Adaptive Bayesian and frequentist data processing for quantum tomography

Giacomo M. D'Ariano [a,b], Daniele F. Magnani [a,c,*], Paolo Perinotti [a,b]

[a] *QUIT Group, Dipartimento di Fisica "A. Volta", via Bassi 6, 27100 Pavia, Italy*
[b] *Istituto Nazionale di Fisica Nucleare (INFN), Sezione di Pavia, via Bassi 6, 27100 Pavia, Italy*
[c] *Consorzio Nazionale Interuniversitario per le Scienze Fisiche della Materia (CNISM), unità di Pavia, via Bassi 6, 27100 Pavia, Italy*

ARTICLE INFO

ABSTRACT

The outcome statistics of an *informationally complete* quantum measurement for a system in a given state can be used to evaluate the ensemble expectation of any linear operator in the same state, by averaging a function of the outcomes that depends on the specific operator. Here we introduce two novel data-processing strategies, non-linear in the frequencies, which lead to faster convergence to theoretical expectations.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

In Quantum Mechanics measuring a single observable provides only partial information about the state of the measured system. According to the Born interpretation, the quantum state is a rule for evaluating the outcome probabilities in all conceivable measurements, and a complete information about the quantum state requires a thorough outcome statistics for a *quorum* of observables, or for a suitable *informationally complete measurement* (shortly info-complete) [1,2], in conjunction with a suitable data-processing, as it is done in quantum tomography (for a review see Ref. [3]). There are two main classes of approaches in quantum tomography: (a) Averaging "patterns functions", a method initiated in Ref. [4]; (b) Maximum likelihood techniques [5].

Method (a) has the advantage of providing any expectation value, e.g. a single density matrix element, without the need of estimating the entire density operator. However, the estimated full matrix is not necessarily positive, which is not a serious drawback, since the non-positivity falls within a small fluctuation for large numbers of data.

Method (b) has the advantage of providing a positive density operator, with smaller fluctuations, however, it has the more serious drawback of needing to estimate the full density matrix, while

is exponentially large versus the number of systems, and, in the infinite-dimensional case needs a dimensionality cutoff which introduce a bias that is under control only if there is some prior knowledge of the state.

In a recent paper [6] the optimal data-processing for evaluating ensemble averages from experimental outcomes was derived for a completely general setting within a Bayesian scheme that assumes a prior probability distribution of states. Using as optimality criterion the rate of estimated-to-theoretical convergence of averages, the optimal data-processing itself depends on the prior distribution of states.

The purpose of the present Letter is to exploit the dependence of the optimal data-processing on the prior distribution of states, in order to improve the convergence rate using an adaptive data-processing scheme. We will consider info-complete measurements — more generally than a quorum of observables — whose statistics allows to reconstruct all possible ensemble averages. Estimation of the quantum state itself is equivalent to the estimation of all possible ensemble averages. We will adopt the natural figure of merit used in Ref. [6], which, in the present context, represents the estimated-to-theoretical convergence rate (in Hilbert–Schmidt distance) of the state. As we will see, exploiting the dependence of the optimal data-processing on the prior state leads to two different data processing strategies, which both improve the convergence rate compared to the standard tomographic procedures, and are easily implementable and computationally efficient:

---

* Corresponding author.
   E-mail address: magnani@qubit.it (D.F. Magnani).

Method 1 (Bayesian iterative procedure): Bayesian update of the prior distribution after the first state reconstruction, with subsequent iteration of the optimization.

Method 2 (Frequentist approach): replace the theoretical probability distribution of the info-complete in the optimal data-processing with the experimental frequencies.

We will see that numerical simulations carried out with both methods show relevant improvement of convergence compared to the plain non-adaptive processing of Ref. [6].

The Letter is organized as follows. In Section 2 we re-derive the optimal data-processing for given prior distribution of Ref. [6] within an improved theoretical framework. In Sections 3 and 4 we introduce Methods 1 and 2, respectively. Finally, in Section 5, we present numerical simulations for testing both methods in comparison with the original plain non-adaptive data-processing, and in Section 6 we end the Letter with concluding remarks.

