneous estimation of states and measurements and its classical analog are closely related to two tasks from computer science: the computation of the nonnegative and positive semidefinite (psd) rank [13,14]. This close connection suggests that fully assumption-free tomography is computationally too challenging because the computation of the nonnegative rank is NP hard.

To proceed, we could either try to come up with approximation algorithms, or we could analyze practically important special cases to avoid the hard instances of the problem, or we could raise the question whether weak assumptions (weaker than the assumptions underlying state tomography) could render the problem tractable. Here, we will pursue the latter strategy which focuses on a middle ground between state tomography and fully assumption-free tomography. One self-evident assumption is that some measurement operators are pairwise orthogonal projectors (projective measurements as opposed to general POVMs). Even though this assumption is strictly speaking violated in real experiments [15], we still believe that this is a good starting point for the exploration of the middle ground mentioned before.

Our objective is the estimation of the Gram matrix associated with the experimental states and measurements. The Gram matrix $G \in \mathbb{R}^{n\times n}$ of vectors $\vec{v}_1,\ldots,\vec{v}_n \in \mathbb{R}^m$ is defined by $G_{ij} = \vec{v}_i^T \vec{v}_j$. The Gram matrix G fixes the pairwise relative geometry of the vectors $(\vec{v}_j)_j$ because it determines via $\|\vec{v}_i - \vec{v}_j\|_2^2 = G_{ii}^2 + G_{jj}^2 - 2G_{ij}$ all the pairwise distances between the vectors generating G.

Projectiveness of the measurements can be enforced by fixing some entries in the Gram matrix G associated to the collection of matrices describing the experimental states and measurements. This observation on the one hand and the observation that G determines the states and the measurements up to simultaneous rotations on the other hand motivates the objective of this work: Given measurement data and assumptions like projectiveness of measurements, find G. The precise list of assumptions can be found in the discussion (Sec. V) at the end of this paper. We introduce a method for G estimation whose output can be certified. Our main contribution is not of a technical nature but the observation that it is possible to analyze systematic errors via estimation of the state-measurement Gram matrix G.

It is shown elsewhere [16] that the Gram estimation opens up the door for heuristic algorithms to search for explicit states and measurement operators describing the experimental data. This is our main motivation for G estimation: Knowing G opens up the door for heuristic algorithms searching for explicit states and measurement operators. However, knowing G allows us also to assess the purity of states because ρ is pure if and only if $\|\rho\|_2 = 1$. So if the diagonal entry of G corresponding to ρ is equal to 1 (consequently, $\rho = |\psi\rangle\langle\psi|$) then the fidelity $F(\rho,\sigma)^2 = \|\rho^{1/2}\sigma^{1/2}\|_1^2 = \mathrm{tr}\,(|\psi\rangle\langle\psi|\sigma)$ is equal to the entry of G corresponding to the inner product between ρ and σ. For general states ρ and σ, the fidelity can be upper bounded [17] via $F(\rho,\sigma) \leqslant \mathrm{tr}\,(\rho\sigma) + \sqrt{[1-\mathrm{tr}\,(\rho^2)][1-\mathrm{tr}\,(\sigma^2)]}$. Again, this upper bound is a function of the entries of the state-measurement Gram matrix G.

Our approach to self-consistent tomography is entirely different from existing approaches which assume that the

unknown quantities (depending on the context, these are states, measurements or processes) are known approximately. For instance, [1] introduces a method for process tomography which is not prone to state preparations and measurement errors if the experimental gates are approximately known beforehand (see also [2,9–12,18]). On the other hand, Kimmel *et al.* showed in their seminal work [19] that randomized benchmarking can be used for self-consistent process tomography if the considered processes are assumed to be unital. Here, we do not need to know any of the experimental states or measurements explicitly. Qualitative features such as projectiveness and nondegeneracy of the measurements are sufficient. Thus, our method can, for instance, be used to detect and correct misalignment errors [5].

We start by describing the setting and by defining the notation. Then, in Sec. III, we introduce the Gram estimation problem, we explain the underlying difficulties, and we show how the Gram matrix generated by the quantum states and measurements can nevertheless be computed by a combination of tools from convex relaxation and rigidity theory. To illustrate the performance of the proposed method, we ran extensive numerical simulations. The results are summarized in Sec. IV. We conclude the paper with a discussion comparing our method for Gram estimation with state tomography and listing the precise set of assumptions.

II. SETTING

We consider an experiment which allows the preparation of W different unknown states $(\rho_w)_{w=1}^W$ and the performance of V different unknown measurements, each of which is described in terms of measurement operators $(E_{vk})_{k=1}^K$. Repeating the measurement of ρ_w with $(E_{vk})_{k=1}^K$ N times, we can count how many times we have measured the outcomes "1", ..., "K". Dividing these numbers by N we obtain frequencies $f_{k|wv}$ for measuring "k" given that we have prepared the state "w" and performed the measurement "v." The data table,

$$\mathcal{D} = \begin{pmatrix} f_{1|11} & \cdots & f_{K|11} & \cdots & f_{1|1V} & \cdots & f_{K|1V} \\ f_{1|21} & \cdots & f_{K|21} & \cdots & f_{1|2V} & \cdots & f_{K|2V} \\ \vdots & & \vdots & & \vdots & & \vdots \\ f_{1|W1} & \cdots & f_{K|W1} & \cdots & f_{1|WV} & \cdots & f_{K|WV} \end{pmatrix},$$

describes the phenomenological content of the experiment. Changing the row index in $\mathcal{D}$ amounts to changing the state while changing the column index in $\mathcal{D}$ amounts to changing the measurement outcome. In the limit $N \to \infty$, by Born's rule, $f_{k|wv} = \mathrm{tr}\,(\rho_w E_{vk})$. The matrices ρ_w and E_{vk} are contained in the space of Hermitian matrices $\mathrm{Herm}(\mathbb{C}^d)$ if the underlying quantum system is d dimensional. With respect to an othonormal basis in $\mathrm{Herm}(\mathbb{C}^d)$ we can express them in terms of vectors $\vec{\rho}_w, \vec{E}_{vk} \in \mathbb{R}^{d^2}$ because $\mathrm{Herm}(\mathbb{C}^d)$ is a real d^2-dimensional vector space. It follows that $\mathrm{tr}\,(\rho_w E_{vk}) = (\vec{\rho}_w)^T \vec{E}_{vk}$. Define the matrices,

