Maximum-likelihood estimation of quantum measurement

Jaromír Fiurášek
Department of Optics, Palacký University, 17 Listopadu 50, 772 07 Olomouc, Czech Republic
(Received 15 January 2001; published 17 July 2001)

Maximum-likelihood estimation is applied to the determination of an unknown quantum measurement. The calibrated measuring apparatus carries out measurements on many different quantum states and the positive operator-valued measure governing the measurement statistics is then inferred from the collected data via the maximum-likelihood principle. In contrast to the procedures based on linear inversion, our approach always provides a physically sensible result. We illustrate the method on the case of the Stern-Gerlach apparatus.

DOI: 10.1103/PhysRevA.64.024102 PACS number(s): 03.65.Ta, 03.65.Wj

I. INTRODUCTION

Let us imagine that we possess an apparatus that performs some measurement on a certain quantum mechanical system such as the spin of an electron. We are not sure which measurement is associated with the measuring device and we would like to calibrate it.

Suppose that the apparatus can respond with k different measurement outcomes. As is well known from the theory of quantum measurement [1], such a device is completely characterized by the positive operator-valued measure (POVM) whose k elements $\hat{\Pi}_l$ govern the measurement statistics. The probability p_{lm} that the apparatus will respond with outcome $\hat{\Pi}_l$ when measuring the quantum state with density matrix $\hat{\varrho}_m$ can be expressed as

$$p_{lm}=\mathrm{Tr}[\hat{\Pi}_l\hat{\varrho}_m], \tag{1}$$

where Tr stands for the trace. The POVM elements are positive semidefinite Hermitian operators,

$$\hat{\Pi}_l \geqslant 0, \tag{2}$$

which decompose the unit operator,

$$\sum_{l=1}^{k}\hat{\Pi}_l=\hat{I}. \tag{3}$$

The condition (2) ensures that $p_{lm}\geqslant 0$ and Eq. (3) follows from the requirement that the total probability is normalized to unity, $\sum_{l=1}^{k}p_{lm}=1$.

In order to determine the POVM we have to perform a set of measurements on various known quantum states and then estimate the $\hat{\Pi}_l$ from the collected experimental data. This strategy belongs to the broad class of quantum reconstruction procedures that have attracted considerable attention recently. Quantum state reconstruction has been widely studied and now represents a well established technique in many branches of quantum physics (for a review, see, e.g., [2,3]). The estimation of quantum mechanical processes, i.e., input-output transformations of quantum devices, was discussed in [4] and the problem of complete characterization of an arbitrary measurement process was recently addressed in [5].

The POVM can be most easily estimated by direct linear inversion of Eq. (1). Let f_{lm} denote the total number of detections of $\hat{\Pi}_l$ for the measurements performed on the quantum state $\hat{\varrho}_m$. Assuming that the theoretical detection probability p_{lm} given by Eq. (1) can be replaced with relative frequency, we may write

$$\mathrm{Tr}[\hat{\Pi}_l\hat{\varrho}_m]\equiv\sum_{i,j=1}^{N}\Pi_{l,ij}\varrho_{m,ji}=\frac{f_{lm}}{\sum_{l'=1}^{k}f_{l'm}}, \tag{4}$$

where N is the dimension of the Hilbert space on which the operators $\hat{\Pi}_l$ act. Formula (4) establishes a system of linear equations for matrix elements of the unknown operators $\hat{\Pi}_l$. If a sufficient amount of data is available then Eq. (4) can be inverted (e.g., by the least squares method) and we can determine $\hat{\Pi}_l$. This approach is a direct analog of linear reconstruction algorithms devised for quantum state and quantum process reconstructions. The linear inversion is simple and straightforward, but it also has one significant disadvantage. The linear procedure cannot guarantee the required properties of $\hat{\Pi}_l$, namely, the conditions (2). Consequently, the linear estimation may lead to an unphysical POVM, predicting negative probabilities p_{lm} for certain input quantum states. To avoid such problems, one should resort to more sophisticated nonlinear reconstruction strategy.