2. Optimization of the data processing

In the modern formulation of Quantum Mechanics, the state of a quantum system associated to a d-dimensional Hilbert space $\mathcal{H} \sim \mathbb{C}^d$ is represented by a *density matrix*, namely a positive operator $\rho \geqslant 0$ with $\text{Tr}[\rho] = 1$. The Born formula provides the probabilities of outcomes in a quantum measurement in terms of the state ρ as follows

$$p(i|\rho) := \text{Tr}[\rho P_i], \tag{1}$$

where the POVM (P_i) (Positive Operator Valued Measure) is a set of (generally non-orthogonal) positive operators $P_i \geqslant 0$ resolving the identity as $\sum_{i=1}^N P_i = I$, thus guaranteeing positivity and normalization of probabilities.

The linear span of the POVM elements P_i, defined as $\mathcal{S} := \text{Span}\{P_i\}_{1 \leqslant i \leqslant n}$, is a linear subspace of the space $\mathcal{L}(\mathcal{H})$ of linear operators on $\mathcal{H}$, and we will take as a canonical basis in $\mathcal{L}(\mathcal{H})$ the operators $|m\rangle\langle n|$, where $|n\rangle$ is an orthonormal basis thus representing operators X by the vectors of their matrix elements $X_{m,n} = \langle m|X|n\rangle$. A POVM is *info-complete* if $\mathcal{S} \equiv \mathcal{L}(\mathcal{H})$, namely all operators $X \in \mathcal{L}(\mathcal{H})$ can be expanded on the POVM elements, and it is possible to determine all ensemble averages $\langle X\rangle_\rho$, as in Quantum Tomography. For each complex operator $X \in \mathcal{S}$ the following decomposition holds

$$X = \sum_{i=1}^N f_i[X] P_i, \tag{2}$$

where $f_i[X]$ is not unique if the set $\{P_i\}$ is over-complete.

With the above expressions we can write the ensemble average of X as follows:

$$\langle X\rangle_\rho := \text{Tr}[X\rho] = \sum_{i=1}^N f_i[X] p(i|\rho), \tag{3}$$

with the following statistical error

$$(\delta X^2)_\rho := \sum_{i=1}^N |f_i[X]|^2 p(i|\rho) - |\langle X\rangle_\rho|^2. \tag{4}$$

In a Bayesian scheme one has an *a priori* ensemble $\mathcal{E} := \{\rho_i, p_i\}$ of possible states ρ_i of the quantum system occurring with probability p_i. We want to minimize the average statistical error on $\mathcal{E}$ in the determination of the expectation value of X, namely the variance

$$(\delta X^2)_\mathcal{E} := \sum_{i=1}^N |f_i[X]|^2 p(i|\rho_\varepsilon) - \overline{|\langle X\rangle|_\varepsilon^2}, \tag{5}$$

where $\rho_\varepsilon = \sum_i p_i \rho_i$ and $\overline{|\langle X\rangle|_\varepsilon^2} = \sum_i p_i |\text{Tr}[\rho_i X]|^2$ is the squared modulus of the expectation of X averaged over the states in the ensemble (since this term depends only on the ensemble it will be neglected from now on). Using Eq. (1) the first term in Eq. (5) can be rewritten as

$$\Sigma_f(X) := \sum_{i=1}^N |f_i[X]|^2 \text{Tr}[P_i \rho_\varepsilon]. \tag{6}$$

Given a POVM (P_i), it is possible to define a linear map Λ from an abstract N-dimensional space $\mathcal{K}$ of coefficient vectors $\mathbf{c} \in \mathcal{K}$ to $\mathcal{L}(\mathcal{H})$, with range $\mathcal{S}$:

$$\Lambda\mathbf{c} = \sum_{i=1}^N c_i P_i, \tag{7}$$

so that using the canonical basis in $\mathcal{K}$, Λ has matrix elements $\Lambda_{mn,i} = (P_i)_{mn}$. A generalized inverse (shortly g-inverse) of Λ is any matrix Γ representing linear operators from $\mathcal{L}(\mathcal{H})$ to $\mathcal{K}$ such that the following identity holds

$$\Lambda\Gamma\Lambda = \Lambda. \tag{8}$$

Notice that the matrix elements $(\Gamma_{i,mn})$ of Γ define a set of operators D_i with matrix elements $(D_i)_{mn} := \Gamma^*_{i,mn}$. The role of g-inverse Γ is assessed by the two following important theorems