$$\begin{aligned} P_{\mathrm{st}} &:= (\vec{\rho}_1 \,|\, \cdots \,|\, \vec{\rho}_W), \\ P_{\mathrm{m}} &:= (\vec{E}_{11} \,|\, \cdots \,|\, \vec{E}_{1K} \,|\, \cdots \,|\, \vec{E}_{V1} \,|\, \cdots \,|\, \vec{E}_{VK}), \\ P &:= (P_{\mathrm{st}} \mid P_{\mathrm{m}}), \end{aligned} \tag{1}$$

and $G = P^T P$. The matrix G is the Gram matrix associated with $\rho_1, \ldots, E_{VK}$. The data table $\mathcal{D}$ appears as an off-diagonal block in the state-measurement Gram matrix:

$$G = \left(\begin{array}{c|c} G_{\mathrm{st}} & \mathcal{D} \\ \hline \mathcal{D}^T & G_{\mathrm{m}} \end{array}\right). \tag{2}$$

III. GRAM ESTIMATION

Our objective is the estimation of G. Let Ω be the set of indices marking the known entries of G, and let $\vec{\mathcal{K}} \in \mathbb{R}^{|\Omega|}$ be the explicit numerical values of the known entries. Thus, $G_\Omega = \vec{\mathcal{K}}$ captures our *a priori* knowledge about the entries of G. In the remainder we assume that at least the phenomenological data $\mathcal{D}$ are known and thus part of G_Ω.

Note that $\mathrm{rank}(G) = \mathrm{rank}(P)$. We can only hope to reconstruct *general* states and measurements if

$$\mathrm{Herm}(\mathbb{C}^d) = \mathrm{span}_{\mathbb{R}}\{\rho_w\}_w = \mathrm{span}_{\mathbb{R}}\{E_{vk}\}_{vk}. \tag{3}$$

Otherwise, we could not even determine the states if all measurements were known and vice versa. Hence, both matrices P_{st} and P_{m} are assumed to be full rank in the remainder [20]. Recalling $\mathcal{D} = P_{\mathrm{st}}^T P_{\mathrm{m}}$, it follows that

$$d^2 = \mathrm{rank}(\mathcal{D}) = \mathrm{rank}(P) = \mathrm{rank}(G). \tag{4}$$

This was first observed in [21]. Let $\mathcal{G}_{\mathrm{QM}}$ denote the set of Gram matrices that can be generated via quantum density matrices and measurement operators. The set $\mathcal{G}_{\mathrm{QM}}$ is contained in the set $S^+(\mathbb{R}^N)$ of symmetric and positive semidefinite matrices on $\mathbb{R}^N$ ($N := W + VK$). By Eq. (4), determining the state-measurement Gram matrix is equivalent to solving the problem,

$$\begin{aligned} \text{find} \quad & G \in \mathcal{G}_{\mathrm{QM}}, \\ \text{subject to} \quad & G_\Omega = \vec{\mathcal{K}}, \\ & \mathrm{rank}(G) = \mathrm{rank}(\mathcal{D}). \end{aligned} \tag{5}$$

The result will automatically be the *lowest-dimensional* model that is favored by Occam's razor. Since $\mathcal{D}$ is a partial matrix of G, $\mathrm{rank}(\mathcal{D}) \leqslant \mathrm{rank}(G)$. Thus, instead of solving (5), we can equally well try to solve the optimization problem [22],

$$\begin{aligned} \text{minimize} \quad & \mathrm{rank}(G), \\ \text{subject to} \quad & G_\Omega = \vec{\mathcal{K}}, \quad G \in \mathcal{G}_{\mathrm{QM}}. \end{aligned} \tag{6}$$

When trying to compute (6) we face two difficulties:

(D1) The optimization problem (6) is not convex because the rank of a matrix is not a convex function [24]. For example,

$$\begin{aligned} 2 &= \mathrm{rank}(p\ |0\rangle\langle 0| + (1-p)\ |1\rangle\langle 1|) \\ &> p\ \mathrm{rank}(|0\rangle\langle 0|) + (1-p)\ \mathrm{rank}(|1\rangle\langle 1|) = 1. \end{aligned}$$

(D2) We do not know how to efficiently characterize the set of quantum Gram matrices $\mathcal{G}_{\mathrm{QM}}$.

Difficulty (D1). One approach to solve nonconvex optimization problems is to relax them to closely related convex optimization problems. This is illustrated in Fig. 1: Instead of trying to find the global minimum of the nonconvex function $f(x)$, we are computing the minimum of the function $g(x)$. The convexity of $g(x)$ guarantees that the found minimum is a global minimum of $g(x)$. A function $g : \mathcal{C} \to \mathbb{R}$ from a convex

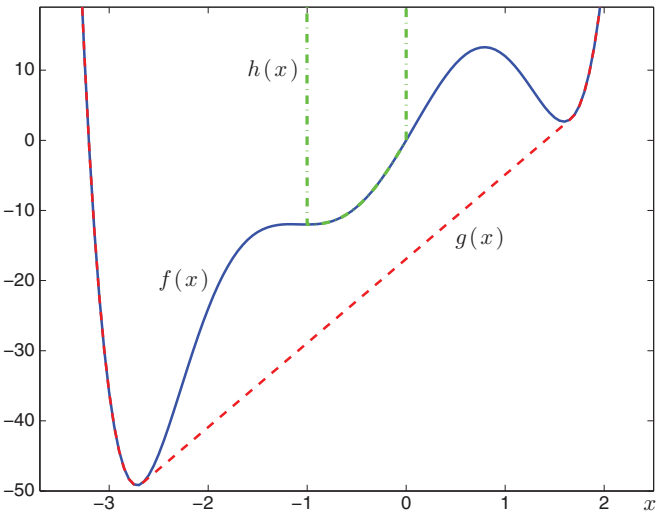

FIG. 1. (Color online) The functions $g(x)$ and $h(x)$ are convex envelopes of $f(x)$ with respect to the intervals $\mathcal{C} = \mathbb{R}$, respectively, $\mathcal{C} = [-1,0]$.