In this paper we show that the maximum-likelihood (ML) estimation is suitable and can be successfully used for the calibration of the measuring apparatus. ML estimation has been recently applied to reconstruction of quantum states [6,7] and quantum processes (completely positive maps between density matrices) [8]. Here we employ it to reconstruct an unknown *quantum measurement*, thereby demonstrating again the remarkable versatility and usefulness of ML estimation. Apart from providing physically sensible results the ML estimation also achieves higher accuracy than linear methods, which is an important practical advantage.

Our method is completely generic and does not rely on any *a priori* assumptions about the measuring apparatus subject to calibration. Of course, in most cases we approximately know what kind of measurement is performed by the apparatus. In some cases we could use this *a priori* information and characterize the device by several parameters which are to be estimated. As a simple example we can mention the photodetector, which can be fully characterized by its quantum efficiency η [9]. However, a calibration restricted by

1050-2947/2001/64(2)/024102(4)/$20.00

some *a priori* assumptions need not provide an exact description of the calibrated apparatus because the device may behave in a slightly different way than assumed. On the other hand, the accuracy of the calibration technique presented in this paper is limited only by the number of performed measurements and can be arbitrarily high. Our calibration method may thus find practical applications in cases when one requires a very precise knowledge of the measurement carried out by the measuring device.

II. MAXIMUM-LIKELIHOOD ESTIMATION

The estimated operators $\hat{\Pi}_l$ should maximize the likelihood functional

$$\mathcal{L}[\{\hat{\Pi}_l\}]=\prod_{l=1}^{k}\prod_{m=1}^{M}(\mathrm{Tr}[\hat{\Pi}_l\hat{\varrho}_m])^{f_{lm}}, \tag{5}$$

where M is the number of different quantum states $\hat{\varrho}_m$ used for the reconstruction. The maximum of the likelihood functional (5) has to be found in the subspace of physically allowed operators $\hat{\Pi}_l$. We can decompose each operator $\hat{\Pi}_l$ as

$$\hat{\Pi}_l=\sum_{q=1}^{N} r_{lq}|\phi_{lq}\rangle\langle\phi_{lq}|, \tag{6}$$

where $r_{lq}\geqslant 0$ are the eigenvalues of $\hat{\Pi}_l$ and $|\phi_{lq}\rangle$ are the corresponding orthonormal eigenstates. The maximum of $\mathcal{L}[\{\hat{\Pi}_l\}]$ can be found from the extremum conditions. It is convenient to work with the logarithm of the original likelihood functional and the constraint (3) has to be incorporated by introducing a Hermitian operator $\hat{\lambda}$ whose matrix elements $\lambda_{ij}=\lambda_{ji}^{*}$ are Lagrange multipliers. The extremum conditions then read

$$\frac{\partial}{\partial\langle\phi_{lq}|}\left[\sum_{l'=1}^{k}\sum_{m=1}^{M} f_{l'm}\ln\left(\sum_{q'=1}^{N} r_{l'q'}\langle\phi_{l'q'}|\hat{\varrho}_m|\phi_{l'q'}\rangle\right) - \sum_{l'=1}^{k}\sum_{q'=1}^{N} r_{l'q'}\langle\phi_{l'q'}|\hat{\lambda}|\phi_{l'q'}\rangle\right]=0. \tag{7}$$

Thus we immediately find

$$r_{lq}|\phi_{lq}\rangle=\hat{R}_l r_{lq}|\phi_{lq}\rangle, \tag{8}$$

where

$$\hat{R}_l=\hat{\lambda}^{-1}\sum_{m=1}^{M}\frac{f_{lm}}{p_{lm}}\hat{\varrho}_m. \tag{9}$$