Theorem 1. *The following statements are equivalent*

1. *Γ is a g-inverse of Λ.*
2. *For all $\mathbf{y} \in \text{Rng}(\Lambda)$, $\mathbf{x} = \Gamma\mathbf{y}$ is a solution of the equation $\Lambda\mathbf{x} = \mathbf{y}$.*

Proof. See Ref. [7]. □

Theorem 2. *For all g-inverse Γ of Λ all solutions of $\Lambda\mathbf{x} = \mathbf{y}$ are of the form*

$$\mathbf{x} = \Gamma\mathbf{y} + (I - \Gamma\Lambda)\mathbf{z}, \tag{9}$$

with arbitrary $\mathbf{z}$.

Proof. See Ref. [7]. □

We now define a norm in $\mathcal{K}$ as follows

$$\|\mathbf{c}\|_\pi^2 := \sum_{i=1}^N |c_i|^2 \pi_{ii}, \tag{10}$$

where $\pi_{ij} = \delta_{ij} \pi_{ii}$ is a positive matrix which is diagonal in the canonical basis in $\mathcal{K}$. In terms of π we define the *minimum norm g-inverses* Γ that satisfy [8]

$$\pi\Gamma\Lambda = \Lambda^\dagger \Gamma^\dagger \pi. \tag{11}$$

Notice that the present definition of minimum norm g-inverse requires that the norm is induced by a scalar product (in our case $\vec{a} \cdot \vec{b} := \sum_{i=1}^N a_i^* \pi_{ii} b_i$). We will now prove the following crucial theorem

Theorem 3. *The following assertions are equivalent*

1. *Γ is a minimum norm g-inverse of Λ.*
2. *For all $\mathbf{y} \in \text{Rng}(\Lambda)$, $\mathbf{x} = \Gamma\mathbf{y}$ is a solution of the equation $\Lambda\mathbf{x} = \mathbf{y}$ with minimum norm.*

Proof. We first prove that $1 \Rightarrow 2$. For Γ g-inverse of Λ, one has due to Theorem 2

$$\|\Gamma\mathbf{y} + (I - \Gamma\Lambda)\mathbf{z}\|_\pi^2$$
$$= [\mathbf{y}^\dagger\Gamma^\dagger + \mathbf{z}^\dagger(I - \Lambda^\dagger\Gamma^\dagger)]\pi[\Gamma\mathbf{y} + (I - \Gamma\Lambda)\mathbf{z}]$$
$$= \|\Gamma\mathbf{y}\|_\pi^2 + \|(I - \Gamma\Lambda)\mathbf{z}\|_\pi^2$$
$$+ \mathbf{z}^\dagger(I - \Lambda^\dagger\Gamma^\dagger)\pi\Gamma\mathbf{y} + \mathbf{y}^\dagger\Gamma^\dagger\pi(I - \Gamma\Lambda)\mathbf{z}. \quad (12)$$

Since by hypothesis $\mathbf{y} \in \text{Rng}(\Lambda)$, then $\mathbf{y} = \Lambda\mathbf{u}$ for some $\mathbf{u}$ in $\mathcal{K}$. For a minimum norm g-inverse Γ as in the hypothesis, due to Eq. (11) one has

$$\mathbf{z}^\dagger(I - \Lambda^\dagger\Gamma^\dagger)\pi\Gamma\Lambda\mathbf{u} + \mathbf{u}^\dagger\Lambda^\dagger\Gamma^\dagger\pi(I - \Gamma\Lambda)\mathbf{z}$$
$$= \mathbf{z}^\dagger(I - \Lambda^\dagger\Gamma^\dagger)\Lambda^\dagger\Gamma^\dagger\pi\mathbf{u} + \mathbf{u}^\dagger\pi\Gamma\Lambda(I - \Gamma\Lambda)\mathbf{z} = 0, \quad (13)$$

where the last equality is due to Eq. (8). Finally, this proves that

$$\|\Gamma\mathbf{y} + (I - \Gamma\Lambda)\mathbf{z}\|_\pi^2 = \|\Gamma\mathbf{y}\|_\pi^2 + \|(I - \Gamma\Lambda)\mathbf{z}\|_\pi^2 \geqslant \|\Gamma\mathbf{y}\|_\pi^2, \quad (14)$$

namely the solution $\mathbf{x} = \Gamma\mathbf{y}$ is minimum-norm.

