

Optimal Data Processing for Quantum Measurements

G. M. D'Ariano and P. Perinotti

QUIT Group, Dipartimento di Fisica "A. Volta," via Bassi 6, I-27100 Pavia, Italy and CNISM
(Received 9 October 2006; published 12 January 2007)

We consider the general measurement scenario in which the ensemble average of an operator is determined via suitable data processing of the outcomes of a quantum measurement described by a positive operator-valued measure. We determine the optimal processing that minimizes the statistical error of the estimation.

DOI: [10.1103/PhysRevLett.98.020403](https://doi.org/10.1103/PhysRevLett.98.020403)

PACS numbers: 03.65.Ta, 03.65.Ca, 03.67.-a

A measurement in quantum mechanics is usually associated to an observable represented by a self-adjoint operator X on the Hilbert space \mathcal{H} of the quantum system [1], with the eigenvalues x_i defining the possible outcomes of the measurement. The probability distribution of the i th outcome is given by the Born rule

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

ρ being the density operator of the state and P_i denoting the orthogonal projectors in the spectral decomposition $X = \sum_{i=1}^N x_i P_i$ (for the sake of illustration here we consider only finite spectrum). Consequently, the expected value for the outcome-averaging over repeated measurements is given by the ensemble average $\langle X \rangle = \text{Tr}[\rho X]$, with statistical error proportional to the rms $\sqrt{\langle \Delta X^2 \rangle}$, with $\Delta X^2 := X^2 - \langle X \rangle^2$.

There are, however, more general kinds of measurements that can be performed in the lab, which are not necessarily associated to any observable, nevertheless enable the experimental determination of ensemble averages: these are the measurements that are described by POVM's. A POVM (positive operator-valued measure) is a set of (generally nonorthogonal) positive operators $P_i \geq 0$, $1 \leq i \leq N$ which resolve the identity $\sum_{i=1}^N P_i = I$ similarly to the orthogonal projectors of an observable, whence with the same Born rule (1). This more general class of quantum measurements includes also the description of optimal joint measurements of noncommuting observables [2,3], along with the measurements of parameters with no corresponding observable such as the phase of a harmonic oscillator [4], and many other practical measurements such as optimized discrimination of states for quantum communications [5], and, most interesting, the so-called informationally complete measurements [6], i.e., measurements that allow one to determine the density matrix of the state or any other desired ensemble average, as for the so-called quantum tomography [7]. Moreover, POVM's also allow to provide a full description of the measurement apparatus, including noisy channels before detection [8]. The POVM's are not just a theoretical tool, since there is a general quantum calibration procedure in order to deter-

mine experimentally the POVM of a measurement device by using a reliable standard [9].

How can we experimentally determine the ensemble average of the (generally complex) operator X using a POVM? Clearly this is possible if X can be expanded over the POVM elements (mathematically we denote this condition as $X \in \text{span}\{P_i\}_{i=1,N}$). This means that there exists a set of coefficients $f_i[X]$ such that

$$X = \sum_{i=1}^N f_i[X] P_i, \quad X \in \mathcal{S} := \text{span}\{P_i\}_{i=1,N}. \quad (2)$$

When $\mathcal{S} \equiv \mathcal{B}(\mathcal{H})$ (i.e., when all operators can be expanded over the POVM), then the measurement is informationally complete. Obviously, once the expansion (2) is established one can obtain the ensemble average of X by the following averaging:

$$\langle X \rangle = \sum_{i=1}^N f_i[X] p(i|\rho), \quad (3)$$

where the probability distribution is given in Eq. (1).

The above general measurement procedure opens the problem of finding the coefficients $f_i[X]$ in Eq. (2), namely, the data processing of the measurement outcomes needed to determine the ensemble average of X . In general the coefficients $f_i[X]$ are not unique [if $N > \dim(\mathcal{S})$], and one then wants to optimize the data processing according to a practical criterion, typically minimizing the statistical error. This problem has never been addressed in the general case, and its solution will be presented in this Letter. Notice that although the processing functions are intrinsically linear in the definition (2), there is no guarantee that the optimal ones are linear in X . However, as we will see, remarkably the optimal processing function is indeed linear in X , and depends only on the POVM and, in a Bayesian scheme, on the ensemble of possible input states (due to the simplicity and popularity of the Bayesian scheme, in this Letter we will restrict the analysis only to this scheme, postponing the analysis of the minimax strategy to another more technical publication: for a comparison between the two frameworks; see, for example, Ref. [10]). The derivation of the optimal data-processing

function requires some notions of frame theory [11,12] and linear algebra, which will be introduced in the first part of the Letter. Actually, for simplicity, instead of presenting the actual derivation we will first prove uniqueness of the optimal processing, then we present the result and prove that it satisfies the equations for optimality. At the end we will also consider a simple example of application for the sake of a quantitative estimation, showing that the optimization can lead to sensible improvements.