Let us now multiply Eq. (8) by $\langle\phi_{lq}|$ from the right and sum over q. Thus we obtain

$$\hat{\Pi}_l=\hat{R}_l\hat{\Pi}_l. \tag{10}$$

It follows from this formula that $\hat{\Pi}_l=\hat{\Pi}_l\hat{R}_l^{\dagger}$. On inserting this expression into the right-hand side of Eq. (10), we obtain

$$\hat{\Pi}_l=\hat{R}_l\hat{\Pi}_l\hat{R}_l^{\dagger}. \tag{11}$$

The constraint (3) provides a formula for the operator of Lagrange multipliers,

$$\hat{\lambda}^{-1}\hat{G}\hat{\lambda}^{-1}=\hat{I}, \tag{12}$$

where $\hat{G}$ is the positive operator

$$\hat{G}=\sum_{l=1}^{k}\sum_{m,m'=1}^{M}\frac{f_{lm}}{p_{lm}}\frac{f_{lm'}}{p_{lm'}}\hat{\varrho}_{m'}\hat{\Pi}_l\hat{\varrho}_m. \tag{13}$$

Upon solving Eq. (12) we get $\hat{\lambda}=\hat{G}^{1/2}$. We fix the branch of the square root of $\hat{G}$ by requiring that $\hat{\lambda}$ should be a positive definite operator. We can factorize the matrix $\hat{G}$ as $\hat{G}=\hat{U}^{\dagger}\hat{\Lambda}\hat{U}$ where $\hat{U}$ is a unitary matrix and $\hat{\Lambda}$ is a diagonal matrix containing eigenvalues of $\hat{G}$. We define $\hat{\Lambda}^{1/2}=\mathrm{diag}(\Lambda_{11}^{1/2},\ldots,\Lambda_{NN}^{1/2})$ and we can write

$$\hat{\lambda}=\hat{U}^{\dagger}\hat{\Lambda}^{1/2}\hat{U}. \tag{14}$$

The extremum Eqs. (11) and (14) can be conveniently solved by means of repeated iterations. Notice that both conditions (2) and (3) are exactly fulfilled at each iteration step because the transformation $\hat{\Pi}_l\rightarrow\hat{R}_l\hat{\Pi}_l\hat{R}_l^{\dagger}$ preserves the positivity of $\hat{\Pi}_l$.

If there exists a POVM whose elements $\hat{\Pi}_l$ meet constraints (2) and (3) and exactly solve the linear Eqs. (4) then the ML estimate agrees with the linear inversion. In this case it holds exactly for all l,m that

$$p_{lm}=\frac{f_{lm}}{\sum_{l'=1}^{k} f_{l'm}}. \tag{15}$$

On inserting this expression into Eq. (10), we find after some algebra that the set of k Eqs. (10) reduces to the formula for the operator of Lagrange multipliers

$$\hat{\lambda}=\sum_{m=1}^{M}\sum_{l=1}^{k} f_{lm}\hat{\varrho}_m. \tag{16}$$

Notice that $\hat{\lambda}$ is positive definite. We emphasize that Eq. (16) holds only in the special case when the ML and linear estimates coincide. Notice that the operators $\hat{R}_l$ given by Eq. (9) contain the inversion of $\hat{\lambda}$. The reconstruction is possible only on such a subspace of the total Hilbert space where the inversion $\hat{\lambda}^{-1}$ exists. This restriction can easily be understood if we make use of Eq. (16). The experimental data contain only information on the Hilbert subspace probed by the density matrices $\hat{\varrho}_m$ and the reconstruction of the POVM must be restricted to this subspace.

The principal advantage of ML estimation lies in its ability to correctly handle any experimental data and provide reliable estimates in cases when linear algorithms fail. As mentioned in the Introduction, the linear inversions may provide unphysical estimates, namely, operators $\hat{\Pi}_l$ that are not positive definite. It should be noted that such a failure of linear inversion is rather typical and can occur with high probability. This is most apparent if the operators $\hat{\Pi}_l$ are projectors; hence $N-1$ eigenvalues of each $\hat{\Pi}_l$ are equal to zero. For a sufficiently large number of measured data, the linear estimate of a matrix element of $\hat{\Pi}_l$ is a random variable with Gaussian distribution centered at the true value. In the basis where the projector $\hat{\Pi}_l$ is diagonal, its $N-1$ diagonal elements fluctuate around zero. It follows that in most cases at least one diagonal element is negative and the linear inversion yields a nonpositive POVM which cannot describe any measuring device.