Now we prove $2 \Rightarrow 1$. If $\mathbf{x} = \Gamma\mathbf{y}$ is a solution of $\Lambda\mathbf{x} = \mathbf{y}$ for all $\mathbf{y} \in \text{Rng}(\Lambda)$, by Theorem 1 Γ is a g-inverse of Λ, namely $\Lambda\Gamma\Lambda = \Lambda$. Then if $\Gamma\mathbf{y}$ is minimum norm solution of $|\Lambda\mathbf{x} = \mathbf{y}|$ then due to Theorem 2

$$\|\Gamma\mathbf{y}\|_\pi^2 \leqslant \|\Gamma\mathbf{y} + (I - \Gamma\Lambda)\mathbf{z}\|_\pi^2 \quad (15)$$

for all $\mathbf{y} \in \text{Rng}(\Lambda)$ and for all $\mathbf{z}$ one has

$$0 \leqslant \|(I - \Gamma\Lambda)\mathbf{z}\|_\pi^2 + \mathbf{z}^\dagger(I - \Lambda^\dagger\Gamma^\dagger)\pi\Gamma\mathbf{y} + \mathbf{y}^\dagger\Gamma^\dagger\pi(I - \Gamma\Lambda)\mathbf{z}. \quad (16)$$

Since an arbitrary $\mathbf{y} \in \text{Rng}(\Lambda)$ is $\Lambda\mathbf{u}$ for arbitrary $\mathbf{u}$, the second term in Eq. (16) becomes

$$\mathbf{z}^\dagger(I - \Lambda^\dagger\Gamma^\dagger)\pi\Gamma\Lambda\mathbf{u} + \mathbf{u}^\dagger\Lambda^\dagger\Gamma^\dagger\pi(I - \Gamma\Lambda)\mathbf{z}$$
$$= 2\Re(\mathbf{z}^\dagger(I - \Lambda^\dagger\Gamma^\dagger)\pi\Gamma\Lambda\mathbf{u}). \quad (17)$$

Let us keep $\mathbf{z}$ fixed and multiply $\mathbf{u}$ by an arbitrary α. If the expression in Eq. (17) is not vanishing then taking $|\alpha|$ sufficiently large, for suitable phase one can contradict the bound in Eq. (16), hence $\Re(\mathbf{z}^\dagger(I - \Lambda^\dagger\Gamma^\dagger)\pi\Gamma\Lambda\mathbf{u}) = 0$ for all $\mathbf{u}$ and $\mathbf{z}$ and by the same reasoning $\Im(\mathbf{z}^\dagger(I - \Lambda^\dagger\Gamma^\dagger)\pi\Gamma\Lambda\mathbf{u}) = 0$ for all $\mathbf{u}$ and $\mathbf{z}$. We can then conclude that $(I - \Lambda^\dagger\Gamma^\dagger)\pi\Gamma\Lambda = \Lambda^\dagger\Gamma^\dagger\pi(I - \Gamma\Lambda) = 0$, and consequently $\pi\Gamma\Lambda = \Lambda^\dagger\Gamma^\dagger\pi$. $\quad\square$

Using Eq. (11), and considering that $\Sigma_f(X)$ is the norm of the vector of coefficients $\mathbf{f}[X]$ with $\pi_{ii} = \text{Tr}[P_i\rho_\varepsilon]$, it has been proved in [6] that the minimum noise is achieved by Γ corresponding to the set of operators D_i given by

$$D_i^{\text{opt}} := \Delta_i - \sum_{j=1}^N \{[(I - M)\pi(I - M)]^\ddagger \pi M\}_{ij}\Delta_j, \quad (18)$$

where Δ_i is the set of operators corresponding to the Moore–Penrose g-inverse Γ_{mp} of Λ, satisfying the properties

$$\Gamma_{\text{mp}}\Lambda = \Lambda^\dagger\Gamma_{\text{mp}}^\dagger, \qquad \Gamma_{\text{mp}}\Lambda\Gamma_{\text{mp}} = \Gamma_{\text{mp}}, \qquad \Gamma_{\text{mp}}^\dagger\Lambda^\dagger = \Lambda\Gamma_{\text{mp}}, \quad (19)$$