In the 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 .

For finite dimension all bounded operators are Hilbert-Schmidt, whence \mathcal{S} is a Hilbert space, and indeed $\mathcal{S} \subseteq \mathcal{H}^{\otimes 2}$ and linear operators can be associated to bipartite vectors as follows [13]:

$$A = \sum_{m,n=1}^d A_{mn} |m\rangle\langle n| \leftrightarrow |A\rangle\rangle = \sum_{m,n=1}^d A_{mn} |m\rangle|n\rangle, \quad (4)$$

with the Hilbert-Schmidt scalar product $\langle\langle A|B\rangle\rangle \equiv \text{Tr}[A^\dagger B]$. In the following, we will retain the double-ket notation as a reminder of the correspondence (4). Completeness of the set of vectors $\{|P_i\rangle\rangle\}_{1 \leq i \leq N}$ with $\mathcal{S} := \text{Span}\{|P_i\rangle\rangle\}_{1 \leq i \leq N}$ can be written as follows:

$$a \|X\|_2^2 \leq \sum_{i=1}^N |\langle\langle P_i|X\rangle\rangle|^2 \leq b \|X\|_2^2, \quad X \in \mathcal{S}, \quad (5)$$

with $0 < a \leq b < \infty$, and the norm $\|Z\|_2$ is the Hilbert-Schmidt norm induced by the scalar product $\|Z\|_2 = \sqrt{\langle\langle Z|Z\rangle\rangle} = \sqrt{\text{Tr}[Z^\dagger Z]}$. In the literature Eq. (5) with $|P_i\rangle\rangle$ regarded as abstract vectors in the linear space \mathcal{S} [14] define a so-called frame of vectors. The main theorem of frame theory states that a set of vectors in \mathcal{S} is a frame iff the operator

$$F = \sum_i |P_i\rangle\rangle\langle\langle P_i|, \quad (6)$$

called frame operator is invertible [11] (here the fact that the set $\{|P_i\rangle\rangle\}_{1 \leq i \leq N}$ is a frame for \mathcal{S} trivially follows from the definition of \mathcal{S}). Since F is invertible, one can obtain suitable coefficients $f_i[X]$ for the expansion of a vector $|X\rangle\rangle$ by the formula

$$f_i[X] = \langle\langle \Delta_i|X\rangle\rangle, \quad (7)$$

where $\{\Delta_i\}$ is the canonical dual [11], which is defined through the identity

$$|\Delta_i\rangle\rangle = F^{-1}|P_i\rangle\rangle. \quad (8)$$

However, if the vectors $\{|P_i\rangle\rangle\}_{1 \leq i \leq N}$ are linearly dependent, the processing rule (7) is not unique, and all different choices of coefficients are provided by $f_i[X] = \langle\langle D_i|X\rangle\rangle$, where $\{D_i\}$ are alternate duals. All alternate duals can be classified as follows [15]:

$$|D_i\rangle\rangle = |\Delta_i\rangle\rangle + |Y_i\rangle\rangle - \sum_j |Y_j\rangle\rangle\langle\langle P_j|\Delta_i\rangle\rangle, \quad (9)$$

where the operators $\{Y_i\}$ are arbitrary elements of \mathcal{S} . Now, one can define a linear map Λ from an abstract N -dimensional space \mathcal{K} of coefficient vectors $|c\rangle$ to \mathcal{S} as follows:

$$\Lambda|c\rangle = \sum_{i=1}^N c_i |P_i\rangle\rangle, \quad (10)$$

and Λ has matrix elements $\Lambda_{mn,i} = (P_i)_{mn}$. By definition any alternate dual must satisfy

$$\sum_{j=1}^N \sum_{m,n=1}^d (P_j)_{pq} (D_j^*)_{mn} (P_i)_{mn} c_i = \sum_{i=1}^N (P_i)_{pq} c_i, \quad (11)$$

for all $|c\rangle \in \mathcal{K}$. Defining the matrix Γ with elements $(\Gamma)_{i,mn} = (D_i^*)_{mn}$ one has

$$\Lambda \Gamma \Lambda = \Lambda, \quad (12)$$

which is the definition of generalized inverse (or pseudoinverse) of Λ . Alternate duals are then in one-to-one correspondence with generalized inverses of Λ . This fact was already noticed in Ref. [16], and will be very useful in the following.