These problems of linear algorithms stem from the difference between recorded relative frequencies and theoretical probabilities, which are assumed to be equal in Eq. (4). The frequencies f_{lm} are fluctuating quantities with a multinomial distribution characterized by probabilities p_{lm}. In an experiment we can, in principle, detect any f_{lm}. However, some sets of relative frequencies do not coincide with any theoretical probabilities (1) calculated for given quantum states $\hat{\varrho}_m$ used for the calibration (i.e., in some cases there does not exist a POVM that would yield probabilities p_{lm} equal to the detected relative frequencies) and direct linear inversion of Eq. (4) may then provide an unphysical result. The observation of several different quantum states by a single measuring apparatus is equivalent to measurement of several noncommuting observables on many copies of a given quantum state. In our scheme, however, the role of the quantum state and the measurement are interchanged, because we employ a known $\hat{\varrho}_m$ to probe $\hat{\Pi}_l$. Thus the ML estimation of the quantum measurement can be interpreted as a synthesis of information from mutually incompatible observations [6].

The determination of the quantum measurement can be simplified considerably if we have some reliable *a priori* information about the apparatus. As an example let us briefly consider a class of optical detectors that are sensitive only to the number of photons in a single mode of the electromagnetic field. The elements $\hat{\Pi}_l$ of the POVM describing a phase-insensitive detector are all diagonal in the Fock basis,

$$\hat{\Pi}_l=\sum_n r_{ln}|n\rangle\langle n|, \tag{17}$$

and the ML estimation reduces to the determination of the eigenvalues $r_{ln}\geqslant 0$. The extremum Eqs. (11) and (14) simplify to

$$r_{ln}=\frac{r_{ln}}{\lambda_n}\sum_{m=1}^{M}\frac{f_{lm}}{p_{lm}}\varrho_{m,nn},$$

$$\lambda_n=\sum_{m=1}^{M}\sum_{l=1}^{k}\frac{f_{lm}}{p_{lm}}\varrho_{m,nn}r_{ln},$$

$$p_{lm}=\sum_n \varrho_{m,nn}r_{ln}. \tag{18}$$

Instead of solving the extremum equations, one may directly search for the maximum of $\mathcal{L}[\{\hat{\Pi}_l\}]$ with the help of a downhill-simplex algorithm [7]. To implement this algorithm successfully, it is necessary to use a minimal parametrization. If we deal with an N-level system, then each $\hat{\Pi}_l$ is parametrized by N^2 real numbers. Since the constraint (3) allows us to determine the operator $\hat{\Pi}_k$ in terms of the remaining $k-1$ operators, the number of independent real parameters reads $N^2(k-1)$. Furthermore, we may take advantage of the Cholesky decomposition,

$$\hat{\Pi}_l=\hat{C}_l^\dagger\hat{C}_l, \tag{19}$$

where $\hat{C}_l$ is a lower triangular matrix with real elements on its main diagonal. The parametrization (19) is used for the first $k-1$ operators, and the last one is calculated from Eq. (3),

$$\hat{\Pi}_k=\hat{I}-\sum_{l=1}^{k-1}\hat{C}_l^\dagger\hat{C}_l, \tag{20}$$

thus achieving minimal parametrization. For each parameter set where $\mathcal{L}[\{\hat{\Pi}_l\}]$ is evaluated, one has to check whether the operator (20) is positive semidefinite. If this does not hold then one sets $\mathcal{L}[\{\hat{\Pi}_l\}]=0$, thereby restricting the numerical search for the maximum to the domain of physically allowed operators. This domain is a finite volume subspace of an $N^2(k-1)$-dimensional space.

III. STERN-GERLACH APPARATUS

In this section we illustrate the developed formalism by means of numerical simulations for a Stern-Gerlach apparatus measuring a spin-1 particle. We compare the linear inversion and ML estimation and demonstrate that the ML algorithm outperforms the linear estimation.