and $M := \Gamma_{\text{mp}}\Lambda = M^\dagger = M^2$. The symbol $X^\ddagger$ denotes the Moore–Penrose g-inverse of X. It is indeed easy to verify that

$$\Gamma_{\text{opt}} := \Gamma_{\text{mp}} - [(I - M)\pi(I - M)]^\ddagger \pi M\Gamma_{\text{mp}} \quad (20)$$

satisfies Eq. (11). Notice that being Γ_{opt} minimum norm independently of X, the statistical error is minimized by the same choice D_i^{opt} for all operators X.

When a N-outcomes POVM on a d-dimensional Hilbert space $\mathcal{H} \sim \mathbb{C}^d$ is info-complete the state ρ can be written as

$$\rho = \sum_{i=1}^N D_i p(i|\rho), \quad (21)$$

where D_i corresponds to any g-inverse Γ. It is then possible to reconstruct any state ρ using the statistics from measurements:

$$\rho = \sum_{i=1}^N p(i|\rho)D_i \cong \sum_{i=1}^N \nu_i D_i^{\text{opt}}, \quad (22)$$

where $\nu_i = \frac{n_i}{n_{\text{tot}}}$ is the experimental frequency of the ith outcome, n_i being the number of occurrence of the ith outcome, and $n_{\text{tot}} = \sum_i n_i$. By the law of large numbers we have that $\lim_{n_{\text{tot}}\to\infty} \nu_i = p(i|\rho)$. However, the convergence rate of $\tilde{\rho}$ to ρ depends on the choice of D_i. It turns out [9] that the choice D_i^{opt}, corresponding to Γ_{opt}, is the one with the fastest convergence (in average over all possible experimental outcomes) in the Hilbert–Schmidt distance, defined as follows

$$\|\tilde{\rho} - \rho_2\|_2^2 := \text{Tr}[(\tilde{\rho} - \rho)^2]. \quad (23)$$

This can be easily proved considering that the Hilbert–Schmidt distance can be written as the sum of the variances $\delta(|m\rangle\langle n|)^2$, and all of the summands are minimized by the choice of minimum-norm $\Gamma = \Gamma_{\text{opt}}$.

3. The Bayesian iterative procedure

In this section we describe the iterative estimation procedure based on the update of the prior information by means of the state reconstruction provided by experimental data. Here we provide an algorithmic description of the procedure, that yields a self-consistent solution:

1. The protocol starts with the choice of *a priori* ensemble $\mathcal{E} := \{\rho_i, p_i\}$ (where ρ_i are states and p_i are their prior probabilities), with the corresponding density matrix $\rho^{(0)} := \rho_{\mathcal{E}}^{(0)} = \sum_i p_i\rho_i$, e.g. the one of the uniform ensemble of all pure states $\rho^{(0)} = I/d$.
2. Using $\rho^{(0)}$ it is possible to calculate the diagonal matrix with the probability of the different outcomes:

$$\pi_{ij} := \delta_{ij}\text{Tr}[P_i\rho^{(0)}]. \quad (24)$$

3. Using π_{ij} in Eq. (18) we can find the optimal g-inverse Γ_{opt} corresponding to D_i^{opt} associated with $\rho^{(0)}$.
4. Now the initial *a priori* density matrix $\rho^{(0)} \equiv \rho_{\mathcal{E}}$ will be updated as follows:

$$\rho^{(1)} = \sum_{i=1}^N \nu_i D_i^{\text{opt}}. \quad (25)$$

5. If $\rho^{(1)} \cong \rho^{(0)}$ within a given tolerable error ε then the average input state is $\tilde{\rho} := \rho^{(1)}$ and the procedure stops.
6. Otherwise after setting $\rho^{(0)} := \rho^{(1)}$ the procedure will go back to the step 2.

It is important to remark that at each step the matrices $\rho^{(1)}$ and D_i^{opt} are automatically self-adjoint and normalized: $\text{Tr}[\rho^{(1)}] = 1$ since for all i: $\text{Tr}[D_i^{\text{opt}}] = 1$ [6], however, they are not necessarily positive.