set $\mathcal{C}$ to $\mathbb{R}$ is the *convex envelope* of a function $f : \mathcal{C} \to \mathbb{R}$ if it is the pointwise largest convex function satisfying $g(x) \leqslant f(x)$ for all $x \in \mathcal{C}$. Note that this property depends on the convex set $\mathcal{C}$. This becomes evident when we replace the choice $\mathcal{C} = \mathbb{R}$ in Fig. 1 with an interval, e.g., to $\mathcal{C} = [-1,0]$. Recall that $\mathcal{G}_{\mathrm{QM}} \subset S^+(\mathbb{R}^N)$. Our goal here is to solve the optimization problem (6) by replacing the rank function by its convex envelope with respect to the convex set,

$$\mathcal{C} := \{X \in \mathbb{R}^{n\times n} \mid X \geqslant 0,\ \|X\| \leqslant R_{\mathrm{QM}}(d)\},$$

where $\|.\|$ denotes the operator norm and

$$R_{\mathrm{QM}}(d) := \sup\{\|X\| \mid X \in \mathcal{G}_{\mathrm{QM}},\ \dim(\mathcal{H}) \leqslant d\}, \qquad (7)$$

i.e., $R_{\mathrm{QM}}(d)$ is the radius of the smallest operator norm ball containing the Gram matrices corresponding to d-dimensional quantum systems. In the appendix we show that $R_{\mathrm{QM}}(d) = W + Vd$. Fazel, Hindi, and Boyd proved [26] that $\|X\|_1/R_{\mathrm{QM}}$ is the convex envelope of the rank function on the larger set,

$$\mathcal{C}' := \{X \in \mathbb{R}^{n\times n} \mid \|X\| \leqslant R_{\mathrm{QM}}(d)\} \supset \mathcal{C}.$$

Here, $\|.\|_1$ denotes the trace norm. Since $\mathcal{C}$ is a proper subset of $\mathcal{C}'$ it is unclear whether or not $\|X\|_1/R_{\mathrm{QM}}$ is also the convex envelope of the rank function with respect to $\mathcal{C}$ (see Fig. 1). The validity of this claim is proven in the appendix by modifying the derivation from [26]. Consequently, we arrive at the following substitution of the nonconvex optimization problem (6):

$$\begin{aligned} &\text{minimize} && \operatorname{tr} G, \\ &\text{subject to} && G_\Omega = \vec{\mathcal{K}}, \\ & && G \in S^+(\mathbb{R}^N),\ \|G\| \leqslant R_{\mathrm{QM}}(d). \end{aligned} \qquad (8)$$

Problem (8) can be cast into a semidefinite program (SDP): The constraint $\|G\| \leqslant R_{\mathrm{QM}}$ in Eq. (8) is equivalent to $R_{\mathrm{QM}}\mathbb{I} - G \geqslant 0$ because $G \geqslant 0$. Hence, the optimization problem (8)

Algorithm 1 Gram estimation

Require: $G \geqslant 0$ is determined uniquely by $G_\Omega = \vec{\mathcal{K}}$ and $\operatorname{rank}(G) = \operatorname{rank}(\mathcal{D})$ [27].
1: Run the optimization (9) to compute a completion $\hat{G}$ of G_Ω.
2: **while** $\operatorname{rank}(\hat{G}) > \operatorname{rank}(\mathcal{D})$ **do**
3: Prepare additional states, or perform additional
4: measurements. Run the optimization (9).
5: **end while**
6: Return $\hat{G}$.

is equivalent to the SDP,

$$\begin{aligned} &\text{minimize} && \operatorname{tr} G, \\ &\text{subject to} && G_\Omega = \vec{\mathcal{K}}, \\ & && Z = (W + Vd)\mathbb{I} - G, \\ & && G, Z \geqslant 0, \end{aligned} \qquad (9)$$

because $M_{\mathrm{QM}}(d)\mathbb{I} - G \geqslant 0$ is implied by $Z \geqslant 0$ and $G + Z = M_{\mathrm{QM}}(d)\mathbb{I}$. Consequently, the optimization problem (8) can be solved efficiently by standard methods [28,29].

Difficulty (D2). Even though closely related, the problems (6) and (8) are not identical. There exists no guarantee that the global optimum of the semidefinite program (8) and the global optimum of the rank minimization (6) agree. In particular, when going from the original rank minimization to its convex relaxation, we extended the feasible set by replacing $\mathcal{G}_{\mathrm{QM}}$ with $\{X \geqslant 0,\ \|X\| \leqslant R_{\mathrm{QM}}(d)\}$. Hence, there is no guarantee that the solution of the relaxed problem (8) lies in $\mathcal{G}_{\mathrm{QM}}$. However, if there existed only one Gram matrix G satisfying $G_\Omega = \vec{\mathcal{K}}$ and $\operatorname{rank}(G) = \operatorname{rank}(\mathcal{D})$ and if the solution $\hat{G}$ of (9) satisfies $G_\Omega = \vec{\mathcal{K}}$ and $\operatorname{rank}(G) = \operatorname{rank}(\mathcal{D})$, then we know that $\hat{G}$ is equal to the correct state-measurement Gram matrix. Is the state-measurement Gram matrix G ever uniquely determined in this way? Generically G is never uniquely fixed if $\mathcal{D}$ is all we know about G. Hence, we need to make sure that our knowledge $G_\Omega = \vec{\mathcal{K}}$ and $\operatorname{rank}(G) = \operatorname{rank}(\mathcal{D})$ is sufficient to uniquely determine G. In [27], a necessary and sufficient criterium for the uniqueness of G—given $G_\Omega = \vec{\mathcal{K}}$ and $\operatorname{rank}(G) = \operatorname{rank}(\mathcal{D})$—is derived for situations where either something is known about G_{st} or G_{m}. Then, G is uniquely determined if and only if

$$\operatorname{rank}[\mathcal{M}(\vec{\mathcal{K}},\Omega)] = d^2(d^2+1)/2,$$

where $\mathcal{M}(\cdot)$ is a matrix-valued function in $\vec{\mathcal{K}}$ and Ω (see [27] (Theorem 2), and [30,31] for related work). Applying this criterion to specific circumstances we observe (for instance) that knowing that the experimental measurements are projective and nondegenerate is sufficient to ensure uniqueness of G if the number of measurements exceeds a critical value.