Let $\hat{s}_x$, $\hat{s}_y$, and $\hat{s}_z$ denote the operators of spin projections to axes x, y, and z, respectively. We choose the three eigenstates of $\hat{s}_z$ as the basis states, $\hat{s}_z|s_z\rangle=s_z|s_z\rangle$, $s_z=-1,0,1$. In our numerical simulations, nine different pure quantum states are used for the calibration: three eigenstates of $\hat{s}_z$ and six superposition states

$$\frac{1}{\sqrt{2}}(|j_z\rangle+|k_z\rangle),\quad \frac{1}{\sqrt{2}}(|j_z\rangle+i|k_z\rangle),$$

where $j_z,k_z=-1,0,1$ and $j_z<k_z$. The measurement on each state is performed $\mathcal{N}$ times. In the simulations, we assumed two slightly different detectors. The first device is an ideal Stern-Gerlach apparatus which measures the projection of the spin component along direction $\vec{n}=(1,1,1)/\sqrt{3}$. The operators $\hat{\Pi}_l$ are projectors

$$\hat{\Pi}_j=|j_{\vec{n}}\rangle\langle j_{\vec{n}}|,\quad j_{\vec{n}}=-1,0,1, \tag{21}$$

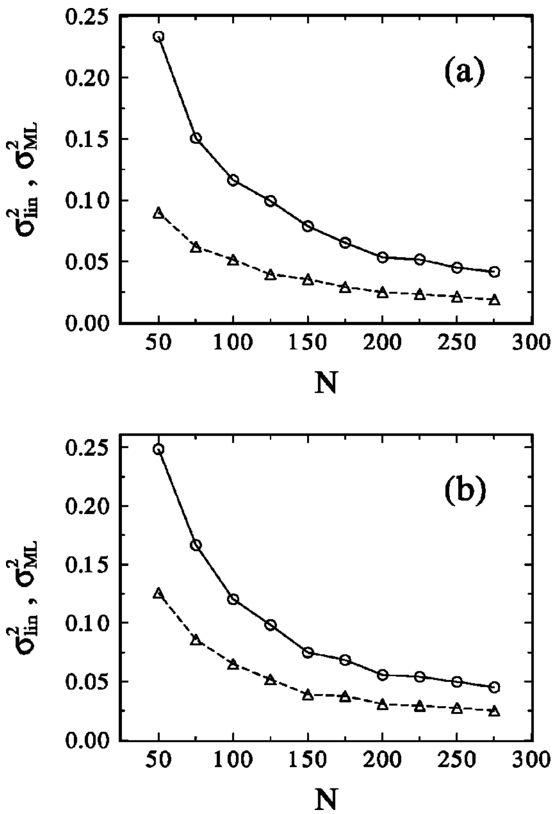

FIG. 1. Variances of linear (○) and ML (△) estimates versus the number of measurements $\mathcal{N}$. The figure shows results for both an ideal (a) and a nonideal (b) Stern-Gerlach apparatus.

where $\hat{s}_{\vec{n}}|j_{\vec{n}}\rangle = j_{\vec{n}}|j_{\vec{n}}\rangle$ and $\hat{s}_{\vec{n}} = (\hat{s}_x + \hat{s}_y + \hat{s}_z)/\sqrt{3}$. As a second example we consider a nonideal Stern-Gerlach apparatus characterized by a POVM whose elements are incoherent mixtures of the projectors (21),

$$\hat{\Pi}_1' = 0.95\hat{\Pi}_1 + 0.025(\hat{\Pi}_{-1} + \hat{\Pi}_0), \tag{22}$$

and the expressions for $\hat{\Pi}_{-1}'$ and $\hat{\Pi}_0'$ can be obtained by cyclic permutations of the subscripts $-1,0,1$.

