Supplementary Material: Projected gradient descent algorithms for quantum state tomography

Eliot Bolduc[1], George C. Knee[2], Erik M. Gauger[1], Jonathan Leach[1]*
[1] *SUPA, Institute of Photonics and Quantum Sciences, Heriot-Watt University, David Brewster Building, Edinburgh, EH14 4AS, UK and*
[2] *Department of Physics, University of Warwick, Coventry, CV4 7AL, UK*
(Dated: August 30, 2017)

In this supporting document, we provide additional background information and context. In particular, we compare log-likelihood functions and describe alternative norms to achieve the projection. Further, we discuss limitations of the CVX software package, and the special case of pure and quasi-pure state tomography.

Supplementary note 1: Multiplication factor

If the unit-trace property is not enforced at every iteration of a tomographic algorithm, the multiplication factor r can be left as a free parameter. The multiplication factor would be incorporated in the density matrix trace during the reconstruction process, that is $\mathrm{Tr}[\rho_{\text{free-trace}}] = r$, and one would normalise the free-trace density matrix only after its recovery. The latter strategy is particularly useful when the elements of the measurement matrix do not form a POVM, such that r cannot be easily estimated *a priori*, which is generally the case for an ill-conditioned measurement matrix.

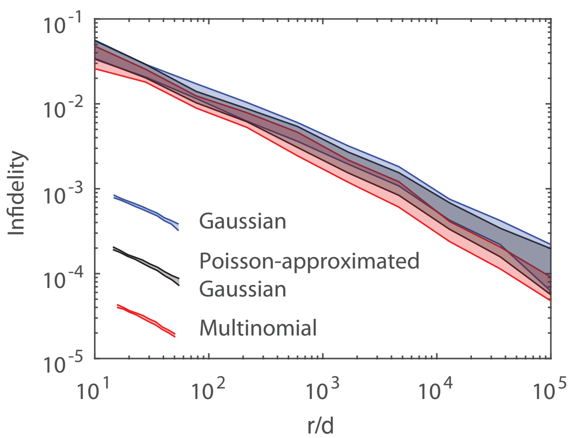

FIG. S1: Simulations of a three-qubit system with various levels of the average number of detector clicks per outcome. Here the noise on the simulated data is multinomial-distributed, and the measurement matrix is composed of canonical Pauli measurements. As expected, the multinomial likelihood (red) provides the best density matrix estimate over the entire range of average number of clicks per outcome. The Gaussian likelihood is implemented with PGDB, the Poisson-approximated Gaussian with SDPT3 and the multinomial likelihood with the DIA.

Supplementary note 2: Comparing log-likelihood functions

The log-likelihood can take many forms in general, and mainly depends on the type of noise affecting the measurement outcomes. Supplementary Figure S1 shows the extent to which deviating from a multinomial likelihood affects the fidelity of the reconstructed state. A multinomial noise model is the most accurate one when using a POVM measurement, and the negative log-likelihood takes the form

$$-\log \mathcal{L}_M(\rho) = -\mathbf{n}^T \log \mathbf{p}. \tag{1}$$

The outcomes that are equal to zero are simply discarded from the calculation to avoid $\log 0$. We normalise this log-likelihood with the number of projections N_+ that have non-zero outcomes. When a significant number of outcomes vanish, the log-likelihoods $\log \mathcal{L}_M$ and $\log \mathcal{L}_P$ become numerically unstable. The gradients of the log-likelihood functions are compactly written

$$\begin{aligned} -\nabla \log \mathcal{L}_M(\rho) &= M^\dagger \mathcal{A}^*, \\ -\nabla \log \mathcal{L}_P(\rho) &= 2P^\dagger \mathcal{A}^*, \end{aligned} \tag{2}$$

where the elements of the M and P matrices are defined as $M_{i,j} = -\mathcal{A}_{i,j}\eta_i$ (with $\eta_i = n_i/p_i$) and $P_{i,j} = \mathcal{A}_{i,j}\nu_i$.