We want now to minimize the statistical error in the determination of the ensemble average. This is provided by the variance

$$\delta_D(X) := \sum_{j=1}^N p(j|\rho_\mathcal{E}) |f_j[X]|^2 - |\overline{\langle X \rangle}|^2_\mathcal{E}, \quad (13)$$

where $\rho_\mathcal{E} = \sum_i p_i \rho_i$, and $|\overline{\langle X \rangle}|^2_\mathcal{E} = \sum_i p_i |\text{Tr}[\rho_i X]|^2$ is the squared modulus of the expectation of X averaged over the states in the ensemble. One has

$$\delta_D(X) = \sum_{i=1}^N |\langle\langle D_i|X\rangle\rangle|^2 \text{Tr}[\rho_\mathcal{E} P_i] - |\overline{\langle X \rangle}|^2_\mathcal{E}. \quad (14)$$

Notice that the term $|\overline{\langle X \rangle}|^2_\mathcal{E}$ depends only on the ensemble, and is independent of the POVM, whence we will focus attention only on the contribution

$$\Sigma_D(X) = \sum_{i=1}^N |\langle\langle D_i|X\rangle\rangle|^2 \text{Tr}[\rho_\mathcal{E} P_i]. \quad (15)$$

A relevant case is that of the uniform ensemble, with all pure states equally distributed, corresponding to $\rho_\mathcal{E} = \frac{1}{d}$ and $|\overline{\langle X \rangle}|^2_\mathcal{E} = \frac{1}{d(d+1)} (\text{Tr}[X^\dagger X] + |\text{Tr}[X]|^2)$ [16].

Equation (15) defines a norm $\|f_i[X]\|_\pi^2$ of the vector of coefficients corresponding to the metric matrix $\pi_{ij} = \text{Tr}[\rho_\mathcal{E} P_i] \delta_{ij}$. Then, minimizing $\Sigma_D(X)$ corresponds to determining the minimum norm generalized inverse Γ of Λ with respect to the norm $\|c\|_\pi = \sum_{i=1}^N |c_i|^2 \pi_{ii}$. The minimum norm condition for $\pi = I$ corresponds to the Moore-Penrose generalized inverse Γ [17], satisfying the three conditions: $\Gamma \Lambda \Gamma = \Gamma$, $\Gamma \Lambda = \Lambda^\dagger \Gamma^\dagger$, and $\Lambda \Gamma = \Gamma^\dagger \Lambda^\dagger$.

The Moore-Penrose generalized inverse of a matrix Z (also denoted as Z^\dagger) turns out to be simply the inverse of Z on its support $\text{supp}(Z)$ [the support $\text{supp}(Z)$ of Z is the orthogonal complement of the kernel $\ker(Z)$ of Z], and acts as the null matrix on $\ker(Z)$.

Following the same lines of derivation for the Moore-Penrose generalized inverse one can show that the minimum norm generalized inverse for a generic π is independent of X , and is defined by the condition [16]

$$\pi\Gamma\Lambda = \Lambda^\dagger\Gamma^\dagger\pi. \quad (16)$$

The matrix $\Gamma\Lambda$ has matrix elements $(\Gamma\Lambda)_{ij} = \langle\langle D_i | P_j \rangle\rangle$. Equation (16) rewritten in terms of the optimal dual $\{\hat{D}_i\}$ becomes

$$\langle\langle \rho_\varepsilon | P_i \rangle\rangle \langle\langle \hat{D}_i | P_j \rangle\rangle = \langle\langle P_i | \hat{D}_j \rangle\rangle \langle\langle P_j | \rho_\varepsilon \rangle\rangle. \quad (17)$$

Upon summing over the index i , and remembering that for any dual $\{D_i\}$ one has $\sum_i |P_i\rangle\langle\langle D_i| = \Pi_{\mathcal{S}}$ where $\Pi_{\mathcal{S}}$ is the projection on \mathcal{S} , one has $\langle\langle \rho_\varepsilon | P_j \rangle\rangle = \text{Tr}[\hat{D}_j \langle\langle P_j | \rho_\varepsilon \rangle\rangle]$, consequently $\text{Tr}[\hat{D}_j] = 1$. This implies that the optimal processing function for the identity operator is $f_i[I] = 1$, whence $\delta_{\hat{D}}(I) = 0$, whereas, remarkably, $f_i[I]$ is generally nonconstant for the canonical dual.