We have performed Monte Carlo simulations of the measurements and have subsequently reconstructed the POVM from the simulated experimental data. The ML estimates $\hat{\Pi}_{j,\rm ML}$ were obtained by iterative solution of the extremum equations (11) and (14). The linear estimates $\hat{\Pi}_{j,\rm lin}$ were found by solving the system of Eqs. (4). In order to compare these two procedures, we define the variances of the estimates as

$$\sigma^2_{\rm ML} = \left\langle \sum_j \mathrm{Tr}[\Delta\hat{\Pi}^2_{j,\rm ML}] \right\rangle_{\rm ens},$$
$$\sigma^2_{\rm lin} = \left\langle \sum_j \mathrm{Tr}[\Delta\hat{\Pi}^2_{j,\rm lin}] \right\rangle_{\rm ens}, \tag{23}$$

where $\Delta\hat{\Pi}_{j,\rm ML} = \hat{\Pi}_{j,\rm ML} - \hat{\Pi}_j$, $\Delta\hat{\Pi}_{j,\rm lin} = \hat{\Pi}_{j,\rm lin} - \hat{\Pi}_j$, and $\langle\rangle_{\rm ens}$ denotes averaging over the ensemble of all possible experimental outcomes.

We have repeated the reconstruction of the POVM for 100 different simulated experimental data sets and the ensemble averages yielded $\sigma^2_{\rm ML}$ and $\sigma^2_{\rm lin}$. The variances were determined for ten different $\mathcal{N}$ and the results are shown in Fig. 1. We can see that the ML estimates exhibit lower fluctuations than the linear ones. Upon comparing Figs. 1(a) and 1(b) we find that the difference between the two methods is more pronounced when the "true" $\hat{\Pi}_j$ are projectors but the ML estimation is significantly better than linear inversion in both cases.

In summary, we have shown how to reconstruct a generic quantum measurement with the use of the maximum-likelihood principle. Our method guarantees that the estimated POVM, which fully describes the measuring apparatus, meets the required positivity constraints. This restriction to physically allowed POVMs significantly improves reconstruction accuracy, which is a considerable practical advantage of the ML estimation over linear inversions.

ACKNOWLEDGMENTS

I would like to thank Z. Hradil, M. Ježek, and J. Řeháček for stimulating discussions. This work was supported by Research Project CEZ: J14/98: 153100009 "Wave and Particle Optics" of the Czech Ministry of Education.

[1] C.W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976); A.S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).

[2] U. Leonhardt, *Measuring the Quantum State of Light* (Cambridge University Press, Cambridge, England, 1997).

[3] D.-G. Welsch, W. Vogel, and T. Opatrný, *Homodyne Detection and Quantum State Reconstruction*, Vol. 39 of *Progress in Optics*, edited by E. Wolf (North-Holland, Amsterdam, 1999).

[4] J.F. Poyatos, J.I. Cirac, and P. Zoller, Phys. Rev. Lett. **78**, 390 (1997); I.L. Chuang and M.A. Nielsen, J. Mod. Opt. **44**, 2455 (1997); G.M. D'Ariano and L. Maccone, Phys. Rev. Lett. **80**, 5465 (1998); A. Luis and L.L. Sánchez-Soto, Phys. Lett. A **261**, 12 (1999).

[5] A. Luis and L.L. Sanchez-Soto, Phys. Rev. Lett. **83**, 3573 (1999).

[6] Z. Hradil, Phys. Rev. A **55**, R1561 (1997); Z. Hradil, J. Summhammer, and H. Rauch, Phys. Lett. A **261**, 20 (1999); Z. Hradil, J. Summhammer, G. Badurek, and H. Rauch, Phys. Rev. A **62**, 014101 (2000); Z. Hradil and J. Summhammer, J. Phys. A **33**, 7607 (2000).

[7] K. Banaszek, G.M. D'Ariano, M.G.A. Paris, and M.F. Sacchi, Phys. Rev. A **61**, 010304(R) (2000).

[8] J. Fiurášek and Z. Hradil, Phys. Rev. A **63**, 020101(R) (2001); M.F. Sacchi, *ibid.* **63**, 054104 (2001).

[9] G.M. D'Ariano, M.G.A. Paris, and M.F. Sacchi, Phys. Rev. A **62**, 023815 (2000).