Supplementary note 3: Alternative projections

In the main text, we chose the simplex projection $\mathcal{S}[\cdot]$ as the key part in our recursion relations, bringing unphysical matrices back into the physical space. There are other projections available, although we found that performance tended to worsen when using them. The *Frobenius norm* provides one such alternative projection $\mathcal{F}[\cdot]$ as derived in Ref. [1]. However, we found that this projection yields suboptimal results when implemented in a PGD algorithm, causing it to stall close to the maximum likelihood state, but before reaching it. Yet another option is the *nuclear norm* with projection $\mathcal{M}(\cdot)$, whose action is to set all negative eigenvalues of the density matrix to zero. However, this type of projection also fails to work as well as $\mathcal{S}[\cdot]$ and $\mathcal{F}[\cdot]$ in practice.

A possible reason for the supremacy of $\mathcal{S}[\cdot]$ is as follows. It is the most sophisticated projection, in the sense that it is based on the knowledge that the diagonal elements of the density matrix constitute probabilities. By contrast, $\mathcal{F}[\cdot]$ and $\mathcal{M}[\cdot]$ only exploit the positivity of the density matrix (and $\mathcal{M}[\cdot]$ requires explicit normalisation afterwards, too).

Supplementary note 4: Limitations of CVX

There exist efficient methods for solving convex problems (convex search set and convex cost function), which guarantee that any local maximum is also a global maximum. Quantum tomography is often, but not always, a convex problem: for example, the Gaussian approximation of a Poisson log-likelihood introduced by James *et al.* in Ref. [2] is non-concave, and so the cost function would be rejected by software packages such as CVX [3, 4], which operate in a 'disciplined convex programming' paradigm. Guarantees of global optimality are therefore not available, although PGD techniques will still work to find a local optimum.

Supplementary note 5: Pure and quasi-pure state tomography

The gradient descent techniques discussed in this Article are most suitable when no prior information is known about the quantum state at hand. In the rare special case where the unknown state is known to be pure, one may recover its state-vector using far fewer projectors (indeed, only $4d$ measurement operators are then sufficient for information completeness [5]). Restricting the search space to state vectors eliminates the need for a projection step, however, gradient descent approaches can still be useful: Candes *et al.* have applied such a gradient descent method to the problem of phase retrieval [6], which differs from state-vector tomography only in notation. Therefore, when the quantum state at hand is pure, one can apply their formalism to the quantum tomography problem. The computational complexity of a single step of their gradient descent algorithm is $O(Nd^2)$, with $N = 4d$ at minimum for informational completeness. A more involved version of the gradient descent approach for quasi-pure states has been developed by Becker *et al.*, who combine gradients and random matrices to converge towards a fixed-rank density matrix [7]. They were able to recover a simulated 16-qubit system by applying a low-memory version of their technique on a multicore computer.

Supplementary note 6: PGD with greater-than-1-rank operators

In the case where PGD is to be run on a measurement matrix with full rank operators, the gradients of the log likelihood functions take the following form

$$\begin{aligned}\nabla \log \mathcal{L}_M[\mathrm{vec}(\rho)] &= -A^T\boldsymbol{\eta}\ ,\\ \nabla \log \mathcal{L}_G[\mathrm{vec}(\rho)] &= 2A^T\boldsymbol{\Delta}\ \text{and}\\ \nabla \log \mathcal{L}_P[\mathrm{vec}(\rho)] &= 2A^T\boldsymbol{\nu}.,\end{aligned} \tag{3}$$

* email: J.Leach@hw.ac.uk

[1] Smolin, J. A., Gambetta, J. M. & Smith, G. Efficient method for computing the maximum-likelihood quantum state from measurements with additive gaussian noise. *Phys Rev Lett* **108**, 070502 (2012).

[2] James, D. F., Kwiat, P. G., Munro, W. J. & White, A. G. Measurement of qubits. *Phys Rev A* **64**, 052312 (2001).

[3] Grant, M., Boyd, S. & Ye, Y. CVX: Matlab software for disciplined convex programming (2008).

[4] Boyd, S. & Vandenberghe, L. *Convex optimization* (Cambridge university press, 2004).

[5] Heinosaari, T., Mazzarella, L. & Wolf, M. M. Quantum tomography under prior information. *Commun. Math. Phys.* **318**, 355–374 (2013).

[6] Candès, E. J., Li, X. & Soltanolkotabi, M. Phase retrieval via Wirtinger flow: Theory and algorithms. *Information Theory, IEEE Transactions on* **61**, 1985–2007 (2015).

[7] Becker, S., Cevher, V. & Kyrillidis, A. Randomized singular value projection. *Proceedings of the 10th International Conference on Sampling Theory and Approximation (SampTA)* (2013).