We will now prove that the solution of Eq. (16) is unique. For not invertible π we can restrict Eq. (16) to $\text{supp}(\pi)$, and from now on we will denote the corresponding blocks of all matrices with the same symbols. Suppose now that there exist two generalized inverses Γ and Γ' satisfying Eq. (16). Upon defining $\Theta = \Gamma - \Gamma'$, we have that

$$\Lambda\Theta\Lambda = 0, \quad \pi\Theta\Lambda = \Lambda^\dagger\Theta^\dagger\pi, \quad (18)$$

and multiplying on the left by $\Lambda\pi^{-1}$ both members of the second equation, and substituting the first equation we obtain $\Lambda\Theta\Lambda = \Lambda\pi^{-1}\Lambda^\dagger\Theta^\dagger\pi = 0$, or equivalently, by invertibility of π , $\Lambda\pi^{-1}\Lambda^\dagger\Theta^\dagger = 0$. The matrix $\Lambda\pi^{-1}\Lambda^\dagger$ can be rewritten as

$$\Lambda\pi^{-1}\Lambda^\dagger = \sum_{i=1}^N (\pi_{ii})^{-1} |P_i\rangle\langle\langle P_i|. \quad (19)$$

Since $\Lambda\pi^{-1}\Lambda^\dagger \geq 0$, a sufficient condition for a vector $X \in \mathcal{S}$ to be in $\ker(\Lambda\pi^{-1}\Lambda^\dagger)$ is that $\langle\langle X | \Lambda\pi^{-1}\Lambda^\dagger | X \rangle\rangle = 0$, namely

$$\sum_{i=1}^N (\pi_{ii})^{-1} |\langle\langle X | P_i \rangle\rangle|^2 = 0, \quad \Leftrightarrow X \in \mathcal{S}, \quad (20)$$

which is possible iff $\langle\langle X | P_i \rangle\rangle = 0$ for all i . By completeness of P_i , this is equivalent to saying that the only vector of \mathcal{S} in $\ker(\Lambda\pi^{-1}\Lambda)$ is $X = 0$. Then $\Lambda\pi^{-1}\Lambda$ is full rank, whence $\Theta = 0$, or equivalently $\Gamma = \Gamma'$.

We will now provide the solution to Eq. (16) in terms of the optimal dual, which is expressed as

$$\hat{D}_i = \Delta_i - \sum_{j=1}^N [(I - M)\pi(I - M)]^\dagger \pi M_{ij} \Delta_j, \quad (21)$$

where Δ_i is the canonical dual, $M_{ij} = \langle\langle \Delta_i | P_j \rangle\rangle = \langle\langle P_i | F^{-1} | P_j \rangle\rangle = \langle\langle P_i | \Delta_j \rangle\rangle = M_{ji}^*$. Since $\Delta_i^\dagger = \Delta_i$, $M_{ij}^* = M_{ij}$ [18] and the optimal dual frame $\{\hat{D}_i\}$ in Eq. (21) is self-adjoint because the matrix $[(I - M)\pi(I - M)]^\dagger \pi M$ has real elements. Notice that $M^2 = M$ and $M^\dagger = M$, namely M is an orthogonal projector, as can be easily verified. Also $(I - M)$ is an orthogonal projector, and $[(I - M)\pi(I - M)]^\dagger (I - M) = [(I - M)\pi(I - M)]^\dagger$. The matrix $\Gamma\Lambda$ for the optimal dual frame can be easily calculated, and is equal to

$$\Gamma\Lambda = M - [(I - M)\pi(I - M)]^\dagger \pi M. \quad (22)$$

We can substitute this expression in Eq. (16) to verify its validity. We have indeed

$$\begin{aligned} \pi\Gamma\Lambda &= \pi M - \pi[(I - M)\pi(I - M)]^\dagger \pi M \\ &= \pi M + \pi[(I - M)\pi(I - M)]^\dagger (I - M)\pi(I - M) \\ &\quad - \pi[(I - M)\pi(I - M)]^\dagger \pi \\ &= \pi - \pi[(I - M)\pi(I - M)]^\dagger \pi, \end{aligned} \quad (23)$$

and analogously

$$\Lambda^\dagger\Gamma^\dagger\pi = \pi - \pi[(I - M)\pi(I - M)]^\dagger \pi. \quad (24)$$