Algorithm. The previous discussions lead to Algorithm 1 for estimating the state-measurement Gram matrix.

IV. NUMERICAL EXAMPLES

We ran Algorithm 1 for different dimensions d assuming that the measurements are projective and nondegenarate. We

TABLE I. Numerical experiments.

Dim($\mathcal{H}$)	Successes	Failures	Start point	Solver
2	1000	0	(5,5)	SeDuMi [28]
3	1000	0	(60,100)	TFOCS [29]
4	1000	0	(65,130)	TFOCS [29]

observed that the completions $\hat{G}$ frequently fail to satisfy $\mathrm{rank}(\hat{G}) = \mathrm{rank}(\mathcal{D})$ whenever Ω, d, W, V, and K are such that G is not overdetermined by G_Ω. The results are summarized in Table I. Here,

$$\text{“failure”} \ :\Leftrightarrow \ \max_{ij} |\hat{G}_{ij} - G_{ij}| \geqslant 10^{-3}, \tag{10}$$

with $\hat{G}$ and G denoting the estimated and the correct Gram matrix, respectively. “Start point” refers to the number of states and measurements we use to start Algorithm 1. These start points are chosen such that P is globally completable. If the trace minimization (8) fails to satisfy $\mathrm{rank}(\hat{G}) = \mathrm{rank}(\mathcal{D})$, we alternatingly add a new state or a new measurement and rerun the trace minimization (8); cf. Algorithm 1. More details are provided in the appendix.

V. DISCUSSION

If we assume the following, then we could prove that Algorithm 1 returns the correct Gram matrix:

(1) Asymptotic limit. That is, unbounded measurement repetitions, so that $f_{k|wv} = \mathrm{tr}(\rho_w E_{vk})$.

(2) Informationally complete states and measurements. The unknown matrices $(\rho_w)_w$ and $(E_{vk})_{vk}$ separately span $\mathrm{Herm}(\mathcal{H})$ [see Eq. (3)].

(3) Uniqueness. Our *a priori* knowledge $G_\Omega = \vec{\mathcal{K}}$ implies uniqueness of the Gram matrix G satisfying $\mathrm{rank}(G) = \mathrm{rank}(\mathcal{D})$ [27].

Note that these assumptions are nonexotic and strictly weaker than the assumptions made in asymptotic state tomography and measurement tomography. For instance, in case of state tomography all measurement operators are assumed to be known and consequently, all of the measurement Gram matrix G_m is known *a priori*. For global completability of the states and the measurements on the other hand it suffices to know (for example) that the measurements we perform are projective and nondegenerate. Such properties are easily enforced by only fixing a block diagonal of G_m (see [27]).

Knowing G is important because G suffices to determine the states and the measurements up to simultaneous rotations in $\mathrm{Herm}(\mathcal{H})$. Experiments involving processes $\{\mathcal{E}_j\}_j$, a single state ρ and a single POVM $(E_k)_k$ can be reduced to the considered prepare and measure setup by focusing on states $(\prod_s \mathcal{E}_{j_s})\rho$ and measurements $(\prod_s \mathcal{E}_{i_s}^\dagger)E_k$. Once these states and POVMs are estimated, they can be used as anchors for conventional process tomography.

Explicit density matrices and measurement operators can be found by heuristic algorithms (e.g., [16]) that take G as input. To deal with finite measurement repetitions, we would need to weaken the sharp constraints $f_{k|wv} = \mathrm{tr}(\rho_w E_{vk})$, e.g., by enforcing $\mathrm{tr}(\rho_w E_{vk}) \in [f_{k|wv} - \varepsilon, f_{k|wv} + \varepsilon]$ for some $\varepsilon > 0$. Assume that under these constraints Algorithm 1 returns $\hat{G}$ satisfying $\mathrm{rank}(\hat{G}) \leqslant \mathrm{Herm}(\mathcal{H}) = d^2$ where d denotes the dimension of the underlying Hilbert space (assumed to be known). Then, as long as $\varepsilon \ll 1$, we suspect that $\hat{G}$ approximates the correct Gram matrix G well if the states and the measurements are somewhat spread out in the cone of positive semidefinite matrices. It is an important open problem to investigate the validity of this intuition and to find convenient upper bounds for the distance between $\hat{G}$ and G. However, this problem is much more challenging than the analogous problem in state tomography.

ACKNOWLEDGMENTS

I would like to thank Johan Åberg for his continuous support. I wish to express my gratitude to Matthias Baur, Matthias Christandl, Matthew Coudron, Dejan Dukaric, Philippe Faist, Hamza and Omar Fawzi, David Gross, Patrick Pletscher, Renato Renner, Lars Steffen, David Sutter, Lídia del Rio, Michael Walter, Sharon Wulff, and Mário Ziman for fruitful discussions. I acknowledge support from the Swiss NSF through a fellowship and through the National Centre of Competence in Research “Quantum Science and Technology.”