This protocol in principle provides reliable state reconstructions, however, its iterative character makes it less efficient than the one introduced in next section, since at any iterative step one has to calculate the Moore–Penrose g-inverse in Eq. (18), which is typically a time-consuming operation, especially for POVMs with a large number N of outcomes.

4. The frequentist approach

In this section we introduce the second processing strategy, based on the substitution of prior probabilities by experimental

frequencies in Eq. (11). While the previous protocol is essentially a Bayesian update, in this case the processing relies on the law of large numbers, namely on the fact that $\lim_{n_{tot}\to\infty}\nu_i = p(i|\rho)$, where the limit has to be understood in probability. We name this approach frequentist because it fits the frequentist interpretation of probabilities as approximations of experimental frequencies, avoiding prior probabilities, which are the signature of the Bayesian approach.

If we substitute the metric matrix π in Eq. (10) with the diagonal matrix of the frequencies ν_i, we get:

$$\nu\Gamma\Lambda = \Lambda^\dagger\Gamma^\dagger\nu \tag{26}$$

and following the same proof as for Eq.(18) we obtain the following expression of the optimal g-inverse Γ_ν satisfying condition Eq. (26), in terms of the corresponding operators $D_i^{(\nu)}$

$$D_i^{(\nu)} := \Delta_i - \sum_{j=1}^{N}\big\{\big[(I-M)\nu(I-M)\big]^\ddagger \nu M\big\}_{ij}\Delta_j \tag{27}$$

that is non-linear in the outcomes frequencies due to the Moore–Penrose g-inverse of $(I-M)\nu(I-M)$.

This protocol has the advantage that it requires only one evaluation of Moore–Penrose g-inverse, and it is then much faster − in terms of computational resources − than the iterative one introduced in the previous section. However, here generally $\mathrm{Tr}[D_i^{(\nu)}] \neq 1$, whence in addition to positivity of the estimated state $\tilde\rho$, also the normalization constraint is lost (but not hermiticity).

5. Numerical simulations

In order to test these two methods and to compare their performances with the plain un-updated procedure some Monte Carlo simulation have been performed. As an example, we considered the info-complete POVM composed by the following six elements

$$P_{\pm i} = \frac{1}{6}(I \pm \sigma_i), \tag{28}$$

$\sigma_0 = I$ and $\vec\sigma = (\sigma_x, \sigma_y, \sigma_z)$ denoting the usual Pauli matrices. The theoretical state is

$$\rho = \begin{pmatrix} \frac{4}{5} & \frac{1}{7}+\frac{i}{3} \\ \frac{1}{7}-\frac{i}{3} & \frac{1}{5} \end{pmatrix}$$
$$= \frac{1}{2}\left(I + \frac{2}{7}\sigma_x - \frac{2}{3}\sigma_y + \frac{3}{5}\sigma_z\right). \tag{29}$$

The simulation consists in 1000 experiments, each consisting in 1000 single-shot measurements, simulated by POVM events extraction according to the theoretical probabilities $p(\pm i|\rho) := \mathrm{Tr}[P_{\pm i}\rho]$. The number of iterations in the Bayesian processing is 10.

In Fig. 1 we show the histograms representing the number of experiments as a function of the Hilbert–Schmidt distance of the resulting state $\tilde\rho$ from the theoretical one ρ. The plots show a well evident shift of the histograms for both new processing methods towards small errors compared to the plain processing without updating. In Table 1 we summarize these considerations by showing the average Hilbert–Schmidt distance obtained with the three kinds of processing, along with the corresponding variance and the relative improvement of the figure of merit.