When $\pi \propto I$ the canonical dual is optimal, since for the canonical dual one has $\Gamma\Lambda = \Lambda^\dagger\Gamma^\dagger$. This is the case, e.g., of the uniform ensemble of pure states with POVM elements with constant trace, which includes all covariant POVMs studied in Ref. [16]. In the general case, one can write the expression of Eq. (15) as follows:

$$\Sigma_{\hat{D}}(X) = \Sigma_{\Delta}(X) - \Psi(X), \quad (25)$$

where Σ_{Δ} is the contribution of the canonical dual

$$\Sigma_{\Delta}(X) = \sum_{i=1}^N |\langle\langle \Delta_i | X \rangle\rangle|^2 \text{Tr}[\rho_\varepsilon P_i], \quad (26)$$

and Ψ is the correction due to the optimization which is given by

$$\Psi(X) = \sum_{i,j=1}^N \langle\langle X | \Delta_i \rangle\rangle (\pi[(I - M)\pi(I - M)]^\dagger \pi)_{ij} \langle\langle \Delta_j | X \rangle\rangle. \quad (27)$$

The relative added noise of the canonical dual compared to the optimal one is given by

$$\epsilon(X) \doteq \frac{\delta_{\Delta}(X) - \delta_{\hat{D}}(X)}{\delta_{\hat{D}}(X)} = \frac{\Psi(X)}{\Sigma_{\Delta}(X) - \Psi(X) - \overline{|\langle\langle X \rangle\rangle|^2}_\varepsilon}. \quad (28)$$

A quantitative estimate of $\epsilon(X)$ can be obtained from the following example in dimension two (see Fig. 1). Consider the following informationally complete POVM

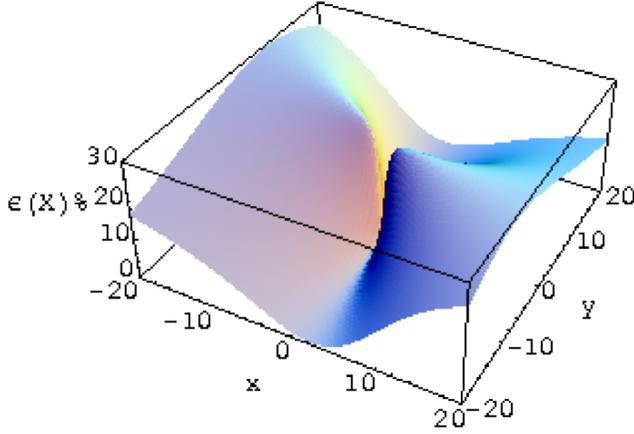


FIG. 1 (color online). Example of optimized data-processing rule for the informationally complete POVM in Eq. (29). The plot shows the relative added noise in Eq. (27) for $X = \sigma_z + x\sigma_x + y\sigma_y$, versus x and y .

$$\begin{aligned}
 P_1 &= \begin{pmatrix} \frac{64}{1197} & -\frac{16}{1197} \\ -\frac{16}{1197} & \frac{40}{1197} \end{pmatrix}, & P_2 &= \begin{pmatrix} \frac{34}{1197} & \frac{2(1-16i)}{1197} \\ \frac{2(1+16i)}{1197} & \frac{34}{1197} \end{pmatrix}, \\
 P_3 &= \begin{pmatrix} \frac{281}{399} & -\frac{18-32i}{1197} \\ -\frac{18+32i}{1197} & \frac{289}{399} \end{pmatrix}, & P_4 &= \begin{pmatrix} \frac{64}{399} & \frac{64(1-i)}{1197} \\ \frac{64(1+i)}{1197} & \frac{32}{399} \end{pmatrix}, & (29) \\
 P_5 &= \begin{pmatrix} \frac{64}{1197} & \frac{-32(1+2i)}{1197} \\ -\frac{32(1-2i)}{1197} & \frac{160}{1197} \end{pmatrix}.
 \end{aligned}$$

The operator X is the following self-adjoint operator

$$X = \begin{pmatrix} 1 & -2.24 + i \\ -2.24 - i & -1 \end{pmatrix}, \quad (30)$$

and for an ensemble of uniformly distributed pure states $\frac{1}{6}(\text{Tr}[X^2] + \text{Tr}[X]^2) = \frac{1}{6} \text{Tr}[X^2] = 2.34$. By direct calculation one obtains $\Sigma_\Delta = 799.66$ and $\Psi = 133.05$, and finally

$$\epsilon(X) \simeq 0.2, \quad (31)$$

which means a relative added noise of about 20%. This example shows that a correct processing can highly improve the statistics of expectation values, and eventually the convergence rate of tomographic state reconstruction. The additional error due to the use of the canonical dual instead of the optimal one is equivalent to a depolarizing channel with depolarization probability 0.09.