APPENDIX A: COMPUTING THE BOUND R_QM

The purpose of this section is the computation of the upper bound R_QM from Eq. (8) in the main text. We start with

$$\|G\| \leqslant \|G\|_2, \tag{A1}$$

where $\|G\|_2^2 = \mathrm{tr}(G^* G)$, which holds true for general matrices. The bound (A1) is tight for $G \in \mathcal{G}_\mathrm{QM}$ because for every Hilbert space dimension d and for every rank of G, we can choose the vectors $\vec{\rho}_1, \ldots, \vec{E}_\mathrm{VK}$ [the columns of P; cf. Eq. (2) in the main text] such that they are almost parallel to $\mathbb{RI}$. Thus, for any choice of $\mathrm{rank}(G) = \mathrm{rank}(P)$, G can become arbitrarily close to a positive semidefinite rank-1 matrix (corresponds to all columns of P being parallel). Thus, the vector of eigenvalues of G becomes arbitrarily close to the vector $(\|G\|, 0, \ldots, 0)^T$, i.e., $\|G\|$ becomes arbitrarily close to $\|G\|_2$. Consequently, the upper bound (A1) is tight. We continue by observing that

$$\|G\|^2 \leqslant \|G\|_2^2 = \|P^T P\|_2^2 \leqslant \|P\|_2^4 = \left(\sum_{j=1}^{W+VK} \|\vec{P}_{j,\cdot}\|_2^2 \right)^2$$
$$= \left(\sum_{w=1}^{W} \|\rho_w\|_2^2 + \sum_{v=1}^{V} \sum_{k=1}^{K} \|E_{vk}\|_2^2 \right)^2. \tag{A2}$$

In the second inequality, we have used the submultiplicativity of the Hilbert-Schmidt norm. The Hilbert-Schmidt norm of quantum states is lower bounded by the norm of the maximally mixed state and upper bounded by the norm of pure states. Consequently,

$$\|\rho_j\|_2 \in [1/\sqrt{d}, 1] \Rightarrow \|\rho_j\|_2^2 \leqslant 1. \tag{A3}$$

The condition $\sum_k E_{vk} = \mathbb{I}$ implies

$$d = \|\mathbb{I}\|_2^2 = \left\| \sum_k E_{vk} \right\|_2^2 = \sum_k \|E_{vk}\|_2^2 + \sum_{k \neq q} \mathrm{tr}\,(E_{vk} E_{vq}), \tag{A4}$$

and therefore,

$$\sum_k \|E_{vk}\|_2^2 \leqslant d, \tag{A5}$$

because $\mathrm{tr}\,(MN) \geqslant 0$ whenever $M, N \geqslant 0$. This upper bound is tight because it is achieved by projective, nondegenerate measurements. Using Eqs. (A5) and (A3) in Eq. (A2), we arrive at

$$R_{\mathrm{QM}} = W + Vd. \tag{A6}$$

APPENDIX B: FINDING THE CONVEX RELAXATION WITH RESPECT TO $S^+ \cap B_{\|.\| \leqslant 1}$

Fazel, Hindi, and Boyd proved [26] that $\|\cdot\|_1$ is the convex envelope of the rank function on the set of matrices X with $\|X\| \leqslant 1$. In the following we present a modification of their argument to show that the trace (and hence still the trace norm $\|\cdot\|_1$) is the convex envelope of the rank function when restricting the above ball of matrices $\|X\| \leqslant 1$ to its intersection with the cone of positive semidefinite matrices, i.e., $X \in S^+ \cap B_{\|.\| \leqslant 1}$.

Recall that for an arbitrary function $f : \mathcal{C} \to \mathbb{R}, \mathcal{C}$ convex,

$$f^*(y) = \sup\{\langle y, x\rangle - f(x) \mid x \in \mathcal{C}\},$$

is its conjugate. The convex envelope of the rank function with respect to the convex set,

$$\mathcal{C} := \{X \in \mathbb{R}^{n \times n} \mid X \geqslant 0,\ \|X\| \leqslant 1\},$$

is rank**, i.e., the double conjugate [32] with respect to $\mathcal{C}$. Observe that

$$\mathrm{rank}^*(Y) = \sup_{X \in \mathcal{C}} \{\mathrm{tr}\,(YX) - \mathrm{rank}(X)\} = \max \left\{ \sup_{\substack{X \in \mathcal{C}, \\ \mathrm{rank}(X) = 1}} \{\mathrm{tr}\,(YX) - 1\}, \ldots, \sup_{\substack{X \in \mathcal{C}, \\ \mathrm{rank}(X) = n}} \{\mathrm{tr}\,(YX) - n\} \right\}. \tag{B1}$$

Here, Y is an arbitrary Hermitian $(n \times n)$ matrix (recall that the Hermitian matrices form the vector space carrying S^+). Due to their Hermiticity, both X and Y can be diagonalized orthogonally,

$$\begin{aligned} X &= \sum_{j=1}^n \varepsilon(X)_j |\varepsilon(X)_j\rangle\langle\varepsilon(X)_j|, \\ Y &= \sum_{j=1}^n \varepsilon(Y)_j |\varepsilon(Y)_j\rangle\langle\varepsilon(Y)_j|. \end{aligned} \tag{B2}$$

In the remainder we are assuming that all the eigenvalues are sorted descendingly. We observe that

$$\begin{aligned} \mathrm{tr}\,(YX) &= \sum_{i=1}^n \varepsilon(Y)_i \left(\sum_{j=1}^n \varepsilon(X)_j |\langle\varepsilon(X)_i|\varepsilon(Y)_j\rangle|^2 \right) \\ &= \vec{\varepsilon}(Y)^T\, Q\, \vec{\varepsilon}(X), \end{aligned} \tag{B3}$$

where Q is the doubly stochastic matrix $Q_{ij} = |\langle\varepsilon(X)_i|\varepsilon(Y)_j\rangle|^2$. Let s be such that $\varepsilon(Y)_j \geqslant 0$ for $j \leqslant s$ and $\varepsilon(Y)_j < 0$ for $j > s$. Consider a term "m", $m \leqslant s$, from Eq. (B1), i.e.,

$$\sup_{\substack{X \in \mathcal{C}, \\ \mathrm{rank}(X) = m}} \left\{\vec{\varepsilon}(Y)^T\, Q\, \vec{\varepsilon}(X) - m\right\}.$$

We claim that

$$\vec{\varepsilon}(Y)^T\, Q\, \vec{\varepsilon}(X) \leqslant \vec{\varepsilon}(Y)^T (\underbrace{1, \ldots, 1}_{m-\text{times}}, 0, \ldots, 0)^T, \forall Q, \vec{\varepsilon}(X), \tag{B4}$$

is a tight upper bound. Consider

$$\begin{aligned} &\text{maximize} \quad \vec{\varepsilon}(Y)^T\, Q\, \vec{\varepsilon}(X), \\ &\text{subject to} \quad Q \text{ doubly stochastic.} \end{aligned} \tag{B5}$$