6. Conclusions

In conclusion, we have presented two novel data-processing strategies to improve convergence of estimation of ensemble average via info-complete measurements. The two approaches adaptively update the data-processing functions in a Bayesian and fre-
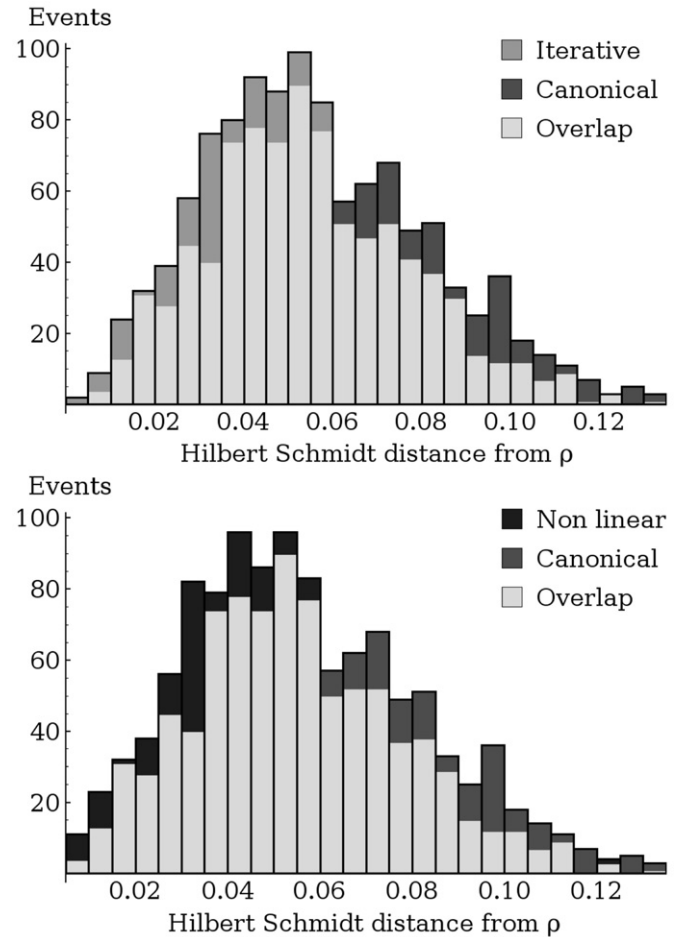

Fig. 1. Histograms representing the number of experiments versus the Hilbert–Schmidt distance of the resulting state from the theoretical one. Upper plot: the light gray bars correspond to the Bayesian processing, the dark grey correspond to the plain processing without updating, the white part is the overlap. Lower plot: the dark grey bars correspond to the frequentist processing method. Both plots show a well visible shift of the histograms corresponding to the new adaptive methods towards small errors compared to the plain processing without update. [For other data concerning plots see text.]

Table 1
Average Hilbert–Schmidt distance, variance σ of the histogram, and relative improvements compared to the plain un-updated procedure of the new data-processing strategies presented in the Letter. [For other data concerning this table see text.]

Procedure	$\langle$H.S. dist.$\rangle$	σ	$\Delta(\langle$H.S. dist.$\rangle)$	$\Delta(\sigma)$
Plain (no update)	0.06	0.03	–	–
Bayesian	0.05	0.02	−17%	−33.3%
Frequentist	0.05	0.02	−17%	−33.3%

quentist fashion, respectively, by substituting the prior probabilities with experimental frequencies (frequentist) and the prior state with the updated state (Bayesian). The two methods have been tested by numerical simulations, and both showed improved convergence rate compared to the original plain un-updated strategy. Clearly, further improvement is possible using both procedure together, however, this would be an higher-order correction.

References

[1] E. Progovečki, Int. J. Theor. Phys. 16 (1977) 321.
[2] P. Bush, Int. J. Theor. Phys. 30 (1991) 1217.
[3] G.M. D'Ariano, M.G.A. Paris, M.F. Sacchi, in: M.G.A. Paris, J. Řeháček (Eds.), Quantum State Estimation, Lecture Notes in Physics, Springer, Berlin, 2004, p. 7.
[4] G.M. D'Ariano, C. Macchiavello, M.G.A. Paris, Phys. Rev. A 50 (1994) 50.
[5] K. Banaszek, G.M. D'Ariano, M.G.A. Paris, M.F. Sacchi, Phys. Rev. A 61 (2000) R010304.
[6] G.M. D'Ariano, P. Perinotti, Phys. Rev. Lett. 98 (2007) 020403.
[7] R.B. Bapat, Linear Algebra and Linear Models, Springer, Berlin, 2000.
[8] G.M. D'Ariano, P. Perinotti, M.F. Sacchi, Phys. Rev. A 72 (2005) 042108.
[9] A.J. Scott, J. Phys. A 39 (2006) 13507.