In conclusion, we considered the general measurement scenario in which the ensemble average of an operator is determined via suitable data processing of the outcomes of a quantum measurement described by a POVM. We have determined the optimal processing that minimizes the statistical error of the estimation. Contrarily to the widespread conviction, the optimal data processing is generally not obtained via the canonical dual of the POVM, and the improvement due to optimization can be substantial. The present analysis has been carried out for finite spectrum

and finite dimensions; however, it can be easily generalized to discrete spectrum in infinite dimensions for bounded operators and bounded duals, and, with more technicalities, even to continuous spectrum (the case of quantum homodyne tomography [7]). We believe that the present result will allow one to improve greatly many relevant experimental analyses of quantum measurements.

P. P. thanks L. Maccone and M. F. Sacchi for interesting discussions and suggestions. This Letter has been supported by Ministero Italiano dell'Università e della Ricerca (MIUR) through PRIN 2005. P. P. acknowledges financial support by EC under Project SECOQC (Contract No. IST-2003-506813).

- [1] J. Von Neumann, *Mathematical Principles of Quantum Mechanics* (Princeton University Press, Princeton, 1955).
- [2] E. Arthurs and J.L. Kelly, *Bell Syst. Tech. J.* **44**, 725 (1965).
- [3] J.P. Gordon and W.H. Louisell, *Physics of Quantum Electronics* (McGraw-Hill, New York, 1966), pp. 833–840.
- [4] C.W. Helstrom *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
- [5] A. Chefles, *Phys. Rev. A* **64**, 062305 (2001).
- [6] G.M. D'Ariano, P. Perinotti, and M.F. Sacchi, *Europhys. Lett.* **65**, 165 (2004); *J. Opt. B* **6**, S487 (2004);
- [7] G.M. D'Ariano, M.G.A. Paris, and M.F. Sacchi, *Adv. Imaging Electron Phys.* **128**, 205 (2003).
- [8] F. Buscemi, G.M. D'Ariano, M. Keyl, P. Perinotti, and R.F. Werner, *J. Math. Phys. (N.Y.)* **46**, 082109 (2005).
- [9] G.M. D'Ariano, P. Lopresti, and L. Maccone, *Phys. Rev. Lett.* **93**, 250407 (2004).
- [10] G.M. D'Ariano, J. Kahn, and M.F. Sacchi, *Phys. Rev. A* **72**, 032310 (2005).
- [11] R.J. Duffin and A.C. Schaeffer, *Trans. Am. Math. Soc.* **72**, 341 (1952); P.G. Casazza, *Taiw. J. Math.* **4**, 129 (2000).
- [12] P.G. Casazza, D. Han, and D.R. Larson, *Contemp Math.* **247**, 149 (1999).
- [13] G.M. D'Ariano, P. Lopresti, and M.F. Sacchi, *Phys. Lett. A* **272**, 32 (2000).
- [14] In infinite dimension generally one considers \mathcal{S} as a Banach space (see Ref. [12]) and the condition $b < \infty$ is generally nontrivial as in the finite dimensional case.
- [15] S. Li, *Numer. Funct. Anal. Optim.* **16**, 1181 (1995).
- [16] G.M. D'Ariano, P. Perinotti, and M.F. Sacchi, *Phys. Rev. A* **72**, 042108 (2005).
- [17] R.B. Bapat, *Linear Algebra and Linear Models* (Springer-Verlag, New York, 2000).
- [18] Notice that since the swap operator E acts on a vector $|A\rangle \in \mathcal{L}(\mathcal{H})$ as $E|A\rangle = |A^T\rangle$, where A^T is the transpose of A on the basis $|n\rangle$ of Eq. (4), by self-adjointness of P_i one has $E|P_i^*\rangle = |P_i\rangle$, and $EF^*E = F$. Similarly $EF^{-1}E = F^{-1}$, and then $E|\Delta_i^*\rangle = EF^{-1}|P_i^*\rangle = F^{-1}E|P_i^*\rangle = |\Delta_i\rangle$, namely $\Delta_i^\dagger = \Delta_i$. As a consequence $\langle\langle \Delta_i | P_j \rangle\rangle = \text{Tr}[\Delta_i P_j] \in \mathbb{R}$.