The optimization problem (B5) is linear. It follows that the optimum is achieved at an extremal point. The doubly stochastic matrices form a polytope whose vertices are the permutation matrices (Birkhoff–von Neumann theorem). Hence, a solution Q to (B5) is a permutation matrix. An optimal choice is $Q = \mathbb{I}$ because $\vec{\varepsilon}(X)$ and $\vec{\varepsilon}(Y)$ are ordered descendingly, and

$$\langle x^{\downarrow}, y\rangle \leqslant \langle x^{\downarrow}, y^{\downarrow}\rangle,$$

for arbitrary vectors $x, y \in \mathbb{R}^n$ (see Corollary II.4.4 in Bathia's book [33]). Consequently, $Q = \mathbb{I}$, e.g., via

$$|\varepsilon(X)_j\rangle := |\varepsilon(Y)_j\rangle, \forall j,$$

solves (B5) independently of the specific values of $\vec{\varepsilon}(X)$ and $\vec{\varepsilon}(Y)$. To conclude the proof that Eq. (B4) describes a tight upper bound, we have to solve

$$\begin{aligned} &\text{maximize} \quad \vec{\varepsilon}(Y)^T\, \vec{\varepsilon}(X), \\ &\text{subject to} \quad X \geqslant 0,\ \|X\| \leqslant 1,\ \mathrm{rank}(X) = m. \end{aligned} \tag{B6}$$

The constraints imply

$$(0, \ldots, 0)^T \leqslant \vec{\varepsilon}(X) \leqslant (\underbrace{1, \ldots, 1}_{m-\text{times}}, 0, \ldots, 0)^T$$

(componentwise). As $m \leqslant s$, the left-hand side of Eq. (B4) becomes maximal for the componentwise maximum of $\vec{\varepsilon}(X)$,

i.e., for

$$\vec{\varepsilon}(X) = (\underbrace{1,\ldots,1}_{m-\text{times}},0,\ldots,0)^T. \tag{B7}$$

This proves that the upper bound in Eq. (B4) is correct and tight. In case of $m > s$, nonzero choices of $\varepsilon(X)_j$, $s < j \leqslant m$, lead to negative contributions to the left-hand side of Eq. (B4). Hence, in the case of $m > s$, the choice,

$$\vec{\varepsilon}(X) = (\underbrace{1,\ldots,1}_{s-\text{times}},0,\ldots,0)^T, \tag{B8}$$

realizes the tight upper bound. Combining Eqs. (B7) and (B8), we arrive at

$$\begin{aligned}&\sup_{\substack{X\in\mathcal{C},\\ \operatorname{rank}(X)=m}} \left\{\vec{\varepsilon}(Y)^T\, Q\, \vec{\varepsilon}(X)\right\} - m\\ &= \begin{cases} \sum_{j=1}^{m}[\vec{\varepsilon}(Y)_j - 1], & \text{for} \quad m \leqslant s,\\ -(m-s) + \sum_{j=1}^{s}[\vec{\varepsilon}(Y)_j - 1], & \text{for} \quad m > s. \end{cases}\end{aligned} \tag{B9}$$

To choose the optimal m [recall Eq. (B1)], we note that $m \mapsto m+1$ is profitable as long as $\vec{\varepsilon}(Y)_m - 1 \geqslant 0$. Using the compact notation $a_+ = \max\{a,0\}$, we conclude

$$\operatorname{rank}^*(Y) = \sum_{j=1}^{n}[\vec{\varepsilon}(Y)_j - 1]_+. \tag{B10}$$

To determine $\operatorname{rank}^{**}(Z)$, we can copy and paste the Fazel-Hindi-Boyd arguments [26]. We repeat them for the reader's convenience:

$$\operatorname{rank}^{**}(Z) = \sup_{Y=Y^T} \left\{\operatorname{tr}(ZY) - \operatorname{rank}^*(Y)\right\}, \tag{B11}$$

for all $Z \geqslant 0$ and $\|Z\| \leqslant 1$. Define

$$\Omega := \left\{\operatorname{tr}(ZY) - \operatorname{rank}^*(Y)\right\}. \tag{B12}$$

We consider the two cases $\|Y\| \leqslant 1$ and $\|Y\| > 1$,

$$\operatorname{rank}^{**}(Z) = \max\left\{\sup_{\substack{Y=Y^T,\\ \|Y\|\leqslant 1}} \Omega, \sup_{\substack{Y=Y^T,\\ \|Y\|> 1}} \Omega\right\}. \tag{B13}$$

Assume $\|Y\| \leqslant 1$. Then, as a consequence of Eq. (B10), $\operatorname{rank}^*(Y) = 0$, and therefore,

$$\sup_{\substack{Y=Y^T,\\ \|Y\|\leqslant 1}} \Omega = \sup_{\substack{Y=Y^T,\\ \|Y\|\leqslant 1}} \left\{\operatorname{tr}(ZY)\right\}. \tag{B14}$$

By von Neumann's trace theorem [34],

$$\operatorname{tr}(ZY) \leqslant \vec{\varepsilon}(Z)^T \vec{\varepsilon}(Y). \tag{B15}$$

This upper bound can be achieved by choosing Y, such that

$$|\varepsilon(Y)_j\rangle := |\varepsilon(Z)_j\rangle, \forall j.$$

Consequently, going back to Eq. (B14),

$$\begin{aligned}\sup \Omega = \max, \quad & \vec{\varepsilon}(Z)^T\, \vec{\varepsilon}(Y),\\ \text{subject to} \quad & |\vec{\varepsilon}(Y)_j| \leqslant 1,\ \forall j.\end{aligned} \tag{B16}$$

Since componentwise $0 \leqslant \vec{\varepsilon}(Z) \leqslant 1$, $\vec{\varepsilon}(Y) = (1,\ldots,1)^T$ is the optimal choice. It follows that

$$\sup_{\substack{Y=Y^T,\\ \|Y\|\leqslant 1}} \Omega = \sum_{j=1}^{n} \vec{\varepsilon}(Z)_j = \operatorname{tr}(Z). \tag{B17}$$

This concludes the discussion of $\|Y\| \leqslant 1$. Assume $\|Y\| > 1$. Note that $\operatorname{rank}^*(Y)$ is independent of our choice of the Y eigenvectors $|\varepsilon(Y)_j\rangle$. Hence, in Eq. (B11), we choose

$$|\varepsilon(Y)_j\rangle := |\varepsilon(Z)_j\rangle, \forall j,$$

as before to reach the von Neumann–upper bound in Eq. (B15). Thus,

$$\sup_{\substack{Y=Y^T,\\ \|Y\|> 1}} \Omega = \sup_{\varepsilon(Y)_1 \geqslant 1} \{\vec{\varepsilon}(Z)^T \vec{\varepsilon}(Y) - \operatorname{rank}^*(Y)\}, \tag{B18}$$

leading to

$$\sup_{\substack{Y=Y^T,\\ \|Y\|> 1}} \Omega = \sup_{\varepsilon(Y)_1 \geqslant 1} \sum_{j=1}^{n}[\varepsilon(Z)_j\varepsilon(Y)_j] - \sum_{j=1}^{s}\big[\varepsilon(Y)_j - 1\big]. \tag{B19}$$

Here, s is chosen such that $\varepsilon(Y)_j \geqslant 1$ for $j \leqslant s$ and $\varepsilon(Y)_j < 1$ for $j > s$. As in the derivation by Fazel and coworkers [26], we continue by the addition and the subtraction of $\sum_{j=1}^{n} \varepsilon(Z)_j$:

$$\begin{aligned}\sup_{\substack{Y=Y^T,\\ \|Y\|> 1}} \Omega = & \sup_{\varepsilon(Y)_1 \geqslant 1} \sum_{j=1}^{n}[\varepsilon(Z)_j\varepsilon(Y)_j] - \sum_{j=1}^{s}[\varepsilon(Y)_j - 1]\\ & - \sum_{j=1}^{n}\varepsilon(Z)_j + \sum_{j=1}^{n}\varepsilon(Z)_j\\ = & \sup_{\varepsilon(Y)_1 \geqslant 1} \sum_{j=1}^{s}[\varepsilon(Y)_j - 1][\varepsilon(Z)_j - 1]\\ & + \sum_{j=s+1}^{n}[\varepsilon(Y)_j - 1]\varepsilon(Z)_j + \sum_{j=1}^{n}\varepsilon(Z)_j.\end{aligned} \tag{B20}$$

In this last expression, the first sum is negative semidefinite because $\|Z\| \leqslant 1$, and the second sum is negative semidefinite because by definition of s, $\varepsilon(Y)_j \leqslant 1$ for all $j > s$. Therefore,

$$\sup_{\substack{Y=Y^T,\\ \|Y\|> 1}} \Omega \leqslant \sum_{j=1}^{n}\varepsilon(Z)_j = \operatorname{tr}(Z). \tag{B21}$$

Hence, using Y with $\|Y\| > 1$ brings no advantage [compare Eqs. (B17) and (B21)]. Going back to Eq. (B13), we conclude

$$\operatorname{rank}^{**}(Z) = \operatorname{tr}(Z), \tag{B22}$$

i.e., the convex envelope of the matrix rank function over the set $S^+ \cap B_{\|.\|\leqslant 1}$ is the matrix trace.

APPENDIX C: NUMERICAL EXPERIMENTS

To perform the numerical experiments, we need to sample explicit states and measurements. We proceeded by choosing

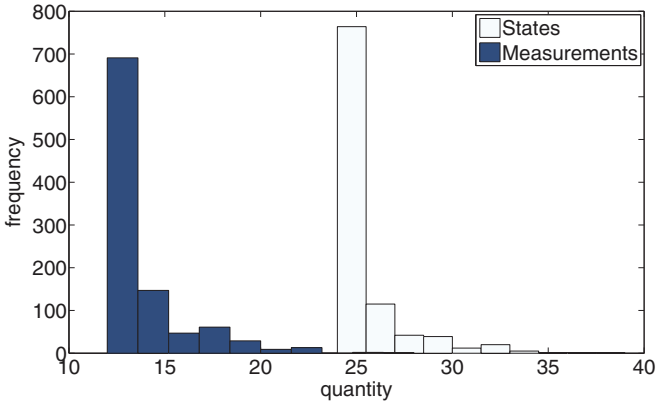

FIG. 2. (Color online) Histogram for $\dim(\mathcal{H}) = 2$ illustrating how many states and measurements were finally needed to find the correct state-measurement Gram matrix.

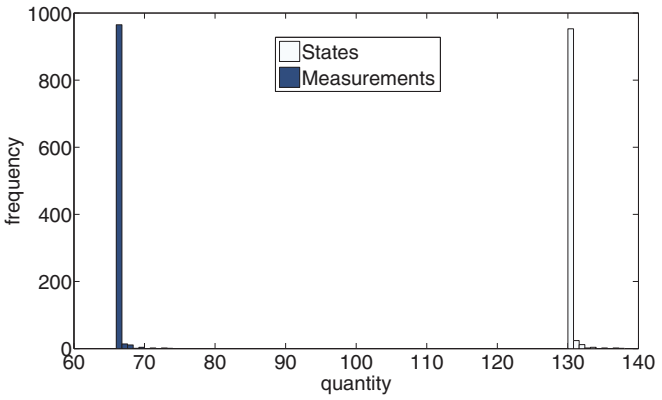

FIG. 3. (Color online) Histogram for $\dim(\mathcal{H}) = 4$ illustrating how many states and measurements were finally needed to find the correct state-measurement Gram matrix.

pure states from the Haar measure and by rotating a reference projective, nondegenerate measurement according to the Haar measure. In Algorithm 1 in the main text we need to compare the ranks of matrices. In principle, the rank of a matrix is equal to the number of its nonzero singular values. However, due to small numerical fluctuations in the solutions G, this definition is too strict. Rather, one should tolerate small variations by setting to zero singular values that are very small. We need to compare the rank of G with the rank of $\mathcal{D}$. We proceed by defining a threshold $\tau := 10^{-4}$ and

$$\check{s} := (s_{\mathrm{rank}\mathcal{D}+1}, \ldots, s_N), \tag{C1}$$

with s_j denoting the singular values of G_{st} (sorted descendingly). Then, we choose the following criterion to decide whether or not the ranks of G and $\mathcal{D}$ agree:

$$\mathrm{rank}(G) \approx_\tau \mathrm{rank}(\mathcal{D}) \;:\Leftrightarrow\; \|\check{s}\|_2 \leqslant \tau. \tag{C2}$$

When running Algorithm 1, we start with a specific number of unknown states and measurements. This is specified by "start point" in Table I. If the trace minimization does not manage to find the solution satisfying the rank condition, then we need to add new states and measurements (see Algorithm 1). The histograms in Figs. 2 and 3 illustrate how often we had to add new states and measurements to the initial situation when the Hilbert space was two dimensional or four dimensional. In case of the three-dimensional simulations we chose the start point large enough (to save computation time) so that we never had to increase the number of states or measurements (the corresponding histogram is trivial). From the perspective of state tomography, the required number of states and measurements appears to be large. However, in the most general cases where all the states and measurements lie in the interior of the cone of positive semidefinite matrices, large numbers of states and measurements are really needed (see [27]) to uniquely specify the quantum model, respectively, the Gram matrix. This changes as soon as enough states and measurements are rank deficient because in these situations, the boundary of the cone of positive semidefinite matrices can help to uniquely specify the state-measurement Gram matrix.

[1] S. T. Merkel, J. M. Gambetta, J. A. Smolin, S. Poletto, A. D. Córcoles, B. R. Johnson, C. A. Ryan, and M. Steffen, Phys. Rev. A **87**, 062119 (2013).
[2] R. Blume-Kohout, J. K. Gamble, E. Nielsen, J. Mizrahi, J. D. Sterk, and P. Maunz, arXiv:1310.4492.
[3] A. Acín, N. Gisin, and L. Masanes, Phys. Rev. Lett. **97**, 120405 (2006).
[4] I. Gerhardt, Q. Liu, A. Lamas-Linares, J. Skaar, V. Scarani, V. Makarov, and C. Kurtsiefer, Phys. Rev. Lett. **107**, 170404 (2011).
[5] D. Rosset, R. Ferretti-Schöbitz, J. D. Bancal, N. Gisin, and Y. C. Liang, Phys. Rev. A **86**, 062325 (2012).
[6] M. Christandl and R. Renner, Phys. Rev. Lett. **109**, 120403 (2012).
[7] R. Blume-Kohout, arXiv:1202.5270.
[8] T. Sugiyama, P. S. Turner, and M. Murao, Phys. Rev. Lett. **111**, 160406 (2013).
[9] J. Medford, J. Beil, J. M. Taylor, S. D. Bartlett, A. C. Doherty, E. I. Rashba, D. P. DiVincenzo, H. Lu, A. C. Gossard, and C. M. Marcus, Nat. Nanotechnol. **8**, 654 (2013).
[10] M. Takahashi, S. D. Bartlett, and A. C. Doherty, Phys. Rev. A **88**, 022120 (2013).
[11] D. Mogilevtsev, J. Řeháček, and Z. Hradil, New J. Phys. **14**, 095001 (2012).
[12] A. M. Brańczyk, D. H. Mahler, L. A. Rozema, A. Darabi, A. M. Steinberg, and D. F. V. James, New J. Phys. **14**, 085003 (2012).
[13] S. Fiorini, S. Massar, S. Pokutta, H. R. Tiwary, and R. de Wolf, in *Proceedings of the 44th Symposium on Theory of Computing Conference (STOC)* (ACM, New York, 2012).
[14] J. Gouveia, P. A. Parrilo, and R. R. Thomas, Mathematics of Operations Research **38**, 248 (2013).
[15] In combination with informational completeness.

[16] C. Stark, arXiv:1210.1105.
[17] Jarosław Adam Miszczak, Zbigniew Puchała, Paweł Horodecki, Armin Uhlmann, and Karol Zyczkowski, Quantum Inf. Comput. **9**, 103 (2009).
[18] T. Moroder, M. Kleinmann, P. Schindler, T. Monz, O. Gühne, and R. Blatt, Phys. Rev. Lett. **110**, 180401 (2013).
[19] S. Kimmel, M. P. da Silva, C. A. Ryan, B. R. Johnson, and T. Ohki, Phys. Rev. X **4**, 011050 (2014).
[20] Regarding the experiment as a d-dimensional quantum mechanical black box, this assumption appears to be weak: as soon as we regard the experimental states and measurements as random matrices drawn from a continuous probability measure on Herm($\mathbb{C}^d$) then d^2 states and d^2 measurement operators will satisfy separate informational completeness with probability 1.
[21] N. Harrigan, T. Rudolph, and S. Aaronson, arXiv:0709.1149.
[22] This is reminiscent of compressed sensing [23].
[23] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, Phys. Rev. Lett. **105**, 150401 (2010).
[24] General rank minimization is NP hard [25].
[25] B. Recht, M. Fazel, and P. A. Parillo, SIAM Rev. **52**, 471 (2010).
[26] M. Fazel, H. Hindi, and S. Boyd, in *Proceedings of the American Control Conference* (IEEE, Washington, DC, 2001).
[27] C. Stark, arXiv:1209.6499.
[28] J. Sturm, Optimization Methods and Software **11–12**, 625 (1999).
[29] S. Becker, E. J. Candès, and M. Grant, Technical Report, Department of Statistics, Stanford University (Stanford University, Stanford, 2010).
[30] A. Singer and M. Cucuringu, SIAM J. Matrix Anal. Appl. **31**, 1621 (2010).
[31] F. J. Király, L. Theran, R. Tomioka, and T. Uno, arXiv:1211.4116.
[32] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms II* (Springer, New York, 2010).
[33] R. Bathia, *Matrix Analysis* (Springer, New York, 1997).
[34] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, 2013).